Title

Identification of Cis-Regulatory Elements and miRNA Targets Involved in Mosquito Reproduction

Permalink

https://escholarship.org/uc/item/3hx3g5m3

Author

HA, JI SU

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Identification of *Cis*-Regulatory Elements and miRNA Targets Involved in
Mosquito Reproduction

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Genetics, Genomics and Bioinformatics

by

Ji Su Ha

August 2016

Dissertation Committee:
      Dr. Alexander S. Raikhel, Chairperson
      Dr. Anupama A. Dahanukar
      Dr. Jason E. Stajich

The Dissertation of Ji Su Ha is approved:

_____

_____

_____

<div align="right">Committee Chairperson</div>

University of California, Riverside

ACKNOWLEDGEMENTS

Lastly, I want to thank to my parents, sister, and brother. They are always supportive and listening to me whenever I talk to them about me. Thank you everybody.

ABSTRACT OF THE DISSERTATION

Identification of *Cis*-Regulatory Elements and miRNA Targets Involved in
Mosquito Reproduction

by

Ji Su Ha

Doctor of Philosophy, Graduate Program in Genetics, Genomics and Bioinformatics
University of California, Riverside, August 2016
Dr. Alexander S. Raikhel, Chairperson

Female *Aedes aegypti* mosquitoes spread pathogens for various vector-borne
diseases while feeding on vertebrate blood. Being anautogenous, these mosquitoes
require a blood meal for vitellogenesis, the process of yolk formation during the
maturation of eggs. A clear understanding of the molecular mechanism of reproduction,
such as detailed transcription mechanisms, should contribute significantly to the vector
control strategies. The aim of this project is to decipher how complex transcription
programs govern genes that are differentially expressed during female mosquito
reproduction. We have identified 89 putative transcription factor biding sites (TFBSs)
using bioinformatics tools, on the promoter regions of more then 1K differentially
regulated genes. The putative TFBSs were screened for positional bias, orientation bias
and evolutionary conservation, properties that are often associated with real TFBSs.
Promoters from orthologous genes in closely related species *Anopheles gambiae* and
Culex quinquefascitus, were used to check for evolutionary conservation. The above-

mentioned screens were used to increase the accuracy of the *in-silico* predictions. JASPAR database was searched to identity the known TFBSs and their corresponding TFs. GeneMANIA webtool was used for the construction of putative regulatory networks. Molecular biology techniques such as RNA-interference mediated depletion of selected TFs and quantitative reverse-transcription polymerase chain reaction have been used for verifying and evaluating the functionality of some of these putative TFBSs.

In addition to identifying *cis*-regulatory elements this dissertation also describes an approach developed for prediction of miRNA targets involved in mosquito reproduction. miRNAs are 21 to 24 nucleotide long non-coding RNAs that degrade or inhibit by coupling with target sites on mRNAs. Since the discovery huge numbers of miRNAs have been reported. However, the functions of only a few of these miRNAs are well understood. The study of miRNA targets is critical step for determining the functionality of miRNAs. There are many miRNA target prediction tools that are readily available, but almost all of these have a problem of producing huge number of false-positive results. Here, we have developed a new approach for computational target prediction that involves five different target prediction programs, the results have been experimentally verified in separate studies and found to be highly reliable.

Overall we have been able to identify binding sites for TFs and targets for miRNAs that are involved in female mosquito reproduction. We realize that future work remains, for a complete understanding of the complex molecular mechanisms, during the reproductive period. However this study can be regarded as a significant step towards the accomplishment of that goal.

TABLE OF CONTENTS

CHAPTER I.
Introduction

CHAPTER II.
Identification of binding sites for transcription factors for differential expression of genes during the vitellogenic period of the female yellow fever mosquito, *Aedes aegypti*

Chapter III.
Construction of gene regulatory networks involved in mosquito reproduction and validation of the functional role of certain components within the networks

CHAPTER IV.
Prediction of targets for miRNAs involved in *Aedes aegypti* mosquito reproduction

CHAPTER V.
Conclusions of the Dissertation

LIST OF FIGURES

CHAPTER II

Identification of binding sites for transcription factors for differential expression of genes during the vitellogenic period of the female yellow fever mosquito, *Aedes aegypti*

CHAPTER III

Construction of gene regulatory networks involved in mosquito reproduction and validation of the functional role of certain components within the networks

CHAPTER IV

Prediction of targets for miRNAs involved in *Aedes aegypti* mosquito reproduction

Figure

LIST OF TABLES

CHAPTER II

Identification of binding sites for transcription factors for differential expression of genes during the vitellogenic period of the female yellow fever mosquito, *Aedes aegypti*

CHAPTER III

Construction of gene regulatory networks involved in mosquito reproduction and validation of the functional role of certain components within the networks

CHAPTER IV

Prediction of targets for miRNAs involved in *Aedes aegypti* **mosquito reproduction**

Chapter I

Introduction

1.1 Introduction to mosquito reproductive biology

Hematophagous female mosquitoes are the most dangerous animal species to humans because they transmit devastating diseases. The negative effects of mosquitoes to global human health and economy are innumerable. *Anopheles gambiae*, the primary vector of malaria, is responsible for the most severe infection cases with high fatalities in the world (584,000 deaths in 2013) according to World Health Organization (WHO, 2014). Another mosquito species, the yellow fever mosquito *Aedes aegypti*, is the primary vector for transmitting Zika (ZIKV), dengue (DENV), yellow fever (YFV), and chikungunya (CHIKV) viruses (Behura et al., 2016). Recent unprecedented rapid spread of ZIKV has added to human suffering from these insects. The WHO declared ZIKV at the level of "Public Health Emergency of International Concern (PHEIC)" on February 1st, 2016 (CDC website). Public health authorities from over sixty countries have been monitoring ZIKV infection cases as of May 18, 2016 (WHO website). The DENV is another global burden for human beings. Approximately 390 million people are being infected with DENV every year and 3.9 billion people are potentially exposed to DENV infection in 128 countries (Bhatt et al., 2013). CHIKV-infected mosquitoes bite humans for blood feeding and transmit the virus to humans (WHO http://www.who.int/mediacentre/factsheets/fs327/en/). According to recently reported data, the CHIKV had been rapidly spreading in the Caribbean and American countries

with reported infection cases over 1.6 million infections and 270 mortalities between the October 2012 and October 2015 (Goupil et al., 2016).

To prevent and to control this tremendous damage resulted from mosquito-transmitted infectious diseases, the investigation of the molecular basis of disease transmission is necessary. In female mosquitoes, blood feeding is required for egg development providing a basis for pathogen acquisition and transmission (Clements, 1992; Raikhel, 2005). Therefore, deciphering the molecular basis of female reproduction is essential for understanding blood-feeding activated pathogen acquisition and transmission. In spite of significant achievements in research of molecular basis of female reproduction, many challenging questions remain. The mosquito fat body, which is a functional analog of the vertebrate liver and adipose tissue, is the organism's metabolic center playing a key role in female reproduction by producing yolk protein precursors (YPP) for developing oocytes in the process called vitellogenesis. Vitellogenin (Vg) is the main and most abundant YPP. In addition, the fat body produces other YPPs, including Vitellogenic cathepsin B (VCB), Vitellogenic Carboxypeptidase (VCP), and Lipophorin (Lp) (Attardo et al., 2005, Raikhel, 2005 and Raikhel et al., 2005). YPPs are secreted into the hemolymph and transported to developing ovaries. In the ovary, Vg binds to its cognate receptor, vitellogenin receptor (VgR), located on the plasma membrane of developing oocytes and is internalized in the process, called receptor-mediated endocytosis. YPPs are deposited in yolk bodies for the future use by developing embryos. Overall, all these YPPs are vital for the development of embryos.

A considerable body of knowledge has been accumulated concerning regulation of vitellogenesis in female mosquitoes (Raikhel, 2005; Raikhel et al., 2005; Roy et al., 2016). After a blood meal, an elevated titer of amino acids (AA) leads to the activation of vitellogenesis providing a link of reproduction with hematophagous blood feeding (Attardo et al., 2005). Ingested blood also stimulates the secretion of an uncharacterized signal from the midgut to the brain. The neuropeptide hormone, ovary ecdysteroidogenic hormone (OEH), from the brain activates the ovary to produce ecdysone, a precursor form of an active steroid hormone (Roy et al., 2016). In target tissues, including the fat body, ecdysone is transformed to a functional form, 20-hydroxyecdysone (20E). One of the key 20E functions is to activate transcription of genes encoding precursors for sex-specific yolk proteins (YPPs) in the female mosquito fat body. The 20E titer reaches its maximum at 18 h post blood meal (PBM), while transcription of *YPP* genes peak at 24-30 h PBM. *YPP* gene expression and synthesis of YPPs comes to the end by 36 h PBM, after which the fat body undergo remodeling for a second gonadotrophic cycle.

The regulatory pathway controlling vitellogenesis has been characterized at the molecular level (Raikhel, 2005; Raikhel et al., 2005; Roy et al., 2016). 20E binds to its receptor, a member of steroid receptor superfamily called Ecdysone receptor (EcR). EcR is active only as a dimer with its partner Ultraspiracle (USP), which is a homolog of the vertebrate retinoid X receptor (RXR). Upon binding to 20E, EcR/USP interacts with its recognition site, EcRE, and activates transcription of so-called early genes, E74, Broad and E75. There are several isoforms of all these transcription factors in the mosquito that play different roles in regulation of vitellogenesis. However, their precise roles are not

known at present. Therefore, further studies of transcription regulatory network in mosquito reproduction are required. Besides, a genome-wide transcriptome study performed in the *Ae. aegypti* mosquito fat body has identified four gene sets, sequentially expressed during the vitellogenic period (Roy et al., 2015). Moreover, hormones and other factors leading to either activation or repression of these gene clusters have been identified (Roy et al., 2015). However, very little is known about underlying regulatory network of transcription factors aiding the differential expression of genes. As the first step towards elucidating such networks we have employed bioinformatics tools for the identification of *cis*-regulatory factors in the promoters of differentially expressed gene sets in the mosquito fat body. After detecting putative TFBSs and their putative transcription factors (TFs), gene regulatory network construction by bioinformatics tools generated for understanding of gene regulation mechanisms.

1.2 Bioinformatics approaches for studying transcriptional and post-transcriptional gene regulation

1.2.1 Identification of putative transcription factor binding sites (TFBSs)

The goal of this study is to identify molecular regulatory networks in the vitellogenic fat body by analyzing the promoters of differentially expressed gene sets in this tissue. Recently, the transcriptome data analysis illustrated significant changes in gene expression during the PBM period, 3 to 72 h (Roy et al., 2015). Results showed that about 7500 transcripts were differentially expressed during this time period with four major sets, exhibiting peaks at 3-12, 18-24, 36 and 48-72 h PBM, respectively. The study has identified the main regulatory factors that are responsible for this sequential patterns of differential gene expression. The 3-12h early gene set is activated by AAs but repressed by 20E; the 18-24h early-mid set is activated by 20E and insulin but repressed by the nuclear receptor HR3; the 36 h late-mid set is repressed by 20E and activated by HR3 and repressed again by juvenile hormone (JH); finally, the 48-72h late gene set is repressed by HR3 and activated by JH (Roy et al., 2015). The goal of this study is to advance further our understanding of regulatory networks underlying this complex pattern of gene expression. Delineation of transcription factor binding sites (TFBSs) of the differentially expressed genes should lead to identification of corresponding transcription factors involved in the regulatory complexity of vitellogenic events in the fat body. The transcription factor binding sites (TFBSs) in most cases are short, 6-10 bp

long sequences with degenerate DNA patterns that represent binding preferences for specific transcription factors (Shlyueva et al., 2014; Tompa et al., 2005).

1.2.1.1 Over-represented TFBS motifs

Since, a co-expressed gene cohort is often regulated by the same factors, there is a high chance of occurrence of similar regulatory patterns in the promoter sequences of the co-expressed genes. The identification of these regulatory elements will be helpful to identify the regulatory networks (Stormo et al., 2010). In computationally driven TFBS prediction, most algorithms follow detection of over-represented motifs (Shlyueva et al., 2014). Because the over-representation of the putative motifs does not guarantee the identification of real TFBSs, we need several steps of filtration processes, such as detection of 1) positional 2) orientation bias and 3) Evolutionary conservation. The high confidence motifs or the putative TFBSs can later be functionally validated by several different molecular techniques.

1.2.1.2 Positional and orientation bias of TFBSs

Authentic TFBSs display likeability for a particular region within promoter sequences (Bellora et al., 2007). It is probably because the transcription machinery associated requires to be at a certain distance from transcription starting sites for regulation of the co-expressed genes. Therefore, measuring the positional bias is a meaningful and efficient evaluation for filtration of the over-represented motifs. Positional bias can be measured by dividing the observed value with statistical expectation (random chance value) (Pan et al., 2004; Smith et al., 2007).

Orientation bias of putative TFBSs is another important trait. Some of the known TFBSs are oriented in a particular direction. Some regulatory motifs have been observed to be functional only when arranged or positioned on one strand, but not so when located on the complementary strand (Elemento et al., 2007).

1.2.1.3 Evolutionary conservation of TFBSs

The evolutionary rate of functional TFBSs in genome sequences is slower than non-functional elements because of the selective pressure to preserve important functions in an organism (Tompa, 2001). Comparative sequence analysis leads to the identification of conserved TFBSs in regulatory regions (Tompa, 2001). Comparison of orthologous sequences can detect evolutionary conserved TFBSs through cross-species searches (Tompa, 2001). The detected well-conserved sites can be excellent candidates for identification of authentic TFBSs (Tompa, 2001; Tagle et al. 1988; Duret and Bucher 1997).

1.2.1.4 TFBSs database

As gene expressions are determined by the binding of transcription factors to their specific TFBSs, accurate model information for the binding preferences of a group of known transcription factors is needed (Mathelier et al., 2014). Although the putative TFBSs would satisfy the criteria of authentic TFBSs within bioinformatics methodologies, those binding sites need to be assessed by comparison with real TFBSs that have been experimentally proven. Because of the development of information retrieval technologies, it is possible to access efficiently, high quality and reliable knowledge gained by studying archives such as TRANSFAC and JASPAR. These

databases maintain the information of experimentally confirmed binding sites, and usually, this information can be used for detection (Shlyueva et al., 2014) of real TFBSs. TRNASFAC was initially released in 1988, and now this database has changed to a "commercial use only" database except for an old version, most of the data has been transferred to BIOBASE company. In comparison, the JASPAR database was originally developed as an "Open-access database for eukaryotic transcription factor binding profiles" (Sandelin et al., 2004) and still maintains the open access status.

1.2.2 Identifying the role of microRNAs in gene regulation

1.2.2.1 microRNA biogenesis

MicroRNAs (miRNAs) are endogenous small non-coding RNAs (20-24 nt) regulating gene expression at the post-transcriptional level by targeting mRNAs (Lucas et al., 2015; Ha and Kim, 2014; Bartel, 2004). The discovery of miRNAs has led to a significant progress in understanding the complexity of gene regulatory networks (Lucas et al., 2013). Pre-miRNA, a transcript with hairpin structure from miRNA genes of RNA Polymerase II (Pol II), is processed by Drosha protein for conformation of pre-miRNA (Lucas et al., 2015; Ha and Kim, 2014). The pre-miRNA is transported by Exportin-5 from the nucleus to the cytoplasm. The Dicer-1 processes the exported pre-miRNA to form miRNA:miRNA duplex. This duplex is needed for both of steps for RNA-induced silencing complex (RISC) assembly: 1) loading the duplex onto Argonaute (Ago) protein, and 2) unwinding the passenger strand of the miRNA duplex to mature miRNA (Lucas et al., 2015; Ha and Kim, 2014). The mature miRNA-RISC complex (miRISC) binds to the

mRNA 3' untranslated region (UTR) pairing with a complementary seed sequence and directing mRNA for translation inhibition or degradation (Lucas et al., 2015; Bartel, 2004). Which leads to the question - How a miRNA can recognize a particular target site within the 3'UTR of mRNAs? This is a critical issue while predicting miRNA targets.

1.2.2.2 miRNA target prediction

The identification of the miRNA targets is a major step in understanding miRNA functionality in genetic regulatory networks. At early stages of target recognition, a complementarity observed between the 5' portion of miRNA and the 3' UTR region was suggested as a focus area for detecting target sites (Bartel, 2004; Lee et al., 1993). Lee et al. (1993) reported lin-4 miRNA and its complementary target sites in the 3' UTR region of lin-14. The "lin-14" mRNA was verified by traditional molecular biology techniques to be an authentic target. After this work, on "lin-14" mRNA our understanding of miRNA target prediction has been enhanced. Because of the laborious procedures for validating targets, effective computational target prediction methods are frequently applied in miRNA studies. The following are the primary criteria for reliable target prediction (Lucas et al., 2013; Min et al., 2010):

1) Seed region matching. The seed region is about 7-8 base lengths from the 5' end of miRNA.

2) Thermodynamic stability checks of the miRNA-mRNA binding site.

3) Evolutionary conservation through cross-species comparison.

Although many algorithms and tools have been developed based on these criteria, the computational prediction results have shown a high rate of false positives (Lucas et al., 2015). The adjustment of thresholds in algorithms is another barrier for true positive target prediction. The low stringency of a threshold of the algorithms leads to an increase in false positive prediction results, such as seed sequence matching with an allowance of mismatch or free energy calculation in miRNA-mRNA hybridization. Therefore, optimization of parameters is of utmost importance. Our primary goal here was to develop an approach for miRNA target prediction that would reduce the false positives. A step-wise approach is quite an efficient solution (Watanebe et al., 2007). Here we developed a multiple-algorithm step-wise approach for robust target predictions.

1.3 Dissertation Objectives and Aims

The purpose of this study is to identify important factors within the regulatory networks involved in female reproduction in the fat body of *Ae. aegypti* mosquito.

Aim 1. At the transcriptional level, I will identify putative *cis*-regulatory elements within the promoters of differentially expressed genes. This study will be based on the existing differentially expressed microarray dataset of the female mosquito fat body after blood feeding from 3h to 72h.

Aim 2. Construction of putative regulatory networks using the elements detected from Aim 1 and functional validation of some of these factors.

Aim 3. I will focus on the most critical step of miRNA studies i.e. target identification, which is essential for understanding of the functionality of miRNAs at the post-transcriptional level. Although multiple tools are available for miRNA target an efficient method for the reduction of false-positive results is required. Our goal is to establish a method for robust predictions and test some of the targets for miRNAs involved in mosquito reproduction.

1.4 References

Attardo GM, Hansen IA, Raikhel AS. Nutritional regulation of vitellogenesis in mosquitoes: implications for anautogeny. Insect Biochem Mol Biol. 2005 Jul;35(7):661-75. Epub 2005 Mar 28. Review.

Raikhel, AS, M. Brown and Belles, X. 2005. Endocrine Control of Reproductive Processes. In: Comprehensive Molecular Insect Science (L Gilbert, S Gill and K. Iatrou eds.) Vol. 3, Endocrinology, pp. 433-491. Elsevier Press.

Raikhel, AS, 2005. Vitellogenesis of disease vectors, from physiology to genes. In: Marquardt, W.C. et al., (Ed.), Biology of Disease Vectors. Elsevier Academic Press, Burlington, San Diego, London, pp. 329–346.

Clements, AN, The Biology of Mosquitoes. CABI 1st Ed. Jan. 1992

Roy S, Saha, TT, Johnson L, Zhao B, Ha J, White KP, Girke T, Zou Z, & Raikhel AS, Regulation of Gene Expression Patterns in Mosquito Reproduction PLoS Genetics 2015 11(8):e1005450

Roy S, Smykal V, Johnson L, Saha TT, Zou Z, Raikhel AS, Regulation of Reproductive Processes in Female Mosquitoes Chapter *in* Advances in insect physiology July 2016 DOI: 10.1016/bs.aiip.2016.05.004

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, Myers MF, George DB, Jaenisch T, Wint GR, Simmons CP, Scott TW, Farrar JJ, Hay SI. Nature. The global distribution and burden of dengue. 2013 Apr 25;496(7446):504-7. doi: 10.1038/nature12060. Epub 2013 Apr 7.

Goupil BA, McNulty MA, Martin MJ, McCracken MK, Christofferson RC, Mores CN. Novel Lesions of Bones and Joints Associated with Chikungunya Virus Infection in Two Mouse Models of Disease: New Insights into Disease Pathogenesis. PLoS One. 2016 May 16;11(5):e0155243. doi: 10.1371/journal.pone.0155243. eCollection 2016.

Pan Y, Smith B, Fang H, Famili FA, Sikorska M, Walker R: Selection of putative cis-regulatory motifs through regional and global conservation. Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference (CSB 2004), August 16–19, 2004. 2004, Stanford, CA, USA, IEEE Computer Society, 684-685.

WHO, World Malaria Report 2014, *WHO Global Malaria Programme*.

WHO website: www.who.int/emergencies/zika-virus/situation-report/19-may-2016/en/

CDC Zika virus website: https://www.cdc.gov/zika/about/index.html

Smith B, Fang H, Pan Y, Walker PR, Famili AF and Sikorska M, Evolution of motif variants and positional bias of the cyclic-AMP response element BMC Evolutionary Biology 2007 7(Suppl 1):S15 DOI: 10.1186/1471-2148-7-S1-S15

Shlyueva D, Stampfel G and Stark A, Transcriptional enhancers: from properties to genome-wide predictions, *Nature Reviews Genetics* **15**, 272–286 (2014) doi:10.1038/nrg3682

Stormo GD and Zhao Y, Determining the specificity of protein–DNA interactions, *Nature Reviews Genetics* **11**, 751-760 (November 2010) | doi:10.1038/nrg2845

Bellora N, Farré D and Mar Albà M, Positional bias of general and tissue-specific regulatory motifs in mouse gene promoters, BMC Genomics 2007 Dec. 13;8:459

Tompa M, Identifying Functional Elements by Comparative DNA equence Analysis, Genome Research 2001 11:1143-1144

Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D91-4.

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, Lim J, Shyr C, Tan G, Zhou M, Lenhard B, Sandelin A, Wasserman WW. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res. 2014 Jan;42(Database issue):D142-7. doi: 10.1093/nar/gkt997. Epub 2013 Nov 4.

Ha M, Kim VN, Regulation of microRNA biogenesis, Nature Reviews 2014 Vol 15 509-524

Lucas KJ, Myles KM, Raikhel AS, Small RNAs: a new frontier in mosquito biology, Trends in Parasitology 2013 June Vol. 29 No. 6 295-303

Chapter II

Identification of binding sites for transcription factors for differential expression of genes during the vitellogenic period of the female yellow fever mosquito,

Aedes aegypti

2.1 Abstract

The yellow fever mosquito, *Aedes aegypti*, is one of the primary vectors that cause devastating vector-borne diseases such as, Zika, Dengue, and Yellow fever. Female *Aedes aegypti* mosquitoes spread diseases while feeding on vertebrate blood. Blood feeding is necessary for vitellogenesis, the process of yolk formation during the maturation of eggs, and is an important event in the mosquito reproductive cycle. The aim of this chapter is to identify putative transcription factor binding sites (TFBSs) within the promoters of the co-expressed genes that are differentially regulated during different time points in vitellogenesis, using bioinformatic tools. Microarray data generated with tissues obtained from fat bodies of female mosquitoes have shown that clusters of genes are up- and down-regulated at different time points post blood meal (PBM) (Roy et al., 2015). We used the microarray data to identify the genes that are differentially expressed between 6 hours and 72 hours PBM. The promoters (2kb regions upstream of the translation start sites) of these co-expressed genes were used for identification of the *cis-regulatory* elements. Previous studies indicated that the vitellogenesis phase is under the influence of the steroid hormone 20-hydroxyecdysone (20E). Very little is known about the molecular mechanisms involved in 20E related regulation of downstream genes. Therefore, a critical step in understanding the regulatory factors governing the *Ae*.

aegypti genome during this period is to investigate the TFBSs within the promoters of these genes and the transcription factors (TFs) related to these TFBSs. We studied the primary features of putative TFBSs, such as positional and orientation bias and evolutionary conservation, as criteria for reliable predictions. We identified 89 putative cluster-specific TFBSs and 34 related TFs from the bioinformatics tests. We also checked for tissue-specificity of these factors.

2.2 Introduction

Gene expression through activation or inhibition is determined by transcription factors recognizing binding sites in promoter sequences of the gene (Zhang, 2015). Thus, identifying the binding sites can lead to identification of their corresponding TFs and this can be a significant step towards understanding the mechanism behind a particular gene expression. While many scientific investigations have tried to understand the secrets of transcriptional initiation for deciphering gene regulatory mechanisms, gene regulation is still not well understood (Zhang, 2015). Promoter mutagenesis, which is the classical molecular strategy for identifying regulatory motifs within gene promoters is laborious and not suited to genome-wide application, especially in non-model systems (Roy et al. 2013). Of late, the chromatin immunoprecipitation (ChIP) method for TFs and specific antibodies on microarray hybridization, or the massively parallel DNA sequencing methods are being used as preferred strategies for identifying binding sites with associated proteins (Rister et al., 2010). But the ChIP method is not easy as one needs specific antibodies and purified proteins for TFs for the ChIP experiments (Zhang, 2015). Also, these tests require high-cost, are time-consuming, and consists of laborious technical steps. Computational motif prediction methods on the other hand have become promising and practical alternate strategies, with the enormous advancement in the fields of biostatistics and bioinformatics.

Many computational tools have been developed for the prediction of transcription factor binding sites (Tompa, 2005). Each of these, run a different algorithm in the background. It can be assumed that a particular motif pattern has more chances of being a

real binding site if multiple prediction tools have identified it. Prediction using multiple tools can lead to a reduction in false positives. Usually, methods that include enumerative search, expectation maximization, or Gibbs Sampling algorithms are employed to search the promoters of co-regulated genes, for the identification of over-represented motifs (Sinha and Tompa, 2003; Liu et al., 2001; Tompa et al., 2005; Bailey et al., 2009). For this study, we have used Multiple EM for Motif Extraction (MEME) that employs Expectation-Maximization algorithm (Bailey et al., 2009), and Weeder that uses an exhaustive searching method (Pavesi et al., 2004), to detect overrepresented motifs or putative TFBSs.

The interaction between transcription factors and their binding sites in the promoter regions control transcriptional initiation for gene expression. A certain distance from the transcription start sites is needed for each transcription factor to be able to construct regulatory complexes such as cis-regulatory modules (CRM), in which transcription factors are coupled with and their specifically arranged binding sites (Roy et al., 2013; Bellora et al., 2007; Elemento et al., 2007). These specifically arranged or positionally biased motif patterns are one of the major aspects for the identification of real transcription factor binding sites.

In addition, some of the transcription factors are oriented in a particular direction relatively more compared to another direction in double strand DNA. Many functional regulatory motifs are arranged on one directional strand as opposed to the other strand (Elemento et al., 2007). The evaluation of positional and orientation bias for motifs depends on the observed occurrences of these motifs when compared to random chance

(expectation) in background genomic DNA sequences. Meanwhile, evolutionary conservation of regulatory sequences is based on phylogenetic footprinting involving orthologs from other species. The idea of checking evolutionary conservation is motivated by the fact that regulation mechanisms and binding sites are preserved in a proper evolutionary distance with the exception of inordinate background conservation (Wang et al., 2003). In other words, an important transcription factor binding site is preserved in other species within a close phylogenetic distance. The evolutionary conservation for transcription factor binding sites should lead to the detection of real transcription factor binding sites.

Information related to real transcription factor binding sites has been stored in databases, such as TRANSFAC or JASPAR. These databases are archives for real transcription factor binding sites with corresponding transcription factors, represented as position weighted matrices based on biological experiments (Wingender et al., 1996; Sandelin et al., 2004). Therefore, it is meaningful to compare computationally detected motif patterns with real transcription factor binding sites in these databases.

In this chapter we have used the above mentioned bioinformatics strategies to identify high-confidence putative TFBSs, that we believe along with their corresponding TFs are involved in regulation of gene expression, during the 72 hr reproductive period in female mosquitoes, post blood meal. The vitellogenic phase of reproduction is under the control of the insect steroid hormone, 20-hydroxyecdysone (20-E). After blood feeding, drastic changes occur within the *Ae. agypti* female mosquito fat body, an important reproductive organ, where vitellogenesis takes place (Raikhel et al, 2005; Raikhel et al.,

18

2002). Recently, a large-scale microarray study related to gene expression during the post blood meal (PBM) period was published (Roy et al., 2015). This data illustrated the changes in gene expression patterns between 6 hr and 72 hr PBM. We have used the same microarray data to select genes that are differentially expressed during different time points PBM and searched their promoters to identify putative *cis*-regulatory elements in order to get an insight into the regulatory mechanism of these differentially regulated, co-expressed gene sets.

2.3 Materials and methods

2.3.1 Selection of genes based on microarray data:

Approximately 200 hundred genes were selected from each of 7 different clusters; All genes in each cluster were sorted by their logFC value, top 200 genes matching the average expression profile of each cluster were selected based on the microarray the data published by Roy et al. in 2015.

2.3.2 Construction of promoter datasets:

Regions 2kb upstream of the translation start sites (ATG) of all selected genes were extracted using an in house R script. The sequences were extracted from vectorbase (https://www.vectorbase.org/), aaegypti.SUPERCONTIGS-Liverpool.AaegL1.fa which is a downloadable file for *Ae. aegypti* genome sequence (gene build 1.3), was the source for the sequences in FASTA format. Information regarding the start sites within the gene feature file (aaegypti.BASEFEATURES_Liverpool-AaegL1.3.gff3), was used for extraction of the required upstream regions. Similarly, *Anopheles gambiae* (genebuild 3.7) and *Culex quinquefasiatus* (gene build 1.3), upstream sequences were also extracted for checking evolutionary conservation from the same database.

2.3.3 Identification of Over-represented motifs:

Two different motif detection programs, MEME (Bailey et al., 2009) and Weeder (Pavesi et al., 2004) that use different algorithms as mentioned above were used to identify the over-represented motifs. These two motif detection programs were installed

in Biocluster maintained by Bioinformatics core facility of University of California, Riverside (UCR). MEME (Multiple EM for Motif Extraction) is operated by Expectation Maximization algorithm whereas Weeder employs an exhaustive searching method. Two different algorithms were used to increase the accuracy and to decrease false positive rate of the putative motifs predictions. The input options about MEME and Weeder were set to detect 6-9 nucleotide length motif patterns. The parameters used were as follows:

Meme: -dna -mod anr -nmotifs 100 -minw 6 -maxw 9 -minsites 20 -maxsize 4000000 -maxiter 5 –revcomp

Weeder:

-R 50 -O AA -W 6 -e 1 -V -S -M -T 100 (6-mer)

-R 50 -O AA -W 8 -e 2 -V -S -M -T 100 (8-mer)

2.3.4 Test of motif enrichment within co-regulated gene clusters

    The motifs from two the different algorithms were selected by two different statistical enrichment tests, equality of proportion test and test for hypergeometric distribution. A two step strategy was used to detect the enrichment of these motifs in different clusters; the equality of proportion test was used confirm the overrepresentation of these motifs by comparing their occurrences in the selected promoter sequences to that in the promoter sequences of all protein-coding genes (background set). Then, the test for hypergeometric distribution was applied to check for the enrichment of these motifs in specific co-regulated gene clusters. The hypergeometric test calculated the statistical significance (p-value) of motif enrichment test in R statistical computing environment (R

21

Development Core Team, 2011). A heatmap was constructed using the P-values from the hypergeometric test between the number of motif occurrences in each cluster and that in all gene clusters.

2.3.5 Detection of Positional and orientation bias

The 2KB promoter regions were divided into ten bins of 200 bases pairs each. The frequency of motifs within each bin was calculated by scanning the bins in both f in forward and reverse directions. Test of equality of proportions was used to detect the region of positional bias for each motif.

Comparing the occurrences of each motif in forward and reverse strands checked the orientation bias. Chi-square test was used to check the statistical significance of the difference between the observed motifs in each strand.

2.3.6 Evolutionary conservation test

As the first step towards checking evolutionary conservation Reciprocal Best Blast was used to identify candidate orthologs of genes, the promoters of which carried the motifs. The reciprocal best blast or reciprocal best hit (RBH) is a method where BLASTP (protein-protein BLAST) search technique is used reciprocally for more reliable orthologous gene search.

The RBH method is more stringent and effective in a selection of conserved orthologs with low false positive rate compared to single best hit (non-reciprocal BLAST search) (Xu & Gogarten, 2008). Once the orthologs were identified in *An. gambiae* and

Cu. quinquefasiatus, their promoter sequences were extracted and aligned with MUSCLE

(Edgar, 2004) to check for evolutionary conservation of the motif.

If a motif, which appeared in the *Ae. aegypti* gene promoter, also appeared within

+/- 100 bps of the said position in the orthologus promoter sequence of either *Anopheles*

gambiae and/or *Culex quinquefasciatus*, the motif we considered it to be conserved (Fig.

2.7.2). Up to five promoters were typically checked, one base degeneracy was allowed

while checking for the conservation.

2.3.7 JASPAR motif database search

JASPAR database is a widely used open-access database maintaining sets of

experimentally proven motif patterns with a matrix-profiled model for DNA binding

preferences, TFs and TFBSs (Bryne et al., 2008). The JASPAR database has many sub-

categories for the motifs information such as the vertebrate, fungi, and insects. The

JASPAR Insecta database was used for our analysis. The over-represented motifs were

queried against the JASPAR Insecta database motifs by using TOMTOM motif

comparison tool at the p-value cutoff of 0.01 (Gupta et al., 2007). The default parameters

of TOMTOM were used for the comparisons.

2.3.8 Determination of tissue specificity:

FlyAtals 2 database (Robinson et al., 2013) was used to check for tissue-specific expression of the transcription factors identified by JASPAR, in *Drosophila*. The transcription factors that were detected by JASPAR with a p-value of < 0.05 and were expressed in the fat body of Drosophila were selected.

2.4 Results

2.4.1 Construction of promoter sequence dataset from the differentially expressed genes:

Roy et al., in a study published in 2015, described 4 sets of genes that are differentially expressed during the 72 h post blood meal period in the fat body of female mosquitoes. We used the published gene expression data (Roy et al., 2015) to select co-expressed genes. We observed that there were subtle variations in gene expression patterns within the early, early-mid and late gene sets. Roy et al., had clubbed certain clusters from the original analyses to come up with the said gene sets. Since, these sets had huge number of genes we went with the original clustering and also divided the largest cluster (cluster 2) into two different clusters based on expression profiles of the genes. Therefore, we had seven gene clusters viz., Clusters 1 and 6 (Early gene clusters), Clusters 4 and 7 (Early-mid gene clusters), Cluster 5 (Late-mid cluster) and Clusters 2A and 2B (Late gene cluster divided into two).

Approximately 200 genes were selected on the basis to their gene expression values ranked according to the log fold change (logFC) and similarity to the average expression profile of a particular cluster (Table 2.7.1). The promoter sequences of *Aedes aegypti* genome sequences i.e. 2Kb upstream from the translation starting site (ATG) were extracted using R programming scripts using information of the start position mentioned in the GFF3 files.

Thus, from this step we had seven datasets with approximately 200 promoters in each set and each promoter with a maximum of 2000 base pairs.

2.4.2 Identification of cluster specific and over-represented motifs:

The promoter datasets generated in the previous step were used as input for the two motif finding programs. MEME was set to identify to 100 motifs for each set, from 6 – 9 base pairs in length. On the other hand Weeder identified top 50 motifs (6 and 8 bp). The intersection between the motifs generated from the two programs was checked and in some cases similar motifs were merged allowing a 2 bp degeneracy. Table 2.7.2 lists the final set of motifs commonly identified by both programs. Maximum number of motifs (22) was detected from Cluster 2B promoter dataset, whereas, a search of cluster 5 and Cluster 6 promoters resulted in identification of five and seven motifs respectively (Table 2.7.2). Each of these motifs, since some of these resulted from manually merging two motifs, were subjected to the test of equality of proportions against a randomly generated set of promoters. All of the motifs passed this test. Now to check for cluster specificity a hypergeometric test was applied. The results plotted with the p-values from the hypergeometric test as a heatmap (Figure 2.7.1) shows a clear demarcation between the motifs from each cluster. For example, the 13 motif patterns detected within the promoters of Cluster 4 genes were tested for over-representation to six other clusters. The results show while that one of the 13 motif patterns were over-represented in any other cluster (Figure 2.7.1).

2.4.3 Positional and Orientation bias:

In order to check whether a motif shows positional bias, the frequencies for occurrences of a motif within each of the ten 200 bp bins (that the 2Kb promoters were broken into) was checked. Based on the raw frequency values, the statistical significance of overrepresentation in one or more bins were checked with the test for equality of proportions when compared to detection by random chance within a similar bin. The table (Table 2.7.5) shows the raw frequency of the over-represented motifs in 5' to 3' and 3' to 5' directions along with the p-values from the statistical tests. The frequencies under the p-value cut off of < 0.01 were selected to be positionally biased. The results of positional bias analysis are shown in the "Pos." column in the Fig. (2.7.1). About 45% of the over represented motifs were found to display positional bias for one or more regions within the promoters.

For checking orientation bias the frequencies of motifs observed in the forward (5' to 3' direction) and the reverse directions (3' to 5') were calculated. These values were then subjected to Chi-square test to check if the differences were statistically significant. The results of the orientation bias are shown in the "Orient." column of Fig. 2.7.1 and in Table 2.7.6 suggests that about 15% of the motifs displayed orientation bias. It was also observed that most of the motifs (~75%) that showed positional bias were found display orientation bias.

2.4.4 Evolutionary conservation

First, genes in which the motifs appeared within the region of positional bias were identified for each motif. Their orthologs were identified in *An. gambiae* and *Cu. quinquefasciatus*. 2kb upstream regions were extracted and were aligned with the multiple sequence alignment program MUSCLE.

A maximum of five genes and their orthologs were checked. A motif was considered to be evolutionarily conserved if the it appeared in the orthologous promoters within +/- 100 bp of position in which it is detected in the *Ae. aegypti* promoter. One base degeneracy was allowed for the identification. Eleven over-represented motifs were found to be evolutionarily conserved in the orthologous genes of either one or both of the mosquito species (Fig. 2.7.2; Table 2.7.4).

2.4.5 JASPAR database search results:

The TOMTOM motif comparison tool (http://meme-suite.org/tools/tomtom) was used to map detected overrepresented motifs onto known transcription factors in the JASPAR database (Gupta et al., 2007). The TOMTOM program confirmed similarities using p-value cut off lower than 0.01. The comparison options of the TOMTOM program were set with default options, and then a distance of two motifs was measured by Pearson correlation coefficient with E-value < 10. Lists of probable transcription factors were obtained from JASPAR search, for each cluster (Table 2.7.3). We used this list to check for tissue specificity.

2.4.6 Detection of tissue specificity:

The list of TFs obtained from JASPAR was used to check against the FlyAtlas 2 database (http://flyatlas.gla.ac.uk/flyatlas/index.html) for tissue specific expression of *Drosophila* TFs. The motivation behind this step was that multiple factors might recognize a particular site but not be active in a particular tissue. Therefore we went down the list of detected transcription factors and indented those expressed in the fat body. A p-value cut-off of 0.01 was used for this purpose. Once the tissue –specificity was determined a final list of TFs were obtained (Fig 2.7.1). The 15 motifs detected from Cluster 1 were mapped to 14 distinct transcription factors that were expressed in the *Drosophila* fat body, whereas the 5 detected from Cluster 6, mapped to five unique TFs. 89 overrepresented motifs were mapped 34 known transcription factors that were expressed in the *Drosophila* fat body. These TFs were used to construct putative cluster specific regulatory networks as described in the next chapter.

2.5 Discussion

The event of transcription factors binding to their corresponding binding sites, is one of the primary biological processes determining gene expression. A clear knowledge of the TFs and TFBSs binding sites can lead to the molecular mechanism behind the expression of each gene. To date, scientists have been trying to understand the mechanisms with the help of various methodologies such as DNA affinity chromatography pull-down assay and mass-spectrometry (Chodosh et al., 2001; Drewett et al., 2001; Deng et al., 2003; Wu et al., 2006). However, these tests are high-cost, time-consuming and laborious tasks. Recently, ChIP-assay, ChIP-chip, and ChIP-seq methods have come into widespread use (Furey et al., 2013; Landt et al., 2012; Ma et al., 2011). Unfortunately, these methods can only be applied restrictively and require the use of a specific antibody. Since molecular biology methods have had their drawbacks, a new research paradigm was needed for a break through. Reliable bioinformatics prediction of TFBSs with functional assessment through molecular biology methods can lead to relatively fast and accurate assessment of regulatory mechanisms.

Various TFBS prediction programs are readily available, but the predictions cannot be completely trusted due to the percentage of false positive results. To minimize these errors, it is wise to employ at least two different algorithm driven programs. In this project, the use of two different algorithm based motif detection programs, MEME and Weeder has contributed to increased accuracy and reliability in the detection of over-represented motifs. All overrepresented motifs may not be transcription factor binding sites, therefore we have employed three filtration steps, where we looked for

characteristics related to real TFBSs. Apart from checking for positional and orientation bias we also checked for evolutionary conservation with a stringent methodology. Evolutionary conservation should be tested within phylogenetically close genome sequences. We are extremely lucky that two other mosquito species have already been sequenced, and the sequences are available.

The JASPAR database when searched provides with a list of probable TFs that can bind to a specific motif, but it is important to set a criterion for detection of the mostly likely ones, the top most hit is always the best option. The gene or the promoter sequences do not change over the course of different stages in an organism's life cycle, neither does it change with respect to tissues. Therefore, it is important to realize that there might be multiple regulatory factors within a promoter. Similarly, one particular element can be detected by different TFs at different time points and also in different tissues. This was the reason behind checking the tissue-specificity of the TFs detected by JASPAR.

Although the multistep approach that we have used here (Fig. 2.7.3) for identifying the putative TFBSs should deliver good initial results, it is still possible that some of the predicted candidate motifs may be false positives. It is also possible that real TFBSs if these are not over-represented, might have not been detected. In spite of these two drawbacks of the computational motif discovery scheme, I am hopeful that the results, in general, from the suggested approach would mostly be real and have significant practical application. Use of two different algorithms along with three different computational screening methods should lead to robust TFBSs prediction. I

have observed from separate RNA-seq analyses that the annotations for the translation start sites in Vectorbase are not always accurate to counter this problem I have used 2Kb upstream regions (Bonizzoni et al., 2012).

This chapter described the bioinformatics prediction for putative transcription factor binding sites for seven different clusters of co-expressed genes. These are used in the following section for network construction and the biological evaluation.

2.6 References

Roy S, Saha TT, Johnson L, Zhao B, Ha J, White KP, Girke T, Zou Z, & Raikhel AS, Regulation of Gene Expression Patterns in Mosquito Reproduction PLoS Genetics 2015 11(8):e1005450

Roy S, Kagda M, & Judelson HS (2013) Genome-wide Prediction and Functional Validation of Promoter Motifs Regulating Gene Expression in Spore and Infection Stages of Phytophthora infastans, *PLoS Pathogens* 9, (Mar), 1-18

Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, Butler R, Campbell KS, Christophides GK, Christley S, Dialynas E, Hammond M, Hill CA, Konopinski N, Lobo NF, MacCallum RM, Madey G, Megy K, Meyer J, Redmond S, Severson DW, Stinson EO, Topalis P, Birney E, Gelbart WM, Kafatos FC, Louis C, Collins FH, VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Research*. 2009;37:D583-D587.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Liw W, & Noble WS, 2009. Meme suite: tools for motif discovery and searching. *Nucleic Acids Research* 37, Web server issue (Jul), 202-208

Raikhel AS, Kokoza VA, Zhu J, Martin D, Wang SF, Li C, Sun G, Ahmed A, Dittmer N, Attardo G, Molecular biology of mosquito vitellogenesis: from basic studies to genetic engineering of antipathogen immunity, *Insect Biochemistry and Molecular Biology*, Vol. 32, Issue 10, Oct. 2002, 1275–1286

Raikhel AS, M. Brown and Belles X, 2005. *Endocrine Control of Reproductive Processes. In: Comprehensive Molecular Insect Science* (L Gilbert, S Gill and K. Iatrou eds.) Vol. 3, Endocrinology, pp. 433-491. Elsevier Press.

Raikhel AS, 2005. Vitellogenesis of disease vectors, from physiology to genes. In: Marquardt, W.C. et al., (Ed.), Biology of Disease Vectors. Elsevier Academic Press, Burlington, San Diego, London, pp. 329–346.

Raikhel AS, Kokoza VA, Zhu J, Martin D, Wang SF, Li C, Sun G, Ahmed A, Dittmer N, Attardo G. Molecular biology of mosquito vitellogenesis: from basic studies to genetic engineering of antipathogen immunity. Insect Biochem Mol Biol. 2002 Oct;32(10):1275-86.

Sinha S, Tompa M. YMF: A program for discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res. 2003 Jul 1;31(13):3586-8.

Liu X, Brutlag DL, Liu JS. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Pac Symp Biocomput. 2001:127-38.

Pavesi, G., Mereghetti, P., Mauri, G., & Pesole, G., Weeder Web : discovery of transcription factor binding sites in a set of sequences from co-regulated genes, *Nucleic Acids Research*. Vol. 32. 2004

Bellora N, Farre D, Alba MM (2007) Positional bias of general and tissue-specific regulatory motifs in mouse gene promoters. *BMC Genomics* 8: 459.

Elemento O., Slonim N, Tavazoie S. (2007) A universal framework for regulatory element discovery across all genomes and data types. *Molec Cell* 28: 337–350.

Hu, X., Cherbas, L., & Chrebas, P., 2003. Transcription activation by the Ecdysone receptor (EcR/USP): Identification of activation functions. *Molecular Endocinology* 17: 716-731

Frith, M. C., Fu, Y., Yu, L., Chen, J., Hansen, U., & Weng, Z., 2004. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acid Research* 32: 1372-1381

Ma W, & Wong WH, 2011. The analysis of ChIP seq data. *Method in Enzymology* 497: 51-73

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K. I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A. J., Hoffman, M. M., Iyer, V. R., Jung, Y. L., Karmakar, S., Kellis, M., Kharchenko, P. V., Li, Q., Liu, T., Liu, X. S., Ma, L., Milosavljevic, A., Myers, R. M., Park, P. J., Pazin, M. J., Perry, M. D., Raha, D., Reddy, T. E. d Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J. A., Tolstorukov, M. Y., White, K. P., Xi, S., Farnham, P. J., Lieb, J. D., Wold, B. J., & Snyder, M., 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia, *Genome Research*, 22, 9 (Sep)1813-1831

Furey, T. S., 2012. ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nature Reviews Genetics*. 13 (Dec) 840-852

Xu, Y. & Gogarten, P., 2008 *Computational Methods for Understanding Bacterial and Archaeal Genomes*, Imperial College Press, pp 220

Bryne J.C., Valen E., Tang M.H., Marstrand T., Winther O., da Piedade I., Krogh A., Lenhard B., Sandelin A.JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update.*Nucleic Acids Res*. 2008 Jan;36(Database issue):D102-6.

Wu, K. K., 2006, Analysis of protein-DNA binding by streptavidin-agarose pulldown, Methods in Molecular Biology Volume 338, 2006, pp 281-290

Deng, W. G., Zhu, Y., Montero, A., & Wu, K. K., 2003, Quantitative analysis of binding of transcription factor complex to biotinylated DNA probe by a streptavidin-agarose pulldown assay, Analytical Biochemistry 323, pp12-18

Drewett, V., Molina, H., Millar, A., Muller, A., Hesler, F. V., & Shaw, P. E., 2001, DNA-bound transcription factor compelxes analysed by mass-spectrometry: binding of novel proteins to the human *c-fos* SRE and related sequences, Nuclear Acids Research, vol. 29, No 2, pp 479-487

Gupta, S., Stamatoyannopolous, J.A., Bailey, T., & Noble, W. S., 2007, Quantifying similarity between motifs, Genome Biology, **8**(2):R24.

Chodosh, L. A., 2001, Purification of DNA-binding proteins using biotin/streptavidin affinity systems, Current Protocols in Molecular Biology, Chapter 12:Unit 12.6

Lodish, H.,BerkA.,Matsudaira,P.,Kaiser,C.A.,Krieger,M.,ScottM.P.,Zipursky,L.,Darnell, J., Molecular Cell Biology 5th ed., WHFreeman and Company, New York, 2004.

Bonizzoni, M., Dunn, W. A., Campbell, C. L., Olson, K. E., Marinotti, O., James, A. A., 2012, Complex modulation of the Aedes aegypti transcriptome in response to dengue virus infection. PLoS One. 7(11)

Wingender, E., Dietze, P., Karas, H. & Knüppel, R. (January 1996), TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res*. **24** (1): 238–41. doi:10.1093/nar/24.1.238. PMC 145586. PMID 8594589.

Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W. & Lenhard, B., JASPAR: Open-access database for eukaryotic transcription factor binding profiles, *Nucleic Acids Res*. (2004) 32 (1) D91-D94

Rister, J. & Desplan, C., 2010, Deciphering the genome's regulatory code: The many languages of DNA, Bioessays 32 (5): 381-384

Zhang, Y., He, Y., Zheng, G. & Wei, C., MOST+: A de novo motif finding approach combining genomic sequence and heterogeneous genome-wide signatures, 2015, BMC Genomics 16 (7):

Robinson SW, Herzyk P, Dow JA, Leader DP. FlyAtlas: database of gene expression in the tissues of Drosophila melanogaster. Nucleic Acids Res. 2013 Jan;41(Database issue):D744-50. doi: 10.1093/nar/gks1141. Epub 2012 Nov 29.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004 Mar 19;32(5):1792-7. Print 2004.

2.7 Figures and tables

Figure 2.7.1

In the Fig. 2.7.1, the cluster specific over-represented motifs are in the left panel. The p-value from the statistical significance test is shown in various colors for better visual illustration. The deeper orange color indicates the more statistically over-represented motif. The cluster-specific motifs were selected with a p-value cut off under 0.01. These selected motifs from one cluster were tested against other clusters to confirm cluster specific motifs.

Figure 2.7.1
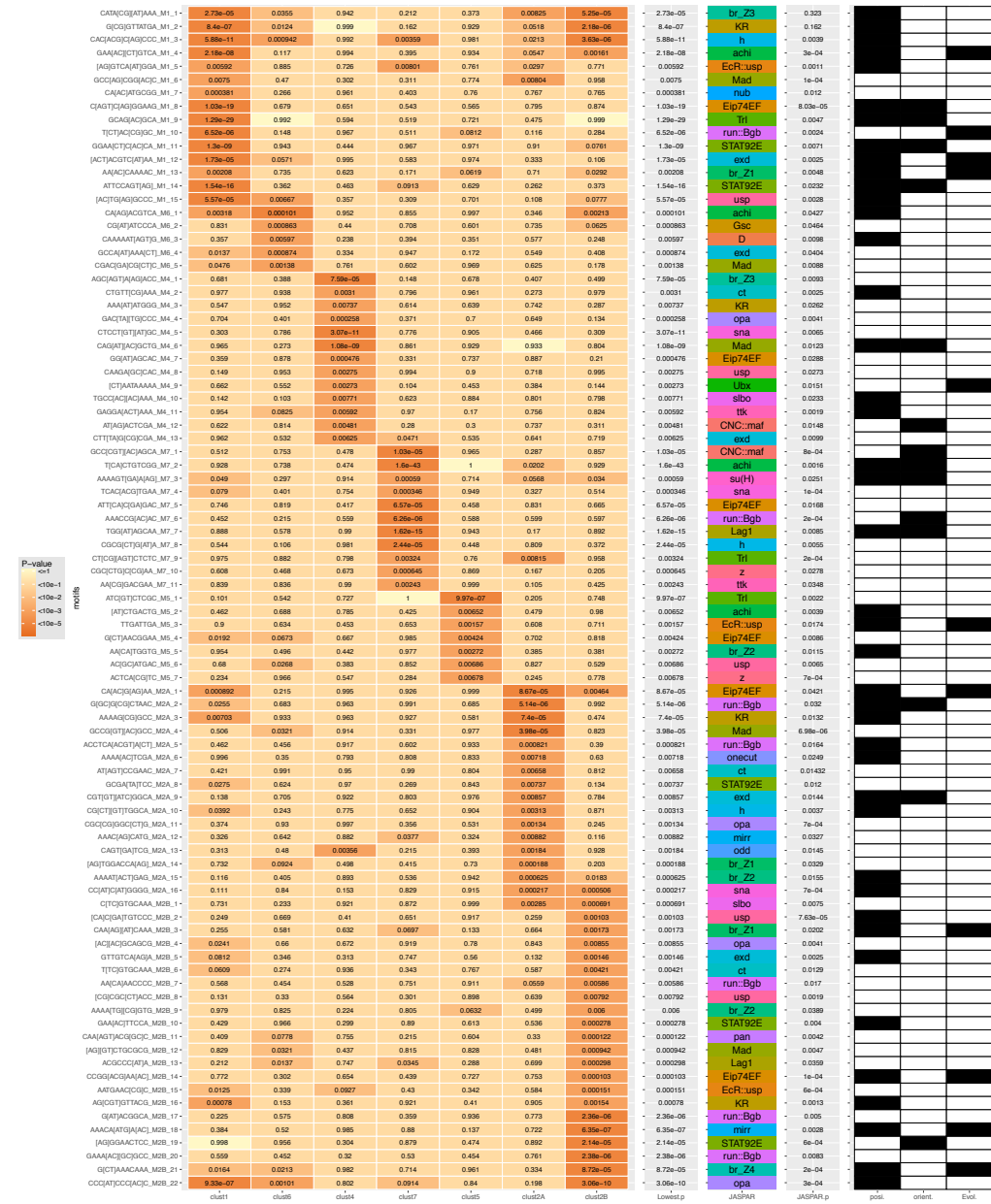
Figure 2.7.2

A schematic diagram of evolutionary conservation test. For example, a putative motif patterns AAAATAGT is detected in a promoter sequence of *Aedes aegypti*. One character mismatch allowed search in other two species promoter regions with +/- 100 position of detected *Aedes aeypti* promoter sequence. Also, two genes, AGAP005733-RA of *Anopheles gambiae* and CPIJ01579-RA of *Culex quinquefasciatus*, are orthologous with AAEL008182-RA.

Figure 2.7.2
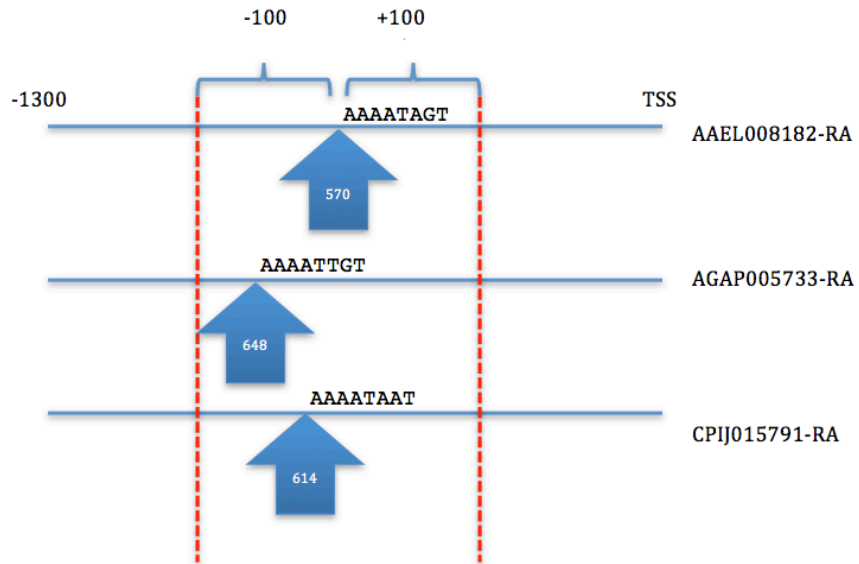
Figure 2.7.3

Overall motif detection pipeline flowchart: The microarray data was analyzed for

detecting cluster specific motifs using bioinformatics tools, MEME and Weeder. Then,

filtration bioinformatics techniques were applied for high confidence and tissue specific

motifs with TFs.

Figure 2.7.3
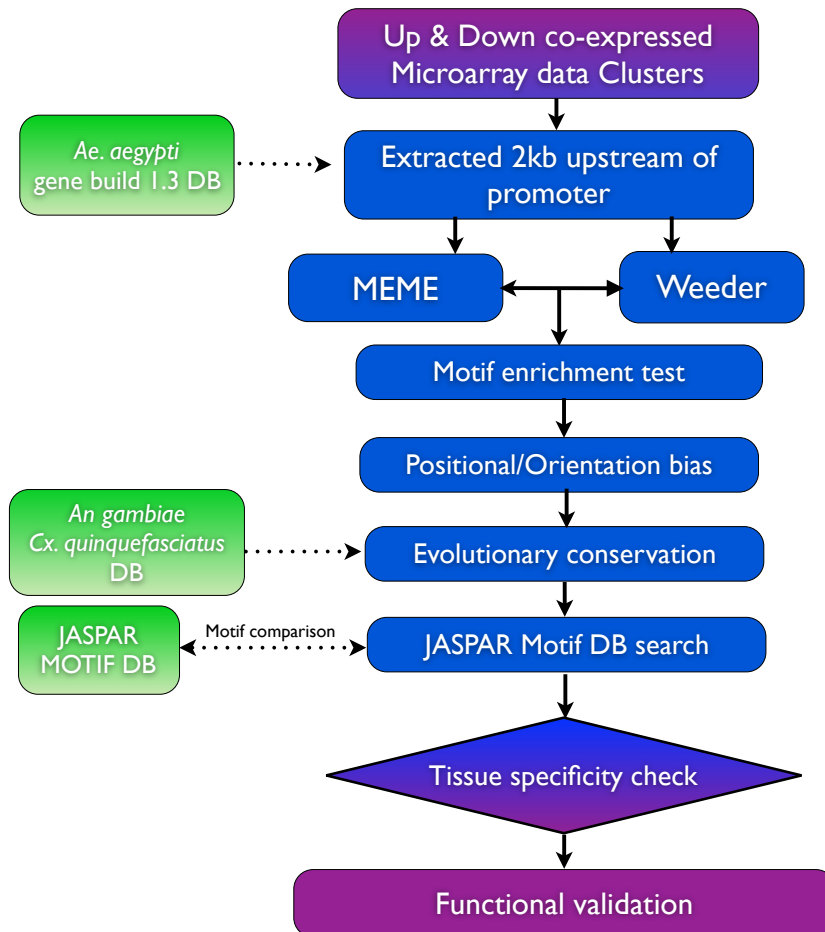
Table 2.7.1

The target genes of each cluster were sorted by the logFC value with the top 200 genes

being selected to analyze their promoter sequences.

Table 2.7.1 Target gene lists

Cluster 1	Cluster 6	Cluster 4	Cluster 7	Cluster 5	Cluster 2A	Cluster 2B
AAEL000088-RA	AAEL000208-RA	AAEL000038-RA	AAEL014284-RA	AAEL000037-RA	AAEL003944-RA	AAEL000076-RA
AAEL000127-RA	AAEL000293-RA	AAEL003904-RA	AAEL004325-RA	AAEL014141-RB	AAEL003950-RA	AAEL000550-RA
AAEL000180-RA	AAEL000407-RA	AAEL014093-RA	AAEL004409-RA	AAEL014196-RA	AAEL014180-RA	AAEL014167-RA
AAEL000383-RA	AAEL000413-RA	AAEL014134-RA	AAEL014635-RA	AAEL014244-RA	AAEL014223-RA	AAEL014171-RA
AAEL000626-RA	AAEL000713-RA	AAEL014162-RA	AAEL004476-RA	AAEL004096-RA	AAEL014222-RA	AAEL004007-RA
AAEL000943-RA	AAEL001065-RA	AAEL014207-RA	AAEL004582-RA	AAEL004088-RC	AAEL014251-RA	AAEL004034-RA
AAEL001011-RA	AAEL001163-RA	AAEL004022-RA	AAEL004575-RA	AAEL004137-RA	AAEL004142-RA	AAEL004059-RA
AAEL001114-RA	AAEL001180-RA	AAEL014258-RA	AAEL014714-RA	AAEL000581-RA	AAEL004172-RA	AAEL004075-RA
AAEL001134-RA	AAEL001596-RA	AAEL004074-RA	AAEL014727-RA	AAEL000567-RA	AAEL000589-RA	AAEL004227-RA
AAEL001177-RA	AAEL001670-RA	AAEL014294-RA	AAEL014762-RA	AAEL004209-RA	AAEL014440-RA	AAEL004256-RA
AAEL001250-RA	AAEL001752-RB	AAEL004103-RA	AAEL014846-RA	AAEL004294-RA	AAEL004302-RA	AAEL004284-RA
AAEL001277-RA	AAEL001930-RB	AAEL004174-RA	AAEL000643-RA	AAEL004277-RA	AAEL004328-RA	AAEL014516-RA
AAEL001280-RA	AAEL002183-RA	AAEL014368-RA	AAEL000642-RA	AAEL014500-RA	AAEL004390-RA	AAEL004389-RA
AAEL001400-RA	AAEL002691-RA	AAEL004242-RA	AAEL000649-RA	AAEL004415-RA	AAEL004386-RA	AAEL004445-RA
AAEL001412-RA	AAEL002691-RB	AAEL014428-RA	AAEL004902-RA	AAEL014612-RA	AAEL004401-RA	AAEL004433-RA
AAEL001682-RA	AAEL002804-RA	AAEL014435-RA	AAEL000700-RA	AAEL004440-RA	AAEL004388-RA	AAEL004482-RA
AAEL001859-RA	AAEL002805-RA	AAEL004307-RA	AAEL005048-RA	AAEL004486-RA	AAEL014825-RA	AAEL004493-RA
AAEL002483-RA	AAEL002873-RA	AAEL014527-RA	AAEL005086-RA	AAEL004502-RA	AAEL004842-RA	AAEL014701-RA
AAEL002488-RA	AAEL002876-RA	AAEL004398-RA	AAEL005071-RA	AAEL014753-RA	AAEL004861-RA	AAEL004571-RA
AAEL002709-RA	AAEL002898-RA	AAEL004464-RA	AAEL000741-RA	AAEL004658-RA	AAEL004976-RA	AAEL014733-RA
AAEL002741-RA	AAEL003365-RB	AAEL014671-RA	AAEL005403-RA	AAEL004653-RA	AAEL004978-RA	AAEL004700-RA
AAEL002748-RA	AAEL003418-RA	AAEL004516-RA	AAEL005505-RA	AAEL014829-RA	AAEL005259-RA	AAEL004775-RA
AAEL002764-RA	AAEL004087-RA	AAEL004588-RA	AAEL005531-RA	AAEL004739-RA	AAEL005259-RC	AAEL014932-RA
AAEL002766-RA	AAEL004300-RA	AAEL004764-RA	AAEL005653-RA	AAEL014864-RA	AAEL005259-RB	AAEL015013-RA
AAEL003016-RA	AAEL004384-RA	AAEL004753-RA	AAEL000823-RA	AAEL004833-RA	AAEL005325-RA	AAEL000685-RA
AAEL003455-RA	AAEL004426-RA	AAEL014867-RA	AAEL005722-RA	AAEL004988-RA	AAEL015143-RA	AAEL000699-RA
AAEL003489-RA	AAEL004497-RA	AAEL014925-RA	AAEL005747-RA	AAEL005026-RA	AAEL005383-RA	AAEL000693-RA
AAEL003536-RA	AAEL004514-RA	AAEL004905-RA	AAEL005763-RA	AAEL000711-RA	AAEL005492-RA	AAEL015060-RA
AAEL003581-RA	AAEL004697-RA	AAEL005030-RA	AAEL015291-RA	AAEL000703-RA	AAEL005484-RA	AAEL015066-RA
AAEL004345-RA	AAEL004749-RA	AAEL005160-RA	AAEL005863-RA	AAEL015036-RA	AAEL000776-RA	AAEL000743-RA
AAEL004490-RA	AAEL004756-RA	AAEL005163-RA	AAEL005921-RA	AAEL005120-RA	AAEL005515-RA	AAEL005326-RA
AAEL004523-RA	AAEL004938-RA	AAEL005249-RA	AAEL005964-RA	AAEL005221-RB	AAEL005849-RA	AAEL005358-RB
AAEL004706-RA	AAEL004969-RA	AAEL005278-RA	AAEL005961-RA	AAEL005458-RA	AAEL015320-RA	AAEL005615-RA
AAEL004914-RA	AAEL005059-RA	AAEL005265-RA	AAEL006002-RA	AAEL005583-RA	AAEL000889-RA	AAEL005606-RA
AAEL005037-RA	AAEL005065-RA	AAEL005266-RA	AAEL006035-RA	AAEL005691-RA	AAEL000923-RA	AAEL005602-RA
AAEL005190-RA	AAEL005119-RA	AAEL015116-RA	AAEL006069-RA	AAEL005687-RA	AAEL000888-RA	AAEL005638-RA
AAEL005279-RA	AAEL005200-RA	AAEL005381-RA	AAEL000955-RA	AAEL005711-RA	AAEL005923-RA	AAEL005628-RA
AAEL005421-RA	AAEL005417-RA	AAEL005581-RA	AAEL000930-RB	AAEL005815-RA	AAEL000961-RA	AAEL000844-RA
AAEL005471-RA	AAEL005457-RA	AAEL015260-RA	AAEL006168-RA	AAEL005839-RA	AAEL006123-RA	AAEL005744-RA
AAEL005822-RA	AAEL005500-RA	AAEL005699-RA	AAEL000119-RA	AAEL005997-RA	AAEL006208-RA	AAEL005826-RA
AAEL005944-RA	AAEL005558-RA	AAEL005773-RA	AAEL000993-RA	AAEL006070-RA	AAEL006271-RC	AAEL005886-RA
AAEL006043-RA	AAEL005705-RA	AAEL005854-RA	AAEL006343-RA	AAEL006180-RA	AAEL006298-RA	AAEL005898-RA
AAEL006079-RA	AAEL005968-RA	AAEL000910-RA	AAEL006397-RA	AAEL006269-RA	AAEL001009-RA	AAEL005950-RA
AAEL006207-RA	AAEL006233-RB	AAEL009573-RA	AAEL006698-RA	AAEL006353-RA	AAEL006331-RA	AAEL006009-RA
AAEL006337-RA	AAEL006258-RA	AAEL006019-RA	AAEL001059-RA	AAEL006364-RA	AAEL006354-RA	AAEL000938-RA
AAEL006466-RA	AAEL006344-RA	AAEL006092-RA	AAEL006848-RA	AAEL006446-RA	AAEL006426-RA	AAEL000950-RA
AAEL006474-RA	AAEL006409-RA	AAEL000937-RA	AAEL006872-RA	AAEL006436-RA	AAEL006407-RA	AAEL006117-RA
AAEL006520-RA	AAEL006438-RA	AAEL006126-RA	AAEL006860-RA	AAEL006568-RB	AAEL006447-RA	AAEL015356-RA
AAEL006521-RA	AAEL006512-RA	AAEL006126-RB	AAEL006870-RA	AAEL006568-RA	AAEL006726-RA	AAEL006286-RA
AAEL007309-RA	AAEL006580-RA	AAEL006222-RA	AAEL006936-RA	AAEL006594-RA	AAEL006775-RA	AAEL000116-RA

