

REPUBLIC OF TÜRKİYE
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

**IDENTIFICATION OF POTENTIAL TARGET GENES IN
HUMAN BY MIRNAS OF CACAO: A CROSS-KINGDOM
IN SILICO APPROACH**

Rümeysa KARAİSMAİLOĞLU

MASTER OF SCIENCE THESIS

Department of Molecular Biology and Genetics

Molecular Biology and Genetics Program

Supervisor

Assoc. Prof. Dr. Sevgi MARAKLI

December, 2024

REPUBLIC OF TÜRKİYE
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

**IDENTIFICATION OF POTENTIAL TARGET GENES IN
HUMAN BY MIRNAS OF CACAO: A CROSS-KINGDOM *IN SILICO* APPROACH**

A thesis submitted by Rümeysa KARAİSMAİLOĞLU in partial fulfilment of the requirements for the degree of **MASTER OF SCIENCE** is approved by the committee on 13.12.2024 in the Department of Molecular Biology and Genetics, Molecular Biology and Genetics Program.

Assoc. Prof. Dr. Sevgi MARAKLI
Yıldız Technical University
Supervisor

Approved By the Examining Committee

Assoc. Prof. Dr. Sevgi MARAKLI, “Supervisor”
Yıldız Technical University

Prof. Dr. Nehir ÖZDEMİR ÖZGENTÜRK, “Member”
Yıldız Technical University

Prof. Dr. Emre YÖRÜK, “Member”
Yeni Yüzyıl University

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I haven't falsified and/or fabricated research data and results of the study, and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of "Identification of potential target genes in human by miRNAs of cacao: A cross-kingdom *in silico* approach" supervised by my supervisor, Assoc. Prof. Dr. Sevgi MARAKLI. In the case of a discovery of a false statement, I am to acknowledge any legal consequence.

Rümeysa KARAİSMAİLOĞLU

Signature



This work was supported by the “TÜBİTAK 2210/A-Domestic Graduate Scholarship Program”.



*Dedicated to the freedom that
leads to change*

ACKNOWLEDGEMENTS

Firstly, I would like to thank my dear advisor Assoc. Prof. Dr. Sevgi MARAKLI has shared her knowledge with me in academic processes since my last year of undergraduate education and helped me start many processes. I would like to thank all MARAKLI Laboratory members and our dear instructor Prof. Dr. Emre YÖRÜK who took the time to train us.

I would like to thank my internship friends at AMGEN for making the internship process easy.

I would like to thank all my dear research assistant instructors for supporting me in my experience of observing academic processes, even if only for a short time. He guided and supported students like me; I remember my dear, late instructor Res. Asst. Ahmet YAĞBASAN with respect and thank him for all his efforts.

I would like to thank all my dear teachers who contributed to my education throughout my life. I would like to thank all the scientists who contribute to the literature and everyone who works to make the world better, more equal, and more livable. I would like to thank all my friends, elders, children, and their families who have supported me and I cannot count more.

Last, but most importantly, I would like to thank my entire family, especially my valuable mother and dear father, who believed in me with all their heart, never stopped supporting me, and stood by me and behind me without questioning every decision I made.

My greatest luck, my star, my valuable mother, who has made the miracle in the skies shine in my heart since the day I was born, I also thank you for teaching me how to turn to ashes and burn again. May this fire always lead to innovation, goodness, and equality in the entire cosmos.

Rümeysa KARAİSMAİLOĞLU

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
LIST OF ABBREVIATIONS	viii
LIST OF FIGURES	x
LIST OF TABLES	xii
ABSTRACT	xiii
ÖZET	xv
1 INTRODUCTION	1
1.1 Literature Review.....	1
1.2 Purpose of Thesis	2
1.3 Hypothesis.....	2
2 THEORETICAL BACKGROUND	3
2.1 RNAi.....	3
2.2 Small non-coding RNAs (sncRNAs).....	7
2.3 miRNA.....	8
2.3.1 miRNA Detection Strategies.....	9
2.3.2 Using miRNAs as a Biomarker for Human Disorders	10
2.3.3 miRNA-based Therapies.....	12
2.3.4 Biogenesis of Plant and Animal miRNAs	14
2.3.5 Cross-kingdom Interactions Between Plant and Human miRNAs...	18
2.4 <i>Theobroma cacao</i> L.	20
3 METHODS	23
3.1 Obtaining Reference miRNA Sequences.....	24
3.2 BLAST Analyses.....	24
3.3 Elimination Process.....	25
3.4 RNAFold Analyses for Prediction of Secondary Structures.....	25
3.5 Determination of miRNA Target Genes.....	26
3.6 Go and KEGG Analyses of Target Cacao Genes.....	27
3.7 Network Interactions.....	27

4 CONCLUSION	28
4.1 Identification of miRNAs.....	29
4.2 Prediction of Secondary Structures.....	32
4.3 Targets and Functions of Potential miRNAs.....	46
4.3.1 RNA Hybrid, GO, and KEGG Analysis of miR166 Target Genes..	46
4.3.2 RNA Hybrid, GO, and KEGG Analysis of miR168 Target Genes..	52
4.3.3 RNA Hybrid, GO, and KEGG Analysis of miR169 Target Genes..	56
4.3.4 RNA Hybrid, GO, and KEGG Analysis of miR1051Target Genes.	60
4.3.5 RNA Hybrid, GO, and KEGG Analysis of miR3476Target Genes.	63
4.3.6 RNA Hybrid, GO, and KEGG Analysis of miR4249Target Genes.	68
4.3.7 RNA Hybrid, GO, and KEGG Analysis of miR6161Target Genes.	70
4.3.8 RNA Hybrid, GO, and KEGG Analysis of miR6279Target Genes.	73
4.3.9 RNA Hybrid, GO, and KEGG Analysis of miR8150Target Genes.	76
4.3.10 RNA Hybrid,GO,and KEGG Analysis of miR9471Target Genes.	82
4.4 Related Pathways, Functions, and Network Interactions of Novel Cacao miRNAs.....	88
REFERENCES	94
PUBLICATIONS FROM THE THESIS	114

LIST OF ABBREVIATIONS

AGO	Argonaute protein
AMFE	Adjusted Minimal Folding Free Energy
APEX1	Apurinic/Apyrimidinic Endodeoxyribonuclease 1
BLAST	Basic Local Alignment Search Tool
BP	Biological Process
CC	Cellular Component
CPH	Cacao Protein Hydrolysates
DAVID	Database for Annotation, Visualization, and Integrated Discovery
DCL 1	Dicer-like 1
DGCR8	DiGeorge Syndrome Critical Region 8
dsRNA	Double-stranded RNA
EST	Expressed Sequence Tags
GeneMANIA	Gene Multiple Relationship Network Integration Algorithm
GI	Gastrointestinal
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
m7G	7-methylguanine Capped
MF	Molecular Function
MFE	Minimum Free Energy
MFEI	Minimum Folding Energy Index
miRISC	miRNA Mediated Silencing Complex
miRNA	microRNA
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
nt	Nucleotide
Pol II	RNA Polymerase II
Pre-miRNA	Precursor-miRNA

Pri-miRNA	Primary miRNA
PTGS	Post-transcriptional Gene Silencing
RISC	RNA-induced silencing complex
RNAi	RNA Interference
SARS-CoV-2	Type 2 Coronavirus
shRNA	Short Hairpin RNA
siRNA	Small Interfering RNA
SNP	Single Nucleotide Polymorphism
TGS	Transcriptional Gene Silencing
WC	Watson-Crick
3'-UTR	3'-Untranslated Regions
5'-UTR	5'-Untranslated Regions

LIST OF FIGURES

Figure 2.1 RNAi mechanism	4
Figure 2.2 Canonical miRNA pathway.....	6
Figure 2.3 Small non-coding RNAs.....	8
Figure 2.4 miRNA: target mRNA relations	9
Figure 2.5 miRNA biogenesis.....	15
Figure 2.6 Comparison of miRNA biogenesis and function in plants and animals	17
Figure 2.7 <i>Theobroma cacao</i> L.....	20
Figure 3.1 Flow chart showing the steps to be followed in <i>in silico</i> miRNA identification in cacao	23
Figure 4.1 Secondary structure of the tcc-miR166b sequence identified in <i>Theobroma cacao</i>	33
Figure 4.2 Secondary structure of the tcc-miR166i sequence identified in <i>Theobroma cacao</i>	34
Figure 4.3 Secondary structure of the tcc-miR168 sequence identified in <i>Theobroma cacao</i>	35
Figure 4.4 Secondary structure of the tcc-miR169 sequence identified in <i>Theobroma cacao</i>	36
Figure 4.5 Secondary structure of the tcc-miR1051 sequence identified in <i>Theobroma cacao</i>	37
Figure 4.6 Secondary structure of the tcc-miR3476 sequence identified in <i>Theobroma cacao</i>	38
Figure 4.7 Secondary structure of the tcc-miR4249 sequence identified in <i>Theobroma cacao</i>	39
Figure 4.8 Secondary structure of the tcc-miR6161 sequence identified in <i>Theobroma cacao</i>	40
Figure 4.9 Secondary structure of the tcc-miR6279 sequence identified in <i>Theobroma cacao</i>	41
Figure 4.10 Secondary structure of the tcc-miR8150 sequence identified in <i>Theobroma cacao</i>	42
Figure 4.11 Secondary structure of the tcc-miR9471 sequence identified in <i>Theobroma cacao</i>	43
Figure 4.12 GO analyses results of targets of newly identified tcc-miR166b and tcc-miR166i.....	50
Figure 4.13 Glycosaminoglycan degradation	51
Figure 4.14 GO analyses results of targets of newly identified tcc-miR1668	54

Figure 4.15 Spliceosome mechanism	55
Figure 4.16 GO analyses results of targets of newly identified tcc-miR169	58
Figure 4.17 Ribosome biogenesis in eukaryotes.....	59
Figure 4.18 GO analyses results of targets of newly identified tcc-miR1051	62
Figure 4.19 GO analyses results of targets of newly identified tcc-miR3476	66
Figure 4.20 Photosynthesis	67
Figure 4.21 GO analyses results of targets of newly identified tcc-miR4249	69
Figure 4.22 GO analyses results of targets of newly identified tcc-miR6161	72
Figure 4.23 GO analyses results of targets of newly identified tcc-miR6279	75
Figure 4.24 GO analyses results of targets of newly identified tcc-miR8150	78
Figure 4.25 mRNA surveillance pathway	79
Figure 4.26 Nucleocytoplasmic transport	80
Figure 4.27 Spliceosome mechanism	81
Figure 4.28 GO analyses results of targets of newly identified tcc-miR9471	85
Figure 4.29 Purine metabolism	86
Figure 4.30 Ubiquitin-mediated proteolysis	87
Figure 4.31 Interaction network of essential target genes (<i>RCHY1</i> , <i>NOL6</i> , <i>DDX5</i> , <i>UPF1</i> , <i>SNCG</i>) in the mechanisms associated with cacao miRNAs with human genes	91

LIST OF TABLES

Table 4.1	Sequences of novel miRNAs in the cacao genome.....	30
Table 4.2	Characteristics of newly identified miRNAs in cacao genome	45
Table 4.3	The targets tcc-miR166b and tcc-miR166i	47
Table 4.4	The targets tcc-miR168	53
Table 4.5	The targets tcc-miR169	57
Table 4.6	The targets tcc-miR1051	61
Table 4.7	The targets tcc-miR3476	64
Table 4.8	The targets tcc-miR4249	69
Table 4.9	The targets tcc-miR6161	71
Table 4.10	The targets tcc-miR6279	74
Table 4.11	The targets tcc-miR8150	77
Table 4.12	The targets tcc-miR9471	83

Identification of potential target genes in human (*Homo sapiens*) by miRNAs of cacao (*Theobroma cacao* L.): A cross-kingdom *in silico* approach

Rümeysa KARAİSMAİLOĞLU

Department of Molecular Biology and Genetics

Molecular Biology and Genetics Program

Master of Science Thesis

Supervisor: Assoc. Prof. Dr. Sevgi MARAKLI

Among agricultural products, cacao (*Theobroma cacao* L.) is one of the world's leading natural resources. Anticarcinogenic and cardioprotective effects of phytochemicals obtained from cacao, neuroprotective effects of flavonoids in cacao, and anti-stress biotechnological properties have been investigated for therapeutic purposes. MicroRNAs have important roles in gene expression regulation with conserved characteristics in plants and mammals. Due to their remarkable importance, many miRNAs have been identified in recent years with increasing *in silico* studies on plants. The impacts of plant miRNAs on human gene expression and their molecular processes have been examined together with the cross-kingdom approach. In this study, we aimed to conduct *in silico* analyses to define miRNA functions in cacao on the mechanism of human diseases, especially cancer. For this purpose, reference miRNA sequences of *Viridiplantae* in the mirBase database were obtained. The obtained reference miRNA sequences

were compared with the EST database of cacao via the BLASTN program in the NCBI database. Secondary structures of suitable pre-miRNA sequences were determined using the "RNAfold " program. The target genes of these miRNA sequences were identified with the “RNA Hybrid” program. The processes and metabolic pathways in which target genes play a role were analyzed with BLAST2GO and KEGG programs. Network and interactions between cacao miRNA targets to human genes were evaluated through the GeneMANIA. 11 new miRNAs from 10 miRNA families were identified and their 47 target genes were determined in this study. We hope that this study, which we believe will bring a new perspective to the cross-kingdom interactions perspective, will shed light on using plants more actively in biomarker studies in cancer shortly by revealing the similarity between humans and cocoa plants and their relationship with proteins that are important in an important disease mechanism such as cancer.

Keywords: Plant-miRNAs, *in silico*, cross-kingdom interactions, gene regulation, human diseases.

Kakao (*Theobroma cacao* L.) miRNA'ları ile insandaki (*Homo sapiens*) potansiyel hedef genlerin tanımlanması: Alemler arası bir *in silico* yaklaşım

Rümeysa KARAİSMAİLOĞLU

Moleküler Biyoloji ve Genetik Departmanı

Moleküler Biyoloji ve Genetik Program

Yüksek Lisans Tezi

Danışman: Doç. Dr. Sevgi MARAKLI

Tarım ürünleri arasında kakao (*Theobroma cacao* L.) dünyanın önde gelen doğal kaynaklarından biri olup, aynı zamanda birçok ürünün üretiminde organoleptik nitelikleri, yüksek kalorisi ve besin değerleri nedeniyle hammadde olarak önem taşımaktadır. Kakaodan elde edilen fitokimyasalların antikarsinojenik ve kardiyoprotektif etkileri, kakaoda bulunan flavonoidlerin nöroprotektif etkileri ve anti-stres biyoteknolojik özellikleri tedavi amaçlı olarak araştırılmıştır. MikroRNA'lar, bitkilerde ve memelilerde korunmuş özellikleriyle gen ekspresyonunun düzenlenmesinde önemli rollere sahiptir. Dikkate değer önemlerinden dolayı, son yıllarda bitkiler üzerinde artan *in silico* çalışmaları ile birçok miRNA tanımlanmıştır. Bitki miRNA'larının insan gen ekspresyonu ve moleküler süreçleri üzerindeki etkileri türler arası yaklaşımla birlikte incelenmiştir. Bu çalışmada kakao miRNA fonksiyonlarını ve başta kanser olmak üzere insan hastalıklarının mekanizması üzerindeki etkisini tanımlamak için *in*

silico analizler yapmayı amaçladık. Bu amaçla mirBase veri tabanındaki *Viridiplantae*'nin referans miRNA dizileri elde edildi. Elde edilen referans miRNA dizileri, NCBI veri tabanındaki BLASTN programı aracılığıyla kakaonun EST veri tabanı ile karşılaştırıldı. Uygun pre-miRNA dizilerinin sekonder yapıları "RNAfold" programı kullanılarak belirlendi. Bu miRNA dizilerinin hedef genleri "RNAHybrid" programı ile belirlendi. Hedef genlerin rol oynadığı süreçler ve metabolik yollar BLAST2GO ve KEGG programları ile analiz edildi. Kakao miRNA hedefleri ile insan genleri arasındaki ağ ve etkileşimler, GeneMANIA aracılığıyla değerlendirildi. Bu çalışmada 10 miRNA ailesinden 11 yeni miRNA tanımlanmış ve bunların 47 hedef gen belirlenmiştir. Krallıklar arası etkileşimler perspektifine yeni bir bakış açısı getireceğine inandığımız bu çalışmanın, insanlarla kakao bitkileri arasındaki benzerlikleri ve kanser gibi önemli bir hastalık mekanizmasında önemli olan proteinlerle ilişkisi ortaya koyarak yakın gelecekte kanserde biyobelirteç çalışmalarında bitkilerin daha aktif kullanılmasına ışık tutacağını umuyoruz.

Anahtar Kelimeler: Bitki miRNA'ları, *in silico*, türler arası etkileşimler, gen düzenlemesi, insan hastalıkları.

1.1 Literature Review

Since ancient times plants as natural sources have been utilized to cure human illnesses, especially cancer [1], [2]. Evolutionarily conserved micro RNAs (miRNAs) in plants, animals, and mammals are small non-coding RNAs (sncRNAs) with, 20-25 nucleotides (nt) in length, playing important roles and have essential roles in gene expression regulation [3]. During the last few decades, studies have shown that some plant/food-derived miRNAs (miRNAs) accumulate in serum, tissues, and feces depending on [4], [5] and [6]. Therefore, the effects of plant miRNAs on human gene expression and their molecular mechanisms have been tried to be clarified in recent years [7].

Cacao (*Theobroma cacao* L.) belonging to the Malvaceae family is a neotropical plant native to the Amazon lowland rainforest. Among agricultural products, cacao is one of the world's leading natural resources due to raw materials in the food industry [8]. Numerous beneficial and variable properties of cacao including antioxidant, antidiabetic, and antibacterial activities, cardioprotective effects, and anticarcinogenic and neuroprotective effects, etc. have been demonstrated for usage in biotechnological applications [9], [10].

Patil et al. identified specific bioactive compounds in cacao and determined their interactions with proteins associated with cancer-related pathways through their computational analyses [11]. These compounds were hypothesized to potentially exhibit anticancer effects by modulating the activity of these proteins. Additionally, the identification of bioactive compounds in cacao and their specific mechanisms of action would require further investigation. While these findings provide preliminary evidence for the potential role of cacao in cancer therapy, it is crucial to interpret them in the context of a comprehensive understanding of cancer treatment and the need for further scientific validation.

Starting from this point, we aimed to define a detailed miRNA in the cacao plant, which has important content, by considering cross-kingdom interactions.

1.2 Purpose of Thesis

The purpose of this study is to identify new miRNA sequences in the cacao plant (*Theobroma cacao* L.) using *in silico* methods and determine their relationship with human genes and mechanisms with a cross-kingdom approach.

1.3 Hypothesis

Based on this point, it is expected to identify new miRNAs and their targets in the cacao plant, and human targets, which have a high phytochemical content, are used in many biotechnological applications and have previously been determined to have an effect on various human diseases, and to determine their effects on human mechanisms with a cross-kingdom approach and to draw attention to the therapeutic potential of the cacaos.

THEORETICAL BACKGROUND

2.1 RNAi

A conserved mechanism in eukaryotic cells called RNA interference (RNAi) controls gene expression by destroying messenger RNA (mRNA) molecules (Figure 2.1). RNA silencing is the repression guided by small RNA molecules, typically 20–30 nucleotides long. This process encompasses various silencing mechanisms, including RNA degradation, translational repression, induction of repressive chromatin, and even DNA deletions [12]. Research on RNA silencing has advanced through studies in diverse model systems, notably flowering plant models, and animal models such as *Caenorhabditis elegans*, *Drosophila melanogaster*, zebrafish, mice, and humans. Almost every eukaryote exhibits some form of RNA silencing [13].

RNAi was initially coined for sequence-specific mRNA degradation induced by long double-stranded RNA (dsRNA) [14]. This canonical RNAi mechanism employs small RNAs produced from long dsRNA. However, the term RNAi is also used more broadly to refer to a range of RNA silencing pathways [12] miRNAs are short RNAs encoded in the genome that regulate gene expression by translational repression and/or degradation of cognate mRNAs. The plant co-suppression was the first discovered RNA silencing pathway, observed as sequence-specific silencing of endogenous genes induced by transgene expression [15].

The miRNA pathway (Figure 2.2) was first identified in 1993 in the nematode *Caenorhabditis elegans* [16]. The conservation of the miRNA pathway from *Caenorhabditis elegans* to mammals became apparent with the discovery of the Let-7 miRNA in 2000 [17]. Around 2000, it was realized that earlier observations, including plant co-suppression, quelling in fungi, and animal RNAi and miRNA pathways, [15], [16], [17], [18] and [19], belonged to one group of related

molecular mechanisms termed RNA silencing. The core principle of RNA silencing (repression mediated by a ribonucleoprotein complex guided by a small RNA), was elucidated between 1998 and 2004 through a combination of genetic and biochemical approaches. Key insights were derived from biochemical studies in *Drosophila* embryo lysates [20], [21] and genetic studies in *Caenorhabditis elegans* and plants [22], [23], [24], [25], [26], [27] and [28].

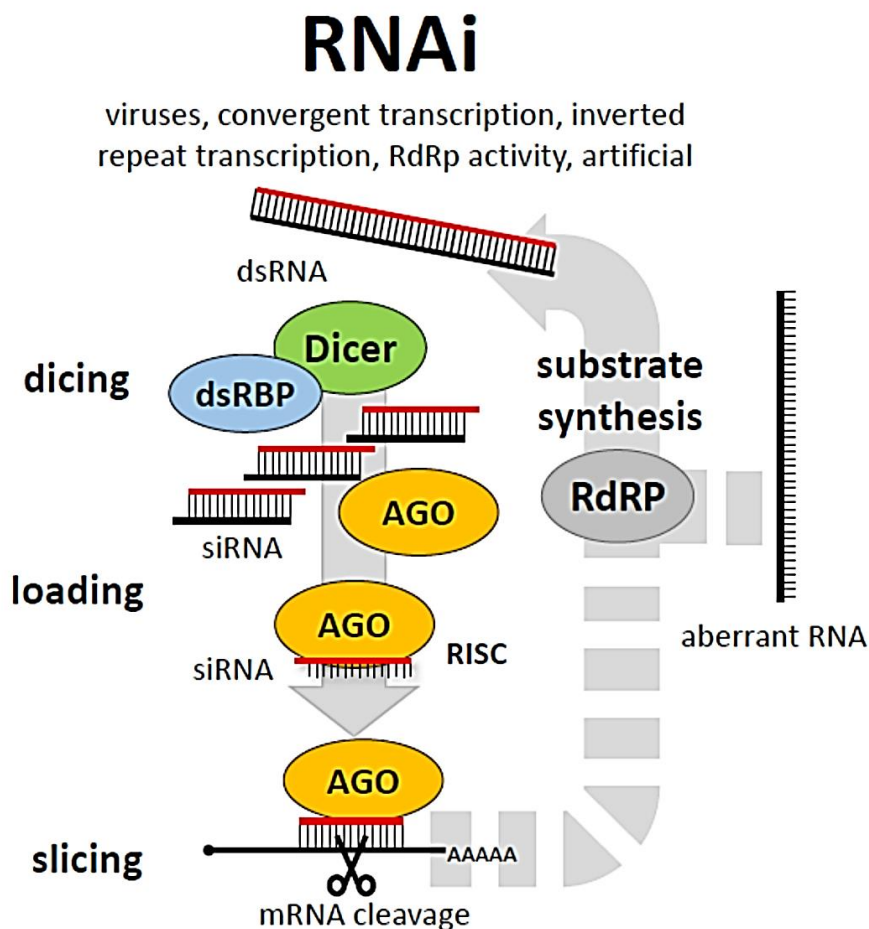


Figure 2.1 RNAi mechanism [13]

Canonical RNA interference (RNAi) is initiated by the presence of long double-stranded RNA (dsRNA). This dsRNA can have diverse origins, including viruses and their replication intermediates, base pairing of RNAs transcribed in the genome (in the form of intramolecular duplex or hairpin dsRNA), or base pairing of RNAs transcribed either in cis (convergent transcription) or in trans (from

interspersed elements, pseudogenes, etc.). The fundamental RNAi mechanism involves three main steps:

1. Dicing: This step involves the cleavage of long dsRNA into small interfering RNA (siRNA) duplexes. This process is facilitated by an enzyme called RNase III Dicer.
2. Loading: In this step, one strand of the siRNA duplex is selected and loaded onto an Argonaut protein from the AGO subfamily. This complex forms the RNA-induced silencing complex (RISC).
3. Slicing: The siRNA guides the RISC to its cognate RNAs. When a perfect duplex is formed with a cognate RNA, Argonaut proteins, specifically AGO proteins, perform endonucleolytic cleavage of the cognate RNA at the middle of the base-paired sequence.

In certain species, RNAi also involves an RNA-dependent RNA polymerase (RdRp). This enzyme may generate initial substrates or participate in amplifying the response by converting cognate RNAs into dsRNA. This additional step contributes to the robustness and efficiency of the RNAi process [13].

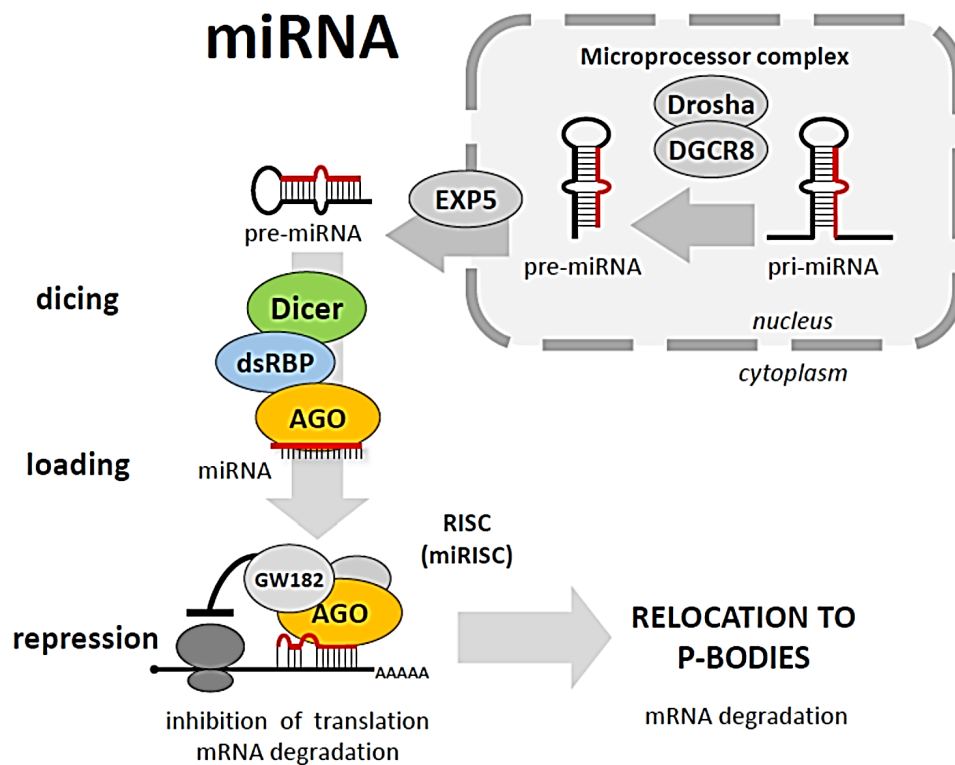


Figure 2.2 Canonical miRNA pathway [13]

MicroRNAs (miRNAs) are encoded in the genome, and their synthesis begins with transcription by RNA polymerase II (Pol II), generating long primary miRNA transcripts (pri-miRNAs). These pri-miRNAs harbor one or more local short hairpins, and a nuclear complex known as the "Microprocessor complex" releases these as precursor miRNAs (pre-miRNAs). Exportin 5 facilitates the transport of pre-miRNAs from the nucleus to the cytoplasm. In the cytoplasm, Dicer plays a crucial role by cleaving pre-miRNAs. Subsequently, one strand of the resulting duplex is loaded onto an Argonaute (AGO) protein, forming the core of the effector complex, also known as the RNA-induced silencing complex (RISC) or miRISC. The effector complex comprises additional proteins that mediate translational repression and RNA degradation. A key link between AGO and the proteins involved in deadenylation and decapping is the GW182 protein. The mRNAs targeted by miRNAs typically localize to P-bodies, which are cytoplasmic foci associated with RNA metabolism. These structures are involved in the regulation of mRNA stability and translation. The entire process, from

transcription to mRNA degradation, constitutes the intricate mechanism by which miRNAs regulate gene expression [13].

Researchers can manipulate the expression of specific genes by using RNAi pathways in plants [29], animals, and [30] even humans [31]. Selectively regulating gene expression has created new opportunities for researching gene function and forming applicable treatment approaches for a range of illnesses [32].

2.2 Small Non-coding RNAs (sncRNAs)

MicroRNAs (miRNAs), small interfering RNAs (siRNAs), short hairpin RNAs (shRNAs), and other small non-coding RNAs (sncRNAs) (Figure 2.3) with lengths of around 20-25 nt have gained significant attention in various applications in plants, animals, and humans [33], [34] and [35]. These sncRNAs, play important roles in gene regulation through both post-transcriptional gene silencing (PTGS) and transcriptional gene silencing (TGS) mechanisms. They can target mRNAs to inhibit their translation or induce their degradation, leading to the suppression of specific genes [36], [37].

To discover novel sncRNAs and their targets involved in various metabolic pathways, researchers have conducted both computational (*in silico*) and experimental investigations. By analyzing sncRNA expression patterns and their interactions with target genes, researchers have gained new insights into the complexity and functions of these small regulatory RNAs including disease diagnosis, therapy, and crop improvement [38], [39], [40], [41] and [42].

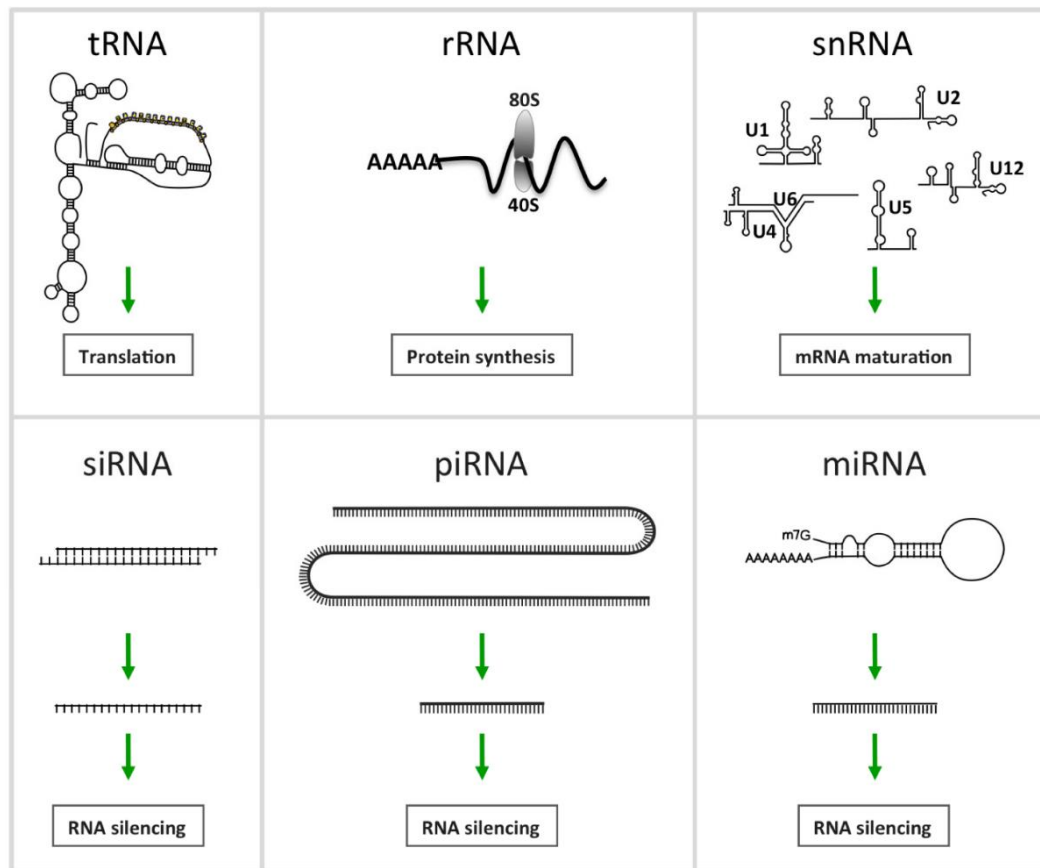


Figure 2.3 Small non-coding RNAs [43]

2.3 miRNA

The importance of miRNAs was understood after the discovery of lin-4 and let-7 miRNAs, which regulate nematode (*Caenorhabditis elegans*) development by imperfect base pairing to the 3' UTRs of target mRNAs to block translation. Reinhart et al. [44] also examined *Arabidopsis* to be a model organism for the detection of miRNAs in plants. They demonstrated that ncRNAs may have evolved early in the history of eukaryotic cells. While one mRNA may contain several miRNA binding sites, numerous mRNAs may be targeted by a single miRNA (Figure 2.4) [45].

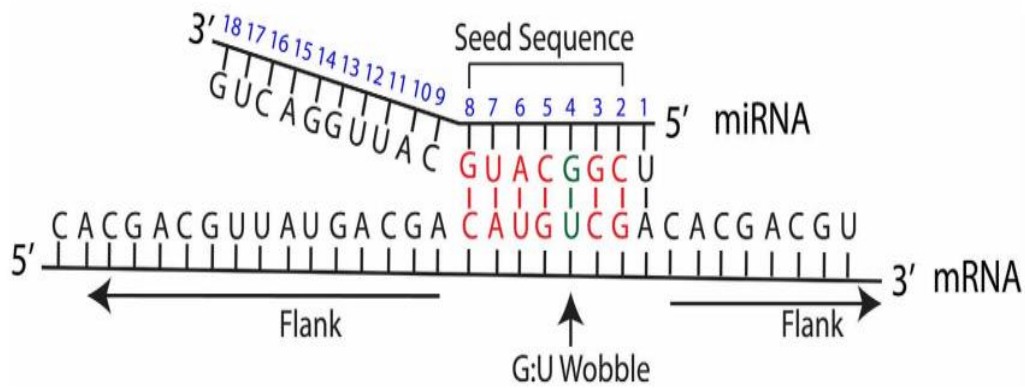


Figure 2.4 miRNA: target mRNA relations. miRNA location is indicated with blue digits. The seed sequence in miRNA is 2- 8 nt in length. The mRNA regions present on either side of the seed region are known as the flank regions. Watson-Crick (WC) matches in the seed region are displayed with a red background and a green background, respectively [46]

2.3.1 miRNA Detection Strategies

Various methods are available for the identification and analysis of novel miRNAs. These methods can generally be categorized into two groups: computational analyses and experimental analyses. Different databases such as miRbase [47], MiRandola (<http://mirandola.iit.cnr.it/>), and human infections (<http://mir2disease.org/>) are commonly utilized to detect novel miRNAs associated with diseases and various metabolic pathways by using *in silico* analyses. Reference sequences obtained from these databases are often employed in conjunction with tools like NCBI-BLAST, RNAfold, RNAHybrid, BLAST2GO, and other related programs to identify miRNAs and their corresponding targets through computational methods [48], [49], [50] and [51]. Experimental analyses, on the other hand, involve laboratory techniques mentioned below. These methods contribute to the discovery and characterization of novel miRNAs, providing insights into their functions and potential implications in biological processes.

