2017

ANTISENSE AFP TRANSCRIPTS IN MOUSE LIVER AND THEIR POTENTIAL ROLE IN AFP GENE REGULATION

Maria S. Dixon
University of Kentucky, mdixon1913@gmail.com
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.356

Recommended Citation
Dixon, Maria S., "ANTISENSE AFP TRANSCRIPTS IN MOUSE LIVER AND THEIR POTENTIAL ROLE IN AFP GENE REGULATION" (2017). Theses and Dissertations--Microbiology, Immunology, and Molecular Genetics. 14.
https://uknowledge.uky.edu/microbio_etds/14

This Doctoral Dissertation is brought to you for free and open access by the Microbiology, Immunology, and Molecular Genetics at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Microbiology, Immunology, and Molecular Genetics by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student’s advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student’s thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Maria S. Dixon, Student

Dr. Martha L. Peterson, Major Professor

Dr. Ken Fields, Director of Graduate Studies
ANTISENSE AFP TRANSCRIPTS IN MOUSE LIVER AND THEIR POTENTIAL ROLE IN AFP GENE REGULATION

DISSertation

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Medicine at the University of Kentucky

By
Maria S. Dixon
Lexington, Kentucky

Director: Dr. Martha L. Peterson, Professor of Microbiology, Immunology & Molecular Genetics
Lexington, Kentucky
2017

Copyright © Maria S. Dixon 2017
ABSTRACT OF DISSERTATION

ANTISENSE AFP TRANSCRIPTS IN MOUSE LIVER AND THEIR POTENTIAL ROLE IN AFP GENE REGULATION

Hepatocellular carcinoma (HCC) is the most common form of primary liver cancer, ranking the sixth most common cancer and third most common cause of cancer mortality worldwide. Alpha-fetoprotein (AFP) is a plasma protein that is highly expressed in the fetal liver and shut off after birth. AFP expression is elevated in regenerating adult liver and HCC and has been used extensively as a diagnostic marker of liver cancer. We have been studying mouse liver gene regulation to better understand mechanisms by which changes in gene expression contribute to liver development, homeostasis and disease. Zinc Fingers and Homeoboxes 2 (Zhx2) has been identified as a repressor of AFP, but the mechanism of this regulation remains unknown. Interestingly, all targets of Zhx2 that have been identified to date, including H19, Glypican 3, Elovl3 and Cytochrome P450 (CYP) genes, are also known to be misregulated in HCC. Thus, a better understanding of the mechanism by which these genes are regulated by Zhx2 will likely lead to new insights into gene regulation during HCC progression.

Antisense transcripts belong to a diverse class of long noncoding RNA molecules >200 nucleotides in length that often structurally resemble mRNAs, but do not encode proteins. While studying AFP mRNA regulation by Zhx2 in the mouse, our lab identified novel antisense AFP (asAFP) RNA transcripts that partially overlap the 3’ half of the mouse AFP gene. ENCODE tracks of ChIP-seq data for histone modifications in mouse liver show that the genomic region around the 5’ end of asAFP RNA has peaks for marks associated with promoters and enhancers. To better understand asAFP regulation, I identified the asAFP RNA 5’ end and the promoter elements that drive transcription. asAFP RNAs are ~5kb alternatively spliced, mainly cytoplasmic transcripts containing 2-4 exons. These transcripts were also detected in adult mouse liver RNA-seq data. asAFP is likely a noncoding RNA because it contains several small open reading frames that are 98 aa or smaller with no known functional domains or homology to known proteins. There is no evidence for similar transcripts in human liver. The abundance of asAFP RNA inversely correlates with AFP mRNA levels during postnatal liver development. Normally, asAFP RNA levels are high and AFP mRNA levels are low in the adult mouse liver. However, in the absence of Zhx2, AFP mRNA levels are higher and asAFP RNA levels are reduced, suggesting asAFP may be involved in the developmental regulation of AFP.

Antisense transcripts function through a variety of mechanisms to positively or negatively regulate the expression of target genes. To explore the role of asAFP RNA in AFP gene regulation, I expressed segments of asAFP RNA in a mouse liver cell line and measured endogenous AFP mRNA levels. My data revealed that all segments of asAFP repressed endogenous AFP mRNA in trans. To determine the mechanism by which asAFP RNA regulates AFP, I expressed asAFP segments that overlapped only with exons or introns of AFP. The asAFP segments that overlap with the exons showed greater repression of endogenous AFP mRNA levels than those overlapping with intronic sequences. Additionally, I considered whether asAFP RNA repression of AFP mRNA may involve
RNA editing by Adenosine deaminase acting on RNA (ADAR). ADARs convert adenosine to inosine in double-stranded RNAs that results in RNA degradation. My data indicate that AFP and asAFP dsRNA is not extensively edited, suggesting ADAR mediated decay is not involved in the regulation of AFP mRNA expression. However, further studies are required to determine the mechanism of cytoplasmic AFP mRNA degradation. Together, my data characterizes the transcriptional regulation of novel mouse asAFP transcripts and provides a model system to investigate how these transcripts regulate AFP mRNA through RNA-RNA interaction.

KEYWORDS: Long Noncoding RNA, asAFP, AFP, RNA-RNA interaction, gene regulation
ANTISENSE AFP TRANSCRIPTS IN MOUSE LIVER AND THEIR POTENTIAL ROLE IN AFP GENE REGULATION

By
Maria S. Dixon

Martha L. Peterson
Director of Dissertation

Ken Fields
Director of Graduate Studies

Date
I dedicate this dissertation to my loving husband.
ACKNOWLEDGEMENTS

“For I know the plans I have for you, declares the Lord,” Jeremiah 29:11. I give all thanks and praises to my Lord and Savior, Jesus Christ for guiding me throughout this journey. I would like to thank my mentor, Dr. Martha Peterson for your support and guidance through the most challenging years of my life. I am grateful to my committee members: Dr. Brett Spear, Dr. Brian Rymond and Dr. Subbarao Bondada for their guidance and feedback throughout my doctoral studies. Special thanks to Dr. John D'Orazio for not only serving as my external examiner, but for graciously allowing me to use his laboratory equipment for the completion of many of my experiments.

I am truly grateful to Drs. Shirley Qiu, Jieyun Jiang and Lilia Turcios for their guidance, patience and expertise, which has molded me into the scientist that I am today. Special thanks to all the members of the Spear lab for your feedback during lab meetings and support. I would also like to thank the Microbiology, Immunology and Molecular Genetics department, Ms. Kate Fresca, and Dr. Beth Garvy for providing an amazing and supportive academic environment.

My friends have been my “rams in the bush” and have given me unwavering support and love during the last five years. I am truly blessed to have such an amazing group of people in my life. Thank you, Cheavar Blair, LaQueta Shelton, Isabel Derera, Will Arnold, Katie Mckenna, Drs. Karine Oben, Sara Alhakeem, and Grant Jones. I am so proud of each of you and I am going to miss you dearly! To my mentees, Maya James and Brittany Rice, thank you for encouraging talks and motivating me to finish the course.

I don't know where I would be without my family. To my parents, Major and Bonnie Smith, thank you for your unconditional love, support, and guidance. To my grandmother, Viola Hughes, you have been a true inspiration in my life. Special thanks to my sister and cousin, Monique Smith and Deedra Hughes for being my best friends and angels on earth. To my Kentucky mother, Ms. Susan Jones, thank you for your love, prayers, and encouraging words. Lastly, I am extremely grateful to my loving husband, William L. Dixon III. You are truly my Miracle on earth! Thank you for being my biggest cheerleader and loving me through thick and thin; you have shown me the true meaning of unconditional love. I love you!
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS........................................................................................................... iii

LIST OF TABLES ....................................................................................................................... vi

LIST OF FIGURES .................................................................................................................. vii

CHAPTER I ................................................................................................................................. 1

INTRODUCTION ......................................................................................................................... 1

IDENTIFICATION of LONG NONCODING RNAs ................................................................. 2
LncRNAs ARE FOUND THROUGHOUT THE GENOME ....................................................... 5
ANTISENSE RNA REGULATES GENE EXPRESSION ......................................................... 9
LncRNA AND HEPATOCELLULAR CARCINOMA .............................................................. 14
AFP DEVELOPMENTAL AND GENE REGULATION ......................................................... 16
ZINC FINGER AND HOMEBOXES 2 (Zhx2) ................................................................. 18

CHAPTER II ............................................................................................................................. 23

MATERIALS AND METHODS ................................................................................................. 23

Mice ..................................................................................................................................... 23
5' Rapid Amplification of Complementary DNA Ends (RACE) .................................. 23
DNA preparations, primers and sequencing ................................................................. 24
Transformation ....................................................................................................................... 25
Plasmid Construction ........................................................................................................... 25
Cell Culture .......................................................................................................................... 29
Transient DNA Transfections ............................................................................................ 29
Dual Luciferase Reporter Assay System ........................................................................... 30
RNA Extraction, cDNA Synthesis, and RT-qPCR ........................................................ 30
Flow Cytometry ................................................................................................................... 31
Bioinformatic Tools .............................................................................................................. 32
Statistical Analysis .............................................................................................................. 32

CHAPTER III ........................................................................................................................... 37

IDENTIFICATION AND CHARACTERIZATION OF NOVEL asAFP RNA TRANSCRIPTS IN MOUSE LIVER ................................................................. 37

Identification of asAFP 5' end ............................................................................................. 39
asAFP is a noncoding RNA ............................................................................................... 40
LIST OF TABLES

Table 2.1 Plasmid constructs used for asAFP promoter study ................................................. 33
Table 2.2 Plasmid constructs used for asAFP functional study .................................................. 34
Table 2.3 Oligos used for 5’RACE .................................................................................................. 34
Table 2.4 Oligos used for asAFP promoter study ....................................................................... 35
Table 2.5 Oligos used for asAFP functional study ....................................................................... 35
Table 2.6 Oligos used for ADAR study .......................................................................................... 36
Table 2.7 Oligos used for RT-qPCR .............................................................................................. 36
LIST OF FIGURES

Figure 1-1 Long non-coding RNAs are found throughout the genome ...................20
Figure 1-2 RNA pol II core promoter elements ............................................21
Figure 1-3 Antisense RNA regulation of gene expression .................................22
Figure 3-1 asAFP 5’ RACE strategy .................................................................44
Figure 3-2 Complete view of characterized asAFP transcripts ...........................45
Figure 3-3 asAFP is a noncoding RNA ..............................................................46
Figure 3-4 Conservation analysis of asAFP .........................................................47
Figure 3-5 AFP mRNA and asAFP RNA are co-expressed in adult mouse tissues....48
Figure 3-6 asAFP RNA levels are not altered in tumors ....................................49
Figure 4-1 ENCODE tracks from ChIP-seq data in the region of the AFP/asAFP/Afm genomic region .................................................................58
Figure 4-2 Identification of the asAFP promoter ..................................................59
Figure 4-3 Analysis of the asAFP promoter .........................................................60
Figure 4-4 Identification of the asAFP core promoter elements ...........................61
Figure 4-5 Zhx2 does not regulate the asAFP promoter ......................................63
Figure 5-1 asAFP RNA regulates AFP in trans ...................................................73
Figure 5-2 Sorted cells confirm that asAFP RNA regulates AFP in trans ............74
Figure 5-3 asAFP does not regulate neighboring genes .......................................76
Figure 5-4 Overexpression of asAFP does not affect the expression of Alb and Afm....77
Figure 5-5 Exonic regions more effectively repress AFP mRNA ..........................78
Figure 5-6 Sorted cells confirm that exonic regions more effectively repress AFP...mRNA .............................................................................................................79
Figure 5-7 Identification and distribution of RNA editing in AFP mRNA and asAFP RNA overlapping region .................................................................81
CHAPTER I

INTRODUCTION

The discovery of noncoding RNA transcripts and their involvement in gene regulation has highlighted the complexity of the mammalian transcriptome. Only 2% of the mammalian genome is transcribed into mature protein coding genes. Interestingly, between 70% to 90% of the genome is transcribed into non-coding RNA (ncRNA) transcripts with little or no protein coding potential, suggesting potential regulatory roles in complex organisms (Wilusz et al., 2009). ncRNAs are grouped based on their transcript size into two classes: small noncoding RNA (<200 base pairs (bp)) and long noncoding RNAs (lncRNA, >200 bp) (Wilusz et al., 2009). Classical small noncoding RNAs include microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), endogenous small interfering (endo-si) RNAs, and some bacterial regulatory RNAs that are highly conserved and are involved in various cellular processes (Cech and Steitz, 2014). miRNAs have been the most extensively studied small noncoding RNA and are posttranscriptional regulators of RNA stability or translation via extensive base pairing (Cai et al., 2009).

Conversely, lncRNAs are a heterogeneous group of RNA molecules that are greater than 200 nt in length, often resemble mRNAs, but lack protein coding potential. (Quan et al., 2015; Yan and Wang, 2012). While the role of pervasive transcription of the genome was not clear when first discovered, it was considered by some as “transcriptional noise”. Yet, multiple stable lncRNA transcripts have been identified and characterized in many organisms. LncRNAs have been shown to play important roles in numerous cellular processes, including regulation of gene expression, genome imprinting and dosage compensation. Despite the widespread attention lncRNAs have
gained in recent years, it is not clear how many exist and whether uncharacterized transcripts carry biological function and significance (Kung et al., 2013).

To better understand IncRNA’s role in liver gene regulation, this dissertation will characterize the function of novel unannotated mouse antisense alpha-fetoprotein (asAFP) RNA transcripts. The introduction will provide a general overview of long noncoding RNAs, with a focus on the current understanding of natural antisense IncRNA regulation, followed by a review of the AFP gene as a model biological system to study liver gene regulation.

IDENTIFICATION of LONG NONCODING RNAs

_H19_ and X-inactive specific transcript (Xist) were among the first discovered IncRNAs from searching cDNA libraries for other developmentally regulated and genomic imprinted genes, respectively (Brown et al., 1992; Pachnis et al., 1984). As expected, a number of genes were protein-coding, yet some transcripts were identified as noncoding (Brannan et al., 1990; Brown et al., 1992). The development of high throughput DNA tiling microarrays and next generation RNA-sequencing (RNA-seq) analysis have expanded the number of identified IncRNAs (Yan and Wang, 2012). Conventional cDNA microarray tiling studies provided the first glimpse of pervasive transcription of non-coding RNAs by detecting transcripts represented by probes on the array. From these studies, it was estimated that more than 70% of protein coding genes have antisense partners that are expressed at lower levels (Faghihi and Wahlestedt, 2009; Katayama et al., 2005). Further advancements of sequencing technology led to the development of next generation RNA-seq technology, which has become a widely-used tool that provides a more accurate and comprehensive view of the transcriptome. RNA-seq studies revealed pervasive transcription of as many IncRNAs as protein coding genes in the genome. Moreover, this technology highlighted IncRNA tissue specific
expression patterns and expanded the view of pervasive transcription of lncRNAs (Rinn and Chang, 2012).

LncRNAs are a diverse class of transcripts that often share structural features with mRNAs. Like mRNAs, they are generally transcribed by RNA polymerase II, capped, polyadenylated, and spliced. However, not all lncRNAs exhibit these mRNA-like features. The genes for lncRNAs, like those for mRNAs, are associated with chromatin marks such as histone H3 lysine 4 trimethylation (H3K4me3) at their promoters and histone H3 lysine 36 trimethylation (H3K36me3) throughout their gene body. These histone modifications indicate evidence of active transcription. Many characterized lncRNAs are preferentially localized in the nucleus, while some transcripts are exported to the cytoplasm, where they exert their function. In comparison to mRNAs, lncRNAs, as a class, are generally shorter in length, less stable, contain fewer exons and are poorly conserved evolutionarily (Lee, 2012; Quinn and Chang, 2016; Zhang et al., 2014). Because lncRNAs are often expressed at lower levels than protein coding genes and display high tissue specificity, functional characterization of newly discovered transcripts remains challenging (Zhang et al., 2014).

LncRNAs are called “noncoding” because they lack significant open reading frames (arbitrarily defined as <100 amino acids) or homology to known protein domains. Protein-coding genes are commonly defined by the presence of an ORF >100 aa. However, a lncRNA may contain such an ORF by chance alone, and many well-characterized long ncRNAs (H19, Xist, and MEG3) do indeed contain longer ORFs (Brannan et al., 1990; Brockdorff et al., 1992; Hagan et al., 2009; Joubel et al., 1996). Various computational tools have been developed to search for predicted ORFs based on genome alignment and homology to known proteins, such as NCBI Open Reading Frame Finder (ORF finder) (Housman and Ulitsky, 2016). Additionally, algorithms such as Coding-Potential Calculator (CPC) and Coding-Potential Assessment Tool (CPAT)
calculate coding probability based on sequence features such as ORF length, ORF coverage, codon frequency, and the presence of sequences encoding known functional domains (Kong et al., 2007; Wang et al., 2013b). These tools provide a general approach to determine coding potential, but are limited in identifying new short proteins that are <50 aa. Therefore, complementing these tools with experimental validation may confirm whether a transcript is indeed noncoding.

Surprisingly, some IncRNAs were found to be associated with ribosomes. However, it is still unclear whether any produced peptides are functional (Ingolia et al., 2011). More recently, several groups have used ribosomal profiling as an approach to answer this question. This technique stabilizes RNAs associated with ribosomes using translation inhibitors, such as cycloheximide or emetine, followed by nuclease digestion of RNAs not engaged with ribosomes. The ribosome-protected RNAs are purified and analyzed by deep sequencing (Ingolia et al., 2011). Ingolia and colleagues performed ribosome profiling experiments from mouse embryonic stem (ES) cells and found that a significant number of IncRNAs had a ribosomal occupancy similar to known protein coding genes. From these studies, they concluded that IncRNAs can be exported to the cytoplasm and engage with ribosomes, but it was unclear whether these IncRNAs were translated (Ingolia et al., 2011). A reanalysis of this same ribosome profiling data focused on the 5’UTR and IncRNA ribosomal occupancy to distinguish between IncRNAs and real coding sequences. During classical translation of protein coding genes, once a ribosome encounters a bona fide stop codon, it results in an immediate release and sharp decrease in ribosomal occupancy. However, when ribosome-associated IncRNA’s encountered a stop codon, they remained engaged, instead of falling off. This observation suggested that ribosome occupancy alone does not imply protein/peptide synthesis (Guttman et al., 2013). Interestingly, some IncRNA have been mistakenly classified as noncoding. For example, the tarsal-less RNA in Drosophila, was
previously thought to be a putative polyadenylated ncRNA (Tupy et al., 2005). In subsequent studies, it was further characterized as a polycistronic mRNA that is conserved among insects. *Tarsal-less* encodes multiple short, 33-nt ORFs, which are translated into 11 aa functional peptides. These results further complicate the debate as to whether IncRNA transcripts can encode functional proteins (Galindo et al., 2007; Kondo et al., 2007; Kondo et al., 2010; Kung et al., 2013).

**LncRNAs are found throughout the genome**

Genome-wide studies have revealed that much of the mammalian genome is pervasively transcribed to produce noncoding and long noncoding RNAs (Cao, 2014; Kung et al., 2013). LncRNAs can originate from intergenic, exonic, intronic, antisense or enhancer regions (Figure 1-1). Long intergenic RNAs (lincRNA) are generally spliced and polyadenylated transcripts transcribed by RNA pol II, from regions between genes and do not overlap protein-coding genes (Brown et al., 1992; Hrdlickova et al., 2014) (Bartolomei et al., 1991; Kung et al., 2013). Many lincRNAs were identified through chromatin signatures of actively transcribed genes that are enriched with H3K4me3 marks at the promoter and H3K36me3 along the gene body (Kung et al., 2013).

Antisense lncRNAs are transcribed from the opposite strand of annotated, protein-coding genes or noncoding sense transcripts, with which they overlap. These transcripts will be discussed in greater detail later in this chapter. Moreover, sense intronic or exonic lncRNAs are transcribed in the same direction as their protein coding sense transcript. If the TSS of lncRNAs that are located on the antisense strand of sense transcripts are <1000 base pairs away from the TSS of protein coding gene, this may affect the expression of their neighboring gene. Enhancer RNAs are generally bidirectionally transcribed from DNA sequences of enhancer regions. Although lncRNAs have been classified based on their genomic location in relation to nearby genes, it is not clear how
this classification reflects biological function or evolutionary origin (Hrdlickova et al., 2014; Lee, 2012).

TRANSCRIPTIONAL REGULATION of LncRNAs

Like mRNAs, lncRNAs are mainly transcribed by RNA polymerase II (pol II) from standard promoters that can be recognized by transcription factors (TFs) to form a transcription initiation complex (Lin et al., 2015). Antisense RNAs also can originate from independent, cryptic, shared bidirectional promoters, or divergent promoters that are positioned within genes (Lin et al., 2015; Pelechano and Steinmetz, 2013). More recently, enhancers have also been shown to direct the synthesis of lncRNAs (Kim et al., 2010).