Continued

Cluster 1	Cluster 6	Cluster 4	Cluster 7	Cluster 5	Cluster 2A	Cluster 2B
AAEL007356-RA	AAEL006757-RA	AAEL006250-RA	AAEL015445-RA	AAEL006815-RA	AAEL006787-RA	AAEL000994-RA
AAEL007512-RA	AAEL007052-RA	AAEL006281-RA	AAEL001132-RA	AAEL006834-RA	AAEL006830-RA	AAEL000978-RA
AAEL007532-RA	AAEL007105-RA	AAEL006292-RA	AAEL006966-RD	AAEL006895-RA	AAEL006865-RA	AAEL006467-RA
AAEL007713-RA	AAEL007687-RA	AAEL015386-RA	AAEL006966-RA	AAEL001128-RA	AAEL001151-RA	AAEL006492-RA
AAEL008044-RA	AAEL007811-RA	AAEL000121-RA	AAEL007060-RA	AAEL007033-RA	AAEL007046-RA	AAEL006508-RA
AAEL008222-RA	AAEL008000-RA	AAEL006412-RA	AAEL007047-RA	AAEL007080-RA	AAEL007097-RA	AAEL006610-RA
AAEL008342-RA	AAEL008064-RA	AAEL006455-RA	AAEL007090-RA	AAEL015454-RA	AAEL007133-RA	AAEL006607-RA
AAEL008406-RA	AAEL008367-RA	AAEL006475-RA	AAEL007143-RA	AAEL007162-RA	AAEL007171-RA	AAEL006600-RB
AAEL008553-RA	AAEL008521-RA	AAEL006482-RA	AAEL001173-RA	AAEL001204-RA	AAEL007323-RA	AAEL006600-RA
AAEL008725-RA	AAEL008656-RA	AAEL006552-RA	AAEL007189-RA	AAEL007370-RA	AAEL007329-RA	AAEL006734-RA
AAEL009039-RA	AAEL008740-RA	AAEL006604-RA	AAEL007191-RA	AAEL007390-RA	AAEL007470-RA	AAEL001090-RA
AAEL009241-RA	AAEL008878-RD	AAEL006668-RA	AAEL007300-RA	AAEL007526-RB	AAEL007474-RA	AAEL006876-RA
AAEL009287-RA	AAEL008889-RA	AAEL006783-RA	AAEL007292-RA	AAEL007537-RA	AAEL007563-RA	AAEL001127-RA
AAEL009369-RA	AAEL009230-RA	AAEL006823-RA	AAEL007296-RA	AAEL007536-RA	AAEL007581-RA	AAEL007138-RA
AAEL009766-RA	AAEL009634-RE	AAEL006832-RA	AAEL007363-RA	AAEL007590-RA	AAEL007584-RA	AAEL007126-RA
AAEL010030-RA	AAEL009748-RA	AAEL015443-RA	AAEL007416-RA	AAEL015493-RA	AAEL007632-RA	AAEL007139-RA
AAEL010135-RA	AAEL010056-RA	AAEL007027-RA	AAEL007439-RA	AAEL007651-RA	AAEL007638-RA	AAEL007128-RA
AAEL010340-RA	AAEL010227-RA	AAEL001184-RA	AAEL007439-RB	AAEL007768-RA	AAEL007657-RA	AAEL007158-RC
AAEL010379-RA	AAEL010243-RA	AAEL007202-RA	AAEL015484-RA	AAEL007874-RA	AAEL007692-RA	AAEL007150-RA
AAEL010468-RB	AAEL010504-RA	AAEL007228-RA	AAEL007602-RA	AAEL008006-RA	AAEL007783-RA	AAEL007201-RA
AAEL010476-RA	AAEL010599-RA	AAEL007373-RA	AAEL007661-RA	AAEL015524-RA	AAEL007810-RA	AAEL007274-RA
AAEL010596-RA	AAEL010901-RB	AAEL007415-RA	AAEL007691-RA	AAEL001325-RB	AAEL007858-RA	AAEL007289-RA
AAEL010716-RA	AAEL010938-RA	AAEL007690-RA	AAEL007689-RB	AAEL008111-RA	AAEL007912-RA	AAEL007321-RA
AAEL011006-RA	AAEL011519-RA	AAEL007721-RA	AAEL007699-RA	AAEL008136-RA	AAEL007917-RA	AAEL007354-RA
AAEL011158-RB	AAEL011912-RA	AAEL001309-RA	AAEL007731-RA	AAEL008144-RA	AAEL008029-RA	AAEL007496-RA
AAEL011395-RA	AAEL011998-RA	AAEL007966-RA	AAEL007763-RA	AAEL008283-RA	AAEL000201-RA	AAEL007641-RA
AAEL011442-RA	AAEL012034-RA	AAEL008043-RA	AAEL007763-RB	AAEL008365-RA	AAEL000217-RA	AAEL007703-RA
AAEL011518-RA	AAEL012138-RB	AAEL008056-RA	AAEL007816-RA	AAEL008452-RA	AAEL000193-RA	AAEL007702-RA
AAEL011536-RA	AAEL012238-RA	AAEL001328-RA	AAEL007826-RA	AAEL008476-RA	AAEL000204-RA	AAEL007718-RA
AAEL011584-RA	AAEL012257-RA	AAEL001349-RA	AAEL001293-RA	AAEL008473-RA	AAEL001327-RA	AAEL007765-RA
AAEL011754-RA	AAEL012320-RA	AAEL008129-RA	AAEL008227-RA	AAEL008574-RA	AAEL008182-RA	AAEL007915-RA
AAEL011773-RA	AAEL012485-RA	AAEL008175-RA	AAEL008296-RA	AAEL008632-RA	AAEL008247-RA	AAEL001315-RA
AAEL011814-RA	AAEL013071-RA	AAEL008198-RA	AAEL008303-RA	AAEL008653-RA	AAEL008256-RA	AAEL007930-RA
AAEL011925-RA	AAEL013406-RA	AAEL008240-RA	AAEL001403-RA	AAEL008734-RA	AAEL008320-RA	AAEL000175-RA
AAEL012025-RA	AAEL013619-RA	AAEL008290-RA	AAEL001429-RA	AAEL008893-RA	AAEL008413-RA	AAEL000196-RA
AAEL012104-RA	AAEL013759-RA	AAEL008294-RA	AAEL008478-RA	AAEL008928-RB	AAEL008431-RA	AAEL001358-RA
AAEL012111-RA	AAEL013857-RA	AAEL001433-RA	AAEL008591-RA	AAEL008928-RC	AAEL008422-RB	AAEL008088-RA
AAEL012278-RA	AAEL014094-RA	AAEL001429-RA	AAEL001480-RA	AAEL009105-RA	AAEL008443-RA	AAEL008098-RA
AAEL012721-RA	AAEL014101-RA	AAEL008426-RA	AAEL008723-RB	AAEL009117-RB	AAEL008500-RA	AAEL008148-RA
AAEL012827-RA	AAEL014548-RA	AAEL008424-RA	AAEL008723-RA	AAEL009132-RA	AAEL008580-RA	AAEL008175-RA
AAEL012880-RA	AAEL014797-RA	AAEL008424-RB	AAEL008876-RA	AAEL009130-RA	AAEL008595-RA	AAEL008390-RA
AAEL013005-RA	AAEL014821-RA	AAEL008525-RA	AAEL008901-RA	AAEL001699-RA	AAEL008678-RA	AAEL015555-RA
AAEL013147-RA	AAEL015049-RA	AAEL008603-RA	AAEL001561-RD	AAEL009483-RA	AAEL008724-RA	AAEL008469-RA
AAEL013314-RA	AAEL015258-RA	AAEL008629-RA	AAEL008959-RA	AAEL000234-RA	AAEL008777-RA	AAEL008572-RA
AAEL013320-RA	AAEL015363-RA	AAEL001499-RA	AAEL015599-RA	AAEL000227-RA	AAEL001511-RA	AAEL008602-RA
AAEL013688-RA	AAEL015572-RA	AAEL008672-RA	AAEL009042-RA	AAEL009598-RA	AAEL008829-RA	AAEL008781-RA
AAEL013964-RA	AAEL015616-RA	AAEL008913-RA	AAEL009111-RA	AAEL009664-RA	AAEL001559-RA	AAEL008852-RA
AAEL013989-RA	AAEL015654-RA	AAEL008924-RA	AAEL015606-RA	AAEL009679-RA	AAEL009022-RA	AAEL008886-RA
AAEL014157-RA	AAEL000049-RA	AAEL008955-RA	AAEL009161-RA	AAEL009742-RA	AAEL009023-RA	AAEL008931-RA

Continued

Cluster 1	Cluster 6	Cluster 4	Cluster 7	Cluster 5	Cluster 2A	Cluster 2B
AAEL014177-RA	AAEL000309-RA	AAEL001615-RA	AAEL009172-RA	AAEL001816-RA	AAEL009084-RA	AAEL008942-RA
AAEL014400-RA	AAEL000309-RB	AAEL009185-RA	AAEL009232-RA	AAEL009826-RA	AAEL009080-RB	AAEL009092-RB
AAEL014401-RA	AAEL000444-RA	AAEL001659-RA	AAEL001692-RA	AAEL009810-RA	AAEL001598-RA	AAEL001613-RA
AAEL014444-RA	AAEL000470-RA	AAEL001650-RA	AAEL001685-RA	AAEL009857-RA	AAEL009118-RA	AAEL009153-RA
AAEL014526-RA	AAEL000687-RA	AAEL009358-RA	AAEL001698-RA	AAEL009872-RA	AAEL001666-RA	AAEL001651-RA
AAEL014742-RA	AAEL000791-RA	AAEL009466-RA	AAEL009362-RA	AAEL009931-RA	AAEL009212-RC	AAEL001656-RA
AAEL014823-RA	AAEL000864-RA	AAEL000242-RA	AAEL009407-RA	AAEL010005-RA	AAEL009238-RA	AAEL009285-RA
AAEL015236-RA	AAEL000901-RA	AAEL000250-RA	AAEL009432-RA	AAEL010037-RA	AAEL009309-RA	AAEL009356-RA
AAEL015489-RA	AAEL000947-RA	AAEL009556-RA	AAEL000236-RA	AAEL001892-RB	AAEL009411-RA	AAEL009380-RB
AAEL015601-RA	AAEL001046-RA	AAEL009588-RA	AAEL009588-RA	AAEL001892-RC	AAEL009457-RA	AAEL009401-RA
AAEL000001-RA	AAEL001275-RA	AAEL009619-RA	AAEL009712-RA	AAEL010102-RA	AAEL009572-RA	AAEL009440-RA
AAEL000077-RA	AAEL001432-RA	AAEL001772-RA	AAEL009747-RA	AAEL010371-RB	AAEL001781-RA	AAEL009469-RA
AAEL000160-RA	AAEL001516-RA	AAEL009745-RA	AAEL009782-RA	AAEL010452-RA	AAEL001749-RA	AAEL001715-RA
AAEL000188-RA	AAEL001541-RA	AAEL009803-RA	AAEL001805-RA	AAEL002042-RA	AAEL009773-RB	AAEL009579-RA
AAEL000199-RA	AAEL001565-RA	AAEL001830-RA	AAEL001813-RA	AAEL002062-RB	AAEL009884-RA	AAEL001751-RA
AAEL000505-RA	AAEL001627-RA	AAEL001812-RA	AAEL009874-RA	AAEL002062-RA	AAEL009983-RA	AAEL009738-RA
AAEL000988-RA	AAEL001767-RB	AAEL009814-RA	AAEL009921-RA	AAEL010479-RA	AAEL010033-RA	AAEL001807-RA
AAEL001005-RA	AAEL001767-RC	AAEL009857-RA	AAEL009938-RA	AAEL010602-RA	AAEL010080-RA	AAEL009859-RA
AAEL001089-RA	AAEL001768-RA	AAEL009937-RB	AAEL015655-RA	AAEL010671-RA	AAEL010094-RA	AAEL009871-RA
AAEL001158-RA	AAEL001793-RA	AAEL001878-RA	AAEL015657-RA	AAEL010684-RA	AAEL010085-RA	AAEL009914-RA
AAEL001580-RA	AAEL001806-RA	AAEL010009-RA	AAEL010053-RA	AAEL010697-RB	AAEL010097-RA	AAEL009959-RA
AAEL001705-RA	AAEL001833-RA	AAEL010103-RA	AAEL010123-RA	AAEL002165-RA	AAEL010240-RA	AAEL010003-RA
AAEL001825-RA	AAEL001944-RA	AAEL010106-RA	AAEL010148-RA	AAEL010731-RA	AAEL001976-RA	AAEL010003-RB
AAEL001869-RA	AAEL002097-RB	AAEL010355-RA	AAEL001951-RA	AAEL010765-RA	AAEL010317-RA	AAEL010069-RA
AAEL001926-RA	AAEL002174-RA	AAEL010415-RA	AAEL010222-RA	AAEL010850-RA	AAEL010342-RA	AAEL010159-RA
AAEL002002-RA	AAEL002198-RA	AAEL002053-RA	AAEL010427-RA	AAEL010881-RA	AAEL010422-RA	AAEL015669-RA
AAEL002061-RA	AAEL002255-RA	AAEL010509-RA	AAEL010404-RA	AAEL010937-RA	AAEL010568-RA	AAEL010327-RA
AAEL002063-RA	AAEL002625-RA	AAEL010536-RA	AAEL010434-RA	AAEL002296-RA	AAEL010583-RA	AAEL010326-RA
AAEL002731-RA	AAEL002996-RA	AAEL010565-RB	AAEL002047-RA	AAEL002347-RA	AAEL010638-RA	AAEL010453-RA
AAEL002757-RA	AAEL003285-RA	AAEL010681-RA	AAEL002082-RA	AAEL011126-RA	AAEL010670-RA	AAEL010486-RA
AAEL003125-RA	AAEL003791-RA	AAEL002142-RA	AAEL010487-RA	AAEL002378-RA	AAEL000300-RA	AAEL010485-RA
AAEL003377-RA	AAEL003859-RA	AAEL010719-RA	AAEL010506-RA	AAEL002397-RA	AAEL010715-RA	AAEL010526-RA
AAEL003547-RA	AAEL004008-RA	AAEL010767-RA	AAEL010555-RA	AAEL011191-RA	AAEL010786-RA	AAEL010637-RA
AAEL003704-RA	AAEL004019-RA	AAEL002191-RA	AAEL000278-RA	AAEL011325-RA	AAEL002250-RA	AAEL002157-RA
AAEL003913-RA	AAEL004023-RA	AAEL002246-RA	AAEL000264-RA	AAEL011327-RA	AAEL011058-RA	AAEL010760-RA
AAEL004327-RA	AAEL004127-RA	AAEL010912-RA	AAEL000297-RA	AAEL011323-RA	AAEL011085-RA	AAEL010905-RA
AAEL004430-RA	AAEL004343-RA	AAEL010921-RA	AAEL002185-RA	AAEL011323-RB	AAEL002422-RA	AAEL010921-RA
AAEL004452-RA	AAEL004449-RA	AAEL002299-RA	AAEL002221-RA	AAEL011408-RA	AAEL011281-RA	AAEL002294-RA
AAEL004546-RA	AAEL004461-RA	AAEL011032-RA	AAEL002216-RA	AAEL011460-RA	AAEL011330-RA	AAEL010993-RA
AAEL004594-RA	AAEL004531-RA	AAEL002358-RA	AAEL002240-RA	AAEL002503-RA	AAEL011358-RA	AAEL002339-RA
AAEL004685-RA	AAEL004603-RA	AAEL011177-RA	AAEL002181-RA	AAEL002576-RA	AAEL011396-RA	AAEL011146-RA
AAEL004708-RA	AAEL004879-RA	AAEL002388-RA	AAEL010838-RA	AAEL011686-RA	AAEL011424-RA	AAEL011163-RA
AAEL005201-RA	AAEL004993-RA	AAEL011204-RA	AAEL010850-RA	AAEL011746-RA	AAEL011596-RA	AAEL011152-RA
AAEL005273-RA	AAEL005015-RA	AAEL011202-RB	AAEL002294-RA	AAEL011739-RA	AAEL011653-RA	AAEL002375-RA
AAEL005730-RA	AAEL005223-RA	AAEL011202-RA	AAEL011078-RA	AAEL002615-RA	AAEL000335-RA	AAEL011198-RA
AAEL005900-RA	AAEL005336-RA	AAEL011347-RA	AAEL011088-RA	AAEL011831-RA	AAEL002551-RA	AAEL011199-RA
AAEL006012-RA	AAEL005429-RA	AAEL011394-RA	AAEL011102-RA	AAEL011863-RA	AAEL002569-RA	AAEL011255-RA
AAEL006049-RA	AAEL005684-RA	AAEL002528-RA	AAEL011206-RB	AAEL011898-RA	AAEL011715-RA	AAEL011282-RA
AAEL006303-RA	AAEL006196-RA	AAEL011604-RA	AAEL002415-RC	AAEL011906-RA	AAEL011735-RA	AAEL011314-RA
AAEL006304-RA	AAEL006450-RA	AAEL011620-RA	AAEL002415-RA	AAEL002669-RA	AAEL011791-RA	AAEL011651-RA
AAEL006367-RA	AAEL006621-RA	AAEL011609-RA	AAEL002426-RA	AAEL002658-RA	AAEL011838-RA	AAEL011696-RA

Continued

Cluster 1	Cluster 6	Cluster 4	Cluster 7	Cluster 5	Cluster 2A	Cluster 2B
AAEL006415-RA	AAEL006919-RA	AAEL002565-RA	AAEL011265-RA	AAEL002687-RA	AAEL012107-RA	AAEL011742-RA
AAEL006519-RA	AAEL007049-RA	AAEL011776-RA	AAEL011264-RA	AAEL012037-RA	AAEL002740-RA	AAEL011738-RA
AAEL006560-RA	AAEL007226-RA	AAEL002637-RA	AAEL011272-RA	AAEL012036-RA	AAEL002751-RA	AAEL011762-RA
AAEL006562-RA	AAEL007229-RA	AAEL011835-RA	AAEL002453-RA	AAEL012162-RA	AAEL012212-RA	AAEL011827-RA
AAEL006645-RA	AAEL007286-RA	AAEL011995-RA	AAEL011466-RA	AAEL012251-RA	AAEL012418-RA	AAEL011851-RA
AAEL006808-RA	AAEL007308-RA	AAEL012041-RA	AAEL002516-RA	AAEL012266-RA	AAEL012499-RA	AAEL011869-RA
AAEL006912-RB	AAEL007506-RA	AAEL012039-RA	AAEL002511-RA	AAEL002781-RB	AAEL002921-RA	AAEL011960-RA
AAEL007366-RA	AAEL007634-RA	AAEL012088-RA	AAEL002520-RA	AAEL002812-RA	AAEL012659-RA	AAEL012073-RA
AAEL007395-RA	AAEL007847-RA	AAEL012124-RA	AAEL011610-RA	AAEL002842-RA	AAEL002968-RA	AAEL012227-RA
AAEL007423-RA	AAEL007854-RA	AAEL012131-RA	AAEL011639-RA	AAEL012430-RA	AAEL012840-RA	AAEL012235-RA
AAEL007485-RA	AAEL007891-RA	AAEL012141-RA	AAEL002557-RA	AAEL012455-RA	AAEL012876-RA	AAEL012300-RA
AAEL008149-RA	AAEL007948-RA	AAEL012192-RA	AAEL002555-RA	AAEL000395-RB	AAEL012935-RA	AAEL002784-RA
AAEL008167-RA	AAEL008211-RA	AAEL012264-RA	AAEL002611-RA	AAEL000395-RA	AAEL012959-RA	AAEL012359-RA
AAEL008474-RA	AAEL008425-RA	AAEL012288-RA	AAEL011764-RA	AAEL012576-RA	AAEL012960-RA	AAEL012361-RA
AAEL008768-RA	AAEL008459-RA	AAEL012334-RA	AAEL011957-RA	AAEL012576-RB	AAEL013025-RA	AAEL000393-RA
AAEL008811-RA	AAEL008461-RA	AAEL002844-RA	AAEL011981-RA	AAEL012636-RA	AAEL013066-RA	AAEL012568-RA
AAEL008834-RC	AAEL008594-RA	AAEL012417-RA	AAEL012101-RA	AAEL012665-RA	AAEL013110-RA	AAEL012618-RB
AAEL009016-RA	AAEL009560-RA	AAEL000384-RA	AAEL012197-RA	AAEL012685-RA	AAEL013168-RA	AAEL012644-RA
AAEL009070-RA	AAEL009570-RA	AAEL012564-RA	AAEL002777-RA	AAEL002961-RA	AAEL000442-RA	AAEL012646-RA
AAEL009281-RA	AAEL009708-RB	AAEL012691-RA	AAEL002803-RA	AAEL002972-RA	AAEL003162-RA	AAEL012746-RA
AAEL009875-RB	AAEL009829-RA	AAEL012736-RA	AAEL012374-RA	AAEL003022-RB	AAEL013288-RB	AAEL012747-RA
AAEL009917-RA	AAEL010398-RA	AAEL003113-RA	AAEL000399-RA	AAEL012862-RA	AAEL013295-RA	AAEL012843-RA
AAEL009946-RA	AAEL010496-RA	AAEL013159-RA	AAEL002862-RA	AAEL012886-RA	AAEL013300-RA	AAEL003039-RA
AAEL010008-RA	AAEL010521-RA	AAEL000417-RA	AAEL012563-RA	AAEL012897-RA	AAEL013338-RA	AAEL012874-RA
AAEL010065-RA	AAEL010538-RA	AAEL003173-RA	AAEL012588-RA	AAEL003086-RA	AAEL013390-RA	AAEL003071-RA
AAEL010377-RA	AAEL010895-RA	AAEL013239-RA	AAEL012671-RA	AAEL003064-RA	AAEL003371-RA	AAEL003090-RA
AAEL010689-RA	AAEL010939-RA	AAEL013303-RA	AAEL012709-RA	AAEL012994-RA	AAEL003382-RA	AAEL003065-RA
AAEL011155-RA	AAEL010952-RA	AAEL013324-RA	AAEL002999-RA	AAEL000445-RB	AAEL013487-RA	AAEL012980-RA
AAEL011224-RA	AAEL011045-RA	AAEL003294-RA	AAEL013037-RA	AAEL000450-RA	AAEL013510-RA	AAEL013144-RA
AAEL011289-RA	AAEL011169-RA	AAEL013499-RA	AAEL000464-RA	AAEL013218-RA	AAEL013546-RA	AAEL000436-RA
AAEL011360-RA	AAEL011292-RA	AAEL013492-RA	AAEL013239-RA	AAEL003209-RA	AAEL013569-RA	AAEL003178-RA
AAEL011666-RA	AAEL011496-RA	AAEL003433-RA	AAEL013309-RC	AAEL003188-RA	AAEL013596-RA	AAEL013179-RA
AAEL011901-RA	AAEL011580-RA	AAEL003444-RA	AAEL003288-RA	AAEL013287-RB	AAEL003510-RA	AAEL013221-RA
AAEL012240-RA	AAEL012016-RA	AAEL003469-RA	AAEL013407-RA	AAEL013341-RA	AAEL013721-RA	AAEL013221-RC
AAEL012282-RC	AAEL012105-RA	AAEL003469-RB	AAEL003433-RA	AAEL013341-RB	AAEL013813-RA	AAEL003235-RA
AAEL012762-RA	AAEL012395-RA	AAEL013686-RA	AAEL003504-RA	AAEL003345-RA	AAEL013819-RA	AAEL013305-RA
AAEL013121-RC	AAEL012511-RA	AAEL013735-RA	AAEL013625-RA	AAEL003347-RB	AAEL013820-RA	AAEL003352-RA
AAEL013521-RA	AAEL012570-RA	AAEL013770-RA	AAEL003548-RA	AAEL003347-RA	AAEL003647-RA	AAEL003385-RA
AAEL013644-RA	AAEL012684-RA	AAEL013771-RA	AAEL013723-RA	AAEL013434-RB	AAEL003673-RA	AAEL013533-RA
AAEL014293-RA	AAEL013494-RA	AAEL013811-RA	AAEL013728-RA	AAEL003402-RB	AAEL003659-RA	AAEL013610-RA
AAEL014381-RA	AAEL013577-RA	AAEL003627-RA	AAEL003628-RA	AAEL013556-RA	AAEL015683-RA	AAEL013725-RA
AAEL014426-RA	AAEL013642-RA	AAEL013881-RA	AAEL003630-RA	AAEL003497-RA	AAEL003685-RA	AAEL013822-RA
AAEL014629-RA	AAEL014029-RA	AAEL013882-RA	AAEL003653-RA	AAEL003643-RA	AAEL003689-RA	AAEL003661-RA
AAEL014709-RA	AAEL014044-RA	AAEL013915-RA	AAEL013879-RA	AAEL013877-RF	AAEL013852-RA	AAEL015682-RA
AAEL014818-RA	AAEL014379-RA		AAEL013915-RA	AAEL003713-RA	AAEL003750-RA	AAEL003676-RA
AAEL014836-RA	AAEL014482-RA		AAEL013935-RA	AAEL003734-RA	AAEL003818-RA	AAEL003769-RA
AAEL015440-RA	AAEL014707-RA		AAEL013994-RA	AAEL013956-RA	AAEL003862-RA	AAEL013950-RA
AAEL015441-RA	AAEL014734-RA			AAEL014005-RA	AAEL015678-RA	AAEL013955-RA
AAEL015571-RA	AAEL015410-RA			AAEL003888-RA	AAEL014025-RA	AAEL003801-RA
AAEL015633-RA	AAEL015438-RA				AAEL003893-RA	AAEL013968-RA
					AAEL014038-RA	AAEL014037-RA
					AAEL014051-RA	AAEL014049-RA

Table 2.7.2 Caption

Detected motifs patterns from MEME and Weeder

Table 2.7.2

Cluster 1	Cluster 6	Cluster 4	Cluster 7	Cluster 5	Cluster 2A	Cluster 2B
CATA[CG][AT]AAA	CA[AG]ACGTCA	AGC[AGT]A[AG]ACC	GCC[CGT][AC]AGCA	ATC[GT]CTCGC	CA[AC]G[AG]AA	C[TC]GTGCAAA
G[CG]GTTATGA	CG[AT]ATCCCA	CTGTT[CG]AAA	T[CA]CTGTCGG	[AT]CTGACTG	G[GC]G[CG]CTAAC	[CA]C[GA]TGTCCC
CAC[ACG]C[AG]CCC	CAAAAAT[AGT]G	AAA[AT]ATGGG	AAAAGT[GA]A[AG]	TTGATTGA	AAAAG[CG]GCC	CAA[AG][AT]CAAA
GAA[AC][CT]GTCA	GCCA[AT]AAA[CT]	GAC[TA][TG]CCC	TCAC[ACG]TGAA	G[CT]AACGGAA	GCCG[GT][AC]GCC	[AC][AC]GCAGCG
[AG]GTCA[AT]GGA	CGAC[GA]CG[CT]C	CTCCT[GT][AT]GC	ATT[CA]C[GA]GAC	AA[CA]TGGTG	ACCTCA[ACGT]A[CT]	GTTGTCA[AG]A
GCC[AG]CGG[AC]C		CAG[AT][AC]GCTG	AAACCG[AC]AC	AC[GC]ATGAC	AAAA[AC]TCGA	T[TC]GTGCAAA
CA[AC]ATGCGG		GG[AT]AGCAC	TGG[AT]AGCAA	ACTCA[CG]TC	AT[AGT]CCGAAC	AA[CA]AACCCC
C[AGT]C[AG]GGAAG		CAAGA[GC]CAC	CGCG[CT]G[AT]A		GCGA[TA]TCC	[CG]CGC[CT]ACC
GCAG[AC]GCA		[CT]AATAAAAA	CT[CG][AGT]CTCTC		CGT[GT][ATC]GGCA	AAAA[TG][CG]GTG
T[CT]AC[CG]GC		TGCC[AC][AC]AAA	CGC[CTG]C[CG]AA		CG[CT][GT]TGGCA	GAA[AC]TTCCA
GGAA[CT]C[AC]CA		GAGGA[ACT]AAA	AA[CG]GACGAA		CGC[CG]GGC[CT]G	CAA[AGT]ACG[GC]C
[ACT]ACGTC[AT]AA		AT[AG]ACTCGA			AAAC[AG]CATG	[AG][GT]CTGCGCG
AA[AC]CAAAAC		CTT[TA]G[CG]CGA			CAGT[GA]TCG	ACGCCC[AT]A
ATTCCAGT[AG]					[AG]TGGACCA[AG]	CCGG[ACG]AA[AC]
[AC]TG[AG]GCCC					AAAAT[ACT]GAG	AATGAAC[CG]C
					CC[AT]C[AT]GGGG	AG[CGT]GTTACG
						G[AT]ACGGCA
						AAACA[ATG]A[AC]
						[AG]GGAACTCC
						GAAA[AC][GC]GCC
						G[CT]AAACAAA
						CCC[AT]CCC[AC]C

Table 2.7.3 Caption

Results of JASPAR search motif patterns of Cluster 1 to Cluster 2B

Table 2.7.3

Cluster 1

Query_ID	Target_ID	Optimal_offset	p_value	E_value	q_value	Overlap	Query_consensus	Target_consensus	Orientation
CATA[CG][AT]AAA	MA0049.1-hb	1	0.00717988	0.940564	1	9	CATACAAAA	GCATAAAAAA	0
CATA[CG][AT]AAA	MA0012.1-br_Z3	0	0.0323174	4.23358	1	9	CATACAAAA	TAAACTAAAAG	0
CATA[CG][AT]AAA	MA0216.2-CAD	3	0.0471952	6.18257	1	8	CATACAAAA	GGCCATAAAAA	0
CATA[CG][AT]AAA	MA0011.1-br_Z2	1	0.0478628	6.27003	1	7	CATACAAAA	AAATAGTA	0
CATA[CG][AT]AAA	MA0013.1-br_Z4	-2	0.0516745	6.76935	1	7	CATACAAAA	TAGTAAACAAA	0
CATA[CG][AT]AAA	MA0010.1-br_Z1	2	0.060809	7.96598	1	9	CATACAAAA	GTAATAAACAAATC	0
G[CG]GTTATGA	MA0165.1-Abd-B	-2	0.00910633	1.19293	0.999989	7	GCGTTATGA	TTTATGA	0
G[CG]GTTATGA	MA0452.2-KR	6	0.0162574	2.12973	0.999989	8	GCGTTATGA	GAAAAAGGGTTAAA	0
G[CG]GTTATGA	MA0174.1-CG42234	-2	0.0239259	3.13429	0.999989	7	GCGTTATGA	TTTATTA	0
G[CG]GTTATGA	MA0016.1-usp	1	0.0257863	3.378	0.999989	9	GCGTTATGA	GGGGTCACGG	0
G[CG]GTTATGA	MA0234.1-oc	0	0.0288119	3.77436	0.999989	6	GCGTTATGA	GGATTA	0
G[CG]GTTATGA	MA0448.1-H2.0	-2	0.0299475	3.92313	0.999989	7	GCGTTATGA	TTAATTA	0
G[CG]GTTATGA	MA0212.1-bcd	0	0.0302155	3.95823	0.999989	6	GCGTTATGA	GGATTA	0
G[CG]GTTATGA	MA0190.1-Gsc	0	0.0389935	5.10815	0.999989	6	GCGTTATGA	GGATTA	0
G[CG]GTTATGA	MA0186.1-Dfd	-2	0.0406224	5.32153	0.999989	7	GCGTTATGA	TTAATGA	0
G[CG]GTTATGA	MA0215.1-btn	-2	0.0406224	5.32153	0.999989	7	GCGTTATGA	TTAATGA	0
G[CG]GTTATGA	MA0216.2-CAD	0	0.0442246	5.79342	0.999989	9	GCGTTATGA	TTTTTATGGCC	0
CAC[ACG]C[AG]CCC	MA0450.1-hkb	-1	0.0008599	0.112647	0.0806606	8	CACGCACCC	TCACGCCCC	0
CAC[ACG]C[AG]CCC	MA0449.1-h	2	0.00093323	0.122252	0.0806606	8	CACGCACCC	GGCACGTGCC	0
CAC[ACG]C[AG]CCC	MA0443.1-btd	0	0.00172885	0.22648	0.112071	9	CACGCACCC	TCCGCCCCCT	0
CAC[ACG]C[AG]CCC	MA0255.1-z	-1	0.0106293	1.39244	0.551228	8	CACGCACCC	AATCACTCAA	0
CAC[ACG]C[AG]CCC	MA0249.1-twi	3	0.0176579	2.31318	0.763103	9	CACGCACCC	CAACATATGCGA	0
CAC[ACG]C[AG]CCC	MA0535.1-Mad	0	0.022155	2.90231	0.799916	9	CACGCACCC	CAGGCGCCGCCGCCG	0
CAC[ACG]C[AG]CCC	MA0086.1-sna	-4	0.0246796	3.23303	0.799916	5	CACGCACCC	CACCTG	0
CAC[ACG]C[AG]CCC	MA0456.1-opa	3	0.0370999	4.86009	0.916617	9	CACGCACCC	GACCCCCCGCTG	0
CAC[ACG]C[AG]CCC	MA0205.1-Trl	1	0.038578	5.05371	0.916617	9	CACGCACCC	TTGCTCTCTC	0
CAC[ACG]C[AG]CCC	MA0213.1-brk	0	0.0418286	5.47955	0.916617	8	CACGCACCC	CTGGCGCC	0
CAC[ACG]C[AG]CCC	MA0531.1-CTCF	5	0.0424202	5.55704	0.916617	9	CACGCACCC	GGCGCCATCTAGCGG	0
CAC[ACG]C[AG]CCC	MA0022.1-dl_1	2	0.0498182	6.52619	0.989681	9	CACGCACCC	CGGAAAAACCCC	0
CAC[ACG]C[AG]CCC	MA0237.2-pan	4	0.06663	8.72853	0.989681	9	CACGCACCC	ATCAAAGGAGCCGA	0
CAC[ACG]C[AG]CCC	MA0247.2-tin	1	0.0751289	9.84189	0.989681	9	CACGCACCC	CCACTTGAAA	0
GAA[AC][CT]GTCA	MA0227.1-hth	-3	9.29792e-06	0.00121803	0.00242026	6	GAAACGTCA	CTGTCA	0
GAA[AC][CT]GTCA	MA0207.1-achi	-3	0.00035778	0.0468696	0.0465658	6	GAAACGTCA	CTGTCA	0
GAA[AC][CT]GTCA	MA0252.1-vis	-3	0.00056069	0.0734509	0.0486498	6	GAAACGTCA	CTGTCA	0
GAA[AC][CT]GTCA	MA0459.1-tll	-1	0.00693444	0.908412	0.451261	8	GAAACGTCA	AAAAGTCAAA	0
GAA[AC][CT]GTCA	MA0085.1-Su(H)	6	0.0137442	1.80049	0.704218	9	GAAACGTCA	CTGTGGGAAACGAGAT	0
GAA[AC][CT]GTCA	MA0243.1-sd	4	0.018372	2.40673	0.704218	8	GAAACGTCA	CCGAGGAATGTC	0
GAA[AC][CT]GTCA	MA0222.1-exd	-4	0.0189664	2.4846	0.704218	5	GAAACGTCA	TGTCAAAA	0
GAA[AC][CT]GTCA	MA0530.1-CNC::maf-S	4	0.0216431	2.83525	0.704218	9	GAAACGTCA	ATTTGCCGAGTCATC	0
GAA[AC][CT]GTCA	MA0217.1-caup	-4	0.0418711	5.48512	0.9935	5	GAAACGTCA	TGTTA	0
GAA[AC][CT]GTCA	MA0199.1-Optix	-4	0.0631379	8.27107	0.9935	5	GAAACGTCA	TATCA	0
[AG]GTCA[AT]GGA	MA0534.1-EcR::usp	2	0.00114454	0.149935	0.297926	9	AGTCAAGGA	AAGGTCAATGAACTC	0
[AG]GTCA[AT]GGA	MA0016.1-usp	2	0.00255639	0.334887	0.332716	8	AGTCAAGGA	GGGGTCACGG	0
[AG]GTCA[AT]GGA	MA0532.1-STAT92E	4	0.0284759	3.73035	0.993517	9	AGTCAAGGA	CGGAATTCCAGGAAA	0
[AG]GTCA[AT]GGA	MA0010.1-br_Z1	-1	0.03579	4.68849	0.993517	8	AGTCAAGGA	GTAATAAACAAATC	0
[AG]GTCA[AT]GGA	MA0253.1-vnd	0	0.0566928	7.42675	0.993517	9	AGTCAAGGA	TTTCAAGTG	0
[AG]GTCA[AT]GGA	MA0216.2-CAD	0	0.056756	7.43504	0.993517	9	AGTCAAGGA	GGCCATAAAAA	0
[AG]GTCA[AT]GGA	MA0459.1-tll	3	0.0661415	8.66454	0.993517	7	AGTCAAGGA	AAAAGTCAAA	0

Continued (cluster 1)

Query_ID	Target_ID	Optimal_offset	p_value	E_value	q_value	Overlap	Query_consensus	Target_consensus	Orientation
GCC[AG]CGG[AC]C	MA0535.1-Mad	5	0.00012633	0.0165486	0.0321905	9	GCCACGGAC	CAGGCGCCGCCGCCG	0
GCC[AG]CGG[AC]C	MA0449.1-h	0	0.0012066	0.158065	0.10249	9	GCCACGGAC	GGCACGTGCC	0
GCC[AG]CGG[AC]C	MA0016.1-usp	3	0.00292405	0.383051	0.186279	7	GCCACGGAC	GGGGTCACGG	0
GCC[AG]CGG[AC]C	MA0456.1-opa	5	0.00933207	1.2225	0.332518	7	GCCACGGAC	GACCCCCCGCTG	0
GCC[AG]CGG[AC]C	MA0450.1-hkb	-1	0.0101712	1.33243	0.332518	8	GCCACGGAC	TCACGCCCC	0
GCC[AG]CGG[AC]C	MA0213.1-brk	3	0.0123564	1.61868	0.332518	5	GCCACGGAC	GGCGCCAG	0
GCC[AG]CGG[AC]C	MA0443.1-btd	0	0.0130157	1.70505	0.332518	9	GCCACGGAC	TCCGCCCCCT	0
GCC[AG]CGG[AC]C	MA0530.1-CNC::maf-S	3	0.0130862	1.7143	0.332518	9	GCCACGGAC	GATGACTCGGCAAAT	0
GCC[AG]CGG[AC]C	MA0534.1-EcR::usp	4	0.040495	5.30485	0.737074	9	GCCACGGAC	AAGGTCAATGAACTC	0
GCC[AG]CGG[AC]C	MA0242.1-run::Bgb	2	0.0495858	6.49574	0.809309	7	GCCACGGAC	TAACCGCAA	0
GCC[AG]CGG[AC]C	MA0085.1-Su(H)	9	0.0560533	7.34299	0.809309	7	GCCACGGAC	ATCTCGGTTCCCACAA	0
GCC[AG]CGG[AC]C	MA0531.1-CTCF	3	0.0591345	7.74662	0.809309	9	GCCACGGAC	GGCGCCATCTAGCGG	0
CA[AC]ATGCGG	MA0249.1-twi	3	7.62072e-05	0.00998314	0.0199663	9	CAAATGCGG	CAACATATGCGA	0
CA[AC]ATGCGG	MA0453.1-nub	4	0.0120328	1.5763	0.966723	8	CAAATGCGG	TATGCAAATTAG	0
CA[AC]ATGCGG	MA0237.2-pan	2	0.0143864	1.88461	0.966723	9	CAAATGCGG	ATCAAAGGAGCCGA	0
CA[AC]ATGCGG	MA0242.1-run::Bgb	-3	0.0170954	2.2395	0.966723	6	CAAATGCGG	TTGCGGTTA	0
CA[AC]ATGCGG	MA0086.1-sna	0	0.0435092	5.69971	0.999823	6	CAAATGCGG	CACCTG	0
CA[AC]ATGCGG	MA0449.1-h	2	0.0468167	6.13298	0.999823	8	CAAATGCGG	GGCACGTGCC	0
CA[AC]ATGCGG	MA0185.1-Deaf1	-4	0.0659703	8.64211	0.999823	5	CAAATGCGG	TTCGGG	0
CA[AC]ATGCGG	MA0026.1-Eip74EF	-2	0.0680407	8.91334	0.999823	7	CAAATGCGG	CTTCCGG	0
CA[AC]ATGCGG	MA0247.2-tin	1	0.0707147	9.26363	0.999823	9	CAAATGCGG	CCACTTGAAA	0
CA[AC]ATGCGG	MA0531.1-CTCF	5	0.0760358	9.96069	0.999823	9	CAAATGCGG	GGCGCCATCTAGCGG	0
C[AGT]C[AG]GGAAG	MA0026.1-Eip74EF	-2	8.02748e-05	0.010516	0.021032	7	CTCAGGAAG	CCGGAAG	0
C[AGT]C[AG]GGAAG	MA0456.1-opa	1	0.0297817	3.9014	0.999998	9	CTCAGGAAG	CAGCGGGGGGTC	0
C[AGT]C[AG]GGAAG	MA0530.1-CNC::maf-S	5	0.0382623	5.01236	0.999998	9	CTCAGGAAG	GATGACTCGGCAAAT	0
C[AGT]C[AG]GGAAG	MA0185.1-Deaf1	-2	0.0640232	8.38703	0.999998	6	CTCAGGAAG	CACGAA	0
GCAG[AC]GCA	MA0451.1-kni	3	0.00166457	0.218058	0.434962	8	GCAGAGCA	AAACTAGAGCAC	0
GCAG[AC]GCA	MA0205.1-Trl	1	0.00476423	0.624114	0.622462	8	GCAGAGCA	GAGAGAGCAA	0
GCAG[AC]GCA	MA0213.1-brk	0	0.030798	4.03454	0.997353	8	GCAGAGCA	CTGGCGCC	0
GCAG[AC]GCA	MA0533.1-SU(HW)	12	0.0405098	5.30678	0.997353	8	GCAGAGCA	GCCCAAAAGTATGCAACA	0
GCAG[AC]GCA	MA0530.1-CNC::maf-S	4	0.0463671	6.07408	0.997353	8	GCAGAGCA	ATTTGCCGAGTCATC	0
T[CT]AC[CG]GC	MA0242.1-run::Bgb	0	0.00249585	0.326957	0.607124	7	TCACCGC	TAACCGCAA	0
T[CT]AC[CG]GC	MA0450.1-hkb	0	0.00464684	0.608736	0.607124	7	TCACCGC	TCACGCCCC	0
T[CT]AC[CG]GC	MA0016.1-usp	4	0.0072841	0.954217	0.634461	6	TCACCGC	GGGGTCACGG	0
T[CT]AC[CG]GC	MA0216.2-CAD	3	0.0147459	1.93172	0.963302	7	TCACCGC	TTTTTATGGCC	0
T[CT]AC[CG]GC	MA0126.1-ovo	2	0.0300895	3.94173	0.995587	7	TCACCGC	AGTAACAGT	0
T[CT]AC[CG]GC	MA0239.1-prd	2	0.0300895	3.94173	0.995587	7	TCACCGC	AGTAACAGT	0
T[CT]AC[CG]GC	MA0026.1-Eip74EF	1	0.0648864	8.50012	0.995587	6	TCACCGC	CTTCCGG	0
T[CT]AC[CG]GC	MA0185.1-Deaf1	-1	0.0718963	9.41841	0.995587	6	TCACCGC	TTCGGG	0
T[CT]AC[CG]GC	MA0443.1-btd	-2	0.0745646	9.76797	0.995587	5	TCACCGC	TCCGCCCCCT	0
GGAA[CT]C[AC]CA	MA0532.1-STAT92E	6	0.00710676	0.930985	0.740416	9	GGAACCACA	TTTCCTGGAATTCCG	0
GGAA[CT]C[AC]CA	MA0023.1-dl_2	1	0.00781318	1.02353	0.740416	9	GGAACCACA	GGAAAACCCC	0
GGAA[CT]C[AC]CA	MA0085.1-Su(H)	5	0.00849049	1.11225	0.740416	9	GGAACCACA	CTGTGGGAAACGAGAT	0
GGAA[CT]C[AC]CA	MA0242.1-run::Bgb	-1	0.0193563	2.53568	0.998515	8	GGAACCACA	TAACCGCAA	0
GGAA[CT]C[AC]CA	MA0022.1-dl_1	1	0.0319427	4.1845	0.998515	9	GGAACCACA	CGGAAAAACCCC	0
GGAA[CT]C[AC]CA	MA0016.1-usp	2	0.0598726	7.84331	0.998515	8	GGAACCACA	CCGTGACCCC	0
GGAA[CT]C[AC]CA	MA0243.1-sd	4	0.0736044	9.64217	0.998515	8	GGAACCACA	CCGAGGAATGTC	0

Continued (cluster 1)

Query_ID	Target_ID	Optimal_offset	p_value	E_value	q_value	Overlap	Query_consensus	Target_consensus	Orientation
[ACT]ACGTC[AT]AA	MA0222.1-exd	-2	0.00250377	0.327994	0.366656	7	TACGTCAAA	TGTCAAAA	0
[ACT]ACGTC[AT]AA	MA0459.1-tll	1	0.00280633	0.367629	0.366656	9	TACGTCAAA	AAAAGTCAAA	0
[ACT]ACGTC[AT]AA	MA0253.1-vnd	0	0.0254973	3.34015	0.99734	9	TACGTCAAA	CACTTGAAA	0
[ACT]ACGTC[AT]AA	MA0173.1-CG11617	-1	0.0369481	4.8402	0.99734	7	TACGTCAAA	ATGTTAA	0
[ACT]ACGTC[AT]AA	MA0227.1-hth	-1	0.0641132	8.39883	0.99734	6	TACGTCAAA	CTGTCA	0
AA[AC]CAAAAC	MA0010.1-br_Z1	5	0.00482087	0.631534	0.65761	9	AAACAAAAC	GTAATAAACAAATC	0
AA[AC]CAAAAC	MA0012.1-br_Z3	1	0.00501993	0.65761	0.65761	9	AAACAAAAC	TAAACTAAAAG	0
AA[AC]CAAAAC	MA0458.1-slp1	0	0.0205043	2.68606	1	9	AAACAAAAC	AATGTAAACAA	0
AA[AC]CAAAAC	MA0049.1-hb	1	0.052066	6.82064	1	9	AAACAAAAC	GCATAAAAAA	0
AA[AC]CAAAAC	MA0013.1-br_Z4	4	0.0575918	7.54452	1	7	AAACAAAAC	TAGTAAACAAA	0
AA[AC]CAAAAC	MA0233.1-mirr	0	0.0619066	8.10977	1	5	AAACAAAAC	AAACA	0
AA[AC]CAAAAC	MA0445.1-D	1	0.068725	9.00297	1	9	AAACAAAAC	AGAACAATGGA	0
ATTCCAGT[AG]	MA0253.1-vnd	0	0.00293808	0.384889	0.769777	9	ATTCCAGTA	TTTCAAGTG	0
ATTCCAGT[AG]	MA0247.2-tin	0	0.00789929	1.03481	0.99923	9	ATTCCAGTA	TTTCAAGTGG	0
ATTCCAGT[AG]	MA0532.1-STAT92E	4	0.0232801	3.04969	0.99923	9	ATTCCAGTA	CGGAATTCCAGGAAA	0
ATTCCAGT[AG]	MA0026.1-Eip74EF	0	0.0254604	3.33532	0.99923	7	ATTCCAGTA	CTTCCGG	0
ATTCCAGT[AG]	MA0244.1-slbo	0	0.0607263	7.95515	0.99923	8	ATTCCAGTA	ATTGCAAA	0
ATTCCAGT[AG]	MA0243.1-sd	3	0.0642036	8.41067	0.99923	9	ATTCCAGTA	GACATTCCTCGA	0
[AC]TG[AG]GCCC	MA0016.1-usp	2	0.00280409	0.367335	0.73467	8	ATGAGCCC	CCGTGACCCC	0
[AC]TG[AG]GCCC	MA0443.1-btd	-1	0.0166851	2.18575	0.9991	7	ATGAGCCC	TCCGCCCCCT	0
[AC]TG[AG]GCCC	MA0456.1-opa	-2	0.0227725	2.9832	0.9991	6	ATGAGCCC	GACCCCCCGCTG	0
[AC]TG[AG]GCCC	MA0204.1-Six4	-1	0.0267583	3.50534	0.9991	6	ATGAGCCC	TGATAC	0
[AC]TG[AG]GCCC	MA0213.1-brk	0	0.0335842	4.39952	0.9991	8	ATGAGCCC	CTGGCGCC	0
[AC]TG[AG]GCCC	MA0452.2-KR	0	0.0378276	4.95542	0.9991	8	ATGAGCCC	TTTAACCCTTTTTC	0
[AC]TG[AG]GCCC	MA0534.1-EcR::usp	7	0.0441526	5.78398	0.9991	8	ATGAGCCC	AAGGTCAATGAACTC	0
[AC]TG[AG]GCCC	MA0234.1-oc	-1	0.0551275	7.2217	0.9991	6	ATGAGCCC	TAATCC	0
[AC]TG[AG]GCCC	MA0255.1-z	0	0.0562883	7.37377	0.9991	8	ATGAGCCC	TTGAGTGATT	0
[AC]TG[AG]GCCC	MA0212.1-bcd	-1	0.0683764	8.95731	0.9991	6	ATGAGCCC	TAATCC	0
[AC]TG[AG]GCCC	MA0244.1-slbo	-1	0.0705123	9.23711	0.9991	7	ATGAGCCC	TTTGCAAT	0
[AC]TG[AG]GCCC	MA0530.1-CNC::maf-S	5	0.07153	9.37043	0.9991	8	ATGAGCCC	GATGACTCGGCAAAT	0
[AC]TG[AG]GCCC	MA0451.1-kni	-2	0.0724787	9.4947	0.9991	6	ATGAGCCC	GTGCTCTAGTTT	0
[AC]TG[AG]GCCC	MA0450.1-hkb	-1	0.0758535	9.93681	0.9991	7	ATGAGCCC	GGGGCGTGA	0
[AC]TG[AG]GCCC	MA0237.2-pan	5	0.0759363	9.94766	0.9991	8	ATGAGCCC	ATCAAAGGAGCCGA	0

Cluster 6

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
CA[AG]ACGTCA	MA0459.1-tll	-1	0.00299717	0.392629	0.785259	8	CAAACGTCA	AAAAGTCAAA	0
CA[AG]ACGTCA	MA0227.1-hth	-3	0.00855743	1.12102	1	6	CAAACGTCA	CTGTCA	0
CA[AG]ACGTCA	MA0252.1-vis	-3	0.0360795	4.72641	1	6	CAAACGTCA	CTGTCA	0
CA[AG]ACGTCA	MA0207.1-achi	-3	0.0427494	5.60017	1	6	CAAACGTCA	CTGTCA	0
CA[AG]ACGTCA	MA0222.1-exd	-4	0.0515849	6.75762	1	5	CAAACGTCA	TGTCAAAA	0
CA[AG]ACGTCA	MA0535.1-Mad	0	0.0596818	7.81832	1	9	CAAACGTCA	CAGGCGCCGCCGCCG	0
CA[AG]ACGTCA	MA0173.1-CG11617	-3	0.0629408	8.24525	1	6	CAAACGTCA	ATGTTAA	0
CG[AT]ATCCCA	MA0085.1-Su(H)	4	0.00177609	0.232668	0.465336	9	CGAATCCCA	ATCTCGGTTCCCACAA	0
CG[AT]ATCCCA	MA0201.1-Ptx1	0	0.0252651	3.30973	0.999921	7	CGAATCCCA	TTAATCC	0
CG[AT]ATCCCA	MA0212.1-bcd	-1	0.0265456	3.47748	0.999921	6	CGAATCCCA	TAATCC	0
CG[AT]ATCCCA	MA0246.1-so	-1	0.0317083	4.15378	0.999921	6	CGAATCCCA	GTATCA	0
CG[AT]ATCCCA	MA0234.1-oc	-1	0.0366297	4.79848	0.999921	6	CGAATCCCA	TAATCC	0
CG[AT]ATCCCA	MA0023.1-dl_2	2	0.0453531	5.94126	0.999921	8	CGAATCCCA	GGGGATTTCC	0
CG[AT]ATCCCA	MA0190.1-Gsc	-1	0.0464642	6.08681	0.999921	6	CGAATCCCA	TAATCC	0
CG[AT]ATCCCA	MA0532.1-STAT92E	1	0.0481735	6.31073	0.999921	9	CGAATCCCA	CGGAATTCCAGGAAA	0
CG[AT]ATCCCA	MA0016.1-usp	1	0.0519282	6.8026	0.999921	9	CGAATCCCA	CCGTGACCCC	0
CG[AT]ATCCCA	MA0249.1-twi	3	0.0585437	7.66922	0.999921	9	CGAATCCCA	CAACATATGCGA	0
CAAAAAT[AGT]G	MA0445.1-D	1	0.00986857	1.29278	1	9	CAAAAATTG	AGAACAATGGA	0
CAAAAAT[AGT]G	MA0237.2-pan	2	0.0669848	8.77501	1	9	CAAAAATTG	ATCAAAGGAGCCGA	0
CAAAAAT[AGT]G	MA0011.1-br_Z2	-3	0.0719456	9.42487	1	9	CAAAAATTG	AAATAGTA	0
GCCA[AT]AAA[CT]	MA0216.2-CAD	1	0.0005446	0.0713432	0.142309	9	GCCAAAAAC	GGCCATAAAAA	0
GCCA[AT]AAA[CT]	MA0165.1-Abd-B	-1	0.00420362	0.550674	0.549216	7	GCCAAAAAC	TCATAAA	0
GCCA[AT]AAA[CT]	MA0533.1-SU(HW)	0	0.0170588	2.23471	0.997335	9	GCCAAAAAC	GCCCAAAAGTATGCAACAAAT	0
GCCA[AT]AAA[CT]	MA0174.1-CG42234	-1	0.0265317	3.47566	0.997335	7	GCCAAAAAC	TAATAAA	0
GCCA[AT]AAA[CT]	MA0248.1-tup	-1	0.0321226	4.20806	0.997335	7	GCCAAAAAC	CAATTAA	0
GCCA[AT]AAA[CT]	MA0010.1-br_Z1	0	0.0379163	4.96704	0.997335	9	GCCAAAAAC	GTAATAAACAAATC	0
GCCA[AT]AAA[CT]	MA0222.1-exd	1	0.0404302	5.29636	0.997335	7	GCCAAAAAC	TGTCAAAA	0
GCCA[AT]AAA[CT]	MA0182.1-CG4328	-1	0.0408187	5.34724	0.997335	7	GCCAAAAAC	CAATAAA	0
GCCA[AT]AAA[CT]	MA0448.1-H2.0	-1	0.0475314	6.22661	0.997335	7	GCCAAAAAC	TAATTAA	0
GCCA[AT]AAA[CT]	MA0214.1-bsh	-1	0.0499456	6.54287	0.997335	7	GCCAAAAAC	CAATTAA	0
GCCA[AT]AAA[CT]	MA0049.1-hb	0	0.0504621	6.61054	0.997335	9	GCCAAAAAC	GCATAAAAAA	0
GCCA[AT]AAA[CT]	MA0243.1-sd	0	0.0636551	8.33882	0.997335	9	GCCAAAAAC	CCGAGGAATGTC	0
CGAC[GA]CG[CT]C	MA0449.1-h	1	0.00363817	0.4766	0.471477	9	CGACACGCC	GGCACGTGCC	0
CGAC[GA]CG[CT]C	MA0535.1-Mad	4	0.0088042	1.15335	0.524523	9	CGACACGCC	CGGCGGCGGCGCCTG	0
CGAC[GA]CG[CT]C	MA0530.1-CNC::maf-S	3	0.0147601	1.93357	0.524523	9	CGACACGCC	ATTTGCCGAGTCATC	0
CGAC[GA]CG[CT]C	MA0204.1-Six4	0	0.0152161	1.99331	0.524523	6	CGACACGCC	TGATAC	0
CGAC[GA]CG[CT]C	MA0450.1-hkb	0	0.01619	2.12089	0.524523	9	CGACACGCC	TCACGCCCC	0
CGAC[GA]CG[CT]C	MA0252.1-vis	0	0.0237285	3.10844	0.680122	6	CGACACGCC	TGACAG	0
CGAC[GA]CG[CT]C	MA0213.1-brk	2	0.0262409	3.43756	0.680122	6	CGACACGCC	GGCGCCAG	0
CGAC[GA]CG[CT]C	MA0456.1-opa	-1	0.0350052	4.58569	0.794015	8	CGACACGCC	GACCCCCCGCTG	0
CGAC[GA]CG[CT]C	MA0016.1-usp	1	0.0398258	5.21718	0.794015	9	CGACACGCC	CCGTGACCCC	0
CGAC[GA]CG[CT]C	MA0249.1-twi	0	0.0433467	5.67841	0.802482	9	CGACACGCC	CAACATATGCGA	0
CGAC[GA]CG[CT]C	MA0237.2-pan	3	0.0570754	7.47687	0.897519	9	CGACACGCC	ATCAAAGGAGCCGA	0
CGAC[GA]CG[CT]C	MA0227.1-hth	0	0.06154	8.06174	0.897519	6	CGACACGCC	TGACAG	0

53

Cluster 4

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
AGC[AGT]A[AG]ACC	MA0012.1-br_Z3	2	0.00936182	1.2264	0.99989	9	AGCTAAACC	TAAACTAAAAG	0
AGC[AGT]A[AG]ACC	MA0023.1-dl_2	-1	0.0111236	1.4572	0.99989	8	AGCTAAACC	GGAAAACCCC	0
AGC[AGT]A[AG]ACC	MA0022.1-dl_1	1	0.0215727	2.82603	0.99989	9	AGCTAAACC	CGGAAAAACCCC	0
AGC[AGT]A[AG]ACC	MA0016.1-usp	-1	0.022138	2.90008	0.99989	8	AGCTAAACC	CCGTGACCCC	0
AGC[AGT]A[AG]ACC	MA0451.1-kni	1	0.0447564	5.86309	0.99989	9	AGCTAAACC	AAACTAGAGCAC	0
AGC[AGT]A[AG]ACC	MA0458.1-slp1	1	0.0487851	6.39085	0.99989	9	AGCTAAACC	AATGTAAACAA	0
AGC[AGT]A[AG]ACC	MA0531.1-CTCF	0	0.0545102	7.14084	0.99989	9	AGCTAAACC	GGCGCCATCTAGCGG	0
CTGTT[CG]AAA	MA0218.1-ct	-2	0.00257112	0.336817	0.673635	6	CTGTTCAAA	GTTCAA	0
CTGTT[CG]AAA	MA0253.1-vnd	0	0.0180991	2.37098	0.999966	9	CTGTTCAAA	CACTTGAAA	0
CTGTT[CG]AAA	MA0217.1-caup	-1	0.0457561	5.99404	0.999966	5	CTGTTCAAA	TGTTA	0
CTGTT[CG]AAA	MA0210.1-ara	-1	0.0536271	7.02515	0.999966	5	CTGTTCAAA	TGTTA	0
CTGTT[CG]AAA	MA0247.2-tin	1	0.0563137	7.3771	0.999966	9	CTGTTCAAA	CCACTTGAAA	0
CTGTT[CG]AAA	MA0173.1-CG11617	0	0.0680985	8.9209	0.999966	7	CTGTTCAAA	ATGTTAA	0
CTGTT[CG]AAA	MA0244.1-slbo	-1	0.0692834	9.07613	0.999966	8	CTGTTCAAA	TTTGCAAT	0
CTGTT[CG]AAA	MA0233.1-mirr	-1	0.0696362	9.12234	0.999966	5	CTGTTCAAA	TGTTT	0
AAA[AT]ATGGG	MA0445.1-D	2	0.00798203	1.04565	1	9	AAAAATGGG	AGAACAATGGA	0
AAA[AT]ATGGG	MA0249.1-twi	2	0.00833386	1.09174	1	9	AAAAATGGG	CAACATATGCGA	0
AAA[AT]ATGGG	MA0452.2-KR	0	0.0262607	3.44015	1	9	AAAAATGGG	GAAAAAGGGTTAAA	0
AAA[AT]ATGGG	MA0451.1-kni	-1	0.0672431	8.80885	1	8	AAAAATGGG	AAACTAGAGCAC	0
GAC[TA][TG]CCC	MA0023.1-dl_2	1	0.00033751	0.0442142	0.088194	8	GACAGCCC	GGAAAACCCC	0
GAC[TA][TG]CCC	MA0456.1-opa	0	0.0041434	0.542785	0.541348	8	GACAGCCC	GACCCCCCGCTG	0
GAC[TA][TG]CCC	MA0227.1-hth	1	0.00880245	1.15312	0.766713	5	GACAGCCC	TGACAG	0
GAC[TA][TG]CCC	MA0243.1-sd	0	0.0156771	2.0537	0.79075	8	GACAGCCC	GACATTCCTCGA	0
GAC[TA][TG]CCC	MA0207.1-achi	1	0.0190481	2.4953	0.79075	5	GACAGCCC	TGACAG	0
GAC[TA][TG]CCC	MA0252.1-vis	1	0.0222068	2.90909	0.79075	5	GACAGCCC	TGACAG	0
GAC[TA][TG]CCC	MA0532.1-STAT92E	7	0.0242091	3.17139	0.79075	8	GACAGCCC	TTTCCTGGAATTCCG	0
GAC[TA][TG]CCC	MA0022.1-dl_1	3	0.0469718	6.1533	0.997348	8	GACAGCCC	CGGAAAAACCCC	0
GAC[TA][TG]CCC	MA0026.1-Eip74EF	-2	0.0527523	6.91055	0.997348	6	GACAGCCC	CTTCCGG	0
GAC[TA][TG]CCC	MA0530.1-CNC::maf-S	3	0.0561806	7.35966	0.997348	8	GACAGCCC	GATGACTCGGCAAAT	0
GAC[TA][TG]CCC	MA0450.1-hkb	1	0.0630575	8.26053	0.997348	8	GACAGCCC	TCACGCCCC	0
GAC[TA][TG]CCC	MA0443.1-btd	-1	0.068128	8.92477	0.997348	7	GACAGCCC	TCCGCCCCCT	0
CTCCT[GT][AT]GC	MA0453.1-nub	0	0.00198801	0.260429	0.520858	9	CTCCTGAGC	CTAATTTGCATA	0
CTCCT[GT][AT]GC	MA0237.2-pan	4	0.00574639	0.752778	0.571463	9	CTCCTGAGC	TCGGCTCCTTTGAT	0
CTCCT[GT][AT]GC	MA0086.1-sna	0	0.00654347	0.857195	0.571463	6	CTCCTGAGC	CACCTG	0
CTCCT[GT][AT]GC	MA0460.1-ttk	3	0.0222605	2.91613	0.999998	6	CTCCTGAGC	ATTATCCTT	0
CTCCT[GT][AT]GC	MA0454.1-odd	3	0.0266833	3.49551	0.999998	8	CTCCTGAGC	CTGCTACTGTT	0
CTCCT[GT][AT]GC	MA0256.1-zen	-1	0.0310142	4.06286	0.999998	7	CTCCTGAGC	TCATTAG	0
CTCCT[GT][AT]GC	MA0533.1-SU(HW)	12	0.0569466	7.46	0.999998	9	CTCCTGAGC	ATTTGTTGCATACTTTTGGGC	0
CTCCT[GT][AT]GC	MA0221.1-eve	-1	0.0620574	8.12953	0.999998	7	CTCCTGAGC	TCATTAG	0
CTCCT[GT][AT]GC	MA0247.2-tin	1	0.063142	8.2716	0.999998	9	CTCCTGAGC	TTTCAAGTGG	0
CTCCT[GT][AT]GC	MA0452.2-KR	5	0.0691996	9.06515	0.999998	9	CTCCTGAGC	TTTAACCCTTTTTC	0
CTCCT[GT][AT]GC	MA0026.1-Eip74EF	1	0.0711321	9.31831	0.999998	6	CTCCTGAGC	CTTCCGG	0
CAG[AT][AC]GCTG	MA0454.1-odd	2	0.00059096	0.0774151	0.15483	9	CAGAAGCTG	AACAGTAGCAG	0
CAG[AT][AC]GCTG	MA0535.1-Mad	0	0.012306	1.61209	0.826167	9	CAGAAGCTG	CAGGCGCCGCCGCCG	0
CAG[AT][AC]GCTG	MA0126.1-ovo	-1	0.0126132	1.65233	0.826167	8	CAGAAGCTG	AGTAACAGT	0
CAG[AT][AC]GCTG	MA0239.1-prd	-1	0.0126132	1.65233	0.826167	8	CAGAAGCTG	AGTAACAGT	0
CAG[AT][AC]GCTG	MA0531.1-CTCF	1	0.0320699	4.20115	0.999972	9	CAGAAGCTG	CCGCTAGATGGCGCC	0
CAG[AT][AC]GCTG	MA0237.2-pan	4	0.047876	6.27176	0.999972	9	CAGAAGCTG	ATCAAAGGAGCCGA	0
CAG[AT][AC]GCTG	MA0213.1-brk	0	0.0576583	7.55324	0.999972	8	CAGAAGCTG	CTGGCGCC	0
CAG[AT][AC]GCTG	MA0249.1-twi	3	0.0715856	9.37772	0.999972	9	CAGAAGCTG	TCGCATATGTTG	0

Continued (cluster 4)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
GG[AT]AGCAC	MA0454.1-odd	3	0.00114668	0.150214	0.300429	8	GGAAGCAC	AACAGTAGCAG	0
GG[AT]AGCAC	MA0022.1-dl_1	1	0.0137643	1.80312	0.999981	8	GGAAGCAC	CGGAAAAACCCC	0
GG[AT]AGCAC	MA0246.1-so	-1	0.0213325	2.79456	0.999981	6	GGAAGCAC	GTATCA	0
GG[AT]AGCAC	MA0026.1-Eip74EF	2	0.0288332	3.77714	0.999981	5	GGAAGCAC	CCGGAAG	0
GG[AT]AGCAC	MA0204.1-Six4	-1	0.0444853	5.82757	0.999981	6	GGAAGCAC	GTATCA	0
GG[AT]AGCAC	MA0458.1-slp1	3	0.0494549	6.47859	0.999981	8	GGAAGCAC	AATGTAAACAA	0
GG[AT]AGCAC	MA0531.1-CTCF	6	0.0577218	7.56155	0.999981	8	GGAAGCAC	CCGCTAGATGGCGCC	0
GG[AT]AGCAC	MA0530.1-CNC::maf-S	5	0.0630498	8.25953	0.999981	8	GGAAGCAC	GATGACTCGGCAAAT	0
GG[AT]AGCAC	MA0023.1-dl_2	0	0.0633048	8.29292	0.999981	8	GGAAGCAC	GGAAAACCCC	0
GG[AT]AGCAC	MA0193.1-Lag1	2	0.0693653	9.08685	0.999981	5	GGAAGCAC	TTGGTAG	0
GG[AT]AGCAC	MA0016.1-usp	0	0.0720737	9.44165	0.999981	8	GGAAGCAC	GGGGTCACGG	0
CAAGA[GC]CAC	MA0451.1-kni	3	0.00089085	0.116701	0.233402	9	CAAGACCAC	AAACTAGAGCAC	0
CAAGA[GC]CAC	MA0454.1-odd	2	0.0192604	2.52311	1	9	CAAGACCAC	AACAGTAGCAG	0
CAAGA[GC]CAC	MA0016.1-usp	1	0.027382	3.58704	1	9	CAAGACCAC	CCGTGACCCC	0
CAAGA[GC]CAC	MA0023.1-dl_2	1	0.0301545	3.95024	1	9	CAAGACCAC	GGAAAACCCC	0
CAAGA[GC]CAC	MA0205.1-Trl	1	0.0353409	4.62966	1	9	CAAGACCAC	GAGAGAGCAA	0
CAAGA[GC]CAC	MA0237.2-pan	2	0.0655996	8.59355	1	9	CAAGACCAC	ATCAAAGGAGCCGA	0
[CT]AATAAAAA	MA0216.2-CAD	2	0.00137766	0.180473	0.345188	9	CAATAAAAA	GGCCATAAAAA	0
[CT]AATAAAAA	MA0049.1-hb	0	0.00321931	0.42173	0.403318	9	CAATAAAAA	GCATAAAAAA	0
[CT]AATAAAAA	MA0182.1-CG4328	0	0.0103646	1.35776	0.657881	7	CAATAAAAA	CAATAAA	0
[CT]AATAAAAA	MA0174.1-CG42234	0	0.0130559	1.71033	0.657881	7	CAATAAAAA	TAATAA	0
[CT]AATAAAAA	MA0165.1-Abd-B	0	0.0151465	1.98419	0.657881	7	CAATAAAAA	TCATAAA	0
[CT]AATAAAAA	MA0094.2-Ubx	0	0.0157537	2.06374	0.657881	8	CAATAAAAA	TAATTAAA	0
[CT]AATAAAAA	MA0448.1-H2.0	0	0.0233164	3.05445	0.705839	7	CAATAAAAA	TAATTAA	0
[CT]AATAAAAA	MA0013.1-br_Z4	0	0.0265715	3.48087	0.705839	9	CAATAAAAA	TAGTAAACAAA	0
[CT]AATAAAAA	MA0176.1-CG15696	0	0.0276612	3.62362	0.705839	7	CAATAAAAA	CAATTAA	0
[CT]AATAAAAA	MA0010.1-br_Z1	1	0.0393324	5.15254	0.705839	9	CAATAAAAA	GTAATAAACAAATC	0
[CT]AATAAAAA	MA0183.1-CG7056	0	0.0400088	5.24115	0.705839	8	CAATAAAAA	TAATTAAA	0
[CT]AATAAAAA	MA0180.1-Vsx2	1	0.0475186	6.22494	0.705839	8	CAATAAAAA	CTAATTAAA	0
[CT]AATAAAAA	MA0226.1-hbn	0	0.0521382	6.83011	0.705839	7	CAATAAAAA	TAATTAA	0
[CT]AATAAAAA	MA0178.1-CG32105	0	0.0538154	7.04982	0.705839	7	CAATAAAAA	TAATTAA	0
[CT]AATAAAAA	MA0229.1-inv	0	0.059183	7.75297	0.705839	8	CAATAAAAA	TAATTAGA	0
[CT]AATAAAAA	MA0240.1-repo	0	0.0593205	7.77098	0.705839	7	CAATAAAAA	TAATTAA	0
[CT]AATAAAAA	MA0179.1-CG32532	0	0.0612143	8.01908	0.705839	7	CAATAAAAA	TAATTAA	0
[CT]AATAAAAA	MA0444.1-CG34031	0	0.0651098	8.52938	0.705839	7	CAATAAAAA	CAATTAA	0
[CT]AATAAAAA	MA0245.1-slou	0	0.0671135	8.79186	0.705839	7	CAATAAAAA	CAATTAA	0
[CT]AATAAAAA	MA0250.1-unc-4	0	0.0671135	8.79186	0.705839	7	CAATAAAAA	CAATTAA	0
[CT]AATAAAAA	MA0170.1-C15	0	0.069195	9.06454	0.705839	7	CAATAAAAA	TAATTAA	0
[CT]AATAAAAA	MA0192.1-Hmx	0	0.069195	9.06454	0.705839	7	CAATAAAAA	CAATTAA	0
[CT]AATAAAAA	MA0196.1-NK7.1	0	0.069195	9.06454	0.705839	7	CAATAAAAA	CAATTAA	0
TGCC[AC][AC]AAA	MA0533.1-SU(HW)	-1	0.002056	0.269336	0.538673	8	TGCCAAAAA	GCCCAAAAGTATGCAACAAAT	0
TGCC[AC][AC]AAA	MA0216.2-CAD	0	0.00916569	1.20071	0.940347	9	TGCCAAAAA	GGCCATAAAAA	0
TGCC[AC][AC]AAA	MA0222.1-exd	0	0.0116286	1.52334	0.940347	8	TGCCAAAAA	TGTCAAAA	0
TGCC[AC][AC]AAA	MA0049.1-hb	-1	0.0143564	1.88069	0.940347	8	TGCCAAAAA	GCATAAAAAA	0
TGCC[AC][AC]AAA	MA0244.1-slbo	1	0.0233241	3.05546	0.999996	7	TGCCAAAAA	ATTGCAAA	0
TGCC[AC][AC]AAA	MA0193.1-Lag1	-2	0.0360244	4.7192	0.999996	7	TGCCAAAAA	CTACCAA	0
TGCC[AC][AC]AAA	MA0451.1-kni	1	0.0387587	5.07738	0.999996	9	TGCCAAAAA	GTGCTCTAGTTT	0
TGCC[AC][AC]AAA	MA0207.1-achi	1	0.0538148	7.04973	0.999996	5	TGCCAAAAA	CTGTCA	0
TGCC[AC][AC]AAA	MA0252.1-vis	1	0.0538148	7.04973	0.999996	5	TGCCAAAAA	CTGTCA	0
TGCC[AC][AC]AAA	MA0227.1-hth	1	0.0603534	7.9063	0.999996	5	TGCCAAAAA	CTGTCA	0
TGCC[AC][AC]AAA	MA0532.1-STAT92E	1	0.0763009	9.99542	0.999996	9	TGCCAAAAA	TTTCCTGGAATTCCG	0