- Northern blot analysis: This method involves the separation of miRNAs by gel electrophoresis, followed by their transfer to a membrane and hybridization with complementary probes to detect specific miRNAs [52], [53].
- *In situ* hybridization: This technique allows the visualization of miRNA expression patterns within cells or tissues by utilizing fluorescently labeled probes that specifically bind to target miRNAs [54].
- Real-time PCR: This quantitative method measures the abundance of specific miRNAs by using target-specific primers and fluorescent probes, enabling precise quantification of miRNA expression levels [55], [56]
- miRNA microarray: Microarray technology allows the simultaneous detection and quantification of multiple miRNAs within a sample, providing a comprehensive miRNA expression profile [57], [58].
- Next-generation sequencing (NGS): NGS techniques, such as Illumina Genome Analyzer, ABI SOLiD, or Roche/454 Genome Sequencer FLX, enable high-throughput sequencing of miRNAs. This approach provides deep sequencing coverage and facilitates the discovery of novel miRNAs and the analysis of their expression profiles [59].

Hong et al. [60] also suggested miRNA-Seq as a powerful approach for miRNA biomarker discovery and the identification of novel miRNAs. It should be noted that miRNA expression profiles in plasma and serum are particularly valuable for their potential use as biomarkers in early disease diagnosis and treatment [61].

2.3.2 Using miRNAs as a Biomarker for Human Disorders

Circulating cell-free miRNAs are stable molecules found in the extracellular compartment and hold promise as non-invasive biomarkers in humans [60]. These miRNAs are enclosed within vesicles with membranes, like exosomes. Extensive research has demonstrated that the expression of circulating miRNAs is associated with various pathophysiological conditions, including cancer, infectious diseases,

cardiovascular diseases, and neurodegenerative diseases [62], [63], [64], [65] and [66].

The stability of circulating miRNAs allows for their detection in bodily fluids such as blood, plasma, serum, urine, saliva, and cerebrospinal fluid. The tissue specificity of miRNAs enables their potential use as specific biomarkers for different diseases or conditions. By analyzing the expression levels and profiles of circulating miRNAs, researchers can gain insights into the underlying molecular mechanisms of these conditions and potentially develop diagnostic, prognostic, and therapeutic applications [58].

Duy et al. [67] focused on Dengue virus infection and explored the utility of circulating miRNAs as diagnostic markers. They analyzed plasma samples from patients with different stages of Dengue infection and identified specific miRNA signatures associated with the disease. These miRNAs showed differential expression patterns in patients compared to healthy controls, suggesting their potential as diagnostic markers for Dengue virus infection.

In another study, Ouyang et al. [68] conducted a study of Ebola virus infection and investigated the role of circulating miRNAs as diagnostic markers. They analyzed plasma samples from Ebola-infected patients and identified differentially expressed miRNAs that could potentially serve as biomarkers for early detection or monitoring of Ebola virus infection. The study provided insights into the potential of miRNAs for diagnosing Ebola and understanding the pathogenesis of the disease.

Biswas et al. [65] also aimed to identify specific miRNAs as biomarkers for the early detection or diagnosis of HIV-1 infection. However, further research and validation are necessary to confirm the diagnostic and prognostic value of these miRNAs and their clinical application in HIV-1 management.

miRNAs possess remarkable stability, as they are resistant to RNase enzymes, temperature fluctuations, long-term storage, extreme pH conditions, and freeze-thaw cycles. Moreover, miRNAs exhibit greater stability in blood compared to other RNA classes [69]. They have been extensively studied in the context of cancer to regulate various oncogenic processes, playing significant roles in tumor progression and prognosis. miRNAs can serve as tumor suppressors or oncogenes, depending on their target genes and the cellular context. Moreover, aberrant expression of miRNAs has been observed in various types of cancer, and their dysregulation can contribute to tumor development, metastasis, angiogenesis, and resistance to therapy [70]. Therefore, miRNAs also hold promise as potential diagnostic and prognostic markers in cancer. The expression levels of certain miRNAs in tumor tissues or body fluids, such as blood or urine, can serve as indicators of disease presence, progression, or response to therapy.

Modulating miRNA expression or activity through synthetic miRNA mimics (miRNA replacement therapy) or anti-miRNA oligonucleotides (miRNA inhibition therapy) can potentially restore the balance of miRNA-mediated gene regulation and inhibit tumor growth [71], [72], [73] and [74].

2.3.3 miRNA-based Therapies

Gastric cancer is a complex disease with various molecular alterations, and the dysregulation of certain genes, such as *APEX1*. In the study, He et al. [75] focused on the interaction between *APEX1* and miR-27a-5p. They found that miR-27a-5p can directly bind to *APEX1* and inhibit its expression. *APEX1* is known to be overexpressed in gastric cancer associated with tumor growth, invasion, and resistance to chemotherapy. By inhibiting *APEX1* through miR-27a-5p, the researchers observed a reduction in cell proliferation, migration, and invasion, suggesting a potential therapeutic strategy for gastric cancer.

The concept of using miRNAs as therapeutic agents to modulate gene expression has gained significant attention in recent years. The ability of miRNAs to regulate

multiple target genes and their involvement in various cellular processes make them attractive candidates for combination therapy with traditional chemotherapy. By targeting specific genes or pathways, miRNAs can potentially enhance the efficacy of chemotherapy and overcome treatment resistance [75].

Another study was carried out by Zhang et al. [76]. They explored the use of miRNA-5119 mimic-engineered dendritic cell (DC) vaccines in mouse breast cancer models. They found that the miRNA-5119 mimic-engineered DC vaccines were able to suppress tumor growth and promote an anti-tumor immune response. This suggests the potential of using miRNA-based therapies, such as miRNA-engineered DC vaccines, in cancer immunotherapy to enhance the immune response against tumors.

In the case of lung cancer, Fang et al. [77] examined the relationship between miRNA variants and the response to platinum-based chemotherapy. They identified several single nucleotide polymorphisms (SNPs) in miRNAs, including rs71428439 (miR-149), rs2910164 (miR-146a), rs928508 (miR-30c-1), rs629367 (let-7a-2), rs11614913 (miR-196a-2), and rs9280508 (miR-30c-1), which significantly influenced the response of patients to chemotherapy. These SNPs serve as potential clinical biomarkers for predicting lung cancer risk and the response to chemotherapy. By analyzing these miRNA variants, it may be possible to identify patients who are more likely to benefit from platinum-based chemotherapy and personalize their treatment accordingly.

These studies highlight the potential of miRNAs as therapeutic agents and clinical biomarkers in various diseases, including gastric cancer, breast cancer, and lung cancer. Combining miRNA-based approaches with chemotherapy and immunotherapy holds promise for improving treatment outcomes and personalized medicine strategies. Further research and clinical trials are needed to validate and optimize the therapeutic potential of miRNA-based interventions in these disease contexts.

2.3.4 Biogenesis of Plant and Animal miRNAs

At the beginning of miRNA biogenesis, RNA pol II and pol III processing affects primary miRNA (pri-miRNA) in the nucleus [79], [80] and [81]. Canonical and non-canonical routes are two distinct types of miRNA biogenesis (Figure 2.5). In the canonical pathway, the target mRNA is bound by the sense miRNA strand of the miRNA-mediated silencing complex (miRISC) complex via the 3'-UTR of the target mRNA [45].

However, it has been observed that approximately 60% of the interactions between the miRISC complex and mRNA in humans follow non-canonical pathways. In these cases, the miRNA-mRNA interactions may not involve perfect complementarity, and the resulting regulatory mechanisms may be different from canonical miRNA-mediated silencing. Non-canonical interactions can involve partial complementarity, binding of the miRNA to the coding sequence or 5'-UTR of the mRNA, or other complex mechanisms that are not fully understood [82]. These non-canonical interactions highlight the versatility and complexity of miRNA-mediated gene regulation and emphasize that miRNAs can have diverse regulatory roles beyond the traditional canonical pathway [83].

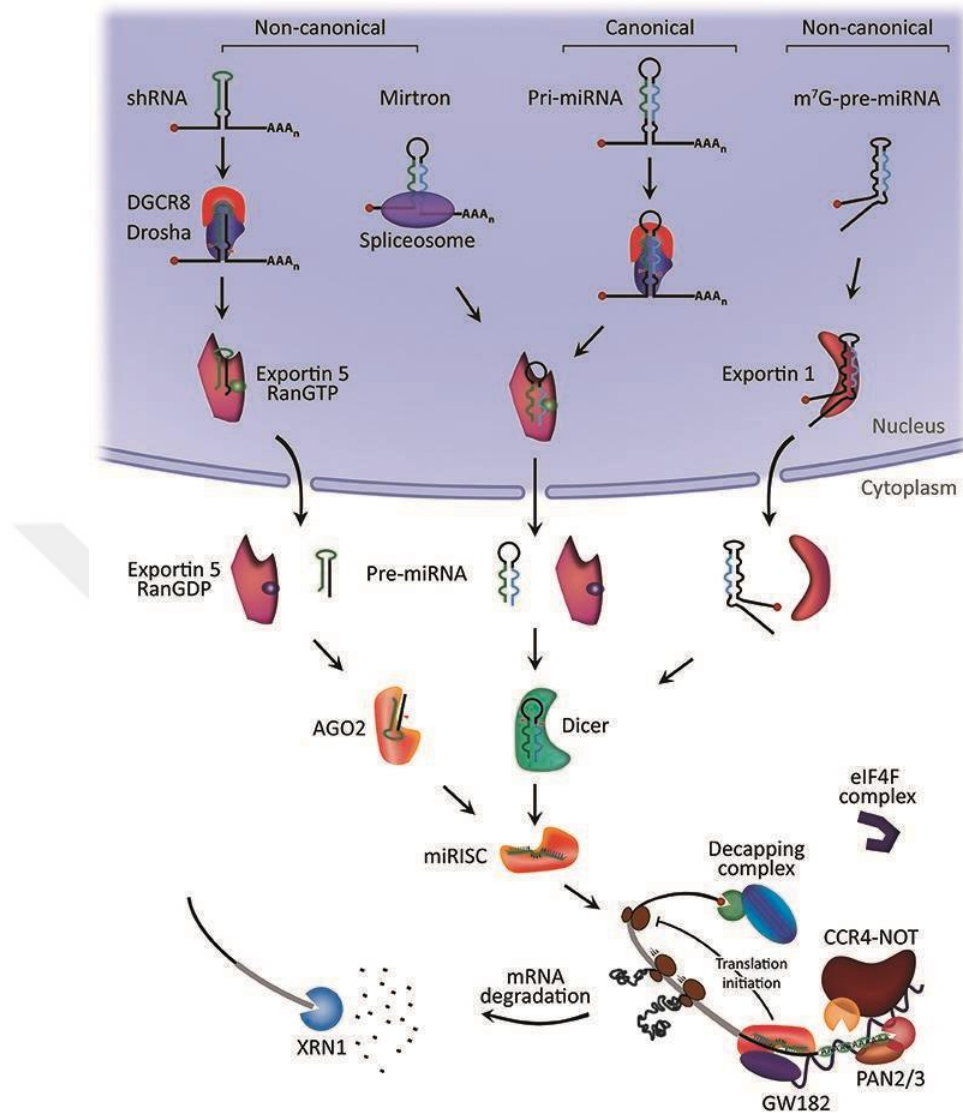


Figure 2.5 miRNA biogenesis [84]; shRNA, short hairpin RNA; pri-miRNA, primary miRNA; pre-miRNA, precursor-miRNA; pre-miRNA, precursor-miRNA; miRISC, miRNA mediated silencing complex; m⁷G, 7-methylguanine capped

The production of a pri-miRNA transcript occurs during the initial stages of standard RNA biogenesis. Subsequently, a complex known as the microprocessor, comprising of Drosha and DiGeorge Syndrome Critical Region 8 (DGCR8), cleaves the pri-miRNA. This cleavage generates a precursor microRNA (pre-miRNA) which is then transported to the cytoplasm through a mechanism dependent on Exportin5 and RanGTP. Within the cytoplasm, the pre-miRNA duplex associates with proteins from the Ago family to form a microRNA-

induced silencing complex (miRISC). Apart from the canonical pathway, there exists an alternative pathway referred to as the non-canonical pathway. This pathway is characterized as being independent of Drosha/DGCR8 and Dicer. In this pathway, small hairpin RNAs (shRNAs) undergo cleavage by the microprocessor complex and are exported to the cytoplasm in a manner like the canonical pathway involving Exportin5 and RanGTP. Furthermore, shRNAs undergo additional processing in an AGO2-dependent but Dicer-independent manner. Mirtrons and 7-methylguanine capped (m7G)-pre-miRNAs are processed by Dicer to generate mature miRNAs. The transportation of Mirtrons is facilitated by Exportin5 and RanGTP, while Exportin1 plays a crucial role in the transportation of m7G-pre-miRNAs [85].

There are notable differences in the biogenesis, function, and evolution of miRNAs between plants and animals [86]. In both plant and animal systems, miRNAs are initially transcribed as pri-miRNAs by RNA polymerase II and III. Pri-miRNAs undergo processing to form pre-miRNAs, which are stable hairpin structures of approximately 70 nt (Figure 2.6). In animals, the processing of pre-miRNAs from pri-miRNAs is carried out by Drosha. On the other hand, plants utilize a protein called Dicer-like 1 (DCL1), which has RNase-III-like activity, for pre-miRNA processing [87].

In plants, DCL1 catalyzes the conversion of pre-miRNAs to miRNA/miRNA* duplexes in the nucleus. In animals, the processing of pre-miRNAs by Dicer occurs in the cytoplasm. Once in the cytoplasm, the miRNA/miRNA* duplexes are separated, and the mature strand of the miRNA is associated with the Argonaute protein (AGO) to form the RNA-induced silencing complex (RISC). The RISC complex guides the mature miRNAs to their target mRNAs, leading to gene regulation and silencing.

The RISC guides the mature miRNA to recognize and bind to complementary sites on target genes [88]. These differences in the processing and localization of miRNA biogenesis highlight the distinct mechanisms and pathways involved in

miRNA regulation in plants and animals. Understanding these variations provides insights into the unique functions and evolutionary aspects of miRNAs in different organisms.

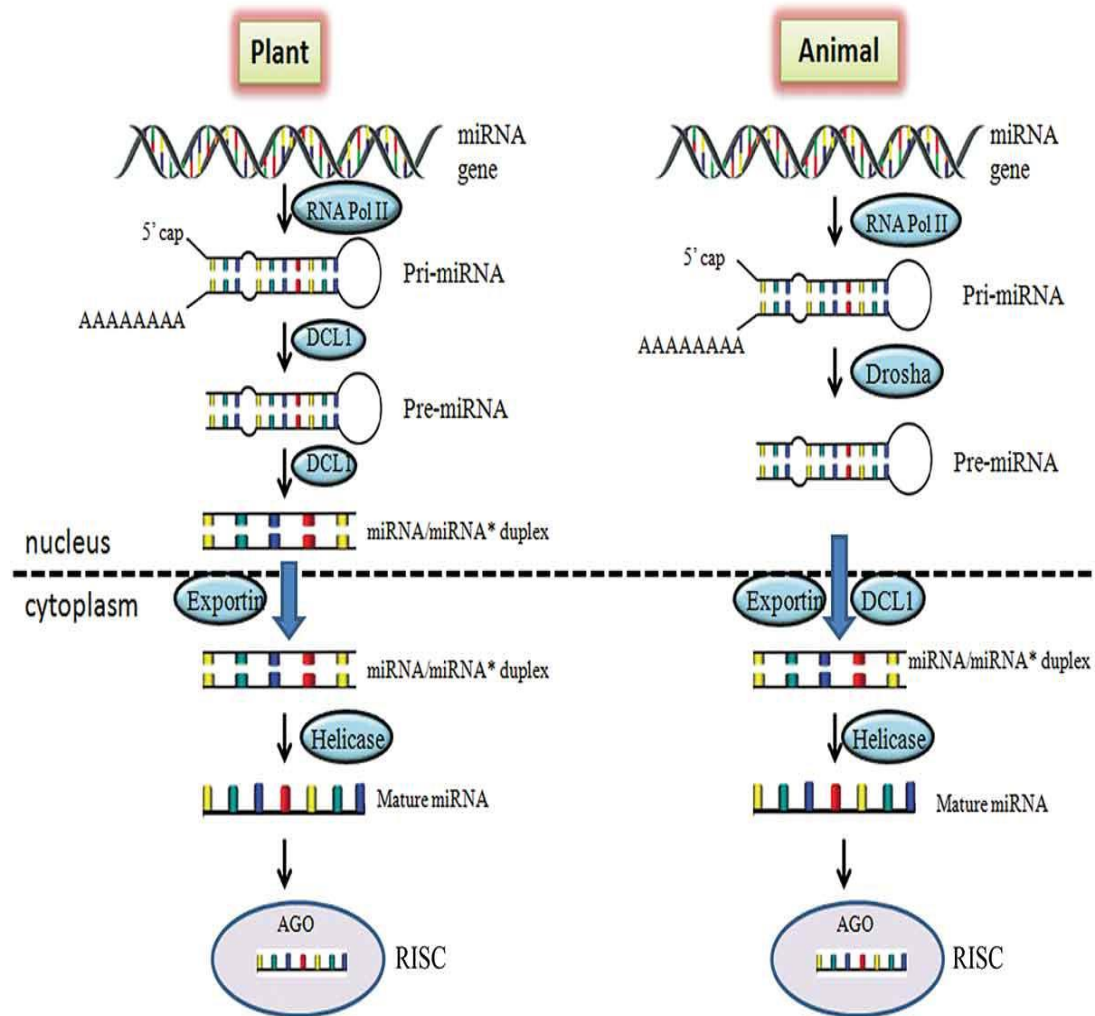


Figure 2.6 Comparison of miRNA biogenesis and function in plants and animals [89]

The binding sites of mature miRNAs within target genes exhibit location specificity in plants and animals. Initially, it was believed that miRNAs predominantly bind to the 3' untranslated region (UTR) of target genes in animals, while, they bind to the protein-coding region of genes in plants [90]. However, subsequent reports have indicated that miRNA binding sites can be found throughout the gene in both kingdoms.

In animals, miRNA binding sites have been observed in the 3' UTR [91], [92], 5' UTR, and coding region of a target gene [91], [92] and [93]. miRNAs regulate the expression of their target mRNAs through mechanisms such as mRNA cleavage or translational repression. Previously, it was thought that mRNA cleavage primarily occurs in plants, while translational repression is the dominant mechanism in animals [90]. However, both mechanisms can operate in plants and animals [94]. The ultimate objective of miRNA interaction is to inhibit the production of the encoded proteins.

2.3.5 Cross-kingdom Interactions Between Plant miRNAs and Human miRNAs

Plants have a long history of being used for therapeutic purposes, including the treatment of cancer, due to their natural compounds [95], [96]. Furthermore, certain plant-derived miRNAs present in foods can accumulate in human serum, tissues, and feces depending on a diet [97], [98] and [99]. Exogenous miRNAs from plants enter the human body through food and can have significant effects on different metabolic processes. Therefore, it is important to understand the interactions of plant miRNAs and their human targets. For this purpose, Rakhmetullina and colleagues identified binding sites between 738 *Oryza sativa* miRNAs (osa-miRNAs) and 17,508 mRNAs of human genes. They observed that osa-miRNA families have binding sites in the mRNAs of several human target genes, which are also candidate genes for cancer, cardiovascular diseases, and neurodegenerative diseases [98].

In addition to *in silico* studies, many plant miRNAs have also been identified. The presence of 2'-O-methylation at the 3' end of plant miRNAs increases their stability, allowing them to survive in various environments [100]. The structural stability of plant miRNAs enables them to withstand adverse conditions within the gastrointestinal (GI) system, such as RNases, phagocytosis, and low pH [101].

Another study conducted by Marzano et al. [7] aimed to explore the similarities between plant and human miRNAs and identify potential functional analogs. The

researchers specifically investigated the effects of certain plant miRNAs on cell proliferation in human colorectal carcinoma cell lines. They collected plant and human miRNA sequences from primary databases and compared their similarities. 1,606 of the 2,588 human miRNAs displayed a perfect sequence match at the 50 end with the 3,172 plant miRNAs. They applied further selections based on the role of human target genes or miRNA in cell cycle regulation (as an oncogene, tumor suppressor, or a biomarker for prognosis or diagnosis in cancer). Their findings revealed that specific plant miRNAs exhibited significant effects on cell proliferation in the human colorectal carcinoma cell lines. This suggests that these plant miRNAs may function as functional analogs of human miRNAs, meaning they have similar regulatory roles and can affect similar biological processes. The identification of plant miRNAs with effects on human cell proliferation provides insights into the potential cross-kingdom regulation of cellular processes. It highlights the conservation and functional relevance of certain miRNAs across different organisms.

Jha et al. [102] also explored miRNAs derived from *Carica papaya* and their potential targets in humans. The researchers identified novel miRNAs from papaya and predicted their regulatory effects on specific human targets. These targets were found to be involved in a range of biological processes and diseases, including cancer, diabetes, mental illness, and platelet disorders. The identification of miRNAs from papaya and their potential targets in humans suggests a cross-kingdom regulatory mechanism where plant-derived miRNAs may interact with human genes.

Similarly, Mangukia et al. [103] examined the interaction of plant miRNAs obtained from fresh fruits against the viral genome of the COVID-19-causing agent, specifically targeting the 3' UTR of SARS-CoV-2. Plant miRNAs that interact against the Omicron (B.1.1.529) variant of SARS-CoV-2 have been identified in 37 countries/regions worldwide. This cross-kingdom virus-plant interaction was determined by the alignment of plant miRNAs on viral genomes found in fruits such as *Citrus sinensis* (orange), *Prunus persica* (peach), *Vitis vinifera* (grape), and *Malus domestica* (apple). Particularly, the interaction of C.

sinensis miRNA - csi miR169-3p and SARS-CoV-2 is remarkable, as the targeted 3' UTR region "CTGCCT" is conserved among all 772 Omicron variants worldwide. Therefore, this region "CTGCCT" and miRNA csi-miR169-3p may become promising therapeutic candidates to induce viral genome silencing. Thus, this study is important in highlighting how fruits tend to fight against viruses such as SARS-CoV-2 and how they can help an individual maintain a strong immune system [32]. It is worth noting that while there is growing interest in the cross-kingdom regulatory effects of plant miRNAs on human gene expression, further studies are needed to fully elucidate the extent and significance of these interactions. Continued research in this area holds the potential for uncovering novel insights into the intricate regulatory networks that govern gene expression in different organisms.

2.4 *Theobroma cacao* L.

Cacao (*Theobroma cacao* L.) (Figure 2.7) is a member of the Malvaceae family and, a neotropical plant native to the Amazon lowland rainforest. Due to their exceptional organoleptic qualities (taste, aroma, texture), high caloric and nutritional value, and abundance in carbohydrates (31%), lipids (84%), proteins (11%), fibers (16%) and their derivatives are among the most consumed food kinds worldwide [105], [106].



Figure 2.7 *Theobroma cacao* L.

Studies have revealed properties in cacao including antioxidant, antidiabetic, and antibacterial activities, cardioprotective effects, and anticarcinogenic and neuroprotective effects which are used for therapeutic purposes [9], [10].

The study conducted by Patil et al. [11] employed computational biology approaches to investigate the potential role of cacao in cancer therapy. The researchers aimed to identify bioactive substances present in cacao that could target proteins involved in cancer-related pathways. These compounds were hypothesized to potentially exhibit anticancer effects by modulating the activity of these proteins. To validate their findings, the researchers conducted in vitro studies using cacao extract. These experiments confirmed the antioxidant activity of cacao extract and also demonstrated cytotoxic effects on cancer cells. The results indicated that cacao extract had the potential to induce cell death in cancer cells, suggesting its possible utility in cancer therapy.

Coronado-Cáceres et al. [107] aimed to investigate the inhibitory effect of cacao protein hydrolysates (CPH) on pancreatic lipase, which is an enzyme involved in the digestion and absorption of dietary fats. The researchers evaluated the potential of cacao protein to act as an anti-obesity agent by inhibiting pancreatic lipase activity, resulting in reduced fat absorption. The findings of the study suggested that cacao protein hydrolysates exhibited inhibitory activity against pancreatic lipase. This implies that cacao protein has the potential to reduce the breakdown of dietary fats into absorbable forms, thereby limiting fat absorption in the body.

In another study, based on the proteases that cacao beans contain antioxidant molecules that have the potential to inhibit the type 2 coronavirus (SARS-CoV-2) that causes COVID-19, Yanez and colleagues analyzed 30 molecules obtained from cacao as a result of- *in silico* studies. They reported that the molecules show a high affinity of SARS-CoV-2, protease inhibitors. It represents an important step in the investigation of natural products obtained from *Theobroma cacao* for the structure-based design of anti-SARS-CoV-2 drugs [108].

Overall, these studies highlight the potential of plant-derived compounds, including miRNAs, in modulating human gene expression, influencing biological processes, and potentially offering therapeutic benefits in various diseases. However, further experimental studies are necessary to validate and expand upon these findings. Starting from this point, we aimed to define a detailed miRNA in the cacao plant, which has important content, by considering cross-kingdom interactions.



3

METHODS

In this thesis, new miRNA sequences were identified in cacao (*Theobroma cacao* L.) using *in silico* methods (Figure 3.1). For this purpose, reference miRNAs belonging to *Viridiplantae* were in the miRbase database. The obtained reference miRNA sequences were compared with the EST database of each cacao using the BLASTN program in the NCBI database. Secondary structures of suitable pre-miRNA sequences were determined using the "RNAfold Web Server" program the target genes of these miRNAs were identified with the "RNA Hybrid" program. The processes and/or pathways of target genes were determined by BLAST2GO and DAVID-KEGG programs. Furthermore, the networks of target genes of the novel cacao miRNAs were evaluated through the GeneMANIA Cytoscape plug-in.

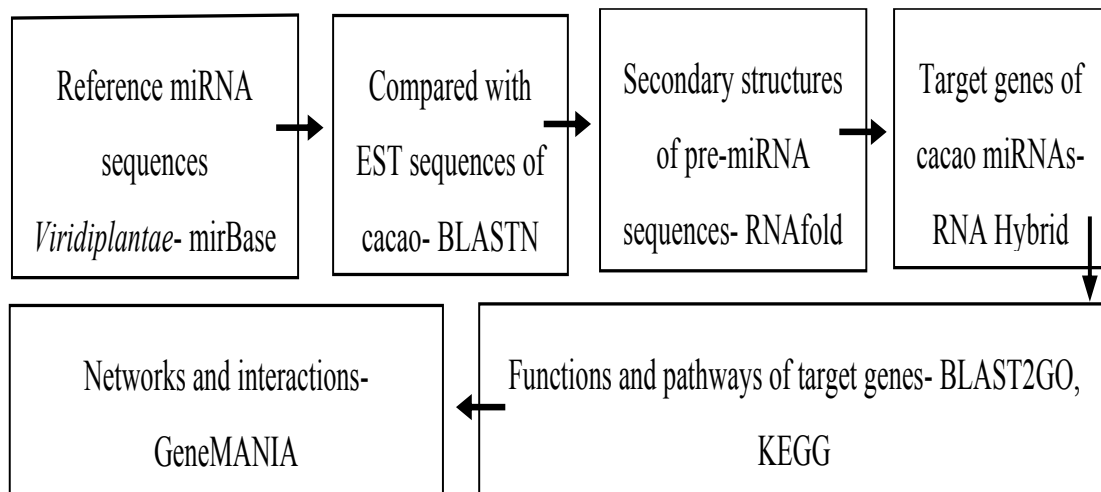


Figure 3.1 Flow chart showing the steps to be followed in *in silico* miRNA identification in cacao

3.1 Obtaining Reference miRNA Sequences

Since mature sequences were more preserved than precursor sequences, analyses were carried out with mature miRNA sequences [109]. References mature miRNAs belonging to eudicotyledon (a total of 6746 sequences) and monocotyledons (a total of 2422 sequences) were obtained from the miRBase database (v22), to identify novel cacao miRNAs [110]. Among sequences, repeating the identical sequences were manually checked and deleted.

3.2 BLAST Analyses

Reference miRNA sequences were compared with the cacao EST sequence by using the BLASTN program in the NCBI database. Analyses were performed using the following parameters.

- *Database: Expressed Sequence Tags (EST)*
- *Organisms: Theobroma cacao (taxid:3641)*
- *Program Selection: Somewhat similar sequences*
- *Expected threshold: 1000*
- *Number of Aligned Sequences: 1000*
- *Word Size: 7*

Other parameters were used as specified in the program [111]. These matching sequences were checked manually by comparing reference miRNA sequences. A maximum of 7 mismatches and/or indels (insertion/deletion, $n=7$, the length of the miRNA query) between reference miRNA sequences and matching EST sequences were determined and then they were used for use in the next stage of the elimination process [112].

3.3 Elimination Process

During the identification of novel miRNAs via *in silico* analyses, researchers utilized different programs, and every program has its parameters. The first step in the elimination process is to eliminate repetitive sequences belonging to the identical gene or EST sequence. As a result of BLAST analysis, more than one gene or EST sequence can be found matching the same miRNA sequence. These sequences belong to the same gene or EST sequence and are available in the NCBI database only as partial sequences with different accession numbers. Results are obtained after analyzing similar sequences, and therefore this causes false positive findings.

For this purpose, repetitive sequences belonging to the same EST sequence determined in the previous stage were eliminated using the BLASTN program to create a 'single tone' for each miRNA sequence [113].

The second important step is the elimination of protein-coding miRNAs. For this purpose, miRNAs were compared with the protein database in NCBI using the BLASTX program to identify protein-coding sequences [113]. As a result of the analysis, all protein-coding sequences were eliminated. The dissimilar sequences were selected as potential candidate miRNAs.

3.4 RNAFold Analyses for Prediction of Secondary Structures

The stem-loop structures of the potential miRNA sequences were identified by using "RNAfold Web Server" (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) in "ViennaRNA Package" (Version 2.4.17) [114] "Minimum Free Energy" algorithm is utilized in this program [115]. The results obtained were evaluated by considering the following criteria.

- The 'stem-loop' hairpin must be in an appropriate position.
- The mature miRNA sequence should be in only one arm of the hairpin structure.

- There should be no loops or breakage in the mature miRNA* (opposite arm) sequences.
- The MFEI (Minimum Folding Energy Index) should be >0.85 and the MFE (Minimum Free Energy) value should be $\leq -18 \text{ kcal mol}^{-1}$.
- Matches between mature miRNA and miRNA sequences must include at least 12 of the Watson-Crick or G/U base pairs [116], [117].

Novel miRNA sequences, EST sequences, EST positions, MFE values, AMFE (Adjusted Minimal Folding Free Energy), GC %, AT %, and MFEI values were saved. The following formulas were used to calculate AMFE and MFEI values [118].

$$\text{AMFE} = ((\text{MFE}/(\text{potential pre-miRNA length})) * 100.$$

$$\text{MFEI} = (\text{AMFE}/(\text{GC}\%))$$

3.5 Determination of miRNA Target Genes

Protein targets of potential miRNAs were determined using a combination of BLASTN and "RNA Hybrid" programs [119], [120]. BLASTN parameters are indicated below:

- *Database: Nucleotide Collection (nr/nt)*
- *Organisms: Theobroma cacao (taxid:3641)*
- *Program Selection: Highly similar sequences (megablast)*

Results with query coverage of $\geq 60\%$ were selected and used for the prediction of target genes. For this purpose, the RNA hybrid program [121] was utilized with the following criteria to reduce false positive targets [122].

- No more than 4 nt mismatches between target miRNA and potential mRNA.
- These 4 mismatches should not be at positions 10 and 11 from the 5' end of the miRNA.

- There should be no more than 1 nt mismatch between positions 1 and 9 from the 5' end of the miRNA.
- There should be no more than 2 nt of mismatches or gaps adjacent to any region between the target miRNA and the potential mRNA.
- The MFE value for the 2-7 and/or 8-13 nt region from the 5' end of the miRNA should be ≤ -20 kcal mol⁻¹. These criteria ensured the elimination of false positive results.

3.6 Go and KEGG Analyses of Target Cacao Genes

Gene Ontology (GO) [123] analyses of cacao miRNA target genes were performed in the BLAST2GO [124] program. Furthermore, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was carried out in the DAVID (<https://david.ncifcrf.gov/>) database (May 2022 version) [125]. GO provides information about molecular function, cellular components, and biological processes, while KEGG [126] provides metabolic pathway information of target genes.

3.7 Network Interactions

Network and interactions between cacao miRNA targets and human genes were also evaluated through the GeneMANIA Cytoscape (<https://genemania.org/>) [127]. The networks are based on co-expression, co-localization, genetic and physical interactions as well as predicted and shared gene protein domains were taken into consideration while building the networks. Light pink indicates physical interactions, orange for predicted interactions purple for possible interactions of co-expressed genes, green for genetic interactions, blue for co-localization, turquoise for interactions on pathways, and yellow for shared protein domain interaction.