RNA pol II transcribed promoters contain DNA sequences that direct accurate and sufficient initiation of transcription by the RNA polymerase II machinery (Butler and Kadonaga, 2002; Juven-Gershon and Kadonaga, 2010; Smale and Kadonaga, 2003). The core promoter can initiate transcription in a dispersed or focused manner. Dispersed promoters are associated with CpG islands and encompass several weak transcriptional start sites (TSS), whereas focused promoters contain a single predominant TSS or a few TSS within a narrow region of several nucleotides. However, some promoters display mixed focused and dispersed modes (Juven-Gershon and Kadonaga, 2010). Moreover, promoters are enriched with chromatin features, such and H3K4me3 as indicators of active transcription.

The RNA pol II core promoter is a structurally and functionally diverse region that contains the minimal stretch of DNA sequences required to initiate transcription (Butler and Kadonaga, 2002; Juven-Gershon and Kadonaga, 2010; Smale and Kadonaga, 2003; Theisen et al., 2010). Accurate and efficient transcription requires the action of basal transcription factors that bind and recognize core promoter elements. The core
promoter contains the transcription initiation start site and may contain one or more key regulatory elements that extend upstream and downstream, including TATA box, initiator (Inr), downstream promoter element (DPE), and motif ten element (MTE) (Figure 1-2) (Butler and Kadonaga, 2002). Each element has specific functions and can be present in a variety of combinations to drive transcription (Butler and Kadonaga, 2002; Juven-Gershon and Kadonaga, 2010; Smale and Kadonaga, 2003).

Some core promoters are TATA-dependent whereas others are considered TATA-less. The TATA box was the first promoter motif found in eukaryotes and is recognized by basal transcription factor for RNA polymerase II D (TFIID) complex, including TATA box binding protein (TBP) and 13 TBP-associated factors (TAFs). The TATA box consensus sequence (TATAAAG) lies about 30 bp upstream of the transcription start site (TSS) and directs the preinitiation complex to the start site that lies within the Inr element. The TFIIB recognition element (BRE) lies immediately upstream and downstream of some TATA boxes and can either increase or decrease transcription initiation in eukaryotes. Although TATA boxes are only present in ~20% of mammalian core promoters such as N-myc gene and Igf2r (Igf2 receptor), many promoters do not contain a TATA box but have other regulatory elements that drive transcription (Butler and Kadonaga, 2002; Juven-Gershon and Kadonaga, 2010; Lin et al., 2015; Smale and Kadonaga, 2003).

The Inr element is the most commonly occurring core promoter element and contains the TSS. Some genes that are transcribed by TATA-less promoters contain multiple TSSs (Lin et al., 2015). Transcription initiates within the Inr, in which the A nucleotide is designated at the +1 position. Mutational analysis revealed sequences spanning the -3 to +5 positions relative to the +1 position are necessary and sufficient for accurate transcription (Smale and Kadonaga, 2003). However, an A at the +1 position is
required for the strict spacing and functional requirements for the downstream MTE and DPE promoter elements (Butler and Kadonaga, 2002; Juven-Gershon and Kadonaga, 2010; Theisen et al., 2010). The MTE is precisely located at +18 to +27 relative to A+1 and is important for basal transcription, but functions independently of a TATA box and DPE (Lim et al., 2004; Theisen et al., 2010). The DPE element lies +28 to +32 bp downstream of the Inr and is functionally similar to the TATA box in recruiting the preinitiation complex to the TSS. Mutations in DPE element spacing relative to the A+1 results in 10-50-fold reduction in basal transcription activity (Butler and Kadonaga, 2002; Juven-Gershon and Kadonaga, 2010; Kutach and Kadonaga, 2000). In summary, core promoters are diverse and include multiple DNA elements that important for transcriptional initiation and regulation.

Transcriptional regulation is not only achieved through the diversity of core promoters, but also by enhancer elements. Transcriptional enhancers are cis-acting DNA regulatory elements that positively regulate transcription of proximal promoters over potentially long distances (Orom and Shiekhattar, 2011a). Enhancers bind general and cell-specific TFs by DNA looping that interact with the promoter-bound transcriptional machinery to activate expression (Melamed et al., 2016). Genome-wide analysis of chromatin modifications suggests that promoters and enhancers may be distinguishable based on their chromatin signatures (Koch et al., 2007). While active promoters can be identified by H3K4me2, H3K4me3, H3K4me2 and H3K4me3, active enhancers are identified by high levels of H3K4me1 and H3K4me2, low levels of H3K4me3, and high levels of H3K27ac. Interestingly, both promoters and enhancers occasionally display overlapping chromatin features, such as binding to RNA pol II, H3K4me1, H3Kme3 and p300/CBP binding (Orom and Shiekhattar, 2011a, b). Although the chromatin state at promoters are largely unchanged across diverse cell-types,
enhancers have been shown to be associated with chromatin marks in a highly cell specific manner (Heintzman et al., 2007; Orom and Shiekhattar, 2011a).

Recent genome-wide studies of transcriptional enhancers based on chromatin modifications and enhancer binding proteins has identified of an emerging subclass of IncRNAs that are transcribed bidirectionally from enhancer elements, called enhancer-derived RNAs (eRNAs). eRNAs are synthesized from enhancers and are identified by chromatin features, such as H3K4me1, H3K27ac and the presence of RNA pol II (Orom and Shiekhattar, 2011a). A study in mouse neuronal cells following depolarization identified transcription of short (<2kb), non-polyadenylated, non-coding, bidirectional eRNA transcripts by mapping the regions associated with enhancer-specific modifications bound by p300/CBP (Kim et al., 2010). eRNA transcription can be bidirectional or unidirectional and their expression levels correlate with those of nearby target genes (Kim et al., 2010; Orom and Shiekhattar, 2011a). A novel enhancer located 13.8 kb downstream the DHRS4-AS1 TSS was shown to produce an eRNA, AS1eRNA, which increases DHRS4-AS1 expression by mediating a chromatin loop between the enhancer and the DHRS4-AS1 promoter (Yang et al., 2016). To date the detailed mechanism of eRNAs activity are ongoing; however, it is proposed that eRNAs may facilitate looping of DNA sequences between enhancers and their target promoters or they may regulate chromatin modifications that keep enhancers and promoters in their active states (Melamed et al., 2016; Orom and Shiekhattar, 2011a).

ANTISENSE RNA REGULATES GENE EXPRESSION

Natural antisense transcripts (NATs) were initially discovered in bacteria and are RNAs transcribed on the opposite strand of sense protein coding or noncoding genes (Pelechano and Steinmetz, 2013; Wagner and Simons, 1994). The development of high-
throughput gene expression analysis identified antisense transcripts widely distributed throughout the mammalian genome in eukaryotes. Global transcriptome analysis estimates that more than 30% and 70% of annotated transcripts in human and mouse, respectively, have antisense transcripts that are noncoding (Katayama et al., 2005). They can originate from the 5’ end, 3 end, or within the gene body of sense transcripts (Faghihi and Wahlestedt, 2009; Lee, 2012).

Although a small portion of natural antisense IncRNAs have been characterized, many more are unlikely to be transcriptional noise, but rather important regulators of gene expression that can control transcription, epigenetic modification, chromatin remodeling, RNA maturation, RNA transport, and protein synthesis (Figure 1-3) (Hrdlickova et al., 2014). They can positively or negatively regulate the expression of their overlapping sense transcript by forming sense-antisense RNA duplexes in the nucleus and/or cytoplasm, suggesting a biological role for antisense RNAs. In the nucleus, antisense RNAs can regulate their sense counterpart mRNAs by blocking transcription, controlling epigenetic modifications, or modulating alternative splicing and stability (Faghihi and Wahlestedt, 2009; Villegas and Zaphiropoulos, 2015; Zhang et al., 2014). Several studies have demonstrated that when an antisense RNA overlaps with a protein coding gene, the act of transcription in the antisense orientation, but not the molecule itself, can interfere with the sense transcription through direct pol II transcriptional interference in cis (Pelechano and Steinmetz, 2013). The transcription collision model proposes that RNA polymerases on opposite DNA strands transcribing towards each other physically crash and stall, blocking further transcription (Faghihi and Wahlestedt, 2009; Pelechano and Steinmetz, 2013). For example, in budding yeast, GAL10 and GAL7 are naturally organized in a tandem gene arrangement. To model the effect of convergent and co-transcribed genes, the orientation of GAL7 was inverted in
the sense orientation, while GAL10 was antisense. During the elongation process, the synthesis of GAL7 halted as a result of synthesis from GAL10 in the antisense orientation, thereby decreasing GAL7 expression (Prescott and Proudfoot, 2002).

Antisense RNAs may facilitate epigenetic changes by recruiting chromatin remodeling complexes or nuclear proteins to specific genomic loci. The antisense non-coding RNA in the INK4 locus (ANRIL) is a 3.8kb transcript that is transcribed on the opposite strand of the INK4A-ARF-INK4B gene cluster. ANRIL is transcribed by RNA pol II and is spliced into several isoforms, most of which are polyadenylated (Kotake et al., 2011; Villegas and Zaphiropoulos, 2015). The first intron of ANRIL overlaps with the two exons of p15INK4B and was shown to be involved in epigenetic silencing of INK4b-ARF-INK4a locus. ANRIL binds two polycomb repressor complexes, PRC1 and PRC2, which are required for the silencing of p15INK4B tumor suppressor gene (Kotake et al., 2011; Villegas and Zaphiropoulos, 2015; Yap et al., 2010).

There is evidence that antisense RNAs can regulate transcription of their sense mRNA counterparts through the formation of antisense RNA:DNA hybrids. The antisense intronic noncoding RASSF1 (ANRASSF1) is a capped, polyadenylated, unspliced IncRNA that is transcribed on the opposite strand of the RASSF1A tumor suppressor gene and overlaps exon 2 of RASSF1A. Through ectopic overexpression and siRNA knockdown of ANRASSF1, this antisense RNA was shown to inversely regulate RASSF1A mRNA and protein levels. The authors proposed that ANRASSF1 interacts with genomic DNA at the transcription site, forming an RNA/DNA hybrid, and recruits the inhibitory PRC2 complex to the RASSF1A promoter region (Beckedorff et al., 2013).

Nuclear sense-antisense RNA duplex formation can regulate gene expression by modulating alternative splicing or altering mRNA localization or stability. Antisense RNAs can influence splicing by masking splice sites, thereby blocking factors that would
initiate splicing and altering the balance between splice variants. For example, the thyroid hormone receptor alpha gene (TRα) encodes two protein isoforms, TRα1 and TRα2, with antagonistic functions. The Rev-ErbA alpha (RevErβα) antisense transcript is complementary to the last exon of TRα2 but not TRα1. Expression of RevErβα prevents splicing of the TRα primary transcript, altering the balance of the two splice variants (Faghihi and Wahlestedt, 2009; Hastings et al., 1997). Nuclear sense-antisense RNA duplexes are also substrates for adenosine deaminases that act on RNA (ADAR) editing enzymes that recognize double stranded RNAs (dsRNAs) and convert adenosines into inosines. A- to-I editing often promotes nuclear retention in which single site mutations can alter mRNA structure, expression and function. Several studies have identified many editing sites to lie within inverted elements, specifically Alu repeats, or intronic sequences. Peters, et al. demonstrated that the interaction between the antisense transcript, sas-10 and the Drosophila melanogaster 4f-rrnp induces A-to-I editing in the overlapping region provides a potential mechanism for modifications of sense-antisense RNA duplexes (Peters et al., 2003; Werner and Sayer, 2009).

LncRNAs that are exported to the cytoplasm can also form RNA duplexes with complementary RNAs. They can regulate gene expression by modulating stability, translation or by functioning as precursors of miRNA or compete for miRNA binding sites (Yoon et al., 2013). Staufen 1 is known to bind double stranded RNAs and promote mRNA decay. Staufen binding sites can be formed by imperfect base pairing between an Alu element in the 3’UTR of a STAU1-mediated decay (SMD) target mRNA and another Alu element in a cytoplasmic, polyadenylated lncRNA (1/2-sbsRNAs). 1/2-sbsRNAs facilitate the binding of STAU1 to mRNA, thus promoting mRNA degradation. However, not all mRNAs containing a 3’UTR Alu elements are targeted for SMD (Gong and Maquat, 2011; Kim et al., 2005; Zhang et al., 2014).
Antisense RNAs that promote mRNA stability is demonstrated by the antisense transcript for β-secretase-1 (BACE1-AS). BACE1-AS is a conserved ~ 2kb spliced, polyadenylated RNA that is transcribed on the opposite strand of the BACE locus. BACE1-AS forms perfect base-pairing with 104 nucleotides of human BACE1 mRNA exon 6. BACE1-AS transcripts promote the stability of BACE mRNA by masking the binding site for miR-485-5p through the formation of sense-antisense RNA duplex, which prevents BACE mRNA degradation (Faghihi et al., 2008; Faghihi et al., 2010).

LncRNAs can promote or inhibit translation of target genes. The human lincRNA-p21 is transcribed in the antisense orientation ~15 kb upstream the Cdkn1a gene (Huart et al., 2010). lincRNA-p21 was recently shown to associate with ribosomes and repress target mRNA translation. LincRNA-p21 was shown to base pair imperfectly with CTNNB1 (β-catenin) and JUNB (JunB) mRNAs in several places throughout the coding and UTR regions. The formation of RNA duplex and association with translation repressor Rck and FMRP leads to translational repression of target genes (Yoon et al., 2012; Zhang et al., 2014).

Previous genome-wide studies have predicted several LncRNAs as precursors of miRNAs (He et al., 2008; Rashid et al., 2016). For example, H19 LncRNA exon 1 gives rise to two conserved microRNAs, miR-675-3p and miR-675-5p that are induced during skeletal muscle differentiation. Thus, H19 has an essential function in skeletal muscle differentiation and regeneration that is mediated by miRNAs (Cai and Cullen, 2007; Dey et al., 2014). Taken together, our understanding of the mechanistic details of antisense RNAs biological function is ongoing, but many have been shown to act in cis or trans at all levels of gene regulation and through a wide variety of mechanisms.
LncRNA AND HEPATOCELLULAR CARCINOMA

Hepatocellular carcinoma (HCC) is the most common form of primary liver cancer, accounting for 90% of cases (Huang et al., 2014). The highest incidences of HCC cases occur in sub-Saharan Africa and Eastern Asia. However, HCC incidence in the United States, United Kingdom, and Australia is increasing primarily due to cirrhosis caused by chronic hepatitis B or C infections (HCV or HBV), excessive alcohol, obesity, diabetes, and hemochromatosis (El-Serag and Rudolph, 2007). The rise of HCC in the United States has been attributed to not only to HCV, but non-alcoholic fatty liver disease (NAFLD), affecting up to 30% of adult population and 90% of those with morbid obesity (Baffy et al., 2012).

LncRNA expression levels have been shown to be dysregulated in several human cancers and to show tissue-specific expression (Huang et al., 2014). In HCC, IncRNAs can contribute to HCC development and disease by influencing proliferation, apoptosis, invasion, metastasis, and angiogenesis (Du et al., 2012; Huang et al., 2014; Kung et al., 2013; Wang et al., 2011; Yu et al., 2015). As mentioned previously, IncRNAs are involved in diverse biological processes including epigenetic silencing, splicing regulation, RNA-miRNA, RNA-protein interactions and genetic variation (Huang et al., 2014).

During embryonic development, the imprinted H19 IncRNA is expressed abundantly in the liver, silenced after birth, and is frequently reactivated during tumorigenesis (Poirier et al., 1991; Serviss et al., 2014). H19 dysregulation has been observed in a variety of cancers, including HCC, colorectal, breast, and bladder cancers. In some cases, increased H19 mRNA levels in HCC tumors were at higher levels than the traditional HCC marker, AFP (Matouk et al., 2007). However, there are several contradictory reports that suggest H19 functions as an oncogene or tumor suppressor.
Matouk, et al. reported that H19 expression enhanced human HCC tumor growth and formation in *in vitro* and *in vivo* models. They showed that hypoxia strongly upregulated H19 RNA levels. When H19 expression was silenced in cells under hypoxic stress, they failed to form colonies in soft agar after hypoxia recovery to the same level as H19-expressing cells. Furthermore, silencing H19 expression significantly retarded tumor growth *in vivo*, suggesting H19's oncogenic phenotype (Matouk et al., 2007).

Conversely, loss and gain of function studies in HCC cells suggest that H19 functions as a tumor suppressor. The suppression of tumor metastasis is proposed to be mediated by the association of H19 and the hnRNP U/PCAF/RNA Pol II complex that activates miR-200 family (Zhang et al., 2013). Further studies are necessary to clarify the molecular mechanism(s) controlling the expression of the imprinted gene H19 and its role in tumor development and metastasis.

Another lncRNA that has been reported to be dysregulated in HCC is highly upregulated in liver cancer (HULC). HULC expression is upregulated in HCC tumors and promotes hepatocyte proliferation via down regulation of the tumor suppressor gene-p18 (Du et al., 2012). In another study, Wang, et al. discovered a self-amplifying, auto-regulatory loop that upregulates HULC through the inhibition of miR-372. The expression of HULC gene is upregulated in liver cancer cells by transcription factor cAMP response element binding protein (CREB). HULC functions as a molecular decoy of miR-372 binding, thus activating its target gene, protein kinase cAMP-activated catalytic subunit beta (PRKACB). This increase in PRKACB can induce the phosphorylation of CREB, which in turn, stimulates HULC expression (Wang et al., 2010). Interestingly, HULC expression levels have been detected in higher frequency in the plasma of HCC patients when compared to healthy patients. These findings suggest that lncRNAs detected in
plasma may serve as potential non-invasive biomarkers for the diagnosis/prognosis and monitoring of disease progression (Xie et al., 2013; Yu et al., 2015).

The maternally expressed gene 3 (MEG3) IncRNA, which belongs to the imprinted DLK1-MEG3 locus is a putative tumor suppressor in HCC and plays an important role in cell development and growth (Qin et al., 2013; Zhou et al., 2012) (Huang et al., 2014). MEG3 RNA levels were lower in four human liver HCC cell lines and tissues in comparison to normal hepatocytes (Zhou et al., 2012). Braconi, et al. demonstrated that MEG3 overexpression in HCC cells decreased cell growth (Braconi et al., 2011; Huang et al., 2014; Yu et al., 2015). More importantly, MEG3 can significantly increase p53 RNA and protein levels, triggering cell cycle arrest or apoptosis in response to oncogene activation and other stress stimuli (Zhang et al., 2010). Altogether, it is important to investigate the molecular mechanisms of IncRNA expression in HCC to provide potential applications in diagnosis and treatment for liver cancer.

**AFP DEVELOPMENTAL AND GENE REGULATION**

Alpha fetoprotein (AFP) was identified as the first oncofetal protein and is the most commonly used serum biomarker for HCC (Abelev, 1971; Behne and Copur, 2012). AFP is a member of the serum albumin gene family that consists of Albumin (Alb), Afamin (Afm), Vitamin D-binding protein (DBP) and AFP-related gene (Arg) (Peterson et al., 2011). AFP is synthesized abundantly in the fetal liver and in the visceral endoderm of the yolk sac and to a lesser extent the fetal gut, but is dramatically repressed by $10^5$ fold during the first several weeks after birth in mice (Belayew and Tilghman, 1982). Postnatal AFP serum levels remain at low basal levels in normal adult liver, but can be reactivated during liver regeneration and HCC (Abelev, 1971).
The mouse AFP gene, located on chromosome 5, is regulated by four distinct regulatory elements: a liver specific 250-bp promoter that is located directly adjacent to exon 1 that contains multiple binding sites for transcription factors and three enhancers (EI, EII, EIII) located further upstream at -2.5 kb, -5.0 kb and -6.5kb respectively. Each 300-bp enhancer is active in varying degrees in tissues where AFP is normally expressed. EI and EII are active in all hepatocytes, whereas EIII activity is restricted to hepatocytes surrounding the central vein (Peyton et al., 2000a). The 22kb mouse AFP gene contains 15 exons, and encodes a 70 Da glycoprotein (Mizejewski, 2004; Peterson et al., 2011). AFP is the most abundant serum protein in the developing mammalian fetus and can bind a variety of ligands, including estrogen, bilirubin, fatty acids, hormones, retinoids, steroids, heavy metals, dyes, dioxins and organic drugs, that could allow it to modulate cell growth in the developing fetus (Mizejewski, 2004; Spear, 1999).

