Regulation Of Fshr And SF-1 In The Hypothalamus-Pituitary-Gonadal (HPG) Axis.

By

Jitu W. George

Submitted to the graduate degree program in Molecular and Integrative Physiology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dissertation Committee

________________________________
Leslie L. Heckert, PhD, Chair

________________________________
Vargheese M. Chennathukuzhi, PhD

________________________________
Lane K. Christenson, PhD

________________________________
Patrick E. Fields, Ph.D.

________________________________
Michael W. Wolfe, Ph.D.

Date Defended: July 10, 2013
The Dissertation Committee for Jitu W. George

certifies that this is the approved version of the following dissertation:

Regulation Of Fshr And SF-1 In The Hypothalamus-Pituitary-Gonadal (HPG) Axis.

________________________________
Leslie L. Heckert, PhD, Chair

Date approved: July 17, 2013
Abstract

Mammalian reproduction is highly dependent on the delicate balance of signals within the hypothalamic-pituitary-gonadal (HPG) axis that maintains proper endocrine environment. One of the key signals is the pituitary glycoprotein follicle-stimulating hormone (FSH) which is critical for gonadal development and fertility. FSH acts via FSHR, a G protein-coupled receptor present on a few distinct cell types, prominently on the Sertoli cells of the testis and granulosa cells of the ovary. Prior studies have identified several key regulatory elements required for Fshr transcription but a clear understanding of what controls its limited cell specific expression remains elusive. Comparative genomics identified a number of evolutionary conserved regions (ECR) distal to the proximal promoter, indicating these regions might harbor regulatory elements. One such distal regulatory element known to play a role in transcriptional regulation is the “multivalent” protein CCCTC-binding factor (CTCF). Computational analysis of ECRs identified multiple CTCF binding sites in the intergenic regions of the Fshr locus and depletion of CTCF in rat granulosa cells led to a two-fold increase in Fshr mRNA. These data indicate that CTCF either by itself or in conjunction with other protein complexes might play a role in transcriptional regulation of the Fshr gene.

Another important component of the HPG axis is the transcription factor, Steroidogenic factor-1 (SF-1) transcribed from the Nr5a1 gene (also known as Ftz-F1 or Ad4bp) which plays a pivotal role in adrenal and gonadal development and regulates genes at all levels of the axis. Comparative genomics identified two conserved sequences associated with Ftz-F1. One was found to encode a long non-coding RNA and the other a regulatory element important for SF-1 expression in the pituitary. A long non-coding RNA named Fast or Ftz-F1 associated transcript was found transcribed in the opposite direction to SF-1, co-expressed, and co-regulated with SF-
and hormonally regulated. Knockdown and over-expression of the *Fast* transcript in MA-10, mouse Leydig cell line, did not alter the transcription or translation of SF-1. Long non-coding RNA have been focus of intense research and have been shown to play diverse role in various physiological and pathological processes. Considering their potential for transcriptional regulation, the identified long non-coding RNA, *Fast*, might play potentially important role in endocrine development and homeostasis.

Comparative sequence analysis identified an evolutionary conserved region, ECR3, approximately 4kb away from the transcriptional start site of SF-1. Transient transfection data revealed ECR3, upregulated SF-1 transcription in alpha T3 (gonadotrope cell line) and Y-1 (adrenal) cells, but downregulated SF-1 in MA-10 (Leydig cell line), MSC-1 (Sertoli cell line) and primary rat peritubular Myoid and Sertoli cells. This region was found to contain an E-box and was bound by upstream stimulatory factors 1 and 2 (USF1 and USF2) and a basic-helix-loop-helix (bHLH) protein E2A. In particular, co-transfection studies identified Inhibitors of DNA binding, Id2 and Id3 to downregulate SF-1 transcription.

In summary, this dissertation looks at the regulation of *Fshr* and SF-1 in the HPG axis and identifies the importance of distal regulatory sequences and non-coding transcripts and its role in gene regulation.
Acknowledgments

This dissertation is the result of a challenging journey, which would not have been possible without the help, guidance, and support of a great number of people. First, and foremost, I am greatly indebted to Dr. Leslie Heckert for accepting me as her graduate student. I am thankful for her undying patience and support and for teaching me to think, plan, and perform as a scientist. I have been fortunate to have learned a lot under her and really wish I could have learned more. I am truly blessed to have had you as my advisor.

I am grateful to all my committee members, Dr. Vargheese Chennathukuzhi, Dr. Lane Christenson, Dr. Michael Wolfe and Dr. Patrick Fields, for your input and guidance, has helped me develop as a scientist. Each of you made time from your busy schedule to counsel and advise me on my scientific research and career. No one could ask for a more dedicated and helpful committee and I thank you for that.

Special thanks go to Daren Rice for teaching me a number of molecular biology techniques. His thorough knowledge of molecular biology is truly admirable. During my time at the lab, he has helped me navigate through a number of technical problems that arose in molecular biology experiments and in my 13-year-old car. Not only was he a wonderful lab manager but also a great friend. I will truly miss the annual zoo outings that were made more enjoyable with the presence of his four wonderful kids, Addy, Zade, Levi, and Finn.

I also thank fellow lab members, Dr. Tatiana Karpova, Dr. Shixin Tao, and my fellow ex-graduate student, Dr. Valentine Agbor, from whom I have learned so much. It was a great pleasure to have worked with you all. I am also grateful to Dr. Ravichandiran Kumarasamy and Dr. Brian Hermann whose research work has helped form this thesis.
I owe a lot to Dr. Paul Cheney, who 5 years ago guided me to Leslie, when I was looking for a graduate student position. He has always been supportive of Physiology students and I thank him and the Molecular and Integrative Physiology Department for supporting my funding and research for over a year. I do realize that this has been very tough especially with the budget cuts and the current NIH funding and I am very appreciative of all the support.

I have been blessed to be part of the Indian community at the KUMC, which comprises of a great number of graduate students and post-doctorates, whom I have been fortunate to know. I am truly thankful to a number of my friends who brought back gifts and packages from India and for their companionship that helped compensate for the five and a half long years I have been away from home.

As I come to the end of my acknowledgments, I am thankful to my parents, for their love, support and all the sacrifices they have made throughout their life so that I could obtain a better education and life.

Finally my dear wife, Ani, I am thankful for your undying love, faith, patience, and extraordinary strength, which kept me going. I cannot even begin to comprehend how I would have made this journey without you.
Table of Contents

Abstract........................................................................................................................................... iii
Acknowledgements.......................................................................................................................... v
Table of Contents............................................................................................................................ vii
List of Figures and Tables................................................................................................................ x
Part 1: Reproduction and transcription regulation...........................................................................

Chapter 1

Introduction....................................................................................................................................... 2

The hypothalamus-pituitary-gonad axis and the gonadotropin hormones........................................ 3
GnRH, GnRHR, and gonadotropins.................................................................................................. 5
Transcriptional regulatory elements................................................................................................. 7
Core and proximal promoters......................................................................................................... 10
Enhancers....................................................................................................................................... 11
Silencers.......................................................................................................................................... 13
Insulators......................................................................................................................................... 14
Locus control regions...................................................................................................................... 17
Non-coding RNA........................................................................................................................... 18
The gonadotropin hormones.......................................................................................................... 33
Fshr expression............................................................................................................................... 34
Transcription of FSHR/Fshr.......................................................................................................... 37
The FSHR gene.............................................................................................................................. 42
Identification of new regulatory elements...................................................................................... 49
Summary......................................................................................................................................... 52
Chapter 2
CTCF and transcription regulation of Fshr in rat granulosa cells

Abstract
Introduction
Materials and Methods
Results
Discussion

Part 2: Steroidogenic Factor 1

Chapter 3
Introduction
Identification of SF-1
Structure of SF-1
SF-1 expression
Function of SF-1
SF-1 gene and transcription

Chapter 4
Analysis of Fast, Ftz-F1-associated transcript, suggests functional cooperation with the nuclear receptor Steroidogenic Factor 1

Abstract
Introduction
Results
Chapter 5

An upstream distal Evolutionary Conserved Region regulates the expression of nuclear Receptor Steroidogenic Factor 1 in the pituitary...126

Abstract...127

Introduction...128

Results...130

Discussion...136

Materials and Methods...140

Chapter 6

Implications and Future Directions...152

Implications...153

Future Directions...154

Defining CTCF regions and its role in regulating Fshr transcription...154

Identifying functional role of Fast...158

Further characterization of ECR3 and role of identified transcription factors...161

References...164
List of Figures and Tables

Chapter 1

Figure 1: The hypothalamus-pituitary-gonadal axis........................................4

Figure 2: Schematic of classical gene regulator elements.................................9

Figure 3: A summary of genomic location of IncRNA.................................24

Figure 4: Summary of FSHR/Fshr transcriptional regulation................40

Figure 5: Organization of FSHR/Fshr and its genomic environment.............45

Figure 6: Annotated region of human FSHR and neighboring genes NRXN1 and LHCGR from the UCSC Genome Browser.................................................................51

Chapter 2

Table 1: Genomic location of ECRs in Rattus norvegicus (Rat) genome........61

Table 2: ECRs predicted to have CTCF binding sites.......................................62

Figure 1: Expression of Fshr and Lhcgr in estrogen treated rat granulosa cells depleted of CTCF.................................................................63

Figure 2: Possible mechanism of CTCF action on Fshr locus.......................67

Chapter 3

Figure 1: Structure of classic nuclear receptor and SF-1............................72
Figure 2: Schematic organization of the mouse *Nr5a1* gene…………………………….79

Figure 3: Schematic representation of SF-1 and its proximal promoter…………………81

Chapter 4

Table 1: Oligodeoxynucleotide primer………………………………………………….104

Figure 1: Two highly conserved ESTs are located between *Nr5A1* (*Ftz-F1*) and *Gcnf* loci………………………………………………………………………………..……110

Figure 2: *Fast* is transcribed opposite to *Ftz-F1*……………………………………….111

Figure 3: Diagrammatic representation of the genomic location of *Fast* and its transcripts……………………………………………………………………….……….112

Figure 4: *Fast* localizes predominantly in the cytoplasm…………………………….113

Figure 5: *Fast* may code for a 7kd protein………………………………………………….115

Figure 6: Expression of *Fast* and SF-1 in adult mouse tissue and mouse Leydig cell line, MA10…………………………………………………………………………….….…117

Figure 7: *Fast* and SF-1 respond to retinoic acid in P19 embryonal carcinoma cells…118

Figure 8: Hormonal response of *Fast* in mouse granulosa cells……………………………119

Figure 9: Knocking down *Fast* does not affect transcription of SF-1………………….120

Figure 10: Knocking down *Fast* does not affect downstream targets of SF-1………121

Figure 11: Over expressing Fast does not affect transcription of SF-1…………………..122
Figure 12: Knocking down Fast does not affect translation of SF-1………………..123

Figure 13: Over expressing Fast does not affect translation of SF-1………………..124

Chapter 5

Table 1. Oligos used to amplify SF-1 Promoter (-734+60) region………………143

Table 2. Oligos used to amplify the ECRs…………………………………………143

Figure 1: Alignment of NR5A1 region using ECR browser, depicting the Evolutionarily conserved regions (ECRs)……………………………………………………………144

Figure 2: Schematic diagram of the vector SF-1(-734+60) Luc and the transcriptional activity of each ECR……………………………………………………………………145

Figure 3: Deletion mutants and analysis of transcriptional activity………………146

Figure 4: Dnase Footprinting of ECR3 region………………………………………147

Figure 5: Transient transfection analysis to check activity of potential transcription factor binding sites……………………………………………………………148

Figure 6: Electrophoretic mobility shift assay (EMSA) of ECR3 region…………149

Figure 7: Evaluation of transcription factor binding by ChIP in αT3 cells…………150

Figure 8: Co-transfection analysis to determine role of Id proteins………………151

Chapter 6

Figure 1: A schematic diagram depicting colony assay…………………………157
Figure 2: Purification of RNA bound proteins by MS2-MBP pull down method…….160
Part 1: Reproduction and transcription regulation
Chapter 1: Introduction
The hypothalamus-pituitary-gonad axis and the gonadotropin hormones

In mammals, fertility depends on precise hormonal regulation of a functional feedback loop that involves the hypothalamus, pituitary, and gonads (HPG), responsible for production of gametes and hormones, and influences the overall reproduction system. The HPG axis was proposed in 1930 by Carl R. Moore and Dorothy Rice, who demonstrated that hormones produced in the testis and ovary feedback onto the pituitary reduced the production of gonadotropins (Moore and Price 1930, Moore and Price 1932). Shortly thereafter, Walter Hohlweg and Karl Junkmann deduced that there was an interdependent relationship between the gonads, the central nervous system, and the pituitary (Hohlweg and Junkmann 1932, Hohlweg 1975). The central thought was that hormones secreted by the gonads, via their effect on the central nervous system; control the secretion of gonadotropins from the pituitary, which in turn regulate hormone production in the gonads. Geoffrey Harris then demonstrated that the hypothalamus was the main control site, by inducing ovulation in rabbits by electric stimulation of either the hypothalamus or the pituitary, thus identifying the components of the HPG axis (Harris 1937).

The intricate relationship starts with the pulsatile secretion of the hypothalamic gonadotropin-releasing hormone (GnRH), leading to the secretion of bioactive gonadotropins, Follicle stimulating hormone (FSH) and luteinizing hormone (LH), which travel through the blood stream and act on specific cell surface receptors on the gonads, leading to hormone production. A complex feedback loop regulates the hormone levels, which is necessary to initiate and support gametogenesis, steriodogenesis, and ovulation. LH and FSH are integral parts of the neural and endocrine interchange between the hypothalamus, pituitary, and gonads that controls
Figure 1: GnRH secreted from the hypothalamic neurons feeds into the anterior pituitary, where it binds its receptor GnRHR, leading to synthesis and secretion of gonadotropin hormones, LH and FSH. These gonadotropin hormones make their way through the blood to the cell surface receptors present on the gonadal cells. FSH binds to its receptor, FSHR present on the Sertoli cells of the testis and the granulosa cells in the ovary. LH binds to its receptor, LHCGR, on the Leydig cells of the testis and in granulosa, theca, luteal, and interstitial cells of the ovary. This leads to production of steroids and inhibin which feedback negatively on the hypothalamus and the pituitary leading to decrease in gonadotrope production.
steroid hormone synthesis and gamete production (reviewed in (Young 1995, Achermann and Jameson 1999, Plant 2008)). At the top of the network is GnRH, which, once released from the hypothalamus, binds specific receptors on pituitary gonadotrophs and induces the synthesis and secretion of LH and the FSH. These dimeric proteins formed by the heterodimerization of a common alpha subunit (α-glycoprotein hormone subunit) with the specific beta subunit (FSHβ or LHβ) confer biological specificity to the hormones. Once in circulation, FSH and LH, bind to their respective receptors on the gonad cells leading to regulation and expression of steroidogenic genes and production of steroid hormones (Figure 1). Expression of a number of steroid hormones in the HPG axis is tightly regulated in a temporal and development manner.

GnRH, GnRHr, and gonadotropins

Gonadotropin releasing hormone is an evolutionarily conserved decapeptide that forms a bridge between the brain and the peripheral reproductive system. Since the first isolation of GnRH, in 1971, from pig and sheep brains, the GnRH family now encompasses 24 molecular isoforms identified in vertebrates and invertebrates (Amoss, Burgus et al. 1971, Matsuo, Baba et al. 1971, Gorbman and Sower 2003). In vertebrates, there are two major isoforms of GnRH; the hypothalamic GnRH1, which acts on the pituitary and extra-hypothalamic GnRH2, which has no known functional involvement to gonadotropin release (Gorbman and Sower 2003). Once released from the hypothalamus, GnRH1 travels via the median eminence to the pituitary where it binds the GnRH receptor, a member of the rhodopsin family of G protein-coupled receptor expressed on the pituitary gonadotropes (Tsutsumi, Zhou et al. 1992, Stojilkovic, Reinhart et al. 1994, Kaiser, Conn et al. 1997, Sealfon, Weinstein et al. 1997). GnRH binds to its receptor causing activation of G-proteins, including Gq and G11, subsequent activation of phospholipase Cβ, release of calcium, and increased activity of protein kinase C and calcium/calmodium kinase.
Pulsatile release of GnRH stimulates the expression of mitogen-activated protein kinase (MAPK) signaling cascades (MAPK1/3 [extracellular signal-regulated kinase, or ERK], MAPK8/9 [c-Jun N-terminal kinase, or JNK], and MAPK14 [p38]). While the role of p38 in transcriptional regulation of LHβ has been controversial, ERK1/2 and JNK are involved in the induction of early growth response-1 (Egr-1), a factor known to regulate the LHβ promoter (Yokoi, Ohmichi et al. 2000, Harris, Bonfil et al. 2002, Burger, Haisenleder et al. 2004, Yamada, Yamamoto et al. 2004). Female Egr-1 knockout mice were infertile and lacked expression of LHβ (Lee, Sadovsky et al. 1996, Topilko, Schneider-Maunoury et al. 1998). The proximal promoter region of GnRH contains binding site for Egr1, and transcription factors SF-1 and Ptx1. GnRH induces gene expression of Egr-1, leading to a synergistic interaction of Egr-1 with transcription factors, SF-1 and Ptx1 inducing LHβ transcription (Dorn, Ou et al. 1999, Tremblay and Drouin 1999).

GnRH also regulates FSH levels, as GnRH null mice reported 60% reduction in serum FSH levels when compared to normal mice (Mason, Hayflick et al. 1986). Furthermore, administration of a single pulse of GnRH led to a fourfold increase in FSHβ gene expression indicating that GnRH is an important regulator of the FSHβ gene (Burger, Dalkin et al. 2001). GnRH mediates FSHβ gene expression via the PKC and MAPK signaling pathways (Vasilyev, Lawson et al. 2002, Bonfil, Chuderland et al. 2004, Coss, Jacobs et al. 2004, Liu, Ruiz et al. 2005). GnRH also regulates FSHβ gene through the induction of activator protein-1 (AP-1), a heterodimeric transcription factor that consists of a variety of dimers of Fos and Jun isoforms (Wurmbach, Yuen et al. 2001, Kakar, Winters et al. 2003). Mice deficient of c-Fos had smaller ovaries with atretic follicles, similar to FSHβ knockout mice (Johnson, Spiegelman et al. 1992, Kumar, Wang et al. 1997). AP-1 also interacts with factors NF-Y and USF-1 to stimulate FSHβ
in response to GnRH (Coss, Jacobs et al. 2004, Ciccone, Lacza et al. 2008). Transcription factor cAMP response element-binding protein (CREB) binds to the CRE/AP-1 site in response to GnRH regulation of rat FSHβ (Ciccone, Lacza et al. 2008). CREB deficient mice did not have reduced FSH levels, indicating that either cAMP-responsive element modulator (de Kok, Merkx et al.) or activating transcription factor (ATF) family members may compensate for the deficiency or CREB activity on FSHβ activity is a species-specific phenomenon (Hummler, Cole et al. 1994, Ciccone, Lacza et al. 2008).

In this dissertation, I will explore the transcriptional regulation of two critical genes involved in the HPG axis. In Part I, I will discuss distal regulatory elements and the role of CCCTC-binding factor (CTCF) in the regulation of the Follicle Stimulating hormone receptor (Fshr) gene. In Part II, I will investigate distal regulatory regions and SF-1 transcription and define a long non-coding RNA transcribed at the SF-1 locus. These distinct regulatory elements will be discussed subsequently under “Transcriptional regulatory elements”.

**Transcriptional regulatory elements**

Various aspects of human development hinge on proper spatial, temporal, and tissue-specific control of gene expression. As we mark 60 years since Watson and Crick proposed the double helical structure of DNA, questions remain on how thousands of genes are orchestrated in a controlled and precise manner (Watson and Crick 1953). Britten and Davidson hypothesized that discrete regulatory sequence not only played a role in regulating gene expression, but also played a major role in governing biological complexity in eukaryotes (Britten and Davidson 1969). This increased level of complexity is understandable with the fruition of the human genome project and identification of ~20,000-25,000 genes (Human Genome Sequencing 2004). It has become
abundantly clear that there are functional regions that extend from the promoters into the vast space of the human genome. Sequencing of human and other vertebrate genomes, coupled with high throughput in-vitro and in-vivo functional screens have facilitated the comprehensive identification of a number of regulatory sequences and helped compile the regulatory architecture required to establish the spatial and temporal expression of genes. Gene expression is regulated by the coordinated expression of distal regulatory elements (Figure 2) which can be crudely classified into a. Core and Proximal promoter, b. Enhancers, c. Silencers, d. Insulators, e. Locus Control regions, and f. non-coding RNA.
Figure 2: Schematic of classical gene regulator elements. The promoter is composed of a core promoter and proximal promoter elements and mediates basal transcriptional control. Upstream regulatory elements, enhancers and locus control regions, mediate positive transcription while silencers negatively affect transcription. Insulators, play a complex role by acting as enhancer-blockers and barrier insulators. These regulatory elements are thought to carry out their function by employing what is known as “DNA looping”. Modified according to (Maston, Evans et al. 2006).
**a. Core and proximal promoters:** The core promoter can be defined as the minimal stretch of DNA that serves as the docking site for the basic transcription machinery and encompasses the site and direction of the transcription start site (Smale and Kadonaga 2003). Accurate and efficient transcription of the core promoter requires RNA polymerase II (RNAP II) along with “basal” transcription factors that include transcription factor (TF) IIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. Assembly of the preinitiation complex starts with the binding of TFIID followed by recruitment of TFIIA and TFIIB. This association leads to the addition of TFIIF bound with RNAP II and of further incorporation of TFIIE and TFIIH leading to the completion of the preinitiation complex and initiation of transcription. An alternate pathway of the preinitiation complex formation is the RNAPII holoenzyme pathway, which comprises of TFIIF, TFIIE, TFIIH and other heterogeneous proteins that are involved in chromatin remodeling, DNA repair, and mRNA processing. Irrespective of the manner in which the preinitiation complex is formed, TFIID is the first TF that binds to the core promoter element and has emerged as the central component of the transcription apparatus (Baumann, Pontiller et al. 2010).

In addition to TATA box, core promoters can consist of Initiator element (Inr), Downstream Promoter Element (DPE), Downstream Core Element (DCE), TFIIB-Recognition Element (BRE), and Motif Ten Element (MTE) (Maston, Evans et al. 2006). A statistical analysis of human core promoters revealed that Inr was the most common core promoter element, followed by DPE and BRE, and surprisingly, TATA being present in the lowest levels. Furthermore, a quarter of the analyzed promoters had neither TATA, Inr, DPE or BRE elements indicating presence of novel promoters (Gershenzon and Ioshikhes 2005). In line with this, recent reports have now identified a new core element labeled as ATG deserts (Lee, Howcroft et al. 2005). This
diverse content of the core promoter possibly play a functional role by contributing to gene specific regulation.

Proximal promoter can be defined as the region immediately upstream of the core promoter and contributes to its basal activity by the presence of multiple binding sites for activators. These transcriptional regulatory elements act synergistically and mutation of any of these sites leads to decreased transcriptional activity (Maston, Evans et al. 2006). While the core and proximal promoter comprise the very basic level of transcriptional regulation, elements that are more distal have now been identified to play a role in this exceeding intricate ballet of gene expression and include enhancers, silencers, insulators, locus control regions and non-coding RNA’s which are discussed in the following sections.

**b. Enhancers:** Enhancers are typically small segments of DNA, typically few hundred base pairs in length, that can reside either in the intergenic regions, introns, or exons and can be located ten to hundred kilobases away from their target genes and operate as transcription factor binding sites, which work cooperatively to enhance transcription. They have the inherent capacity to act in a modular manner such that a single promoter can be acted upon by distinct enhancer elements in a spatial and temporal specific manner (Maston, Evans et al. 2006, Spitz and Furlong 2012). How do these enhancers carry out their function in a spatio-temporal manner? Analysis of transcription factor occupancy at different stages of development in both Drosophila and mammals have identified that precise timing of DNA occupancy is what controls the temporal nature of gene expression driving development (Sandmann, Jensen et al. 2006, Jakobsen, Braun et al. 2007, Cao, Yao et al. 2010, Lin, Jhunjhunwala et al. 2010). Temporal regulation was thought to correspond to relative affinity or number of transcription factor binding sites, affecting change in occupancy relative to concentration of transcription factor over time (Gaudet and
Mango 2002, Sandmann, Jensen et al. 2006). However, evidence now indicate that enhancers with context-dependant occupancy with differential motif enrichment are co-occupied with different transcription factors in a time-dependant manner and this co-occupancy is needed for recruitment of additional transcriptional factors (Zeitlinger, Simon et al. 2003, Jakobsen, Braun et al. 2007, Sandmann, Girardot et al. 2007, Lin, Jhunjhunwala et al. 2010, Wilczynski and Furlong 2010, Mullen, Orlando et al. 2011, Trompouki, Bowman et al. 2011, Yanez-Cuna, Dinh et al. 2012).

Based on the relative order, orientation, and spacing of transcription factors, three models have been proposed to explain enhancer activity. In the enhanceosome model, as exemplified by $IFN-\beta$ and $TCRa$, recruited transcription factors form an ordered and specific position relative to each other, such that only one arrangement can lead to gene expression (Thanos and Maniatis 1995, Spicuglia, Payet et al. 2000, Merika and Thanos 2001). While this ordered positioning is thought to be essential for rapid response, most developmental enhancers are more flexible, with a subset of transcription factors binding co-operatively (Senger, Armstrong et al. 2004). This flexible positioning known as the “billboard model” in which there are fewer constraints on relative binding of transcription factors but requires a subset of factors to be active, such as the stripe 2 enhancer of $Drosophila$ even-skipped (Arnosti, Barolo et al. 1996, Ludwig, Patel et al. 1998, Kulkarni and Arnosti 2003, Arnosti and Kulkarni 2005, Hare, Peterson et al. 2008). The third model of regulation, known as the “transcription factor collective”, suggests that the same set of transcription factors can bind to multiple enhancers and do so in a cooperative manner. In a recent example when one transcription factor was removed, all other transcription factors failed to activate the enhancer in vitro indicating cooperative binding occurs in vivo, with all or a sub-
set of transcription factor binding to the DNA and other transcription factors recruited via protein-protein interaction (Junion, Spivakov et al. 2012).

Since enhancers often reside at significant distances, either upstream or downstream of their target promoters, mechanisms must exist for them to communicate over large genomic distances. Current data indicates that chromatin “looping” permits the interaction between the enhancer and the core promoter (Blackwood and Kadonaga 1998, Bulger and Groudine 1999, de Laat, Klous et al. 2008). An alternative model supports a diffusion or “tracking” mechanism, in which the enhancer scans along the chromatin until it arrives at the core promoter (Blackwood and Kadonaga 1998). With the advent of techniques such as chromosome conformation capture (3C), in which the interactions between two genomic regions are determined by cross-linking, restriction enzyme digestion, followed by intramolecular ligation and PCR analysis of the resulted ligated products, data has now accumulated in favor of the looping model (Cullen, Kladde et al. 1993, Dekker, Rippe et al. 2002, Miele and Dekker 2009).

c. Silencers: Silencers are sequence specific elements that silence or repress transcriptional activity of a gene and act typically in a distance and orientation independent manner (Brand, Breeden et al. 1985, Ogbourne and Antalis 1998). Silencers can be situated at the proximal promoter, as part of distal elements, or independently far away from its target gene in the intron or the 3’-untranslated region (Maston, Evans et al. 2006).