Continued (cluster 4)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
GAGGA[ACT]AAA	MA0460.1-ttk	0	0.00193494	0.253478	0.506955	9	GAGGATAAA	AAGGATAAT	0
GAGGA[ACT]AAA	MA0201.1-Ptx1	-2	0.0185416	2.42895	1	7	GAGGATAAA	GGATTAA	0
GAGGA[ACT]AAA	MA0049.1-hb	0	0.0187011	2.44984	1	9	GAGGATAAA	GCATAAAAAA	0
GAGGA[ACT]AAA	MA0212.1-bcd	-2	0.0445568	5.83695	1	6	GAGGATAAA	GGATTA	0
GAGGA[ACT]AAA	MA0022.1-dl_1	-1	0.0483248	6.33055	1	8	GAGGATAAA	CGGAAAAACCCC	0
GAGGA[ACT]AAA	MA0234.1-oc	-2	0.0704877	9.23389	1	6	GAGGATAAA	GGATTA	0
AT[AG]ACTCGA	MA0255.1-z	1	0.00923761	1.21013	0.999976	9	ATAACTCGA	AATCACTCAA	0
AT[AG]ACTCGA	MA0530.1-CNC::maf-S	1	0.0148312	1.94288	0.999976	9	ATAACTCGA	GATGACTCGGCAAAT	0
AT[AG]ACTCGA	MA0229.1-inv	-1	0.0672719	8.81262	0.999976	8	ATAACTCGA	TAATTAGA	0
AT[AG]ACTCGA	MA0183.1-CG7056	-1	0.0750417	9.83046	0.999976	8	ATAACTCGA	TAATTAAA	0
CTT[TA]G[CG]CGA	MA0185.1-Deaf1	-4	0.00643037	0.842379	0.99938	5	CTTAGCCGA	CACGAA	0
CTT[TA]G[CG]CGA	MA0222.1-exd	0	0.0099567	1.30433	0.99938	8	CTTAGCCGA	TTTTGACA	0
CTT[TA]G[CG]CGA	MA0249.1-twi	3	0.0322143	4.22008	0.99938	9	CTTAGCCGA	CAACATATGCGA	0
CTT[TA]G[CG]CGA	MA0446.1-fkh	1	0.0589301	7.71985	0.99938	9	CTTAGCCGA	TGTTTGCTTAA	0
CTT[TA]G[CG]CGA	MA0530.1-CNC::maf-S	0	0.0675764	8.85251	0.99938	9	CTTAGCCGA	ATTTGCCGAGTCATC	0

Cluster 7

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
GCC[CGT][AC]AGCA	MA0530.1-CNC::maf-S	3	0.00082479	0.108047	0.214693	9	GCCTAAGCA	GATGACTCGGCAAAT	0
GCC[CGT][AC]AGCA	MA0535.1-Mad	5	0.00297963	0.390331	0.258534	9	GCCTAAGCA	CAGGCGCCGCCGCCG	0
GCC[CGT][AC]AGCA	MA0458.1-slp1	1	0.025902	3.39316	0.993512	9	GCCTAAGCA	AATGTAAACAA	0
GCC[CGT][AC]AGCA	MA0185.1-Deaf1	0	0.0318103	4.16715	0.993512	6	GCCTAAGCA	CACGAA	0
GCC[CGT][AC]AGCA	MA0213.1-brk	1	0.0451722	5.91756	0.993512	7	GCCTAAGCA	GGCGCCAG	0
GCC[CGT][AC]AGCA	MA0454.1-odd	1	0.0484922	6.35248	0.993512	9	GCCTAAGCA	AACAGTAGCAG	0
GCC[CGT][AC]AGCA	MA0049.1-hb	0	0.0548848	7.1899	0.993512	9	GCCTAAGCA	GCATAAAAAA	0
GCC[CGT][AC]AGCA	MA0446.1-fkh	2	0.0571491	7.48654	0.993512	9	GCCTAAGCA	TTAAGCAAACA	0
GCC[CGT][AC]AGCA	MA0456.1-opa	2	0.0712192	9.32971	0.993512	9	GCCTAAGCA	GACCCCCCGCTG	0
T[CA]CTGTCGG	MA0207.1-achi	-2	0.00161818	0.211982	0.224375	6	TACTGTCGG	CTGTCA	0
T[CA]CTGTCGG	MA0227.1-hth	-2	0.00258595	0.338759	0.224375	6	TACTGTCGG	CTGTCA	0
T[CA]CTGTCGG	MA0252.1-vis	-2	0.00258595	0.338759	0.224375	6	TACTGTCGG	CTGTCA	0
T[CA]CTGTCGG	MA0026.1-Eip74EF	-2	0.010332	1.35349	0.672359	7	TACTGTCGG	CTTCCGG	0
T[CA]CTGTCGG	MA0533.1-SU(HW)	10	0.0142319	1.86437	0.740915	9	TACTGTCGG	ATTTGTTGCATACTTTTGGGC	0
T[CA]CTGTCGG	MA0531.1-CTCF	6	0.017642	2.31111	0.765375	9	TACTGTCGG	GGCGCCATCTAGCGG	0
T[CA]CTGTCGG	MA0536.1-pnr	0	0.0278462	3.64785	0.841371	9	TACTGTCGG	AACTATCGATA	0
T[CA]CTGTCGG	MA0085.1-Su(H)	-2	0.0316586	4.14728	0.841371	7	TACTGTCGG	CTGTGGGAAACGAGAT	0
T[CA]CTGTCGG	MA0213.1-brk	-2	0.0323229	4.2343	0.841371	7	TACTGTCGG	CTGGCGCC	0
T[CA]CTGTCGG	MA0529.1-BEAF-32	4	0.0391851	5.13325	0.927267	9	TACTGTCGG	TCGAAACTATCGATA	0
T[CA]CTGTCGG	MA0205.1-Trl	1	0.0561939	7.36141	0.99351	9	TACTGTCGG	TTGCTCTCTC	0
T[CA]CTGTCGG	MA0454.1-odd	4	0.0571685	7.48907	0.99351	7	TACTGTCGG	CTGCTACTGTT	0
T[CA]CTGTCGG	MA0185.1-Deaf1	-4	0.0621735	8.14473	0.99351	5	TACTGTCGG	TTCGGG	0
AAAAGT[GA]A[AG]	MA0459.1-tll	0	0.00158734	0.207942	0.415883	9	AAAAGTAAA	AAAAGTCAAA	0
AAAAGT[GA]A[AG]	MA0533.1-SU(HW)	4	0.0251208	3.29082	1	9	AAAAGTAAA	GCCCAAAAGTATGCAACAAAT	0
AAAAGT[GA]A[AG]	MA0012.1-br_Z3	0	0.0569901	7.46571	1	9	AAAAGTAAA	TAAACTAAAAG	0
TCAC[ACG]TGAA	MA0086.1-sna	-1	0.00017472	0.0228885	0.0457771	6	TCACGTGAA	CACCTG	0
TCAC[ACG]TGAA	MA0253.1-vnd	-1	0.00522902	0.685001	0.685001	8	TCACGTGAA	CACTTGAAA	0
TCAC[ACG]TGAA	MA0247.2-tin	0	0.0136853	1.79278	0.999997	9	TCACGTGAA	CCACTTGAAA	0
TCAC[ACG]TGAA	MA0449.1-h	1	0.0312842	4.09824	0.999997	9	TCACGTGAA	GGCACGTGCC	0
TCAC[ACG]TGAA	MA0211.1-bap	-1	0.040823	5.34781	0.999997	7	TCACGTGAA	CACTTAA	0
TCAC[ACG]TGAA	MA0185.1-Deaf1	-3	0.0745144	9.76139	0.999997	6	TCACGTGAA	CACGAA	0
ATT[CA]C[GA]GAC	MA0244.1-slbo	0	0.00535353	0.701313	0.91732	8	ATTACAGAC	ATTGCAAA	0
ATT[CA]C[GA]GAC	MA0026.1-Eip74EF	0	0.0168439	2.20655	0.91732	7	ATTACAGAC	CTTCCGG	0
ATT[CA]C[GA]GAC	MA0447.1-gt	0	0.0239956	3.14342	0.91732	9	ATTACAGAC	ATTACGTAAT	0
ATT[CA]C[GA]GAC	MA0224.1-exex	2	0.0560661	7.34465	0.91732	5	ATTACAGAC	TAATTAC	0
ATT[CA]C[GA]GAC	MA0243.1-sd	3	0.0584611	7.65841	0.91732	9	ATTACAGAC	GACATTCCTCGA	0
ATT[CA]C[GA]GAC	MA0187.1-Dll	2	0.0727447	9.52956	0.91732	5	ATTACAGAC	TAATTAC	0
ATT[CA]C[GA]GAC	MA0227.1-hth	-1	0.0751949	9.85053	0.91732	6	ATTACAGAC	TGACAG	0

Continued (cluster 7)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
AAACCG[AC]AC	MA0242.1-run::Bgb	0	0.00023245	0.0304502	0.0609005	9	AAACCGAAC	TAACCGCAA	0
AAACCG[AC]AC	MA0185.1-Deaf1	-2	0.0254539	3.33446	1	6	AAACCGAAC	CACGAA	0
AAACCG[AC]AC	MA0085.1-Su(H)	6	0.0357071	4.67763	1	9	AAACCGAAC	CTGTGGGAAACGAGAT	0
AAACCG[AC]AC	MA0012.1-br_Z3	1	0.054003	7.07439	1	9	AAACCGAAC	TAAACTAAAAG	0
AAACCG[AC]AC	MA0126.1-ovo	2	0.0633823	8.30308	1	7	AAACCGAAC	AGTAACAGT	0
AAACCG[AC]AC	MA0239.1-prd	2	0.0633823	8.30308	1	7	AAACCGAAC	AGTAACAGT	0
AAACCG[AC]AC	MA0234.1-oc	1	0.0658534	8.6268	1	5	AAACCGAAC	TAATCC	0
AAACCG[AC]AC	MA0011.1-br_Z2	0	0.0658546	8.62695	1	8	AAACCGAAC	AAATAGTA	0
AAACCG[AC]AC	MA0023.1-dl_2	3	0.0667748	8.74749	1	7	AAACCGAAC	GGAAAACCCC	0
AAACCG[AC]AC	MA0212.1-bcd	1	0.0703038	9.20979	1	5	AAACCGAAC	TAATCC	0
TGG[AT]AGCAA	MA0193.1-Lag1	1	0.00853951	1.11868	1	6	TGGAAGCAA	TTGGTAG	0
TGG[AT]AGCAA	MA0454.1-odd	2	0.0128061	1.67759	1	9	TGGAAGCAA	AACAGTAGCAG	0
TGG[AT]AGCAA	MA0246.1-so	-2	0.0243276	3.18691	1	6	TGGAAGCAA	GTATCA	0
TGG[AT]AGCAA	MA0458.1-slp1	2	0.0462564	6.05959	1	9	TGGAAGCAA	AATGTAAACAA	0
TGG[AT]AGCAA	MA0204.1-Six4	-2	0.0565283	7.40521	1	6	TGGAAGCAA	GTATCA	0
TGG[AT]AGCAA	MA0013.1-br_Z4	1	0.0575608	7.54046	1	9	TGGAAGCAA	TAGTAAACAAA	0
CGCG[CT]G[AT]A	MA0450.1-hkb	2	0.00507235	0.664478	0.486854	7	CGCGCGAA	GGGGCGTGA	0
CGCG[CT]G[AT]A	MA0449.1-h	2	0.00558285	0.731353	0.486854	8	CGCGCGAA	GGCACGTGCC	0
CGCG[CT]G[AT]A	MA0247.2-tin	1	0.0268954	3.5233	0.998527	8	CGCGCGAA	CCACTTGAAA	0
CGCG[CT]G[AT]A	MA0255.1-z	1	0.0282561	3.70155	0.998527	8	CGCGCGAA	TTGAGTGATT	0
CGCG[CT]G[AT]A	MA0086.1-sna	0	0.0311725	4.0836	0.998527	6	CGCGCGAA	CAGGTG	0
CGCG[CT]G[AT]A	MA0535.1-Mad	7	0.0401639	5.26148	0.998527	8	CGCGCGAA	CGGCGGCGGCGCCTG	0
CGCG[CT]G[AT]A	MA0185.1-Deaf1	0	0.0413946	5.42269	0.998527	6	CGCGCGAA	TTCGGG	0
CGCG[CT]G[AT]A	MA0213.1-brk	2	0.0471274	6.17369	0.998527	6	CGCGCGAA	CTGGCGCC	0
CGCG[CT]G[AT]A	MA0253.1-vnd	0	0.0587708	7.69898	0.998527	8	CGCGCGAA	CACTTGAAA	0
CGCG[CT]G[AT]A	MA0085.1-Su(H)	1	0.069626	9.12101	0.998527	8	CGCGCGAA	CTGTGGGAAACGAGAT	0
CGCG[CT]G[AT]A	MA0249.1-twi	1	0.0725514	9.50424	0.998527	8	CGCGCGAA	TCGCATATGTTG	0
CT[CG][AGT]CTCTC	MA0205.1-Trl	1	0.00020485	0.0268357	0.0536714	9	CTCTCTCTC	TTGCTCTCTC	0
CT[CG][AGT]CTCTC	MA0255.1-z	1	0.00767218	1.00506	1	9	CTCTCTCTC	AATCACTCAA	0
CT[CG][AGT]CTCTC	MA0207.1-achi	-4	0.0239633	3.1392	1	5	CTCTCTCTC	CTGTCA	0
CT[CG][AGT]CTCTC	MA0213.1-brk	0	0.0268869	3.52218	1	8	CTCTCTCTC	CTGGCGCC	0
CT[CG][AGT]CTCTC	MA0227.1-hth	-4	0.0346058	4.53336	1	5	CTCTCTCTC	CTGTCA	0
CT[CG][AGT]CTCTC	MA0252.1-vis	-4	0.0346058	4.53336	1	5	CTCTCTCTC	CTGTCA	0
CT[CG][AGT]CTCTC	MA0459.1-tll	1	0.0625297	8.19139	1	9	CTCTCTCTC	TTTGACTTTT	0
AAAAC[GC][TC]GA	MA0022.1-dl_1	4	0.00367967	0.482037	0.964074	8	AAAACCCGA	CGGAAAAACCCC	0
AAAAC[GC][TC]GA	MA0023.1-dl_2	2	0.0231848	3.03721	1	8	AAAACCCGA	GGAAAACCCC	0
AAAAC[GC][TC]GA	MA0242.1-run::Bgb	0	0.0407471	5.33787	1	9	AAAACCCGA	TAACCGCAA	0
AAAAC[GC][TC]GA	MA0085.1-Su(H)	6	0.0559522	7.32974	1	9	AAAACCCGA	CTGTGGGAAACGAGAT	0
AAAAC[GC][TC]GA	MA0237.2-pan	5	0.0583876	7.64878	1	9	AAAACCCGA	ATCAAAGGAGCCGA	0
AAAAC[GC][TC]GA	MA0086.1-sna	-2	0.0666882	8.73615	1	6	AAAACCCGA	CACCTG	0
[AT]AACCCGA	MA0022.1-dl_1	5	0.00303275	0.39729	0.539275	7	AAACCCGA	CGGAAAAACCCC	0
[AT]AACCCGA	MA0185.1-Deaf1	-3	0.00534153	0.69974	0.539275	5	AAACCCGA	CACGAA	0
[AT]AACCCGA	MA0242.1-run::Bgb	0	0.00617491	0.808913	0.539275	8	AAACCCGA	TAACCGCAA	0
[AT]AACCCGA	MA0023.1-dl_2	3	0.027448	3.59568	0.999994	7	AAACCCGA	GGAAAACCCC	0
[AT]AACCCGA	MA0086.1-sna	-1	0.034853	4.56574	0.999994	6	AAACCCGA	CACCTG	0
[AT]AACCCGA	MA0452.2-KR	2	0.0400073	5.24096	0.999994	8	AAACCCGA	TTTAACCCTTTTTC	0
[AT]AACCCGA	MA0234.1-oc	0	0.04246	5.56225	0.999994	6	AAACCCGA	TAATCC	0
[AT]AACCCGA	MA0212.1-bcd	0	0.0451182	5.91049	0.999994	6	AAACCCGA	TAATCC	0
[AT]AACCCGA	MA0190.1-Gsc	0	0.055897	7.32251	0.999994	6	AAACCCGA	TAATCC	0
[AT]AACCCGA	MA0026.1-Eip74EF	-3	0.0598787	7.84412	0.999994	5	AAACCCGA	CCGGAAG	0
[AT]AACCCGA	MA0451.1-kni	0	0.0682219	8.93706	0.999994	8	AAACCCGA	AAACTAGAGCAC	0
[AT]AACCCGA	MA0085.1-Su(H)	7	0.0706892	9.26029	0.999994	8	AAACCCGA	CTGTGGGAAACGAGAT	0

Continued (cluster 7)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
CGC[CTG]C[CG]AA	MA0185.1-Deaf1	-2	0.00646332	0.846695	0.774751	6	CGCTCCAA	CACGAA	0
CGC[CTG]C[CG]AA	MA0450.1-hkb	3	0.00778851	1.02029	0.774751	6	CGCTCCAA	TCACGCCCC	0
CGC[CTG]C[CG]AA	MA0443.1-btd	2	0.0103916	1.3613	0.774751	8	CGCTCCAA	TCCGCCCCCT	0
CGC[CTG]C[CG]AA	MA0213.1-brk	0	0.0125681	1.64642	0.774751	8	CGCTCCAA	GGCGCCAG	0
CGC[CTG]C[CG]AA	MA0247.2-tin	1	0.0148469	1.94494	0.774751	8	CGCTCCAA	CCACTTGAAA	0
CGC[CTG]C[CG]AA	MA0255.1-z	3	0.0278072	3.64275	0.99584	7	CGCTCCAA	AATCACTCAA	0
CGC[CTG]C[CG]AA	MA0451.1-kni	1	0.0295109	3.86593	0.99584	8	CGCTCCAA	GTGCTCTAGTTT	0
CGC[CTG]C[CG]AA	MA0531.1-CTCF	0	0.0430307	5.63702	0.99584	8	CGCTCCAA	GGCGCCATCTAGCGG	0
CGC[CTG]C[CG]AA	MA0533.1-SU(HW)	-1	0.0449583	5.88953	0.99584	7	CGCTCCAA	GCCCAAAAGTATGCAACAAAT	0
CGC[CTG]C[CG]AA	MA0449.1-h	4	0.0467139	6.11952	0.99584	6	CGCTCCAA	GGCACGTGCC	0
CGC[CTG]C[CG]AA	MA0530.1-CNC::maf-S	5	0.0542165	7.10236	0.99584	8	CGCTCCAA	GATGACTCGGCAAAT	0
CGC[CTG]C[CG]AA	MA0253.1-vnd	0	0.0563017	7.37552	0.99584	8	CGCTCCAA	CACTTGAAA	0
CGC[CTG]C[CG]AA	MA0535.1-Mad	2	0.0567728	7.43723	0.99584	8	CGCTCCAA	CAGGCGCCGCCGCCG	0
AA[CG]GACGAA	MA0185.1-Deaf1	-3	0.00170108	0.222842	0.445684	6	AACGACGAA	CACGAA	0
AA[CG]GACGAA	MA0460.1-ttk	0	0.0348095	4.56005	1	9	AACGACGAA	AAGGATAAT	0
AA[CG]GACGAA	MA0459.1-tll	1	0.0473805	6.20684	1	9	AACGACGAA	AAAAGTCAAA	0
AA[CG]GACGAA	MA0243.1-sd	-1	0.0583055	7.63802	1	8	AACGACGAA	CCGAGGAATGTC	0
AA[CG]GACGAA	MA0201.1-Ptx1	-2	0.0671579	8.79768	1	7	AACGACGAA	GGATTAA	0
AA[CG]GACGAA	MA0193.1-Lag1	-2	0.0741751	9.71694	1	7	AACGACGAA	CTACCAA	0

Cluster 5

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
ATC[GT]CTCGC	MA0205.1-Trl	1	0.00222209	0.291093	0.582187	9	ATCGCTCGC	TTGCTCTCTC	0
ATC[GT]CTCGC	MA0255.1-z	1	0.00511578	0.670167	0.670167	9	ATCGCTCGC	AATCACTCAA	0
ATC[GT]CTCGC	MA0536.1-pnr	1	0.0105643	1.38392	0.922613	9	ATCGCTCGC	TATCGATAGTT	0
ATC[GT]CTCGC	MA0531.1-CTCF	6	0.0202888	2.65783	0.999999	9	ATCGCTCGC	GGCGCCATCTAGCGG	0
ATC[GT]CTCGC	MA0456.1-opa	1	0.030005	3.93065	0.999999	9	ATCGCTCGC	GACCCCCCGCTG	0
ATC[GT]CTCGC	MA0243.1-sd	3	0.031275	4.09703	0.999999	9	ATCGCTCGC	GACATTCCTCGA	0
ATC[GT]CTCGC	MA0529.1-BEAF-32	1	0.054545	7.14539	0.999999	9	ATCGCTCGC	TATCGATAGTTTCGA	0
ATC[GT]CTCGC	MA0450.1-hkb	-3	0.0662829	8.68305	0.999999	6	ATCGCTCGC	TCACGCCCC	0
[AT]CTGACTG	MA0207.1-achi	-2	0.00396575	0.519513	0.574213	6	ACTGACTG	TGACAG	0
[AT]CTGACTG	MA0227.1-hth	-2	0.00537704	0.704392	0.574213	6	ACTGACTG	TGACAG	0
[AT]CTGACTG	MA0252.1-vis	-2	0.00658461	0.862584	0.574213	6	ACTGACTG	TGACAG	0
[AT]CTGACTG	MA0459.1-tll	0	0.0185479	2.42977	0.854482	8	ACTGACTG	TTTGACTTTT	0
[AT]CTGACTG	MA0126.1-ovo	0	0.019597	2.56721	0.854482	8	ACTGACTG	AGTAACAGT	0
[AT]CTGACTG	MA0239.1-prd	0	0.019597	2.56721	0.854482	8	ACTGACTG	AGTAACAGT	0
[AT]CTGACTG	MA0235.1-onecut	-1	0.0280867	3.67936	0.997275	7	ACTGACTG	TTGATTT	0
[AT]CTGACTG	MA0255.1-z	-1	0.0309032	4.04832	0.997275	7	ACTGACTG	TTGAGTGATT	0
[AT]CTGACTG	MA0530.1-CNC::maf-S	0	0.0527747	6.91349	0.997275	8	ACTGACTG	GATGACTCGGCAAAT	0
[AT]CTGACTG	MA0198.1-OdsH	-1	0.0574307	7.52343	0.997275	7	ACTGACTG	CTAATTA	0
[AT]CTGACTG	MA0175.1-CG13424	-1	0.0623324	8.16554	0.997275	7	ACTGACTG	TTAATTG	0
[AT]CTGACTG	MA0229.1-inv	0	0.0650606	8.52294	0.997275	8	ACTGACTG	TCTAATTA	0
[AT]CTGACTG	MA0250.1-unc-4	-1	0.0677403	8.87398	0.997275	7	ACTGACTG	TTAATTG	0
[AT]CTGACTG	MA0202.1-Rx	-1	0.0706283	9.25231	0.997275	7	ACTGACTG	CTAATTA	0
[AT]CTGACTG	MA0026.1-Eip74EF	-1	0.0738017	9.66803	0.997275	7	ACTGACTG	CTTCCGG	0
G[CT]AACGGAA	MA0126.1-ovo	1	0.0034819	0.456129	0.454921	8	GCAACGGAA	AGTAACAGT	0
G[CT]AACGGAA	MA0239.1-prd	1	0.0034819	0.456129	0.454921	8	GCAACGGAA	AGTAACAGT	0
G[CT]AACGGAA	MA0026.1-Eip74EF	-3	0.00866907	1.13565	0.617529	6	GCAACGGAA	CCGGAAG	0
G[CT]AACGGAA	MA0447.1-gt	0	0.0118162	1.54792	0.617529	9	GCAACGGAA	ATTACGTAAT	0
G[CT]AACGGAA	MA0533.1-SU(HW)	12	0.0244454	3.20235	0.997324	9	GCAACGGAA	GCCCAAAAGTATGCAACAAAT	0
G[CT]AACGGAA	MA0013.1-br_Z4	2	0.04817	6.31026	0.997324	9	GCAACGGAA	TAGTAAACAAA	0
G[CT]AACGGAA	MA0242.1-run::Bgb	-1	0.0557022	7.29698	0.997324	8	GCAACGGAA	TAACCGCAA	0
G[CT]AACGGAA	MA0016.1-usp	3	0.0559145	7.3248	0.997324	7	GCAACGGAA	GGGGTCACGG	0
G[CT]AACGGAA	MA0085.1-Su(H)	6	0.0719381	9.42389	0.997324	9	GCAACGGAA	CTGTGGGAAACGAGAT	0
AA[CA]TGGTG	MA0086.1-sna	-2	0.00969445	1.26997	1	6	AAATGGTG	CAGGTG	0
AA[CA]TGGTG	MA0011.1-br_Z2	0	0.0115568	1.51395	1	8	AAATGGTG	AAATAGTA	0
AA[CA]TGGTG	MA0193.1-Lag1	-2	0.0181596	2.37891	1	6	AAATGGTG	TTGGTAG	0
AA[CA]TGGTG	MA0451.1-kni	1	0.0307756	4.03161	1	8	AAATGGTG	AAACTAGAGCAC	0
AA[CA]TGGTG	MA0188.1-Dr	1	0.0360247	4.71924	1	6	AAATGGTG	TAATTGG	0
AA[CA]TGGTG	MA0249.1-twi	4	0.0591735	7.75173	1	8	AAATGGTG	TCGCATATGTTG	0
AA[CA]TGGTG	MA0452.2-KR	3	0.0662715	8.68157	1	8	AAATGGTG	GAAAAAGGGTTAAA	0
AA[CA]TGGTG	MA0531.1-CTCF	5	0.0745273	9.76308	1	8	AAATGGTG	CCGCTAGATGGCGCC	0
AC[GC]ATGAC	MA0016.1-usp	-1	0.00650383	0.852002	0.999991	7	ACCATGAC	CCGTGACCCC	0
AC[GC]ATGAC	MA0530.1-CNC::maf-S	-2	0.0290851	3.81015	0.999991	6	ACCATGAC	GATGACTCGGCAAAT	0
AC[GC]ATGAC	MA0207.1-achi	-4	0.044261	5.79819	0.999991	4	ACCATGAC	TGACAG	0
AC[GC]ATGAC	MA0227.1-hth	-4	0.044261	5.79819	0.999991	4	ACCATGAC	TGACAG	0
AC[GC]ATGAC	MA0252.1-vis	-4	0.044261	5.79819	0.999991	4	ACCATGAC	TGACAG	0
AC[GC]ATGAC	MA0126.1-ovo	0	0.0486691	6.37565	0.999991	8	ACCATGAC	ACTGTTACT	0
AC[GC]ATGAC	MA0239.1-prd	0	0.0486691	6.37565	0.999991	8	ACCATGAC	ACTGTTACT	0
AC[GC]ATGAC	MA0214.1-bsh	-1	0.050439	6.60752	0.999991	7	ACCATGAC	CAATTAA	0
AC[GC]ATGAC	MA0445.1-D	3	0.057512	7.53407	0.999991	8	ACCATGAC	AGAACAATGGA	0
AC[GC]ATGAC	MA0248.1-tup	-1	0.0661544	8.66622	0.999991	7	ACCATGAC	CAATTAA	0
ACTCA[CG]TC	MA0255.1-z	0	0.00073969	0.0968998	0.1938	8	ACTCACTC	AATCACTCAA	0
ACTCA[CG]TC	MA0205.1-Trl	2	0.00701705	0.919234	0.919234	8	ACTCACTC	TTGCTCTCTC	0

Continuedc (cluster 5)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
TTGATTGA	MA0235.1-onecut	0	0.00042739	0.0559884	0.072328	7	TTGATTGA	TTGATTT	0
TTGATTGA	MA0176.1-CG15696	0	0.00326568	0.427805	0.139226	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0192.1-Hmx	0	0.00326568	0.427805	0.139226	7	TTGATTGA	TTAATTG	0
TTGATTGA	MA0444.1-CG34031	0	0.00411348	0.538866	0.139226	7	TTGATTGA	TTAATTG	0
TTGATTGA	MA0171.1-CG11085	0	0.00433646	0.568077	0.139226	7	TTGATTGA	TTAATTG	0
TTGATTGA	MA0250.1-unc-4	0	0.00503667	0.659804	0.139226	7	TTGATTGA	TTAATTG	0
TTGATTGA	MA0169.1-B-H2	0	0.00577747	0.756848	0.139226	7	TTGATTGA	TTAATTG	0
TTGATTGA	MA0196.1-NK7.1	0	0.00658161	0.86219	0.139226	7	TTGATTGA	TTAATTG	0
TTGATTGA	MA0255.1-z	0	0.00802323	1.05104	0.150864	8	TTGATTGA	TTGAGTGATT	0
TTGATTGA	MA0168.1-B-H1	0	0.00970806	1.27176	0.152411	7	TTGATTGA	TTAATTG	0
TTGATTGA	MA0175.1-CG13424	0	0.0100655	1.31858	0.152411	7	TTGATTGA	TTAATTG	0
TTGATTGA	MA0248.1-tup	0	0.0108073	1.41576	0.152411	7	TTGATTGA	TTAATTG	0
TTGATTGA	MA0226.1-hbn	0	0.0119792	1.56927	0.154392	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0245.1-slou	0	0.012813	1.6785	0.154392	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0179.1-CG32532	0	0.0136847	1.7927	0.154392	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0214.1-bsh	0	0.0171195	2.24265	0.173246	7	TTGATTGA	TTAATTG	0
TTGATTGA	MA0534.1-EcR::usp	3	0.0174034	2.27984	0.173246	8	TTGATTGA	GAGTTCATTGACCTT	0
TTGATTGA	MA0220.1-en	0	0.0194305	2.5454	0.180621	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0170.1-C15	0	0.0206873	2.71004	0.180621	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0182.1-CG4328	0	0.0213461	2.79634	0.180621	7	TTGATTGA	TTTATTA	0
TTGATTGA	MA0240.1-repo	0	0.0265192	3.47402	0.213708	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0251.1-unpg	0	0.0282142	3.69606	0.217032	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0174.1-CG42234	0	0.0328752	4.30665	0.231057	7	TTGATTGA	TTTATTA	0
TTGATTGA	MA0178.1-CG32105	0	0.0328752	4.30665	0.231057	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0181.1-Vsx1	0	0.0354987	4.65033	0.231057	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0208.1-al	-1	0.0354987	4.65033	0.231057	7	TTGATTGA	TAATTAA	0
TTGATTGA	MA0202.1-Rx	0	0.0381692	5.00017	0.236518	7	TTGATTGA	CTAATTA	0
TTGATTGA	MA0172.1-CG11294	-1	0.0410569	5.37846	0.236518	7	TTGATTGA	TAATTAA	0
TTGATTGA	MA0191.1-HGTX	0	0.0425863	5.5788	0.236518	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0457.1-PHDP	0	0.0441271	5.78065	0.236518	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0167.1-Awh	0	0.0457008	5.98681	0.236518	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0094.2-Ubx	1	0.0489845	6.41697	0.236518	7	TTGATTGA	TTTAATTA	0
TTGATTGA	MA0195.1-Lim3	0	0.0490122	6.4206	0.236518	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0194.1-Lim1	0	0.0507561	6.64904	0.236518	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0198.1-OdsH	0	0.0525344	6.88201	0.236518	7	TTGATTGA	CTAATTA	0
TTGATTGA	MA0188.1-Dr	-1	0.054338	7.11827	0.236518	7	TTGATTGA	TAATTGG	0
TTGATTGA	MA0236.1-otp	0	0.0564186	7.39083	0.236518	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0183.1-CG7056	-1	0.0575801	7.54299	0.236518	7	TTGATTGA	TAATTAAA	0
TTGATTGA	MA0448.1-H2.0	0	0.0590646	7.73746	0.236518	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0206.1-abd-A	0	0.0618232	8.09884	0.236518	7	TTGATTGA	TTAATTA	0
TTGATTGA	MA0166.1-Antp	-1	0.0645384	8.45453	0.236518	7	TTGATTGA	TCATTAA	0
TTGATTGA	MA0225.1-ftz	0	0.0698897	9.15555	0.236518	7	TTGATTGA	TTAATGA	0
TTGATTGA	MA0230.1-lab	-1	0.0698897	9.15555	0.236518	7	TTGATTGA	TAATTAA	0
TTGATTGA	MA0186.1-Dfd	-1	0.0724958	9.49694	0.236518	7	TTGATTGA	TCATTAA	0
TTGATTGA	MA0201.1-Ptx1	-1	0.0724958	9.49694	0.236518	7	TTGATTGA	GGATTAA	0

Cluster 2A

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
CA[AC]G[AG]AA	MA0185.1-Deaf1	0	0.0086113	1.12808	1	6	CAAGAAA	CACGAA	0
CA[AC]G[AG]AA	MA0026.1-Eip74EF	-1	0.0421039	5.51561	1	6	CAAGAAA	CCGGAAG	0
CA[AC]G[AG]AA	MA0532.1-STAT92E	8	0.0439938	5.76319	1	7	CAAGAAA	CGGAATTCCAGGAAA	0
CA[AC]G[AG]AA	MA0534.1-EcR::usp	5	0.0467334	6.12208	1	7	CAAGAAA	AAGGTCAATGAACTC	0
CA[AC]G[AG]AA	MA0243.1-sd	1	0.0559571	7.33037	1	7	CAAGAAA	CCGAGGAATGTC	0
CA[AC]G[AG]AA	MA0444.1-CG34031	0	0.0566256	7.41796	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0192.1-Hmx	0	0.0701344	9.18761	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0211.1-bap	0	0.0741061	9.7079	1	7	CAAGAAA	CACTTAA	0
CA[AC]G[AG]AA	MA0171.1-CG11085	0	0.0821092	10.7563	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0182.1-CG4328	0	0.0821092	10.7563	1	7	CAAGAAA	CAATAAA	0
CA[AC]G[AG]AA	MA0016.1-usp	5	0.0894159	11.7135	1	5	CAAGAAA	GGGGTCACGG	0
CA[AC]G[AG]AA	MA0196.1-NK7.1	0	0.0903722	11.8388	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0049.1-hb	1	0.0931079	12.1971	1	7	CAAGAAA	GCATAAAAAA	0
CA[AC]G[AG]AA	MA0250.1-unc-4	0	0.0947115	12.4072	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0176.1-CG15696	0	0.0991875	12.9936	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0169.1-B-H2	0	0.103805	13.5985	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0248.1-tup	0	0.136976	17.9439	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0253.1-vnd	3	0.140216	18.3683	1	6	CAAGAAA	TTTCAAGTG	0
CA[AC]G[AG]AA	MA0168.1-B-H1	0	0.149285	19.5563	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0447.1-gt	2	0.152629	19.9944	1	7	CAAGAAA	ATTACGTAAT	0
CA[AC]G[AG]AA	MA0533.1-SU(HW)	13	0.158487	20.7617	1	7	CAAGAAA	GCCCAAAAGTATGCAACAAAT	0
CA[AC]G[AG]AA	MA0216.2-CAD	2	0.160099	20.973	1	7	CAAGAAA	GGCCATAAAAA	0
CA[AC]G[AG]AA	MA0175.1-CG13424	0	0.173783	22.7655	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0237.2-pan	2	0.187162	24.5183	1	7	CAAGAAA	ATCAAAGGAGCCGA	0
CA[AC]G[AG]AA	MA0193.1-Lag1	1	0.192823	25.2598	1	6	CAAGAAA	CTACCAA	0
CA[AC]G[AG]AA	MA0165.1-Abd-B	0	0.199381	26.1189	1	7	CAAGAAA	TCATAAA	0
CA[AC]G[AG]AA	MA0226.1-hbn	0	0.199381	26.1189	1	7	CAAGAAA	TAATTAA	0
CA[AC]G[AG]AA	MA0247.2-tin	3	0.203944	26.7167	1	7	CAAGAAA	CCACTTGAAA	0
CA[AC]G[AG]AA	MA0174.1-CG42234	0	0.212809	27.878	1	7	CAAGAAA	TAATAAA	0
CA[AC]G[AG]AA	MA0245.1-slou	0	0.212809	27.878	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0179.1-CG32532	0	0.219683	28.7784	1	7	CAAGAAA	TAATTAA	0
CA[AC]G[AG]AA	MA0170.1-C15	0	0.226728	29.7013	1	7	CAAGAAA	TAATTAA	0
CA[AC]G[AG]AA	MA0445.1-D	4	0.230869	30.2438	1	7	CAAGAAA	AGAACAATGGA	0
CA[AC]G[AG]AA	MA0448.1-H2.0	0	0.233937	30.6458	1	7	CAAGAAA	TAATTAA	0
CA[AC]G[AG]AA	MA0458.1-slp1	0	0.248742	32.5852	1	7	CAAGAAA	AATGTAAACAA	0
CA[AC]G[AG]AA	MA0214.1-bsh	0	0.248778	32.59	1	7	CAAGAAA	CAATTAA	0
CA[AC]G[AG]AA	MA0235.1-onecut	0	0.248778	32.59	1	7	CAAGAAA	AAATCAA	0
CA[AC]G[AG]AA	MA0023.1-dl_2	-3	0.255647	33.4898	1	4	CAAGAAA	GGAAAACCCC	0
CA[AC]G[AG]AA	MA0013.1-br_Z4	4	0.267485	35.0406	1	7	CAAGAAA	TAGTAAACAAA	0
CA[AC]G[AG]AA	MA0220.1-en	0	0.280737	36.7766	1	7	CAAGAAA	TAATTAA	0
CA[AC]G[AG]AA	MA0452.2-KR	-1	0.293174	38.4058	1	6	CAAGAAA	GAAAAAGGGTTAAA	0
CA[AC]G[AG]AA	MA0460.1-ttk	0	0.301933	39.5532	1	7	CAAGAAA	AAGGATAAT	0
CA[AC]G[AG]AA	MA0249.1-twi	0	0.314946	41.2579	1	7	CAAGAAA	CAACATATGCGA	0
CA[AC]G[AG]AA	MA0255.1-z	3	0.317243	41.5588	1	7	CAAGAAA	AATCACTCAA	0
CA[AC]G[AG]AA	MA0178.1-CG32105	0	0.32478	42.5462	1	7	CAAGAAA	TAATTAA	0
CA[AC]G[AG]AA	MA0240.1-repo	0	0.32478	42.5462	1	7	CAAGAAA	TAATTAA	0
CA[AC]G[AG]AA	MA0254.1-vvl	0	0.329455	43.1586	1	6	CAAGAAA	TATGCA	0
CA[AC]G[AG]AA	MA0012.1-br_Z3	4	0.329811	43.2053	1	7	CAAGAAA	TAAACTAAAAG	0
CA[AC]G[AG]AA	MA0094.2-Ubx	0	0.332648	43.5769	1	7	CAAGAAA	TAATTAAA	0
CA[AC]G[AG]AA	MA0022.1-dl_1	-2	0.337203	44.1736	1	5	CAAGAAA	CGGAAAAACCCC	0
CA[AC]G[AG]AA	MA0126.1-ovo	-2	0.358015	46.9	1	5	CAAGAAA	AGTAACAGT	0
CA[AC]G[AG]AA	MA0239.1-prd	-2	0.358015	46.9	1	5	CAAGAAA	AGTAACAGT	0
CA[AC]G[AG]AA	MA0251.1-unpg	0	0.364983	47.8127	1	7	CAAGAAA	TAATTAA	0

Continued (cluster 2A)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
G[GC]G[CG]CTAAC	MA0242.1-run::Bgb	2	0.0320396	4.19719	0.993481	7	GCGCCTAAC	TTGCGGTTA	0
G[GC]G[CG]CTAAC	MA0217.1-caup	-5	0.0399186	5.22934	0.993481	4	GCGCCTAAC	TAACA	0
G[GC]G[CG]CTAAC	MA0450.1-hkb	1	0.0425721	5.57695	0.993481	8	GCGCCTAAC	GGGGCGTGA	0
G[GC]G[CG]CTAAC	MA0023.1-dl_2	0	0.0427276	5.59732	0.993481	9	GCGCCTAAC	GGGGATTTCC	0
G[GC]G[CG]CTAAC	MA0460.1-ttk	0	0.0526779	6.90081	0.993481	9	GCGCCTAAC	AAGGATAAT	0
G[GC]G[CG]CTAAC	MA0016.1-usp	-1	0.0545164	7.14165	0.993481	8	GCGCCTAAC	GGGGTCACGG	0
G[GC]G[CG]CTAAC	MA0452.2-KR	5	0.0592181	7.75757	0.993481	9	GCGCCTAAC	GAAAAAGGGTTAAA	0
G[GC]G[CG]CTAAC	MA0022.1-dl_1	0	0.0654487	8.57378	0.993481	9	GCGCCTAAC	CGGAAAAACCCC	0
G[GC]G[CG]CTAAC	MA0446.1-fkh	1	0.0711698	9.32324	0.993481	9	GCGCCTAAC	TTAAGCAAACA	0
G[GC]G[CG]CTAAC	MA0201.1-Ptx1	-1	0.0764406	10.0137	0.993481	7	GCGCCTAAC	GGATTAA	0
G[GC]G[CG]CTAAC	MA0529.1-BEAF-32	2	0.083334	10.9168	0.993481	9	GCGCCTAAC	TCGAAACTATCGATA	0
G[GC]G[CG]CTAAC	MA0233.1-mirr	-5	0.0984797	12.9008	0.993481	4	GCGCCTAAC	AAACA	0
G[GC]G[CG]CTAAC	MA0213.1-brk	3	0.0988267	12.9463	0.993481	5	GCGCCTAAC	CTGGCGCC	0
G[GC]G[CG]CTAAC	MA0536.1-pnr	-2	0.0996989	13.0606	0.993481	7	GCGCCTAAC	AACTATCGATA	0
G[GC]G[CG]CTAAC	MA0443.1-btd	2	0.101751	13.3294	0.993481	8	GCGCCTAAC	AGGGGGCGGA	0
G[GC]G[CG]CTAAC	MA0210.1-ara	-5	0.102006	13.3627	0.993481	4	GCGCCTAAC	TAACA	0
G[GC]G[CG]CTAAC	MA0216.2-CAD	0	0.105218	13.7835	0.993481	9	GCGCCTAAC	GGCCATAAAAA	0
G[GC]G[CG]CTAAC	MA0222.1-exd	-1	0.111024	14.5442	0.993481	8	GCGCCTAAC	TGTCAAAA	0
G[GC]G[CG]CTAAC	MA0449.1-h	1	0.116249	15.2287	0.993481	9	GCGCCTAAC	GGCACGTGCC	0
G[GC]G[CG]CTAAC	MA0185.1-Deaf1	-2	0.138652	18.1634	0.993481	6	GCGCCTAAC	CACGAA	0
G[GC]G[CG]CTAAC	MA0211.1-bap	-1	0.157149	20.5865	0.993481	7	GCGCCTAAC	CACTTAA	0
G[GC]G[CG]CTAAC	MA0454.1-odd	-1	0.159199	20.8551	0.993481	8	GCGCCTAAC	CTGCTACTGTT	0
G[GC]G[CG]CTAAC	MA0231.1-lbe	-5	0.162145	21.241	0.993481	4	GCGCCTAAC	TAACTA	0
G[GC]G[CG]CTAAC	MA0085.1-Su(H)	2	0.166795	21.8501	0.993481	9	GCGCCTAAC	CTGTGGGAAACGAGAT	0
G[GC]G[CG]CTAAC	MA0234.1-oc	-1	0.177388	23.2379	0.993481	6	GCGCCTAAC	GGATTA	0
G[GC]G[CG]CTAAC	MA0212.1-bcd	-2	0.188117	24.6433	0.993481	6	GCGCCTAAC	GGATTA	0
G[GC]G[CG]CTAAC	MA0237.2-pan	0	0.194757	25.5131	0.993481	9	GCGCCTAAC	TCGGCTCCTTTGAT	0
G[GC]G[CG]CTAAC	MA0049.1-hb	-2	0.19869	26.0283	0.993481	7	GCGCCTAAC	GCATAAAAAA	0
G[GC]G[CG]CTAAC	MA0535.1-Mad	8	0.203058	26.6006	0.993481	7	GCGCCTAAC	CGGCGGCGGCGCCTG	0
G[GC]G[CG]CTAAC	MA0173.1-CG11617	-4	0.209542	27.4499	0.993481	5	GCGCCTAAC	TTAACAT	0
G[GC]G[CG]CTAAC	MA0193.1-Lag1	-1	0.215146	28.1842	0.993481	7	GCGCCTAAC	CTACCAA	0
G[GC]G[CG]CTAAC	MA0190.1-Gsc	-1	0.216544	28.3672	0.993481	6	GCGCCTAAC	GGATTA	0
G[GC]G[CG]CTAAC	MA0458.1-slp1	0	0.218589	28.6352	0.993481	9	GCGCCTAAC	AATGTAAACAA	0
G[GC]G[CG]CTAAC	MA0126.1-ovo	-3	0.221578	29.0267	0.993481	6	GCGCCTAAC	AGTAACAGT	0
G[GC]G[CG]CTAAC	MA0239.1-prd	-3	0.221578	29.0267	0.993481	6	GCGCCTAAC	AGTAACAGT	0
G[GC]G[CG]CTAAC	MA0256.1-zen	-4	0.226618	29.6869	0.993481	5	GCGCCTAAC	CTAATGA	0
G[GC]G[CG]CTAAC	MA0241.1-ro	-4	0.232803	30.4972	0.993481	5	GCGCCTAAC	CTAATTA	0
G[GC]G[CG]CTAAC	MA0221.1-eve	-4	0.239034	31.3135	0.993481	5	GCGCCTAAC	CTAATGA	0
G[GC]G[CG]CTAAC	MA0530.1-CNC::maf-S	6	0.250795	32.8541	0.993481	9	GCGCCTAAC	GATGACTCGGCAAAT	0
G[GC]G[CG]CTAAC	MA0198.1-OdsH	-4	0.25154	32.9517	0.993481	5	GCGCCTAAC	CTAATTA	0
G[GC]G[CG]CTAAC	MA0209.1-ap	-4	0.25154	32.9517	0.993481	5	GCGCCTAAC	CTAATTA	0
G[GC]G[CG]CTAAC	MA0228.1-ind	-4	0.25795	33.7914	0.993481	5	GCGCCTAAC	CTAATTA	0
G[GC]G[CG]CTAAC	MA0249.1-twi	0	0.259328	33.9719	0.993481	9	GCGCCTAAC	TCGCATATGTTG	0
G[GC]G[CG]CTAAC	MA0531.1-CTCF	0	0.2624	34.3744	0.993481	9	GCGCCTAAC	GGCGCCATCTAGCGG	0
G[GC]G[CG]CTAAC	MA0184.1-CG9876	-4	0.264425	34.6396	0.993481	5	GCGCCTAAC	CTAATTA	0
G[GC]G[CG]CTAAC	MA0456.1-opa	7	0.271245	35.5331	0.993481	5	GCGCCTAAC	CAGCGGGGGGTC	0
G[GC]G[CG]CTAAC	MA0168.1-B-H1	-1	0.277602	36.3659	0.993481	7	GCGCCTAAC	CAATTAA	0
G[GC]G[CG]CTAAC	MA0205.1-Trl	-1	0.280659	36.7663	0.993481	8	GCGCCTAAC	TTGCTCTCTC	0
G[GC]G[CG]CTAAC	MA0202.1-Rx	-4	0.284342	37.2489	0.993481	5	GCGCCTAAC	CTAATTA	0
G[GC]G[CG]CTAAC	MA0200.1-Pph13	-4	0.305113	39.9698	0.993481	5	GCGCCTAAC	CTAATTA	0
G[GC]G[CG]CTAAC	MA0533.1-SU(HW)	-2	0.309668	40.5665	0.993481	7	GCGCCTAAC	GCCCAAAAGTATGCAACAAAT	0
G[GC]G[CG]CTAAC	MA0229.1-inv	-3	0.324324	42.4864	0.993481	6	GCGCCTAAC	TCTAATTA	0
G[GC]G[CG]CTAAC	MA0177.1-CG18599	-4	0.326798	42.8106	0.993481	5	GCGCCTAAC	CTAATTA	0
G[GC]G[CG]CTAAC	MA0447.1-gt	1	0.337826	44.2552	0.993481	9	GCGCCTAAC	ATTACGTAAT	0
G[GC]G[CG]CTAAC	MA0218.1-ct	-2	0.343216	44.9613	0.993481	6	GCGCCTAAC	GTTCAA	0
G[GC]G[CG]CTAAC	MA0244.1-slbo	0	0.36084	47.27	0.993481	8	GCGCCTAAC	ATTGCAAA	0
G[GC]G[CG]CTAAC	MA0189.1-E5	-4	0.364602	47.7628	0.993481	5	GCGCCTAAC	TTAATTA	0
G[GC]G[CG]CTAAC	MA0238.1-pb	-4	0.364602	47.7628	0.993481	5	GCGCCTAAC	TTAATTA	0
G[GC]G[CG]CTAAC	MA0255.1-z	1	0.372057	48.7395	0.993481	9	GCGCCTAAC	AATCACTCAA	0
G[GC]G[CG]CTAAC	MA0457.1-PHDP	-4	0.380403	49.8328	0.993481	5	GCGCCTAAC	TTAATTA	0

Continued (cluster 2A)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
AAAAG[CG]GCC	MA0237.2-pan	3	0.0001135	0.014869	0.0296945	9	AAAAGCGCC	ATCAAAGGAGCCGA	0
AAAAG[CG]GCC	MA0213.1-brk	-3	0.0060324	0.79025	0.540275	6	AAAAGCGCC	GGCGCCAG	0
AAAAG[CG]GCC	MA0022.1-dl_1	3	0.0061954	0.811601	0.540275	9	AAAAGCGCC	CGGAAAAACCCC	0
AAAAG[CG]GCC	MA0452.2-KR	2	0.0132816	1.7399	0.868674	9	AAAAGCGCC	GAAAAAGGGTTAAA	0
AAAAG[CG]GCC	MA0023.1-dl_2	1	0.0346632	4.54088	0.998535	9	AAAAGCGCC	GGAAAACCCC	0
AAAAG[CG]GCC	MA0242.1-run::Bgb	-1	0.0409867	5.36926	0.998535	8	AAAAGCGCC	TAACCGCAA	0
AAAAG[CG]GCC	MA0205.1-Trl	0	0.0460037	6.02648	0.998535	9	AAAAGCGCC	GAGAGAGCAA	0
AAAAG[CG]GCC	MA0531.1-CTCF	-3	0.0517157	6.77476	0.998535	6	AAAAGCGCC	GGCGCCATCTAGCGG	0
AAAAG[CG]GCC	MA0443.1-btd	-1	0.0855442	11.2063	0.998535	8	AAAAGCGCC	AGGGGGCGGA	0
AAAAG[CG]GCC	MA0449.1-h	1	0.100032	13.1041	0.998535	9	AAAAGCGCC	GGCACGTGCC	0
AAAAG[CG]GCC	MA0451.1-kni	2	0.113349	14.8487	0.998535	9	AAAAGCGCC	AAACTAGAGCAC	0
AAAAG[CG]GCC	MA0249.1-twi	2	0.131928	17.2825	0.998535	9	AAAAGCGCC	CAACATATGCGA	0
AAAAG[CG]GCC	MA0085.1-Su(H)	7	0.15004	19.6552	0.998535	9	AAAAGCGCC	CTGTGGGAAACGAGAT	0
AAAAG[CG]GCC	MA0459.1-tll	0	0.200486	26.2637	0.998535	9	AAAAGCGCC	AAAAGTCAAA	0
AAAAG[CG]GCC	MA0445.1-D	3	0.243868	31.9467	0.998535	8	AAAAGCGCC	AGAACAATGGA	0
AAAAG[CG]GCC	MA0535.1-Mad	-1	0.261806	34.2966	0.998535	8	AAAAGCGCC	CAGGCGCCGCCGCCG	0
AAAAG[CG]GCC	MA0450.1-hkb	0	0.2741	35.9071	0.998535	9	AAAAGCGCC	TCACGCCCC	0
AAAAG[CG]GCC	MA0530.1-CNC::maf-S	-2	0.285112	37.3496	0.998535	7	AAAAGCGCC	ATTTGCCGAGTCATC	0
AAAAG[CG]GCC	MA0011.1-br_Z2	-1	0.295976	38.7729	0.998535	8	AAAAGCGCC	AAATAGTA	0
AAAAG[CG]GCC	MA0248.1-tup	0	0.377791	49.4906	0.998535	7	AAAAGCGCC	TTAATTG	0
GCCG[GT][AC]GCC	MA0535.1-Mad	4	6.98374e-06	0.000915	0.0017867	9	GCCGGAGCC	CGGCGGCGGCGCCTG	0
GCCG[GT][AC]GCC	MA0213.1-brk	-1	5.31134e-05	0.006958	0.00452947	8	GCCGGAGCC	CTGGCGCC	0
GCCG[GT][AC]GCC	MA0531.1-CTCF	6	0.0030745	0.402755	0.158153	9	GCCGGAGCC	CCGCTAGATGGCGCC	0
GCCG[GT][AC]GCC	MA0449.1-h	1	0.0037091	0.485887	0.158153	9	GCCGGAGCC	GGCACGTGCC	0
GCCG[GT][AC]GCC	MA0454.1-odd	1	0.0082111	1.07566	0.262589	9	GCCGGAGCC	AACAGTAGCAG	0
GCCG[GT][AC]GCC	MA0443.1-btd	0	0.0119667	1.56763	0.340169	9	GCCGGAGCC	TCCGCCCCCT	0
GCCG[GT][AC]GCC	MA0016.1-usp	-1	0.014428	1.89007	0.369122	8	GCCGGAGCC	CCGTGACCCC	0
GCCG[GT][AC]GCC	MA0237.2-pan	-1	0.0204846	2.68349	0.476431	8	GCCGGAGCC	TCGGCTCCTTTGAT	0
GCCG[GT][AC]GCC	MA0530.1-CNC::maf-S	3	0.0235158	3.08057	0.501353	9	GCCGGAGCC	GATGACTCGGCAAAT	0
GCCG[GT][AC]GCC	MA0026.1-Eip74EF	-1	0.0372262	4.87664	0.732605	7	GCCGGAGCC	CCGGAAG	0
GCCG[GT][AC]GCC	MA0205.1-Trl	0	0.0655794	8.5909	0.953459	9	GCCGGAGCC	GAGAGAGCAA	0
GCCG[GT][AC]GCC	MA0085.1-Su(H)	2	0.0727569	9.53116	0.953459	9	GCCGGAGCC	ATCTCGGTTCCCACAA	0
GCCG[GT][AC]GCC	MA0185.1-Deaf1	0	0.0865469	11.3376	0.976479	6	GCCGGAGCC	CACGAA	0
GCCG[GT][AC]GCC	MA0456.1-opa	2	0.0975199	12.7751	0.976479	9	GCCGGAGCC	CAGCGGGGGGTC	0
GCCG[GT][AC]GCC	MA0126.1-ovo	0	0.11014	14.4284	0.976479	9	GCCGGAGCC	ACTGTTACT	0
GCCG[GT][AC]GCC	MA0239.1-prd	0	0.11014	14.4284	0.976479	9	GCCGGAGCC	ACTGTTACT	0
GCCG[GT][AC]GCC	MA0450.1-hkb	0	0.11014	14.4284	0.976479	9	GCCGGAGCC	TCACGCCCC	0
GCCG[GT][AC]GCC	MA0533.1-SU(HW)	1	0.115446	15.1235	0.976479	9	GCCGGAGCC	ATTTGTTGCATACTTTTGGGC	0
GCCG[GT][AC]GCC	MA0023.1-dl_2	1	0.131462	17.2215	0.976479	9	GCCGGAGCC	GGAAAACCCC	0
GCCG[GT][AC]GCC	MA0227.1-hth	1	0.13802	18.0806	0.976479	5	GCCGGAGCC	TGACAG	0
GCCG[GT][AC]GCC	MA0086.1-sna	-1	0.144197	18.8898	0.976479	6	GCCGGAGCC	CAGGTG	0
GCCG[GT][AC]GCC	MA0536.1-pnr	2	0.210327	27.5528	0.976479	9	GCCGGAGCC	TATCGATAGTT	0
GCCG[GT][AC]GCC	MA0451.1-kni	2	0.213635	27.9862	0.976479	9	GCCGGAGCC	GTGCTCTAGTTT	0
GCCG[GT][AC]GCC	MA0242.1-run::Bgb	2	0.215271	28.2005	0.976479	7	GCCGGAGCC	TAACCGCAA	0
GCCG[GT][AC]GCC	MA0534.1-EcR::usp	3	0.222825	29.1901	0.976479	9	GCCGGAGCC	GAGTTCATTGACCTT	0
GCCG[GT][AC]GCC	MA0204.1-Six4	-3	0.272878	35.747	0.976479	6	GCCGGAGCC	GTATCA	0
GCCG[GT][AC]GCC	MA0216.2-CAD	1	0.277257	36.3207	0.976479	9	GCCGGAGCC	GGCCATAAAAA	0
GCCG[GT][AC]GCC	MA0246.1-so	-3	0.283194	37.0984	0.976479	6	GCCGGAGCC	GTATCA	0
GCCG[GT][AC]GCC	MA0447.1-gt	2	0.287836	37.7065	0.976479	8	GCCGGAGCC	ATTACGTAAT	0
GCCG[GT][AC]GCC	MA0010.1-br_Z1	0	0.324607	42.5235	0.976479	9	GCCGGAGCC	GTAATAAACAAATC	0
GCCG[GT][AC]GCC	MA0207.1-achi	-1	0.342626	44.884	0.976479	6	GCCGGAGCC	CTGTCA	0
GCCG[GT][AC]GCC	MA0458.1-slp1	0	0.347723	45.5517	0.976479	9	GCCGGAGCC	AATGTAAACAA	0
GCCG[GT][AC]GCC	MA0247.2-tin	0	0.378741	49.6151	0.976479	9	GCCGGAGCC	CCACTTGAAA	0
GCCG[GT][AC]GCC	MA0252.1-vis	1	0.379427	49.705	0.976479	5	GCCGGAGCC	TGACAG	0

Continued (cluster 2A)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
ACCTCA[ACGT]A[CT]	MA0242.1-run::Bgb	2	0.0164964	2.16103	1	7	ACCTCAAAC	TAACCGCAA	0
ACCTCA[ACGT]A[CT]	MA0253.1-vnd	1	0.0181983	2.38398	1	8	ACCTCAAAC	CACTTGAAA	0
ACCTCA[ACGT]A[CT]	MA0086.1-sna	1	0.0317577	4.16026	1	5	ACCTCAAAC	CACCTG	0
ACCTCA[ACGT]A[CT]	MA0247.2-tin	2	0.0473348	6.20086	1	8	ACCTCAAAC	CCACTTGAAA	0
ACCTCA[ACGT]A[CT]	MA0222.1-exd	-1	0.0631337	8.27052	1	8	ACCTCAAAC	TGTCAAAA	0
ACCTCA[ACGT]A[CT]	MA0459.1-tll	2	0.0799437	10.4726	1	8	ACCTCAAAC	AAAAGTCAAA	0
ACCTCA[ACGT]A[CT]	MA0211.1-bap	1	0.127533	16.7068	1	6	ACCTCAAAC	CACTTAA	0
ACCTCA[ACGT]A[CT]	MA0255.1-z	3	0.151246	19.8132	1	7	ACCTCAAAC	AATCACTCAA	0
ACCTCA[ACGT]A[CT]	MA0249.1-twi	-1	0.158106	20.7118	1	8	ACCTCAAAC	CAACATATGCGA	0
ACCTCA[ACGT]A[CT]	MA0227.1-hth	0	0.167226	21.9067	1	6	ACCTCAAAC	CTGTCA	0
ACCTCA[ACGT]A[CT]	MA0012.1-br_Z3	2	0.182237	23.873	1	9	ACCTCAAAC	TAAACTAAAAG	0
ACCTCA[ACGT]A[CT]	MA0085.1-Su(H)	8	0.196381	25.7259	1	8	ACCTCAAAC	CTGTGGGAAACGAGAT	0
ACCTCA[ACGT]A[CT]	MA0016.1-usp	5	0.197783	25.9096	1	5	ACCTCAAAC	CCGTGACCCC	0
ACCTCA[ACGT]A[CT]	MA0447.1-gt	3	0.197783	25.9096	1	7	ACCTCAAAC	ATTACGTAAT	0
ACCTCA[ACGT]A[CT]	MA0244.1-slbo	0	0.210629	27.5924	1	8	ACCTCAAAC	ATTGCAAA	0
ACCTCA[ACGT]A[CT]	MA0243.1-sd	5	0.226021	29.6087	1	7	ACCTCAAAC	GACATTCCTCGA	0
ACCTCA[ACGT]A[CT]	MA0235.1-onecut	0	0.235633	30.8679	1	7	ACCTCAAAC	AAATCAA	0
ACCTCA[ACGT]A[CT]	MA0199.1-Optix	-1	0.253158	33.1637	1	5	ACCTCAAAC	TATCA	0
ACCTCA[ACGT]A[CT]	MA0456.1-opa	1	0.281025	36.8143	1	9	ACCTCAAAC	GACCCCCCGCTG	0
ACCTCA[ACGT]A[CT]	MA0204.1-Six4	-3	0.282767	37.0424	1	6	ACCTCAAAC	TGATAC	0
ACCTCA[ACGT]A[CT]	MA0193.1-Lag1	-1	0.322269	42.2172	1	7	ACCTCAAAC	CTACCAA	0
ACCTCA[ACGT]A[CT]	MA0246.1-so	0	0.332247	43.5244	1	6	ACCTCAAAC	GTATCA	0
ACCTCA[ACGT]A[CT]	MA0530.1-CNC::maf-S	7	0.341316	44.7124	1	8	ACCTCAAAC	ATTTGCCGAGTCATC	0
ACCTCA[ACGT]A[CT]	MA0183.1-CG7056	1	0.35926	47.0631	1	7	ACCTCAAAC	TAATTAAA	0
ACCTCA[ACGT]A[CT]	MA0023.1-dl_2	5	0.378833	49.6271	1	5	ACCTCAAAC	GGAAAACCCC	0
AAAA[AC]TCGA	MA0529.1-BEAF-32	4	0.0071331	0.934441	0.97486	9	AAAAATCGA	TCGAAACTATCGATA	0
AAAA[AC]TCGA	MA0459.1-tll	0	0.0074417	0.97486	0.97486	9	AAAAATCGA	AAAAGTCAAA	0
AAAA[AC]TCGA	MA0536.1-pnr	0	0.0114785	1.50368	1	9	AAAAATCGA	AACTATCGATA	0
AAAA[AC]TCGA	MA0235.1-onecut	-2	0.0249254	3.26522	1	7	AAAAATCGA	AAATCAA	0
AAAA[AC]TCGA	MA0451.1-kni	-1	0.0321699	4.21426	1	8	AAAAATCGA	AAACTAGAGCAC	0
AAAA[AC]TCGA	MA0022.1-dl_1	4	0.0350683	4.59395	1	8	AAAAATCGA	CGGAAAAACCCC	0
AAAA[AC]TCGA	MA0445.1-D	2	0.0536969	7.03429	1	9	AAAAATCGA	AGAACAATGGA	0
AAAA[AC]TCGA	MA0012.1-br_Z3	0	0.0556631	7.29187	1	9	AAAAATCGA	TAAACTAAAAG	0
AAAA[AC]TCGA	MA0011.1-br_Z2	-2	0.0592256	7.75855	1	7	AAAAATCGA	AAATAGTA	0
AAAA[AC]TCGA	MA0023.1-dl_2	2	0.161102	21.1043	1	8	AAAAATCGA	GGAAAACCCC	0
AAAA[AC]TCGA	MA0255.1-z	1	0.177346	23.2323	1	9	AAAAATCGA	AATCACTCAA	0
AAAA[AC]TCGA	MA0237.2-pan	5	0.204073	26.7336	1	9	AAAAATCGA	ATCAAAGGAGCCGA	0
AAAA[AC]TCGA	MA0010.1-br_Z1	2	0.210558	27.5831	1	9	AAAAATCGA	GTAATAAACAAATC	0
AAAA[AC]TCGA	MA0247.2-tin	-1	0.236732	31.0119	1	8	AAAAATCGA	CCACTTGAAA	0
AAAA[AC]TCGA	MA0185.1-Deaf1	-4	0.271412	35.5549	1	5	AAAAATCGA	CACGAA	0
AAAA[AC]TCGA	MA0242.1-run::Bgb	0	0.282608	37.0217	1	9	AAAAATCGA	TAACCGCAA	0
AAAA[AC]TCGA	MA0085.1-Su(H)	6	0.293863	38.496	1	9	AAAAATCGA	CTGTGGGAAACGAGAT	0
AAAA[AC]TCGA	MA0190.1-Gsc	-2	0.347727	45.5523	1	6	AAAAATCGA	TAATCC	0
AAAA[AC]TCGA	MA0532.1-STAT92E	6	0.352769	46.2127	1	9	AAAAATCGA	TTTCCTGGAATTCCG	0
AAAA[AC]TCGA	MA0168.1-B-H1	-1	0.355819	46.6122	1	7	AAAAATCGA	TTAATTG	0
AAAA[AC]TCGA	MA0253.1-vnd	-2	0.358705	46.9904	1	7	AAAAATCGA	CACTTGAAA	0
AAAA[AC]TCGA	MA0049.1-hb	4	0.363694	47.644	1	6	AAAAATCGA	GCATAAAAAA	0
AAAA[AC]TCGA	MA0229.1-inv	-1	0.368688	48.2982	1	8	AAAAATCGA	TAATTAGA	0

Continued (cluster 2A)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
AT[AGT]CCGAAC	MA0185.1-Deaf1	-2	0.0039616	0.518965	0.999952	6	ATTCCGAAC	CACGAA	0
AT[AGT]CCGAAC	MA0244.1-slbo	0	0.0120961	1.58459	0.999952	8	ATTCCGAAC	ATTGCAAA	0
AT[AGT]CCGAAC	MA0218.1-ct	-3	0.0143212	1.87608	0.999952	6	ATTCCGAAC	TTGAAC	0
AT[AGT]CCGAAC	MA0243.1-sd	3	0.0438427	5.74339	0.999952	9	ATTCCGAAC	GACATTCCTCGA	0
AT[AGT]CCGAAC	MA0222.1-exd	-1	0.0751383	9.84312	0.999952	8	ATTCCGAAC	TGTCAAAA	0
AT[AGT]CCGAAC	MA0026.1-Eip74EF	-3	0.118798	15.5625	0.999952	6	ATTCCGAAC	CCGGAAG	0
AT[AGT]CCGAAC	MA0242.1-run::Bgb	0	0.141704	18.5632	0.999952	9	ATTCCGAAC	TAACCGCAA	0
AT[AGT]CCGAAC	MA0447.1-gt	0	0.160816	21.0669	0.999952	9	ATTCCGAAC	ATTACGTAAT	0
AT[AGT]CCGAAC	MA0530.1-CNC::maf-S	4	0.175088	22.9366	0.999952	9	ATTCCGAAC	GATGACTCGGCAAAT	0
AT[AGT]CCGAAC	MA0210.1-ara	-5	0.178438	23.3754	0.999952	4	ATTCCGAAC	TAACA	0
AT[AGT]CCGAAC	MA0217.1-caup	-5	0.186354	24.4124	0.999952	4	ATTCCGAAC	TAACA	0
AT[AGT]CCGAAC	MA0451.1-kni	1	0.203895	26.7103	0.999952	9	ATTCCGAAC	AAACTAGAGCAC	0
AT[AGT]CCGAAC	MA0237.2-pan	7	0.261498	34.2563	0.999952	7	ATTCCGAAC	ATCAAAGGAGCCGA	0
AT[AGT]CCGAAC	MA0233.1-mirr	-5	0.293997	38.5136	0.999952	4	ATTCCGAAC	AAACA	0
AT[AGT]CCGAAC	MA0446.1-fkh	3	0.300432	39.3566	0.999952	8	ATTCCGAAC	TGTTTGCTTAA	0
AT[AGT]CCGAAC	MA0443.1-btd	-2	0.343588	45.01	0.999952	7	ATTCCGAAC	TCCGCCCCCT	0
AT[AGT]CCGAAC	MA0453.1-nub	1	0.34394	45.0561	0.999952	9	ATTCCGAAC	TATGCAAATTAG	0
AT[AGT]CCGAAC	MA0193.1-Lag1	0	0.360521	47.2283	0.999952	7	ATTCCGAAC	CTACCAA	0
AT[AGT]CCGAAC	MA0536.1-pnr	2	0.361812	47.3974	0.999952	9	ATTCCGAAC	AACTATCGATA	0
AT[AGT]CCGAAC	MA0254.1-vvl	1	0.363745	47.6506	0.999952	5	ATTCCGAAC	TATGCA	0
AT[AGT]CCGAAC	MA0013.1-br_Z4	-1	0.381464	49.9718	0.999952	8	ATTCCGAAC	TAGTAAACAAA	0
AT[AGT]CCGAAC	MA0458.1-slp1	0	0.381464	49.9718	0.999952	9	ATTCCGAAC	AATGTAAACAA	0
GCGA[TA]TCC	MA0023.1-dl_2	0	0.0076794	1.00601	0.686708	8	GCGAATCC	GGAAAACCCC	0
GCGA[TA]TCC	MA0532.1-STAT92E	1	0.0120572	1.57949	0.686708	8	GCGAATCC	CGGAATTCCAGGAAA	0
GCGA[TA]TCC	MA0201.1-Ptx1	-1	0.0122226	1.60116	0.686708	7	GCGAATCC	TTAATCC	0
GCGA[TA]TCC	MA0212.1-bcd	-2	0.0158159	2.07188	0.686708	6	GCGAATCC	TAATCC	0
GCGA[TA]TCC	MA0234.1-oc	-2	0.0158159	2.07188	0.686708	6	GCGAATCC	TAATCC	0
GCGA[TA]TCC	MA0190.1-Gsc	-2	0.0209572	2.7454	0.723404	6	GCGAATCC	TAATCC	0
GCGA[TA]TCC	MA0085.1-Su(H)	3	0.0222148	2.91013	0.723404	8	GCGAATCC	ATCTCGGTTCCCACAA	0
GCGA[TA]TCC	MA0243.1-sd	0	0.0397497	5.20721	0.994212	8	GCGAATCC	GACATTCCTCGA	0
GCGA[TA]TCC	MA0237.2-pan	0	0.0445202	5.83214	0.994212	8	GCGAATCC	TCGGCTCCTTTGAT	0
GCGA[TA]TCC	MA0016.1-usp	2	0.0546478	7.15887	0.994212	8	GCGAATCC	CCGTGACCCC	0
GCGA[TA]TCC	MA0453.1-nub	3	0.0562966	7.37486	0.994212	8	GCGAATCC	TATGCAAATTAG	0
GCGA[TA]TCC	MA0242.1-run::Bgb	2	0.1043	13.6633	0.994212	7	GCGAATCC	TTGCGGTTA	0
GCGA[TA]TCC	MA0022.1-dl_1	3	0.104799	13.7287	0.994212	8	GCGAATCC	GGGGTTTTTCCC	0
GCGA[TA]TCC	MA0255.1-z	4	0.128975	16.8957	0.994212	6	GCGAATCC	TTGAGTGATT	0
GCGA[TA]TCC	MA0126.1-ovo	0	0.129206	16.9259	0.994212	8	GCGAATCC	ACTGTTACT	0
GCGA[TA]TCC	MA0239.1-prd	0	0.129206	16.9259	0.994212	8	GCGAATCC	ACTGTTACT	0
GCGA[TA]TCC	MA0187.1-Dll	0	0.136605	17.8952	0.994212	7	GCGAATCC	GTAATTA	0
GCGA[TA]TCC	MA0185.1-Deaf1	1	0.181037	23.7158	0.994212	5	GCGAATCC	CACGAA	0
GCGA[TA]TCC	MA0188.1-Dr	0	0.184261	24.1382	0.994212	7	GCGAATCC	CCAATTA	0
GCGA[TA]TCC	MA0536.1-pnr	2	0.200003	26.2004	0.994212	8	GCGAATCC	TATCGATAGTT	0
GCGA[TA]TCC	MA0531.1-CTCF	1	0.201659	26.4173	0.994212	8	GCGAATCC	GGCGCCATCTAGCGG	0
GCGA[TA]TCC	MA0168.1-B-H1	-1	0.205733	26.951	0.994212	7	GCGAATCC	CAATTAA	0
GCGA[TA]TCC	MA0214.1-bsh	-1	0.213512	27.9701	0.994212	7	GCGAATCC	CAATTAA	0
GCGA[TA]TCC	MA0529.1-BEAF-32	0	0.25234	33.0565	0.994212	8	GCGAATCC	TCGAAACTATCGATA	0
GCGA[TA]TCC	MA0011.1-br_Z2	1	0.259178	33.9523	0.994212	7	GCGAATCC	TACTATTT	0
GCGA[TA]TCC	MA0535.1-Mad	3	0.273166	35.7848	0.994212	8	GCGAATCC	CAGGCGCCGCCGCCG	0
GCGA[TA]TCC	MA0246.1-so	-2	0.273947	35.887	0.994212	6	GCGAATCC	GTATCA	0
GCGA[TA]TCC	MA0249.1-twi	8	0.285132	37.3523	0.994212	4	GCGAATCC	CAACATATGCGA	0
GCGA[TA]TCC	MA0192.1-Hmx	-1	0.304794	39.9281	0.994212	7	GCGAATCC	CAATTAA	0
GCGA[TA]TCC	MA0026.1-Eip74EF	-3	0.339646	44.4936	0.994212	5	GCGAATCC	CTTCCGG	0
GCGA[TA]TCC	MA0169.1-B-H2	-1	0.339646	44.4936	0.994212	7	GCGAATCC	CAATTAA	0
GCGA[TA]TCC	MA0171.1-CG11085	-1	0.339646	44.4936	0.994212	7	GCGAATCC	CAATTAA	0
GCGA[TA]TCC	MA0213.1-brk	0	0.343898	45.0506	0.994212	8	GCGAATCC	CTGGCGCC	0
GCGA[TA]TCC	MA0530.1-CNC::maf-S	9	0.346151	45.3458	0.994212	6	GCGAATCC	GATGACTCGGCAAAT	0
GCGA[TA]TCC	MA0452.2-KR	0	0.346308	45.3664	0.994212	8	GCGAATCC	TTTAACCCTTTTTC	0
GCGA[TA]TCC	MA0248.1-tup	-1	0.363963	47.6792	0.994212	7	GCGAATCC	CAATTAA	0
GCGA[TA]TCC	MA0250.1-unc-4	-1	0.363963	47.6792	0.994212	7	GCGAATCC	CAATTAA	0
GCGA[TA]TCC	MA0449.1-h	1	0.3763	49.2953	0.994212	8	GCGAATCC	GGCACGTGCC	0