Much of the work to understand AFP gene regulation has been done in mice. Olsson and Ruoslahti performed initial studies to understand the mechanisms of AFP postnatal silencing by analyzing AFP serum levels in adult livers of 27 different inbred mouse strains (Olsson et al., 1977). BALB/cJ mice expressed 20-fold higher AFP serum levels compared to other mouse strains. The elevated serum AFP levels in BALB/cJ mice were due to higher levels of steady-state hepatic AFP mRNA (Belayew and Tilghman, 1982; Olsson et al., 1977). The phenotype for this post-natal persistence of AFP mRNA levels in BALB/cJ mice was named raf (regulator of AFP), later named Afr1 (Alpha-fetoprotein regulator 1) (Belayew and Tilghman, 1982; Pachnis et al., 1984).

Linkage analysis initially mapped the Afr1 gene to mouse chromosome 15 (Blankenhorn et al., 1988). By positional cloning, Afr1 was identified as the Zinc-fingers and homeoboxes 2 (Zhx2) gene that is responsible for the increased AFP levels in BALB/cJ mice (Perincheri et al., 2005). The persistent expression of AFP in adult BALB/cJ mice is due to a hypomorphic mutation in the Zhx2 gene, which contains an
ETnIIα mouse endogenous retroviral element (MERV) in intron 1. Normal transcription initiation of the Zhx2 gene in BALB/cJ mice occurs, however most pre-mRNAs splice into the retroviral element, dramatically reducing expression of Zhx2.

A molecular screen for other genes expressed similarly to the developmental regulation of AFP in adult BALB/cJ liver identified H19 as another Zhx2 target (Pachnis et al., 1984). Low expression of Zhx2 in adult BALB/cJ mice correlated with increased expression of liver AFP and H19 mRNA levels compared to wild-type mice. Further studies showed that the expression of a Zhx2 transgene driven by a liver specific transthyretin enhancer/promoter restored wild-type Zhx2 expression in BALB/cJ mice and repressed AFP and H19 mRNA, providing further evidence that Zhx2 is involved in the postnatal silencing of AFP and H19 in the adult liver (Perincheri et al., 2005). Moreover, Zhx2 mRNA levels increase-while AFP mRNA levels decrease after birth in wild-type mice, consistent with Zhx2’s role in AFP mRNA repression (Jiang et al., 2017). These data suggest that Zhx2 functions as a repressor of target genes in the liver and the reduced Zhx2 levels in BALB/cJ livers leads to increased postnatal expression of these genes.

ZINC FINGER AND HOMEBOXES 2 (Zhx2)

Zinc fingers and homeoboxes 2 (Zhx2) is a member of a small family of highly conserved vertebrate-specific transcription factors that include Zhx1 and Zhx3. The Zhx family share similar gene structure in which the proteins are encoded by an unusually large internal exon. Each Zhx protein contains two amino terminal C2-H2-type zinc fingers and four or five homeodomains, protein domains typically associated with DNA, RNA, and protein interactions, suggesting that these proteins may function as transcriptional regulators. Analysis of mouse Zhx mRNAs by Northern blot revealed that these transcripts are expressed ubiquitously and levels vary among different tissues.
Zhx1 and Zhx2 are tightly linked on mouse chromosome 15, whereas Zhx3 is located on chromosome 20. Each Zhx family member forms homodimers and heterodimers with each other and NF-YA and luciferase assays suggest that ZHX proteins function as transcriptional repressors (Kawata et al., 2003a; Kawata et al., 2003b).

Transgenic mouse studies showed that Zhx2 acts through the 250 bp AFP promoter; however, it is not known if this is a direct or indirect interaction (Peyton et al., 2000a). Zhx2 is considered as a repressor of developmentally regulated genes including AFP, H19, Glypican 3 (GPC3) and Lipoprotein lipase (Lpl) (Gargalovic et al., 2010; Morford et al., 2007; Pachnis et al., 1984). Interestingly, all these Zhx2 targets are also mis-regulated in HCC. More recently, in contrast to previously identified Zhx2 targets that appear to be repressed by Zhx2, cytochrome P450 (CYP) enzymes and the mouse major urinary proteins are positively regulated by Zhx2 (Creasy et al., 2016; Jiang et al., 2017). Creasy, et al. performed gender specific studies using conditional Zhx2 deletion in mice and demonstrated that Cyp2a4 expression in adult male liver increased in the absence of Zhx2 (Creasy et al., 2016). In another study, Zhx2 was shown to bind the Mup24 promoter and is required for normal levels of Mup expression in the adult liver; BALB/cJ mice have abnormally low levels of Mup expression (Jiang et al., 2017). Taken together, these data indicate Zhx2 may positively or negatively regulate target genes; however, the exact mechanism by which Zhx2 regulates AFP and other target genes is not fully understood.
Long non-coding RNAs are found throughout the genome. IncRNAs are defined based on their genomic location relative to nearby protein-coding genes. Intergenic IncRNAs are located in between genes and do not overlap nearby protein coding genes. IncRNAs that are bi-directionally transcribed include enhancer RNAs synthesized from enhancer regions or located on the opposite strand; their transcription start sites (TSS) lie close to the TSS protein of coding sense transcripts. Exonic and Intronic sense IncRNAs share exons or introns with, and are transcribed in the same direction as, the protein coding genes. Antisense RNAs are found on the opposite strand and can completely or partially overlap nearby protein coding genes.
RNA Pol II core promoter elements. RNA Pol II promoters contain a set of core DNA elements that can be present in a variety of combinations to facilitate transcription. Some promoters contain a TATA box that lies about 30 bp upstream of the transcription start site that directs the preinitiation complex to the TSS, which lies within the Initiator (Inr) element. The TFIIB Recognition elements (BRE) that lie upstream and downstream the TATA box can either increase or decrease transcription rates in eukaryotes. The Initiator element contains a consensus sequence where the A is at the +1 position. In TATA-less promoters, the downstream promoter elements that are dependent of the Initiator include the motif ten element (MTE) Downstream Promoter element (DPE). The MTE is precisely located at +18 to +27 relative to A+1 in the Inr. The DPE element lies at +28 to +33 bp downstream of the Inr and is functionally similar to the TATA box in directing preinitiation complex to the TSS. Adapted from (Butler and Kadonaga, 2002)
Figure 1-3. Antisense RNA regulation of gene expression. Antisense RNAs can regulate gene expression by modulating transcription, RNA processing, and/or translation. The act of transcription in the antisense orientation can modulate transcription of sense RNA as in the case of transcriptional interference, when RNA polymerases from both strands collide in the overlapping region, thus blocking transcription. Antisense RNAs may modulate epigenetic regulation through DNA methylation or by recruiting chromatin modifying complexes to activate or repress gene expression. Posttranscriptionally, antisense RNAs can bind to the overlapping sense transcript and mask splice sites resulting in alternative splicing. The formation of nuclear sense-antisense dsRNAs may serve as substrates for RNA editing enzymes, thereby altering nuclear mRNA retention, stability, and transport. The formation of cytoplasmic sense-antisense RNA duplex can alter mRNA stability by masking miRNA binding sites or by Staufen-mediated decay. Alternatively, antisense RNAs can alter translation by physical interaction with the translational machinery or binding to sense transcripts, thereby blocking initiation/elongation or ribosomal entry.
CHAPTER II
MATERIALS AND METHODS

Mice

Mouse liver RNA samples used in this dissertation were obtained from mice that were housed in the University of Kentucky Division of Laboratory Animal Research (DLAR) facility and kept according to Institutional Animal Care and Use Committee (IACUC) approved protocols. I obtained tissues or RNA from others in the lab for my analyses. Several types of mice with altered Zhx2 expression were used. The TTR-Flag-Zhx2 transgenic mice express a liver-specific tagged Zhx2 protein on a BALB/cJ background. Zhx2 protein levels in the transgenic animals are similar to wild-type mice. Liver RNA was prepared from BALB/cJ littermates with or without the transgene. Wild-type Zhx2 (Zhx2 +/+) , liver-specific knock-out (Zhx2 Δhep) or whole-body Zhx2 knock-out (Zhx2 -/-) littermates on a C57/Bl6 background in which the Zhx2 gene had been deleted by Cre recombinase were also used in some experiments. These mice have been described previously (Creasy et al., 2016; Jiang et al., 2017).

For the liver tumor model, male C3H/Bl6 livers were harvested 36 weeks after a single DEN treatment. These samples have been described previously (Creasy et al., 2016).

5’ Rapid Amplification of Complementary DNA Ends (RACE)

5’ RACE was performed using a 5′/3′ 2nd Generation RACE kit (catalog no. 03353621001; Roche Applied Science) following the manufacturer’s instruction. Briefly, 1ug of purified total RNA from BALB/cJ mice that express a liver-specific Zhx2 transgene was reverse transcribed by (SuperScript™ II) using RT primer #725 (see Table 2-1 for all the oligos used in this thesis). The resulting cDNA was purified using the PCR
cleanup kit (Promega), and a homopolymeric (dA) tail was added at the 3’ end of the cDNA (corresponding to the 5’ end of the RNA) using terminal transferase. The (dA)-tailed cDNA was amplified by PCR using the provided forward oligo (dT)-anchor primer and the gene-specific reverse primer #760. Finally, the resulting PCR product was used as a template to generate a nested PCR product using the provided anchor primer and another gene-specific reverse primer #396. The final PCR product was cloned in pGEM-T-Easy vector (Promega). Following bacterial transformation, colony PCR (primers #386 and #387) was performed to screen individual clones for the presence and size of the insert. Minipreps were prepared for plasmids with various sized (400bp, 550bp, 600bp, and 1kb) inserts and sequenced (ACGT). Only 5’RACE clone #7, with a 400 bp insert, corresponded to the AFP gene locus. The remaining 5’ RACE clones were either found on a different locus or were repetitive elements.

A second nested PCR reaction was performed to identify additional 5’ RACE products using the provided anchor primer and two independent gene-specific reverse primers: #824 and #726. However, after cloning and sequencing additional fragment, no additional clones from the AFP locus were found.

A second 5’ RACE reaction was performed with the previously described cDNA and amplified by PCR using the provided forward oligo (dT)-anchor primer and three different gene-specific reverse primers #835, #831, and #776. From this, two additional clones that matched the previously identified 5’ RACE clone #7, but contained sequences that did not align to AFP gene locus at the ends, were found. None of the 5’RACE clones contained the alternative 219 bp exon 2.

**DNA preparations, primers and sequencing**

Plasmid DNA used for sequencing or transfections was prepared by Thermo Scientific GeneJet Plasmid Miniprep and Midiprep Kit (Thermo Fisher #K0502, #K0481), per
the manufacturer's instructions. DNA was quantified by an optical density at 260nm using a NanoDrop Spectrophotometer (Thermo Scientific). Primers were purchased from Integrated DNA Technologies, Inc. (IDT; Coralville, IA). DNA sequencing was performed by ACGT, Inc. (Germantown, MD) or the University of Kentucky Genomics Core.

**Transformation**

Plasmid DNA was transformed into DH5α E. coli competent cells by incubating 3uL of ligation reaction with 100uL of competent cells on ice for 30 minutes. Cells were heat shocked for 45 seconds at 42°C, and incubated on ice for 2 minutes. 1mL of Luria Bertani LB Broth (AMRESCO, #J106-500G) was added and the cells were shaken at 37°C for 1 hour. Cells were directly plated (100uL) and the remaining cells were pelleted for one minute at 16,000 x g. LB media was decanted and the cells were resuspended in the remaining liquid (50 uL). Cells were plated on LB Agar plates containing 50mg/mL of Ampicillin, (LB/Amp) (Sigma) and incubated overnight at 37°C. For plasmid DNA reactions ligated into pGEM-T Easy, 100 uL of ChromoMax™ IPTG/X-Gal Solution (Fisher BioReagents # BP4200-1) was spread on LB/Amp plates and incubated for 1 hour at 37°C prior to plating the bacteria.

**Plasmid Construction**

*Generation of asAFP promoter reporter plasmids*

Various sized asAFP promoter fragments, between -2160 (#779), -830 (#874), -480 (#875), -320 (#876), -175 (#726), and -42 (#941) to +112 (#877) relative to transcription start site (TSS), were generated from mouse genomic DNA (from M12 B cell line, derived from a BALB/c mouse) using the Q5 high fidelity PCR amplification (NEB # M0491S). The amplicons were purified, A-tailed with Taq Polymerase (NEB #M0267) and cloned into pGEM-T Easy. The orientation of the insert within the plasmid was determined by PCR using primers #387 and #861, for the SP6 orientation and primers
#386 and #860 for the T7 orientation. The T7 oriented miniprep DNA plasmids were sequenced (ACGT). Fragments were isolated by SpeI and BglII restriction digest and subcloned into the promoter-less pGL3-basic (no control elements) and pGL3-promoter (promoter element only) luciferase reporter vectors (Promega) that had been linearized by Nhel and BglIII restriction digest. The asAFP -90 to +112 pGL3 basic promoter reporter plasmid was generated by KpnI and BstXI restriction enzyme digest from the asAFP -175 to +112 pGL3 basic reporter plasmid. The digested DNA was treated with Klenow (NEB #MO210) to form blunt ends followed by T4 DNA ligation and transformation into DH5α competent cells. Deletion of the putative Downstream Promoter Element (DPE) promoter element was generated by PCR reactions with primers #726/#942 and #941/#942, respectively, from mouse genomic DNA (from M12 B cell line). Amplicons were A-tailed and cloned into pGEM-T Easy. The orientation of the insert within the plasmid was determined by PCR primers #386 and #941 for Sp6 orientation and primers #387 and #942 for T7 orientation. T7 oriented clones were sequenced and subcloned into pGL3 basic vector as described above.

**Generation of asAFP RNA expression plasmids**

To generate asAFP RNA expression plasmids, four segments that cover the entire asAFP RNA transcript were amplified by high fidelity PCR amplification. To generate the asAFP mid expression plasmid that contains an 1800bp fragment, primers #879 and #836 were used and to make the asAFP 3’end expression plasmid that contains a 2000bp fragment, primers #835 and 870 were used in high fidelity PCR reactions from mouse genomic DNA, since the genomic and cDNA were the same sequence. Two asAFP 5’ alternatively spliced isoform expression plasmids were PCR amplified from C3H/BL6 cDNA, primers #878 and #880-Xba. The asAFP 5-1 expression plasmid that contains a 900 bp fragment, includes exon 219 and the asAFP 5-2 expression plasmid that contains a 600bp fragment, skips exon 219. The amplicons were purified, A-tailed,
cloned into pGEM-T Easy and sequenced. Fragments were excised by XbaI and Xhol (asAFP mid), XhoI and EcoRI (asAFP 3’end) and XbaI and Nhel (asAFP 5-1 and 5-2) restriction digests and subcloned into pcDNA3.1(-) expression vector (Invitrogen V790-20) that had been cut with appropriate restriction enzymes for directional cloning. Primers were designed with restriction sites so that segments could be reassembled to generate an asAFP full-length expression plasmid as diagrammed in Figure 5-1.

To generate pcDNA constructs that express AFP/asAFP overlapping cDNA sequences, AFP exons 9-14 cDNA expression plasmid was generated from Zhx2Δhep mouse liver RNA that was reverse transcribed by iScript cDNA synthesis and PCR amplified using primers #328 and #761. The asAFP mid pGEM-T plasmid construct was used to generate just exon 12 by PCR amplification with primers #266/#916. Amplicons were purified, A-tailed, and cloned into pGEM-T Easy and sequenced. Fragments were excised by EcoRI restriction digest, and subcloned into pcDNA. The pcDNA vector was prepared for ligation by digestion with EcoRI to generate ends complementary to the DNA insert. The digested vector was then treated with Antarctic Alkaline Phosphatase (NEB, #M0289S) to remove the 5’-phosphate group from the exposed ends and ligated to the prepared fragments. The orientation of the insert within the plasmid was determined by PCR using primers #909 and #266 for (-) orientation and #909 and #267 for (+) orientation. Both orientations were isolated so that the sense strand could serve as a control for the antisense strand.

A negative control expression plasmid that contained a 1147 bp fragment from an intron within the IgM gene was generated by high fidelity PCR amplification using primers #706 and #635. The amplicon was A-tailed, cloned into pGEM-T Easy and sequenced. The fragment was excised by EcoRI restriction digest, and subcloned into pcDNA treated with Antarctic Alkaline Phosphatase. The orientation of the insert within
the plasmid was determined by PCR using primers #909 and #706 for (-) orientation and #909 and #644 for (+) orientation.

Single primer mutagenesis was performed to generate expression plasmids that contain AFP/asAFP overlapping intron sequences by removing the exon sequences from the plasmids. The asAFP mid pGEM-T construct was used to delete exon 12 (Δe12) by PCR mutagenesis; reactions were performed per the protocol described by Makarova, et al (Makarova et al., 2000). In this method, the single primer contains sequences on either side of the desired deletion. Reactions included: 300 ng template DNA, a single primer: #915, 200um dNTP, 1.25 U Q5 high fidelity polymerase and 5X Q5 reaction buffer. The preliminary denaturation was at 95°C for 3 minutes, followed by 18 cycles of PCR cycles: 15 seconds of denaturation at 95°C. 1 minute of annealing at 56°C, and 12 minutes of extension at 68°C. The PCR reactions were digested with 1 uL of DpnI endonuclease for 2 hours at 37°C to remove template DNA. Transformations were performed using the following ratios of digest reaction to DH5α competent cells: 2uL /50 uL, 5ul/50ul, and 5uL/100UL. The cells were combined, pelleted for one minute at 16,000 x g and plated out on one LB/Amp Agar plate each. PCR was performed using primers flanking the deletions to analyze 24 clones for exon 12 deletion (#872 /#759) and 18 clones for exon 13 deletion (#815/ #432). The exon 12 and exon 13 deletion efficiencies were 33% and 22% respectively. The fragment that contains the exon 12 deletion was excised by Xbal and Xhol restriction digest and cloned into pcDNA.

**Miniprep plasmids for ADAR study**

To distinguish RNA editing changes from DNA polymorphisms, various AFP/asAFP overlapping regions were PCR amplified using primers #811 and #870 and #873 and #836, respectively from total BL/6 Zhx2fl/fl wild-type liver cDNA and tail genomic DNA. A 2kb fragment that contains overlapping AFP/asAFP regions was PCR amplified (primers
from total and nuclear BL/6 Zhx2\textsuperscript{Alb}/Zhx2\textsuperscript{Alb} wild-type liver cDNA. Amplicons were purified, cloned into pGEM-T Easy and sequenced with universal primers SP6 and T7. An asAFP-internal primer #836 was used to sequence the 2kb clones. A 782 bp fragment that contains the AFP exon 9-14 cDNA that overlap with asAFP RNA was PCR amplified using primers #328/#761 from total BL/6 Zhx2 Alb-Cre expression (Zhx2\textsuperscript{Alb}/Zhx2\textsuperscript{shep}) liver cDNA. The amplicon was cloned into pGEM-T Easy, sequenced and aligned to AFP mRNA and asAFP compiled RNA sequences for potential RNA editing (Geneious software). Multiple clones were aligned to asAFP compiled RNA sequence for potential RNA editing as diagrammed in Figure 5-7.

Cell Culture

Human Embryonic Kidney 293 (HEK 293), human hepatoma (HepG2 and Hep3B) and C57/L (Leaden) derived mouse hepatoma (Hepa 1.6) mammalian cells were grown in T25 flasks. Hepa 1.6, HEK 293, and Hep3B cells were maintained with Dulbecco’s minimal eagle’s media (DMEM,) supplemented with 10% fetal bovine serum (FBS, Gibco). HepG2 cells were supplemented with DMEM, 10% FBS, and 0.1 % Insulin. All cells were maintained in an incubator at 37°C and 5% CO\textsubscript{2}.

Transient DNA Transfections

For asAFP promoter studies, HEK 293 and Hep3B cells were seeded 2.0 x 10\textsuperscript{5} and Hepa 1.6 cells were seeded 1.0 x 10\textsuperscript{5} cells/well in 24-well plates supplemented with 500 uL of the appropriate media for 24 hours prior to transfection. The next day, serum free DMEM media (50 uL) was diluted with TK/Renilla (10 ng) and reporter plasmid DNA (500ng). Turbofect reagent (Thermo Scientific, #R0532) was added to the diluted DNA, vortexed, and incubated for 20 minutes at room temperature. The complex was added to cells in a dropwise manner, gently swirled to mix and incubated for 48 hours. After 48-hour incubation, cells were harvested for Dual Luciferase Reporter assay.
For transient transfections with asAFP expression vectors, Hepa 1.6 cells were
seeded 2.0 x 10^5 cells/well in 6 well plates supplemented with 2mL of the appropriate
media for 24 hours prior to transfection. The next day, serum-free DMEM (200 uL)
media was diluted with 5ug of expression plasmid DNA. Turbofect reagent (10uL) was
added to the diluted DNA and added to cells as previously described. After 48-hour
incubation, cells were harvested for total RNA isolation.