Silencers repress promoter activity by binding of negative transcription factors called repressors and recruitment of negative cofactors, called corepressors (Privalsky 2004). Based on distance of action, two classes of silencer elements have been identified in Drosophila, short-range silencers that reside within ~100 bp of the target gene and repress its transcription, and long-range
silencers, which can affect multiple enhancers or promoters over a span of few kilobases. The current understanding is that this distance dependant repression is due to the difference in recruitment of different co-factors (Kulkarni and Arnosti 2005).

These silencer elements repress gene transcription by blocking the binding of a nearby activator or by competing for activator binding sites (Li, He et al. 2004, Harris, Mostecki et al. 2005). Other models that have been proposed to explain the repressor activity suggests that they may establish a repressive chromatin structure by recruiting histone modifications or chromatin stabilizing factors or by inhibiting assembly of the preinitiation complex (Srinivasan and Atchison 2004, Chen and Widom 2005).

d. Insulators: Insulators, also known as boundary elements are a family of DNA sequence elements typically ~0.5-3kb in length, that protect genes from transcriptional activity of its neighboring genes, in a position-dependent and orientation independent manner (Maston, Evans et al. 2006). Originally described in Drosophila, insulators have now been found in most eukaryotes, from yeast to human (Phillips and Corces 2009). Insulators that disrupt communication between enhancer and promoter are known as enhancer-blocking insulators and those that protect against heterochromatin silencing are known as barrier insulators (Sun and Elgin 1999, Valenzuela and Kamakaka 2006).

The first vertebrate insulator identified was a Dnase1 hypersensitive site (HS), 5’ HS4, sequence of the chicken $\beta$-globin locus. The $\beta$-globin locus was an ideal starting ground due to the extensive work on local and distant regulatory elements and the fact that chicken erythrocyte nuclei were relatively free of proteases and nucleases. The folate receptor gene located upstream is separated from the $\beta$-globin locus by a 16-kb long region of condensed chromatin that bear
silenced heterochromatin marks of H3K9 and K27 methylation (Litt, Simpson et al. 2001, Litt, Simpson et al. 2001). This sudden transition of heterochromatin to euchromatin at the active β-globin locus suggested the presence of a boundary sequence that separated the domains. Dnase 1 hypersensitivity and H3 lysine 9 acetylation histone modification identified the HS4 insulator element (Litt, Simpson et al. 2001, Litt, Simpson et al. 2001, Ma, Heath et al. 2011). Further work identified a core ~250bp sequence of the HS4 insulator element with properties of both enhancer blocking and barrier activity (Chung, Whiteley et al. 1993, Pikaart, Recillas-Targa et al. 1998, Bell, West et al. 1999, Recillas-Targa, Pikaart et al. 2002). The transcription factor mediating this activity was identified as CCCTC-binding factor (CTCF), an eleven-zinc finger protein, both necessary and sufficient for the enhancer-blocking activity of the 5’ HS4 (Bell, West et al. 1999). Another CTCF binding site was identified in the 3’ end of the chicken β-globin locus (Bell, West et al. 1999). In parallel, CTCF binding sites were subsequently identified within the imprinted control region (ICR) of the mammalian H19/Igf2 locus (Bell and Felsenfeld 2000, Hark, Schoenherr et al. 2000, Szabo, Tang et al. 2000). To date, CTCF has been the only protein identified to show enhancer-blocking activity in vertebrates.

CTCF is a multivalent factor, originally identified, owing to its ability to bind to a variety of sequences as well as co regulatory proteins using multiple zinc fingers (Filippova, Fagerlie et al. 1996). This unique structural feature suggested that CTCF plays a varied role in genome regulation. Global knockout of CTCF led to embryonic lethality prior to implantation (Splinter, Heath et al. 2006, Heath, Ribeiro de Almeida et al. 2008). Depletion of CTCF in oocytes caused transcriptional misregulation of hundreds of genes (Wan, Pan et al. 2008). Taken together this data suggests that CTCF plays an important and varied role in gene regulation that include
traditional enhancer blocking, promoter activation/repression, hormone-responsive silencing, long-range chromatin interactions and genomic imprinting (Phillips and Corces 2009).

The H19/Igf2, one of the first identified imprinted loci, is a classical example wherein genetic information is stored in two copies, one from each parent, with expression of one parental copy while silencing the other (Bartolomei, Zemel et al. 1991, DeChiara, Robertson et al. 1991, Ferguson-Smith, Cattanach et al. 1991). Gene knockout studies have identified that Igf2 is crucial for placental and fetal growth, whereas the H19 gene encodes for a non-coding RNA, and retards fetal growth. The same enhancer activates H19 on the maternal allele and Igf2 on the paternal allele. In the paternal allele, the ICR region immediately upstream of the H19 gene is methylated preventing CTCF binding, abrogating the insulator activity, resulting in the functional communication between the Igf2 promoter and enhancer. In the maternal allele, ICR is unmethylated, which allows CTCF to bind to the ICR region and prevents the enhancers from accessing the Igf2 promoter (Herold, Bartkuhn et al. 2012). Proper expression of the parental allele is crucial, as mutant H19 gene is one of the causes of Beckwith-Wiedemann syndrome (BWS); a condition associated with fetal overgrowth and increased risk of tumor formation. BWS phenotype has been observed by increased DNA methylation at ICR resulting in biallelic Igf2 expression and reduced H19 activity and also in patients with micro-deletions at the CTCF binding site at the ICR locus, indicating that loss of CTCF binding at ICR either due to mutations or due to DNA methylation causes Igf2 activation and H19 inactivation (Sparago, Cerrato et al. 2004, Azzi, Rossignol et al. 2009). Conversely, loss of methylation at the paternal ICR, results in biallelic expression of the H19 gene and reduced expression of Igf2, is observed in patients diagnosed with Silver-Russell syndrome (SRS), a condition characterized by growth retardation during fetal and postnatal development (Herold, Bartkuhn et al. 2012).
With the use of techniques such as fluorescence in situ hybridization (FISH) and chromosome capture conformation (3C), a greater understanding has now emerged on the importance of nuclear architecture (Fraser and Bickmore 2007, Miele and Dekker 2009). For the mouse \(H19/Igf2\) locus, silencing of the \(Igf2\) occurs by the formation of a tightly coiled loop by the interaction of CTCF bound at the ICR and the upstream differentially methylated region (DMR) and a downstream matrix attachment region (MAR), which prevent the interaction between the maternal-specific enhancer and the promoter (Yoon, Jeong et al. 2007, Li, Hu et al. 2008). The overlapping distribution patterns of genome-wide analysis of CTCF and cohesin as well as functional tests have now identified cohesin to be an important mediator for CTCF dependent loop function, with cohesin conferring the linking function and CTCF responsible for the sequence-specific binding (Parelho, Hadjur et al. 2008, Wendt, Yoshida et al. 2008, Hadjur, Williams et al. 2009, Nativio, Wendt et al. 2009). In fact, the cohesin component SA2 directly contacts CTCF, with the ring-forming components recruited by SA2 (Xiao, Wallace et al. 2011). Other players involved in the looping and insulation functions of CTCF include RNA helicase p68 and the non-coding RNA steroid receptor RNA activator (SRA), which is thought to help stabilize the CTCF-cohesin interaction (Yao, Brick et al. 2010).

e. Locus control regions: Locus control regions (LCR) are groups of regulatory sequences that regulate transcriptional programming of either an entire locus or gene cluster. They are operationally defined as elements that are able to enhance tissue-specific physiological expression of linked genes in a position-independent and copy-number dependent manner (Li, Peterson et al. 2002). LCRs are typically composed of multiple regulatory elements such as enhancers, silencers, insulators, and matrix attachment regions that are bound by transcription factors, which differentially affect gene expression. These collective activities define a LCR and
regulate spatio-temporal gene expression (Maston, Evans et al. 2006). Similar to enhancers and silencers, LCRs can regulate gene transcription from a distance. While typically located upstream, LCRs are reported to be present in introns, downstream of genes and even in introns of neighboring genes (Lang, Mamalaki et al. 1991, Aronow, Silbiger et al. 1992, Adlam and Siu 2003, Lee, Fields et al. 2003).

The first LCR region to be identified and the best studied to date, is the mammalian β-globin gene (Grosveld, van Assendelft et al. 1987). Transgenic mice containing 1.5kb of the mammalian β-globin promoter region failed to express β-globin gene to the extent seen in vivo, indicating that major regulatory regions required for proper expression were missing (Magram, Chada et al. 1985, Townes, Lingrel et al. 1985, Kollias, Wrighton et al. 1986). Further evidence from cases of β-thalassemia and later from transgenic mouse studies indicated that a LCR region upstream of the promoter was required for proper expression of the β-globin gene in vivo (Kioussis, Vanin et al. 1983, Grosveld, van Assendelft et al. 1987, Driscoll, Dobkin et al. 1989).

Although a number of models have been proposed to understand how LCRs function over long distance, a series of recent studies showed that in the case of the β-globin LCR, DNA looping is the primary mechanism employed by LCRs to make long-range physical contacts that lead to active chromatin formation and transcription of the β-globin gene (Liu, Austin et al. 2002, Tolhuis, Palstra et al. 2002, Bank 2006). The looping mechanism is also currently established for the Th2 LCR (Spilianakis, Lalioti et al. 2005).

**f. non-coding RNA:** The advent of high throughput assays and next generation sequencing has expanded the catalog of functional non-coding RNA. Among the growing classes of functional RNA, non-coding RNAs can be classified as short non-coding RNA, which are less than 200
base pairs (bps) in length, and the long-non coding RNAs (IncRNA) which are larger than 200 bps in length (Gibb, Brown et al. 2011).

Small non-coding RNA are functionally sub-divided into housekeeping non-coding RNA and regulatory non-coding RNA (Ponting, Oliver et al. 2009). The housekeeping non-coding RNA consists of ribosomal, transfer, small nuclear and small nucleolar RNAs. While a plethora of regulatory non-coding RNA have been identified (see (Esteller 2011) for review), the three main classes of small regulatory non-coding RNAs that have been well studied are the small interfering RNA (siRNA), micro RNA (miRNA), and the piwi-interacting RNA (piRNA).

Small interfering RNA’s are derived from long double stranded RNA (dsRNA) precursors that are either endogenous or exogenous in origin (Mello and Conte 2004). Dicer, a ribonuclease III enzyme, processes the dsRNA, leading to the formation of a short double stranded intermediate RNA. One strand, labeled as the guide strand is loaded onto the multi protein RNA-silencing complex (Driscoll, Dobkin et al.), leading to the posttranscriptional repression by endonucleolytic cleavage of the mRNA (McCaffrey, Meuse et al. 2002, Vagin, Klenov et al. 2004, Siomi, Sato et al. 2011). Small interfering RNA’s are also know to induce transcriptional gene silencing by inducing heterochromatin formation (Carthew and Sontheimer 2009).

micro RNA are a family of endogenous, short, non-coding RNA that consist of 21-25 nucleotides that regulate gene expression at the post-transcriptional level (Du and Zamore 2005, Cannell, Kong et al. 2008, Suh, Baehner et al. 2010). RNA polymerase II transcribes miRNA into 2-4kb long single stranded RNA which are capped and polyadenylated (Bartel 2004, Bartel 2005). These primary transcripts called the pri-mRNA form a stem-loop structure with an imperfectly paired ~33bp stem, a terminal loop and flanking segment. Drosha, a RNase III enzyme, bound to
co-factor DiGeorge Syndrome Critical region 8 (DGCR8) processes the pri-miRNA in the nucleus by excising the stem-loop to create the pre-miRNA. The pre-miRNA is exported to the cytoplasm by exportin-5 together with RAN-GTPase. Once exported into the cytoplasm, Dicer processes the pre-miRNA, in concert with Tar RNA-binding protein to form the ~22 nucleotide mature duplex miRNA (Liu, Fortin et al. 2008, Kim, Hwang do et al. 2009). One strand of the miRNA incorporates onto the RISC to function as mature miRNA and guide the RISC to target the specific mRNA (Lund and Dahlberg 2006, Chua, Armugam et al. 2009). The miRNA binds to the 3’-UTR “seed sequence” of the target mRNA transcript. If perfect or near-perfect base pairing occurs between the seed sequence and the miRNA, the mRNA transcript is degraded, inhibiting translation (Jackson and Standart 2007). However, multiple partial complementarity between miRNA and target mRNA can lead to decreased mRNA and protein levels (Pillai 2005).

miRNA are also implicated in regulation of gene expression by increasing translation in some biological systems (Vasudevan, Tong et al. 2007). Since the discovery of miRNA in *Caenorhabditis elegans*, miRNA have been shown to have diverse expression and act as master regulators of a number of fundamental biological progresses, such as proliferation, metabolism, aging, and cell death (Pasquinelli and Ruvkun 2002, He and Hannon 2004, Brennecke, Aravin et al. 2007, Melton, Judson et al. 2010). In the female reproductive tract, miRNAs have been shown to be essential for proper development and function, with loss of *Dicer1* resulting in female infertility (Nothnick 2012).

P-element induced wimpy testis (Piwi)-interacting RNA (piRNA), discovered in 2006, are small non-coding RNAs, ~26-30 nucleotides in length. PiRNAs were isolated by sequencing the small non-coding RNAs pulled down by immunoprecipitation of the Piwi protein in mammalian testes (Aravin, Gaidatzis et al. 2006, Girard, Sachidanandam et al. 2006, Grivna, Pyhtila et al. 2006,
Unlike the miRNA and the siRNA, these piRNAs are generated in a RnaseIII independent manner (Houwing, Kamminga et al. 2007). They also differ from miRNA and siRNA in origin, length, and structure. piRNAs are generated from nascent transcripts that arise from the piRNA clusters, intergenic repetitive elements in the genome, which are loaded onto the PIWI proteins. On maturation, the 3’ ends of the piRNAs are 2’-O-methylated by Hen1/Pime, leading to the stability of the piRNAs in vivo (Horwich, Li et al. 2007, Billi, Alessi et al. 2012). The ping-pong pathway amplifies primary piRNAs, wherein the antisense piRNA are associated with the Piwi subfamily of Argonaute proteins including Mili and Miwi, leading to the production of sense piRNAs, which then associate with Argonaute protein, Ago 3, and target anti-sense transcripts (Figure 2) (Siomi, Sato et al. 2011).

While the main role of piRNAs been associated with curbing the silencing of transposable elements in the germ cell line, recent studies have also implicated them in regulating memory storage in the brain (Siomi, Sato et al. 2011, Rajasethupathy, Antonov et al. 2012). In *Drosophila melanogaster*, piRNAs are transmitted maternally and are able to mount an effective silencing response in the progeny (Brennecke, Malone et al. 2008).

Noncoding RNAs that are greater than 200 nucleotides in length and do not have any coding potential are defined as long noncoding RNA (lncRNA). The completion of the Human Genome project and advancements in RNA sequencing, microarray technology, and cDNA cloning led to the conclusion that the number of genes coding for RNA were far greater than protein-coding genes. It was now clear that a vast majority of the genome was transcribed, however, the function of these noncoding RNAs was elusive, leading to the belief that these were products of translational noise (Shoemaker, Schadt et al. 2001, Birney, Stamatoyannopoulos et al. 2007, Ponjavic, Ponting et al. 2007, Struhl 2007, Clark, Amaral et al. 2011). Furthermore, RNA
polymerase II is known to initiate transcription spuriously in more than 90% of the cases, and transcription events can “spill over” or ripple to the neighboring genes leading to leaky expression (Ebisuya, Yamamoto et al. 2008).

Identifying functional IncRNA from the pervasive transcription of the genome is a daunting task and tantamount to finding a needle in a haystack. The massive parallel sequencing and understanding of different chromatin modifications identified clear signatures of polymerase II binding, namely Histone H3 lysine 4 trimethylation (H3K4me3) at the gene promoters, and histone H3 lysine 36 trimethylation (K4-K36 chromatin domain) at the transcribed region. Coupled with conserved patterns, coding patterns, and anatomical properties, great progress is now being made in identifying IncRNA (Mikkelsen, Ku et al. 2007, Marson, Levine et al. 2008, Guttman, Amit et al. 2009, Khalil, Guttman et al. 2009).

Based on their geographical location in the genome, relative to nearby protein-coding gene IncRNA can be grouped in: a. stand-alone IncRNA: b. natural antisense transcripts, c. pseudogenes, d. long intronic ncRNA and e. divergent transcripts, promoter-associated and enhancer RNA (Figure 3) (Kung, Colognori et al. 2013).

a. Stand-alone IncRNA: Also referred to as large intergenic noncoding RNAs (lincRNA), are located in regions that do not overlap protein-coding genes (Guttman, Amit et al. 2009, Cabili, Trapnell et al. 2011, Ulitsky, Shkumatava et al. 2011). Largely identified through active chromatin signatures of H3Kme3 at the promoter and H3K36me3 along the transcribed region, these noncoding RNA are usually 1kb in length, are transcribed by RNA pol II, are polyadenylated, and spliced. Examples of this noncoding RNA are *Xist, H19, HOTAIR*, and
b. Natural antisense transcript: As the name suggests, these transcripts occur opposite the sense DNA strand and are enriched at either the 5’ promoter or the 3’ terminator end of the sense transcript. Rarely are the lncRNA belonging to this class spliced or polyadenylated when compared to the stand-alone lncRNA (Kung, Colognori et al. 2013). Well-documented transcripts of this class are *Kcnq1ot1* and *Air* (Lyle, Watanabe et al. 2000, Kanduri, Thakur et al. 2006).

c. Pseudogenes: The term “pseudogene” coined in 1977, refers to genes that have lost their ability to code for functional proteins due to nonsense, frameshift and other mutations (Jacq, Miller et al. 1977, Pink, Wicks et al. 2011). Estimated to be numerically equal to protein-coding genes, these pseudogenes are by majority, transcriptionally silent. However, there are pseudogenes that are transcriptionally active and exhibit high levels of sequence conservation. These expressed pseudogenes maybe on their way to be completely silenced or are possibly resurrected for a functional purpose (Pink, Wicks et al. 2011). Long noncoding RNA produced from pseudogenes have added another interesting layer for gene
Figure 3: A summary of genomic location of lncRNA: LncRNA can be transcribed from enhancer, associated with the promoter or could be stand-alone transcripts. They are also intronic, or can be antisense to other genes, completely anti-sense (nested) or partially (terminal). They can also arise from pseudogenes or could host small RNA. Adapted from (Kung, Colognori et al. 2013)
regulation. A recent example is the PTENpg1, a *PTEN* pseudogene that regulates *PTEN* by encoding for three different lncRNA; two functional antisense (as) RNA and one sense PTENpg1. The sense PTENpg1 regulates *PTEN* by acting as a sponge for several miRNA that target *PTEN*. The PTENpg1 asRNA α is a trans acting transcript that binds to the promoter of *PTEN*, and inhibits transcription by recruiting epigenetic repressor complexes. The PTENpg1 as RNA β, stabilizes the sense PTENpg1 by binding to the 5’end (Johnsson, Ackley et al. 2013).

d. Long Intrinsic RNA: RNAs with regulatory roles have long been known to reside in intronic regions, indicating that intronic regions maybe more stable than previously thought. While a substantial fraction of the long intrinsic RNA correspond to the antisense transcript, a recent survey revealed noncoding RNA to have a strong preference to be associated with the transcribing sense strand (Nakaya, Amaral et al. 2007, Valen, Preker et al. 2011). Many of these are implicated in a myriad of biological functions and have been observed to respond to stimuli or misregulated in cancer (Guil, Soler et al. 2012). An example of long intrinsic RNA is COLDAIR, transcribed from the sense strand, it has been implicated in plant vernalization by recruiting PRC2 complex and inducing epigenetic silencing (Heo and Sung 2011).

e. divergent transcripts, promoter-associated and enhancer RNA: As a result of Pol II pausing a number of transcripts, such as transcription start site-associated (TSSa-) RNA, upstream antisense (ua) RNA or promoter upstream transcripts (PROMPTS), are formed at the transcription start site, both in the sense and anti-sense direction (Buratowski 2008, Core and Lis 2008). These transcripts are usually capped, polyadenylated, low in abundance, and degraded rapidly by exosomes. The current thought is that the act of transcription of these heterogeneous transcripts helps maintain the open chromatin status.
Another class of regulatory long noncoding RNA is the short, bidirectional transcripts formed at enhancer regions and hence called enhancer (e) RNA. These transcripts are postulated to function by activating the promoter or by keeping the chromatin in an open state, or by recruiting and interacting with enhancer-associated proteins (Orom, Derrien et al. 2010).

The ubiquitous expression of lncRNA indicates that these transcripts engage in diverse mechanisms to regulate gene expression. Over the past 15 years, multiple studies have shown that lncRNA are regulated during development, epigenetic regulation, imprinting, and are also associated with human disease (Kung, Colognori et al. 2013). Broadly, lncRNA can be classified according to their mode of action into four archetypes, a. signals, b. decoys, c. scaffold, and d. guide (Wang and Chang 2011). While this classification will help in demarcating their mode of regulation, an lncRNA may utilize multiple modes to bring about their function indicating that these archetypes are not mutually exclusive.

a. Signals: LncRNA can act as signaling molecules as they show cell-specific expression and respond to various stimuli. While some lncRNA belonging to this category have regulatory function, even as by-products, the act of transcription of lncRNA in itself indicates that the chromatin is in an active state. The advantage of using RNA as a medium is that the cell can skip protein translation and as a result, the regulatory functions can be performed quickly (Wang and Chang 2011). One example where lncRNA can act as signaling molecules is during imprinting. Kcnq1ot1 and Air transcribed at the Kcnq1 and Igf2r locus respectively, mediate transcriptional silencing by recruiting chromatin modifying complexes. Kcnq1ot1, a 90kb lncRNA directs silencing of a cluster of genes at the Kcnq1 domain, in the paternal allele. Kcnq1ot1 recruits histone methyltransferases G9a and PRC2, leading to bidirectional silencing of genes in the Kcnq1 domain (Pandey, Mondal et al. 2008). Similarly, the lncRNA Air, transcribed from the
second intron of the \emph{Igf2r} gene, represses several genes at the paternal chromosome in a tissue specific manner. The \emph{Air} IncRNA mediates silencing by recruiting G9a, a histone methyltransferase (Nagano, Mitchell et al. 2008).

X inactivation, a process in which one of the two chromosomes in the female is silenced, such that only one of the X chromosome is expressed is now known to be controlled by a cluster of IncRNA termed X-inactivation center. The 17kb X inactive specific transcript (\emph{Xist}), is expressed from the inactive X. It coats the X chromosome and forms a “Xist cloud” leading to the recruitment of the PRC2 complex (Zhao, Sun et al. 2008). \emph{Xist} is regulated by an antisense IncRNA called \emph{Tsix}, which reverses the action of \emph{Xist} by repressing the silencing caused by \emph{Xist} and \emph{Tsix} accomplishes this by several different mechanisms. Recruitment of the polycomb complex to the 5’ end of \emph{Xist} is accomplished by a 1.6kb ncRNA called RepA. \emph{Tsix} blocks loading of this complex to \emph{Xist}, inhibiting induction of the IncRNA, \emph{Xist} (Zhao, Sun et al. 2008). Transcription of \emph{Tsix} also causes silencing of \emph{Xist} activity by recruiting DNA (cytosine-5)-methyltransferase 3A (Dnmt3a) enzyme leading to the formation of repressive histone modifications inhibiting transcription of \emph{Xist} (Sado, Hoki et al. 2005). More recently, it has been shown that the very act of transcription of \emph{Tsix} can lead to suppression of \emph{Xist} as insertion of poly A cassette leading to the truncation of \emph{Tsix}, releases the repressive environment at the Xist promoter (Ohhata, Hoki et al. 2008). Another IncRNA \emph{Jpx}, regulated in trans activates \emph{Xist} on the inactive X chromosome (Lee, Davidow et al. 1999, Chureau, Prissette et al. 2002, Johnston, Newall et al. 2002, Tian, Sun et al. 2010).

b. Decoys: In this mechanism, IncRNA act as molecular decoys by binding and pirating a protein away from its intended target and restricting the functional outcome. In this case, IncRNA act as a repressor of function and its action can be negated by knocking down the IncRNA, allowing
the protein to bind and leads to gain of function (Wang and Chang 2011). The human dihydrofolate reductase \((DHFR)\) gene is an example of RNA dependant repression. The minor upstream promoter of the \(DHFR\) gene initiates transcription of the DHFR lncRNA, inhibiting the formation of the Pre-initiation complex (PIC) at the major complex by forming a stable ncRNA-DNA complex and by directly interacting with the general transcription factor IIB (TFIIB). Knockdown of the lncRNA, removed the repressor effect of the lncRNA, leading to high occupancy of the TFIIB at the major promoter (Martianov, Ramadass et al. 2007).

The “molecular sink” mechanism of these lncRNA is not restricted to proteins but also encompasses miRNAs and splicing factors. The tumor suppressor gene PTENP1 functions as a decoy and sequesters miRNA that would affect the transcriptional regulation of the PTEN gene. The 3’ UTR sequence of the PTENP1 RNA is similar to that of the PTEN gene, allowing the miRNA to bind to the PTENP1 RNA, and in the process, allowing the PTEN gene to be transcribed and translated (Poliseno, Salmena et al. 2010).

Metastasis-associated lung adenocarcinoma transcript (MALAT1) is a nuclear lncRNA abundantly present in nuclear speckles. This lncRNA binds to several serine/arginine (SR) splicing factors and sequesters them into nuclear speckles. Depletion of MALAT1 leads to altered splicing pattern for pre-mRNA (Tripathi, Ellis et al. 2010). MALAT1 regulates splicing factors in hippocampal neurons and is important for synapse formation (Bernard, Prasanth et al. 2010). Thus, lncRNA can act as decoys and sequester away regulatory factors both in the cytoplasmic and nuclear domains.

c. Guides: The third mechanism of lncRNA is as a guide leading to the proper localization of specific complexes to their target regions. This can occur both in \textit{cis} (on neighboring genes) or in
trans (distantly located genes) and can include both repressive and activating complexes. The underlining concept is that these groups of lncRNA convey regulatory information and control gene expression. The functional role of these lncRNA can be predicted by knocking down the lncRNA leading to the loss of proper localization of the effector molecule or loss of function of the effector, or both.

One of the best-studied “guide” lncRNA is at the X-inactivation center (XIC), which controls silencing of one of the X chromosomes. Coating of Xist on the X chromosome marks it for repression. This silencing is initiated by the recruitment of the PRC2 complex by RepA, a 1.5kb lncRNA, which originates from the 5’ end of Xist (Wutz, Rasmussen et al. 2002, Sun, Deaton et al. 2006). A similar mechanism is employed by the lncRNA Air, which recruits G9a leading to H3K9 methylation and silencing (Nagano, Mitchell et al. 2008). In the plant lncRNA, COLDAIR, guides the PRC2 complex to the floral repressor gene FLC, during vernalization, leading to its repression by trimethylation of H3K27 (Heo and Sung 2011).

LncRNA act both in cis as well as in trans as seen in the case of the Hox lncRNA HOTAIR. HOTAIR plays a critical role in cancer cells, as depletion of this lncRNA reduces invasiveness of cells that express high levels of PRC2 (Gupta, Shah et al. 2010). Long intergenic non-coding RNA-p21 induces gene expression changes across multiple sites in the genome. Ectopic expression of LincRNA-p21 leads to apoptosis by bypassing the upstream p53 (Huart, Guttman et al. 2010).

d. Scaffolds: Originally thought to be the function of proteins, recent evidence now point to the possibility that lncRNA may play similar roles (Good, Zalatan et al. 2011). Highly intricate and complex, the lncRNA binds to multiple effectors at the same time leading to either gene
activation or repression. Knockdown of these lncRNA could change the spatial occupancy of the effector leading to dismantling of the complex, loss of phenotype, or both.