65

Continued (cluster 2A)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
CGT[GT][ATC]GGCA	MA0222.1-exd	-1	0.0144578	1.89398	0.999922	8	CGTGTGGCA	TTTTGACA	0
CGT[GT][ATC]GGCA	MA0244.1-slbo	-2	0.0323612	4.23931	0.999922	7	CGTGTGGCA	TTTGCAAT	0
CGT[GT][ATC]GGCA	MA0185.1-Deaf1	-2	0.0348064	4.55963	0.999922	6	CGTGTGGCA	TTCGGG	0
CGT[GT][ATC]GGCA	MA0449.1-h	2	0.0438825	5.74861	0.999922	8	CGTGTGGCA	GGCACGTGCC	0
CGT[GT][ATC]GGCA	MA0533.1-SU(HW)	1	0.0611068	8.00499	0.999922	9	CGTGTGGCA	ATTTGTTGCATACTTTTGGGC	0
CGT[GT][ATC]GGCA	MA0530.1-CNC::maf-S	3	0.0642095	8.41144	0.999922	9	CGTGTGGCA	GATGACTCGGCAAAT	0
CGT[GT][ATC]GGCA	MA0536.1-pnr	3	0.0718351	9.4104	0.999922	8	CGTGTGGCA	TATCGATAGTT	0
CGT[GT][ATC]GGCA	MA0085.1-Su(H)	4	0.0878588	11.5095	0.999922	9	CGTGTGGCA	ATCTCGGTTCCCACAA	0
CGT[GT][ATC]GGCA	MA0207.1-achi	-4	0.102523	13.4305	0.999922	5	CGTGTGGCA	TGACAG	0
CGT[GT][ATC]GGCA	MA0252.1-vis	-4	0.102523	13.4305	0.999922	5	CGTGTGGCA	TGACAG	0
CGT[GT][ATC]GGCA	MA0026.1-Eip74EF	0	0.107543	14.0881	0.999922	7	CGTGTGGCA	CTTCCGG	0
CGT[GT][ATC]GGCA	MA0243.1-sd	-1	0.108013	14.1497	0.999922	8	CGTGTGGCA	CCGAGGAATGTC	0
CGT[GT][ATC]GGCA	MA0205.1-Trl	1	0.113838	14.9128	0.999922	9	CGTGTGGCA	GAGAGAGCAA	0
CGT[GT][ATC]GGCA	MA0227.1-hth	-4	0.116501	15.2617	0.999922	5	CGTGTGGCA	TGACAG	0
CGT[GT][ATC]GGCA	MA0016.1-usp	1	0.12211	15.9964	0.999922	9	CGTGTGGCA	CCGTGACCCC	0
CGT[GT][ATC]GGCA	MA0255.1-z	1	0.139908	18.3279	0.999922	9	CGTGTGGCA	TTGAGTGATT	0
CGT[GT][ATC]GGCA	MA0447.1-gt	-1	0.139908	18.3279	0.999922	8	CGTGTGGCA	ATTACGTAAT	0
CGT[GT][ATC]GGCA	MA0446.1-fkh	-2	0.151698	19.8724	0.999922	7	CGTGTGGCA	TTAAGCAAACA	0
CGT[GT][ATC]GGCA	MA0218.1-ct	-1	0.162499	21.2874	0.999922	6	CGTGTGGCA	GTTCAA	0
CGT[GT][ATC]GGCA	MA0254.1-vvl	-3	0.183544	24.0442	0.999922	6	CGTGTGGCA	TATGCA	0
CGT[GT][ATC]GGCA	MA0247.2-tin	1	0.205572	26.9299	0.999922	9	CGTGTGGCA	CCACTTGAAA	0
CGT[GT][ATC]GGCA	MA0213.1-brk	-3	0.209513	27.4462	0.999922	6	CGTGTGGCA	CTGGCGCC	0
CGT[GT][ATC]GGCA	MA0204.1-Six4	-3	0.213565	27.9771	0.999922	6	CGTGTGGCA	GTATCA	0
CGT[GT][ATC]GGCA	MA0249.1-twi	3	0.215579	28.2408	0.999922	9	CGTGTGGCA	CAACATATGCGA	0
CGT[GT][ATC]GGCA	MA0049.1-hb	2	0.246848	32.3371	0.999922	8	CGTGTGGCA	TTTTTTATGC	0
CGT[GT][ATC]GGCA	MA0451.1-kni	2	0.250391	32.8012	0.999922	9	CGTGTGGCA	AAACTAGAGCAC	0
CGT[GT][ATC]GGCA	MA0450.1-hkb	4	0.253797	33.2474	0.999922	5	CGTGTGGCA	GGGGCGTGA	0
CGT[GT][ATC]GGCA	MA0256.1-zen	1	0.258375	33.8471	0.999922	6	CGTGTGGCA	TCATTAG	0
CGT[GT][ATC]GGCA	MA0210.1-ara	0	0.274399	35.9463	0.999922	5	CGTGTGGCA	TGTTA	0
CGT[GT][ATC]GGCA	MA0237.2-pan	6	0.276499	36.2213	0.999922	8	CGTGTGGCA	TCGGCTCCTTTGAT	0
CGT[GT][ATC]GGCA	MA0456.1-opa	3	0.288985	37.8571	0.999922	9	CGTGTGGCA	CAGCGGGGGGTC	0
CGT[GT][ATC]GGCA	MA0126.1-ovo	2	0.296682	38.8654	0.999922	7	CGTGTGGCA	ACTGTTACT	0
CGT[GT][ATC]GGCA	MA0239.1-prd	2	0.296682	38.8654	0.999922	7	CGTGTGGCA	ACTGTTACT	0
CGT[GT][ATC]GGCA	MA0233.1-mirr	0	0.319974	41.9166	0.999922	5	CGTGTGGCA	TGTTT	0
CGT[GT][ATC]GGCA	MA0454.1-odd	1	0.322787	42.2851	0.999922	9	CGTGTGGCA	AACAGTAGCAG	0
CGT[GT][ATC]GGCA	MA0535.1-Mad	2	0.325679	42.6639	0.999922	9	CGTGTGGCA	CGGCGGCGGCGCCTG	0
CGT[GT][ATC]GGCA	MA0217.1-caup	0	0.33269	43.5824	0.999922	5	CGTGTGGCA	TGTTA	0
CGT[GT][ATC]GGCA	MA0242.1-run::Bgb	-1	0.345598	45.2734	0.999922	8	CGTGTGGCA	TTGCGGTTA	0
CGT[GT][ATC]GGCA	MA0221.1-eve	1	0.355819	46.6123	0.999922	6	CGTGTGGCA	TCATTAG	0
CGT[GT][ATC]GGCA	MA0531.1-CTCF	8	0.372141	48.7504	0.999922	7	CGTGTGGCA	GGCGCCATCTAGCGG	0
CGT[GT][ATC]GGCA	MA0452.2-KR	6	0.373404	48.916	0.999922	8	CGTGTGGCA	TTTAACCCTTTTTC	0
CGT[GT][ATC]GGCA	MA0193.1-Lag1	-3	0.376825	49.3641	0.999922	6	CGTGTGGCA	TTGGTAG	0
CGT[GT][ATC]GGCA	MA0453.1-nub	1	0.378455	49.5776	0.999922	9	CGTGTGGCA	CTAATTTGCATA	0

Continued (cluster 2A)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
CG[CT][GT]TGGCA	MA0449.1-h	2	0.0037866	0.496044	0.492065	8	CGCGTGGCA	GGCACGTGCC	0
CG[CT][GT]TGGCA	MA0222.1-exd	-1	0.0096239	1.26072	0.582495	8	CGCGTGGCA	TTTTGACA	0
CG[CT][GT]TGGCA	MA0207.1-achi	-4	0.0115765	1.51652	0.582495	5	CGCGTGGCA	TGACAG	0
CG[CT][GT]TGGCA	MA0252.1-vis	-4	0.0115765	1.51652	0.582495	5	CGCGTGGCA	TGACAG	0
CG[CT][GT]TGGCA	MA0227.1-hth	-4	0.0134474	1.76161	0.582495	5	CGCGTGGCA	TGACAG	0
CG[CT][GT]TGGCA	MA0533.1-SU(HW)	1	0.0212388	2.78228	0.621333	9	CGCGTGGCA	ATTTGTTGCATACTTTTGGGC	0
CG[CT][GT]TGGCA	MA0085.1-Su(H)	-1	0.0215161	2.81861	0.621333	8	CGCGTGGCA	CTGTGGGAAACGAGAT	0
CG[CT][GT]TGGCA	MA0247.2-tin	1	0.0280795	3.67842	0.684321	9	CGCGTGGCA	CCACTTGAAA	0
CG[CT][GT]TGGCA	MA0204.1-Six4	-3	0.0289633	3.7942	0.684321	6	CGCGTGGCA	GTATCA	0
CG[CT][GT]TGGCA	MA0244.1-slbo	-3	0.0425544	5.57462	0.831409	6	CGCGTGGCA	TTTGCAAT	0
CG[CT][GT]TGGCA	MA0530.1-CNC::maf-S	3	0.0511836	6.70506	0.831409	9	CGCGTGGCA	GATGACTCGGCAAAT	0
CG[CT][GT]TGGCA	MA0249.1-twi	3	0.057439	7.52451	0.878136	9	CGCGTGGCA	CAACATATGCGA	0
CG[CT][GT]TGGCA	MA0193.1-Lag1	-3	0.0618081	8.09686	0.892435	6	CGCGTGGCA	TTGGTAG	0
CG[CT][GT]TGGCA	MA0253.1-vnd	0	0.0715759	9.37645	0.911286	9	CGCGTGGCA	CACTTGAAA	0
CG[CT][GT]TGGCA	MA0213.1-brk	-3	0.0791416	10.3676	0.911286	6	CGCGTGGCA	CTGGCGCC	0
CG[CT][GT]TGGCA	MA0243.1-sd	-1	0.0961811	12.5997	0.991978	8	CGCGTGGCA	CCGAGGAATGTC	0
CG[CT][GT]TGGCA	MA0217.1-caup	-4	0.117699	15.4186	0.991978	5	CGCGTGGCA	TAACA	0
CG[CT][GT]TGGCA	MA0016.1-usp	-1	0.124746	16.3417	0.991978	8	CGCGTGGCA	CCGTGACCCC	0
CG[CT][GT]TGGCA	MA0454.1-odd	1	0.127007	16.6379	0.991978	9	CGCGTGGCA	AACAGTAGCAG	0
CG[CT][GT]TGGCA	MA0173.1-CG11617	-3	0.133086	17.4342	0.991978	6	CGCGTGGCA	TTAACAT	0
CG[CT][GT]TGGCA	MA0255.1-z	1	0.133928	17.5446	0.991978	9	CGCGTGGCA	TTGAGTGATT	0
CG[CT][GT]TGGCA	MA0531.1-CTCF	5	0.147485	19.3205	0.991978	9	CGCGTGGCA	GGCGCCATCTAGCGG	0
CG[CT][GT]TGGCA	MA0536.1-pnr	3	0.162409	21.2756	0.991978	8	CGCGTGGCA	TATCGATAGTT	0
CG[CT][GT]TGGCA	MA0199.1-Optix	-4	0.164212	21.5117	0.991978	5	CGCGTGGCA	TGATA	0
CG[CT][GT]TGGCA	MA0246.1-so	-3	0.178121	23.3339	0.991978	6	CGCGTGGCA	GTATCA	0
CG[CT][GT]TGGCA	MA0010.1-br_Z1	0	0.179063	23.4573	0.991978	9	CGCGTGGCA	GATTTGTCTATTAC	0
CG[CT][GT]TGGCA	MA0453.1-nub	1	0.182025	23.8453	0.991978	9	CGCGTGGCA	CTAATTTGCATA	0
CG[CT][GT]TGGCA	MA0237.2-pan	4	0.191262	25.0553	0.991978	9	CGCGTGGCA	TCGGCTCCTTTGAT	0
CG[CT][GT]TGGCA	MA0456.1-opa	1	0.207119	27.1326	0.991978	9	CGCGTGGCA	CAGCGGGGGGTC	0
CG[CT][GT]TGGCA	MA0450.1-hkb	2	0.220292	28.8583	0.991978	7	CGCGTGGCA	GGGGCGTGA	0
CG[CT][GT]TGGCA	MA0452.2-KR	6	0.239367	31.3571	0.991978	8	CGCGTGGCA	TTTAACCCTTTTC	0
CG[CT][GT]TGGCA	MA0012.1-br_Z3	0	0.261266	34.2258	0.991978	9	CGCGTGGCA	CTTTTAGTTTA	0
CG[CT][GT]TGGCA	MA0233.1-mirr	-4	0.270052	35.3768	0.991978	5	CGCGTGGCA	AAACA	0
CG[CT][GT]TGGCA	MA0185.1-Deaf1	-2	0.284052	37.2109	0.991978	6	CGCGTGGCA	TTCGGG	0
CG[CT][GT]TGGCA	MA0210.1-ara	-4	0.297106	38.9209	0.991978	5	CGCGTGGCA	TAACA	0
CG[CT][GT]TGGCA	MA0451.1-kni	2	0.297441	38.9647	0.991978	9	CGCGTGGCA	AAACTAGAGCAC	0
CG[CT][GT]TGGCA	MA0446.1-fkh	0	0.311377	40.7904	0.991978	9	CGCGTGGCA	TGTTTGCTTAA	0
CG[CT][GT]TGGCA	MA0529.1-BEAF-32	7	0.317259	41.5609	0.991978	8	CGCGTGGCA	TATCGATAGTTTCGA	0
CG[CT][GT]TGGCA	MA0086.1-sna	0	0.332521	43.5602	0.991978	6	CGCGTGGCA	CACCTG	0
CG[CT][GT]TGGCA	MA0254.1-vvl	-5	0.342587	44.8789	0.991978	4	CGCGTGGCA	TGCATA	0
CG[CT][GT]TGGCA	MA0535.1-Mad	2	0.373614	48.9434	0.991978	9	CGCGTGGCA	CGGCGGCGGCGCCTG	0
CG[CT][GT]TGGCA	MA0013.1-br_Z4	2	0.378259	49.5519	0.991978	9	CGCGTGGCA	TTTGTTTACTA	0
CG[CT][GT]TGGCA	MA0216.2-CAD	2	0.378259	49.5519	0.991978	9	CGCGTGGCA	TTTTTATGGCC	0

Continued (cluster 2A)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
CGC[CG]GGC[CT]G	MA0535.1-Mad	4	0.0004326	0.056676	0.0940447	9	CGCCGGCCG	CAGGCGCCGCCGCCG	0
CGC[CG]GGC[CT]G	MA0456.1-opa	3	0.0007397	0.096903	0.0940447	9	CGCCGGCCG	GACCCCCCGCTG	0
CGC[CG]GGC[CT]G	MA0213.1-brk	2	0.0017462	0.228747	0.101104	6	CGCCGGCCG	GGCGCCAG	0
CGC[CG]GGC[CT]G	MA0531.1-CTCF	2	0.0034468	0.45153	0.121755	9	CGCCGGCCG	GGCGCCATCTAGCGG	0
CGC[CG]GGC[CT]G	MA0449.1-h	2	0.0038307	0.501823	0.121755	8	CGCCGGCCG	GGCACGTGCC	0
CGC[CG]GGC[CT]G	MA0443.1-btd	0	0.0049013	0.642075	0.138474	9	CGCCGGCCG	AGGGGGCGGA	0
CGC[CG]GGC[CT]G	MA0086.1-sna	0	0.0266641	3.49299	0.564993	6	CGCCGGCCG	CACCTG	0
CGC[CG]GGC[CT]G	MA0026.1-Eip74EF	-2	0.0310639	4.06937	0.60759	7	CGCCGGCCG	CCGGAAG	0
CGC[CG]GGC[CT]G	MA0185.1-Deaf1	0	0.0461604	6.04701	0.771136	6	CGCCGGCCG	TTCGGG	0
CGC[CG]GGC[CT]G	MA0249.1-twi	3	0.0485237	6.3566	0.771136	9	CGCCGGCCG	CAACATATGCGA	0
CGC[CG]GGC[CT]G	MA0243.1-sd	1	0.0590545	7.73614	0.883287	9	CGCCGGCCG	CCGAGGAATGTC	0
CGC[CG]GGC[CT]G	MA0237.2-pan	1	0.068735	9.00429	0.970501	9	CGCCGGCCG	TCGGCTCCTTTGAT	0
CGC[CG]GGC[CT]G	MA0450.1-hkb	2	0.082549	10.8139	0.970501	7	CGCCGGCCG	GGGGCGTGA	0
CGC[CG]GGC[CT]G	MA0205.1-Trl	1	0.0854865	11.1987	0.970501	9	CGCCGGCCG	TTGCTCTCTC	0
CGC[CG]GGC[CT]G	MA0536.1-pnr	3	0.124422	16.2993	0.970501	8	CGCCGGCCG	TATCGATAGTT	0
CGC[CG]GGC[CT]G	MA0016.1-usp	1	0.132739	17.3888	0.970501	9	CGCCGGCCG	CCGTGACCCC	0
CGC[CG]GGC[CT]G	MA0454.1-odd	0	0.138329	18.1212	0.970501	9	CGCCGGCCG	CTGCTACTGTT	0
CGC[CG]GGC[CT]G	MA0529.1-BEAF-32	3	0.140077	18.35	0.970501	9	CGCCGGCCG	TATCGATAGTTTCGA	0
CGC[CG]GGC[CT]G	MA0242.1-run::Bgb	1	0.170689	22.3603	0.970501	8	CGCCGGCCG	TTGCGGTTA	0
CGC[CG]GGC[CT]G	MA0530.1-CNC::maf-S	3	0.184816	24.2109	0.970501	9	CGCCGGCCG	ATTTGCCGAGTCATC	0
CGC[CG]GGC[CT]G	MA0247.2-tin	3	0.199426	26.1248	0.970501	7	CGCCGGCCG	TTTCAAGTGG	0
CGC[CG]GGC[CT]G	MA0451.1-kni	3	0.250836	32.8595	0.970501	9	CGCCGGCCG	AAACTAGAGCAC	0
CGC[CG]GGC[CT]G	MA0255.1-z	3	0.27851	36.4848	0.970501	7	CGCCGGCCG	TTGAGTGATT	0
CGC[CG]GGC[CT]G	MA0227.1-hth	0	0.290137	38.0079	0.970501	6	CGCCGGCCG	TGACAG	0
CGC[CG]GGC[CT]G	MA0533.1-SU(HW)	14	0.330231	43.2603	0.970501	7	CGCCGGCCG	ATTTGTTGCATACTTTTGGGC	0
CGC[CG]GGC[CT]G	MA0022.1-dl_1	-2	0.381526	49.9799	0.970501	7	CGCCGGCCG	GGGGTTTTTCCC	0
AAAC[AG]CATG	MA0249.1-twi	0	0.002229	0.292001	0.584002	9	AAACACATG	CAACATATGCGA	0
AAAC[AG]CATG	MA0086.1-sna	-3	0.0125663	1.64618	1	6	AAACACATG	CACCTG	0
AAAC[AG]CATG	MA0233.1-mirr	0	0.0327405	4.289	1	5	AAACACATG	AAACA	0
AAAC[AG]CATG	MA0533.1-SU(HW)	4	0.0369109	4.83532	1	9	AAACACATG	GCCCAAAAGTATGCAACAAAT	0
AAAC[AG]CATG	MA0210.1-ara	0	0.0437389	5.72979	1	5	AAACACATG	TAACA	0
AAAC[AG]CATG	MA0242.1-run::Bgb	1	0.0515811	6.75712	1	8	AAACACATG	TAACCGCAA	0
AAAC[AG]CATG	MA0010.1-br_Z1	5	0.0531394	6.96126	1	9	AAACACATG	GTAATAAACAAATC	0
AAAC[AG]CATG	MA0217.1-caup	0	0.107413	14.0711	1	5	AAACACATG	TAACA	0
AAAC[AG]CATG	MA0012.1-br_Z3	1	0.108434	14.2048	1	9	AAACACATG	TAAACTAAAAG	0
AAAC[AG]CATG	MA0458.1-slp1	5	0.124911	16.3633	1	6	AAACACATG	AATGTAAACAA	0
AAAC[AG]CATG	MA0218.1-ct	2	0.138295	18.1166	1	4	AAACACATG	TTGAAC	0
AAAC[AG]CATG	MA0445.1-D	1	0.148195	19.4136	1	9	AAACACATG	AGAACAATGGA	0
AAAC[AG]CATG	MA0013.1-br_Z4	4	0.163127	21.3696	1	7	AAACACATG	TAGTAAACAAA	0
AAAC[AG]CATG	MA0022.1-dl_1	3	0.221708	29.0438	1	9	AAACACATG	CGGAAAAACCCC	0
AAAC[AG]CATG	MA0449.1-h	-1	0.231973	30.3884	1	8	AAACACATG	GGCACGTGCC	0
AAAC[AG]CATG	MA0085.1-Su(H)	7	0.237461	31.1074	1	9	AAACACATG	CTGTGGGAAACGAGAT	0
AAAC[AG]CATG	MA0253.1-vnd	-3	0.271486	35.5646	1	6	AAACACATG	CACTTGAAA	0
AAAC[AG]CATG	MA0459.1-tll	1	0.305719	40.0492	1	9	AAACACATG	AAAAGTCAAA	0
AAAC[AG]CATG	MA0243.1-sd	1	0.314014	41.1359	1	9	AAACACATG	CCGAGGAATGTC	0
AAAC[AG]CATG	MA0247.2-tin	-2	0.322682	42.2714	1	7	AAACACATG	CCACTTGAAA	0
AAAC[AG]CATG	MA0015.1-Cf2_II	0	0.367045	48.0829	1	9	AAACACATG	GTATATATAC	0
AAAC[AG]CATG	MA0531.1-CTCF	1	0.376209	49.2834	1	9	AAACACATG	CCGCTAGATGGCGCC	0
AAAC[AG]CATG	MA0254.1-vvl	-3	0.378704	49.6102	1	6	AAACACATG	TGCATA	0

Continued (cluster 2A)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
CAGT[GA]TCG	MA0204.1-Six4	-2	0.0040373	0.528885	0.9998	6	CAGTATCG	GTATCA	0
CAGT[GA]TCG	MA0246.1-so	-2	0.0119716	1.56828	0.9998	6	CAGTATCG	GTATCA	0
CAGT[GA]TCG	MA0454.1-odd	2	0.014555	1.9067	0.9998	8	CAGTATCG	AACAGTAGCAG	0
CAGT[GA]TCG	MA0529.1-BEAF-32	4	0.0165753	2.17136	0.9998	8	CAGTATCG	TCGAAACTATCGATA	0
CAGT[GA]TCG	MA0536.1-pnr	0	0.0267317	3.50186	0.9998	8	CAGTATCG	AACTATCGATA	0
CAGT[GA]TCG	MA0199.1-Optix	-3	0.0433742	5.68202	0.9998	5	CAGTATCG	TATCA	0
CAGT[GA]TCG	MA0126.1-ovo	1	0.0494328	6.47569	0.9998	8	CAGTATCG	ACTGTTACT	0
CAGT[GA]TCG	MA0239.1-prd	1	0.0494328	6.47569	0.9998	8	CAGTATCG	ACTGTTACT	0
CAGT[GA]TCG	MA0535.1-Mad	0	0.0512023	6.70751	0.9998	8	CAGTATCG	CGGCGGCGGCGCCTG	0
CAGT[GA]TCG	MA0530.1-CNC::maf-S	5	0.0607194	7.95424	0.9998	8	CAGTATCG	ATTTGCCGAGTCATC	0
CAGT[GA]TCG	MA0252.1-vis	-2	0.0705636	9.24383	0.9998	6	CAGTATCG	CTGTCA	0
CAGT[GA]TCG	MA0460.1-ttk	-1	0.117714	15.4205	0.9998	7	CAGTATCG	ATTATCCTT	0
CAGT[GA]TCG	MA0456.1-opa	0	0.121516	15.9186	0.9998	8	CAGTATCG	CAGCGGGGGGTC	0
CAGT[GA]TCG	MA0445.1-D	2	0.12785	16.7483	0.9998	8	CAGTATCG	TCCATTGTTCT	0
CAGT[GA]TCG	MA0207.1-achi	3	0.195498	25.6103	0.9998	3	CAGTATCG	TGACAG	0
CAGT[GA]TCG	MA0213.1-brk	-2	0.210208	27.5372	0.9998	6	CAGTATCG	CTGGCGCC	0
CAGT[GA]TCG	MA0211.1-bap	2	0.213806	28.0086	0.9998	5	CAGTATCG	TTAAGTG	0
CAGT[GA]TCG	MA0197.1-Oct	1	0.229473	30.061	0.9998	7	CAGTATCG	TAATTAAA	0
CAGT[GA]TCG	MA0227.1-hth	-2	0.234152	30.674	0.9998	6	CAGTATCG	CTGTCA	0
CAGT[GA]TCG	MA0016.1-usp	0	0.238552	31.2503	0.9998	8	CAGTATCG	CCGTGACCCC	0
CAGT[GA]TCG	MA0533.1-SU(HW)	0	0.286354	37.5123	0.9998	8	CAGTATCG	ATTTGTTGCATACTTTTGGGC	0
CAGT[GA]TCG	MA0215.1-btn	1	0.328984	43.0969	0.9998	6	CAGTATCG	TCATTAA	0
CAGT[GA]TCG	MA0249.1-twi	4	0.335519	43.953	0.9998	8	CAGTATCG	TCGCATATGTTG	0
CAGT[GA]TCG	MA0534.1-EcR::usp	5	0.342639	44.8857	0.9998	8	CAGTATCG	AAGGTCAATGAACTC	0
CAGT[GA]TCG	MA0237.2-pan	5	0.363215	47.5812	0.9998	8	CAGTATCG	ATCAAAGGAGCCGA	0
CAGT[GA]TCG	MA0253.1-vnd	4	0.370016	48.4721	0.9998	5	CAGTATCG	TTTCAAGTG	0
CAGT[GA]TCG	MA0243.1-sd	5	0.372497	48.7972	0.9998	7	CAGTATCG	CCGAGGAATGTC	0
CAGT[GA]TCG	MA0182.1-CG4328	-1	0.376167	49.2779	0.9998	7	CAGTATCG	TTTATTA	0

Continued (cluster 2A)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
[AG]TGGACCA[AG]	MA0451.1-kni	3	0.0164951	2.16085	0.994163	9	ATGGACCAA	AAACTAGAGCAC	0
[AG]TGGACCA[AG]	MA0010.1-br_Z1	3	0.0329809	4.3205	0.994163	9	ATGGACCAA	GTAATAAACAAATC	0
[AG]TGGACCA[AG]	MA0201.1-Ptx1	-2	0.0351442	4.60389	0.994163	7	ATGGACCAA	GGATTAA	0
[AG]TGGACCA[AG]	MA0212.1-bcd	-2	0.0409969	5.3706	0.994163	6	ATGGACCAA	GGATTA	0
[AG]TGGACCA[AG]	MA0234.1-oc	-2	0.0436025	5.71193	0.994163	6	ATGGACCAA	GGATTA	0
[AG]TGGACCA[AG]	MA0246.1-so	-2	0.0484876	6.35188	0.994163	6	ATGGACCAA	GTATCA	0
[AG]TGGACCA[AG]	MA0190.1-Gsc	-2	0.055987	7.33429	0.994163	6	ATGGACCAA	GGATTA	0
[AG]TGGACCA[AG]	MA0193.1-Lag1	-2	0.0585269	7.66703	0.994163	7	ATGGACCAA	CTACCAA	0
[AG]TGGACCA[AG]	MA0244.1-slbo	0	0.0626792	8.21097	0.994163	8	ATGGACCAA	ATTGCAAA	0
[AG]TGGACCA[AG]	MA0204.1-Six4	-2	0.0848167	11.111	0.994163	6	ATGGACCAA	GTATCA	0
[AG]TGGACCA[AG]	MA0126.1-ovo	-1	0.092451	12.1111	0.994163	8	ATGGACCAA	AGTAACAGT	0
[AG]TGGACCA[AG]	MA0239.1-prd	-1	0.092451	12.1111	0.994163	8	ATGGACCAA	AGTAACAGT	0
[AG]TGGACCA[AG]	MA0218.1-ct	-1	0.0948412	12.4242	0.994163	6	ATGGACCAA	TTGAAC	0
[AG]TGGACCA[AG]	MA0533.1-SU(HW)	10	0.108293	14.1864	0.994163	9	ATGGACCAA	GCCCAAAAGTATGCAACAAAT	0
[AG]TGGACCA[AG]	MA0185.1-Deaf1	-3	0.121593	15.9287	0.994163	6	ATGGACCAA	CACGAA	0
[AG]TGGACCA[AG]	MA0531.1-CTCF	7	0.125082	16.3858	0.994163	8	ATGGACCAA	CCGCTAGATGGCGCC	0
[AG]TGGACCA[AG]	MA0013.1-br_Z4	2	0.139301	18.2484	0.994163	9	ATGGACCAA	TAGTAAACAAA	0
[AG]TGGACCA[AG]	MA0012.1-br_Z3	-1	0.149818	19.6261	0.994163	8	ATGGACCAA	TAAACTAAAAG	0
[AG]TGGACCA[AG]	MA0205.1-Trl	1	0.158739	20.7949	0.994163	9	ATGGACCAA	GAGAGAGCAA	0
[AG]TGGACCA[AG]	MA0016.1-usp	2	0.17018	22.2935	0.994163	8	ATGGACCAA	CCGTGACCCC	0
[AG]TGGACCA[AG]	MA0022.1-dl_1	-1	0.1719	22.5189	0.994163	8	ATGGACCAA	CGGAAAAACCCC	0
[AG]TGGACCA[AG]	MA0210.1-ara	-3	0.184203	24.1306	0.994163	5	ATGGACCAA	TAACA	0
[AG]TGGACCA[AG]	MA0446.1-fkh	-1	0.185677	24.3237	0.994163	8	ATGGACCAA	TTAAGCAAACA	0
[AG]TGGACCA[AG]	MA0217.1-caup	-3	0.193458	25.343	0.994163	5	ATGGACCAA	TAACA	0
[AG]TGGACCA[AG]	MA0443.1-btd	1	0.20109	26.3428	0.994163	9	ATGGACCAA	AGGGGGCGGA	0
[AG]TGGACCA[AG]	MA0460.1-ttk	0	0.209907	27.4978	0.994163	9	ATGGACCAA	AAGGATAAT	0
[AG]TGGACCA[AG]	MA0534.1-EcR::usp	6	0.234764	30.7541	0.994163	9	ATGGACCAA	GAGTTCATTGACCTT	0
[AG]TGGACCA[AG]	MA0207.1-achi	-2	0.259113	33.9438	0.994163	6	ATGGACCAA	TGACAG	0
[AG]TGGACCA[AG]	MA0252.1-vis	-2	0.27663	36.2386	0.994163	6	ATGGACCAA	TGACAG	0
[AG]TGGACCA[AG]	MA0213.1-brk	0	0.276882	36.2716	0.994163	8	ATGGACCAA	CTGGCGCC	0
[AG]TGGACCA[AG]	MA0199.1-Optix	-3	0.294337	38.5582	0.994163	5	ATGGACCAA	TATCA	0
[AG]TGGACCA[AG]	MA0233.1-mirr	-3	0.294337	38.5582	0.994163	5	ATGGACCAA	AAACA	0
[AG]TGGACCA[AG]	MA0445.1-D	-2	0.311396	40.7929	0.994163	7	ATGGACCAA	AGAACAATGGA	0
[AG]TGGACCA[AG]	MA0235.1-onecut	-1	0.314276	41.1702	0.994163	7	ATGGACCAA	TTGATTT	0
[AG]TGGACCA[AG]	MA0086.1-sna	-3	0.315258	41.2988	0.994163	6	ATGGACCAA	CACCTG	0
[AG]TGGACCA[AG]	MA0222.1-exd	1	0.316577	41.4716	0.994163	7	ATGGACCAA	TTTTGACA	0
[AG]TGGACCA[AG]	MA0231.1-lbe	-2	0.327896	42.9544	0.994163	6	ATGGACCAA	TAACTA	0
[AG]TGGACCA[AG]	MA0458.1-slp1	3	0.328337	43.0121	0.994163	8	ATGGACCAA	AATGTAAACAA	0
[AG]TGGACCA[AG]	MA0085.1-Su(H)	4	0.336329	44.0591	0.994163	9	ATGGACCAA	CTGTGGGAAACGAGAT	0
[AG]TGGACCA[AG]	MA0216.2-CAD	5	0.363526	47.6218	0.994163	6	ATGGACCAA	TTTTTATGGCC	0

Continued (cluster 2A)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
AAAAT[ACT]GAG	MA0451.1-kni	0	0.0018537	0.242839	0.485678	9	AAAATTGAG	AAACTAGAGCAC	0
AAAAT[ACT]GAG	MA0011.1-br_Z2	-1	0.0155006	2.03058	1	8	AAAATTGAG	AAATAGTA	0
AAAAT[ACT]GAG	MA0235.1-onecut	-1	0.0856065	11.2144	1	7	AAAATTGAG	AAATCAA	0
AAAAT[ACT]GAG	MA0012.1-br_Z3	1	0.0916446	12.0054	1	9	AAAATTGAG	TAAACTAAAAG	0
AAAAT[ACT]GAG	MA0169.1-B-H2	0	0.127896	16.7543	1	7	AAAATTGAG	TTAATTG	0
AAAAT[ACT]GAG	MA0529.1-BEAF-32	5	0.130722	17.1246	1	9	AAAATTGAG	TCGAAACTATCGATA	0
AAAAT[ACT]GAG	MA0536.1-pnr	1	0.140659	18.4264	1	9	AAAATTGAG	AACTATCGATA	0
AAAAT[ACT]GAG	MA0445.1-D	3	0.150166	19.6718	1	8	AAAATTGAG	AGAACAATGGA	0
AAAAT[ACT]GAG	MA0168.1-B-H1	0	0.181665	23.7982	1	7	AAAATTGAG	TTAATTG	0
AAAAT[ACT]GAG	MA0176.1-CG15696	-1	0.188003	24.6284	1	7	AAAATTGAG	CAATTAA	0
AAAAT[ACT]GAG	MA0248.1-tup	0	0.188003	24.6284	1	7	AAAATTGAG	TTAATTG	0
AAAAT[ACT]GAG	MA0459.1-tll	1	0.215328	28.2079	1	9	AAAATTGAG	AAAAGTCAAA	0
AAAAT[ACT]GAG	MA0182.1-CG4328	0	0.230243	30.1618	1	7	AAAATTGAG	TTTATTA	0
AAAAT[ACT]GAG	MA0183.1-CG7056	-1	0.232482	30.4551	1	8	AAAATTGAG	TAATTAAA	0
AAAAT[ACT]GAG	MA0229.1-inv	0	0.239182	31.3328	1	8	AAAATTGAG	TAATTAGA	0
AAAAT[ACT]GAG	MA0094.2-Ubx	-1	0.253068	33.1519	1	8	AAAATTGAG	TAATTAAA	0
AAAAT[ACT]GAG	MA0085.1-Su(H)	7	0.253273	33.1788	1	9	AAAATTGAG	CTGTGGGAAACGAGAT	0
AAAAT[ACT]GAG	MA0010.1-br_Z1	0	0.258029	33.8018	1	9	AAAATTGAG	GTAATAAACAAATC	0
AAAAT[ACT]GAG	MA0205.1-Trl	-2	0.274769	35.9947	1	7	AAAATTGAG	GAGAGAGCAA	0
AAAAT[ACT]GAG	MA0208.1-al	-1	0.277537	36.3574	1	7	AAAATTGAG	TAATTAA	0
AAAAT[ACT]GAG	MA0253.1-vnd	-1	0.277563	36.3607	1	8	AAAATTGAG	CACTTGAAA	0
AAAAT[ACT]GAG	MA0171.1-CG11085	0	0.297126	38.9235	1	7	AAAATTGAG	TTAATTG	0
AAAAT[ACT]GAG	MA0453.1-nub	5	0.306515	40.1535	1	7	AAAATTGAG	TATGCAAATTAG	0
AAAAT[ACT]GAG	MA0247.2-tin	0	0.30672	40.1804	1	9	AAAATTGAG	CCACTTGAAA	0
AAAAT[ACT]GAG	MA0172.1-CG11294	-1	0.307095	40.2295	1	7	AAAATTGAG	TAATTAA	0
AAAAT[ACT]GAG	MA0196.1-NK7.1	0	0.307095	40.2295	1	7	AAAATTGAG	TTAATTG	0
AAAAT[ACT]GAG	MA0444.1-CG34031	0	0.307095	40.2295	1	7	AAAATTGAG	TTAATTG	0
AAAAT[ACT]GAG	MA0457.1-PHDP	-1	0.307095	40.2295	1	7	AAAATTGAG	TAATTAA	0
AAAAT[ACT]GAG	MA0192.1-Hmx	0	0.317184	41.5511	1	7	AAAATTGAG	TTAATTG	0
AAAAT[ACT]GAG	MA0174.1-CG42234	-1	0.327301	42.8764	1	7	AAAATTGAG	TAATAAA	0
AAAAT[ACT]GAG	MA0175.1-CG13424	0	0.327301	42.8764	1	7	AAAATTGAG	TTAATTG	0
AAAAT[ACT]GAG	MA0237.2-pan	3	0.33311	43.6375	1	9	AAAATTGAG	ATCAAAGGAGCCGA	0
AAAAT[ACT]GAG	MA0226.1-hbn	-1	0.348264	45.6225	1	7	AAAATTGAG	TAATTAA	0
AAAAT[ACT]GAG	MA0197.1-Oct	-1	0.351331	46.0243	1	8	AAAATTGAG	TAATTAAA	0
AAAAT[ACT]GAG	MA0255.1-z	-4	0.356308	46.6764	1	5	AAAATTGAG	TTGAGTGATT	0
AAAAT[ACT]GAG	MA0448.1-H2.0	-1	0.359005	47.0297	1	7	AAAATTGAG	TAATTAA	0
AAAAT[ACT]GAG	MA0188.1-Dr	-1	0.369865	48.4523	1	7	AAAATTGAG	TAATTGG	0

Continued (cluster 2A)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
CC[AT]C[AT]GGGG	MA0086.1-sna	-3	0.0007863	0.103009	0.120357	6	CCACAGGGG	CAGGTG	0
CC[AT]C[AT]GGGG	MA0456.1-opa	0	0.0009329	0.122209	0.120357	9	CCACAGGGG	CAGCGGGGGGGTC	0
CC[AT]C[AT]GGGG	MA0247.2-tin	0	0.0039671	0.519686	0.341209	9	CCACAGGGG	CCACTTGAAA	0
CC[AT]C[AT]GGGG	MA0531.1-CTCF	5	0.007675	1.00543	0.488763	9	CCACAGGGG	GGCGCCATCTAGCGG	0
CC[AT]C[AT]GGGG	MA0016.1-usp	-5	0.0221385	2.90014	0.637066	4	CCACAGGGG	GGGGTCACGG	0
CC[AT]C[AT]GGGG	MA0213.1-brk	-3	0.0224658	2.94302	0.637066	6	CCACAGGGG	CTGGCGCC	0
CC[AT]C[AT]GGGG	MA0023.1-dl_2	-5	0.0258451	3.38571	0.637066	4	CCACAGGGG	GGGGATTTCC	0
CC[AT]C[AT]GGGG	MA0450.1-hkb	-5	0.0269787	3.53421	0.637066	4	CCACAGGGG	GGGGCGTGA	0
CC[AT]C[AT]GGGG	MA0443.1-btd	-4	0.0271584	3.55776	0.637066	5	CCACAGGGG	AGGGGGCGGA	0
CC[AT]C[AT]GGGG	MA0454.1-odd	3	0.0311984	4.08699	0.641862	8	CCACAGGGG	CTGCTACTGTT	0
CC[AT]C[AT]GGGG	MA0535.1-Mad	6	0.032338	4.23627	0.641862	9	CCACAGGGG	CAGGCGCCGCCGCCG	0
CC[AT]C[AT]GGGG	MA0249.1-twi	2	0.0371836	4.87106	0.676241	9	CCACAGGGG	CAACATATGCGA	0
CC[AT]C[AT]GGGG	MA0237.2-pan	1	0.0393116	5.14982	0.676241	9	CCACAGGGG	ATCAAAGGAGCCGA	0
CC[AT]C[AT]GGGG	MA0533.1-SU(HW)	12	0.0621883	8.14667	0.806056	9	CCACAGGGG	ATTTGTTGCATACTTTTGGGC	0
CC[AT]C[AT]GGGG	MA0242.1-run::Bgb	-3	0.062669	8.20964	0.806056	6	CCACAGGGG	TTGCGGTTA	0
CC[AT]C[AT]GGGG	MA0452.2-KR	1	0.0626701	8.20978	0.806056	9	CCACAGGGG	GAAAAAGGGTTAAA	0
CC[AT]C[AT]GGGG	MA0253.1-vnd	-1	0.0653605	8.56222	0.806056	8	CCACAGGGG	CACTTGAAA	0
CC[AT]C[AT]GGGG	MA0026.1-Eip74EF	0	0.0656013	8.59377	0.806056	7	CCACAGGGG	CTTCCGG	0
CC[AT]C[AT]GGGG	MA0022.1-dl_1	-5	0.0795686	10.4235	0.865005	4	CCACAGGGG	GGGGTTTTTCCC	0
CC[AT]C[AT]GGGG	MA0185.1-Deaf1	-3	0.0828462	10.8529	0.865005	6	CCACAGGGG	TTCGGG	0
CC[AT]C[AT]GGGG	MA0445.1-D	1	0.089642	11.7431	0.868377	9	CCACAGGGG	TCCATTGTTCT	0
CC[AT]C[AT]GGGG	MA0451.1-kni	0	0.093557	12.256	0.868377	9	CCACAGGGG	AAACTAGAGCAC	0
CC[AT]C[AT]GGGG	MA0193.1-Lag1	-3	0.0942311	12.3443	0.868377	6	CCACAGGGG	TTGGTAG	0
CC[AT]C[AT]GGGG	MA0530.1-CNC::maf-S	4	0.148992	19.5179	0.984851	9	CCACAGGGG	GATGACTCGGCAAAT	0
CC[AT]C[AT]GGGG	MA0449.1-h	1	0.190618	24.971	0.984851	9	CCACAGGGG	GGCACGTGCC	0
CC[AT]C[AT]GGGG	MA0205.1-Trl	2	0.286132	37.4832	0.984851	8	CCACAGGGG	TTGCTCTCTC	0
CC[AT]C[AT]GGGG	MA0085.1-Su(H)	-1	0.288386	37.7786	0.984851	8	CCACAGGGG	CTGTGGGAAACGAGAT	0
CC[AT]C[AT]GGGG	MA0255.1-z	2	0.296043	38.7816	0.984851	8	CCACAGGGG	AATCACTCAA	0
CC[AT]C[AT]GGGG	MA0012.1-br_Z3	0	0.355495	46.5699	0.984851	9	CCACAGGGG	CTTTTAGTTTA	0

Cluster 2B

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
C[TC]GTGCAAA	MA0244.1-slbo	-1	0.00756214	0.99064	0.796844	8	CCGTGCAAA	TTTGCAAT	0
C[TC]GTGCAAA	MA0446.1-fkh	0	0.00905167	1.18577	0.796844	9	CCGTGCAAA	TTAAGCAAACA	0
C[TC]GTGCAAA	MA0453.1-nub	-1	0.00983971	1.289	0.796844	8	CCGTGCAAA	TATGCAAATTAG	0
C[TC]GTGCAAA	MA0218.1-ct	-2	0.01296	1.69776	0.796844	6	CCGTGCAAA	GTTCAA	0
C[TC]GTGCAAA	MA0530.1-CNC::maf-S	5	0.0152293	1.99503	0.796844	9	CCGTGCAAA	GATGACTCGGCAAAT	0
C[TC]GTGCAAA	MA0085.1-Su(H)	0	0.0189348	2.48046	0.825607	9	CCGTGCAAA	CTGTGGGAAACGAGAT	0
C[TC]GTGCAAA	MA0254.1-vvl	-3	0.0311405	4.07941	0.998459	6	CCGTGCAAA	TGCATA	0
C[TC]GTGCAAA	MA0222.1-exd	-1	0.0558003	7.30983	0.998459	8	CCGTGCAAA	TGTCAAAA	0
C[TC]GTGCAAA	MA0458.1-slp1	-1	0.0743944	9.74567	0.998459	8	CCGTGCAAA	AATGTAAACAA	0
[CA]C[GA]TGTCCC	MA0016.1-usp	0	7.63352e-05	0.00999991	0.0199469	9	ACATGTCCC	CCGTGACCCC	0
[CA]C[GA]TGTCCC	MA0085.1-Su(H)	3	0.00457748	0.59965	0.598063	9	ACATGTCCC	ATCTCGGTTCCCACAA	0
[CA]C[GA]TGTCCC	MA0449.1-h	1	0.0101941	1.33542	0.665945	9	ACATGTCCC	GGCACGTGCC	0
[CA]C[GA]TGTCCC	MA0249.1-twi	2	0.0266046	3.4852	0.997353	9	ACATGTCCC	CAACATATGCGA	0
[CA]C[GA]TGTCCC	MA0450.1-hkb	0	0.0325307	4.26153	0.997353	9	ACATGTCCC	TCACGCCCC	0
[CA]C[GA]TGTCCC	MA0252.1-vis	-2	0.0454891	5.95907	0.997353	6	ACATGTCCC	CTGTCA	0
[CA]C[GA]TGTCCC	MA0531.1-CTCF	5	0.057349	7.51271	0.997353	9	ACATGTCCC	CCGCTAGATGGCGCC	0
[CA]C[GA]TGTCCC	MA0204.1-Six4	-2	0.071726	9.39611	0.997353	6	ACATGTCCC	GTATCA	0
[CA]C[GA]TGTCCC	MA0227.1-hth	-2	0.071726	9.39611	0.997353	6	ACATGTCCC	CTGTCA	0
[CA]C[GA]TGTCCC	MA0530.1-CNC::maf-S	5	0.0760455	9.96196	0.997353	9	ACATGTCCC	ATTTGCCGAGTCATC	0
CAA[AG][AT]CAAA	MA0459.1-tll	1	0.00295006	0.386458	0.539645	9	CAAAACAAA	AAAAGTCAAA	0
CAA[AG][AT]CAAA	MA0235.1-onecut	-1	0.00413036	0.541077	0.539645	7	CAAAACAAA	AAATCAA	0
CAA[AG][AT]CAAA	MA0010.1-br_Z1	3	0.0202305	2.6502	0.997353	9	CAAAACAAA	GTAATAAACAAATC	0
CAA[AG][AT]CAAA	MA0049.1-hb	1	0.0439624	5.75908	0.997353	9	CAAAACAAA	GCATAAAAAA	0
CAA[AG][AT]CAAA	MA0445.1-D	-1	0.0491173	6.43437	0.997353	8	CAAAACAAA	AGAACAATGGA	0
CAA[AG][AT]CAAA	MA0180.1-Vsx2	0	0.0547003	7.16574	0.997353	9	CAAAACAAA	CTAATTAAA	0
CAA[AG][AT]CAAA	MA0222.1-exd	-2	0.0587462	7.69575	0.997353	7	CAAAACAAA	TGTCAAAA	0
CAA[AG][AT]CAAA	MA0446.1-fkh	0	0.0670461	8.78304	0.997353	9	CAAAACAAA	TTAAGCAAACA	0
[AC][AC]GCAGCG	MA0456.1-opa	-3	0.00412459	0.540321	0.748907	5	AAGCAGCG	CAGCGGGGGGTC	0
[AC][AC]GCAGCG	MA0530.1-CNC::maf-S	2	0.0133746	1.75208	0.748907	8	AAGCAGCG	ATTTGCCGAGTCATC	0
[AC][AC]GCAGCG	MA0213.1-brk	-2	0.0140822	1.84477	0.748907	6	AAGCAGCG	CTGGCGCC	0
[AC][AC]GCAGCG	MA0454.1-odd	2	0.0158862	2.0811	0.748907	8	AAGCAGCG	AACAGTAGCAG	0
[AC][AC]GCAGCG	MA0535.1-Mad	0	0.0172271	2.25675	0.748907	8	AAGCAGCG	CGGCGGCGGCGCCTG	0
[AC][AC]GCAGCG	MA0237.2-pan	5	0.020009	2.62118	0.748907	8	AAGCAGCG	ATCAAAGGAGCCGA	0
[AC][AC]GCAGCG	MA0450.1-hkb	-2	0.0515615	6.75456	1	6	AAGCAGCG	GGGGCGTGA	0
[AC][AC]GCAGCG	MA0449.1-h	-2	0.066813	8.75251	1	6	AAGCAGCG	GGCACGTGCC	0
[AC][AC]GCAGCG	MA0443.1-btd	0	0.0753812	9.87494	1	8	AAGCAGCG	AGGGGGCGGA	0
GTTGTCA[AG]A	MA0222.1-exd	-2	0.00256028	0.335396	0.280284	7	GTTGTCAAA	TGTCAAAA	0
GTTGTCA[AG]A	MA0207.1-achi	-1	0.00276429	0.362122	0.280284	6	GTTGTCAAA	CTGTCA	0
GTTGTCA[AG]A	MA0252.1-vis	-1	0.0032303	0.423169	0.280284	6	GTTGTCAAA	CTGTCA	0
GTTGTCA[AG]A	MA0227.1-hth	-1	0.00935121	1.22501	0.608533	6	GTTGTCAAA	CTGTCA	0
GTTGTCA[AG]A	MA0217.1-caup	-2	0.0237231	3.10772	0.993421	5	GTTGTCAAA	TGTTA	0
GTTGTCA[AG]A	MA0204.1-Six4	-1	0.0284969	3.7331	0.993421	6	GTTGTCAAA	GTATCA	0
GTTGTCA[AG]A	MA0199.1-Optix	-2	0.0416178	5.45193	0.993421	5	GTTGTCAAA	TATCA	0

Continued (cluster 2B)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
T[TC]GTGCAAA	MA0446.1-fkh	0	0.00227503	0.298029	0.596058	9	TCGTGCAAA	TTAAGCAAACA	0
T[TC]GTGCAAA	MA0244.1-slbo	-1	0.0078718	1.03121	0.850777	8	TCGTGCAAA	TTTGCAAT	0
T[TC]GTGCAAA	MA0453.1-nub	-1	0.0107255	1.40504	0.850777	8	TCGTGCAAA	TATGCAAATTAG	0
T[TC]GTGCAAA	MA0218.1-ct	-2	0.012989	1.70155	0.850777	6	TCGTGCAAA	GTTCAA	0
T[TC]GTGCAAA	MA0254.1-vvl	-3	0.0310011	4.06114	0.999983	6	TCGTGCAAA	TGCATA	0
T[TC]GTGCAAA	MA0085.1-Su(H)	0	0.0400946	5.25239	0.999983	9	TCGTGCAAA	CTGTGGGAAACGAGAT	0
T[TC]GTGCAAA	MA0222.1-exd	-1	0.0586292	7.68042	0.999983	8	TCGTGCAAA	TGTCAAAA	0
T[TC]GTGCAAA	MA0530.1-CNC::maf-S	0	0.0650185	8.51743	0.999983	9	TCGTGCAAA	ATTTGCCGAGTCATC	0
AA[CA]AACCCC	MA0022.1-dl_1	3	1.37254e-05	0.00179803	0.0035866	9	AAAAACCCC	CGGAAAAACCCC	0
AA[CA]AACCCC	MA0023.1-dl_2	1	0.00046837	0.0613561	0.0611937	9	AAAAACCCC	GGAAACCCC	0
AA[CA]AACCCC	MA0242.1-run::Bgb	-2	0.0170558	2.23431	0.997353	7	AAAAACCCC	TAACCGCAA	0
AA[CA]AACCCC	MA0237.2-pan	3	0.0338308	4.43183	0.997353	9	AAAAACCCC	ATCAAAGGAGCCGA	0
AA[CA]AACCCC	MA0452.2-KR	0	0.0412494	5.40367	0.997353	9	AAAAACCCC	TTTAACCCTTTTTC	0
AA[CA]AACCCC	MA0534.1-EcR::usp	6	0.0642231	8.41323	0.997353	9	AAAAACCCC	AAGGTCAATGAACTC	0
AA[CA]AACCCC	MA0456.1-opa	-3	0.0697528	9.13761	0.997353	6	AAAAACCCC	GACCCCCCGCTG	0
[CG]CGC[CT]ACC	MA0531.1-CTCF	1	2.81702e-05	0.0036903	0.0071991	8	CCGCCACC	GGCGCCATCTAGCGG	0
[CG]CGC[CT]ACC	MA0443.1-btd	1	0.00072307	0.0947222	0.0702087	8	CCGCCACC	TCCGCCCCCT	0
[CG]CGC[CT]ACC	MA0213.1-brk	1	0.00082419	0.107969	0.0702087	7	CCGCCACC	GGCGCCAG	0
[CG]CGC[CT]ACC	MA0016.1-usp	0	0.00191472	0.250828	0.110749	8	CCGCCACC	CCGTGACCCC	0
[CG]CGC[CT]ACC	MA0535.1-Mad	3	0.00225765	0.295752	0.110749	8	CCGCCACC	CAGGCGCCGCCGCCG	0
[CG]CGC[CT]ACC	MA0454.1-odd	0	0.00260019	0.340625	0.110749	8	CCGCCACC	CTGCTACTGTT	0
[CG]CGC[CT]ACC	MA0193.1-Lag1	-3	0.00666929	0.873677	0.243482	5	CCGCCACC	CTACCAA	0
[CG]CGC[CT]ACC	MA0237.2-pan	0	0.00826607	1.08285	0.264055	8	CCGCCACC	TCGGCTCCTTTGAT	0
[CG]CGC[CT]ACC	MA0456.1-opa	0	0.0385889	5.05515	0.819486	8	CCGCCACC	GACCCCCCGCTG	0
[CG]CGC[CT]ACC	MA0204.1-Six4	-1	0.0463974	6.07806	0.819486	6	CCGCCACC	TGATAC	0
[CG]CGC[CT]ACC	MA0451.1-kni	2	0.0477552	6.25593	0.819486	8	CCGCCACC	GTGCTCTAGTTT	0
[CG]CGC[CT]ACC	MA0205.1-Trl	2	0.0584601	7.65827	0.819486	8	CCGCCACC	TTGCTCTCTC	0
[CG]CGC[CT]ACC	MA0242.1-run::Bgb	3	0.0600597	7.86782	0.819486	6	CCGCCACC	TAACCGCAA	0
[CG]CGC[CT]ACC	MA0450.1-hkb	2	0.0600597	7.86782	0.819486	7	CCGCCACC	TCACGCCCC	0
[CG]CGC[CT]ACC	MA0530.1-CNC::maf-S	7	0.0600946	7.8724	0.819486	8	CCGCCACC	ATTTGCCGAGTCATC	0
[CG]CGC[CT]ACC	MA0126.1-ovo	1	0.0679041	8.89544	0.826346	8	CCGCCACC	ACTGTTACT	0
[CG]CGC[CT]ACC	MA0239.1-prd	1	0.0679041	8.89544	0.826346	8	CCGCCACC	ACTGTTACT	0
AAAA[TG][CG]GTG	MA0452.2-KR	2	0.00552615	0.723925	1	9	AAAAGCGTG	GAAAAAGGGTTAAA	0
AAAA[TG][CG]GTG	MA0011.1-br_Z2	-1	0.0389623	5.10406	1	8	AAAAGCGTG	AAATAGTA	0
AAAA[TG][CG]GTG	MA0533.1-SU(HW)	4	0.041387	5.4217	1	9	AAAAGCGTG	GCCCAAAAGTATGCAACAAAT	0
AAAA[TG][CG]GTG	MA0451.1-kni	0	0.0713545	9.34744	1	9	AAAAGCGTG	AAACTAGAGCAC	0
GAA[AC]TTCCA	MA0532.1-STAT92E	1	0.00406654	0.532717	0.99999	9	GAAATTCCA	CGGAATTCCAGGAAA	0
GAA[AC]TTCCA	MA0243.1-sd	0	0.00895737	1.17342	0.99999	9	GAAATTCCA	GACATTCCTCGA	0
GAA[AC]TTCCA	MA0023.1-dl_2	1	0.0152917	2.00322	0.99999	9	GAAATTCCA	GGAAACCCC	0
GAA[AC]TTCCA	MA0459.1-tll	0	0.0636528	8.33851	0.99999	9	GAAATTCCA	AAAAGTCAAA	0
CAA[AGT]ACG[GC]C	MA0237.2-pan	3	0.0042497	0.556711	0.999999	9	CAATACGCC	ATCAAAGGAGCCGA	0
CAA[AGT]ACG[GC]C	MA0023.1-dl_2	1	0.0120813	1.58266	0.999999	9	CAATACGCC	GGAAAACCCC	0
CAA[AGT]ACG[GC]C	MA0022.1-dl_1	3	0.0296226	3.88056	0.999999	9	CAATACGCC	CGGAAAAACCCC	0
CAA[AGT]ACG[GC]C	MA0531.1-CTCF	3	0.0497962	6.5233	0.999999	9	CAATACGCC	CCGCTAGATGGCGCC	0
CAA[AGT]ACG[GC]C	MA0012.1-br_Z3	4	0.0716508	9.38625	0.999999	7	CAATACGCC	TAAACTAAAAG	0

Continued (cluster 2B)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
[AG][GT]CTGCGCG	MA0535.1-Mad	4	0.00470551	0.616422	0.6366	9	AGCTGCGCG	CGGCGGCGGCGCCTG	0
[AG][GT]CTGCGCG	MA0456.1-opa	-2	0.00487244	0.63829	0.6366	7	AGCTGCGCG	CAGCGGGGGGTC	0
[AG][GT]CTGCGCG	MA0213.1-brk	-1	0.0125272	1.64107	0.84535	8	AGCTGCGCG	CTGGCGCC	0
[AG][GT]CTGCGCG	MA0237.2-pan	-1	0.0157184	2.05911	0.84535	8	AGCTGCGCG	TCGGCTCCTTTGAT	0
[AG][GT]CTGCGCG	MA0450.1-hkb	-3	0.0219709	2.87819	0.84535	6	AGCTGCGCG	GGGGCGTGA	0
[AG][GT]CTGCGCG	MA0242.1-run::Bgb	-2	0.0261458	3.4251	0.84535	7	AGCTGCGCG	TTGCGGTTA	0
[AG][GT]CTGCGCG	MA0255.1-z	-2	0.0300095	3.93125	0.84535	7	AGCTGCGCG	TTGAGTGATT	0
[AG][GT]CTGCGCG	MA0443.1-btd	-1	0.035586	4.66177	0.84535	8	AGCTGCGCG	AGGGGGCGGA	0
[AG][GT]CTGCGCG	MA0449.1-h	-1	0.035586	4.66177	0.84535	8	AGCTGCGCG	GGCACGTGCC	0
[AG][GT]CTGCGCG	MA0451.1-kni	-2	0.040232	5.2704	0.846118	7	AGCTGCGCG	GTGCTCTAGTTT	0
[AG][GT]CTGCGCG	MA0205.1-Trl	-2	0.0420944	5.51436	0.846118	7	AGCTGCGCG	TTGCTCTCTC	0
[AG][GT]CTGCGCG	MA0249.1-twi	4	0.0496238	6.50072	0.92247	8	AGCTGCGCG	CAACATATGCGA	0
[AG][GT]CTGCGCG	MA0185.1-Deaf1	-3	0.0529533	6.93688	0.92247	6	AGCTGCGCG	TTCGGG	0
[AG][GT]CTGCGCG	MA0026.1-Eip74EF	-2	0.0588005	7.70287	0.931385	7	AGCTGCGCG	CTTCCGG	0
[AG][GT]CTGCGCG	MA0531.1-CTCF	6	0.0702984	9.20909	0.977393	9	AGCTGCGCG	GGCGCCATCTAGCGG	0
[AG][GT]CTGCGCG	MA0085.1-Su(H)	-2	0.0710678	9.30988	0.977393	7	AGCTGCGCG	CTGTGGGAAACGAGAT	0
[AG][GT]CTGCGCG	MA0529.1-BEAF-32	7	0.0762292	9.98602	0.995959	8	AGCTGCGCG	TATCGATAGTTTCGA	0
ACGCCC[AT]A	MA0450.1-hkb	2	0.00024077	0.0315411	0.0629153	7	ACGCCCAA	TCACGCCCC	0
ACGCCC[AT]A	MA0443.1-btd	1	0.00492461	0.645124	0.643416	8	ACGCCCAA	TCCGCCCCCT	0
ACGCCC[AT]A	MA0193.1-Lag1	-1	0.035985	4.71403	0.997353	7	ACGCCCAA	CTACCAA	0
ACGCCC[AT]A	MA0456.1-opa	1	0.0375025	4.91282	0.997353	8	ACGCCCAA	GACCCCCCGCTG	0
ACGCCC[AT]A	MA0255.1-z	2	0.0446909	5.85451	0.997353	8	ACGCCCAA	AATCACTCAA	0
ACGCCC[AT]A	MA0451.1-kni	0	0.0511871	6.70551	0.997353	8	ACGCCCAA	GTGCTCTAGTTT	0
CCGG[ACG]AA[AC]	MA0026.1-Eip74EF	0	0.00016995	0.0222627	0.044066	7	CCGGGAAA	CCGGAAG	0
CCGG[ACG]AA[AC]	MA0532.1-STAT92E	7	0.00155756	0.20404	0.176636	8	CCGGGAAA	CGGAATTCCAGGAAA	0
CCGG[ACG]AA[AC]	MA0022.1-dl_1	-1	0.00493245	0.646151	0.319742	7	CCGGGAAA	CGGAAAAACCCC	0
CCGG[ACG]AA[AC]	MA0535.1-Mad	6	0.0328741	4.30651	0.989681	8	CCGGGAAA	CAGGCGCCGCCGCCG	0
CCGG[ACG]AA[AC]	MA0185.1-Deaf1	0	0.0410366	5.37579	0.989681	6	CCGGGAAA	CACGAA	0
CCGG[ACG]AA[AC]	MA0242.1-run::Bgb	3	0.0418261	5.47922	0.989681	6	CCGGGAAA	TAACCGCAA	0
CCGG[ACG]AA[AC]	MA0085.1-Su(H)	2	0.0507507	6.64834	0.989681	8	CCGGGAAA	CTGTGGGAAACGAGAT	0
CCGG[ACG]AA[AC]	MA0243.1-sd	0	0.0562084	7.36331	0.989681	8	CCGGGAAA	CCGAGGAATGTC	0
CCGG[ACG]AA[AC]	MA0086.1-sna	2	0.062439	8.17951	0.989681	4	CCGGGAAA	CACCTG	0
CCGG[ACG]AA[AC]	MA0213.1-brk	4	0.0752083	9.85229	0.989681	4	CCGGGAAA	GGCGCCAG	0
CCGG[ACG]AA[AC]	MA0216.2-CAD	2	0.0752786	9.86149	0.989681	8	CCGGGAAA	GGCCATAAAAA	0
AATGAAC[CG]C	MA0534.1-EcR::usp	6	0.00068488	0.0897188	0.142498	9	AATGAACCC	AAGGTCAATGAACTC	0
AATGAAC[CG]C	MA0218.1-ct	-1	0.0281802	3.69161	0.486247	6	AATGAACCC	TTGAAC	0
AATGAAC[CG]C	MA0016.1-usp	1	0.0344606	4.51434	0.486247	9	AATGAACCC	CCGTGACCCC	0
AATGAAC[CG]C	MA0186.1-Dfd	2	0.0353639	4.63267	0.486247	5	AATGAACCC	TTAATGA	0
AATGAAC[CG]C	MA0203.1-Scr	2	0.0407502	5.33828	0.486247	5	AATGAACCC	TTAATGA	0
AATGAAC[CG]C	MA0256.1-zen	2	0.0407502	5.33828	0.486247	5	AATGAACCC	CTAATGA	0
AATGAAC[CG]C	MA0215.1-btn	2	0.0468097	6.13208	0.486247	5	AATGAACCC	TTAATGA	0
AATGAAC[CG]C	MA0180.1-Vsx2	2	0.0468484	6.13714	0.486247	7	AATGAACCC	CTAATTAAA	0
AATGAAC[CG]C	MA0452.2-KR	-1	0.0601212	7.87588	0.486247	8	AATGAACCC	TTTAACCCTTTTTC	0
AATGAAC[CG]C	MA0166.1-Antp	2	0.0605667	7.93424	0.486247	5	AATGAACCC	TTAATGA	0
AATGAAC[CG]C	MA0194.1-Lim1	1	0.0605667	7.93424	0.486247	6	AATGAACCC	TAATTAA	0
AATGAAC[CG]C	MA0235.1-onecut	1	0.0631049	8.26675	0.486247	6	AATGAACCC	AAATCAA	0
AATGAAC[CG]C	MA0208.1-al	1	0.0656208	8.59633	0.486247	6	AATGAACCC	TAATTAA	0
AATGAAC[CG]C	MA0197.1-Oct	1	0.0672562	8.81056	0.486247	7	AATGAACCC	TAATTAAA	0
AATGAAC[CG]C	MA0172.1-CG11294	1	0.0682244	8.9374	0.486247	6	AATGAACCC	TAATTAA	0
AATGAAC[CG]C	MA0176.1-CG15696	1	0.0682244	8.9374	0.486247	6	AATGAACCC	CAATTAA	0
AATGAAC[CG]C	MA0192.1-Hmx	1	0.0682244	8.9374	0.486247	6	AATGAACCC	CAATTAA	0
AATGAAC[CG]C	MA0226.1-hbn	1	0.0682244	8.9374	0.486247	6	AATGAACCC	TAATTAA	0
AATGAAC[CG]C	MA0240.1-repo	1	0.0682244	8.9374	0.486247	6	AATGAACCC	TAATTAA	0
AATGAAC[CG]C	MA0094.2-Ubx	1	0.0692829	9.07605	0.486247	7	AATGAACCC	TAATTAAA	0
AATGAAC[CG]C	MA0178.1-CG32105	1	0.0734002	9.61543	0.486247	6	AATGAACCC	TAATTAA	0
AATGAAC[CG]C	MA0221.1-eve	2	0.0734002	9.61543	0.486247	5	AATGAACCC	CTAATGA	0
AATGAAC[CG]C	MA0250.1-unc-4	1	0.0760311	9.96007	0.486247	6	AATGAACCC	CAATTAA	0
AATGAAC[CG]C	MA0022.1-dl_1	2	0.0761707	9.97836	0.486247	9	AATGAACCC	CGGAAAAACCCC	0