**Dual Luciferase Reporter Assay System**

Cells were harvested 48 hrs post transfection and washed with sterile 1X phosphate
buffered saline (PBS), followed by lysis with 1x Passive Lysis Buffer (100 uL) for 15
minutes, while rocking. Cell lysates (20uL) were placed in duplicate wells of 96-well
luciferase plates (CoStar). Analysis was performed using the GloMax Explorer
(Promega). Both dual luciferase substrates (Promega) were each injected (100uL) into
the designated wells. Luciferase values were corrected relative to Renilla values. The
pGL3 negative control luciferase values were set to 1 and asAFP promoter luciferase
values were normalized to these control values. The average and standard deviation are
shown from at least three independent analyses.

**RNA Extraction, cDNA Synthesis, and RT-qPCR**

Total RNA from 2.0 x 10^5 transiently transfected Hepa 1.6 cells was extracted with
0.5mL RNAzol RT (Molecular Research Center, Inc., Cincinnati, Ohio). Total RNA from
mouse tissues were obtained by homogenization of about 50 mg in 1mL RNAzol RT, per
manufacturer’s protocol. RNA was eluted with 50 µl of Nuclease -free water and
quantified by OD measurement. Some RNA preparations were DNase treated using the
Turbo DNA-free kit (Ambion #AM1907) to remove contaminating genomic DNA or
residual plasmid DNA. Briefly, total RNA (5ug) was incubated with 10x Turbo DNase
buffer and Turbo DNase enzyme (1uL). Half of the Turbo DNase enzyme was initially
added to the reaction and incubated for 30 minutes at 37°C and the remaining enzyme was added incubated for 30 minutes at 37°C. DNase Inactivation reagent (2uL) and Nuclease-free water (3uL) was added to the reaction and incubated for 5 minutes at room temperature, mixing occasionally. The reaction was centrifuged at 10,000 x g for 1.5 minutes and the supernatant RNA was transferred to a clean 0.5mL Eppendorf tube. cDNA was synthesized from 1 ug of RNA using the iScript cDNA Synthesis Kit (BioRad #170-8891) per manufacturer’s protocol. Quantitative Real-Time PCR (qPCR) reactions using SYBR Green PCR Supermix (BioRad #172-5275) were performed using the CFX96 Touch Real-Time PCR Detection System and analyzed using the CFX Manager software (BioRad). To measure detectable endogenous AFP mRNA levels from Hepa 1.6 cells by RT-qPCR, 2 ul of undiluted RT input was required and 1 ul was used for all other RNAs. qPCR Ct values were normalized to the ribosomal protein L30 and were calculated using the ∆∆Ct method.

Flow Cytometry
Hepa 1.6 cells were co-transfected with 4.5 ug of asAFP RNA plasmids and 0.5 ug of a plasmid expressing GFP as a reporter. Reporter expression was determined 48 hr after transfection by flow cytometry. To prepare cells for flow analysis, cells were trypsinized for 3 minutes, resuspended in sterile 1X PBS, and centrifuged at 1000 x g for minutes. Cells were washed once with sterile 1X PBS, filtered through a 0.2um sterile filter and resuspended in Basic Sorting Buffer (1X PBS, 0.5 M EDTA, 0.5M HEPES (pH 7.0)), and 1X Fetal Bovine Serum. Cells were analyzed Becton-Dickinson FACSCalibur flow cytometer and analyzed by the BD CellQuest™ Pro software (San Jose, CA). Total RNA isolated from unsorted, sorted GFP positive, and GFP negative cells was reverse transcribed and changes in endogenous AFP was measured by RT-qPCR.
Bioinformatic Tools

The UCSC Genome Browser on the Mouse Dec. 2011 (GRCm38/mm10) (http://genome.ucsc.edu) was used to analyze the evolutionary conservation of mouse asAFP RNA with other mammalian species. The NCBI Open Reading Frame Finder (ORF finder) (https://www.ncbi.nlm.nih.gov/orffinder) was used to search for asAFP RNA predicted ORFs. The Coding Potential Assessment Tool (CPAT) (http://lilab.research.bcm.edu/cpat/index.php) was used to assess the coding potential of AFP mRNA and asAFP RNA. The Jaspar database (jaspar.genereg.net/ ) was used to identify candidate transcription factors that could bind to asAFP promoter sequences. The ElemeNT navigation tool (http://lifefaculty.biu.ac.il/gershon-tamar/index.php/resources) was used to identify putative asAFP core promoter elements. The asAFP -175 to +106, relative to TSS, was used as the input sequence. Geneious was used for the alignment of asAFP cDNA clones with asAFP RNA and AFP mRNA consensus templates.

Statistical Analysis

All values within a group were averaged and plotted as mean +/- standard deviation. p-values were calculated between two groups using student’s t-test and between three or more groups by ANOVA followed by Tukey’s test. A p-value < 0.05 was considered significant. Data was graphed and analyzed using GraphPad Prism 6 software.
<table>
<thead>
<tr>
<th>plasmid</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEM-T Easy vector</td>
<td>cloning vector for PCR products (Promega)</td>
</tr>
<tr>
<td>pGL3- Basic</td>
<td>empty vector driving luciferase (Promega)</td>
</tr>
<tr>
<td>pGL3- Promoter</td>
<td>SV40 promoter upstream of the luciferase (Promega)</td>
</tr>
<tr>
<td>asAFP -2160 to +112</td>
<td>2-kb fragment from (M12, B cell line) mouse genomic DNA was inserted into Spe1-Nhel/BglII pGL3 basic vector</td>
</tr>
<tr>
<td>asAFP -830 to +112</td>
<td>943 bp fragment from M12 mouse genomic DNA was inserted into Spe1-Nhel/BglII pGL3 basic and pGL3 promoter vector, respectively</td>
</tr>
<tr>
<td>asAFP -480 to +112</td>
<td>594bp fragment from M12 mouse genomic DNA was inserted into Spe1-Nhel/BglII pGL3 basic vector</td>
</tr>
<tr>
<td>asAFP -320 to +112</td>
<td>430 bp fragment from M12 mouse genomic DNA was inserted into Spe1-Nhel/BglII pGL3 basic vector</td>
</tr>
<tr>
<td>asAFP -175 to +112</td>
<td>256 bp fragment from M12 mouse genomic DNA was inserted into Spe1-Nhel/BglII pGL3 basic vector</td>
</tr>
<tr>
<td>asAFP -175 to +29</td>
<td>204 bp fragment from M12 mouse genomic DNA was inserted into Spe1-Nhel/BglII pGL3 basic vector</td>
</tr>
<tr>
<td>asAFP -42 to +112</td>
<td>154 bp fragment from M12 mouse genomic DNA was inserted into Spe1-Nhel/BglII pGL3 basic vector</td>
</tr>
<tr>
<td>asAFP -42 to +29</td>
<td>71 bp fragment from M12 mouse genomic DNA was inserted into Spe1-Nhel/BglII pGL3 basic vector</td>
</tr>
<tr>
<td>Tk/Renilla</td>
<td>HSV-thymidine kinase promoter Renilla luciferase control reporter (Promega)</td>
</tr>
<tr>
<td>HA-Zhx2</td>
<td>full-length Zhx2 with 5’ HA epitope inserted into pcDNA 3.1(-) (Spear/Peterson lab)</td>
</tr>
<tr>
<td>Mup20 promoter</td>
<td>~1.4-kb fragment inserted into the pGL4.14[luc2/Hydro] luciferase vector (Promega) (Jiang, Creasy et al. 2017)</td>
</tr>
</tbody>
</table>
Table 2.2. Plasmid constructs used for asAFP functional study

<table>
<thead>
<tr>
<th>plasmid</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pcDNA 3.1(-)</td>
<td>empty expression vector (Invitrogen)</td>
</tr>
<tr>
<td>asAFP 5-1 pcDNA</td>
<td>900 bp fragment inserted 3'-5' orientation Nhel and Xbal site of pcDNA 3.1(-)</td>
</tr>
<tr>
<td>asAFP 5-2 pcDNA</td>
<td>600 bp fragment inserted 3'-5' orientation Nhel and Xbal site of pcDNA 3.1(-)</td>
</tr>
<tr>
<td>asAFP mid pcDNA</td>
<td>1800 bp fragment inserted 3'-5' orientation between Xbal and Xhol site of pcDNA3.1 (-)</td>
</tr>
<tr>
<td>asAFP 3’ pcDNA</td>
<td>3-kb fragment inserted 3'-5’ orientation between Xhol and EcoRI sites of pcDNA3.1(-)</td>
</tr>
<tr>
<td>asAFP 3+mid pcDNA</td>
<td>3.8-kb fragment inserted 3'-5’ orientation between Xbal and EcoRI site of pcDNA3.1(-)</td>
</tr>
<tr>
<td>e12 (-) pcDNA</td>
<td>224 bp fragment inserted 3'-5' orientation in pcDNA 3.1 (-) EcoR1 site</td>
</tr>
<tr>
<td>delta e12 pcDNA</td>
<td>1576 bp fragment inserted 3'-5' orientation in pcDNA 3.1 (-) EcoR1 site</td>
</tr>
<tr>
<td>AFP e9-14 (+) pcDNA</td>
<td>782 bp fragment inserted 5'-3' orientation in pcDNA 3.1 (-) EcoR1 site</td>
</tr>
<tr>
<td>IgM pcDNA</td>
<td>1147 bp fragment inserted 3'-5' orientation in pcDNA 3.1 (-) EcoR1 site</td>
</tr>
</tbody>
</table>

Table 2.3. Oligos used for 5’RACE

<table>
<thead>
<tr>
<th>lab oligo</th>
<th>oligo name</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>725</td>
<td>i13-F (used for RT)</td>
<td>CTC CTG ACC TCA CTT AAG AGC GAG</td>
</tr>
<tr>
<td>kit</td>
<td>oligo (dT)-anchor primer 1 (provided in kit)</td>
<td>GAC CAC GCG TAT CGA TGT CGA TTT TTT TTT TTV (V = A, C, or T)</td>
</tr>
<tr>
<td>760</td>
<td>AFPi13/e14-Fq</td>
<td>TAA GAG CGA GCA CAC TGA CA</td>
</tr>
<tr>
<td>kit</td>
<td>anchor primer 2 (provided in kit)</td>
<td>GAC CAC GCG TAT CGA TGT CGA C</td>
</tr>
<tr>
<td>396</td>
<td>Afp 3'-F</td>
<td>TTG ATT TCC AAA ACT CGT GAG C</td>
</tr>
<tr>
<td>824</td>
<td>AFPds6.7-F</td>
<td>CTC ATC TCT TCA TAC CGC TTT</td>
</tr>
<tr>
<td>776</td>
<td>AFPds6.9-F</td>
<td>CCC AAA AGC TGC ATC ACC AC</td>
</tr>
<tr>
<td>831</td>
<td>Ex219-R</td>
<td>GTT CTC CTT GCT AGT GGG TAT CA</td>
</tr>
</tbody>
</table>
Table 2.4. Oligos used for asAFP promoter study

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>386</td>
<td>pGEMU</td>
<td>CACTATAGGGCGAATTGG</td>
</tr>
<tr>
<td>387</td>
<td>pGEML</td>
<td>CAAGCTATGCATCCACGC</td>
</tr>
<tr>
<td>726</td>
<td>AFP AS-F</td>
<td>CCG CTA CAG AGA AAG TGG AAT GA</td>
</tr>
<tr>
<td>779</td>
<td>AFP ds8.8-R</td>
<td>TTC GAG ATG CTC ACA AGG GT</td>
</tr>
<tr>
<td>860</td>
<td>AS5'Ex-R2</td>
<td>GTG AGA CCT TGT TCC TGT CTC C</td>
</tr>
<tr>
<td>861</td>
<td>AS5'Ex-F2</td>
<td>GGC AAC CAT ACA ACC AGC AAA G</td>
</tr>
<tr>
<td>874</td>
<td>asAFP-P1</td>
<td>GGA TGC AGG TGG ACC TTA AA</td>
</tr>
<tr>
<td>875</td>
<td>asAFP-P2</td>
<td>CCT CTC AGG CCC ATT GAT ATG</td>
</tr>
<tr>
<td>876</td>
<td>asAFP-P3</td>
<td>AAC GAT AGT AGG CTG ATG GAG AAG G</td>
</tr>
<tr>
<td>877</td>
<td>850-Bgl</td>
<td>AGA TCT GCG GTA TGA AGA GAT GAG ACC TTAA</td>
</tr>
<tr>
<td>914</td>
<td>asP-42-F</td>
<td>ACT AGT CAT GCC CTG TAA GCC TAG AC</td>
</tr>
<tr>
<td>942</td>
<td>asP-42-R</td>
<td>AGA TCTCAAGGTCTCACCCCATCGA</td>
</tr>
</tbody>
</table>

Table 2.5. Oligos used for asAFP functional study

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>266</td>
<td>AFPE12S</td>
<td>CCA GGC CGA CAT TTT CAT TGG ACA</td>
</tr>
<tr>
<td>266</td>
<td>AFPE12S</td>
<td>CCA GGC CGA CAT TTT CAT TGG ACA</td>
</tr>
<tr>
<td>267</td>
<td>AFPE12A</td>
<td>TGA GCT TGG CAC AGA TCC TTG TGG A</td>
</tr>
<tr>
<td>328</td>
<td>E9E10-F</td>
<td>CTT CAT GAA TAC TCA AGA ACT CAC C</td>
</tr>
<tr>
<td>386</td>
<td>pGEMU</td>
<td>CACTATAGGGCGAATTGG</td>
</tr>
<tr>
<td>387</td>
<td>pGEML</td>
<td>CAAGCTATGCATCCACGC</td>
</tr>
<tr>
<td>432</td>
<td>nest 12-13 BF</td>
<td>TTC TCC CAG TTT TCT TCA CCC CAG</td>
</tr>
<tr>
<td>635</td>
<td>Igμmbr</td>
<td>TTC TCA AAG CCT TCC TCC TCA GCA</td>
</tr>
<tr>
<td>706</td>
<td>8303 fw</td>
<td>TGG ACA CTT ACA AAC GCC TTC AGA G</td>
</tr>
<tr>
<td>761</td>
<td>AFPi13/e14-Rq</td>
<td>GCC CAA AGC ATC ACG AGT TT</td>
</tr>
<tr>
<td>835</td>
<td>i11-R1</td>
<td>CCT CAA CAA GAC CAG CAC AT</td>
</tr>
<tr>
<td>836</td>
<td>i11-F2</td>
<td>CCA AGC AAA GTA CAG GAG TAA GA</td>
</tr>
<tr>
<td>870</td>
<td>i8-F3</td>
<td>CTG AGC CTG GCT TAT TCA CTT A</td>
</tr>
<tr>
<td>872</td>
<td>i11-F</td>
<td>GGA CCG GAG TTT GGA CCT T</td>
</tr>
<tr>
<td>878</td>
<td>860-Nhe</td>
<td>GCT AGC GTG AGA CCT TGT TCC TGT CTC C</td>
</tr>
<tr>
<td>879</td>
<td>1012-Xba-T</td>
<td>GTG CTC TCT GTC TAG ATT ACT ACA TTA TCC</td>
</tr>
<tr>
<td>880</td>
<td>1012-Xba-B</td>
<td>GGA TAA TGT AGT AAT CTA GAC AGA GAG</td>
</tr>
<tr>
<td>909</td>
<td>3.1U</td>
<td>TTAATACGACTCAGTCTAGGGAGACC</td>
</tr>
<tr>
<td>915</td>
<td>e12del</td>
<td>TTT TGT CTG CCT GCT TCC TTC GTA AGA CAT TGT TAT TTT GAG CTG AGA C</td>
</tr>
<tr>
<td>916</td>
<td>E12-rev</td>
<td>ACT CTT GTT TCA TGG TCT GTA GGG</td>
</tr>
</tbody>
</table>

35
Table 2.6. Oligos used for PCR and sequencing in ADAR study

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>328</td>
<td>E9E10-F</td>
<td>CTT CAT GAA TAC TCA AGA ACT CAC C</td>
</tr>
<tr>
<td>386</td>
<td>pGEML</td>
<td>CACTATAGGGCGAATTTGG</td>
</tr>
<tr>
<td>387</td>
<td>pGEMU</td>
<td>CAAGCTATGCATCCAACG</td>
</tr>
<tr>
<td>714</td>
<td>i10-F2</td>
<td>GGT TCC AAC ACC CAA AGT ACA CGG T</td>
</tr>
<tr>
<td>759</td>
<td>AFPe/i12-Rq</td>
<td>AGG GCA GCA ATG TTC TAC TGA</td>
</tr>
<tr>
<td>761</td>
<td>AFPi13/e14-Rq</td>
<td>GCC CAA AGC ATC ACG AGT TT</td>
</tr>
<tr>
<td>811</td>
<td>AFPi9-R</td>
<td>CTC TGT GGA GTG CAG TTC TT</td>
</tr>
<tr>
<td>836</td>
<td>i11-F2</td>
<td>CCA AGC AAA GTA CAG GAG TAA GA</td>
</tr>
<tr>
<td>836</td>
<td>i11-F2 (sequencing)</td>
<td>CCA AGC AAA GTA CAG GAG TAA GA</td>
</tr>
<tr>
<td>840</td>
<td>i8-F3</td>
<td>CTG AGC CTG GCT TAT TCA CTT A</td>
</tr>
<tr>
<td>870</td>
<td>i11-R</td>
<td>CAG ATG AGA GTT GGG TAC CTG TG</td>
</tr>
<tr>
<td>SP6</td>
<td>universal primer</td>
<td>GAT TTA GGT GAC ACT ATA G</td>
</tr>
<tr>
<td>T7</td>
<td>universal primer</td>
<td>TAA TAC GAC TCA CTA TAG GG</td>
</tr>
</tbody>
</table>

Table 2.7 Oligos used for RT-qPCR

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>137</td>
<td>L30F</td>
<td>ATG GTG GCC GCC AAG AAG ACG AA</td>
</tr>
<tr>
<td>138</td>
<td>L30R</td>
<td>CCT CAA AGC TGG ACA GTT GTT GGC A</td>
</tr>
<tr>
<td>267</td>
<td>AFPE12A</td>
<td>TGA GCT TGG CAC AGA TCC TTG TGG A</td>
</tr>
<tr>
<td>324</td>
<td>E2E3-F</td>
<td>TCC ACG TTA GAT TCC TCC CAG TGC G</td>
</tr>
<tr>
<td>325</td>
<td>E2E3-R</td>
<td>TGT TCA CTT CCT CCT CGG TGG CTT C</td>
</tr>
<tr>
<td>432</td>
<td>nest 12-13 BF</td>
<td>TTC TCC CAG TTT TCT TCA CCC CAG</td>
</tr>
<tr>
<td>752</td>
<td>mAlbmlpF</td>
<td>TGA AAT GCG AGG TAA GTA TGG T</td>
</tr>
<tr>
<td>753</td>
<td>mAlbmlpR</td>
<td>AGG TTA CCC ACT TCA TTT TGC C</td>
</tr>
<tr>
<td>825</td>
<td>AS-RACE-2</td>
<td>GCA ACC ATA CAA CCA GCA AAG A</td>
</tr>
<tr>
<td>833</td>
<td>e9-R</td>
<td>AAT GAC TGA GAC AGG AAG GTT G</td>
</tr>
<tr>
<td>909</td>
<td>3.1U</td>
<td>TTAATACGACTCTATAGGGAGACC</td>
</tr>
<tr>
<td>910</td>
<td>3.1L</td>
<td>TAG AAG GCA CAG TCG AGG CTG ATC AG</td>
</tr>
<tr>
<td>1283</td>
<td>AFM-138</td>
<td>GGA TCC AGA AAT TTG CAC CAG GAC CGA A</td>
</tr>
<tr>
<td>1284</td>
<td>AFM-158</td>
<td>GGA TCC CCT AGT TAA TAA TTA CTC AGA</td>
</tr>
</tbody>
</table>

Copyright © Maria S. Dixon 2017
CHAPTER III
IDENTIFICATION AND CHARACTERIZATION OF NOVEL asAFP RNA TRANSCRIPTS IN MOUSE LIVER

Recent advances in high throughput RNA-sequencing (RNA-seq) have led to the discovery of thousands of unannotated noncoding RNA transcripts in the mammalian genome. Long noncoding RNAs (IncRNAs) are a heterogeneous group of RNA molecules that are > 200 nt in length and share similar features of mRNAs (Quan et al., 2015; Yan and Wang, 2012). Many identified IncRNAs are transcribed by RNA polymerase II (RNA pol II) and are polyadenylated, but lack significant open-reading frames. They can be identified through bioinformatics prediction or a high-throughput analysis such as microarrays and transcriptome analysis (Yan and Wang, 2012). They are generally poorly conserved, tissue-specific transcripts. IncRNAs can function in the nucleus or cytoplasm to regulate gene expression at the transcriptional or post-transcriptional level in a positive (concordant) or negative (discordant) manner. The expression of some IncRNAs are correlated with expression of neighboring genes, while others have inverse expression patterns during development and disease (Zhang et al., 2014). A number of well-characterized IncRNAs have provided clues about the biology of these transcripts and their proposed mechanisms, but a key question for each identified transcript is whether it has a biological function.