The scaffold mechanism can be seen in telomerase activity, a fundamental process that maintains genomic stability by adding back telomere DNA repeats lost from chromosome ends. This catalytic activity requires the association of TERT, a catalytic protein subunit and a telomerase RNA (TERC). TERC provides a template, contributes to binding of TERT and its catalytic activity, and plays a major role in stability of the complex (Lustig 2004). In the disease condition of dyskeratosis congenita, mutations affect the conformation of TERC leading to loss of RNA scaffold structure underlining the functional importance of TERC (Chen and Greider 2004).

Recent work has identified that a 300-nucleotide fragment present at the 5’ end of the lncRNA HOTAIR, is responsible for the binding of the PRC2 complex (Tsai, Manor et al. 2010). In addition, the 700 nucleotides of the 3’end of HOTAIR was found to interact with Lysine specific demethylase (LSD1), RE1 silencing transcription factor (REST), and its co-repressor protein, CoREST that demethylates H3 on K4 leading to gene repression (Tsai, Manor et al. 2010). These findings identify HOTAIR as a scaffold and bridge between the PRC2 and the LSD1/CoREST/REST complex leading to suppression of gene expression.

While these regulatory mechanisms encompass the majority of known functional lncRNAs, the growing numbers of lncRNA are certain to reveal more mechanistic and functional aspects of lncRNA. While the argument remains as to if these merely represent transcriptional noise, support is increasing for their functional significance and thus adds another level of complexity to the mechanisms of gene regulation. The varied sub-cellular expression, nuclear or cytoplasmic, cell-specific expression patterns, and varying levels of conservation among the
lncRNA add chaos to the understanding of these transcripts. Another intriguing possibility is that lncRNAs may be tools for evolution to tinker in generating mechanisms for survival and advancement of the species (Kung, Colognori et al. 2013). One thing is for sure, we have barely scratched the surface of the lncRNA world and as more are discovered, their versatility as regulators of gene expression will be revealed.
The writing in this chapter has been previously published in Biology of Reproduction.

Current concepts Of follicle stimulating hormone receptor gene regulation.
George JW, Dille EA, Heckert LL.
The gonadotropin hormones

Luteinizing hormone (LH) and follicle-stimulating hormone (FSH), are integral parts of the neural and endocrine interchange between the hypothalamus, pituitary, and gonads that controls steroid hormone synthesis and gamete production (reviewed in (Plant 2008)). At the top of the network is gonadotropin releasing hormone (GnRH), which, once released from the hypothalamus, binds receptors on pituitary gonadotrophs and induces the synthesis and secretion of LH and the FSH. Once in circulation, LH and FSH finalize the communication by binding their receptors and transmitting signals to the gonads. These signals are at the hub of the regulatory network, relaying neuronal signals from the hypothalamus to the gonads and inducing feedback signals returned to the hypothalamus and pituitary. The receptors for FSH and LH, FSHR and LHR (LHCGR), reside on the surface of somatic cells in the gonads and are members of the Rhodopsin receptor family of G-protein coupled receptors, but unlike the other members, LHR and FSHR have extended NH2-terminal extracellular domains with numerous leucine-rich repeats that assist ligand specificity (Braun, Schofield et al. 1991, Dias, Cohen et al. 2002, Vassart, Pardo et al. 2004, Bogerd 2007, Lagerstrom and Schiotth 2008). FSH binding elicits several diverse signaling events, but the most characterized is that initiated by adenylyl cyclase, followed by induction of cAMP, PKA activation, and protein phosphorylation (Heindel, Rothenberg et al. 1975, Dorrington and Armstrong 1979). FSH binding is also associated with increased intracellular calcium, activation of mitogen activated protein kinase (MAPK), and stimulation of inositol triphosphate (IP3) (Flores, Veldhuis et al. 1990, Tena-Sempere, Manna et al. 1999, Seger, Hanoch et al. 2001).

Since FSH acts exclusively through FSHR, mechanisms controlling receptor expression determines the FSH-responsive cell population and influences their sensitivity to hormone. Thus,

**FSHR expression**

Expression of FSHR, both protein and mRNA, is remarkably limited with respect to its cellular profile, with Sertoli and granulosa cells by far the predominant expressing cell types
FSHR/Fshr transcripts are first observed in embryonic gonads, around embryonic day 14.5 in males and 20.5 in females (Dankbar, Brinkworth et al. 1995, Rannikki, Zhang et al. 1995). These initial transcripts are incomplete and represent only the extracellular portion of the receptor, with full-length mRNA expressed several days later (Richards, Ireland et al. 1976, Sokka and Huhtaniemi 1990, Rannikki, Zhang et al. 1995). In the rodent ovary, FSHR expression coincides with primary follicle formation and follicular development through the pre-antral stage, with initial full-length transcripts and hormone binding observed shortly after birth (around postnatal day 3) and continuing to increase through day 21 (Sokka and Huhtaniemi 1990, Dunkel, Tilly et al. 1994, O'Shaughnessy, Marsh et al. 1994, Rannikki, Zhang et al. 1995, Drummond, Dyson et al. 1996). In the testis, full-length FSHR mRNA initiates during fetal development (around embryonic day 16.5 in the rat) and expression is maintained throughout development and in the adult testis (Steinberger, Thanki et al. 1974, Nimrod, Erickson et al. 1976, Sprengel, Braun et al. 1990, Heckert and Griswold 1991, O'Shaughnessy, Marsh et al. 1994, Dankbar, Brinkworth et al. 1995, Rannikki, Zhang et al. 1995). Once the spermatogenic cycle is initiated, FSHR levels change with the cycle, with levels highest at stages X-II and lowest at VI-VII (Heckert and Griswold 1991, Kliesch, Penttila et al. 1992, Heckert and Griswold 1993, Rannikko, Penttila et al. 1996). Several signals that regulate ovarian and testicular physiology also influence FSHR expression. In the ovary, FSHR is regulated by a combination of transcriptional and posttranscriptional mechanisms induced by FSH and activin and indirectly by follistatin through its influence on activin (Knecht, Ranta et al. 1983, Woodruff, D'Agostino et al. 1988, Sanford and Batten 1989, Nakatani, Shimasaki et al. 1991, Themmen, Blok et al. 1991, Nakamura, Minegishi et al. 1993, Sites, Patterson et al. 1994,

While FSHR expression is considered gonad-specific and restricted to Sertoli cells in the testis and granulosa cells in the ovary, there are a few notable reports of its presence elsewhere; in particular, uterus, prostate, bone, and the ovarian surface epithelia (Zheng, Magid et al. 1996, Ben-Josef, Yang et al. 1999, Mariani, Salvatori et al. 2006, Sun, Peng et al. 2006). In prostate and ovarian epithelial cells, FSH signaling is implicated in cell proliferation and tumor invasiveness in precancerous and malignant cells, and thus reports of receptor expression are often within the same cell context (Ben-Josef, Yang et al. 1999, Choi, Choi et al. 2004, Ji, Liu et al. 2004, Zhang, Chen et al. 2009). In bone, there is strong evidence linking FSH and FSHR to hypogonadal bone loss in women but there is still uncertainty as to the site of FSH action (Sun, Peng et al. 2006, Prior 2007, Zaidi, Blair et al. 2007, Ritter, Thuering et al. 2008, Robinson, Tourkova et al. 2010). Consequently, it has been difficult to assess from the literature the degree to which normally differentiated cells, other than Sertoli and granulosa cells, actively transcribe FSHR/Fshr. However, its specificity can be greatly appreciated using publically available, high-throughput expression data to examine distribution of FSHR/Fshr mRNA. As an example, data derived from more than 30 tissues, using massively parallel signature sequencing, (FSHR query with GEO Series accession number GSE1581 http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1581 and dataset record GDS868)
showed \textit{FSHR/Fshr} mRNA present only in the testis and ovary; a remarkable finding given the high sensitivity of the detection method (Edgar, Domrachev et al. 2002, Barrett, Troup et al. 2007, Barrett, Troup et al. 2009). What is the mechanism responsible for this remarkable cell-specificity? Currently, it is unknown but as reviewed below, the evidence indicates that \textit{FSHR/Fshr} expression is directed by transcription factors that function through elements located at significant distances from the gene itself.

\textbf{Transcription of \textit{FSHR/Fshr}}

\textit{FSHR/Fshr} transcription contributes to receptor levels and directs its cell-specificity, indicating components of the underlying mechanisms are important for both FSH responsiveness and target cell identity. Our current understanding of \textit{FSHR/Fshr} transcription is derived largely from studies on the rat, murine, ovine, and human \textit{FSHR/Fshr} genes, which focused on characterization of the 5' flanking region. The resulting data provided significant insight, revealing both similarities and differences between promoters of the four species. Since most promoter characteristics were detailed in a previous review, the discussion here will be limited to its prominent features (Heckert 2005). The accumulated information represents primarily transient transfection and DNA/protein binding results from Sertoli and granulosa cells. These studies identified regulatory elements and their associated binding proteins within promoters represented by various species and lengths and revealed both similarities and differences in promoter function (Huhtaniemi, Eskola et al. 1992, Gromoll, Dankbar et al. 1994, Goetz, Lloyd et al. 1996, Heckert, Daggett et al. 1998, Heckert, Sawadogo et al. 2000, Heckert 2001, Kim and Griswold 2001, Xing and Sairam 2001, Xing, Danilovich et al. 2002, Xing and Sairam 2002, Xing and Sairam 2002). Promoter sequence comparisons between several species showed significant conservation of approximately 1000bp 5' to the translational start, a reference point...

Located just 5’ to the transcriptional start sites, the E box and its binding proteins have, by far, received the greatest attention with respect to FSHR/Fshr transcription. USF1 and USF2 are members of the helix-loop-helix family that form both homo- and heterodimers (USF1, USF2, USF1/2, respectively) and considerable evidence, from studies in rodents, document their role directing Fshr promoter activity via the E box (Shieh, Sparkes et al. 1993, Goetz, Lloyd et al. 1996, Viollet, Lefrancois-Martinez et al. 1996, Heckert, Daggett et al. 1998, Findlay and Drummond 1999, Heckert, Sawadogo et al. 2000, Griswold and Kim 2001, Heckert 2001, Rodriguez, Girones et al. 2003). In both testis and ovary, in vitro (electrophoretic mobility shift
assay) and *in vivo* (chromatin immunoprecipitation, ChIP) studies showed USF1, USF2 and USF1/2 bind the promoter, but dimer composition differed between males and females, with USF1/2 and USF1 favored in the testis and USF1, USF2, and USF1/2 equally matched in the ovary (Figure 1B) (Hermann, Hornbaker et al. 2008). Furthermore, *Fshr* mRNA expression and *in vivo* promoter binding evaluated in *Usf1* and *Usf2*-null mice, revealed differences between testis and ovary in their response to loss of either USF protein (Hermann, Hornbaker et al. 2008). Thus, in testis, *Fshr* expression was unchanged by either *Usf1* or *Usf2* deletion and showed compensatory increases in promoter-bound USF homodimers (Hermann, Hornbaker et al. 2008, Viswanathan, Wood et al. 2009). In contrast, ovarian *Fshr* expression declined in both and compensatory change in homodimer binding was not indicated. Additional studies on the *Fshr* promoter in Sertoli cells, showed USF binding to the E box increases during differentiation and decreases, together with promoter activity, upon FSH treatment (Scobey, Fix et al. 2004, Viswanathan, Wood et al. 2009, Wood and Walker 2009). The predicted mediator of these FSH-induced changes is the inhibitor of DNA binding/differentiation protein, ID2, which increases with FSH treatment and inhibits both E box binding and promoter activity (Figure 4B).
FIGURE 4: Summary of FSHR/Fshr transcriptional regulation. A) Identified elements within the FSHR/Fshr promoter region are depicted (rectangles) and their identified binding proteins indicated as shapes above the element: SF1 (Steroidogenic factor 1), USF (Upstream stimulatory factor), GATA binding protein (GATA), and Octamer-binding protein 1 (OCT1, POU2F1). DHS#3, noted in the first intron, is bound by a GATA-containing complex, which contains the OCT1 in non-expressing cell but not FSHR-expressing Sertoli cells. The presence of OCT1 within the DHS#3-bound GATA complex correlates with DHS#3 silencing activity observed in non-expressing cells. Also indicated are distal regulatory elements predicted by studies in transgenic mice, which failed to recapitulate Fshr expression from using its promoter region (up to 5.0Kbp in size) or a yeast artificial chromosome containing the entire gene plus considerable flanking sequence. B) Binding of USF1 and USF2 to the Fshr promoter occurs only in expressing (Sertoli cells and granulosa) cells. In Sertoli cells, dimer composition favors heterodimers and USF1 homodimers, while in granulosa cells, the relative composition of heterodimer and homodimer binding is similar. In Sertoli cells, FSH induces the inhibitory protein ID2, which reduces USF binding and Fshr expression.
Despite numerous studies on the FSHR/Fshr promoter, no mechanism has evolved to explain the gene’s remarkable cell specific expression. This deficiency in promoter specificity was also demonstrated by studies that evaluated 16 distinct transgenic mouse lines for cell-specific expression of reporters directed by either 5.0kbp or 143bp (8 lines each) of rat Fshr promoter sequence, none of which showed Sertoli or granulosa cell expression (Heckert, Sawadogo et al. 2000). Similar findings were reported for 1.5kbp of the human promoter (Nordhoff, Gromoll et al. 2003). The recognition that sequences beyond the 5.0Kbp promoter were required for expression, together with emerging data on regulatory elements that act from distal positions, led to studies using yeast artificial chromosomes (YACs) as transgenes, in order to define the region required for FSHR/Fshr specificity. The absent transgene expression in Sertoli and granulosa cells of mice carrying a YAC with the entire rat Fshr gene, plus bordering sequences stretching more than 50kbp 5’ and 30kbp 3’, further supports involvement of distal regulation and suggests regulatory regions extend well beyond the gene itself (Hermann, Hornbaker et al. 2007).

With the evidence that proper expression of FSHR/Fshr requires contributions from regulatory elements located outside the promoter region, most likely at significant distances, the challenge became identifying these sites within a vast amount of potential sequence. Initially, with only a small amount of available sequence, this was tackled using conventional DNase I hypersensitivity mapping to identify regions of accessible chromatin and thus potential sites of regulatory importance (Hermann and Heckert 2005). This revealed four hypersensitive sites located within a 45kbp region surrounding the first exon, three of which showed significant sequence conservation, a predictive feature of important regulatory elements (Hermann and Heckert 2005). One of these, DHS#3, was located approximately 4kb downstream in intron 1 and showed much greater sensitivity to DNase I in non-expressing cells (myoid) than expressing
cells (Sertoli), suggesting an association with gene silencing (Figure 1). Further functional characterization, using transient transfections and *in vitro* and *in vivo* binding assays, revealed important elements that attenuated gene expression and their cognate binding proteins, OCT1, GATA4, and GATA1. The studies also implicated OCT1 in selective binding to this element in order to maintain its silent state in non-expressing cells (Hermann and Heckert 2005). While the approach proved to be a valid means to identify important regulatory elements, it was evident that scanning extensive regions of the genome without better knowledge of its DNA content or genomic landscape was impractical. This has since been remedied by the infusion of genomics data that brought, not only new sequence information, but also a wealth of insight on the *FSHR/Fshr* gene and it residing landscape.

**The *FSHR* gene**

The initial characterization of the *Fshr* gene in 1992, which revealed a pronounced structural similarity to the *LHCGR/Lhcgr* gene that suggested the two evolved through duplication of a common ancestral gene (Heckert, Daley et al. 1992). Similarities between *FSHR/Fshr* and *LHCGR/Lhcgr* and the genes for other G-protein coupled receptors (GPCR) also suggested the predecessor for *FSHR/Fshr* and *LHCGR/Lhcgr* was formed by combining a common GPCR ancestral gene. The ancestor for these genes presumably arose by encoding the characteristic transmembrane and intracellular domains, with multiple repeated exons derived from tandem duplications of a module for a leucine-rich motif, a featured attribute of the extracellular domain of glycoprotein hormone receptors (Heckert, Daley et al. 1992). Formation of the gonadotropin receptor genes through tandem duplication of an ancestral gene was further substantiated by recent genomics data that show they are arranged in tandem on chromosomes from nearly all annotated Tetrapoda genomes (Heckert, Daley et al. 1992, Montgomery, Tate et al. 1995,
Chauvigne, Tingaud-Sequeira et al. 2010). While the two ancestral descendents, *FSHR/Fshr* and *LHCGR/Lhcgr*, differ by one exon (10 for *FSHR/Fshr* and 11 for *LHCGR/Lhcgr*), the coding scheme is largely the same, with the carboxy-terminal, transmembrane-intracellular domain encoded by the last exon and the amino-terminal, extracellular domain by all preceding exons (Koo, Ji et al. 1991, Tsai-Morris, Buczko et al. 1991, Heckert, Daley et al. 1992). Exons that partition the receptors’ extracellular domains also share a repeated structure that delineates seven of the leucine-rich motifs into exons 2-8 and two into exon 9 (Figure 5).

However, despite knowing the structure of *FSHR/Fshr* for nearly two decades, it was not until the various genome-sequencing projects greatly expanded the available sequence data that there was accurate knowledge of its size or chromosome habitat. Now, *FSHR/Fshr* chromosome locations and annotated sequences are reported for more than 40 vertebrate species through the University of California, Santa Cruz Genome Browser (http://genome.ucsc.edu/) and similar web sites. In human, rat, and mouse *FSHR/Fshr* are located on chromosomes 2, 6 and 17, respectively, and span roughly 200kb, a size much larger than originally predicted (Heckert, Daley et al. 1992, Huhtaniemi, Eskola et al. 1992, Rousseau-Merck, Atger et al. 1993, Gromoll, Ried et al. 1994, Rhead, Karolchik et al. 2010) (Figure 2). This wealth of sequence information also disclosed several defining features of *FSHR/Fshr* and its surrounding neighborhood that indicate its associated regulatory environment is strongly influenced by evolutionary constraints that retain regulatory sequences directing distally-located genes (Hermann and Heckert 2005, Hermann, Hornbaker et al. 2007). This includes the tandem placement of *FSHR/Fshr* between its closest 5’ neighbor *NRXN1/Nrxn1*, which encodes the synaptic neuronal adhesion protein NEUREXIN 1, and its closest 3’ neighbor *LHCGR/Lhcgr*, the gene encoding the luteinizing hormone receptor, to form a highly conserved syntenic block, a feature identified by several
studies to indicate evolutionary constraints intended to preserve the relative positions between noncoding sequences and their target genes (Hermann and Heckert 2005, Engstrom, Ho Sui et al. 2007, Kikuta, Laplante et al. 2007, Akalin, Fredman et al. 2009). The genomic relationship between \textit{FSHR/Fshr} and \textit{LHCGR/Lhcgr}, when examined together with their similar expression profiles and functions in gonadotropin signaling, raises the intriguing possibility that function of one or both genes depends on their relative positions. In considering potential mechanisms, a link between transcriptional regulation and the shared synteny can be readily conceived if \textit{FSHR/Fshr} and/or \textit{LHCGR/Lhcgr} expression depends on regulatory sequences that would be lost if the genes separated. This implies the sequences are either; 1) required for expression of one gene and reside within the locus defined by the other (gene plus regulatory domain) or 2) required by both genes (i.e. identical or overlapping sites) and function depends on their position relative to \textit{FSHR/Fshr} and \textit{LHCGR/Lhcgr}. The second mechanism can be expanded to include the use of sequences for concurrent gene regulation, which, for co-regulation of \textit{FSHR/Fshr} and \textit{LHCGR/Lhcgr} activity is limited to granulosa cells of growing follicles, the only cells that express both genes. So despite similar gonad-specific expression profiles, the receptors are largely confined to distinct cell types, where any simultaneous use of a regulatory sequence will require opposite transcriptional effects on the two genes. Thus, most gonadotropin-responsive cells do not co-express \textit{FSHR/Fshr} and \textit{LHCGR/Lhcgr} and FSH response is restricted to Sertoli and granulosa cells of the testis and ovary, respectively, and LH response is confined to testicular Leydig cells and ovarian theca, maturing granulosa, and luteal cells (Camp, Rahal et al. 1991, Peng, Hsueh et al. 1991, Kaminski, Gawronska et al. 2000, Zhang, Shi et al. 2001, Dickinson, Stewart et al. 2009).
FIGURE 5: Organization of FSHR/Fshr and its genomic environment. A) Structure of FSHR/Fshr and LHCGR/Lhcg, depicting the exon distribution of the receptor’s protein domains. Exons are indicated by rectangles, intervening regions by lines, and transcriptional direction by arrows. B) Position of the NRXN1;FSHR;LH/CGR syntenic region on human chromosome 2 (red box indicated by arrow). C) Syntenic region of NRXN1/Nrxn1, FSHR/Fshr, and LHCGR/Lhcg from human, rat, mouse, and chicken, with sizes indicated for the genes and intervening regions. Genes are indicated by rectangles, intervening regions by lines, and transcriptional direction by arrows.
To date, there are no identified regulatory elements shared by *FSHR/Fshr* and *LHCGR/Lhcgr* or located within the other’s defined locus. However, our knowledge of their transcription is represented almost entirely by promoter characteristics and, therefore, insufficient to conclude the elements do not exist. Regardless, promoter characteristics do suggest divergent regulation, as they differ significantly with respect to sequence, identified regulatory elements, and activity in transgenic mice, i.e. *LHCGR/Lhcgr*, but not *FSHR/Fshr*, promoter directs cell-specific expression *in vivo* (Heckert and Griswold 1991, Heckert, Daley et al. 1992, Tsai-Morris, Xie et al. 1993, Tsai-Morris, Geng et al. 1994, Tsai-Morris, Geng et al. 1995, Goetz, Lloyd et al. 1996, Heckert, Daggett et al. 1998, Hamalainen, Poutanen et al. 1999, Hamalainen, Poutanen et al. 2001, Kim and Griswold 2001, Apaja, Aatsinki et al. 2005, Hermann, Hornbaker et al. 2007).

On the other hand, there are a few common promoter features worth noting. First is the core structure, which, for both, lacks a TATA box and has multiple transcription initiation sites within a region similar to an initiator element (Tsai-Morris, Buczko et al. 1991, Heckert, Daley et al. 1992, Dufau, Tsai-Morris et al. 1995, Goetz, Lloyd et al. 1996, Juven-Gershon, Hsu et al. 2008). Second is the ubiquitous nature of each promoter’s main functional element(s), which depend largely on non-specific, widely expressed but distinct, transcription factors; *LHCGR/Lhcgr* with two GC-rich sequences that bind SP1 and SP3 and *FSHR/Fshr* with a single E box bound by USF1 and USF2 (Tsai-Morris, Geng et al. 1995, Goetz, Lloyd et al. 1996, Geng, Tsai-Morris et al. 1999, Heckert, Sawadogo et al. 2000). While these similarities may reflect the common ancestry and/or association via an undisclosed fundamental mechanism, the predominant promoter characteristics emphasize functional divergence of key regulatory features that highlight the promoter for *LHCGR/Lhcgr* and distal regulatory sites for *FSHR/Fshr*. 
Another notable feature is the large intergenic distances between FSHR/Fshr and its neighbors, which span more than 750kbp on its 5’ side and 200kbp on its 3’ side (Hermann and Heckert 2005). The uninhabited region between NRXN1/Nrxn1 and FSHR/Fshr is characteristic of a gene desert, or long genomic stretch devoid of protein-coding sequences or other obvious biological functions. Such regions, when located within a conserved syntenic block, are associated with areas of enhanced sequence conservation, suggesting the content was under evolutionary pressure to retain functional and contextual information of resident elements (Venter, Adams et al. 2001, Ovcharenko, Loots et al. 2005, Akalin, Fredman et al. 2009). While evidence shows not all gene deserts have measurable activity, numerous risk loci and regulatory elements are documented within these regions, confirming their importance to the genome (Lodder, Eussen et al. 2009, Xu, Tsumagari et al. 2009, Kiltie 2010). Comparative sequence analysis has been consistently used in genome-wide studies to evaluate evolutionary constraint as a means to understanding genome structure, biological function, and evolution (Waterston, Lindblad-Toh et al. 2002). The initial whole-genome sequence comparisons between mouse and human provided considerable insight and estimated 5% of the human genome was conserved over 70-100 million years (Lander, Linton et al. 2001, Venter, Adams et al. 2001, Dermitzakis, Reymond et al. 2002, Mural, Adams et al. 2002, Waterston, Lindblad-Toh et al. 2002, Pennacchio 2003). What was remarkable in this finding was that only 1/3 of the conserved sequences were located in coding regions. Thus, the genome’s non-coding sector represents the largest portion under evolutionary selection, which suggests there is considerable functional information within conserved non-coding sequences, aka evolutionary conserved regions (ECRs) (Waterston, Lindblad-Toh et al. 2002, Dermitzakis, Reymond et al. 2005).
Expansion of genome sequence data and species representation added significantly to the power of cross-species sequence comparison, the genomic landscape and sites of potential biological functions. These breakthroughs have led to identification of many non-coding sequences involved in gene regulation, including enhancers, insulators, silencers and matrix attachment regions (Glazko, Koonin et al. 2003, Koonin 2003, Nobrega, Ovcharenko et al. 2003, de la Calle-Mustienes, Feijoo et al. 2005, Woolfe, Goodson et al. 2005, Pennacchio, Ahituv et al. 2006, Prabhakar, Poulin et al. 2006, Visel, Prabhakar et al. 2008). A connection between sequence conservation and regulatory sequences was nicely illustrated in a recent study that used ChIP with massively parallel sequencing to map DNA sites linked to p300, a co-regulator for many transcription factors (Visel, Blow et al. 2009). Results from mouse embryonic tissues (forebrain, midbrain and limb) showed p300 highly enriched at sites containing conserved non-coding sequence and, of those evaluated, nearly all were functional. The study also demonstrated that most p300-associated sites/enhancers were located at least 10kbp from a potential target gene, which suggests distal elements are commonly involved in transcriptional control. Direct evidence for long-distance gene regulation is found in numerous studies that demonstrate transcriptional effects from distant sequences on specific target genes, such as those for β-globin, IFγ, SOX9, GATA3, FGF4, IL-10, and CD69 (Martinez-Jimenez, Gomez-Lechon et al. 2005, Bejerano, Lowe et al. 2006, Schoenborn, Dorschner et al. 2007, Vazquez, Laguna et al. 2009). The significance of non-coding regulatory sequences to human disorders has also gained recognition with reports linking them to diseases and developmental disorders, as noted in cases of X-linked deafness, preaxial polydactyl, campomelic dysplasia, sex reversal, postaxial polydactyly (de Kok, Merkx et al. 1995, Bishop, Whitworth et al. 2000, Jamieson, Perveen et al. 2002, Epstein 2009, Lodder, Eussen et al. 2009). In the case of preaxial polydactyly, the reported
mutation was located within a regulatory element approximately 1Mbp from its target gene that emphasizes not only the remarkable linear distance in which the transcriptional signal passes but also the importance of nuclear architecture in positioning regions that collaborate in the transcriptional signal (reviewed in (Schneider and Grosschedl 2007)).

