January 2017

The Molecular Mechanisms Of Sex Determination In Vertebrates

Lei Guo

THE MOLECULAR MECHANISMS OF SEX DETERMINATION IN VERTEBRATES

by

Lei Guo
Bachelor of Science, Xinjiang Normal University, China, 2004
Master of Science, Shanghai Normal University, China, 2010

A Dissertation

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota

May

2017

This dissertation, submitted by Lei Guo in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Turk Rhen, Chairperson

Diane Darland

Manu Manu

Jeffery Carmichael

Peter Meberg

Roxanne Vaughan

This dissertation is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

Grant McGimpsey
Dean of the School of Graduate Studies

April 19, 2017
Date

PERMISSION

Title The Molecular Mechanisms of Sex Determination in Vertebrates

Department Biology

Degree Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Lei Guo
13 April 2017

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to my advisor, Dr. Turk Rhen, and my committee, Dr. Diane Darland, Dr. Manu Manu, Dr. Jeffery Carmichael, Dr. Peter Meberg and Dr. Roxanne Vaughan for their guidance and support during my time in the doctoral program at the University of North Dakota. Thanks to graduate students and postdocs of Dr. Rhen's lab for comments and insights on this work. Thanks to the Biology Department and graduate school of the University of North Dakota, NSF and NIH for funding this research. Thanks to the faculty and staff of the biology Department University of North Dakota for assisting my research. Finally, I thank my wife Ling Guo and my son Grayson who have been supporting me by my side throughout this work.

ABSTRACT

Many reptiles display temperature-dependent sex determination (TSD), in which the primary sex is determined by incubation temperatures rather than sex chromosomes. However, temperature is not the only factor that play critical roles in sex determination in the species with TSD. Previous studies in the snapping turtle, a species with TSD, showed that dihydrotestosterone (DHT) induces ovary development at temperatures that normally produce males or mixed sex ratios. In addition, the feminizing effect of DHT was found to be associated with increased expression of the ovary-determining gene *Foxl2*, suggesting a potential androgen-*Foxl2* regulatory mechanism. This dissertation aims to clarify the molecular mechanisms underlying TSD in several aspects. First, determine the role of androgen in TSD; second, identify novel thermosensitive genes involved in TSD and lastly, reconstruct gene regulatory networks underlying sex determination.

To test the hypothetical androgen-*Foxl2* interaction, I cloned the proximal promoter (1.6 kb) and coding sequence for snapping turtle *Foxl2* (t*Foxl2*) in frame with mCherry, a red fluorescent protein. The t*Foxl2*-mCherry fusion plasmid or mCherry plasmid were stably transfected into mouse KK1 granulosa cells. Although expression of t*Foxl2*-mCherry was not affected by androgen treatment in KK1 cells, androgen inhibited expression of the endogenous mouse *Foxl2* gene, suggesting the androgen-*Foxl2* interaction does exist but it differs between species. We also found t*Foxl2*-

mCherry potentiated low dose DHT effects on aromatase expression, which has not been reported in any other studies.

To identify novel sex-determining genes in TSD, I first *de novo* assembled and annotated the transcriptome of the snapping turtle using next-generation sequencing (NGS) and then performed RNA-seq analyses on the newly assembled reference transcriptome. With the differential gene expression analyses, I identified 293 thermosensitive genes. Among these genes, I find AEBP2, JARID2, and KDM6B of particular interest because these genes could influence expression of many other genes via epigenetic modifications.

To further investigate the molecular mechanisms underlying sex determination, I reconstructed gene regulatory networks using an entropy based network reconstructing algorithm – ARACNE with public microarray experiments in mouse gonads. The subsequent hub gene analyses revealed the basic molecular pathways underlying gonadal development and the master regulator analyses identified 110 candidate sex-determining genes including both known sex-determining genes and novel candidate genes.

My findings demonstrate that androgens can influence expression of key ovarian genes but further studies are needed to understand the androgen signaling in TSD. Furthermore, my study provides a first description of the snapping turtle transcriptome and the effects of temperature on transcriptome-wide patterns of gene expression during the TSP. In addition, hub genes and master regulators identified for mammalian gonad determination will guide the direction of future studies in the field of sex determination. However, additional studies are needed to validate the computational findings.

CHAPTER I

SEX DETERMINATION AND DIFFERENTIATION IN AMINIOTIC VERTEBRTES

Sexual dimorphism, where males and females of the same species exhibit different characteristics, has attracted researchers' attention for centuries. Phenotypes, behaviors and even diseases of animals can diverge enormously between the sexes. For example, *Drosophila melanogaster* body size, wing shape, sensory bristles, and color are sexually dimorphic (David et al., 2011). In birds, feather patterns, wing size and songs differ between males and females (Owens and Hartley, 1998). In humans, tooth size, amount of subcutaneous fat and muscle fibers, pre/postnatal hormone levels, growth rate and diseases vary between males and females. In addition, reproductive behaviors, such as courtship, sexual behavior, parturition, and the care of young, are sexually dimorphic in mammals, amphibians, birds and insects. A better understanding of sex differences among species helps to elucidate evolution and find new treatments for disorders of sexual development in humans. To study the mechanisms underlying sexual dimorphisms, biologists must investigate sexual differentiation when it starts early in embryogenesis.

Sex determination and sexual differentiation occur sequentially during vertebrate embryogenesis. Sex determination in vertebrates involves commitment of the undifferentiated gonads to develop as sexually dimorphic ovaries and testes. Sexual

differentiation is a developmental process in which traits diverge between male and female after sex has been determined. Numerous studies in vertebrates have shown that sex can be determined either by chromosomes (genotypic sex determination, GSD) or environmental factors, such as temperature or social variables (environment sex determination, ESD) (Gamble and Zarkower, 2012) (Figure 1).

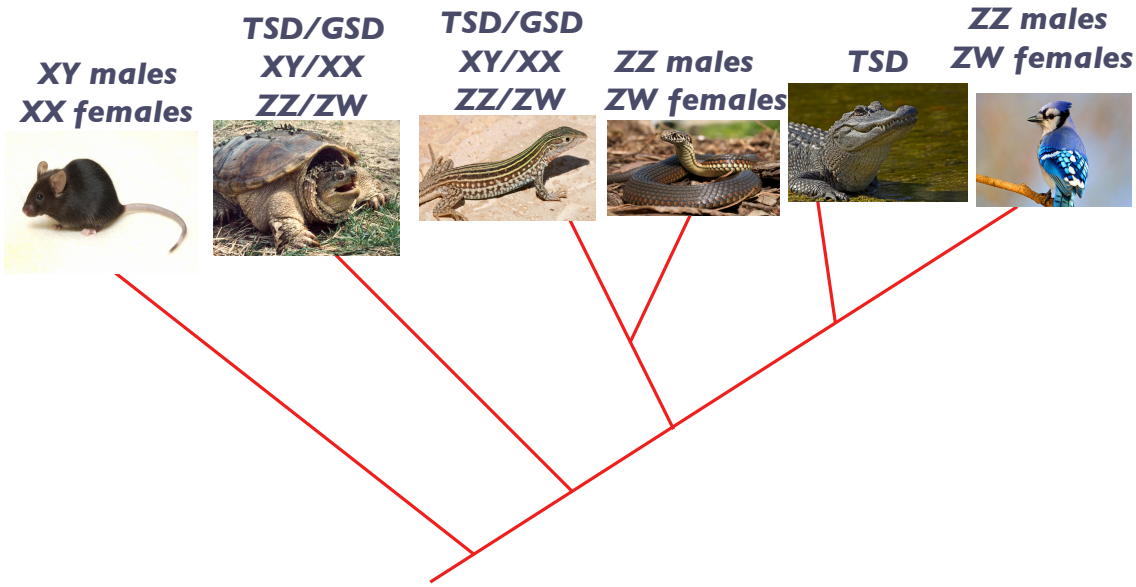


Figure 1. Sex determination in vertebrates varies among species. From left to right, sex in mammals is determined by sex chromosome X and Y; sex in turtles is determined either by temperature (TSD) or by genotype (GSD); sex in lizards can be determined by temperature, genotype, sex chromosome X and Y or sex chromosome Z and W; sex in snakes is determined by sex chromosome Z and W; sex in alligators is determined by temperature; sex in birds is determined by sex chromosome Z and W. The evolution of sex-determining mechanisms is not displayed in this phylogenetic tree. (From left to right, the pictures are from http://www.yourgenome.org/sites/default/files/images/photos/Black%20mouse_Credit_Wellcome%20Library,%20London_cropped.jpg, http://www.marshall.edu/herp/images/SNAPPER.JPG, http://thehigherlearning.com/wp-content/uploads/2014/07/whiptail-lizard.png, https://aos.iacpublishinglabs.com/question/aq/700px-394px/moth-balls-keep-snakes-away_d731c368d3991a0e.jpg?domain=cx.aos.ask.com, http://refugeassociation.org/wp-content/uploads/2011/08/alligator-ding-darling-michael-dougherty.jpg, http://d2fbmjy3x0sdua.cloudfront.net/cdn/farfuture/xX2dO2IN71t0tfGOITDQ0HSLNOml6xiRu_z3MU6Xx5M/mtime:1486669862/sites/default/files/styles/engagement_card/public/sfw_apa_2013_28342_232388_briankushner_blue_jay_kk_high.jpg?itok=ttMfUhUu)

In mammals, birds and some reptiles, sex is determined by heteromorphic chromosomes (XY for males and XX for females in mammals; ZW for females and ZZ for males in birds and snakes). It is worth noting that Z and W chromosomes in birds and snakes are not homologous but analogous. According to Ohno's law, sex chromosomes derive from autosomes that acquire a new sex-determining gene. Sexually antagonistic selection on genes near the new sex-determining locus favors suppression of recombination. This leads to linkage disequilibrium between the sex-determining gene and alleles that are favored in the corresponding sex. For instance, a male-determining allele would be linked to alleles that increase male fitness. Depletion of heterochromatin accounts for the different size of sex chromosomes (Modi and Crews, 2005). For some animals that don't have distinct sex chromosomes, sex is determined by environmental factors. Temperature is one of the most common environmental factors involved in sex determination. This is known as temperature-dependent sex determination (TSD) and is observed primarily in reptiles, such as lizards, turtles and crocodilians (Shoemaker and Crews, 2009).

Sex-determining mechanisms in vertebrates show little conservation in invertebrates (Figure 2).

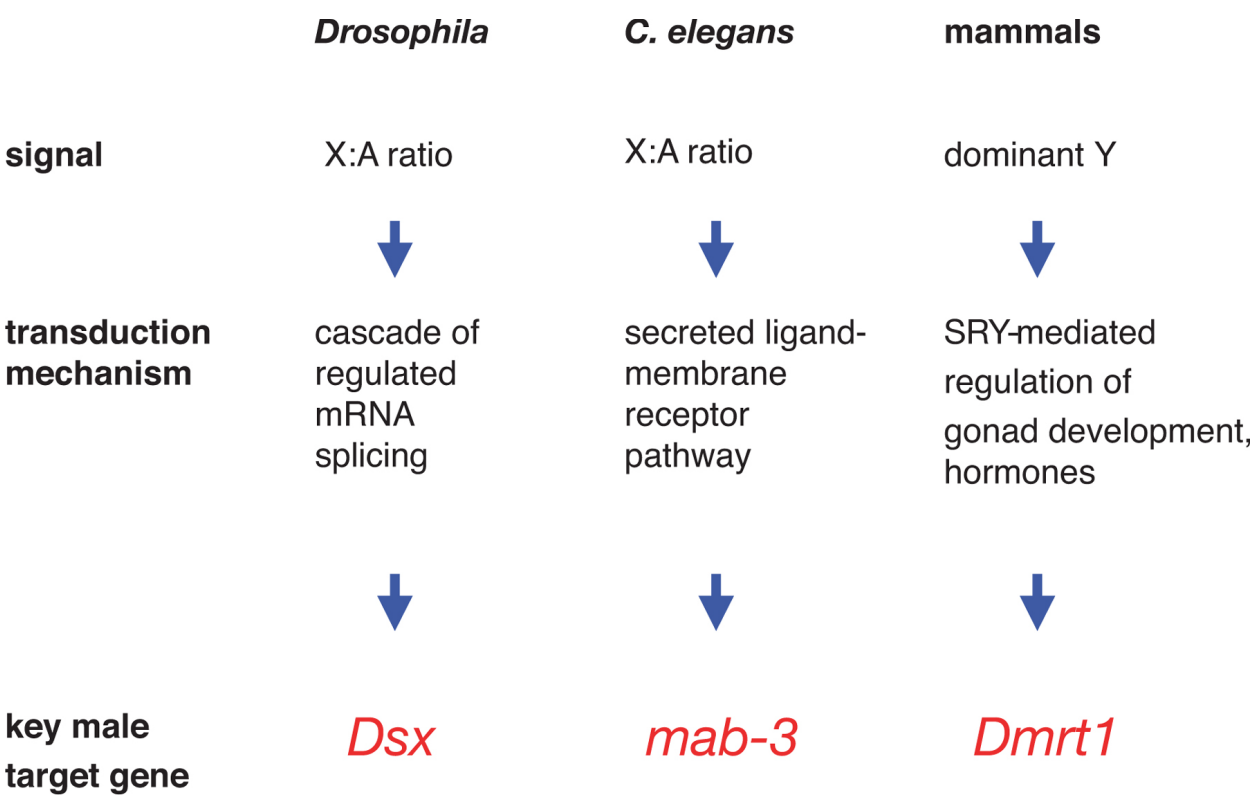


Figure 2. Sex determination in Drosophila, C. elegans *and mammals (Haag, 2005). In* Drosophila *and* C. elegans*, sex is determined by the ratio of X chromosome to autosome while in mammals, sex is determined by sex chromosomes. Only key male genes are showed in this figure.*

In *Drosophila melanogaster*, sex is determined by the ratio of X chromosomes to autosomes (Parkhurst et al., 1990). Individual *Drosophila* with high X chromosome to autosome ratios activate their master sex-determining gene *Sxl* (Sawanth et al., 2016). In contrast, in mammals, the master gene for sex determination is *Sry* (sex-determining region of the Y) on the Y chromosome, which is not found in *Drosophila*. The sex-determining gene *SRY* is also completely different from *sxl* in *Drosophila* (Sinclair et al., 1990). The downstream targets of master sex-determining genes between *Drosophila* and mice differ as well. The direct target of *Sxl* is *tra* which splices *dsx* into a female specific form (Valcárcel et al., 1993) while the direct target of *Sry* is *Sox9* (Sekido et al., 2004) which is not found in *Drosophila melanogaster.* Although a homolog of *dsx* named

Dmrt1 is found in mice, *Dmrt1* doesn't appear to be involved in primary sex determination like *dsx* in *Drosophila melanogaster* (Raymond et al., 2000). Huge differences in sex-determining mechanisms between phyla make investigation of the evolution of sex-determining mechanisms challenging.

Among vertebrates, reptiles are suitable models for studying the interaction between environment and sexual development as well as the interplay between different genes and cellular events during sexual development. The reason lies in the special evolutionary position that reptiles occupy as sister groups to mammals and birds. In reptiles, sex is determined either genotypically (GSD) or environmentally (ESD) or by both mechanisms. Some turtles, lizards and all snakes exhibit GSD while other reptiles employ TSD (Angelopoulou et al., 2012; Sarre et al., 2004). In TSD reptiles, the temperature sensitivity of the gonad during development varies among species. For example, in alligators, low (30°C) and high (35°C) incubation temperatures produce females while intermediate temperatures (32.5°C~33°C) produce males (Lance et al., 2000; Lang and Andrews, 1994). In contrast, in snapping turtles, low temperatures (23°C~27°C) produce males while high temperatures ($\geq$29.5°C) and intermediate temperatures (28.2) produce a roughly 1:1 mixed sex ratio (Lang and Andrews, 1994; Rhen and Lang, 1998; Yntema, 1979). Of note, TSD species are not sensitive to temperature throughout gonadal development. Sex determination only occurs in a specific developmental window, called the thermosensitive period (TSP), which also varies among TSD species (Bull, 1987; Burke and Calichio, 2014; Pieau and Dorizzi, 1981; Siroski et al., 2007; Yntema, 1979). This intriguing process has been intensively investigated, but the mechanism underlying TSD remains unknown.

Several critical cellular events have been distinguished in the timeline of gonadogenesis in TSD species. First, during the bipotential gonad phase, individuals can become either sex. Second, unknown temperature sensitive molecules initiate determination of gonad fate (Merchant-Larios and Díaz-Hernández, 2013). Sex determination occurs before the bipotential gonads start to differentiate or at the earliest stages of differentiation depending on species (Rhen and Schroeder, 2010). Sex is determined and cannot be reversed at the end of the TSP. Differentiation of several sets of cells, such as Sertoli cells, germ cells, peritubular myoid cells and Leydig cells, occurs after testis fate has been determined. When female fate has been determined, epithelial and germ cells proliferate, leading to the thickening of gonadal cortex and the differentiation of theca cells and granulosa cells (Merchant-Larios and Díaz-Hernández, 2013).

Gonad Morphogenesis in Vertebrates

Divergent sex-determining mechanisms converge towards the same end. In all vertebrates, regardless the sex-determining mechanisms they adopt, testes and ovaries develop from a bipotential primordium that is morphologically indistinguishable between the sexes. The bipotential gonads, or genital ridges, have the potential to develop into either testes or ovaries. The genital ridge consists of an outer cortex and an inner medulla. Under the influence of testis-determining genes, the inner medullary region grows and differentiates into testes whilst the outer cortex regresses (Figure 3).

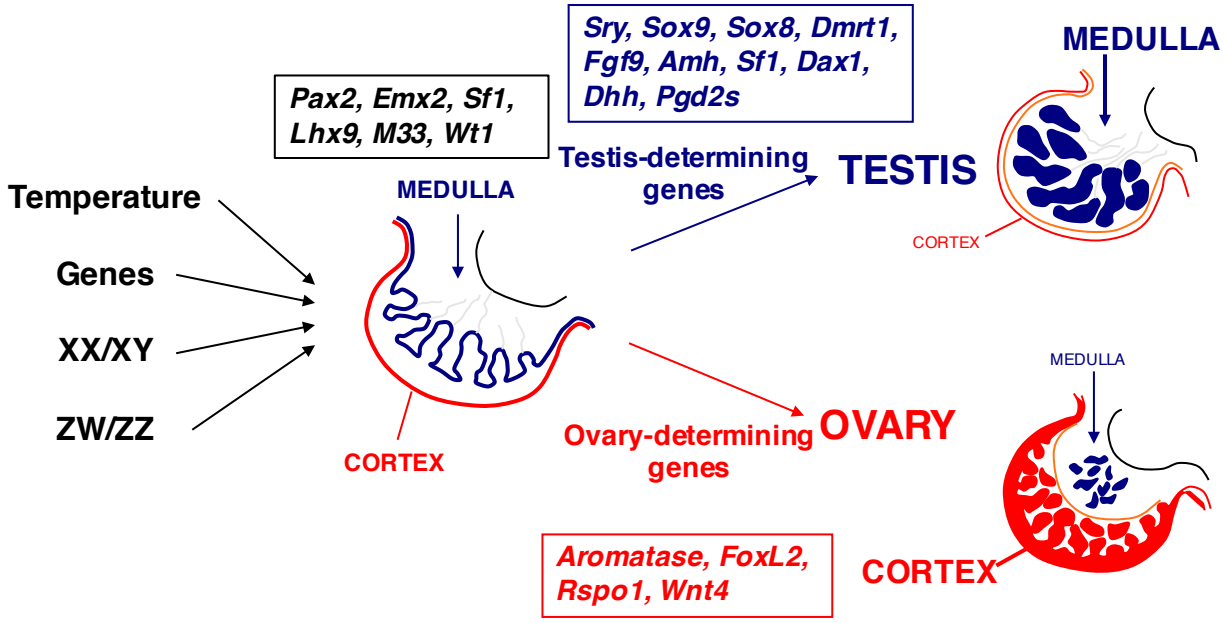


Figure 3. Sexual differentiation of gonad in vertebrates. From left to right, environmental factors, genes or sex chromosomes determine the differentiation of bipotential gonad (middle) either to testis or to ovary by initiating different signaling pathways. Genes in black font are involved in the formation of the bipotential gonad; genes in the blue box are testis specific; genes in the red box are ovary specific.

During testis development, Sertoli cells are the first to differentiate. These cells surround germ cells and adhere to each other to form the seminiferous cords. Meanwhile, steroidogenic Leydig cells and a functional vasculature start to differentiate in the interstitial space of the testis to produce and export hormones (Brennan and Capel, 2004). In XX gonads, the outer cortex of the genital ridge grows and differentiates into ovaries and the inner medulla regresses under the influence of ovary-determining genes. During ovarian development, oocytes, which are derived from primordial germ cells, are surrounded by somatic granulosa cells and the extracellular matrix to for follicles. As the follicle develops, theca cells are recruited for hormone production (Sarraj and Drummond, 2012). After sex determination and gonadal differentiation, testes and ovaries release sex hormones that regulate development of the reproductive tract, brain

and all other non-gonadal tissues. For example, testosterone secreted by Leydig cells promotes survival and differentiation of the Wolffian ducts into the male internal reproductive tract. Sertoli cells secrete AMH, which triggers regression of the Mullerian ducts and loss of the female internal reproductive tract. Estrogen secreted by theca cells and granulosa cells in ovaries promotes the development of female reproductive structures.

Molecular Models of Sex Determination in Vertebrates

After the discovery of the master switch in mammals (i.e., *Sry* gene), sex determination was thought of as an active process of testis determination while ovarian differentiation was a "default" pathway (Ottolenghi et al., 2007b). This model was based on studies in which *SRY* was demonstrated to be both necessary and sufficient to initiate the testis determination (Koopman et al. 1990). However, this male-determining pathway seems to be antagonized by some ovarian genes (e.g., *Wnt4, Foxl2*), indicating there may be a master ovary-determining gene that can switch the male pathway to the female pathway just as *SRY* does in males (Vainio et al., 1999). Thus, an ovarian determinant (Od) located on the X chromosome or an autosome was postulated to initiate ovarian determination by activating its target genes (Eicher and Washburn, 1986). A decade later, McElreavey et al. indicated that an anti-testis activity (Z) is necessary for ovarian determination (McElreavey et al., 1993). Therefore, an Od/Z model was established, i.e., ovarian development requires not only ovary-determining genes (Od) but also testis-repressing genes (Z). The Od/Z model is supported by the fact that mutation of some ovarian genes, such as *Dax1*, *Wnt4*, *Rspo1* and *Foxl2*, cause female to male sex reversal in mice (Ottolenghi et al., 2007b).

In the Od/Z model, maleness is determined in a "default" way. It takes the primary sex-determining genes *Wnt4* and *Foxl2* as "Z genes". Double knockout of *Wnt4* or knockout of *Foxl2* results in perinatal sex reversal in somatic cells of ovary in mammalian model, indicating *Wnt4* and *Foxl2* may compensate for each other or they extend the bipotential status of gonads (Ottolenghi et al., 2007b). Due to its function of maintaining gonadal vasculature in both testes and ovaries, *Wnt4* is more like a gene crucial to the bipotential gonad rather than an anti-testis gene, which leaves *Foxl2* as the best candidate for a "Z gene".

Evidence supporting *Foxl2* as the best candidate for "Z gene" comes from the studies of ovarian failure in mammals due to abnormal expression of *Foxl2*. Heterozygous mutation of *Foxl2* leads to Blepharophimosis, ptosis, and epicanthus inversus syndrome (BPES) and ovarian failure in humans; homozygous mutation of *Foxl2* leads to sex reversal in mice and goats (Ottolenghi et al., 2007b). Mutation in *Foxl2* coding sequence or in *Foxl2* cis-regulatory regions (some are 100 ~ 200 kb upstream/downstream of the coding sequence) leads to BPES I or BPES II in both human and mouse. In BPES I, craniofacial abnormalities and premature ovarian failure (POF) occur and in BPES II patients are infertile (Uhlenhaut and Treier, 2006). Uhlenhaut and Treier (2006) indicated that *Foxl2* is the only ovarian gene found so far that antagonizes male-determining genes and maintains high expression throughout a female's life. Furthermore, *SRY* and *Sox9* are not needed for testis maintenance while *Foxl2* is required to maintain the ovary and to antagonize testis-specific genes. *Foxl2* activates gonadotropin releasing hormone receptor (*Gnrhr*) and represses steroidogenic acute regulatory gene (StAR) which controls the rate limiting step of steroidogenesis (Cheng et

al., 2013; Pisarska et al., 2004). By repressing StAR, *Foxl2* prevents premature follicle development. *Foxl2* is also able to repress the testis-specific gene *sox9*, because its expression increases when *Foxl2* is absent. Thus, *Foxl2* is now considered one of the major female-determining genes and a good "Z factor" candidate.

However, recent studies have demonstrated neither of the models alone is sufficient to explain the mechanisms of sex determination by discovering a Z gene counterpart – *Dmrt1* in testis development. *Dmrt1* activates *Sox9* and *Sox8* or represses Wnt4 and *Foxl2* in postnatal testes by binding near these genes (Matson et al., 2011). The deletion of *Dmrt1* in the developing gonad results in the failure of Sertoli cell differentiation while the over expression of *Dmrt1* leads to female-to-male sex reversal (Raymond et al., 2000; Zhao et al., 2015). A reasonable explanation that reconciles these two models is that interactions among mutually antagonistic genes determine sex. Excess or insufficient activity of the antagonistic sex-determining genes will tip the balance towards the opposite sex.

Molecular and Cellular Events Underlying Sex Determination in Reptiles

Homologs of mammalian sex-determining genes are primary candidates to investigate in reptile sex determination, although solely relying on the discovery of sex-determining genes in mammals may slow the study of sex determination in reptiles. This approach can also be misleading, as some sex-determining genes in mammals may not be related to sex determination in reptiles. Genes that are differentially expressed between the sexes during mammalian sexual development, such as *Sox9, Sox8, Fgf9, Dmrt1, Foxl2*, etc., were hypothesized to be part of the gene network underlying TSD (Lance et al., 2004; Rhen and Schroeder, 2010; Shoemaker and Crews, 2009). Data collected from

different studies need to be integrated because the timing of expression of these putative sex-determining genes varies among TSD species and experimental designs. Many of the genes involved in sex determination are broadly conserved among vertebrates although the timing or location of expression may differ among species. Here I review some genes that are differentially expressed between the sexes during gonadogenesis.

Several genes are involved in the formation and maintenance of bipotential gonads in both mammals and reptiles. For example, in mammals *Emx2* is expressed in urogenital system and is crucial to the formation of kidney and genital tracks. *Sf1*, *Lhx9* and *M33* are involved in the formation of the bipotential gonad and proliferation of somatic cells within gonads (Biason-Lauber, 2010). Among the factors involved in the formation of the bipotential gonad, genes such as *Wt1* (Wilms tumor 1), *Sf1* (steroidogenic factor 1) and *Lhx9* (LIM Homeobox 9) have been examined in TSD species (Rhen and Schroeder, 2010). *Sf1* is expressed throughout the bipotential gonad at both male and female incubation temperatures but its expression pattern differs from species to species. This difference may result from the technique or tissue utilized in different experiments. Whether *Sf1* has a testis-specific role or not is still not clear in TSD species. *Wt1* is required for the development of the kidney and the bipotential gonad in mice (Kreidberg et al., 1993; Moore et al., 1999). Two splice variants (-KTS and +KTS) of WT1 play different roles in kidney and gonad development. The +KTS variant is involved in testicular development by regulating the expression of male-determining genes such as *Sry*, *Sox9* and *Fgf9* (Bradford et al., 2009; Hammes et al., 2001). In the snapping turtle, the ratio of +KTS:-KTS variants was found to be significantly higher in bipotential gonads at male-producing temperature (MPT) than it was at female-producing

temperature (FPT), indicating the importance of *Wt1* in male determination in TSD (Rhen et al., 2015).

Factors involved in testis development in mammals are *SRY, SOX9, SOX8, FGF9, AMH, SF1, DAX1, DMRT1, DHH, ATRX, TSPYL1, PGD2S* (Biason-Lauber, 2010) (Fig.1). Among these factors, *SOX9, SOX8, FGF9, AMH, SF1, DAX1,* and *DMRT1* have been studied in TSD species. Reptiles do not have *Sry* (Lance, 1997), but *Sox9* appears to play a role in testis development and may act as a master sex-determining gene in male development in TSD species. Studies in turtles, lizards and alligators have found monomorphic expression pattern for *Sox9* at early stages of the TSP and elevated expression in testis at the end of the TSP (Rhen and Schroeder, 2010; Shoemaker and Crews, 2009). Other conserved male-determining genes in TSD reptiles include *Dmrt1*, whose expression was detected in the early bipotential gonad and was gradually increased at MPT but suppressed at FPT during TSP (Rhen and Schroeder, 2010; Rhen et al., 2007) and *Amh,* whose expression was demonstrated to be significantly up-regulated at MPT and suppressed at FPT during TSP (Shoemaker-Daly et al., 2010).