Continued (cluster 2B)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
AG[CGT]GTTACG	MA0126.1-ovo	0	0.00130171	0.170524	0.119639	9	AGTGTTACG	ACTGTTACT	0
AG[CGT]GTTACG	MA0239.1-prd	0	0.00130171	0.170524	0.119639	9	AGTGTTACG	ACTGTTACT	0
AG[CGT]GTTACG	MA0452.2-KR	5	0.001388	0.181828	0.119639	9	AGTGTTACG	GAAAAAGGGTTAAA	0
AG[CGT]GTTACG	MA0016.1-usp	0	0.00787695	1.03188	0.409524	9	AGTGTTACG	GGGGTCACGG	0
AG[CGT]GTTACG	MA0447.1-gt	-3	0.00950223	1.24479	0.409524	6	AGTGTTACG	ATTACGTAAT	0
AG[CGT]GTTACG	MA0217.1-caup	-2	0.0411004	5.38415	0.986961	5	AGTGTTACG	TGTTA	0
AG[CGT]GTTACG	MA0234.1-oc	-1	0.0729871	9.56131	0.986961	6	AGTGTTACG	GGATTA	0
G[AT]ACGGCA	MA0242.1-run::Bgb	0	0.00506309	0.663264	0.999999	8	GAACGGCA	TAACCGCAA	0
G[AT]ACGGCA	MA0451.1-kni	3	0.0248379	3.25377	0.999999	8	GAACGGCA	AAACTAGAGCAC	0
G[AT]ACGGCA	MA0026.1-Eip74EF	-2	0.0393109	5.14972	0.999999	6	GAACGGCA	CCGGAAG	0
G[AT]ACGGCA	MA0244.1-slbo	-2	0.0488427	6.3984	0.999999	6	GAACGGCA	ATTGCAAA	0
AAACA[ATG]A[AC]	MA0233.1-mirr	0	0.00288987	0.378573	0.543185	5	AAACATAA	AAACA	0
AAACA[ATG]A[AC]	MA0210.1-ara	0	0.00415745	0.544626	0.543185	5	AAACATAA	TAACA	0
AAACA[ATG]A[AC]	MA0013.1-br_Z4	4	0.00844857	1.10676	0.65123	7	AAACATAA	TAGTAAACAAA	0
AAACA[ATG]A[AC]	MA0217.1-caup	0	0.011214	1.46904	0.65123	5	AAACATAA	TAACA	0
AAACA[ATG]A[AC]	MA0012.1-br_Z3	1	0.012461	1.6324	0.65123	8	AAACATAA	TAAACTAAAAG	0
AAACA[ATG]A[AC]	MA0458.1-slp1	5	0.0176538	2.31265	0.768841	6	AAACATAA	AATGTAAACAA	0
AAACA[ATG]A[AC]	MA0010.1-br_Z1	5	0.0261249	3.42236	0.97523	8	AAACATAA	GTAATAAACAAATC	0
AAACA[ATG]A[AC]	MA0446.1-fkh	6	0.0299675	3.92574	0.978838	5	AAACATAA	TTAAGCAAACA	0
AAACA[ATG]A[AC]	MA0011.1-br_Z2	0	0.0555283	7.27421	0.997353	8	AAACATAA	AAATAGTA	0
[AG]GGAACTCC	MA0023.1-dl_2	-1	0.00031872	0.0417526	0.0829638	8	AGGAACTCC	GGAAAACCCC	0
[AG]GGAACTCC	MA0532.1-STAT92E	5	0.00065331	0.0855833	0.0850285	9	AGGAACTCC	TTTCCTGGAATTCCG	0
[AG]GGAACTCC	MA0022.1-dl_1	1	0.0122241	1.60136	0.641988	9	AGGAACTCC	CGGAAAAACCCC	0
[AG]GGAACTCC	MA0085.1-Su(H)	4	0.0269105	3.52527	0.993508	9	AGGAACTCC	CTGTGGGAAACGAGAT	0
[AG]GGAACTCC	MA0534.1-EcR::usp	7	0.0490023	6.4193	0.993508	8	AGGAACTCC	AAGGTCAATGAACTC	0
[AG]GGAACTCC	MA0126.1-ovo	0	0.0491522	6.43894	0.993508	9	AGGAACTCC	AGTAACAGT	0
[AG]GGAACTCC	MA0239.1-prd	0	0.0491522	6.43894	0.993508	9	AGGAACTCC	AGTAACAGT	0
GAAA[AC][GC]GCC	MA0023.1-dl_2	1	0.00043013	0.0563471	0.112396	9	GAAAACGCC	GGAAAACCCC	0
GAAA[AC][GC]GCC	MA0022.1-dl_1	2	0.0009108	0.119315	0.118999	9	GAAAACGCC	CGGAAAAACCCC	0
GAAA[AC][GC]GCC	MA0242.1-run::Bgb	-1	0.00832305	1.09032	0.60397	8	GAAAACGCC	TAACCGCAA	0
GAAA[AC][GC]GCC	MA0085.1-Su(H)	6	0.0156329	2.04791	0.60397	9	GAAAACGCC	CTGTGGGAAACGAGAT	0
GAAA[AC][GC]GCC	MA0237.2-pan	3	0.0160112	2.09746	0.60397	9	GAAAACGCC	ATCAAAGGAGCCGA	0
GAAA[AC][GC]GCC	MA0126.1-ovo	1	0.0161794	2.1195	0.60397	8	GAAAACGCC	AGTAACAGT	0
GAAA[AC][GC]GCC	MA0239.1-prd	1	0.0161794	2.1195	0.60397	8	GAAAACGCC	AGTAACAGT	0
GAAA[AC][GC]GCC	MA0535.1-Mad	5	0.0649422	8.50743	0.997353	9	GAAAACGCC	CAGGCGCCGCCGCCG	0
G[CT]AAACAAA	MA0013.1-br_Z4	2	0.00022834	0.0299119	0.0591121	9	GCAAACAAA	TAGTAAACAAA	0
G[CT]AAACAAA	MA0458.1-slp1	3	0.00182262	0.238763	0.235923	8	GCAAACAAA	AATGTAAACAA	0
G[CT]AAACAAA	MA0446.1-fkh	4	0.00574855	0.75306	0.495868	7	GCAAACAAA	TTAAGCAAACA	0
G[CT]AAACAAA	MA0233.1-mirr	-2	0.00874014	1.14496	0.495868	5	GCAAACAAA	AAACA	0
G[CT]AAACAAA	MA0010.1-br_Z1	3	0.0104726	1.37191	0.495868	9	GCAAACAAA	GTAATAAACAAATC	0
G[CT]AAACAAA	MA0210.1-ara	-2	0.0119459	1.56491	0.495868	5	GCAAACAAA	TAACA	0
G[CT]AAACAAA	MA0012.1-br_Z3	-1	0.0134079	1.75643	0.495868	8	GCAAACAAA	TAAACTAAAAG	0
G[CT]AAACAAA	MA0049.1-hb	0	0.0206759	2.70855	0.669082	9	GCAAACAAA	GCATAAAAAA	0
G[CT]AAACAAA	MA0217.1-caup	-2	0.0291612	3.82012	0.838817	5	GCAAACAAA	TAACA	0
G[CT]AAACAAA	MA0235.1-onecut	-1	0.0343003	4.49334	0.887978	7	GCAAACAAA	AAATCAA	0
G[CT]AAACAAA	MA0453.1-nub	3	0.0526547	6.89777	0.988104	9	GCAAACAAA	TATGCAAATTAG	0
G[CT]AAACAAA	MA0216.2-CAD	1	0.070274	9.2059	0.988104	9	GCAAACAAA	GGCCATAAAAA	0

Continued (cluster 2B)

#Query ID	Target ID	Optimal offset	p-value	E-value	q-value	Overlap	Query consensus	Target consensus	Orientation
CCC[AT]CCC[AC]C	MA0443.1-btd	0	2.09406e-05	0.00274322	0.0054509	9	CCCACCCAC	TCCGCCCCCT	0
CCC[AT]CCC[AC]C	MA0456.1-opa	1	0.00035257	0.0461864	0.045887	9	CCCACCCAC	GACCCCCCGCTG	0
CCC[AT]CCC[AC]C	MA0450.1-hkb	1	0.00190533	0.249598	0.16532	8	CCCACCCAC	TCACGCCCC	0
CCC[AT]CCC[AC]C	MA0255.1-z	1	0.00595544	0.780163	0.387552	9	CCCACCCAC	AATCACTCAA	0
CCC[AT]CCC[AC]C	MA0205.1-Trl	1	0.0121413	1.59051	0.632081	9	CCCACCCAC	TTGCTCTCTC	0
CCC[AT]CCC[AC]C	MA0193.1-Lag1	-1	0.0192361	2.51993	0.658238	7	CCCACCCAC	CTACCAA	0
CCC[AT]CCC[AC]C	MA0452.2-KR	5	0.0200643	2.62843	0.658238	9	CCCACCCAC	TTTAACCCTTTTTC	0
CCC[AT]CCC[AC]C	MA0085.1-Su(H)	9	0.02023	2.65013	0.658238	7	CCCACCCAC	ATCTCGGTTCCCACAA	0
CCC[AT]CCC[AC]C	MA0016.1-usp	0	0.0312048	4.08783	0.902518	9	CCCACCCAC	CCGTGACCCC	0
CCC[AT]CCC[AC]C	MA0531.1-CTCF	3	0.0489892	6.41759	0.993517	9	CCCACCCAC	GGCGCCATCTAGCGG	0

Table 2.7.4

Results of evolution conservation tests from Cluster 1 to Cluster 2B

Table 2.7.4

Cluster 1

motif_seq	Evol_test	AG	CX	Both	location	NumGene
CATA[CG][AT]AAA	0	0	1	1	600-800R	83
G[CG]GTTATGA	0	0	0	0	0	28
CAC[ACG]C[AG]CCC	0	0	0	0	0	35
GAA[AC][CT]GTCA	1	2	1	0	200-400F	68
[AG]GTCA[AT]GGA	0	0	0	1	0	25
GCC[AG]CGG[AC]C	0	0	0	0	0	14
CA[AC]ATGCGG	0	0	0	0	0	8
C[AGT]C[AG]GGAAG	0	0	1	0	0	6
GCAG[AC]GCA	0	1	0	0	0	25
T[CT]AC[CG]GC	1	1	1	1	200-400F	107
GGAA[CT]C[AC]CA	0	0	0	0	0	16
[ACT]ACGTC[AT]AA	1	0	3	1	200-400F	56
AA[AC]CAAAAC	1	0	3	0	0-200R	53
ATTCCAGT[AG]	0	0	0	0	0	14
[AC]TG[AG]GCCC	0	0	0	0	0	38

Cluster 6

motif_seq	Evol_test	AG	CX	Both	location	numGene
CA[AG]ACGTCA	0	2	0	0	1000-1200F	40
CG[AT]ATCCCA	0	0	0	0	0	21
CAAAAAT[AGT]G	0	2	0	0	600-800R	65
GCCA[AT]AAA[CT]	0	1	1	0	400-600F	46
CGAC[GA]CG[CT]C	0	0	0	0	0	11

Cluster 4

motif_seq	Evol_test	AG	CX	Both	location	numGene
AGC[AGT]A[AG]ACC	0	0	0	0	0	30
CTGTT[CG]AAA	0	0	0	0	0	38
AAA[AT]ATGGG	0	0	0	0	0	24
GAC[TA][TG]CCC	0	0	0	0	0	47
CTCCT[GT][AT]GC	0	0	0	0	0	19
CAG[AT][AC]GCTG	0	0	0	0	0	9
GG[AT]AGCAC	0	0	0	0	0	27
CAAGA[GC]CAC	0	0	0	0	0	14
[CT]AATAAAAA	1	1	2	0	600-800F	65
TGCC[AC][AC]AAA	0	0	0	0	0	34
GAGGA[ACT]AAA	0	0	2	0	0-200F	28
AT[AG]ACTCGA	0	0	0	0	0	12
CTT[TA]G[CG]CGA	0	0	0	0	0	22

Cluster 7

motif_seq	Evol_test	AG	CX	Both	location	numGene
GCC[CGT][AC]AGCA	0	2	0	0	400-600F	15
T[CA]CTGTCGG	0	0	0	0	0	3
AAAAGT[GA]A[AG]	0	0	0	0	0	71
TCAC[ACG]TGAA	0	0	0	0	0	28
ATT[CA]C[GA]GAC	0	0	0	0	0	11
AAACCG[AC]AC	0	0	0	0	0	12
TGG[AT]AGCAA	0	0	0	0	0	9
CGCG[CT]G[AT]A	0	0	0	0	0	25
CT[CG][AGT]CTCTC	0	0	0	0	0	25
CGC[CTG]C[CG]AA	0	0	0	0	0	32
AA[CG]GACGAA	0	0	0	0	0	21

Cluster 5

motif_seq	Evol_test	AG	CX	Both	location	numGene
ATC[GT]CTCGC	0	0	0	0	0	9
[AT]CTGACTG	0	0	0	0	0	33
TTGATTGA	0	0	0	0	0	53
G[CT]AACGGAA	0	0	0	0	0	15
AA[CA]TGGTG	0	1	0	1	200-400F	40
AC[GC]ATGAC	0	0	0	0	0	33
ACTCA[CG]TC	0	0	0	0	0	24

Cluster 2A

motif_seq	Evol_test	AG	CX	Both	location	numGene
CA[AC]G[AG]AA	1	6	5	1	200-400F	170
G[GC]G[CG]CTAAC	0	0	0	0	0	4
AAAAG[CG]GCC	0	0	0	0	0	13
GCCG[GT][AC]GCC	0	0	0	0	0	10
ACCTCA[ACGT]A[CT]	0	0	0	0	0	25
AAAA[AC]TCGA	0	0	0	0	0	45
AT[AGT]CCGAAC	0	0	0	0	0	13
GCGA[TA]TCC	0	0	0	0	0	31
CGT[GT][ATC]GGCA	0	0	0	0	0	26
CG[CT][GT]TGGCA	0	0	0	0	0	28
CGC[CG]GGC[CT]G	0	0	0	0	0	11
AAAC[AG]CATG	0	0	0	0	0	20
CAGT[GA]TCG	0	0	0	0	0	29
[AG]TGGACCA[AG]	0	0	0	0	0	25
AAAAT[ACT]GAG	0	0	0	0	0	51
CC[AT]C[AT]GGGG	0	0	0	0	0	16

Cluster 2B

motif_seq	Evol_test	AG	CX	Both	location	numGene
C[TC]GTGCAAA	0	0	0	0	0	21
[CA]C[GA]TGTCCC	0	0	0	0	0	18
CAA[AG][AT]CAAA	1	0	5	2	0-200F	81
[AC][AC]GCAGCG	0	0	0	0	0	46
GTTGTCA[AG]A	0	0	0	0	0	25
T[TC]GTGCAAA	0	0	0	0	0	24
AA[CA]AACCCC	0	0	0	0	0	22
[CG]CGC[CT]ACC	0	0	0	0	0	36
AAAA[TG][CG]GTG	0	0	0	0	0	38
GAA[AC]TTCCA	0	0	0	0	0	28
CAA[AGT]ACG[GC]C	0	0	0	0	0	26
[AG][GT]CTGCGCG	0	0	0	0	0	21
ACGCCC[AT]A	0	0	0	0	0	26
CCGG[ACG]AA[AC]	1	1	0	2	0-200F	67
AATGAAC[CG]C	0	0	0	0	0	24
AG[CGT]GTTACG	0	0	0	0	0	28
G[AT]ACGGCA	0	0	0	0	0	32
AAACA[ATG]A[AC]	1	3	12	6	0-200F	174
[AG]GGAACTCC	0	0	0	0	0	3
GAAA[AC][GC]GCC	0	0	0	0	0	28
G[CT]AAACAAA	1	1	2	1	0-200F	58
CCC[AT]CCC[AC]C	0	0	0	0	0	46

Table 2.7.5

Results of position bias tests from Cluster 1 to Cluster 2B

Table 2.7.5

Cluster 1

motif_seq_fwd	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
CATA[CG][AT]AAA	3	6	4	7	7	9	10	8	4	11
G[CG]GTTATGA	1	5	5	2	1	2	3	3	1	3
CAC[ACG]C[AG]CCC	2	7	5	5	3	5	2	4	1	4
GAA[AC][CT]GTCA	4	8	6	5	5	8	0	5	5	1
[AG]GTCA[AT]GGA	0	0	1	3	3	2	2	2	2	2
GCC[AG]CGG[AC]C	1	1	0	2	1	2	0	1	1	1
CA[AC]ATGCGG	0	0	2	0	0	3	1	0	0	1
C[AGT]C[AG]GGAAG	0	1	9	7	9	7	6	5	5	2
GCAG[AC]GCA	1	4	0	1	1	1	2	2	1	2
T[CT]AC[CG]GC	12	7	7	8	5	10	12	6	9	9
GGAA[CT]C[AC]CA	2	1	1	0	2	0	11	11	9	2
[ACT]ACGTC[AT]AA	4	7	5	3	3	6	0	5	3	2
AA[AC]CAAAAC	2	5	4	4	6	4	4	3	0	3
ATTCCAGT[AG]	2	3	1	0	0	1	11	10	10	0
[AC]TG[AG]GCCC	1	3	2	2	6	1	3	4	2	2
	0.46034838	0.094907495	0.2732347	0.05584836	0.05575738	0.0192127	0.01132817	0.03261857	0.27137255	0.00662959
	0.56551461	0.043461438	0.04346144	0.27619773	0.56525079	0.27554082	0.14428797	0.14428797	0.5641462	0.1439344
	0.20563817	0.010344248	0.03236662	0.03236662	0.10755923	0.03232833	0.20538268	0.05830879	0.42437879	0.058239
	0.10762927	0.011036846	0.03381177	0.0599482	0.05989938	0.01099332	0.45412285	0.05972136	0.05969531	0.72292547
	0.55314173	0.553141735	0.57711546	0.14801712	0.14793638	0.28175127	0.28175127	0.28175127	0.28167159	0.28105863
	0.38261623	0.382616233	0.76554605	0.18607695	0.38254541	0.1859277	0.7661158	0.38228685	0.38224897	0.38195748
	0.71404935	0.714049349	0.20156026	0.71404935	0.71419634	0.10538474	0.41578334	0.71473377	0.7148126	0.41528397
	0.63722048	0.481852829	0.00384579	0.01166702	0.00384437	0.01164685	0.02050886	0.0364923	0.03648499	0.23227714
	0.64414498	0.091725471	0.5048182	0.64414498	0.64380485	0.64256261	0.31898975	0.31898975	0.64238058	0.31807876
	0.17896206	0.83851237	0.83851237	0.64452817	0.70910056	0.34885973	0.17623419	0.94829926	0.47659957	0.47181116
	0.28416991	0.58029765	0.58029765	0.55068577	0.28401866	0.55168364	0.00158016	0.00158016	0.0046978	0.28276399
	0.10862234	0.019437078	0.06051674	0.19844543	0.1982937	0.03400562	0.45132051	0.06028561	0.19765917	0.37002222
	0.36433053	0.05898258	0.10594068	0.10594068	0.03323026	0.1055564	0.1055564	0.19307832	0.45911762	0.19240559
	0.27950492	0.14642691	0.57167211	0.55738559	0.55759751	0.57041621	0.00155252	0.00267193	0.00267114	0.55936283
	0.84757018	0.244180009	0.44920177	0.44920177	0.0430346	0.8451963	0.24316559	0.13465232	0.44735502	0.44589012

motif_seq_rcp	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
TTT[AT][CG]TATG	2	4	7	8	3	13	1	5	5	7
TCATAAC[CG]C	0	3	2	3	3	4	2	2	2	2
GGG[CT]G[CGT]GTG	0	4	3	1	2	5	4	4	2	2
TGAC[AG][GT]TTC	7	7	6	6	5	7	2	4	7	4
TCC[AT]TGAC[CT]	0	0	5	1	1	2	2	2	1	0
G[GT]CCG[CT]GGC	1	0	0	0	2	0	0	0	0	1
CCGCAT[GT]TG	0	2	0	0	1	4	0	1	0	0
CTTCC[CT]G[ACT]G	2	0	0	0	0	0	0	0	0	1
TGC[GT]CTGC	2	9	22	22	16	1	0	1	0	0
GC[CG]GT[AG]A	9	8	5	8	8	14	5	5	2	8
TG[GT]G[AG]TTCC	0	1	0	2	0	0	0	1	1	1
TT[AT]GACGT[AGT]	1	5	5	3	5	7	8	4	4	4
GTTTTG[GT]TT	8	2	4	1	1	2	4	4	4	1
[CT]ACTGGAAT	1	0	1	1	1	0	0	0	0	1
GGGC[CT]CA[GT]	1	1	3	2	4	3	4	4	1	3
	0.76436481	0.273234704	0.05584836	0.03287713	0.45982279	0.00233719	0.77973633	0.16012855	0.16000426	0.05500613
	0.56225328	0.144642002	0.27619773	0.144642	0.14456585	0.07825396	0.27554082	0.27554082	0.27546531	0.27488444
	0.70242543	0.058378635	0.10758735	0.42498439	0.20558323	0.03232833	0.05830879	0.05830879	0.2053533	0.20512732
	0.01924859	0.019248591	0.03381177	0.03381177	0.05989938	0.01917343	0.36825623	0.10723285	0.0191648	0.1068373
	0.55314173	0.553141735	0.04449763	0.57711546	0.57684006	0.28175127	0.28175127	0.28175127	0.57568699	0.55513289
	0.38261623	0.765546055	0.76554605	0.76554605	0.18604486	0.7661158	0.7661158	0.7661158	0.76618142	0.38195748
	0.71404935	0.201560258	0.71404935	0.71404935	0.41617535	0.05720308	0.71473377	0.41578334	0.7148126	0.71542015
	0.23309514	0.637220482	0.63722048	0.63722048	0.6374015	0.63806346	0.63806346	0.63806346	0.63816057	0.48020076
	0.31990152	0.005370269	5.3528482004	5.3528482004	0.00012484	0.64256261	0.50588518	0.64256261	0.50600813	0.50695604
	0.48265711	0.644528175	0.70769172	0.64452817	0.64317223	0.08387892	0.71425987	0.71425987	0.14081081	0.6319356
	0.55068577	0.58029765	0.55068577	0.28416991	0.55090003	0.55168364	0.55168364	0.57900185	0.5788528	0.57770554
	0.73158552	0.060516743	0.06051674	0.19844543	0.060467	0.01936041	0.0111016	0.10821875	0.1081724	0.10781607
	0.01085179	0.364330533	0.10594068	0.71852692	0.71811973	0.36315971	0.1055564	0.1055564	0.10551226	0.71473578
	0.57167211	0.557385592	0.57167211	0.57167211	0.57140213	0.55837254	0.55837254	0.55837254	0.55848626	0.56915986
	0.84757018	0.847570182	0.24418001	0.44920177	0.13512359	0.24316559	0.13465232	0.13465232	0.84492306	0.24215301

Cluster 6

motif_seq_fwd	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
CA[AG]ACGTCA	3	3	3	1	2	6	2	2	1	1
CG[AT]ATCCCA	1	0	3	2	2	3	0	3	0	0
CAAAAAT[AGT]G	4	5	7	3	3	4	4	3	4	4
GCCA[AT]AAA[CT]	2	1	5	5	2	1	2	2	4	3
CGAC[GA]CG[CT]C	1	1	0	2	1	0	1	2	0	0
	0.11875574	0.118755743	0.11858067	0.46931252	0.22691735	0.01999357	0.22654127	0.22654127	0.46854449	0.46777642
	0.41121171	0.721093463	0.10419923	0.19897825	0.19897825	0.10413309	0.72243354	0.10407537	0.72243354	0.72310643
	0.10475055	0.058302863	0.01864084	0.19100402	0.19100402	0.10419287	0.10399988	0.19034519	0.10399988	0.10362577
	0.36629071	0.721691445	0.05915194	0.05912862	0.3651085	0.71885831	0.36392726	0.36392726	0.1058082	0.19284087
	0.41228646	0.41228646	0.72018454	0.1994752	0.41180721	0.72058202	0.41132798	0.1992524	0.72092952	0.72160574

motif_seq_rcp	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
TGACGT[CT]TG	2	2	3	5	2	0	5	2	3	0
TGGGAT[AT]CG	0	1	2	0	1	1	1	0	1	0
C[ACT]ATTTTTG	3	5	6	8	2	4	5	4	4	3
[AG]TTT[AT]TGGC	2	1	5	4	4	2	3	4	1	0
G[AG]CG[TC]GTCG	0	1	1	1	1	1		0	0	1
	0.22729367	0.227293665	0.11858067	0.03559966	0.22691735	0.65069749	0.03554169	0.22654127	0.11836471	0.65194462
	0.72109346	0.411211714	0.19900136	0.72176255	0.41073788	0.41050808	0.41026407	0.72243354	0.41026407	0.72310643
	0.19166395	0.058302863	0.0327574	0.0106808	0.35954909	0.10419287	0.05787447	0.10399988	0.10399988	0.18968746
	0.36629071	0.721691445	0.05915194	0.10619625	0.10619625	0.36453548	0.19351997	0.1058082	0.7178748	0.45938605
	0.71958284	0.41228646	0.41185753	0.41180721	0.41180721	0.41157478	0.72092952	0.72092952	0.72092952	0.41084877

Cluster 4

motif_seq_fwd	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
AGC[AGT]A[AG]ACC	2	3	2	0	2	2	3	0	1	2
CTGTT[CG]AAA	2	2	0	0	2	5	7	0	2	3
AAA[AT]ATGGG	1	0	1	1	2	0	2	1	4	0
GAC[TA][TG]CCC	0	2	4	0	3	0	5	1	3	1
CTCCT[GT][AT]GC	1	1	0	0	0	1	4	3	4	4
CAG[AT][AC]GCTG	1	0	0	0	0	0	0	0	0	1
GG[AT]AGCAC	2	2	3	0	1	2	0	2	1	1
CAAGA[GC]CAC	0	0	0	0	3	2	1	0	1	0
[CT]AATAAAAA	5	3	4	6	2	6	7	4	3	6
TGCC[AC][AC]AAA	0	3	3	7	1	1	0	0	3	2
GAGGA[ACT]AAA	5	0	1	1	0	2	2	2	5	1
AT[AG]ACTCGA	0	0	0	0	2	0	0	2	0	1
CTT[TA]G[CG]CGA	2	0	1	1	2	3	3	3	3	3
	0.27976913	0.146569552	0.27976913	0.55700067	0.27976913	0.27976913	0.14656955	0.55700067	0.57161159	0.27907052
	0.35876307	0.35876307	0.46326525	0.46326525	0.35876307	0.05794182	0.01858495	0.46326525	0.35826414	0.18988034
	0.55187907	0.573299536	0.55187907	0.57329954	0.2689414	0.57329954	0.2689414	0.55187907	0.07622957	0.57428953
	0.39665004	0.436989107	0.13085249	0.39665004	0.23672901	0.39665004	0.07337139	0.82994074	0.23630653	0.82755571
	0.48057492	0.48057492	0.63852563	0.63852563	0.63852563	0.48057492	0.0658197	0.12144733	0.06576965	0.06570198
	0.47124161	0.648247051	0.64824705	0.64824705	0.64824705	0.64824705	0.64824705	0.64824705	0.64860659	0.47044396
	0.30695912	0.306959122	0.16141602	0.52043	0.62145936	0.30695912	0.52043	0.30695912	0.62080996	0.61993111
	0.64782853	0.647828526	0.64782853	0.64782853	0.11915244	0.22805686	0.47163677	0.64782853	0.47129691	0.64867612
	0.15022246	0.435252738	0.25619919	0.08801347	0.73107794	0.08801347	0.05159391	0.25619919	0.4342509	0.08737871
	0.45760918	0.193903279	0.19390328	0.01894508	0.71895188	0.71895188	0.45760918	0.45760918	0.19360745	0.36338464
	0.04211503	0.57477484	0.55009319	0.55009319	0.57477484	0.26799788	0.26799788	0.26799788	0.04207144	0.54890254
	0.56195354	0.561953545	0.56195354	0.56195354	0.27639917	0.56195354	0.56195354	0.27639917	0.56238226	0.56462396
	0.27841663	0.558975978	0.5696501	0.5696501	0.27841663	0.14583892	0.14583892	0.14583892	0.14568044	0.14546617

motif_seq_rcp	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
GGT[CT]T[ACT]GCT	3	1	1	1	1	2	2	3	1	2
TTT[CG]AACAG	3	2	2	4	1	1	2	1	1	1
CCCAT[AT]TTT	3	0	0	1	2	0	3	2	3	2
GGG[CA][TA]GTC	5	4	3	1	2	3	1	4	2	3
GC[AT][AC]AGGAG	1	3	1	1	1	1	0	1	0	0
CAGC[GT][AT]CTG	0	0	0	3	5	6	7	4	7	5
GTGCT[AT]CC	3	2	1	2	2	1	3	0	2	1
GTG[GC]TCTTG	1	1	1	1	1	1	2	2	2	0
TTTTTATT[AG]	4	4	5	4	1	1	4	3	5	3
TTT[GT][GT]GGCA	3	1	5	2	3	3	0	3	0	1
TTT[AGT]TCCTC	3	1	0	1	1	2	1	1	1	0
TCGAGT[CT]AT	0	3	1	0	1	4	2	3	1	3
TCG[CG]C[TA]AAG	0	0	0	2	1	1	1	2	1	3
	0.14656955	0.572162802	0.5721628	0.5721628	0.5721628	0.27976913	0.27976913	0.14656955	0.57161159	0.27907052
	0.19055305	0.35876307	0.35876307	0.10411805	0.70948722	0.70948722	0.35876307	0.70948722	0.70867335	0.7075717
	0.14074105	0.573299536	0.57329954	0.55187907	0.2689414	0.57329954	0.14074105	0.2689414	0.1405966	0.26830756
	0.07337139	0.130852486	0.23672901	0.82994074	0.43698911	0.23672901	0.82994074	0.13085249	0.43629338	0.23573546
	0.48057492	0.121447333	0.48057492	0.48057492	0.48057492	0.48057492	0.63852563	0.48057492	0.63889354	0.63939209
	0.64824705	0.648247051	0.64824705	0.11905159	0.03574537	0.02009078	0.01141024	0.06452796	0.01140211	0.03568497
	0.16141602	0.306959122	0.62145936	0.30695912	0.30695912	0.62145936	0.16141602	0.52043	0.30659268	0.61993111
	0.47163677	0.471636766	0.47163677	0.47163677	0.47163677	0.47163677	0.22805686	0.22805686	0.22789006	0.64867612
	0.25619919	0.256199191	0.15022246	0.25619919	0.81045421	0.81045421	0.25619919	0.43525274	0.1497899	0.43289641
	0.19390328	0.718951876	0.05903166	0.36459323	0.19390328	0.19390328	0.45760918	0.19390328	0.45810227	0.71699683
	0.14023539	0.550093192	0.57477484	0.55009319	0.55009319	0.26799788	0.55009319	0.55009319	0.54958721	0.57576222
	0.56195354	0.144750453	0.56589099	0.56195354	0.56589099	0.0785079	0.27639917	0.14475045	0.56535256	0.14438477
	0.55897598	0.558975978	0.55897598	0.27841663	0.5696501	0.5696501	0.5696501	0.27841663	0.569104	0.14546617

Cluster 7

motif_seq_fwd	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
GCC[CGT][AC]AGCA	0	0	2	0	0	1	1	0	0	0
T[CA]CTGTCGG	1	0	1	7	8	8	9	7	7	1
AAAAGT[GA]A[AG]	5	3	6	4	9	4	1	9	11	7
TCAC[ACG]TGAA	0	0	1	3	1	1	3	1	2	0
ATT[CA]C[GA]GAC	0	0	4	4	2	1	3	5	1	0
AAACCG[AC]AC	3	3	2	2	2	2	0	2	2	1
TGG[AT]AGCAA	0	0	1	7	7	6	7	4	6	0
CGCG[CT]G[AT]A	1	4	3	3	4	2	2	3	2	0
CT[CG][AGT]CTCTC	2	4	1	0	0	1	4	2	4	1
CGC[CTG]C[CG]AA	1	5	4	3	4	4	5	4	5	0
AA[CG]GACGAA	1	1	0	0	1	1	0	1	2	2
	0.63382992	0.634580158	0.23433107	0.63469244	0.63469244	0.48434459	0.48405378	0.63555766	0.63555766	0.63624311
	0.41709	0.713553483	0.41657889	0.01015914	0.00581836	0.00581836	0.00335126	0.01014795	0.01014795	0.41566405
	0.16334056	0.462948925	0.09554423	0.27479152	0.0195168	0.27479152	0.77464397	0.01935469	0.00673838	0.05546572
	0.55335878	0.554236142	0.575536	0.14755435	0.575536	0.575536	0.14742508	0.57423619	0.28088759	0.55618173
	0.44697205	0.447977182	0.10936032	0.10936032	0.37468049	0.73513046	0.19948643	0.06070038	0.7331295	0.45020789
	0.12169251	0.121504314	0.23251744	0.23251744	0.23251744	0.23251744	0.63870287	0.2321033	0.2321033	0.47918816
	0.64265074	0.643385781	0.47576275	0.01151906	0.01151906	0.02028309	0.01151238	0.06503918	0.02024842	0.64501497
	0.84128139	0.133147518	0.24048887	0.24048887	0.13306993	0.44316535	0.44260236	0.23947938	0.44150983	0.39427304
	0.28943912	0.082204631	0.58859676	0.5443645	0.5443645	0.58859676	0.08209946	0.28794953	0.08195192	0.58615986
	0.89133276	0.126632901	0.21877807	0.37868828	0.21877807	0.21877807	0.12625893	0.21750602	0.1257349	0.27618941
	0.57372983	0.572612148	0.55677952	0.55677952	0.57244511	0.57244511	0.5571219	0.57115991	0.27922896	0.27868238

motif_seq_rcp	(0,200].2	(200,400].2	(400,600].2	(600,800].2	(800,1000].2	(1000,1200].2	(1200,1400].2	(1400,1600].2	(1600,1800].2	(1800,2000].2
TGCT[GT][ACG]GGC	0	0	2	0	0	1	1	0	0	0
CCGACAG[TG]A	1	0	1	7	8	8	9	7	7	1
[CT]T[TC]ACTTTT	5	3	6	4	9	4	1	9	11	7
TTCA[CGT]GTGA	0	0	1	3	1	1	3	1	2	0
GTC[TC]G[TG]AAT	0	0	4	4	2	1	3	5	1	0
GT[GT]CGGTTT	3	3	2	2	2	2	0	2	2	1
TTGCT[AT]CCA	0	0	1	7	7	6	7	4	6	0
T[AT]C[AG]CGCG	1	4	3	3	4	2	2	3	2	0
GAGAG[ACT][CG]AG	2	4	1	0	0	1	4	2	4	1
TT[CG]G[CAG]GCG	1	5	4	3	4	4	5	4	5	0
TTCGTC[CG]TT	1	1	0	0	1	1	0	1	2	2
	0.234756	0.234386295	0.23433107	0.12242319	0.23433107	0.23433107	0.23418666	0.23390643	0.12220134	0.23357118
	0.71294696	0.713553483	0.71364424	0.71364424	0.41657889	0.71364424	0.71388178	0.71434355	0.71434355	0.71489744
	0.77024618	0.767772462	0.77352583	0.05624267	0.03311865	0.27479152	0.76622837	0.76405891	0.27307485	0.09428951
	0.08025949	0.575704936	0.55436747	0.28158987	0.08004994	0.28158987	0.07997878	0.55537964	0.55537964	0.2803332
	0.44697205	0.735390428	0.73513046	0.37468049	0.10936032	0.10936032	0.73445038	0.44928796	0.7331295	0.45020789
	0.63755633	0.480795266	0.6384115	0.48068646	0.6384115	0.6384115	0.48040188	0.47984945	0.47984945	0.63994893
	0.64265074	0.475868337	0.47576275	0.47576275	0.47576275	0.47576275	0.64378369	0.64434345	0.22968588	0.64501497
	0.44482194	0.133147518	0.83889303	0.39210251	0.44316535	0.44316535	0.83808011	0.39331297	0.83650082	0.83460825
	0.28943912	0.082204631	0.58859676	0.28869401	0.5443645	0.28869401	0.58813403	0.5453931	0.58723559	0.58615986
	0.65596804	0.378943226	0.65319589	0.21877807	0.65319589	0.37868828	0.27440394	0.89755715	0.65042293	0.37518322
	0.57372983	0.556648718	0.07955326	0.57244511	0.07955326	0.55677952	0.5571219	0.57115991	0.55778767	0.14598243

Cluster 5

motif_seq_fwd	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
ATC[GT]CTCGC	0	2	0	1	2	2	2	1	1	0
[AT]CTGACTG	2	1	3	1	1	1	1	2	1	0
TTGATTGA	6	6	1	3	5	4	9	1	1	1
G[CT]AACGGAA	2	0	2	2	1	0	2	1	1	0
AA[CA]TGGTG	3	7	3	4	1	2	5	1	1	4
AC[GC]ATGAC	2	0	2	4	0	2	0	3	2	4
ACTCA[CG]TC	1	0	1	3	1	0	1	2	0	2
	0.7113064	0.202502035	0.7113064	0.41829841	0.20250203	0.20250203	0.20250203	0.41829841	0.41829841	0.7113064
	0.43792462	0.831301372	0.23729758	0.83130137	0.83130137	0.83130137	0.83130137	0.43792462	0.83130137	0.39595583
	0.07339002	0.073390023	0.89423511	0.37881754	0.12658202	0.21886215	0.01453731	0.89423511	0.89423511	0.89423511
	0.23174482	0.640014883	0.23174482	0.23174482	0.47912416	0.64001488	0.23174482	0.47912416	0.47912416	0.64001488
	0.36958842	0.04128951	0.36958842	0.21288441	0.90904733	0.64027241	0.12285466	0.90904733	0.90904733	0.21288441
	0.64945885	0.275643263	0.64945885	0.21706498	0.27564326	0.64945885	0.27564326	0.37604872	0.64945885	0.21706498
	0.8909918	0.272633041	0.8909918	0.38086687	0.8909918	0.27263304	0.8909918	0.65627281	0.27263304	0.65627281

motif_seq_rcp	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
GCGAG[AC]GAT	0	2	0	1	2	2	2	1	1	0
CAGTCAG[AT]	2	1	3	1	1	1	1	2	1	0
TCAATCAA	6	6	1	3	5	4	9	1	1	1
TTCCGTT[AG]C	2	0	2	2	1	0	2	1	1	0
CACCA[TG]TT	3	7	3	4	1	2	5	1	1	4
GTCAT[GC]GT	2	0	2	4	0	2	0	3	2	4
GA[CG]TGAGT	1	0	1	3	1	0	1	2	0	2
	0.41829841	0.202502035	0.20250203	0.41829841	0.7113064	0.7113064	0.7113064	0.7113064	0.7113064	0.7113064
	0.01369437	0.041691594	0.23729758	0.43792462	0.23729758	0.39595583	0.43792462	0.39595583	0.43792462	0.39595583
	0.65337848	0.378817542	0.12658202	0.21886215	0.89423511	0.21886215	0.37881754	0.21886215	0.65337848	0.37881754
	0.47912416	0.640014883	0.64001488	0.64001488	0.47912416	0.64001488	0.23174482	0.64001488	0.64001488	0.23174482
	0.27976481	0.909047334	0.07111081	0.64027241	0.90904733	0.64027241	0.90904733	0.36958842	0.90904733	0.90904733
	0.89864328	0.217064978	0.64945885	0.27564326	0.64945885	0.27564326	0.89864328	0.89864328	0.37604872	0.37604872
	0.65627281	0.380866866	0.2201944	0.65627281	0.8909918	0.65627281	0.8909918	0.8909918	0.27263304	0.8909918

Cluster 2A

motif_seq_fwd	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
CA[AC]G[AG]AA	16	30	31	26	20	19	18	15	20	18
G[GC]G[CG]CTAAC	1	0	3	0	0	5	0	9	1	0
AAAAG[CG]GCC	0	0	3	3	0	6	0	4	1	1
GCCG[GT][AC]GCC	1	0	0	2	2	2	2	1	2	1
ACCTCA[ACGT]A[CT]	1	0	3	2	0	3	3	0	0	2
AAAA[AC]TCGA	1	3	4	3	6	0	8	2	2	5
AT[AGT]CCGAAC	3	0	0	1	0	1	2	2	1	2
GCGA[TA]TCC	1	0	2	0	2	2	0	2	3	3
CGT[GT][ATC]GGCA	1	1	4	1	2	1	0	1	0	2
CG[CT][GT]TGGCA	0	1	7	2	1	1	0	1	1	2
CGC[CG]GGC[CT]G	1	0	2	0	0	2	1	0	0	1
AAAC[AG]CATG	0	3	2	0	1	4	0	1	1	0
CAGT[GA]TCG	0	0	2	1	2	1	4	5	2	0
[AG]TGGACCA[AG]	0	0	1	0	2	2	1	3	4	1
AAAAT[ACT]GAG	4	4	6	11	3	1	0	2	1	9
CC[AT]C[AT]GGGG	0	0	2	0	1	1	2	0	0	0
	0.27973035	0.00204895	0.00136927	0.00974253	0.08186144	0.11299519	0.15419738	0.36211616	0.08186144	0.15419738
	0.59503054	0.54023427	0.15305587	0.54023427	0.54023427	0.04605282	0.54023427	0.00484703	0.59410453	0.54023427
	0.63072173	0.63130261	0.12330134	0.12330134	0.63130261	0.02079882	0.63130261	0.06682107	0.48772083	0.48772083
	0.3842707	0.76309914	0.76309914	0.18672308	0.18672308	0.18672308	0.18672308	0.38404127	0.18672308	0.38404127
	0.76133428	0.43402049	0.2089358	0.3904469	0.43402049	0.2089358	0.2089358	0.43402049	0.43402049	0.3904469
	0.74907423	0.20435917	0.11201496	0.20435917	0.03525819	0.44091299	0.01152087	0.38262795	0.38262795	0.06246278
	0.123724	0.63027241	0.63027241	0.48875482	0.63027241	0.48875482	0.23652584	0.23652584	0.48875482	0.23652584
	0.66213078	0.49364216	0.32969992	0.49364216	0.32969992	0.32969992	0.49364216	0.32969992	0.17406785	0.17406785
	0.48963065	0.48904367	0.06700756	0.48904367	0.2366699	0.48904367	0.62998528	0.48904367	0.62998528	0.2366699
	0.62940268	0.48904367	0.01184369	0.2366699	0.48904367	0.48904367	0.62998528	0.48904367	0.48904367	0.2366699
	0.38363427	0.76418461	0.18643542	0.76418461	0.76418461	0.18643542	0.38340702	0.76418461	0.76418461	0.38340702
	0.54180074	0.15215367	0.2900608	0.54248781	0.59109107	0.08258667	0.54248781	0.59109107	0.59109107	0.54248781
	0.38641624	0.38722395	0.44992345	0.84860282	0.44992345	0.84860282	0.13551485	0.07609835	0.44992345	0.38722395
	0.53954479	0.54023427	0.59410453	0.54023427	0.29171619	0.29171619	0.59410453	0.15305587	0.0830853	0.59410453
	0.1114974	0.111209	0.03499168	0.00222269	0.20295666	0.74391534	0.44307368	0.38022274	0.74391534	0.00659403
	0.76379506	0.76418461	0.18643542	0.76418461	0.38340702	0.38340702	0.18643542	0.76418461	0.76418461	0.76418461

Continued (cluster 2A)

motif_seq_rcp	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
TT[CT]C[GT]TG	16	30	31	26	20	19	18	15	20	18
GTTAG[CG]C[GC]C	1	0	3	0	0	5	0	9	1	0
GGC[CG]CTTTT	0	0	3	3	0	6	0	4	1	1
GGC[GT][AC]CGGC	1	0	0	2	2	2	2	1	2	1
[AG]T[ACGT]TGAGGT	1	0	3	2	0	3	3	0	0	2
TCGA[GT]TTTT	1	3	4	3	6	0	8	2	2	5
GTTCGG[ACT]AT	3	0	0	1	0	1	2	2	1	2
GGA[TA]TCGC	1	0	2	0	2	2	0	2	3	3
TGCC[GAT][AC]ACG	1	1	4	1	2	1	0	1	0	2
TGCCA[AC][AG]CG	0	1	7	2	1	1	0	1	1	2
C[AG]GCC[CG]GCG	1	0	2	0	0	2	1	0	0	1
CATG[CT]GTTT	0	3	2	0	1	4	0	1	1	0
CGA[TC]ACTG	0	0	2	1	2	1	4	5	2	0
[CT]TGGTCCA[CT]	0	0	1	0	2	2	1	3	4	1
CTC[AGT]ATTTT	4	4	6	11	3	1	0	2	1	9
CCCC[AT]G[AT]GG	0	0	2	0	1	1	2	0	0	0
	0.21057742	0.00665549	0.02931663	0.11299519	0.00204895	0.02045846	0.74939249	0.02045846	0.02931663	0.04166435
	0.53954479	0.54023427	0.59410453	0.29171619	0.54023427	0.54023427	0.54023427	0.54023427	0.54023427	0.54023427
	0.63072173	0.48772083	0.48772083	0.2360105	0.48772083	0.63130261	0.48772083	0.48772083	0.63130261	0.2360105
	0.76270791	0.76309914	0.38404127	0.38404127	0.38404127	0.76309914	0.18672308	0.76309914	0.76309914	0.76309914
	0.76133428	0.75993331	0.2089358	0.3904469	0.75993331	0.3904469	0.3904469	0.02058914	0.2089358	0.3904469
	0.74907423	0.38262795	0.38262795	0.74770535	0.74770535	0.38262795	0.11201496	0.44091299	0.38262795	0.11201496
	0.62969019	0.48875482	0.63027241	0.06696681	0.12357099	0.23652584	0.63027241	0.63027241	0.48875482	0.12357099
	0.17443881	0.49364216	0.17406785	0.32969992	0.17406785	0.66100324	0.66100324	0.66100324	0.17406785	0.32969992
	0.48963065	0.2366699	0.00677873	0.2366699	0.2366699	0.06700756	0.48904367	0.03711297	0.06700756	0.48904367
	0.62940268	0.2366699	0.06700756	0.2366699	0.12364639	0.06700756	0.48904367	0.03711297	0.06700756	0.48904367
	0.76379506	0.76418461	0.18643542	0.76418461	0.18643542	0.18643542	0.76418461	0.38340702	0.38340702	0.76418461
	0.54180074	0.2900608	0.15215367	0.59109107	0.59109107	0.2900608	0.54248781	0.59109107	0.54248781	0.59109107
	0.85022099	0.38722395	0.84860282	0.38722395	0.44992345	0.07609835	0.84860282	0.44992345	0.44992345	0.44992345
	0.53954479	0.59410453	0.0830853	0.15305587	0.29171619	0.29171619	0.29171619	0.29171619	0.59410453	0.29171619
	0.74527417	0.111209	0.01143156	0.74391534	0.74391534	0.20295666	0.74391534	0.06199996	0.06199996	0.38022274
	0.76379506	0.76418461	0.76418461	0.76418461	0.01662276	0.38340702	0.09782758	0.05318397	0.38340702	0.18643542

Cluster 2B

motif_seq_fwd	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
C[TC]GTGCAAA	1	3	2	1	0	1	1	2	4	2
[CA]C[GA]TGTCCC	0	1	0	0	2	0	0	1	2	4
CAA[AG][AT]CAAA	16	3	7	2	5	4	4	1	5	7
[AC][AC]GCAGCG	2	3	4	4	1	2	3	4	1	4
GTTGTCA[AG]A	2	1	4	2	1	5	0	0	0	1
T[TC]GTGCAAA	1	3	1	3	1	1	0	1	1	1
AA[CA]AACCCC	0	2	2	1	2	0	2	1	3	1
[CG]CGC[CT]ACC	1	2	0	4	3	1	5	2	2	5
AAAA[TG][CG]GTG	3	1	2	1	3	2	1	4	0	4
GAA[AC]TTCCA	0	1	1	3	2	3	0	0	2	2
CAA[AGT]ACG[GC]C	1	0	2	1	6	1	1	1	2	3
[AG][GT]CTGCGCG	1	2	0	3	1	2	1	0	3	1
ACGCCC[AT]A	2	0	2	5	2	2	1	1	1	2
CCGG[ACG]AA[AC]	10	1	4	0	5	6	3	1	7	7
AATGAAC[CG]C	1	2	0	0	0	4	2	1	2	2
AG[CGT]GTTACG	1	3	2	0	2	4	5	4	1	0
G[AT]ACGGCA	2	2	1	2	3	2	1	5	3	1
AAACA[ATG]A[AC]	42	19	29	15	17	27	20	18	18	27
[AG]GGAACTCC	0	0	1	1	0	0	0	0	0	0
GAAA[AC][GC]GCC	2	0	4	2	2	0	3	1	3	3
G[CT]AAACAAA	10	5	5	1	4	3	1	3	2	4
CCC[AT]CCC[AC]C	2	9	3	4	2	5	6	2	5	3
	0.483425984	0.122157213	0.233821717	0.483318212	0.635731105	0.483318212	0.483156553	0.233410518	0.066087051	0.233118734
	0.620730641	0.498391533	0.620843553	0.620843553	0.241352538	0.620843553	0.621013006	0.497489115	0.240898697	0.068115325
	0.000242752	0.290501864	0.030350195	0.522809109	0.092472866	0.163144197	0.162986188	0.945369735	0.091987961	0.030059533
	0.468483651	0.255937008	0.142244611	0.142244611	0.874503777	0.46825558	0.255724498	0.141594467	0.872053015	0.141133889
	0.283750277	0.579357521	0.080667568	0.283659651	0.579357521	0.04470031	0.5516026	0.552401207	0.552401207	0.577159804
	0.577291433	0.148020173	0.577125747	0.148020173	0.577125747	0.577125747	0.553326368	0.575850963	0.575850963	0.574945574
	0.634361473	0.234440164	0.234440164	0.484563995	0.234440164	0.634471016	0.234359277	0.483728856	0.122263622	0.483135861
	0.88280331	0.473992876	0.370616582	0.144380586	0.259508931	0.882481722	0.081235897	0.472207423	0.472207423	0.080641021
	0.288699717	0.94405351	0.519752206	0.94405351	0.288529286	0.519752206	0.943535007	0.16113753	0.343129034	0.160570735
	0.551402978	0.579199031	0.579199031	0.148628929	0.283573669	0.148628929	0.551724719	0.552523167	0.282877368	0.28238334
	0.593278351	0.540981437	0.291165595	0.59310318	0.025837213	0.59310318	0.592840405	0.59175538	0.2904253	0.152066089
	0.430733446	0.208314976	0.695131869	0.108958435	0.430659503	0.208314976	0.43054859	0.695853247	0.108820557	0.429686889
	0.33495943	0.487992427	0.334830148	0.05356437	0.334830148	0.334830148	0.669395917	0.668040129	0.668040129	0.333130913
	0.008935778	0.874716871	0.227010083	0.266287717	0.131688422	0.076524393	0.390919597	0.877791278	0.044263634	0.04404716
	0.577134113	0.28236496	0.553255598	0.553255598	0.553255598	0.080281153	0.282230501	0.57569446	0.281675655	0.281186591
	0.483528374	0.122183732	0.233872488	0.635627394	0.233872488	0.066217336	0.036664336	0.066101282	0.482590984	0.63706854
	0.331764078	0.331637209	0.664304901	0.331637209	0.175155639	0.331637209	0.663981081	0.052817391	0.174607701	0.661463692
	4.86E-08	0.003345039	2.87E-05	0.020506002	0.008365169	7.55E-05	0.002094145	0.005198025	0.005198025	7.25E-05
	0.70388242	0.703973614	0.423801226	0.423801226	0.703973614	0.703973614	0.704110469	0.704676239	0.704676239	0.705176323
	0.290066396	0.542610664	0.082559651	0.289971014	0.289971014	0.542610664	0.152026858	0.589589458	0.151705442	0.151422254
	0.003577378	0.05822502	0.05822502	0.711959975	0.104614148	0.191424373	0.711597649	0.190769306	0.359139883	0.103976973
	0.189957296	0.003174599	0.099596626	0.054122866	0.189935309	0.030038488	0.016905025	0.18976621	0.030013651	0.099450294

Continued (cluster 2B)

motif_seq_rcp	(0,200]	(200,400]	(400,600]	(600,800]	(800,1000]	(1000,1200]	(1200,1400]	(1400,1600]	(1600,1800]	(1800,2000]
TTTGCAC[GA]G	1	3	2	1	0	1	1	2	4	2
GGGACA[TC]G[TG]	0	1	0	0	2	0	0	1	2	4
TTTG[AT][CT]TTG	16	3	7	2	5	4	4	1	5	7
CGCTGC[GT][GT]	2	3	4	4	1	2	3	4	1	4
T[CT]TGACAAC	2	1	4	2	1	5	0	0	0	1
TTTGCAC[GA]A	1	3	1	3	1	1	0	1	1	1
GGGGTT[TG]TT	0	2	2	1	2	0	2	1	3	1
GGT[AG]GCG[CG]	1	2	0	4	3	1	5	2	2	5
CAC[CG][CA]TTTT	3	1	2	1	3	2	1	4	0	4
TGGAA[GT]TTC	0	1	1	3	2	3	0	0	2	2
G[GC]CGT[ACT]TTG	1	0	2	1	6	1	1	1	2	3
CGCGCAG[AC][CT]	1	2	0	3	1	2	1	0	3	1
T[AT]GGGCGT	2	0	2	5	2	2	1	1	1	2
[GT]TT[CGT]CCGG	10	1	4	0	5	6	3	1	7	7
G[CG]GTTCATT	1	2	0	0	0	4	2	1	2	2
CGTAAC[ACG]CT	1	3	2	0	2	4	5	4	1	0
TGCCGT[AT]C	2	2	1	2	3	2	1	5	3	1
[GT]T[CAT]TGTTT	42	19	29	15	17	27	20	18	18	27
GGAGTTCC[CT]	0	0	1	1	0	0	0	0	0	0
GGC[GC][GT]TTTC	2	0	4	2	2	0	3	1	3	3
TTTGTTT[AG]C	10	5	5	1	4	3	1	3	2	4
G[GT]GGG[AT]GGG	2	9	3	4	2	5	6	2	5	3
	0.635621878	0.483318212	0.233821717	0.635731105	0.635731105	0.483318212	0.483156553	0.233410518	0.233410518	0.233118734
	0.620730641	0.620843553	0.021267812	0.241352538	0.241352538	0.620843553	0.498215571	0.621713629	0.125863184	0.496848328
	0.017543129	0.522809109	0.030350195	0.163144197	0.290501864	0.290501864	0.017510395	0.05252588	0.05252588	0.091644667
	0.874822059	0.255937008	0.874503777	0.874503777	0.255937008	0.255937008	0.874026217	0.872053015	0.079669918	0.141133889
	0.551280795	0.579357521	0.283659651	0.579357521	0.551409475	0.579357521	0.5516026	0.578072482	0.578072482	0.148030094
	0.577291433	0.282450078	0.282450078	0.553133784	0.553133784	0.282450078	0.576877202	0.147646675	0.28176028	0.554826915
	0.634361473	0.634471016	0.122480303	0.234440164	0.634471016	0.484563995	0.634635412	0.483728856	0.483728856	0.233731211
	0.370461899	0.882481722	0.473992876	0.259508931	0.144380586	0.473992876	0.473644639	0.258395731	0.880005441	0.878244568
	0.520016577	0.94405351	0.94405351	0.94405351	0.341925362	0.288529286	0.161781487	0.287219608	0.517718988	0.286291166
	0.28366423	0.579199031	0.579199031	0.579199031	0.283573669	0.283573669	0.578948629	0.552523167	0.282877368	0.004680014
	0.291261878	0.291165595	0.59310318	0.540981437	0.59310318	0.291165595	0.082875833	0.541989713	0.59175538	0.289900064
	0.108976367	0.208314976	0.695131869	0.695131869	0.695131869	0.695131869	0.695272375	0.43009072	0.695853247	0.429686889
	0.487851769	0.334830148	0.669724214	0.487992427	0.669724214	0.487992427	0.096357681	0.489076674	0.333836147	0.333130913
	0.003089476	0.076524393	0.131688422	0.670931728	0.131688422	0.131688422	0.02598385	0.389315555	0.044263634	0.666166179
	0.577134113	0.553255598	0.553255598	0.576968521	0.576968521	0.28236496	0.576720117	0.281675655	0.281675655	0.281186591
	0.635518141	0.233872488	0.233872488	0.066217336	0.483420538	0.483420538	0.122141806	0.482590984	0.482590984	0.121816351
	0.331764078	0.331637209	0.664304901	0.664304901	0.664304901	0.664304901	0.095348221	0.662643802	0.095100257	0.493341654
	1.10E-08	0.000122236	0.005301285	0.013134516	0.003345039	0.000122236	0.00031709	0.020145624	1.72E-05	0.000493452
	0.107319462	0.703973614	0.423801226	0.107302753	0.205081847	0.423801226	0.423696993	0.204830676	0.058158027	0.422887199
	0.591101257	0.082559651	0.59092737	0.59092737	0.082559651	0.59092737	0.590666522	0.589589458	0.543616388	0.151422254
	0.018687709	0.05822502	0.711959975	0.360281469	0.05822502	0.191424373	0.058183479	0.710101192	0.190769306	0.057861418
	0.391147902	0.030038488	0.099596626	0.189935309	0.751385461	0.054122866	0.005507042	0.099511037	0.030013651	0.054045136

Table 2.7.6

Results of orientation bias tests from Cluster 1 to Cluster 2B

Table 2.7.6
Cluster 1

motif_seq	fwd	rcp	ori_bias_diff	ori_bias	p.value	chisq_bias
CATA[CG][AT]AAA	69	55	14	1	0.24307723	0
G[CG]GTTATGA	26	23	3	0	0.77511041	0
CAC[ACG]C[AG]CCC	38	27	11	1	0.21486937	0
GAA[AC][CT]GTCA	47	55	8	0	0.4882805	0
[AG]GTCA[AT]GGA	17	14	3	0	0.71944799	0
GCC[AG]CGG[AC]C	10	4	6	0	0.18145457	0
CA[AC]ATGCGG	7	8	1	0	1	0
C[AGT]C[AG]GGAAG	51	3	48	1	1.60E-10	1
GCAG[AC]GCA	15	73	58	1	1.23E-09	1
T[CT]AC[CG]GC	85	72	13	1	0.33826738	0
GGAA[CT]C[AC]CA	39	6	33	1	1.84E-06	1
[ACT]ACGTC[AT]AA	38	46	8	0	0.44503954	0
AA[AC]CAAAAC	35	31	4	0	0.71194213	0
ATTCCAGT[AG]	38	5	33	1	1.06E-06	1
[AC]TG[AG]GCCC	26	26	0	0	1	0

Cluster 6

motif_seq	fwd	rcp	ori_bias_diff	ori_bias	p_value	chisq_bias
CA[AG]ACGTCA	24	24	0	0	1	0
CG[AT]ATCCCA	14	7	7	0	0.19043807	0
CAAAAAT[AGT]G	41	44	3	0	0.82828062	0
GCCA[AT]AAA[CT]	27	26	1	0	1	0
C[AG]CAGAAAA	16	16	0	0	1	0
CGAC[GA]CG[CT]C	8	6	2	0	0.78927346	0

Cluster 4

motif_seq	fwd	rcp	ori_bias_diff	ori_bias	p.value	chisq_bias
AGC[AGT]A[AG]ACC	17	17	0	0	1	0
CTGTT[CG]AAA	23	18	5	0	0.53218749	0
AAA[AT]ATGGG	12	16	4	0	0.57076198	0
GAC[TA][TG]CCC	19	28	9	0	0.24326056	0
CTCCT[GT][AT]GC	18	9	9	0	0.12366588	0
CAG[AT][AC]GCTG	2	37	35	1	5.20E-08	1
GG[AT]AGCAC	14	17	3	0	0.71944829	0
CAAGA[GC]CAC	7	12	5	0	0.35880419	0
[CT]AATAAAAA	46	34	12	1	0.21878609	0
TGCC[AC][AC]AAA	20	21	1	0	1	0
GAGGA[ACT]AAA	19	11	8	0	0.20125383	0
AT[AG]ACTCGA	5	18	13	1	0.01234487	1
CTT[TA]G[CG]CGA	21	11	10	1	0.1116208	0

92

Cluster 7

motif_seq	fwd	rcp	ori_bias_diff	ori_bias	p.value	chisq_bias
GCC[CGT][AC]AGCA	4	22	18	1	0.00085634	1
T[CA]CTGTCGG	49	1	48	1	3.00E-11	1
AAAAGT[GA]A[AG]	59	37	22	1	0.03210029	1
TCAC[ACG]TGAA	12	19	7	0	0.2812108	0
ATT[CA]C[GA]GAC	20	14	6	0	0.39118692	0
AAACCG[AC]AC	19	5	14	1	0.0079646	1
TGG[AT]AGCAA	38	7	31	1	7.75E-06	1
CGCG[CT]G[AT]A	24	14	10	1	0.14430396	0
CT[CG][AGT]CTCTC	19	14	5	0	0.48624811	0
CGC[CTG]C[CG]AA	35	22	13	1	0.11197659	0
AA[CG]GACGAA	9	14	5	0	0.40425885	0

Cluster 5

motif_seq	fwd	rcp	ori_bias_diff	ori_bias	p.value	chisq_bias
ATC[GT]CTCGC	11	6	5	0	0.33198321	0
[AT]CTGACTG	13	26	13	1	0.05467085	0
TTGATTGA	37	31	6	0	0.5443137	0
G[CT]AACGGAA	11	6	5	0	0.33198321	0
AA[CA]TGGTG	31	18	13	1	0.08648757	0
AC[GC]ATGAC	19	17	2	0	0.86764123	0
ACTCA[CG]TC	11	17	6	0	0.3447161	0

Cluster 2A

motif_seq	fwd	rcp	ori_bias_diff	ori_bias	p.value	chisq_bias
CA[AC]G[AG]AA	213	218	5	0	0.84727553	0
G[GC]G[CG]CTAAC	19	3	16	1	0.00138403	1
AAAAG[CG]GCC	18	9	9	0	0.12366559	0
GCCG[GT][AC]GCC	13	5	8	0	0.09896501	0
ACCTCA[ACGT]A[CT]	14	24	10	1	0.14430364	0
AAAA[AC]TCGA	34	19	15	1	0.05448313	0
AT[AGT]CCGAAC	12	14	2	0	0.84452671	0
GCGA[TA]TCC	15	19	4	0	0.6069181	0
CGT[GT][ATC]GGCA	13	30	17	1	0.01469118	1
CG[CT][GT]TGGCA	16	26	10	1	0.1649283	0
CGC[CG]GGC[CT]G	7	8	1	0	1	0
AAAC[AG]CATG	12	11	1	0	1	0
CAGT[GA]TCG	17	16	1	0	1	0
[AG]TGGACCA[AG]	14	19	5	0	0.48624774	0
AAAAT[ACT]GAG	41	31	10	1	0.28887075	0
CC[AT]C[AT]GGGG	6	17	11	1	0.03705947	1

Cluster 2B

motif_seq	fwd	rcp	ori_bias_diff	ori_bias	p.value	chisq_bias
C[TC]GTGCAAA	17	12	5	0	0.45762626	0
[CA]C[GA]TGTCCC	10	15	5	0	0.42372163	0
CAA[AG][AT]CAAA	54	52	2	0	0.92263956	0
[AC][AC]GCAGCG	28	23	5	0	0.57542123	0
GTTGTCA[AG]A	16	10	6	0	0.32681054	0
T[TC]GTGCAAA	13	13	0	0	1	0
AA[CA]AACCCC	14	10	4	0	0.54030147	0
[CG]CGC[CT]ACC	25	19	6	0	0.45099969	0
AAAA[TG][CG]GTG	21	20	1	0	1	0
GAA[AC]TTCCA	14	21	7	0	0.3105085	0
CAA[AGT]ACG[GC]C	18	15	3	0	0.7277343	0
[AG][GT]CTGCGCG	14	7	7	0	0.190438	0
ACGCCC[AT]A	18	12	6	0	0.36132299	0
CCGG[ACG]AA[AC]	44	55	11	1	0.31491446	0
AATGAAC[CG]C	14	12	2	0	0.84452673	0
AG[CGT]GTTACG	22	18	4	0	0.63527034	0
G[AT]ACGGCA	22	17	5	0	0.52185456	0
AAACA[ATG]A[AC]	232	245	13	1	0.58283264	0
[AG]GGAACTCC	2	18	16	1	0.00079637	1
GAAA[AC][GC]GCC	20	17	3	0	0.74231985	0
G[CT]AAACAAA	38	37	1	0	1	0
CCC[AT]CCC[AC]C	41	35	6	0	0.56630531	0

Chapter III.

Construction of gene regulatory networks involved in mosquito reproduction and validation of the functional role of certain components within the networks

3.1 Abstract

In the previous chapter we discussed how cluster-specific overrepresented motifs were detected with the help of motif finding programs and subjected to *in-silico* tests for characteristics related to real transcription factor binding sites. JASPAR identified putative transcription factors (TFs) related to each of these motifs and we searched the Flyatlas2 database to look for fat body specific TFs. So we had a list of cluster-specific motifs that could be linked to tissue specific TFs. Now, we wanted to check if the TFs that we identified were functional and if so then what other factors were these related to.

We started off by building putative gene regulatory networks on the basis of experimental data available from *Drosophila* using the GeneMANIA network-building tool. Once the connections between the different TFs that acted as nodes were established we went on to check if the depletion of one factor affected the expression of another to which it was connected in other words we checked the epistatic relationship between some of the major factors within the network.

Results showed that STAT and Achi are related to activation of EcR in the fat body of a female *Ae. aegypti* at 24 hr PBM. There was also evidence for indirect regulation of certain genes by EcR through E74, which is one of the early genes in the ecdysone regulatory cascade. A new EcR:USP binding site was identified, certain genes

carrying this site in there promoters showed de-repression in EcR depleted tissues suggesting direct negative regulation by EcR. The epistatic relationship between Broad and Mirror transcription factors and evidence for possible co-regulation was also observed. Overall the results displayed the usefulness of the bioinformatics analyses in the previous chapter.

3.2 Introduction

Bioinformatics plays a critical role in analyzing and deciphering the huge amount of data generated by transcriptomics, genomics and proteomics studies (Kanehisa & Bork, 2003). In the post-sequencing era, the role of bioinformatics has complemented traditional biological methods to push the boundaries. Every process of life is dependent upon one or more regulatory networks of genes. It is critically important to understand the dynamics of these networks, as that would help us in understanding the molecular mechanisms behind each of these processes. Accurate prediction of gene regulatory networks (GRNs) can speed up functional analysis and have already been proven to be a valuable research tool (Karlebach and Shamir, 2008).

Since the construction of GRN requires integration of numerous quantitative or high-throughput experimental data, the construction of the GRN is challenging. However, due to the importance of GRNs, there have been considerable efforts towards establishing robust and reliable methods for the prediction and modeling of GRNs in the field of "System biology" or "Network biology".

In this section we have used GeneMANIA (Warde-Farley et al., 2010), which is a flexible and user-friendly web interface that generates putative GRNs that are useful for "generating hypotheses about gene function and prioritizing genes for functional assays" (Warde-Farley et al., 2010). GeneMANIA uses available genomics and proteomics data to identify functionally similar genes and extend a list of queried genes (Warde-Farley et al., 2010).

We had a list of cluster-specific, overrepresented motifs from all the bioinformatics analyses described in the previous chapter. We queried the JASPAR database with these putative TFBSs to identify the TFs that might bind to these sites. We also looked for and selected those factors that are expressed in the fat body of *Drosophila*, to come up with a list of TFs for cluster-specific motifs. We used these lists of TFs as an input for the GeneMANIA web interface to come up with probable GRN for each cluster. This helped us in the process of selecting genes for functional analysis, the process of gene selection is described in details in the results section.

Once we were able to select the genes that we would like to test for their functions we used two different molecular biology techniques for the analyses described in the results section, viz. depletion of genes by RNA-interference (RNAi) mechanism and quantitative reverse transcription polymerase chain reaction (RT-qPCR). RNA interference (RNAi) is a regulatory mechanism in which small double-stranded RNA (dsRNA) molecules inhibit gene expression, typically by destroying specific mRNA molecules (Kim and Rossi, 2008). Loss-of-function studies commonly use RNAi virtually in all eukaryotic organisms. We used this technique to deplete the TFs that we selected after the bioinformatics analyses.

The real-time, fluorescence-based reverse transcription polymerase chain reaction (RT-qPCR) is one of the most commonly used mRNA detection techniques (Bustin 2000). RT-qPCR is method in which the RNA obtained either from total RNA or messenger RNA (mRNA) is first transcribed into complementary DNA (cDNA) which

then is then used as the template for the qPCR reaction. We used RT-qPCR to analyze the

effects of RNAi depletion of certain factors on other factors and certain target genes.

3.3 Methods and Materials

3.3.1 Construction of regulatory networks

The GRNs were constructed using the GeneMANIA prediction server
(www.genemania.org) for biological network integration and predicting gene function
(Warde-Farley et al., 2010). *Drosophila melanogaster* transcription factor information in
the server was used to construct networks with transcription factors identified from
cluster-specific motifs. The GeneMANIA prediction tool generates network topology
based on information already available from published datasets. The input data were the
transcription factors identified from each cluster.

3.3.2 Functional validation of network components:

3.3.2.1 Experimental animals

Aedes aegypti Ugal strain mosquitoes were reared at 27ºC and 80% humidity as
described previously. The female *Aedes aegypti* used after hatching 24h (24PE) and
dissected with 1% PBS buffer solution. White Leghorn chickens were used for blood
feeding. All procedures were followed University of California, Riverside Animal care
protocol.

3.3.2.2 PCR and dsRNA synthesis

PCR performed using Platinum Supermix (Invitrogen) for dsRNA synthesis.
Primer sequences were designed by using Vectorbase sequence information.
dsRNA was synthesized by using Megascript T7 kit (Ambion).

3.3.2.3 Microinjection to mosquito

0.3μl of about 4 μg/μl dsRNA was injected to each the anesthetized female

mosquitoes for RNAi-depletion.

3.3.2.4 Total RNA extraction and cDNA synthesis

RNA was extracted from the dissected 7-10 female mosquito fat bodies using

TRIzol (Invitrogen) following manufacturer's protocol. The cDNA was synthesized with

~2 μg of extracted total RNA by using SuperScript III Reverse Transcriptase kit

(Invitrogen) with DNase I (Invitrogen) treatment.

3.3.2.5 qPCR analysis

qPCR was conducted using iCycler iQ System (Bio-Rad) and IQ SYBR Green

Supermix (Bio-Rad) following manufacturer's protocol. The housekeeping gene, S7

ribosomal protein (RPS7) was used as reference gene for normalization. Experiments

were done with triplicate samples and relative expression (RE) to RPS7 was checked for

the analysis. The RE was calculated as $RE=2^{-\Delta\Delta Ct}$ in Microsoft Office Excel. Standard

deviation was calculated in Microsoft Office Excel.

3.3.2.6 Microscopy

The ovaries were visualized using the Leica M165FC stereomicroscope. The

Zeiss Axio Observer.A1 microscope was used to measure the follicle size of eggs. Scale

bar is 200 μm.

3.3.2.7 Statistical significance test

P-values were computed with unpaired t-test using the online version of GraphPad (http://www.graphpad.com/quickcalcs/ttest1.cfm).

3.4 Results

3.4.1 Construction of Gene Regulatory Networks (GRNs):

To understand the possible regulatory mechanism of the genes at a particular time point we used the TFs identified by JASPAR to create putative regulatory networks for each of the 7 clusters using GeneMANIA (Warde-Farley, 2010).

First we created networks for the two early gene clusters – Cluster 1 and Cluster 6. We used the 15 TFs predicted for the cluster 1 motifs to build a network. All of the nodes were connected to one or more different nodes, as a result there were no orphan nodes detected. Of the interactions shown, 69.24% were predicted, 15.55% were based on co-expression, 13.60% were based on physical (8.04%) and genetic interactions (5.56%) and only 1.62% was based on co-localization (Figure 3.7.1, Table 3.7.1).