**Identification of new regulatory elements**

The search for new regulatory elements is clearly facilitated by predictions based on sequence conservation. However, the evidence is also clear that sequence conservation does not detect all regulatory elements nor provide any assurance that the predicted sites are functional, and thus, most effective when used in conjunction with other corroborating techniques, particularly ones that can assess chromatin changes linked to transcriptional activity (Nobrega, Ovcharenko et al. 2003, Dermitzakis, Reymond et al. 2005, Rizzolio, Bione et al. 2008, D'Haene, Attanasio et al. 2009, Liska, Snajdr et al. 2009, Vazquez, Laguna et al. 2009, Visel, Zhu et al. 2010). In studies to identify regulatory sequences that direct FSHR/Fshr transcription, results from transgenic and promoter studies shifted the experimental focus away from the promoter to the region encompassing all of FSHR/Fshr and its adjoining intergenic regions, which required a new set of tools and resources that fortunately evolved from the collection of genome sequences and efforts to understand their content. Initial reports on the FSHR/Fshr conservation profile compared human and rat sequences by direct pairwise comparison and analysis of precompiled LAGAN alignments through the web-based VISTA genome browser (Hermann and Heckert 2005). This revealed over 150 conserved sites, which, when matched together with DNase I hypersensitivity data, was instrumental in the identification of an important silencing region in the first intron (discussed above). However, it was also evident that greater constraints were required to improve functional prediction and therefore, once more distant genomes (e.g. chicken) were
available and included in computations, ECRs having greater predictive power could be distinguished and seven of the most conserved were selected for functional testing by transient transfection (Hermann, Hornbaker et al. 2007). With continuous enhancements in genome data and resources that improve regulatory element prediction, the number of selected sites has grown to more than twenty, which includes the original seven ECRs. These sequences were identified using the human genome and the UCSC Genome Browser (http://genome.ucsc.edu/) to reveal highly conserved (vertebrate consensus and net alignment with chicken as the target), non-coding sequences with a significant regulatory potential score (7X regulatory potential; (Waterston, Lindblad-Toh et al. 2002, 2004, Gibbs, Weinstock et al. 2004, Havlak, Chen et al. 2004, Kolbe, Taylor et al. 2004, King, Taylor et al. 2007)). Figure 6 provides an example of the FSHR locus spanning from LHCG to NRXN1 that was modified from the UCSC Genome Browser results to show the top predicted sites (marked ECRs) for FSHR/Fshr regulatory sequences and the key features used in their selection.
FIGURE 6: Annotated region of human FSHR and neighboring genes NRXN1 and LHCGR from the UCSC Genome Browser (http://genome.ucsc.edu/). With human as the reference genome, noncoding evolutionary conserved regions (ECRs, top) were selected for their potential regulatory activity using sequence conservation, both vertebrate consensus and net alignment (chicken as the target), and 7X regulatory potential. Vertebrate consensus represents alignments of 44 vertebrate species generated using multiz and other tools and measurements of evolutionary conservation determined by phastCons and phyloP (Siepel, Bejerano et al. 2005, Pollard, Hubisz et al. 2010). The 7X regulatory potential track represents scores computed from human, chimpanzee, macaque, mouse, rat, dog, and cow by comparing short alignment pattern frequencies between known regulatory elements and neutral DNA (Waterston, Lindblad-Toh et al. 2002, Gibbs, Weinstock et al. 2004, Havlak, Chen et al. 2004, Kolbe, Taylor et al. 2004, King, Taylor et al. 2007). The net alignment tracks show the best mouse/human or chicken/human chains for every part of the human genome, with ungapped alignments represented by boxes and gaps by lines. Annotated sequences are from the chicken May 2006 (WUGSC 2.1/galGal3) (galGal3) assembly, mouse July 2007
Summary

The evidence to date shows that sequences directing FSHR/Fshr expression lie far from the start of transcription in a regulatory environment without defined boundaries, which complicates their detection using standard molecular approaches. While computational genomics has helped narrow the search, limitations due to false positives and undetected sequences caution its use without additional methods to substantiate the data (Giresi, Kim et al. 2007). Fortunately, many technologies have adapted to the influx of sequence data by developing high-throughput and genome-wide strategies. Two such strategies offer considerable promise for regulatory element identification and that the FSHR/Fshr transcriptional mechanism is within reach. Both strategies reveal chromatin signatures featured in regulatory sequences; one identifies sequences bound to modified histones linked to transcriptional activity by chromatin immunoprecipitation, the other identifies open regions of chromatin, similar to DNase I hypersensitivity, by formaldehyde associated identification of regulatory elements. (Wu, Smith et al. 2006, McGaughey, Stine et al. 2009). Implementation of such strategies together with comparative genomics will significantly enhance the probability of relevant sequence identification and the mechanistic understanding of the regulatory landscape. When combined with high-throughput strategies, such as DNase-Chip, high-density tiling arrays and next generation sequencing, to canvass the genome without the bias of conservation, additional insight is likely on mechanisms that employ non-conserved regulatory elements and possible contributions to species-specific regulatory features (Crawford, Davis et al. 2006, Roh, Wei et al. 2007, Chen, Lin et al. 2008).
Chapter 2

CTCF and transcription regulation of Fshr in rat granulosa cells
Abstract

The anterior pituitary in response to gonadotropin releasing hormone (GnRH) stimulates the release of follicle stimulating hormone (FSH). FSH targets the gonad by binding specifically to its cell-surface receptor, FSHR, present on Sertoli cells of the testis and granulosa cells of the ovary. FSHR hence forms a bridge for FSH action and plays a crucial role in mediating gonadal development and fertility. Prior studies investigating the transcriptional regulation of Fshr focused on the promoter region and identified a number of crucial elements required for the proper transcription of the gene. However, *in vitro* transfection studies and *in vivo* YAC transgenic mice soon identified that regions at considerable distance from the promoter are needed for correct spatio-temporal regulation. This thought is supported by numerous studies showing that many regulatory elements reside far from the target gene. Comparative genomics coupled with CTCF binding prediction tools identified multiple highly conserved regions both 5’ and 3’ to the Fshr gene that indicate binding sites for the versatile CCCTC-binding factor (CTCF). CTCF is an eleven-zinc finger protein known to play varied roles in genome regulation, including transcription, chromatin insulation, and high order chromatin structure. To identify the role of CTCF in Fshr regulation, granulosa cells harvested from estrogen treated female rats were depleted of CTCF by siRNA transfection, leading to a two-fold increase in Fshr mRNA expression while transcription of the nearby Lhcgr gene was unaffected. These data indicate that CTCF either by itself or in conjunction with other protein complexes might play a role in transcriptional regulation of the Fshr gene.
Introduction

Gonadotropin releasing hormone (GnRH), synthesized by the peptidergic neurons of the hypothalamus, binds to receptors present on the surface of the gonadotroph cells of the anterior pituitary gland, initiating the synthesis and secretion of Lutenizing hormone (LH) and Follicle stimulating hormone (FSH). FSH is a heterogeneous glycoprotein that recognizes and binds to FSHR, a G-protein coupled receptor present on the Sertoli cells of the testis, granulosa cells of the ovary, and osteoclasts of the bone (Richards and Midgley 1976, Heckert and Griswold 1991, Dankbar, Brinkworth et al. 1995, Sun, Peng et al. 2006). Binding to the receptor elicits a number of cellular activities, especially the activation of adenyl cyclase, increased cAMP levels, activation of PKA and phosphorylation of a number of transcriptional activators (Flores, Veldhuis et al. 1990, Tena-Sempere, Manna et al. 1999, Seger, Hanoch et al. 2001). Since FSH acts exclusively through FSHR, greater understanding of its regulation can provide insight on how the cells respond to the hormone, as changes in the receptor level will influence response. Furthermore, absence of FSH or its receptor is known to cause arrest in folliculogenesis and infertility (Huhtaniemi and Themmen 2005). Over the past two decade studies conducted in rat, mouse, porcine, and human have identified a number of regulatory elements and proteins that are crucial for the proper transcription of Fshr (Huhtaniemi, Eskola et al. 1992, Gromoll, Dankbar et al. 1994, Goetz, Lloyd et al. 1996, Sairam and Subbarayan 1997, Simoni, Gromoll et al. 1997, Heckert, Daggett et al. 1998, Kim and Griswold 2001). However, transgenic studies directed by either 5.0kbp or 143bp of rat Fshr promoter sequence were unable to replicate Sertoli and granulosa cell specific expression (Heckert, Sawadogo et al. 2000). Furthermore, transgenic mice carrying yeast artificial chromosomes containing 413-kbp region of the rat Fshr gene and its encompassing regions were found to be insufficient to direct proper spatial and temporal
expression (Hermann, Hornbaker et al. 2007). These data suggest that distal \textit{cis}-acting elements present outside this region were required for proper specific expression.

Evolutionary conserved regions of DNA indicate that they are under pressure to retain functional significance. These highly conserved non-coding regions imply that they might play a role in gene regulation (Bejerano, Pheasant et al. 2004, Shin, Priest et al. 2005, Pennacchio, Ahituv et al. 2006). They are often home to regulatory elements involved in spatial and temporal gene regulation (Maston, Evans et al. 2006). Among these regulatory elements is the CCCTC-binding factor (CTCF), a highly conserved and versatile regulatory factor that plays a varied role in genome regulation. CTCF can act as enhancer blocking, gene activator/repressor, hormone-responsive silencing, long-range chromatin interactions, X-chromosome inactivation, genomic imprinting and in regulating chromatin architecture (Phillips and Corces 2009, Ohlsson, Bartkuhn et al. 2010). Recently microarray data analysis of CTCF depleted mice identified \textit{Fshr} was among the genes that were up regulated (Wan, Pan et al. 2008). The current study explores the possible CTCF binding sites in the highly conserved regions identified by computational analysis and looks at the regulation of \textit{Fshr} in rat granulosa cells depleted of CTCF.
Materials and Methods

Animal Use.
All experiments using animals were approved by the Institutional Animal Care and Use Committee of the University of Kansas Medical Center and performed in accordance with National Institute of Health guide for the care and use of laboratory animals.

Rat granulosa Cell culture
23 day old Immature Sprague Dawley female rats (Harlan, Indianapolis, IN) were injected subcutaneously with 1.5mg 17β-estradiol once daily at 24,25, and 26 days of age. Ovaries were isolated on day 27 and granulosa cells harvested as previously described (Alliston, Maiyar et al. 1997). Briefly, ovaries were harvested, cleaned of fat pads and placed into the M199 collection media containing M199 media (Sigma M2520), 10mM HEPES and 0.2% BSA. The collection media was then slowly aspirated and the ovaries were incubated in 1ml of Sucrose solution media (M199 collection media, 1.8mM EGTA and 0.5M Sucrose) for 15 minutes at 37°C. The ovaries were rinsed three times with M199 collection media and poked gently with a 22-gauge needle to isolate the granulosa cells. The cells were centrifuged, counted and 100,000 cells were plated in each well of a six well plate coated with fibronectin and grown with Dulbecco modified Eagle medium/F12 (10% fetal bovine serum, 1% gentamicin).

siRNA transfection
Various amounts of CTCF siRNA and scrambled siRNA control ( siGenome, rat CTCF, NM_031824, Dharmacon) were transfected into each well with 5μl of Lipofectamine RNAimax (Invitrogen). The transfection media was aspirated 6 hours later and replaced with rat granulosa media and the cells were harvested 48 hours post transfection.
RNA isolation and quantitative real time PCR

Total RNA was isolated using Trizol according to manufactures protocol (Life Technology). Total RNA (100ng) was reverse transcribed using iScript cDNA synthesis kit (Bio-Rad) according to manufactures protocol. 1ul of the cDNA was assayed using the SYBR Green PCR master mix (Applied Biosystems) and primers for rat CTCF (F: TGCCAGTGTAGAAGTCAGCAAATT; R: TGTATGTGTCCCTGCTGGCATA); Fshr (F: GCCAAGACAGCAAGGTGACA; R: GAGCACAAACCTCAGTTCAATGG); Lhcgr (F: GGTCGCCACGCTGACCTA; R: TCTGTTCTTTCTCGCAAATTC) and normalized to the internal control L7 (F: GCTAGGATGGCGAGGAAAGC; R: TGATACCTCGGATTCTGATGACA). Real time was carried out in a 7900HT Sequence Detection real-time analyzer (Applied Biosystems) as described (Gifford, Racicot et al. 2007). Samples and negative control (Hatano, Takayama et al.) were run in triplicate and gene expression was quantified using the delta delta C(T) method in comparison with L7.
Results

*Computational identification of conserved regions and CTCF binding sites*

To identify distal conserved regions, *Fshr* locus and its flanking areas encompassing a region of 1.4 Mb was analyzed for conservation using the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway) and associated software (King, Taylor et al. 2005, King, Taylor et al. 2007). Potential regulatory elements were identified by using human *FSHR* as the base genome and Gallus gallus (chicken) as the most distant. In addition to alignment, 7X regulatory potential was employed to increase predictive measure of regulatory regions. 7X regulatory potential displays potential predictive scores by aligning the genome of human, chimpanzee, macaque, mouse, rat, dog, and cow and comparing the frequencies of short alignment patterns between known regulatory elements and neutral DNA, allowing for the identification of new putative regulatory element(Kolbe, Taylor et al. 2004). This analysis led to the identification of 30 evolutionary conserved regions (ECR) of which seventeen were mapped upstream to *Fshr*, seven to the intergenic region, six downstream (Table 1).

These sequences were then analyzed in the CTCF database CTCFBS 2.0 using the CTCFBS prediction tool (http://insulatordb.uthsc.edu/)(Bao, Zhou et al. 2008, Ziebarth, Bhattacharya et al. 2013). CTCF employs different combinations of its zinc finger proteins to bind to divergent regulatory DNA sequences. Recent studies have identified the core regulatory regions bound by CTCF and these motifs were assigned position weight matrices (PWM). Six such PWM were identified and the CTCFBS prediction tool scores these PWM in the query sequence. Motifs with an overall high positive score indicating favorable match were chosen as probable CTCF binding sites(Ziebarth, Bhattacharya et al. 2013). Of the 30 conserved regions, ECR1, ECR1d, ECR1f,
ECR2, ECR5, and ECR15 had overall positive scores greater than 3, indicating high probability of CTCF binding (Table 2).
Table 1: Genomic location of ECRs in *Rattus norvegicus* (Rat) genome.

<table>
<thead>
<tr>
<th>ECR</th>
<th>Chromosome Location</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECR 1</td>
<td>Chr6: 23036890---23037477</td>
<td>588</td>
</tr>
<tr>
<td>ECR 1c</td>
<td>Chr6: 22585376---22586088</td>
<td>713</td>
</tr>
<tr>
<td>ECR 1d</td>
<td>Chr6: 22597052---22597262</td>
<td>211</td>
</tr>
<tr>
<td>ECR 1f</td>
<td>Chr6: 22614761---22616669</td>
<td>1909</td>
</tr>
<tr>
<td>ECR 1g</td>
<td>Chr6: 22590321---22590502</td>
<td>182</td>
</tr>
<tr>
<td>ECR 2</td>
<td>Chr6: 22626739---22629962</td>
<td>3224</td>
</tr>
<tr>
<td>ECR 3</td>
<td>Chr6: 22596865---22597324</td>
<td>460</td>
</tr>
<tr>
<td>ECR 4</td>
<td>Chr6: 23287153---23287391</td>
<td>239</td>
</tr>
<tr>
<td>ECR 5</td>
<td>Chr6: 23423202---23423600</td>
<td>399</td>
</tr>
<tr>
<td>ECR 6</td>
<td>Chr6: 23219667---23220065</td>
<td>399</td>
</tr>
<tr>
<td>ECR 7</td>
<td>Chr6: 23437816---23438244</td>
<td>429</td>
</tr>
<tr>
<td>ECR 8</td>
<td>Chr6: 22771058---22771652</td>
<td>595</td>
</tr>
<tr>
<td>ECR 9</td>
<td>Chr6: 22776439---22777106</td>
<td>668</td>
</tr>
<tr>
<td>ECR 9b</td>
<td>Chr6: 22785969---22786300</td>
<td>332</td>
</tr>
<tr>
<td>ECR 10</td>
<td>Chr6: 22813283---22813805</td>
<td>523</td>
</tr>
<tr>
<td>ECR 11</td>
<td>Chr6: 22826398---22826718</td>
<td>321</td>
</tr>
<tr>
<td>ECR 12</td>
<td>Chr6: 22826285---22826757</td>
<td>473</td>
</tr>
<tr>
<td>ECR 13</td>
<td>Chr6: 22845802---22846436</td>
<td>635</td>
</tr>
<tr>
<td>ECR 14</td>
<td>Chr6: 23024900---23025427</td>
<td>528</td>
</tr>
<tr>
<td>ECR 15</td>
<td>Chr6: 23048424---23049218</td>
<td>795</td>
</tr>
<tr>
<td>ECR 16</td>
<td>Chr6: 23055601---23056494</td>
<td>894</td>
</tr>
<tr>
<td>ECR 17</td>
<td>Chr6: 23148636---23149418</td>
<td>783</td>
</tr>
<tr>
<td>ECR 18</td>
<td>Chr6: 23153766---23154371</td>
<td>606</td>
</tr>
<tr>
<td>ECR 19</td>
<td>Chr6: 23167660---23168222</td>
<td>563</td>
</tr>
<tr>
<td>ECR 20</td>
<td>Chr6: 23173974---23174670</td>
<td>697</td>
</tr>
<tr>
<td>ECR 21</td>
<td>Chr6: 23180610---23181272</td>
<td>663</td>
</tr>
<tr>
<td>ECR 22</td>
<td>Chr6: 23188053---23188745</td>
<td>693</td>
</tr>
<tr>
<td>ECR 23</td>
<td>Chr6: 23195435---23196040</td>
<td>606</td>
</tr>
<tr>
<td>ECR 24</td>
<td>Chr6: 23212514---23213040</td>
<td>527</td>
</tr>
<tr>
<td>ECR 25</td>
<td>Chr6: 23259302---23260347</td>
<td>1046</td>
</tr>
</tbody>
</table>
Table 2: ECRs predicted to have CTCF binding sites

<table>
<thead>
<tr>
<th>Input Sequence</th>
<th>Motif Sequence</th>
<th>Motif Start Location</th>
<th>Motif Length</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECR 1</td>
<td>CACCATCAGCTGCC</td>
<td>1104</td>
<td>14</td>
<td>7.5</td>
</tr>
<tr>
<td>ECR 1d</td>
<td>AAAACAATAATAGGGATGTG</td>
<td>165</td>
<td>20</td>
<td>22.9</td>
</tr>
<tr>
<td>ECR 1f</td>
<td>TGGCAAGCAGAGGGAGTCT</td>
<td>44</td>
<td>20</td>
<td>13.7</td>
</tr>
<tr>
<td>ECR 2</td>
<td>TTCCATCTGCTGGA</td>
<td>98</td>
<td>14</td>
<td>15.0</td>
</tr>
<tr>
<td>ECR 5</td>
<td>AGCCATCTCCTGGGC</td>
<td>69</td>
<td>14</td>
<td>14.3</td>
</tr>
<tr>
<td>ECR 15</td>
<td>TAGCCACAGGTTGGCATT</td>
<td>139</td>
<td>19</td>
<td>16.0</td>
</tr>
<tr>
<td>ECR 19</td>
<td>GGAACAGGCC</td>
<td>512</td>
<td>9</td>
<td>10.7</td>
</tr>
</tbody>
</table>
Figure 1: Expression of *Fshr* and *Lhcgr* in estrogen treated rat granulosa cells depleted of CTCF. Granulosa cells harvested from estrogen treated immature female rats were transfected with various amount of CTCF siRNA 60nm, 80nm, and 100nm. RNA was isolated 48 hours post transfection, reverse transcribed, and real time was done using primers against CTCF; *Fshr; Lhcgr* and normalized to the internal control L7. CTCF was downregulated by approximately 70% (A), and *Fshr* levels was found to be increased roughly two-fold (B). No change was seen in *Lhcgr* levels (C). * p< 0.05
Knockdown of CTCF in rat granulosa cells increases transcription of Fshr

To test if CTCF plays a role in regulation of Fshr, rat granulosa cells were transiently transfected with CTCF siRNA or control siRNA and qPCR was performed using primers specific for each gene (Figure 1). CTCF was downregulated by approximately 70% and the results showed that granulosa cells in which CTCF was knocked down, expression of Fshr increased two-fold, while the levels of Lhcgr remained unchanged. This data indicate that CTCF can modulate the level of Fshr mRNA.

Discussion

Following the initial characterization of the Fshr gene, a number of studies have identified regulatory elements at the promoter and its proximal regions (Heckert, Sawadogo et al. 2000, Heckert and Griswold 2002). These studies while highlighting important factors required for proper transcription of the Fshr gene, were unable to explain the mechanism for the genes cell specific expression (Heckert, Sawadogo et al. 2000, Nordhoff, Gromoll et al. 2003). In a bid to identify regions required for the cell specific expression, efforts were made encompassing the Fshr promoter and its surrounding regions, this body of work summarized that Fshr belonged to the class of genes that employed distal regions for transcriptional control. Previous studies from our laboratory identified seven such distal ECR regions, which were shown to possess transcriptional activity in vitro (Hermann, Hornbaker et al. 2007). With the advent of the UCSC genome browser and inclusion of genome sequences from a number of species, ranging from human to chicken, we were able to identify thirty ECRs, which were highly conserved indicating that these sequences likely harbor regulatory elements. Among the distal regulatory elements identified to play a role in distal regulation is the ubiquitous, multivalent protein CTCF that binds
to multiple DNA sequences by using different combinations of the twelve zinc finger proteins (Filippova, Fagerlie et al. 1996). Earlier work in CTCF depleted transgenic mouse using microarrays a number of misregulated genes, among which $Fshr$ was found to be upregulated (Wan, Pan et al. 2008). Building on this finding and identification of highly conserved regions, we utilized the CTCF binding site prediction tool. Among the 30 ECR regions tested, seven were found to have overall high positive scores indicative of probable CTCF binding. Among the ECR regions predicted to be bound by CTCF, ECR1d, ECR1f, and ECR2 are located 3’ to the $Fshr$ gene, while ECR1, ECR5, ECR15 and ECR 19 are 5’ to the gene. The intergenic location of these predicted binding sites corresponds to a study done by Kim et al, who mapped global distribution of CTCF and reported 46% of CTCF binding sites to be intergenic, similar to where our predicted CTCF bound ECR regions are located (Kim, Abdullaev et al. 2007, Jothi, Cuddapah et al. 2008).

To analyze if CTCF does play a role in regulating $Fshr$, CTCF was knocked down in rat granulosa cells leading to relief of inhibition and increase of $Fshr$ transcription by two-fold. This data indicates that CTCF binds within the $Fshr$ locus, possibly at the predicted regions, inhibiting transcription, thus acting as a repressor. CTCF is a well-known repressor of gene transcription as documented in the chicken-$myc$ gene and $hTERT$ transcription (Filippova, Fagerlie et al. 1996, Renaud, Loukinov et al. 2005). It is possible that in vivo, CTCF activity keeps $Fshr$ transcription level in check, optimizing binding of FSH to the receptor thus limiting ovarian response to gonadotropin stimulation. CTCF could mediate its repressor activity by homodimerizing as shown in Figure 2, isolating enhancers, and bound activators, physically blocking the interaction between these activator elements and the $Fshr$ promoter. CTCF plays a major role in transcriptional activity, silencing, insulation and imprinting control (Phillips and Corces 2009,
Concerning $Fshr$, CTCF possibly keeps $Fshr$ regulation in check and thus indirectly regulating the delicate balance of folliculogenesis and oocyte development.

In summary, computational analysis identified putative CTCF binding sites within highly conserved regions. Knockdown of CTCF relieved repressor activity on $Fshr$ in granulosa cells, thus establishing that CTCF could regulate transcription of $Fshr$. Future experiments designed to identify CTCF binding regions and the chromatin architecture both in the presence and absence of CTCF will reveal the role played by this protein in regulating transcription of $Fshr$. 
Figure 2: Possible mechanism of CTCF action on Fshr locus: CTCF (green ovals) homodimerize leading to chromatin looping and altering its conformation. This altered conformation inhibits enhancers (blue ovals) from interacting with $Fshr$ (red solid line), leading to repressed transcriptional activity.
Part 2: Steroidogenic Factor 1
Chapter 3: Introduction
Identification of Steroidogenic Factor -1

In the early 1990’s, analysis of the 5’ flanking region of genes encoding steroid hydroxylases, identified a common AGGTCA DNA recognition motif that interacted with the same DNA-binding protein. This protein was designated as steroidogenic factor-1 (SF-1) or adrenal 4-binding protein (Ad4BP) (Rice, Mouw et al. 1991, Morohashi, Honda et al. 1992). The unique expression profile of SF-1 in the adrenal gland, gonad and the pituitary, and its regulation of distinct genes encoding steroid hydroxylases provided the first clue to its important role in directing cellular expression of steroidogenic enzymes (Rice, Mouw et al. 1991, Morohashi, Honda et al. 1992, Hatano, Takayama et al. 1994, Morohashi and Omura 1996, Sadovsky and Crawford 1998). Cloning of SF-1 cDNA from mouse and bovine adrenals revealed that it belongs to the nuclear hormone receptor family and shared homology to the Drosophila nuclear receptor fushi tarazu factor 1 (FTz-F1), hence, the gene encoding SF-1 was originally designated but is now referred to as Nr5A1, in compliance with Nuclear Receptors Nomenclature Committee of 1999 (Scott and Weiner 1984, Lala, Rice et al. 1992, Morohashi, Honda et al. 1992, Honda, Morohashi et al. 1993, 1999).

Structure of SF-1

Like other nuclear hormone receptors, SF-1 harbors a DNA binding domain (DBD), a hinge region, a ligand-binding domain, and activation function 2 (AF-2) sequence activation domain (Figure 1). The DNA binding domain is the most conserved part of SF-1 and is comprised of two zinc-chelating modules, which coordinates binding to its DNA response element. The hinge region plays a role in homo-or heterodimerization of SF-1, and is a docking site for miscellaneous cofactors that affect transcriptional activity of SF-1 (Tan, Hall et al. 2002). The C-
terminal ligand-binding domain is composed of a ligand binding pocket, dimerization site, activation function 2 (AF-2) sequence and co-factor binding site. This highly conserved site mediates dimerization with other receptors and contains ligand-induced activation as well as ligand-reversed transcriptional silencing domains (Sadovsky and Crawford 1998). Unlike classic members of the nuclear receptor family, SF-1 lacks a ligand-independent activation function 1 sequence (AF-1) and depends on activation of AF-2 sequence for full transcriptional activity. Initially thought to function in absence of an exogenous ligand, X-ray crystallographic studies on bacterially expressed SF-1 has now identified phospholipids as ligands for SF-1 (Krylova, Sablin et al. 2005, Li, Choi et al. 2005, Wang, Zhang et al. 2005). Mass spectrometry studies on SF1 immunoprecipitated from cAMP-stimulated H295R adrenocortical cells identified endogenous ligands, sphingosine (SPH) and lyso-sphingomyelin (lysoSM) which bound to the receptor. These ligands bind to SF-1 under basal conditions with decreased binding of sphingolipids to the receptor under cAMP treatment (Urs, Dammer et al. 2006). Hence, SF-1 follows the basic structure of a classic nuclear receptor with the exception of the N-terminal AF-1 domain and is a constitutively active lipid binding protein.
**Figure 1:** Structure of classic nuclear receptor and SF-1. The classic nuclear receptor has two activation function (AF) domains, a DNA-binding domain (DBD), a hinge region and a ligand-binding domain (LBD). In contrast, SF-1 lacks a functional AF-1 and transcriptional function rests on the conserved AF-2 (black box) located on the C-terminal. SF-1 has a DBD containing two Zn fingers (gray boxes) followed by a Hinge region and a ligand-binding domain.
SF-1 Expression

In developing embryos, *in situ* hybridization studies identified SF-1 expression as early as embryonic day 9.5 (E9.5) in the urogenital ridge, corresponding to the adrenogonadal primordium. During this time, the developing testes and the ovary are histologically indistinguishable and thus SF-1 represents the earliest marker of adrenal and gonadal differentiation. These groups of cells later give rise to two distinct populations of SF-1 expressing cells; a group of cells adjacent to the dorsal aorta representing the adrenocortical precursors and cells adjacent to the coelomic epithelium that give rise to the bipotential gonad (Hatano, Takayama et al. 1994).