Our research has focused on the role of developmentally regulated protein coding genes including alpha-fetoprotein (AFP). AFP was the first HCC-associated oncofetal protein and is the most commonly used serum biomarker for liver damage and Hepatocellular Carcinoma (HCC). AFP is expressed abundantly in fetal liver, dramatically repressed at birth, and reactivated in liver damage and cancer. More than forty years ago, BALB/cJ mice were shown to express 20-fold higher AFP serum and
adult liver mRNA levels compared to 27 other inbred mouse strains (Belayew and Tilghman, 1982; Olsson et al., 1977). The locus controlling this post-natal persistence of AFP levels in BALB/cJ mice was named \textit{raf} (regulator of AFP), later named \textit{Afr1} (Alpha-fetoprotein regulator 1) (Belayew and Tilghman, 1982).

Our lab identified Zinc-fingers and homeoboxes 2 (Zhx2) as the gene that is responsible for continued expression of AFP and H19, another \textit{Afr1}-regulated gene, in adult BALB/cJ liver (Perincheri et al., 2005). The \textit{Afr1} allele of Zhx2 in BALB/cJ mice contains an ETnI\textalpha mouse endogenous retroviral element in intron 1 that drastically reduces Zhx2 expression in this strain. Low levels of Zhx2 in BALB/cJ mice correlates with increased levels of AFP and other developmentally regulated target genes including \textit{H19}, \textit{Glypican 3 (Gpc3)}, and \textit{Lipoprotein lipase}. However, it is not well known which genes are direct vs indirect targets of Zhx2 (Gargalovic et al., 2010; Morford et al., 2007; Pachnis et al., 1984).

Transgenic mouse studies showed that Zhx2 acts through the 250 bp AFP promoter and that a Zhx2 transgene driven by a liver specific promoter restored normal AFP silencing in adult BALB/cJ liver (Perincheri et al., 2005; Peyton et al., 2000a). However, transcriptional rates on the AFP and H19 genes, measured by nuclear run-on assays, were the same in the presence or absence of Zhx2 expression, indicating that Zhx2 may act at the post-transcriptional level (Morford, unpublished; Vacher et al., 1992). Therefore, the Peterson lab explored splicing across the AFP gene as possible post-transcriptional step that could be affected by Zhx2 regulation. In the presence of Zhx2, low levels of mature mRNA were detected, as expected, however, elevated levels of unspliced RNA from the AFP exon 10/11 region was also detected and this was not expected (Lilia Turcios, unpublished). To further characterize the “unspliced RNA”, strand-specific reverse transcriptase (RT) primers were used, followed by PCR. The
strand-specific RT primers showed that the “unspliced” RNAs were actually transcribed from the antisense DNA strand. Thus, these RNAs are now named antisense AFP (asAFP). To identify the ends of asAFP RNA, multiple overlapping RT-PCR reactions were performed, starting from the originally identified exon 10/11 region and extending in both the 5’ and 3’ directions. When a PCR product was not detected, this region was used to anchor 3’ and 5’ Rapid Amplification of cDNA ends (RACE) reactions. One product was smaller than predicted and identified a 338 bp intron in the asAFP RNA. The 3’ RACE results identified a major 3’ end in the region of AFP intron 8 (Qiu et al, in prep).

RESULTS

Identification of asAFP 5’ end

To identify the 5’ end of the asAFP RNA, I performed 5’ RACE on liver RNA from BALB/cJ mice that express a liver-specific Zhx2 transgene. These transgenic mice express similar levels of Zhx2 as BALB/c mice with wild type Zhx2 expression and postnatal repression of AFP. For these experiments, I used a gene-specific RT primer and two nested gene-specific PCR primers located within the confirmed region of asAFP but where standard PCR reactions using additional upstream primers did not detect products (Figure 3-1A). The final products from the second nested PCR reaction were cloned and sequenced (Figure 3-1B). I aligned various sized RACE clones to the UCSC genome database (NCBI37/mm9). Only one RACE clone mapped to a region ~6.8 kb downstream of the AFP gene in the opposite orientation. The 5’ RACE identified a 123 bp asAFP exon 1 (123bp) (Figure 3-1C). The remaining 5’ RACE clones were either found on a different locus or were from repetitive elements, suggesting artifacts produced by spurious priming in the PCR. Additional RT-PCR reactions, using a primer in this new 5’ exon and primers within the previously identified part of asAFP, identified a
219bp asAFP exon 2 that is spliced into some of the RNAs. This analysis also identified asAFP exon 3 as an alternative exon (Qiu et al, in prep). Based on these analyses asAFP is a 5kb transcript that contains 4 exons and 3 major spliced isoforms that partially overlaps the 3’ half of the AFP gene (Figure 3-2).

**asAFP is a noncoding RNA**

To investigate asAFP RNA coding potential, I first searched for open reading frames (ORFs) using NCBI Open Reading Frame Finder (ORF finder). This tool identified multiple predicted small ORFs of 98 amino acids or smaller (Figure 3-3A). BLAST analysis of the longer putative ORFs failed to reveal any known functional domains or homology to known proteins. To determine coding potential, I used the Coding Potential Assessment tool (CPAT) and analyzed asAFP RNA along with the AFP and albumin mRNAs as known coding genes and H19 RNA as a known non-coding gene. The CPAT tool predicts coding potential based on ORF features to distinguish between coding and noncoding RNAs. The protein coding genes AFP and albumin mRNAs were scored as coding RNAs with coding probabilities of 0.99. Like H19 RNA, asAFP RNA was scored as a noncoding RNA with coding probabilities of 0.30 and 0.04 respectively, suggesting that it is highly unlikely that any of the ORFs in asAFP encode proteins (Figure 3-3C).

To analyze the evolutionary conservation of mouse asAFP RNA with other mammalian species, I utilized UCSC Genome Browser on the Mouse Dec. 2011 (GRCm38/mm10) Assembly (Kent et al., 2002). I performed BLAT alignment of the asAFP genomic sequence that mapped to the mouse AFP genome. The multiple sequence alignment track (Multiz) and the Placental Mammal Basewise Conservation by PhyloP tracks were selected to display analysis of evolutionary conservation. My analysis revealed that the asAFP primary sequence is poorly conserved between all
species except for rat (Figure 3-4). Altogether, my results demonstrate that mouse asAFP is a novel, poorly conserved noncoding RNA.

**AFP and asAFP are co-expressed in adult mouse tissues**

I analyzed AFP and asAFP transcript levels, by RT-qPCR using total RNA isolated from a panel of tissues from adult BL/6 mice with a hepatocyte-specific deletion of Zhx2 (Zhx2\(^{\text{hep}}\)), including liver, spleen, small intestine, kidney, heart and lung. Both transcripts were predominantly expressed in the liver and to a lesser extent in the small intestine (Figure 3-5A). We also analyzed levels of asAFP and AFP mRNA in liver RNA from multiple sets of adult littermate BALB/cJ mice with or without a liver-specific Zhx2 transgene. These mice express normal or low Zhx2 levels, respectively. As expected, AFP mRNA levels were lower in the presence of Zhx2 and higher in the absence of Zhx2. However, the asAFP RNA levels were higher in the presence of Zhx2 and lower in the absence of Zhx2 (Figure 3.5B). Taken together, we demonstrate AFP and asAFP transcripts are co-expressed in mouse liver and the expression differs with the Zhx2 status of the mice, suggesting there may be a regulatory connection.

**asAFP RNA levels are not altered in liver tumors**

Previous studies have shown the expression of several antisense RNAs are dysregulated in HCC (Qin et al., 2016; Yuan et al., 2014; Zhu et al., 2016). I explored the expression profiles of hepatic AFP mRNA, Zhx2 mRNA and asAFP RNA levels by RT-qPCR using total RNA isolated from mouse model of HCC, as described (Creasy et al., 2016). Consistent with the findings of Creasy, et al., I observed increased hepatic AFP mRNA levels in tumors compared to normal livers. However, I observed no differences in Zhx2 mRNA or asAFP RNA expression between normal liver or liver tumors (Figure 3-6). These findings indicate that asAFP RNA expression is not altered in tumors and suggests that it may not be involved in regulating AFP mRNA during liver cancer.
Discussion

Alpha fetoprotein (AFP) was the first identified HCC-associated oncofetal protein and is still the most commonly used serum biomarker for HCC (Abelev, 1971; Behne and Copur, 2012). AFP is a member of the serum albumin gene family and is highly expressed in the fetal liver and yolk sac; serum AFP levels decrease dramatically during the first several weeks after birth due to the significant repression of AFP mRNA (~10^5-fold) in the liver (Belayew and Tilghman, 1982; Peterson et al., 2011). Postnatal AFP mRNA levels remain at low basal levels in normal adult liver, but AFP expression can be reactivated during liver regeneration and in HCC (Abelev, 1971). Our laboratory has been studying mouse liver gene regulation, focused on AFP, to better understand mechanisms by which changes in gene expression contribute to liver development, homeostasis and disease.

In this study, we identified novel antisense transcripts that are transcribed on the opposite strand of the AFP gene, named asAFP. We performed 5’ and 3’ RACE to characterize the genomic structure of asAFP. asAFP are ~5kb transcripts that partially overlap the 3’ half of AFP mRNA. asAFP contains 4 exons and 3 major spliced isoforms with similar abundance. In addition to the characterization data shown here, others in the lab have shown RNA-seq data from adult mouse liver confirmed the 5’ end, aligned to our experimentally validated splice junctions and supported the quantitative analysis of the relative isoform abundance. This data also identified additional asAFP RNAs that splice to locations further downstream from the major asAFP 3’ end. The subcellular location of asAFP RNA was assessed from cytoplasmic and nuclear RNA extracted from adult BALB/cJ mice with or without the Zhx2 transgene. asAFP RNA accumulated to significantly higher levels in the cytoplasm compared to the nucleus. The difference in
AFP mRNA accumulation in the presence or absence of Zhx2 was observed in the cytoplasmic RNA, but not in the nuclear RNA, suggesting that AFP mRNA regulation occurs at a step after nuclear mRNA processing (Qiu et al., in prep). In addition, cytoplasmic asAFP RNA and AFP mRNA abundance were in a similar range, but correlated with the Zhx2 status of the mice (Qiu et al., in prep). AFP mRNA levels were lower in the presence of Zhx2, whereas asAFP RNA levels were higher in the presence of Zhx2, suggesting a regulatory connection between Zhx2 and asAFP.

LncRNAs are considered noncoding because they lack open reading frames that are >100 amino acids and any ORFs present have no evolutionary conservation or homology to known protein domains (Zhang et al., 2014). I identified several predicted ORFs in asAFP RNA that are 98 aa or smaller but coding potential analysis suggests that asAFP is a noncoding RNA. Moreover, further characterization of asAFP sequences showed no evidence of evolutionary conservation with human. In addition, we demonstrated AFP and asAFP are co-expressed in adult mouse tissues and their levels of abundance differ with the status of Zhx2 in adult mouse liver. Our developmental studies revealed that both asAFP RNA and Zhx2 mRNA are expressed at low levels in the fetal liver and increase gradually after birth while AFP expression decreases (Qiu et al., in prep). To determine whether asAFP plays a role in regulating AFP mRNA reactivation in tumors, I measured AFP, asAFP, and Zhx2 expression in tumor and nontumor samples from a mouse model of HCC. Consistent with previous findings, AFP expression was increased in tumors. However, asAFP and Zhx2 mRNA levels were not altered. This suggests that asAFP is not directly involved in the mechanism by which AFP is reactivated in tumors. While Zhx2 mRNA levels do not change in liver tumors, ZHX2 protein levels have been shown to be frequently localized to the cytoplasm in human tumors, compared to its nuclear location in normal cells (Shen et al., 2008). Altogether, we have identified novel unannotated asAFP transcripts that partially overlap...
the AFP gene. Our observation that both AFP and asAFP are co-expressed in mouse liver and are inversely correlated depending on Zhx2, suggests a regulatory connection between these transcripts and Zhx2 during liver development. Therefore, it is important to understand asAFP transcriptional regulation and its potential biological function.
Figure 3-1. asAFP 5’ RACE strategy. Map of the AFP gene from exon 8 to about 7 kb downstream of exon 15. Above the line are the AFP exons (open boxes), with the exon and intron sizes shown. Below the line are shown the exons (grey boxes) and introns identified for the antisense transcripts, with their sizes shown. The transcriptional orientation of the two transcription units is shown above the map. (A) Schematic representation of asAFP 5’ RACE strategy. The primers used for the 5’ RACE are shown directly below the diagram. i13 was used for RT and e13 and e14 were used for the PCR reactions. (B) Gel from the 5’ RACE experiment. Multiple bands were cloned and sequenced; the arrow identified by (5’ RACE) is gene-specific and identified an asAFP exon located about 6.8 kb downstream from AFP exon 15. The other bands in the gel were from repetitive elements. (C) Visualization of asAFP 5’ RACE clone in UCSC Genome Browser, mm9 assembly. asAFP is transcribed beginning ~6.8 kb downstream from AFP gene and ~2kb upstream the adjacent, Afamin (Afm) gene.
Figure 3-2. Complete view of characterized asAFP transcripts. Map of the AFP gene from intron 6 to about 7 kb downstream of exon 15 (map not drawn to scale). Above the line are the AFP exons (open boxes), with the exon and intron sizes shown. Below the line are shown the asAFP exons (grey boxes) and introns identified for the antisense transcripts, with their sizes shown; the dark grey boxes are exons identified by RNA-seq. The transcriptional orientation of the two transcription units is shown above the map. The alternative splicing reactions identified for the asAFP transcripts are identified by the angled lines connecting alternative exons. RNA-seq data from adult mouse liver confirmed the asAFP exons 1-4 and identified two additional alternative exons located downstream from the major asAFP 3’ end (dark grey boxes) which were detected by RT-PCR and RT-qPCR and found to be low abundance RNAs (dotted lines). RNA-seq data received from Joe Locker (University of Pittsburgh) and analysis performed by Jinze Liu (University of Kentucky).
Figure 3-3. asAFP is a noncoding RNA. (A) Schematic representation of asAFP open reading frames (ORFs) adapted from https://www.ncbi.nlm.nih.gov/orffinder, oriented 5'-3' relative to the asAFP RNA. ORFs are delimited by AUG start codons and UAG, UAA or UGA termination codons. The ORFs within each reading frame are color coded in boxes with arrows and the numbers indicate the size, in amino acids (aa). The ORF cutoff was set at ≥ 30 aa. The two largest ORFs are 90 aa (5') and 98 aa (3'). (B) A screenshot of the Coding Potential Assessment Tool (CPAT) output when the AFP mRNA and asAFP RNA sequences were submitted to http://lilab.research.bcm.edu/cpat/index.php. Based on the coding probability (> 0.364 suggests coding sequence), this tool predicts that AFP mRNA clearly codes for a protein while asAFP RNA is highly likely to be noncoding.
Figure 3-4. Sequence conservation analysis of asAFP (chromosome 5, bp 90505000–90510000) in the UCSC genome browser (UCSC Genome Browser on the Mouse Dec. 2011 (GRCm38/mm10) Assembly (http://genome.ucsc.edu) (Kent, Sugnet et al. 2002). The RefSeq AFP transcript exon 8-15 (below) and asAFP cDNA (above). MultiZ Align, multiple sequence alignment of 30 vertebrates and Mammalian conservation tracks. The most magnified view at the bottom depicts species alignment with primates (green) and placental mammals (blue).
Figure 3-5. AFP mRNA and asAFP RNA are co-expressed in mouse tissues. (A) Expression of AFP mRNA and asAFP RNA were measured from BL/6 Zhx2\textsuperscript{hep} mouse tissues including small intestine, liver, testes, heart, muscle, and kidney, by RT-qPCR. Data shown represents one mouse. (B) RT-qPCR of total RNA of adult BALB/cJ littermates that express a liver-specific Zhx2 transgene (Zhx2\textsuperscript{Tg}) or without the transgene (Zhx2\textsuperscript{-Tg}); data from a single pair of animals but is representative of multiple animals. Ribosomal protein L30 was used as an internal control.
Figure 3-6. asAFP RNA levels are not altered in tumors. Male C3H-B6 mice were injected with phosphate buffer saline, control (PBS, n=4) or diethylnitrosamine (DEN, n=8) to initiate liver tumor development in a mouse model of HCC. The total liver RNA isolated from normal liver (PBS-injected) or HCC tumors (DEN-treated) mice was analyzed by RT-qPCR for expression of AFP, asAFP, and Zhx2. AFP mRNA levels are higher in tumors, but asAFP and Zhx2 mRNA levels are not altered. Expression was calculated relative to the ribosomal protein L30 internal control and the expression in normal liver tissue set to 1. Samples were provided by Kate Townsend Creasy.
CHAPTER IV
TRANSCRIPTIONAL REGULATION OF asAFP

The discovery of lncRNAs has gained extensive attention for their potential role in regulating gene expression. Like protein coding genes, lncRNAs display chromatin features associated with active transcription, such as trimethylation of histone H3 lysine 4 trimethylation (H3K4me3) at their promoters and histone H3 lysine 36 trimethylation (H3K36me3) throughout their gene body (Lin et al., 2015). The transcriptional activity of lncRNA’s can be regulated by standard promoters or enhancers to influence gene expression. Therefore, it is important to identify the transcriptional regulators that promote the expression of lncRNAs.

Eukaryotic gene expression is regulated by the interaction between the basal transcriptional machinery and the core promoter. The basal transcriptional machinery is composed of RNA pol II and basal transcription factors that direct accurate initiation of transcription. The core promoter is composed of diverse cis-acting regulatory elements that function independently or together to regulate transcriptional activity, such as the TATA box, an initiator (Inr), downstream promoter element (DPE), and motif ten element (MTE) (Figure 1-3) (Butler and Kadonaga, 2002). Only 20% of eukaryotic promoters contain a TATA box that directs the preinitiation complex to the transcriptional start site (TSS) within the Inr element. Transcription from TATA-less promoters also usually initiate within the Inr in conjunction with other core promoter elements. The Inr encompasses the TSS; the A within this sequence is designated as the +1 position. Transcription does not always begin at the A, but the Inr A+1 is the position that is used to count the spacing to other promoter elements, including the DPE and MTE. The function of these elements depends on strict distances from the Inr. In summary, the
core promoter is structurally and functionally diverse and multiple combinations of promoter elements can initiate transcription.

Enhancers have been shown to generate non-coding RNAs (eRNAs), many of which are short and unstable. However, some of these eRNAs can also be spliced, polyadenylated and stable (Kim et al., 2010; Lin et al., 2015; Orom and Shiekhattar, 2011a). eRNA levels often correlate with mRNA expression levels at nearby genes (Kim et al., 2010). While there are a few examples of eRNAs that contribute to enhancer function, possibly by looping the DNA sequences between enhancers to their target promoters or by regulating chromatin modifications that keeps enhancers and promoters in their active states, the vast majority have not been studied. Some eRNAs also may not play a direct role beyond marking the enhancer DNA as an open chromosomal region (Orom and Shiekhattar, 2011a).

We have characterized novel asAFP transcripts that partially overlap the 3’ half of the mouse AFP gene. We found that AFP mRNA and asAFP RNA are co-expressed in mouse liver and their abundance is inversely correlated in mouse strains that contain wild-type and low levels of Zhx2. To test the hypothesis that Zhx2 regulates the expression of asAFP RNA, which then contributes to AFP mRNA regulation, requires that we characterize the asAFP transcriptional control elements.

Results

Identification of cis-acting sequences required for asAFP transcriptional activity

To identify the genomic regions that promote asAFP transcription, we began with publicly available ENCODE data from the UCSC Genome Browser (http://genome.ucsc.edu/) (Raney et al., 2014). The mouse 8-week liver ChIP-seq data of chromatin modifications associated with active promoters and enhancers was visualized in the region of the AFP, asAFP, and Afamin (Afm) genomic region. The
regulatory elements that drive AFP expression are contained within a 7.6 kb region upstream of the AFP transcription start site and include the promoter and three enhancer elements. The three enhancers: E1, E2, and E3 are positioned about -2.5, -5.0 and -6.5 kb, respectively, and are each about 200-300 bp in length (Godbout et al., 1988). E1 and E2 are highly similar suggesting they arose from a duplication event; moreover, E1 is present in rodent but it is absent in several other species analyzed (Long et al., 2004). Transgenic studies showed that all three enhancers are active in both fetal and adult mouse livers (Ramesh et al., 1995). In the adult liver, E1 and E2 are active in all hepatocytes while E3 is active only in hepatocytes around the central vein (Peyton et al., 2000b).