Network construction with five TFs detected by JASPAR on the basis of cluster-specific putative TFBSs for Cluster 6 resulted in a network where all nodes other than that for Dichaete (D), were shown to be connected to Extradenticle (Exd) and Achintya (Achi). The network was built on the basis of 28.57% of co-expression data, 27.11% predicted interactions, 30.58% genetic (19.66%) and physical (10.92%) interactions and 1.93% based on shared protein domains (Figure 3.7.2, Table 3.7.1).

Next we used the TFs identified from the early-mid gene clusters. Cluster 4 had 14 nodes, all of which were connected to one or more nodes. Co-expression data (51.35%), was the major source for construction of this network. Physical interactions constituted 18.86% of the data whereas Genetic interactions with 14.66% followed

closely. 12.21% of the network construction was based on information available from shared protein domains and 2.89% from Co-localization data (Figure 3.7.3, Table 3.7.1).

Cluster 7 comprising of 13 nodes was mainly built around predicted interactions (51.48%). Co-expression (23.09%), Physical (12.24%) and Genetic (9.61%) interactions followed with Co-localization data contributing 3.58% (Figure 3.7.4, Table 3.7.1).

Network for the late-mid cluster - Cluster 5, with 7 nodes, was built mainly upon known Genetic (40.97%) and physical (33%) interactions. Shared protein domain (9.8%), Co-expression (9.61%) and Co-localization (6.62%) were other contributing factors (Figure 3.7.5, Table 3.7.1).

Co-expression was (49.59%) was the major source of information for the regulatory network constructed with constructed with 15 TFs for cluster 2A, one of the two late gene clusters. Predicted interactions at 35.57% were the major source of information. Physical (5.16%) and Genetic (4.97%) interactions constituted approximately 5% each, contribution from Co-localization (2.78%) and Shared protein domains (1.93%) were even lower (Figure 3.7.6, Table 3.7.1).

Cluster 2B regulatory network consisted of 16 nodes was very similar to 2A in terms of the information used for construction. Information from Co-expression was a little lower at 40.32% and that from Genetic interactions (8.64%) was a bit higher (Figure 3.7.7, Table 3.7.1). In case of both late gene clusters no orphan nodes could be detected.

3.4.2   Functional validation of the effects of the transcription factors involved within the regulatory networks:

3.4.2.1 Functional validation of the role of a subset of transcription factors in Cluster1:

In order to study if the bioinformatics analyses are helpful towards deciphering the role of the transcription factors, identified within a cluster, we first selected a subset of Cluster1 TFs for functional validation. As mentioned previously, searching of JASPAR insecta database with Cluster1 motifs resulted in identification of 15 different transcription factors that are expressed in the fat body of an *Ae. aegypti* (Figure 2.7.1). We went back to the bioinformatics data and found that six out of these fifteen motifs passed two of the three bioinformatics tests viz. positional bias, orientation bias and evolutionary conservation. JASPAR search suggested that these were putative TFBSs for the following TFs: Achi, Eip74EF (E74), Trl, Exd, BrZ1 (Br) and Stat92E(STAT) (Figure 2.7.1).

Next we checked how these transcription factors are linked to EcR within the regulatory network, built GeneMANIA, which as depends on experimental data available from *Drosophila*. EcR as mentioned previously along USP forms a heterodimer that is the functional ecdysteroid receptor capable of binding the 20-Hydroxyecdysone (20E), hormone. 20E is known to be the major regulator of PBM events in the fat body, which produces yolk protein precursors (YPPs) for subsequent egg development (Roy et al., 2015; Raikhel, 2005, Hagedorn, 2005). Since our data is from the PBM fat body transcriptome, EcR surely is one of the major factors and we wanted to look at what other factors are effecting gene regulation in conjunction with EcR.

Of the six TFs mentioned above Stat92E, Achi and Broad displayed direct connections to EcR within the network, whereas Trithorax-like (Trl) and Exd did not show any direct connection (Figure 3.7.1). Eip74EF (E74) did not show any direct connection but was linked to EcR through Broad, also it is a well known that E74 is one of the downstream genes in the ecdysone regulatory cascade. Two other transcription factors viz. Mothers against dpp (Mad) and Nubbin (Nub) were linked directly to EcR within the network but the putative TFBSs for these factors did not test positive for any of the bioinformatics tests (Figure 2.7.1) as a result we did not have much confidence on these factors and did not pursue with these factors any further. We decided knockdown Stat92E (STAT), Broad (Br), Achi and E74 along with EcR to check epistatic relationships, if any, among these factors. We also checked the effects of these knockdowns on genes targeted by E74.

3.4.2.1.1 STAT activates EcR while EcR negatively regulates STAT at the transcriptional level:

To test if there is any epistatic relationship between EcR and Stat both of these factors were depleted (Figure 3.7.8.1, Figure 3.7.8.2) by using RNAi and the effect of the depletion of one factor on the other was checked at the transcriptomic level by quantitative real-time polymerase chain reactions (qRT-PCR). The depletion of STAT resulted in a decrease of the EcR transcripts suggesting that EcR is activated by STAT within the fat body transcriptome (Figure 3.7.9.1). On the other hand depletion of EcR

resulted in an increase of STAT transcripts suggesting that EcR negatively regulates the transcription of STAT (Figure 3.7.9.2).

3.4.2.1.2 Achi activates EcR at the transcriptional level:

Another factor that was found to have direct connection with EcR within the network is Achi, therefore, we tested the epistatic relationship between EcR and Achi in the same way we checked the relationship between Stat and EcR. Achi was depleted with the help of RNAi (Figure 3.7.10), the level of the EcR transcripts were significantly down after Achi depletion suggesting that Achi has an activating effect on EcR (Figure 3.7.11.1). On the other hand we could not detect any significant change in the level of Achi transcripts post EcR knockdown, suggesting that EcR has no effect on Achi at the transcriptional level (Figure 3.7.11.2).

3.4.2.1.3 EcR activates downstream factors Br and E74:

Broad (Br) is known to play important roles not only in insect metamorphosis and pupal stage specificity (Bayer et al., 2003; Zhou and Riddiford, 2002) but also in adult female mosquito reproduction (Chen et al., 2004; Zhu et al., 2007). Expression of the 20E regulatory hierarchy early genes like E74, is dependent upon a family of C2H2-type, zinc-finger DNA-binding transcription factors, that Br encodes (Karim et al., 1993).

E74 belongs to the ETS family of transcription factors and is involved in controlling stage- and tissue-specific expression of 20E- regulated genes in Drosophila (Fletcher and Thummel, 1995). The E74 gene is located within the E74EF puff (Fletcher

et al., 1995) and in *Ae. aegypti*, is known to play important roles at different time points during the PBM phase (Sun et al., 2002; 2005).

We checked the effects of RNAi depletion of EcR (Figure 3.7.8.1) on the Br and E74 transcripts with the help of qRT-PCR. Significant decline in the transcript levels of both Br and E74 confirmed the activation of these factors within the 24 hr PBM fat body tissue (Figure 3.7.12.1, Figure 3.7.12.2). Next, we wanted to check the effects of these depletions on some target genes, which actually carried one of the putative TFBSs for one of the TFs, in their promoters.

3.4.2.1.4 Effects of Transcription factor knock downs on E74 target genes:

We selected three E74 target genes, viz. Lipase (AAEL008222), Sideroflexin 1,2,3 (AAEL014526) and Ankyrin 2,3/unc44 (AAEL014742) that carries the Eip74EF motifs within their promoters between 1850 and 2000 bases upstream of the translation start site. This motif showed a positional bias for the region between 1800 and 2000 bases upstream of ATG, in the bioinformatics analysis. It should be mentioned that all of these are early genes, not in terms of the ecdysone regulatory cascade, but those that are activated early within the post blood meal fat body transcriptome. These genes are up-regulated between 6 hr and 12 hr PBM then starts to go down reaching their lowest levels between 18 hr and 24 hr before showing some increase in their expression between 48 hr and 72 hr. Therefore, between 18 and 24 hr PBM, during the peak of 20E, when both Br and E74 (early genes in the ecdysone regulatory cascade) show high expressions, the expression levels of these genes are at their lowest. Therefore we wanted to check if E74,

108

the binding sites for which were detected in the promoters of these genes, is responsible for negative regulation of these genes.

First we checked the effects of E74 knockdown (Figure 3.7.13) on these target genes. As hypothesized there was significant increase in the level of transcripts for all three genes (Figures 3.7.14.1 – 3.7.14.3), confirming that these genes are indeed being negatively regulated by E74.

Br was shown to be the intermediary link between E74 and EcR, therefore we wanted to check the effects of Br knockdown (Figure 3.7.15) on E74. We found that Br has an activating effect on E74 (Figure 3.7.16) since the E74 transcripts showed a decline within the Br depleted tissues. Now if Br activates E74 and E74 negatively regulates the target genes then a knockdown of Br should result in a decrease of E74 levels and as a result an increase in the level of transcripts of the target genes. The results confirmed that the target genes are de-repressed within the Br depleted 24 hr PBM fat body tissues (Figures 3.7.17.1 – 3.7.17.3).

Similarly, if EcR is responsible for activation of E74 either directly or through Br, then depletion of EcR should also have the same effects as that of Br depletion on the E74 target genes. We checked the E74 target genes on EcR depleted tissue and the results showed that the transcript levels were up (Figures 3.7.18.1 – 3.7.18.3), confirming that depletion of EcR results in a decrease of the E74 levels which in turn results in de-repression of its target genes.

Our results suggested that both STAT and Achi could activate EcR (Figures 3.7.9.1, 3.7.10); therefore, we wanted to check if these factors were involved in this

negative regulatory pathway. We found that STAT knockdown resulted in a reduction of E74 transcripts (Figure 3.7.19.1). While E74 knockdown did not affect STAT at the transcriptomic level (Figure 3.7.19.2). The effects of STAT knockdown on the target genes showed that these genes are de-repressed within the STAT depleted tissues (Figures 3.7.20.1-3.7.20.3), adding further proof that STAT might in someway be related to this 20E-EcR negative regulatory cascade. We could not detect any effects of Achi depletion on the E74 target genes.

Therefore, the above section provides important information regarding how EcR might negatively regulate certain target genes indirectly and what other factors might be involved in this negative regulatory cascade. Next we wanted to check if we could gather some evidence regarding direct negative regulation by EcR.

3.4.2.2 Identification of a novel binding factor for EcR:USP within the promoters of genes that are not activated at the peak of the 20E titer:

Late-mid or Cluster 5 genes are those, which shows low level of expression prior to 30 h PBM followed by a sudden increase between 36–48 h PBM, and a sharp decline post 48 h (Roy et al. 2015). These genes have very low expression levels at the peak of 20E titer between 18 h and 24 h PBM. Only seven cluster-specific overrepresented motifs could be identified within the upstream regions of the genes within this cluster. One of these motifs TTGATTGA, that is present within the promoters of 53 out of the 199 transcripts checked, was identified to be a putative target for EcR:USP by JASPAR. Sequences with inverted repeat of TTGA (5'-CGTTGAATCAATG-3') and that with

TTGA half sites (5'-GG<u>TTGA</u>ATGAATT-3', 5'-ATTTCT<u>TTGA</u>ATT-3') have previously been identified as binding sites for EcR:USP (Gauhar et al., Genome Research, 2009). We selected three genes that have the TTGATTGA motif within 1kb upstream regions, based on their expression profiles. These were AAEL002658 – amp dependent ligase (motif 998 bases upstream of the translation start site), AAEL006568 – serine protease (motif 315 bases upstream of the translation start site) and AAEL012037 – sulphate transporter (motif 242 bases upstream of the translation start site). We checked the effects of EcR depletion on these genes by qRT-PCR, and all three transcripts showed significant increase (Figures 3.7.21.1-3.7.21.3) when compared to the control (iluc). This suggested that they were de-repressed with the depletion of EcR within the fat body at 24 h PBM. Therefore, the TTGATTGA binding site might allow EcR:USP to bind to the promoters of these genes and might repress these genes directly or by employing other factors to build a repression complex.

3.4.2.3 Broad-Mirror epistasis and possible co-regulation within cluster 2B:

The expression of the Late or Cluster 2B genes decline after a blood meal, and are maintained at low levels during most of the vitellogenic period. An increase in their expression is observed between 48 h and 72 h (Roy et al., 2015). These genes show low levels of expression during the peak of 20E, but when the representatives from this group were checked in 20E depleted tissues at 24 h PBM, no de-repression could be detected (Roy et al., 2015). These genes were found to be activated by Juvenile Hormone (JH) and its receptor Methoprene tolerant (Met) (Roy et al., 2015). Therefore, we wanted to check if the low levels of expression at 24 h PBM for these genes was due to the lack of an

activator or there was an involvement of some repressor. Looking at the heatmap (Figure 2.7.1) described in the previous chapter we observed that the putative TFBSs for Br-Z1, Br-Z4, Eip74EF and Mirror (Mirr) were the only ones to pass two out of the three bioinformatics tests. Looking at the regulatory network we could see that Br and Mirr were linked (Figure 3.7.7) through Cut, which did not pass a single bioinformatics test. Also, there have been multiple studies suggesting interaction between Mirr and Br (Atkey et al., 2006; Boisclair Lachance et al., 2009; Fuchs et al., 2012). Two of these studies show that Mirr acts as an activator of Br whereas the other study (Fuchs et al., 2001) describes how Mirr can recognize two different regulatory elements within the Br promoter and can either activate or repress the Br gene (Fuchs et al., 2012).

Mirr is an Iroquois (Iro) transcription factor, with a highly conserved family of DNA-binding homeodomain proteins. Iro proteins can function as transcriptional repressors (Bilioni et al., 2005). Mirr has also been implicated to link signaling pathways in *Drosophila* oogenesis (Zhao et al. 2000). Therefore, we wanted to check the role Mirr in this cluster, we also wanted to check if Mirr had any effects on Br expression. Mirr and Br are both are highly expressed at 24 h PBM.

Depletion of Mirr with RNAi (Figure 3.7.22.1) resulted in significant reduction of Br transcripts (Figure 3.7.22.2) suggesting that Mirr activates Br within the fat body at 24 hr PBM. We also checked if Br had any effects on Mirr expression, the results suggested that the depletion of Br did not affect Mirr at the transcriptional level (Figure 3.7.22.3).

Next we wanted to check if RNAi depletion of these two factors resulted in phenotypic changes by looking at the size of the ovaries and the ovarian follicles. It was

observed that while ovarian development was effected with the depletion of both Br and Mirr, Mirr knockdown had a more pronounced effect (Figure 3.7.23.1) when compared to the control (iluc). The size of the ovarian follicles was also more reduced in case of Mirr depletion (Figures 3.7.23.2, 3.7.23.3).

We selected two Cluster 2B genes, which had both BrZ4 and Mirr sites within the 1.5 Kb upstream regions and checked for the effects of Br and Mirr depletion on these genes. These were AAEL008886 – conserved hypothetical, carrying the Br-Z4 and Mirr motifs at 974 and 1231 bases upstream of ATG, and AAEL011651 – l-caldesmon, putative, carrying the Br-Z4 and Mirr motifs at 140 and 772 bases upstream of ATG.

It was observed that the transcripts of these genes went up with the depletion of either Br (Figure 3.7.24.1) or Mirr (Figure 3.7.24.2), suggesting that Mirr and Br, are either by themselves, directly involved in negative regulation of these genes, or these factors co-regulate the expression of these genes negatively within the fat body at 24 h PBM.

3.5 Discussion

Our results suggest that STAT activates EcR and in turn affects the downstream genes in the EcR regulatory cascade. Previous studies have suggested that same factors interact with both JAK/STAT and Ecdysone pathways (Kugler et al., 2011). Jang et al., 2009 should that the *Abrupt* gene is a repressor of the Ecdysone signaling whereas it itself is down-regulated in response to the JAK/STAT pathway during border cell migration in *Drosophila* ovary (Jang et al., 2009). Therefore, the JAK/STAT pathway has already been implicated of indirectly activating the Edysone signaling by repressing a gene the product of which can bind to *Taiman* (Tai) a co-activator, and thereby repress ecdysone signaling.

On the other hand *Achi* is one of the two tightly linked and nearly identical homeobox genes of the TGIF (TG-interacting factor) subclass essential for spermatogenesis in *Drosophila* (Wang and Mann, 2003). *Achi* has not been implicated to have a role in female mosquito reproduction.

Our results suggest that Br activates E74, which again is new knowledge in terms female mosquito reproduction. However, similar function for Br has been suggested in case of Drosophila development. It has been reported that prior expression of Br proteins is required for the induction of the early primary-response genes and the ecdysone-receptor protein complex by itself is not sufficient for the activation (Karim et al., 1993).

Similarly, one of the E74 isoforms – E74B, has been shown to be involved in partial repression of downstream genes in the ecdysone regulatory cascade (Fletcher et al.

1997). Therefore, our results, which suggest that E74 represses certain target genes, is in sync with the above-mentioned report.

Therefore, based on the results we can hypothesize a possible negative regulatory pathway where Ecdysone/EcR/USP would activate E74 either directly or through Br, and E74 in turn would identify the recognition sites in the promoters of certain target genes to repress those directly or by organizing a repression complex (Figure 3.7.25). Further work is needed for experimental validation of this proposed pathway.

The novel biding site identified for EcR:USP, and the presence of this site in a big number of genes along with its evolutionary conservation, suggest that this might be an authentic binding site. The depletion of EcR resulted in an increase of the transcript levels of the few genes tested at 24 PBM, which carries this element in their promoters. Further experiments like electrophorectic mobility shift assay (EMSA) is required to prove that EcR recognizes the "TTGATTGA" site. Also, if mutation of the site results in the loss of binding activity, then that would add further evidence to the hypothesis that EcR might directly bind to the promoters of these genes in order to negatively regulate their expression.

In case of the co-repression of representative late genes by Br and Mirr it was observed that the binding sites of BrZ4 (the binding site of this isoform was detected in the promoters of the genes) of Mirr are quite similar except for the 5' and 3' ends. Now, it is possible that Br and Mirr can recognize the same sites and this can be tested by gel super-shift assays using the antibodies specific to these factors. But the activation of Br by Mirr seems to be in congruence with the previous studies related to these two factors

(Atkey et al., 2006; Boisclair Lachance et al., 2009; Fuchs et al., 2012). Negative

regulation by Br has previously been reported, it would be interested to test if this

pathway is dependent/linked to the Ecdysone pathway or functions independent of it.

3.6 References

Kanehisa M, Bork P, Bioinformatics in the post-sequence era. Nat Genet. 2003 Mar;33 Suppl:305-10.

Roy S, Saha TT, Johnson L, Zhao B, Ha J, White KP, Girke T, Zou Z, & Raikhel AS, Regulation of Gene Expression Patterns in Mosquito Reproduction   PLoS Genetics 2015 11(8):e1005450

Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010 Jul;38(Web Server issue):W214-20. doi: 10.1093/nar/gkq537.

Karlebach G, Shamir R, Modelling and analysis of gene regulatory networks. Nat Rev Mol Cell Biol. 2008 Oct;9(10):770-80. doi: 10.1038/nrm2503. Epub 2008 Sep 17.

Kim D, Rossi J. RNAi mechanisms and applications. Biotechniques. 2008 Apr;44(5):613-6. doi: 10.2144/000112792.

Bayer C, Zhou X, Zhou B, Riddiford LM, von Kalm L, 2003. Evolution of the Drosophila broad locus: the Manduca sexta broad Z4 isoform has biological activity in Drosophila. Dev. Genes Evol. 213, 471–476.

Zhou XF, Riddiford LM, 2002. Broad specifies pupal development and mediates the "status quo" action of juvenile hormone on the pupal-adult transformation in Drosophila and Manduca. Development 129, 2259–2269.

Karim F, Guild G, Thummel C, 1993. The drosophila broad-complex plays a key role in controlling ecdysone-regulated gene-expression at the onset of metamorphosis. Development 118, 977–988.

Chen L, Zhu J, Sun G, Raikhel AS, 2004. The early gene Broad is involved in the ecdysteroid hierarchy governing vitellogenesis of the mosquito Aedes aegypti. J. Mol. Endocrinol. 33, 743–761.

Zhu J, Chen L, Raikhel AS, 2007. Distinct roles of Broad isoforms in regulation of the 20-hydroxyecdysone effector gene, Vitellogenin, in the mosquito Aedes aegypti. Mol. Cell. Endocrinol. 267, 97–105.

Sun G, Zhu J, Li C, Tu Z, Raikhel AS, 2002. Two isoforms of the early E74 gene, an Ets transcription factor homologue, are implicated in the ecdysteroid hierarchy governing

vitellogenesis of the mosquito, Aedes aegypti. Mol. Cell. Endocrinol. 190, 147–157.

Sun G, Zhu J, Raikhel AS, 2004. The early gene E74B isoform is a transcriptional activator of the ecdysteroid regulatory hierarchy in mosquito vitellogenesis. Mol. Cell. Endocrinol. 218, 95–105.

Sun G, Zhu J, Chen L, Raikhel AS, 2005. Synergistic action of E74B and ecdysteroid receptor in activating a 20-hydroxyecdysone effector gene. Proc. Natl. Acad. Sci. U.S.A. 102, 15506–15511.

Fletcher J, Thummel C, 1995. The Drosophila E74 gene is required for the proper stage-specific and tissue-specific transcription of ecdysone-regulated genes at the onset of metamorphosis. Development 121, 1411–1421.

Fletcher J, Burtis K, Hogness D, Thummel C, 1995. The Drosophila E74 gene is required for metamorphosis and plays a role in the polytene chromosome puffing response to ecdysone. Development 121, 1455–1465.

Atkey MR, Lachance JF, Walczak M, Rebello T, Nilson LA. Capicua regulates follicle cell fate in the Drosophila ovary through repression of mirror. Development. 2006 Jun;133(11):2115-23. Epub 2006 May 3.

Boisclair Lachance JF, Fregoso Lomas M, Eleiche A, Bouchard Kerr P, Nilson LA. Graded Egfr activity patterns the Drosophila eggshell independently of autocrine feedback. Development. 2009 Sep;136(17):2893-902. doi: 10.1242/dev.036103. Epub 2009 Jul 29.

Fuchs G, Diges C, Kohlstaedt LA, Wehner KA, Sarnow P. Proteomic analysis of ribosomes: translational control of mRNA populations by glycogen synthase GYS1. J Mol Biol. 2011 Jul 1;410(1):118-30. doi: 10.1016/j.jmb.2011.04.064. Epub 2011 May 5.

Fuchs A, Cheung LS, Charbonnier E, Shvartsman SY, Pyrowolakis G. Transcriptional interpretation of the EGF receptor signaling gradient. Proc Natl Acad Sci U S A. 2012 Jan 31;109(5):1572-7. doi: 10.1073/pnas.1115190109. Epub 2012 Jan 17.

Bilioni A, Craig G, Hill C, McNeill H. Iroquois transcription factors recognize a unique motif to mediate transcriptional repression in vivo. Proc Natl Acad Sci U S A. 2005 Oct 11;102(41):14671-6. Epub 2005 Oct 3.

Zhao D, Clyde D, Bownes M. Expression of fringe is down regulated by Gurken/Epidermal Growth Factor Receptor signalling and is required for the morphogenesis of ovarian follicle cells. J Cell Sci. 2000 Nov;113 Pt 21:3781-94.

Kugler SJ, Gehring EM, Wallkamm V, Krüger V, Nagel AC., The Putzig-NURF nucleosome remodeling complex is required for ecdysone receptor signaling and innate immunity in Drosophila melanogaster. Genetics. 2011 May;188(1):127-39. doi: 10.1534/genetics.111.127795. Epub 2011 Mar 8.

Jang AC, Chang YC, Bai J, Montell D. Border-cell migration requires integration of spatial and temporal signals by the BTB protein Abrupt. Nat Cell Biol. 2009 May;11(5):569-79. doi: 10.1038/ncb1863. Epub 2009 Apr 6.

Wang Z, Mann RS. Requirement for two nearly identical TGIF-related homeobox genes in Drosophila spermatogenesis. Development. 2003 Jul;130(13):2853-65.

Gauhar Z, Sun LV, Hua S, Mason CE, Fuchs F, Li TR, Boutros M, White KP. Genomic mapping of binding regions for the Ecdysone receptor protein complex. Genome Res. 2009 Jun;19(6):1006-13. doi: 10.1101/gr.081349.108. Epub 2009 Feb 23.

3.7 Figures and tables

Figure 3.7.1 Regulatory network for Cluster 1 transcription factors:

The regulatory network consists of 15 transcription factors identified by JASPAR, based on the overrepresented and cluster-specific motifs or putative transcription factor binding sites, detected by bioinformatics analyses. The node, FBgn0001168, represents the hairy (h) gene, for which GeneMANIA recognizes only the flybase id.

Figure 3.7.1   Regulatory network for Cluster 1 transcription factors
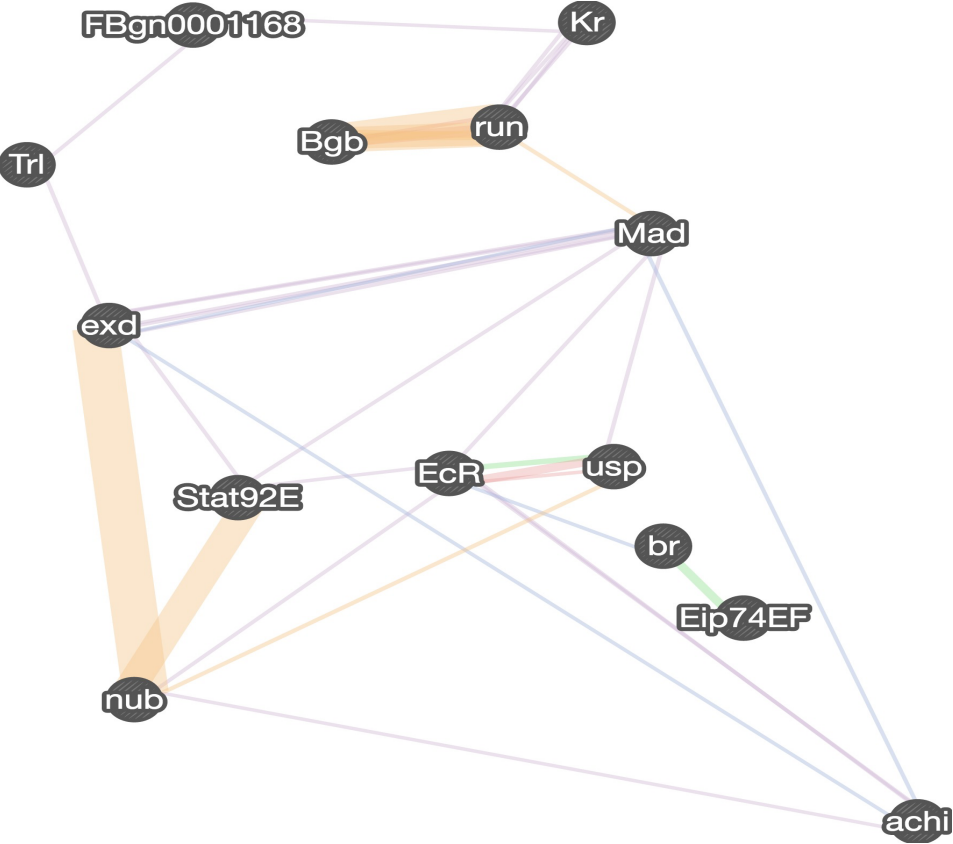
Figure 3.7.2 Regulatory network for cluster 6 transcription factors: Five nodes of

transcription factor genes were drawn for the network. FBgn0000411 is Dichaete (D)

transcription factor.

Figure 3.7.2   Regulatory network for Cluster 6 transcription factors
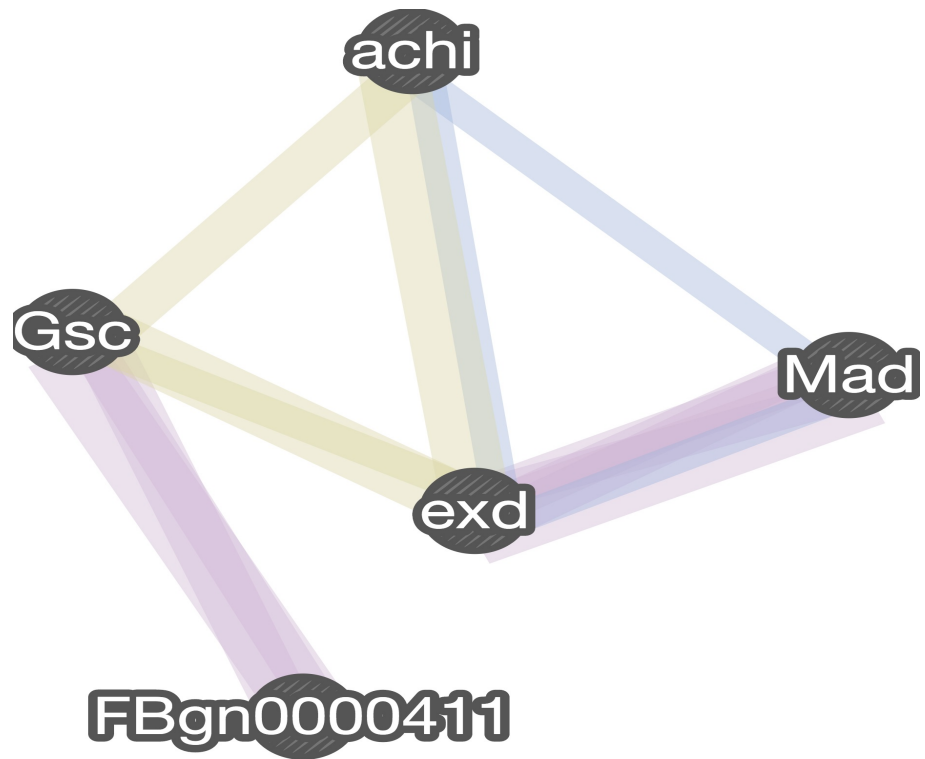
Figure 3.7.3 Regulatory network for cluster 4 transcription factors: Fourteen nodes of

transcription factor were suggested as a regulatory network using GeneMANIA. All

nodes were connected and there is no orphan node.

Figure 3.7.3   Regulatory network for Cluster 4 transcription factors
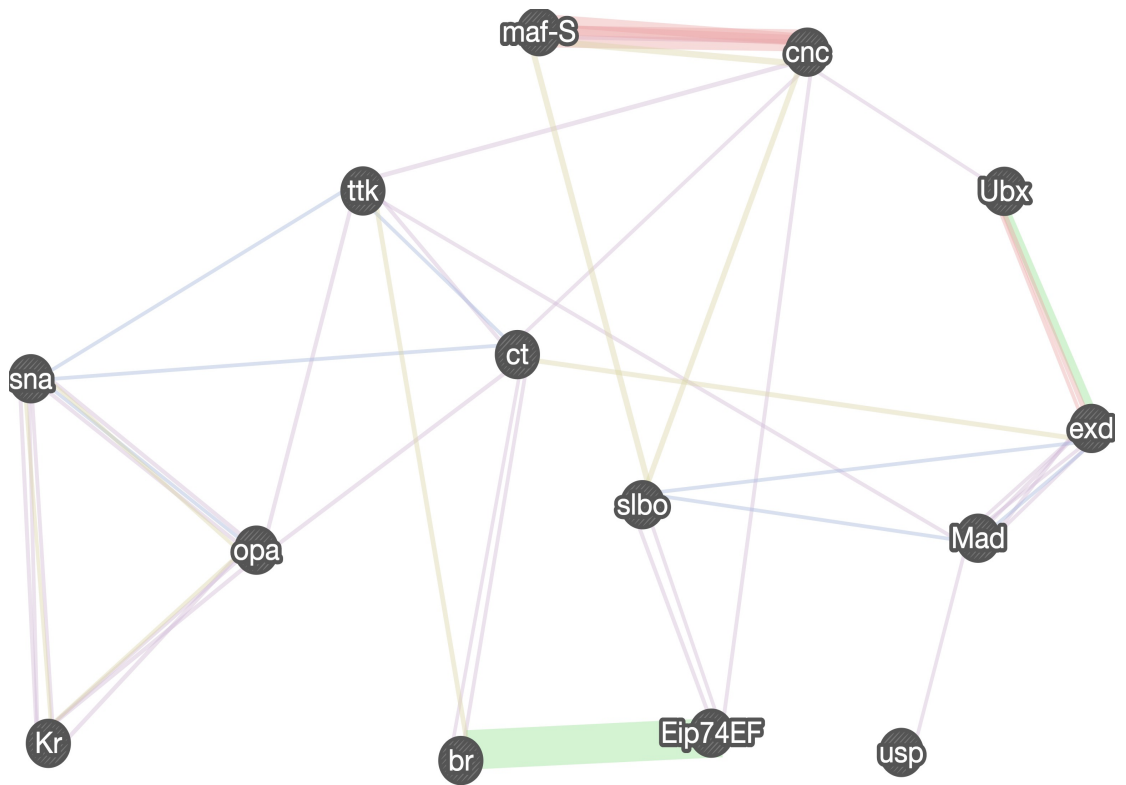
Figure 3.7.4 Regulatory network for cluster 7 transcription factors: Thirteen nodes of

transcription factor gene were shown. The schlank is identical to Lag1.

Figure 3.7.4   Regulatory network for Cluster 7 transcription factors
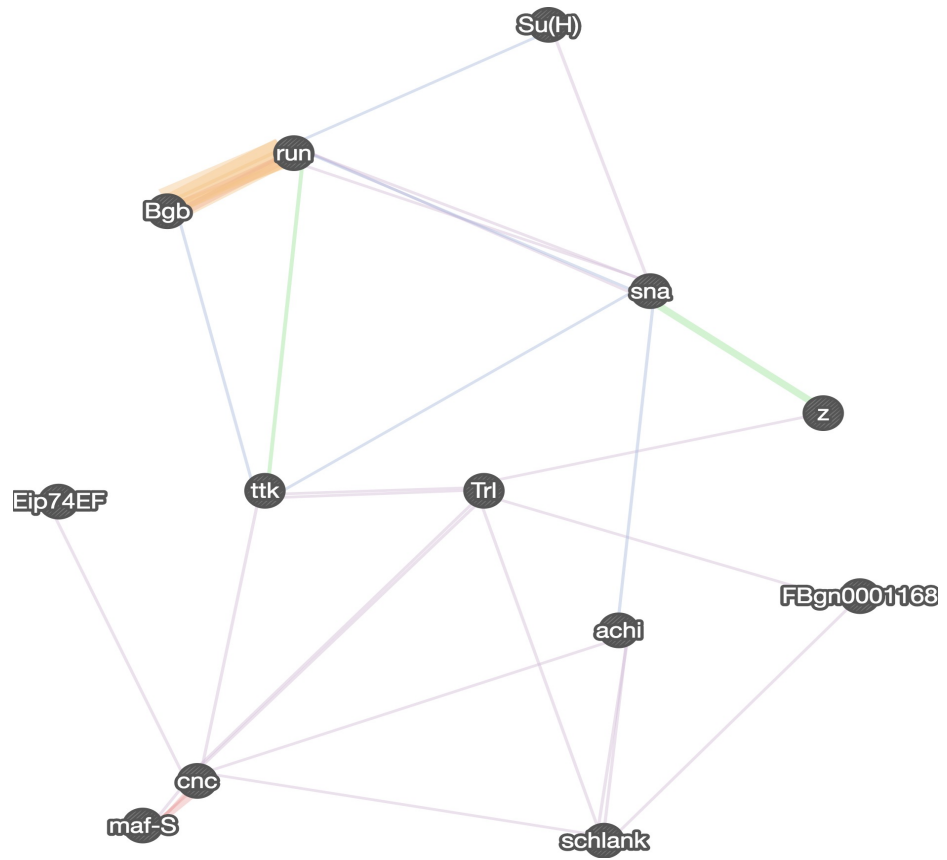
Figure 3.7.5 Regulatory network for cluster 5 transcription factors: Regulatory network of cluster 5 was consisted of seven nodes of transcription factor. EcR::usp connected Br with Eip74EF node.

Figure 3.7.5   Regulatory network for Cluster 5 transcription factors
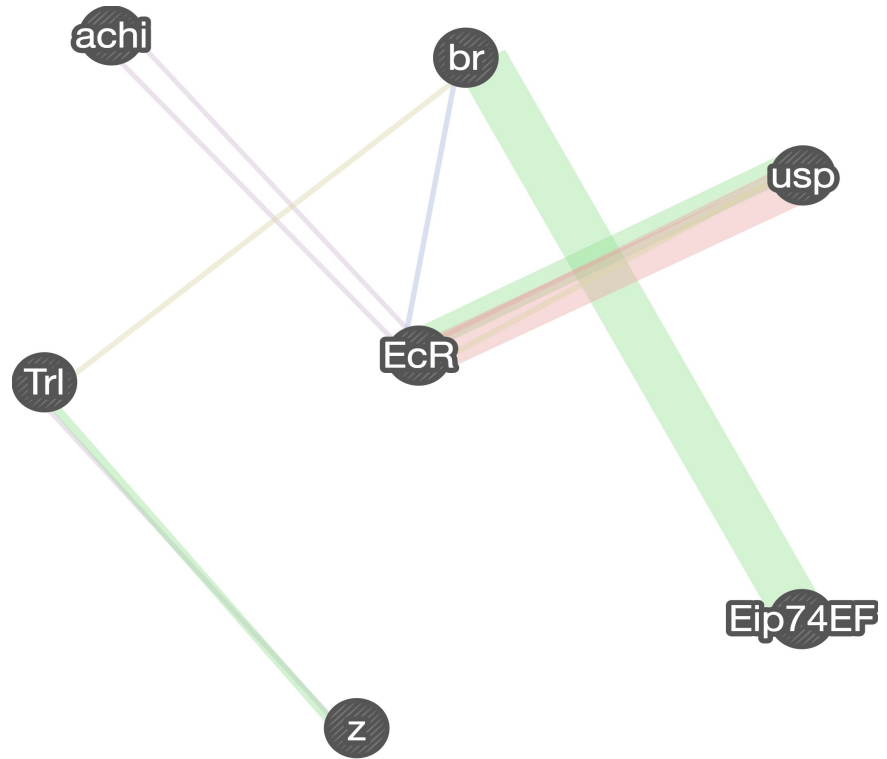
Figure 3.7.6 Regulatory network for Cluster 2a transcription factors: Fifteen nodes of transcription factor gene were illustrated. The genetic interaction between Br and E74 was detected in this cluster as well.

Figure 3.7.6   Regulatory network for Cluster 2a transcription factors
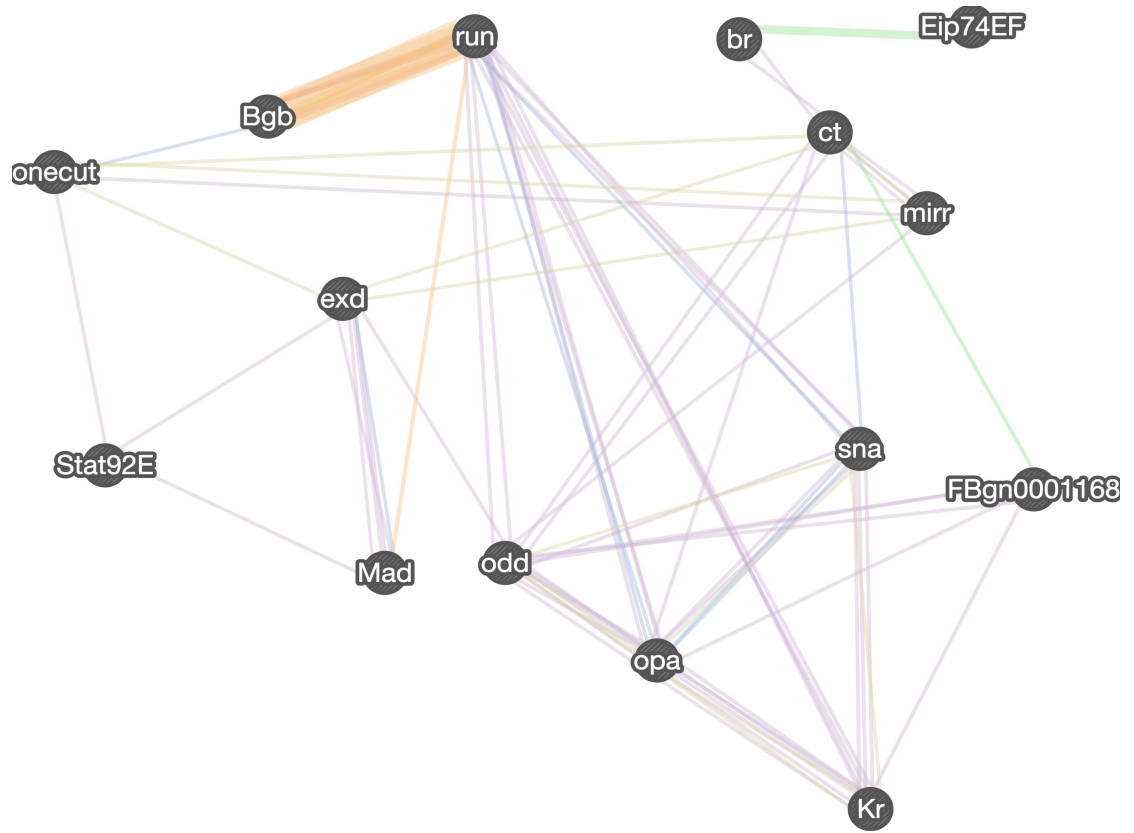
Figure 3.7.7 Regulatory network for Cluster 2b transcription factors: Sixteen nodes of transcription factor gene used for construction of regulatory network.

Figure 3.7.7   Regulatory network for Cluster 2b transcription factors
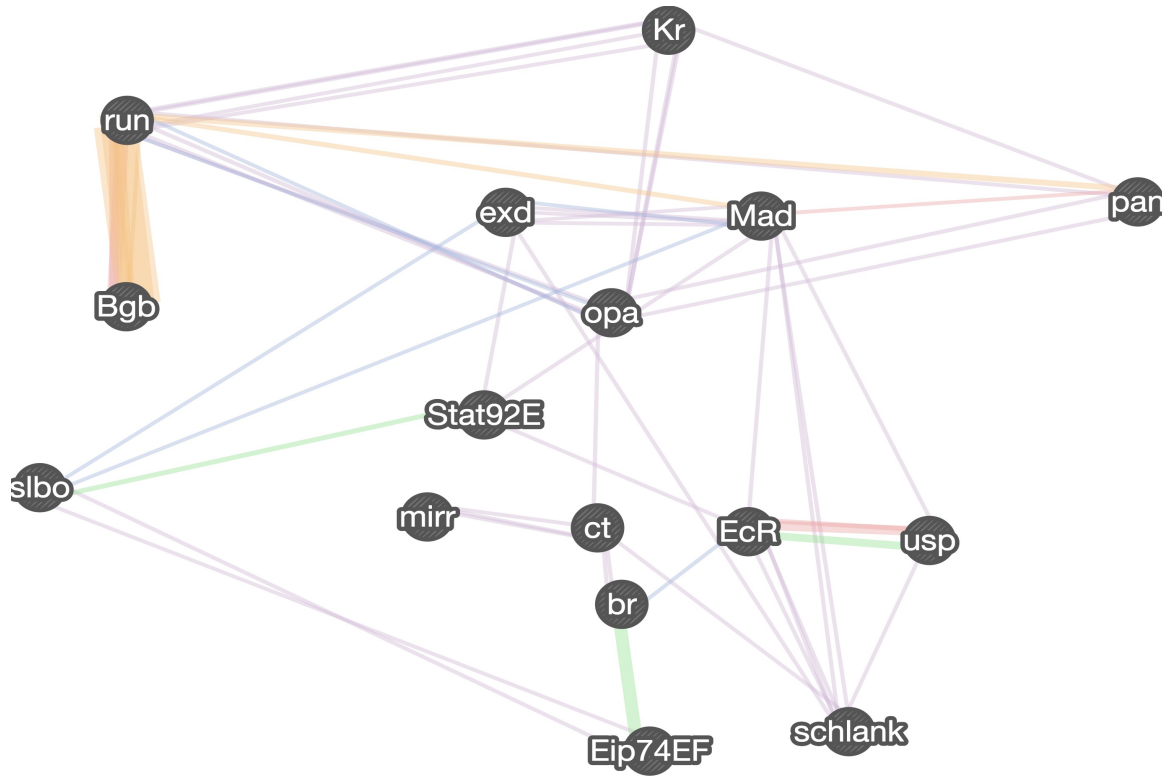
Figure 3.7.8.1 EcR knockdown: Triplicate samples of EcR depletion by RNAi for testing epistatic relation with Stat using qRT-PCR of transcriptomic level compared with negative control gene (Luciferase gene)

Figure 3.7.8.2 Stat knockdown: Triplicate samples of Stat depleted by RNAi for testing epistatic relation with EcR

Figure 3.7.8.1

EcR knockdown

1st   2nd   3rd

Figure 3.7.8.2

STAT knockdown

1st   2nd   3rd
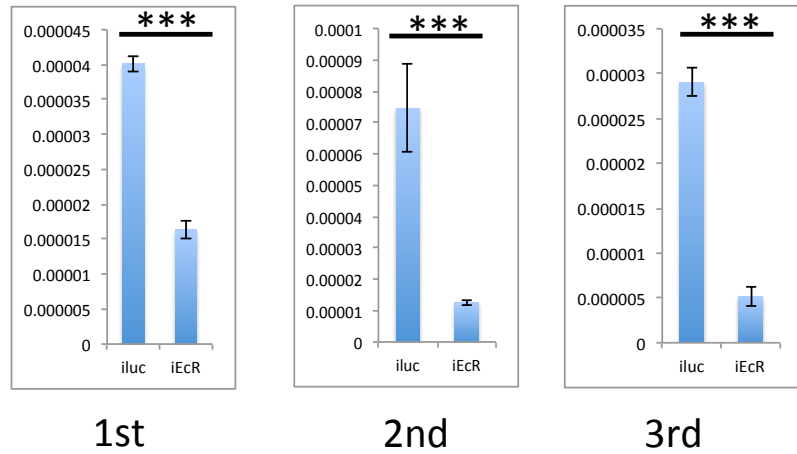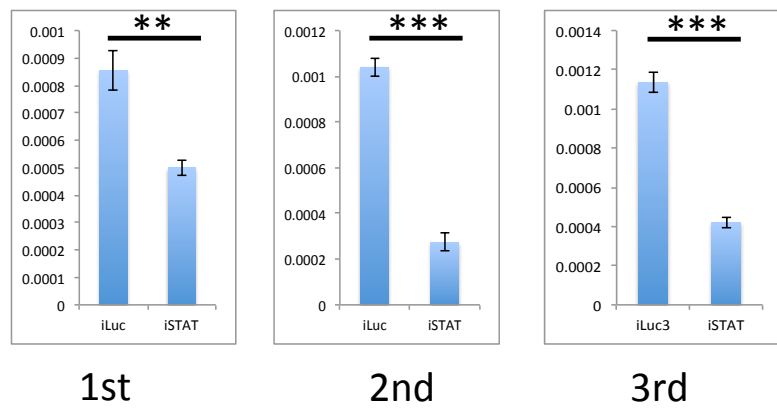
Figure 3.7.9.1: Triplicate qRT-PCR tests for EcR transcription level check on Stat knockdown. EcR transcription level decreased on Stat depletion, which EcR is originally activated by Stat at fat body tissue.

Figure 3.7.9.2: Triplicate qRT-PCR tests for Stat transcription level check on EcR knockdown. Stat transcription level increased on the EcR knockdown, which EcR represses Stat normally.

Figure 3.7.9.1   Effects of Stat knockdown on EcR

1st   2nd   3rd
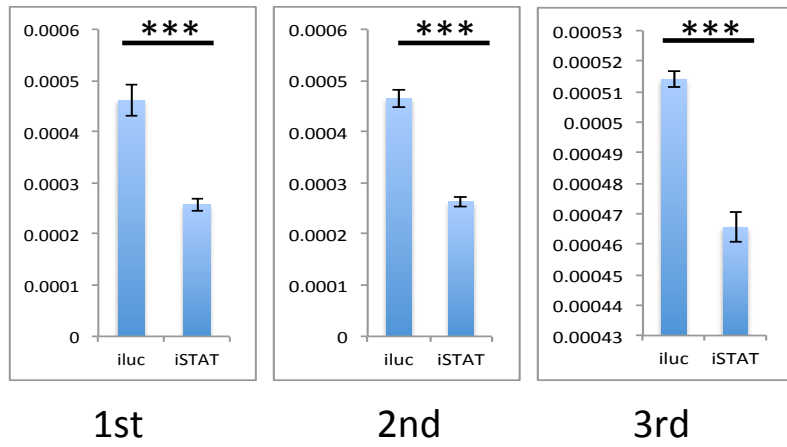
Figure 3.7.9.2

Effects of EcR knockdown on Stat

1st   2nd   3rd
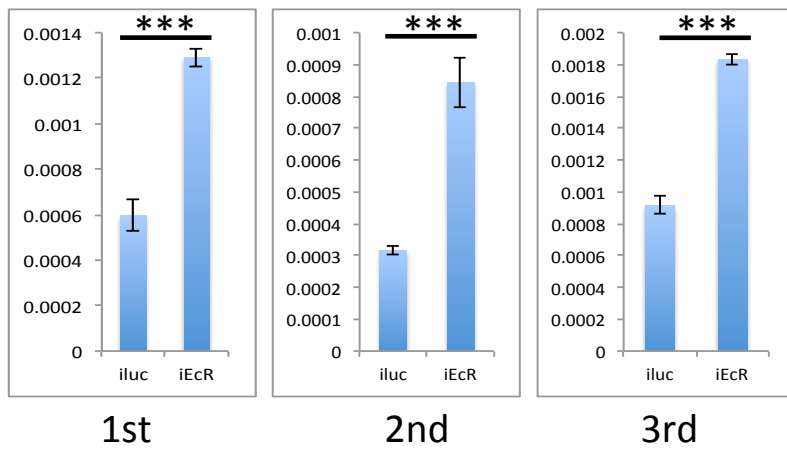
Figure 3.7.10 Achi knockdown: Triplicated Achi knockdown depleted by RNAi for

epistatic test with EcR.

Figure 3.7.10

Achi knockdown

1st   2nd   3rd
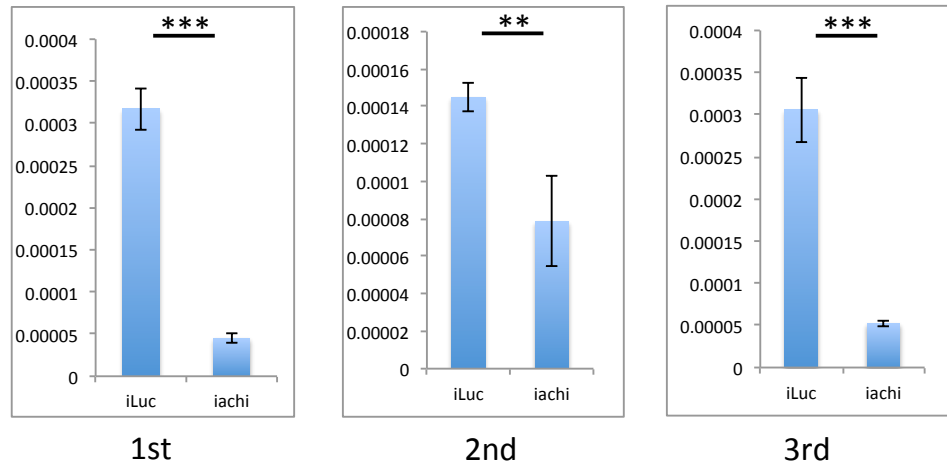
Figure 3.7.11.1 Effects of Achi knockdown on EcR: EcR transcription level on Achi knockdown was drastically decreased. Achi gene activates EcR gene expression.

Figure 3.7.11.2 Effects of EcR knockdown on Achi: The Achi gene does not respond under the EcR depletion status. EcR does not effect on Achi gene expression.

Figure 3.7.11.1   Effects of Achi knockdown on EcR

1st   2nd   3rd

Figure 3.7.11.2

Effects of EcR knockdown on Achi

1st   2nd   3rd
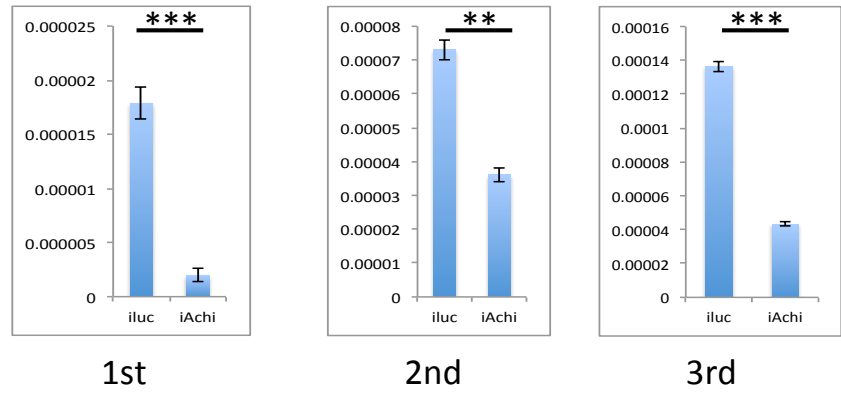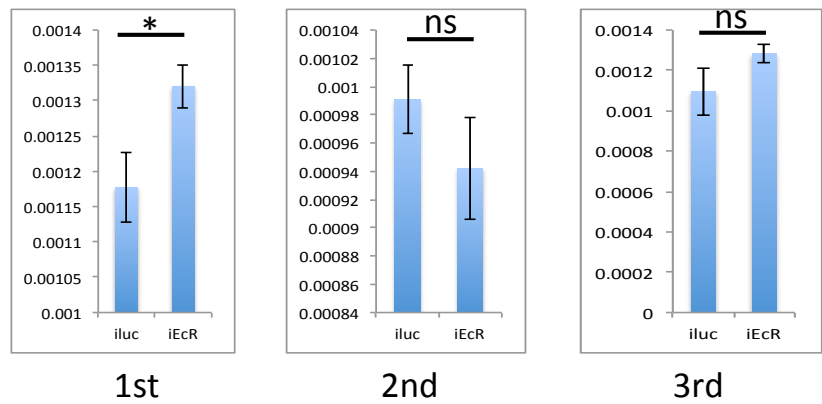
Figure 3.7.12.1 Effect of EcRc knockdown on Br: Br transcription level is dramatically decreased than negative control gene.

Figure 3.7.12.2 Effect of EcRc knockdown on E74: Similarly, E74 transcription level is shown significantly decreased.

Figure 3.7.12.1

Effect of EcRc knockdown on Br

1st   2nd   3rd

Figure 3.7.12.2

Effect of EcRc knockdown on E74

1st   2nd   3rd
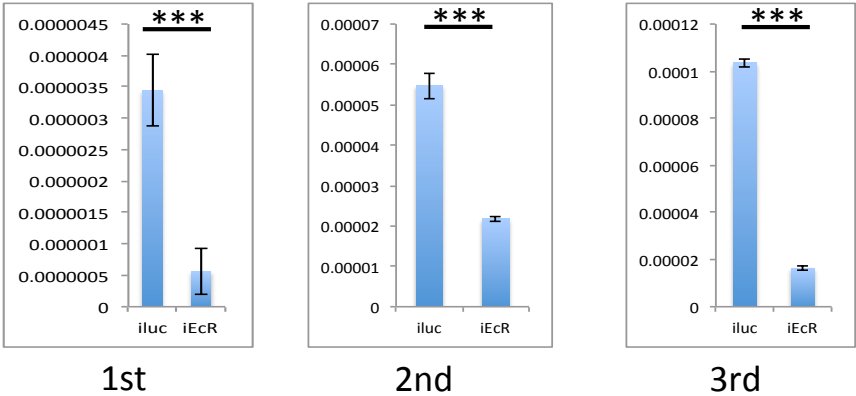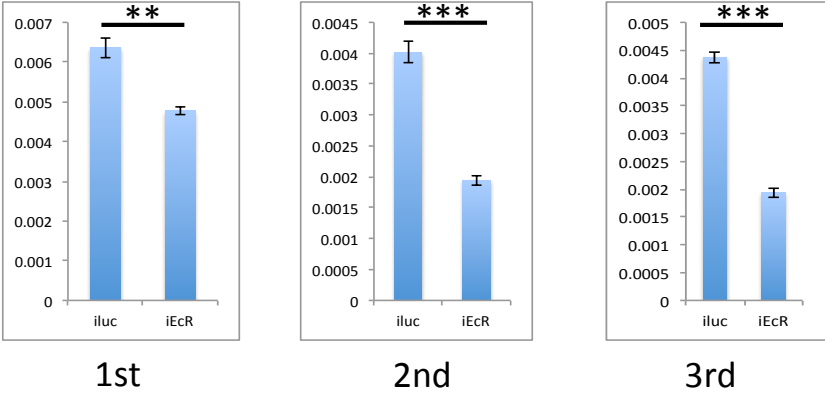
Figure 3.7.13 E74 knockdown: Triplicated qRT-PCR tests of E74 knockdown depleted

by RNAi for epistatic test with three target genes (Figure 3.7.14.1-3.7.14.3).

Figure 3.7.13

E74 knockdown
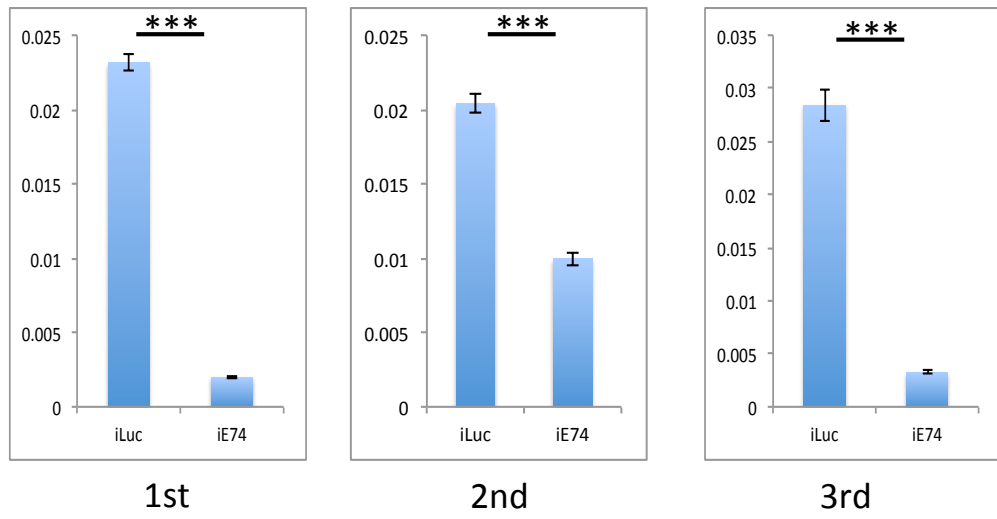
Figure 3.7.14.1, 3.7.14.2 and 3.7.14.3: All three selected target genes shown their significantly increased transcription levels at 24h PBM of fat body tissue. These genes are repressed by E74.

Figure 3.7.14.1 Transcription level of Ankyrin 2,3/unc44 (AAEL014742) on E74 knockdown: The transcription level of Ankyrin 2,3/unc44 (AAEL014742) shown significantly increased.

Figure 3.7.14.2 Transcription level of Lipase (AAEL008222) on E74 knockdown: The transcription level was increased on the E74 knockdown.

Figure 3.7.14.3 Transcription level of Sideroflexin 1,2,3 (AAEL014526) on E74 knockdown: The qRT-PCR test results shown that the transcription of AAEL014526 highly expressed than control genes.

Effect of E74 knock down on early genes with Eip74EF motif in their promoters

Figure 3.7.14.1

1st   2nd   3rd

Ankyrin 2,3/unc44 (AAEL014742)

Figure 3.7.14.2

1st   2nd   3rd

Lipase (AAEL008222)

Figure 3.7.14.3

1st   2nd   3rd

Sideroflexin 1,2,3 (AAEL014526)
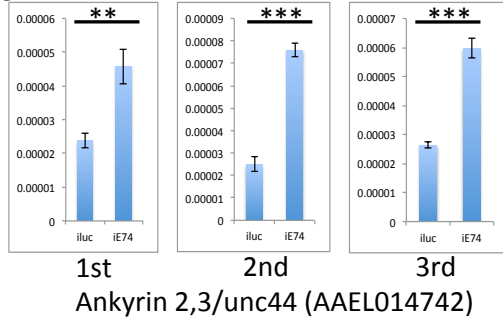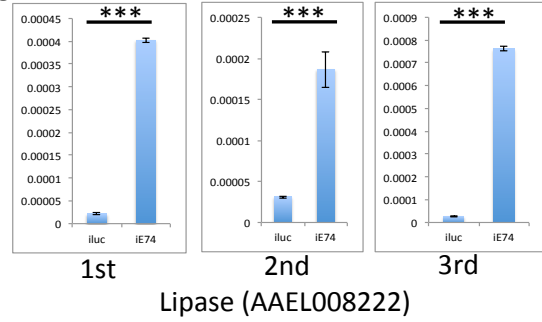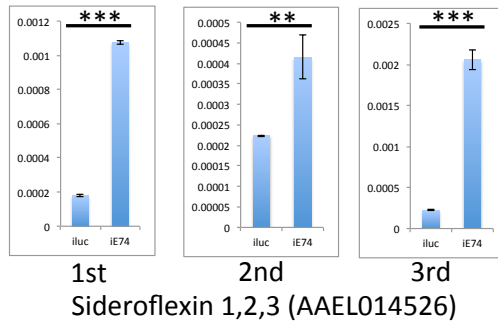
Figure 3.7.15 and 3.7.16: The three negatively regulated target genes of E74 were

checked on the Br knockdown.

Figure 3.7.15 Br knockdown: Triplicated qRT-PCR tests of E74 knockdown depleted by

RNAi for epistatic test with E74 and three target genes of E74 (Figure 3.7.16 and

3.7.17.1-3.7.17.3).

Figure 3.7.16 E74 transcriptional level check on Br knockdown: Under Br knockdown

condition, the E74 transcription level decreased significantly. It means that Br activates

E74.

Br knockdown

Figure 3.7.15

1st   2nd   3rd
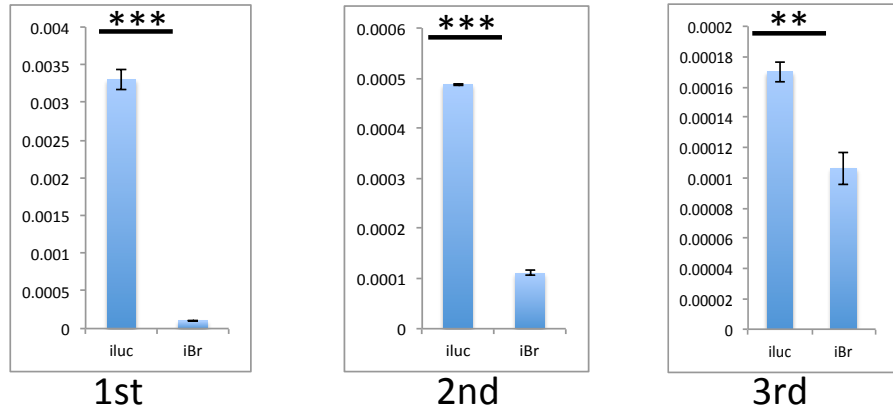
Figure 3.7.16   Effects of Br knockdown on E74

1st   2nd   3rd
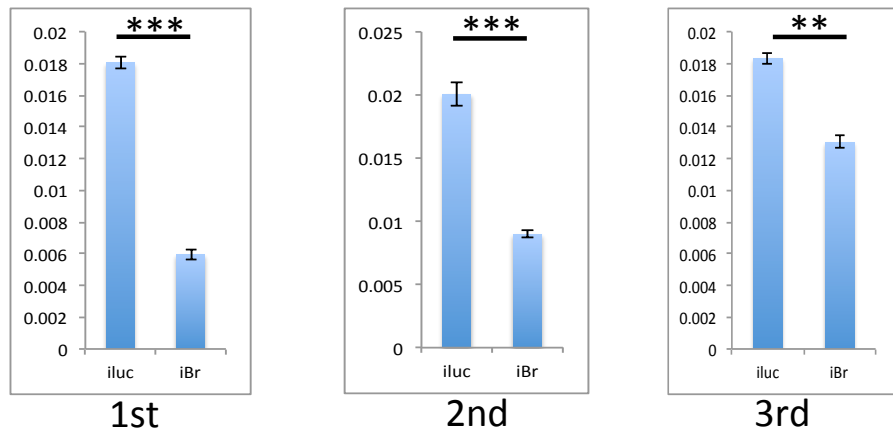
Figure 3.7.17.1, 3.7.17.2 and 3.7.17.3: Effect of Br knock down on early genes with

Eip74EF motif in their promoters: The overall results were quite similar E74 gene

knockdown case.

Figure 3.7.17.1 Transcription level of Ankyrin 2,3/unc44 (AAEL014742) on Br

knockdown: The transcription level of Ankyrin 2,3/unc44 (AAEL014742) shown

significantly increased.

Figure 3.7.17.2 Transcription level of Lipase (AAEL008222) on Br knockdown: The

transcription level was increased on the Br knockdown.

Figure 3.7.17.3 Transcription level of Sideroflexin 1,2,3 (AAEL014526) on Br

knockdown: The repressed gene AAEL014526 by E74 shown that its transcription level

was increased on Br knockdown as similar as E74 knockdown.

Effect of Br knock down on early genes with Eip74EF motif in their promoters

Figure 3.7.17.1

Ankyrin 2,3/unc44 (AAEL014742)

Figure 3.7.17.2

Lipase (AAEL008222)

Figure 3.7.17.3

Sideroflexin 1,2,3 (AAEL014526)
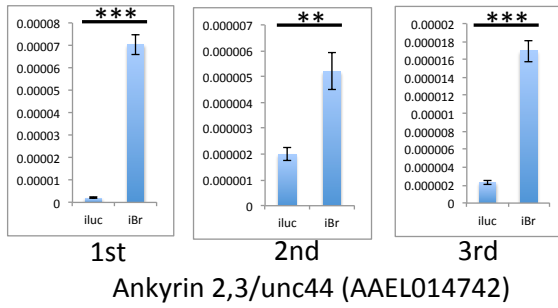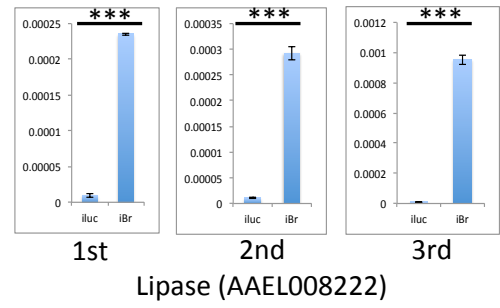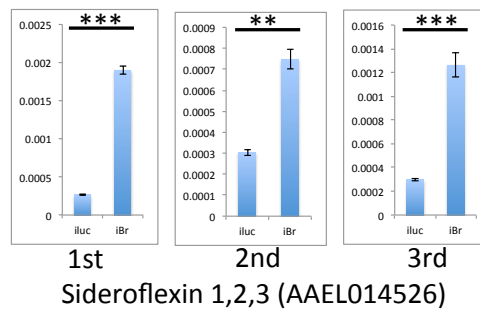
Figure 3.7.18.1, 3.7.18.2 and 3.7.18.3: Effect of EcR knock down on early genes with E74 motif in their promoters: The three negatively regulated genes by E74 were checked their transcription level on EcR knockdown condition. EcR represses E74, then E74 de-represses the three target genes.

Figure 3.7.18.1 Transcription level of Ankyrin 2,3/unc44 (AAEL014742) on EcR knockdown: The transcription level of Ankyrin 2,3/unc44 (AAEL014742) shown significantly increased.

Figure 3.7.18.2 Transcription level of Lipase (AAEL008222) on EcR knockdown: The transcription level was increased on the EcR knockdown.

Figure 3.7.18.3 Transcription level of Sideroflexin 1,2,3 (AAEL014526) on Br knockdown: The repressed gene AAEL014526 by E74 shown that its transcription level was increased on EcR knockdown as similar as E74 and Br knockdown.

Effect of EcR knock down on early genes with Eip74EF motif in their promoters

Figure 3.7.18.1

1st　　　2nd　　　3rd
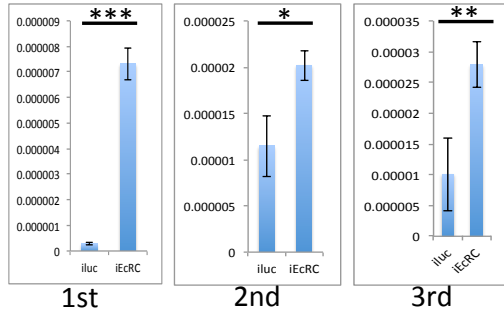
Ankyrin 2,3/unc44 (AAEL014742)

Figure 3.7.18.2

1st　　　2nd　　　3rd
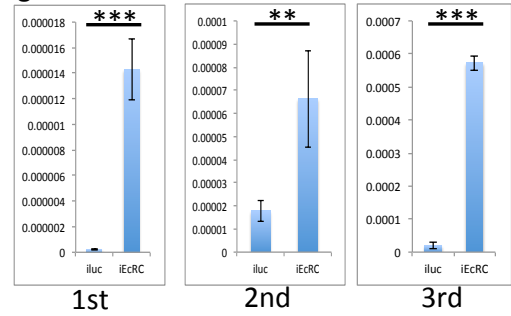
Lipase (AAEL008222)

Figure 3.7.18.3

1st　　　2nd　　　3rd
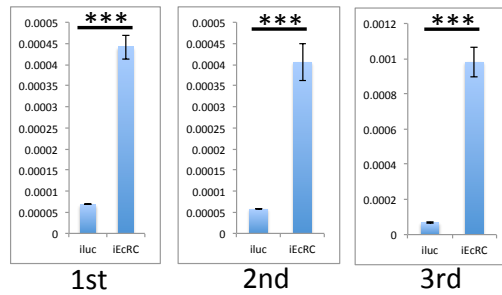
Sideroflexin 1,2,3 (AAEL014526)

Figure 3.7.19.1 Effects of Stat knockdown on E74: Triplicate qRT-PCR results shown

clearly that E74 transcription level significantly increased.

Figure 3.7.19.2 Effects of E74 knockdown on Stat: E74 knockdown did not affect STAT

at the transcriptomic level.