SF-1 expression is first detected at E10-10.5 in the adrenal primordium preceding P450scc, which was not detected until E11, indicating that SF-1 is pivotal for the expression of steroid hydroxylases. The chromaffin cell precursors migrate to the adrenal primordium at E12.5-E13.5 and SF-1 expression is limited to the outer cortical cells wherein it is expressed throughout gestation and postnatal life. SF-1 expression in the urogenital ridge at E9 precedes the onset of Sry expression during which the testes and the ovaries are indistinguishable and are termed as bipotential gonads (Parker and Schimmer 1997). Expression of Sry triggers testes specific gene to initiate differentiation of the testis. SF-1 expression persists in the testes, in both the interstitial region where steroidogenic Leydig cells reside and in the testicular cords, which contain fetal Sertoli cells and primordial germ cells. In contrast, SF-1 expression diminishes in the ovaries, both in transcript and protein levels, and does not resume until just before birth, when it is first detected in the theca and granulosa cells, indicating that decrease in SF-1 expression facilitates ovarian development (Hatano, Takayama et al. 1994, Takayama, Sasano et al. 1995).
SF-1 transcripts were also detected in the fetal spleen, anterior pituitary gland, and hypothalamus (Hatano, Takayama et al. 1994, Ikeda, Shen et al. 1994, Hatano, Takakusu et al. 1996, Hanley, Ball et al. 1999, Morohashi, Tsuboi-Asai et al. 1999, Hanley, Rainey et al. 2001). In adults, SF-1 is highly expressed in major steroidogenic cells; the three zones of adrenal cortex, testicular Leydig cells, and ovarian theca cells and granulosa cells and luteal cells (Honda, Morohashi et al. 1993, Ikeda, Lala et al. 1993, Morohashi, Iida et al. 1994, Ramayya, Zhou et al. 1997, Morohashi 1999). It is also expressed in pituitary gonadotrophs (Barnhart and Mellon 1994), ventromedial hypothalamic neurons, a subset of hippocampal neurons that co-express steroidogenic acute regulatory protein and aromatase, and the endothelial linings of the venous sinuses and pulp veins of the spleen (Ramayya, Zhou et al. 1997, Morohashi, Tsuboi-Asai et al. 1999). Low levels of SF-1 transcripts were also detected in the placenta (Morohashi, Hatano et al. 1995). Taken together these findings indicate that SF-1 plays a major role in all components of the hypothalamus-pituitary-adrenal (HPA) and gonadal (HPG) axis.

Function of SF-1

Global knockout of SF-1

The expression pattern of SF-1 suggests that it acts at multiple levels of the HPG axis, and is required for adrenal and gonadal steroidogenesis. To understand the in-vivo role of SF-1, three separate laboratories utilized gene disruption models in embryonic stem cells to generate SF-1 knockout mice, all of which showed similar findings (Luo, Ikeda et al. 1995, Luo, Ikeda et al. 1995, Sadovsky, Crawford et al. 1995, Shinoda, Lei et al. 1995). Homozygous SF-1 knockout mice were born at expected Mendelian frequency of 1:4, establishing that SF-1 is not essential for prenatal survival. SF-1 null mice exhibited male to female sex reversal and adrenocortical
insufficiency due to complete absence of the adrenal glands and the gonads. Embryonic analysis of SF-1 null mice revealed an initial development of adrenogonadal primordia followed by apoptosis resulting in complete loss of adrenal glands and gonads. SF-1 is a known regulator of the Mullerian inhibiting substance (MIS), which is crucial for the regression of the Mullerian duct. These SF-1 null mice also lack testosterone resulting in regression of the Wolfian duct. The combination of these causes male-to-female sex reversal and these SF-1 knockout mice are born female irrespective of genetic sex. Interestingly, the gonadotropes of these SF-1 null mice showed impaired expression of a number of genes including LH-β, FSH-β, αGSU, and GnRHR, which are essential for reproduction. One of the surprising features was these knockout mice lacked the Ventral Medial Hypothalamus (VMH), a homeostatic relay center linked to metabolic and female reproductive behavior. Finally, these knockout mice also had defects in their splenic parenchyma. (Ingraham, Lala et al. 1994, Luo, Ikeda et al. 1994, Sadovsky, Crawford et al. 1995, Shinoda, Lei et al. 1995). These global knockout mice died shortly after birth owing to adrenocortical insufficiency since they were rescued by administration of exogenous steroids (Ikeda, Luo et al. 1995).

Tissue-specific knockout of SF-1

Combined with the expression pattern and global knockout of SF-1, mice further underlined the role of SF-1 in adrenal and gonadal development and function. However, these SF-1 knockout mice died soon after birth complicating the effort to identify the role of SF-1 at specific sites. To define the role of SF-1 at specific tissues, Cre/LoxP system was used to produce tissue specific knockouts. Pituitary specific deletion of SF-1 mice was carried out by cre recombinase expression directed to the anterior pituitary gland by the 5’ flanking sequences of the α-subunit of glycoprotein hormones. These αGSU-Cre/loxP mice lacked SF-1 in the anterior pituitary but
displayed normal levels in other tissues. These mice displayed diminished levels of pituitary gonadotropins and exhibited severe gonadal hypoplasia. In males, testis showed some differentiation, however, the germ cells were severely decreased in number, and lacked mature sperm. The Leydig cells were low in number and devoid of steriodogenesis. In females, ovaries develop follicles through the antral stage but do produce preovulatory follicles and corpora lutea (Zhao, Bakke et al. 2001). The females had hypoplastic uteri, indicating a severe impairment on sex steroid production in pituitary specific SF-1 knockout mice. Administration of exogenous pregnant mares serum gonadotropins (PMSG) to these pituitary specific knockout mice induced sperm maturation and increased proliferation of Leydig cells. Female mice showed PMSG induced follicular maturation and formation of corpora lutea, establishing that gonads of these mice were functional. These findings indicate that SF-1 plays an important role in regulation of a number of genes involved in the HPG axis and is essential for normal pituitary gonadotrope function (Parker and Schimmer 1997, Schimmer and White 2010).

Anti-Müllerian hormone type 2 receptor-Cre (Amhr2-Cre) driven recombinase transgene was used to generate mice with SF-1 deletion specifically in the testicular Leydig cells and ovarian granulosa cells. These SF-1 adult male mice had hypoplastic testes and the testes failed to descend (Jeyasuria, Ikeda et al. 2004). These mice also had decreased expression of cholesterol side-chain cleavage enzyme (Cyp11a) and absence of steroidogenic acute regulatory (StAR) protein, which form the essential components of testosterone biosynthesis. Female mice with Amhr2 Cre specific deletion of SF-1 were infertile, with ovaries lacking corpora lutea and hemorrhagic cysts. Granulosa cell-specific SF-1 knockout mice had reduced estrogen levels, due to decreased expression of Cyp19a1 or reduced follicle numbers, thus providing conclusive
evidence that SF-1 was required for normal reproductive function (Jeyasuria, Ikeda et al. 2004, Pelusi, Ikeda et al. 2008).

Ablation of SF-1 from the central nervous system (CNS) was directed by using a nestin-Cre transgene. These mice displayed normal development and function of the adrenal glands, pituitary, and gonads but the VMH nuclei were disrupted. In multiple behavioral tests, these CNS-specific knockout mice, both males and females, displayed decreased locomotor activity, exhibited heightened anxiety and later onset of obesity (Majdic, Young et al. 2002, Zhao, Kim et al. 2008). CNS-specific SF-1 knockout female mice showed impaired follicular maturation and decrease in lordosis, ovulation, and fertility, thus highlighting the crucial role of SF-1 in female reproduction (Kim, Li et al. 2010).

**SF-1 gene and transcription**

The *Nr5a1/NR5A1* is located on mouse chromosome 2 and human chromosome 9. The gene is flanked on its 5’ side by *NR6A1*, which codes for germ cell nuclear factor, a member of the nuclear receptor family and 3’ side by *GPR144* that encodes for probable G Protein-Coupled Receptor 144 (Figure 2). In addition to SF-1, the genomic sequence of *Nr5A1* also encodes for embryonal long-terminal repeat-binding protein, ELP 1 and its isoforms ELP2 and ELP3 through alternate promoter usage and splicing (Figure 1). Among these four transcripts, SF-1 has emerged as a key regulator of endocrine homeostasis. Function analysis of ELP transcripts have revealed ELP1 acts as a repressor while ELP2 and ELP3 function as transcriptional activators (Ninomiya, Okada et al. 1995).

While there is no evidence of a TATA box in the first 110bp in the 5’-flanking region of NR5A1, a number of other regulatory elements such as SRY (sex determining region Y)-box (SOX)
binding site, an E box, a CCAAT box and Sp1/Sp3 site have been identified (Schimmer and White 2010). In addition, two other Sp1/Sp3 sites are located between the +10 and +30 region upstream of the transcriptional start site (Figure 3) (Scherrer, Rice et al. 2002). Numerous other transcription factors that regulate SF-1 transcription such as GATA-4, WT1, Lhx9, SOX15, SOX30, TEAD-4 and CBX2, have been identified (Schimmer and White 2010). To examine transcriptional requirements for \textit{Nr5a1} in vivo, short promoter fragments were introduced into transgenic mice. Earlier work identified a 90bp proximal promoter fragment reporter construct that was sufficient to direct SF-1 expression in Y1 adrenocortical cells (Woodson, Crawford et al. 1997).
Figure 2: Schematic organization of the mouse Nr5a1 gene. Located on chromosome 2, Nr5a1 is flanked by Nr6a1, encoding germ cell nuclear factor and Gpr144 probable G-protein coupled receptor 144. SF-1 is encoded by seven exons as depicted in B, and codes for 462 amino acids and bears 94% homology to the human SF-1. In addition to SF-1, Nr5a1 also gives rise to three other mRNA transcripts named named embryonal long-terminal repeat binding proteins (ELP1-3). These ELP proteins are encoded by exons as depicted in the figure. ELP3 bears a very close resemblance to SF-1 except for the starting exon. In mouse, ELP1 is expressed in the gonads and in Y1 mouse adrenocortical cell line. ELP2 is not expressed in any mouse tissue, but similar to ELP1 is expressed in Y1 cell line. ELP3 bears an expression profile very similar to SF-1 and is expressed in the pituitary, adrenal, ovary, testis, and spleen. ELP3 expression has also been found in the cerebrum, heart, kidney, spleen and Y-1. The potential biological role of these isoforms are unknown although in vitro studies have shown that ELP1 acts as a repressor while ELP2 and ELP3 function as transcriptional activators. Adapted from (Parker, 1997)
Figure 3: Schematic representation of SF-1 and its proximal promoter. The transcription start site (TSS) is depicted at base pair 0 and indicated by the bend arrow. The box regions represent conserved elements and their location from the TSS is depicted by the base pairs. Transcription factors that bind to these conserved regions are in ovals. Different SF-1 expressing tissues utilize multiple combinations of transcription factors and include many more proteins than depicted here. Sx is the binding site for Sox which is required for SF-1 expression in the Sertoli cells of the testes; The E-box is bound by the upstream stimulatory factors 1 and 2; C represents a CCAAT box that binds to nuclear factor Y (NFY) and Sp represents binding site for transcription factors Sp1 and Sp3. Adapted from (Schimmer and White 2010).
To analyze regions required for proper SF-1 expression, bacterial artificial chromosome (BAC) transgenic mouse were generated with the enhanced green fluorescent protein reporter gene (eGFP) under the control of 50kb of the Nr5a1 locus encoding for SF-1. Analysis of transgenic SF-1/eGFP mouse embryos indicated that this transgene was able to replicate the developmental profile of SF-1 in the gonads, adrenal cortex and VMH. In adult transgenic mouse, this fragment was able to replicate SF-1 expression in the adrenal cortex, testes, ovaries, and the VMH. However, expression was not seen in the corpora lutea or the anterior pituitary gland indicating that regulatory regions beyond those expressed in the construct were required (Stallings, Hanley et al. 2002). A somewhat larger fragment that encompassed the -590 to +85 region cloned in front of the Lac Z gene was able to recapitulate expression of SF-1 only in the gonads (Wilhelm and Englert 2002). Transgenic mouse that expressed mouse genomic DNA extending from exon2 of Nr5a1 into the upstream Nr6a1 was created. This transgene showed SF-1 expression in the adrenal cortex, testicular Leydig cells, ovarian theca cells, the ventromedial hypothalamus and spleen, but not in the pituitary gland or corpus luteum (Stallings, Hanley et al. 2002). To recapitulate SF-1 expression, transgenic mouse expressing a much larger 500-kb yeast artificial chromosome (YAC) that contains the entire rat Nr5a1 locus plus 5’ and 3’ sequence extending into the Nr6a1 and the Psmb7 genes was developed. These mice were able to duplicate endogenous expression of SF-1 both spatially and quantitatively indicating that the transgene contained all necessary sequences for proper SF-1 expression (Karpova, Presley et al. 2005).

In addition to the identified regulatory sequences and transcription factors, Nr5a1 expression is regulated by epigenetic modifications. In particularly DNA methylation appears to act as another layer of regulation. In endometriotic tissues, the CPG islands in the Nr5a1 promoter were found to be hypomethylated when compared to normal endometrial tissues leading to aberrant
expression of a number of steroidogenic genes including *Nr5a1* (Xue, Lin et al. 2007). When normal endometrial cells were treated with 5-aza-deoxycytidine, a demethylating reagent, the *Nr5a1* promoter was demethylated leading to transcription of *Nr5a1*, whereas methylation led to loss of activity (Hoivik, Aumo et al. 2008). The correlation between methylation of the *Nr5a1* promoter and activity is also seen in steroid-secreting cell lines and normal steroid-secreting tissues. The *Nr5a1* proximal promoter is unmethylated in developing testis and ovary and hypermethylated in tissues that do not express for SF-1 (Mohn and Schubeler 2009).
Chapter 4: Analysis of Fast, Ftz-F1-associated transcript, suggests functional cooperation with the nuclear receptor Steroidogenic Factor 1
Abstract

An evolutionarily conserved expressed sequence tag lacking an identifiable open reading frame was identified and named *Ftz-F1 associated transcript* (*Fast*) for its proximity and functional associations with *Ftz-f1* (*Nr5a1*), the gene encoding steroidogenic factor-1 (SF-1). SF-1 is a nuclear receptor and a key determinant and regulator of the adrenal and reproductive axes. Mammals with SF-1 mutations display a range of phenotypes, including absence of adrenal glands and gonads, diminished pituitary gonadotropins, disruption of the ventromedial hypothalamus, ovarian failure, and XY-sex reversal. Directional RT-PCR demonstrated *Fast* is transcribed in opposite direction to that of *Ftz-F1* and 5’ and 3’ RACE showed *Fast* is composed of three exons, polyadenylated, and derived from multiple transcriptional start sites that border and extend into exon 1g of *Ftz-F1*. In addition, sequence analysis revealed two possible spliced variants that lacked part of exon 3. RT-PCR, using RNA from multiple mouse tissues and cell lines, revealed *Fast* and *Ftz-f1* share the same expression profile. Furthermore, *Fast* transcript was located predominantly in cytoplasm, when analyzed in MA-10 Leydig cell lines. In addition, in P-19 embryonal-carcinoma cells, SF-1 and *Fast* transcript levels were similarly attenuated in response to retinoic acid, while in granulosa cells, both were induced by *in vivo* treatment with pregnant mare's serum gonadotropin. To evaluate function, *Fast* knockdown and overexpression paradigms were used. Unexpectedly, neither strategy caused a change in SF-1 protein or mRNA levels. RNA structural analysis and preliminary northern blot suggests *Fast* gives rise to a small RNA and could possibly regulate other genes. Thus, *Fast* is a long non-coding RNA that is regulated like *Ftz-F1* with respect to its tissue profile and hormone response, suggesting it acts within the same biological pathway and potentially, in conjunction with SF-1, regulate development and function of adrenal glands and gonads.
Introduction

Steroidogenic factor 1 (SF-1, Ad4BP, NR5A1) is a nuclear hormone receptor and encoded by one of four known transcripts derived from *Ftz-F1* gene, officially known as *Nr5a1*. Of the four *Ftz-F1* gene products, SF-1 has received the most interest due to its critical roles in endocrine homeostasis and organ development (Luo, Ikeda et al. 1995, Ikeda 1996, Parker and Schimmer 1997). SF-1, first identified as a protein that bound a common regulatory motif (PyCAAGGTCA) within several genes associated with steroid biosynthesis, has emerged as a major regulator of the endocrine system (Lala, Rice et al. 1992, Morohashi, Honda et al. 1992, Schimmer and White 2010, Gardiner, Shima et al. 2012). Expression of SF-1 is limited to a discrete set of cells that are functionally linked by their role in the endocrine system. More specifically, SF-1 is produced in supporting cells of the testis (Sertoli) and ovary (granulosa cells), steroidogenic cells of the adrenal gland (adrenocortical), testis (Leydig), and ovary (theca), gonadotrope cells of the pituitary, and cells within the ventromedial hypothalamus (Honda, Morohashi et al. 1993, Ikeda, Lala et al. 1993, Barnhart and Mellon 1994, Ingraham, Lala et al. 1994, Morohashi, Iida et al. 1994, Morohashi 1999, Ngan, Cheng et al. 1999). During development, SF-1 is first observed on embryonic day 9 (e9) in a single population of cells in the urogenital ridge that subsequently resolve into two discrete populations that give rise to adrenocortical cells and gonadal cells (Ikeda, Shen et al. 1994, Morohashi, Hatano et al. 1995, Hanley, Ball et al. 1999, Hanley, Rainey et al. 2001). By e12.5, SF-1 transcripts are limited to the gonad, the adrenal primordium, and the diencephalon, as development continues, SF-1 expression in the gonads becomes sexually-dimorphic with higher levels observed in the developing testis (Ikeda, Shen et al. 1994, Morohashi, Hatano et al. 1995, Hanley, Ball et al. 1999). While SF-1’s expression profile implicated it in transcriptional roles unrelated to
steroidogenesis, it was not until mice lacking \textit{Ftz-F1} were generated that it was recognized as a mediator of organ development. Four different SF-1 knockout models were created and the results from each demonstrated the requirement for SF-1 in development of the adrenal glands and gonads (Ingraham, Lala et al. 1994, Luo, Ikeda et al. 1994, Sadovsky, Crawford et al. 1995). Accordingly, SF-1 deficiency resulted in male-to-female sex reversal and early postnatal death, due to the loss of testicular and adrenal steroids, respectively. Further evaluation of null embryos indicated that organ development initiated but regressed soon after mesenchymal thickening of the genital ridge (Luo, Ikeda et al. 1994).

Considering the critical role of SF-1 in development and function of the endocrine system, there has been much enthusiasm to elucidate the regulatory mechanisms that specify SF-1 transcription in the appropriate temporal and spatial manner. While DNA sequences comprising the SF-1 promoter region have been thoroughly evaluated, transgenic analysis of the promoter region in mice, revealed it was insufficient to establish the tissue-, cell- and developmental specificity of SF-1 expression (Nomura, Bartsch et al. 1995, Woodson, Crawford et al. 1997, Daggett, Rice et al. 2000, Wilhelm and Englert 2002). Other studies, using increasingly larger amounts of DNA as transgenes, expanded the location of regulatory sequences required for transcriptional competency to nearly 150kb spanning the \textit{Ftz-f1} locus (Stallings, Hanley et al. 2002, Karpova, Presley et al. 2005). Results from these studies revealed a mechanism of SF-1 regulation that emphasizes the role of distal regulatory sequences in cell-specific expression.

In an effort to resolve the location of \textit{cis}-acting regulatory elements that control SF-1 expression, comparative genomics was used to identify highly conserved sequences between the mouse and human loci that serve as candidate regulatory elements. While several conserved, non-coding sequences were identified in this analysis and are reported elsewhere, a conserved EST located
between *Ftz-f1* and *Gcnf* was uncovered (Karpova, Presley et al. 2005). The EST was 1377 bp long, encoded by three exons and was transcribed in the opposite orientation from *Ftz-f1*. Expression profile of *Fast* was nearly identical to SF-1 and lacked open reading frames, indicating it was a long noncoding RNA (lncRNA). This study further characterizes the non-coding RNA and provides preliminary evidence that it is a precursor to a small RNA of unknown function.

**Results**

**Comparative genomics revealed two conserved ESTs upstream of *Ftz-f1*.** To identify conserved, non-coding sequences in proximity to *Ftz-F1*, genomic sequences of human and mouse *Ftz-F1* were compared using pair-wise blast analysis. The identified conserved sequences were next compared to sequences in the expressed sequence tag databases, revealing that two of the conserved sequences corresponded to novel transcripts cataloged in a RIKEN testis cDNA library (Fig. 1, Accession #AK017050 and AK007201). BLAST analysis of AK007201 confirmed its genomic location approximately 10kb 5’ to *Ftz-F1* exon 1a and revealed its identity with another EST (Accession #AF390897S4), which indicated it was expressed from *Nr6a1*, the 5’ proximal gene to *Ftz-F1* that encodes Germ-cell nuclear factor, Gcnf (Fig. 1). Alignment of AK017050 to mouse *Ftz-F1* revealed two exons. The larger exon (790bp) contained the identified conserved sequence, which localized 5.3kb 5’ to *Ftz-F1*. The smaller exon (125bp) localized just 5’ to *Ftz-F1* exon 1g (Fig. 1). The genomic location and transcript processing revealed EST AK017050 as a novel transcript derived from the *Ftz-F1* locus and it was named *Ftz-F1-associated-transcript* or *Fast*. 
**Fast transcriptional orientation opposes that of Ftz-f1.** The novelty of Fast and its association with SF-1 prompted further investigation to determine its transcript characteristics and expression profile. Transcriptional orientation of Fast was determined by directional RT-PCR, using cDNA generated with oligo-dT or primers of different orientations corresponding to sequences within the two identified Fast exons (Fig.2A). Following PCR of cDNA generated from mouse testis RNA, Fast products were visualized by Southern blot analysis, using an internally located probe (Fig. 2B). Two distinct amplified products were observed, when template cDNA was synthesized using oligo-dT or primer 2, but not primer 1, indicating Fast is transcribed in opposite orientation to Ftz-F1, proceeding from the shorter exon, closest to Ftz-F1 exon 1g, towards the larger exon, located midway between Gcnf and Ftz-F1 (Fig. 2 A&B). Fast expression and transcriptional orientation were also examined from human testis RNA, using primers directed to human sequence syntenic to the identified mouse exons. No amplified product for human FAST was observed using primers located in the two putative exons (primers 1 & 2, data not shown). However, amplification of human FAST was observed using primers (primers 1 & 3) located within the single conserved exon and cDNA synthesized with either oligo-dT or primer 1 (Fig. 2 C&D). Taken together, these data indicate that Fast is an evolutionarily conserved, poly-adenylated transcript that arises from syntenic regions of mouse chromosome 2 and human chromosome 9 in the opposite orientation to that of Ftz-F1 (Brian Hermann- Regulation of FSH-receptor and SF-1: transcriptional control in reproduction). Furthermore, the studies revealed that mature mouse Fast (mFast) is produced through splicing of at least two exons, while the complete exon structure of human FAST is currently unresolved.

**Fast is a noncoding RNA consisting of multiple transcript variants.** Further sequence analysis performed on PCR-amplified Fast products confirmed mFast as a poly-adenylated,
spliced transcript, and revealed a third exon and several potential splice variants within its transcript pool (Fig. 3). 5’RACE was used to map the transcriptional start site for mFast and revealed several distinct 5’ ends, three of which overlapped Ftz-f1 exon 1g (Fig. 3). Similarly, 3'RACE was used to identify the 3’ end of mFast transcripts and revealed the presence a non-canonical poly-A signal sequence (CATAAA) that likely directs transcript poly-adenylation. In addition to identification of a third mFast exon, sequence analysis revealed two amplified products lacking portions of exon 3. Variant 1 represents loss of 182bp from exon 3, while variant 2 lacks 71bp. While it is currently unknown if these represent true splice variants or amplification artifacts caused by strong secondary structure, it is interesting to note that the deleted sequences reside within the regions of greatest conservation. Regardless, the studies demonstrate the existence of multiple mFast transcript variants that result from the use of multiple transcriptional starts sites and at least one splice variant based on inclusion of the small exon 2 (Fig. 3). Similar experiments showed the existence of human FAST and revealed that it contains only the large exon homologous to mouse exon 3 and has a non-canonical poly-A signal similar to mouse Fast (Brian Hermann- Regulation of FSH-receptor and SF-1: transcriptional control in reproduction).

**Fast is predominantly cytoplasmic.** To delineate the function of the lncRNA, its sub-cellular location was determined (Fig. 4). RNA from mouse MA-10 cells was isolated from cytoplasmic and nuclear fractions, which were assayed by RT-PCR for Fast and Gapdh as control. RT-PCR was done subsequently using the gene-specific primers for the noncoding RNA (Table 1). PCR products of Fast were obtained predominantly in the cytoplasmic fraction, with a no-RT control to eliminate any false amplification that might arise due to DNA contamination.
**Fast can produce a 7kd protein.** To determine if Fast is capable of producing smaller peptides, the RNA was used in an *in vitro* translation system and labeled proteins visualized by $^{35}$S methionine (Fig.5). A ~7 kd protein was detected in pcDNA3 Fast construct, indicating that a small protein maybe translated under *in vitro* conditions. The control plasmid that harbors the luciferase and DMRT1 was transcribed in full length *in vitro*. Inspection of cDNA sequences for both human and mouse Fast failed to identify any open reading frames (ORFs) coding for proteins larger than 260 or 194 nucleotides, respectively. Furthermore, none of these small ORFs were conserved between mouse and human Fast, nor did they share homology with any known protein. This characteristic is consistent with transcripts designated as noncoding RNAs (ncRNAs), in which the average length of the longest ORF is approximately 200 nucleotides (Numata, Kanai et al. 2003). Thus, the structural studies of Fast reveals it to be a conserved, polyadenylated, spliced transcript that does not code for a functional protein, suggesting it is a member of a growing class of ncRNAs that contribute to or are implicated in numerous biological processes.

**Fast is co-expressed with SF-1.** Fast expression was examined by RT-PCR of RNA isolated from adult mouse tissues and cell lines. Interestingly, the tissue expression pattern for Fast was closely identical to that of SF-1, with both transcripts detected in adrenal, testis, pituitary, and ovary (Fig. 6). Neither Fast nor SF-1 was consistently detected in mouse spleen, in contrast to previous reports of SF-1 in this tissue (Morohashi 1999, Kimura, Yoshii et al. 2000, Karpova, Presley et al. 2005). The conflict may reflect differences in the methodology used to detect Ftz-F1 transcripts. To determine specific Fast-expressing cell types, mRNA *in situ* hybridization was performed. However, high background hybridization precluded identification of cellular expression patterns, using this method. To help circumvent this obstacle an alternative approach
was employed that examined *Fast* in SF-1-positive cell lines using RT-PCR. The results showed *Fast* and SF-1 expression in many of the same tissues and cell line, including MA10 Leydig cells (Fig. 6) (Brian Hermann- Regulation of FSH-receptor and SF-1: transcriptional control in reproduction). In summary, the expression data show identical expression patterns for *Fast* and SF-1, at both the tissue and cellular level. The similar expression patterns of SF-1 and *Fast* suggested they share similar regulatory mechanisms.