CONCLUSION

In this study, we aimed to conduct *in silico* analyses to define miRNA functions in cacao on the mechanism of human diseases, especially cancer. For this purpose, reference miRNA sequences of *Viridiplantae* in the mirBase database were obtained. The obtained reference miRNA sequences were compared with the EST (Expressed Sequence Tag) database of cacao via the BLASTN (Basic Local Alignment Search Tool Nucleotide) program in the NCBI (National Center for Biotechnology Information) database. The RNAfold Web Server program was used to construct secondary structures of suitable pre-miRNA (precursor-miRNA) sequences. Secondary structures of suitable pre-miRNA (precursor-miRNA) sequences were determined using the "RNAfold Web Server" program. The target genes of these miRNA sequences were identified with the "RNA Hybrid" program. The processes and metabolic pathways in which target genes play a role were analyzed with BLAST2GO and KEGG (Kyoto Encyclopedia of Genes and Genomes) programs. Network and interactions between cacao miRNA targets to human genes were evaluated through the GeneMANIA (Gene Multiple Relationship Network Integration Algorithm)- Cytoscape plugin. 11 new miRNAs from 10 miRNA families were identified and their 47 target genes were determined in this study. It was determined that these genes play a role in different metabolic pathways. Important proteins specific to human cancer were detected in the KEGG pathways, which are related to the target genes of miRNAs. Based on these important proteins (*UPFI*, *SNCG*, *RCHY1*, *DDX5*, *NOL6*), cacao plant-human interaction was detected with the Cytoscape Genemania extension.

4.1 Identification of miRNAs

Similar to previous studies [128], EST sequences with ≤ 7 mismatches with the reference sequence were determined as a result of comparing the reference miRNA sequences and the EST databases of the cacao plant BLASTN program. Then similarity between the obtained EST sequences was analyzed by using the BLASTN program and a 'single tone' was created for each miRNA sequence by deleting the repetitive EST sequences (Table 4.1). Then, protein-coding sequences were identified among these sequences using the BLASTX program, and these sequences were deleted. The remaining sequences were recorded to use further analyses.

Table 4.1 Sequences of novel miRNAs in the cacao genome, red letters show mismatches against reference miRNA(s)

EST ACCESSION	MATCHING	REFERENCE SEQUENCE	MISMATCH	E VALUE	MATCH VALUE
CU621030.1	Query 5 GGACCAGGCTTCATTCC 21	>crt-miR166b MIMAT0014085	TCGGACCAGGCTTCATTCCCTT	0.026	17/17
	Sbjct 139 GGACCAGGCTTCATTCC 155	UCGGACCAGGCUUCAUUCCCUU			
CU601757.1	Query 5 GGACCAGGCTTCATTCC 21	>crt-miR166b MIMAT0014085	TCGGACCAGGCTTCATTCCCTT	0.026	17/17
	Sbjct 589 GGACCAGGCTTCATTCC 605	UCGGACCAGGCUUCAUUCCCUU			
CU535889.1	Query 5 GGACCAGGCTTCATTCC 21	>crt-miR166b MIMAT0014085	TCGGACCAGGCTTCATTCCCTT	0.026	17/17
	Sbjct 216 GGACCAGGCTTCATTCC 232	UCGGACCAGGCUUCAUUCCCUU			
CU530930.1	Query 1 TCTTGACCAGGCTTCATTCC 21	>csi-miR166i-3p MIMAT0048881	GGAATGAAGCCTGGTCCAAGA	0,0001	21/21
	Sbjct 393 TCTTGACCAGGCTTCATTCC 373	UCUUGGACCAGGCUUCAUUC			
CU526475.1	Query 1 ATTCAGTTGATGCAAGGCGGGATC 24	>aau-miR168 MIMAT0022678	GATCCCGCCTTGCATCAACTGAAT	0,000002	24/24
	Sbjct 248 ATTCAGTTGATGCAAGGCGGGATC 225	AUUCAGUUGAUGCAAGGCGGGAUC			
CU544356.1	Query 1 AAGCCAAGGATGACTTGCCG 20	>gma-miR169c MIMAT0007357	AAGCCAAGGATGACTTGCCGA	0.10	19/20
	Sbjct 166 AAGCCAAGGATGAGTTGCCG 185	AAGCCAAGGAUGACUUGCCGA			
CU522244.1	Query 1 TTCCTGTCTACTTGA 15	>ppt-miR1051-3p MIMAT0005168	TTCCTGTCTACTTGA GTGAAG	0.41	15/15
	Sbjct 43 TTCCTGTCTACTTGA 57	UUCCUGUCUACUUGAAGCCAC			
CU619562.1	Query 6 CTGGATTTGTTGACAA 21	>gra-miR3476 MIMAT0033982	TGGTCCTGGATTTGTTGACAA	0.10	16/16
	Sbjct 357 CTGGATTTGTTGACAA 372	UCGGACUGGAUUUGUUGACAA			
CU498864.1	Query 6 CTGGATTTGTTGACA 20	>gra-miR3476 MIMAT0033982	CAAGTCTGGATTTGTTGACAT	0.41	15/15
	Sbjct 32 CTGGATTTGTTGACA 46	UCGGACUGGAUUUGUUGACAA			

Table 4.1 Sequences of novel miRNAs in the cacao genome (continued)

EST ACCESSION	MATCHING	REFERENCE SEQUENCE	MISMATCH	E VALUE	MATCH VALUE
CU630909.1	Query 7 TGGATTGTTGACAA 21	>gra-miR3476 MIMAT0033982	TTGTCAACAAATCCAATAAGC	0.41	15/15
	Sbjct 28 TGGATTGTTGACAA 14	UCGGACUGGAUUUGUUGACAA			
CU550802.1	Query 4 ATTTGAGAAGTTGAG 18	>aly-miR4249 MIMAT0017940	TATCTCAACTTCTCAAATCCT	0.41	15/15
	Sbjct 518 ATTTGAGAAGTTGAG 504	UGAAUUUGAGAAGUUGAGGUA			
CU500027.1	Query 4 ACTCCAGCATATTATA 19	>nta-miR6161d MIMAT0024773	GCGACTCCAGCATATTATAAT	0.10	16/16
	Sbjct 90 ACTCCAGCATATTATA 105	UGAACUCCAGCAUAUUUAUACU			
CU583121.1	Query 1 TAGACAAGAATTCCA 15	>ppe-miR6279 MIMAT0027308	CTAGCATGGAATTCTTGTCTA	0.41	15/15
	Sbjct 311 TAGACAAGAATTCCA 297	UAGACAAGAAUCCAGAGACC			
CU491311.1	Query 3 AACCTGAGTCAGATGATGA 21	>cpa-miR8150 MIMAT0031847	CGCTCATCATCTGACTCAGGTTT	0.53	18/19
	Sbjct 110 AACCTGAGCCAGATGATGA 92	AAAACCUGAGUCAGAUGAUGAGCG			
CU601517.1	Query 5 CTGAGTGAGCATCAC 19	>sly-miR9471b-3p MIMAT0035442	CAGTGATGCTCACTCAGCAA	0.41	15/15
	Sbjct 479 CTGAGTGAGCATCAC 465	UUGGCUGAGUGAGCAUCACUG			

4.2 Prediction of Secondary Structures

Secondary structures of pre-miRNAs were constructed by using strict criteria recommended in previous studies [129], [130], [131] and [132]. For this purpose, "RNAfold Web Server" program was used and 11 novel miRNAs belonging to 10 different miRNA families including tcc-miR166b, tcc-miR166i, tcc-miR168, tcc-miR169, tcc-miR1051, tcc-miR3476, tcc-miR4249, tcc-miR6161, tcc-miR6279, tcc-miR8150 and tcc-miR9471 were identified (Figure 4.1, Figure 4.2, Figure 4.3, Figure 4.4, Figure 4.5, Figure 4.6, Figure 4.7, Figure 4.8, Figure 4.9, Figure 4.10, Figure 4.11).



Figure 4.1 Secondary structure of the tcc-miR166b sequence identified in *Theobroma cacao*

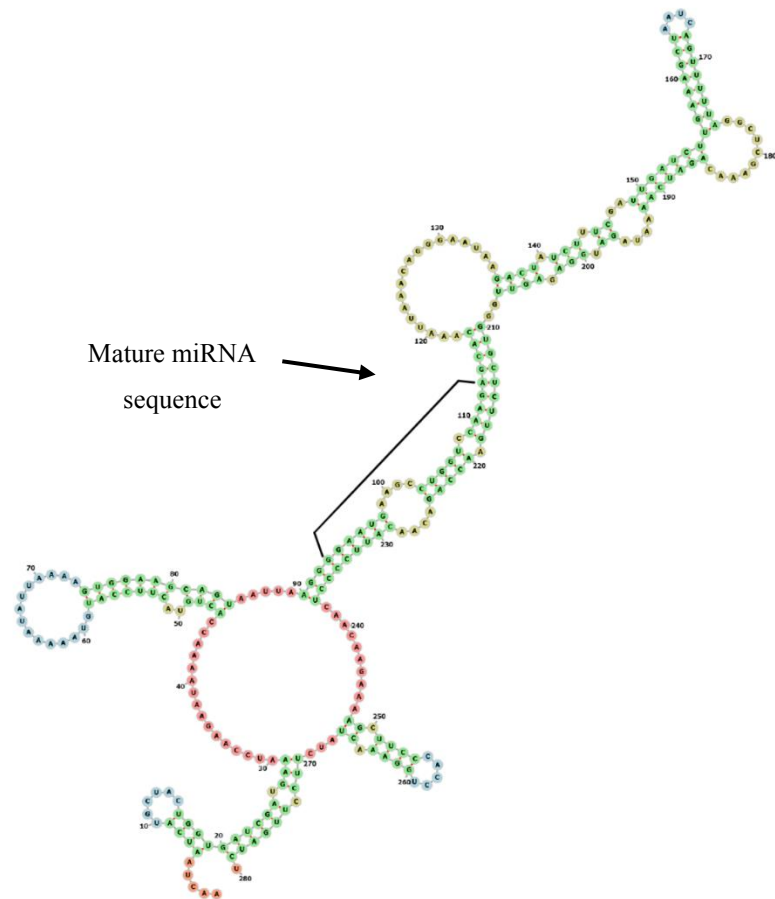


Figure 4.2 Secondary structure of the tcc-miR166i sequence identified in *Theobroma cacao*

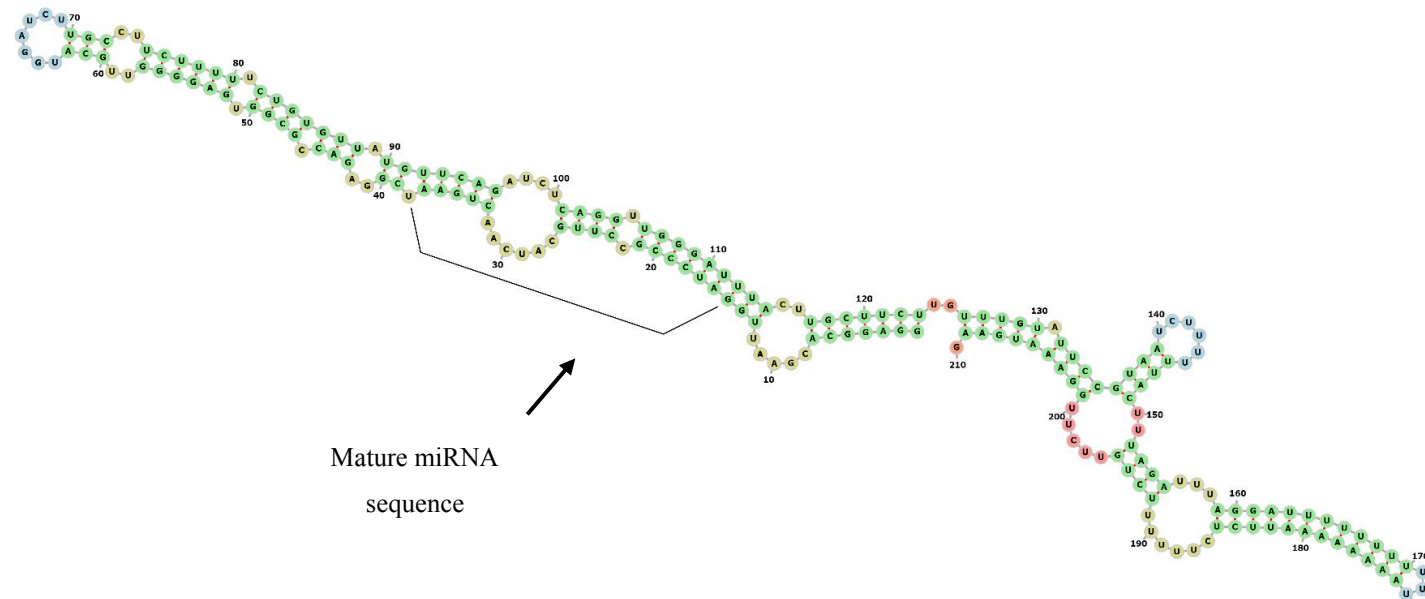


Figure 4.3 Secondary structure of the tcc-miR168 sequence identified in *Theobroma cacao*

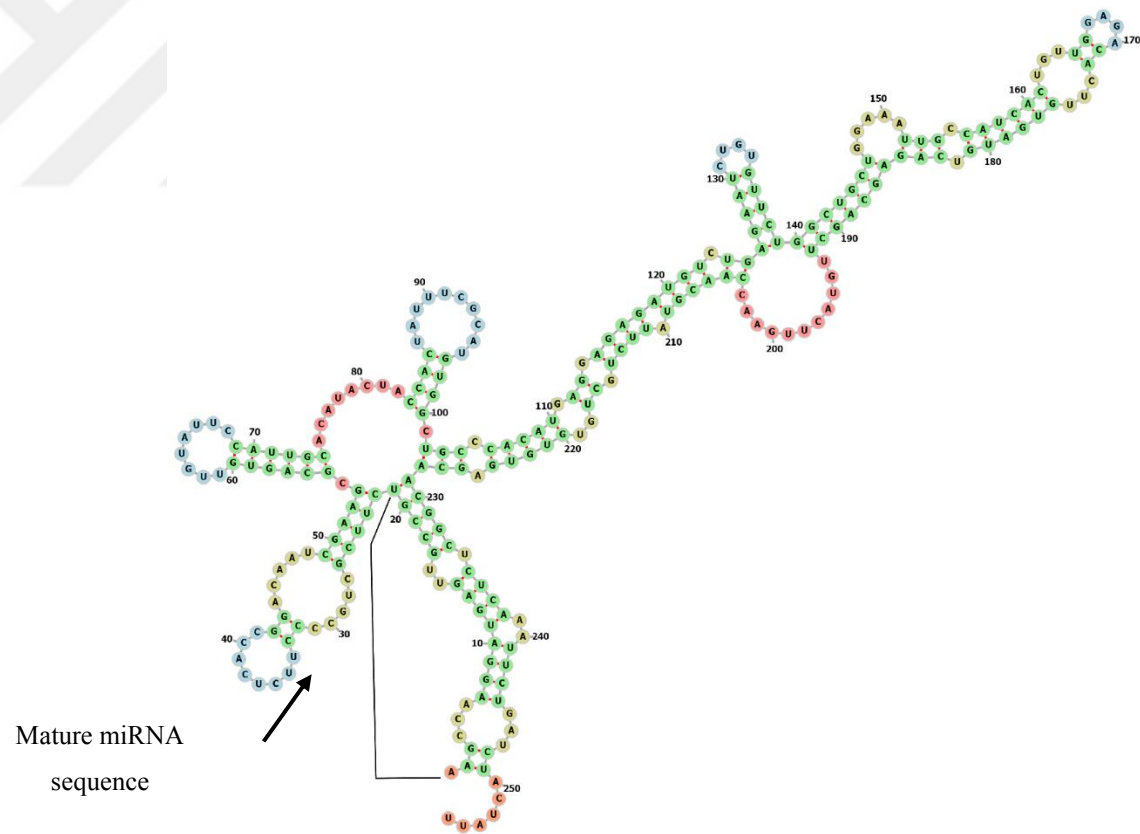


Figure 4.4 Secondary structure of the tcc-miR169 sequence identified in *Theobroma cacao*



Mature miRNA
sequence



Figure 4.5 Secondary structure of the tcc-miR1051 sequence identified in *Theobroma cacao*

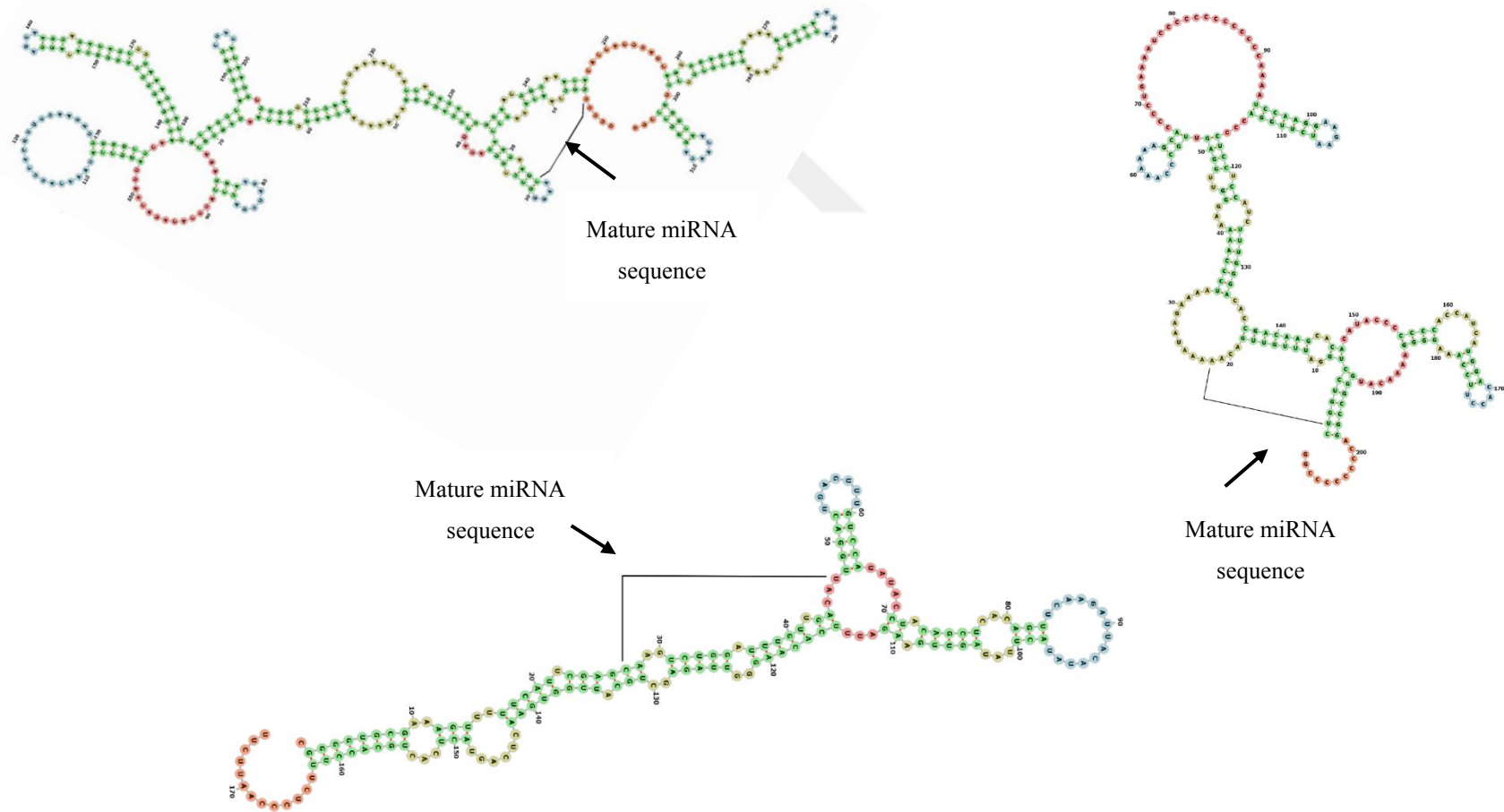


Figure 4.6 Secondary structure of the tcc-miR3476 sequence identified in *Theobroma cacao*

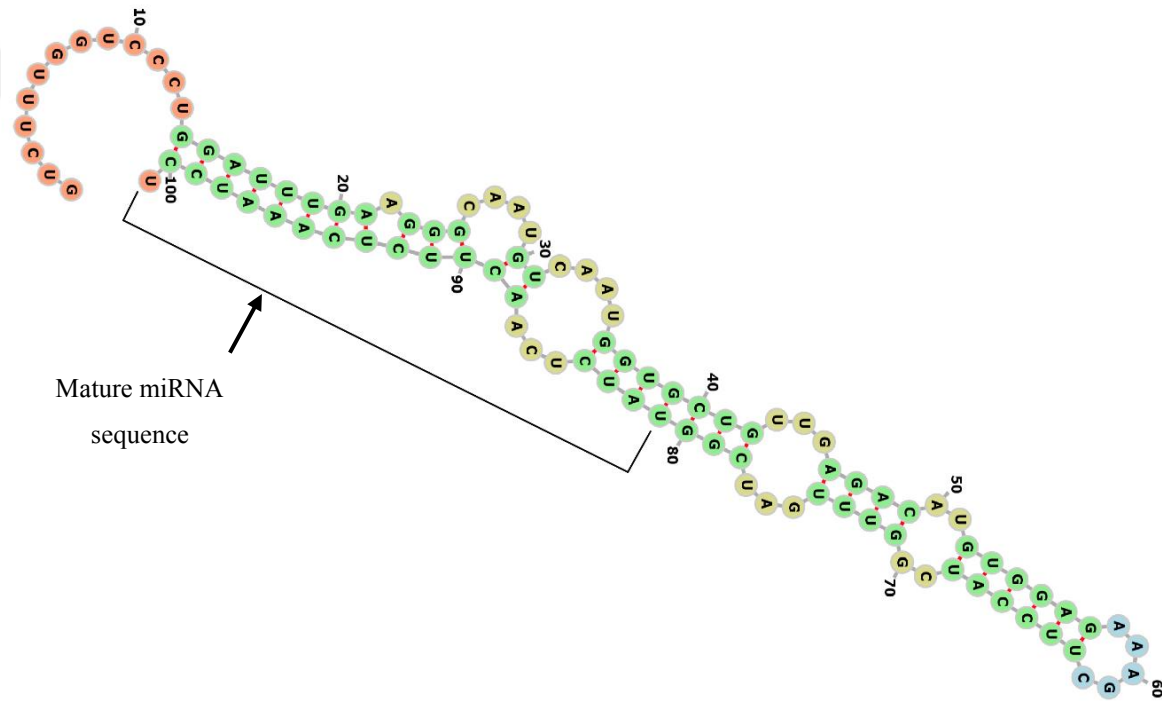


Figure 4.7 Secondary structure of the tcc-miR4249 sequence identified in *Theobroma cacao*

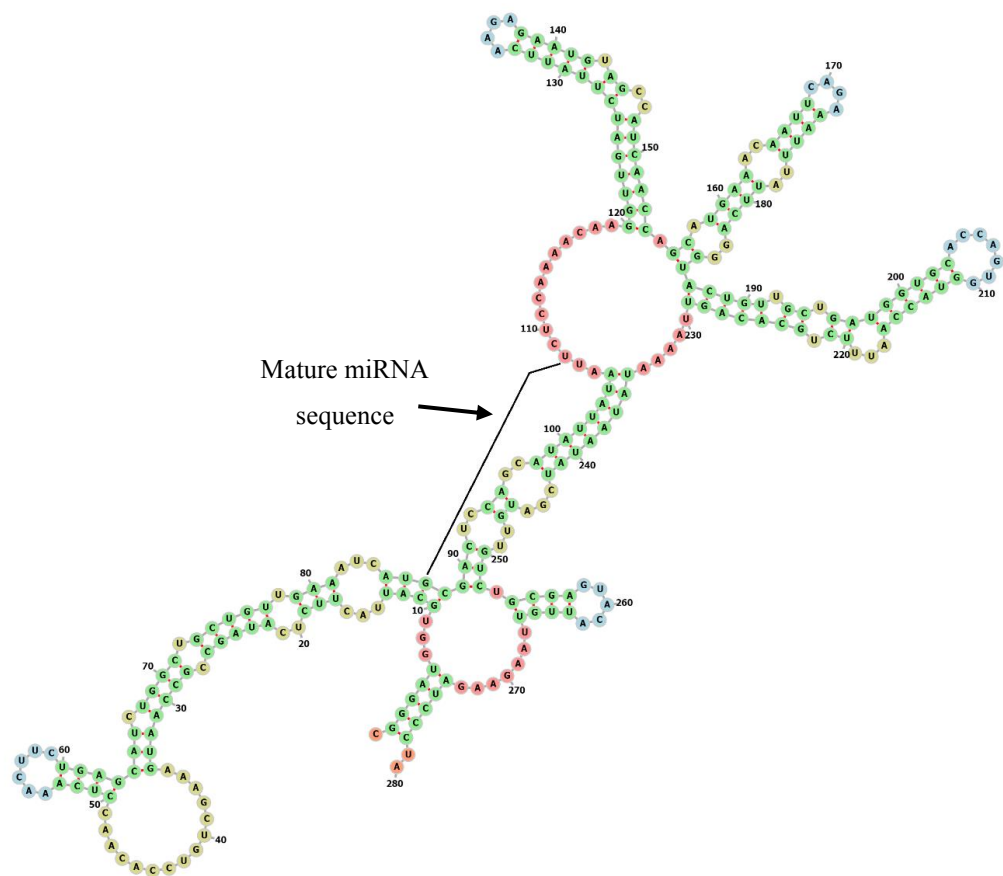


Figure 4.8 Secondary structure of the tcc-miR6161 sequence identified in *Theobroma cacao*

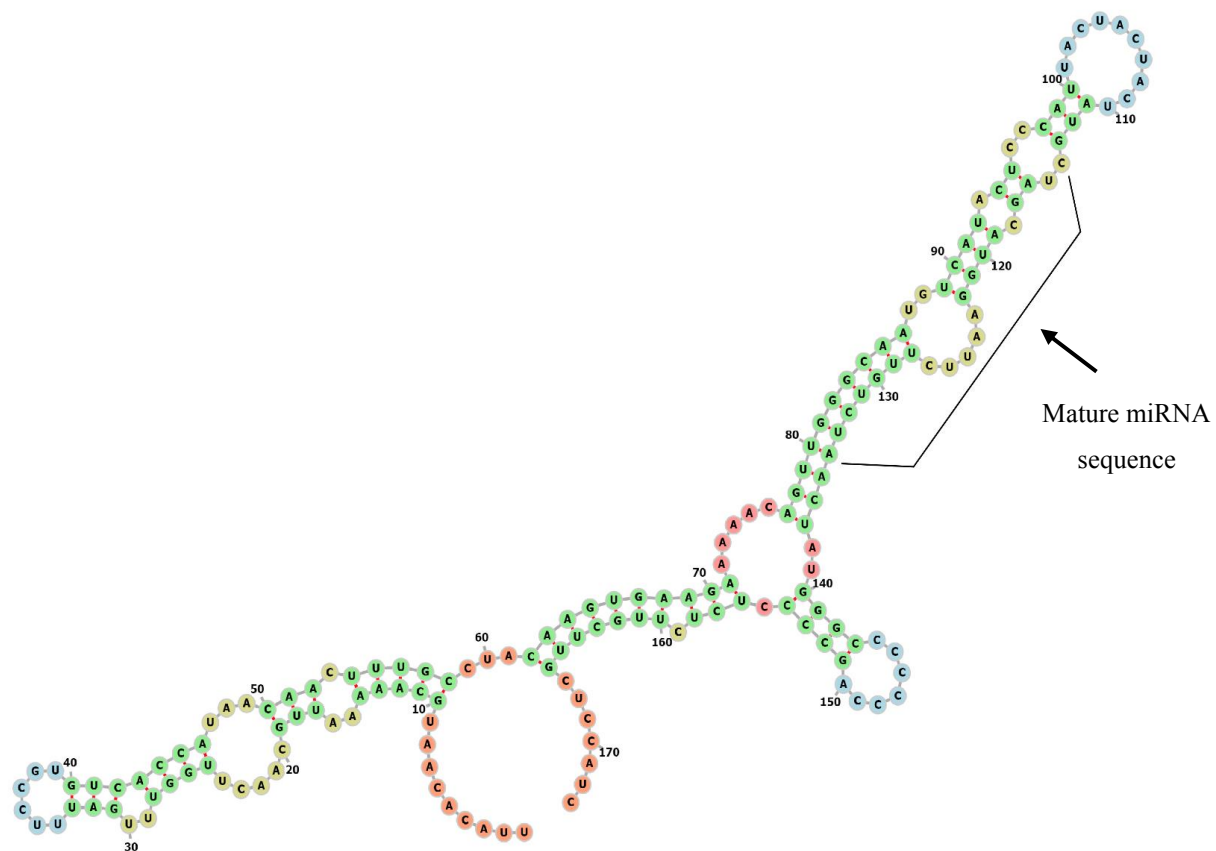


Figure 4.9 Secondary structure of the tcc-miR6279 sequence identified in *Theobroma cacao*

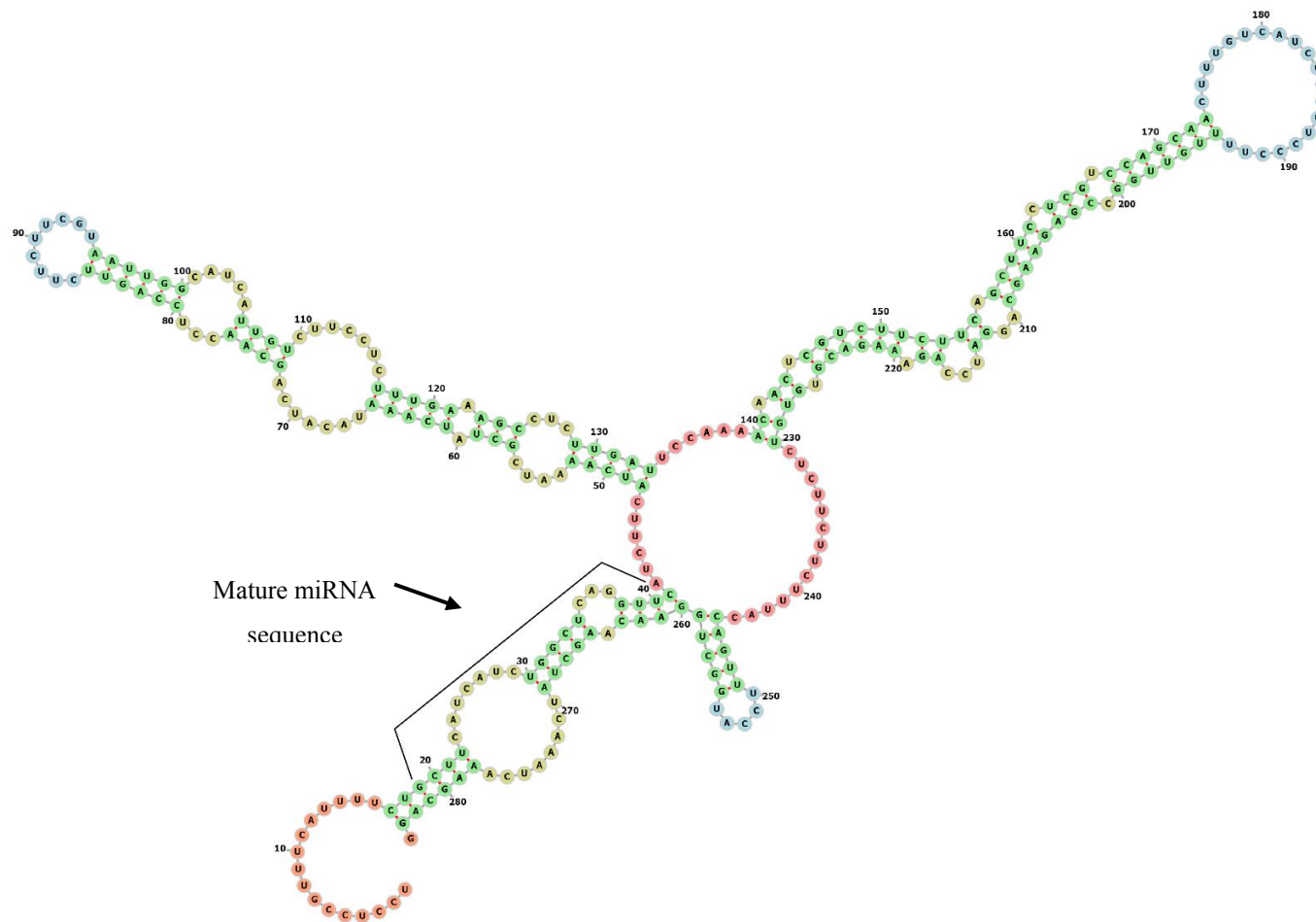


Figure 4.10 Secondary structure of the tcc-miR8150 sequence identified in *Theobroma cacao*

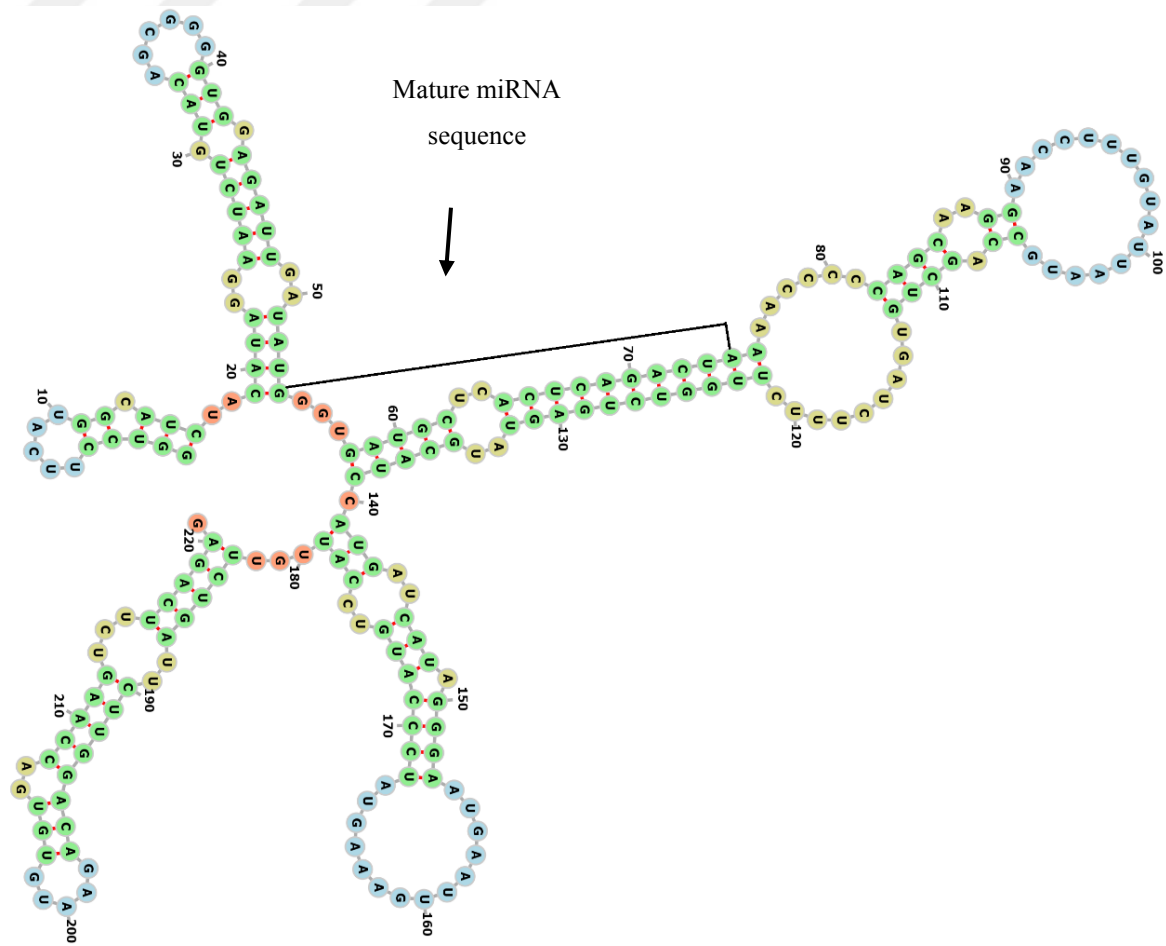


Figure 4.11 Secondary structure of the tcc-miR9471 sequence identified in *Theobroma cacao*

The newly discovered miRNAs in cacao exhibit stem-loop structures with a length of at least 12 nt, with typical Watson-Crick or G/U base pairings between the mature miRNA and its complementary sequence (miRNA*) in the stem region. These findings are consistent with previous studies [133], [134]. The length of the identified cacao pre-miRNAs ranges from 101 to 319 nt (Table 4.2), with an average length of 241 nt, which is in line with previous reports on mature miRNA sequences [135], [136]. Moreover GC% content of the miRNAs varies from 2.19% to 6.93%, with an average of 4.43%. The minimum free energy index (MFEI) values of the predicted miRNA precursor sequences range from 3.51 to 8.97, with an average of 5.53. These findings are consistent with research conducted by different research groups on a variety of plant and animal species [137], [138]. The MFE of the pre-miRNA is an important characteristic used for characterization. The rule that MFE should be $\leq -18 \text{ kcal mol}^{-1}$ was also considered when creating secondary structures. The newly identified cacao pre-miRNAs exhibit MFE values ranging from -99.3 to -25.8 kcal mol^{-1} , with an average of -60.35 kcal mol^{-1} . Concordant with these results, [139], [140] reported similar results.