Factors like *Rspo1, Wnt4, Foxl2, HoxA* and *Lim1* play important roles in the vertebrate ovary development (Fig.1). In TSD, *Wnt4, Rspo1* and *Foxl2* appear to play a conserved role in reptile sexual development (Merchant-Larios and Díaz-Hernández, 2013; Rhen and Schroeder, 2010; Shoemaker and Crews, 2009). In mammals, *Wnt4* regulates germ cell viability and formation of kidney and adrenal glands by influencing steroid genesis through the up-regulation of *Dax1* which inhibits the production of steroidgenic enzymes through interfering with *Sf1*(Mizusaki et al., 2003). *Rspo1* reinforces *Wnt4* signaling pathway by activating β-catenin, thereby promoting ovarian

development (Chassot et al., 2008). Double knockout of *Foxl2* and *Wnt4* results in complete female to male sex reversal in mammals (Ottolenghi et al., 2007a). *Foxl2,* a member of the forkhead box gene family is critical to ovarian development and is the earliest marker of the differentiation of ovarian somatic cells (Uhlenhaut and Treier, 2006). Mutations in *Foxl2* lead to gonadal dysgenesis and ovarian failure in mice and goats (Pailhoux et al., 2001; Uda et al., 2004).

Steroid Signaling in Reptilian Sex Determination

Steroid hormones not only regulate the sexual differentiation of somatic cells after sex determination, but are equally important in directing gonad fate of ESD species. Estrogens are well-studied hormones that regulate ovarian determination. In European pond turtles, exogenous estrogen treatment of developing embryos at MPT causes male to female sex reversal, indicating estrogen is able to override the effect of temperature thereby redirecting the fate of gonad (Pieau, 1974). Ramsey and Crews (2009) reported that warm temperature acts in concert with estrogen since less estrogen is required to reverse sex at FPT than at MPT. Inhibition of aromatase, an enzyme which converts androgens into estrogens, at FPT induces testis development in turtles (Dorizzi et al., 1994; Rhen and Lang, 1994; Wibbels and Crews, 1994).

Aromatase and estrogens have been reported to influence ovarian development in many TSD species such as reptiles, fishes, amphibians and some other non-TSD species such as birds and marsupials. However, in mammals, estrogen only helps to maintain ovarian phenotype at later stages (Pieau and Dorizzi, 2004). Aromatase and estrogens also play important roles in ovarian differentiation in snapping turtles. Although an aromatase inhibitor (AI) has no effect on sex ratio at MPT, it is able to induce testis

differentiation at a temperature that produces a mixed sex ratio (Rhen and Lang, 1994). This indicates aromatase plays a role in sex determination in snapping turtles, although AI alone appears inefficient in inducing testis differentiation at strictly FPT. The inefficiency of AI may result from the production of large amount of aromatase at FPT or different affinities between AI and aromatase in different tissues or at different temperatures.

Some researchers believe that estrogen may not be involved in early ovarian differentiation and suggest that temperature may not act directly on the gonad in TSD species but on extra-gonadal tissues during the TSP. This idea comes from studies in which aromatase, which directly regulates estrogen levels, was not differentially expressed in adrenal-kidney-gonad complexes (AKG) (Pieau and Dorizzi, 2004). However, studies show that genes involved in sex determination are also expressed in the adrenal gland and kidney. Therefore, subtle expression changes of aromatase in gonads could be masked by aromatase expression in adrenal gland and kidney (Ramsey and Crews, 2007). Ramsey and Crews pointed out that the expression of 5 genes (*Ar, Er-α, Er-β,* aromatase*, Sf1*) in the gonad during TSP was masked by their expression in adrenal gland and kidney in slider turtles. Only genes with large changes, such as Dmrt1, could be distinguished in the gonad between MPT and FPT. This masking effect was also displayed by another study of fresh water turtle, *Emys orbicularis* (Pieau and Dorizzi, 2004). In that study, synthesis of estrogen was shown to occur only in the gonad and *Er* were expressed throughout the gonad during sexual development. Studies based on the entire AKG may result in skewed data.

Apart from estrogen, androgens play important roles in sex determination through AR in TSD species. AR is a ligand-activated transcription factor of the steroid hormone subfamily of nuclear receptors (NR) and it is the only NR situated on X chromosome in mammals (Lavery and Bevan, 2011). AR protein is composed of three different parts: the N-terminus, the hinge region, and the C-terminus. The N-terminal domain contains some secondary structures involved in protein-protein interactions. The DNA binding domain (DBD) is situated in the center of the AR and binds to specific DNA sequences termed androgen response elements (ARE). C-terminus contains a ligand-binding domain (LBD), where androgens are recognized and docked. Two transactivation domains, activation function 1 and activation function 2, are located in N- and C-terminals respectively. AR is found in the cytoplasm in association with a set of heat-shock or heat-shock-related proteins. Binding of androgen leads to a conformational change in the AR. The AR then enters the nucleus, binds to AREs, and recruits co-activators and co-repressors, thereby regulating gene expression. During gonadogenesis in chickens, expression of *Ar* is higher in ovary than in testis and disturbing *Ar* function leads to ovarian disorganization (Katoh et al., 2006). Similarly, in *Anguilla australis*, a New Zealand short-finned eel, *Ar* is able to increase expression of ovarian follicle-stimulating hormone (FSH) receptor and plasma levels of 17β-estradiol, thereby stimulating the development of follicles (Setiawan et al., 2012). Studies also indicate that sheep embryos exposed to large doses of testosterone develop a phenotype similar to polycystic ovary syndrome (PCOS) in humans (Padmanabhan and Veiga-Lopez, 2013).

Testosterone and dihydrotestosterone (DHT) are two forms of androgen, whose synthesis is regulated by the hypothalamic-pituitary-gonadal axis (Heemers and Tindall,

2007). Both testosterone and DHT exert their effects by binding to AR. DHT has higher binding affinity for AR than does testosterone. In slider turtles, exogenous DHT treatment at a pivotal developmental stage leads to 100% female to male sex reversal although DHT cannot override all female temperature (Wibbels and Crews, 1992; 1995); inhibiting DHT synthesis leads to male to female sex reversal (Wibbels and Crews, 1994); combined estrogen and DHT treatment at specific time leads to ovotesis (Wibbels and Crews, 1994; 1995; Wibbels et al., 1992). In contrast, androgens appear to be playing a role in ovarian development rather than testicular development in snapping turtles (Rhen and Lang, 1994; Rhen and Schroeder, 2010). Understanding the feminizing effect of androgen in gonad development of snapping turtles is one of my study objectives. However, the study of molecular mechanism underlying TSD is still at its infant stage. For this reason, this study is limited to only a few well-studied *Foxl2* targets. The impact of sex steroids on the developing gonad at larger scales needs to be revealed. To reach this goal, we decided to bring our study of TSD to a genome-wide scale. By doing so, we will be able to provide a strong foundation for future studies in TSD.

Identification of New Candidate Sex-Determining Genes for TSD Using High Throughput Sequencing Data

Studies of the common snapping turtle have revealed genes that are involved in TSD. These genes include *Wt1* (Rhen et al., 2015), *Pdgf* (Rhen et al., 2009), *Dmrt1, Sox9*, aromatase, *Ar* and *Foxl2* (Rhen et al., 2007). It is rather common to identify a core set of genes that are presumably conserved in the process of sex determination by comparing closely related vertebrates and testing whether the genes are differentially expressed *in vitro* or *in vivo*. However, this approach lacks the ability to discover novel sex-determining genes and can be time consuming and misleading. For example, *Dax1, Fgf9,*

and *Sf1* are involved in sex determination in GSD species, but are not differentially expressed between MPT and FPT during the TSP in snapping turtles (Rhen et al., 2007).

To further study the molecular mechanism of TSD and overcome the limitations of using well-studied sex-determining genes from mammals, we initiated a transcriptome study on the snapping turtle. This study sequenced the entire gonad transcriptome of the snapping turtle during the TSP. Assembly and annotation of the transcriptome along with differential gene expression (DGE) analysis provides novel insight into TSD in the snapping turtle from a transcriptome-wide perspective.

The development of Next Generation Sequencing (NGS) techniques allows researchers to conduct transcriptome-wide analyses of gene expression, which immensely accelerates research progress in many fields of biology. Although new technology has brought down sequencing costs, sequencing of large vertebrate genomes is still quite expensive. According to NCBI Genome Database, 325 out of 13525 published genomes are from vertebrates and only 11 genomes are from reptiles, reflecting the high cost and challenges of sequencing large and complex vertebrate genomes. Compared to whole genome sequencing, *de novo* transcriptome sequencing is a cost-efficient method that sequences all the transcripts from a given sample. This process is ideal for acquiring information about gene function and expression in non-model organisms, such as the common snapping turtle.

Although assembly and annotation of the snapping turtle transcriptome and DGE analysis significantly improves our understanding of the molecular mechanism of TSD, more work needs to be done in order to better understand this mechanism. Network reverse engineering is a great way to elucidate the interactions between genes and how

these interactions are influenced by the environment (temperatures in this case) during TSD. Reconstructing the gene regulatory network during sex determination enables us to further reveal the molecular mechanism underlying this biological process.

Network Reverse Engineering

In well-studied model species, such as the mouse, efforts have also been made to identify novel genes involved in sex determination and the transcriptional cascade controlling this process. Some of the studies used high-throughput whole-mount in situ hybridization to identify genes specifically expressed in the developing gonad (Wertz and Herrmann, 2000). Some used microarrays to determine the expression profiles of whole embryonic mouse gonads and identified candidate sex-determining genes through differential expression analysis (Munger et al., 2009; Small et al., 2005). Some went further by examining gene expression profiles in separate cell lineages from the developing gonad (Jameson et al., 2012; Munger et al., 2013). However, none of these studies revealed how these genes are regulated specifically in gonads. Even less is known about how they interact with each other. In other words, to fully understand the molecular mechanism of sex determination, gene regulatory networks need to be reconstructed.

High-throughput technologies such as microarray and RNA-Seq provide us with powerful means of identifying differentially expressed genes at a transcriptome-wide scale. Reconstructing transcriptional regulatory networks based on gene expression profiles generated by these tools has proven to be a promising approach in many biological and medical fields (Cho et al., 2012; Thompson et al., 2015). Transcription networks in both prokaryotes and eukaryotes exhibit a hierarchical scale-free nature, characterized by vertices with a degree that greatly exceeds the average (Albert, 2005).

Numerous computational algorithms have been developed to dissect genome-wide gene regulatory networks (Margolin et al., 2006b). Though some of these methods were successfully applied to infer regulatory modules from gene expression data in simple eukaryotes, model limitations confine their application to small and less complex networks (Margolin et al., 2006b). A great challenge in computational biology involves organization of large number of genes into complex networks in higher eukaryotes (Jiang et al., 2004). A number of algorithms have been proposed in the past few years, which include entropy-based network modeling (Margolin et al., 2006a; Villaverde et al., 2014; Wang et al., 2013), networks based on marginal dependencies (Liu et al., 2016), network reconstruction by integrating prior biological knowledge (Li and Jackson, 2015), and integration of predictions from multiple inference methods (Ceci et al., 2015).

Interactions among genes are not always linear and straightforward. They can be nonlinear, condition dependent, or time-lagged dependent (Liu et al., 2016). Previously proposed linear models in most studies are restricted not only by the need for estimating linear high-dimensional dependency structures but also suffer from the limitation of capturing nonlinear interactions (Hausser and Strimmer, 2009). To loosen the linearity assumption and capture the nonlinear associations among genes, entropy-based network reconstructing algorithms, such as ARACNE, MRNET, MIDER, CLR, C3NET and TINGe, were proposed (Altay and Emmert-Streib, 2010; Aluru et al., 2013; Faith et al., 2007; Margolin et al., 2006a; Meyer et al., 2007; Villaverde et al., 2014). These methods rely on computing the mutual information (MI) between genes, a concept borrowed from probability theory and information theory. Mutual information is always positive if two variables are related and zero if they are independent regardless whether their relationship

is linear or nonlinear (Kraskov et al., 2003). This makes MI a robust measure of gene interactions. To reconstruct accurate interactomes, this study used a mutual information based algorithm – ARACNE, which was widely used in inferring transcriptional regulatory networks (Agnelli et al., 2011; Carro et al., 2010; Remo et al., 2015).

Study Objectives

The research described in this dissertation is focused on understanding the genetic and molecular mechanisms of TSD in the common snapping turtle and identifying gene interactions in developing mouse gonads from publicly available data sets. The main objectives of this study are:

1. Determine the role of androgens in regulating *Foxl2* expression in the snapping turtle and testing for interactions between androgens and *Foxl2*.
2. Identify candidate genes involved in TSD at a transcriptome-wide scale.
3. Analyze mammalian sex determination by reconstructing and comparing gene regulatory networks in developing mouse gonads.

The first objective is an extension of previous work that suggests androgens play a role in TSD in the snapping turtle (Rhen and Lang, 1994; Rhen and Schroeder, 2010). More specifically, we will examine androgen signaling in ovarian Granulosa cells. Objective 2 will focus on deciphering the molecular mechanisms underlying TSD based on RNA-Seq analysis of transcriptome-wide patterns of gene expression. Objective 3 aims to discover novel genes and interactions that are involved in sex determination in mice. The results of objective 2 and 3 will be used to guide future studies of TSD.

REFERENCES

Agnelli, L., Forcato, M., Ferrari, F., Tuana, G., Todoerti, K., Walker, B.A., Morgan, G.J., Lombardi, L., Bicciato, S., and Neri, A. (2011). The reconstruction of transcriptional networks reveals critical genes with implications for clinical outcome of multiple myeloma. Clin. Cancer Res. 17, 7402–7412.

Albert, R. (2005). Scale-free networks in cell biology. Journal of Cell Science 118, 4947–4957.

Altay, G., and Emmert-Streib, F. (2010). Inferring the conservative causal core of gene regulatory networks. BMC Systems Biology 4, 132.

Aluru, M., Zola, J., Nettleton, D., and Aluru, S. (2013). Reverse engineering and analysis of large genome-scale gene networks. Nucleic Acids Research 41, e24–e24.

Angelopoulou, R., Lavranos, G., and Manolakou, P. (2012). Sex determination strategies in 2012: towards a common regulatory model? Reprod. Biol. Endocrinol. 10, 13.

Biason-Lauber, A. (2010). Control of sex development. Best Practice and Research: Clinical Endocrinology and Metabolism 24, 163–186.

Bradford, S.T., Wilhelm, D., Bandiera, R., Vidal, V., Schedl, A., and Koopman, P. (2009). A cell-autonomous role for WT1 in regulating Sry in vivo. Human Molecular Genetics 18, 3429–3438.

Brennan, J., and Capel, B. (2004). One tissue, two fates: molecular genetic events that underlie testis versus ovary development. Nature Reviews. Genetics 5, 509–521.

Bull, J.J. (1987). Temperature-sensitive periods of sex determination in a lizard: Similarities with turtles and crocodilians. The Journal of Experimental Zoology 241, 143–148.

Burke, R.L., and Calichio, A.M. (2014). Temperature-Dependent Sex Determination in the Diamond-backed Terrapin (Malaclemys terrapin). Journal of Herpetology 48, 466–470.

Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature 463, 318–325.

Ceci, M., Pio, G., Kuzmanovski, V., and Džeroski, S. (2015). Semi-Supervised Multi-View Learning for Gene Network Reconstruction. Plos One 10, e0144031.
Chassot, A.-A., Ranc, F., Gregoire, E.P., Roepers-Gajadien, H.L., Taketo, M.M., Camerino, G., de Rooij, D.G., Schedl, A., and Chaboissier, M.-C. (2008). Activation of beta-catenin signaling by Rspo1 controls differentiation of the mammalian ovary. Human Molecular Genetics 17, 1264–1277.

Cheng, J.-C., Klausen, C., and Leung, P.C.K. (2013). Overexpression of wild-type but not C134W mutant *FOXL2* enhances GnRH-induced cell apoptosis by increasing GnRH receptor expression in human granulosa cell tumors. Plos One 8, e55099.

Cho, D.-Y., Kim, Y.-A., and Przytycka, T.M. (2012). Chapter 5: Network Biology Approach to Complex Diseases. PLoS Comput Biol 8, e1002820.

David, J.R., Yassin, A., Moreteau, J.-C., Legout, H., and Moreteau, B. (2011). Thermal phenotypic plasticity of body size in Drosophila melanogaster: sexual dimorphism and genetic correlations. Journal of Genetics 90, 295–302.

Dorizzi, M., Richard-Mercier, N., Desvages, G., Girondot, M., and Pieau, C. (1994). Masculinization of gonads by aromatase inhibitors in a turtle with temperature-dependent sex determination. Differentiation 58, 1–8.

Eicher, E.M., and Washburn, L.L. (1986). Genetic control of primary sex determination in mice. Annual Review of Genetics 20, 327–360.

Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J., and Gardner, T.S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biology 5, e8–e0066.

Gamble, T., and Zarkower, D. (2012). Sex determination. Curr. Biol. 22, R257–R262.

Hammes, A., Guo, J.K., Lutsch, G., Leheste, J.R., Landrock, D., Ziegler, U., Gubler, M.C., and Schedl, A. (2001). Two splice variants of the Wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. Cell 106, 319–329.

Hausser, J., and Strimmer, K. (2009). Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks. Journal of Machine Learning Research 10, 1469–1484.

Heemers, H.V., and Tindall, D.J. (2007). Androgen receptor (AR) coregulators: A diversity of functions converging on and regulating the AR transcriptional complex. Endocrine Reviews 28, 778–808.

Jameson, S.A., Natarajan, A., Cool, J., DeFalco, T., Maatouk, D.M., Mork, L., Munger, S.C., and Capel, B. (2012). Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad. PLOS Genet 8, e1002575.

Jiang, D., Tang, C., and Zhang, A. (2004). Cluster Analysis for Gene Expression Data: A Survey. IEEE Trans. on Knowl. and Data Eng. 16, 1370–1386.

Katoh, H., Ogino, Y., and Yamada, G. (2006). Cloning and expression analysis of androgen receptor gene in chicken embryogenesis. FEBS Letters 580, 1607–1615.

Kraskov, A., Stögbauer, H., Andrzejak, R.G., and Grassberger, P. (2003). Hierarchical Clustering Based on Mutual Information. arXiv:Q-Bio/0311039.

Kreidberg, J.A., Sariola, H., Loring, J.M., Maeda, M., Pelletier, J., Housman, D., and Jaenisch, R. (1993). WT-1 is required for early kidney development. Cell 74, 679–691.

Lance, V.A. (1997). Sex determination in reptiles: An update. American Zoologist 37, 504–513.

Lance, V.A., Elsey, R.M., and Lang, J.W. (2000). Sex ratios of American alligators (Crocodylidae): male or female biased? Journal of Zoology 252, 71–78.

Lance, V.A., Elsey, R.M., Butterstein, G., and Trosclair, P.L. (2004). Rapid suppression of testosterone secretion after capture in male American alligators (Alligator mississippiensis). General and Comparative Endocrinology 135, 217–222.

Lang, J.W., and Andrews, H.V. (1994). Temperature-dependent sex determination in crocodilians. Journal of Experimental Zoology 270, 28–44.

Lavery, D.N., and Bevan, C.L. (2011). Androgen Receptor Signalling in Prostate Cancer: The Functional Consequences of Acetylation. Journal of Biomedicine and Biotechnology 2011, 1–7.

Li, Y., and Jackson, S.A. (2015). Gene Network Reconstruction by Integration of Prior Biological Knowledge. G3 (Bethesda) 5, 1075–1079.

Liu, H., Li, P., Zhu, M., Wang, X., Lu, J., and Yu, T. (2016). Nonlinear Network Reconstruction from Gene Expression Data Using Marginal Dependencies Measured by DCOL. Plos One 11, e0158247.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla-Favera, R., and Califano, A. (2006a). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7 Suppl 1, S7.

Margolin, A.A., Wang, K., Lim, W.K., Kustagi, M., Nemenman, I., and Califano, A. (2006b). Reverse engineering cellular networks. Nature Protocols 1, 662–671.

Matson, C.K., Murphy, M.W., Sarver, A.L., Griswold, M.D., Bardwell, V.J., and Zarkower, D. (2011). DMRT1 prevents female reprogramming in the postnatal mammalian testis. Nature 476, 101–104.

McElreavey, K., Vilain, E., Abbas, N., Herskowitz, I., and Fellous, M. (1993). A regulatory cascade hypothesis for mammalian sex determination: SRY represses a negative regulator of male development. Proceedings of the National Academy of Sciences of the United States of America 90, 3368–3372.

Merchant-Larios, H., and Díaz-Hernández, V. (2013). Environmental sex determination mechanisms in reptiles. Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation 7, 95–103.

Meyer, P.E., Kontos, K., Lafitte, F., and Bontempi, G. (2007). Information-theoretic inference of large transcriptional regulatory networks. EURASIP Journal on Bioinformatics \& Systems Biology 2007, 79879–79879.

Mizusaki, H., Kawabe, K., Mukai, T., Ariyoshi, E., Kasahara, M., Yoshioka, H., Swain, A., and Morohashi, K.-I. (2003). Dax-1 (dosage-sensitive sex reversal-adrenal hypoplasia congenita critical region on the X chromosome, gene 1) gene transcription is regulated by wnt4 in the female developing gonad. Molecular Endocrinology (Baltimore, Md.) 17, 507–519.

Modi, W.S., and Crews, D. (2005). Sex chromosomes and sex determination in reptiles: Commentary. Curr. Opin. Genet. Dev. 15, 660–665.

Moore, A.W., McInnes, L., Kreidberg, J., Hastie, N.D., and Schedl, A. (1999). YAC complementation shows a requirement for Wt1 in the development of epicardium, adrenal gland and throughout nephrogenesis. Development (Cambridge, England) 126, 1845–1857.

Munger, S.C., Aylor, D.L., Syed, H.A., Magwene, P.M., Threadgill, D.W., and Capel, B. (2009). Elucidation of the transcription network governing mammalian sex determination by exploiting strain-specific susceptibility to sex reversal. Genes \& Development 23, 2521–2536.

Munger, S.C., Natarajan, A., Looger, L.L., Ohler, U., and Capel, B. (2013). Fine time course expression analysis identifies cascades of activation and repression and maps a putative regulator of mammalian sex determination. PLOS Genet 9, e1003630.

Ottolenghi, C., Pelosi, E., Tran, J., Colombino, M., Douglass, E., Nedorezov, T., Cao, A., Forabosco, A., and Schlessinger, D. (2007a). Loss of Wnt4 and *Foxl2* leads to female-to-male sex reversal extending to germ cells. Human Molecular Genetics 16, 2795–2804.

Ottolenghi, C., Uda, M., Crisponi, L., Omari, S., Cao, A., Forabosco, A., and Schlessinger, D. (2007b). Determination and stability of sex. BioEssays 29, 15–25.

Owens, I.P.F., and Hartley, I.R. (1998). Sexual dimorphism in birds: why are there so many different forms of dimorphism? Proceedings of the Royal Society B: Biological Sciences 265, 397–407.

Padmanabhan, V., and Veiga-Lopez, A. (2013). Sheep models of polycystic ovary syndrome phenotype. Molecular and Cellular Endocrinology 373, 8–20.

Pailhoux, E., Vigier, B., Chaffaux, S., Servel, N., Taourit, S., Furet, J.P., Fellous, M., Grosclaude, F., Cribiu, E.P., Cotinot, C., et al. (2001). A 11.7-kb deletion triggers intersexuality and polledness in goats. Nature Genetics 29, 453–458.

Parkhurst, S.M., Bopp, D., and Ish-Horowicz, D. (1990). X:A ratio, the primary sex-determining signal in Drosophila, is transduced by helix-loop-helix proteins. Cell 63, 1179–1191.

Pieau, C. (1974). Sex differentiation according to the temperature in embryos of Emys arbicularis L. (Chelonia): effect of sex hormones. Ann.Embryol.Morphogen. 7, 365–394.

Pieau, C., and Dorizzi, M. (2004). Oestrogens and temperature-dependent sex determination in reptiles: all is in the gonads. The Journal of Endocrinology 181, 367–377.

Pieau, C., and Dorizzi, M. (1981). Determination of temperature sensitive stages for sexual differentiation of the gonads in embryos of the turtle, Emys orbicularis. Journal of Morphology 170, 373–382.

Pisarska, M.D., Bae, J., Klein, C., and Hsueh, A.J.W. (2004). Forkhead L2 Is Expressed in the Ovary and Represses the Promoter Activity of the Steroidogenic Acute Regulatory Gene. Endocrinology 145, 3424–3433.

Ramsey, M., and Crews, D. (2007). Steroid signaling system responds differently to temperature and hormone manipulation in the red-eared slider turtle (Trachemys scripta elegans), a reptile with temperature-dependent sex determination. Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation 1, 181–196.

Raymond, C.S., Murphy, M.W., O'Sullivan, M.G., Bardwell, V.J., and Zarkower, D. (2000). Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. Genes \& Development 14, 2587–2595.

Remo, A., Simeone, I., Pancione, M., Parcesepe, P., Finetti, P., Cerulo, L., Bensmail, H., Birnbaum, D., Van Laere, S.J., Colantuoni, V., et al. (2015). Systems biology analysis reveals NFAT5 as a novel biomarker and master regulator of inflammatory breast cancer. J Transl Med 13, 138.

Rhen, T., and Lang, J.W. (1994). Temperature-dependent sex determination in the snapping turtle: Manipulation of the embryonic sex steroid environment. General and Comparative Endocrinology 96, 243–254.

Rhen, T., and Schroeder, A. (2010). Molecular mechanisms of sex determination in reptiles. Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation 4, 16–28.

Rhen, T., Metzger, K., Schroeder, A., and Woodward, R. (2007). Expression of putative sex-determining genes during the thermosensitive period of gonad development in the snapping turtle, Chelydra serpentina. Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation 1, 255–270.

Rhen, T., and Lang, J.W. (1998). Among-family variation for environmental sex determination in reptiles. Evolution 52, 1514–1520.

Rhen, T., Fagerlie, R., Schroeder, A., Crossley, D.A., and Lang, J.W. (2015). Molecular and morphological differentiation of testes and ovaries in relation to the thermosensitive period of gonad development in the snapping turtle, Chelydra serpentina. Differentiation; Research in Biological Diversity 89, 31–41.

Rhen, T., Jangula, A., Schroeder, A., and Woodward-Bosh, R. (2009). The platelet-derived growth factor signaling system in snapping turtle embryos, Chelydra serpentina: potential role in temperature-dependent sex determination and testis development. General and Comparative Endocrinology 161, 335–343.

Sarraj, M.A., and Drummond, A.E. (2012). Mammalian foetal ovarian development: consequences for health and disease. Reproduction 143, 151–163.

Sarre, S.D., Georges, A., and Quinn, A. (2004). The ends of a continuum: genetic and temperature-dependent sex determination in reptiles. BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology 26, 639–645.

Sawanth, S.K., Gopinath, G., Sambrani, N., and Arunkumar, K.P. (2016). The autoregulatory loop: A common mechanism of regulation of key sex determining genes in insects. J. Biosci. 41, 283–294.

Sekido, R., Bar, I., Narváez, V., Penny, G., and Lovell-Badge, R. (2004). SOX9 is up-regulated by the transient expression of SRY specifically in Sertoli cell precursors. Developmental Biology 274, 271–279.

Setiawan, A.N., Ozaki, Y., Shoae, A., Kazeto, Y., and Lokman, P.M. (2012). Androgen-specific regulation of FSH signalling in the previtellogenic ovary and pituitary of the New Zealand shortfinned eel, Anguilla australis. General and Comparative Endocrinology 176, 132–143.

Shoemaker, C.M., and Crews, D. (2009). Analyzing the coordinated gene network underlying temperature-dependent sex determination in reptiles. Seminars in Cell \& Developmental Biology 20, 293–303.

Shoemaker-Daly, C.M., Jackson, K., Yatsu, R., Matsumoto, Y., and Crews, D. (2010). Genetic network underlying temperature-dependent sex determination is endogenously regulated by temperature in isolated cultured Trachemys scripta gonads. Developmental Dynamics 239, 1061–1075.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R., and Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346, 240–244.

Siroski, P., Larriera, A., Piña, C., Lance, V., and Verdade, L. (2007). The temperature-sensitive period (TSP) during incubation of broad-snouted caiman (Caiman latirostris) eggs. Amphibia-Reptilia 28, 123–128.

Small, C.L., Shima, J.E., Uzumcu, M., Skinner, M.K., and Griswold, M.D. (2005). Profiling gene expression during the differentiation and development of the murine embryonic gonad. Biology of Reproduction 72, 492–501.

Thompson, D., Regev, A., and Roy, S. (2015). Comparative Analysis of Gene Regulatory Networks: From Network Reconstruction to Evolution. Annual Review of Cell and Developmental Biology 31, 399–428.

Uda, M., Ottolenghi, C., Crisponi, L., Garcia, J.E., Deiana, M., Kimber, W., Forabosco, A., Cao, A., Schlessinger, D., and Pilia, G. (2004). *Foxl2* disruption causes mouse ovarian failure by pervasive blockage of follicle development. Human Molecular Genetics 13, 1171–1181.