To further explore this avenue, the response of SF-1 and *Fast* to retinoic acid (RA) was evaluated in P-19 embryonal carcinoma cells. Previous studies have shown time-dependent declines in SF-1 expression in P-19 cells exposed to RA (Barnea and Bergman 2000, Gu, Goodwin et al. 2005). RNA samples from P-19 treated cells obtained from Dr. Austin Cooney were evaluated for SF-1 and *Fast* expression by semi-quantitative PCR. Results from the analysis revealed that both SF-1 and *Fast* transcripts decreased in response to RA, with notable decreases in transcript levels after 36 hours of treatment and nearly undetectable levels by 72 hours (Fig. 7). No change was observed in the L7 mRNA control (Brian Hermann- Regulation of FSH-receptor and SF-1: transcriptional control in reproduction).

In addition, hormonal response of *Fast* in mouse granulosa cells was examined. Expression analysis of *Fast* in RNA samples isolated from hormone-treated granulosa cells (graciously provided by Dr. Lane K. Christenson) was evaluated. This revealed an increase in *Fast* transcripts 48 hours after PMSG stimulation (Fig. 8). In summary the expression studies show *Fast* expression is hormonally regulated, which together with its restricted tissue profile, strongly support its role as a functional ncRNA. Its striking similarities to SF-1 suggest *Fast* and SF-1 share an important functional relationship and act within the same biological pathways.
**Fast does not regulate SF-1 mRNA stability.** Because shared expression and regulatory profiles of SF-1 and Fast, suggested a functional relationship, the effect of Fast knockdown on SF-1 mRNA was examined. short hairpin RNAs (shRNAs) were introduced in MA10 Leydig cells to decrease Fast levels (Figure 9A). Semi-quantitative RT-PCR showed that despite reduction of Fast mRNA by >80%, there was no notable decrease in SF-1 mRNA. (Fig. 9B). Further RT-PCR analysis of genes under SF-1 regulation showed no change in Cyp17 and StAR, mRNA levels indicating that they are not regulated by Fast (Figure 10). Furthermore, over expression of the Fast transcript in MA 10 cells did not alter mRNA levels of SF-1 (Figure 11).

**Fast does not regulate SF-1 protein levels.** SF-1 regulates steriodogenesis which is regulated by gonadotropins via cAMP. Thus, Fast may act in conjunction with the cAMP pathway to regulate SF-1. MA-10 cells were transfected with mu6 shRNA control and Fast shRNA 6, and 48 hrs post transfection cells were treated with cAMP and harvested at various time points of 0hr, 1hr, 2hr, and 6 hrs (Figure 10). No change in SF-1 protein was detected by Western blot with either mU6 or Fast shRNA or Fast over expression (Fig. 12, 13)

**Discussion**

In the current study, we characterized a conserved transcript within the Ftz-F1 locus that lacks an identifiable ORF expressed and is regulated in a manner that resembles SF-1. A ~7kd protein was detected in the pcDNA3 Fast construct in our in vitro transcription translation assay. On closer analysis of the Fast transcript we were able to identify an open reading frame that may contribute for the expressed protein. However, owing to the fact that it is an in vitro transcription translation assay, it is possible that what we are seeing is a spurious reaction (Dinger, Pang et al. 2008). A number of bifunctional RNAs have been identified that have been found to function at
both the transcript level and at the protein level. Examples of such bifunctional RNA include *Steroid Receptor Activator*, *VegT* and *Oskar* mRNA [reviewed in (Dinger, Pang et al. 2008)]. While the data are intriguing on several levels, *Fast*’s close association with SF-1 and its status as a noncoding RNA (ncRNA) are features, which reveal its greatest potential. In particular, its expression profile and response to retinoic acid and FSH suggest that, like SF-1, it too acts to control normal development and endocrine homeostasis. Furthermore, the molecular characteristics of *Fast* place it in the family of mRNA-like ncRNAs, which not only contribute significantly to the transcribed genome but whose members are increasingly identified with functions in a variety of biological processes.

With completion of the human genome came the surprising finding that only 25,000 or so protein-coding genes were required to direct such a complex biological program (Venter, Adams et al. 2001, 2004). It was then revealed that the number of expressed transcripts was much greater than the number of protein coding genes predicted from genome annotation (Okazaki, Furuno et al. 2002, Carninci, Kasukawa et al. 2005, Mattick 2005). Resolution of this discrepancy came with the discovery that a significant portion of the genome gives rise to transcripts that do not code for proteins. Thus, tiling experiments across ten human chromosomes revealed approximately 10% of the genome is expressed as polyadenylated transcripts, a number ten times greater than that for protein-coding genes and extensive analysis of mouse transcriptional units revealed a similar difference in the number of transcripts to estimated protein-coding genes (Carninci, Kasukawa et al. 2005, Cheng, Kapranov et al. 2005). While there is evidence that many of these putative transcripts represent library contamination from genomic DNA or unprocessed pre-mRNA, or alternatively nonspecific transcriptional noise, a wealth of studies substantiate important biological roles for polyadenylated ncRNAs.
(Ravasi, Suzuki et al. 2006). The few well-characterized ncRNAs identify functions in X-chromosome inactivation, chromatin structure, DNA imprinting, DNA methylation, transcription, environmental response, development, and cell differentiation, and has led to recognition of the potential magnitude of their biological importance [reviewed in (Kung, Colognori et al. 2013)]. Well known examples of such ncRNAs include *Xist*, *TsiX*, *H19*, *Kcnq1ot1*, *Air*, and *MALAT1*. *Xist*, found in both mice and humans, localizes to the inactive X chromosome in females to control dosage compensation (Lyon 1961, Lee 2011). It is also regulated by the ncRNA *TsiX* (Lee, Davidow et al. 1999). *H19*, *Kcnq1ot1*, and *Air* are mammalian RNAs associated with genomic imprinting (Brannan, Dees et al. 1990, Bartolomei, Zemel et al. 1991, Lee, DeBaun et al. 1999, Lyle, Watanabe et al. 2000). *MALAT1* a nuclear IncRNA, is associated with various cancers and regulates a number of cytoskeletal and extracellular matrix genes (Ji, Diederichs et al. 2003, Tano, Mizuno et al. 2010, Lin, Roychowdhury-Saha et al. 2011).

While the concept of noncoding RNAs (ncRNAs) is not novel, their extensive contribution to the expressed portion of the genome came as a surprise and the resulting biological implications has brought new appreciation to these RNAs and intensified research in order to understand their functional potential. In addition to the traditional RNA polymerase III-derived transcripts, tRNA and rRNA, there are numerous classes of noncoding RNAs including; microRNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small non-mRNA (snmRNA, a.k.a small ncRNA), small nucleolar RNA (snoRNA), small temporal RNA (stRNA), and large, mRNA-like ncRNA [reviewed in (Guil and Esteller 2012)] . While the mechanisms and characteristics of these RNAs are still being revealed, these non-coding RNA can be classified according to their nucleotide size and function. For small noncoding RNAs, the most
extensively studied are the 21–25 nucleotide non-mRNAs, which include miRNAs, siRNAs, and the 100-200 nucleotide snoRNAs. In animals, many miRNAs and all snoRNAs are located within introns [reviewed in (Mattick and Makunin 2005)]. Both siRNAs and miRNAs act to suppress gene expression by interacting in a sequence-specific manner with target mRNAs to regulate their stability and translation. They are also both excised from larger ncRNAs by RNase III digestion, requiring the parent RNA to have double-stranded secondary structure [reviewed in (Ghildiyal and Zamore 2009)]. Small nucleolar RNAs function within the nucleolus to guide ribosome biogenesis (Eddy 2001). These RNAs fall into two general classes, C/D box and H/ACA snoRNAs, which direct ribose methylation and pseudo-uridylation of rRNA, respectively. Among small ncRNAs, miRNAs have received the greatest attention due to their participation in diverse regulatory pathways, including control of developmental timing, haematopoietic cell differentiation, apoptosis, cell proliferation and organ development (Pasquinelli 2012).

While the above studies demonstrate important functional roles for these specific ncRNAs, they do resolve the question of function with respect to the numerous uncharacterized ncRNAs identified in expression libraries and if they, in general, are biologically important or an artifact of genome analysis. However, studies have begun to take more global approaches to determine if these ncRNAs represent functional RNAs. In a study by Cawley et al. binding sites for the transcription factors Sp1, cMyc, and p53 were mapped along human chromosomes 21 and 22 to help identify promoters regulated by these proteins (Cawley, Bekiranov et al. 2004). The results revealed the surprising finding that only 22% of the binding sites associated with protein-coding genes, while 36% of the sites were correlated with ncRNAs. Further examination showed that the ncRNAs on these chromosomes acted similar to the coding genes in response to retinoic
acid-induced differentiation. Likewise, studies in mice showed that many cloned cDNAs lacking open reading frames (ORFs) were derived from genuine transcripts of unknown function and that these ncRNAs, in general, have larger exons and fewer introns than protein-coding transcripts. The study also revealed that a significant proportion of ncRNAs are expressed in a tissue-specific manner and were regulated by extracellular signals (Ravasi, Suzuki et al. 2006). More recently, a study by Willingham et al. used shRNAs to knockdown expression of selected ncRNAs from the RIKEN Fantom2 mouse cDNA collection that showed significant sequence conservation to human genomic sequence. Eight new functional ncRNAs were identified, six essential for cell viability, one a repressor of Hedgehog signaling, and one a repressor of the transcription factor NF-κB called NRON (Willingham, Orth et al. 2005). Together, the findings indicated that ncRNAs are an important, regulated component of the mammalian genome and represent a largely uncovered level of gene regulation in complex organisms.

Importantly, transcription of one Fast exon and its orientation was conserved between species, suggesting that Fast was not a merely non-functional artifact of abundant, nearby SF-1 transcription. The most intriguing feature of Fast was its shared expression profile with SF-1, which strongly implicates a functional relationship between their respective protein and RNA products. Since Fast and SF-1 are expressed and regulated similarly, a concerted regulatory scheme for these two genes would be most efficient, and thus, the two genes may share some of the same transcriptional regulatory mechanisms. While its expression is clear in the human testis, it is not known if the full extent of Fast expression completely mimics human SF-1 expression.

Although a functional relationship between Fast and SF-1 were suggested by their shared expression patterns, a cellular or molecular role for Fast was not immediately apparent. At least
one possibility, however, was based on reports of other non-coding RNAs, which suggest that Fast mRNA could be processed to form a miRNA or siRNA. These two types of small ncRNAs (generally 20-25nt) constitute a growing class of small regulatory RNAs that control expression of protein-coding genes via an antisense-RNA mechanism by blocking mRNA translation or directly targeting mRNA degradation by hybridization with the homologous sequence in other genes (reviewed in (Huttenhofer, Schattner et al. 2005)). Although experiments in this study indicate that Fast does not regulate SF-1 expression via its homology to Ftz-f1 exon 1g, it is possible that another portion of Fast produces a small RNA that is homologous to one or more other genes.

In support of this theory, RNA secondary structure modeling indicated several highly stable stem-loop and hairpin structures which in the Fast cDNA, which would form the precursors necessary for generating a small RNA by RNase III digestion. Additional evidence was obtained experimentally by detection of a small RNA containing Fast sequence in Fast expressing cells within the highly conserved segment of the gene (data not shown). Thus, these results raise the possibility that secondary structure in Fast is processed to form a small RNA which serves some function in SF-1 positive cells. Although this possibility is intriguing, a significant number of non-coding RNAs are emerging that have no known associated function, raising the possibility that a large number of non-functional ncRNAs are produced in eukaryotic cells. Although we cannot exclude Fast as a member of this latter non-functional class, its high level of evolutionary conservation, intriguing expression pattern, and opposite transcriptional orientation to Ftz-f1 suggest that Fast is not merely coincidental. Ongoing experiments are seeking to clarify the existence of a small RNA produced from Fast and to determine its function in expressing cells.
The extensive biological importance of small ncRNAs to a wide array of critical developmental and physiological processes has emerged over the last several years. If subsequent studies demonstrate that *Fast* produces a small ncRNA, it may therefore play a fundamental regulatory role in expressing tissues. Some examples of miRNAs which are expressed at specific developmental time-points or during particular processes include those shown to participate in the mechanisms controlling stem cell and hematopoietic lineage differentiation, insulin secretion, and some which are imprinted, just to name a few (Seitz, Youngson et al. 2003, Seitz, Royo et al. 2004, Suh, Lee et al. 2004). Based on these exciting findings, there is tremendous potential that *Fast* participates in a novel physiological or developmental process that is key to endocrine development and function.
Materials and Methods

Assembly, annotation and comparative sequence analysis of the human and mouse Ftz-f1 loci. A 4187947bp subsequence of human chromosome 9 genomic contig (Accession #NT_008470) containing the entire human FTZ-F1 gene and roughly 2Mb each of 5’ and 3’ flanking sequence was identified in the NCBI Entrez Nucleotide search engine (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide). The mouse Ftz-f1 gene was identified within a single chromosome 2 genomic contig (Accession #NT_039206) in the same way. Repetitive sequences were identified and masked using the Repeatmasker web server (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker). Exons were assigned using the Spidey mRNA-to-genomic alignment program within NCBI (http://www.ncbi.nlm.nih.gov/Sitemap/index.html#Spidey). Two expressed sequence tags (ESTs) (Accession #AK017050 & #AK007201) were identified by sequence comparison of the rat and human FTZ-F1 loci using pairwise Blast alignment through the NCBI web site (http://www.ncbi.nlm.nih.gov/; Tatusova and Madden, 1999) under the program’s default parameters (Cost to open a gap [5], Cost to extend a gap [2], Penalty for a mismatch [-2], Reward for a match [1], Expectation value (E) [10], Word size, 11 for blastn) and a minimum conservation level of 75%.

The analysis used a FTZ-F1-containing subsequence of a human chromosome 9 contig (Accession #NT_008470) and a mouse chromosome 2 contig (Accession #NT_039206) containing Ftz-F1. Repetitive sequences were identified and masked using the web-based program Repeatmasker, exons assigned, and the sequences aligned using the BLAST sequence alignment program through the National Center for Biotechnology Information (NCBI). All
identified conserved sequences were examined against the nonredundant and expressed sequence tag databases at NCBI.

**Cell culture conditions.** Mouse P19 embryonal carcinoma (EC) cells were maintained in Dulbecco’s modified Eagle medium supplemented with 10% fetal calf serum, 2 mM glutamine, 100 U/ml penicillin, and 100 mg/ml streptomycin. For retinoic acid (RA) treatment of P19 cells, 1 μM RA was added to growth media for 12, 24, 36, 48 or 72 hours prior cell harvest. MA-10 cells, a mouse Leydig tumor cell line were cultured as described (Ascoli 1981). Mouse granulosa cells were isolated from ovaries of 19-day-old CF-1 mice at various time points after treatment with PMSG at various time points were a gift from Dr. Lane K. Christenson.

**RNA isolation and RT-PCR.** C57BL/6J mice were sacrificed at 6 weeks of age and total RNA was isolated from liver, kidney, spleen, stomach, heart, brain, lung, adrenal, testis, pituitary, and ovary, from immortalized cell lines MA-10, and from primary mouse Sertoli cells, mouse granulosa cells, and P-19 EC cells (with and without RA treatment) using Trizol reagent (Invitrogen Life Technologies, Carlsbad, CA) according to manufacturer recommendations. Human testis total RNA was purchased from BD Biosciences (San Jose, CA). Samples of polyadenylated mRNA were enriched from MA10 Leydig cell total RNA by oligo-dT cellulose chromatography as described (Sambrook and Russell 2006). Complementary DNA (cDNA) was synthesized from each RNA sample using 2 μg total RNA as described (Daggett, Rice et al. 2000).

For directional RT-PCR to determine transcriptional orientation, cDNA was synthesized from 2μg adult mouse Testis total RNA using species-specific primer 1 or primer 2 (Table 1), and separately with oligo-dT17 in the presence and absence of reverse transcriptase as described.
(Heckert et al., 2000). PCR was then performed with the cDNA templates and species-specific oligodeoxynucleotide primers 1 and 2. PCR-amplified products were visualized by agarose gel electrophoresis and analyzed by Southern blot hybridization using species-specific internal primers (primer 3, Table 1) radiolabeled using T4-polynucleotide kinase (New England Biolabs, Ipswich, MA) according to manufacturer recommendations.

For expression profiling, PCR-amplification of Fast (primers 1 and 2) cDNAs were performed as described above, while SF-1 cDNA containing exon 1g was detected by PCR using a primer against exon 1g (primer 1; Table 1) and another against exon 5 (primer 2; Table 1). Amplification of L7 cDNA controlled for cDNA synthesis as described (Lei and Heckert, 2002). Amplified products were visualized by agarose gel electrophoresis and, in some cases, Fast and SF-1 products were analyzed by Southern blot hybridization with internal oligodeoxynucleotide primers (primer 3; Table 1).

**Fast cloning.** Mouse EST transcripts were also PCR-amplified with oligodeoxynucleotide primers (primers 4 and 5, Table 1), and cloned using the pGEM T-easy vector system (Promega). The PCR product was cleaned of residual reactions by GeneJet PCR purification kit (Thermo Scientific) and ligation reaction was set up as described by the manufacturer. 2 μl of the ligation reaction was used for transformation and grown on agar plate containing ampicillin/X-gal/IPTG. The white colonies containing the PCR product was chosen and plasmid DNAs were prepared from overnight bacterial cultures using DNA plasmid columns according to the supplier’s protocol (GeneJet Plasmid Mini-prep kit, Thermo Scientific). All resulting clones were sequenced. The PCR-amplified product from human testis cDNA was sequenced with human primers 1 and 2.
Mouse *Fast* transcripts were PCR-amplified with oligodeoxynucleotide primer sets engineered with Kpn1 (GCG CGG TAC CCT CGG CCT TCA CCC TCA CCT CCT GGC CCT CCA GTT CCA GCT CGA TC) /Xba1 (CGG CTC TAG ACC AGT TCT GTG CAC CCA CTT TAT GTC TGG) restriction endonuclease sites and double restriction digest performed using Kpn1/ Xba1. The cloning vectors, pcDNA3 (Invitrogen, Carlsbad, CA) was restriction digested with Kpn1/ Xba1; dephosphorylated with calf intestinal phosphatase (New England Biolabs, Ipswich, MA), and ligated with the Kpn1/ Xba1 digested PCR products, respectively, using T4 ligase (New England Biolabs, Ipswich, MA).
<table>
<thead>
<tr>
<th>Target</th>
<th>Primer name and sequence</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>AK017050 &amp; mouse Fast</td>
<td>Primer 1; 5’-CACTCTACGGCATCCCAAGG-3’&lt;br&gt;Primer 2; 5’-CCAGGCAGGCACCAGACC-3’&lt;br&gt;Primer 3; 5’-CCATGTGGGACAGGGAGGT-3’</td>
<td>Directional RT- PCR, RT-PCR, southern blot</td>
</tr>
<tr>
<td>human FAST</td>
<td>Primer 1; 5’-ACTTTATGTTGCAAGGTTCC-3’&lt;br&gt;Primer 2; 5’-GGAAGCTGGGGCTGGAGGTCT-3’&lt;br&gt;Primer 3; 5’-gaaggagctctcagctggg-3’</td>
<td>Directional RT- PCR, southern blot</td>
</tr>
<tr>
<td>mouse Fast</td>
<td>Primer 4; 5’-CTCGGCCCCTCACCTCAC-3’&lt;br&gt;Primer 5; 5’-CCAGTTCGTCGACCACCTTACCT -3’</td>
<td>Cloning</td>
</tr>
<tr>
<td>mouse SF-1 (exon1g)</td>
<td>Primer 1; 5’-GTCGAGTTTTTCCTTGCTACC-3’&lt;br&gt;Primer 2; 5’-GCGGTTAAGAGGGCCAGGATAG-3’&lt;br&gt;Primer 3; 5’-GGGGTGCTAGAGACCTGGACGCTGTCG-3’</td>
<td>RT-PCR</td>
</tr>
<tr>
<td>mouse Fast §</td>
<td>5’RACE1; 5’-GATGTCCCCCGAGATTTGGTC-3’&lt;br&gt;5’RACE2; 5’-AGACCTTTGGGATGCCCATGCAG-3’&lt;br&gt;5’RACE3; 5’-gcgcaagcttGTGGCCACAGGAGCTGTCG-3’</td>
<td>5’RACE</td>
</tr>
<tr>
<td>human FAST #</td>
<td>5’RACE1h; 5’-GCCAGCACCTCATTAGAGAGGAGG-3’&lt;br&gt;5’RACE2h; 5’-CTGAACTATTTTAGGGCCAGG-3’&lt;br&gt;5’RACE3h; 5’-gcgcaagcttGTGGGCACAGGAGGCTGTCG-3’</td>
<td>5’RACE</td>
</tr>
<tr>
<td>mouse Fast</td>
<td>3’RACE1; 5’-ACTTTTGGCTTCGTCGCTGCACC-3’&lt;br&gt;3’RACE2; 5’-gcgcaagcttGCCTGGCTTCCGACAAATG-3’</td>
<td>3’RACE</td>
</tr>
<tr>
<td>Mouse Fast</td>
<td>Primer 6; 5’-TCGCCACAGTCTGACTCTTC -3’</td>
<td>Southern blot</td>
</tr>
<tr>
<td>mouse SF-1 exon 1g *</td>
<td>SF-1 e1g; 5’-gcgcaagcttGTCCAGTTTTTCCTTGCTACC-3’&lt;br&gt;SF-1 e4; 5’-geggagctcGTTGATGCTTTGTGACGG-3’</td>
<td>RNase protection probe</td>
</tr>
<tr>
<td>mouse SF-1 exon 1a *</td>
<td>SF-1 e1a; 5’-geggagcttcGAGTTTTTGAGAGCGCCC-3’&lt;br&gt;SF-1 e4; 5’-geggagctcCGTGAATGCTTTGAGTGTA-3’</td>
<td>RNase protection probe</td>
</tr>
<tr>
<td>Gene</td>
<td>Forward Primer</td>
<td>Reverse Primer</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Mouse Gapdh</td>
<td>mGapdh exon-exon primer F: 5’-AACCTTTGGCATTGTGGAAGG-3’</td>
<td>mGapdh exon-exon primer R: 5’-TGTGAGGGGAGATGCTCAGTG-3’</td>
</tr>
<tr>
<td></td>
<td>mGapdh splice primer F: 5’-GTGCAGGACCTCACTCATTG-3’</td>
<td>mGapdh splice primer R: 5’-CACATTGGGGGTAGGAACAC-3’</td>
</tr>
<tr>
<td>Mouse L7</td>
<td>F: GGGGGAAGCTTCAAGCAAGCAAGGAAGCT</td>
<td>R: GGGGGGTCGACTCCATGCAGATGATGCC</td>
</tr>
<tr>
<td>Mouse StAR</td>
<td>F: CCG GAG CAG AGT GGT GTC A</td>
<td>R: GCC AGT GGA TGA AGC ACC AT</td>
</tr>
<tr>
<td>Mouse Cyp17</td>
<td>F: GCC TGA CAG ACA TTC TG</td>
<td>R: TCG TGA TGC AGT GCC CAG</td>
</tr>
</tbody>
</table>
Determination of transcriptional start site, polyadénylation site and transcript size. 5’ Rapid amplification of cDNA ends (5’RACE) was performed as described elsewhere (Frohman, 1993; Frohman, 1990) with modifications using nested oligodeoxynucleotide primers generated against mouse Fast. cDNA was synthesized from mouse testis total RNA using a primer within the first exon of Fast (5’RACE1) and enhanced AMV reverse transcriptase (Promega) using a ramped synthesis incubation (55°C to 58°C over 55 minutes; Table 1). The cDNA was used for two rounds of PCR amplification using gene-specific nested primers 5’RACE2 and 5’RACE3, and products were cloned and sequenced (Table 1). Transcriptional start sites were assigned based on DNA sequence analysis of the RACE clones. 5’RACE to detect the transcriptional start site of human FAST was performed from human testis total RNA as above with primers 5’RACE1h, 5’RACE2h and 5’RACE3h. 5’RACE was also performed to detect 5’ 7-methylguanosine capped mRNAs using the same primers described above with the GeneRacer kit (Invitrogen) according to manufacturer recommendations. 3’RACE was performed to determine the polyadenylation site as described (Frohman, 1993; Frohman, 1990) with modifications. In short, cDNA was synthesized from mouse testis RNA as described and used in two rounds of nested amplification using 3’RACE1 and 3’RACE2, and products were cloned and sequenced.

RNAi-mediated Fast knockdown: Targets for shRNAi were selected from within the Fast cDNA sequence as described (Elbashir et al., 2002) and short hairpin RNA (shRNA) double-stranded oligodeoxynucleotide inserts for each target were designed and cloned into the mU6pro vector as described (Table 2; Yu et al., 2002). A Bgl I-Pvu II fragment from each clone containing the U6 promoter-shRNA cassette or the control U6 promoter-GFP cassette was blunted with Klenow fragment (Roche). Likewise, pcDNA3 (Invitrogen) was digested with Bgl
I to remove the CMV promoter, blunted with Klenow, dephosphorylated with calf intestinal phosphatase as above, and ligated with the U6 expression cassettes with Quick Stick ligase (Bioline). Each U6 cassette- pcDNA3 plasmid was transiently transfected into MA10 Leydig cells as follows. MA-10 cells were plated at a density of 400,000 cells per well in 6-well plates and transfected using 2µg vector DNA, 2µg vector DNA, 12µl lipofectamine (Invitrogen) and 10µl PLUS reagent (Invitrogen). Lipid/DNA complexes were removed after 24h, and the cells were incubated with complete growth media (Waymouth's supplemented with 20 mM Hepes, 15% horse serum, and 50 µg/ml gentamicin). Total RNA was harvested from the transfected MA10 cells and cDNA was synthesized as above and diluted 1:20 in nuclease-free H2O. Expression of Fast was analyzed by RT-PCR.

**Overexpression studies:** Fast was cloned in a pcDNA3 vector under the CMV promoter. MA-10 cells were plated at a density of 400,000 cells per well in 6-well plates and transfected using 2µg vector DNA, 12µl lipofectamine (Invitrogen) and 10µl PLUS reagent (Invitrogen). The next day cells were fed with complete growth media. Cells were harvested 48 hr post-transfection, and protein was extracted in 50ul of freshly made RIPA solution.

**Western Blot:** Western blot analysis was performed as described (Heckert, Sawadogo et al. 2000). Briefly, whole cell extracts from transfected MA-10 cells were extracted in RIPA buffer and resolved by SDS-PAGE, transferred to polyvinylidene fluoride (Millipore Corp., Billerica, MA), blocked, and probed overnight at 4 C with SF-1 antibody (1:3,000 dilution) a kind gift from Dr. Kenichirou Morohashi, Japan,. After incubation with secondary antibody, protein complexes were visualized by chemiluminescence using the enhanced chemiluminescence (ECL) system (Amersham Life Sciences, Arlington Heights, IL). Horseradish peroxidase-
conjugated secondary antibodies were donkey antirabbit IgG (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA). Blots were stripped [100 mM 2-mercaptoethanol, 2% sodium dodecyl sulfate, and 62.5 mM Tris-HCl (pH 6.7)] for 30 min at 55°C according to manufacturer’s recommendations (Millipore), and reprobed for Mouse anti-α-Tubulin (Calbiochem, CP06) which was used as loading control.