We observed enrichment for histone modifications indicative of active transcription from promoter and enhancers, including: H3K4me1, H3K4me3, H3K9ac, and H3K27ac (Figure 4-1). The marks include H3K4me3 and H3K9ac that identify active promoters and a signal for H3K4me1, which is indicative for active enhancers in the absence of H3K4me3. Moreover, we observed a signal for H3K27ac, which associates with active promoters and enhancers (Raney et al., 2014). The gene downstream of AFP, Afm, is expressed in adult liver and the promoter, but enhancers have not been characterized (Liu et al., 2011). Since the transcription start site of asAFP is located ~2kb upstream of the Afm gene, it was possible that the DNA sequence driving asAFP transcription was also an enhancer of Afm transcriptional activity. To determine whether asAFP is transcribed by a standard promoter or originate as enhancer-associated RNAs (eRNAs), a 940 bp DNA fragment from BALB/c genomic DNA, spanning the region -830 to +112 bp relative to the asAFP TSS was PCR amplified and inserted into two different pGL3 luciferase plasmid vectors. One contains no transcriptional control elements and will detect promoter activity present in the inserted fragment while the other vector contains an SV40 early promoter and will detect
enhancer activity (Figure 4-2, top panel). I used the mouse liver hepatoma cell line, Hepa 1.6, as a model to study the functional and molecular characteristics of the mouse asAFP promoter. The Hepa 1.6 cells, derived from the BW7756 mouse hepatoma cells that arose from a C57/L mouse, secrete several serum albumin proteins including AFP, Albumin (Alb), alpha 1 antitrypsin (alpha-1-antitrypsin) and Amylase (Darlington et al., 1980). I transiently transfected the asAFP-luciferase plasmids into Hepa 1.6 cells with a Renilla luciferase-expressing control vector to normalize for variation in transfection efficiency. Luciferase and Renilla values were measured and compared to the control promoter-less and SV40 promoter vectors. The asAFP DNA segment increased luciferase levels 4-fold compared to the promoter-less plasmid, showing that this element had promoter activity. In contrast, this element did not affect the expression of the SV40 promoter, indicating that it does not have enhancer activity. Thus, asAFP RNA cannot be considered an eRNA and I have identified a fragment that has promoter activity (Figure 4-2, bottom panel).

**Analysis of asAFP promoter region**

To further localize the asAFP promoter activity, I constructed a series of 5’ deletions of the -830 to +112 promoter in the pGL3 basic luciferase vector (Figure 4-3, top panel). The asAFP promoter luciferase plasmids were transiently transfected into Hepa 1.6 cells and their luciferase activity was analyzed, as previously described. The 5’ deletion constructs showed no significant differences in promoter activity; the -90 to +112 construct maintained activity, suggesting that this segment contains the active promoter sequences that are important for asAFP expression (Figure 4-3, bottom panel). I also measured asAFP promoter activity in different cell lines including human embryonic kidney 293 (HEK 293) and the human hepatoma cell lines Hep3B and HepG2. The trends of transcriptional activity in all cells were similar to those observed in the Hepa 1.6 cells (data not shown).
Identification of asAFP core promoter elements

To identify the regulatory elements that influence asAFP expression, I used the core promoter element prediction tool, ElemeNT (http://lifefaculty.biu.ac.il/gershon-tamar/index.php/resources) to search for potential core promoter elements within the asAFP promoter (-90 to +112) sequence. The ElemeNT tool uses experimentally-validated sequences with known biological functions to predict core promoter elements (Sloutskin et al., 2015). Based on this analysis, the asAFP promoter contains predicted Inr and DPE elements in the region of the identified transcription start site (Figure 4-4A). Since the asAFP -90°luc maintained promoter activity, I examined the minimal sequences required for full promoter activity by measuring promoter activity of a 154 bp DNA region containing the asAFP promoter inserted into the pGL3 basic luciferase reporter vector to generate asAFP-42°luc. Transcription driven by the asAFP-42°luc declined relative to the -90 and -175 constructs, but did retain substantial promoter activity, suggesting important core promoter regulatory element(s) lie within this region (Figure 4-4C). To examine the function of the predicted Inr and DPE elements, additional constructs were generated that further reduced the size of the promoter-containing fragment (Figure 4-4B). I compared luciferase activity of intact asAFP promoters with constructs that contain sequence deletions up to position +29 within the DPE. Disruption of the DPE element significantly decreased the -175 to +29 promoter activity by ~2-fold, while the asAFP -42 to +29 promoter activity was also reduced by ~2-fold, resulting in loss of promoter activity (Figure 4-4C). These results indicate asAFP transcription is driven by a DPE-dependent promoter and the sequences within the DPE element are critical for full promoter strength. In addition, an element between -90 and -42 likely also contributes to promoter activity.
Zhx2 does not regulate the asAFP promoter

Previous studies showed that Zhx2 negatively regulates AFP, Cyclin A and Cyclin E promoters (Shen et al., 2008; Yue et al., 2012). However, a recent study showed that Zhx2 activates several of the mouse urinary protein (Mup) promoters, suggesting that Zhx2 may be a positive and negative transcriptional regulator of target genes (Jiang et al., 2017). During mouse liver development, both asAFP RNA and Zhx2 mRNA expression increases after birth, raising the possibility that asAFP is positively regulated by Zhx2. To determine whether Zhx2 regulates the asAFP promoter, I co-transfected the asAFP -830° luc plasmid with a pcDNA plasmid expressing Zhx2 using a 1:3 (luciferase:driver) ratio in Hepa 1.6 cells. I used the pGL3 promoter-less and Mup20°-luc plasmids as negative and positive controls, respectively. Data in Figure 4-5 shows the expression of Zhx2 increased Mup promoter activity about 1.5-fold, while no affect is observed on asAFP -830° promoter activity. Since a consensus Zhx2 binding site has not been identified for its targets genes, I explored Zhx2 regulation of an extended sequence of the asAFP promoter. A 2-kb region containing the asAFP promoter (-2160 to +112) was inserted into the pGL3 basic promoter-less luciferase reporter vector, to generate asAFP -2160°luc. When I co-transfected the asAFP -2160°luc with Zhx2, there was still no effect on luciferase expression (Figure 4-5). I tested several other asAFP promoter luciferase constructs with a range of Zhx2 plasmid ratios (1:0.5 to 1:10) in multiple cell lines (HEK293, HepG2, Hep3B), but did not observe any changes in promoter activity (data not shown). Although I observed a correlation between Zhx2 and asAFP in mice, my in vitro data indicates that Zhx2 does not appear to be involved in regulating the asAFP promoter.

Discussion

In this chapter, I showed that asAFP transcriptional activity is driven by a standard promoter and not an enhancer, as is the case with eRNAs. Deletion analysis
of the asAFP promoter identified the regulatory regions that are important for asAFP transcriptional activity. There was no significant difference in activity between the promoter fragment beginning at -830 and the 5’ deletion constructs down to -90. Since the asAFP 5’ deletion constructs displayed similar promoter activity, I performed computational analysis that identified an Inr and DPE element that comprise the asAFP core promoter. I tested the function of predicted core promoter elements by comparing activity of asAFP promoters that contained the deletions of the DPE sequences up to position +29. Mutational analysis showed that transcription driven by the asAFP promoter (-175 to +29) resulted in ~2-fold reduction in promoter activity, while a complete loss of activity was observed in the -42 to -29 promoter. My data clearly demonstrates that the asAFP promoter requires an intact DPE sequence that is important for full promoter strength and transcription initiation.

Previous studies suggest Zhx2 is a transcriptional regulator of target genes; yet the mechanism by which Zhx2 regulates target gene expression is not fully understood. We previously showed the expression of asAFP RNA and Zhx2 are positively correlated in animals expressing normal and low levels of Zhx2. In addition, data from others in the Spear and Peterson lab indicate asAFP and Zhx2 transcript levels are positively correlated during liver development, suggesting asAFP may be a potential target of Zhx2. However, when a pcDNA plasmid expressing Zhx2 was co-transfected the asAFP830 luc plasmid, Zhx2 had no effect on promoter activity. I also cloned a larger promoter fragment, in case a Zhx2 responsive site was located further from the start site. However, consistent with the asAFP830 results, I did not observe changes in promoter activity, suggesting Zhx2 does not regulate asAFP transcriptional activity.

Multiple liver-enriched transcription factors are involved in the regulation of hepatic gene expression, such as the hepatic nuclear factors (HNFs), CCAAT/enhancer-binding protein beta (C/EBP), and D-binding protein (DBP) (Schrem et al., 2002).
identify potential liver-enriched transcription factors that may regulate asAFP promoter activity, I used the JASPAR transcription factor (TF) prediction program (jaspar.genereg.net) to screen for potential binding sites in the asAFP -90 to +112 promoter sequence. Several predicted TF’s were identified, including putative binding sites for Transcription Factor 4 (TCF4), Hepatocyte nuclear factor 4 alpha (HNF4a), and Nuclear factor 1 alpha (NF1a). I examined their effect on the asAFP promoter by co-transfecting the -90\(^{\text{th}}\) luc promoter with individual pcDNA plasmids expressing TCF4, HNF4a, and NF1a into Hepa 1.6 cells. My preliminary data suggested that the expression of HNF4a and NF1a did not influence asAFP promoter activity, whereas TCF4 reduced promoter activity (data not shown). Curiously, the predicted TCF4 binding site is within the Inr motif of the asAFP promoter, suggesting TCF4 binding may interfere with initiation of asAFP transcription. However, the mechanism by which TCF4 influences asAFP transcriptional repression will require further investigation.
Figure 4-1. ENCODE tracks from ChIP-seq data of the AFP/asAFP/Atm genomic region. Histone tracks for the following chromatin modification marks for active promoters (H3K4me3, H3K9ac) enhancers (H3K4me1, in the absence of H3K4me3) or promoters/enhancers (H3K27ac) were extracted from the publicly available data (http://genome.ucsc.edu/). The transcriptional directions of the genes and asAFP are shown; the asAFP exon 1 is identified and corresponds to a clear peak of histone modifications. Two of the three well-characterized AFP enhancers (E1, E2) are also identified (Raney et al., 2014).
will detect promoter activity

will detect enhancer activity

Luciferase/Renilla normalized to control vector
Figure 4-2. Identification of the asAFP promoter. Top panel. Schematic of asAFP putative promoter/enhancer cloning strategy. Bottom panel. A 940 bp fragment of asAFP predicted promoter/enhancer from -830 to +112 was cloned upstream of luciferase in the promoter-less and SV40 early promoter pGL3 luciferase plasmids. The luciferase constructs were transfected together with TK-Renilla, which served as an internal control for transfection efficiency, into Hepa 1.6 cells. pGL3-basic and pGL3-SV40-promoter were also transfected as controls. 48 hours later, the relative luciferase values were determined. The promoter activity is expressed relative to the activity of the pGL3 basic and pGL3 promoter luciferase vectors. The results represent the mean +/- SD of 8 independent transfections, *** p < 0.001.
**Figure 4-3.** Deletion analysis of the asAFP promoter. (A) Schematic illustration of a series of 5' deletions of the mouse asAFP promoter. (B) asAFP 5' deletion promoter luciferase constructs were transfected together with TK-Renilla into Hepa 1.6 cells that served as an internal control for transfection efficiency. 48 hours later, the relative luciferase values were determined. The promoter activity is expressed relative to the activity of the pGL3 basic luciferase vector. The results represent the mean +/- SD of 5 independent transfection experiments.
A predicted asAFP promoter elements

-90

Initiator

+1

DPE

Downstream Promoter Element

+112

B

-175

-65

-24

+29

+112

Inr

DPE

-175 to +112

-42 to +112

-175 to +29

-42 to +29

C

Normalized to pcDNA (Luciferase/Renilla)

basic

-175 to +112

-175 to +29

-42 to +112

-42 to +29

***

NS
**Figure 4-4.** Identification of the asAFP core promoter elements. asAFP -90 to +112 sequence was used as the input sequence for ElemeNT computational analysis of asAFP putative core promoter elements (http://lifefaculty.biu.ac.il/gershon-tamar/index.php/resources). (A) Schematic of asAFP predicted core promoter elements. (B) Schematic illustration of the intact asAFP promoter (-175 to +112) and deletion constructs (-42 to +112), (-175 to +29) and (-42 to +29). (C) asAFP promoter constructs were transfected together with TK-Renilla into Hepa 1.6 cells that served as an internal control for transfection efficiency. 48 hours later, the relative luciferase values were determined. The promoter activity is expressed relative to the activity of the pGL3 basic luciferase vector and was set to 1. The results represent the mean +/- SD of 3 independent transfection experiments, ** p <0.01, *** p < 0.001.
Figure 4-5. Zhx2 does not regulate the asAFP promoter. Hepa 1.6 cell were transfected with pGL3 basic, -830<sup>P</sup>-Luc, -2160<sup>P</sup>-Luc or Mup20<sup>P</sup>-Luc reporter genes along with pcDNA 3.1 empty vector or the Zhx2 expression vector at a 3:1 ratio; Renilla luciferase was also included to control for variations in transfection efficiency. After 48 hrs, cells were harvested, and lysates were analyzed for firefly/Renilla luciferase. The asAFP -830 and -2160 promoters were not responsive to Zhx2, while the Mup20 promoter was activated by Zhx2. The results represent the mean +/- SD of 3 independent transfection experiments, *** p < 0.001.
CHAPTER V
CHARACTERIZATION OF asAFP RNA MECHANISM OF ACTION

LncRNAs have gained widespread attention in recent years as new and potentially crucial biological regulators. While only a small number of LncRNAs have been fully characterized to date, it is believed that LncRNAs are involved in a range of developmental processes and diseases. Understanding the mechanisms by which LncRNAs act is still surprisingly limited. It is estimated that more than 70% of mammalian sense transcripts have antisense partners, suggesting a central role of antisense transcripts in regulating eukaryotic gene expression (Kung et al., 2013). Several studies have shown antisense RNAs can negatively or positively modulate expression of their overlapping sense transcript in cis or in trans at the transcriptional and posttranscriptional level (Pelechano and Steinmetz, 2013). There are several proposed mechanisms by which antisense RNAs can regulate their sense partner, including transcriptional interference, epigenetic modification and chromatin remodeling. Additionally, the genomic arrangement of some antisense RNAs allow them to base pair with their sense partner mRNAs in the nucleus or cytoplasm to potentially regulate processes such as RNA stability, transport and translation (Cech and Steitz, 2014; Mahmoudi et al., 2016; Pelechano and Steinmetz, 2013; Wilusz et al., 2009).

Antisense RNAs that regulate gene expression in the nucleus can alter transcription, modulate alternative splicing or recruit editing enzymes to dsRNAs, leading to alterations in subcellular localization, transport and RNA stability (Faghihi and Wahlestedt, 2009). For example, the expression of antisense RevErbAα inhibits splicing of the thyroid hormone receptors isoforms, TRα2 and TRα1. The interaction between RevErbAα and TRα2, negatively regulates TRα2 mRNA expression, which alters the balance and function of the two isoforms (Hastings et al., 1997). Antisense RNAs have
been linked to influencing mRNA editing by altering RNA structure, coding potential, or stability. In *Drosophila melanogaster*, the interaction between 4f-rnp (RNA-binding protein 4F) mRNA and its antisense sas-10 recruits adenosine deaminases that act on RNA (ADAR) enzymes that convert adenosines to inosines (A-to-I) in the overlapping region. A-to-I RNA editing decreased 4f-rnp mRNA abundance by decreasing the stability of the 4f-rnp mRNA (Peters et al., 2003).

Antisense RNAs that regulate gene expression in the cytoplasm are mainly involved in modulating processes such as mRNA stability or translation (Faghihi and Wahlestedt, 2009). For instance, the formation of the RNA duplex between BACE1-AS and BACE mRNA masks the binding site for miR-485-5p, resulting in an increase of BACE mRNA stability (Faghihi et al., 2010). In contrast, antisense RNAs can facilitate the degradation of mRNAs by promoting Staufen-mediated mRNA decay (SMD) via binding of Stau1 dsRNA binding protein. Gong and Maquat demonstrated that Stau1 binding sites are formed by imperfect base-pairing between an Alu element in the 3'UTR of the target mRNA and another Alu element in a cytoplasmic, poly-adenylated IncRNA (1/2-sbsRNA) that recruits Stau1 for mRNA degradation (Gong and Maquat, 2011).

Antisense RNAs can also promote or repress translation of target mRNAs. Carrieri, et al. demonstrated that cellular stress conditions caused the nuclear antisense ubiquitin carboxy-terminal hydrolase L1 (AS Uchl1) RNA translocation to the cytoplasm where it bound to the 5' end of Uchl1 mRNA to increase translation (Carrieri et al., 2012). Conversely, it has been reported that lincRNA-p21 inhibits translation of CTNNB1 (β-catenin) and JUNB transcripts through imperfect base pairing with several sites in the coding and UTRs of both transcripts (Huarte et al., 2010). Despite the widespread attention IncRNAs have gained in recent years, the full extent of their biological significance is not well understood (Kung et al., 2013). We have identified novel antisense AFP (asAFP) RNA transcripts that partially overlap the 3' half of the mouse
AFP gene. AFP mRNA and asAFP RNA are co-expressed in the adult mouse liver and are inversely correlated in the presence or absence of Zhx2. This suggests a regulatory connection between these two RNAs, which will be examined here.

Results

asAFP RNA regulates AFP in trans

To investigate whether asAFP regulates AFP mRNA expression, various asAFP expression constructs that together cover the full-length asAFP transcript were generated as shown in Figure 5-1A. As a negative control, a segment of an IgM Cµ intron was inserted into pcDNA expression vector. These constructs were transiently transfected in Hepa 1.6 cells and changes in endogenous AFP mRNA was measured by RT-qPCR. To confirm the expression of asAFP RNAs, I analyzed Hepa 1.6 cells transfected with asAFP constructs by RT-qPCR using asAFP-specific primers. To avoid amplification of contaminating plasmid DNA, total RNA was treated with RNase-free DNase I and reverse transcribed. As controls, a no reverse transcription (NRT) and no template controls (NTC) were used to confirm the absence of contamination. Expression of transfected asAFP RNA segments was detected in transfected Hepa 1.6 cells although the levels varied among experiments (Figure 5-1B). Interestingly, overexpression of all segments of asAFP RNA significantly decreased endogenous AFP mRNA levels, while with IgM expression, levels were not affected (Figure 5-1C). These results suggest asAFP negatively regulates AFP mRNA expression in trans.

Because all asAFP RNA constructs downregulated AFP mRNA to similar levels, I utilized flow cytometry analysis to separate the transfected cells that express asAFP RNA from the untransfected cells using a plasmid expressing Green Fluorescent protein (GFP) as a reporter. The asAFP mid expression construct was co-transfected with GFP plasmid into Hepa 1.6 cells. As negative controls, cells were co-transfected with pcDNA
empty vector-GFP and an expression vector that contains AFP exons 9-14 (+) cDNA-GFP in the sense orientation. Since changes in endogenous AFP mRNA is measured using primers in AFP exons 1 and 2, the expression of AFP e9-14(+) plasmid should not affect AFP mRNA expression or detection. 48 hrs after transfection, cells were harvested and sorted into GFP-positive and GFP-negative populations and endogenous AFP mRNA levels in each were analyzed by RT-qPCR. Flow analysis showed that 49%, 34%, and 40% of cells transfected with pcDNA, asAFP mid, and AFP e9-14 (+), respectively were GFP positive (Figure 5-2A). Consistent with my previous transient transfection data, the expression of asAFP mid plasmid reduced AFP mRNA levels in the unsorted population, while expression of AFP exon 9-14(+) did not affect AFP mRNA expression, as expected (Figure 5-2B). The asAFP mid-expressing GFP positive cells showed reduced AFP mRNA levels in comparison to the unsorted transfected cells, whereas the GFP-negative cells did not have decreased AFP mRNA levels (Figure 5-2B). When the data was plotted to compare the AFP mRNA from the GFP+ and GFP- populations from a single sort experiment, the asAFP mid plasmid substantially decreased of AFP mRNA levels (Figure 5-2C). These findings confirm that asAFP RNA downregulates AFP mRNA expression because the reduced endogenous AFP mRNA is only seen in the cells that are actually transfected with the asAFP constructs.