Uhlenhaut, N.H., and Treier, M. (2006). *Foxl2* function in ovarian development. Molecular Genetics and Metabolism 88, 225–234.

Vainio, S., Heikkilä, M., Kispert, A., Chin, N., and McMahon, A.P. (1999). Female development in mammals is regulated by Wnt-4 signalling. Nature 397, 405–409.

Valcárcel, J., Singh, R., Zamore, P.D., and Green, M.R. (1993). The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. Nature 362, 171–175.

Villaverde, A.F., Ross, J., Morán, F., and Banga, J.R. (2014). MIDER: network inference with mutual information distance and entropy reduction. Plos One 9, e96732.

Wang, J., Chen, B., Wang, Y., Wang, N., Garbey, M., Tran-Son-Tay, R., Berceli, S.A., and Wu, R. (2013). Reconstructing regulatory networks from the dynamic plasticity of gene expression by mutual information. Nucleic Acids Research 41, e97–e97.

Wertz, K., and Herrmann, B.G. (2000). Large-scale screen for genes involved in gonad development. Mechanisms of Development 98, 51–70.

Wibbels, T., and Crews, D. (1992). Specificity of steroid hormone-induced sex determination in a turtle. The Journal of Endocrinology 133, 121–129.

Wibbels, T., and Crews, D. (1994). Putative aromatase inhibitor induces male sex determination in a female unisexual lizard and in a turtle with temperature-dependent sex determination. The Journal of Endocrinology 141, 295–299.

Wibbels, T., and Crews, D. (1995). Steroid-Induced Sex Determination at Incubation Temperatures Producing Mixed Sex Ratios in a Turtle with TSD. General and Comparative Endocrinology 100, 53–60.

Wibbels, T., Bull, J.J., and Crews, D. (1992). Steroid hormone-induced male sex determination in an amniotic vertebrate. The Journal of Experimental Zoology 262, 454–457.

Yntema, C.L. (1979). Temperature levels and periods of sex determination during incubation of eggs of Chelydra serpentina. Journal of Morphology 159, 17–27.

Zhao, L., Svingen, T., Ng, E.T., and Koopman, P. (2015). Female-to-male sex reversal in mice caused by transgenic overexpression of Dmrt1. Development 142, 1083–1088.

CHAPTER II

SNAPPING TURTLE (CHELYDRA SERPENTINA) *FOXL2* AND LOW DOSES OF DIHYDROTESTOSTERONE SYNERGISTICALLY REGULATE AROMATASE EXPRESSION IN MOUSE KK1 GRANULOSA CELLS

Abstract

Sex is determined by temperature during embryogenesis in the snapping turtle, *Chelydra serpentina*. Previous studies show that dihydrotestosterone (DHT) induces ovary development at temperatures that normally produce males or mixed sex ratios. The feminizing effect of DHT is associated with increased expression of the ovary-determining gene *Foxl2*, suggesting that androgens may regulate transcription of *Foxl2*. To test this hypothesis, we cloned the proximal promoter (1.6 kb) and coding sequence for snapping turtle *Foxl2* (t*Foxl2*) in frame with mCherry, a red fluorescent protein. The t*Foxl2*-mCherry fusion plasmid or mCherry plasmid were stably transfected into mouse KK1 granulosa cells. These cells were then treated with 0, 1, 10, or 100 nM DHT to assess androgen effects on t*Foxl2*-mCherry expression as well as the combined effects of DHT and t*Foxl2*-mCherry on endogenous target genes. In contrast to the main hypothesis, expression of t*Foxl2*-mCherry was not affected by DHT treatment. However, DHT inhibited expression of the endogenous mouse *Foxl2* gene, suggesting that androgen effects on *Foxl2* 1) require regulatory sequences outside the proximal

promoter/coding sequence, 2) depend on genomic context, and/or 3) differ between species. We also found that t*Foxl2*-mCherry influenced expression of *Fshr, Gnrhr*, and *Star*. Our most interesting discovery was that t*Foxl2*-mCherry potentiated low dose DHT effects on aromatase expression. In addition, we found newborn calf serum (NCS) suppressed expression of the transfected t*Foxl2*-mCherry plasmid and endogenous aromatase, *Gnrhr, Star*, *Foxl2* when compared to charcoal-stripped NCS.

Introduction

Sex-determining systems are remarkably diverse among vertebrates. In some cases, sex is determined chromosomally, which is known as genotypic sex determination (GSD) while in other cases sex is determined environmentally, which is known as environmental sex determination (ESD) (Manolakou et al., 2006). Different types of GSD and ESD exist in reptiles and sometimes both occur together in the same species (Conover and Heins 1987; Radder et al., 2008; Holleley et al. 2015). Temperature is the only natural environmental factor that affects sexual development in reptiles (Janzen and Paukstis, 1991). Sex of many turtles and all crocodilians examined so far is determined by ambient temperature during a specific period of embryonic development known as the temperature-sensitive period (TSP) (Ciofi and Swingland, 1997). In temperature-dependent sex determination (TSD), temperature serves as a switch that initiates a cascade of changes in gene expression that determines gonad fate (Rhen and Schroeder, 2010; Rhen et al., 2015; Schroeder et al., 2016).

In addition, developing embryos of TSD species respond to steroid hormones and the timing of their sensitivity to steroids coincides with the TSP. Manipulation of either

incubation temperature or exposure to steroid hormones during the TSP will redirect the sex of the embryo (Ramsey and Crews, 2009). Steroid-induced sex determination has been extensively studied in the red-eared slider turtle (*Trachemys scripta*) (Crews et al., 1996; Wibbels et al., 1991). During embryonic development, exogenous estrogen treatment or inhibition of estrogen production can override the temperature cue thereby redirecting the sexual fate of embryos (Crews and Bergeron, 1994; Wibbels et al., 1993). However, estrogen and incubation temperatures do not work independently in the process of sex determination. In fact, they act synergistically – more estrogen is needed to sex-reverse an embryo at an extreme male-producing temperature than at a temperature closer to the female-producing range of temperatures (Ramsey and Crews, 2009). Administration of non-aromatizable androgens to slider turtle embryos cannot override temperatures that produce exclusively females, but can induce more males at temperatures that produce mixed sex ratios (Wibbels and Crews, 1992; Wibbels and Crews, 1995). Conversely, inhibition of 5α-reductase and DHT synthesis can induce more females than expected at male-biased temperatures in the red-eared slider turtle (Crews and Bergeron, 1994).

Both estrogens and androgens play crucial roles in sexual development of all vertebrates. In mammals, estrogens are involved in the development of female secondary sex characteristics, though they are not considered to be necessary for ovary determination and ovarian development in placental mammals (Couse et al., 2000; Fisher et al., 1998). However, knockout of estrogen receptors or aromatase in mice leads to postnatal sex reversal (Couse et al., 1999; Dupont et al., 2003; Britt et al., 2001). Androgens also play important roles in normal ovary development and differentiation.

During chicken gonadogenesis, expression of androgen receptor is higher in ovary than in testis and disturbing its function leads to ovarian disorganization (Katoh et al., 2006). In *Anguilla australis*, a New Zealand short-finned eel, androgen receptor is able to increase expression of follicle-stimulating hormone (FSH) receptor and plasma levels of estradiol-17β, thereby stimulating ovarian follicle development (Setiawan et al., 2012). However, too much androgen causes ovarian dysfunction. In rhesus monkey, females exposed to excess androgen early in gestation display polycystic ovary syndrome (Abbott et al., 2005).

In species with TSD, steroid hormones and incubation temperature play critical roles in ovary determination and differentiation as discussed above. Androgens and estrogens influence ovarian development through binding to their respective receptors – androgen receptor (AR) and estrogen receptors alpha and beta (ERα and ERβ). Aromatase, *Erα, Erβ* and *Ar* are expressed at higher levels at female-producing temperatures than at male-producing temperatures during gonadal development of the slider turtle (Ramsey and Crews, 2007). Aromatase is regulated by *Foxl2*, a key female-determining gene that is highly conserved among vertebrates, during ovarian development of fish, reptiles, mammals, and chickens (Cocquet et al., 2003; Pannetier et al., 2006; Batista et al., 2007; Hudson et al., 2005; Baroiller et al., 2009; Guiguen et al., 2010). In the snapping turtle, a TSD species, expression of *Foxl2* and aromatase is significantly higher in gonads at a female-producing temperature than at a male-producing temperature (Rhen et al., 2007), suggesting a potential *Foxl2*-aromatase regulatory relationship similar to the one in mammals and chickens. In addition, DHT has a feminizing effect on developing snapping turtle embryos (Rhen and Lang, 1994; Rhen

and Schroeder, 2010). Expression of *Foxl2* and aromatase is higher in gonads from DHT treated embryos when compared to control embryos (Rhen and Schroeder, 2010). These findings suggest a novel AR-*Foxl2*-aromatase regulatory interaction during ovarian development in the snapping turtle.

This study tests the proposed AR-*Foxl2* interaction through cloning and analysis of the snapping turtle *Foxl2* promoter. We also tested whether androgen and *Foxl2* co-regulate other genes involved in ovary and follicle development, including follicle-stimulating hormone receptor (*Fshr*), gonadotropin releasing hormone receptor (*Gnrhr*), steroidogenic acute regulatory protein (*Star*), and aromatase (*Cyp19*) (Yamaguchi et al., 2007; Escudero et al., 2010; Pisarska et al., 2004). During ovarian development, *Foxl2* is exclusively expressed in granulosa cells (Garzo and Dorrington, 1984; Schmidt et al., 2004). The other genes examined here are also expressed in granulosa cells (Garzo and Dorrington, 1984; Schmidt et al., 2004; Pollack et al., 1997; Chu et al., 2002; Tetsuka and Hillier, 1996). We used the mouse granulosa cell line KK1 for our studies because turtle granulosa cell lines are not commercially available and because protocols have not been developed to isolate purified granulosa cells from turtles.

Materials and Methods

Cell Culture

KK1 granulosa cells (a gift from Dr. Joseph Marino, University of Toledo, OH, USA) were cultured in Dulbecco's Modified Eagle's Medium/Nutrient Mixture F-12 Ham (D6434 SIGMA) with 20% newborn calf serum (N4762 SIGMA), 10000 U/mL Penicillin-Streptomycin (P4333 SIGMA) and 365 mg/mL L-glutamine (G3126 SIGMA).

Cells frozen in liquid nitrogen were thawed and initially cultured in a 150mm petri dish in a humidified atmosphere of 5% CO2 at 37ºC to 100% confluence and then were dissociated and suspended using trypsin/EDTA solution (Life technologies). After cell isolation and wash, the cell-medium mixture was evenly dispensed to a 6-well cell culture plate. Cells in the plate were cultured under the same conditions until the confluence of cells in each plate reached 100%.

Foxl2-mCherry Vector Construction and Sequencing

The coding sequence and 1.6 kb upstream flanking region of *Foxl2* (i.e., proximal promoter) was cloned from the genome of the common snapping turtle, *Chelydra serpentina,* using inverse PCR (Ochman et al., 1988). Restriction sites for AseI and BamHI were added to the 5' end and 3' end of the clone with PCR primers: AseI sense 5'-CATGACATTAATGCTGTAGCTATAAACGACGGCTCA-3' and BamHI antisense 5'-ACATATGGATCCGAGATGTCTATCCGGGAGTGCAAG.

The PCR product was gel purified and digested with AseI and BamHI. The pEF1α-mCherry-N1 plasmid (Clontech) was also digested with AseI and BamHI, which removed the constitutive human EF1α promoter. The digested t*Foxl2* amplicon was then ligated into the cut mCherry plasmid. After plasmid ligation and bacterial transformation, the t*Foxl2*-mCherry fusion gene was sequenced to verify the position and orientation of the insert in the plasmid. This vector allows expression of the t*Foxl2*-mCherry fusion gene to be driven by the proximal turtle *Foxl2* promoter rather than the human EF1α promoter (Figure 4). The clone was sequenced using ABI PRISM® 3100 Genetic Analyzer and sequences were aligned using SEQUENCHER 5.3 and BioEdit v7.2.5.

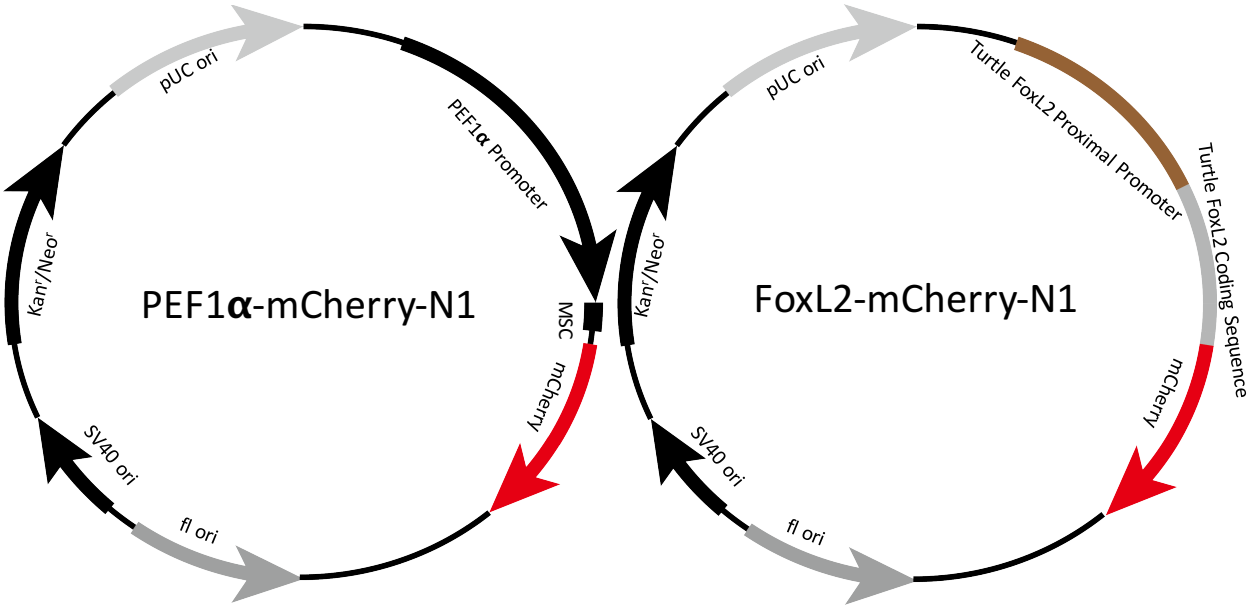


Figure 4. pEF1α-mCherry-N1 vector (left) and Foxl2-mCherry fusion construct (right). In the Foxl2-mCherry fusion construct, the pEF1α promoter was replaced with the proximal promoter of the snapping turtle Foxl2 gene (1,500 bp upstream from the start codon)

Phylogenetic Footprinting for The Foxl2 Proximal Promoter

The 1.6 kb proximal promoter of *Foxl2* was subjected to phylogenetic footprinting analysis which is used to identify regulatory elements conserved between different species. The same length of *Foxl2* proximal promoter from 3 turtle species (*Chelonia mydas*, *Chrysemys picta* and *Pelodiscus sinensis*) and 5 other vertebrates (*Monodelphis domestica*, *Anolis carolinensis*, *Alligator mississippiensis, Gallus gallus*, *Monodelphis_domestica* and *Mus musculus*) were compared to the snapping turtle *Foxl2* promoter. Sequences for each species were obtained from the National Center for Biotechnology Information (NCBI) database. The transcription start site (TSS) in the snapping turtle *Foxl2* promoter region was identified by a combination of read mapping from a separate RNA-Seq study and computational prediction using Promoter 2.0 Prediction Server (Rhen et al. 2015; Knudsen, 1999).

Sequences were aligned and the conservation of the region was evaluated with MEME suite 4.11.2, an expectation maximization-based motif-finding algorithm (Bailey et al. 2009). The minimum and maximum width of the motif were set to 4 and 30 respectively to reflect the widths of most established position weight matrices (Mathelier et al. 2016; Hume et al. 2015; Jolma et al. 2013; Matys et al. 2006). The motif E-value threshold was set to 1e-10 for highly significant motifs. To identify potential transcription factor binding sites (TFBS) in the conserved regions, we compared the resulting motifs against the Vertebrates database (*in vivo* and *in silico*) of known motifs using Tomtom in MEME suite. The resulted TFBSs with $P \leq 5e\text{-}3$ and E-value < 10 were considered as statistically significant for the *Foxl2* promoter. Androgen response elements (AREs) in *Foxl2* among 4 turtle species (*Chelydra serpentina*, *Chelonia mydas*, *Chrysemys picta* and *Pelodiscus sinensis*) were also predicted with PROMO (Messeguer et al. 2002).

Foxl2-mCherry Stable Transfection

Transfections were performed using Lipofectamine® 2000 Transfection Reagent from Life technologies. When cell confluence reached 100%, 1ml Opti-MEM medium (Catalog number 11058021 ThermoFisher Scientific) with 2.5µg t*Foxl2*-mCherry plasmid and 5µl of lipofectamine 2000 (Catalog number 11668019 ThermoFisher Scientific) were added to each well on the plate. On another plate, the same amount of pEF1α-mCherry-N1 plasmid and lipofectamine 2000 were added to each well as a control. After 6h incubation, 2ml of DMEM medium with 10% NCS and 500 µg/ml G418 (Catalog number 11811023 ThermoFisher Scientific) were added to each well. Selection lasted for 1 week, during which the selecting medium was changed every 2

days. Weaker selection medium containing 200 µg/ml G418 was then used to maintain stably transfected cell lines containing the t*Foxl2*-mCherry fusion gene or mCherry alone.

DHT Treatment of Stably Transfected Cells

Stably transfected cells were equally distributed to the wells on a 24-well plate and grown to 100% confluence before treatment. DHT was dissolved in 100% ethanol and was further diluted with cell culture medium to four final concentrations, i.e. 0nM, 1nM, 10nM and 100nM. Each row on the 24-well plate was subject to one DHT concentration (6 wells per dose). We treated two plates of cells with *Foxl2*-mCherry fusion plasmid and two plates of cells with pEF1α-mCherry-N1 plasmid. Culture medium was supplemented with NCS or charcoal-stripped NCS for a fully factorial design (2 plasmids x 2 types of serum x 4 DHT doses = 16 treatment groups). NCS was charcoal-stripped using a previously described protocol (Cao et al., 2009). We examined six biological replicates for each combination of DHT dose, NCS treatment, and plasmid (16 treatment groups x 6 biological replicates = 96 samples). Cells were incubated at 37ºC for 48h before collection.

RNA Extraction and cDNA Synthesis

Transfected cells were visualized with Olympus IX70 Fluorescence Microscope and pictures were taken before cells were lysed for RNA extraction. Cells were lysed and total RNA was extracted with RNAzol®RT (Molecular Research Center, Inc.). The average concentration of total RNA extracted from each of the 96 wells (4 x 24-well plates) was 60 ng/µl. The A260/A280 ratio was between 1.8 and 2.0 for all samples. We used 50 ng RNA as template in quantitative PCR reactions to test for genomic DNA contamination. No amplification was observed from RNA template in any sample,

indicating that RNA was pure. We then used 200 ng of pure RNA from each sample for cDNA synthesis with the High-Capacity cDNA Reverse Transcription Kit (Life technologies).

Quantitative PCR and Statistics

Primers for mCherry, mouse *Foxl2*, mouse aromatase (*CYP19*), mouse *Gnrhr*, mouse FshR, mouse StAR, and 18S rRNA were designed using Primer Express v2.0 software (Table 1). Standard curves for absolute quantitative measurement of gene expression were made as previously described (Rhen et al., 2007). qPCR was performed using SsoFast™ EvaGreen®Supermix and CFX384 Touch™ Real-Time PCR Detection System (BIO-RAD).

Table 1: Primers for quantitative PCR

mCherry forward	GACTACTTGAAGCTGTCCTTCC
mCherry reverse	CGCAGCTTCACCTTGTAGAT
Mouse *Foxl2* forward	GCTATGGCTACCTGGCGC
Mouse *Foxl2* reverse	GAGTTGTTGAGGAACCCCGAT
Mouse aromatase forward	CCTGACACCATGTCGGTCAC
Mouse aromatase reverse	GGATTGCTGCTTCGACCTCT
Mouse *Gnrhr* forward	GCCATCAACAACAGCATCCC
Mouse *Gnrhr* reverse	CGGTCACTCGGATCTTTCCA
Mouse *Fshr* forward	AAAGTGAGCATCTGCCTGCC
Mouse *Fshr* reverse	TTGAGTACGAGGAGGGCCATA
Mouse *Star* forward	CTCGAGACTTCGTGAGCGTG
Mouse *Star* reverse	AAATGTGTGGCCATGCCTG

We used JMP for all statistical analyses. We used a three-way ANCOVA with DHT dose, serum (normal versus charcoal-stripped), and plasmid (t*Foxl2*-mCherry plasmid versus pEF1α-mCherry-N1 plasmid) as independent variables. Expression of 18S rRNA was used as a covariate to control for random sample-to-sample variation in efficiency of RNA extraction and/or cDNA synthesis. Differences were considered statistically significant when $P < 0.05$. We used the Dunn-Sidák correction for multiple

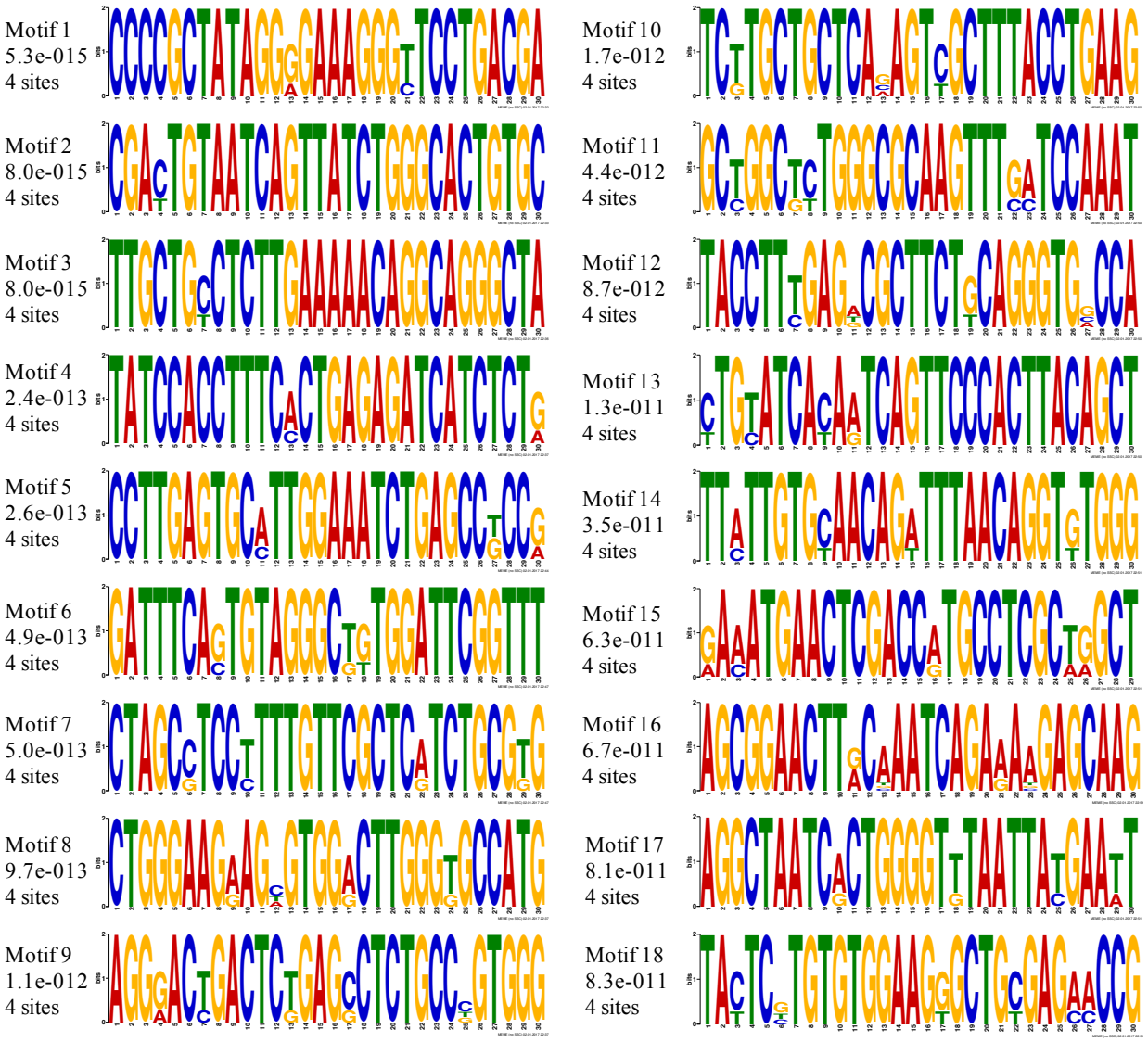
comparisons: the nominal significance level was $\alpha'=1-(1-\alpha)^{1/k}$, where k is the number of comparisons for an experiment wise $\alpha=0.05$. Sample sizes are shown in each figure. All means are presented as least squares means $\pm$ 1 standard error from the ANCOVA analysis.

Results

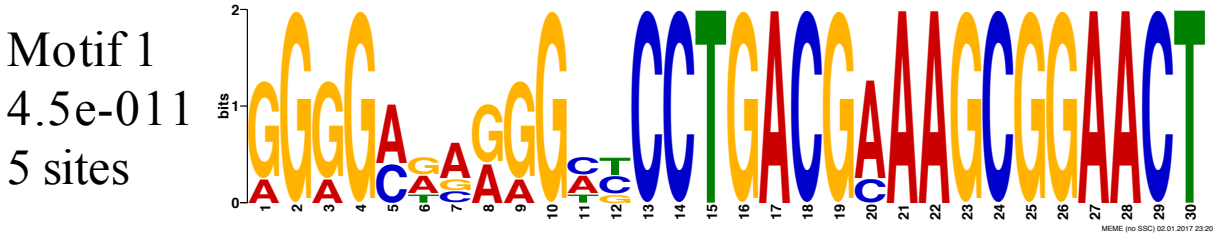
Potential Transcription Factor Binding Sites in The Foxl2 Proximal Promoter

Phylogenetic footprinting is a technique used to identify regulatory elements within a non-coding region of DNA sequence by comparing it to orthologous sequences in different species. This technique assumes important regulatory elements are conserved between species because they are required for gene expression. The coding sequence for *Foxl2*, known for its critical role in ovarian development, is highly conserved among vertebrates. Therefore, identification of conserved elements in its promoter may help reveal the regulation of this gene and clarify interactions between *Foxl2* and sex steroids.

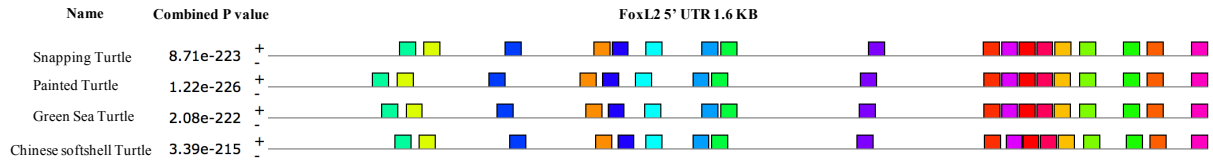
Motifs detected by MEME in the 1.6 kb *Foxl2* promoter from 6 species and 4 turtle species are shown in Figure 5A and Figure 5B, respectively. Relative positions of the motifs in the 1.6 kb *Foxl2* promoter across the species and the combined significance of motif co-occurrence are shown in Figure 5C and Figure 5D. Motifs discovered in closely related species (4 turtle species) showed similar distribution patterns and frequency (Figure 5C). However, there were fewer motifs when comparing the sequences between distantly related species and the distribution of the motifs also varied (Figure 5D).