**In vitro transcription translation:** The in vitro coupled transcription/translation was done using the T7 rabbit reticulolysate system (Promega). Briefly, the gene was cloned in a pcDNA3 vector downstream of T7 polymerase and was added to rabbit reticulolysate along with $^{35}$S-methionine, and incubated for 90 min at 30°C. The in vitro translated products were separated in 15% SDS-PAGE. The gel was dried and exposed to a PhosphorImager screen overnight.

**Subnuclear fractionation:** Subnuclear fractionation of RNA was done using Norgen Biotek’s Cytoplasmic and Nuclear RNA isolation kit. In brief, MA-10, mouse Leydig cells were grown in a monolayer in 100cm plates. Confluent plates were washed with PBS, lysed, and lysates was spun down. The supernatant containing the cytoplasmic RNA was transferred to a new tube, and the nuclear RNA pellet was washed and eluted from the spin columns to a new eppendorf tube.

**Southern Blot:** RNA was isolated from mouse granulosa cells at various time points post-PMSG, and cDNA was generated as described earlier. RT-PCR was performed by using primers 4 and 5 that span the entire *Fast* locus and products were resolved in a 1% Ethidium Bromide gel for two hours. The bands were then transferred to a Protran nitrocellulose membrane for 3 hours. The membrane was washed twice, pre-hybridized, and probed overnight with Primer 6 located on *Fast* exon3, which was 5’ end labeled with $\gamma^{32}$P-ATP, by T4 Polynucleotide Kinase (New
England Biolabs). The following day the membrane was washed twice, and bands were detected by exposing it to film overnight.
Figure 1: Two highly conserved ESTs are located between *Nr5A1* (*Ftz-F1*) and *Gcnf* loci. Two conserved ESTs located between the *Ftz-f1* and *Gcnf* loci. Representation of mouse chromosome 2 in the proximity of *Ftz-f1* (*Nr5a1*; blue rectangles) and *Gcnf* (*Nr6a1*; grey rectangles). Conserved, non-coding ESTs identified by pairwise alignment of mouse and human genomic sequences are marked (Accession #AK007201 = EST 12840611; red rectangles; and Accession #AK007201 = EST 12840611; green rectangle, ). A *Gcnf* variant transcript with an extended exon 12 is depicted by the hatched line extending the *Nr6a1* locus.
**Figure 2:** Fast is transcribed opposite to Ftz-F1 (A) Diagram of mouse Fast and Ftz-f1 5' sequence, indicating exon positions of Ftz-f1 (blue) and FAT (red). (B) PCR amplification of Fast using primers 1 and 2 (locations shown in A) and cDNA synthesized with the indicated primer (top) in the presence (+) or absence (-) of reverse transcriptase. Upper panel shows the resulting ethidium bromide stained agarose gel and the lower panel an autoradiogram of its Southern analysis hybridized with an internal primer (shown in A). (C) Human FAST and FTZ-F1 5' flanking sequence, indicating positions of FTZ-F1 (blue) and FAST exons (red). (D) PCR-amplification of FAST using primers 1 and 3 (locations shown in C) and cDNA synthesized with the indicated primer (top) in the presence (+) or absence (-) of reverse transcriptase. Upper panel is the ethidium bromide stained agarose gel and the lower panel an autoradiogram of its Southern analysis hybridized with an internal primer (shown in C). Bent arrows indicate transcriptional orientation and start sites. The mouse SF-1 gene has multiple promoters accompanied by first exons, Ia, Ig, and Io, which have been identified in the figure below. (Brian Hermann-Regulation of FSH-receptor and SF-1: transcriptional control in reproduction)
**Figure 3:** Diagrammatic representation of the genomic location of *Fast* and its transcripts. Top shows the relative position of *Ftz-f1* exons 1-4 (blue) and *Fast* exons 1-3 (orange). Transcriptional orientation and start sites are shown as bent arrows. Human *FAST* contains only the exon that corresponds to mouse exon 3. Bottom shows the two known *Fast* splice variants; one, containing all three known exons, indicates introns as solid lines and the other, lacking exon 2, indicates the intron as a hatched line. (Brian Hermann- Regulation of FSH-receptor and SF-1: transcriptional control in reproduction)
**Figure 4:** *Fast* localizes predominantly in the cytoplasm. Nuclear and cytoplasmic RNA fractions were isolated from MA-10, mouse Leydig cells, and examined for the presence of *Fast* by RT-PCR (Top). Mouse *Gapdh* primers were designed located in different exons (middle) and within an intron and exon (bottom), to check for purity of cytoplasmic and nuclear fractions, respectively. Experimental repeats shown below with L7 as control.
**Figure 5:** *Fast* may code for a 7kd protein. Top: In vitro transcription-coupled translation assay of no plasmid control; empty vector negative control; triplicate of *Fast* pcDNA3 (1.7-kb gene cloned into pcDNA3 vector downstream from T7 promoter); DMRT1 pcDNA3 encoding 39-kDa DMRT1 protein; and the positive control luciferase encoding 61-kDa protein. Below: The longest ORF highlighted in red, which may possibly encode for the 7kd protein.
LGLHPHLALQFQLDPKPASSKRGDEHPRPVRNERGFGPPAPGGLDLPHPASPVStopGCRGLNAGRLLAStopKWLKVStopVRGRStopARKNWTFELALPHLPLQAAGTHSSPVFStopPEASSPRVVW

LKMetDVVTARVPGGEPHYQESPSCPGNGMetYGQFHTASQGLHSCKNPQLFIFPPSSKPEPRILStopPALELPGLSLNStopTKSR

GHLCSYDStopStopPTS Met GVRVLRWEDQRARESRYCPGNNGVT

GTFRPPSTHLGSHPStopStopIHEARSGGLFHVVEEARAStopGGR

GEGRKLGLEVLAStopGPSSCLStopGSDCLRKLGERREAMetVTSG

KKAQPPCAH Met GFSSPGLKStopCNSAIDAAGPSLNLLEGLOPWFVLPVGAVVMetCEGLPSYLVQIRLCPTATGTKTNKQTNKQPHRSStop

TSFSGCKFQLLHStopSASRLWSGAACLAFQNENSSGMetARSGAFStopPNEDRALPDIKWVHRT
Figure 6: Expression of Fast and SF-1 in adult mouse tissue and mouse Leydig cell line, MA10. (A) Top panel, adult mouse tissues and MA-10, cell line were examined for the presence of Fast, SF-1 and ribosomal protein, L7, in the presence (+) and (-) of reverse transcriptase. Southern blot utilizing an internal primer for Fast is shown in the lower panel. (Brian Hermann- Regulation of FSH-receptor and SF-1: transcriptional control in reproduction)
**Figure 7:** *Fast* and SF-1 respond to retinoic acid in P19 embryonal carcinoma cells. P19 embryonal carcinoma cells were treated with 1μM retinoic acid. RNA was isolated at various time points (top). RT-PCR-amplified products were resolved by agarose gel electrophoresis and stained with ethidium bromide. Assays were performed in the presence (+) or absence (-) of reverse transcriptase (RT). (Brian Hermann- Regulation of FSH-receptor and SF-1: transcriptional control in reproduction)

<table>
<thead>
<tr>
<th>hrs + RA</th>
<th>0</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>72</th>
<th>Testis</th>
<th>H₂O</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**Fast**

**SF-1**

**L7**
Figure 8: Hormonal response of Fast in mouse granulosa cells. Granulosa cells were isolated from ovaries of 19-day-old CF-1 mice at various time points after treatment with PMSG at various time points as indicated. RNA was isolated and assayed for fast by RT-PCR followed by Southern blot analysis.
Figure 9: Knocking down Fast does not affect transcription of SF-1: (A) Position of Fast shRNA 6 target used for RNAi-mediated Fast knockdown (blue boxes) is shown on this diagram of the Fat cDNA sequence. Fat and SF-1 exons (shaded boxes), EST sequence (red boxes), the highly conserved segment of exon3 (85% identity, orange box), and the poly-adenylation signal (red balloon) are shown. Expression vectors containing each shRNA target were transfected into MA10 Leydig cells and cells were harvested 48 hours post-transfection. (B) RNA was isolated, reverse transcribed and assayed for the presence of Fast and mouse SF-1. Mouse Gapdh was used as internal control.
**Figure 10:** Knocking down *Fast* does not affect downstream targets of SF-1: RNA was isolated, reverse transcribed and assayed for the presence of Cyp17, Cyp11a, and StAR. Mouse Gapdh was used as internal control.
**Figure 11:** Over expressing *Fast* does not affect SF-1 mRNA levels: RNA was isolated from MA-10 Leydig cells transfected with empty pcDNA3 control and *Fast* pcDNA3. RNA was isolated, reverse transcribed and assayed for the *Fast* and SF-1 Gapdh served as internal control.
**Figure 12:** Knocking down *Fast* does not affect translation of SF-1: Protein was isolated from MA-10 Leydig cells transfected with mu6shRNA control and *Fast* shRNA#6 and probed for SF-1 levels at various time points, post cAMP as indicated. Tubulin was used as internal control.
Figure 13: Over expressing *Fast* does not affect SF-1 protein levels: Protein was isolated from MA-10 Leydig cells transfected with empty pcDNA3 control and *Fast* pcDNA3 and probed for SF-1 levels 48 hours post-transfection. Tubulin was used as internal control.
Figure 13: Over expressing *Fast* does not affect SF-1 protein levels: Protein was isolated from MA-10 Leydig cells transfected with empty pcDNA3 control and *Fast* pcDNA3 and probed for SF-1 levels 48 hours post-transfection. Actin was used as internal control.
Chapter 5: An upstream distal Evolutionary Conserved Region regulates the expression of nuclear Receptor Steroidogenic Factor 1 in the pituitary
ABSTRACT

Steroidogenic factor 1 (SF-1/Ad4BP/NR5A1) is a nuclear receptor with a pivotal role in the development of adrenal glands and gonads. Despite many studies that describe transcriptional features of \( Nr5a1 \), the mechanisms associated with cell-specificity are incompletely resolved for many expressing cell types. As revealed by studies in transgenic mice, these mechanisms require distal regulatory elements residing within a 153kb transgene containing the \( Nr5a1 \) locus. This 153kb transgene defined the initiation point for the current studies. Thus, comparative sequence analysis of this region was used identify sequences of high evolutionary conservation as a means to identify candidate regulatory elements of \( Nr5a1 \). This identified multiple evolutionarily conserved non-coding regions (ECRs) and four of the most highly conserved were selected for additional analysis to test their transcriptional potential. Each ECR was cloned upstream of the -734/+60 SF-1 promoter directing expression of a luciferase reporter and its transcriptional activity assessed in various cell types by transient transfection analysis. One of the ECRs, ECR3, upregulated SF-1 transcription in alpha T3 (gonadotrope cell line) and Y-1 (adrenal) cells, but downregulated SF-1 in MA-10 (Leydig cell line), MSC-1 (Sertoli cell line) and primary rat Peritubular Myoid and Sertoli cells. Our data indicate that ECR3 contains control elements necessary for SF-1 transcriptional induction in pituitary and adrenal cells and repression in MA-10, MSC-1 cell lines and primary rat Peritubular Myoid and Sertoli cells. Mutagenesis and DNA/protein interaction studies of ECR3 identified sequences important for activity in alpha T3 cells. The ECR3 sequence contains an E box that binds class A basic-helix-loop-helix (bHLH) protein E2A, suggesting cell-specific class B bHLH protein as its likely dimeric partner. Co-transfection studies identified Inhibitors of DNA binding, Id2 and Id3 downregulate SF-1 transcription possibly by interacting with E2A, and preventing its binding to the E-box. Thus,
these studies show that an E box present in the ECR is required for SF-1 expression in the pituitary and involves members of the bHLH protein family.

**INTRODUCTION**

Steroidogenic factor 1 (SF-1), also known as Ad4BP (adrenal 4 binding protein) and officially designated NR5A1, belongs to the nuclear receptor superfamily. The gene, *NR5A1*, encodes four different proteins, ELP1, ELP2, ELP3 and SF-1. In particular, the transcription factor SF-1 has emerged as a key regulator of reproductive tract development and endocrine homeostasis (Sadovsky, Crawford et al. 1995, Shinoda, Lei et al. 1995, Parker and Schimmer 1997). SF-1 binds to a common regulatory motif (AGGTCA) within promoters and initiates and regulates the expression of various genes. SF-1 was first identified by its ability to regulate the promoter activity of several cytochrome P450 steroid hydroxylase genes (Rice, Mouw et al. 1991, Morohashi, Honda et al. 1992)

The necessity of the SF-1 protein emerged from the development of the SF-1 knockout (KO) mice. The SF-1 KO mice lacked adrenal glands and gonads and died shortly after birth due to adrenal insufficiency (Luo, Ikeda et al. 1994, Hammer, Parker et al. 2005). SF-1 KO mice also exhibited male to female sex reversal, impaired gene expression within pituitary gonadotropes and structural abnormalities of the ventromedial hypothalamic nucleus (VMH) (Ingraham, Lala et al. 1994). Data obtained from both global and pituitary-specific SF-1 KO mice support a role for SF-1 in gonadotrope function and transcriptional regulation of the gonadotropins, luteinizing hormone and follicle-stimulating hormone, as both are absent in mice deficient of SF-1(Zhao, Bakke et al. 2001). Gonad specific knockout mice demonstrated that SF-1 action in the gonads is essential for proper gonadal function in both male and female mice (Jeyasuria, Ikeda et al. 2004).
In humans, mutations in SF-1 resulting in a loss of activity are also associated with male to female sex reversal and adrenal failure (Achermann, Ito et al. 1999, Biason-Lauber and Schoenle 2000, Achermann, Ozisik et al. 2002). Thus, both in vitro and in vivo evidence establish SF-1 as a critical regulator of development and endocrine homeostasis in humans and mice.

Given the critical role played by SF-1 in development and endocrine homeostasis, many studies have focused on uncovering the regulatory mechanisms that specify SF-1 transcription in the appropriate spatio-temporal manner. However, despite the efforts, little has been revealed about the specific transcriptional mechanisms that drive tissue-specific SF-1 expression. Transient transfection studies, using the SF-1 promoter in front of a reporter, have revealed several cis-regulatory elements that control basal transcription. In particular, the promoter has an E box, a CCAAT box, and three Sp1 binding sites (Barnhart and Mellon 1994, Nomura, Bartsch et al. 1995, Daggett, Rice et al. 2000). Further studies on the proteins binding to these elements uncovered a variety of binding complexes as well as cell-specific interactions (Daggett, Rice et al. 2000, Scherrer, Rice et al. 2002). While these findings have shed light on basal function of the SF-1 promoter, they have not revealed any mechanisms to explain how SF-1 expression is restricted to different target tissues or how its transcriptional regulation contributes to its role in development and endocrine homeostasis.

The development of transgenic mice has lead to the identification of specific DNA elements, which direct SF-1 transcription in vivo. Accumulated evidence indicates that proper expression of SF-1 requires a large genomic fragment that spans a 153kb region of the Nr5a1 locus. (Wilhelm and Englert 2002) identified a 674bp region on the SF-1 promoter that partially directs expression to the gonads. However, the study also revealed that the 674bp promoter region
lacked key regulatory sequences needed for SF-1 expression in most of its target tissues (Wilhelm and Englert 2002). A transgenic study by Stallings et al., targeted a 50kb region of mouse \textit{Nr5a1}, including 45kb of 5’ flanking sequence, exon 1 and part of exon 2, to direct expression of green fluorescent protein (Stallings, Hanley et al. 2002). This 50kb transgene was able to direct expression to many but not all SF-1-expressing cells. Our laboratory demonstrated that a 153kb yeast artificial chromosome (YAC) transgene containing rat \textit{Nr5a1} completely mimicked expression of endogenous \textit{Nr5a1} and rescued all known defects of SF-1 null mice (Karpova, Presley et al. 2005). Together, these studies indicate that tissue-specific expression of SF-1 is directed by regulatory elements located at significant distances from the transcription start site. Since the 153kb transgene used in our previous studies contains all the sequences necessary for proper SF-1 expression, we used this transgene to identify and characterize essential distal regulatory elements. Here, we report the identification and characterization of an Evolutionary Conserved Region (ECR), ECR 3. ECR3 contains regulatory elements that enhance transcriptional activity of the SF-1 promoter specifically in gonadotrope and adrenocortical-derived cells.

RESULTS

Comparison of \textit{NR5A1} between human and chicken genomic sequences identifies Six Major Evolutionarily Conserved Regions.

To help identify potential regulatory elements essential for SF-1 expression, a web-based sequence analysis tool, ECR browser was used (Ovcharenko, Nobrega et al. 2004). The ECR browser identifies sequence conservations between various species that have remained unchanged for millions of years, suggesting that the predicted sites are functionally important.
Comparison of rat, mouse, frog and chicken *Nr5a1* identified more than twenty-five ECRs. (Figure 1). To improve functional prediction, a more distant genome (chicken) was included in the comparative analysis. This sequence conservation analysis between chicken, mouse, and human led to a more stringent identification of 6 ECRs which were chosen for further analysis, as they represent the most conserved, and thus most likely, regions to contain essential regulatory elements.

**ECR3 differentially regulates SF-1 transcriptional activity in αT3 and Y-1 cells versus MA-10, Myoid, MSC-1 and primary Sertoli cells.**

To identify functional importance of the ECR’s, four ECRs (ECRs 1-4) were cloned into a vector that contains the SF-1 promoter, from -734 to +60, directing expression of the firefly luciferase reporter. Transcriptional activity was tested by transiently transfecting the vector in mouse Leydig tumor cell line MA-10, mouse sertoli cell line MSC-1, Y-1 murine adrenocortical cells, and mouse gonadotroph-derived alpha T3 cells, as well as primary cultures of rat Sertoli and Peritubular Myoid cells. With TK-renilla as an internal control, firefly and renilla luciferase activities were compared to determine transcriptional activity. The luciferase/renilla values of the vectors containing ECRs were normalized to the values for the SF-1 promoter alone. Of the four ECRs, ECR3 had the most dramatic effect (Figure 2). ECR3 increased SF-1 transcriptional activity in αT3 and Y-1 cells, but decreased transcriptional activity in MA-10, MSC-1 and primary rat Peritubular Myoid and Sertoli cells. This data indicates that ECR 3 likely contains elements important for SF-1 regulation and acts in a cell-specific manner, activating SF-1 transcription in pituitary and adrenal cells, while repressing its activity in Leydig, Sertoli, and myoid cells.
ECR3 contains potential regulatory elements and key regions necessary for full enhancer activity.

Next, we decided to identify regions of ECR3 that are important for the enhancer activity seen in αT3 and Y-1 cells. To this effect, deletion mutants were generated by cloning varying lengths of ECR3 based on conservation profile and size, namely ECR3 LC (less conserved), ECR C (highly conserved), ECR3-300, ECR3-350, ECR3-504, and ECR3-580 into SF-1(-734+60)Luc vector. The transcriptional activities of these mutants were then assayed in αT3 pituitary cells, as ECR3 showed highest enhancer activity in these cells (Figure 2). The less conserved region within ECR3 showed very minimal activity when compared to that of full-length ECR3 (Figure 3). ECR3 C showed increased activity compared to ECR3 LC, but not to the level of full-length ECR3. None of the additional constructs tested (ECR3-300, ECR3-350, ECR3-504, and ECR3-580) reached the activity of full-length ECR3. These results suggest that regulatory elements reside within the ECR3 locus and their interactions are necessary for complete activity.

To further categorize ECR3, DNase I footprinting was performed with and without αT3 cell nuclear extracts using a probe as indicated in Figure 4. Addition of nuclear proteins revealed a prominent hypersensitive site and two protected footprints. In silico sequence analysis using the web-based program rvista (http://rvista.dcode.org/) identified several evolutionarily conserved sequences that are potential transcription factor binding sites, E box, p53 tumor suppressor protein and E twenty-six (ETS)-like transcription factor 1, also known as Elk1 were chosen based on the scores.
An E box is required for full transcriptional activity of ECR3.

To evaluate the effects of these transcriptional sites in αT3 cells, potential transcription factor-binding sites identified above, E box, p53 and Elk1, were mutated and were transiently transfected in αT3 cells. These cells were then assayed for relative luciferase activity as before, and activities were graphed (Figure 5). Of the three transcription factor binding sites tested, mutation of the E box element significantly decreased the transcriptional activity, suggesting functional importance.

The E box site identified was **CATCTG** (E box elements have the consensus sequence **CANNTG**), a sequence that binds basic helix-loop-helix (bHLH) proteins. bHLH proteins are transcription factors, divided into two broad functional groups by their patterns of expression. Class A bHLH members, often also referred to as E-proteins, which are ubiquitously expressed and bind DNA as homo- and hetero-dimers. In contrast, Class B proteins are expressed in cells of particular lineage (e.g., MyoD and Myogenin) and mainly bind as heterodimers. These proteins have a helix–loop–helix domain that mediates homo- and/or hetero-dimerization with other HLH proteins. To help identify the bHLH protein that binds to E box, we performed an electrophoretic mobility shift assay (EMSA) using αT3 nuclear extracts and a radiolabeled probe containing the ECR3 E box. Unlabeled competitors with sequence homology were added to the reaction at a concentration 100X that of probe. Similarly, antibodies to bHLH proteins, Upstream stimulatory factors 1 and 2, (USF1 and USF2) or transcription factor E2A, a ubiquitous bHLH protein, were added to specific reactions. Notably, formation of binding complexes was inhibited by inclusion of unlabeled homologous competitor DNA but not a non-specific (N sequence, indicating the proteins bind specifically to the ECR3 E-box probe sequence (Figure 6). However, a competitor containing a mutation in the E box did not compete for the fastest and slowest migrating
complexes and was a poorer competitor for some of the complexes in the middle. Thus, the top and bottom complexes appear to be dependent on the E box sequence for binding and are candidates for eliciting the transcriptional activity associated with E box. As indicated by their cross-reactivity to included antibodies, these candidate proteins are USF1, USF2, and E2A, plus an unidentified binding partner.

Chromatin immunoprecipitation (ChIP) analysis was employed to analyze transcription factor binding to SF-1 ECR3 site, in vivo. Formaldehyde cross-linked chromatin from αT3 cells was used for immunoprecipitation with antibodies to E2A, USF1, USF2, RNA polymerase II, p300, histone active mark H3Ac, and repressive histone marker H3K27me3. With chromatin prepared from αT3 cells, E2A, USF1 and USF2 were found to bind to the ECR3 region and were congruent with EMSA results (Fig 7). In contrast, p300 an enhancer marker was found not enriched at Sf-1 ECR3 site. Active histone mark H3Ac was found to be significantly enriched with lower binding of H3K27me3 indicating that the region was in an “open” chromatin state.

**Id2 and Id3 inhibit transcriptional activity through the ECR3 E box**

Having identified that bHLH proteins bind to the E-box, we then decided to establish whether Id proteins inhibit transcriptional activity by binding to the E-box. Ids are a group of HLH proteins that lack a DNA binding domain and act as negative regulators of bHLH proteins (Ruzinova and Benezra 2003, Perk, Iavarone et al. 2005). To investigate Id proteins that inhibit transcriptional activity, Id1, Id2, and Id3 were cloned and co-transfected in αT3 cells. Increasing concentrations of empty vector or expression vectors for Id1, Id2, and Id3, were co-transfected into αT3 cells with luciferase reporters driven by wild type (Trompouki, Bowman et al.) or E box mutant
(mEbox) SF-1(-734+60)-ECR3. Relative reporter activity was assayed for each line (Figure 7). Co-transfection of SF-1(-734+60)-ECR3 with plasmids expressing Id2 and Id3 showed a dose dependant diminished SF-1 promoter activity and this change in transcription levels was completely lost when Id plasmids were transfected with mEbox SF-1(-734+60)-ECR3 construct. In contrast, expression of Id1 expression plasmid had little or no impact on the transcription level of SF-1. This data demonstrates that Id2 and Id3 proteins inhibit the activity of the SF-1 transcription and it does so by inhibiting the binding of the bHLH proteins to the E-box in the ECR3 distal region.
DISCUSSION

Nuclear hormone receptor, SF-1, is a key regulator of endocrine homeostasis and development. Many studies have investigated the expression profile of SF-1 and the mechanisms associated with its strict tissue specificity. However, our knowledge of the transcriptional regulatory mechanisms responsible for proper spatiotemporal expression of SF-1 has been limited. A previous study from our lab established that a 153kb region of the NR5A1 locus contains all the necessary control elements for SF-1 activity (Karpova, Presley et al. 2005). We identified Evolutionarily Conserved Regions that serve as potential regulatory sequences important for directing the spatiotemporal expression of SF-1.

DNA sequence conservation is a characteristic that has long been associated with functional regions of the genome. More recently, the importance of conserved non-coding sequences in transcriptional regulation has been validated by numerous studies (Frazer, Sheehan et al. 2001, Visel, Blow et al. 2009, Blow, McCulley et al. 2010). Sequence comparisons between different species have often shown that functional non-coding sequences are highly conserved, whereas sequences that are not functional diverge (Hardison 2000). The location and function of regulatory sequences that orchestrate gene expression remain obscure, which makes it difficult to study their role in developmental processes (Pennacchio, Ahituv et al. 2006, Prabhakar, Poulin et al. 2006, Holland, Albalat et al. 2008, Visel, Prabhakar et al. 2008). A widely used approach is comparative genomics of both closely related and highly divergent organisms to identify specific control elements (Gottgens, Barton et al. 2000, Thomas, Touchman et al. 2003, Chapman, Donaldson et al. 2004). We employed the same approach in comparing different genomes ranging from human to chicken to identify six ECRs as potential regulators of SF-1.
While non-coding sequences under evolutionary pressure could predict the location of enhancer elements in the genome, they do not reveal when and where these enhancers are active in vivo (Visel, Blow et al. 2009). To test the functional importance of these sequences, one can use different experimental strategies, including in vitro or in vivo assays, like transient transfection or transgenic studies. In the present study, we tested the transcriptional activity of SF-1 as regulated by four ECRs, using transient transfection reporter assays in various cell types. Of these ECRs, ECR3 showed strong cell-specific up regulation in αT3 pituitary and Y-1 adrenal cell types; however, transcriptional repression was observed in testis cell types.

ECR3, like other critical regulatory elements, contains elements identified through sequence conservation. For example, DNaseI hypersensitivity studies of cis-regulatory elements showed that about 70% of regulatory sequences are found at least 5kb from the transcriptional start site of the gene (Sabo, Humbert et al. 2004). Interestingly, ECR3 is located 8kb upstream of the transcriptional start site. The observed characteristics of ECR3, including evolutionary conservation, cell-specific activity and distance relative to SF-1 transcription start site, triggered us to further characterize ECR3.

Previous studies employing transgenics from the laboratory of Ken Morohashi identified many tissue-specific regulatory regions within the SF-1 gene (Shima, Zubair et al. 2005, Zubair, Oka et al. 2009, Shima, Miyabayashi et al. 2012). Using a 5.8kb construct containing Exon 2 and the upstream promoter region, enhancers for fetal adrenal gonad and VMH were found in introns 4 and 6, respectively, in two different studies (Shima, Zubair et al. 2008). Interestingly, these regions correspond to two of our ECRs, ECR2 and ECR1. Another transgenic study from the same group identified a 5kb fragment that contained Rathkes pouch-specific enhancer activity
Further studies, uncovered a 486bp enhancer element in intron 6 of the gene, that is necessary for gonadotrope-specific enhancer. This enhancer element was reported to contain several conserved sequences. In contrast, however, the ECR we identified is a non-coding region (ECR3) that contains regulatory sequences specific for SF-1 expression in pituitary.