**asAFP does not regulate neighboring genes**

AFP is a member of the serum albumin gene family that is comprised of albumin, alpha-fetoprotein, alpha-albumin (afamin), and the more distantly related Vitamin D binding protein (Peterson et al., 2011). These genes arose from a common ancestor through a series of duplication events, and are expressed primarily in the liver (Alexander et al., 1984; Naidu et al., 2010). AFP is part of a gene cluster, with the
neighboring genes Alb and Afm located 14.1 kb upstream and 10.0 kb downstream of the AFP gene, respectively (Figure 5-3A). Some antisense RNAs can mediate widespread transcription to modulate the expression of neighboring genes in cis or trans by the act of transcription or through the recruitment of chromatin-modifying complexes, such as Polycomb repressive complex 2 (PRC2) to genomic loci (Pelechano and Steinmetz, 2013). To test the extent of asAFP mediated regulation, I analyzed changes in AFP, Afm and Alb mRNA levels in liver RNA from multiple sets of adult C57/Bl6 wild-type (Zhx2 +/+) and whole-body Zhx2 knock-out (Zhx2 -/-) littermate mice. As expected, AFP mRNA levels were lower in the presence of Zhx2, while asAFP RNA levels were higher in the presence of Zhx2. Alb and Afm mRNA levels were not affected by the status of Zhx2 or differences in asAFP RNA in the mice (Figure 5-3B). I also measured changes in AFP, Alb, and Afm mRNA after overexpressing the asAFP mid expression construct in Hepa 1.6 cells (Figure 5-4). My data show that while elevated expression of asAFP significantly reduced AFP mRNA levels, Afm and Alb mRNA levels were not affected. Together, my in vivo and in vitro data demonstrate that asAFP does not regulate the expression the neighboring albumin genes and confirms the specificity of asAFP’s downregulation of AFP mRNA.

asAFP segments that overlap with exons more effectively repress AFP

Because all asAFP RNA segments were able to downregulate AFP mRNA, I further examined which segments may be more effective in AFP downregulation. I analyzed whether the portions of asAFP RNAs that overlap with AFP introns or exons were more effective in decreasing AFP mRNA levels. If the sequences of asAFP RNA that can base-pair with AFP introns are more active, this would likely require the regulatory mechanism to occur in the nucleus before the AFP mRNA is processed. If, however, the portion of the asAFP RNA that can base-pair with the AFP exons is more
active, this means the regulation likely occurs after mRNA processing and before or during cytoplasmic accumulation. Using the asAFP mid pcDNA construct, I performed PCR mutagenesis to remove the exon 12 sequences (delta e12) and also generated a pcDNA plasmid that expresses just exon 12 in the antisense orientation, (e12-) (Figure 5-5 A). The delta e12 expression plasmid contains the overlapping intron sequences and the e12- expression plasmid contains the overlapping exonic sequences from the original asAFP mid construct. I transiently transfected these plasmids into Hepa 1.6 cells along with the starting asAFP mid plasmid and pcDNA empty vector control and measured endogenous AFP mRNA levels by RT-qPCR. As previously described, the expression of asAFP RNAs were confirmed by RT-qPCR and exhibited higher expression levels than the pcDNA control (Figure 5-5B). Figure 5-5C shows that expression of both delta e12 and e12 (-) constructs significantly downregulated endogenous AFP mRNA. However, the construct that overlaps the exon, e12(-), had a greater effect on AFP mRNA downregulation than the construct that overlaps the intron sequences, suggesting this regulation possibly occurs during nuclear export or in the cytoplasm.

I also separated the transfected cells that express asAFP from untransfected cells by flow cytometry as previously described. I co-transfected this set of asAFP RNA plasmids with a plasmid expressing GFP in Hepa 1.6 cells. As negative controls, cells were transfected with pcDNA empty vector-GFP empty vector. 48 hrs after transfection, the cells were sorted into GFP-positive and GFP-negative populations by flow cytometry and endogenous AFP mRNA levels in each were analyzed by RT-qPCR. Flow analysis showed that 50% to 60% of cells transfected with the pcDNA, asAFP mid, delta e12, and e12 (-) expression plasmids were GFP positive (Figure 5-6A). In both the unsorted and sorted GFP+ populations, delta e12, and e12 (-)-transfected samples decreased AFP mRNA to similar levels (Figure 5-6B). Curiously, the GFP- populations
showed higher levels of AFP mRNA compared to the pcDNA empty vector control (Figure 5-6B). However, when I plotted the data to compare the AFP mRNA from the GFP+ and GFP- populations from this single sort experiment, while all the asAFP plasmids substantially decreased AFP mRNA, e12 (-) showed a marginally greater decrease than delta e12 (Figure 5-6C). Together, my data suggest that the complementary regions between AFP mRNA and asAFP RNA may contribute to AFP mRNA downregulation and the overlapping exonic sequences are more effective.

**RNA editing of AFP and asAFP complementary regions**

To explore the possibility that asAFP regulates AFP by affecting mRNA decay, I tested whether overlapping sequences of AFP and asAFP RNA form dsRNA duplexes that could serve as substrates for A-to-I RNA editing by ADAR. A-to-I edited sites can be detected as A>G conversions in cDNA. To investigate the possible role of RNA editing, we examined the sequences of the cDNA clones that were generated to characterize the extent of asAFP genomic structure. These clones were generated by RT-PCR from adult liver RNA of BALB/cJ mice that express the Zhx2 transgene but did not use a high-fidelity polymerase. To distinguish between RNA editing changes and DNA polymorphisms, we screened for unique edited sites in one clone that did not appear in another and mapped the sequences to the compiled asAFP cDNA sequence. This analyses identified four potential A>G changes due to editing distributed randomly throughout on the antisense strand (Figure 5-7, above dashed line). This observation led me to determine whether ADAR activity may result in RNA editing throughout the complementary sequences between AFP and asAFP. Although editing mostly occurs in noncoding regions such as UTRs or introns, a small fraction of editing occurs in coding regions that could change amino acid sequences. Because my data suggest that the overlapping exons are likely more effective in mRNA downregulation, I examined the presence of editing in the sequences that overlap with AFP exons 9-14. I used a high-
fidelity DNA polymerase enzyme in the PCR reactions to avoid mutations introduced during the amplification, followed by cloning and Sanger sequencing. In addition, to control for possible errors introduced by DNA polymerase infidelity, I used the lowest number of PCR cycles that resulted in detectable PCR products. My sequencing analysis of 51 cDNA clones did not show any evidence of sequence differences that could be attributed to RNA editing in the AFP coding regions (Figure 5-7). I also assessed the presence of A-to-I RNA editing in the AFP/asAFP overlapping genomic regions, focusing on those where potential A>G changes were observed previously in single clones. I generated multiple cDNA segments within the overlapping sequences between AFP and asAFP transcripts (Figure 5-7). My analysis of 142 cDNA clones that spanned ~600 bp to >2kb sequence overlap did not show any indication A>G sequence changes, but did detect two clones with T>C conversions, suggesting editing on the antisense RNA strand of the sense-antisense duplex. After second-strand cDNA synthesis during PCR, these sites would then be visualized as T>C conversions on the antisense strand (Peters et al., 2003). Additionally, I found multiple single nucleotide polymorphisms (SNPs) within these regions; however their biological significance is unknown. Previous studies by Peters et al, reported extensive ADAR-mediated RNA editing within the overlapping regions of Drosophila 4f-rnp and sas-10 transcripts during embryo development. During development, the frequency of RNA editing detected was 7.8% in the late embryo stage (4/51 clones that covered 162 bp of overlap), increased to 24.4% in adult heads (11/45 clones) and 16.7% in adult bodies (8/48 clones) (Peters et al., 2003). In contrast, the frequency of potential RNA editing sites found here detected was 0% of the 23 clones that cover ~600 bp overlap and 7.4% of the 27 clones that cover more than 2kb of the overlap. The pooled results indicate 2 of the 50 cDNA clones sequenced had evidence of potential RNA editing on the antisense strand. These low levels of potential RNA editing prompted me to analyze nuclear cDNA clones in
attempts to detect higher levels of editing in the AFP/asAFP overlap regions. Again, I was unable to detect evidence of editing in the 41 cDNA clones that covered more than 2kb of the overlap (Figure 5-7). Even though I can not determine how many unique transcripts are represented by the multiple clones I sequenced, similar approaches in other labs have detected higher levels of potentially edited sites. Thus, I conclude form this analysis that the complementary regions between AFP and asAFP are not extensively edited by ADAR enzymes.

Discussion

Sense and antisense transcripts can interact through their complementary regions, resulting in dsRNA that can alter mRNA stability, transport, and translation (Werner and Sayer, 2009). I show here that overexpression of all segments of asAFP RNA negatively regulate AFP mRNA expression in trans. Because asAFP RNA is expressed at very low levels in Hepa 1.6 cells, I utilized this approach to study the effects of increased levels of asAFP RNA on the endogenous expression of AFP mRNA. Although overexpression of individual genes has been widely used as a tool for connecting genes to biological pathways, this approach does not always produce biologically relevant results and should be confirmed by alternate methods (Prelich, 2012). To complement my overexpression studies, cell lines that stably express asAFP would allow knock-down studies of asAFP RNA using both siRNA and antisense oligos (ASO) to measure the changes on AFP mRNA levels. Because siRNAs function mainly in the cytoplasm and ASO effects depend on nuclear RNase H, these experiments would further help distinguish whether the regulation of AFP mRNA by asAFP RNA occurs in the cytoplasm or the nucleus. In attempts to complement my overexpression experiments, I used G418 as a selection marker to generate stable cells lines.
overexpressing asAFP 5-1 and asAFP mid constructs. However, I was unable to generate clones that substantially express asAFP RNA when measured by RT-qPCR.

Antisense transcription has been shown to regulate neighboring genes that are transcribed on the same genomic locus; therefore, I determined whether asAFP RNA could regulate AFP’s neighboring genes. I showed that the expression of asAFP does not affect Alb or Afm mRNA levels both in vitro and in vivo, confirming the specificity of asAFP for AFP mRNA regulation.

My Hepa 1.6 transfection data showed that the expression of various segments of asAFP RNA decreased AFP mRNA levels. Because only a portion of the cells are transfected, to confirm my results with representative constructs, I separated transfected from non-transfected cells by sorting for GFP+ cells. Clearly, AFP mRNA was downregulated by the asAFP mid construct in the transfected, but not the non-transfected cell populations. As an independent way to identify the cellular location by which asAFP regulates of AFP mRNA, I examined whether the asAFP sequences that were complementary to AFP intronic or exonic sequences had a greater effect on AFP expression. My transient transfection data showed that while both complementary intronic and exonic sequences downregulated AFP mRNA expression, the exonic sequences were more effective. Curiously, I observed that the asAFP mid and exon 12 (−) constructs decreased AFP mRNA to similar levels even though the amount of asAFP sequence in each differ greatly (1800 bp in asAFP mid and 224 bp in e12(−)). One interpretation of this result is that the exonic sequences, which are the same in these two constructs, are the active sequences. Together these findings suggest that asAFP RNA is exerting its effect on AFP mRNA at a step after RNA splicing, possibly during nuclear-cytoplasmic transport or in the cytoplasm. Because the segments that overlapped with intronic sequences decreased AFP mRNA levels, I also explored the asAFP regulation
of AFP mRNA stability via ADAR mediated RNA editing. I analyzed multiple clones that contained complementary sequences between AFP and asAFP but there was no indication of extensive editing in AFP and asAFP dsRNA. Previous studies suggested ADAR proteins can affect gene expression independently of its deamination activity (Wang et al., 2013a). Future studies would examine changes in asAFP RNA and AFP mRNA expression following siRNA-mediated knockdown of the catalytic active ADAR 1 and ADAR 2 enzymes in Hepa 1.6 cells.
A

![Diagram of AFP mRNA with EcoRI, Xhol, and Xbal digestion sites.](image)

- asAFP 3' end (2000 bp)
- asAFP mid (1800 bp)
- asAFP 3' + mid (3800 bp)

B

<table>
<thead>
<tr>
<th>Construct</th>
<th>Transfection Date</th>
<th>asAFP/L30</th>
<th>pcDNA/L30</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>asAFP 5-1</td>
<td>5/16/2015</td>
<td>0.000006</td>
<td>0.000002</td>
<td>909/825</td>
</tr>
<tr>
<td></td>
<td>7/30/2015</td>
<td>0.000012</td>
<td>0.000003</td>
<td></td>
</tr>
<tr>
<td>asAFP 5-2</td>
<td>5/16/2015</td>
<td>0.135954</td>
<td>0.000002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7/30/2015</td>
<td>0.000047</td>
<td>0.000003</td>
<td></td>
</tr>
<tr>
<td>asAFP mid</td>
<td>5/16/2015</td>
<td>0.019669</td>
<td>0.000197</td>
<td>758/759</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.000931</td>
<td>0.000036</td>
<td>909/432</td>
</tr>
<tr>
<td></td>
<td>11/24/2015</td>
<td>0.011134</td>
<td>0.002357</td>
<td>758/759</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.000266</td>
<td>0.000044</td>
<td>909/432</td>
</tr>
<tr>
<td></td>
<td>10/17/16-2</td>
<td>0.000235</td>
<td>0.000001</td>
<td></td>
</tr>
<tr>
<td>asAFP 3' end</td>
<td>5/16/2015</td>
<td>0.016324</td>
<td>0.000197</td>
<td>910-833</td>
</tr>
<tr>
<td></td>
<td>7/30/2015</td>
<td>0.000003</td>
<td>0.000000</td>
<td></td>
</tr>
<tr>
<td>asAFP 3+mid</td>
<td>11/24/2015</td>
<td>32.742320</td>
<td>0.000000</td>
<td></td>
</tr>
</tbody>
</table>

C

![Bar chart showing normalized AFP/L30 levels.](image)

- pcDNA
- 5-1
- 5-2
- mid
- 3' end
- 3' + mid
- IgM

- **NS**: Not significant
- *******: Significant at p < 0.001
Figure 5-1. asAFP RNA regulates AFP in trans. (A) asAFP pcDNA cloning strategy. Segments of asAFP cDNA (blue open boxes) that together cover the entire asAFP transcript were cloned into pcDNA3.1 and transiently transfected into the mouse liver cell line Hepa-1.6 along with an empty vector (pcDNA). The numbers indicate sizes, in nt, of the asAFP sequence included in the construct. (B) Total RNA was isolated, DNAsε-treated, and reverse-transcribed. The expression of asAFP RNAs and pcDNA control were measured by RT-qPCR using the asAFP-specific primers indicated and normalized to ribosomal protein gene L30 mRNA. (C) Endogenous AFP mRNA levels were analyzed by RT-qPCR 48 hrs after transfection with the asAFP constructs shown and normalized to L30 mRNA. A segment from the IgM gene cloned into pcDNA was used as a negative control. The fold-change was determined by comparison of asAFP RNA levels with those of the pcDNA empty vector control. The results represent the mean +/- SD of ≥7 independent transfection experiments, *** p <0.001.
A

Control (pcDNA)  pcDNA-GFP  asAFP Mid-GFP  AFP exon 9-14(+) -GFP

Green Fluorescence
B

![Graph showing fold change normalized to pcDNA (AFP/L30)]

- **Unsorted**
- Sorted GFP+
- Sorted GFP-

C

![Graph showing fold change of AFP]
Figure 5-2. Sorted cells confirm that asAFP RNA regulates AFP in *trans*. (A) Density plots of Hepa 1.6 cells transfected with pcDNA alone (control), or pcDNA, asAFP mid, and AFP e9-14(+) co-transfected with pEGFP. The GFP+ (red boxes) and GFP-negative (black boxes) cell populations were sorted by flow cytometry. (B) RT-qPCR analysis of endogenous AFP mRNA levels of unsorted, GFP positive and GFP negative populations using L30 mRNA as a control. The data were normalized to the pcDNA empty vector control. (C) Fold-change was determined by comparison of AFP mRNA levels from GFP-positive sorted cells with those to GFP-negative sorted cells. This data is from one transfection and sorting experiment; the cDNA was analyzed twice in duplicates, ** p <0.01.
**Figure 5-3.** asAFP does not regulate neighboring genes. (A) Map of the mouse Alb, AFP and Afm locus, showing the relative position of the asAFP. The location and distance between the genes are shown, numbers below indicate the size of the genes and intergenic regions (in kb). (B) AFP, asAFP, Alb, and Afm transcript levels were measured by RT-qPCR from four sets of adult liver RNA from C57/Bl6 wild-type (Zhx2 +/-, grey bars) and whole-body Zhx2 knock-out (Zhx2 -/-, blue bars) littermate mice. Ribosomal protein L30 was used as an internal control, *p <0.05, *** p<0.001.
Figure 5-4. Overexpression of asAFP does not affect the expression of Alb and Afm. Map of the AFP gene from exon 7 to about 7 kb downstream of exon 15. Above the line are the AFP exons (open boxes) and below are the asAFP exons (grey boxes). Top panel. Schematic of asAFP RNA and relative location of Alb and Afm (not drawn to scale). Bottom panel. asAFP mid pcDNA construct and pcDNA empty vector control were transiently transfected in Hepa 1.6 cells. Alb, AFP, and Afm mRNA levels were measured by RT-qPCR. Ribosomal gene L30 was used as a normalizing control for all samples. The results represent the mean +/- SD of 3 independent transfection experiments, * p <0.05
A

asAFP mid

delta e12

e12(-)

(1800 bp)

(1576 bp)

(224 bp)

B

<table>
<thead>
<tr>
<th>construct</th>
<th>transfection date</th>
<th>asAFP/L30</th>
<th>pcDNA/L30</th>
<th>primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>asAFP mid</td>
<td>11/24/2015</td>
<td>0.01113</td>
<td>0.00236</td>
<td>758/759</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00027</td>
<td>0.00004</td>
<td>909/432</td>
</tr>
<tr>
<td>delta e12</td>
<td></td>
<td>0.00014</td>
<td>0.00004</td>
<td></td>
</tr>
<tr>
<td>e12 (-)</td>
<td></td>
<td>0.00269</td>
<td>0.00000</td>
<td>910-833</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00002</td>
<td>0.00000</td>
<td>909-267</td>
</tr>
</tbody>
</table>

C

Normalized to pcDNA (AFP/L30)
Figure 5-5. Exonic regions more effectively repress AFP mRNA. (A) Schematic of pcDNA cloning strategy to generate segments that overlap AFP intron and exonic sequences. Single primer mutagenesis was performed to remove exon 12 sequences from the asAFP mid pcDNA3.1 construct and to make a pcDNA construct that expresses just sequences that are antisense to exon 12. The numbers indicate sizes of the asAFP sequence included in the construct. Plasmids were transfected into Hepa 1.6 cells along with pcDNA empty vector control and the asAFP mid parental expression vector. (B) Total RNA was isolated, DNase-treated, and reverse-transcribed. The expression of asAFP RNAs and pcDNA control were measured by RT-qPCR using the asAFP-specific primers indicated and normalized to ribosomal protein gene L30 mRNA. The fold-change was determined by comparison of asAFP RNA levels with those of the pcDNA empty vector control. (C) Endogenous AFP mRNA levels, relative to L30 mRNA, were analyzed by RT-qPCR. The fold-change was determined by comparison of asAFP RNA levels with those of the pcDNA empty vector control. The results represent the mean +/- SD of >7 independent transfection experiments, *** p< 0.001.
A

Control (pcDNA)  pcDNA- GFP  asAFP mid-GFP  delta e12-GFP  e12 (-)-GFP

Green Fluorescence
Figure 5-6. Cell sorting to compare effectiveness of overlapping exonic and intronic regions of AFP mRNA. (A) Density plots of Hepa 1.6 cells transfected with pcDNA alone (control), or pcDNA, asAFP mid, delta e12 and e12 (-) co-transfected with pEGFP. The GFP-positive (red boxes) and GFP-negative (back boxes) cell populations were sorted by flow cytometry. (B) RT-qPCR analysis of endogenous AFP mRNA levels of unsorted, GFP positive and GFP negative populations using L30 mRNA as a control. The data were normalized to the pcDNA empty vector control. (C) Fold-change was determined by comparison of AFP mRNA levels from GFP-positive sorted cells with those to GFP-negative sorted cells. This data is from one transfection and sorting experiment; the cDNA was analyzed twice in duplicates, *** p< 0.001.
AFP mRNA

• A>G on antisense strand
• T>C on antisense strand

714-761 3424 bp (2/1)
714-759 2115 bp (1/1)
716-3'RACE 1200 bp (1/1)