Motif 1
5.3e-015
4 sites
Motif 2
8.0e-015
4 sites
Motif 3
8.0e-015
4 sites
Motif 4
2.4e-013
4 sites
Motif 5
2.6e-013
4 sites
Motif 6
4.9e-013
4 sites
Motif 7
5.0e-013
4 sites
Motif 8
9.7e-013
4 sites
Motif 9
1.1e-012
4 sites
Motif 10
1.7e-012
4 sites
Motif 11
4.4e-012
4 sites
Motif 12
8.7e-012
4 sites
Motif 13
1.3e-011
4 sites
Motif 14
3.5e-011
4 sites
Motif 15
6.3e-011
4 sites
Motif 16
6.7e-011
4 sites
Motif 17
8.1e-011
4 sites
Motif 18
8.3e-011
4 sites

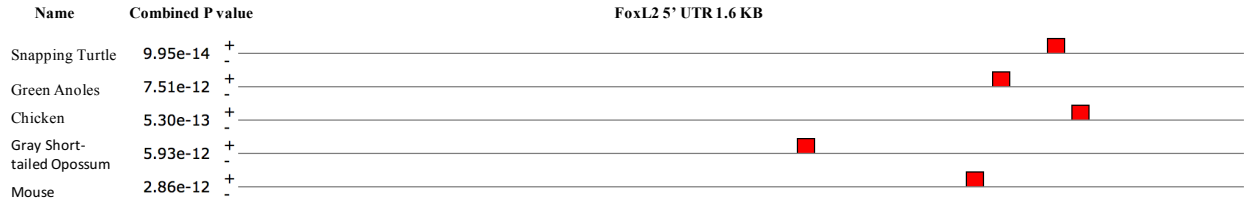
A


Motif 1
4.5e-011
5 sites
bits
MEME (no SSC) 02.01.2017 23:20

B



C

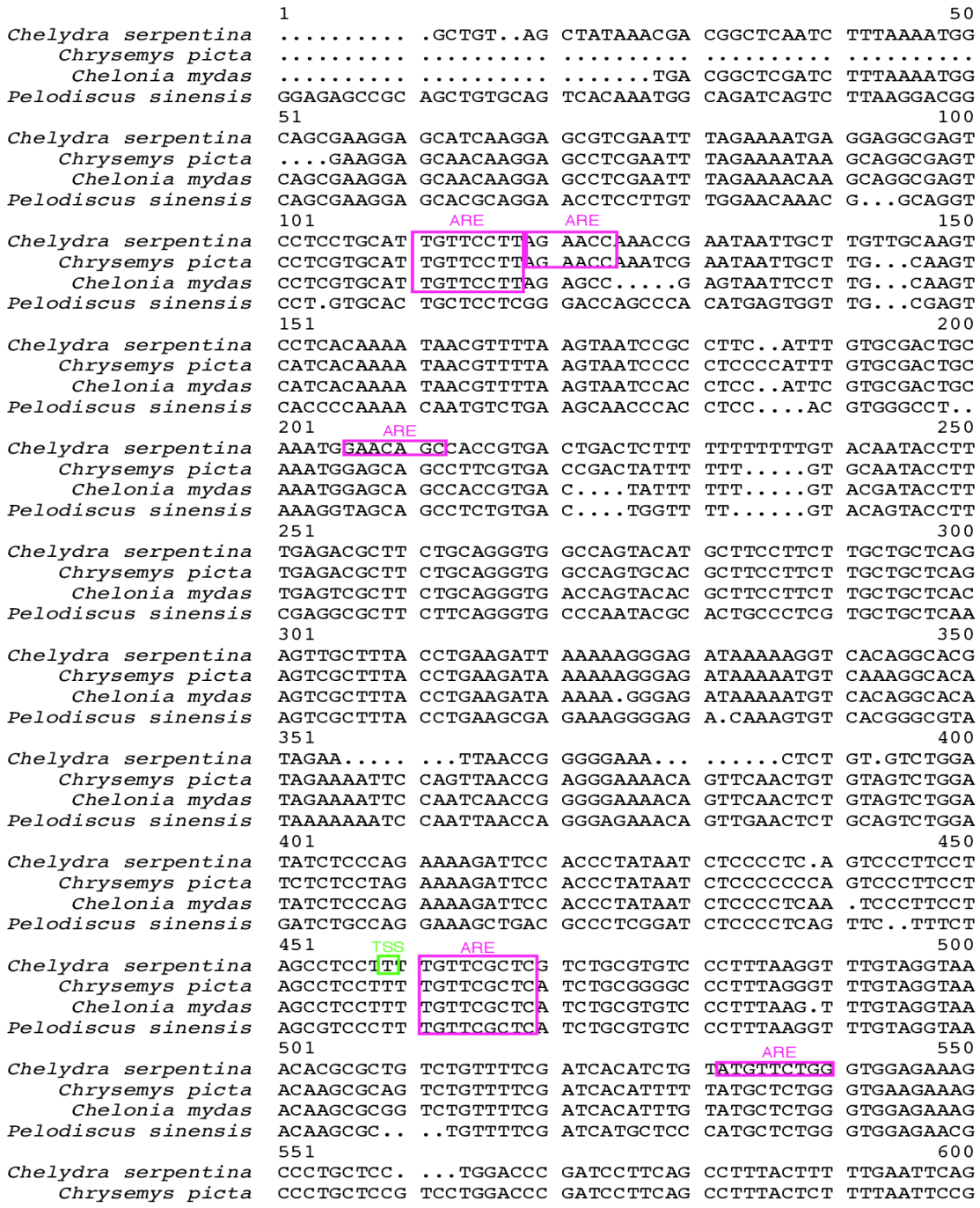


D

Figure 5. Identification of potential binding motifs by phylogenetic footprinting of 1.6 kb upstream regulatory regions of Snapping turtle Foxl2 *gene. (A) By comparing Foxl2 1.6 kb upstream sequences across 4 turtle species (snapping turtle, painted turtle, green sea turtle and Chinese softshell turtle), MEME identified 18 phylogenetically conserved and statistically significant (indicated by e-value) motifs. The number of sites contributing to their identification were also displayed. These motifs were displayed as sequence LOGOs representing position weight matrices of each possible letter code occurring at particular position of motif and its height representing the probability of the letter at that position multiplied by the total information content of the stack in bits. (B) One phylogenetically conserved and statistically significant motif was identified bycomparing* Foxl2 *1.6 kb upstream sequences across 6 species (snapping turtle, green anoles, chicken, gray short-tailed opossum and mouse). (C) Location of 18 motifs identified and their distribution in 1.6 kb upstream sequences across turtle* Foxl2 *and its orthologs in 3 other turtle species were shown in the block diagram. The combined best matches of a sequence to a group of motifs were shown by combined p value. Sequence strand specified as "+" (input sequence was read from left to right) and "-" (input sequence was read on its complementary strand from right to left) with respect to the occurrence of motifs. (D) Location of 1 motif identified and their distribution in 1.6 kb upstream sequences across turtle* Foxl2 *and its orthologs in 5 other species.*

All predicted TFBSs residing in the conserved motifs are shown in Table 2. High confidence set of TFBSs predicted to regulate the expression of *Foxl2* with their

associated transcription factors included sites for *Irf, FoxO, Etv, Pax, Esr, Hox, Sry* and *Sox*. Predicted AREs in *Foxl2* are shown in Figure 6.



```
     Chelonia mydas  CCCTGCTCC. ...TGGACCT GATTCTTCAG CCTTTACTCT TTGAATTCAG
Pelodiscus sinensis  TCCATCCCC. ...CGCACCA AATCCGTCAG CCTTTACTTT TCCAAATCGG
                     601                                                650
Chelydra serpentina  TGTTTTTT.A CTCCTGTGTG GAAGGGCTGC GAG....... CCCCGTTTTA
    Chrysemys picta  TGTTTTTT.A CTCGTGTGTG GAAGGGCTGC GAGAACCGAG CCCCGTTTTC
     Chelonia mydas  GGTTTTTTTA CTCTTGTGTG GAAGTGCTGC GAGAACCGAG CCCCGTTTTA
Pelodiscus sinensis  GGTTTTT..A TTCGTGTGTG GAAGGGCTGT GAGAACCGGG CCCCCATTTA
                     651                                                700
Chelydra serpentina  TTGTGCAACA GATTTAACAG GTGTGGGTTG GGTTTTTTGT TTGTTTTTAC
    Chrysemys picta  TTGTGCAACA GATTTAACAG GTGTGGGTTG GGATTTT.GT TTGTTTTTAC
     Chelonia mydas  TTGTGTAACA GTTTTAACAG GTGTGGGTTG GGTTTTTTGT TTGTTTTTAC
Pelodiscus sinensis  TTGTGCAACA GATTTAACAG GTTTGGGG.. ........GA TGGTTGTTAC
                     701                                                750
Chelydra serpentina  TTTTTT.... .CGACTGTAA TCAGTTATCT GGGCACTGTG CCAAGACT..
    Chrysemys picta  TTTTTT.... .CGACTGTAA TCAGTTATCT GGGCACTGTG CCAGGACTCT
     Chelonia mydas  TTTTTTTTTT TCGACTGTAA TCAGTTATCT GGGCACTGTG CCAAGACTCT
Pelodiscus sinensis  TTTTTT.... .CGATTGTAA TCAGTTATCT GGGCACTGTG CCAAAA....
                     751                                                800
Chelydra serpentina  AGTGAGTATA CATGCAAATG TACATATAGA CATAGACATC TGAGAATGGG
    Chrysemys picta  AGTGAGTATA AATGCAAATG TATATATAGA CATAGACATC TGAGAATGGA
     Chelonia mydas  AGTGAGTATA AATGTAAATG TACAGATAGG CAGAGACATC TGAGAATGGG
Pelodiscus sinensis  ...GGGAAAA AATGCGAATG TACGTACACA CATGAAC... ........GG
                     801                                                850
Chelydra serpentina  GAAAGAGTGC CTTGTATCAC AATCAGTTCC CACTTACAGC TATTGATTTC
    Chrysemys picta  GAAAGAGTAC CCTGTATCAC AATCAGTTCC CACTTACAGC TATTGATTTC
     Chelonia mydas  GAAAGAGTAC CCTGTATCAT AATCAGTTCC CACTTACAGC TATTGATTTC
Pelodiscus sinensis  GAAAGCGTCC TCTGCATCAC AGTCAGTTCC CACTTACAGC TGTGGATTTC
                     851                                                900
Chelydra serpentina  ACTGTAGGGC TGTGGATTCG GTTTCTAGGC TGGGATTGGA AG....TTTG
    Chrysemys picta  AGTGTAGGGC TGTGGATTCG GTTTCTAGGC TGGGAGTGGA AG....TTTG
     Chelonia mydas  AGTGTAGGGC GGTGGATTCG GTTTCTAGGC TGGGAGTGGA AGGAAGTTTG
Pelodiscus sinensis  AGTGTAGGGC TTTGGATTCG GTTTCTAAAC TGGGATTATA CG....TTTG
                     901                                                950
Chelydra serpentina  CAGTTTACTG CACAGTAGGC ATGTTTGAGC TCACTTCAGT GGTGCTAAAA
    Chrysemys picta  CAGCTTACTG CACAGTAGGC ATGTTTGAGC TCACTTCCCT GGTGCTAAAA
     Chelonia mydas  CAGCTTACTG CACAGTAGGC ATGTTTGAGC TCATTTCCCT GGTGCTAAAA
Pelodiscus sinensis  CAGTTTACTG CAAGATAGGC ATGTG.GAAC TCACTTTCCT GGTGC....A
                     951                                               1000
Chelydra serpentina  TGAATGCCTG TGCCAGCCAA GCCCAAATAA CATGGTAGGT TTGCCATGAC
    Chrysemys picta  TGAATGCCTG TGCCAGCCAA ACCCAAATAA CGTGGTAGGT TTGCAATGAA
     Chelonia mydas  TGAATGCCTA TGCCAGCCAA ACCCAAATAA CATGGTAGGT TTGCAATGAA
Pelodiscus sinensis  CGCCTGTCA. .GCCAAACAA ACCCAAAGAA CACGCTAGAT TTGCCACGAG
                     1001                                              1050
Chelydra serpentina  ATTTGGTAGC TCC..TTTAC CGCTGATAAT ATGGTAGCCT TCGGAACCTC
    Chrysemys picta  ATTTGGTAGC TTCCTTTTAC TGCTGATAGT ATGGTAGCCT TCAGAACCTC
     Chelonia mydas  ATTTGGACGC TTCCTTTTAC TGCTGATAG. .......... ....AACCTC
Pelodiscus sinensis  ATGAGGTTGC TTCCTTTTAC TGGCGATAAT ATGGTAGCCC TCGGGACCTC
                     1051                                              1100
Chelydra serpentina  ACGTTACTGT AGTACTTAGC ATTTTCCCCT CACGTAACTG GACATCTGGG
    Chrysemys picta  ATGTTACTAT AGTACTTAGC ATTTTCCCCT CACTTAACTG GACATCTGGG
     Chelonia mydas  AAGTTACTGT AA.......C ATTTCCCCCT CACTTAACTG GACACCCGGG
Pelodiscus sinensis  ATGTGCCTAT AGTATTTAGC ATTTCCCCCT CACTTGGTTG GACACCTGGA
                     1101                                              1150
Chelydra serpentina  .TAATAATGT ATCCACCTTT CACTGAGAGA TCATCTCTGT CTATCAGGAC
    Chrysemys picta  .TAATAATTT ATCCACCTTT CACTGAGAGA TCATCTCTGT CTATCAGGAC
     Chelonia mydas  .TAATAATTT ATCCACCTTT CACTGAGAGA TCATCTCT.. ..ATCAGGAC
Pelodiscus sinensis  ATGATAATTT ATCCACCTTT CCCTGAGAGA TCATCTCTGA CCATCAGGAA
                     1151                                              1200
```

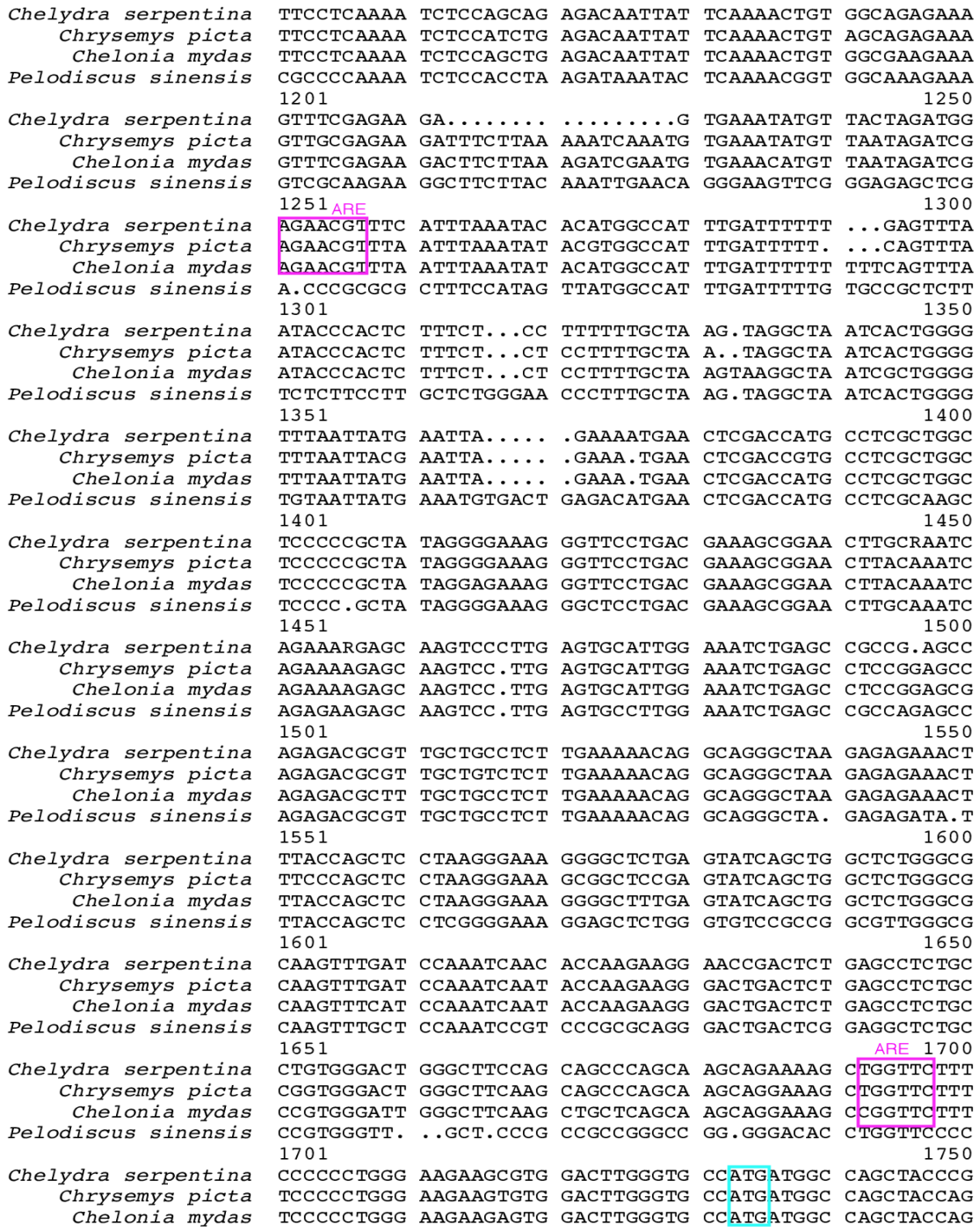


Figure 6. Predicted androgen response elements (AREs) in Foxl2 *promoter region and coding sequence among 4 turtle species (Chelydra serpentina, Chelonia mydas, Chrysemys picta and Pelodiscus sinensis)*

Turtle Promoter Drives tFoxl2-mCherry Expression at Physiological Levels

The t*Foxl2*-mCherry fusion gene was expressed at a much lower level (177 ± 7 ag of mRNA/2.5 ng total RNA) than the pEF1α-mCherry-N1 control plasmid (4,118 ± 194 ag of mRNA/2.5 ng total RNA), which contains the constitutive human EF1α promoter. The difference in mRNA levels between plasmids translated to the protein level (Figure 7). The turtle promoter drove expression of the t*Foxl2*-mCherry fusion gene (177 ± 7 ag/2.5 ng total RNA) at a level very similar to the endogenous mouse *Foxl2* gene (195 ± 11 ag/2.5 ng total RNA). The mCherry protein was distributed throughout the cytoplasm and the nucleus (Figure 7B). In contrast, the t*Foxl2*-mCherry fusion protein was primarily found in the nucleus (Figure 7D). Thus, the t*Foxl2*-mCherry fusion gene was expressed in a manner comparable to the endogenous mouse *Foxl2* gene.

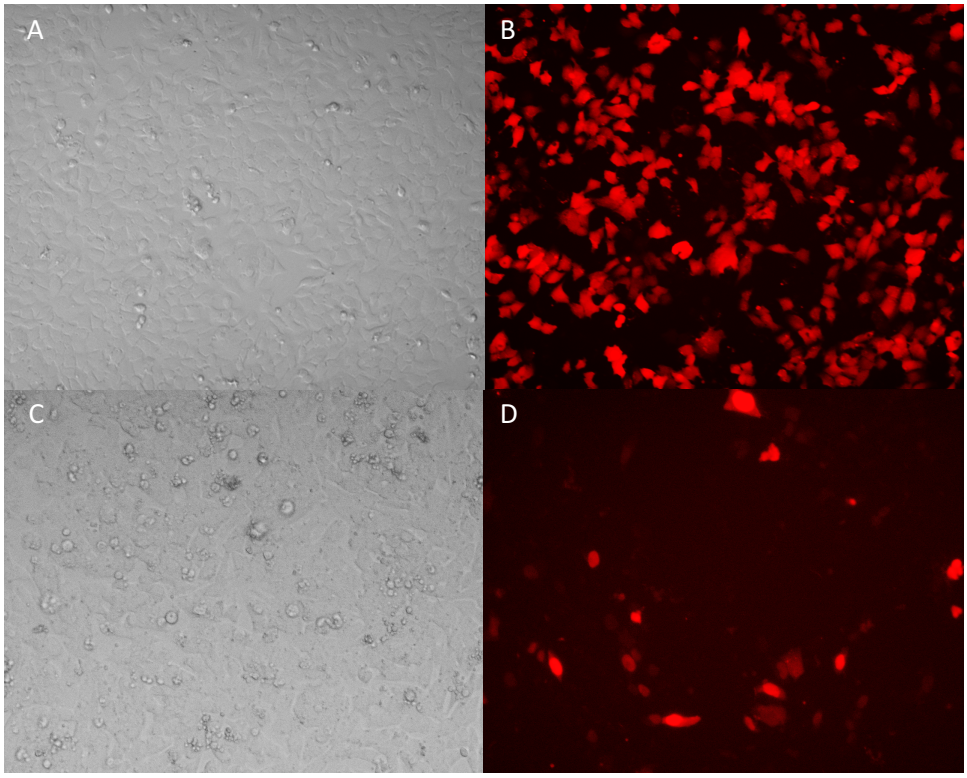


Figure 7. KK1 cells transfected with pEF1α-mCherry-N1 (A, B) and turtle Foxl2*-mCherry (C, D). Cells confluence was 100% at transfection (A, C). The pEF1α-mCherry-N1 plasmid was strongly expressed throughout the entire cell while the* Foxl2*-mCherry plasmid was only expressed in cell nuclei.*

Androgen and Serum Effects on tFoxl2-mCherry Expression

As described above, there was a highly significant difference in expression between the two plasmids ($F_{1,77}$ = 8,484, $p < 0.0001$). However, DHT treatment had no detectable effect on expression of either plasmid: DHT dose ($F_{3,77}$ = 2.45, p = 0.07), DHT dose x plasmid interaction ($F_{3,77}$ = 0.08, p = 0.97), and DHT dose x plasmid x serum interaction ($F_{3,77}$ = 0.79, p = 0.50).

On the other hand, the two promoters responded differently to normal versus charcoal-stripped serum: plasmid x serum interaction ($F_{1,77}$ = 77.1, $p <0.0001$). Normal serum increased expression of the t*Foxl2*-mCherry fusion gene (214 $\pm$ 7 ag/2.5 ng total RNA) when compared to stripped serum (182 $\pm$ 6 ag/2.5 ng total RNA). In contrast, normal serum decreased expression of the pEF1α-mCherry-N1 control gene (3162 $\pm$ 104 ag/2.5 ng total RNA) versus stripped serum (4677 $\pm$ 153 ag/2.5 ng total RNA). There was a significant DHT dose x serum interaction ($F_{3,77}$ = 7.1, $p < 0.0001$): DHT had no effect in the presence of normal serum, but slightly increased reporter expression at 1 nM and 100 nM doses in stripped serum. Levels of 18S rRNA were a significant covariate ($F_{1,77}$ = 929, $p < 0.0001$).

tFoxl2-mCherry, Androgen, and Serum Effects on Endogenous Foxl2 Expression

The t*Foxl2*-mCherry plasmid and the mCherry control plasmid had differential effects on the endogenous mouse *Foxl2* gene ($F_{1,78}$ = 7.3, p = 0.009). Expression of mouse *Foxl2* was significantly higher in cells transfected with t*Foxl2*-mCherry plasmid

(187 ± 5 ag/2.5 ng total RNA) compared to the mCherry plasmid (168 ± 5 ag/2.5 ng total RNA). Serum also had a significant effect on the endogenous mouse *Foxl2* gene ($F_{1,78}$ = 64.4, $p < 0.0001$). Mouse *Foxl2* was expressed at a lower level in cells exposed to normal serum (150 ± 5 ag/2.5 ng total RNA) compared to charcoal-stripped serum (205 ± 5 ag/2.5 ng total RNA). In contrast to the t*Foxl2*-mCherry fusion gene, DHT treatment had a significant effect on expression of the endogenous mouse *Foxl2* gene ($F_{3,78}$ = 3.33, p = 0.02). Although DHT generally decreased mouse *Foxl2* expression, only the highest dose had a significant effect after correction for multiple comparisons (Figure 8). Levels of 18S rRNA were a significant covariate ($F_{1,77}$ = 140, $p < 0.0001$). Effects of t*Foxl2*-mCherry, DHT, and serum on mouse *Foxl2* were independent of each other because their interactions were not significant (p's > 0.05).

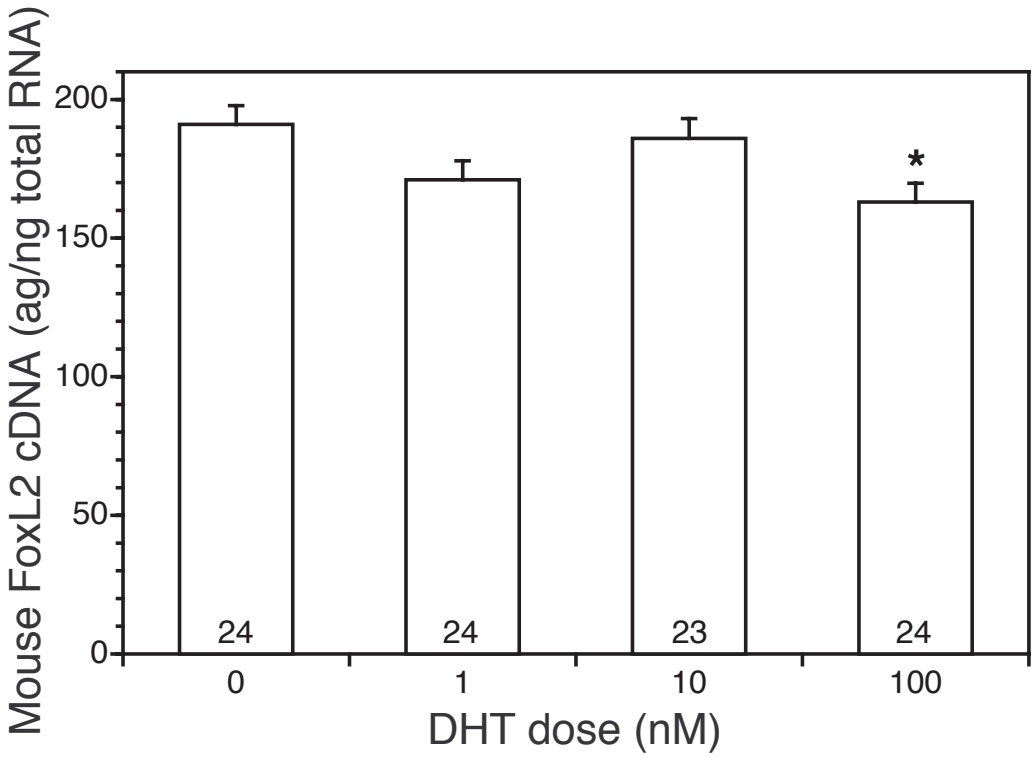


Figure 8. Androgen effects on expression of endogenous Foxl2 *in mouse granulosa (KK1) cells. Cells were grown to confluence and then treated with the indicated dose of DHT for 48 hours. Asterisk (*) indicates a significant difference from the untreated controls at $p \leq 0.017$. The significance level was adjusted for multiple comparisons using the Dunn-Sidák correction (3 DHT doses versus control). Data is presented as least squares means ± 1 SE from the ANCOVA described in the text. Sample sizes for each group are shown within the bars.*

Foxl2-mCherry, Androgen, and Serum Effects on Fshr and Gnrhr Expression

The t*Foxl2*-mCherry plasmid and the mCherry control plasmid had differential effects on *Fshr* expression ($F_{1,77} = 4.42$, $p = 0.04$). Expression of *Fshr* was higher in cells transfected with the t*Foxl2*-mCherry plasmid (135 $\pm$ 6 ag/2.5 ng total RNA) compared to the mCherry plasmid (116 $\pm$ 6 ag/2.5 ng total RNA). Levels of 18S rRNA were a significant covariate ($F_{1,77} = 4.53$, $p = 0.04$). *Fshr* expression was not affected by androgen treatment, the type of serum added to culture media, or interactions among these factors (p's > 0.05).

Serum ($F_{1,75} = 38.4$, $p < 0.0001$) and the serum x plasmid interaction ($F_{1,75} = 30.1$, $p < 0.0001$) influenced *Gnrhr* expression. The t*Foxl2*-mCherry fusion gene had opposite effects on *Gnrhr* expression in the presence of normal versus charcoal-stripped serum. In cells exposed to normal serum, the t*Foxl2*-mCherry plasmid decreased *Gnrhr* expression compared to the mCherry control (Figure 9). In stripped serum, the t*Foxl2*-mCherry plasmid increased *Gnrhr* expression compared to the mCherry control (Figure 9). In contrast, serum had no effect on *Gnrhr* expression in cells transfected with the mCherry control plasmid showing that the serum effect on *Gnrhr* was mediated by t*Foxl2*. No other main effects or interactions influenced *Gnrhr* expression (p's > 0.05).