Table 4.2 Characteristics of newly identified miRNAs in cacao genome

miRNA	EST ACCESSION	MISMATCH	MFE ($\leq$ -18 kcal/mol)	PRECURSOR HAIRPIN LENGHT (nt)	AMFE	MATURE miRNA G+C NUMBER	GC%	MFEI (>0.85)	EST POSITION
miR166b	CU621030.1	TCGGACCAGGCTTCATTCCC TT	-68,1	291	23,40206186	14	4,81	4,86	117-138 nt
miR166b	CU601757.1	TCGGACCAGGCTTCATTCCC TT	-99,3	299	33,21070234	12	4,01	8,28	270-291 nt
miR166b	CU535889.1	TCGGACCAGGCTTCATTCCC TT	-88,9	291	30,54982818	13	4,47	6,84	118-139 nt
miR166i	CU530930.1	GGAATGAAGCCTGGTCCAAGA	-72,6	280	25,92857143	11	3,93	6,60	93-113 nt
miR168	CU526475.1	GATCCCGCCTTGCATCAACTGAAT	-47,4	210	22,57142857	12	5,71	3,95	15-38 nt
miR169	CU544356.1	AAGCCAAGGATGACTTGCCGA	-67	255	26,2745098	11	4,31	6,09	1-21 nt
miR1051	CU522244.1	TTCCTGTCTACTTGA AGCCAC	-31,6	227	13,92070485	9	3,96	3,51	1-21 nt
miR3476	CU619562.1	TCGGACT GGATTGTGACAA	-43,8	209	20,9569378	9	4,31	4,87	2-22 nt
miR3476	CU498864.1	TCGGACT GGATTGTGACAA	-34,5	175	19,71428557	9	5,14	3,83	27-47 nt
miR3476	CU630909.1	TTGTCAACAAATCCA GTCCGA	-62,8	319	19,68652038	7	2,19	8,97	5-25 nt
miR4249	CU550802.1	TACCTCAACTTCTCAAATTCA	-25,8	101	25,54455446	7	6,93	3,69	81-100 nt
miR6161	CU500027.1	TGAACTCCAGCATATTATACT	-60,3	280	21,53571429	8	2,86	7,54	87-107 nt
miR6279	CU583121.1	GGTCTCTGGAATTCTGTGCTA	-31,1	173	17,9768786	8	4,62	3,89	114-134 nt
miR8150	CU491311.1	CGCTCATCATCTGACTCAGGTTTT	-56,6	284	19,92957746	12	4,23	4,72	19-42 nt
miR9471	CU601517.1	CAGTGATGCTCACTCAGCCAA	-57,8	221	26,1538462	11	4,98	5,25	54-74 nt

¹Red letters indicate mismatches against reference miRNA(s), GC% GC content percentage, AMFE adjusted MFE, MFEI minimal folding free energy index, MFE minimum.

4.3 Targets and Functions of Potential miRNAs

After identifying miRNA sequences in cacao and creating secondary structures, the targets of potential novel miRNAs were determined in the cacao genome. After the identification of target genes, GO analyses were performed to figure out which processes the target genes were associated with. In addition, KEGG analyses were also carried out to evaluate which metabolic pathways the genes played a role in.

4.3.1 RNA Hybrid, GO, and KEGG Analysis of miR166 Target Genes

The target gene of the tcc-miR166b and tcc-miR166i were identified in *Theobroma cacao* genome. A total of 12 target genes were identified and 1 of which was uncharacterized. It was observed that these miRNAs play a role in the regulation of different genes that play a role in different metabolic pathways (Table 4.3).

Table 4.3 The targets tcc-miR166b and tcc-miR166i

miRNA	GENBANK ACCESSION	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc-miR166i	XM_007042601.2 XM_007042602.2	Telomere repeat-binding factor 1(LOC18608101)	GGAATGAAGCCTGGTCCAAGA	target 5' G CU U 3' UUUACAGGCUUUCUUG GAUGG AGGUGUCCGAAAGAAC UUGCC miRNA 3' 5'	-32.7 kcal/mol
tcc-miR166b	XM_018116764.1	Telomere repeat-binding factor 1(LOC18608101)	TCGGACCAGGCTTCATTCCCTT	target 5' C AUU AGG UU A 3' UCCA GGGCU UCUUGAGC GG AGGU UCCGA AGAACUUG CC miRNA 3' G A 5'	-28.2 kcal/mol
tcc-miR166b	XM_007042601.2 XM_007042602.2	Telomere repeat-binding factor 1(LOC18608101)	TCGGACCAGGCTTCATTCCCTT	target 5' G CU U 3' UUUACAGGCUUUCUUG GAUGG AGGUGUCCGAAAGAAC UUGCC miRNA 3' 5'	-32.7 kcal/mol
tcc-miR166b	XM_007042601.2 XM_007042602.2	Telomere repeat-binding factor 1(LOC18608101)	TCGGACCAGGCTTCATTCCCTT	target 5' G CU U 3' UUUACAGGCUUUCUUG GAUGG AGGUGUCCGAAAGAAC UUGCU miRNA 3' 5'	-30.9 kcal/mol
tcc-miR166i	XM_018116764.1	Growth-regulating factor 10(LOC18606399)	GGAATGAAGCCTGGTCCAAGA	target 5' C AUU AGG UU A 3' UCCA GGGCU UCUUGAGC GG AGGU UCCGA AGAACUUG CC miRNA 3' G A 5'	-28.2 kcal/mol
tcc-miR166i	XM_018127495.1	Callose synthase 3(LOC18589684)	GGAATGAAGCCTGGTCCAAGA	target 5' G C U C 3' CUGCAGGCUUUC UUGA AUGG GGUGUCCGAAAG AACU UGCC miRNA 3' A 5'	-30.1 kcal/mol

Table 4.3 The targets tcc-miR166b and tcc-miR166i (continued)

miRNA	GENBANK ACCESSION	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc- miR166b	XM_018127495.1	Callose synthase 3 (LOC18589684)	TCGGACCAGGCTTCATTCCCTT	target 5' G C U C 3' CUGCAGGCUUUC UUGA AUGG GGUGUCCGAAAG AACU UGCC miRNA 3' A 5'	-30.1 kcal/mol
tcc- miR166i	XM_007031567.2	GDSL esterase/lipase At5g14450 (LOC18600859)	GGAATGAAGCCTGGTCCAAGA	target 5' C UC CAAGC C 3' UCC CAGGCUUUCUUGA AUGG AGG GUCCGAAAGAACU UGCC miRNA 3' U 5'	-31.7 kcal/mol
tcc- miR166b	XM_007031567.2	GDSL esterase/lipase At5g14450 (LOC18600859)	TCGGACCAGGCTTCATTCCCTT	target 5' C UC CAAGC C 3' UCC CAGGCUUUCUUGA AUGG AGG GUCCGAAAGAACU UGCC miRNA 3' U 5'	-31.7 kcal/mol
tcc- miR166b	XR_001926196.1, XR_001926197.1	Heparan-alpha- glucosaminide N- acetyltransferase (LOC18613653)	TCGGACCAGGCTTCATTCCCTT	target 5' U A GGGUAA A U 3' UCUAC AGGC UUUC UGGGCG AGGUG UCCG AAAG ACUUGC A C 5'	-23.6 kcal/mol
tcc- miR166b	XM_007026095.2	Lysine-specific demethylase JMJ25 (LOC18597208)	TCGGACCAGGCTTCATTCCCTT	target 5' G C GUAAG AAUU C 3' CC CGGGC UUUUUGAA CGA GG GUCCG AAGAACUU GCU miRNA 3' A U A 5'	-23.5 kcal/mol
tcc- miR166b	XM_007037819.2	Uncharacterized (LOC18605043)	TCGGACCAGGCTTCATTCCCTT	target 5' G UCUGUAAAAGCUUGG AU U 3' GACAAGGCUUUCU UGAA UGA CUGUUCCGAAAGA ACUU GCU miRNA 3' A 5'	-28.2 kcal/mol

GO analysis of target genes showed that genes were divided into 3 categories: biological process (BP), cellular component (CC), and molecular function (MF). It consists of the Biological process (BP), regulation of cellular process (4,35%), demethylation (4,35%), regulation of metabolic process (4,35%), cellular component organization or biogenesis (8,7%), nitrogen compound metabolic process (13,04%), biosynthetic process (13,04%), organic substance metabolic process (17,39%), primary metabolic process (17,39%) and cellular metabolic process (17,39%). Within the Molecular Function (MF), catalytic activity, acting on a protein (6,25%), structural constituent of ribosome (6,25%), hydrolase activity (6,25%), carbohydrate derivative binding (6,25%), demethylase activity (6,25%), small molecule binding (6,25%), ion binding (6,25%), heterocyclic compound binding (18,75%), organic cyclic compound binding (18,75%) and transferase activity (18,75%) was observed. Membrane protein complex (4,54%), cell periphery (4,55%), catalytic complex (4,55%), intrinsic component of membrane (9,09%), chromatin (9,09%), DNA packaging complex (9,09%), protein-DNA complex (9,09%), membrane (13,64%), intracellular anatomical structure (18,18%) and organelle (18,18%) were also detected in Cellular Component (CC) (Figure 4.12). After GO analysis, KEGG pathway searches were performed in the DAVID database to identify biological pathways of genes. The results showed that tcc-miR166b was involved in Glycosaminoglycan degradation (Figure 4.13).

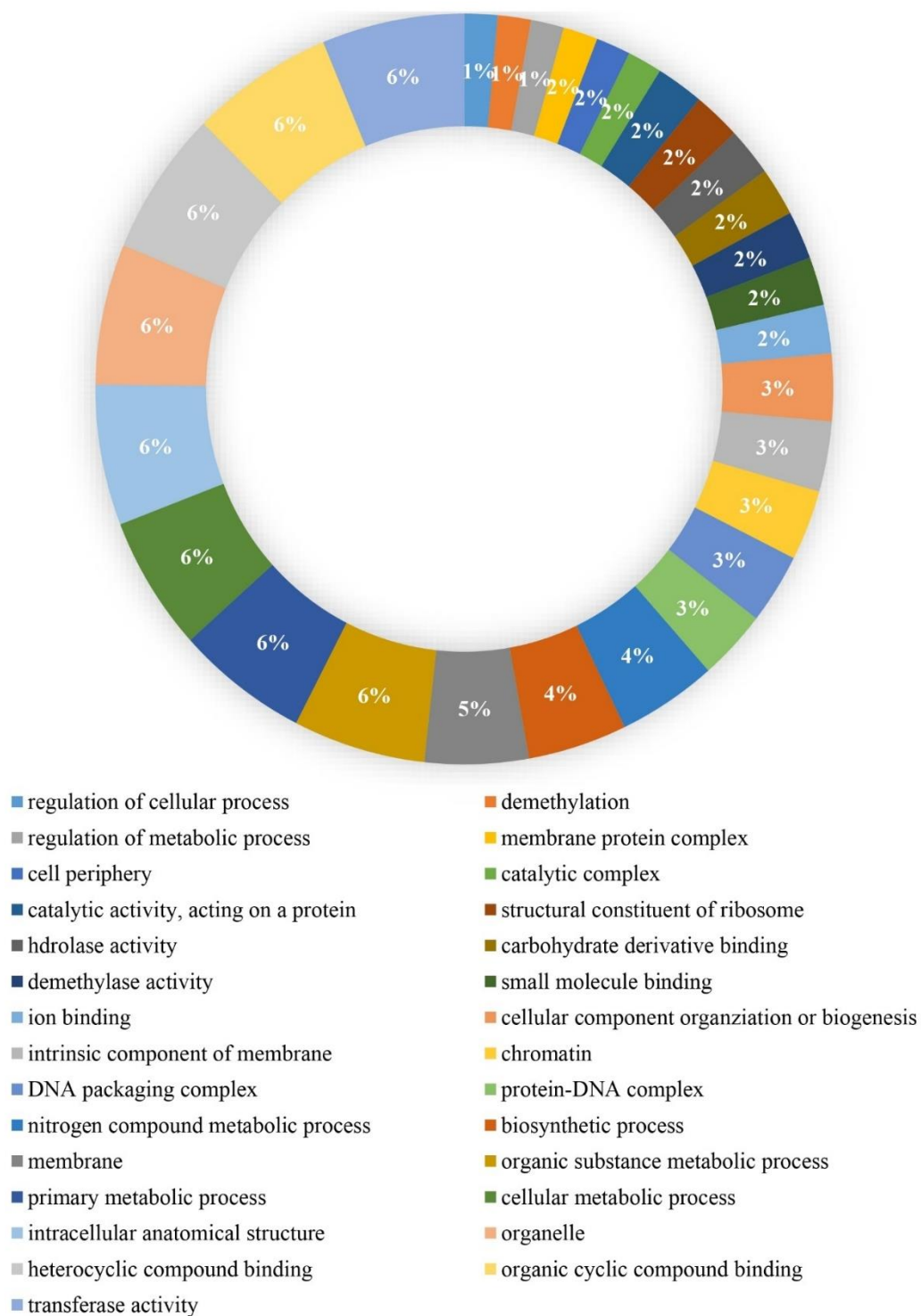


Figure 4.12 GO analyses results of targets of newly identified tcc-miR166b and tcc-miR166i

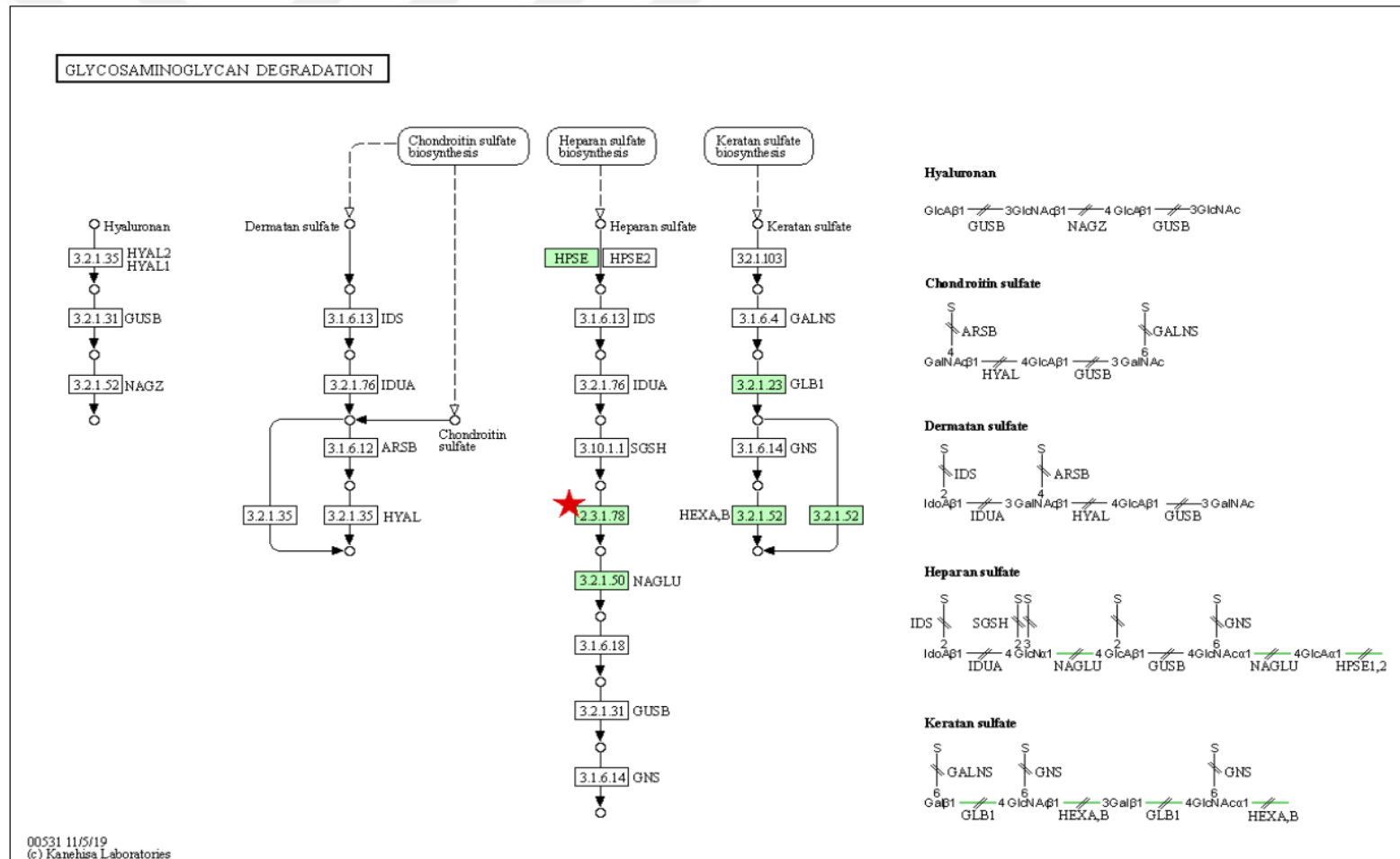


Figure 4.13 Glycosaminoglycan degradation

²The pathway to which tcc-miR166b is associated is indicated by a red star.

4.3.2 RNA Hybrid, GO, and KEGG Analysis of miR168 Target Genes

It was determined that there are 3 target genes of the tcc-miR168 sequence in *Theobroma cacao*. Similar to the tcc-miR166 sequence, it was observed that one miRNA sequence was associated with the regulation of more than one mRNA sequence (Table 4.4).

As a result of GO analyses, biological process (BP) consists of regulation of cellular process (5%), cellular response to stimulus (5%), signal transduction (5%), cell communication (5%), catabolic process (10%), nitrogen compound metabolic process (10%), organic substance metabolic process (20%), primary metabolic process (20%) and cellular metabolic process (20%). Within the Molecular Function (MF), catalytic activity, acting on a protein (25%), heterocyclic compound binding (25%), organic cyclic compound binding (25%), and hydrolase activity (25%) were observed. Membrane-enclosed lumen (7,69%), nuclear body (7,69%), nucleoplasm (7,69%), cytoplasm (15,39%), intracellular anatomical structure (30,77%) and organelle (30,77%) were also detected in Cellular Component (CC) (Figure 4.14).

After GO, KEGG pathway searches were performed in the DAVID database to identify biological pathways of genes. The results showed that tcc-miR168 was involved in the spliceosome mechanism (Figure 4. 15).

Table 4.4 The targets to

miRNA	GENBANK ACCESSION	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc-miR168	XM_007032323.2	Pentatri-copeptide repeat-containing protein Atlg71460, chloroplastic (LOC18601392)	GATCCCGCCTTGCATCAACTGAAT	target 5' G GAUAUGUAAU GGGAGGCUUAA U 3' AUGUCAG CAAAU CCAGGCCUU UACAGUU GUUUA GGUCUGAA miRNA 3' C 5'	-21.0 kcal/mol
tcc-miR168	XM_018129310.1 XM_007010202.2	Neutral/alkaline invertase 3, chloroplastic (LOC18586692)	GATCCCGCCTTGCATCAACTGAAT	target 5' G AU ACAGAGGAGGUAAUUG U U 3' AUG UCAGCA GAUCC GACUU UAC AGUUGU UUAGG CUGAA miRNA 3' U C 5'	-21.6 kcal/mol
tcc-miR168	XM_007038526.2	Serine/arginine-rich splicing factor RS40 (LOC18605496)	GATCCCGCCTTGCATCAACTGAAT	target 5' A GA UG C 3' GUG UA AAGUCUGGAUUUG UAC GU UUUAGGUCUGAAC miRNA 3' A UG 5'	-21.5 kcal/mol

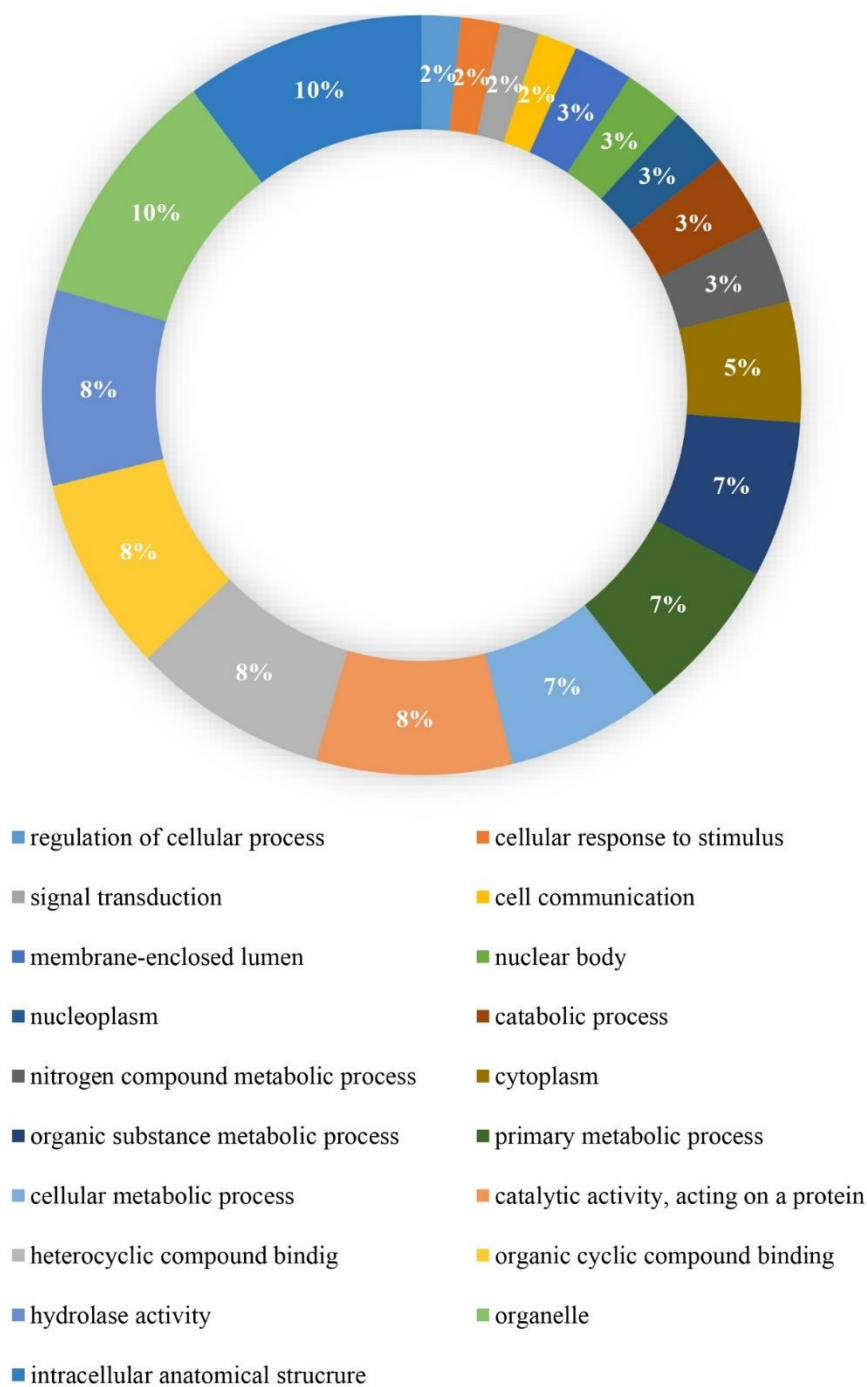


Figure 4.14 GO analyses results of targets of newly identified tcc-miR168

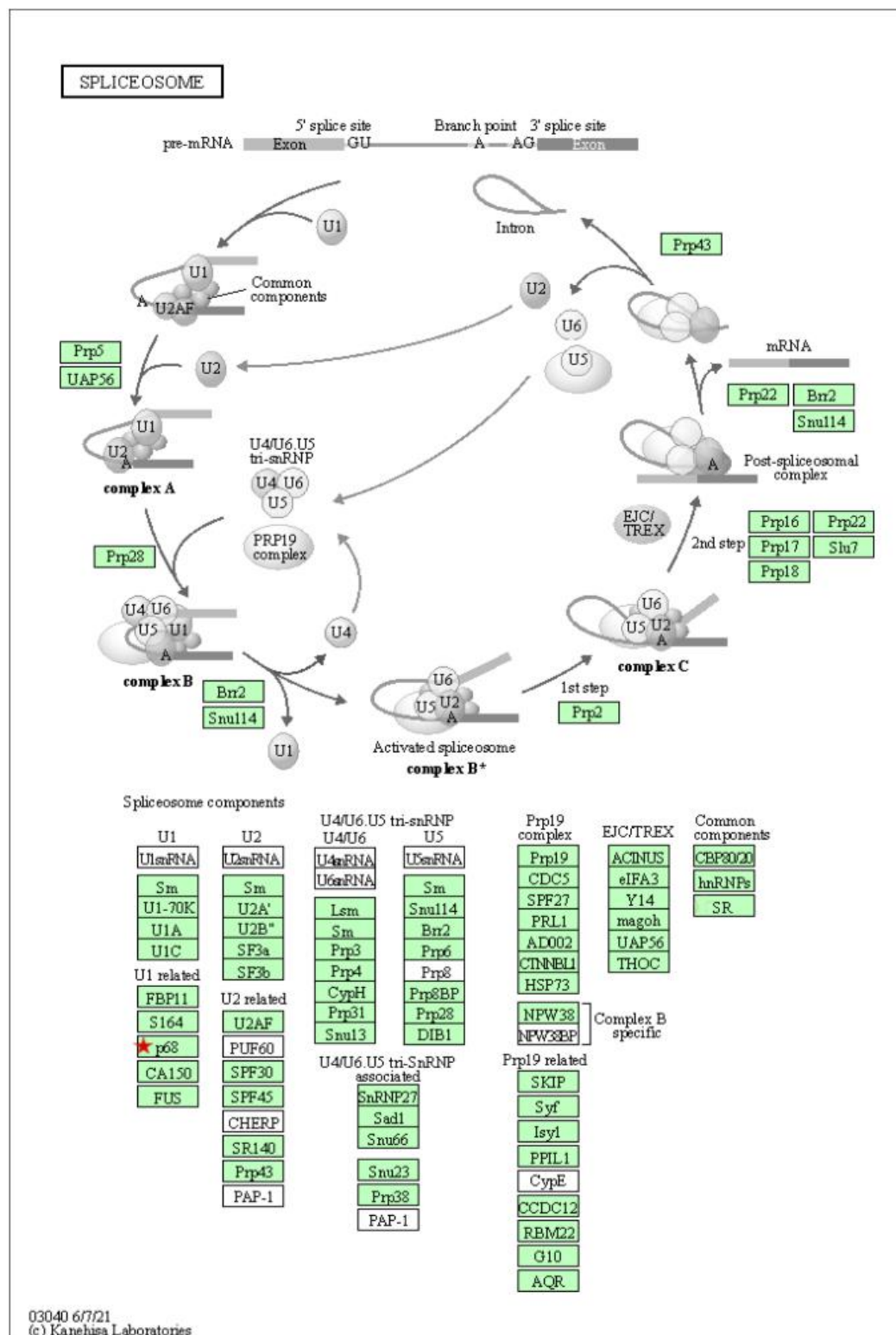


Figure 4.15 Spliceosome mechanism

³The protein to which tcc-miR168 is associated is indicated by a red star.