714-759 2115 bp (2/27; 1 change in 4 clones)
714-759 2115 bp (0/41; cDNA from nuclear RNA)
811-870 599 bp (0/13)
873-836 612 bp (0/10)
328-761 762 bp cDNA (0/51)
Figure 5-7. Identification and distribution of RNA editing in AFP mRNA and asAFP RNA overlapping region. Map of the AFP gene from exon 7 to about 7 kb downstream of exon 15. Above the line are the AFP exons (open boxes) and below are the asAFP exons (grey boxes). The transcriptional orientation of the two transcription units is shown above the map. RNA editing sites showing the location of each A>G (green) or T>C (red) conversion. The open black box represents the segment that was cloned and sequenced. A description of the primers used for PCR amplification, amplicon size, the number of changes and the number of clones examined are indicated on the right.
CHAPTER VI
SUMMARY AND FUTURE DIRECTIONS

LncRNAs have emerged as important regulators of gene expression. While only a small number of LncRNAs have been fully characterized, several are involved in various developmental processes and diseases, including HCC (Prensner and Chinnaiyan, 2011; Yu et al., 2015). The data reported in this dissertation contributes to the understanding by which antisense RNAs control the expression of developmental genes that are misregulated in liver cancer. The AFP gene was used as the model biological system to study gene regulation. Previous studies suggested that Zhx2 functions as a transcriptional repressor of AFP postnatal expression through the AFP promoter (Peyton et al., 2000a). However, despite differences in steady state AFP mRNA levels in adult livers of BALB/cJ and wild-type mice, the transcription rates across the AFP gene were similar between the two strains of mice (Morford, unpublished; Vacher et al., 1992). This result is not consistent with Zhx2 being a transcriptional repressor. Thus, based on this observation, it was proposed that Zhx2 functions in a promoter-dependent manner but acts at the post-transcriptional level, perhaps by connecting a transcriptional to a post-transcriptional step of gene expression. To date, the mechanism by which Zhx2 regulates AFP is not well understood; it is also not yet clear if AFP is a direct target of Zhx2. Other genes, such as H19, Gpc3, Lpl, and several Cyp genes (Creasy et al., 2016; Gargalovic et al., 2010; Morford et al., 2007; Pachnis et al., 1984), are also differentially expressed in the presence or absence of Zhx2; which of these are direct or indirect targets of Zhx2 is not yet clear. There is evidence from our lab that Zhx2 positively regulates Mup genes through a site about 800 bp upstream of the promoter (Jiang et al., 2017). How Zhx2 may act to activate some genes and repress others is not known.
While studying the mechanism by which Zhx2 regulates AFP, our studies identified novel unannotated asAFP transcripts transcribed on the opposite strand of the AFP gene. In this work, I characterized asAFP RNAs and their role in regulating AFP gene expression. asAFP RNAs are ~5kb alternatively spliced transcripts containing 2-4 exons that partially overlap the 3’ half of AFP. Analysis of asAFP’s largest predicted open reading frame (ORF) and coding potential suggest that asAFP is a noncoding RNA. In addition, my data suggest that asAFP RNAs are not conserved with no evidence of sequence homology with other species, which is consistent with the idea that lncRNAs are generally less conserved than protein-coding genes (Derrien et al., 2012). Moreover, the asAFP transcription initiation region exhibits marks for active promoters and enhancers. Therefore, I examined asAFP transcriptional regulation and showed that asAFP is driven by a standard promoter comprised of an Inr and DPE motif that are required for full promoter strength. Because asAFP expression is positively correlated with Zhx2 in adult mouse liver and during liver development, I examined whether Zhx2 regulates the asAFP promoter. However, despite using multiple promoter constructs, multiple ratios of promoter to Zhx2 plasmid and multiple different cell lines, I found no evidence that Zhx2 was able to regulate the asAFP promoter. Although Zhx2 may not directly regulate asAFP, data shown here indicated that AFP and asAFP are co-expressed in mouse liver and their expression is inversely correlated in adult littermate Zhx2 wild-type and Zhx2 mutant mice and during liver development. Based on these findings, I hypothesized that asAFP contributes to the developmental regulation of AFP gene expression.

LncRNAs are emerging as new players in developmental and cancer paradigms, demonstrating their potential roles in embryonic development and both oncogenic and tumor suppressive pathways. These novel transcripts are frequently overexpressed in
cancer cells in comparison to normal tissue of the same origin, suggesting IncRNAs may be used as diagnostic markers or potential therapeutic targets in the treatment of cancer (Braconi et al., 2011; Matouk et al., 2007; Panzitt et al., 2007). Curiously, my data indicated that asAFP RNA expression levels were not altered in HCC tumors compared to non-tumor tissues, suggesting that asAFP RNA may not be directly involved in AFP misregulation in liver cancer. The rise of HCC in the United States has been linked to chronic hepatitis and more recently to NAFLD that can lead to hepatic fibrotic scarring that can interfere with normal liver function. Extensive hepatic damage can progress to cirrhosis and declining liver function (Baffy et al., 2012; El-Serag and Rudolph, 2007). Because the progression of HCC occurs through a step-wise process, future studies will examine changes in asAFP expression in a liver fibrosis model. Studies from others in the Peterson lab demonstrated that the expression of AFP and asAFP are inversely correlated during liver development, suggesting that asAFP may contribute to AFP repression during postnatal development. During fetal liver maturation, the function of the liver significantly changes from a hematopoietic organ to a metabolic organ (Si-Tayeb et al., 2010). While the roles of IncRNAs in liver development are not understood, one hypothesis is that developmental cues may induce asAFP RNA expression. A better understanding of how asAFP expression is regulated may provide insight into a mechanism by which asAFP contributes to AFP mRNA regulation during liver maturation.

Here, I showed that asAFP acts as a negative regulator of AFP mRNA expression mediated possibly through a base-pairing mechanism. Antisense RNAs are functionally diverse and can potentially regulate target genes at the transcriptional, posttranscriptional, and translational level (Figure 1-3). At the transcriptional level, the act of transcription in the antisense orientation can affect gene regulation through
transcriptional collision of RNA polymerases on convergent genes (Prescott and Proudfoot, 2002). However, several lines of evidence suggest that transcriptional interference is not involved in asAFP regulation of AFP which includes: 1. asAFP and AFP are co-expressed in mouse tissues 2. Data from nuclear run-on assays demonstrated that transcription rates of AFP from adult littermate Zhx2 wild-type or Zhx2 mutant mice that were similar. Studies done by others in our lab analyzed splicing across the AFP gene using liver RNA from adult BALB/cJ littermates with or without the Zhx2 transgene. These studies did not find evidence of alternatively spliced AFP transcripts, thus eliminating asAFP’s role in regulating alternative splicing. Also, IncRNAs have been shown to alter local chromatin modifications to affect gene expression within a specific chromosomal domain (Gupta et al., 2010; Kotake et al., 2011). I have no evidence that asAFP affects either of AFP’s neighboring genes, so this is an unlikely mechanism of action.

Previous studies showed the formation of nuclear sense-antisense transcripts could be edited by adenosine deaminases, which convert adenosine residues into inosines, leading to codon changes, nuclear retention, or alterations in mRNA stability (Peters et al., 2003). I proposed that AFP and asAFP RNA/RNA duplexes may form and serve as substrates for ADAR mediated decay. While I anticipated potential RNA editing within the complementary regions would lead to AFP mRNA degradation, the detected low levels of nucleotide changes attributed to editing suggests that ADAR activity does not appear to play a role in AFP mRNA stability. Recent studies suggest that ADAR enzymes can affect gene expression independently of their deamination activity. Following ADAR1 knockdown in human B cells, the expression of zinc-finger proteins were altered, without being edited (Wang et al., 2013a). Based on this finding, future studies would examine ADAR effects on AFP and asAFP expression.
As previously mentioned, the expression of various sized asAFP segments that together cover the full-length transcript downregulated AFP mRNA to similar levels. To better understand asAFP regulation, I examined whether the asAFP segments that overlapped with AFP intronic or exonic sequences were more effective. Although both overlapping intronic and exonic segments significantly decreased AFP mRNA, my data showed that overlapping exons had greater activity. This data suggests that asAFP regulation of AFP possibly occurs after nuclear export in the cytoplasm, which is correlative with other studies in my lab that showed asAFP RNA accumulated higher in the cytoplasm compared to the nucleus. In addition, the difference in AFP mRNA accumulation in the presence or absence of Zhx2 was observed more prominently in the cytoplasmic RNA but much less in the nuclear RNA, suggesting that AFP mRNA regulation occurs in the cytoplasm. One hypothesis regarding AFP regulation in the cytoplasm is that the formation of AFP and asAFP dsRNA duplexes serves as a substrate for Staufen-mediated decay, which modulates AFP mRNA stability. Future studies to determine whether AFP mRNA stability is altered include measuring AFP mRNA half-life in the presence or absence of robust levels of asAFP RNA. Alternatively, Staufen-mediated mRNA decay mechanism can be explored by measuring changes in AFP mRNA levels after siRNA knockdown of Stau1 binding protein.

To complement our overexpression studies, knockdown of asAFP RNA using both siRNA and antisense oligos (ASO) will help distinguish whether the regulation occurs in the cytoplasm or the nucleus. Because asAFP RNA is expressed in low abundance in our cell culture system, I made several attempts to generate stable cells lines expressing asAFP to carry out loss of function studies. However, I was not able to generate cell lines that expressed asAFP RNA above background levels. This is possibly due to clonal selection of low asAFP expressing cells or false-positives during the
selection process. To address this limitation, monoclonal selection may ensure that high expressing clones are obtained. In addition, it will be important to determine whether the results obtained with cultured cell lines reliably reflect regulatory events in mice. Because developmental silencing of AFP cannot be replicated in tissue culture cells, animal models are essential for a full understanding of regulatory mechanisms. To further investigate asAFP regulation of AFP in a mouse model, future studies will use recombinant adeno-associated viral vector, rAAV8 to overexpress asAFP or knock-down endogenous asAFP by noninvasive tail-vein injection of AAV8 vectors expressing the sequences that were shown to have maximal function in tissue culture cells, such as asAFP mid or exon 12 (-). From these studies, I anticipate an inverse correlation of AFP mRNA levels when asAFP RNA is manipulated. These mouse studies may establish a biological role for asAFP in regulating AFP expression and may provide additional insight into the mechanism of developmentally regulated AFP expression.

While this project began with observed differences in AFP mRNA expression in the presence or absence of Zhx2, data in this dissertation identified a functional role of novel asAFP transcripts in regulating AFP expression. While not shown in this work, I tested whether co-expression of asAFP RNA and Zhx2 would further regulate AFP mRNA. Since Zhx2 and asAFP regulate AFP expression independently, I proposed that asAFP and Zhx2 would together further downregulate AFP mRNA; for example perhaps Zhx2 could facilitate asAFP:AFP dsRNA formation. When the asAFP mid construct was co-transfected with as FLAG-epitope tagged Zhx2 (FLAG-Zhx2) expression plasmid, I did not observe a greater effect on AFP downregulation. Because AFP continues to be a clinically relevant biomarker for HCC, it is important that we understand all aspects of AFP regulation during development and disease. The identification of asAFP as a regulator of AFP adds an additional level of the coupled transcriptional and post-
transcription regulation of AFP and contributes to the current understanding of cell regulatory and disease mechanisms mediated by lncRNAs.
REFERENCES


### APPENDIX A

**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>amino acid</td>
</tr>
<tr>
<td>ADAR</td>
<td>adenosine deaminases that act of RNA</td>
</tr>
<tr>
<td>Af m</td>
<td>Afamin</td>
</tr>
<tr>
<td>AFP</td>
<td>Alpha fetoprotein</td>
</tr>
<tr>
<td>Afr1</td>
<td>Alpha-fetoprotein regulator 1</td>
</tr>
<tr>
<td>Alb</td>
<td>Albumin</td>
</tr>
<tr>
<td>alpha-1-antitrypsin</td>
<td>alpha 1 antitrypsin</td>
</tr>
<tr>
<td>ANRASSF1</td>
<td>antisense non-coding RNA in the INK4 locus</td>
</tr>
<tr>
<td>Arg</td>
<td>ANRIL</td>
</tr>
<tr>
<td>AS Uchl1</td>
<td>ubiquitin carboxy-terminal hydrolase L1</td>
</tr>
<tr>
<td>asAFP</td>
<td>antisense alpha-fetoprotein</td>
</tr>
<tr>
<td>ASO</td>
<td>antisense oligos</td>
</tr>
<tr>
<td>A-to-I</td>
<td>adenosines to inosines</td>
</tr>
<tr>
<td>BACE1-AS</td>
<td>antisense transcript for β-secretase-1</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BRE</td>
<td>TFIIB recognition element</td>
</tr>
<tr>
<td>C/EBP</td>
<td>CCAAT/enhancer-binding protein beta</td>
</tr>
<tr>
<td>CPAT</td>
<td>Coding-Potential Assessment Tool</td>
</tr>
<tr>
<td>CPC</td>
<td>Coding-Potential Calculator</td>
</tr>
<tr>
<td>CREB</td>
<td>cAMP response element binding protein</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>β-catenin</td>
</tr>
<tr>
<td>CYP</td>
<td>cytochrome P450</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco's minimal eagle's media</td>
</tr>
<tr>
<td>DBP</td>
<td>Vitamin D-binding protein</td>
</tr>
<tr>
<td>DPE</td>
<td>downstream promoter element</td>
</tr>
<tr>
<td>dsRNAs</td>
<td>double stranded RNAs</td>
</tr>
<tr>
<td>endo-si RNA</td>
<td>endogenous small interfering RNA</td>
</tr>
<tr>
<td>eRNAs</td>
<td>enhancer-derived RNAs</td>
</tr>
<tr>
<td>ES</td>
<td>embryonic stem</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent protein</td>
</tr>
<tr>
<td>GPC3</td>
<td>Glypican 3</td>
</tr>
<tr>
<td>H3K36me3</td>
<td>histone H3 lysine 36 trimethylation</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>histone H3 lysine 4 trimethylation</td>
</tr>
<tr>
<td>HBV</td>
<td>hepatitis B</td>
</tr>
<tr>
<td>HCC</td>
<td>Hepatocellular carcinoma</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>HCV</td>
<td>hepatitis C</td>
</tr>
<tr>
<td>HEK 293</td>
<td>Human Embryonic Kidney 293</td>
</tr>
<tr>
<td>Hep3B</td>
<td>human hepatoma</td>
</tr>
<tr>
<td>Hepa 1.6</td>
<td>mouse hepatoma</td>
</tr>
<tr>
<td>HepG2</td>
<td>human hepatoma</td>
</tr>
<tr>
<td>HNFs</td>
<td>hepatic nuclear factors</td>
</tr>
<tr>
<td>HNF4a</td>
<td>Hepatocyte nuclear factor 4 alpha</td>
</tr>
<tr>
<td>HULC</td>
<td>highly upregulated in liver cancer</td>
</tr>
<tr>
<td>Igf2r</td>
<td>Igf2 receptor</td>
</tr>
<tr>
<td>Inr</td>
<td>initiator</td>
</tr>
<tr>
<td>IncRNA</td>
<td>Long noncoding RNA</td>
</tr>
<tr>
<td>Lpl</td>
<td>Lipoprotein lipase</td>
</tr>
<tr>
<td>MEG3</td>
<td>maternally expressed gene 3</td>
</tr>
<tr>
<td>MERV</td>
<td>mouse endogenous retroviral element</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>MTE</td>
<td>motif ten element</td>
</tr>
<tr>
<td>Multiz</td>
<td>multiple sequence alignment track</td>
</tr>
<tr>
<td>Mup</td>
<td>mouse urinary protein</td>
</tr>
<tr>
<td>NAFLD</td>
<td>non-alcoholic fatty liver disease NAFLD</td>
</tr>
<tr>
<td>NATs</td>
<td>Natural antisense transcripts</td>
</tr>
<tr>
<td>NF1ona</td>
<td>Nuclear factor 1 alpha</td>
</tr>
<tr>
<td>NRT</td>
<td>no reverse transcription</td>
</tr>
<tr>
<td>NTC</td>
<td>no template control</td>
</tr>
<tr>
<td>ORFs</td>
<td>open reading frames</td>
</tr>
<tr>
<td>Piwi RNA</td>
<td>piRNAs</td>
</tr>
<tr>
<td>PRC2</td>
<td>Polycomb repressive complex 2</td>
</tr>
<tr>
<td>PRKACB</td>
<td>protein kinase cAMP-activated catalytic subunit beta</td>
</tr>
<tr>
<td>RACE</td>
<td>Rapid Amplification of Complementary DNA Ends</td>
</tr>
<tr>
<td>raf</td>
<td>regulator of AFP</td>
</tr>
<tr>
<td>RevErbα</td>
<td>Rev-ErbA alpha</td>
</tr>
<tr>
<td>RNA pol II</td>
<td>RNA polymerase II (pol II)</td>
</tr>
<tr>
<td>RNA-seq</td>
<td>RNA-sequencing</td>
</tr>
<tr>
<td>SMD</td>
<td>STAU1-mediated decay</td>
</tr>
<tr>
<td>SNPs</td>
<td>single nucleotide polymorphisms</td>
</tr>
<tr>
<td>TAFs</td>
<td>TBP-associated factors</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA box binding protein</td>
</tr>
<tr>
<td>TCF4</td>
<td>Transcription Factor 4</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>TFs</td>
<td>transcription factors</td>
</tr>
<tr>
<td>TFIID</td>
<td>transcription factor for RNA polymerase II D</td>
</tr>
<tr>
<td>TRα</td>
<td>thyroid hormone receptor alpha gene</td>
</tr>
<tr>
<td>TSS</td>
<td>transcriptional start sites</td>
</tr>
<tr>
<td>Xist</td>
<td>X-inactive specific transcript</td>
</tr>
<tr>
<td>Zhx2</td>
<td>Zinc-fingers and homeoboxes 2</td>
</tr>
<tr>
<td>4f-rnp</td>
<td>RNA-binding protein 4F</td>
</tr>
<tr>
<td>1/2-sbsRNAs</td>
<td>1/2-Staufen binding site RNAs</td>
</tr>
</tbody>
</table>
VITA

Maria S. Dixon
Jackson, MS

EDUCATION

2012-2017 University of Kentucky
   ▪ Ph.D. Biomedical Sciences
   Lexington, KY

2010-2012 Jackson State University
   ▪ M.S. Biology
   Jackson, MS

2005-2007 Jackson State University
   ▪ B.S., Biology, Cum Laude
   Jackson, MS

2004-2005 Dillard University
   New Orleans, LA

PROFESSIONAL POSITIONS

2012-Present Graduate Research Assistant, University of Kentucky
2011-2012 Teaching Assistant, Introduction to Biology Lab-101 (BIOL-101), Jackson State University
2009-2010 Inorganic Environmental Chemistry Laboratory Technician, MS Public Health Laboratory, Jackson, MS

RESEARCH SUPPORT

▪ University of Kentucky
  Lyman T. Johnson Academic Year Fellowship  08/12-05/15
  Stipend- $7,500

▪ R25 GM067592-08 and R56-AI077705  Hill, Glake and Brutkiewicz, Randy R. (PI)
  NIH Bridges to Doctorate Training Grant  08/10-04/12
  Stipend - $24,000; Research Support - $5,000

ACADEMIC AND PROFESSIONAL HONORS

2016- Women in Medicine and Science Mentoring and Demographics & Statistics Subcommittees
2016 ASBMB Minority Affairs Graduate Student Travel Award ($1000)
2016 Markey Cancer Center Research Day Poster Award 2nd place winner ($100)
2016 University of Kentucky Women in Medicine Science Rising Star Finalist
2015 Microbiology, Immunology & Molecular Genetics 3-minute thesis 3rd place winner

114
2014 RNA Rustbelt Meeting undergraduate student and minority travel scholarship

2012 University of Kentucky Graduate Student Minority Academic Award ($1000)

2011 Peoples Education Society Institute of Technology exchange fellow (Bangalore, India)

MEMBERSHIPS

2014-2017 Student Mentor: University of Kentucky/Eastern Kentucky University Bridges to Doctorate Program

2015-2016 American Society for Biochemistry and Molecular Biology

2014-2016 MIMG Departmental Retreat Planning Committee

2010-2012 President, Biology Graduate Student Association, Jackson State University

2006- Delta Sigma Theta Sorority, Inc.

COMMUNITY INVOLVEMENT

2011-2012 MS State Regional Science Fair Judge, Jackson, MS

2012-2017 Morton Middle School Science Fair Judge (Lexington, KY)

MANUSCRIPTS IN PREPARATION


ABSTRACTS


Dixon, M. and Ndebele, K., (2011) Validation of IRAK1 and IRAK4 in Glioblastomas. Annual Biomedical Research Conference for Minority Students, St. Louis, MO.


**ORAL PRESENTATIONS**

Dixon, M. AFP anti-sense transcripts in mouse liver and their potential role in gene regulation. Jackson State University