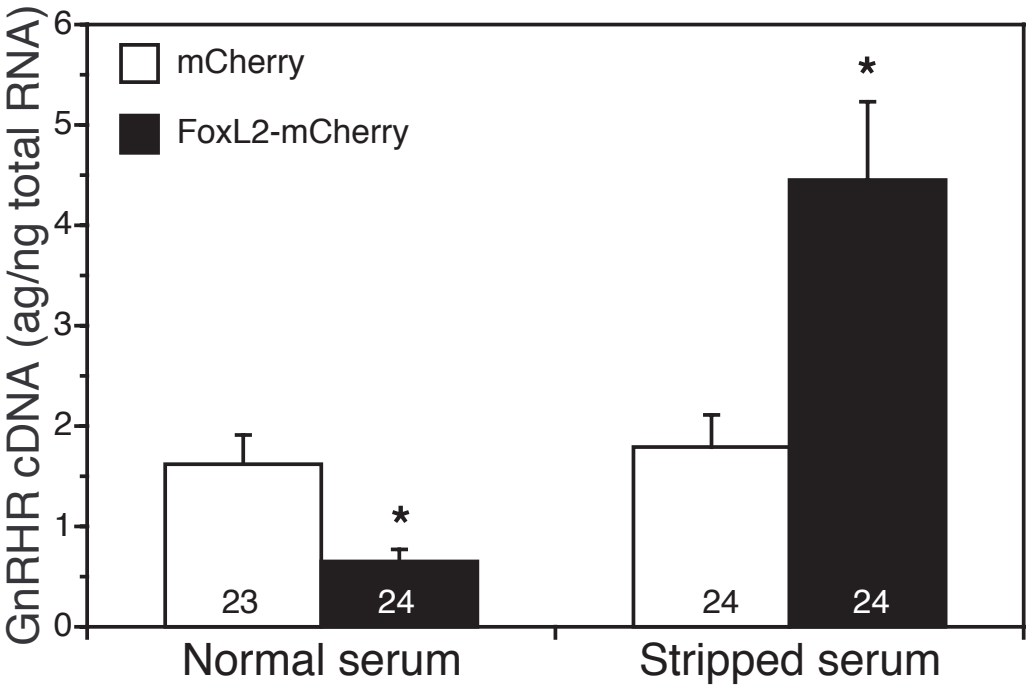


Figure 9. Effects of serum and turtle Foxl2-*mCherry fusion gene on* Gnrhr *expression in mouse granulosa (KK1) cells. Cells were stably transfected with mCherry control plasmid or turtle* Foxl2-*mCherry plasmid and then cultured in normal serum or charcoal stripped serum for 48 hours. Asterisks (*) indicate a significant difference between* Foxl2-*mCherry and the mCherry control (within serum type). The significance level was adjusted for multiple comparisons using the Dunn-Sidák correction ($p \leq 0.017$ for three comparisons;* Foxl2-*mCherry versus mCherry in normal serum;* Foxl2-*mCherry versus mCherry in stripped serum; normal versus stripped serum for mCherry controls). Data is presented as least squares means ± 1 SE from the ANCOVA described in the text. Sample sizes for each group are shown within the bars.*

tFoxl2-mCherry, Androgen, and Serum Effects on Star and Cyp19 Expression

All three treatments had significant effects on *Star* expression: plasmid ($F_{1,78}$ = 22.9, $p < 0.0001$), serum ($F_{1,78} = 808$, $p < 0.0001$), and DHT dose ($F_{3,78} = 11.0$, $p < 0.0001$). On average, *Star* expression was 2.6 times higher in cells exposed to charcoal-stripped versus normal serum (Figures 10 and 11). More importantly, there were significant interactions between plasmid and serum ($F_{1,78} = 68.1$, $p < 0.0001$) and between DHT and serum ($F_{3,78} = 4.3$, $p = 0.008$). The t*Foxl2*-mCherry fusion gene had

different effects on *Star* expression in the presence of normal versus stripped serum. In normal serum, the t*Foxl2*-mCherry plasmid had no effect on *Star* expression when compared to the mCherry control (Figure 10). In stripped serum, the t*Foxl2*-mCherry plasmid significantly decreased *Star* expression compared to the mCherry control (Figure 10). DHT had no effect on *Star* expression when cells were cultured in normal serum (Figure 11). However, the 1nM dose of DHT significantly increased *Star* expression when cells were in stripped serum (Figure 11). Some component of normal serum blocked t*Foxl2*-mCherry and DHT effects on *Star* expression. No other interactions influenced *Star* expression (p's > 0.05). Levels of 18S rRNA were a significant covariate ($F_{1,78} = 184$, $p < 0.0001$).

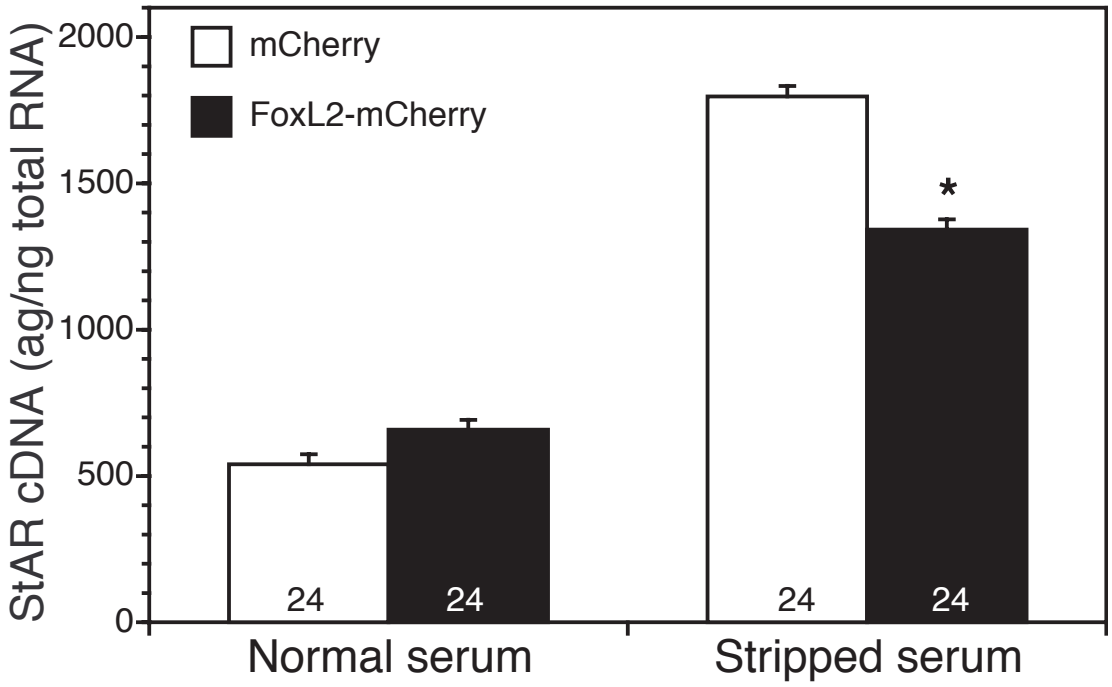


Figure 10. Effects of serum and turtle Foxl2-*mCherry fusion gene on* Star *expression in mouse granulosa (KK1) cells. Cells were stably transfected with mCherry control plasmid or turtle* Foxl2-*mCherry plasmid and then cultured in normal serum or charcoal stripped serum for 48 hours. Asterisk (*) indicates a significant difference between* Foxl2-*mCherry and the mCherry control within serum type (2 comparisons). The significance level was adjusted for multiple comparisons using the Dunn-Sidák correction ($p \leq 0.006$ for 8 total comparisons). Data is presented as least squares means ± 1 SE from the ANCOVA described in the text.*

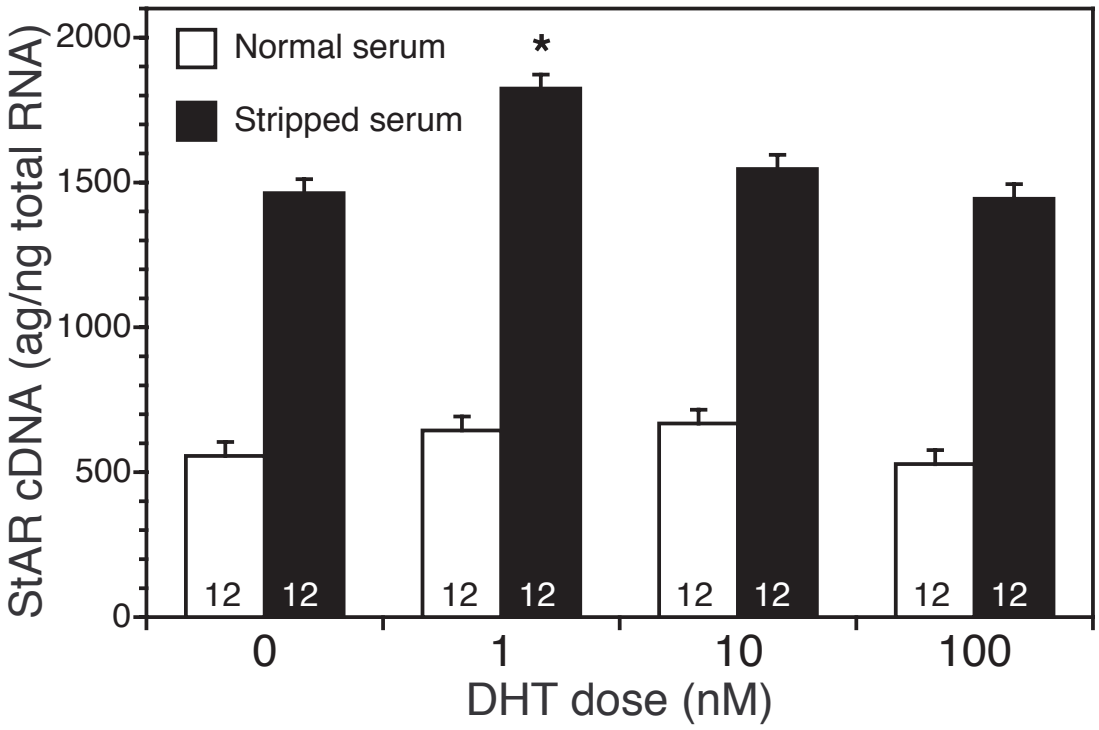


Figure 11. Effects of serum and DHT treatment on Star *expression in mouse granulosa (KK1) cells. Cells were cultured in normal serum or charcoal stripped serum and treated with DHT at various concentrations for 48 hours. Asterisk (*) indicates a significant difference between the indicated DHT dose and the untreated control within serum type (6 comparisons). The significance level was adjusted for multiple comparisons using the Dunn-Sidák correction ($p \leq 0.006$ for 8 total comparisons). Data is presented as least squares means ± 1 SE from the ANCOVA described in the text. Sample sizes for each group are shown within the bars.*

All three treatments and all their interactions influenced *Cyp19* expression: plasmid ($F_{1,78} = 241$, $p < 0.0001$), serum ($F_{1,78} = 169$, $p < 0.0001$), DHT dose ($F_{3,78} = 21.4$, $p < 0.0001$), plasmid x serum ($F_{1,78} = 54$, $p < 0.0001$), plasmid x DHT dose ($F_{3,78} = 5.2$, $p = 0.003$), serum x DHT dose ($F_{3,78} = 4.6$, $p = 0.005$), and plasmid x serum x DHT

dose ($F_{3,78} = 3.6$, $p = 0.017$). Levels of 18S rRNA were a significant covariate ($F_{1,78} = 47.6$, $p < 0.0001$). These complex interactions can be summarized as follows. The t*Foxl2*-mCherry plasmid increased *Cyp19* expression and made cells more responsive to stripped serum and to the lowest DHT doses. On average, the t*Foxl2*-mCherry fusion gene doubled *Cyp19* expression in comparison to the mCherry control (compare the same serum/DHT treatments in Figure 12A versus Figure 12B). Serum and DHT did not influence *Cyp19* expression in cells transfected with the mCherry control plasmid after correcting for multiple comparisons (Figure 12A). In stark contrast, stripped serum significantly increased expression of *Cyp19* in cells transfected with the t*Foxl2*-mCherry plasmid (arrows in Figure 12B). The 1nM and 10 nM doses of DHT significantly increased expression of *Cyp19* in cells transfected with t*Foxl2*-mCherry and cultured in stripped serum (black bars with asterisks in Figure 12B). DHT treatments had no detectable effect in cells incubated in normal serum (white bars in Figure 12B).

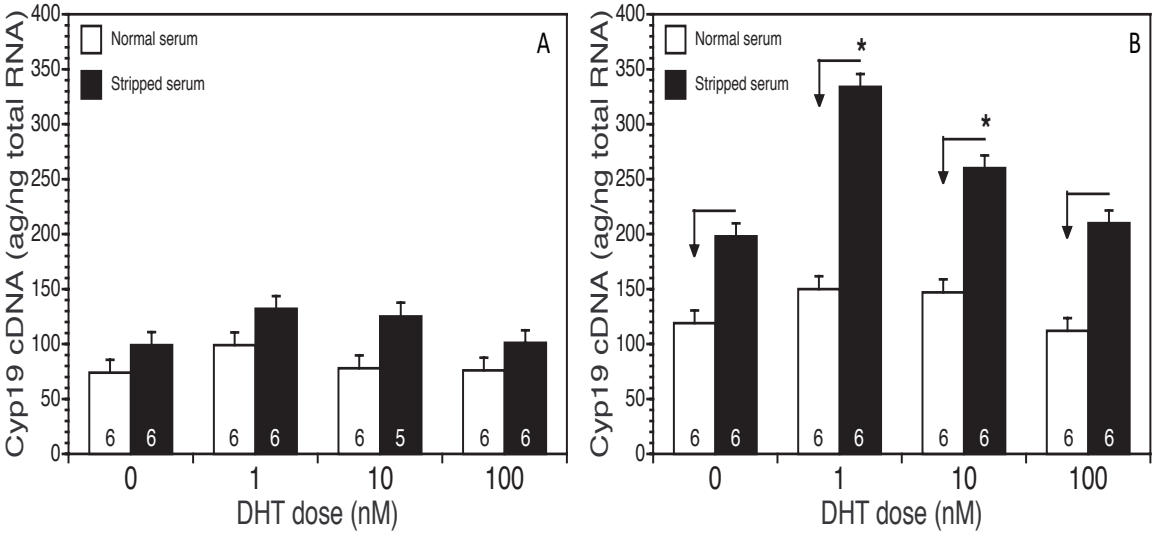


Figure 12. Effects of plasmid, serum, and androgen treatment on Cyp19 *expression in mouse granulosa (KK1) cells. Cells were stably transfected with (A) mCherry control plasmid or (B) turtle* Foxl2*-mCherry plasmid. These cells were then cultured in normal serum or charcoal stripped serum and treated with DHT at various concentrations for 48 hours. Asterisks (*) indicate a significant difference between the indicated DHT dose and the untreated control within serum type. Arrows indicate a significant difference between cells in stripped serum versus normal serum at the same DHT dose. The significance level was adjusted for multiple comparisons using the Dunn-Sidák correction ($p \leq 0.0026$ for 20 total comparisons; 3 DHT doses versus controls x 2 plasmids x 2 serum types = 12 comparisons; stripped versus normal serum x 4 DHT doses x 2 plasmids = 8 comparisons). Data is presented as least squares means ± SE from the ANCOVA described in the text. Sample sizes are shown within the bars.*

Discussion

Androgen action has been well studied in male reproductive development and prostate cancer (Quigley et al., 1995). However, the direct involvement of AR and androgen in female reproduction was not firmly established until recently (Walters, 2015). One source of confusion in studying androgen action in ovarian development is that some androgens, such as testosterone, can be converted into estrogens, which act via ERs. This problem was solved by creating AR knockout mouse models (ARKO) or using a non-aromatizable androgen like DHT. Studies based on ARKO mice have found that androgens in granulosa cells regulate follicle development and function (Sen and Hammes, 2010; Cheng et al., 2013). In these studies, granulosa cell-specific ARKO female mice had altered estrus cycles, produced fewer oocytes and displayed reduced fertility. To the contrary, oocyte-specific AR-null female mice had normal fertility and follicle morphology at early ages, indicating that AR-mediated effects on follicle development are confined to granulosa cells. The essential role of androgens in ovarian development has been confirmed in ARKO mice, but the molecular mechanism and gene regulatory cascade behind AR-mediated effects are still unclear.

In reptiles with TSD, androgens not only influence ovarian development but also play a role in sex determination. Treatment of red-eared slider turtles with DHT during the TSP can induce testis development at temperatures that produce mixed sex ratios (Ramsey and Crews, 2009). On the contrary, DHT treatments in snapping turtles during the TSP induce ovary development at temperatures that produce males or mixed sex ratios (Rhen and Lang, 1994; Rhen and Schroeder, 2010; Schroeder and Rhen, 2017). The feminizing effect of DHT in snapping turtle embryos is associated with induction of both *Foxl2* and aromatase, suggesting these genes are androgen targets. Furthermore, co-administration of the androgen receptor antagonist flutamide completely blocked DHT induction of *Foxl2*, suggesting that this effect is specifically mediated by AR.

Here we examine two main hypotheses. First, we tested the hypothesis that androgens regulate *Foxl2* expression. Second, we tested the hypothesis that androgens and *Foxl2* co-regulate *Fshr*, *Gnrhr*, *Star*, and *Cyp19* expression in granulosa cells. Our results partially support the hypothesis that androgens influence expression of *Foxl2*. The t*Foxl2*-mCherry reporter construct containing the snapping turtle *Foxl2* proximal promoter was not affected by DHT treatment, but the endogenous mouse *Foxl2* gene was significantly influenced by DHT treatment. There are several potential explanations for this difference. Putative AREs were identified in the snapping turtle *Foxl2* gene (Figure. 6), but key regulatory elements could be in more distal enhancers. Second, genomic and chromatin context may influence the activity of AREs in the cloned region of *Foxl2*. A third hypothesis is that there are species differences in androgen regulation of *Foxl2* (i.e., up-regulation in turtle versus down-regulation in mouse). Although we did not elucidate the mechanism underlying DHT induction of the snapping turtle *Foxl2* gene that we

previously observed in embryos, we found that the t*Foxl2*-mCherry fusion gene and DHT co-regulate *Star* and *Cyp19*, which are two key steroidogenic genes.

Foxl2 belongs to the Forkhead box family of transcription factors, which share a common DNA binding domain approximately 110 amino acids long (Carlsson and Mahlapuu, 2002). *Foxl2* plays important roles in many biological processes, such as apoptosis (Carlsson and Mahlapuu, 2002), cell differentiation (Cocquet et al., 2003), eyelid morphogenesis (Crisponi et al., 2001), female somatic cell sex determination (Uhlenhaut et al., 2009), and granulosa cell differentiation (Schmidt et al., 2004). *Foxl2* is highly conserved in vertebrates, such as human, goat, mouse, chicken, turtle and pufferfish, both at protein level and nucleotide level (Cocquet et al., 2003; Loffler et al., 2003). During embryonic development of vertebrates, *Foxl2* is expressed at the earliest stage of ovary differentiation (Loffler et al., 2003). Although *Foxl2* is not required for early ovarian development in mice (Uda et al., 2004), it is required for ovary determination and normal ovarian development in goats (Pailhoux et al., 2005). In addition, *Foxl2* interacts with estrogen receptor to suppress *Sox9* through the *cis*-regulatory sequence TESCO in mice thereby preventing trans-differentiation of ovaries to testes, suggesting cross-talk between steroids and *Foxl2* (Uhlenhaut et al., 2009). In slider turtles, expression of *Foxl2* coincides with ovarian determination (Loffler et al., 2003). In snapping turtles, expression of *Foxl2* in the developing gonads increases at a female determining temperature precisely when ovarian fate is determined (Rhen et al., 2007; Rhen et al., 2015).

Previous research has shown that DHT treatments increase *Foxl2* expression in gonads during snapping turtle embryogenesis (Rhen and Schroeder, 2010; Schroeder and

Rhen, 2017). Yet, the snapping turtle proximal promoter did not respond to DHT treatments when the t*Foxl2*-mCherry plasmid was transfected into mouse KK1 cells. In contrast, the endogenous *Foxl2* gene in mouse KK1 cells was down-regulated by the highest DHT dose. This contradiction may be explained by species differences (mouse vs. turtle) in the proximal promoter, regulatory elements outside the cloned fragment, or differences in chromatin context. It is also possible the state of cell differentiation (undifferentiated embryonic cells vs. immortalized granulosa cell line) or developmental environment (*in vivo* vs. *in vitro*) contribute to this difference. The potential impact of DHT on *Foxl2* expression, either positively or negatively, merits further study. The effect of DHT on *Foxl2* expression may be achieved through direct AR binding to AREs in the *Foxl2* gene. Alternatively, AR could alter *Foxl2* expression indirectly through interaction with other transcription factors (or co-regulators) that bind to regulatory sequences in the *Foxl2* gene (McKenna et al., 1999; Robyr et al., 2000). We cannot exclude the possibility that AR regulates *Foxl2* expression through an indirect pathway (i.e., by regulating expression of a gene that in turn influences *Foxl2* expression).

One of our most interesting and novel findings was that t*Foxl2*-mCherry potentiated the effect of the lowest DHT dose on *Cyp19* expression when cells were cultured in stripped serum. Expression of *Cyp19* was 2x higher than would be expected if *Foxl2*-mCherry, 1 nM DHT, and stripped serum had strictly additive effects. The level of potentiation was not as strong at 10 nM DHT (1.75x higher than expected) and was no longer significant at 100 nM (1.44x higher than expected). *Cyp19* encodes aromatase, which converts androgens to estrogens. Aromatase and estrogens have been shown to play a key role in sex determination in many TSD species (Pieau and Dorizzi, 2004; Rhen

and Schroeder, 2010). Stimulation of *Cyp19* expression by DHT in our experiment is consistent with studies in other species, such as fish (González et al., 2015), mice (Roselli and Resko, 1984), and humans (Eriksen et al., 2014). Activation of *Cyp19* by the t*Foxl2*-mCherry fusion gene mirrors findings in other vertebrates (Pannetier et al., 2006; Batista et al., 2007; Wang et al., 2007; Baroiller et al., 2009; Guiguen et al., 2010). However, the current study is the first to explicitly test for interactions between androgens and *Foxl2*. Synergistic regulation of *Cyp19* by low doses of DHT and *Foxl2*, as observed here, could have implications for understanding TSD in the snapping turtle. A small increase in androgen synthesis in gonads at female-producing temperatures could synergize with *Foxl2* to activate *Cyp19* and estrogen synthesis, thereby inducing ovarian development. A test of this hypothesis will require development of techniques for isolation and efficient transfection of primary cells from embryonic turtle gonads.

In addition to this hypothetical feed-forward mechanism, androgens may act through a positive feedback loop to increase steroidogenesis via activation of *Star* expression. *Star* encodes a protein that plays a critical role in steroid synthesis by regulating the delivery of cholesterol from the outer mitochondrial membrane to the inner mitochondrial membrane where the cholesterol side chain cleavage enzyme resides (Kallen et al., 1998). It is postulated that *Star* may stimulate follicle development by regulating production of steroids, i.e. androgens and estrogens (Ronen-Fuhrmann et al., 1998). Here we show that 1nM DHT increased *Star* expression in mouse granulosa cells. In agreement with this finding, DHT increases *Star* expression in embryonic turtle gonads (Schroeder and Rhen, 2017).

The current study demonstrated that some basic features of snapping turtle *Foxl2* gene expression and protein function were similar to mammalian *Foxl2*. We found that t*Foxl2*-mCherry mRNA was expressed at the same level as mouse *Foxl2* mRNA, suggesting that the turtle proximal promoter for *Foxl2* was functional in mouse granulosa cells. When the t*Foxl2*-mCherry plasmid or the pEF1α-mCherry-N1 plasmid were stably transfected into KK1 cells, the t*Foxl2*-mCherry fusion protein was exclusively found in nuclei while mCherry was found in both cytoplasm and nuclei. Thus, sub-cellular localization of the t*Foxl2*-mCherry fusion protein is the same as reported for mammalian *Foxl2* protein. The current study also confirms some other important regulatory interactions in granulosa cells. The t*Foxl2*-mCherry plasmid increased expression of the endogenous mouse *Foxl2* gene, supporting a previous report of positive feedback by *Foxl2* (Benayoun et al., 2009). We examined two other *Foxl2* target genes, *Fshr* and *Gnrhr*, to test whether t*Foxl2*-mCherry could also regulate their expression. We found that t*Foxl2*-mCherry was able to induce *Fshr*, which supports previous studies (Escudero et al., 2010; Fortin et al., 2014). The t*Foxl2*-mCherry plasmid also regulated *Gnrhr* expression, but the direction of the effect depended on whether cells were in normal or charcoal-stripped serum. The *Foxl2*-mCherry plasmid repressed *Gnrhr* in normal serum, but induced *Gnrhr* in stripped serum. The latter result is consistent with reports that *Foxl2* can activate a *Gnrhr* promoter-luciferase reporter plasmid (Escudero et al., 2010). Taken together, these results suggest activity of turtle *Foxl2* protein is not altered by mCherry at its carboxyl terminus.

Comparison of gene expression in cells cultured with normal versus charcoal-stripped serum suggests the presence of a signaling factor that dramatically modulates

Foxl2 activity. For instance, t*Foxl2*-mCherry repressed *Gnrhr* in normal serum, but induced *Gnrhr* in charcoal-stripped serum. Likewise, t*Foxl2*-mCherry had no effect on *Star* expression in normal serum, but repressed *Star* in stripped serum. The latter finding is consistent with a previous report that *Foxl2* can directly repress the activity of the *Star* promoter (Pisarska et al., 2004). Finally, t*Foxl2*-mCherry had a much weaker effect on *Cyp19* expression in normal serum versus stripped serum. NCS contains a large number of factors, including steroid hormones, vitamins, enzymes, and chemicals that are either removed or decreased in concentration by charcoal stripping (Cao et al., 2009). It is not clear which of these components interacts with *Foxl2* to alter gene expression, but at least one clearly changes *Foxl2* activity.

Serum effects could be due to post-translational modification of *Foxl2*, changes in expression of other genes that interact with *Foxl2*, or changes in expression of *Foxl2* itself (Benayoun et al., 2008; Caburet et al., 2012). Phylogenetic footprinting revealed potential TFBSs in the *Foxl2* promoter region (Table 2). These TFs are involved in a wide range of biological functions, such as immune response, sex determination, endocrine signaling, cell cycle, and cell death. It is possible that expression of *Foxl2* was directly influenced by serum components that trigger signaling cascades that impact transcription factor binding to the core *Foxl2* promoter and/or distal enhancers. In this regard, it is noteworthy that normal and charcoal-stripped serum had opposing effects on the t*Foxl2*-mCherry plasmid versus the endogenous mouse *Foxl2* gene. Normal serum increased expression of t*Foxl2*-mCherry, but decreased expression of mouse *Foxl2* when compared to charcoal-stripped serum. Serum effects on t*Foxl2*-mCherry expression could be due to CCAAT/enhancer-binding protein (C/EBPbeta) binding sites found in the turtle

proximal promoter. C/EBPbeta is known to interact with serum response factor to regulate expression of serum responsive genes (Hanlon and Sealy, 1999). Many studies have been done to identify *Foxl2* targets, but very few studies have examined factors that regulate *Foxl2* expression (Georges et al., 2014). This is an important area of study because relatively subtle changes in *Foxl2* expression (i.e., doubling by transfection of t*Foxl2*-mCherry) can have dramatic effects on expression of key target genes involved in follicle development and sex determination (i.e., t*Foxl2* potentiates low dose DHT effects on aromatase expression).

Our study demonstrates that androgens can influence expression of key ovarian genes and that snapping turtle *Foxl2* is capable of regulating these genes in mouse granulosa cells. The most interesting finding was that t*Foxl2*-mCherry potentiated the effect of low DHT doses on aromatase expression in mouse granulosa cells. It will be especially interesting to test whether this also occurs in embryonic gonads of snapping turtles. Interactions between androgens, *Foxl2*, and an un-identified serum factor(s) have a major impact on key genes in granulosa cells. The mechanisms underlying these interactions need further investigation.