Effects of Stat knockdown on E74

Figure 3.7.19.1

1st 2nd 3rd

Figure 3.7.19.2   Effects of E74 knockdown on Stat

1st 2nd 3rd
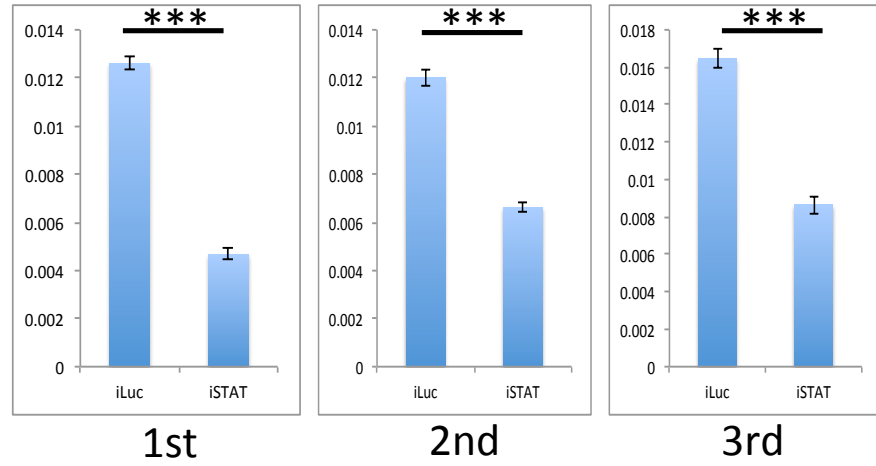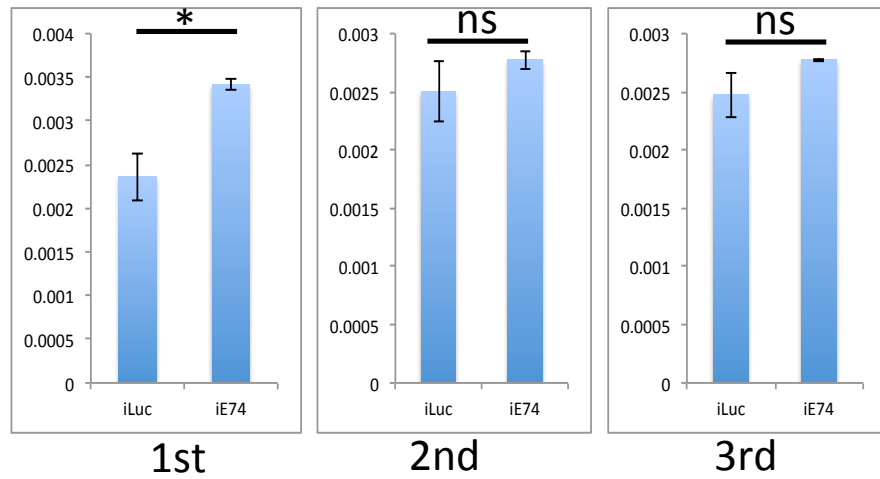
Figure 3.7.20.1 Effects of Stat knockdown on Ankyrin 2,3/unc44 (AAEL014742):

Negatively regulated gene by E74Ankyrin 2,3/unc44 (AAEL014742) shown significantly

increased.

Figure 3.7.20.2 Effects of Stat knockdown on Lipase (AAEL008222): Negatively

regulated gene by E74 Lipase (AAEL008222) shown increased.

Figure 3.7.20.3 Effects of Stat knockdown on Sideroflexin 1,2,3 (AAEL014526):

Negatively regulated gene by E74 Sideroflexin 1,2,3 (AAEL014526) shown increased.

Effect of Stat knock down on early genes with Eip74EF motif in their promoters

Figure 3.7.20.1
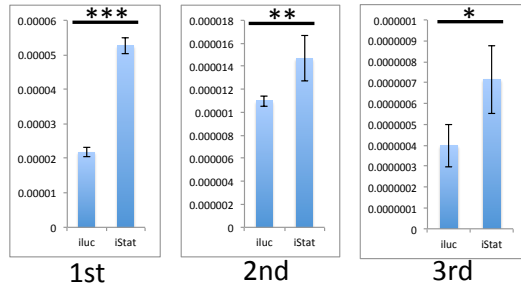
Ankyrin 2,3/unc44 (AAEL014742)

Figure 3.7.20.2
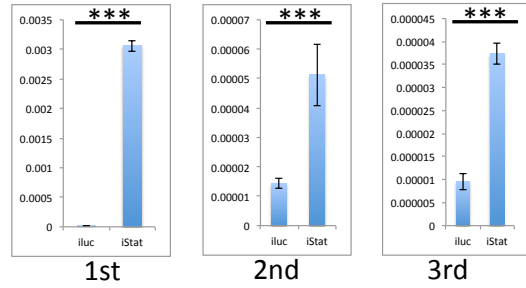
Lipase (AAEL008222)

Figure 3.7.20.3
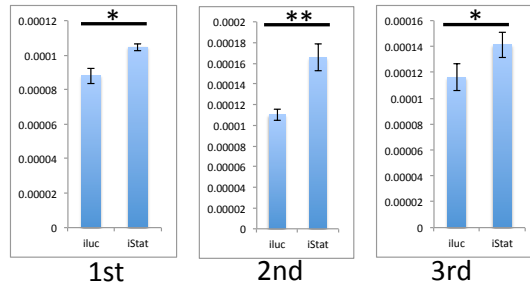
Sideroflexin 1,2,3 (AAEL014526)

Figure 3.7.21.1, 3.7.21.2, and 3.7.21.3: Effect of EcR kd on Cluster 5 genes with TTGATTGA motifs

Figure 3.7.21.1 Effects of transcription level of Amp dependent ligase (AAEL002568): On the EcR knockdown, AAEL002568 gene shown increased transcription level.

Figure 3.7.21.2 Transcription level of Serine protease (AAEL006568) on EcR knockdown: The selected negatively regulated AAEL006568 at 24 PBM shown increased transcription level on EcR depletion.

Figure 3.7.21.3 Transcription level of Sulfate transporter (AAEL012037) on EcR knockdown: Sulfate transporter shown increased transcription level on EcR knockdown.

Effect of EcR kd on Cluster 5 genes with TTGATTGA motifs

Figure 3.7.21.1

AAEL002658 – amp dependent ligase

Figure 3.7.21.2

AAEL006568 – serine protease

Figure 3.7.21.3

AAEL012037 – Sulphate transporter
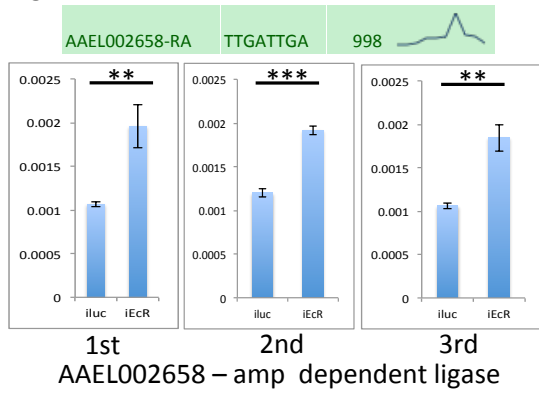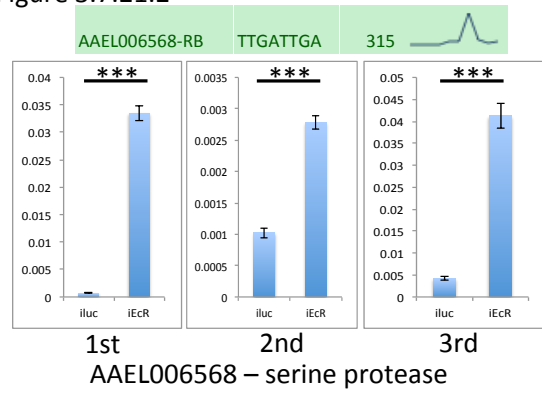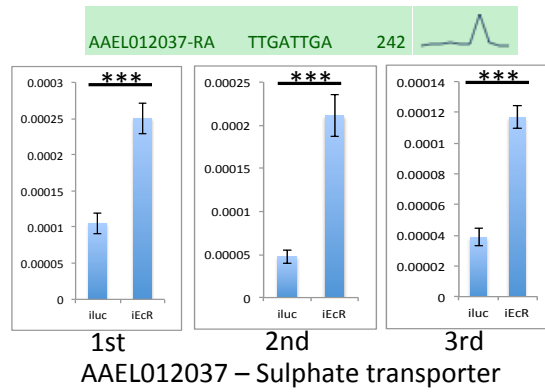
Figure 3.7.22.1 Mirr knockdown: Mirr knockdown was conducted for epitatic test with Br. (Figure 3.7.22.2)

Figure 3.7.22.2 Effects of Mirr knockdown on Br: Br transcription level significantly decreased on Mirr depletion. Mirr activate Br.

Figure 3.7.22.3 Effectc of Br knockdown on Mirr: Mirr transcription level check on Br knockdown: Mirr dose not repond Br depletion.

Figure 3.7.22.1

Mirr knockdown

1st

2nd

3rd

Figure 3.7.22.2

Effects of Mirr knockdown on Br

1st

2nd

3rd
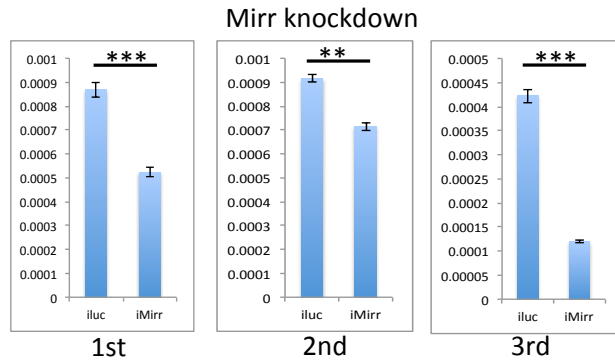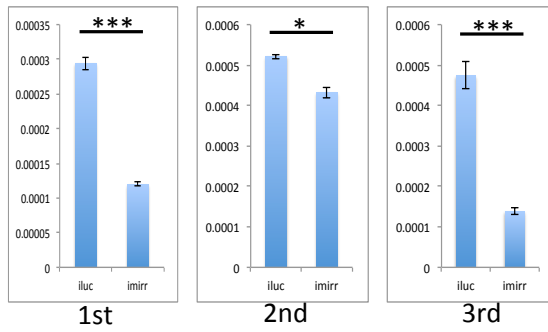
Figure 3.7.22.3

Effects of Br knockdown on Mirr

1st

2nd

3rd
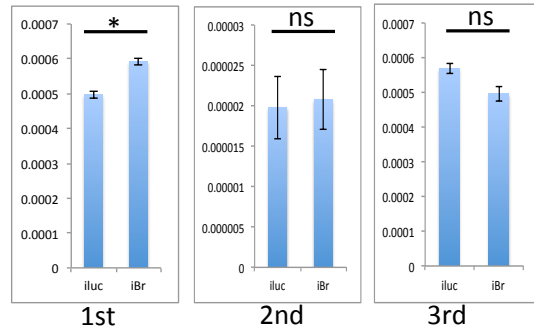
Figure 3.7.23.1, 3.7.23.2 and 3.7.23.3: Phenotype checking on Br and Mirr gene depletion

Figure 3.7.23.1 Ovary development on Br and Mirr knockdown: Ovary development was inhibited on Br and Mirr knockdown.

Figure 3.7.23.2-3.7.23.3 Egg follicle size measures: Significantly different egg sizes from Br and Mirr depletion

Figure 3.7.23.1

iLuc　　　　　　　　　　iBr　　　　　　　　　　iMirr

Figure 3.7.23.2

iLuc　　　　　　　　　　iBr　　　　　　　　　　iMirr

Figure 3.7.23.3

Follicle
diameter
(uM)

Genes depleted by RNAi
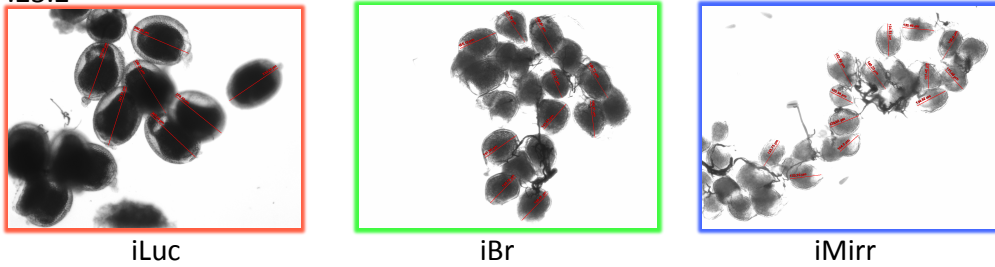
163
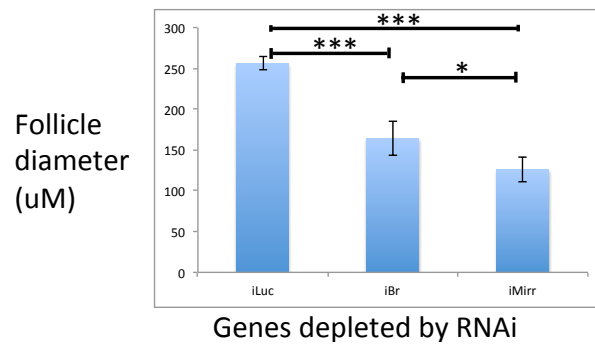
Figure 3.7.24.1 and 3.7.24.2: Effects of Br knockdown on Br target genes

Figure 3.7.24.1 Effects of Br knockdown on two selected genes: Negatively regulated two genes by Br, AAEL008886 and AAEL011651, shown their increased transcription level on Br knockdown.

Figure 3.7.24.2 Effects of Mirr knockdown on two selected genes: Negatively regulated two genes by Br, AAEL008886 and AAEL011651, shown their increased transcription level on Mirr knockdown.

Figure 3.7.24.1

Effects of Br knockdown on Br target genes

1st   2nd   3rd
AAEL008886 – conserved hypothetical

1st   2nd   3rd
AAEL011651 – l-caldesmon, putative

Figure 3.7.24.2

Effects of Mirr knockdown on Br target genes

1st   2nd   3rd
AAEL008886 – conserved hypothetical

1st   2nd   3rd
AAEL011651 – l-caldesmon, putative
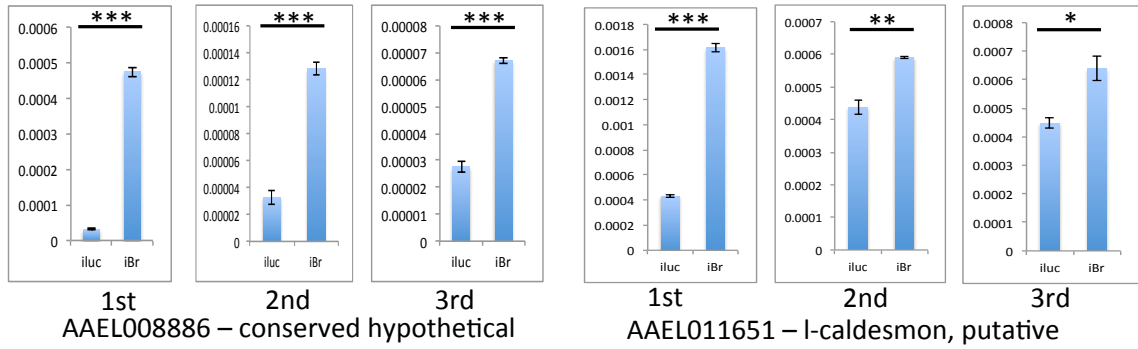
Figure 3.7.25 Experimentally proven gene regulatory network
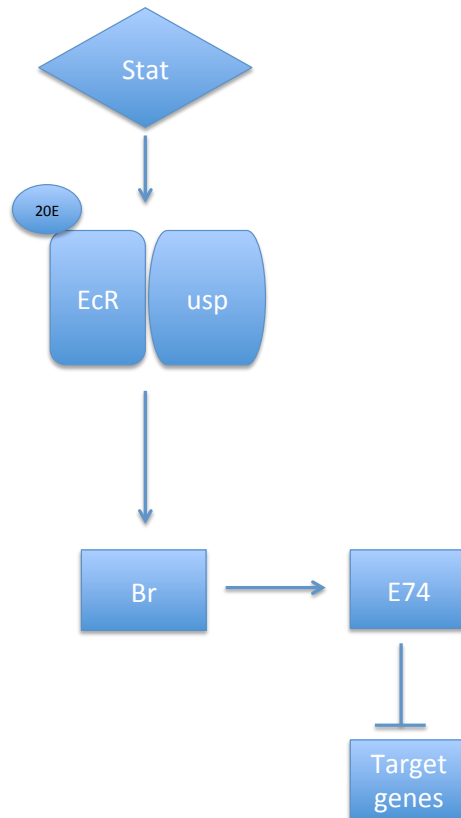
Tables 3.7.1

Table 3.7.1 Caption: Gene regulatory network statistics from GeneMANIA

3.7.1 Tables

	Cluster	CO-expression	Physical interactions	Genetic interactions	Shared protein domains	Co-localization	predicted interaction
Egs	1	13.60%	8.04%	5.56%	0.00%	1.62%	69.24%
	6	28.57%	19.66%	30.58%	1.93%	0.00%	27.11%
EMGs	4	51.35%	18.86%	14.66%	12.21%	2.89%	0.00%
	7	23.09%	12.24%	9.61%	0.00%	3.58%	51.48%
LMGs	5	9.61%	33%	40.97%	9.80%	6.62%	0.00%
LGs	2A	49.59%	5.16%	4.97%	1.93%	2.78%	35.57%
	2B	40.32%	11.44%	8.64%	0.00%	2.82%	36.78%

* EGs : Eearly Genes

* EMGs : Eearly Mid Genes

* LMGs : Late Mid Genes

* LGs : Late Genes

CHAPTER IV

Prediction of targets for miRNAs involved in *Aedes aegypti* mosquito reproduction

4.1 Abstract

miRNAs are 21 to 24 nucleotide long non-coding RNAs that modulate gene expression at the post-transcriptional level by coupling with target sites on mRNAs (Lucas et al., 2015). Enormous numbers of miRNAs have been reported from mammalian, plants and virus species, since miRNAs were first discovered (Fire et al., 1998; Griffith-Jones et al., 2008). According to miRBase, 28645 miRNAs have been registered to this day (Release 21, June 2014, http://www.mirbase.org). However, the functions of only a few of these miRNAs have been understood. The study of genes targeted by the miRNAs is one of the best ways to understand their biological functions. Therefore, miRNA target identification is a critical step towards determining the functionality of miRNAs. Many target prediction tools are readily available, but the primary concern for computational prediction methods is the false-positive results. We have devised and applied a multi-algorithm approach for computational target prediction of miRNAs that involves five different target prediction programs. We have focused on the critical properties of miRNA-mRNA binding mechanism such as thermodynamic energy between mRNA and stem sequence of miRNA, evolutionary conservation with phylogenetically similar species, etc. Using this multi-algorithm approach we have been able to identify significant molecular regulators for gene expression involved in female mosquito reproduction. Our studies have unraveled some of the most important factors related

female mosquito reproduction, the results have been experimentally verified and found to be reliable (Liu et al., 2014, Lucas et al., a: 2015, Zhao et al., 2016). We realize that miRNA target prediction is only a preliminary, but critical step towards a complete understanding of molecular mechanisms when the ultimate goal is to understand different regulatory mechanisms, at the molecular level. Identifying the critical components of various pathways involved in vitellogenesis can lead to effective and novel control mechanisms for mosquito reproduction, by blocking or transforming these components, in future, thereby curbing pathogen transmission and human diseases.

4.2 Introduction

MicroRNAs (miRNAs), play a significant role in repression for gene expression at the post-transcriptional level by binding mostly to the 3' UTR of mRNA, this is due to the fact that the coupled miRNA either leads to translational inhibition of the mRNA or degradation of the target mRNA. There have been multiple studies suggesting the role of miRNAs in various biological aspects like metabolism, developmental timing, cell death, differentiation and proliferation (Bushati et al., 2007; Kloosterman et al., 2006 and Gangaraju et al., 2009).

Although numerous miRNAs are reported in animals, plants and viruses species, it is not easy to identify their functions. Their target genes characterize the functionalities of miRNAs, as the roles played by these genes within the regulatory networks, are easier to identify (Watanabe et al., 2007). Therefore, identification of target genes is one of the most significant tasks in the miRNA research.

The miRNA target identification in early stages of miRNA research was entirely dependent on the classical molecular biology techniques (Lee et al., 1993). There were significant obstacles for these studies such as laborious and tedious experimental work procedures. The scarcity of high-throughput automated biological screening techniques resulted in a large amount of workloads (Min et al., 2010). Therefore, with traditional experimental research can only search for limited number of miRNA targets, on the other hand bioinformatics methodologies, such as string processing algorithm, have been found to predict target genes effectively. Except for the contiguous 7-8-mer miRNA binding sites (traditional canonical binding sites), there are many positions for non-canonical

binding sites which have been shown to degrade mRNA for gene repression (Lucas et al. 2013). Identification of these non-canonical sites through traditional methods is even more difficult.

Multiple target prediction tools based on various algorithms have been developed for miRNA target prediction. Although there are many algorithms and tools, it is vital to use them effectively and efficiently. Because each algorithm has its own pros and cons, it is hard to determine if a specific algorithm would be better than the others. Some of the different criteria used to build the most popular algorithms are: sequence alignment for seed region, thermodynamic energy calculation, evolutionary conservation search and statistical inference methods. It should be mentioned that the accuracy of the implemented algorithms are around 60% when tested under the same condition (Alexiou et al., 2009), even though every algorithm has different and distinctive features to predict authentic targets. Witcos et al. in 2011 summarized the problems regarding target prediction algorithms: "There is no universal miRNA target prediction algorithm that can be used routinely and efficiently for every 3' UTR sequence since not all rules of mRNA-miRNA interactions have been seen discovered yet."

Therefore, it has been suggested that the practical approach for miRNA target prediction should involve the use of multiple algorithms (Watanabe et al., 2007; Witcos et al., 2011). Based on this idea, we established an approach for miRNA target prediction the uses multiple algorithms. The approach is explained in the results section and Fig. 4.7.1 illustrates the flow chart for this working approach (Fig. 4.7.1). Before selecting algorithms, we focused on different criteria of miRNA-mRNA interaction such as

171

complementary sequence searching, optimal thermodynamic stability calculation, and evolutionary conservation analysis with orthologous sequences alignment of 3'UTR. Based on these criteria, we selected four prediction tools. These were selected on the basis of the background algorithms running these programs.

The Probability of Interaction by Target Accessibility (PITA) algorithm aims at highly accessible target sites in mRNA and miRNA seed sequence by combining thermodynamic and seed finding (pattern matching) procedures (PITA reference here). The RNAHybrid algorithm employs RNA secondary structure prediction algorithms, Mfold (Mattews et al., 1999) and RNAfold (Hofacker, 2003). These two secondary structure prediction programs are based on the thermodynamic free energy calculation for stabilization. The TargetScan algorithm employs the thermodynamics between miRNA seed sequence and mRNA target site. First, this algorithm searches the seed sequence against the mRNA sequences (Dynamic programming). Then, using RNAfold program from Vienna package calculates the thermodynamics properties of binding between miRNA and the sequence matching sites of mRNA. The miRanda algorithm runs in three steps with different algorithms running at every step. The first phase is the detection of the target sites that may make duplexes. The second step is a calculation step to optimize thermodynamic stabilities. The last stage is filtering out false positive results based on evolutionary conservation. Candidate target sites from step 2 are compared with relevant parts of mRNA sequences from other species (Betel et al., 2008).

We also used a program that was developed in-house. The algorithm for the "in-house" program is a relaxed implementation of the five features proposed by Grimson et

172

al. 2007 and has been describe in details in Liu et al. 2014. In short this program identifies all seed matches and uses five different criteria for scoring points viz. (*i*) perfect seed-matching; (*ii*) additional Watson–Crick base pairing in nucleotides 12–17 (complementary seed site); (*iii*) location in locally Adenine-Uracil (AU)-rich regions; (*iv*) sites in the first 15 nucleotides of the 3′ UTR are seemingly less effective; and (*v*) when long 3′ UTR's are divided into quartiles, sites that reside within the first and last quartiles of the 3′UTR are more efficient. Then the sites are ranked according to the scores—the higher the score, the more likely it is that the seed match is a putative miRNA target site.

In the results section we discuss the results obtained using this approach for the target prediction of three mosquito miRNAs viz. miR-1174, miR-8 and miR-275, these results have been experimentally validated and the targets along with the respective miRNAs were found have significant roles in female mosquito reproduction.

4.3 Materials and methods

4.3.1 miRNA sequence extraction:

Ae. aegypti and *An. gambiae* miRNA-1174, miR-8 and miR-275 sequences were downloaded from the miRBase database (mirbase.org). Mature miRNA sequence or the seed sequence of a particular miRNA was used, depending on the program requirements.

4.3.2 Extraction of 3'-UTR sequences

The 3'-UTR sequences of *Ae. aegypti* and *An. gambiae* were extracted from the Vectorbase Gene Feature File ver. 3. (GFF3). *Aedes aegypti* Genebuild 1.3 and Anopheles gambiae Genebuild 3.7 were used for this purpose and the sequences were extracted using R script and Bioconductor packages.

4.3.3. Establishment of orthologous gene pairs between *Aedes aegypti* and *Anopheles gambiae*

One of the target prediction tools (TargetScan), needed information regarding orthologous sequences of 3'UTRs, as an input. Reciprocal BLAST search method was used to detect orthologous genes between *Ae. aegypti* and *An. gambiae* mosquito species. The reciprocal BLAST search method was described in a previous section.

4.3.4 Program selection for micro RNA target prediction

Four miRNA target prediction programs were selected for target prediction. All programs downloaded and locally installed in a Linux cluster at UCR Bioinformatics core

facility are shown with version names: 1) miRanda: 3.3a, 2) PITA: 6, 3) RNAhybrid: 2.1.1, and 4) TargetScan: 6.0. The "in-house" target prediction program written with Perl scripting language was also used.

4.3.5 Execution of micro RNA target prediction programs

All programs were executed following each user's manual instruction. Miranda program was run with default parameter settings. However determination of optimal parameters is a difficult task (NP-hard problem) due to enormous parameter combinations, we solved this issue by conducting empirical executions.

Following command line statements were used to execute the programs:

Miranda (default parameters used): miranda <input_microRNAs.fa> < input_utr.fa >

PITA: ./pita_prediction.pl -utr <input_utr.fa> -mir <input_microRNAs.fa> -prefix example -gxp

RNAhybrid: ./RNAhybrid -d 2,0.1 -t <input_utr.fa> -q <input_microRNAs.fa> -p 0.001

TargetScan: ./targetscan_60.pl < input_microRNAs_info.txt> <UTR_align_AA_AG.fasta>

In-house Perl script: perl find_miRNA_target.pl <Transcript file> <miRNA sequence> <Seed region length>

4.4 Results

As mentioned in the introduction section of this chapter we used a multi-algorithm approach for the prediction of targets of the following miRNAs. The flow chart 4.7.1 describes our approach. In brief either the seed sequence or the mature miRNA sequence of a particular miRNA was downloaded from the miRbase database. All available 3' UTR sequences, of *Ae. aegypti* and *An. gambiae*, were extracted from vectorbase. *An. gambiae* was used to evaluate the cross-species conservation of the miRNA targets. The target prediction programs were run locally and the results coming out of these programs were parsed using either excel or PERL scripts. Lists of target genes were made by identifying the overlap between the results from all the programs.

4.4.1 Computational analyses for identification of targets for miR-1174:

The preliminary results of target prediction for miRNA-1174 are shown in the tables 4.7.2.1 through 4.7.2.10. The first step towards analyzing the data from all five programs, for both *Ae. aegypti* and *An. gambiae* 3'UTRs, was to clean up the results by getting rid of the duplicate and isoform information. Once this was done the results were compared to one another to check for overlaps. It was observed that putative miR-1174 targets were detected by all five programs within the 3' UTRs of two genes in *Ae. Aegypti* (AAEL002510 - Serine hydroxymethyl transferase and AAEL012079 - Heat Repeat Containing 5B) and within the 3'UTR of the ortholog of one other (AAEL001779 - Apoptosis related Bax inhibitor) in *An. gambiae* (AGAP005775). Putative targets were detected within the 3' UTRs of two Ae. *Aegypti* genes (AAEL005411 - Equilibrative

176

nucleoside transporter and AAEL010558 - Conserved hypothetical protein) and also within the 3' UTRs of the orthologs of these target genes in *An. gambiae* (AGAP003892, AGAP000964), by four out of five programs. The final list for functional validation consisted of five putative target genes (Table - 4.7.2.11, Liu et al. 2014), three of which were detected by all five programs either in *Ae. aegypti* or in *An. gambiae* when their 3'UTRs were searched. The two other genes were selected as putative target genes, even though all five programs did not detect them. The reason behind this was that four out of the five programs viz. the "in-house" program, TargetScan, Miranda and PITA were able to detect putative miR-1174 targets within the 3' UTRs of the said genes in *Ae. aegypti* and their orthologs in *An. gambiae*.

4.4.2 Computational analyses for identification of targets for miR-8:

Similar to that of miR-1174, *in-silico* target prediction was performed for identifying the genes targeted by miR-8. 3'UTR sequences of *Ae. aegypti* (Aaeg 1.3) and *Anopheles gambiae* (Agam 3.7) were extracted from Vectorbase and were searched for this purpose. The preliminary target prediction data is shown in tables 4.7.3.1 trough 4.7.3.10 where, 4.7.3.1 through 4.7.3. 5 lists the putative targets identified by the different prediction programs for *Ae. aegypti* miRNA-8 within the 3'UTRs of the *Ae. aegypti* genes. Similarly, tables' 4.7.3.6 through 4.7.3.10 lists the *An. gambiae* putative targets. *An. gambiae* targets were predicted to check which of the *Ae. aegypti* targets were evolutionarily conserved in both mosquitoes. The overlap between the results from the five different programs was checked to come up with a list of high confidence putative targets for miR-8 (Table 4.7.3.11, Lucas et al. 2015), for functional validation.

Only one gene (AAEL003834-RA - Metalloproteinase, putative) was detected as a putative target of miR-8 by five different programs. But, there were six others that were detected by four out of five different programs. Three out of these six viz. AAEL000454 - Isocitrate dehydrogenase, AAEL001232 - Tubulointerstitial nephritis antigen and AAEL008347- Monocarboxylate transporter were detected by the "in-house" program, TargetScan, Miranda and PITA. The orthologs of these genes in *An. gambiae* were detected as putative targets for miR-8 by the same four programs. Three other genes viz. AAEL005958 – Oxidoreductase, AAEL009059 - arp2/3 complex 16 kd subunit (P16-arc) and AAEL009530 - tmc6 protein (evin) were detected as putative targets of miR-8 only in *Ae. aegypti* by the "in-house" program, RNAhybrid, Miranda and PITA.

4.4.3 Computational analyses for identification of targets for miR-275:

Putative targets for miR-275 were identified in the same way as that for miR-1174 and miR-8. The same five programs were used to search the 3' UTRs of *Ae. aegypti* and *An. gambiae* genes. The initial results (Tables 4.7.4.1 through 4.7.4.10) were cleaned and analyzing the overlaps identified high confidence putative targets.

Like for miR-8 there was only one gene AAEL006852, Sarco/endoplasmic reticulum Ca2$^+$ adenosine triphosphatase (SERCA), which was detected as a putative target by all five programs. Four more genes were included in the final list of the putative targets (Table 4.7.4.11; Zhao et. al. unpublished data) as four out of the five programs detected these. Two of these viz. AAEL000577- DNA binding protein elf-1 and AAEL006126 - conserved hypothetical protein similar to Vitellogenin A1, were detected by "in-house" program, TargetScan, Miranda and PITA. The other two viz.

AAEL002714- kinesin-like protein KIF23 (mitotic kinesin-like protein 1) and

AAEL005191 - cdk10/11, were detected by the "in-house" program, TargetScan,

Miranda and PITA.

4.5 Discussion

miRNAs have been implicated to be involved in virtually all biological processes and pathways (Li and Zhang, 2015). miRNAs function by controlling gene expression of their target genes at the post-transcription level. Therefore, identification of their targets is absolutely essential for functional characterization of the respective miRNAs. But, there are problems related to miRNA target prediction, accurate prediction is highly challenging due to factors like imperfect base-pairing and condition-specific miRNA regulatory dynamics (Li and Zhang, 2015). There are numerous programs available for the prediction of miRNA targets, which employs highly diverse algorithms, each looking for certain different features related to the miRNA-mRNA interaction. It has been suggested the use of multiple algorithms and a cross check between the results coming out of these is necessary for accurate true positive target predictions (Ekimler and Sahin, 2014). We have designed an approach for miRNA target prediction where we have used five different target prediction programs. Even though most of these are sequence-based methods, the underlying algorithms operating each of the programs are diverse and use different features of the miRNA-mRNA interaction in order to predict targets. Since, each of the programs are looking at different features, targets identified by more than one program suggests that those targets possess more than one feature required for the miRNA-mRNA interactions. Therefore, a target detected by five programs is more likely to be an accurate true positive than detected by one or even three programs. This was the logic behind comparing the results from each of the five programs and identifying the overlaps before coming up with a final list of high confidence putative miRNA targets.

In the previous section we have described the results from prediction of targets for

three miRNAs viz. miR-1174, miR-8 and miR-275. The ultimate proof of a good

prediction method is functional validation of the results. Our predictions have

subsequently been validated through multiple experiments and the results of experimental

validation of targets for miR-11174 and miR-8 have already been published (Liu et al.,

2014 and Lucas et al., 2015), and that of miR-275 would soon be submitted (Zhao et al.,

unpublished data).

In case of mir-1174, Dual Luciferase Reporter Assay was used to check its

response on the 3′UTRs of the top five potential targets. One of the five predicted targets

(AAEL002510 - Serine Hydroxymethyltransferase, SHMT) was found to yield less than

50% luciferase activity (Liu et al., 2014). This in-vitro analysis was followed by in-vivo

experiments to confirm that SHMT is an authentic target of mir-1174. The results also

showed that miR-1174 targeted SHMT to control important physiological functions like

sugar absorption, fluid excretion, blood intake in the gut, and, consequently, egg

maturation and survival in *Ae. aegypti* and *An. Gambiae* (Liu et al. 2014).

miR-8 microRNA family is widely studied, and this has been found to be a

critical regulator for development in animals (Lucas et al., 2015). miR- 8 depletion

revealed that this miRNA affects mosquito ovarian development and thereby

reproduction. To understand the mechanism identification of its targets were absolutely

necessary. As mentioned in the results section, seven potential targets were identified

after analyzing the results from five different programs. AAEL001232 - Tubulointerstitial

nephritis antigen, which is the ortholog for *Drosophila* Swim, was chosen for further

analyses by Lucas et al. (Lucas et al, 2015). This was due to the fact that

Drosophila Swim has been shown to function in the wingless (Wg) signaling pathway.

Drosophila Swim is essential for long-range Wg signaling as it maintains the stability of

the Wg ligand (Mulligan et al., 2012). It is also known that *Drosophila* miR-8 targets

multiple factors in the Wg signaling pathway (Kennell et al., 2008). Another reason

behind this selection was the conservation of this target between *Ae*. *Aegypti* and *An*.

Gambiae, this was one of the three genes identified by four out of five programs in both

mosquitos (Lucas et al., 2015). *In-vivo* functional analysis of miR-8 confirmed that this

miRNA targets Ae. Swim (AAEL001232) and thereby acts on the long-range Wingless

signaling pathway to control the secretory activity of yolk protein precursors required for

oocyte development (Lucas et al. 2015).

For the validation of miR-275 targets Zhao et al. took an approach that was

similar to that of Lucas et al., 2015. They looked for prior information regarding the

predicted targets. Bryant et al. in 2010 had already demonstrated that miR-275 is

essential for blood digestion. There was also information available regarding the ortholog

of AAEL006852-SERCA gene, regarding suggested its relation to contraction and

relaxation of the muscle in mammalian species. Therefore, Zhao et al. (unpublished data)

selected the SERCA gene for further experimental validation. In-vitro dual luciferase

assay demonstrated that miR-275 targets SERCA directly. Further in-vivo experiments

like RNAi mediated depletion of *SERCA* or application of SERCA-specific inhibitor

Thapsigargin (TG) and application of miR-275 mimic to restore SERCA levels were

performed. It was confirmed that miR-275 targets SERCA to effect blood digestion, subsequent egg development and multiple gut functions.

Thus targets predicted for all three miRNAs have been experimentally validated and their involvement in either blood digestion or ovarian development or both have been demonstrated. This clearly shows that the method used for miRNA target prediction is quite robust. It should be mentioned that only one each, of the predicted miR-8 and miR-275 targets, were tested based on information available from other species and both were found to be authentic targets. In case of miR-1174 only one out of the five predicted targets was found to respond in the dual luciferase assay with miR-1174. The said target was one of the three targets predicted by all five programs but did not show evolutionary conservation. There were two other predicted targets that were evolutionarily conserved but were not found to be authentic targets. On the other hand the target for mir-8 was not detected by all five but four programs and was conserved in *An. gambiae*. In case of miR-275 the validated target was the only one to be detected by all five programs. Therefore, it is worth mentioning that while determining targets for functional validation, one should take both evolutionary conservation and overlap between predicted targets, into account. Also, prior information about the targets from other species can be very important as shown in case of both miR-8 and miR-275.

The method we have developed and used here for miRNA target prediction helps in reduction of the number of potential targets that are to be tested functionally, and thus saves time resource and money. The approach is similar to the one we have used for the prediction of Transcription Factor Binding Sites (TFBSs) in Chapter II. It should also be

mentioned that this approach could be applied for any other organism, for which

sequence information related to 3'UTRs and miRNAs is available.

4.6 References

Liu S., Lucas KJ, Roy S, Ha, J, & Raikhel AS, Mosquito-specific microRNA-1174 targets serine hydroxymethyltransferase to control key functions in the gut, Proc. Natl. Acad. Sci. 2014 Vol. 111 14460-14465

Lucas KJ, Roy S, Ha J, Gervaise AL, Kokoza VA, & Raikhel AS, MicroRNA-8 targets the Wingless signaling pathway in the female mosquito fat body to regulate reproductive processes, Proc. Natl. Acad. Sci. 2015 Vol. 112(3): 1440-1445

Zhao B, Lucas KJ, Saha TT, Ha J, Chen C, Roy S, & Raikhel AS, MicroRNA-275 directly targets sarco/endoplasmic reticulum Ca2$^+$ adenosine triphosphatase (SERCA) to control key functions in the mosquito gut, ICE 2016 Orlando, Florida Sep. 25-30, 2016.

Bushati N, Cohen SM: MicroRNA functions. Annu Rev Cell Dev Biol. 2007, 23: 175-205. 10.1146/annurev.cellbio.23.090506.123406.

Kloosterman WP, Plasterk RH: The diverse functions of microRNAs in animal development and disease. Dev Cell. 2006, 11: 441-450. 10.1016/j.devcel.2006.09.009.

Gangaraju VK, Lin H: MicroRNAs: key regulators of stem cells. Nat Rev Mol Cell Biol. 2009, 10: 116-125. 10.1038/nrm2621.

mirSVR predicted target site scoring method: Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Betel D, Koppal A, Agius P, Sander C, Leslie C., Genome Biology 2010 11:R90

microRNA target predictions: The microRNA.org resource: targets and expression. Betel D, Wilson M, Gabow A, Marks DS, Sander C., Nucleic Acids Res. 2008 Jan; 36(Database Issue): D149-53.
miRanda application: Human MicroRNA targets. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS., PLoS Biol. 2005 Jul;3(7):e264.

miRanda algorithm: MicroRNA targets in Drosophila. Enright AJ, John B, Gaul U, Tuschl T, Sander C and Marks DS., Genome Biology 2003 5;R1

Rehmsmeier, Marc and Steffen, Peter and Hoechsmann, Matthias and Giegerich, Robert Fast and effective prediction of microRNA/target duplexes RNA, RNA, 2004
Author: M. Rehmsmeier
RNAhybrid is a tool for finding the minimum free energy hybridization of a long and a short RNA. The hybridization is performed in a kind of domain mode, ie. the short sequence is hybridized to the best fitting part of the long one. The tool is primarily meant as a means for microRNA target prediction.

Kertesz M, Iovino N, Unnerstall U, Gaul U & Segal E, The role of site accessibility in microRNA target recognition, Nature Genetics 2007 39, 1278 - 1284 Published online: 23 September 2007 | doi:10.1038/ng2135

Bartel DP. MicroRNA: Genomics, Biogenesis, Mechanism, and Function, 2004, Cell, Vol. 116, 281-297, Jan. 23

Lewis BP, Burge CB, Bartel DP. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. Cell, 120:15-20 (2005).

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell 2003; 115(7): 787-798

Witokos, T. M., Koscianska, E., and Krzyzosiak, W. J., 2011, Practical Aspects of microRNA Target Prediction, Current Molecular Medicine Vol. 11, 93-109

Watanabe Y, Tomita M, Kanai A, Computational Methods for microRNA Target Prediction, Methods in Enzymology 2007, Vol. 427: 65-86

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ., miRBase: tools for microRNA genomics
NAR 2008 36:D154-D158

Li L, Xu J, Yang D, Tan X, Computational approaches for microRNA studies: a review, Mamm Genome, 2010, 21:1-12

Mulligan KA, Fuerer C, Ching W, Fish M, Willert K, Nusse R., Secreted Wingless-interacting molecule (Swim) promotes long-range signaling by maintaining Wingless solubility. 2012 Proc Natl Acad Sci U S A. 2012 Jan 10;109(2):370-7. doi: 10.1073/pnas.1119197109. Epub 2011 Dec 27.

Grimson A, Farth KK, Johnston WK, et al., Garrette-Engele P, Lim LP, Bartel DP, MicroRNA targeting specificity in mammals: determinants beyond seed paring Mol. Cell 2007; 27(1): 91-105

Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG, Lost in translation: an assessment and perspective for computational microRNA target identification. Bioinformatics 2009; 25(23): 3049-3055

Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschi T, Manoharan M, Stoffel M, Silencing of microRNAs in vivo with 'antagomirs' Nature 2005; 438(7068):685-689

Mulligan KA, et al. (2012) Secreted Wingless-interacting molecule (Swim) promotes long-range signaling by maintaining Wingless solubility. Proc Natl Acad Sci USA **109**(2):370–377

Kennell JA, Gerin I, MacDougald OA, Cadigan KM (2008) The microRNA miR-8 is a conserved negative regulator of Wnt signaling. Proc Natl Acad Sci USA **105**(40):15417–15422

4.7 Figures & tables

Figure 4.7.1 miRNA target prediction pipeline:

Multiple algoritm based tools were used to predict authentic miRNA targets. miRNA mature sequences were downloaded from miRbase. Then, 3' UTR regions were extracted from Vectorbase genome sequences.

Figure 4.7.1 miRNA target prediction flow chart
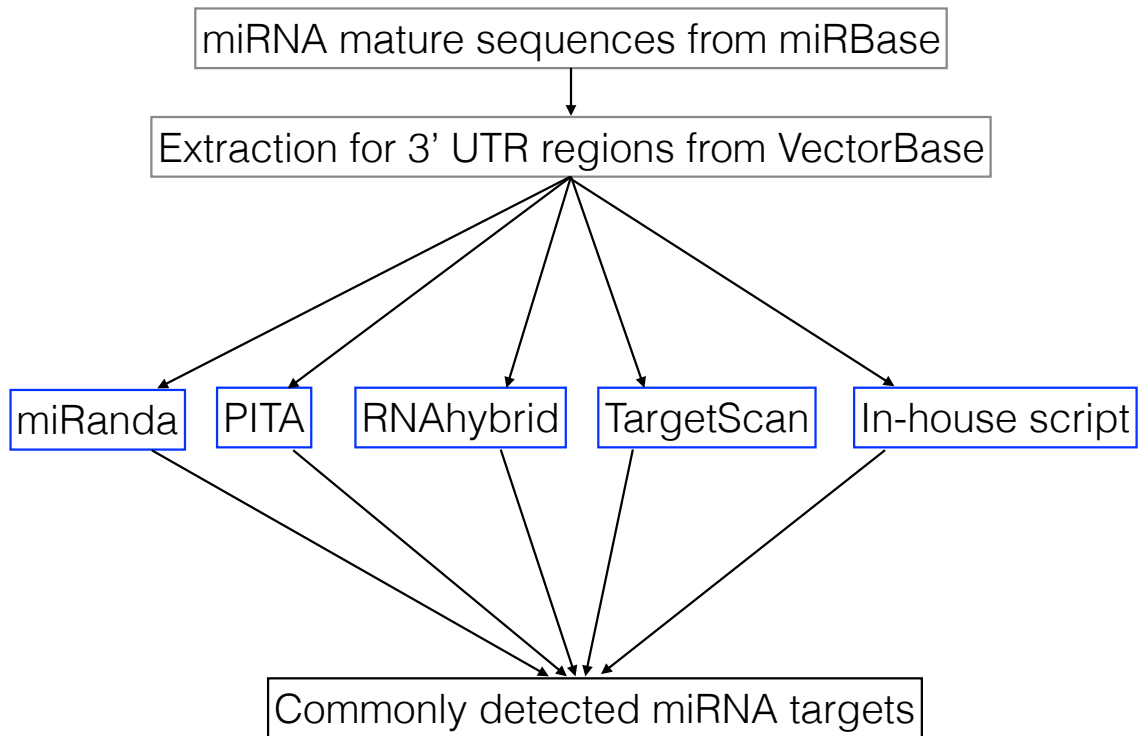
Table 4.7.2 miRNA-1174 target prediction results from five programs (4.7.2.1-4.7.2.10)

Table 4.7.2.1: Target prediction results from In-house program in Ae. aegypti
Table 4.7.2.2: Target prediction results from TargetScan program in Ae. Aegypti
Table 4.7.2.3: Target prediction results from miRanda program in Ae. Aegypti
Table 4.7.2.4: Target prediction results from PITA program in Ae. Aegypti
Table 4.7.2.5: Target prediction results from RNAhybrid program in Ae. aegypti
Table 4.7.2.6: Target prediction results from In-house program in An. gambiae
Table 4.7.2.7: Target prediction results from TargetScan program in An. gambiae
Table 4.7.2.8: Target prediction results from miRanda program in An. gambiae
Table 4.7.2.9: Target prediction results from PITA program in An. gambiae
Table 4.7.2.10: Target prediction results from RNAhybrid program in An. gambiae
Table 4.7.2.11 Final list of miR-8 target genes

Table 4.7.2 miR-1174 targets detected by each of the five programs:

Aedes aegypti results
Table 4.7.2.1

In-house targets

AAEL000111-RA	AAEL005391-RB	AAEL003959-RA	AAEL006126-RB	AAEL007534-RA	AAEL010184-RA	AAEL013382-RA
AAEL000452-RA	AAEL005400-RA	AAEL004327-RA	AAEL006240-RA	AAEL007927-RA	AAEL010268-RA	AAEL013408-RB
AAEL000508-RA	AAEL002078-RA	AAEL004356-RA	AAEL006298-RA	AAEL008041-RA	AAEL010306-RA	AAEL013537-RA
AAEL000804-RA	AAEL002375-RA	AAEL004521-RA	AAEL006346-RA	AAEL008041-RB	AAEL010470-RA	AAEL013537-RB
AAEL000804-RB	AAEL002411-RA	AAEL004586-RA	AAEL006346-RC	AAEL008342-RA	AAEL010558-RA	AAEL013546-RA
AAEL000958-RA	AAEL002411-RB	AAEL004621-RA	AAEL006346-RD	AAEL008391-RA	AAEL010765-RA	AAEL014280-RA
AAEL001188-RA	AAEL002475-RA	AAEL004653-RA	AAEL006356-RA	AAEL008489-RA	AAEL010814-RC	AAEL014847-RA
AAEL001264-RA	AAEL002510-RA	AAEL004688-RA	AAEL006413-RA	AAEL008688-RA	AAEL011338-RA	AAEL015122-RA
AAEL001264-RB	AAEL002510-RB	AAEL004691-RA	AAEL006447-RA	AAEL008723-RA	AAEL011441-RA	AAEL017530-RA
AAEL001394-RA	AAEL002525-RA	AAEL004709-RA	AAEL006454-RA	AAEL008723-RB	AAEL011470-RA	
AAEL001447-RA	AAEL002673-RA	AAEL005217-RA	AAEL006489-RA	AAEL009086-RA	AAEL011688-RA	
AAEL001592-RA	AAEL002983-RA	AAEL005391-RA	AAEL006489-RB	AAEL009295-RA	AAEL011711-RA	
AAEL001656-RA	AAEL003078-RA	AAEL005400-RC	AAEL006578-RA	AAEL009795-RA	AAEL011766-RA	
AAEL001722-RA	AAEL003229-RA	AAEL005411-RA	AAEL006958-RA	AAEL009810-RA	AAEL012002-RA	
AAEL001751-RA	AAEL003366-RA	AAEL005428-RA	AAEL007114-RA	AAEL009872-RA	AAEL012076-RA	
AAEL001896-RA	AAEL003424-RA	AAEL005579-RA	AAEL007114-RB	AAEL010023-RA	AAEL012079-RA	
AAEL001896-RB	AAEL003512-RA	AAEL005870-RA	AAEL007353-RA	AAEL010123-RA	AAEL012114-RA	
AAEL001982-RA	AAEL003919-RA	AAEL006107-RA	AAEL007449-RA	AAEL010165-RA	AAEL012267-RA	

Table 4.7.2.2

TargetScan targets

AAEL000111-RA	AAEL002411-RB	AAEL004691-RA	AAEL006454-RA	AAEL008489-RA	AAEL011302-RB	AAEL013382-RA
AAEL000248-RA	AAEL002475-RA	AAEL004709-RA	AAEL006489-RA	AAEL008723-RA	AAEL011302-RB	AAEL013408-RB
AAEL000452-RA	AAEL002510-RA	AAEL005140-RA	AAEL006489-RA	AAEL008723-RB	AAEL011302-RB	AAEL014280-RA
AAEL000804-RA	AAEL002510-RB	AAEL005217-RA	AAEL006489-RB	AAEL008959-RA	AAEL011338-RA	AAEL014301-RA
AAEL000804-RB	AAEL002525-RA	AAEL005258-RA	AAEL006489-RB	AAEL009086-RA	AAEL011470-RA	AAEL014847-RA
AAEL000958-RA	AAEL002673-RA	AAEL005258-RB	AAEL006578-RA	AAEL009295-RA	AAEL011711-RA	AAEL015122-RA
AAEL001264-RA	AAEL002983-RA	AAEL005391-RA	AAEL006958-RA	AAEL009810-RA	AAEL012079-RA	AAEL015122-RA
AAEL001264-RB	AAEL003203-RA	AAEL005391-RB	AAEL006993-RC	AAEL009872-RA	AAEL012114-RA	AAEL017098-RA
AAEL001394-RA	AAEL003203-RB	AAEL005411-RA	AAEL007114-RA	AAEL010119-RA	AAEL012243-RA	AAEL017333-RA
AAEL001592-RA	AAEL003229-RA	AAEL005428-RA	AAEL007114-RB	AAEL010123-RA	AAEL012293-RA	
AAEL001656-RA	AAEL003326-RA	AAEL005579-RA	AAEL007353-RA	AAEL010165-RA	AAEL012293-RA	
AAEL001667-RA	AAEL003366-RA	AAEL006107-RA	AAEL007534-RA	AAEL010306-RA	AAEL012655-RA	
AAEL001722-RA	AAEL003581-RA	AAEL006126-RB	AAEL007535-RA	AAEL010470-RA	AAEL013119-RA	
AAEL001896-RA	AAEL003581-RB	AAEL006240-RA	AAEL007927-RA	AAEL010558-RA	AAEL013119-RB	
AAEL001896-RB	AAEL003704-RA	AAEL006298-RA	AAEL008041-RA	AAEL010765-RA	AAEL013119-RC	
AAEL001982-RA	AAEL003919-RA	AAEL006346-RA	AAEL008041-RB	AAEL010814-RC	AAEL013207-RA	
AAEL002181-RA	AAEL004356-RA	AAEL006346-RC	AAEL008319-RA	AAEL010895-RA	AAEL013272-RA	
AAEL002375-RA	AAEL004621-RA	AAEL006346-RD	AAEL008342-RA	AAEL011133-RA	AAEL013307-RA	
AAEL002411-RA	AAEL004653-RA	AAEL006413-RA	AAEL008391-RA	AAEL011302-RB	AAEL013382-RA	

Table 4.7.2.3

miRANDA targets

AAEL000804-RA	AAEL006126-RB	AAEL013139-RG	AAEL000053-RA	AAEL004832-RA	AAEL001264-RB	AAEL001357-RA
AAEL000804-RB	AAEL005009-RA	AAEL013139-RI	AAEL011116-RA	AAEL003424-RA	AAEL006298-RA	AAEL006346-RA
AAEL011779-RA	AAEL002475-RA	AAEL010094-RA	AAEL011116-RB	AAEL014847-RA	AAEL003512-RA	AAEL006346-RC
AAEL010097-RA	AAEL005391-RA	AAEL006554-RA	AAEL011116-RC	AAEL010268-RA	AAEL000015-RA	AAEL006346-RD
AAEL014212-RA	AAEL005391-RB	AAEL007191-RA	AAEL004586-RB	AAEL004653-RA	AAEL006454-RA	AAEL004860-RA
AAEL012079-RA	AAEL004621-RA	AAEL010762-RA	AAEL001896-RA	AAEL006240-RA	AAEL011688-RA	AAEL011172-RA
AAEL013382-RA	AAEL001751-RA	AAEL005579-RA	AAEL001896-RB	AAEL010306-RA	AAEL011470-RA	AAEL000508-RA
AAEL013408-RB	AAEL005428-RA	AAEL007927-RA	AAEL012267-RA	AAEL008041-RA	AAEL001656-RA	AAEL008688-RA
AAEL007466-RA	AAEL001722-RA	AAEL005870-RA	AAEL006413-RA	AAEL008041-RB	AAEL013546-RA	AAEL003229-RA
AAEL015122-RA	AAEL012035-RA	AAEL009470-RA	AAEL002673-RA	AAEL002375-RA	AAEL003366-RA	AAEL011478-RA
AAEL004376-RA	AAEL006370-RA	AAEL003129-RA	AAEL003108-RA	AAEL012002-RA	AAEL009872-RA	AAEL011478-RB
AAEL011112-RA	AAEL002983-RA	AAEL010558-RA	AAEL001982-RA	AAEL007534-RA	AAEL012095-RA	AAEL009645-RA
AAEL011112-RB	AAEL003078-RA	AAEL005458-RA	AAEL011338-RA	AAEL000111-RA	AAEL004699-RA	AAEL009645-RB
AAEL011112-RC	AAEL009295-RA	AAEL000119-RA	AAEL001394-RA	AAEL006578-RA	AAEL005400-RA	AAEL004709-RA
AAEL003824-RA	AAEL005733-RB	AAEL014526-RA	AAEL014280-RA	AAEL007449-RA	AAEL005400-RC	AAEL012117-RA
AAEL005881-RA	AAEL005733-RA	AAEL009086-RA	AAEL000452-RA	AAEL002525-RA	AAEL004688-RA	AAEL017098-RA
AAEL010023-RA	AAEL008342-RA	AAEL004327-RA	AAEL001447-RA	AAEL004691-RA	AAEL002411-RB	AAEL000297-RA
AAEL004875-RA	AAEL006688-RA	AAEL005411-RA	AAEL008723-RA	AAEL007114-RA	AAEL002411-RA	AAEL006116-RA
AAEL004369-RA	AAEL006489-RA	AAEL014530-RA	AAEL008723-RB	AAEL007114-RB	AAEL011711-RA	AAEL003739-RA
AAEL011471-RA	AAEL006489-RB	AAEL007353-RA	AAEL007669-RA	AAEL010765-RA	AAEL008489-RA	AAEL010123-RA
AAEL011471-RB	AAEL002078-RA	AAEL012341-RA	AAEL006447-RA	AAEL012071-RA	AAEL017530-RA	AAEL007731-RA
AAEL011471-RC	AAEL005662-RA	AAEL003193-RB	AAEL009795-RA	AAEL001853-RA	AAEL001592-RA	AAEL003453-RA
AAEL011471-RD	AAEL000081-RA	AAEL004548-RA	AAEL009770-RA	AAEL003959-RA	AAEL009810-RA	AAEL003572-RA
AAEL002510-RA	AAEL000682-RA	AAEL010470-RA	AAEL009770-RB	AAEL011441-RA	AAEL006958-RA	AAEL006609-RA
AAEL002510-RB	AAEL012109-RA	AAEL013537-RA	AAEL004521-RA	AAEL005515-RE	AAEL001188-RA	AAEL010509-RA
AAEL000054-RA	AAEL013139-RA	AAEL013537-RB	AAEL011766-RA	AAEL005217-RA	AAEL012114-RA	AAEL012730-RA
AAEL012360-RA	AAEL013139-RB	AAEL010165-RA	AAEL010814-RC	AAEL002537-RA	AAEL006107-RA	AAEL004865-RA
AAEL009285-RA	AAEL013139-RC	AAEL003919-RA	AAEL004586-RA	AAEL010184-RA	AAEL011240-RA	AAEL001799-RA
AAEL001830-RA	AAEL013139-RD	AAEL006356-RA	AAEL004356-RA	AAEL000958-RA	AAEL011240-RB	
AAEL013675-RA	AAEL013139-RE	AAEL007160-RA	AAEL008391-RA	AAEL001264-RA	AAEL012076-RA	

Table 4.7.2.4

PITA targets

AAEL012079-RA	AAEL010558-RA	AAEL006240-RA	AAEL008041-RA	AAEL006958-RA	AAEL004688-RA	AAEL001896-RA
AAEL010123-RA	AAEL002510-RA	AAEL002475-RA	AAEL008041-RB	AAEL002411-RA	AAEL011338-RA	AAEL001896-RB
AAEL001982-RA	AAEL002510-RB	AAEL013382-RA	AAEL005400-RA	AAEL010306-RA	AAEL004586-RA	AAEL001751-RA
AAEL003078-RA	AAEL002078-RA	AAEL009086-RA	AAEL005400-RC	AAEL012002-RA	AAEL003512-RA	AAEL001264-RA
AAEL005579-RA	AAEL009795-RA	AAEL009810-RA	AAEL011470-RA	AAEL010165-RA	AAEL006578-RA	AAEL001264-RB
AAEL013408-RB	AAEL004327-RA	AAEL000804-RA	AAEL006298-RA	AAEL005411-RA	AAEL006447-RA	AAEL000508-RA
AAEL008391-RA	AAEL004653-RA	AAEL010184-RA	AAEL010268-RA	AAEL007353-RA	AAEL010765-RA	AAEL008688-RA
AAEL007114-RA	AAEL006413-RA	AAEL006489-RA	AAEL001447-RA	AAEL017530-RA	AAEL001188-RA	AAEL010814-RC
AAEL007114-RB	AAEL007534-RA	AAEL006489-RB	AAEL008723-RA	AAEL012267-RA	AAEL013546-RA	AAEL014280-RA
AAEL004521-RA	AAEL004621-RA	AAEL007449-RA	AAEL008723-RB	AAEL005217-RA	AAEL003919-RA	AAEL013537-RA
AAEL002673-RA	AAEL002525-RA	AAEL006346-RA	AAEL015122-RA	AAEL008489-RA	AAEL001656-RA	AAEL013537-RB
AAEL005428-RA	AAEL011441-RA	AAEL006346-RC	AAEL000804-RB	AAEL006107-RA	AAEL011711-RA	AAEL011688-RA
AAEL008342-RA	AAEL012076-RA	AAEL006346-RD	AAEL003229-RA	AAEL004709-RA	AAEL002375-RA	AAEL006454-RA
AAEL001722-RA	AAEL005391-RA	AAEL009295-RA	AAEL001592-RA	AAEL002983-RA	AAEL010470-RA	
AAEL003366-RA	AAEL005391-RB	AAEL003424-RA	AAEL000452-RA	AAEL009872-RA	AAEL011766-RA	
AAEL010023-RA	AAEL000111-RA	AAEL001394-RA	AAEL012114-RA	AAEL014847-RA	AAEL003959-RA	
AAEL007927-RA	AAEL006126-RB	AAEL006356-RA	AAEL004356-RA	AAEL004691-RA	AAEL002411-RB	

Table 4.7.2.5

RNA hybrid targets

AAEL005043-RA	AAEL010142-RA	AAEL011929-RA	AAEL008800-RA	AAEL013749-RA	AAEL005639-RA	AAEL006879-RA
AAEL011112-RA	AAEL002510-RA	AAEL000988-RA	AAEL013061-RA	AAEL011462-RA	AAEL014212-RA	AAEL007820-RA
AAEL011112-RB	AAEL002510-RB	AAEL007820-RB	AAEL006688-RA	AAEL011251-RA	AAEL008182-RA	AAEL000339-RA
AAEL011112-RC	AAEL011059-RA	AAEL012828-RA	AAEL005763-RA	AAEL015231-RA	AAEL002155-RA	AAEL002759-RA
AAEL012341-RA	AAEL003431-RA	AAEL007130-RA	AAEL011782-RA	AAEL008763-RA	AAEL002155-RB	AAEL002759-RB
AAEL003552-RA	AAEL002606-RA	AAEL014704-RA	AAEL004908-RA	AAEL009214-RA	AAEL002437-RA	
AAEL007466-RA	AAEL012079-RA	AAEL002809-RA	AAEL003211-RA	AAEL008962-RA	AAEL006466-RA	
AAEL013134-RA	AAEL005316-RA	AAEL002809-RB	AAEL004868-RA	AAEL002573-RA	AAEL007790-RA	

Anophles gambiae results:

Table 4.7.2.6

In-house targets

AGAP005062-RA	AGAP007393-RA	AGAP002685-RA	AGAP003121-RA	AGAP010462-RA	AGAP009879-RA	AGAP007975-RA
AGAP005775-RA	AGAP007393-RB	AGAP002837-RA	AGAP003192-RA	AGAP010480-RA	AGAP012515-RA	AGAP007975-RB
AGAP006462-RA	AGAP007393-RC	AGAP002858-RB	AGAP003352-RC	AGAP011370-RA	AGAP000313-RA	
AGAP006462-RB	AGAP001548-RB	AGAP002858-RC	AGAP003352-RA	AGAP012296-RA	AGAP000343-RA	
AGAP006656-RA	AGAP001548-RA	AGAP002858-RA	AGAP003352-RB	AGAP007963-RA	AGAP000693-RA	
AGAP006942-RA	AGAP001548-RC	AGAP002858-RD	AGAP003892-RA	AGAP008321-RA	AGAP000964-RA	
AGAP007374-RA	AGAP001624-RA	AGAP002858-RE	AGAP004227-RA	AGAP028164-RA	AGAP003794-RB	

Table 4.7.2.7

miRANDA targets

AGAP004854-RA	AGAP007374-RA	AGAP001874-RB	AGAP003352-RA	AGAP004657-RB	AGAP008152-RA	AGAP000964-RA
AGAP005062-RA	AGAP007393-RA	AGAP002685-RA	AGAP003352-RB	AGAP010462-RA	AGAP008321-RA	AGAP003794-RB
AGAP005775-RA	AGAP007393-RB	AGAP002837-RA	AGAP003398-RC	AGAP010480-RA	AGAP028164-RA	AGAP007975-RA
AGAP006031-RA	AGAP007393-RC	AGAP002858-RB	AGAP003475-RA	AGAP010534-RA	AGAP008916-RA	AGAP007975-RB
AGAP006462-RA	AGAP007683-RA	AGAP002858-RC	AGAP003475-RB	AGAP011352-RA	AGAP009464-RA	AGAP011677-RA
AGAP006462-RB	AGAP001548-RB	AGAP002858-RA	AGAP003857-RA	AGAP011370-RA	AGAP009579-RA	
AGAP006656-RA	AGAP001548-RA	AGAP002858-RD	AGAP003892-RA	AGAP012296-RA	AGAP009879-RA	
AGAP006733-RA	AGAP001548-RC	AGAP002858-RE	AGAP004059-RA	AGAP007864-RB	AGAP012515-RA	
AGAP006885-RA	AGAP001624-RA	AGAP003121-RA	AGAP004227-RA	AGAP007864-RA	AGAP000313-RA	
AGAP006942-RA	AGAP001874-RA	AGAP003192-RA	AGAP004524-RA	AGAP007864-RC	AGAP000343-RA	
AGAP007198-RB	AGAP001874-RC	AGAP003352-RC	AGAP004657-RA	AGAP007963-RA	AGAP000693-RA	

Table 4.7.2.8

TargetScan targets

AGAP000223-RA	AGAP001548-RA	AGAP003192-RA	AGAP005062-RA	AGAP006942-RA	AGAP007975-RA	AGAP012515-RA
AGAP000313-RA	AGAP001548-RB	AGAP003352-RA	AGAP005331-RA	AGAP007374-RA	AGAP007975-RB	AGAP013294-RA
AGAP000700-RA	AGAP001548-RC	AGAP003352-RB	AGAP005391-RA	AGAP007393-RA	AGAP008036-RA	AGAP013730-RA
AGAP000700-RB	AGAP001586-RA	AGAP003352-RC	AGAP005413-RB	AGAP007393-RB	AGAP008036-RB	AGAP028164-RA
AGAP000700-RC	AGAP001683-RA	AGAP003790-RB	AGAP005595-RA	AGAP007393-RC	AGAP008321-RA	
AGAP000964-RA	AGAP002685-RA	AGAP003790-RB	AGAP005618-RA	AGAP007713-RA	AGAP008730-RA	
AGAP001110-RA	AGAP002837-RA	AGAP003860-RA	AGAP005775-RA	AGAP007713-RB	AGAP009879-RA	
AGAP001110-RB	AGAP002858-RA	AGAP003860-RB	AGAP006023-RA	AGAP007713-RC	AGAP010462-RA	
AGAP001168-RA	AGAP002858-RB	AGAP003860-RC	AGAP006023-RB	AGAP007713-RD	AGAP010462-RA	
AGAP001168-RB	AGAP002858-RC	AGAP003892-RA	AGAP006204-RA	AGAP007731-RA	AGAP011370-RA	
AGAP001168-RB	AGAP002858-RD	AGAP004227-RA	AGAP006462-RA	AGAP007836-RA	AGAP011386-RA	
AGAP001452-RA	AGAP002858-RE	AGAP004631-RA	AGAP006462-RB	AGAP007836-RB	AGAP011677-RA	

4.7.2.9

PITA targets

AGAP003121-RA	AGAP001548-RB	AGAP007975-RA	AGAP010480-RA	AGAP002858-RA	AGAP005775-RA	AGAP006462-RB
AGAP003892-RA	AGAP001548-RC	AGAP007975-RB	AGAP001624-RA	AGAP002858-RB	AGAP004227-RA	AGAP000693-RA
AGAP002837-RA	AGAP007393-RA	AGAP009879-RA	AGAP002685-RA	AGAP002858-RC	AGAP011370-RA	AGAP003794-RB
AGAP012515-RA	AGAP007393-RB	AGAP000964-RA	AGAP003352-RA	AGAP002858-RD	AGAP028164-RA	AGAP003192-RA
AGAP005062-RA	AGAP007393-RC	AGAP007374-RA	AGAP003352-RB	AGAP002858-RE	AGAP007963-RA	AGAP008321-RA
AGAP001548-RA	AGAP006656-RA	AGAP010462-RA	AGAP003352-RC	AGAP006942-RA	AGAP006462-RA	

4.7.2.10

RNAhybrid targets

AGAP005427-RA	AGAP006031-RA	AGAP007493-RA	AGAP003719-RA	AGAP008108-RA	AGAP028045-RA	AGAP006569-RA
AGAP005928-RA	AGAP006802-RA	AGAP001976-RA	AGAP011298-RA	AGAP008384-RA	AGAP000561-RA	
AGAP005928-RB	AGAP007140-RA	AGAP002346-RA	AGAP008073-RA	AGAP008523-RA	AGAP000598-RA	

Table 4.7.2.11 Final list of putative miR-1174 target genes:

Gene ID	Ortholog ID	Gene name	Abbreviation	Species	Programs
AAEL002510	AGAP004900	Serine hydroxymethyl transferase	SHMT	Aae	IN; TS; PITA; MR; RH
AAEL012079	AGAP002215	Heat Repeat Containing 5B	HR5B	Aae	IN; TS; PITA; MR; RH
AAEL001779	AGAP005775	Apoptosis-related Bax inhibitor	ARBI	Aga	IN; TS; PITA; MR; RH
AAEL005411	AGAP003892	Equilibrative nucleoside transporter	ENT	Aae, Aga	IN; TS; PITA; MR
AAEL010558	AGAP000964	Conserved hypothetical	CH	Aae, Aga	IN; TS; PITA; MR

Abbreviation, gene name abbreviation; Species, species in which the target was detected; Programs, programs that detected a gene target: Aae, *A. aegypti*; Aga, *A. gambiae*; IN, "in-house"; TS, TargetScan; PITA, Probability of Interaction by Target Accessibility; MR, miRanda; RH, RNAhybrid.