4.3.3 RNA Hybrid, GO, and KEGG Analysis of miR169 Target Genes

There are 3 target genes of tcc-miR169. Among them, 1 of which was uncharacterized (Table 4.5). GO analyses showed that the Biological process (BP) consists of, anatomical structure morphogenesis (3,22%), response to endogenous stimulus (3,22%), cellular response to stimulus (3,22%), anatomical structure development (3,22%), response to chemical (3,22%), regulation of cellular process (3,23%), regulation of metabolic process (3,23%), biosynthetic process (3,23%), signal transduction (3,23%), cellular development process (3,23%), cellular localization (6,45%), cell communication (6,45%), establishment of localization (6,45%), cellular component organization or biogenesis (6,45%), macromolecule localization (6,45%), nitrogen compound metabolic process (9,68%), organic substance metabolic process (9,68%), primary metabolic process (9,68%) and cellular metabolic process (9,68%). Within the Molecular Function (MF), ion binding (10%), DNA-binding transcription factor activity (10%), heterocyclic compound binding (40%), and organic cyclic compound binding (40%) were observed. An intrinsic component of membrane (5,88%), cytoplasm (5,88%), membrane (5,88%), membrane-enclosed lumen (11,76%), nuclear-protein-containing complex (11,77%), ribonucleoprotein complex (11,77%), intracellular anatomical structure (23,53%) and organelle (23,53%) were also detected in Cellular Component (CC) (Figure 4.16). KEGG results indicated that tcc-miR169c was involved in ribosome biogenesis in eukaryotes (Figure 4.17)

Table 4.5 The targets tcc-miR169

miRNA	GENBANK ACCESSION	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc-miR169	XM_018114390.1 XM_018114391.1	Nucleolar protein 6 (LOC18610924)	AAGCCAAGGATGA CTTGCCGA	target 5' U CCAG A C 3' UUUGG UUGUU GGCUUGC AAAA AAUCC AACAA CCGAACGUUUU miRNA 3' U A 5'	-30.0 kcal/mol
tcc-miR169	XM_007044434.2	Zinc finger protein 5 (LOC18609363)	AAGCCAAGGATGA CTTGCCGA	target 5' C A UGAGCUUAACCCGA A 3' UUAG UUGUUUGGCUU GUAGA AAUC AACAAACCGAA CGUUUU miRNA 3' U C 5'	-24.5 kcal/mol
tcc-miR169	XR_001926460.1	Uncharacterized (LOC18609804)	AAGCCAAGGATGA CTTGCCGA	target 5' U GAGAGGAAAAAUG U 3' AUUGGG UGUUUUGGCUUGCAAA UAAUCC ACAAAACCGAACGUUU miRNA 3' A U 5'	-32.3 kcal/mol

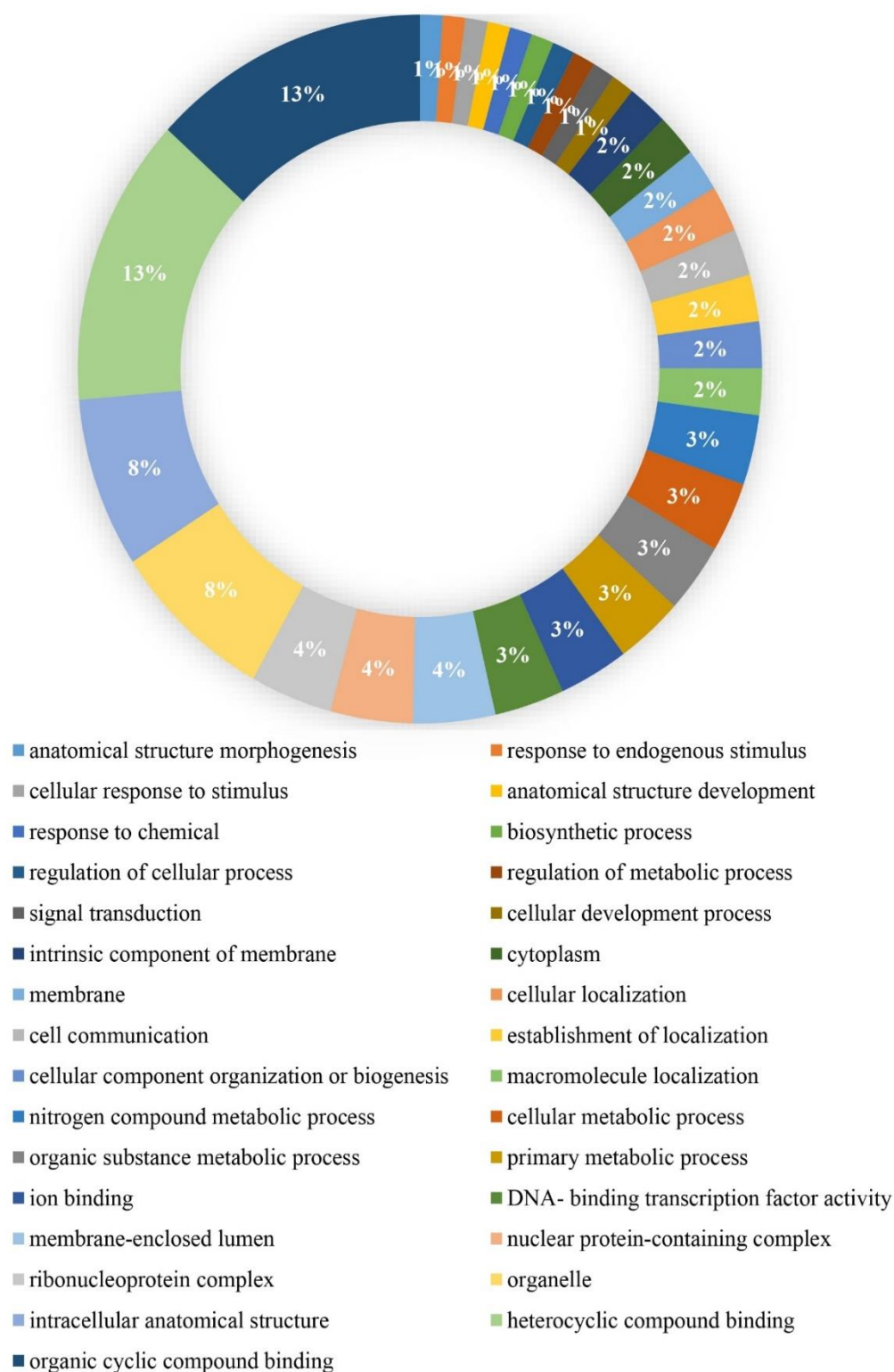


Figure 4.16 GO analyses results of targets of newly identified tcc-miR169

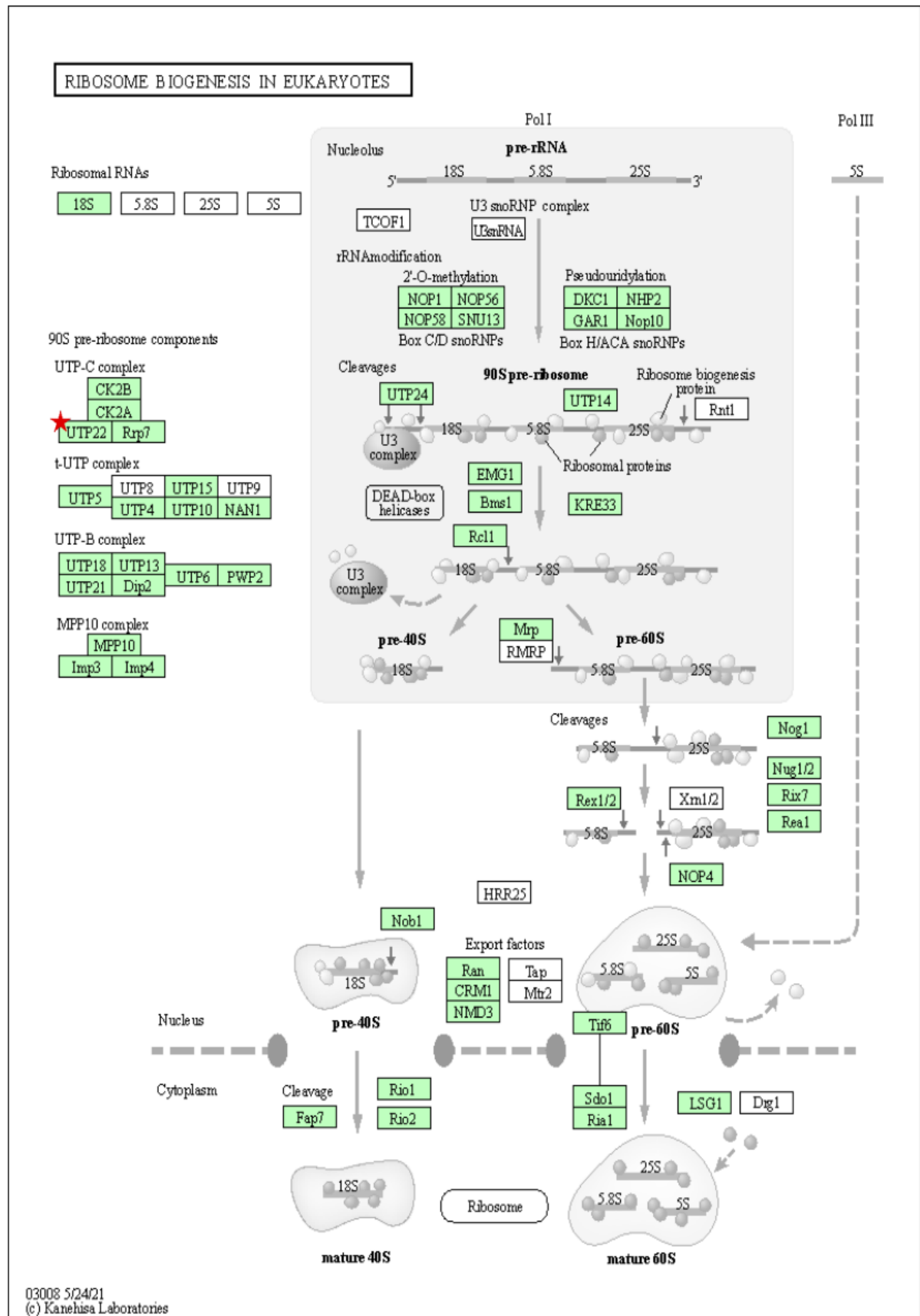


Figure 4.17 Ribosome biogenesis in eukaryotes

⁴The protein to which tcc-miR169 is associated is indicated by a red star.

4.3.4. RNA Hybrid, GO, and KEGG Analysis of miR1051 Target Genes

Only 1 target gene of the tcc-miR 1051 was identified in *Theobroma cacao* was detected (Table 4.6). Go analyses indicate that Biological process (BP) consist of regulation of cellular process (8,33%), cellular response to stimulus (8,33%), response to other organism (8,33%), primary metabolic process (8,33%), organic substance metabolic process (8,33%), response to external stimulus (8,33%), cellular metabolic process (8,33%), nitrogen compound metabolic process (8,33%), response to biotic stimulus (8,34%), stress response (8,34%), signal transduction (8,34%) and cell communication (8,34%). Within the Molecular Function (MF), catalytic activity, acting on a protein (11,11%), signaling receptor activity (11,11%), carbohydrate derivative binding (11,11%), ion binding (11,11%), heterocyclic compound binding (11,11%), organic cyclic compound binding (11,11%), carbohydrate binding (11,11%), transferase activity (11,11%) and small molecule binding (11,12%) was observed. Cell periphery (16,66%), an intrinsic component of membrane (16,66%), membrane (16,67%), intracellular anatomical structure (16,67%), organelle (16,67%) and cytoplasm (16,67%) were also detected in Cellular Component (CC) (Figure 4.18). However, it was observed that the target gene was not associated with any biological pathway as a result of KEGG analysis.

Table 4.6 The targets for

miRNA	GENBANK ACCESSION	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc-miR105l	XM_007020426.2	Probable L-type lectin-domain containing receptor kinase S.7 (LOC18593279)	TTCCTGTCTACTTGA GTGAAG	target 5' U AUCA C 3' UUCACUCAAGUAGA UAGGA AAGUGAGUUCAUCU GUCCU miRNA 3' G U 5'	-32.1 kcal/mol

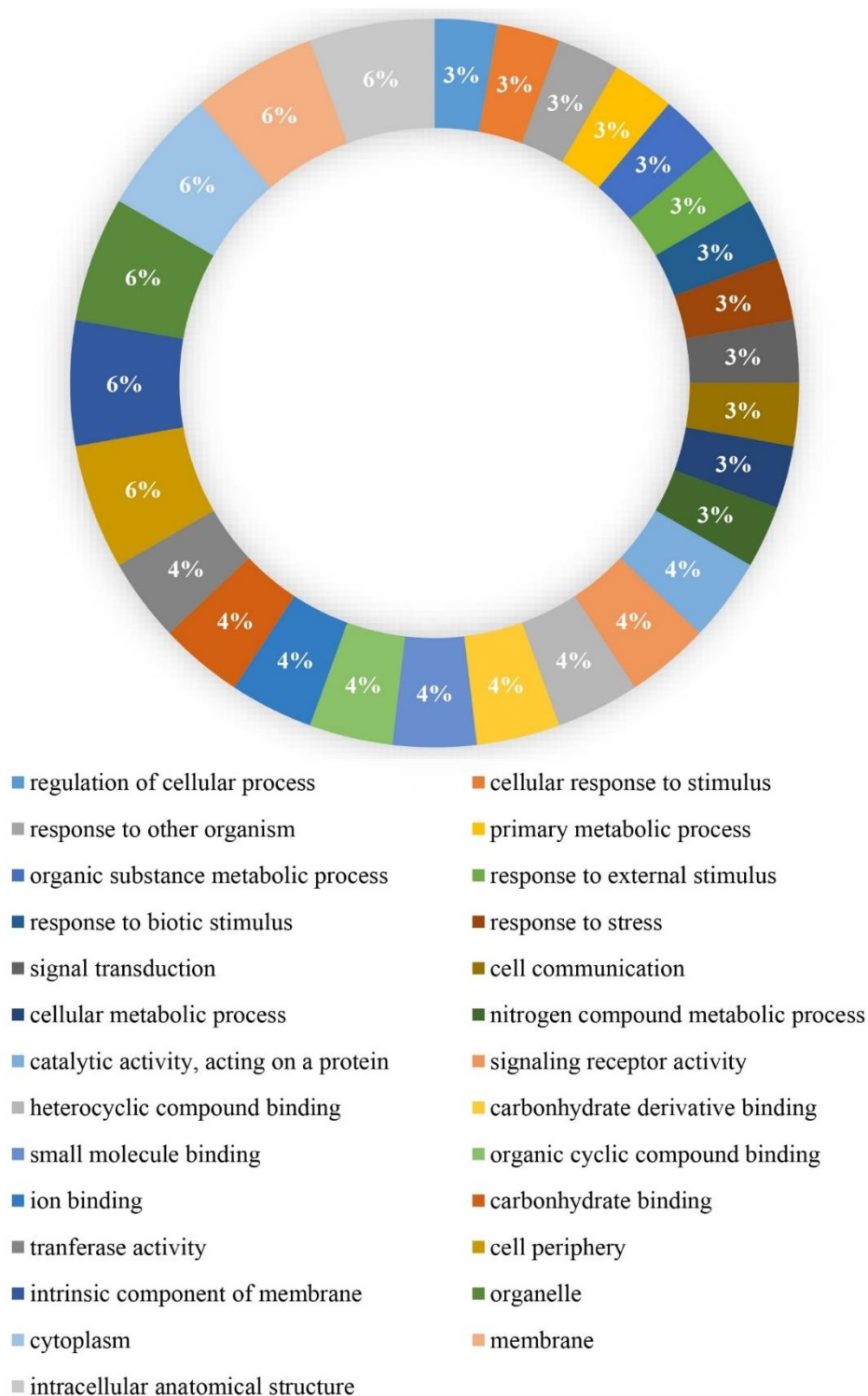


Figure 4.18 GO analyses results of targets of newly identified tcc-miR1051

4.3.5 RNA Hybrid, GO, and KEGG Analysis of miR3476 Target Genes

As a result of RNA Hybrid analyses 8 target genes of the tcc-miR3476 sequence were identified (Table 4.7). When GO results were evaluated, it was observed that, Biological process (BP) consist of , anatomical structure morphogenesis (3,57%), anatomical structure development (3,57%), response to external stimulus (3,57%), response to chemical (3,57%), small molecule metabolic process (3,57%), cellular localization (3,57%), response to stress (3,57%), regulation of multicellular organismal process (3,57%), multicellular organismal development (3,57%), regulation of cellular process (3,57%), post-embryonic development (3,57%), response to other organism (3,57%), response to biotic stimulus (3,57%), macromolecule localization (3,57%), cell wall organization or biogenesis (3,57%), anatomical structure formation involved in morphogenesis (3,57%), regulation of development process (3,57%), nitrogen compound metabolic process (5,36%), cellular component organization or biogenesis (5,36%), cellular metabolic process (7,15%), organic substance metabolic process (10,72%) and primary metabolic process (10,72%). Within the Molecular Function (MF), isomerase activity (4%), small molecule binding (4%), ATP-dependent activity, acting on DNA (4%), carbohydrate derivative binding (4%), chromatin binding (8%), ion binding (12%), catalytic activity, acting on a nucleic acid (16%), hydrolase activity (16%), heterocyclic compound binding (16%) and organic cyclic compound binding (16%) was observed. Membrane protein complex (3,45%), phagophore (6,49%), an intrinsic component of membrane (10,34%), cytoplasm (13,79%), membrane (13,79%), organelle (24,14%) and intracellular anatomical structure (27,59%) were also detected in Cellular Component (CC) (Figure 4.19). After GO analysis, KEGG pathway searches were performed in the DAVID database to identify biological pathways of genes. The results showed that tcc-miR3476 was involved in photosynthesis (Figure 4. 20).

Table 4.7 The targets of tcc-miR3476

miRNA	GENBANK ACCESSIO N	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc- miR3476	XR_001929 582.1 XR_001929 583.1	Uncharacterized (LOC18587984)	TGGTCCTGGATTGTGACAA	target 5' A GAGC CACAU A 3' GUUGGA UUG GGCUUGAGA UAGCUU AAC CCGAACUCU miRNA 3' C A 5'	-22.8 kcal/mol
tcc- miR3476	XM_018129 572.1	RING-H2 finger protein ATL74 (LOC18587473)	TGGTCCTGGATTGTGACAA	target 5' G AG U AA A 3' GA GGAUUUGG GCU AGA CU CUUAAACC CGA UCU miRNA 3' AG AC 5'	-21.5 kcal/mol
tcc- miR3476	XM_018118 347.1 XM_007039 662.2	Polycomb group protein EMBRYONIC FLOWER 2 (LOC18606194)	CAAGTCTGGATTGTGACAT	target 5' U G 3' GCAGAAUGGAUGGCAUGGUGUG CGUCUUACCUACCGUACCACAC miRNA 3' 5'	-48.2 kcal/mol
tcc- miR3476	XM_018121 588.1	DNA gyrase subunit B, chloroplastic/mitoch ondrial (LOC18598680)	CAAGTCTGGATTGTGACAT	target 5' A UGUCUCU U G 3' CAG GGUGGAU GCAUGGUGUG GUC UUACCUA CGUACCACAC miRNA 3' C C 5'	-32.1 kcal/mol

Table 4.7 The targets tcc-miR3476 (continued)

miRNA	GENBANK ACCESSION	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc-miR3476	XM_018122608.1 XM_007025347.2	Uncharacterized (LOC18596707)	CAAGTCTGGATTGTGACAT	target 5' U U AG A C 3' GCA GG GUGGAUG CAUGGUGUG CGU CU UACCUAC GUACCACAC miRNA 3' C 5'	-32.2 kcal/mol
tcc-miR3476	XM_018124776.1	Photosystem I reaction center subunit V, chloroplastic (LOC108662932)	TTGTCAACAAATCCAATAAGC	target 5' G A 3' UUGGUCAAAGAAAUGGC AAUGGG AAC CAGUUUCUUUUACCGUUA CCC miRNA 3' 5'	-46.0 kcal/mol
tcc-miR3476	XM_018122098.1	Uncharacterized (LOC18600366)	TTGTCAACAAATCCAATAAGC	target 5' A GU A 3' UGG UUAAGGAGAAUGGC AAUGG ACC AGUUUCUUUUACCGUUA ACC miRNA 3' A C 5'	-35.6 kcal/mol
tcc-miR3476	XM_007016349.2	Glucan endo- 1,3-beta-glucosidase (LOC18590685)	TTGTCAACAAATCCAATAAGC	target 5' G G CU CCGG CUUC U 3' UGGU G GAAA GUGGC GGUGGG ACCA U CUUU UACCG UUACCC miRNA 3' A G UU 5'	-26.6 kcal/mol

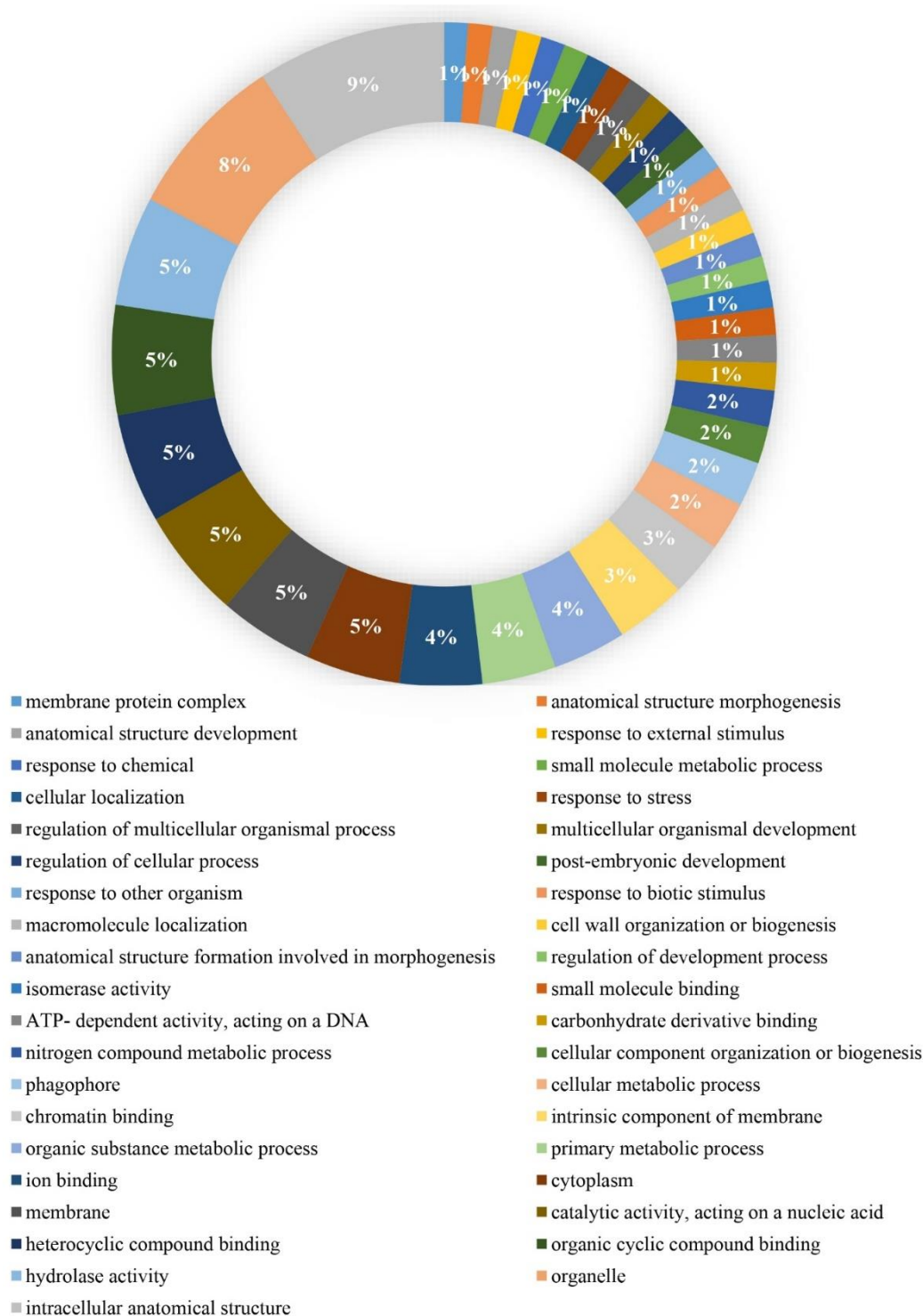


Figure 4.19 GO analyses results of targets of newly identified tcc-miR3476

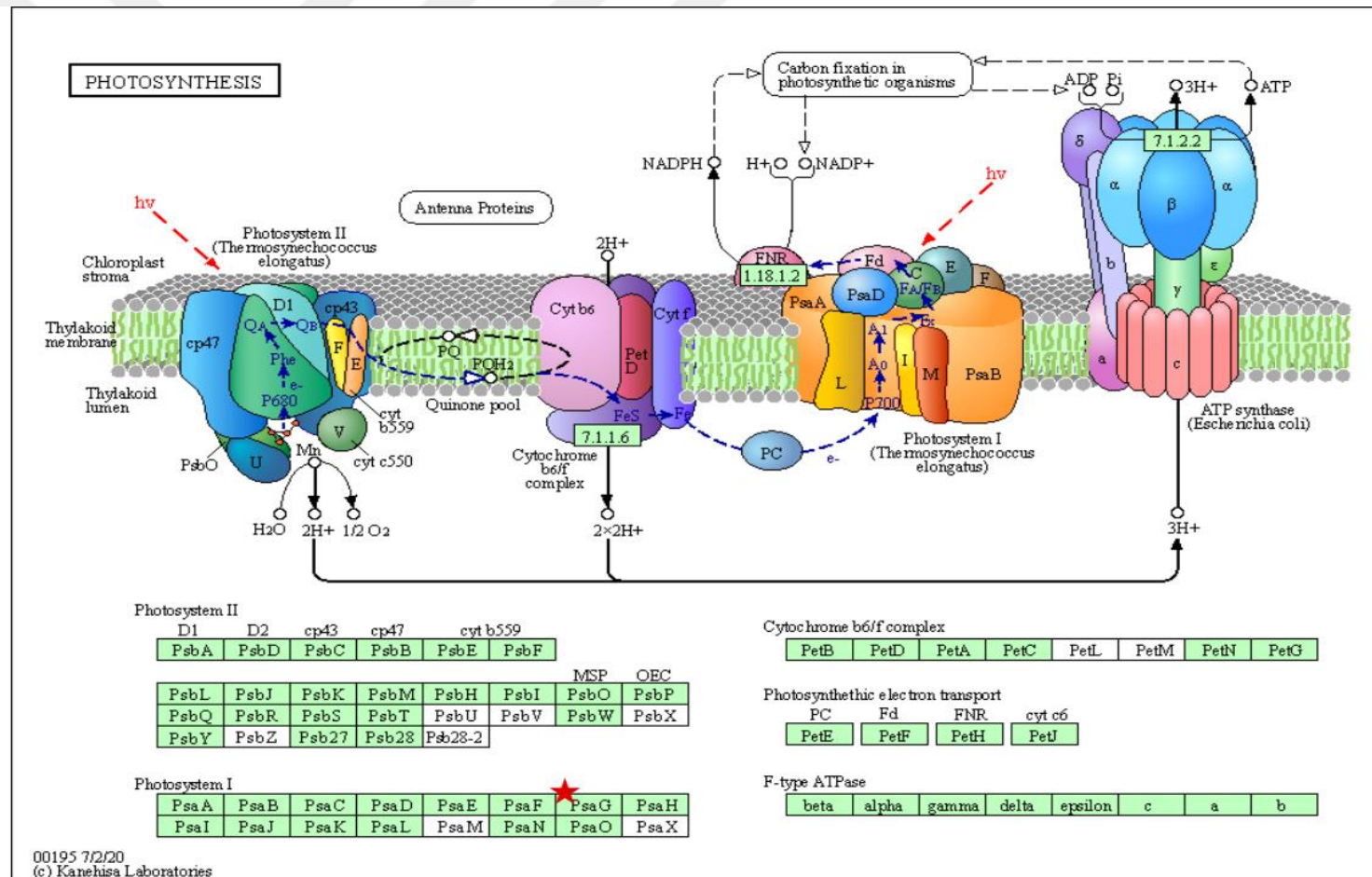


Figure 4.20 Photosynthesis

⁵The protein to which tcc-miR3476 is associated is indicated by a red star.

4.3.6 RNA Hybrid, GO, and KEGG Analysis of miR4249 Target Genes

Similar to tcc-miR1051, tcc-miR4249 has only 1 target gene in *Theobroma cacao* genome (Table 4.8). In GO analysis, the Biological process (BP) consists of, cell division (50%) and cell cycle (50%). Within the Molecular Function (MF), lipid binding (100%) was observed. Intrinsic components of the membrane (25%), intracellular anatomical structure (25%), cytoplasm (25%), and membrane (25%) were also detected in the Cellular Component (CC) (Figure 4.21). With these limited profiles, it was also determined that the target gene was not associated with any biological pathway in KEGG analysis.



Table 4.8 The targets tcc-miR4249

miRNA	GENBANK ACCESSION	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc- miR4249	XM_007026493.2	Patellin-6 (LOC18597447)	TATCTCAACTTCTCAAATCCT	target 5' C AAA GCAGACAAU C G A 3' GG GAGUUCUCC GUUG G AGG CC CUCAAGAGG UAAC C UCC miRNA 3' C A G 5'	-26.5 kcal/mol

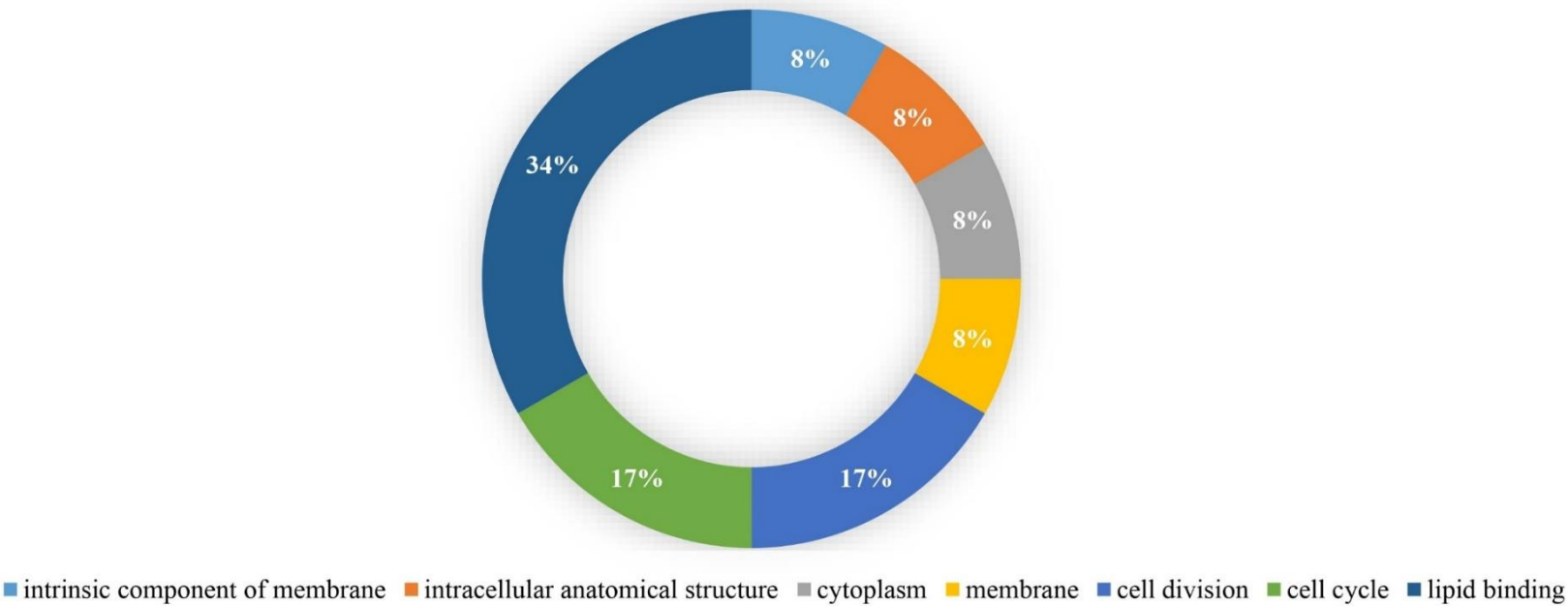


Figure 4.21 GO analyses results of targets of newly identified tcc-miR4249

4.3.7 RNA Hybrid, GO, and KEGG Analysis of miR6161 Target Genes

RNA Hybrid results indicated that 5 target genes (3 of which were uncharacterized) of tcc-miR6161 were detected in the cacao genome (Table 4.9). In GO analysis of target genes, miRNA target genes; It was determined that it was divided into 2 categories: biological process and molecular function. No data related to the Cellular component was found in the GO analysis. It consists of the Biological process (BP), multi-multicellular organism process (10%), pollen-pistil interaction (10%), cell recognition (10%), primary metabolic process (10%), organic substance metabolic process (10%), cell communication (10%), pollination (10%), cellular metabolic process (10%), recognition of pollen (10%) and nitrogen compound metabolic process (10%). Within the Molecular Function (MF), catalytic activity, acting on a protein (14,28%), heterocyclic compound binding (14,28%), carbohydrate derivative binding (14,28%), small molecule binding (14,29%), organic cyclic compound binding (14,29%), ion binding (14,29%) and transferase activity (14,29%) was observed (Figure 4.22). Similar to tcc-miR4249, the target genes of tcc-miR6161 were not associated with any biological pathway.

Table 4.9 The targets to

miRNA	GENBANK ACCESSION	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc-miR616l	XR_001927715.1 XR_001927716.1 XR_001927718.1	Uncharacterized (LOC18603070)	GCGACTCCAGCATATTATAAT	target 5' U A C 3' CAGAUCAU UGGCAGUUUCA GUCUAGUA ACCGUCAAAAGU miRNA 3' C A 5'	-34.0 kcal/mol
tcc-miR616l	XM_007013132.2 XM_018128505.1	G-type lectin S-receptor-like serine/threonine- protein kinase CES101 (LOC18588610)	GCGACTCCAGCATATTATAAT	target 5' G UA AGAAGAACACAGU U 3' AGAUUGU GUGGCA GUUUCAU UCUAGUA CACCGU CAAAGUA miRNA 3' G 5'	-23.9 kcal/mol

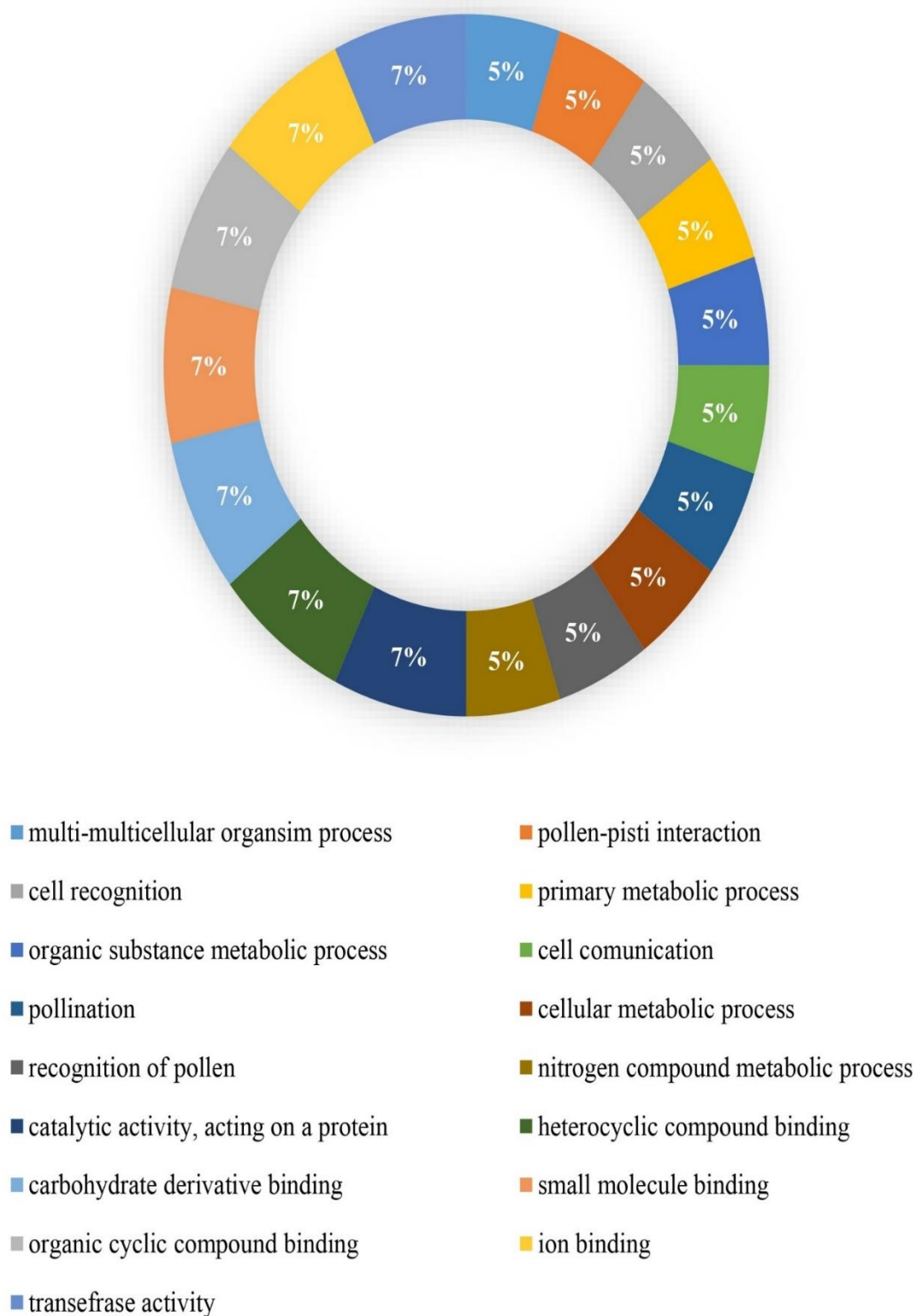


Figure 4.22 GO analyses results of targets of newly identified tcc-miR6161

4.3.8 RNA Hybrid, GO, and KEGG Analysis of miR6279 Target Genes

Four target genes of tcc-miR6279 were sequence identified in *Theobroma cacao*. (Table 4.10). It was reported that these genes regulate different metabolic pathways. When they were categorized into groups, Biological process (BP) consists of regulation of cellular process (4,35%), demethylation (4,35%), regulation of metabolic process (4,35%), cellular component organization or biogenesis (8,7%), nitrogen compound metabolic process (13,04%), biosynthetic process (13,04%), organic substance metabolic process (17,39%), primary metabolic process (17,39%) and cellular metabolic process (17,39%). Within the Molecular Function (MF), catalytic activity, acting on a protein (6,25%), structural constituent of ribosome (6,25%), hydrolase activity (6,25%), carbohydrate derivative binding (6,25%), demethylase activity (6,25%), small molecule binding (6,25%), ion binding (6,25%), heterocyclic compound binding (18,75%), organic cyclic compound binding (18,75%) and transferase activity (18,75%) was observed. Membrane protein complex (4,54%), cell periphery (4,55%), catalytic complex (4,55%), intrinsic component of membrane (9,09%), chromatin (9,09%), DNA packaging complex (9,09%), protein-DNA complex (9,09%), membrane (13,64%), intracellular anatomical structure (18,18%) and organelle (18,18%) were also detected in Cellular Component (CC) (Figure 4.23). Similar to tcc-miR4249 and tcc-miR6161, target genes of tcc-miR6279, were not associated with any biological pathway.