Table 2: Predicted androgen response elements (AREs) in *Foxl2* promoter region and coding sequence among 4 turtle species (*Chelydra serpentina, Chelonia mydas, Chrysemys picta* and *Pelodiscus sinensis*)

Binding Sites	p-value	consensus
ZNF282_DBD	0.000670799	GTCGTGTTGTGGGAAAG
ZNF524_full_2	0.00187374	GGCACGGGTTCGAG
UP00082_2	0.00203645	CAAGGGACAAGGGCTC
UP00067_1	0.002219	GATAGATCAAAGGGATT
VENTX_DBD_2	0.00300249	CGCTAATCGGAAAACGATTAG
Zfp652_DBD	0.00444117	AGAAAGGGTTAAT
UP00225_1	0.00183095	TGTAATTAATTATGG
PAX7_DBD	0.00185559	TAATCGATTA
PAX7_full	0.00194736	TAATCGATTA
PAX3_DBD	0.00224748	TAATCGATTA

Binding Sites	p-value	consensus
UP00237_1	0.00276899	CGTAATTAATTAATTGG
UP00238_1	0.00481737	CAAAGTAATTAATTATC
LHX6_full_3	0.00483085	TGATTGCAATCA
SRY_DBD_4	0.00135944	TGAATAACATTCA
MA0473.1	0.0019003	GAACCAGGAAGTG
MA0080.3	0.0023949	AAAAAGAGGAAGTGA
SOX21_DBD_4	0.00315767	TGAATAACATTCA
UP00013_1	0.00324835	CAATACCGGAAGTGTAA
MAFF_DBD	0.0033235	TTGCTGACTCAGCAA
SOX21_DBD_2	0.0049467	AACAATGTGCAGTGTT
MA0520.1	0.00320808	CATTTCCTGAGAAAT
ZNF143_DBD	0.00360649	TACCCACAATGCATTG
MA0130.1	0.00449298	ATCCAC
MA0159.1	0.00497294	AGGTCATGGAGAGGTCA
MA0503.1	0.000267209	CTTGAGTGGCT
MA0483.1	0.00113754	AAATCACAGCA
EBF1_full	0.00147358	ATTCCCTTGGGAAT
ISL2_DBD	0.003613	TTAAGTGC
IRF5_full_2	7.60973e-05	TGGTTTCGGTT
MA0479.1	9.33388e-05	TGTGGATTGGA
HNF4A_DBD_1	0.00110783	GATGGACTTTGGACTC
UP00040_1	0.00121517	TTGGTTTCGGTTTAT
IRF4_full	0.00144589	TAGTTTCGGTTTCGG
HNF4A_full_1	0.00167004	ATTGGACTTTGGACCC
MA0073.1	0.00261415	TGGGGGGGGGTGGTTTGGGG
UP00018_1	0.00389881	TTTGGTTTCGATACG
MA0002.2	0.00408211	GTCTGTGGTTT
MA0048.1	0.000753874	GCGCAGCTGCGT
SCRT1_DBD	0.00334857	AACCACCTGTTGCTC
SOX9_DBD	0.00345584	CCATTGTTC
MA0006.1	0.0035267	TGCGTG
NHLH1_DBD	0.00400777	CGCAGCTGCG
MA0514.1	0.00400777	CCTTTGTTTT
MA0077.1	0.00402518	CCATTGTTC
NHLH1_full	0.00473576	CGCAGCTGCG
E2F4_DBD_1	0.00174063	TTTGGCGCCATT
TBX20_DBD_3	0.0018299	GAAAAGGTGTGAAAG
MA0472.1	0.0025938	GTGCGTGGGCGGGGG
SP3_DBD	0.00285632	GGGGGCGTGGC
NFIA_full_2	0.00285887	GGTGCCAAGT
MA0516.1	0.00325767	GGGAGGGGGCGGGGC
UP00023_1	0.00436741	ATTGAACAATGGAATT
HINFP1_full_3	0.00120756	GCGGACGTTGAACGTCCGC
MA0141.2	0.0024499	TGACCTTGACCT
HNF4A_DBD_2	0.00437758	AATGGACTTTGACCCC
MA0144.2	0.00458281	TTTCCCAGAAG
MA0486.1	0.00125509	CTTCTAGAAGGTTCT
UP00057_2	0.00401254	TCTCCTGCTGTGTGG
UP00110_1	0.0023658	TCGCTATAATTACCGAC
MA0076.2	0.00351737	CCACTTCCGGC

Binding Sites	p-value	consensus
MA0475.1	0.00466174	CCACTTCCTGT
NKX3-1_full	0.00495239	ACCACTTAA
MA0103.2	0.000128902	CAGGTGAGG
GRHL1_DBD_1	0.00018314	AACCGGTTTAACCGGTT
SNAI2_DBD	0.000417714	AACAGGTGT
MA0102.3	0.000419135	TATTGTGCAAT
TCF4_DBD	0.000511494	AGCAGGTGCG
MA0466.1	0.000549125	ATTGTGCAATA
TCF4_full	0.000812347	TGCAGGTGTG
TFCP2_full_2	0.00108134	ACCGGTTTAAACCGGT
TBX21_full_3	0.00111001	TCACACCTAAAAGGTGTGA
FIGLA_DBD	0.00122302	AACAGGTGGT
TCF3_DBD	0.001868	AGCAGGTGTT
TBX15_DBD_2	0.00231803	AGGTGTGA
TBX1_DBD_3	0.00248291	AGGTGTGA
CEBPE_DBD	0.00265123	ATTGCGCAAT
MGA_DBD_1	0.00266151	AGGTGTGA
TBX4_DBD_1	0.00266151	AGGTGTGA
TBX5_DBD_1	0.00266151	AGGTGTGA
CEBPG_DBD	0.00278576	ATTGCGCAAT
MESP1_DBD	0.00278576	CACAGGTGTT
CEBPB_DBD	0.00307859	ATTGCGCAAT
UP00075_1	0.00371862	TAGTGAACAATAGATTT
CEBPD_DBD	0.00377004	ATTGCGCAAT
MSC_full	0.00377004	AACAGCTGTT
CEBPG_full	0.00396845	ATTGCGCAAT
Meis2_DBD_2	0.00399541	TGACAGGTGTCA
MYBL2_DBD_2	0.00408352	TAACGGTTTTAACGGT
SCRT2_DBD	0.00408871	ATGCAACAGGTGG
MEIS3_DBD_2	0.00417883	TGACAGGTGTCA
PKNOX2_DBD	0.00417883	TGACAGGTGTCA
TEF_FL	0.00417883	TGTTATGTAATA
CEBPB_full	0.00463551	ATTGCGCAAT
Cebpb_DBD	0.00463551	ATTGCGCAAT
Pknox2_DBD	0.00499981	TGACAGGTGTCA
VDR_full	0.00162687	TGAACTCAATGAACTC
Vdr_DBD	0.00251476	TGAACCCGATGAACTC
MA0074.1	0.0029633	TGAACTCGTTGACCC
SOX2_DBD_1	0.00300392	GAACAATACCATTGTTC
ESRRG_full_3	0.00358013	ATGACCTTGA
Esrra_DBD_2	0.00414171	ATGACCTTGAA
UP00097_2	0.000407139	AAATAAGAAAAAAC
MA0081.1	0.00356427	AGAGGAA
POU1F1_DBD_2	0.00429983	AATATGCAAATTAG
MA0502.1	0.004302	AAATGGACCAATCAG
UP00078_1	0.000251712	GGGTTTAATTAAAATTC
UP00061_2	0.00156294	TGTTTTGTTTTGATAT
POU3F1_DBD_2	0.00207155	TAAATTATGCAT
POU3F3_DBD_2	0.00261962	TAATTTATGCAT
HOXB2_DBD	0.00290205	GTTAATTACT

Binding Sites	p-value	consensus
MA0038.1	0.00290205	CAAATCACTG
SOX8_full_3	0.00305989	AATCACTGCAATTGATT
HNF1B_full_2	0.00309434	AGTTAATCATTAACT
HOXB3_DBD	0.00320282	GCTAATTAGT
MA0075.1	0.00336157	AATTA
UP00229_1	0.00342703	GGAGGGGATTAATTTAT
SOX9_full_5	0.00369392	AATCACTGAAATTGATT
UP00132_1	0.00383441	AACGCTAATTAGCGGTG
HNF1A_full	0.0044573	AGTTAATCATTAACT
GSC_full	0.00452762	GCTAATCCCC
FOXO1_DBD_3	0.000167985	CGTGTGGGGAAA
FOXO4_DBD_3	0.000224106	CGTGTGGGGAAA
FOXO6_DBD_3	0.000262273	GTCGTGTGGGGAAA
FOXO3_full_3	0.000747644	GTGTGGGGAAA
IRF7_DBD_1	5.12017e-06	ACGAAAGCGAAAGT
IRF8_DBD	6.64622e-06	ACGAAACCGAAACT
IRF9_full	1.7807e-05	AACGAAACCGAAACT
IRF8_full	3.41553e-05	TCGAAACCGAAACT
IRF5_full_1	0.000100304	CCGAAACCGAAACT
IRF4_full	0.000163577	CCGAAACCGAAACTA
ETV6_full_1	0.000511526	CCGGAAGCGGAAGTG
SPI1_full	0.000819	AAAAAGCGGAAGTA
SPIB_DBD	0.000870785	AAAAAGCGGAAGTA
Spic_DBD	0.00133788	AAAAAGCGGAAGTA
MA0081.1	0.00147309	AGAGGAA
ETV6_full_2	0.00194891	AGCGGAAGTG
SPIC_full	0.00229075	AAAAAGAGGAAGTA
ETV2_DBD	0.0047249	AACCGGAAATA
UP00074_1	0.00491601	CAAAATCGAAACTAA

REFERENCES

Abbott, D.H., Barnett, D.K., Bruns, C.M., and Dumesic, D.A. (2005). Androgen excess fetal programming of female reproduction: a developmental aetiology for polycystic ovary syndrome? Hum. Reprod. Update *11*, 357–374.

Bailey, Timothy L, Mikael Boden, Fabian A Buske, Martin Frith, Charles E Grant, Luca Clementi, Jingyuan Ren, Wilfred W Li, and William S Noble. 2009. "MEME SUITE: Tools for Motif Discovery and Searching." Nucleic Acids Research 37 (Web Server issue). Oxford University Press: W202–8. doi:10.1093/nar/gkp335.

Baroiller, J.F., D'Cotta, H., and Saillant, E. (2009). Environmental effects on fish sex determination and differentiation. Sexual Development *3,* 118-135.

Batista, F., Vaiman, D., Dausset, J., Fellous, M., and Veitia, R.A. (2007). Potential targets of *FOXL2*, a transcription factor involved in craniofacial and follicular development, identified by transcriptomics. Proc. Natl. Acad. Sci. *104*, 3330–3335.

Benayoun, B.A., Batista, F., Auer, J., Dipietromaria, A., L'Hôte, D., Baere, E.D., and Veitia, R.A. (2009). Positive and negative feedback regulates the transcription factor *FOXL2* in response to cell stress: evidence for a regulatory imbalance induced by disease-causing mutations. Hum. Mol. Genet. *18*, 632–644.

Benayoun, B.A., Caburet, S., Dipietromaria, A., Bailly-Bechet, M., Batista, F., Fellous, M., Vaiman, D., and Veitia, R.A. (2008). The identification and characterization of a *FOXL2* response element provides insights into the pathogenesis of mutant alleles. Human Molecular Genetics *17*, 3118–3127

Britt, K.L., Drummond, A.E., Dyson, M., Wreford, N.G., Jones, M.E., Simpson, E.R., and Findlay, J.K. (2001). The ovarian phenotype of the aromatase knockout (ArKO) mouse. J. Steroid Biochem. Mol. Biol. *79*, 181–185.

Caburet, S., Georges, A., L'Hôte, D., Todeschini, A.-L., Benayoun, B.A., and Veitia, R.A. (2012). The transcription factor *FOXL2*: At the crossroads of ovarian physiology and pathology. Molecular and Cellular Endocrinology *356*, 55–64.

Cao, Z., West, C., Norton-Wenzel, C.S., Rej, R., Davis, F.B., Davis, P.J., and Rej, R. (2009). Effects of Resin or Charcoal Treatment on Fetal Bovine Serum and Bovine Calf Serum. Endocr. Res. *34*, 101–108.

Carlsson, P., and Mahlapuu, M. (2002). Forkhead Transcription Factors: Key Players in Development and Metabolism. Dev. Biol. *250*, 1–23.

Cheng, X.B., Jimenez, M., Desai, R., Middleton, L.J., Joseph, S.R., Ning, G., Allan, C.M., Smith, J.T., Handelsman, D.J., and Walters, K.A. (2013). Characterizing the neuroendocrine and ovarian defects of androgen receptor-knockout female mice. Am. J. Physiol. Endocrinol. Metab. *305*, E717–E726.

Chu, S., Rushdi, S., Zumpe, E.T., Mamers, P., Healy, D.L., Jobling, T., Burger, H.G., and Fuller, P.J. (2002). FSH-regulated gene expression profiles in ovarian tumours and normal ovaries. Mol. Hum. Reprod. *8*, 426–433.

Ciofi, C., and Swingland, I.R. (1997). Environmental sex determination in reptiles. Appl. Anim. Behav. Sci. *51*, 251–265.

Cocquet, J., De Baere, E., Gareil, M., Pannetier, M., Xia, X., Fellous, M., and Veitia, R.A. (2003). Structure, evolution and expression of the *FOXL2* transcription unit. Cytogenet. Genome Res. *101*, 206–211.

Conover, D.O., and Heins, S.W. (1987). Adaptive variation in environmental and genetic sex determination in a fish. Nature *326*, 496–498.

Couse, J.F., Hewitt, S.C., Bunch, D.O., Sar, M., Walker, V.R., Davis, B.J., and Korach, K.S. (1999). Postnatal sex reversal of the ovaries in mice lacking estrogen receptors alpha and beta. Science *286*, 2328–2331.

Couse, J.F., Curtis Hewitt, S., and Korach, K.S. (2000). Receptor null mice reveal contrasting roles for estrogen receptor alpha and beta in reproductive tissues. J. Steroid Biochem. Mol. Biol. *74*, 287–296.

Crews, D., and Bergeron, J.M. (1994). Role of reductase and aromatase in sex determination in the red-eared slider (Trachemys scripta), a turtle with temperature-dependent sex determination. J. Endocrinol. *143*, 279–289.

Crews, D., Cantú, A.R., Rhen, T., and Vohra, R. (1996). The Relative Effectiveness of Estrone, Estradiol-17β, and Estriol in Sex Reversal in the Red-Eared Slider (Trachemys scripta), a Turtle with Temperature-Dependent Sex Determination. Gen. Comp. Endocrinol. *102*, 317–326.

Crisponi, L., Deiana, M., Loi, A., Chiappe, F., Uda, M., Amati, P., Bisceglia, L., Zelante, L., Nagaraja, R., Porcu, S., et al. (2001). The putative forkhead transcription factor *FOXL2* is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. Nat. Genet. *27*, 159–166.

Dupont, S., Dennefeld, C., Krust, A., Chambon, P., and Mark, M. (2003). Expression of Sox9 in granulosa cells lacking the estrogen receptors, ERalpha and ERbeta. Dev. Dyn. Off. Publ. Am. Assoc. Anat. *226*, 103–106.

Eriksen, M.B., Glintborg, D., Nielsen, M.F.B., Jakobsen, M.A., Brusgaard, K., Tan, Q., and Gaster, M. (2014). Testosterone treatment increases androgen receptor and aromatase gene expression in myotubes from patients with PCOS and controls, but does not induce insulin resistance. Biochemical and Biophysical Research Communications *451*, 622–626.

Escudero, J.M., Haller, J.L., Clay, C.M., and Escudero, K.W. (2010). Microarray analysis of *Foxl2* mediated gene regulation in the mouse ovary derived KK1 granulosa cell line: Over-expression of *Foxl2* leads to activation of the gonadotropin releasing hormone receptor gene promoter. J. Ovarian Res. *3*, 4.

Fisher, C.R., Graves, K.H., Parlow, A.F., and Simpson, E.R. (1998). Characterization of mice deficient in aromatase (ArKO) because of targeted disruption of the *Cyp19* gene. Proc. Natl. Acad. Sci. U. S. A. *95*, 6965–6970.

Fortin, J., Boehm, U., Deng, C.-X., Treier, M., and Bernard, D.J. (2014). Follicle-stimulating hormone synthesis and fertility depend on SMAD4 and *FOXL2*. FASEB J. *28*, 3396–3410.

Garzo, V.G., and Dorrington, J.H. (1984). Aromatase activity in human granulosa cells during follicular development and the modulation by follicle-stimulating hormone and insulin. Am. J. Obstet. Gynecol. *148*, 657–662.

Georges, A., Auguste, A., Bessiere, L., Vanet, A., Todeschini, A.-L., and Veitia, R.A. (2014). *Foxl2*: a central transcription factor of the ovary. J. Mol. Endocrinol. *52,* R17-R33.

González, A., Fernandino, J.I., and Somoza, G.M. (2015). Effects of 5α-dihydrotestosterone on expression of genes related to steroidogenesis and spermatogenesis during the sex determination and differentiation periods of the pejerrey, *Odontesthes bonariensis*. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology *182*, 1–7.

Guiguen, Y., Fostier, A., Piferrer, F., and Chang, C.-F. (2010). Ovarian aromatase and estrogens: a pivotal role for gonadal sex differentiation and sex change in fish. Gen. Comp. Endocrinol. *165,* 352-366.

Hanlon, M., and Sealy, L. (1999). Ras regulates the association of serum response factor and CCAAT/enhancer-binding protein beta. J. Biol. Chem. 274, 14224-14228.

Holleley, C. E., D. O'Meally, S. D. Sarre, J. A. Marshall Graves, T. Ezaz, et al., 2015. Sex reversal triggers the rapid transition from genetic to temperature-dependent sex. Nature 523:79-82.

Hudson, Q.J., Smith, C.A., and Sinclair, A.H. (2005). Aromatase inhibition reduces expression of *FOXL2* in the embryonic chicken ovary. Dev. Dyn. *233*, 1052–1055.

Hume, Maxwell A, Luis A Barrera, Stephen S Gisselbrecht, and Martha L Bulyk. 2015. "UniPROBE, Update 2015: New Tools and Content for the Online Database of Protein Binding Microarray Data on Protein-DNA Interactions." *Nucleic Acids Research* 43 (Database issue): D117–22. doi:10.1093/nar/gku1045.

Janzen, F.J., and Paukstis, G.L. (1991). Environmental Sex Determination in Reptiles: Ecology, Evolution, and Experimental Design. Q. Rev. Biol. *66*, 149–179.

Jolma, Arttu, Jian Yan, Thomas Whitington, Jarkko Toivonen, Kazuhiro R Nitta, Pasi Rastas, Ekaterina Morgunova, et al. 2013. "DNA-Binding Specificities of Human Transcription Factors." *Cell* 152 (1-2): 327–39. doi:10.1016/j.cell.2012.12.009.

Kallen, C.B., Billheimer, J.T., Summers, S.A., Stayrook, S.E., Lewis, M., and Strauss, J.F. (1998). Steroidogenic Acute Regulatory Protein (StAR) Is A Sterol Transfer Protein. J. Biol. Chem. *273*, 26285–26288.

Katoh, H., Ogino, Y., and Yamada, G. (2006). Cloning and expression analysis of androgen receptor gene in chicken embryogenesis. FEBS Lett. *580*, 1607–1615.

Knudsen, S. (1999). Promoter2.0: for the recognition of PolII promoter sequences. Bioinformatics, 15(5), 356–361. http://doi.org/10.1093/bioinformatics/15.5.356

Loffler, K.A., Zarkower, D., and Koopman, P. (2003). Etiology of ovarian failure in blepharophimosis ptosis epicanthus inversus syndrome: *FOXL2* is a conserved, early-acting gene in vertebrate ovarian development. Endocrinology *144*, 3237–3243.

Manolakou, P., Lavranos, G., and Angelopoulou, R. (2006). Molecular patterns of sex determination in the animal kingdom: a comparative study of the biology of reproduction. Reprod. Biol. Endocrinol. *4*, 59.

Mathelier, Anthony, Oriol Fornes, David J Arenillas, Chih Yu Chen, Grégoire Denay, Jessica Lee, Wenqiang Shi, et al. 2016. "JASPAR 2016: a Major Expansion and Update of the Open-Access Database of Transcription Factor Binding Profiles." *Nucleic Acids Research* 44 (D1). Oxford University Press: D110–15. doi:10.1093/nar/gkv1176.

Matys, V, O V Kel-Margoulis, E Fricke, I Liebich, S Land, A Barre-Dirrie, I Reuter, et al. 2006. "TRANSFAC and Its Module TRANSCompel: Transcriptional Gene Regulation in Eukaryotes." Nucleic Acids Research 34 (Database issue). Oxford University Press: D108–10. doi:10.1093/nar/gkj143.

McKenna, N.J., Lanz, R.B., and O'Malley, B.W. (1999). Nuclear receptor coregulators: cellular and molecular biology. Endocr. Rev. *20*, 321–344.

Messeguer, Xavier, Ruth Escudero, Domènec Farré, Oscar Núñez, Javier Martínez, and M Mar Albà. 2002. "PROMO: Detection of Known Transcription Regulatory Elements Using Species-Tailored Searches.." Bioinformatics 18 (2): 333–34.

Ochman, H., Gerber, A.S., and Hartl, D.L. (1988). Genetic applications of an inverse polymerase chain reaction. Genetics *120*, 621–623.

Pailhoux, E., Vigier, B., Schibler, L., Cribiu, E.P., Cotinot, C., and Vaiman, D. (2005). Positional cloning of the PIS mutation in goats and its impact on understanding mammalian sex-differentiation. Genet. Sel. Evol. *37*, S55.

Pannetier, M., Fabre, S., Batista, F., Kocer, A., Renault, L., Jolivet, G., Mandon-Pepin, B., Cotinot, C., Veitia, R., and Pailhoux, E. (2006). *Foxl2* activates P450 aromatase gene transcription: towards a better characterization of the early steps of mammalian ovarian development. J. Mol. Endocrinol. *36,* 399-413.

Pieau, C., and Dorizzi, M. (2004). Oestrogens and temperature-dependent sex determination in reptiles: all is in the gonads. J. Endocrinol. *181*, 367–377.

Pisarska, M.D., Bae, J., Klein, C., and Hsueh, A.J.W. (2004). Forkhead L2 is expressed in the ovary and represses the promoter activity of the steroidogenic acute regulatory gene. Endocrinology *145*, 3424–3433.

Pollack, S.E., Furth, E.E., Kallen, C.B., Arakane, F., Kiriakidou, M., Kozarsky, K.F., and Strauss, J.F. (1997). Localization of the steroidogenic acute regulatory protein in human tissues. J. Clin. Endocrinol. Metab. *82*, 4243–4251.

Quigley, C.A., Bellis, A.D., Marschke, K.B., El-Awady, M.K., Wilson, E.M., and French, F.S. (1995). Androgen Receptor Defects: Historical, Clinical, and Molecular Perspectives. Endocr. Rev. *16*, 271–321.

Radder, R.S., Quinn, A.E., Georges, A., Sarre, S.D., and Shine, R. (2008). Genetic evidence for co-occurrence of chromosomal and thermal sex-determining systems in a lizard. Biology Letters *4*, 176–178.

Ramsey, M., and Crews, D. (2007). Steroid signaling system responds differently to temperature and hormone manipulation in the red-eared slider turtle (Trachemys scripta elegans), a reptile with temperature-dependent sex determination. Sex. Dev. *1*, 181–196.

Ramsey, M., and Crews, D. (2009). Steroid signaling and temperature-dependent sex determination—Reviewing the evidence for early action of estrogen during ovarian determination in turtles. Semin. Cell Dev. Biol. *20*, 283–292.

Rhen, T., and Lang, J.W. (1994). Temperature-dependent sex determination in the snapping turtle: Manipulation of the embryonic sex steroid environment. Gen. Comp. Endocrinol. *96*, 243–254.

Rhen, T., and Schroeder, A. (2010). Molecular Mechanisms of Sex Determination in Reptiles. Sex. Dev. *4*, 16–28.

Rhen, T., Fagerlie, R., Schroeder, A., Crossley, D.A., and Lang, J.W. (2015). Molecular and morphological differentiation of testes and ovaries in relation to the thermosensitive period of gonad development in the snapping turtle, Chelydra serpentina. Differentiation *89*, 31–41.

Rhen, T., Metzger, K., Schroeder, A., and Woodward, R. (2007). Expression of putative sex-determining genes during the thermosensitive period of gonad development in the snapping turtle, Chelydra serpentina. Sex. Dev. *1*, 255–270.

Robyr, D., Wolffe, A.P., and Wahli, W. (2000). Nuclear hormone receptor coregulators in action: diversity for shared tasks. Mol. Endocrinol. *14*, 329–347.

Ronen-Fuhrmann, T., Timberg, R., King, S.R., Hales, K.H., Hales, D.B., Stocco, D.M., and Orly, J. (1998). Spatio-Temporal Expression Patterns of Steroidogenic Acute Regulatory Protein (StAR) During Follicular Development in the Rat Ovary. Endocrinology *139*, 303–315.

Roselli, C.E., and Resko, J.A. (1984). Androgens Regulate Brain Aromatase Activity in Adult Male Rats through a Receptor Mechanism. Endocrinology *114*, 2183–2189.

Schmidt, D., Ovitt, C.E., Anlag, K., Fehsenfeld, S., Gredsted, L., Treier, A.-C., and Treier, M. (2004). The murine winged-helix transcription factor *Foxl2* is required for granulosa cell differentiation and ovary maintenance. Dev. Camb. Engl. *131*, 933–942.

Schroeder, A.L., Metzger, K.J., Miller, A., and Rhen, T. (2016). A Novel Candidate Gene for Temperature-Dependent Sex Determination in the Common Snapping Turtle. Genetics genetics.115.182840.

Sen, A., and Hammes, S.R. (2010). Granulosa cell-specific androgen receptors are critical regulators of ovarian development and function. Mol. Endocrinol. *24*, 1393–1403.

Setiawan, A.N., Ozaki, Y., Shoae, A., Kazeto, Y., and Lokman, P.M. (2012). Androgen-specific regulation of FSH signalling in the previtellogenic ovary and pituitary of the New Zealand shortfinned eel, Anguilla australis. Gen. Comp. Endocrinol. *176*, 132–143.

Tetsuka, M., and Hillier, S.G. (1996). Androgen receptor gene expression in rat granulosa cells: the role of follicle-stimulating hormone and steroid hormones. Endocrinology *137*, 4392–4397.

Uda, M., Ottolenghi, C., Crisponi, L., Garcia, J.E., Deiana, M., Kimber, W., Forabosco, A., Cao, A., Schlessinger, D., and Pilia, G. (2004). *Foxl2* disruption causes mouse ovarian failure by pervasive blockage of follicle development. Hum. Mol. Genet. *13*, 1171–1181.

Uhlenhaut, N.H., Jakob, S., Anlag, K., Eisenberger, T., Sekido, R., Kress, J., Treier, A.-C., Klugmann, C., Klasen, C., Holter, N.I., et al. (2009). Somatic Sex Reprogramming of Adult Ovaries to Testes by *FOXL2* Ablation. Cell *139*, 1130–1142.

Walters, K.A. (2015). Role of androgens in normal and pathological ovarian function. Reproduction *149*, R193–R218.

Wang, D.-S., Kobayashi, T., Zhou, L.-Y., Paul-Prasanth, B., Ijiri, S., Sakai, F., Okubo, K., Morohashi, K., and Nagahama, Y. (2007). *Foxl2* up-regulates aromatase gene transcription in a female-specific manner by binding to the promoter as well as interacting with ad4 binding protein/steroidogenic factor 1. Mol. Endocrinol. *21*, 712–725.

Wibbels, T., and Crews, D. (1992). Specificity of steroid hormone-induced sex determination in a turtle. J. Endocrinol. *133*, 121–129.

Wibbels, T., and Crews, D. (1995). Steroid-Induced Sex Determination at Incubation Temperatures Producing Mixed Sex Ratios in a Turtle with TSD. General and Comparative Endocrinology *100*, 53–60.

Wibbels, T., Bull, J.J., and Crews, D. (1991). Synergism between temperature and estradiol: a common pathway in turtle sex determination? J. Exp. Zool. *260*, 130–134.

Wibbels, T., Gideon, P., Bull, J.J., and Crews, D. (1993). Estrogen- and temperature-induced medullary cord regression during gonadal differentiation in a turtle. Differ. Res. Biol. Divers. *53*, 149–154.

Yamaguchi, T., Yamaguchi, S., Hirai, T., and Kitano, T. (2007). Follicle-stimulating hormone signaling and *Foxl2* are involved in transcriptional regulation of aromatase gene during gonadal sex differentiation in Japanese flounder, Paralichthys olivaceus. Biochem. Biophys. Res. Commun. *359*, 935–940.

CHAPTER III

DE NOVO GONAD TRANSCRIPTOME ASSEMBLY AND ANALYSIS OF THE COMMON SNAPPING TURTLE, CHELYDRA SERPENTINA, REVEALS POTENTIAL SEX-DETERMINING GENES

Abstract

The snapping turtle (*Chelydra serpentina*) is a species whose sex is determined by incubation temperature during embryonic development. How temperature participates in signal transduction during this biological process is still largely unknown. With Next Generation Sequencing techniques, we were able to shed some light on this mystery by conducting transcriptome-wide analyses of gene expression during temperature-dependent sex determination (TSD). We performed high-throughput RNA sequencing (RNA-seq) on gonads collected from snapping turtle embryos incubated at both a male and a female producing temperature (26.5 °C and 31 °C respectively) during the sex-determining period. With a total of 360.4 million single-ended reads from RNA-seq, we assembled and annotated a reference transcriptome which was then used to characterize differential gene expression. We identified 725 differentially expressed genes (DEGs) in total. Among them, 293 DEGs were significantly affected by incubation temperature and included genes such as *Kdm6b, Aebp2, Crabp, Star, Cyp11a1, Hsd17b, Cyp17, Inhbb, Jarid2* and *Sox9*, which were demonstrated to be differentially expressed in TSD in previous studies. We find *Aebp2, Jarid2,* and *Kdm6b* of particular interest

because these genes could influence expression of many other genes via epigenetic modifications. Our findings provide a first description of the snapping turtle transcriptome and the effects of temperature on transcriptome-wide patterns of gene expression during the sex-determining period and expand our understanding of vertebrate sex determination.

Introduction

Vertebrates adopt diverse sex-determining mechanisms. In mammals, sex is determined by heritable genetic elements carried by sex chromosomes at fertilization (Wilhelm et al., 2007) whereas in certain reptiles, sex is determined by incubation temperature during embryogenesis (Ewert et al., 1999). These represent the two major types of sex determination – genotypic sex determination (GSD) and environmental sex determination (ESD). Temperature-dependent sex determination (TSD) is the most well studied form of ESD. Sex determination occurs during a specific developmental window, called the thermosensitive period (TSP), which varies among TSD species (Bull, 1987; Burke and Calichio, 2014; Pieau and Dorizzi, 1981; Siroski et al., 2007; Yntema, 1979). During the TSP, developing gonads respond to temperature differently from species to species. In American alligators, low (29-31°C) and high (35°C) incubation temperatures produce females, while intermediate temperatures (33°C) produce males, and 32°C and 34°C produce both sexes (Lance et al., 2000; Lang and Andrews, 1994). In the common snapping turtle, low temperatures (23-27°C) produce males while high temperatures (above 29.5°C) produce females and intermediate temperatures (27-29.5°C) produce both

sexes (Rhen and Lang, 1998). In addition, the TSD pattern in the snapping turtle varies geographically and among clutches within populations (Ewert et al., 1999).