Table 4.7.3 miRNA-8 target prediction results from five programs (4.7.3.1-4.7.3.10)

Table 4.7.3.1: Target prediction results from In-house program in Ae. aegypti
Table 4.7.3.2: Target prediction results from miRanda program in Ae. Aegypti
Table 4.7.3.3: Target prediction results from PITA program in Ae. Aegypti
Table 4.7.3.4: Target prediction results from RNAhybrid program in Ae. Aegypti
Table 4.7.3.5: Target prediction results from TargetScan program in Ae. aegypti
Table 4.7.3.6: Target prediction results from In-house program in An. gambiae
Table 4.7.3.7: Target prediction results from miRanda program in An. gambiae
Table 4.7.3.8: Target prediction results from PITA program in An. gambiae
Table 4.7.3.9: Target prediction results from RNAhybrid program in An. gambiae
Table 4.7.3.10: Target prediction results from TargetScan program in An. gambiae
Table 4.7.3.11 Final list of miR-8 target genes

Table 4.7.3 miR-8 results of each program

Aedes aegypti results

Table 4.7.3.1

In-house

AAEL014053-RA	AAEL004246-RA	AAEL005515-RA	AAEL010081-RA	AAEL009658-RC	AAEL008853-RA	AAEL001708-RA
AAEL010379-RA	AAEL001910-RA	AAEL005515-RC	AAEL010465-RA	AAEL007160-RA	AAEL008847-RA	AAEL004121-RA
AAEL015122-RA	AAEL010912-RA	AAEL008654-RA	AAEL010814-RB	AAEL013683-RA	AAEL007018-RA	AAEL004121-RB
AAEL006553-RA	AAEL010576-RJ	AAEL012549-RA	AAEL012139-RA	AAEL003024-RA	AAEL004930-RA	AAEL006891-RA
AAEL006452-RA	AAEL007579-RA	AAEL012035-RA	AAEL012631-RA	AAEL004114-RA	AAEL003651-RA	AAEL000454-RA
AAEL013390-RA	AAEL010124-RA	AAEL000983-RC	AAEL009367-RA	AAEL011350-RA	AAEL002163-RA	AAEL000454-RB
AAEL013390-RB	AAEL006267-RA	AAEL012997-RA	AAEL009700-RA	AAEL007374-RA	AAEL006666-RA	AAEL000608-RA
AAEL000893-RA	AAEL003834-RA	AAEL003972-RA	AAEL009700-RB	AAEL001215-RA	AAEL006675-RA	AAEL002550-RA
AAEL011302-RB	AAEL011556-RA	AAEL015143-RB	AAEL003145-RA	AAEL004301-RA	AAEL001921-RA	AAEL002550-RD
AAEL011302-RE	AAEL006057-RA	AAEL004583-RC	AAEL003989-RA	AAEL003388-RA	AAEL013704-RA	AAEL002709-RA
AAEL008921-RA	AAEL005337-RA	AAEL006581-RA	AAEL005140-RA	AAEL015022-RA	AAEL010050-RA	AAEL002885-RA
AAEL008921-RD	AAEL006322-RA	AAEL006581-RA	AAEL008046-RA	AAEL008758-RA	AAEL010844-RA	AAEL004149-RA
AAEL002294-RA	AAEL006322-RB	AAEL014421-RA	AAEL008058-RA	AAEL001091-RA	AAEL007193-RA	AAEL005017-RA
AAEL005998-RB	AAEL001429-RA	AAEL013875-RA	AAEL008058-RB	AAEL017161-RA	AAEL007532-RA	AAEL005221-RB
AAEL005564-RA	AAEL017300-RA	AAEL013875-RB	AAEL008058-RC	AAEL014891-RA	AAEL007546-RA	AAEL005998-RA
AAEL013138-RA	AAEL009295-RA	AAEL010306-RA	AAEL015059-RA	AAEL002550-RB	AAEL012215-RA	AAEL008836-RA
AAEL001569-RA	AAEL006931-RA	AAEL003066-RA	AAEL001152-RA	AAEL002550-RC	AAEL006460-RA	AAEL009949-RA
AAEL001540-RA	AAEL006931-RC	AAEL011702-RA	AAEL001185-RA	AAEL012360-RA	AAEL012310-RA	AAEL013092-RB
AAEL012472-RA	AAEL009530-RA	AAEL003754-RA	AAEL004351-RA	AAEL005408-RA	AAEL011093-RA	AAEL013614-RA
AAEL003464-RA	AAEL011184-RA	AAEL003754-RB	AAEL004351-RB	AAEL005407-RB	AAEL013092-RA	AAEL013614-RB
AAEL005795-RA	AAEL008129-RA	AAEL007843-RA	AAEL004351-RC	AAEL005407-RC	AAEL000720-RA	
AAEL005821-RA	AAEL007888-RA	AAEL005676-RA	AAEL017098-RA	AAEL002203-RA	AAEL000682-RA	
AAEL008347-RA	AAEL013184-RA	AAEL001232-RA	AAEL009782-RA	AAEL012561-RA	AAEL011412-RA	
AAEL001462-RA	AAEL013678-RA	AAEL011979-RA	AAEL009174-RA	AAEL005958-RA	AAEL013534-RA	
AAEL001644-RA	AAEL002848-RA	AAEL017253-RC	AAEL015454-RA	AAEL000109-RA	AAEL001709-RA	
AAEL002513-RA	AAEL008228-RA	AAEL009043-RA	AAEL007915-RA	AAEL011989-RB	AAEL001708-RB	

Table 4.7.3.2

Miranda_target

AAEL014054-RA	AAEL003453-RA	AAEL009537-RA	AAEL003754-RB	AAEL004351-RA	AAEL004042-RC	AAEL013092-RA
AAEL014053-RA	AAEL003487-RA	AAEL014238-RA	AAEL007843-RA	AAEL004351-RB	AAEL004048-RA	AAEL012109-RA
AAEL010379-RA	AAEL005795-RA	AAEL014238-RB	AAEL005676-RA	AAEL004351-RC	AAEL002550-RB	AAEL012106-RA
AAEL010382-RA	AAEL005821-RA	AAEL011184-RA	AAEL001232-RA	AAEL003348-RA	AAEL002550-RC	AAEL000720-RA
AAEL015122-RA	AAEL005457-RA	AAEL008129-RA	AAEL001372-RA	AAEL003348-RB	AAEL012360-RA	AAEL000682-RA
AAEL006553-RA	AAEL008347-RA	AAEL003548-RA	AAEL001378-RA	AAEL001746-RA	AAEL008936-RA	AAEL011412-RA
AAEL009501-RA	AAEL001462-RA	AAEL003512-RA	AAEL011979-RA	AAEL017098-RA	AAEL012158-RA	AAEL005688-RA
AAEL009500-RA	AAEL001644-RA	AAEL003530-RA	AAEL017253-RC	AAEL017263-RA	AAEL012154-RB	AAEL004457-RA
AAEL011133-RA	AAEL001656-RA	AAEL007888-RA	AAEL011205-RA	AAEL010778-RA	AAEL005408-RA	AAEL004457-RB
AAEL006452-RA	AAEL002513-RA	AAEL010343-RA	AAEL009043-RA	AAEL010778-RB	AAEL005407-RB	AAEL004457-RC
AAEL013390-RA	AAEL004668-RA	AAEL013184-RA	AAEL010081-RA	AAEL017451-RA	AAEL005407-RC	AAEL013534-RA
AAEL013390-RB	AAEL007209-RA	AAEL013678-RA	AAEL010465-RA	AAEL004869-RA	AAEL017081-RA	AAEL001709-RA
AAEL001607-RA	AAEL006234-RA	AAEL012608-RA	AAEL010819-RA	AAEL009782-RA	AAEL017081-RB	AAEL001708-RB
AAEL001618-RA	AAEL017301-RA	AAEL002848-RA	AAEL010819-RB	AAEL011637-RA	AAEL002203-RA	AAEL001708-RA
AAEL000893-RA	AAEL000643-RA	AAEL008228-RA	AAEL010814-RB	AAEL009846-RA	AAEL014325-RA	AAEL004121-RA
AAEL004688-RA	AAEL004246-RA	AAEL012723-RA	AAEL012265-RA	AAEL012116-RA	AAEL012561-RA	AAEL004121-RB
AAEL007227-RA	AAEL011137-RA	AAEL011688-RA	AAEL007384-RA	AAEL009174-RA	AAEL014365-RA	AAEL011876-RA
AAEL007228-RA	AAEL001910-RA	AAEL012655-RA	AAEL012139-RA	AAEL015454-RA	AAEL014365-RB	AAEL011870-RA
AAEL014309-RA	AAEL010911-RA	AAEL005515-RA	AAEL012631-RA	AAEL007915-RA	AAEL005958-RA	AAEL011081-RB
AAEL014309-RB	AAEL010912-RA	AAEL005515-RC	AAEL001357-RA	AAEL007898-RA	AAEL002464-RA	AAEL011081-RA
AAEL008677-RA	AAEL010576-RJ	AAEL009059-RA	AAEL010877-RA	AAEL009658-RB	AAEL000109-RA	AAEL011087-RA
AAEL008674-RA	AAEL007579-RA	AAEL008654-RA	AAEL012272-RA	AAEL009658-RC	AAEL011989-RB	AAEL001472-RA
AAEL014901-RA	AAEL009880-RA	AAEL012549-RA	AAEL012272-RB	AAEL007160-RA	AAEL008853-RA	AAEL010170-RA
AAEL007799-RA	AAEL010124-RA	AAEL000792-RA	AAEL009367-RA	AAEL013683-RA	AAEL008847-RA	AAEL010170-RB
AAEL007799-RB	AAEL010116-RA	AAEL009357-RA	AAEL014813-RA	AAEL006632-RD	AAEL007018-RA	AAEL010170-RC
AAEL006253-RA	AAEL006267-RA	AAEL012035-RA	AAEL009681-RA	AAEL003024-RA	AAEL006993-RC	AAEL006891-RA
AAEL014114-RA	AAEL012819-RA	AAEL000983-RC	AAEL014738-RA	AAEL003043-RA	AAEL001513-RA	AAEL000454-RA
AAEL011302-RB	AAEL003845-RA	AAEL000824-RA	AAEL009700-RA	AAEL012169-RA	AAEL001516-RB	AAEL000454-RB
AAEL011302-RE	AAEL003834-RA	AAEL000824-RB	AAEL009700-RB	AAEL000081-RA	AAEL004485-RA	AAEL000608-RA
AAEL008921-RA	AAEL003424-RA	AAEL012997-RA	AAEL008078-RA	AAEL004089-RA	AAEL005712-RA	AAEL002550-RA
AAEL008921-RD	AAEL010942-RA	AAEL003972-RA	AAEL012918-RA	AAEL004083-RA	AAEL004930-RA	AAEL002550-RD
AAEL002294-RA	AAEL002578-RA	AAEL003955-RA	AAEL013939-RA	AAEL004114-RA	AAEL003701-RA	AAEL002709-RA
AAEL008351-RA	AAEL002614-RA	AAEL007252-RA	AAEL009932-RA	AAEL011376-RA	AAEL003651-RA	AAEL002885-RA
AAEL005998-RB	AAEL011556-RA	AAEL015143-RB	AAEL003145-RA	AAEL011350-RA	AAEL012960-RA	AAEL003348-RC
AAEL005564-RA	AAEL007718-RA	AAEL000212-RA	AAEL003989-RA	AAEL003285-RA	AAEL002163-RA	AAEL004149-RA
AAEL000595-RA	AAEL008630-RA	AAEL000191-RA	AAEL005140-RA	AAEL007374-RA	AAEL006666-RA	AAEL004972-RB
AAEL000581-RA	AAEL011074-RA	AAEL013461-RA	AAEL010452-RA	AAEL001215-RA	AAEL006675-RA	AAEL005017-RA
AAEL000564-RA	AAEL012397-RA	AAEL004583-RC	AAEL008046-RA	AAEL004301-RA	AAEL001919-RA	AAEL005221-RB
AAEL000564-RB	AAEL000343-RA	AAEL006581-RA	AAEL006330-RA	AAEL003388-RA	AAEL001922-RA	AAEL005901-RB
AAEL013079-RA	AAEL006057-RA	AAEL013510-RA	AAEL006330-RB	AAEL013468-RA	AAEL001921-RA	AAEL005998-RA
AAEL007562-RA	AAEL005337-RA	AAEL014421-RA	AAEL006356-RA	AAEL006731-RA	AAEL013704-RA	AAEL006575-RB
AAEL013138-RA	AAEL013232-RA	AAEL013875-RA	AAEL001787-RA	AAEL006731-RB	AAEL010050-RA	AAEL007029-RA
AAEL013199-RA	AAEL006786-RA	AAEL013875-RB	AAEL001779-RA	AAEL015022-RA	AAEL010844-RA	AAEL007029-RB
AAEL002771-RA	AAEL006786-RB	AAEL007281-RA	AAEL007187-RA	AAEL003739-RA	AAEL007193-RA	AAEL007173-RC
AAEL007439-RA	AAEL006786-RC	AAEL010306-RA	AAEL007173-RA	AAEL008758-RA	AAEL007521-RA	AAEL007521-RB
AAEL007439-RB	AAEL006322-RA	AAEL010308-RA	AAEL007173-RB	AAEL001091-RA	AAEL007532-RA	AAEL008351-RB
AAEL014490-RA	AAEL006322-RB	AAEL003066-RA	AAEL008058-RA	AAEL017161-RA	AAEL011810-RA	AAEL008836-RA
AAEL001569-RA	AAEL005475-RA	AAEL004621-RA	AAEL008058-RB	AAEL010062-RA	AAEL007546-RA	AAEL009949-RA
AAEL001540-RA	AAEL001429-RA	AAEL004623-RA	AAEL008058-RC	AAEL014891-RA	AAEL012215-RA	AAEL012207-RC
AAEL001559-RA	AAEL017300-RA	AAEL015199-RA	AAEL015059-RA	AAEL013863-RA	AAEL006460-RA	AAEL013092-RB
AAEL014949-RA	AAEL009295-RA	AAEL013121-RC	AAEL011240-RA	AAEL007067-RA	AAEL012310-RA	AAEL013614-RA
AAEL003883-RA	AAEL006931-RA	AAEL011702-RA	AAEL011240-RB	AAEL005871-RA	AAEL014771-RA	AAEL013614-RB
AAEL012472-RA	AAEL006931-RC	AAEL008719-RA	AAEL001152-RA	AAEL004042-RA	AAEL011093-RA	
AAEL003464-RA	AAEL009530-RA	AAEL003754-RA	AAEL001185-RA	AAEL004042-RB	AAEL001104-RA	

Table 4.7.3.3

Pita_target

AAEL005676-RA	AAEL004351-RC	AAEL008847-RA	AAEL006581-RA	AAEL013534-RA	AAEL004583-RC	AAEL010379-RA
AAEL011350-RA	AAEL010050-RA	AAEL007915-RA	AAEL000682-RA	AAEL007579-RA	AAEL012472-RA	AAEL004246-RA
AAEL013184-RA	AAEL005407-RB	AAEL001232-RA	AAEL008228-RA	AAEL009174-RA	AAEL009043-RA	AAEL012035-RA
AAEL011302-RB	AAEL005407-RC	AAEL013875-RA	AAEL010124-RA	AAEL009782-RA	AAEL004121-RA	AAEL007888-RA
AAEL011302-RE	AAEL002203-RA	AAEL013875-RB	AAEL009700-RA	AAEL000983-RC	AAEL004121-RB	AAEL008347-RA
AAEL015022-RA	AAEL008058-RA	AAEL003754-RA	AAEL009700-RB	AAEL013683-RA	AAEL004149-RA	AAEL007160-RA
AAEL004301-RA	AAEL008058-RB	AAEL003754-RB	AAEL011412-RA	AAEL017098-RA	AAEL015059-RA	AAEL006891-RA
AAEL003464-RA	AAEL008058-RC	AAEL007374-RA	AAEL000454-RA	AAEL005515-RA	AAEL008836-RA	AAEL011184-RA
AAEL003388-RA	AAEL012997-RA	AAEL011093-RA	AAEL000454-RB	AAEL005515-RC	AAEL006553-RA	AAEL005140-RA
AAEL004114-RA	AAEL013678-RA	AAEL012561-RA	AAEL002294-RA	AAEL005821-RA	AAEL017161-RA	AAEL003145-RA
AAEL009367-RA	AAEL008921-RA	AAEL005337-RA	AAEL002550-RA	AAEL011979-RA	AAEL008853-RA	AAEL001215-RA
AAEL006931-RA	AAEL008921-RD	AAEL008758-RA	AAEL002550-RB	AAEL013138-RA	AAEL012139-RA	AAEL010465-RA
AAEL006931-RC	AAEL012215-RA	AAEL001910-RA	AAEL002550-RC	AAEL005958-RA	AAEL012631-RA	AAEL006267-RA
AAEL002709-RA	AAEL001708-RA	AAEL014421-RA	AAEL002550-RD	AAEL008046-RA	AAEL009658-RC	AAEL011702-RA
AAEL007843-RA	AAEL001708-RB	AAEL002163-RA	AAEL001462-RA	AAEL013704-RA	AAEL000109-RA	AAEL010844-RA
AAEL007018-RA	AAEL003651-RA	AAEL017253-RC	AAEL000893-RA	AAEL006675-RA	AAEL012310-RA	AAEL010306-RA
AAEL006452-RA	AAEL010576-RJ	AAEL002513-RA	AAEL001540-RA	AAEL013390-RA	AAEL001709-RA	AAEL010814-RB
AAEL003024-RA	AAEL004930-RA	AAEL001152-RA	AAEL015454-RA	AAEL013390-RB	AAEL010081-RA	AAEL002885-RA
AAEL012360-RA	AAEL010912-RA	AAEL014053-RA	AAEL005408-RA	AAEL001185-RA	AAEL000720-RA	AAEL009295-RA
AAEL005221-RB	AAEL009530-RA	AAEL001644-RA	AAEL006322-RA	AAEL001921-RA	AAEL015122-RA	AAEL002848-RA
AAEL003834-RA	AAEL015143-RB	AAEL011989-RB	AAEL006322-RB	AAEL003989-RA	AAEL005998-RA	AAEL001091-RA
AAEL006057-RA	AAEL003066-RA	AAEL006460-RA	AAEL008654-RA	AAEL013092-RA	AAEL005998-RB	AAEL014891-RA
AAEL004351-RA	AAEL007532-RA	AAEL013614-RA	AAEL000608-RA	AAEL013092-RB	AAEL011556-RA	AAEL005564-RA
AAEL004351-RB	AAEL009949-RA	AAEL013614-RB	AAEL003972-RA	AAEL005795-RA	AAEL012549-RA	AAEL005017-RA

Table 4.7.3.4

Rnahybrid_target

AAEL001050-RA	AAEL002834-RA	AAEL001827-RA	AAEL006731-RA	AAEL014367-RA	AAEL001100-RA	AAEL002834-RB
AAEL001618-RA	AAEL010653-RA	AAEL001241-RA	AAEL006731-RB	AAEL005958-RA	AAEL001100-RB	AAEL002935-RB
AAEL001466-RA	AAEL009059-RA	AAEL008320-RA	AAEL015022-RA	AAEL002464-RA	AAEL001100-RC	AAEL008719-RA
AAEL004266-RA	AAEL000988-RA	AAEL010520-RA	AAEL004063-RA	AAEL008852-RA	AAEL009308-RA	AAEL010340-RB
AAEL010576-RJ	AAEL000181-RA	AAEL004102-RA	AAEL006386-RA	AAEL008175-RA	AAEL013979-RB	
AAEL003834-RA	AAEL013620-RA	AAEL007375-RA	AAEL006386-RB	AAEL002436-RA	AAEL013979-RC	
AAEL002935-RA	AAEL003083-RA	AAEL007375-RD	AAEL004165-RA	AAEL002436-RB	AAEL013979-RD	
AAEL009530-RA	AAEL005676-RA	AAEL004301-RA	AAEL017232-RA	AAEL002436-RC	AAEL002392-RA	

Table 4.7.3.5

Targetscan_target

AAEL000109-RA	AAEL002194-RB	AAEL004276-RA	AAEL006809-RA	AAEL009174-RA	AAEL011302-RE	AAEL013635-RA
AAEL000222-RA	AAEL002203-RA	AAEL004457-RA	AAEL006895-RA	AAEL009367-RA	AAEL011309-RA	AAEL013678-RA
AAEL000304-RA	AAEL002280-RA	AAEL004457-RB	AAEL006929-RA	AAEL009380-RA	AAEL011314-RA	AAEL013700-RA
AAEL000304-RB	AAEL002318-RA	AAEL004457-RC	AAEL007184-RA	AAEL009380-RC	AAEL011739-RA	AAEL013771-RA
AAEL000304-RC	AAEL002318-RB	AAEL004484-RA	AAEL007191-RA	AAEL009427-RA	AAEL011739-RB	AAEL013771-RB
AAEL000304-RD	AAEL002390-RA	AAEL004500-RA	AAEL007292-RA	AAEL009510-RA	AAEL011748-RA	AAEL013864-RA
AAEL000304-RE	AAEL002411-RB	AAEL004500-RB	AAEL007322-RA	AAEL009607-RB	AAEL011748-RB	AAEL013870-RA
AAEL000339-RA	AAEL002436-RA	AAEL004559-RA	AAEL007402-RA	AAEL009682-RA	AAEL011781-RA	AAEL013875-RA
AAEL000343-RA	AAEL002436-RB	AAEL004583-RC	AAEL007427-RA	AAEL009785-RA	AAEL011901-RA	AAEL013875-RB
AAEL000378-RA	AAEL002436-RC	AAEL004583-RD	AAEL007427-RC	AAEL009825-RA	AAEL011993-RB	AAEL013979-RA
AAEL000454-RA	AAEL002469-RA	AAEL004621-RA	AAEL007439-RA	AAEL009887-RA	AAEL011998-RA	AAEL013979-RE
AAEL000454-RB	AAEL002469-RB	AAEL004623-RA	AAEL007439-RB	AAEL009936-RA	AAEL012139-RA	AAEL014053-RA
AAEL000569-RA	AAEL002557-RA	AAEL004691-RA	AAEL007494-RA	AAEL009955-RA	AAEL012175-RA	AAEL014054-RA
AAEL000682-RA	AAEL002625-RA	AAEL004743-RB	AAEL007562-RA	AAEL010062-RA	AAEL012209-RA	AAEL014178-RA
AAEL000713-RA	AAEL002723-RA	AAEL005019-RA	AAEL007881-RA	AAEL010116-RB	AAEL012262-RA	AAEL014309-RA
AAEL000767-RB	AAEL002759-RD	AAEL005175-RA	AAEL008131-RA	AAEL010116-RC	AAEL012262-RC	AAEL014309-RB
AAEL001069-RA	AAEL002759-RA	AAEL005476-RA	AAEL008347-RA	AAEL010143-RA	AAEL012262-RD	AAEL014318-RA
AAEL001100-RA	AAEL002759-RB	AAEL005481-RA	AAEL008351-RA	AAEL010308-RA	AAEL012262-RB	AAEL014825-RA
AAEL001100-RB	AAEL002759-RC	AAEL005485-RA	AAEL008351-RB	AAEL010348-RA	AAEL012443-RA	AAEL015010-RA
AAEL001100-RC	AAEL002860-RA	AAEL005493-RA	AAEL008381-RA	AAEL010382-RA	AAEL012515-RA	AAEL015059-RA
AAEL001112-RA	AAEL002862-RA	AAEL005515-RA	AAEL008497-RA	AAEL010423-RA	AAEL012608-RA	AAEL015065-RA
AAEL001232-RA	AAEL002885-RA	AAEL005515-RC	AAEL008502-RA	AAEL010470-RA	AAEL012655-RA	AAEL015300-RA
AAEL001252-RA	AAEL002886-RA	AAEL005544-RA	AAEL008556-RA	AAEL010509-RA	AAEL012819-RA	AAEL017081-RA
AAEL001293-RC	AAEL002886-RB	AAEL005704-RA	AAEL008680-RA	AAEL010610-RA	AAEL012832-RA	AAEL017081-RB
AAEL001352-RA	AAEL002891-RA	AAEL006002-RA	AAEL008701-RA	AAEL010691-RA	AAEL012832-RB	AAEL017098-RA
AAEL001381-RA	AAEL003022-RB	AAEL006057-RA	AAEL008723-RA	AAEL010704-RA	AAEL012996-RA	AAEL017300-RA
AAEL001411-RA	AAEL003083-RA	AAEL006231-RA	AAEL008723-RB	AAEL010778-RA	AAEL013092-RA	AAEL017400-RA
AAEL001411-RB	AAEL003154-RA	AAEL006234-RA	AAEL008828-RA	AAEL010778-RB	AAEL013092-RB	AAEL017418-RA
AAEL001472-RA	AAEL003193-RA	AAEL006249-RA	AAEL008853-RA	AAEL010784-RA	AAEL013111-RA	
AAEL001499-RA	AAEL003203-RA	AAEL006267-RA	AAEL008859-RA	AAEL010787-RA	AAEL013220-RA	
AAEL001516-RA	AAEL003203-RB	AAEL006364-RA	AAEL008871-RA	AAEL010787-RB	AAEL013252-RB	
AAEL001516-RB	AAEL003325-RA	AAEL006478-RA	AAEL008928-RA	AAEL010787-RD	AAEL013252-RC	
AAEL001656-RA	AAEL003355-RA	AAEL006489-RA	AAEL008928-RB	AAEL010809-RA	AAEL013314-RA	
AAEL001752-RB	AAEL003492-RA	AAEL006489-RB	AAEL008928-RC	AAEL010814-RB	AAEL013341-RA	
AAEL001765-RA	AAEL003705-RA	AAEL006529-RA	AAEL009041-RB	AAEL010819-RA	AAEL013341-RB	
AAEL001779-RA	AAEL003739-RA	AAEL006531-RA	AAEL009041-RA	AAEL010819-RB	AAEL013353-RA	
AAEL001799-RA	AAEL003834-RA	AAEL006531-RB	AAEL009041-RC	AAEL010905-RA	AAEL013353-RC	
AAEL001946-RA	AAEL003968-RA	AAEL006582-RA	AAEL009041-RD	AAEL011133-RA	AAEL013353-RB	
AAEL002107-RA	AAEL004089-RA	AAEL006582-RB	AAEL009057-RA	AAEL011180-RA	AAEL013353-RD	
AAEL002123-RA	AAEL004178-RA	AAEL006659-RA	AAEL009059-RA	AAEL011242-RA	AAEL013386-RA	
AAEL002194-RA	AAEL004178-RB	AAEL006791-RA	AAEL009092-RA	AAEL011302-RB	AAEL013490-RA	

Anopheles gambiae results

Table 4.7.3.6

In-house targets

AGAP005747-RA	AGAP001650-RA	AGAP003212-RA	AGAP004159-RA	AGAP007791-RB	AGAP008310-RB	AGAP012956-RB
AGAP005747-RB	AGAP002030-RA	AGAP003243-RA	AGAP004159-RB	AGAP007791-RC	AGAP008310-RD	AGAP000128-RA
AGAP006115-RB	AGAP002224-RA	AGAP003389-RA	AGAP004372-RA	AGAP007791-RA	AGAP008310-RF	AGAP000699-RA
AGAP006115-RA	AGAP013481-RD	AGAP003485-RA	AGAP004584-RA	AGAP007791-RD	AGAP008310-RA	AGAP000699-RB
AGAP006733-RA	AGAP002527-RA	AGAP003513-RA	AGAP004603-RA	AGAP007791-RE	AGAP008641-RA	AGAP000717-RA
AGAP006830-RA	AGAP002728-RB	AGAP003586-RA	AGAP004631-RA	AGAP007896-RA	AGAP008930-RA	AGAP000870-RA
AGAP007029-RA	AGAP002728-RA	AGAP003722-RA	AGAP010561-RA	AGAP008017-RB	AGAP009305-RB	AGAP000927-RB
AGAP007309-RA	AGAP002883-RA	AGAP003867-RA	AGAP010598-RA	AGAP008017-RA	AGAP010062-RA	AGAP000927-RA
AGAP007391-RA	AGAP003021-RB	AGAP003921-RA	AGAP010887-RA	AGAP008017-RC	AGAP010145-RA	AGAP001886-RA
AGAP007684-RA	AGAP003021-RA	AGAP003940-RA	AGAP011334-RA	AGAP008020-RA	AGAP012934-RA	AGAP004833-RA
AGAP007684-RB	AGAP003038-RA	AGAP003968-RA	AGAP011426-RA	AGAP008196-RA	AGAP000021-RA	AGAP005549-RA
AGAP012990-RA	AGAP003191-RA	AGAP004059-RA	AGAP012334-RA	AGAP008304-RA	AGAP012956-RA	

Table 4.7.3.7

miRanda targets

AGAP004725-RA	AGAP007160-RA	AGAP002379-RA	AGAP003685-RA	AGAP011341-RA	AGAP008113-RA	AGAP012934-RA
AGAP004765-RA	AGAP007160-RB	AGAP013481-RD	AGAP003722-RA	AGAP011426-RA	AGAP013730-RA	AGAP000021-RA
AGAP004768-RB	AGAP007160-RC	AGAP013481-RE	AGAP003860-RA	AGAP011532-RA	AGAP008184-RA	AGAP012956-RA
AGAP004768-RA	AGAP007309-RA	AGAP002509-RA	AGAP003860-RB	AGAP012140-RA	AGAP008184-RB	AGAP012956-RB
AGAP004890-RA	AGAP007361-RA	AGAP002527-RA	AGAP003860-RC	AGAP012177-RA	AGAP008196-RA	AGAP000128-RA
AGAP004890-RB	AGAP007391-RA	AGAP002564-RA	AGAP003867-RA	AGAP012334-RA	AGAP008264-RA	AGAP000304-RB
AGAP005046-RB	AGAP007503-RA	AGAP002728-RB	AGAP003921-RA	AGAP012413-RA	AGAP008304-RA	AGAP000304-RA
AGAP005132-RA	AGAP007684-RA	AGAP002728-RA	AGAP003940-RA	AGAP007737-RA	AGAP008310-RB	AGAP000304-RC
AGAP005263-RA	AGAP007684-RB	AGAP002767-RA	AGAP003968-RA	AGAP007740-RA	AGAP008310-RD	AGAP000412-RA
AGAP005405-RA	AGAP001196-RA	AGAP002767-RB	AGAP003984-RA	AGAP007791-RB	AGAP008310-RF	AGAP000550-RA
AGAP005747-RA	AGAP001217-RA	AGAP002883-RA	AGAP004031-RA	AGAP007791-RC	AGAP008310-RA	AGAP000699-RA
AGAP005747-RB	AGAP001267-RA	AGAP002884-RA	AGAP004059-RA	AGAP007791-RA	AGAP008321-RA	AGAP000699-RB
AGAP006115-RB	AGAP001477-RA	AGAP003021-RB	AGAP004159-RA	AGAP007791-RD	AGAP008641-RA	AGAP000717-RA
AGAP006115-RA	AGAP012990-RA	AGAP003021-RA	AGAP004159-RB	AGAP007791-RE	AGAP008895-RA	AGAP000870-RA
AGAP006148-RA	AGAP001650-RA	AGAP003028-RA	AGAP004197-RA	AGAP007827-RA	AGAP008921-RA	AGAP000927-RB
AGAP006328-RA	AGAP001713-RD	AGAP003038-RA	AGAP004294-RA	AGAP007827-RB	AGAP008930-RA	AGAP000927-RA
AGAP006389-RA	AGAP001713-RA	AGAP003191-RA	AGAP004372-RA	AGAP007827-RC	AGAP028187-RA	AGAP001030-RA
AGAP006569-RA	AGAP001713-RB	AGAP003212-RA	AGAP004422-RA	AGAP007896-RA	AGAP009305-RA	AGAP001886-RA
AGAP006609-RA	AGAP001713-RC	AGAP003243-RA	AGAP004502-RA	AGAP013764-RA	AGAP009305-RB	AGAP002356-RA
AGAP006645-RA	AGAP001730-RA	AGAP003372-RA	AGAP004584-RA	AGAP007928-RA	AGAP009650-RA	AGAP004833-RA
AGAP006645-RB	AGAP001731-RA	AGAP003372-RB	AGAP004603-RA	AGAP007942-RA	AGAP009745-RA	AGAP005549-RA
AGAP006645-RC	AGAP002030-RA	AGAP003389-RA	AGAP004631-RA	AGAP007942-RB	AGAP009976-RA	AGAP007934-RB
AGAP006645-RD	AGAP002076-RA	AGAP003485-RA	AGAP010389-RA	AGAP008016-RA	AGAP010062-RA	AGAP010716-RA
AGAP006733-RA	AGAP002093-RA	AGAP003513-RA	AGAP010561-RA	AGAP008017-RB	AGAP010135-RA	AGAP013481-RA
AGAP006780-RA	AGAP002182-RA	AGAP003583-RA	AGAP010598-RA	AGAP008017-RA	AGAP010145-RA	AGAP013481-RC
AGAP006830-RA	AGAP002224-RA	AGAP003586-RA	AGAP010887-RA	AGAP008017-RC	AGAP010211-RA	AGAP013481-RF
AGAP007029-RA	AGAP002286-RA	AGAP003596-RA	AGAP011318-RA	AGAP008020-RA	AGAP010260-RA	AGAP013481-RG
AGAP007041-RA	AGAP002357-RA	AGAP003598-RA	AGAP011334-RA	AGAP008113-RB	AGAP010276-RB	

Table 4.7.3.8

Pita targets

AGAP003940-RA	AGAP002527-RA	AGAP008641-RA	AGAP008196-RA	AGAP000128-RA	AGAP007029-RA	AGAP008310-RB
AGAP008304-RA	AGAP008017-RA	AGAP003389-RA	AGAP003513-RA	AGAP013481-RD	AGAP008930-RA	AGAP008310-RD
AGAP003191-RA	AGAP008017-RB	AGAP000699-RA	AGAP001650-RA	AGAP007309-RA	AGAP007391-RA	AGAP008310-RF
AGAP003485-RA	AGAP008017-RC	AGAP000699-RB	AGAP003212-RA	AGAP003921-RA	AGAP007684-RA	AGAP011334-RA
AGAP009305-RB	AGAP004159-RA	AGAP004059-RA	AGAP002728-RA	AGAP006115-RA	AGAP007684-RB	AGAP012334-RA
AGAP000927-RA	AGAP004159-RB	AGAP000870-RA	AGAP002728-RB	AGAP006115-RB	AGAP008020-RA	AGAP003021-RA
AGAP000927-RB	AGAP010062-RA	AGAP006830-RA	AGAP010598-RA	AGAP010145-RA	AGAP010561-RA	AGAP003021-RB
AGAP004631-RA	AGAP003867-RA	AGAP002224-RA	AGAP007791-RA	AGAP006733-RA	AGAP002883-RA	AGAP003586-RA
AGAP003038-RA	AGAP003968-RA	AGAP012956-RA	AGAP007791-RB	AGAP004584-RA	AGAP003243-RA	AGAP004372-RA
AGAP003722-RA	AGAP010887-RA	AGAP012956-RB	AGAP007791-RC	AGAP005747-RA	AGAP002030-RA	
AGAP001886-RA	AGAP005549-RA	AGAP004603-RA	AGAP007791-RD	AGAP005747-RB	AGAP012990-RA	
AGAP000717-RA	AGAP007896-RA	AGAP004833-RA	AGAP007791-RE	AGAP000021-RA	AGAP008310-RA	

Table 4.7.3.9

RNAHybrid targets

AGAP004698-RA	AGAP006046-RA	AGAP007713-RB	AGAP002332-RA	AGAP004069-RA	AGAP008881-RA	AGAP012802-RA
AGAP005046-RB	AGAP006610-RA	AGAP007713-RA	AGAP002969-RA	AGAP004246-RA	AGAP008916-RA	AGAP000734-RA
AGAP005153-RA	AGAP007163-RA	AGAP001138-RA	AGAP003028-RA	AGAP004292-RA	AGAP009020-RA	AGAP000870-RA
AGAP005153-RB	AGAP007163-RB	AGAP001595-RA	AGAP003366-RA	AGAP011374-RA	AGAP009077-RA	AGAP007934-RB
AGAP005226-RA	AGAP007529-RA	AGAP001645-RA	AGAP003583-RA	AGAP012185-RA	AGAP009274-RB	AGAP011305-RA
AGAP005545-RA	AGAP007713-RD	AGAP001972-RA	AGAP003746-RA	AGAP012397-RA	AGAP009274-RA	AGAP007565-RA
AGAP005802-RA	AGAP007713-RC	AGAP002035-RA	AGAP003796-RA	AGAP008264-RA	AGAP010187-RA	AGAP003940-RA

Table 4.7.3.10

TargetScan targets

AGAP003350-RB	AGAP003429-RB	AGAP005990-RA	AGAP002968-RA	AGAP001935-RB	AGAP006532-RA	AGAP004239-RA
AGAP003350-RA	AGAP002473-RA	AGAP003181-RA	AGAP003921-RA	AGAP001935-RA	AGAP008684-RA	AGAP003860-RA
AGAP012533-RA	AGAP004766-RB	AGAP002081-RA	AGAP007668-RA	AGAP011352-RA	AGAP003464-RA	AGAP003860-RB
AGAP010479-RA	AGAP004766-RD	AGAP008184-RA	AGAP007642-RA	AGAP004737-RA	AGAP000949-RA	AGAP003860-RC
AGAP006573-RA	AGAP007520-RA	AGAP008184-RB	AGAP001966-RA	AGAP003937-RB	AGAP002509-RA	AGAP002356-RA
AGAP002728-RB	AGAP013481-RB	AGAP009992-RA	AGAP002687-RA	AGAP003937-RC	AGAP010175-RB	AGAP007713-RD
AGAP002728-RA	AGAP013481-RD	AGAP011341-RA	AGAP009896-RA	AGAP003937-RA	AGAP010175-RA	AGAP007713-RC
AGAP004283-RA	AGAP001435-RA	AGAP003809-RB	AGAP010276-RB	AGAP013293-RA	AGAP010175-RC	AGAP007713-RB
AGAP002929-RA	AGAP007130-RA	AGAP003809-RC	AGAP003742-RA	AGAP002364-RB	AGAP010175-RD	AGAP007713-RA
AGAP008931-RA	AGAP008143-RA	AGAP003809-RA	AGAP003742-RB	AGAP002364-RA	AGAP006935-RA	AGAP011190-RA
AGAP007684-RA	AGAP011938-RA	AGAP006101-RA	AGAP003738-RA	AGAP004098-RA	AGAP006935-RB	AGAP002647-RA
AGAP007684-RB	AGAP011426-RA	AGAP006235-RA	AGAP007896-RA	AGAP004098-RB	AGAP006935-RC	AGAP009305-RA
AGAP005263-RA	AGAP011334-RA	AGAP010587-RA	AGAP002011-RA	AGAP004963-RA	AGAP012960-RA	AGAP009091-RA
AGAP010211-RA	AGAP001373-RA	AGAP003717-RA	AGAP003968-RA	AGAP007382-RA	AGAP002655-RA	AGAP009091-RB
AGAP009156-RA	AGAP007348-RB	AGAP001110-RA	AGAP000750-RA	AGAP007786-RA	AGAP010356-RA	AGAP000769-RB
AGAP003857-RA	AGAP004892-RA	AGAP001110-RB	AGAP000717-RA	AGAP008294-RA	AGAP002767-RA	AGAP000769-RA
AGAP006441-RB	AGAP007249-RA	AGAP008598-RA	AGAP001477-RA	AGAP002622-RA	AGAP002767-RB	AGAP007887-RA
AGAP006441-RA	AGAP028023-RA	AGAP002557-RA	AGAP009464-RA	AGAP002247-RA	AGAP007667-RA	AGAP006569-RA
AGAP010728-RA	AGAP009537-RA	AGAP007538-RA	AGAP010716-RA	AGAP003598-RA	AGAP004492-RA	AGAP007867-RA
AGAP003430-RA	AGAP009616-RA	AGAP002017-RA	AGAP002935-RA	AGAP002359-RA	AGAP007834-RA	AGAP004631-RA
AGAP003429-RC	AGAP004136-RA	AGAP009486-RA	AGAP001070-RA	AGAP005405-RA	AGAP007162-RB	
AGAP003429-RA	AGAP011532-RA	AGAP007925-RA	AGAP005245-RD	AGAP010062-RA	AGAP007162-RA	

Table 4.7.3.11 Final list of miR-8 target genes

Gene ID	Gene name	Species	Algorithm*
AAEL003834-RA	Metalloproteinase, putative	*Aedes*	IN; TS; PITA; R; RH
AAEL000454-RA; RB	Isocitrate dehydrogenase	*Aedes; Anopheles*	IN; TS; PITA; R
AAEL001232-RA	Tubulointerstitial nephritis antigen[†]	*Aedes; Anopheles*	IN; TS; PITA; R
AAEL008347-RA	Monocarboxylate transporter	*Aedes; Anopheles*	IN; TS; PITA; R
AAEL005958-RA	Oxidoreductase	*Aedes*	IN; PITA; R; RH
AAEL009059-RA	arp2/3 complex 16 kd subunit (P16-arc)	*Aedes*	IN; PITA; R; RH
AAEL009530-RA	tmc6 protein (evin)	*Aedes*	IN; PITA; R; RH

*Algorithms used include in-house (IN), TargetScan (TS), PITA; miRanda (R), and RNAhybrid (RH).
[†]Homolog to *Drosophila* Swim.

Table 4.7.4 miRNA-275 target prediction results from five programs (4.7.4.1-4.7.4.10)

Table 4.7.4.1: Target prediction results from In-house program in Ae. aegypti
Table 4.7.4.2: Target prediction results from miRanda program in Ae. Aegypti
Table 4.7.4.3: Target prediction results from PITA program in Ae. Aegypti
Table 4.7.4.4: Target prediction results from TargetScan program in Ae. Aegypti
Table 4.7.4.5: Target prediction results from RNAhybrid program in Ae. aegypti
Table 4.7.4.6: Target prediction results from In-house program in An. gambiae
Table 4.7.4.7: Target prediction results from miRanda program in An. gambiae
Table 4.7.4.8: Target prediction results from PITA program in An. gambiae
Table 4.7.4.9: Target prediction results from TargetScan program in An. gambiae
Table 4.7.4.10: Target prediction results from RNAhybrid program in An. gambiae
Table 4.7.4.11 Final list of miR-8 target genes

4.7.4 miR-275 results

Aedes aegypti results

Table 4.7.4.1

In-house targets

AAEL000577-RA	AAEL002714-RA	AAEL005191-RA	AAEL006582-RA	AAEL007883-RA	AAEL011901-RA	AAEL014819-RA
AAEL000964-RA	AAEL003651-RA	AAEL006126-RA	AAEL006582-RB	AAEL007883-RB	AAEL012215-RA	AAEL017232-RA
AAEL001796-RA	AAEL004178-RA	AAEL006126-RB	AAEL006834-RA	AAEL007883-RC	AAEL013092-RB	AAEL017240-RB
AAEL002078-RA	AAEL004178-RB	AAEL006532-RA	AAEL007476-RA	AAEL008738-RA	AAEL013955-RA	AAEL017481-RA

Table 4.7.4.2

miRANDA targets

AAEL000077-RA	AAEL002597-RA	AAEL005252-RA	AAEL006320-RA	AAEL008569-RA	AAEL011092-RA	AAEL014755-RA
AAEL000088-RA	AAEL002597-RB	AAEL005324-RA	AAEL006532-RA	AAEL008607-RA	AAEL011872-RA	AAEL014819-RA
AAEL000343-RA	AAEL002714-RA	AAEL005569-RA	AAEL006582-RA	AAEL008674-RA	AAEL011901-RA	AAEL017232-RA
AAEL000577-RA	AAEL002791-RA	AAEL006001-RA	AAEL006582-RB	AAEL008738-RA	AAEL012081-RA	AAEL017240-RB
AAEL000815-RA	AAEL002791-RB	AAEL006001-RB	AAEL006834-RA	AAEL009154-RA	AAEL012184-RA	AAEL017481-RA
AAEL000964-RA	AAEL003651-RA	AAEL006001-RC	AAEL007150-RA	AAEL009154-RB	AAEL012215-RA	AAEL017508-RA
AAEL001056-RA	AAEL004178-RA	AAEL006001-RD	AAEL007476-RA	AAEL009484-RA	AAEL013092-RB	AAEL017508-RB
AAEL001190-RA	AAEL004178-RB	AAEL006057-RA	AAEL007717-RA	AAEL009484-RB	AAEL013656-RA	
AAEL001252-RA	AAEL005092-RA	AAEL006126-RA	AAEL007883-RA	AAEL010032-RA	AAEL013700-RA	
AAEL001796-RA	AAEL005191-RA	AAEL006126-RB	AAEL007883-RB	AAEL010167-RA	AAEL013955-RA	
AAEL002078-RA	AAEL005217-RA	AAEL006240-RA	AAEL007883-RC	AAEL011006-RA	AAEL014026-RA	

Table 4.7.4.3

PITA targets

AAEL000577-RA	AAEL002714-RA	AAEL006126-RB	AAEL006834-RA	AAEL007883-RB	AAEL011901-RA	AAEL013955-RA
AAEL000964-RA	AAEL003651-RA	AAEL006532-RA	AAEL007476-RA	AAEL007883-RC	AAEL012215-RA	AAEL017240-RB
AAEL001796-RA	AAEL005191-RA	AAEL006582-RA	AAEL007883-RA	AAEL008738-RA	AAEL013092-RB	AAEL017481-RA
AAEL002078-RA	AAEL006126-RA	AAEL006582-RB				

Table 4.7.4.4

TargetScan targets

AAEL000566-RA	AAEL003216-RA	AAEL005704-RA	AAEL006879-RA	AAEL008719-RA	AAEL011444-RA	AAEL013207-RA
AAEL000577-RA	AAEL003326-RA	AAEL005727-RA	AAEL007029-RA	AAEL008928-RA	AAEL011518-RA	AAEL013410-RA
AAEL000848-RA	AAEL003454-RA	AAEL005733-RA	AAEL007029-RB	AAEL008928-RC	AAEL011739-RA	AAEL013530-RA
AAEL001360-RA	AAEL003769-RB	AAEL005733-RB	AAEL007151-RB	AAEL009148-RA	AAEL011739-RB	AAEL013635-RA
AAEL001360-RB	AAEL003873-RA	AAEL005832-RA	AAEL007151-RC	AAEL009295-RA	AAEL011978-RA	AAEL014702-RA
AAEL001394-RA	AAEL003873-RB	AAEL006028-RA	AAEL007151-RE	AAEL009380-RB	AAEL011993-RA	AAEL015065-RA
AAEL001642-RA	AAEL004040-RA	AAEL006032-RA	AAEL007288-RB	AAEL009380-RC	AAEL011993-RD	AAEL015099-RD
AAEL001642-RB	AAEL004049-RA	AAEL006126-RA	AAEL007290-RB	AAEL009392-RA	AAEL012055-RA	AAEL017135-RA
AAEL001913-RA	AAEL004378-RA	AAEL006126-RB	AAEL007454-RA	AAEL009874-RA	AAEL012055-RB	AAEL017293-RA
AAEL001982-RA	AAEL004471-RA	AAEL006181-RA	AAEL007787-RA	AAEL010159-RA	AAEL012154-RB	
AAEL002700-RA	AAEL004623-RA	AAEL006582-RA	AAEL007826-RA	AAEL010430-RA	AAEL012262-RA	
AAEL002700-RB	AAEL004808-RA	AAEL006582-RB	AAEL007841-RA	AAEL010834-RA	AAEL012262-RD	
AAEL003143-RA	AAEL005140-RA	AAEL006721-RA	AAEL008391-RA	AAEL010975-RA	AAEL012686-RA	
AAEL003154-RA	AAEL005410-RA	AAEL006809-RA	AAEL008620-RA	AAEL011309-RA	AAEL012944-RB	

Table 4.7.4.5

RNAhybrid targets

AAEL000010-RA	AAEL001580-RA	AAEL003716-RA	AAEL006544-RA	AAEL008928-RB	AAEL010896-RA	AAEL013278-RA
AAEL000010-RB	AAEL001798-RA	AAEL003716-RB	AAEL006566-RA	AAEL008947-RA	AAEL011126-RA	AAEL013408-RB
AAEL000034-RA	AAEL001798-RB	AAEL003743-RA	AAEL006575-RC	AAEL008954-RA	AAEL011340-RB	AAEL013487-RA
AAEL000068-RA	AAEL001830-RA	AAEL003864-RA	AAEL006582-RA	AAEL009048-RA	AAEL011527-RA	AAEL013531-RA
AAEL000068-RB	AAEL001859-RA	AAEL003871-RA	AAEL006582-RB	AAEL009379-RA	AAEL011575-RA	AAEL013536-RA
AAEL000080-RA	AAEL002014-RA	AAEL004071-RA	AAEL006660-RA	AAEL009389-RA	AAEL011684-RA	AAEL013704-RA
AAEL000138-RA	AAEL002055-RA	AAEL004097-RA	AAEL006733-RA	AAEL009451-RA	AAEL011701-RA	AAEL013783-RA
AAEL000257-RA	AAEL002194-RA	AAEL004131-RA	AAEL006833-RA	AAEL009484-RA	AAEL011717-RA	AAEL013783-RB
AAEL000257-RB	AAEL002194-RB	AAEL004347-RA	AAEL006833-RB	AAEL009484-RB	AAEL011739-RA	AAEL013979-RB
AAEL000276-RA	AAEL002211-RA	AAEL004364-RA	AAEL006833-RC	AAEL009658-RA	AAEL011739-RB	AAEL013979-RC
AAEL000301-RA	AAEL002310-RA	AAEL004526-RA	AAEL006980-RA	AAEL009731-RA	AAEL011778-RA	AAEL013979-RD
AAEL000395-RA	AAEL002450-RA	AAEL004554-RC	AAEL007056-RA	AAEL009764-RA	AAEL011872-RA	AAEL013994-RA
AAEL000395-RB	AAEL002474-RA	AAEL004583-RB	AAEL007102-RA	AAEL009764-RB	AAEL011881-RA	AAEL014222-RA
AAEL000404-RA	AAEL002483-RA	AAEL004623-RA	AAEL007150-RA	AAEL009772-RB	AAEL011925-RA	AAEL014490-RA
AAEL000418-RA	AAEL002488-RA	AAEL004694-RA	AAEL007203-RA	AAEL009775-RA	AAEL011936-RA	AAEL014524-RA
AAEL000509-RA	AAEL002714-RA	AAEL004740-RA	AAEL007638-RA	AAEL009825-RA	AAEL012055-RA	AAEL014536-RA
AAEL000575-RA	AAEL002781-RA	AAEL004839-RA	AAEL007760-RA	AAEL009859-RA	AAEL012055-RB	AAEL014819-RA
AAEL000605-RA	AAEL002791-RA	AAEL004854-RA	AAEL007771-RA	AAEL009883-RA	AAEL012081-RA	AAEL014903-RA
AAEL000693-RA	AAEL002808-RA	AAEL005037-RA	AAEL007771-RB	AAEL009944-RA	AAEL012156-RA	AAEL015022-RA
AAEL000746-RA	AAEL002903-RA	AAEL005140-RA	AAEL007848-RA	AAEL010030-RA	AAEL012184-RA	AAEL017059-RA
AAEL001158-RA	AAEL002912-RA	AAEL005191-RA	AAEL007877-RA	AAEL010037-RA	AAEL012283-RA	AAEL017213-RA
AAEL001158-RB	AAEL002972-RA	AAEL005269-RA	AAEL007939-RA	AAEL010047-RA	AAEL012328-RA	AAEL017232-RA
AAEL001201-RA	AAEL002995-RA	AAEL005269-RB	AAEL007943-RA	AAEL010159-RA	AAEL012562-RA	AAEL017258-RA
AAEL001294-RA	AAEL003116-RA	AAEL005269-RC	AAEL008025-RA	AAEL010226-RA	AAEL012629-RA	AAEL017397-RA
AAEL001294-RB	AAEL003184-RB	AAEL005466-RA	AAEL008216-RA	AAEL010251-RA	AAEL012683-RA	AAEL017503-RA
AAEL001324-RA	AAEL003205-RA	AAEL005742-RA	AAEL008227-RA	AAEL010338-RA	AAEL012686-RA	AAEL017571-RA
AAEL001466-RA	AAEL003216-RA	AAEL005785-RA	AAEL008248-RA	AAEL010340-RA	AAEL012801-RA	
AAEL001472-RA	AAEL003285-RA	AAEL005886-RA	AAEL008373-RA	AAEL010496-RA	AAEL012880-RA	
AAEL001482-RA	AAEL003396-RA	AAEL006002-RB	AAEL008388-RA	AAEL010576-RN	AAEL012904-RA	
AAEL001501-RA	AAEL003418-RA	AAEL006053-RA	AAEL008468-RA	AAEL010598-RA	AAEL013009-RA	
AAEL001516-RB	AAEL003509-RA	AAEL006249-RA	AAEL008485-RA	AAEL010742-RA	AAEL013137-RA	
AAEL001554-RA	AAEL003530-RA	AAEL006298-RA	AAEL008734-RA	AAEL010827-RA	AAEL013228-RA	

Anopheles gambiae results

Table 4.7.4.6
In-house targets

AGAP005747-RA	AGAP001650-RA	AGAP003212-RA	AGAP004159-RA	AGAP007791-RB	AGAP008310-RB	AGAP012956-RB
AGAP005747-RB	AGAP002030-RA	AGAP003243-RA	AGAP004159-RB	AGAP007791-RC	AGAP008310-RD	AGAP000128-RA
AGAP006115-RB	AGAP002224-RA	AGAP003389-RA	AGAP004372-RA	AGAP007791-RA	AGAP008310-RF	AGAP000699-RA
AGAP006115-RA	AGAP013481-RD	AGAP003485-RA	AGAP004584-RA	AGAP007791-RD	AGAP008310-RA	AGAP000699-RB
AGAP006733-RA	AGAP002527-RA	AGAP003513-RA	AGAP004603-RA	AGAP007791-RE	AGAP008641-RA	AGAP000717-RA
AGAP006830-RA	AGAP002728-RB	AGAP003586-RA	AGAP004631-RA	AGAP007896-RA	AGAP008930-RA	AGAP000870-RA
AGAP007029-RA	AGAP002728-RA	AGAP003722-RA	AGAP010561-RA	AGAP008017-RB	AGAP009305-RB	AGAP000927-RB
AGAP007309-RA	AGAP002883-RA	AGAP003867-RA	AGAP010598-RA	AGAP008017-RA	AGAP010062-RA	AGAP000927-RA
AGAP007391-RA	AGAP003021-RB	AGAP003921-RA	AGAP010887-RA	AGAP008017-RC	AGAP010145-RA	AGAP001886-RA
AGAP007684-RA	AGAP003021-RA	AGAP003940-RA	AGAP011334-RA	AGAP008020-RA	AGAP012934-RA	AGAP004833-RA
AGAP007684-RB	AGAP003038-RA	AGAP003968-RA	AGAP011426-RA	AGAP008196-RA	AGAP000021-RA	AGAP005549-RA
AGAP012990-RA	AGAP003191-RA	AGAP004059-RA	AGAP012334-RA	AGAP008304-RA	AGAP012956-RA	

Table 4.7.4.7
miRanda targets

AGAP004819-RC	AGAP006645-RA	AGAP007713-RD	AGAP002161-RB	AGAP003243-RA	AGAP004080-RA	AGAP009891-RA
AGAP004819-RA	AGAP006645-RB	AGAP007713-RC	AGAP002161-RA	AGAP003315-RA	AGAP011896-RA	AGAP010239-RB
AGAP004819-RB	AGAP006645-RC	AGAP007713-RB	AGAP002198-RA	AGAP003623-RA	AGAP012397-RA	AGAP010239-RC
AGAP004967-RA	AGAP006645-RD	AGAP007713-RA	AGAP002342-RA	AGAP003623-RB	AGAP007732-RB	AGAP010239-RD
AGAP005618-RA	AGAP006647-RA	AGAP001417-RA	AGAP002362-RA	AGAP003623-RC	AGAP007732-RA	AGAP010239-RA
AGAP006101-RA	AGAP006647-RB	AGAP001417-RB	AGAP002525-RA	AGAP003854-RA	AGAP008132-RA	AGAP006637-RA
AGAP006380-RA	AGAP007700-RA	AGAP013434-RA	AGAP003177-RA	AGAP003892-RA	AGAP008385-RA	AGAP007897-RA
AGAP006637-RB						

Table 4.7.4.8
PITA targets

AGAP001417-RA	AGAP002161-RB	AGAP003243-RA	AGAP006637-RB	AGAP007732-RB	AGAP010239-RB	AGAP013434-RA
AGAP001417-RB	AGAP002342-RA	AGAP003315-RA	AGAP007700-RA	AGAP008132-RA	AGAP010239-RC	AGAP006637-RA
AGAP002161-RA	AGAP002362-RA	AGAP006101-RA	AGAP007732-RA	AGAP010239-RA	AGAP010239-RD	

Table 4.7.4.9
TargetScan targets

AGAP000436-RA	AGAP001281-RA	AGAP004437-RA	AGAP005360-RB	AGAP009284-RA	AGAP010147-RI	AGAP012397-RA
AGAP000440-RA	AGAP001281-RB	AGAP004437-RB	AGAP005437-RA	AGAP010147-RA	AGAP010147-RJ	AGAP013481-RA
AGAP000532-RA	AGAP001452-RA	AGAP004437-RC	AGAP005585-RA	AGAP010147-RB	AGAP010147-RK	AGAP013481-RC
AGAP000532-RB	AGAP002359-RA	AGAP004437-RD	AGAP005585-RB	AGAP010147-RB	AGAP010164-RA	AGAP013481-RE
AGAP001027-RA	AGAP003836-RA	AGAP004592-RA	AGAP005585-RC	AGAP010147-RC	AGAP010164-RB	AGAP013481-RF
AGAP001027-RB	AGAP003903-RA	AGAP004592-RE	AGAP006479-RA	AGAP010147-RD	AGAP010164-RC	AGAP013481-RG
AGAP001027-RC	AGAP003937-RA	AGAP004592-RH	AGAP007249-RB	AGAP010147-RE	AGAP010216-RA	AGAP013742-RA
AGAP001091-RB	AGAP003937-RB	AGAP004592-RI	AGAP008156-RA	AGAP010147-RF	AGAP010286-RA	
AGAP001148-RA	AGAP003937-RC	AGAP005245-RE	AGAP008288-RA	AGAP010147-RG	AGAP010286-RD	
AGAP001148-RB	AGAP003940-RA	AGAP005360-RA	AGAP009182-RA	AGAP010147-RH	AGAP011092-RA	

Table 4.7.4.10

RNAhybrid targets

AGAP004737-RA	AGAP006573-RA	AGAP001765-RA	AGAP003463-RA	AGAP010835-RA	AGAP008432-RB	AGAP000438-RA
AGAP004967-RA	AGAP006575-RA	AGAP001778-RA	AGAP003464-RA	AGAP010940-RA	AGAP008432-RD	AGAP000448-RA
AGAP004992-RA	AGAP006637-RB	AGAP001779-RB	AGAP003490-RA	AGAP010941-RA	AGAP008446-RA	AGAP000453-RA
AGAP005003-RA	AGAP006784-RA	AGAP001779-RA	AGAP003494-RA	AGAP011131-RA	AGAP008503-RB	AGAP000488-RA
AGAP005038-RA	AGAP006799-RA	AGAP001779-RC	AGAP003542-RA	AGAP011286-RA	AGAP008518-RA	AGAP013283-RA
AGAP005079-RI	AGAP006828-RA	AGAP001796-RA	AGAP003560-RA	AGAP011308-RA	AGAP008574-RA	AGAP000533-RA
AGAP005079-RJ	AGAP006830-RA	AGAP001935-RB	AGAP013333-RA	AGAP011344-RA	AGAP008647-RA	AGAP000696-RA
AGAP005079-RK	AGAP007002-RA	AGAP001935-RA	AGAP003685-RA	AGAP011350-RA	AGAP008702-RA	AGAP000697-RB
AGAP005079-RF	AGAP007082-RA	AGAP002022-RA	AGAP003730-RA	AGAP011373-RA	AGAP008719-RA	AGAP000697-RA
AGAP005079-RG	AGAP007120-RA	AGAP002028-RA	AGAP003789-RA	AGAP011374-RA	AGAP008724-RA	AGAP000719-RA
AGAP005079-RH	AGAP007163-RA	AGAP002032-RA	AGAP003789-RC	AGAP011421-RA	AGAP008884-RA	AGAP000726-RA
AGAP005124-RC	AGAP007163-RB	AGAP002109-RA	AGAP003789-RB	AGAP011580-RA	AGAP008925-RA	AGAP000808-RA
AGAP005131-RA	AGAP007242-RA	AGAP002198-RA	AGAP003790-RC	AGAP028105-RA	AGAP009019-RA	AGAP000820-RA
AGAP005136-RA	AGAP007249-RB	AGAP002236-RA	AGAP003790-RB	AGAP011717-RA	AGAP009023-RA	AGAP000851-RA
AGAP005201-RA	AGAP007325-RA	AGAP002262-RA	AGAP003802-RA	AGAP011733-RA	AGAP009034-RA	AGAP000904-RA
AGAP005213-RA	AGAP007325-RB	AGAP002269-RA	AGAP003808-RA	AGAP011756-RA	AGAP009044-RA	AGAP000930-RA
AGAP005213-RD	AGAP007333-RA	AGAP002320-RA	AGAP003836-RA	AGAP011816-RA	AGAP009072-RA	AGAP000941-RA
AGAP005213-RF	AGAP007365-RA	AGAP002332-RA	AGAP003843-RA	AGAP011829-RA	AGAP009075-RA	AGAP000941-RB
AGAP005300-RB	AGAP007375-RA	AGAP002355-RB	AGAP003937-RB	AGAP011888-RA	AGAP009117-RB	AGAP000995-RA
AGAP005300-RA	AGAP007529-RA	AGAP002355-RA	AGAP003937-RC	AGAP011890-RA	AGAP009117-RA	AGAP001021-RB
AGAP005300-RC	AGAP007574-RA	AGAP002401-RA	AGAP003937-RA	AGAP028038-RA	AGAP009193-RA	AGAP001021-RA
AGAP005431-RA	AGAP007579-RA	AGAP002421-RA	AGAP004011-RA	AGAP012009-RA	AGAP009290-RA	AGAP001044-RA
AGAP005654-RA	AGAP007713-RD	AGAP002525-RA	AGAP004080-RA	AGAP012054-RA	AGAP009459-RA	AGAP007135-RA
AGAP005843-RA	AGAP007713-RC	AGAP002570-RA	AGAP004086-RA	AGAP012100-RA	AGAP009516-RA	AGAP007306-RA
AGAP005899-RA	AGAP007713-RB	AGAP002579-RA	AGAP004101-RA	AGAP012177-RA	AGAP009926-RA	AGAP007897-RA
AGAP005961-RA	AGAP007713-RA	AGAP002603-RA	AGAP004134-RA	AGAP012180-RA	AGAP009929-RA	AGAP008653-RA
AGAP006085-RA	AGAP001140-RA	AGAP002606-RA	AGAP004143-RA	AGAP012397-RA	AGAP009945-RA	AGAP009781-RA
AGAP006103-RC	AGAP001217-RA	AGAP002661-RA	AGAP004148-RA	AGAP007839-RA	AGAP009998-RA	AGAP011677-RA
AGAP006103-RB	AGAP001285-RA	AGAP002672-RA	AGAP004256-RA	AGAP007849-RB	AGAP010137-RA	AGAP012090-RA
AGAP006103-RA	AGAP001343-RA	AGAP002752-RA	AGAP004256-RB	AGAP007852-RB	AGAP010239-RB	AGAP007636-RA
AGAP006186-RE	AGAP001364-RA	AGAP002927-RA	AGAP004296-RA	AGAP007852-RA	AGAP010239-RC	AGAP004846-RA
AGAP006186-RC	AGAP001504-RA	AGAP003017-RA	AGAP013294-RA	AGAP013764-RA	AGAP010239-RD	AGAP004846-RB
AGAP006186-RD	AGAP012990-RA	AGAP003049-RC	AGAP004489-RA	AGAP007923-RA	AGAP010239-RA	AGAP004846-RC
AGAP006346-RB	AGAP013288-RA	AGAP003049-RA	AGAP004533-RA	AGAP007978-RA	AGAP012570-RA	AGAP005132-RA
AGAP006346-RC	AGAP001563-RA	AGAP003090-RA	AGAP004548-RA	AGAP013742-RA	AGAP012792-RA	AGAP007565-RA
AGAP006353-RA	AGAP001573-RA	AGAP003093-RA	AGAP004560-RA	AGAP008137-RB	AGAP000104-RA	AGAP003669-RA
AGAP006376-RA	AGAP001621-RB	AGAP003177-RA	AGAP004592-RB	AGAP008273-RA	AGAP000110-RA	
AGAP006376-RB	AGAP001621-RA	AGAP003289-RA	AGAP004592-RF	AGAP008296-RA	AGAP000278-RA	
AGAP006380-RA	AGAP001651-RA	AGAP003296-RA	AGAP004659-RA	AGAP008322-RA	AGAP000312-RA	
AGAP006383-RA	AGAP001752-RB	AGAP003301-RA	AGAP010386-RA	AGAP008393-RA	AGAP000359-RA	
AGAP006444-RA	AGAP001761-RA	AGAP003431-RC	AGAP010672-RA	AGAP008432-RA	AGAP000407-RA	

Table 4.7.4.11 Final list of putative miR-275 target genes

Gene ID	Gene name	Species	Algorithm*
AAEL006582-RA; RB	calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type	Aedes	IN; TS; PITA; R; RH
AAEL000577-RA	DNA binding protein elf-1/transcription factor CP2 and related proteins	Aedes	IN; TS; PITA; R
AAEL006126-RA; RB	conserved hypothetical protein similar to Vitellogenin A1	Aedes	IN; TS; PITA; R
AAEL002714-RA	kinesin-like protein KIF23 (mitotic kinesin-like protein 1)	Aedes	IN; PITA; R; RH
AAEL005191-RA	cdk10/11 (cell division protein kinase 10/11)	Aedes	IN; PITA; R; RH

*Algorithm used include in-house (IN), TargetScan (TS), miRanda (R), and RNAhybrid (RH).

Chapter V

Conclusions of the Dissertation

5.1 Conclusions

Hematophagous female mosquitos require a large enough vertebrate blood meal for the maturation and development of their eggs. The blood meal triggers swift and synchronized reproductive events in the two main reproductive organs fat body and ovary. The nutritional amino acids in vertebrate blood acts as signals for vitellogenesis, the process of yolk formation during maturation of their eggs - one of the most important events in the mosquito reproductive cycle (Roy et al., 2015). It is during blood-feeding that the female mosquitoes transmit the disease pathogens. Therefore, a better understanding of the molecular mechanisms behind these events should lead to new and innovative control mechanisms. The goal of this dissertation was to identify important factors related to female mosquito reproduction with the help of interdisciplinary studies involving bioinformatics tools and molecular biology techniques.

I focused on two different levels viz. 1) transcriptional and 2) post-transcriptional, in order to accomplish this goal. At the transcriptional level my goal was to identify *cis-*regulatory elements on the promoters of genes that are co-expressed at a particular time point during the reproductive period, in female mosquito fat body, post blood meal. This was the first aim for this dissertation, because we believe that a significant step towards understanding the regulatory mechanisms governing the *Ae. aegypti* transcriptome during

this period, is to identify the transcription factor binding sites (TFBSs) and their

respective transcription factors (TFs).

Analyses of microarray data generated from female fat bodies have previously

shown that clusters of genes are differentially expressed during different time points

between 6 and 72 hours post blood meal (PBM) (Roy et al. 2015). Major regulators

involved in gene regulation during this period have also been identified (Roy et al. 2015).

This study was aimed at identifying other unknown factors and their relation to the major

regulators, during vitellogenesis. We have used multiple bioinformatics tools to search

the non-coding regulatory regions (2kb upstream of the translation start sites) of the

differentially regulated, co-expressed genes, for identification of *cis*-regulatory elements.

We have identified 89 putative TFBSs, most of which are cluster specific, and studied the

primary features associated with real TFBSs, such as, positional and orientation bias and

evolutionary conservation. Thirty-four unique TFs related to these TFBSs have also been

identified using the JASPAR database.

The second aim for this dissertation was use these cluster-specific factors to build

putative regulatory networks and then test some of these factors for their functionality

within the network. To accomplish the first part of this aim we used the GeneMANIA

webtool, which builds putative regulatory networks based on prior experimental data

from different animals stored in its databases. For the second part i.e. functional analyses

of some of the factors, we used molecular techniques like RNA-interference mediated

depletion of these factors and quantitative reverse transcription polymerase chain

reactions to evaluate the effects of the said depletions on other factors and certain target

genes. Depletion of six TFs which includes one of the major factors EcR, provided interesting information regarding the regulation of different gene sets during this period. We found evidences for possible indirect and direct negative regulatory pathways controlled by the receptor for the steroid hormone 20-hydroxyecdysone. Our results also suggest a possible co-regulation of certain genes by Broad, which encodes a family of C2H2-type, zinc-finger DNA-binding transcription factors and the Iriquois protein Mirror.

The third and final goal of this dissertation was aimed at the post-tarnscriptional level of events occurring during the reproductive period. The specific aim was to develop a reliable approach for miRNA target prediction and use the same for predicting targets for miRNAs involved in mosquito reproduction. Prediction of miRNAs targets is a critical step towards understanding the functionality of this group of small RNAs (Watanabe et al., 2007). A large number of miRNAs have already been identified in mosquitoes, but very little is known about functions of these miRNAs. The scarcity of high-throughput analysis in this field, is one of the reasons behind this problem. Functional tests using traditional molecular techniques are time-consuming, laborious and expensive. Therefore, an approach that can lead to reliable target predictions can help a lot. Here we have developed an approach where we use multiple prediction tools that use distinctly different background algorithms for miRNA target predictions. As a next step we compared the results from the different tools and looked for the overlaps. The idea behind this was - since the different algorithms use different criteria related to possible miRNA::mRNA interactions, targets identified by more programs should be more

reliable. Finally, we used *An. gambiae* sequences to look for evolutionary conservation of these targets, as authentic targets should be the subjects of positive selection pressure.

We used this approach to predict targets for three miRNAs viz. miR-1174, miR-8 and miR-275. The prediction resulted in five, seven and five targets respectively. Concurrent studies confirmed that Serine Hydroxymethyltransferase, Tubulointerstitial nephritis antigen, which is the ortholog for *Drosophila* Swim and SERCA are authentic targets for miR-1174, miR-8 and miR-275 (Liu et al., 2014 and Lucas et al., 2015, Zhao et al., unpublished data) and involved in blood digestion or mosquito reproduction. All of these were identified to be putative targets by the bioinformatics prediction method.

Overall, I have fulfilled all three aims proposed for this dissertation. I have identified multiple cluster-specific, *cis*-regulatory elements, constructed putative regulatory networks with those and tested the function of the TFs related to some of these elements. I have also developed an approach for reliable miRNA target prediction and used that to predict targets for miRNAs that have been subsequently validated and found to be authentic. The *cis*-regulatory elements tested and the miRNA targets validated were found to be involved in different processes related to mosquito reproduction. Therefore, I truly believe that my results have contributed towards acquiring important knowledge related to the molecular mechanisms involved in mosquito reproduction, even though future work remains to be done. I hope that my study would contribute towards the development of new and efficient vector control strategies that would ultimately save some lives from mosquito-borne diseases.

5.2 References

Roy S., Saha, TT, Johnson, L, Zhao B., Ha J., White, KP, Girke T, Zou Z, Raikhel AS., Regulation of Gene Expression Patterns in Mosquito Reproduction PLoS Genetics 2015 11(8):e1005450

Liu S., Lucas KJ, Roy S, Ha, J, Raikhel AS, Mosquito-specific microRNA-1174 targets serine hydroxymethyltransferase to control key functions in the gut, Proc. Natl. Acad. Sci. 2014 Vol. 111 14460-14465

Lucas KJ, Roy S, Ha J, Gervaise AL, Kokoza VA, Raikhel AS, MicroRNA-8 targets the Wingless signaling pathway in the female mosquito fat body to regulate reproductive processes, Proc. Natl. Acad. Sci. 2015 Vol. 112(3): 1440-1445

Zhao B, Lucas KJ, Saha TT, Ha J, Chen C, Roy S, Raikhel AS, MicroRNA-275 directly targets sarco/endoplasmic reticulum Ca2+ adenosine triphosphatase (SERCA) to control key functions in the mosquito gut, ICE 2016 Orlando, Florida Sep. 25-30, 2016.

Watanabe Y, Tomita M, Kanai A, Computational Methods for microRNA Target Prediction, Methods in Enzymology 2007, Vol. 427: 65-86