Table 4.10 The targets to

miRNA	GENBANK ACCESSION	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc- miR6279	XM_018116764.1	Growth-regulating factor 10 (LOC18606399)	CTAGCATGGAATTCTTGTCTA	target 5' C AUU AGG UU A 3' UCCA GGGCU UCUUGAGC GG AGGU UCCGA AGAACUUG CC miRNA 3' G A 5'	-28.2 kcal/mol
tcc- miR6279	XM_018127495.1	Callose synthase 3 (LOC18589684)	CTAGCATGGAATTCTTGTCTA	target 5' G C U C 3' CUGCAGGCUUUC UUGA AUGG GGUGUCCGAAAG AACU UGCC miRNA 3' A 5'	-30.1 kcal/mol
tcc- miR6279	XM_007031567.2	GDSL esterase/lipase At5g14450 (LOC18600859)	CTAGCATGGAATTCTTGTCTA	target 5' C UC CAAGC C 3' UCC CAGGCUUUCUUG AUGG AGG GUCCGAAAGAACU UGCC miRNA 3' U 5'	-31.7 kcal/mol
tcc- miR6279	XM_007042601.2 XM_007042602.2	Telomere repeat-binding factor 1 (LOC18608101)	CTAGCATGGAATTCTTGTCTA	target 5' G CU U 3' UUUACAGGCUUUCUUG GAUGG AGGUGUCCGAAAGAAC UUGCC miRNA 3' 5'	-32.7 kcal/mol

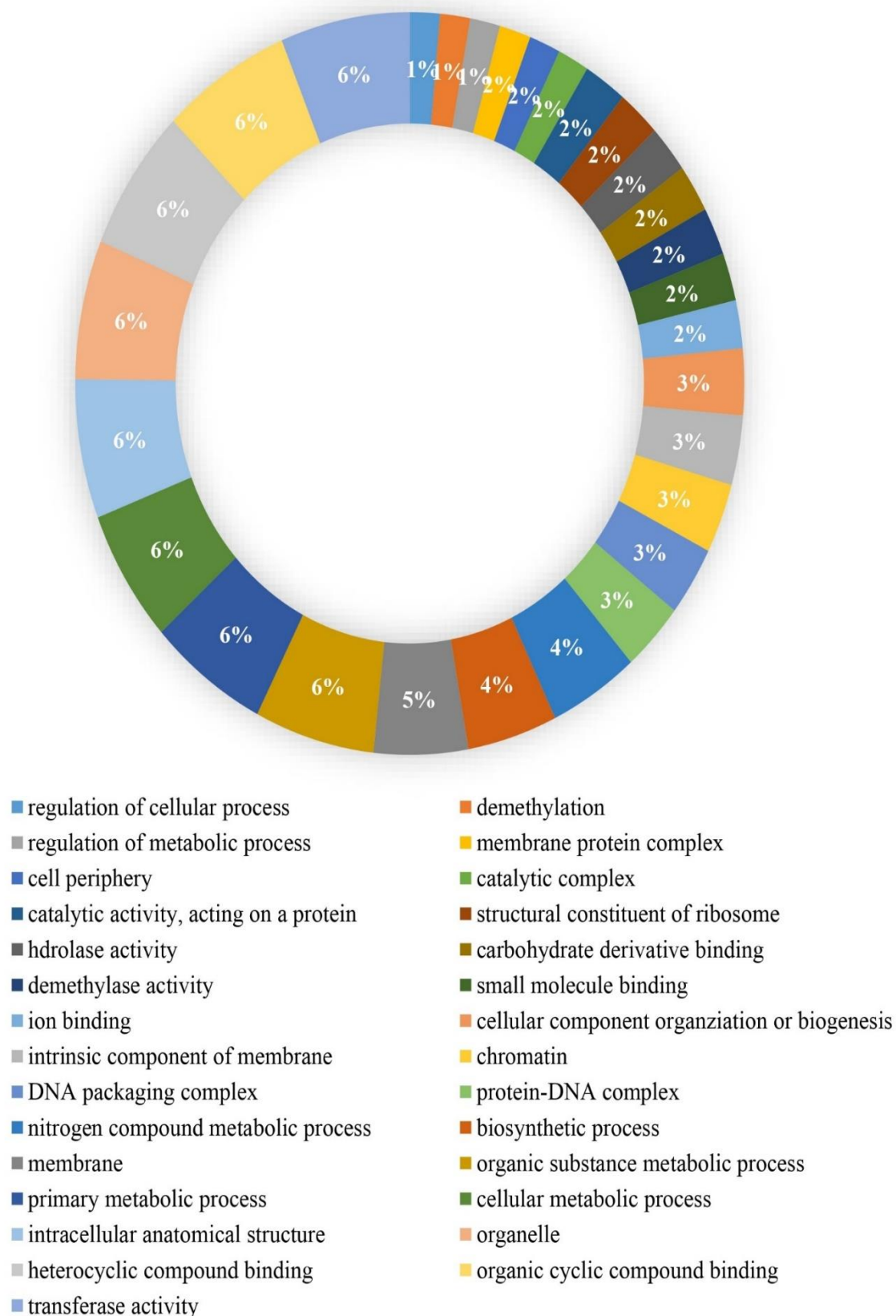


Figure 4.23 GO analyses results of targets of newly identified tcc-miR6279

4.3.9 RNA Hybrid, GO, and KEGG Analysis of miR8150 Target Genes

Six target genes (2 of which were uncharacterized) of tcc-miR8150 were identified in the cacao genome (Table 4.11). GO analyses indicated that Biological process (BP) consist of negative regulation of metabolic process (11,11%), organic substance metabolic process (11,11%), primary metabolic process (11,11%), nitrogen compound metabolic process (11,11%), catabolic process (11,11%), regulation of metabolic process (11,11%), cellular metabolic process (11,11%) and cellular component organization or biogenesis (22,23%). Within the Molecular Function (MF), oxidoreductase activity (2,94%), ATP-dependent activity, acting on DNA (5,88%), small molecule binding (5,88%), ATP-dependent activity, acting on RNA (5,88%), carbohydrate derivative binding (5,88%), ion binding (8,82%), catalytic activity, acting on a nucleic acid (11,76%), helicase activity (11,77%), hydrolase activity (11,77%), heterocyclic compound binding (14,71%) and organic cyclic compound binding (14,71%) was observed. Cytoplasm (5,88%), an intrinsic component of membrane (11,76%), nuclear-protein-containing complex (11,76%), membrane (11,77%), organelle (17,65%), ribonucleoprotein complex (17,65%) and intracellular anatomical structure (23,53%) were also detected in Cellular Component (CC) (Figure 4.24). After GO analysis, KEGG pathway searches were performed in the DAVID database to identify biological pathways of genes. The results showed that tcc-miR8150 was involved in the mRNA surveillance pathway (Figure 4.25), nucleocytoplasmic transport (Figure 4.26), and spliceosome mechanism (Figure 4.27).

77

miRNA	GENBANK ACCESSION	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc-miR8150	XM_018116910.1	11-beta-hydroxysteroid dehydrogenase-like 4A (LOC18605142)	CGCTCATCATCTGACTCAGGTTTT	target 5' U ACGGUGAA CUUCGGGCAGUU U 3' UUGUUGACAAG CCA GGACCA AACAGUUGUUU GGU CCUGGU miRNA 3' A 5'	-22.5 kcal/mol
tcc-miR8150	XM_007039543.2	Regulator of nonsense transcripts 1 homolog (LOC18606113)	CGCTCATCATCTGACTCAGGTTTT	target 5' U UAUGCCA A A 3' UUGUUAACAA GUCC AGGGCCA AACAGUUGUU UAGG UCCUGGU miRNA 3' 5'	-30.2 kcal/mol
tcc-miR8150	XM_007029731.2 XM_018121810.1	DNA-binding protein SMUBP-2 (LOC18599669)	CGCTCATCATCTGACTCAGGTTTT	target 5' C C AA G 3' UGUUAGCA AUC CAGGAUUA ACAGUUGU UAG GUCCUGGU miRNA 3' A U 5'	-25.3 kcal/mol
tcc-miR8150	XM_007010967.2	Uncharacterized (LOC18587260)	CGCTCATCATCTGACTCAGGTTTT	target 5' A GUUCAGCUGU CUUUCUCAC G 3' UGUCA GCAG AUUCAGGGCCA ACAGU UGUU UAGGUCCUGGU miRNA 3' A 5'	-27.0 kcal/mol
tcc-miR8150	XM_007031450.2	DEAD-box ATP-dependent RNA helicase 46 (LOC18600784)	CGCTCATCATCTGACTCAGGTTTT	target 5' A GA CAGGCUGACA U 3' UG GACAAAUUCU GGGACUA AC UUGUUUAGG UCCUGGU miRNA 3' A AG 5'	-23.2 kcal/mol
tcc-miR8150	XM_018123605.1	Uncharacterized (LOC18596072)	CGCTCATCATCTGACTCAGGTTTT	target 5' A UCA AGAA C C 3' UUGUCAAU GAAU GGG GCCA AACAGUUG UUUAG UCC UGGU miRNA 3' G 5'	-23.3 kcal/mol

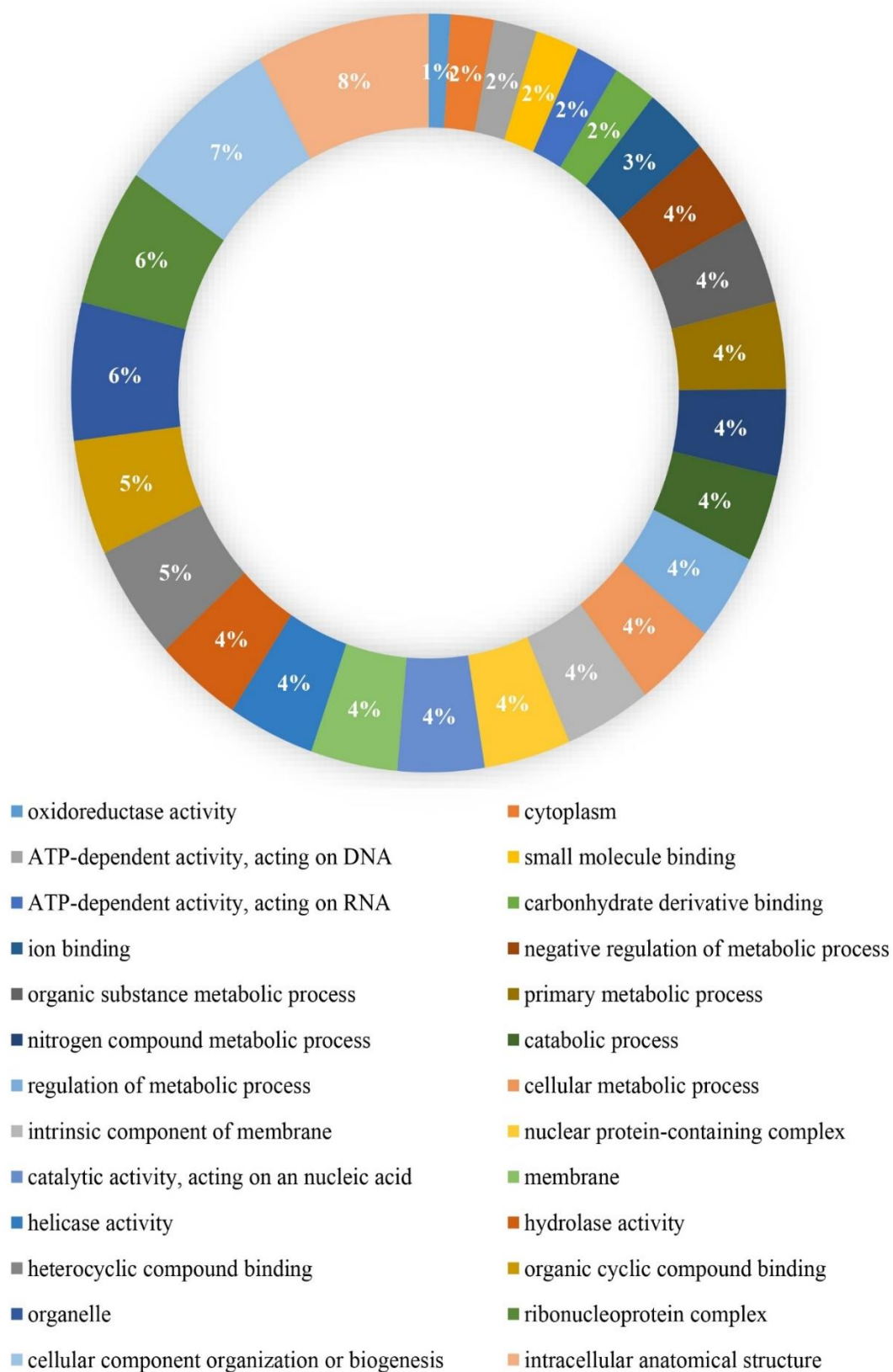


Figure 4.24 GO analyses results of targets of newly identified tcc-miR8150

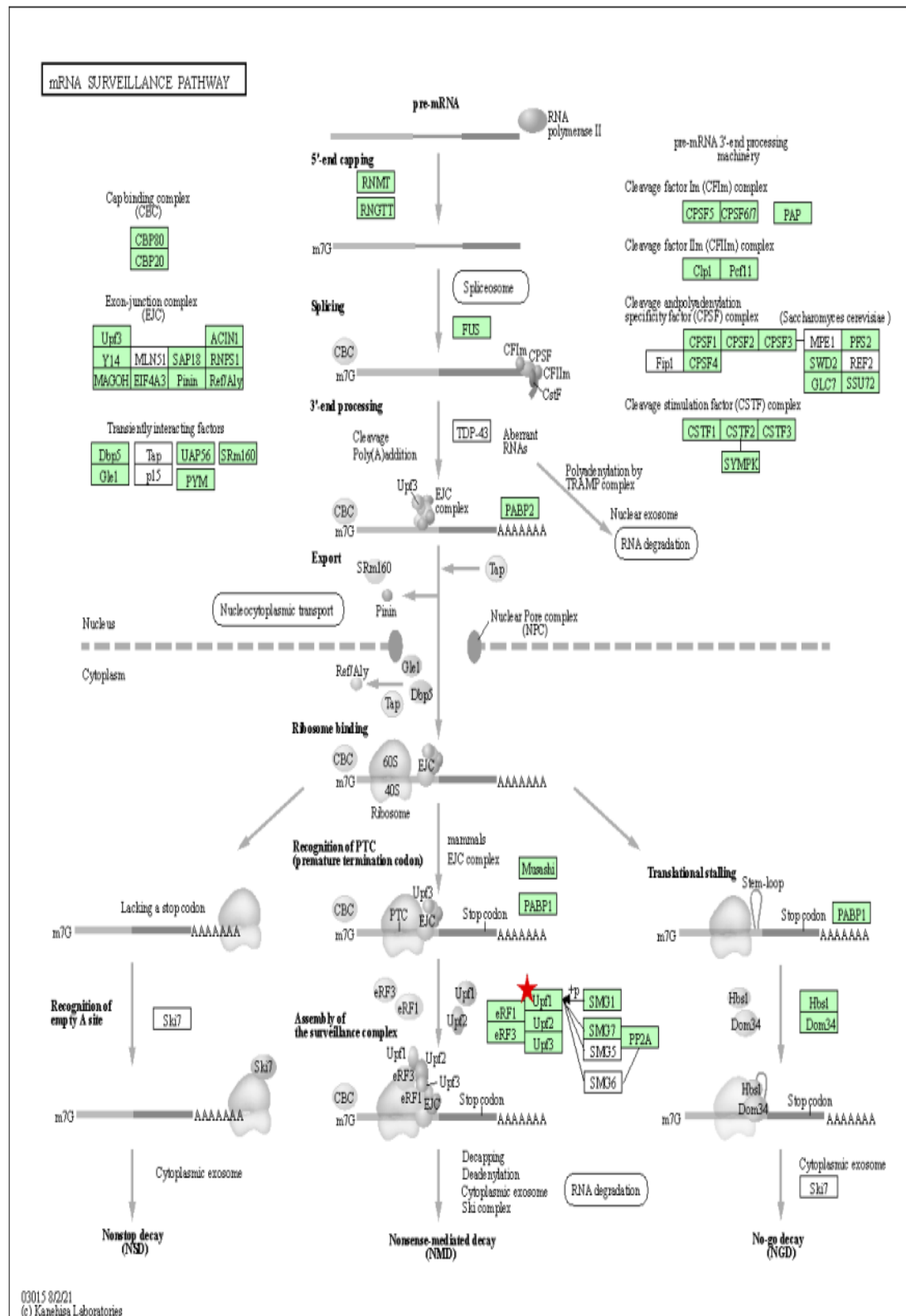


Figure 4.25 mRNA surveillance pathway

⁶The protein to which tcc-miR8150 is associated is indicated by a red star.

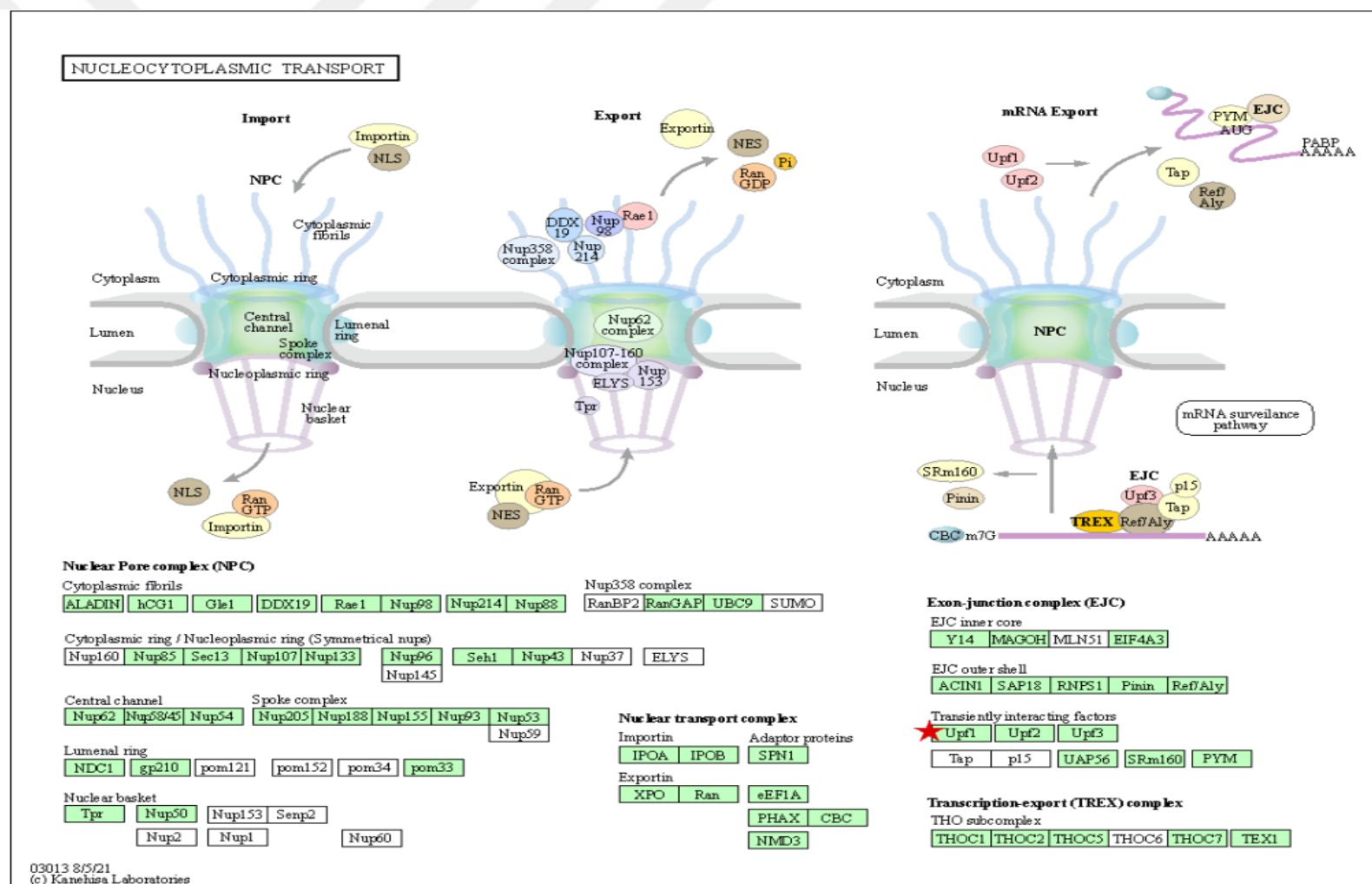


Figure 4.26 Nucleocytoplasmic transport

⁷The protein to which tcc-miR8150 is associated is indicated by a red star.

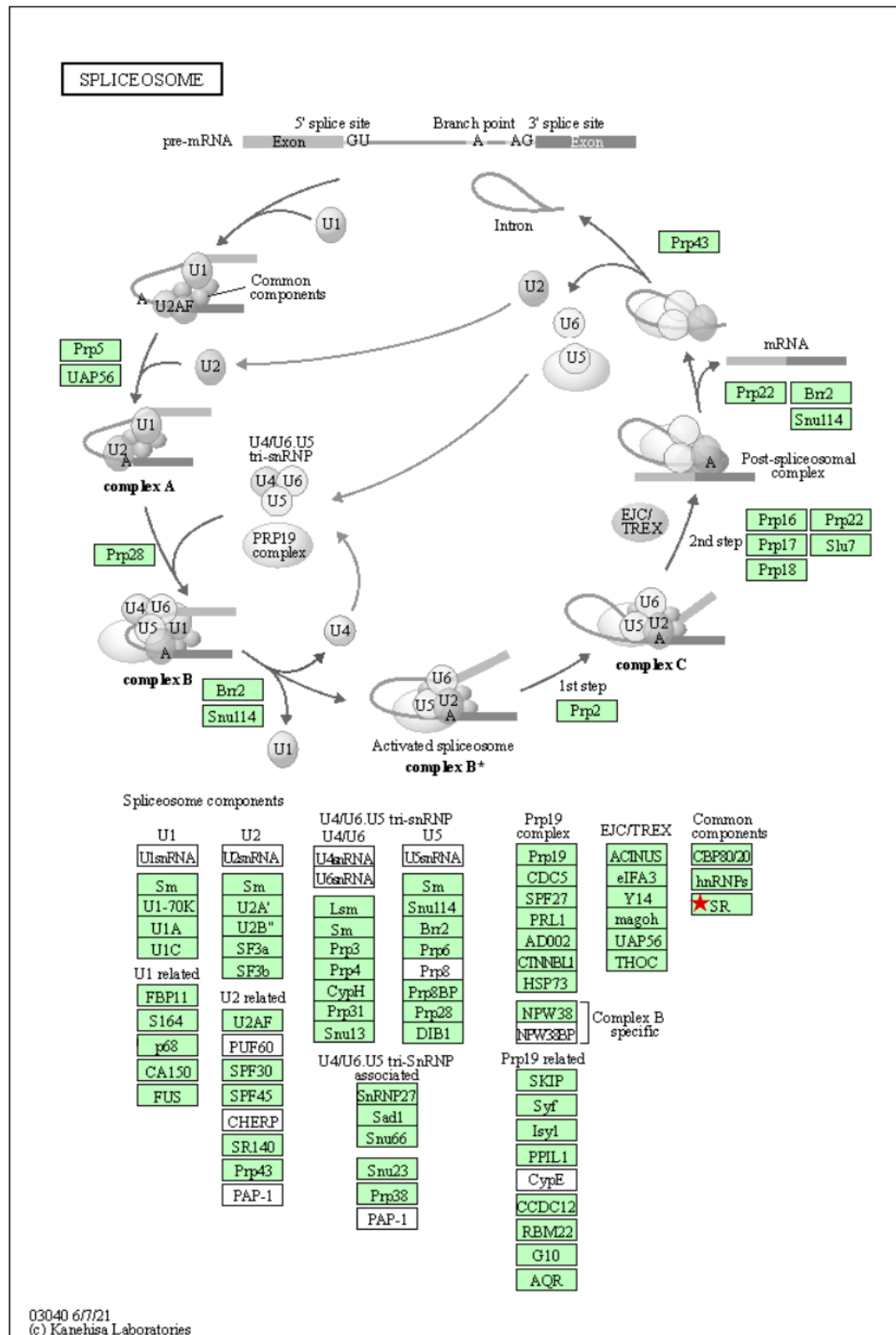


Figure 4.27 Spliceosome mechanism

⁸The protein to which tcc-miR8150 is associated is indicated by a red star.

4.3.10 RNA Hybrid, GO, and KEGG Analysis of miR9471 Target Genes

Seven targets of tcc-miR9471 were detected in the *Theobroma cacao* genome (Table 4.12). In GO analysis of target genes, miRNA target genes; It was determined that it was divided into 3 categories: biological process, cellular component, and molecular function. It consists of the Biological process (BP), cellular response to stimulus (3,33%), response to chemical (3,33%), small molecule metabolic process (3,33%), stress response (3,33%), cell communication (3,33%), pigment metabolic process (3,33%), regulation of signaling (3,33%), regulation of response to stimulus (3,34%), cell death (3,34%), signal transduction (3,34%), regulation of metabolic process (6,67%), nitrogen compound metabolic process (10%), biosynthetic process (10%), regulation of cellular process (10%), cellular metabolic process (10%), primary metabolic process (10%) and organic substance metabolic process (10%). Within the Molecular Function (MF), protein binding (10%), ligase activity (10%), heterocyclic compound binding (10%), carbohydrate derivative binding (10%), small molecule binding (10%), organic cyclic compound binding (10%), transcription coregulator activity (20%) and ion binding (20%) was observed. Nuclear protein-containing complex (33,33%), organelle (33,33%), and intracellular anatomical structure (33,34%) were also detected in the Cellular Component (CC) (Figure 4.28). It was also found that these genes play important roles in purine metabolism (Figure 4.29) and ubiquitin-mediated proteolysis (Figure 4.30).

Table 4.12 The targets to

Table 4.12 The targets to

Table 4.12 The targets tcc-miR947 (continued)

miRNA	GENBANK ACCESSION	FUNCTION	MISMATCH	RNA HYBRID RESULT	MFE VALUE
tcc-miR9471	XM_018119858.1	MACPF domain- containing protein CAD1 (LOC18602231)	CAGTGATGCTCACTCAGCAA	target 5' U CG GGUUUCCGCUU C 3' GUC GU GGCAGUUUCAU UAG CA CCGUCAAGUA miRNA 3' C UA A 5'	-24.0 kcal/mol
tcc-miR9471	XM_018116619.1	Pentatricopeptide repeat -containing protein At3g13880 (LOC108661004)	CAGTGATGCTCACTCAGCAA	target 5' C AAUCUGAUA G 3' GGUUAUG UGGCAGUUUCAU CUAGUAC ACCGUCAAAGUA miRNA 3' A 5'	-27.5 kcal/mol
tcc-miR9471	XM_018115965.1	Phosphoribosylamine-- glycine ligase(LOC18607134)	CAGTGATGCTCACTCAGCAA	target 5' U CG A A 3' GAUU G UGGCAGUUUCAU CUAG C ACCGUCAAAGUA miRNA 3' UA A 5'	-28.3 kcal/mol

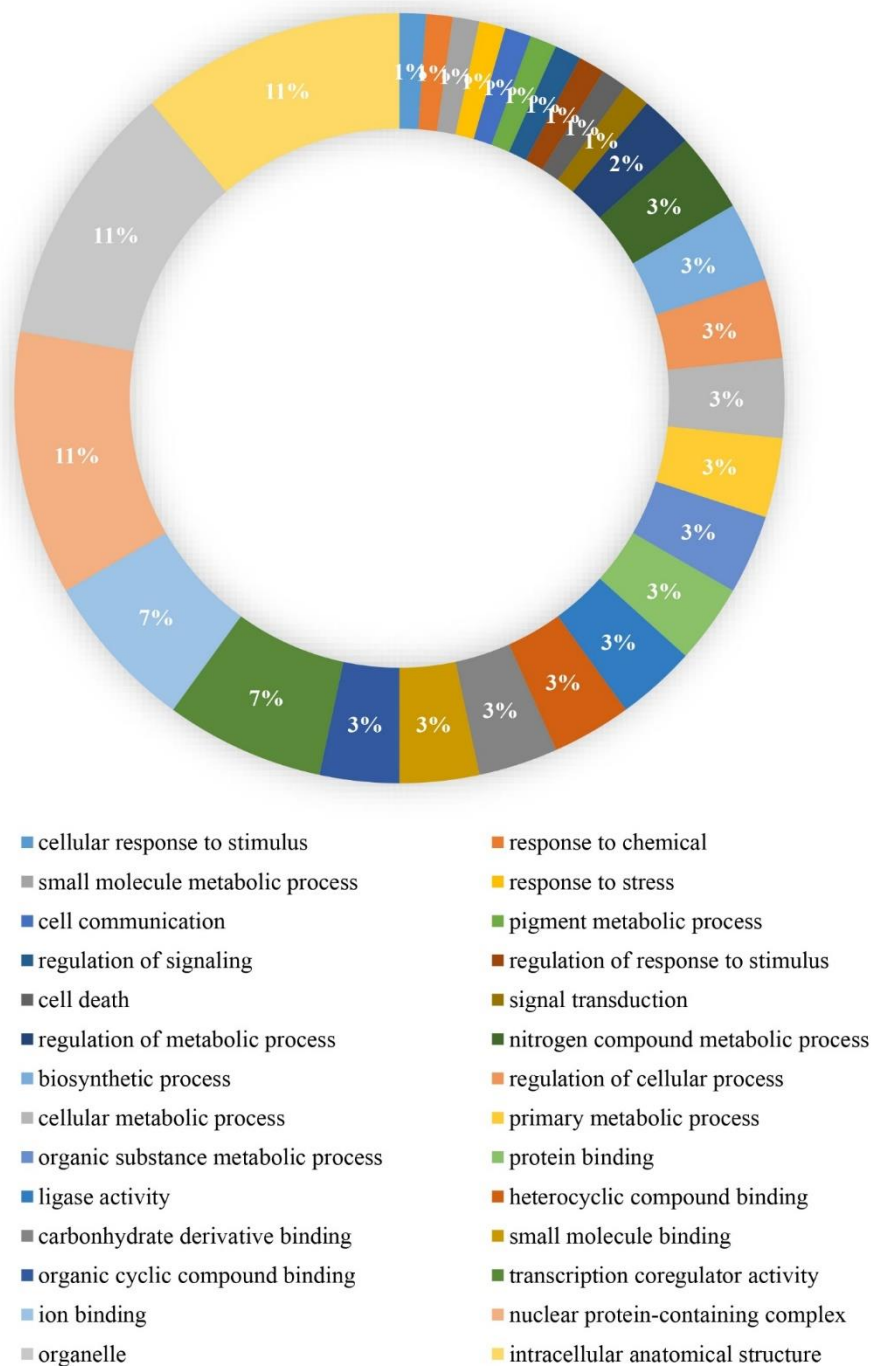


Figure 4.28 GO analyses results of targets of newly identified tcc-miR9471

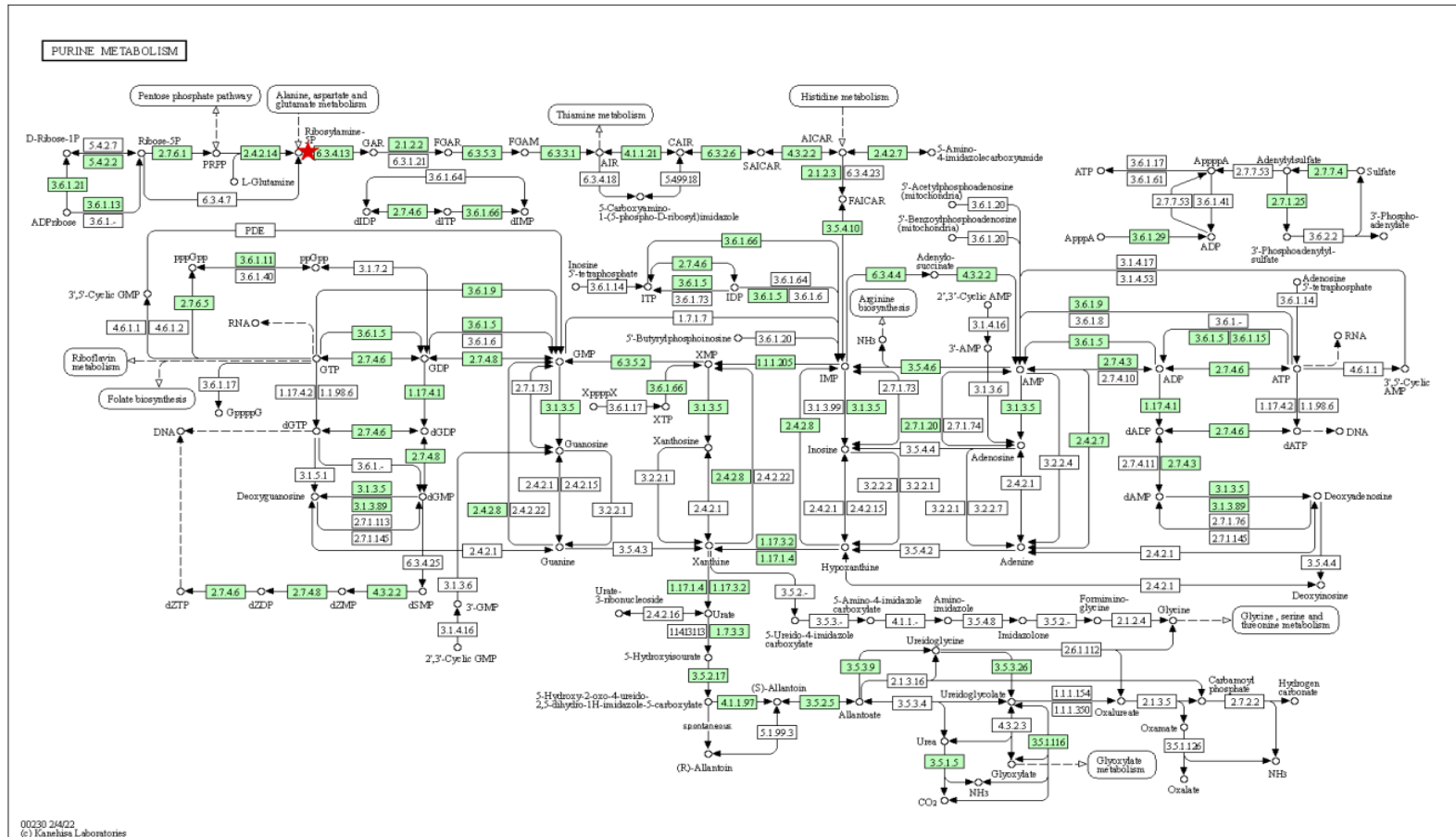


Figure 4.29 Purine metabolism

⁹The pathway to which tcc-miR9471 is associated is indicated by a red star.