With respect to GSD and TSD, it seems the two sex-determining mechanisms are unrelated. Even at the molecular level, the key sex-determining genes differ between GSD species and TSD species. The master male-determining gene in mammals is sex-determining region Y or *Sry*, which resides on the Y chromosome (Wilhelm et al., 2007). This gene evolved in the last common ancestor of therian mammals (Graves, 2016) and is absent in TSD species. However, other sex-determining genes appear to be conserved across vertebrates. For example, *Sox9, Amh* and *Dmrt1* are expressed at a higher level in the incipient testes than in developing ovaries of both GSD and TSD species, although the timing of expression differs somewhat (Kent et al., 1996; Münsterberg and Lovell-Badge, 1991; Raymond et al., 2000; Rhen et al., 2015; Western et al., 1999). In contrast, *Foxl2* is the earliest ovarian marker during gonadal differentiation in both mammals and non-mammalian species (Hudson et al., 2005; Loffler et al., 2003; Shoemaker et al., 2007). Expression of this gene is induced by exposure to a female-determining temperature during embryonic development of the snapping turtle (Rhen et al., 2007). Aromatase, a key enzyme that converts testosterone (T) to 17β-estradiol (E2), plays a crucial role in sexual development of both GSD and TSD species (Pieau et al., 2001). In addition, the development pattern and morphological differentiation of the gonad are evolutionarily conserved in amniotic vertebrates.

In all vertebrates, both testes and ovaries develop from a bipotential primordium that is morphologically indistinguishable between the sexes (Witschi, 1959). Therefore, studying sexual development in one species may shed light on the process in other

species. Due to their evolutionary relationship with mammals and birds, reptiles with TSD serve as a good model to understand sex determination and sexual differentiation. In this study, we use the common snapping turtle (*Chelydra serpentina*), a TSD species that is widespread in North America, as a model species to study the mechanism of TSD. The sex-determining period of this species was defined by shifting eggs between male-producing temperatures (MPT) and female-producing temperatures (FPT), which provides a foothold for more mechanistic studies of TSD (Rhen et al., 2015; Yntema, 1979).

Previous studies on the common snapping turtle have revealed candidate genes that may be involved in TSD: *Cirbp* (Schroeder et al., 2016), *Wt1* (Rhen et al., 2015), *PdgfB* (Rhen et al., 2009), *Dmrt1, Sox9*, aromatase, *Ar* and *Foxl2* (Rhen et al., 2007). It is rather common to study a core set of genes that are presumably conserved in the process of sex determination across vertebrates and validate differential expression either *in vitro* or *in vivo*. However, this process will not reveal novel sex-determining genes and can be time consuming and sometimes misleading (Rhen and Schroeder, 2010). For example, *Dax1, Fgf9,* and *Sf1* are involved in sex determination in mammals but are not differentially expressed between MPT and FPT in snapping turtle embryos during the TSP (Rhen et al., 2007). A new approach is needed to identify potential TSD genes more efficiently and in an un-biased manner.

Next Generation Sequencing (NGS) techniques enable researchers to conduct transcriptome-wide analyses of gene expression. However, sequencing of large vertebrate genomes is still quite complex and costly. According to NCBI Genome Database, 325 out of 13525 genomes published so far (2.4%, all levels included, i.e. complete,

chromosome, scaffold and contig) are of vertebrates and only 11 of them are reptiles. Compared to genome sequencing, *de novo* transcriptome sequencing is a cost-efficient method that sequences all the transcripts in a cell or tissue type. In this study, we used *de novo* transcriptome sequencing of the common snapping turtle to acquire information about gene function and expression at a transcriptome-wide scale.

NGS platforms include Roche/454 FLX, the Illumina/Solexa Genome Analyzer and the Applied Biosystems SOLiDTM System (Mardis, 2008). A single 454/Roche-system run generates an average of 800,000 reads at lengths of up to 600 bp (Renaut et al., 2010), while systems like Illumina/Solexa and Applied Biosystems SOLiDTM produce millions of reads per lane with the sequences up to 125 bp long (Crawford et al., 2010). Each sequencing platform has its pros and cons. The 454/Roche system produces longer reads while Illumina/Solexa generates more reads with higher accuracy. To take full advantage of NGS technology, we used both 454/Roche and Illumina/Solexa platforms for *de novo* transcriptome sequencing, assembly, and expression analysis in embryonic gonads from the common snapping turtle. Our goals were to produce a reference transcriptome for this species and to identify novel candidate genes that are potentially involved in TSD. Availability of a reference transcriptome will also facilitate future studies of population genetics in this species and evolution of sex-determining mechanisms.

Materials and Methods

Egg Collection and Incubation

Snapping turtle eggs (32 clutches) were collected in Minnesota, USA in June of 2009 and 2010. Eggs were transported to the Biology Department at the University of North Dakota in Grand Forks, North Dakota, USA. Eggs were cleaned in tepid water and infertile eggs were removed based on the result of candling. Eggs from 7 clutches were assigned for 454/Roche sequencing. Eggs from 25 clutches were assigned for Illumina RNA-seq sequencing. Eggs were covered by moist vermiculite (mix of 1 part vermiculite to 1 part water by mass) and incubated at 26.5°C, a temperature that produces 100% males, until embryos reached stage 17.5 (Yntema, 1968). Embryos are very sensitive to brief exposure to female-producing temperatures at this developmental stage (Rhen et al., 2015; Yntema, 1979). Half of the eggs for both sequencing methods were then shifted to 31°C for 6 days, a temperature treatment that produces 100% females. The other half of the eggs were kept at 26.5°C throughout this 6-day period.

RNA Preparation and Quality Controls

Approximately equal numbers of embryos were collected at 26.5°C and 31°C on day 1, 2, 3, 4, and 5 of the temperature shift. Adrenal-kidney-gonad complexes were dissected and immediately placed in RNAlater solution (Sigma) and stored at -20°C. Gonads were micro-dissected from the underlying kidney prior to RNA extraction from pure gonadal tissue. To get better representation of all transcripts in the snapping turtle, we also collected hypothalamus-pituitary and intestinal tissues to represent tissues derived from all three germ layers. Total RNA was isolated from each tissue and treated with DNase as described previously (Rhen et al., 2007). The integrity of the total RNA

was assessed by agarose gel electrophoresis. A NanoDrop ND-1000 spectrophotometer was used to quantify the isolated RNA. The 260/280 absorbance ratio of total RNA was between 1.8 and 2.0.

Next Generation Sequencing

RNA from gonads was sent to the University of Illinois Urbana-Champaign for 454/Roche sequencing and Illumina sequencing. We combined equal amount of RNA from days 1-5 into two pools (26.5°C and 31°C) for 454/Roche sequencing. Two sequencing libraries were synthesized and normalized to produce as many unique cDNA sequences as possible regardless of abundance (i.e, low, medium, and high abundance transcripts). Each library generated 1.4 million reads with an average read length of 350 bp (2.8 million reads in total). For Illumina sequencing, 20 libraries were synthesized without normalization to enable expression analysis (2 temperatures x 5 days x 2 biological replicates). The Illumina HiSeq 2000 platform produced 156.4 million reads (100 bp, single-end reads).

RNA samples from hypothalamus-pituitary and intestinal tissues were sent to University of Utah for Illumina sequencing. Eight libraries were generated for hypothalamus-pituitary (2 temperatures x 2 stages x 2 biological replicates). Two libraries were made for intestine (1 male hatchling and 1 female hatchling). The Illumina platform produced 172.2 million reads for the hypothalamus-pituitary and 31.8 million reads for intestine (50 bp, single-end reads).

De Novo Sequence Assembly and Sequence Clustering

Sequence assembly was performed by CLC Genomics Workbench (CLC bio, Cambridge, MA) on a Mac Pro with 12 cores and 96GB RAM. The *de novo* assembly

and annotation pipeline is shown in Figure 13. We used a combination of *de novo* assembly (word size = 64, auto bubble size) and reference transcriptomes to assemble the snapping turtle transcriptome. Reads were mapped (length fraction = 1.0, similarity 0.9) to reference transcriptomes from chicken (Gallus_gallus.WASHUC2.65.cdna.all), green anole (Anolis_carolinensis.AnoCar2.0.65.cdna.all), duck-billed platypus (Ornithorhynchus_anatinus.OANA5.65.cdna.all), and zebra finch (Taeniopygia_guttata.taeGut3.2.4.65.cdna.all). Snapping turtle contigs from the initial assemblies were then used as references for another round of mapping and *de novo* assembly. This process was repeated 6 more times with snapping turtle contigs used as references. Contigs less than 200bp were filtered from the transcriptome. Similar sequences were clustered with CD-HIT-EST (version 4.6.5) at 95% similarity threshold (Fu et al., 2012). The resulting sequences were then subjected to TransDecoder (https://transdecoder.github.io) to predict coding regions.

To produce a reference gene set, we generated a gene list in which each sequence represents a unique protein coding gene. To accomplish this, we first aligned the predicted coding sequences from the snapping turtle transcriptome to the human, chicken, Chinese softshell turtle and painted turtle protein databases individually using NCBI-BLAST-2.4.0+ suite. Then we generated four tentative unique gene sets based on the BLAST results against each protein database, using a Perl script (Zeng et al., 2011). The final unique gene set was generated by extracting the common sequences from the four tentative unique gene sets. To verify and complement this method, we compared these candidate genes with over 3,500 manually annotated contigs. This verified unique gene

set was used as a reference gene set for gene ontology and for gene set enrichment analyses.

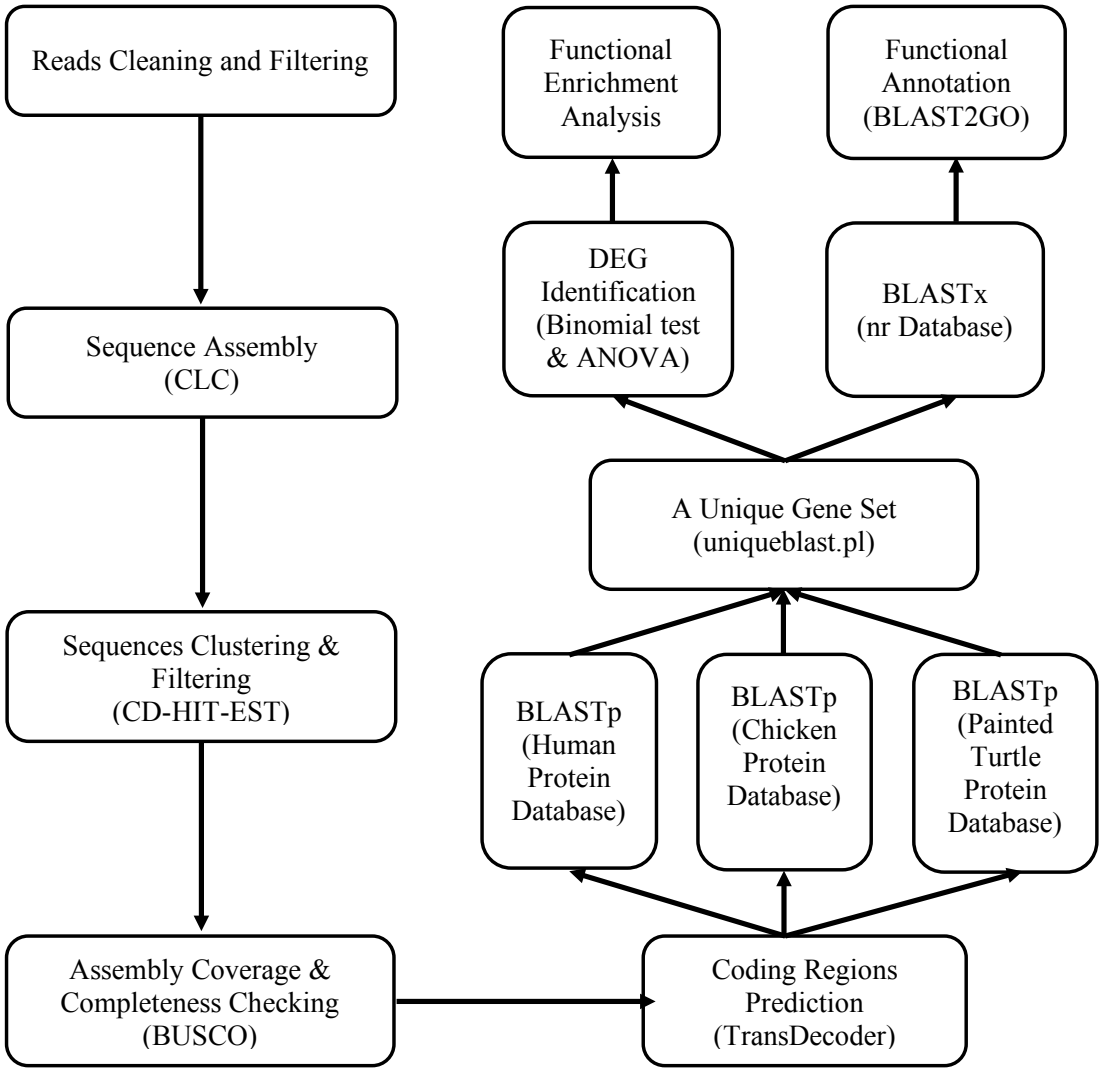


Figure 13. De novo assembly and annotation workflow. Reads from 454 and Illumina sequencing were cleaned and assembled with CLC genomics workbench. A unique protein coding transcriptome was then generated by comparing the assemblies with sequences in three different protein databases (human, chicken and painted turtle), which was used as a reference for the following DEG identification and functional annotation.

Assembly Validation and Estimation of Trancriptome Completeness

We manually blasted and annotated approximately 3,500 contigs to assess the accuracy of the assembly. We also compared the assembled contigs to cDNA sequences

that were independently determined via Sanger sequencing in previous studies from our lab. BLASTN in the NCBI-BLAST-2.4.0+ suite was performed to determine the homology between Sanger sequences and sequences that were assembled *de novo* from Illumina and 454 reads. We used BUSCO v1.2 to assess the completeness of our transcriptome. BUSCO includes comprehensive lineage-specific sets of Benchmarking Universal Single-Copy Orthologs for arthropods, vertebrates, metazoans, fungi, eukaryotes, and bacteria (Simão et al., 2015).

Similarity Search and Functional Annotation

Sequences from the unique gene set were then aligned to NCBI non-redundant (Nr) protein database using BLASTx in the NCBI-BLAST-2.4.0+ suite locally on a cluster with 256 processors. Settings of the BLASTx search were the same as the default BLASTx settings in BLAST2GO. The resulting XML files from the BLAST search were imported into BLAST2GO for further analyses. Gene ontology (GO) terms were retrieved and assigned to sequences using the default settings of BLAST2GO. To make gene ontology annotation graphs, GO-slim was used to simplify the GO annotation. For the combined GO graphs, GO terms containing less than 10 sequences were removed (Miller et al., 2012).

GC Content Analysis and Retroelements Identification

We used RepeatMasker 4.0.6 (http://www.repeatmasker.org) to measure GC content of the transcriptome and identify retroelements (Tarailo-Graovac and Chen, 2009)**.** We used MISA (MIcroSAtellite; http://pgrc.ipk-gatersleben.de/misa/) to identify microsatellite sequences. We set the following MISA search criteria: mono-nucleotide repeats greater than 10, di-nucleotides repeats greater than 6, tri-, tetra-, penta- and hexa-

nucleotide repeats greater than 5. The maximal number of bases interrupting 2 microsatellites in a compound microsatellite was set as 100.

Differential Expression Analysis and Gene Enrichment Analysis

The assembled transcriptome was used as a reference to identify differentially expressed genes (DEGs) between MPT and FPT. To generate a comprehensive DEG list, we compared gene expression between all ten experimental groups (2 temperatures x 5 days = 10 groups for 45 total pairwise comparisons) using the beta-binomial test at a false discovery rate (FDR) of 0.05 (Baggerly et al., 2003). To confirm differential expression of these genes, we carried out two-way ANOVA on RPKM values for each gene with incubation temperature, sampling day, and the temperature by day interaction as independent variables in SAS JMP (version 12). Our final DEG list is conservative because it only contains genes that were statistically significant in both analyses (Baggerley's test and ANOVA). Gene expression values (RPKM) were used for hierarchical clustering which was visualized in a heat map. We also performed gene enrichment analysis for DEGs using our reference gene set in BLAST2GO.

Results and Discussion

Sequence Clustering, Transcriptome Completeness Assessment and Assembly Validation

We assembled 2.8 million 454 reads and 360.4 million Illumina reads from three distinct tissues into 421,738 contigs. Sequences less than 200bp were removed, resulting in 307,745 contigs. We used CD-HIT-EST to cluster different transcripts from the same locus (i.e., splice variants). We then extracted the longest contig from each cluster. With a 95% similarity cut off, the number of unique assembled transcripts was further reduced

to 270,094 contigs, which were used as the start point for all subequent analyses. We used BUSCO (Benchmarking Universal Single-Copy Orthologs) to estimate the completeness of this set of transcripts. We identified 367 complete BUSCOs and 25 fragmented BUSCOs in the snapping turtle transcriptome, yielding 91% of the 429 total BUSCOs expected to be found in vertebrates. To evaluate the quality of the *de novo* assembled sequences, 270,094 contigs were aligned using BLASTN to 30 sequences independently derived via Sanger sequencing in our lab. The de novo assembled sequences showed high similarity (average = 99.1% identity) with the 30 sequences derived from Sanger sequencing, indicating the high quality of our assembly.

GC Content and Retroelements

It is generally believed that GC content is enriched in coding regions compared to surrounding genomic regions (Eyre-Walker and Hurst, 2001). The GC content of snapping turtle transcriptome is 45.4%, which is in range reported for 7 reptiles, including *Pogona vitticeps* (41.8%)*, Anolis carolinensis* (40.3%)*, Crocodylus porosus* (44.2%)*, Pelodiscus sinensis* (44.1%)*, Chrysemys picta* (43.7%)*, Python bivittatus* (39.6%), and *Ophiophagus hannah* (38.6%) (Georges et al., 2015). The GC content of the predicted coding regions is 45.6% which is similar to other vertebrates such as *Danio rerio* (47.9%), *Xenopus laevis* (48.1%), *Mus musculus* (53.2%) and *Gallus gallus* (55.1%) (Zhou et al., 2004).

Retroelements are components of eukaryotic genomes that are able to copy and translocate themselves to other locations within a genome and are abundant in some eukaryotic genomes (Deininger and Batzer, 2002). About 42% of the human genome is made up of retroelements (Lander et al., 2001). The total interspersed repeats in our

assembled transriptome account for 5.73% of the total length. Among them, 1.3% are SINEs, 3.37% are LINEs, 0.05% are LTR elements and 0.44% are small RNAs (Table 3). We identified a total of 45,088 microsatellites in 33,713 transcripts (12.5% of 270,094 transcripts) with frequency of one microsatellite per 4.80 kb of sequence (Table 4). Mono-nucleotide repeats represented the largest fraction (62.8%) of microsatellites identified followed by di-nucleotide (27%) and tri-nucleotide (8.4%) repeats. Only a small number of tetra- (691), penta- (63) and hexa-nucleotide (17) microsatellites were identified in the assembled transcripts (Table 5).

Table 3: Retroelements identified in the assembled Chelydra serpentina contigs

Elements Type	Number of Elements	Length Occupied	Percentage of Sequence
SINES	21064	2825249 bp	1.30%
SINES: ALUs	13	794 bp	0.00%
SINES: MIRs	12033	1401882 bp	0.65%
LINEs	33906	7296878 bp	3.37%
LINEs: LINE1	104	18934 bp	0.01%
LINEs: LINE2	3313	527269 bp	0.24%
LINEs: L3/CR1	30156	6673548 bp	3.08%
LTR Elements	449	106661 bp	0.05%
LTR Elements: ERVL	27	2318 bp	0.00%
LTR Elements: ERVL-MaLRs	7	532 bp	0.00%
LTR Elements: ERVL_classI	218	57350 bp	0.03%
LTR Elements: ERVL_classII	9	441 bp	0.00%
DNA Elements	14420	2014996 bp	0.93%
DNA Elements: hAT-Charlie	1705	152083 bp	0.07%
DNA Elements: TcMar-Tigger	1081	141987 bp	0.07%
Unclassified	1102	161186 bp	0.07%
Total interspersed repeats:		12404970 bp	5.73%
Small RNA	6571	960286 bp	0.44%

Table 4: Microsatellite identified in the assembled Chelydra serpentina contigs

Total number of sequences examined	270094
Total size of examined sequences (bp)	216669289
Total number of identified SSRs	45088
Number of SSR containing sequences	33713
Number of sequences containing more than 1 SSR	7620
Number of SSRs present in compound formation	4311

Table 5: Distribution of different repeat type classes in the assembled Chelydra serpentina contigs

Repeat type	Number of SSRs
Mono-nucleotide repeats	28316
di-nucleotide repeats	12207
tri-nucleotide repeats	3794
tetra- nucleotide repeats	691
Penta- nucleotide repeats	63
Hexa-nucleotide repeats	17

Homology Search Against the Non-Redundant Protein Database

TransDecoder detected 51,289 transcripts that contain potential coding regions longer than 100 amino acids in the reduced redundancy transcriptome (270,094 transcripts). For ORFs predicted by TransDecoder, only the longest single ORF for each transcript was kept for BLAST analysis against NCBI protein databases. Combining BLAST results and manually annotated contigs (2,204 protein coding transcripts), we generated a set of 19,602 unique protein-coding sequences that we used as a reference gene set. However, these 19,602 transcripts may not represent unique loci. As for all *de novo* transcriptome assemblies, sequences transcribed from the same gene as a single RNA may not be assembled into a single contig due to low coverage, e.g. non-overlapping sequences. Such sequences can have different top BLAST hits, leading to misidentification of different parts of the same transcript as different genes. As a result, this number may overestimate the true number of unique protein coding transcripts in the

assembled transcriptome. Second, many predicted coding regions did not yield any BLAST hits. Unmatched sequences may contain novel genes, non-coding sequences, or incorrectly assembled transcripts. Therefore, it is hard to determine if the 19,602 genes in our reference gene set accurately represent the actual number of unique protein coding genes in the snapping turtle genome. Other studies have estimated the number of unique protein coding genes from turtle genome sequences: there are 21,796 predicted protein coding genes in *Chrysemys picta*, 19,327 predicted protein coding genes in *Pelodiscus sinensis*, and 19, 633 predicted protein coding genes in *Chelonia mydas*. Based on these numbers, we conclude that the snapping turtle reference gene set is sufficiently complete to be used in further analyses.

With an E-value cutoff of 1e-5, we used BLASTx to compare our reference gene sequences against the nr database and kept the 20 highest scoring alignments. Top hits were dominated by three turtle species, *Chrysemys picta bellii* (49%), *Chelonia mydas* (25%), *Pelodiscus sinensis* (4%) (Figure 14). Remaining hits were mainly from other reptiles, including alligators, lizards and snakes.

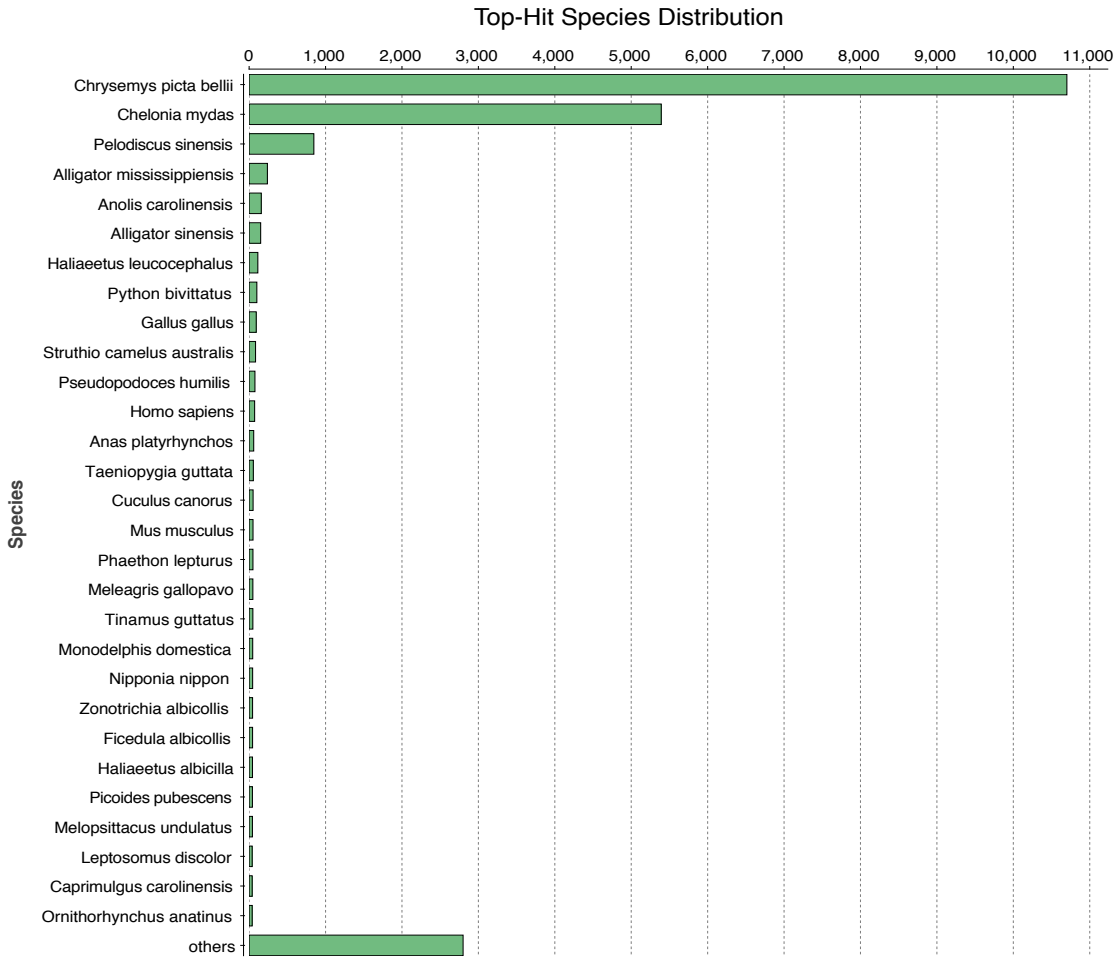


Figure 14. Top hit species distribution of BLASTX of Chelydra serpentina transcripts against Nr database. Proportion of Chelydra serpentina *transcripts with similarity to sequences from Nr protein database.*

Functional Annotation Based on Gene Ontology

Gene Ontology (GO) is a standardized classification system for describing particular attributes of genes or gene products (Ashburner et al., 2000). The GO database provides three general ontologies: “molecular function” describes gene product activity at the molecular level, “cellular component” describes where the gene product is located at the sub-cellular level, and “biological process” describes a series of events with a defined

beginning and end that are influenced by the gene product (Ashburner et al., 2000). Results of BLASTx for 19,602 unique protein-coding transcripts were fed into BLAST2GO to obtain gene ontology terms. A total of 16,966 contigs were assigned one or more GO terms. GO annotations for each contig were merged to eliminate redundancy. To summarize the results of GO annotation of our assembled transcriptome, we grouped the GO classes into GO-slim terms (a subset of GO terms). Among all GO terms, 47% belong to Biological Process, followed by Cellular Component (39%) and Molecular Function (14%). The top 20 sub-categories from GO level 2 classification are shown in Figure 15. The total functional annotation is provided in Figure 16. Among three GO categories (biological process, molecular function and cellular component), the largest number of assigned terms are biological processes, followed by cellular component and molecular function. The most commonly assigned GO terms in the biological process category included biosynthetic process and signal transduction (Figure 16A). Protein complex and cytoplasm were the most commonly assigned GO terms in the cellular component category (Figure 16B). In the molecular function category, binding was the top assigned GO terms, which included ion binding, DNA binding, enzyme binding and RNA binding (Figure 16C).