*In vitro* foot printing identified binding sites for several proteins that, when mutated, revealed their importance for transcriptional activity of ECR3. In addition, *in silico* analysis of footprinted sequences identified potential transcription factor binding sites for p53 and members of the ETS- and bHLH families. Further analysis of these sites revealed that an E-box element is essential for the activity of ECR3 in pituitary gonadotropes. E-box elements have been known to play an important role in the expression of many genes in different organs. Our current study, combined with the study of Shima et al, demonstrates that, pituitary-specific expression of SF-1 involves a highly complex transcriptional regulatory network and involves multiple transcriptional control sequences (Shima, Zubair et al. 2008).

We hoped to identify the protein that binds to the E-box (critical for pituitary-specific SF-1 expression), and the most likely candidate is the ubiquitous E2A, a basic helix-loop-helix protein (bHLH). bHLH proteins are transcription factors that play important roles in various developmental processes, including sex determination (Zheng, Wang et al. 2009). They contain a basic DNA binding domain and two helices that are involved in interaction with other proteins during the formation of homo- or hetero-dimers (Massari and Murre 2000). E2A heterodimerizes with other tissue-specific bHLH proteins before binding to DNA to regulate transcription. E2A has been extensively studied and is widely believed to play a central role in
transcriptional regulatory networks (Kee 2009). Although, E2A is expressed in pituitary, no studies account for the pituitary functions of E2A. Our current study indicates that E2A plays a transcriptional regulatory role in pituitary gonadotropes (Roberts, Steenbergen et al. 1993).

Inhibitors of DNA binding/differentiation are another type of bHLH proteins, which inhibit helix-loop-helix activators by acting as repressors (Ruzinova and Benezra 2003, Perk, Iavarone et al. 2005). To test whether Ids repress SF-1 activity through ECR3, we performed a co-transfection with ECR3, ECR3 mutant and increasing amounts of Id proteins. Out of three Ids tested Id1, Id2 and Id3 only Id2 and Id3 showed inhibitory effects on SF-1 promoter activity. As expected, co-transfection with the mutant ECR3 did not alter the activity of SF-1. Although, all Id proteins are closely related in structure and were thought to show similar binding affinity to other bHLH proteins, a recent study using gene targeting revealed the difference between these Id proteins. Id1 null mice were normal. However, Id2 null mice lacked lymph nodes and had severe defects in development of natural killer cell lineage, while Id3 null mice had defects in B-cell immune response (Lyden, Young et al. 1999, Pan, Sato et al. 1999, Yokota, Mori et al. 2001). Id2 and Id3, therefore have greater functional importance than Id1. The difference in the inhibitory effects of different Id proteins in the present study could be attributed to interactions of different Ids with E2A and its unidentified binding partner.

In summary, our study identified a highly evolutionarily conserved region namely ECR3. Further characterization of the region revealed that an E-box within ECR3 is essential for pituitary expression of SF-1. Additional studies will identify more or all of the regulatory sequences that are important for pituitary expression of SF-1 and increase our understanding on SF-1 gene regulation in pituitary.
MATERIALS AND METHODS

SF-1(-734+60)Luc vector preparation: The SF-1 promoter region from approximately -734 to +60 was amplified by polymerase chain reaction (PCR) from rat primary Sertoli cell DNA using primers shown in Table 1 and Bio-X-act DNA polymerase (Bioline USA Inc.MA, USA) and standard procedures. An XhoI site was incorporated into the 5’ primer and a HindIII site was incorporated into the 3’ primer. The resulting PCR product was digested with XhoI and HindIII and cloned into pGL3 basic (Promega).

Identification and cloning of ECRs:
ECRs were identified using a web-based sequence analysis tool, ECR browser (http://ecrbrowser.dcode.org/),(Ovcharenko, Nobrega et al. 2004). The ECRs were amplified by PCR from rat genomic DNA isolated from primary Sertoli cells using primers shown in Table 2 and Bio-X-act DNA polymerase (Bioline USA Inc.MA, USA) and standard procedures. The primers used contained sites for the restriction endonuclease KpnI. The amplified ECRs, following digestion with the KpnI enzyme, were cloned into SF-1(-734+60) Luc which contains the rat SF-1 promoter, from -734 to +60, directing expression of the firefly luciferase reporter. The ECR DNA fragments were cloned upstream of the SF-1 promoter insert. All clones were confirmed by sequence analysis.

Identification of potential transcription factor binding sites:
The web-based tool “rVISTA,” which combines TFBS predictions, sequence comparisons and cluster analysis to identify non-coding DNA regions that are evolutionarily conserved, was used to analyze ECR3 sequence for potential transcription factor binding sites (Ovcharenko, Nobrega et al. 2004).
Site directed mutagenesis:

For deletion mutagenesis, specific lengths (Highly conserved, Less conserved, 300, 363, 500, and 580 bp) of the sequences of ECR3 were PCR amplified using primers containing KpnI sites and cloned into SF-1(-734+60)luc, as described above. For PCR based site-directed mutagenesis, a HindIII site was introduced into the potential transcription factor binding sites (E box element, p53 and Elk1) and the mutant ECRs were amplified using primers containing KpnI sites and cloned into SF-1(-734+60)Luc as for the full length ECR3. All clones were confirmed by sequence analysis.

DNase I footprint and EMSA analysis:

The preparation of nuclear extracts and the generation of DNA probes were as described elsewhere (Lei and Heckert 2004, Hermann and Heckert 2005). DNase I footprinting assay using nuclear extracts from αT3 cells (containing 20 micrograms of protein) was performed as previously described (Lei and Heckert 2004). EMSAs were performed as previously described (Lei and Heckert 2004, Hermann and Heckert 2005).

ChIP

ChIP was performed as described (Hiroi, Christenson et al. 2004, Hermann and Heckert 2005). Cross-linked chromatin was prepared from αT3 gonadotrope cells. Antibodies were purchased from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA), and 5 μg each was used for immunoprecipitation: rabbit anti-USF1 IgG (sc-229), rabbit anti-USF2 IgG (sc-862), rabbit E2A.E12 (V-18) X (sc-349X), RNA polymerase II (sc-899), p300 (SC-6149), H3Ac (Millipore 06-599), H3K27me3 (Millipore 07-449), and normal rabbit IgG (sc-2027). PCR was performed
using 5 μl immunoprecipitated DNA or 1 μl input material, and primers were directed to sequences within the mouse SF-1 ECR3 loci. For mouse SF-1 ECR3, the primer set spanned the ECR3 region, ECR F: AGCAGCAGGGCTGTTTCC and ECR R: AGCTGAGGCTCCCTCCAC. Products were resolved by agarose gel electrophoresis.

**Cell Culture and Transfections:** Primary cultures of rat Sertoli and peritubular myoid cells were prepared, and transfections performed, as previously described (Karl and Griswold 1990, Heckert, Daggett et al. 1998, Chen and Heckert 2001). Preparation and transient transfection of mouse MA-10 Leydig cells was performed as described, except that cells were plated at a density of 52,000 cells/well (Ascoli 1981). Preparation and transient transfection of the αT3 gonadotrope cells were performed as described, with modifications (Windle, Weiner et al. 1990, Wolfe and Call 1999). Preparation and transient transfection of the MSC-1 cells were performed as described earlier (Heckert, Daggett et al. 1998). Co-transfection of Ids was performed as described, with modifications (Scherrer, Rice et al. 2002). Transfection experiments included TK-renilla as an internal control and the firefly and renilla luciferase activities were measured. The luciferase/renilla values of the vectors containing ECRs were made relative to the values for the SF-1 promoter alone.
### Table 1. Oligos used to amplify SF-1 Promoter (-734+60) region

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF-1 F</td>
<td>GGGCTCGAGATCCGTCTAGGCCAGTTCAG</td>
</tr>
<tr>
<td>SF-1 R</td>
<td>GGGAAGCTTCTATCGGGCTGTCAGGAACT</td>
</tr>
</tbody>
</table>

### Table 2. Oligos used to amplify the ECRs

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECR 1 F</td>
<td>GCGCGGTACCACTTCCAGTCCGCCTGCTCGT</td>
</tr>
<tr>
<td>ECR 1 R</td>
<td>GCGCGGTACCGGACTGGGACCCTTGCCGAG</td>
</tr>
<tr>
<td>ECR 2 F</td>
<td>GCGCGGTACCTCGGAGAGATGGTTTATTA</td>
</tr>
<tr>
<td>ECR 2 R</td>
<td>GCGCGGTACCCCTGGCTTGGGGTCCCTGGC</td>
</tr>
<tr>
<td>ECR 3 F</td>
<td>GCGCGGTACCGGTACATTGGCTGGCAGGTTGGGAT</td>
</tr>
<tr>
<td>ECR 3 R</td>
<td>GCGCGGTACCTGGAGCAGAAAATGAAT</td>
</tr>
<tr>
<td>ECR 4 F</td>
<td>GCGCGGTACCCTCTGCCAGGGATTTCCATTTAC</td>
</tr>
<tr>
<td>ECR 4 R</td>
<td>GCGCGGTACCCTCTCCAGAACGAAACCC</td>
</tr>
</tbody>
</table>

143
**Figure 1:** Alignment of NR5A1 region using ECR browser, depicting the Evolutionarily conserved regions (ECRs). In this figure, red peaks depict regions of non-coding evolutionarily conserved sequence, *NR5A1* and its neighboring gene *NR6A1* are annotated at the top (arrows indicate transcription direction), and conserved exons and untranslated sequences are marked as blue and yellow peaks, respectively.
Figure 2: Schematic diagram of the vector SF-1(-734+60) Luc and the transcriptional activity of each ECR. Schematic diagram of the vector SF-1(-734+60) Luc and the transcriptional activity of each ECR shown as determined in multiple cell types (noted to the right of the graph) by transient transfection analysis. The bar graph shows promoter activity as the ratio of firefly/renilla luciferase activity of each construct relative to that of the SF-1 promoter without the ECR and error bars represent the SD of mean.
**Figure 3:** Deletion mutants and analysis of transcriptional activity. Deletion mutants were generated by cloning varying lengths of highly conserved sequences (shown on the top of the diagram) in SF-1(-734/+60)Luc and their transcriptional activities tested individually in alpha T3 cell type and graphed (bottom).
Figure 4: Dnase Footprinting gel of ECR3 region. Diagram showing the identified binding sites, hypersensitive site and footprints on the sequence of ECR3 (left and below). A picture of the footprinting gel is shown along with a hypersensitive site, footprints and the associated sequence ladder. +NE and -NE represent with and without nuclear extract respectively.
**Figure 5:** Transient transfection analysis to check activity of potential transcription factor binding sites. Schematic diagram of the relative positions of the generated mutations and potential transcription factor binding sites within the ECR3 fragment (top). The activities of E-box, p53 and Elk 1 mutants were tested by transient transfection analysis and compared to that of the base vector (bottom).
Figure 6: Electrophoretic mobility shift assay (EMSA) of ECR3 region. Electrophoretic mobility shift assay (EMSA) of ECR3 was performed using αT3 nuclear extracts and a radiolabeled probe containing the ECR3 E box (bottom). Where indicated competitors were added to the reaction at a concentration 100X that of probe. Similarly, indicated lanes contain antibodies to basic-helix-loop-helix (bHLH) proteins (E2A, USF1 and USF2), transcription factors that bind to E box elements. The results are summarized at the top with a schematic diagram indicating E box binding of E2A and its binding partner, USF1 and USF2.
Figure 7: Evaluation of transcription factor binding by ChIP in αT3 cells. Interactions between E2A, USF1, USF2, RNA polymerase II, p300, H3Ac, H3K27me3 and ECR3, \textit{in vivo}, were evaluated using ChIP with cross linked chromatin from αT3 cells. Immunoprecipitations for each antibody were compared with normal rabbit IgG.
Figure 8: Co-transfection analysis to determine role of Id proteins. Negative regulators of bHLH proteins Id1, Id2, and Id3 were co transfected into αT3 cells with luciferase reporters driven by wild type (Trompouki, Bowman et al.) or E box mutant (mEbox) SF-1(-734+60)-ECR3. Relative activities were assessed and are shown below.
Chapter 6: Implications and Future Directions
**Implications**

This thesis delves in the transcriptional regulation of *Fshr* and *SF-1* that form an important component for the proper functioning of the HPA/HPG axis. It is now clear that regions beyond the promoter and its proximal elements are required for proper gene transcription. To identify these distal regulatory elements, sequence conservation was employed to identify genomic sequences that have remained unchanged over millions of years of evolution, and hence indicate that these regions are under evolutionary pressure to remain constant and harbor transcriptional factor binding regions. Chapter 2 of this thesis identifies highly conserved regions in the *Fshr* locus and potential binding sites for CTCF, a protein known to regulate transcription and chromatin architecture, and insinuates to a role of CTCF in transcriptional regulation of the *Fshr* gene. Chapter 4 and Chapter 5 of this thesis deals with conserved elements identified at SF-1 locus. A long non-coding RNA labeled *Fast* was found transcribed in opposite orientation of SF-1 and displayed similar tissue expression and regulatory profile as of SF-1. Knockdown and over-expression of the *Fast* transcript did not modulate mRNA or protein level of SF-1, indicating that the very act of transcription of the non-coding RNA might contribute to the opening of the chromatin for transcription of SF-1. Comparative sequence analysis identified an evolutionary conserved non-coding region, ECR3, approximately 4kb from the transcriptional start site of SF-1. ECR3 were cloned upstream of a -734/+60 SF-1 promoter directing expression of a luciferase reporter and its transcriptional activity assessed in various cell types by transient transfection analysis. Transient transfection analysis identified an ECR in the SF-1 locus, which regulates expression of SF-1 in the pituitary. Further characterization of the ECR region identified an E-box and binding of USF-1 and USF-2, along with E2A transcription factor,
confirmed by EMSA and ChIP. In addition, co-transfection studies identified Inhibitors of DNA binding, Id2 and Id3 that interfere with bHLH proteins to downregulate SF-1 transcription.

These studies highlight the importance of conserved elements and sequence conservation as a tool to identify regulatory elements distal to the promoter and lay the groundwork for further characterization of these identified regions and the role they might play in modulating gene transcription.

Future directions

Defining CTCF regions and its role in regulating Fshr transcription

Initiation of transcription starts from the promoter and involves recruitment of general transcriptional factors (Lee and Young 2000, Butler and Kadonaga 2002). Since the characterization of Fshr, twenty years ago, we have come a long way in identifying transcriptional mechanisms that regulate gene expression (Heckert, Daley et al. 1992, Gromoll, Pekel et al. 1996, George, Dille et al. 2011). The extensive characterization of the Fshr gene, while underlining the importance of the promoter and its proximal regions, fall short in revealing cell-specific expression of the receptor. Multiple transgenic reporter studies expressing the promoter and large swathe of the surrounding region were unable to replicate the cell-specific expression, supporting the theory that distal regulatory elements are required for proper spatio-temporal expression (Heckert, Sawadogo et al. 2000, Nordhoff, Gromoll et al. 2003, Hermann, Hornbaker et al. 2007). However, identification of distal regulatory elements possess great challenge as they can act over large distances and the Fshr gene is located in a long genomic

To facilitate the search of these regulatory elements comparative genomics and web based prediction tools were employed as described in Chapter 2. Among the varied distal regulatory elements known to regulate gene transcription, we looked at the ubiquitously expressed “multivalent” protein CTCF, with known multiple functions, including transcription (Filippova, Fagerlie et al. 1996, Phillips and Corces 2009). Evaluation of these evolutionary conserved regions (ECR) in the CTCF prediction software led to the identification of six ECRs; ECR1, ECR1d, ECR1f, ECR2 and ECR15, with high probability of CTCF binding. Depletion of CTCF in rat granulosa cells increased transcription of $Fshr$ two-fold while the $Lhcgr$ remained unchanged; indicating that depletion of CTCF removed its repressor effect on $Fshr$ transcription. While these findings implicate CTCF in regulating $Fshr$ transcription, a series of experiments will have to be performed to evaluate the location and the mechanism involved.

A good starting point will be to identify if CTCF binds to the ECR regions using electrophoretic mobility shift assay (EMSA) and further confirmation by chromatin immunoprecipitation (ChIP). ChiP-chip, a technique that combines ChIP with microarray technology (chip), can be used to identify a number of important DNA-protein interactions leading to important discoveries on transcriptional regulation (Ren, Robert et al. 2000). This technology will help identify CTCF binding regions on the $Fshr$ locus.

Colony assay is an established assay to determine enhancer-blocking activity of insulator elements (Gombert, Farris et al. 2003). This assay employs a bacterial neomycin resistance gene (neo) reporter construct, stably integrated into the genomic site and measures the ability of the proposed insulator element to repress or stimulate transcription of selectable marker construct.
As shown in Figure 1, the insulator element will be placed between the enhancer and the promoter region or upstream of the enhancer. The number of colonies obtained after transfection into mouse transformed granulosa cells, GRMO2, will be proportional to the number of cells expressing the neomycin resistance gene (Neo') at levels sufficient to confer resistance to the drug (G418). CMV promoter and the enhancer run the Neo' transcription. In absence of the enhancer, there will be reduced number of colonies. To test the ability of the construct to measure enhancer blocking activity, Drosophila scs and scs’ boundary elements will be introduced at E-scs-P-neo-scs’ and also as scs-E-P-Neo-scs’. Insertion of the scs should insulate the promoter activity and reduce the number of colonies. However, when the scs will be inserted upstream of the enhancer, the colony number will remain high. Insertion of the scs element will increase the distance between the enhancer and the promoter and this might be the cause of lower number of colonies.
To verify that the reduction in colony number seen is a result of enhancer blocking and not the increased distance between the enhancer and promoter, a similar length control (C) fragment will be inserted between the enhancer and the promoter. A reduction in the number of colonies will indicate that enhancement is blocked by the control fragment. This control is necessary to confirm that the reduction in the number of colonies is due to enhancer blocking and not simply a distance effect. A reduction in the number of colonies will indicate that enhancement is blocked by the control fragment. Another positive control is the Drosophila scs and scs’ boundary elements, which will affect colonies when placed between the enhancer and the Promoter-Neo construct. cHS4 is a known insulator, and it will have a similar effect as seen with Drosophila scs. This insulator effect shall be relieved when cHS4 is placed upstream of the enhancer. To check if the regions identified function as insulators, they will be placed in both orientations and in both upstream and in between the enhancer and promoter-Neo construct to check if they block enhancer activity, which will be measured by the number of resistant colonies.

**Figure 1:** A schematic diagram depicting colony assay. As shown, the CMV promoter and neomycin resistance gene (neo) reporter construct denotes control and bar graph denotes expected trend for each reporter construct. Enhancer element placed upstream of the Promoter-Neo construct should yield in increased G418 resistant mouse GRMO2 cell colonies. C; control should not affect the colony number. *Drosophila* scs and scs’ boundary elements will affect colonies when placed between the enhancer and the Promoter-Neo construct and will be another positive control. cHS4 is a known insulator, when placed between the enhancer and promoter should have a similar effect as seen with *Drosophila* scs. This insulator effect shall be relieved when cHS4 will be placed upstream of the enhancer. To check if the regions identified function as insulators, they will be placed in both orientations and both upstream and in between the enhancer and promoter-Neo construct to check if they block enhancer activity, which will be measured by the number of resistant colonies.
ascertain the enhancer blocking property of the insulator when placed between the CMV enhancer and the promoter. As a control, 1.2-kb chicken β-globin HS4 insulator sequence will be placed between the enhancer and the promoter. As HS4 is a previously reported enhancer-blocking element, it will reduce the number of colonies (Chung, Bell et al. 1997).

Chromosome conformation capture (3C) analysis is another method to evaluate the interaction between distal regulatory cis-acting elements and the promoter (Dekker, Rippe et al. 2002). In 3C, transient chromatin interactions are stabilized by formaldehyde cross-linking, followed by restriction digestion and intramolecular ligation and subsequent analysis of the ligation products by PCR. 3C has been used successfully to analyze spatial organization of small genomic domains such as mouse interferon gamma gene domain to larger genomic areas as in the case of mammalian alpha and beta globin gene domains (Palstra, Tolhuis et al. 2003, Eivazova and Aune 2004, Zhou, Xin et al. 2006). These techniques can identify CTCF binding regions on the Fshr locus, determine if there is a physical interaction between these distal regulatory elements and the Fshr promoter, and identify its role in modulating transcriptional regulation of Fshr.

**Identifying functional role of Fast**

The long noncoding RNA, *Fast*, identified at the *Nr5a1* locus did not affect transcription or translation of SF-1 indicating that it could function by acting in trans as in the case of HOX antisense intergenic RNA (*HOTAIR*). *HOTAIR* does not affect at the site of its expression at the HOX C locus, but interacts with PRC2 and the (LSD1)–CoReST–ReST complex and represses HOXD and several other loci (Rinn, Kertesz et al. 2007, Tsai, Manor et al. 2010). To identify targets of *Fast*, mU6 ShRNA and *Fast* shRNA plasmids will be transfected into MA-10 Leydig cell line, selected for neomycin resistance, expanded, and RNA extracted using Trizol. RT-PCR will be utilized to analyze the efficacy of the RNAi construct, and constructs that decrease
expression of Fast by 80-90% will be used for further experimentation. Six samples will be used, three for Fast knockdown and three controls and expression profiling can be performed using Affymetrix Gene Chip Mouse Genome 430 2.0 platform. Ingenuity Pathway Analysis (IPA) can then be utilized to analyze the raw data and promising gene candidates can be identified on basis of minimum two-fold difference and p-value less than 0.05. Once identified, these candidate genes can be confirmed by utilizing RT-PCR.

Fast shares similar expression and regulatory functions as SF-1 indicating a functional role of the non-coding RNA, which it might relay through interaction with cellular proteins. It has been frequently observed that a number of lncRNA recruit or interact with protein complexes to bring about their associated function. The lncRNA, RepA recruits the Polycomb Repressive Complex 2 (PRC2) to the future inactivated X chromosome resulting in trimethylation of Histone 3 lysine27 (H3K27me3)(Zhao, Sun et al. 2008). Similarly, lncRNA Air and Kcnqlot1 interact with H3K9 histone methyltransferases G9a to epigenetically silence transcription. Identification of the proteins that interact with the non-coding RNA will help elucidate its functional significance and any potential mechanism involved in regulation (Nagano, Mitchell et al. 2008, Pandey, Mondal et al. 2008).

In order to identify proteins interacting with Fast, a biochemical approach can be taken as described previously for purification of proteins interacting with non-coding RNA, NRON (Willingham, Orth et al. 2005). Full length Fast will be cloned into pGEM5z and in vitro transcribed, following which a 3-hairpin RNA (MS2 loop) epitope bound to a MS2-MBP will be added to the 3’ termini of the full length Fast RNA (Figure 2). MS2 is a bacteriophage coat protein that is bound to the maltose binding protein (MBP). MS2-MBP, Fast and a nonspecific (NS) RNA without the hairpin loop
Figure 2: Purification of RNA bound proteins by MS2-MBP pull down method. The *Fast* RNA will be tagged with the three MS2 stem loops and then bound with the MS2:MBP fusion protein. Nuclear extracts are now incubated with the *Fast* RNA-MS2 stem loop and MS2:MBP fusion will be then affinity selected by binding to amylose resin and eluted.
will be prebound and incubated with protein extract from mice granulosa cells (*Fast* expressing) and NIH3T3 cells (*Fast* non-expressing). Samples will then be washed and eluted in an amylose binding resin. The resulting proteins can then be concentrated, resolved, analyzed by Mass Spectrometry and peptide sequence identified.

Over the past decade, the role of non-coding RNA in gene transcription and regulation has been appreciated. The so called “dark matter” has been resolved to play important biological role in proper regulation of the cellular machinery and our appreciation grows as new functions are described. The identification of a long non-coding RNA transcribed from the same locus as SF-1, in the opposite direction, with similar expression profile and tissue-specific transcript variants indicates a functional role. The data generated will provide insight to the functional role of the non-coding RNA and the role it may play in adrenal and gonadal development and will reveal a new chapter both in the role of the non-coding RNA and of SF-1 function.

**Further characterization of ECR3 and role of identified transcription factors**

Comparative genome sequencing between chicken, mouse, and human led to the identification and characterization of ECR3 located upstream of SF-1. An E-box was identified and it was found to be bound by upstream stimulatory factors 1 and 2 (USF1 and USF2), and transcription factor E2A, a ubiquitous bHLH protein. bHLH proteins contain a basic DNA binding domain and two helices that are involved in interaction with other proteins during the formation of homo- or hetero-dimers and play important roles in various developmental processes, including sex determination and is widely believed to play a central role in transcriptional regulatory networks (Massari and Murre 2000, Kee 2009, Zheng, Wang et al. 2009). Although, E2A is expressed in pituitary, no studies account for the pituitary functions of E2A. Our current study
indicates that E2A plays a transcriptional regulatory role in pituitary gonadotropes (Roberts, Steenbergen et al. 1993). The bHLH protein, E2A hetero-dimerizes with other tissue-specific bHLH proteins before binding to DNA to regulate transcription. This protein can be identified by employing proteomics of isolated chromatin segments (PICh), that allows isolation and identification of proteins bound at a chromatin locus (Dejardin and Kingston 2009). PICh involves fixing cells with formaldehyde, followed by solubilization of chromatin, hybridization of a specific probe, capture using magnetic beads, elution of hybrids and identifying associated proteins by Mass spectrometric analysis. Once the proteins are identified, they can be confirmed by standard ChIP. PICh has been used successfully to identify binding of orphan nuclear receptor COUP-TF2 at telomeres (Dejardin and Kingston 2009).

Earlier work in mouse and rat, identified an E box motif in the proximal promoter region of SF-1 and it was found to be bound by USF1 and USF2 (Harris and Mellon 1998, Daggett, Rice et al. 2000). In the current study, another E-box was identified in the distal regions upstream of SF-1. ECR3 lies ~4kb upstream of the SF-1 promoter region, indicating that distal looping is involved and the two E-box might work in concert to regulate SF-1 transcription in the gonadotrope pituitary. 3C can be employed, as described earlier, to evaluate the interaction between ECR3 and the SF-1 promoter.

This is the first time that E2A has been implicated to play a role in transcriptional regulation of SF-1 in the pituitary gonadotropes. To identify the role of E2A in the transcriptional regulation of SF-1 in the pituitary gonadotropes, α-T3 cells stably transfected with shRNA directed against E2A, can be analyzed to identify the function of E2A on SF-1 transcription.
In summary, our study identified a highly evolutionarily conserved region namely ECR3 and an E-box within ECR3 that is essential for pituitary expression of SF-1. Additional studies will identify more or all of the regulatory sequences that are important for pituitary expression of SF-1 and increase our understanding on SF-1 gene regulation in pituitary.
REFERENCES


adrenal glands and gonads but express P450 side-chain-cleavage enzyme in the placenta and have normal embryonic serum levels of corticosteroids." Proc Natl Acad Sci U S A 92(24): 10939-10943.


enzyme expression in cycling human ovaries and sex cord stromal tumors." J Clin Endocrinol Metab 80(9): 2815-2821.


202