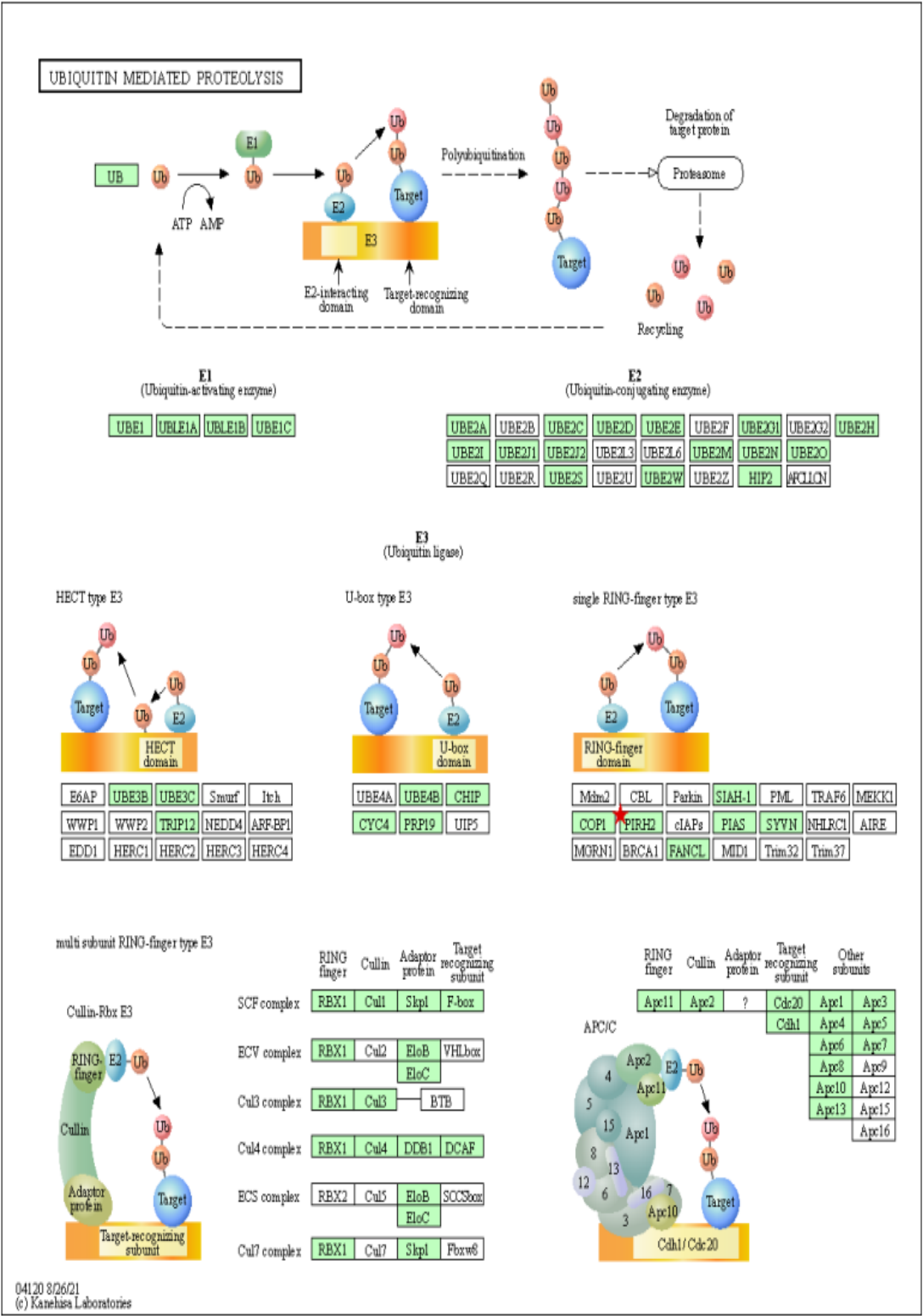


Figure 4.30 Ubiquitin-mediated proteolysis

¹⁰The protein to which tcc-miR9471 is associated is indicated by a red star.

4.4 Related Pathways, Functions, and Network Interactions of Novel Cacao miRNAs

Kegg pathway analyses showed the association of some cacao miRNAs with essential mechanisms and proteins, as mentioned in section 4.3. For example, tcc-miR166b is related to glycosaminoglycan degradation and metabolic pathways, and tcc-miR3476 is related to photosynthesis [141], metabolic pathways, and the *PsaG* gene; it has been predicted to be associated. Apart from these, some genes and mechanisms identified in previous studies are thought to be important oncogenes in cancer prognosis and targets for new therapeutics [142], [143], [144], [145] and [146]. Notice these important detected cacao miRNAs, their target genes, and their mechanisms.

tcc-miR8150, mRNA surveillance pathway [147], nucleocytoplasmic transport [148], and *UPFI* gene. tcc-miR8150, the relationship between *UPFI* in the mRNA surveillance pathway and nucleocytoplasmic transport. Nonsense-mediated mRNA degradation (NMD) is a highly conserved mRNA quality control process that eliminates aberrant transcripts [144]. As a master regulator of NMD, Up-frameshift protein 1 (*UPFI*) was first identified in 1996. It is an evolutionarily conserved and ubiquitously expressed phosphoprotein with RNA/DNA-dependent ATPase and RNA helicase activity [149]. Studies in various cancers have shown that *UPFI* acts as a tumor suppressor, and the expression level of *UPFI* is downregulated in most cell lines of cancers [144].

tcc-miR8150 and tcc-miR168, are related to spliceosome mechanism and *p68* and *SR* genes [150], [151]. Splicing mRNA precursors (pre-mRNAs) is a fundamental mechanism in the post-transcriptional regulation of gene expression [152]. Type-specific alternative splicing is essential for various biological processes, such as tissue development and cell differentiation [153]. Recent research has shown that pre-mRNA splicing also plays a role in various diseases, particularly cancer. It has been observed that the expressions and activities of splicing factors differ between various types of cancer, resulting in different splicing patterns of multiple transcripts that contribute to disease pathophysiologies such as proliferation,

migration, and hormone response [154], [155], [156] and [157]. Based on these results, it is thought that splicing factors, which are irregular in cancer, are potential targets for prognosis, diagnosis, and treatment [152]. Splicing factors (*SRs*) (*SNCG*) are the basis of the spliceosome mechanism and play critical roles in regulating alternative splicing [158]. Studies have proven that *SRs* are involved in almost every step of spliceosome assembly [159], and *SRs* are associated with genomic stability, mRNA export, mRNA stability, and translation [160], [161]. Based on its mechanism of action and mentioned properties, many studies have shown that *SRs* are abnormally expressed in human cancers and participate in tumorigenesis and metastasis [162]. *p68* (*DDX5*) is a DAD-box family's representative nuclear member, discovered via an antibody cross-reaction with the simian virus SV-40 large T antigen [163]. The first identified RNA is stranded and carries extensive amino acid sequence homology to the eukaryotic translation initiation factor eIF-4A. *p68* is a multifunctional protein with diverse roles in transcription, RNA processing, and miRNA processing [164]. Studies have shown that aberrant expression of *p68* exists in various types of cancer due to its potential roles in cell proliferation and transformation. It also draws attention due to its transcriptional coactivator roles that lead to its oncogenic feature [165].

Ubiquitin-mediated proteolysis and the *PIRH2* gene [166] are related to another miRNA identified in this study, tcc-miR9471, purine metabolism, metabolic pathways, and biosynthesis of secondary metabolites. *UPS* (ubiquitin-proteasome system) is a proteolysis system essential for cellular protein homeostasis and regulates protein degradation. This system consists of ubiquitin, ubiquitin-activating enzymes (E1s), ubiquitin-conjugated enzymes (E2s), ubiquitin ligases (E3s), deubiquitinating enzymes (DUBs), and 26S proteasome components [167]. An E3 ubiquitin ligase *Pirh2* (p53-induced RING-H2 protein) encoded by the *RCHY1* gene has different names such as CHIMP (CH-rich-interacting match with PLAG1), *RNF199* (RING finger protein 199) ZNF363 (Zinc finger protein 363). It was first described as *ARNIP* (Androgen receptor N-terminal-interacting protein) in 2002 [168]. *Pirh2* is a transcriptional target of the p53 tumor suppressor protein and targets *p53* for proteasomal degradation [169]. *Pirh2* is

known to cleave several proteins and *p53* family members that play a crucial role in cellular processes such as DNA damage response, cell cycle progression, gene expression regulation, and tumor transformation. Previous studies have observed that the expression of *Pirh2* is increased or decreased in different cancer types; it has a prognostic role as a biomarker [170].

tcc-miR169, ribosome biogenesis in eukaryotes and *UTP22* gene [171]. tcc-miR169c, the relationship between *UTP22(NOL6)* in ribosome biogenesis in eukaryotes. Nucleolar protein 6 (*NOL6*), also known as *NRAP*, *UTP22*, is encoded by the *NOL6* gene and encodes nucleolar RNA-related proteins, highly conserved across species [172] in the establishment of RNA binding and localization [173]. From recent *in silico* studies, Meng et al. found that *NOL6* promotes cell proliferation and is a potential new prognostic marker for hepatocellular carcinoma [143]; Dong et al. have determined that *NOL6* is an oncogene in prostate cancer [173].

At the same time, using the GeneMANIA Cytoscape plug-in, we created an interaction network of essential target genes (*RCHY1*, *NOL6*, *DDX5*, *UPF1*, *SNCG*) in the mechanisms associated with cacao miRNAs with human genes (Figure 4.31). Figure 4.31 showed that between detected genes and other human genes physical interactions (77.64%), co-expression (8.01%), predicted (5.37%), co-localization (3.63%), genetic interactions (2.87%), pathway (1.88%) and shared protein domains (0.60%). According to Genemania, it is shown in Figure 4.31 that 5 genes common to cacao and humans, which are important in the human mechanisms associated with the new cacao miRNAs identified because of the analysis, may interact with each other at a high level with 20 different genes in humans. The fact that the related human genes were genes involved in the detected mechanisms also supported the cross-kingdom approach.

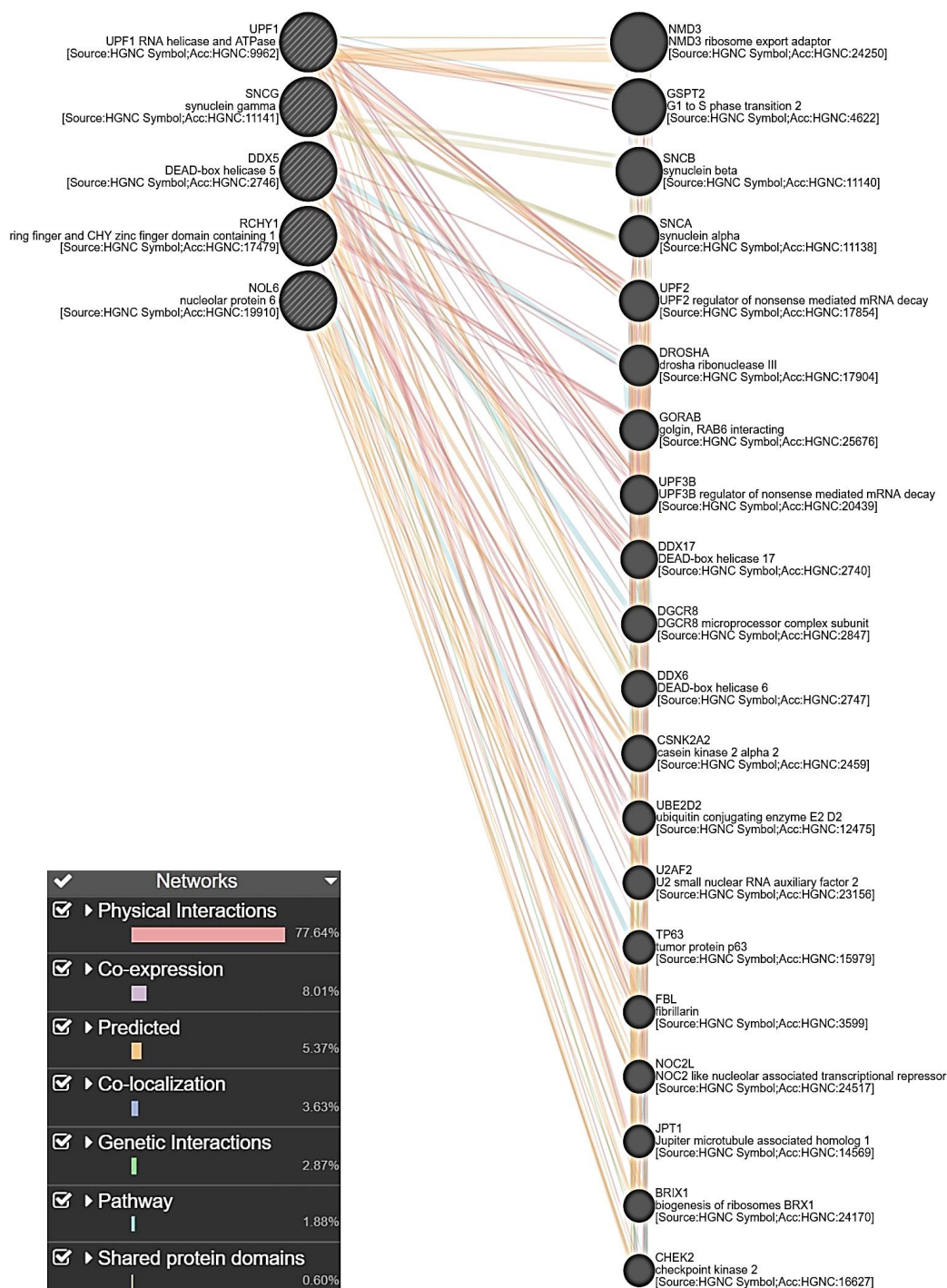


Figure 4.31 Interaction network of essential target genes (*RCHY1*, *NOL6*, *DDX5*, *UPF1*, *SNCG*) in the mechanisms associated with cacao miRNAs with human genes; light pink indicates physical interactions, orange for predicted interactions, purple for possible interactions of co-expressed genes, green for genetic interactions, blue for co-localization, turquoise for interactions on pathways, and yellow for shared protein domain interaction

There are other very important supportive studies conducted with a cross-kingdom approach, such as this *in silico* study we conducted.

For example, Jha et al. conducted a study on *Carica papaya*, specifically investigating miRNAs derived from papaya and their potential targets in humans. The identification of miRNAs from papaya and their potential interaction with human genes suggests the presence of a cross-kingdom regulatory mechanism. The researchers discovered novel miRNAs in papaya and assessed their regulatory effects on specific targets in the human genome. These targets are involved in a range of biological processes and diseases, encompassing cancer, diabetes, mental illness, and platelet disorders. Notably, the tcc-miR8150 that we detected in cacao is highly conserved and shared within the papaya family [102].

A significant study by Zhang et al. sheds light on cross-kingdom interactions by revealing the presence of plant-based miRNAs in human serum. In their research, they specifically focused on the tcc-miR168 family, which we also identified in our cacao analyses, and found that miR-168a, derived from the rice plant (*Oryza sativa*), can be absorbed through the intestines. They demonstrated that this plant-derived miRNA inhibits the expression of its target gene, low-density lipoprotein receptor 1 (*LDLRAP1*), in the liver. This finding highlights the ability of plant miRNAs to affect human gene regulation [4].

Mangukia et al. investigated the interaction between fruit-derived plant microRNAs (miRNAs) and the viral genome of SARS-CoV-2, with a particular focus on the Omicron variant. The researchers aimed to identify plant miRNAs that could potentially target conserved regions of the viral genome and explore their potential as therapeutic candidates for silencing the viral genome. The interaction of *C. sinensis* miRNA - csi-miR169-3p and SARS-CoV-2 from the same family as tcc-miR169c, which we also identified in cacao, was noted in their analyses, because the 3' UTR region "CTGCCT" targeted for all 772 Omicron variants curated worldwide is preserved among them. Thus, this site "CTGCCT" and miRNA csi-miR169-3p may become promising therapeutic candidates for triggering viral genome silencing. Thus, this study reveals the mechanistic way in

which fruits tend to initiate a fight against viruses such as SARS-CoV-2 and help maintain an individual's strong immune system [104].

In this thesis, 11 novel miRNAs of the *Theobroma cacao* including tcc-miR166b, tcc-miR166i, tcc-miR168, tcc-miR169, tcc-miR1051, tcc-miR3476, tcc-miR4249, tcc-miR6161, tcc-miR6279, tcc-miR8150 and tcc-miR9471b and their 47 target genes. The targets of miRNA were not only detected in only cacao genome but also in humans in a cross-kingdom approach. This study has reached a remarkable conclusion due to the correlation between the mechanisms and the target genes of miRNAs, particularly those linked to human cancer. The conclusions drawn from previous investigations along with our study's findings point to the substantial and widespread advantages of miRNA screening, not only for plants in general but also for the cacao plant. The high similarity of most human miRNAs with plant miRNAs in studies reveals the critical potential of plants for human disease prognosis, diagnosis, and treatment. Therefore, this study provides detailed *in silico* analysis to demonstrate the potential of cacao and its miRNAs in the prognosis, diagnosis, and treatment of cancer. By utilizing cross-kingdom interactions, we can enhance our understanding of human mechanisms and create molecular dockings between miRNA and associated proteins.

REFERENCES

- [1] G. M. Cragg, D. J. Newman, and K. M. Snader, “Natural products in drug discovery and development,” *J Nat Prod*, vol. 60, no. 1, pp. 52–60, Jan. 1997, doi: 10.1021/NP9604893.
- [2] G. M. Cragg and D. J. Newman, “Natural products: a continuing source of novel drug leads,” *Biochim Biophys Acta*, vol. 1830, no. 6, pp. 3670–3695, Jun. 2013, doi: 10.1016/J.BBAGEN.2013.02.008.
- [3] B. Zhang, X. Pan, G. P. Cobb, and T. A. Anderson, “Plant microRNA: a small regulatory molecule with big impact,” *Dev Biol*, vol. 289, no. 1, pp. 3–16, Jan. 2006, doi: 10.1016/J.YDBIO.2005.10.036.
- [4] L. Zhang *et al.*, “Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA,” *Cell Res*, vol. 22, no. 1, pp. 107–126, Jan. 2012, doi: 10.1038/CR.2011.158.
- [5] L. García-Segura, M. Pérez-Andrade, and J. Miranda-Ríos, “The emerging role of microRNAs in the regulation of gene expression by nutrients,” *J Nutrigenet Nutrigenomics*, vol. 6, no. 1, pp. 16–31, 2013, doi: 10.1159/000345826.
- [6] J. Link *et al.*, “Food-Derived Xeno-microRNAs: Influence of Diet and Detectability in Gastrointestinal Tract-Proof-of-Principle Study,” *Mol Nutr Food Res*, vol. 63, no. 2, Jan. 2019, doi: 10.1002/MNFR.201800076.
- [7] F. Marzano *et al.*, “Plant miRNAs Reduce Cancer Cell Proliferation by Targeting MALAT1 and NEAT1: A Beneficial Cross-Kingdom Interaction,” *Front Genet*, vol. 11, Sep. 2020, doi: 10.3389/fgene.2020.552490.
- [8] J. Delgado-Ospina *et al.*, “Bioactive compounds and techno-functional properties of high-fiber co-products of the cacao agro-industrial chain,” *Heliyon*, vol. 7, no. 4, p. e06799, Apr. 2021, doi: 10.1016/J.HELİYON.2021.E06799.
- [9] S. E. S. T. Panjikaran, S. C. L., and R. P. R., “Health benefits of bioactive compounds from cocoa (*Theobroma cacao*),” *Agricultural Reviews*, no. of, Aug. 2019, doi: 10.18805/AG.R-1851.

- [10] A. W. Indrianingsih, M. P. Wulanjati, A. Windarsih, D. K. Bhattacharjya, T. Suzuki, and T. Katayama, “In vitro studies of antioxidant, antidiabetic, and antibacterial activities of *Theobroma cacao*, *Annona muricata* and *Clitoria ternatea*,” *Biocatal Agric Biotechnol*, vol. 33, May 2021, doi: 10.1016/J.BCAB.2021.101995/REFERENCES.
- [11] P. P. Patil *et al.*, “Network pharmacology and in vitro testing of *Theobroma cacao* extract’s antioxidative activity and its effects on cancer cell survival,” *PLoS One*, vol. 17, no. 4, p. e0259757, Apr. 2022, doi: 10.1371/JOURNAL.PONE.0259757.
- [12] R. F. Ketting, “Developmental Cell The Many Faces of RNAi”, doi: 10.1016/j.devcel.2011.01.012.
- [13] P. Svoboda, “Introduction to RNAi and miRNA pathways,” *Introduction to RNAi and miRNA pathways*, Jul. 2020, doi: 10.14712/9788024643724.
- [14] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello, “Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*,” *Nature*, vol. 391, no. 6669, pp. 806–811, Feb. 1998, doi: 10.1038/35888.
- [15] C. Napoli, C. Lemieux, and R. Jorgensen, “Introduction of a Chimeric Chalcone Synthase Gene into *Petunia* Results in Reversible Co-Suppression of Homologous Genes *‘In trans*,” *Plant Cell*, vol. 2, pp. 279–289, 1990, Accessed: Jan. 08, 2024. [Online]. Available: <https://academic.oup.com/plcell/article/2/4/279/5983193>
- [16] R. C. Lee, R. L. Feinbaum, and V. Ambrost, “The *C. elegans* Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*,” *Cell*, vol. 75, pp. 843–854, 1993.
- [17] A. E. Pasquinelli *et al.*, “Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA,” *Nature 2000 408:6808*, vol. 408, no. 6808, pp. 86–89, Nov. 2000, doi: 10.1038/35040556.
- [18] A. R. Van Der Krol, L. A. Mur, M. Beld, J. N. M. Moiré, and A. R. Stuitje, “Flavonoid Genes in *Petunia*: Addition of a Limited Number of Gene Copies May Lead to a Suppression of Gene Expression,” *Plant Cell*, vol. 2, pp. 291–299, 1990,

Accessed: Jan. 08, 2024. [Online]. Available:
<https://academic.oup.com/plcell/article/2/4/291/5983192>

- [19] N. Romano and G. Macino, “Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences,” *Mol Microbiol*, vol. 6, no. 22, pp. 3343–3353, Nov. 1992, doi: 10.1111/J.1365-2958.1992.TB02202.X.
- [20] T. Tuschl, P. D. Zamore, R. Lehmann, D. P. Bartel, and P. A. Sharp, “fruit flies (Kennerdell and Carthew 1998; Misquitta and Paterson 1999), trypanosomes (Ngo et al. 1998), plants (Waterhouse et al. 1998), planaria (Sánchez-Alvarado and Newmark 1999), hydra,” *Hunter*, 1998, Accessed: Jan. 08, 2024. [Online]. Available: www.genesdev.org
- [21] P. D. Zamore, T. Tuschl, P. A. Sharp, and D. P. Bartel, “RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals,” *Cell*, vol. 101, no. 1, pp. 25–33, Mar. 2000, doi: 10.1016/S0092-8674(00)80620-0.
- [22] K. Bohmert, I. Camus, C. Bellini, D. Bouchez, M. Caboche, and C. Benning, “AGO1 defines a novel locus of *Arabidopsis* controlling leaf development,” *EMBO J*, vol. 17, no. 1, pp. 170–180, 1998.
- [23] H. Tabara, A. Grishok, and C. C. Mello, “RNAi in *C. elegans*: Soaking in the Genome Sequence,” *Science (1979)*, vol. 282, no. 5388, pp. 430–431, Oct. 1998, doi: 10.1126/SCIENCE.282.5388.430.
- [24] K. Lynn et al., “The PINHEAD/ZWILLE gene acts pleiotropically in *Arabidopsis* development and has overlapping functions with the ARGONAUTE1 gene,” *Development*, vol. 126, no. 3, pp. 469–481, Feb. 1999, doi: 10.1242/DEV.126.3.469.
- [25] T. Dalmay, A. Hamilton, S. Rudd, S. Angell, and D. C. Baulcombe, “An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus,” *Cell*, vol. 101, no. 5, pp. 543–553, May 2000, doi: 10.1016/S0092-8674(00)80864-8.

- [26] M. Fagard, S. Boutet, J. B. Morel, C. Bellini, and H. Vaucheret, “AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals,” *Proc Natl Acad Sci U S A*, vol. 97, no. 21, pp. 11650–11654, Oct. 2000, doi: 10.1073/PNAS.200217597.
- [27] A. Grishok, H. Tabara, and C. C. Mello, “Genetic Requirements for Inheritance of RNAi in *C. elegans*,” *Science (1979)*, vol. 287, no. 5462, pp. 2494–2497, Mar. 2000, doi: 10.1126/SCIENCE.287.5462.2494.
- [28] A. Smardon, J. M. Spoerke, S. C. Stacey, M. E. Klein, N. Mackin, and E. M. Maine, “EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*,” 2000.
- [29] S. Liu, M. Jaouannet, M. A. Dempsey, J. Imani, C. Coustau, and K.-H. Kogel, “RNA-based technologies for insect control in plant production,” 2019, doi: 10.1016/j.biotechadv.2019.107463.
- [30] O. Christiaens, J. Niu, C. Nji, and T. Taning, “insects Editorial RNAi in Insects: A Revolution in Fundamental Research and Pest Control Applications”, doi: 10.3390/insects11070415.
- [31] S. Kumar *et al.*, “RNA-targeting strategies as a platform for ocular gene therapy,” *Prog Retin Eye Res*, vol. 92, p. 101110, Jan. 2023, doi: 10.1016/J.PRETEYERES.2022.101110.
- [32] A. Kotowska-Zimmer, M. Pewinska, and M. Olejniczak, “Artificial miRNAs as therapeutic tools: Challenges and opportunities,” *Wiley Interdiscip Rev RNA*, vol. 12, no. 4, p. e1640, Jul. 2021, doi: 10.1002/WRNA.1640.
- [33] Q. Liu and Z. Paroo, “Biochemical Principles of Small RNA Pathways,” <http://dx.doi.org/10.1146/annurev.biochem.052208.151733>, vol. 79, pp. 295–319, Jun. 2010, doi: 10.1146/ANNUREV.BIOCHEM.052208.151733.
- [34] S. E. Castel and R. A. Martienssen, “RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond,” *Nature Reviews Genetics* 2013 14:2, vol. 14, no. 2, pp. 100–112, Jan. 2013, doi: 10.1038/nrg3355.

- [35] A. Movahedi *et al.*, “Plant small RNAs: definition, classification and response against stresses,” *Biologia (Poland)*, vol. 73, no. 3, pp. 285–294, Mar. 2018, doi: 10.2478/S11756-018-0034-5.
- [36] M. J. Axtell and J. L. Bowman, “Evolution of plant microRNAs and their targets,” *Trends Plant Sci*, vol. 13, no. 7, pp. 343–349, Jul. 2008, doi: 10.1016/J.TPLANTS.2008.03.009.
- [37] A. A. Covarrubias and J. L. Reyes, “Post-transcriptional gene regulation of salinity and drought responses by plant microRNAs,” *Plant Cell Environ*, vol. 33, no. 4, pp. 481–489, Apr. 2010, doi: 10.1111/J.1365-3040.2009.02048.X.
- [38] M. Chetta, L. Di Pietro, N. Bukvic, and W. Lattanzi, “Rising Roles of Small Noncoding RNAs in Cotranscriptional Regulation: In Silico Study of miRNA and piRNA Regulatory Network in Humans,” *Genes (Basel)*, vol. 11, no. 5, May 2020, doi: 10.3390/GENES11050482.
- [39] Y. Zhang *et al.*, “Identifying Cleaved and Noncleaved Targets of Small Interfering RNAs and MicroRNAs in Mammalian Cells by SpyCLIP,” *Mol Ther Nucleic Acids*, vol. 22, pp. 900–909, Dec. 2020, doi: 10.1016/J.OMTN.2020.10.009.
- [40] S. A. El-Kafrawy, S. S. Sohrab, Z. Mirza, A. M. Hassan, F. Alsaqaf, and E. I. Azhar, “In Vitro Inhibitory Analysis of Rationally Designed siRNAs against MERS-CoV Replication in Huh7 Cells,” *Molecules*, vol. 26, no. 9, May 2021, doi: 10.3390/MOLECULES26092610.
- [41] A. Garcia, S. Dunoyer-Geindre, S. Nolli, J. L. Reny, and P. Fontana, “An Ex Vivo and In Silico Study Providing Insights into the Interplay of Circulating miRNAs Level, Platelet Reactivity and Thrombin Generation: Looking beyond Traditional Pharmacogenetics,” *J Pers Med*, vol. 11, no. 5, p. 323, 2021, doi: 10.3390/JPM11050323.
- [42] N. K. Rana, N. Srivastava, and B. Koch, “Identification of the key miRNA; hsa-miR-1269a targeting TP53, Caspase-9 and FOXO3a in breast cancer cells under hypoxia by integrated bioinformatics analysis,” *Gene Rep*, vol. 25, p. 101408, Dec. 2021, doi: 10.1016/J.GENREP.2021.101408.

- [43] F. Morceau, S. Chateauvieux, A. Gaigneaux, M. Dicato, and M. Diederich, “Long and Short Non-Coding RNAs as Regulators of Hematopoietic Differentiation,” *International Journal of Molecular Sciences* 2013, Vol. 14, Pages 14744-14770, vol. 14, no. 7, pp. 14744–14770, Jul. 2013, doi: 10.3390/IJMS140714744.
- [44] B. J. Reinhart, E. G. Weinstein, M. W. Rhoades, B. Bartel, and D. P. Bartel, “MicroRNAs in plants,” 2002, doi: 10.1101/gad.1004402.
- [45] J. R. Chevillet, I. Lee, H. A. Briggs, Y. He, and K. Wang, “molecules Issues and Prospects of microRNA-Based Biomarkers in Blood and Other Body Fluids,” *Molecules*, vol. 19, pp. 6080–6105, 2008, doi: 10.3390/molecules19056080.
- [46] S. M. Peterson *et al.*, “Common features of microRNA target prediction tools,” 2014, doi: 10.3389/fgene.2014.00023.
- [47] A. Kozomara, M. Birgaoanu, and S. Griffiths-Jones, “miRBase: from microRNA sequences to function,” *Nucleic Acids Res*, vol. 47, pp. 155–162, 2018, doi: 10.1093/nar/gky1141.
- [48] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” *J Mol Biol*, vol. 215, no. 3, pp. 403–410, Oct. 1990, doi: 10.1016/S0022-2836(05)80360-2.
- [49] F. Bibi, M. Y. K. Barozai, and M. Din, “Bioinformatics profiling and characterization of potential microRNAs and their targets in the genus Coffea,” *Turkish Journal of Agriculture and Forestry*, vol. 41, no. 3, pp. 191–200, 2017, doi: 10.3906/TAR-1612-121.
- [50] A. Conesa and S. G. “Otz, “Methodology Report Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics,” *Int J Plant Genomics*, vol. 12, 2008, doi: 10.1155/2008/619832.
- [51] R. Lorenz *et al.*, “ViennaRNA Package 2.0,” 2011, Accessed: Jan. 09, 2024. [Online]. Available: <http://www.tbi.univie.ac.at/RNA>.
- [52] L. F. Sempere, S. Freemantle, I. Pitha-Rowe, E. Moss, E. Dmitrovsky, and V. Ambros, “Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal

- differentiation,” *Genome Biol*, vol. 5, no. 3, 2004, doi: 10.1186/GB-2004-5-3-R13.
- [53] A. Vá ló czi, C. Hornyik, N. ra Varga, J. zsef Burgyá, S. Kauppinen, and Z. Havelda, “Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes”, doi: 10.1093/nar/gnh171.
- [54] “_”, doi: 10.1242/dev.025080.
- [55] “MicroRNA sequences”, Accessed: Jan. 09, 2024. [Online]. Available: <https://academic.oup.com/nar/article/33/20/e179/1082564>
- [56] J. Wang *et al.*, “MicroRNAs in Plasma of Pancreatic Ductal Adenocarcinoma Patients as Novel Blood-Based Biomarkers of Disease”, doi: 10.1158/1940-6207.CAPR-09-0094.
- [57] J. M. Thomson, J. Parker, C. M. Perou, and S. M. Hammond, “A custom microarray platform for analysis of microRNA gene expression,” *Nature Methods* 2004 1:1, vol. 1, no. 1, pp. 47–53, Sep. 2004, doi: 10.1038/nmeth704.
- [58] J. Wang *et al.*, “Circulating microRNAs in pancreatic juice as candidate biomarkers of pancreatic cancer,” *J Cancer*, vol. 5, no. 8, pp. 696–705, 2014, doi: 10.7150/JCA.10094.
- [59] Y. Shirakami, T. Iwashita, S. Uemura, H. Imai, K. Murase, and M. Shimizu, “Micro-rna analysis of pancreatic cyst fluid for diagnosing malignant transformation of intraductal papillary mucinous neoplasm by comparing intraductal papillary mucinous adenoma and carcinoma,” *J Clin Med*, vol. 10, no. 11, p. 2249, Jun. 2021, doi: 10.3390/JCM10112249/S1.
- [60] L. Z. Hong *et al.*, “Systematic evaluation of multiple qPCR platforms, NanoString and miRNA-Seq for microRNA biomarker discovery in human biofluids,” *Scientific Reports* |, vol. 11, p. 4435, 123AD, doi: 10.1038/s41598-021-83365-z.
- [61] N. N. S. B. N. M. Kamal and W. N. S. Shahidan, “Non-exosomal and exosomal circulatory MicroRNAs: Which are more valid as biomarkers?,” *Front Pharmacol*, vol. 10, p. 487717, Jan. 2020, doi: 10.3389/FPHAR.2019.01500/BIBTEX.

- [62] Z. He *et al.*, “MicroRNA-137 reduces stemness features of pancreatic cancer cells by targeting KLF12”, doi: 10.1186/s13046-019-1105-3.
- [63] M. El Aamri, G. Yammouri, H. Mohammadi, A. Amine, and H. Korri-Youssoufi, “Electrochemical Biosensors for Detection of MicroRNA as a Cancer Biomarker: Pros and Cons”, doi: 10.3390/bios10110186.
- [64] Z. Aghili, N. Nasirizadeh, A. Divsalar, S. Shoeibi, and P. Yaghmaei, “A highly sensitive miR-195 nanobiosensor for early detection of Parkinson’s disease,” *Artif Cells Nanomed Biotechnol*, vol. 46, no. sup1, pp. 32–40, Oct. 2018, doi: 10.1080/21691401.2017.1411930.
- [65] S. Biswas, M. Haleyurgirisetty, S. Lee, I. Hewlett, and K. Devadas, “Development and validation of plasma miRNA biomarker signature panel for the detection of early HIV-1 infection,” 2019, doi: 10.1016/j.ebiom.2019.04.023.
- [66] N. Coban, D. Pirim, A. F. Erkan, B. Dogan, and B. Ekici, “Hsa-miR-584-5p as a novel candidate biomarker in Turkish men with severe coronary artery disease,” *Mol Biol Rep*, vol. 47, no. 2, pp. 1361–1369, Feb. 2020, doi: 10.1007/S11033-019-05235-2/METRICS.
- [67] J. Duy *et al.*, “Circulating microRNA profiles of Ebola virus infection OPEN,” 2016, doi: 10.1038/srep24496.
- [68] X. Ouyang *et al.*, “Dysregulated Serum MiRNA Profile and Promising Biomarkers in Dengue-infected Patients,” *Int. J. Med. Sci*, vol. 13, no. 3, pp. 195–205, 2016, doi: 10.7150/ijms.13996.
- [69] M. Smolarz and P. Widlak, “cancers Serum Exosomes and Their miRNA Load-A Potential Biomarker of Lung Cancer,” 2021, doi: 10.3390/cancers13061373.
- [70] M. I. Qadir and A. Faheem, “miRNA: A Diagnostic and Therapeutic Tool for Pancreatic Cancer,” *Critical Reviews™ in Eukaryotic Gene Expression*, vol. 27, no. 3, pp. 197–204, 2017, doi: 10.1615/CRITREVEUKARYOTGENEEXPR.2017019494.

- [71] H.-P. Tsai, S.-F. Huang, C.-F. Li, H.-T. Chien, and S.-C. Chen, “Differential microRNA expression in breast cancer with different onset age,” 2018, doi: 10.1371/journal.pone.0191195.
- [72] F. Urabe *et al.*, “Large-scale circulating microRNA profiling for the liquid biopsy of prostate cancer,” *Clinical Cancer Research*, vol. 25, no. 10, pp. 3016–3025, May 2019, doi: 10.1158/1078-0432.CCR-18-2849/74176/AM/LARGE-SCALE-CIRCULATING-MICRORNA-PROFILING-FOR-THE.
- [73] G. Du, C. Ren, J. Wang, and J. Ma, “Research Article The Clinical Value of Blood miR-654-5p, miR-126, miR-10b, and,” 2022, doi: 10.1155/2022/8225966.
- [74] G. Jung, E. Hernández-Illán, L. Moreira, F. Balaguer, and A. Goel, “Epigenetics of colorectal cancer: biomarker and therapeutic potential,” *Nature Reviews Gastroenterology & Hepatology* 2020 17:2, vol. 17, no. 2, pp. 111–130, Jan. 2020, doi: 10.1038/s41575-019-0230-y.
- [75] H. He *et al.*, “The APEX1/miRNA-27a-5p axis plays key roles in progression, metastasis and targeted chemotherapy of gastric cancer,” *Int J Pharm*, vol. 599, p. 120446, Apr. 2021, doi: 10.1016/J.IJPHARM.2021.120446.
- [76] M. Zhang *et al.*, “miRNA-5119 regulates immune checkpoints in dendritic cells to enhance breast cancer immunotherapy,” *Cancer Immunology, Immunotherapy*, vol. 69, no. 6, pp. 951–967, Jun. 2020, doi: 10.1007/S00262-020-02507-W/METRICS.
- [77] C. Fang *et al.*, “Functional miRNA variants affect lung cancer susceptibility and platinum-based chemotherapy response,” *J Thorac Dis*, vol. 10, no. 6, pp. 3329–3340, Jun. 2018, doi: 10.21037/JTD.2018.05.145.
- [78] Y. Lee *et al.*, “MicroRNA genes are transcribed by RNA polymerase II,” *EMBO J*, vol. 23, no. 20, pp. 4051–4060, Oct. 2004, doi: 10.1038/SJ.EMBOJ.7600385.
- [79] J. Krol, I. Loedige, and W. Filipowicz, “The widespread regulation of microRNA biogenesis, function and decay,” *Nat Rev Genet*, vol. 11, no. 9, pp. 597–610, Sep. 2010, doi: 10.1038/NRG2843.

- [80] S. W. S. Young, M. Stenzel, and Y. Jia-Lin, “Nanoparticle-siRNA: A potential cancer therapy?,” *Crit Rev Oncol Hematol*, vol. 98, pp. 159–169, Feb. 2016, doi: 10.1016/j.critrevonc.2015.10.015.
- [81] Y. K. Kim and V. N. Kim, “Processing of intronic microRNAs,” *EMBO J*, vol. 26, no. 3, pp. 775–783, Feb. 2007, doi: 10.1038/SJ.EMBOJ.7601512.
- [82] A. Helwak, G. Kudla, T. Dudnakova, and D. Tollervey, “Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding”, doi: 10.1016/j.cell.2013.03.043.
- [83] S. Jonas and E. Izaurralde, “Towards a molecular understanding of microRNA-mediated gene silencing,” *Nature Reviews Genetics* 2015 16:7, vol. 16, no. 7, pp. 421–433, Jun. 2015, doi: 10.1038/nrg3965.
- [84] J. O’Brien, H. Hayder, Y. Zayed, and C. Peng, “Overview of microRNA biogenesis, mechanisms of actions, and circulation,” *Front Endocrinol (Lausanne)*, vol. 9, no. AUG, p. 402, Aug. 2018, doi: 10.3389/FENDO.2018.00402/BIBTEX.
- [85] H. Hayder, J. O’Brien, U. Nadeem, and C. Peng, “MicroRNAs: crucial regulators of placental development,” *Reproduction*, vol. 155, no. 6, pp. R259–R271, Jun. 2018, doi: 10.1530/REP-17-0603.
- [86] M. J. Axtell, J. O. Westholm, and E. C. Lai, “Vive la différence: Biogenesis and evolution of microRNAs in plants and animals,” *Genome Biol*, vol. 12, no. 4, pp. 1–13, Apr. 2011, doi: 10.1186/GB-2011-12-4-221/FIGURES/2.
- [87] S. R. Liu, J. J. Zhou, C. G. Hu, C. L. Wei, and J. Z. Zhang, “MicroRNA-mediated gene silencing in plant defense and viral counter-defense,” *Front Microbiol*, vol. 8, no. SEP, p. 1801, Sep. 2017, doi: 10.3389/FMICB.2017.01801/BIBTEX.
- [88] Z. H. Wang and C. J. Xu, “Research Progress of MicroRNA in Early Detection of Ovarian Cancer,” *Chin Med J (Engl)*, vol. 128, no. 24, p. 3363, Dec. 2015, doi: 10.4103/0366-6999.171459.
- [89] Sanchita, R. Trivedi, M. H. Asif, and P. K. Trivedi, “Dietary plant miRNAs as an augmented therapy: cross-kingdom gene regulation,”

- <https://doi.org/10.1080/15476286.2018.1551693>, vol. 15, no. 12, pp. 1433–1439, Dec. 2018, doi: 10.1080/15476286.2018.1551693.
- [90] A. A. Millar and P. M. Waterhouse, “Plant and animal microRNAs: similarities and differences,” *Funct Integr Genomics*, vol. 5, no. 3, pp. 129–135, Jul. 2005, doi: 10.1007/S10142-005-0145-2.
 - [91] A. Brümmer and J. Hausser, “MicroRNA binding sites in the coding region of mRNAs: extending the repertoire of post-transcriptional gene regulation,” *Bioessays*, vol. 36, no. 6, pp. 617–626, 2014, doi: 10.1002/BIES.201300104.
 - [92] S. Ch, “microRNAs: a newly described class of encoded molecules that play a role in health and disease,” vol. 14, pp. 236–240, 2010.
 - [93] D. Didiano and O. Hobert, “Molecular architecture of a miRNA-regulated 3' UTR,” *RNA*, vol. 14, no. 7, p. 1297, Jul. 2008, doi: 10.1261/RNA.1082708.
 - [94] H. oki Iwakawa and Y. Tomari, “Molecular insights into microRNA-mediated translational repression in plants,” *Mol Cell*, vol. 52, no. 4, pp. 591–601, Nov. 2013, doi: 10.1016/J.MOLCEL.2013.10.033.
 - [95] G. M. Cragg, D. J. Newman, and K. M. Snader, “Natural products in drug discovery and development,” *J Nat Prod*, vol. 60, no. 1, pp. 52–60, Jan. 1997, doi: 10.1021/NP9604893.
 - [96] G. M. Cragg and D. J. Newman, “Natural products: A continuing source of novel drug leads,” *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1830, no. 6, pp. 3670–3695, Jun. 2013, doi: 10.1016/J.BBAGEN.2013.02.008.
 - [97] L. García-Segura, M. Pérez-Andrade, and J. Miranda-Ríos, “The emerging role of MicroRNAs in the regulation of gene expression by nutrients,” *J Nutrigenet Nutrigenomics*, vol. 6, no. 1, pp. 16–31, 2013, doi: 10.1159/000345826.
 - [98] J. Link *et al.*, “Food-Derived Xeno-microRNAs: Influence of Diet and Detectability in Gastrointestinal Tract-Proof-of-Principle Study,” *Mol Nutr Food Res*, vol. 63, no. 2, Jan. 2019, doi: 10.1002/MNFR.201800076.

- [99] L. Zhang *et al.*, “Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA,” *Cell Res*, vol. 22, no. 1, pp. 107–126, Jan. 2012, doi: 10.1038/CR.2011.158.
- [100] K. Tatiparti, S. Sau, S. K. Kashaw, and A. K. Iyer, “siRNA Delivery Strategies: A Comprehensive Review of Recent Developments,” *Nanomaterials 2017, Vol. 7, Page 77*, vol. 7, no. 4, p. 77, Apr. 2017, doi: 10.3390/NANO7040077.
- [101] M. Smolarz and P. Widlak, “Serum Exosomes and Their miRNA Load—A Potential Biomarker of Lung Cancer,” *Cancers 2021, Vol. 13, Page 1373*, vol. 13, no. 6, p. 1373, Mar. 2021, doi: 10.3390/CANCERS13061373.
- [102] N. Jha *et al.*, “Small RNA sequencing and identification of papaya (*Carica papaya* L.) miRNAs with potential cross-kingdom human gene targets,” *Molecular Genetics and Genomics 2022 297:4*, vol. 297, no. 4, pp. 981–997, May 2022, doi: 10.1007/S00438-022-01904-3.
- [103] N. Mangukia, P. Rao, K. Patel, H. Pandya, and R. M. Rawal, “Unveiling the nature’s fruit basket to computationally identify *Citrus sinensis* csi-mir169–3p as a probable plant miRNA against Reference and Omicron SARS-CoV-2 genome,” *Comput Biol Med*, vol. 146, p. 105502, Jul. 2022, doi: 10.1016/J.COMPBIOMED.2022.105502.
- [104] N. Mangukia, P. Rao, K. Patel, H. Pandya, and R. M. Rawal, “Unveiling the nature’s fruit basket to computationally identify *Citrus sinensis* csi-mir169–3p as a probable plant miRNA against Reference and Omicron SARS-CoV-2 genome,” *Comput Biol Med*, vol. 146, p. 105502, Jul. 2022, doi: 10.1016/J.COMPBIOMED.2022.105502.
- [105] S. E. S. T. Panjikkaran, S. C. L., and R. P. R., “Health benefits of bioactive compounds from cocoa (*Theobroma cacao*),” *Agricultural Reviews*, no. of, Aug. 2019, doi: 10.18805/AG.R-1851.
- [106] M. Santander Muñoz, J. Rodríguez Cortina, F. E. Vaillant, and S. Escobar Parra, “An overview of the physical and biochemical transformation of cocoa seeds to beans and to chocolate: Flavor formation,”

- <https://doi.org/10.1080/10408398.2019.1581726>, vol. 60, no. 10, pp. 1593–1613, May 2019, doi: 10.1080/10408398.2019.1581726.
- [107] L. J. Coronado-Cáceres, G. Rabadán-Chávez, L. Mojica, B. Hernández-Ledesma, L. Quevedo-Corona, and E. L. Cervantes, “Cocoa (*Theobroma cacao* L.) Seed Proteins’ Anti-Obesity Potential through Lipase Inhibition Using In Silico, In Vitro and In Vivo Models,” *Foods* 2020, Vol. 9, Page 1359, vol. 9, no. 10, p. 1359, Sep. 2020, doi: 10.3390/FOODS9101359.
- [108] O. Yañez *et al.*, “*Theobroma cacao* L. compounds: Theoretical study and molecular modeling as inhibitors of main SARS-CoV-2 protease,” *Biomedicine & Pharmacotherapy*, vol. 140, p. 111764, Aug. 2021, doi: 10.1016/J.BIOPHA.2021.111764.
- [109] B. Zhang, X. Pan, C. H. Cannon, G. P. Cobb, and T. A. Anderson, “Conservation and divergence of plant microRNA genes,” *Plant J*, vol. 46, no. 2, pp. 243–259, Apr. 2006, doi: 10.1111/J.1365-313X.2006.02697.X.
- [110] A. Kozomara, M. Birgaoanu, and S. Griffiths-Jones, “miRBase: from microRNA sequences to function,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D155–D162, Jan. 2019, doi: 10.1093/NAR/GKY1141.
- [111] B. H. Zhang, X. P. Pan, Q. L. Wang, G. P. Cobb, and T. A. Anderson, “Identification and characterization of new plant microRNAs using EST analysis,” *Cell Res*, vol. 15, no. 5, pp. 336–360, 2005, doi: 10.1038/SJ.CR.7290302.
- [112] S. Marakli, “Identification and functional analyses of new sesame miRNAs (*Sesamum indicum* L.) and their targets,” *Mol Biol Rep*, vol. 45, no. 6, pp. 2145–2155, Dec. 2018, doi: 10.1007/S11033-018-4373-7.
- [113] S. F. Altschul *et al.*, “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.,” *Nucleic Acids Res*, vol. 25, no. 17, p. 3389, Sep. 1997, doi: 10.1093/NAR/25.17.3389.
- [114] R. Lorenz, I. L. Hofacker, and P. F. Stadler, “RNA folding with hard and soft constraints,” *Algorithms Mol Biol*, vol. 11, no. 1, 2016, doi: 10.1186/S13015-016-0070-Z.

- [115] M. Zuker and P. Stiegler, “Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information.,” *Nucleic Acids Res*, vol. 9, no. 1, p. 133, Jan. 1981, doi: 10.1093/NAR/9.1.133.
- [116] B. C. Meyers *et al.*, “Criteria for annotation of plant MicroRNAs,” *Plant Cell*, vol. 20, no. 12, pp. 3186–3190, 2008, doi: 10.1105/TPC.108.064311.
- [117] M. Y. K. Barozai, M. Din, and I. A. Baloch, “Structural and functional based identification of the bean (*Phaseolus*) microRNAs and their targets from expressed sequence tags,” *J Struct Funct Genomics*, vol. 14, no. 1, pp. 11–18, Mar. 2013, doi: 10.1007/S10969-013-9152-Z.
- [118] B. H. Zhang, X. P. Pan, S. B. Cox, G. P. Cobb, and T. A. Anderson, “Evidence that miRNAs are different from other RNAs,” *Cell Mol Life Sci*, vol. 63, no. 2, pp. 246–254, Jan. 2006, doi: 10.1007/S00018-005-5467-7.
- [119] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” *J Mol Biol*, vol. 215, no. 3, pp. 403–410, 1990, doi: 10.1016/S0022-2836(05)80360-2.
- [120] S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura, “MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms,” *Mol Biol Evol*, vol. 35, no. 6, pp. 1547–1549, Jun. 2018, doi: 10.1093/MOLBEV/MSY096.
- [121] J. Krüger and M. Rehmsmeier, “RNAhybrid: microRNA target prediction easy, fast and flexible,” *Nucleic Acids Res*, vol. 34, no. suppl_2, pp. W451–W454, Jul. 2006, doi: 10.1093/NAR/GKL243.
- [122] S. Marakli, “In silico determination of transposon-derived miRNAs and targets in *Aegilops* species,” *J Biomol Struct Dyn*, vol. 38, no. 10, pp. 3098–3109, Jul. 2020, doi: 10.1080/07391102.2019.1654409.
- [123] H. Mi, A. Muruganujan, D. Ebert, X. Huang, and P. D. Thomas, “PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D419–D426, Jan. 2019, doi: 10.1093/NAR/GKY1038.

- [124] S. Götz *et al.*, “High-throughput functional annotation and data mining with the Blast2GO suite,” *Nucleic Acids Res*, vol. 36, no. 10, pp. 3420–3435, Jun. 2008, doi: 10.1093/NAR/GKN176.
- [125] B. T. Sherman *et al.*, “DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update),” *Nucleic Acids Res*, vol. 50, no. W1, pp. W216–W221, Jul. 2022, doi: 10.1093/NAR/GKAC194.
- [126] M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, and M. Tanabe, “KEGG: integrating viruses and cellular organisms,” *Nucleic Acids Res*, vol. 49, no. D1, pp. D545–D551, Jan. 2021, doi: 10.1093/NAR/GKAA970.
- [127] D. Warde-Farley *et al.*, “The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function”, doi: 10.1093/nar/gkq537.
- [128] S. Marakli, “Identification and functional analyses of new sesame miRNAs (*Sesamum indicum* L.) and their targets,” *Mol Biol Rep*, vol. 45, no. 6, pp. 2145–2155, Dec. 2018, doi: 10.1007/S11033-018-4373-7.
- [129] R. Singh Tomar, R. Kumar Mishra, G. Jyot Kaur, and S. Kaushik, “In-Silico identification of miRNAs and their targets using expressed sequence tags (ESTs) in plants *Mimosa pudica*-A broad spectrum antimicrobial herb View project Suppression of Wound associated Biofilm forming pathogens using high efficiency metallic nanoparticles. View project In-Silico identification of miRNAs and their targets using Expressed Sequence Tags (ESTs) in Plants,” *Article in Journal of Pharmaceutical Sciences and Research*, 2017, Accessed: Jun. 10, 2023. [Online]. Available: <https://www.researchgate.net/publication/317752492>
- [130] A. A. Sabana, G. Antony, C. U. Rahul, and M. K. Rajesh, “In silico identification of microRNAs and their targets associated with coconut embryogenic calli,” *Agri Gene*, vol. 7, pp. 59–65, Mar. 2018, doi: 10.1016/J.AGGENE.2018.01.002.
- [131] K. Hussain, K. Mungikar, A. Kulkarni, and A. Kamble, “Identification, characterization and expression analysis of pigeonpea miRNAs in response to *Fusarium wilt*,” *Gene*, vol. 653, pp. 57–64, May 2018, doi: 10.1016/J.GENE.2018.02.017.

- [132] I. A. Baloch, M. Y. K. Barozai, and M. Din, "Identification and characterization of 25 microRNAs and their targeted proteins in apricot (*Prunus armeniaca* L.)," *JAPS, Journal of Animal and Plant Sciences*, vol. 25, no. 5, pp. 1466–1476, 2015.
- [133] J. Wang, X. Yang, H. Xu, X. Chi, M. Zhang, and X. Hou, "Identification and characterization of microRNAs and their target genes in *Brassica oleracea*," *Gene*, vol. 505, no. 2, pp. 300–308, Sep. 2012, doi: 10.1016/J.GENE.2012.06.002.
- [134] M. Younas and K. Barozai, "IDENTIFICATION OF MICRORNAS AND THEIR TARGETS IN *ARTEMISIA ANNUA* L," *Pak. J. Bot*, vol. 45, no. 2, pp. 461–465, 2013, Accessed: Jun. 10, 2023. [Online]. Available: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.
- [135] M. Y. K. Barozai, M. Din, and I. A. Baloch, "Structural and functional based identification of the bean (*Phaseolus*) microRNAs and their targets from expressed sequence tags," *J Struct Funct Genomics*, vol. 14, no. 1, pp. 11–18, Mar. 2013, doi: 10.1007/S10969-013-9152-Z.
- [136] Z. Gul, M. Y. K. Barozai, and M. Din, "In-silico based identification and functional analyses of miRNAs and their targets in Cowpea (*Vigna unguiculata* L.)," *AIMS Genet*, vol. 4, no. 2, pp. 138–165, Aug. 2017, doi: 10.3934/GENET.2017.2.138.
- [137] H. Joshi and M. K. Mandavia, "In silico identification and target prediction of micornas in Sesame (*sesamum indicum* L.) expressed sequence tags," *Genetics and Molecular Research*, vol. 17, no. 2, p. 16039911, Apr. 2018, doi: 10.4238/GMR16039911.
- [138] J. Wang, X. Yang, H. Xu, X. Chi, M. Zhang, and X. Hou, "Identification and characterization of microRNAs and their target genes in *Brassica oleracea*," *Gene*, vol. 505, no. 2, pp. 300–308, Sep. 2012, doi: 10.1016/J.GENE.2012.06.002.
- [139] F. Aydinoglu and S. J. Lucas, "Identification and expression profiles of putative leaf growth related microRNAs in maize (*Zea mays* L.) hybrid ADA313," *Gene*, vol. 690, pp. 57–67, Mar. 2019, doi: 10.1016/J.GENE.2018.12.042.

- [140] B. Bhan *et al.*, “Identification and expression profiling of miRNAs in two color variants of carrot (*Daucus carota* L.) using deep sequencing,” 2019, doi: 10.1371/journal.pone.0212746.
- [141] H. H. Dixon and T. G. Mason, “The Primary Sugar of Photosynthesis,” *Nature* 1916 97:2425, vol. 97, no. 2425, pp. 160–160, 1916, doi: 10.1038/097160b0.
- [142] Z. Xu, W. Wu, H. Yan, Y. Hu, Q. He, and P. Luo, “Regulation of p53 stability as a therapeutic strategy for cancer,” *Biochem Pharmacol*, vol. 185, p. 114407, Mar. 2021, doi: 10.1016/J.BCP.2021.114407.
- [143] L. Meng *et al.*, “Nucleolar protein 6 promotes cell proliferation and acts as a potential novel prognostic marker for hepatocellular carcinoma,” *Chin Med J (Engl)*, vol. 134, no. 21, pp. 2611–2618, Nov. 2021, doi: 10.1097/CM9.0000000000001655/ASSET/4ECEC274-39E9-45AF-9E10-D2FBCA5F5B04/ASSETS/GRAPHIC/0366-6999-134-21-017-F004.PNG.
- [144] B. L. Chen, H. M. Wang, X. S. Lin, and Y. M. Zeng, “UPF1: A potential biomarker in human cancers,” *Frontiers in Bioscience - Landmark*, vol. 26, no. 5, pp. 76–84, Apr. 2021, doi: 10.52586/4925/HTM.
- [145] F. Li, C. Fountzilias, I. Puzanov, K. M. Attwood, C. Morrison, and X. Ling, “Multiple functions of the DEAD-box RNA helicase, DDX5 (p68), make DDX5 a superior oncogenic biomarker and target for targeted cancer therapy,” *Am J Cancer Res*, vol. 11, no. 10, p. 5190, 2021, Accessed: Dec. 02, 2022. [Online]. Available: /pmc/articles/PMC8569338/
- [146] I. P. Nikas, S. C. Themistocleous, S. A. Paschou, K. I. Tsamis, and H. S. Ryu, “Serine-Arginine Protein Kinase 1 (SRPK1) as a Prognostic Factor and Potential Therapeutic Target in Cancer: Current Evidence and Future Perspectives,” *Cells* 2020, Vol. 9, Page 19, vol. 9, no. 1, p. 19, Dec. 2019, doi: 10.3390/CELLS9010019.
- [147] C. A. Beelman and R. Parker, “Degradation of mRNA in eukaryotes,” *Cell*, vol. 81, no. 2, pp. 179–183, Apr. 1995, doi: 10.1016/0092-8674(95)90326-7.

- [148] H. E. Neuman De Vegvart and J. E. Dahlberg, "Nucleocytoplasmic transport and processing of small nuclear RNA precursors," *Mol Cell Biol*, vol. 10, no. 7, pp. 3365–3375, Jul. 1990, doi: 10.1128/MCB.10.7.3365-3375.1990.
- [149] S. E. Applequist, M. Selg, C. Raman, and H. M. Jäck, "Cloning and characterization of HUPF1, a human homolog of the *Saccharomyces cerevisiae* nonsense mRNA-reducing UPF1 protein," *Nucleic Acids Res*, vol. 25, no. 4, pp. 814–821, Feb. 1997, doi: 10.1093/NAR/25.4.814.
- [150] S. M. Berget, C. Moore, and P. A. Sharp, "Spliced segments at the 5' terminus of adenovirus 2 late mRNA," *Proc Natl Acad Sci U S A*, vol. 74, no. 8, pp. 3171–3175, 1977, doi: 10.1073/PNAS.74.8.3171/ASSET/09233CCD-834C-41EB-941D-AE627986990A/ASSETS/GRAPHIC/PNAS.3171FIG05.JPEG.
- [151] L. T. Chow, J. M. Roberts, J. B. Lewis, and T. R. Broker, "A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids," *Cell*, vol. 11, no. 4, pp. 819–836, Aug. 1977, doi: 10.1016/0092-8674(77)90294-X.
- [152] T. Takeiwa, Y. Mitobe, K. Ikeda, K. Horie-Inoue, and S. Inoue, "Roles of Splicing Factors in Hormone-Related Cancer Progression," *International Journal of Molecular Sciences* 2020, Vol. 21, Page 1551, vol. 21, no. 5, p. 1551, Feb. 2020, doi: 10.3390/IJMS21051551.
- [153] F. E. Baralle and J. Giudice, "Alternative splicing as a regulator of development and tissue identity," *Nat Rev Mol Cell Biol*, vol. 18, no. 7, pp. 437–451, Jul. 2017, doi: 10.1038/NRM.2017.27.
- [154] O. Anczukow and A. R. Krainer, "Splicing-factor alterations in cancers," *RNA*, vol. 22, no. 9, pp. 1285–1301, Sep. 2016, doi: 10.1261/RNA.057919.116.
- [155] A. A. Agrawal, L. Yu, P. G. Smith, and S. Buonomici, "Targeting splicing abnormalities in cancer," *Curr Opin Genet Dev*, vol. 48, pp. 67–74, Feb. 2018, doi: 10.1016/J.GDE.2017.10.010.
- [156] X. Song, Z. Zeng, H. Wei, and Z. Wang, "Alternative splicing in cancers: From aberrant regulation to new therapeutics," *Semin Cell Dev Biol*, vol. 75, pp. 13–22, Mar. 2018, doi: 10.1016/J.SEMCDB.2017.09.018.

- [157] K. I. Takayama, “Splicing Factors Have an Essential Role in Prostate Cancer Progression and Androgen Receptor Signaling,” *Biomolecules* 2019, Vol. 9, Page 131, vol. 9, no. 4, p. 131, Apr. 2019, doi: 10.3390/BIOM9040131.
- [158] S. Jeong, “SR Proteins: Binders, Regulators, and Connectors of RNA,” *Mol Cells*, vol. 40, no. 1, p. 1, Jan. 2017, doi: 10.14348/MOLCELLS.2017.2319.
- [159] R. Savisaar and L. D. Hurst, “Purifying Selection on Exonic Splice Enhancers in Intronless Genes,” *Mol Biol Evol*, vol. 33, no. 6, pp. 1396–1418, Jun. 2016, doi: 10.1093/MOLBEV/MSW018.
- [160] Y. Huang, R. Gattoni, J. Stévenin, and J. A. Steitz, “SR Splicing Factors Serve as Adapter Proteins for TAP-Dependent mRNA Export,” *Mol Cell*, vol. 11, no. 3, pp. 837–843, Mar. 2003, doi: 10.1016/S1097-2765(03)00089-3.
- [161] X. Li and J. L. Manley, “Inactivation of the SR Protein Splicing Factor ASF/SF2 Results in Genomic Instability,” *Cell*, vol. 122, no. 3, pp. 365–378, Aug. 2005, doi: 10.1016/J.CELL.2005.06.008.
- [162] X. Zheng *et al.*, “Serine/arginine-rich splicing factors: the bridge linking alternative splicing and cancer,” *Int J Biol Sci*, vol. 16, no. 13, p. 2442, 2020, doi: 10.7150/IJBS.46751.
- [163] ac, “EXPRESSION OF THE ‘DEAD BOX’ RNA HELICASE p68 IS DEVELOPMENTALLY AND GROWTH REGULATED AND CORRELATES WITH ORGAN DIFFERENTIATION/MATURATION IN THE FETUS,” *J. Pathol*, vol. 184, pp. 351–359, 1998, doi: 10.1002/(SICI)1096-9896(199804)184:4.
- [164] F. V. Fuller-Pace and H. C. Moore, “RNA helicases p68 and p72: multifunctional proteins with important implications for cancer development,” <http://dx.doi.org/10.2217/fon.11.1>, vol. 7, no. 2, pp. 239–251, Feb. 2011, doi: 10.2217/FON.11.1.
- [165] S. Tabassum and M. K. Ghosh, “DEAD-box RNA helicases with special reference to p68: Unwinding their biology, versatility, and therapeutic opportunity in cancer,” *Genes Dis*, Mar. 2022, doi: 10.1016/J.GENDIS.2022.02.008.

- [166] A. Hershko, A. Ciechanover, H. Heller, A. L. Haas, and I. A. Rose, "Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis," *Proc Natl Acad Sci U S A*, vol. 77, no. 4, pp. 1783–1786, 1980, doi: 10.1073/PNAS.77.4.1783.
- [167] X. Zhang, S. Linder, and M. Bazzaro, "Drug Development Targeting the Ubiquitin-Proteasome System (UPS) for the Treatment of Human Cancers," *Cancers 2020, Vol. 12, Page 902*, vol. 12, no. 4, p. 902, Apr. 2020, doi: 10.3390/CANCERS12040902.
- [168] M. Trifiro and H.-Y. Kang, "Cloning and Characterization of Androgen Receptor Coactivator, ARA55, in Human Prostate," 2002, Accessed: Dec. 05, 2022. [Online]. Available: <http://www.endocrinology.org>
- [169] R. P. Leng *et al.*, "Pirh2, a p53-Induced Ubiquitin-Protein Ligase, Promotes p53 Degradation," *Cell*, vol. 112, no. 6, pp. 779–791, Mar. 2003, doi: 10.1016/S0092-8674(03)00193-4.
- [170] A. Daks *et al.*, "The RNA-binding protein HuR is a novel target of Pirh2 E3 ubiquitin ligase," *Cell Death & Disease 2021 12:6*, vol. 12, no. 6, pp. 1–14, Jun. 2021, doi: 10.1038/s41419-021-03871-w.
- [171] A. A. Hadjiolov, "Patterns of ribosome biogenesis in eukaryotes," *Trends Biochem Sci*, vol. 2, no. 4, pp. 84–86, Apr. 1977, doi: 10.1016/0968-0004(77)90044-5.
- [172] B. Utama, D. Kennedy, K. Ru, and J. S. Mattick, "Isolation and characterization of a new nucleolar protein, Nrap, that is conserved from yeast to humans," *Genes to Cells*, vol. 7, no. 2, pp. 115–132, Feb. 2002, doi: 10.1046/J.1356-9597.2001.00507.X.
- [173] D. Dong, M. Song, X. Wu, and W. Wang, "NOL6, a new founding oncogene in human prostate cancer and targeted by miR-590-3p," *Cytotechnology*, vol. 72, no. 3, p. 469, Jun. 2020, doi: 10.1007/S10616-020-00394-8.

PUBLICATIONS FROM THE THESIS

Conference Papers

1. Karaismailoğlu R. and Maraklı S, “ Detection of new miRNAs in cacao (*Theobroma cacao* L.), a raw material with high biotechnological potential,” in *6th International Plant Science and Technology Congress*, Eskişehir, Turkey, November 10, 2023, pp. 9, ISBN: 978-625-98726-2-9.