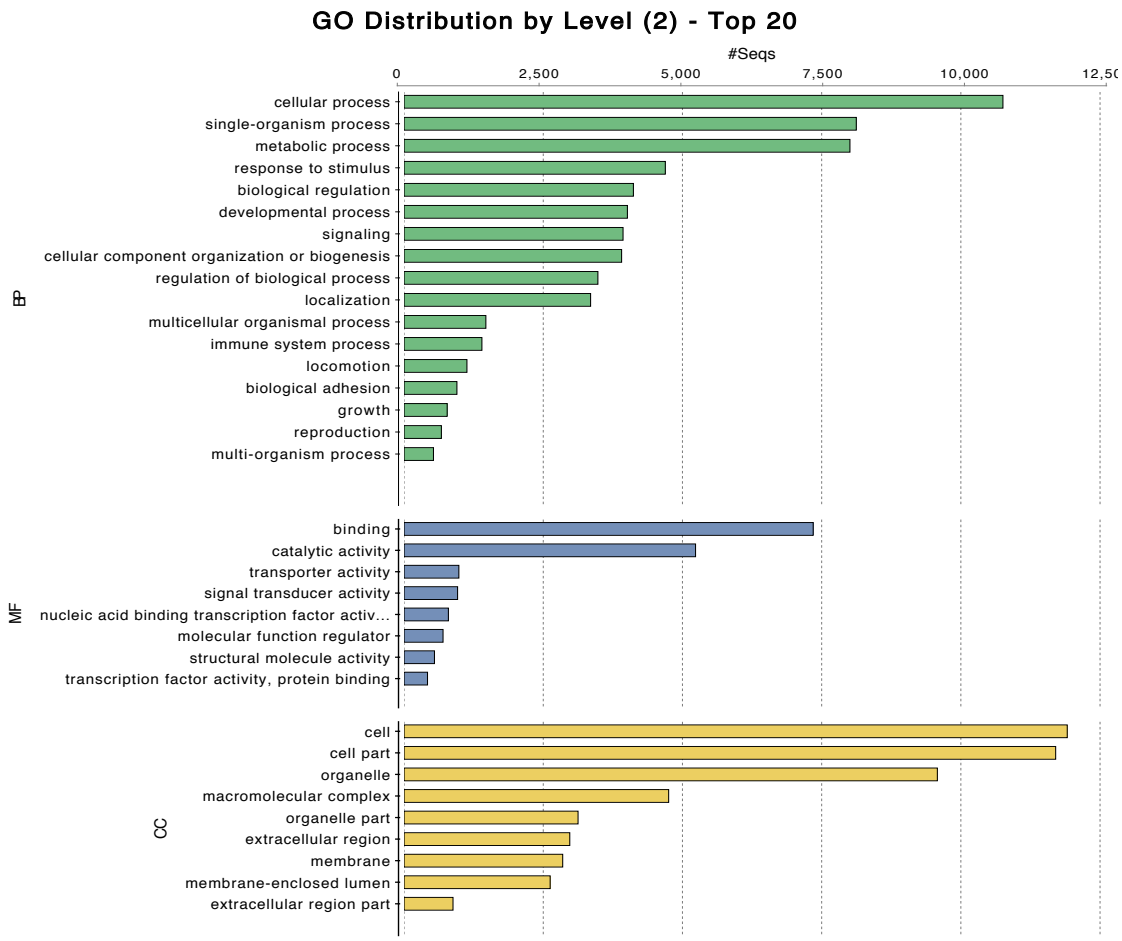
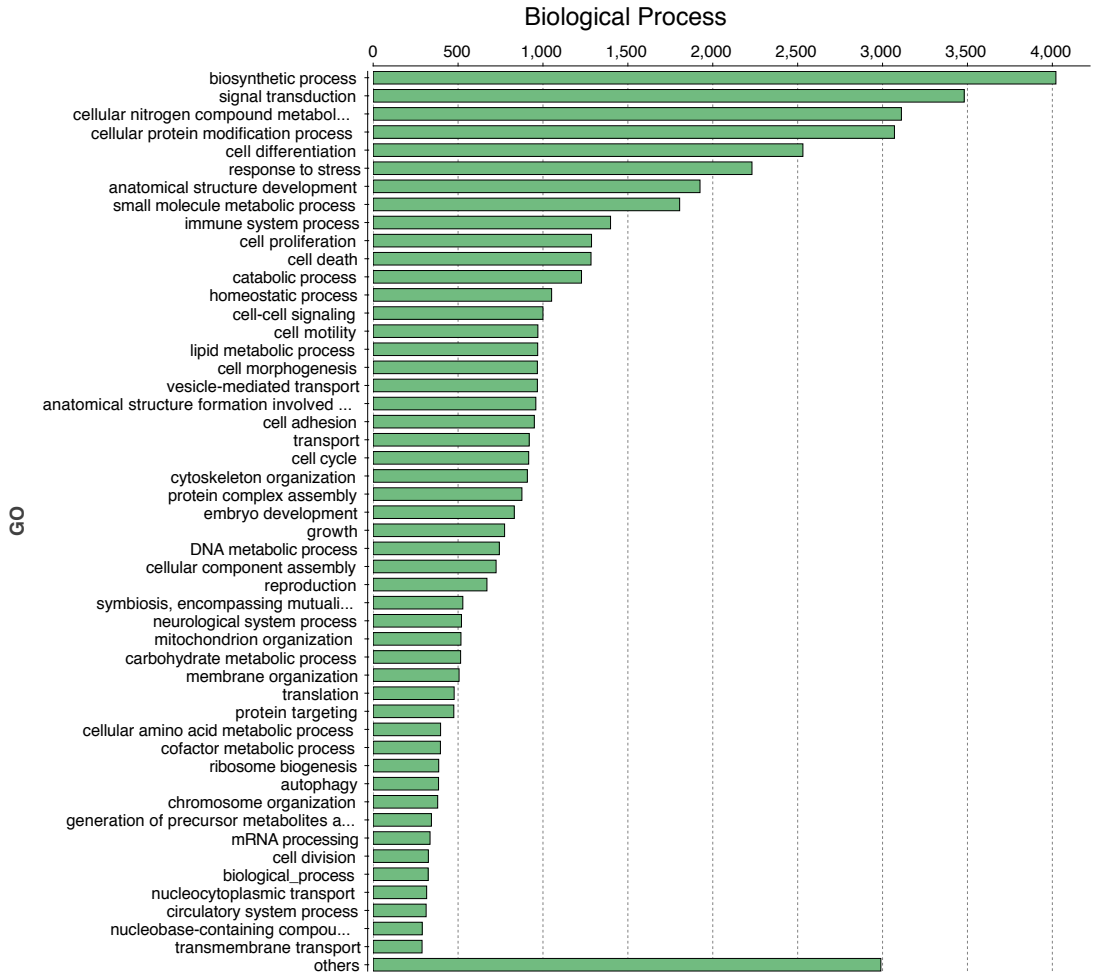
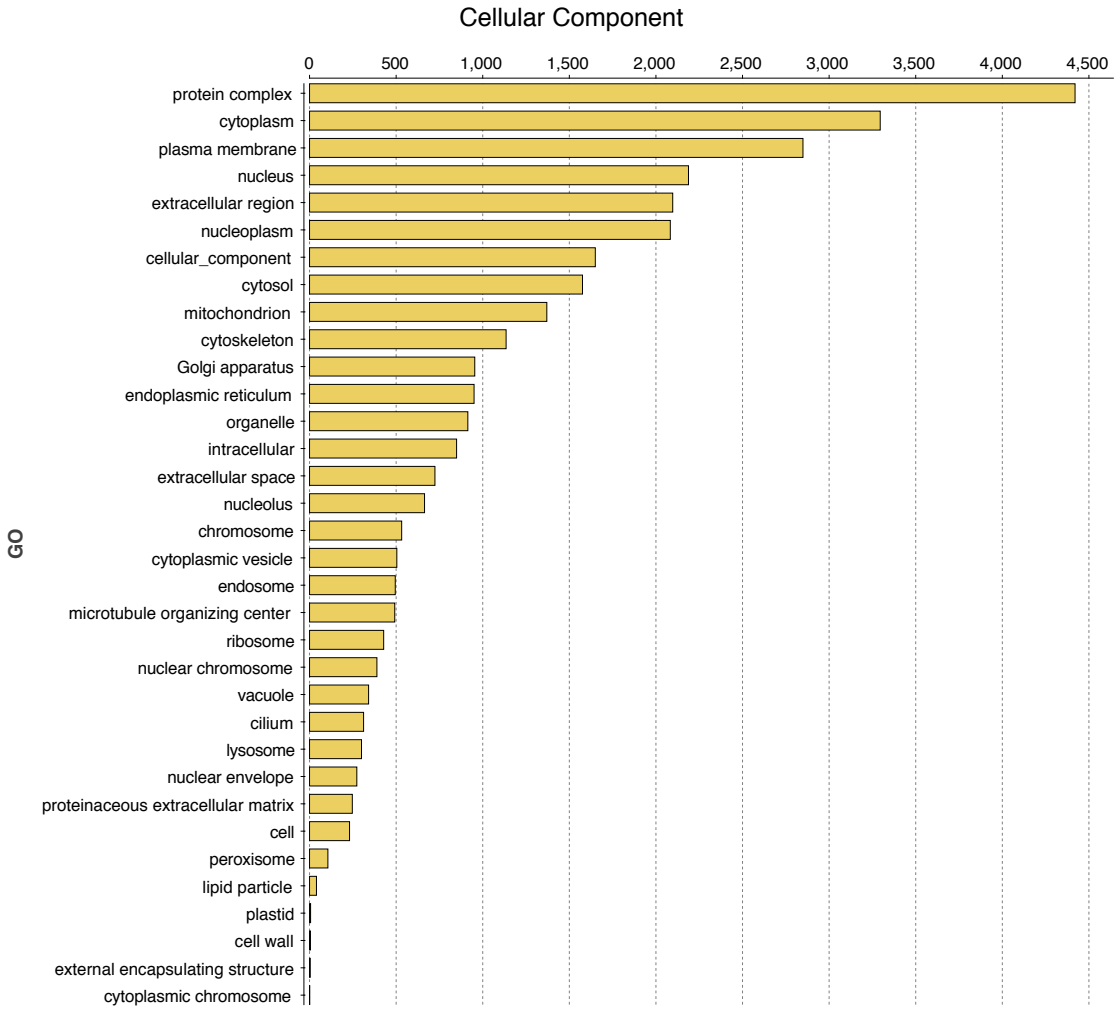


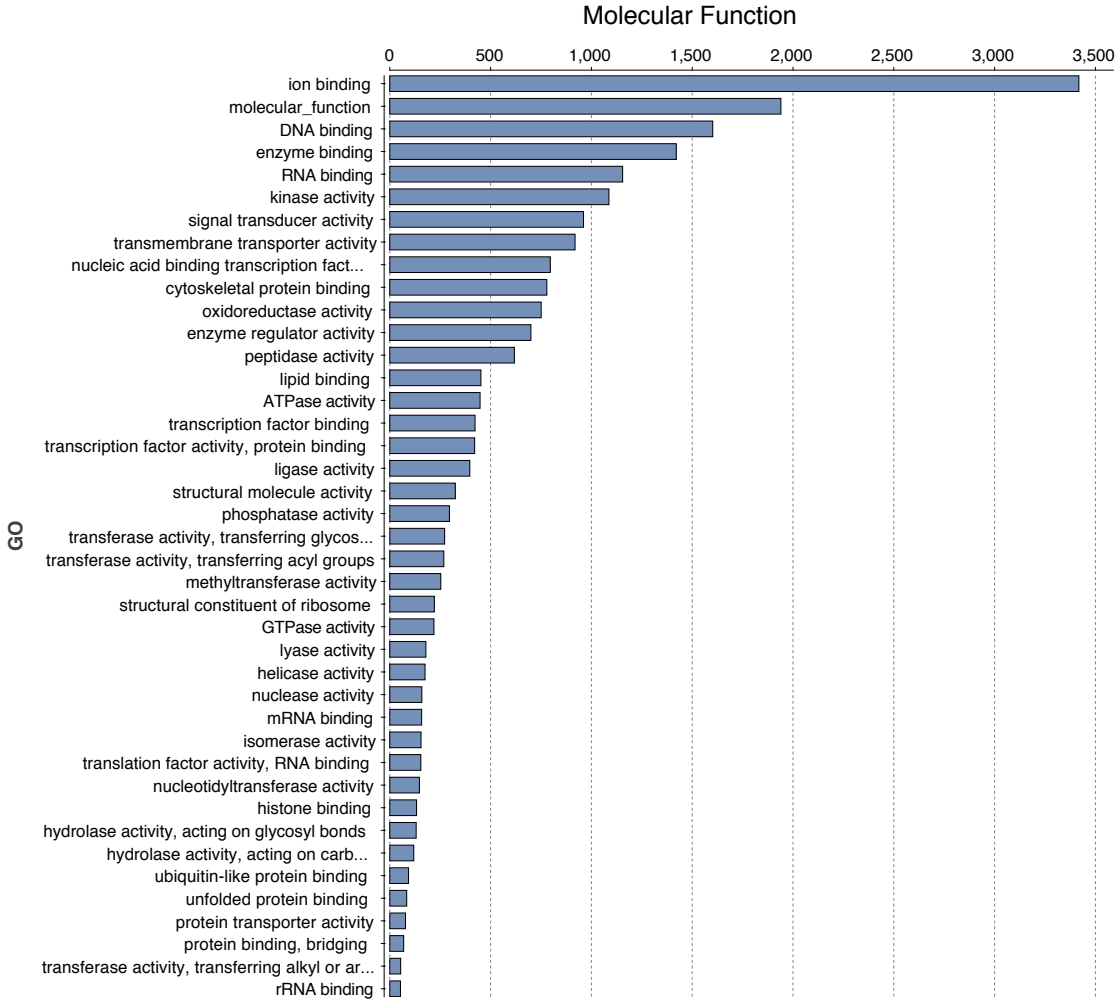
Figure 15. Gene ontology (GOslim) assignments for Chelydra serpentina *transcripts. Level 2 annotations are shown in three main categories: Biological Process, Molecular Function and Cellular Component.*


Biological Process
0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
GO
biosynthetic process
signal transduction
cellular nitrogen compound metabol...
cellular protein modification process
cell differentiation
response to stress
anatomical structure development
small molecule metabolic process
immune system process
cell proliferation
cell death
catabolic process
homeostatic process
cell-cell signaling
cell motility
lipid metabolic process
cell morphogenesis
vesicle-mediated transport
anatomical structure formation involved ...
cell adhesion
transport
cell cycle
cytoskeleton organization
protein complex assembly
embryo development
growth
DNA metabolic process
cellular component assembly
reproduction
symbiosis, encompassing mutuali...
neurological system process
mitochondrion organization
carbohydrate metabolic process
membrane organization
translation
protein targeting
cellular amino acid metabolic process
cofactor metabolic process
ribosome biogenesis
autophagy
chromosome organization
generation of precursor metabolites a...
mRNA processing
cell division
biological_process
nucleocytoplasmic transport
circulatory system process
nucleobase-containing compou...
transmembrane transport
others

A


Cellular Component
0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
GO
protein complex
cytoplasm
plasma membrane
nucleus
extracellular region
nucleoplasm
cellular_component
cytosol
mitochondrion
cytoskeleton
Golgi apparatus
endoplasmic reticulum
organelle
intracellular
extracellular space
nucleolus
chromosome
cytoplasmic vesicle
endosome
microtubule organizing center
ribosome
nuclear chromosome
vacuole
cilium
lysosome
nuclear envelope
proteinaceous extracellular matrix
cell
peroxisome
lipid particle
plastid
cell wall
external encapsulating structure
cytoplasmic chromosome

B



C

Figure 16. The total functional annotation for Chelydra serpentina *transcripts. Annotations are shown in three main categories: Biological process (A), Cellular component (B) and Molecular function (C).*

Differential Expression and Cluster Analysis

At a FDR ≤ 0.05, we identified 913 transcripts that were differentially expressed between at least 2 groups (45 pair-wise comparisons among 10 experimental groups). Subsequent two-way ANOVA on these transcripts confirmed that expression of 725 genes was significantly influenced by incubation temperature, sampling day, and/or the

interaction between temperature and day. To detect the functional characteristics of these 725 DEGs, we performed a GO term enrichment analysis. The most significantly enriched GO terms were related to translation and DNA metabolism, such as ribosome biogenesis, translation, nucleobase-containing compound catabolic process, protein targeting and protein maturation (Figure 17). Reproduction and embryo development were also significantly enriched GO categories (Figure 17). The most enriched GO term involved the ribosome within each of the three GO categories. This is the first report for candidate TSD genes related to translation. Interestingly, we found two GO categories – transferase activity and ligase activity were underrepresented in our DEG set.

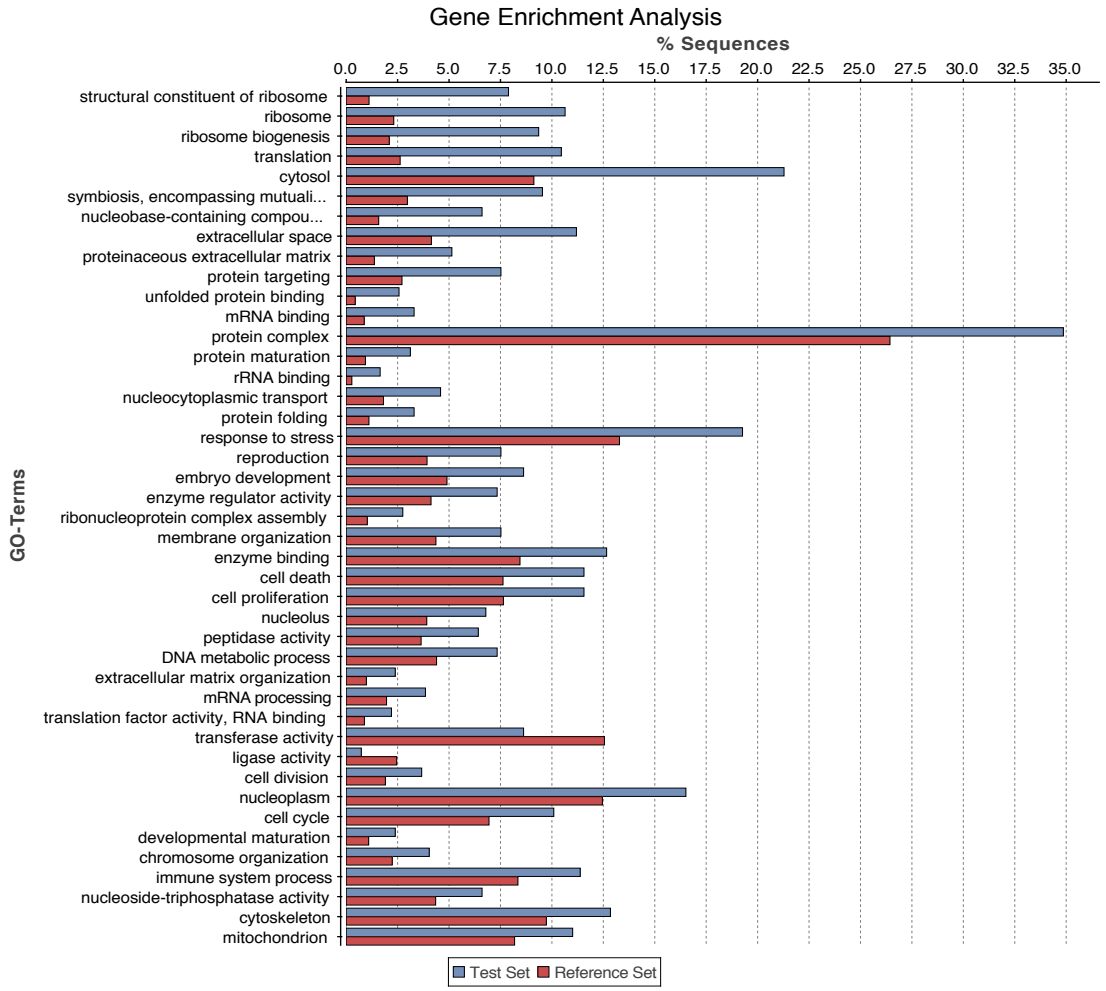


Figure 17. Gene enrichment analysis (Fisher's Exact Test) for the 725 differentially expressed genes. The unique protein coding genes from the annotated Chelydra serpentina transcriptome were used as the reference set (19,602 sequences) and the differentially expressed genes was the test set.

We performed two-way hierarchical clustering to identify groups of genes that display similar patterns of expression within temperatures and across time. Gene expression patterns were visualized in a heat map (Figure 18). Temperature and day have significant impacts on the transcriptome of the embryonic gonad during the TSP, with three major clusters of transcripts (Figure 18). The upper branch (in red) in the dendrogram contains samples that were incubated at 26.5^{o}C. Samples that were shifted to 31^{o}C for one day or two days form a cluster in the middle branch of the dendrogram (in green). The bottom branch (in blue) contains samples that were shifted to 31^{o}C for three, four or five days. These results indicate that temperature had a very rapid effect on expression of some genes, with an increasing number of changes at later stages. Early response genes may play a key role in regulating the entire TSD gene network. Gene expression was randomly clustered by geographic origin at the male producing temperature. In contrast, transcripts at the female temperature were much more organized. Southern and northern population were paired at most time points, but not on days 3 and 4 of the temperature shift. This suggests that snapping turtles in southern and northern Minnesota differ slightly in their responsiveness to incubation temperature, as previously described ((Rhen et al., 2015; Schroeder et al., 2016).

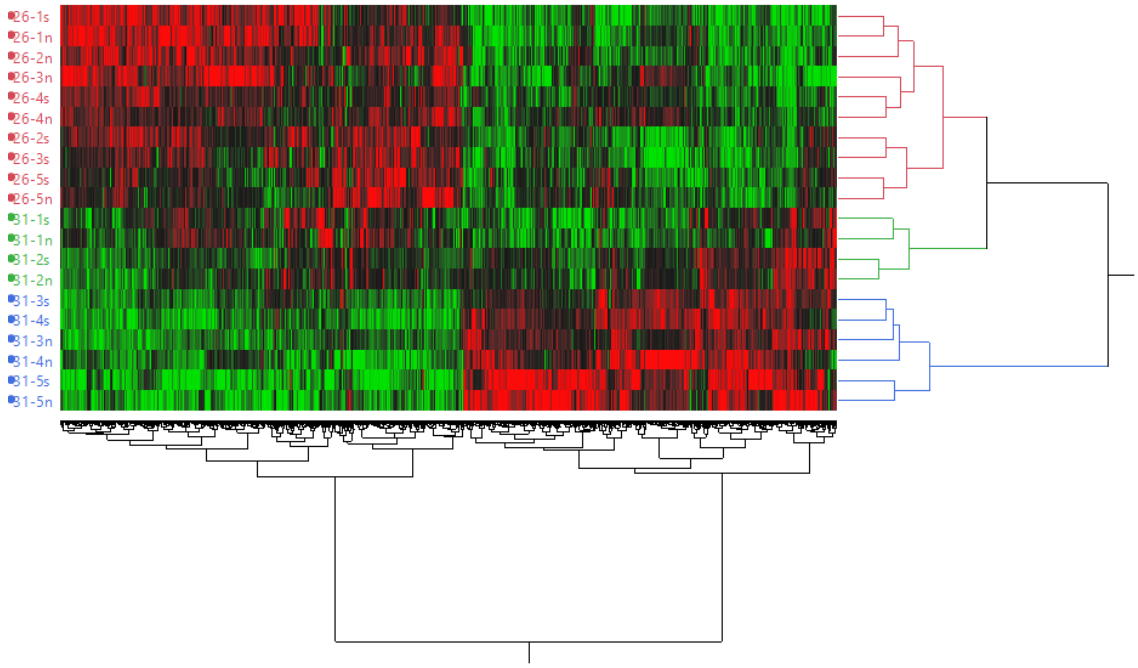


Figure 18. Expression levels of 725 differentially expressed genes in turtle gonads during the 5-day temperature shift. The upper branch (in red) in the dendrogram contains samples that were incubated at 26.5°C. Samples that were shifted to 31°C for one day or two days form a cluster in the middle branch of the dendrogram (in green). The bottom branch (in blue) contains samples that were shifted to 31°C for three, four or five days.

Identify Temperature-Responsive Genes

Two-way ANOVA revealed that 293 DEGs were significantly affected by incubation temperature alone. GO term enrichment analysis was performed on these putative temperature-responsive genes. Similar to the above enrichment analysis, ribosome and translation were the most significantly enriched GO categories (Figure 19).

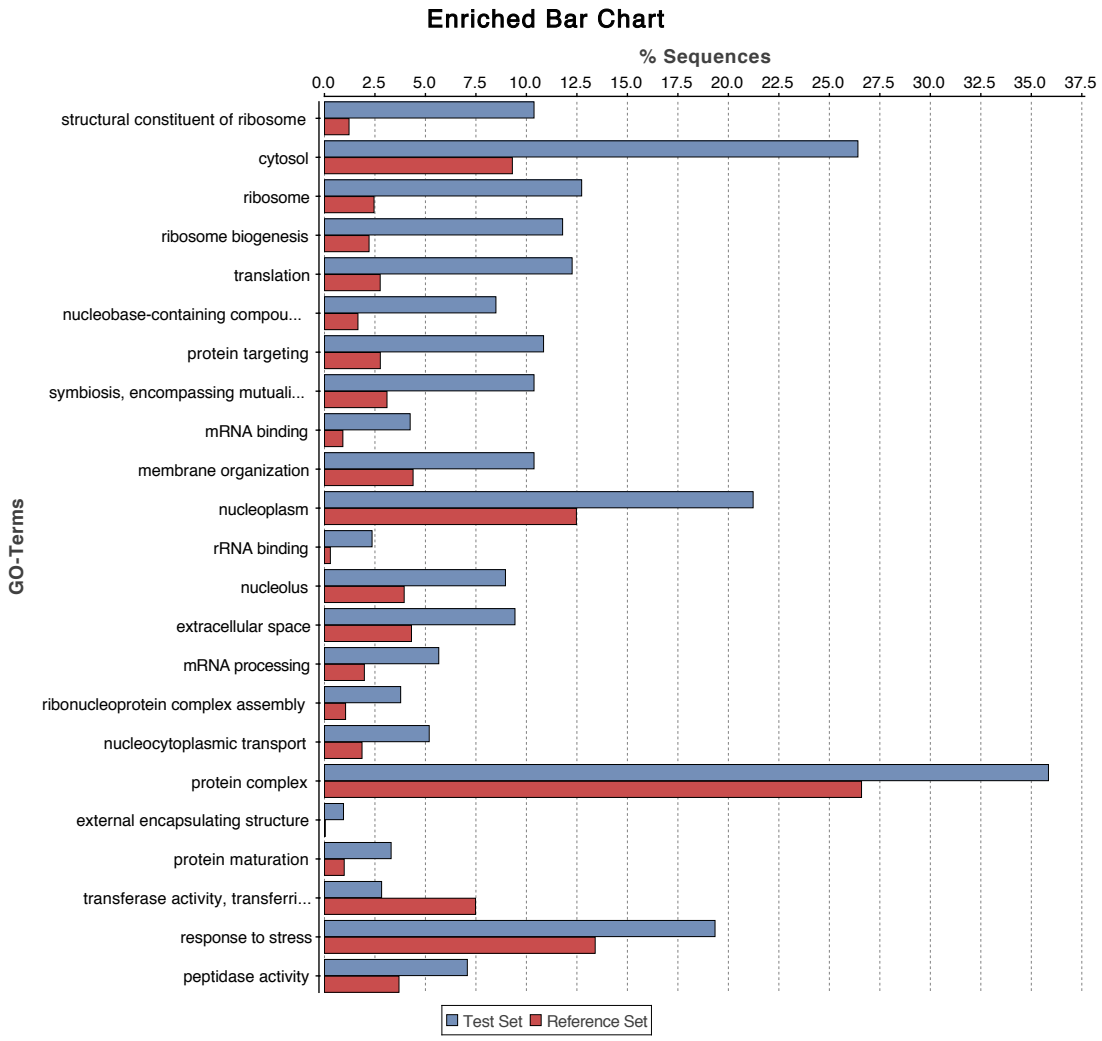


Figure 19. Gene enrichment analysis (Fisher's Exact Test) for the 293 temperature responsive genes. The annotated Chelydra serpentina transcriptome is used as the reference set and the differentially expressed genes is the test set.

To understand how temperature impacts gene expression during TSP, we further examined the expression profile of these putative temperature-responsive genes. The up-regulated genes at FPT after the temperature shift include *Aebp2, Jarid2, Kdm6b, Lmx1b, Axin2, Cirbp* and so on, some of which were also identified as important temperature-responsive candidate genes in TSD in previous studies (Czerwinski et al., 2016; Schroeder et al., 2016; Yatsu et al., 2016). We find *Aebp2, Jarid2,* and *Kdm6b* of

particular interest because these genes could influence expression of many other genes via epigenetic modifications. *Kdm6b* encodes a histone demethylase that removes methyl groups from lysine 27 on histone H3. Numerous studies have shown histone demethylases are involved in cell fate choices and cell differentiation (Lan et al., 2007; Sen et al., 2008; Shi, 2007; Ye et al., 2012). *Kdm6b* was identified as a temperature-responsive gene in American alligators, another TSD species (Yatsu et al., 2016). *Aebp2* and *Jarid2* are proteins that recruit Polycomb repressive complex 2 (PRC2), which is believed to repress transcription by methylating lysine 27 on histone H3 (Pasini et al., 2010). Further studies of epigenetic mechanisms in the snapping turtle have been started in our lab, which will help elucidate the role of epigenetics in TSD.

ANOVA indicated there were 47 genes that were developmentally regulated, but not affected by temperature. These genes may play a general role in gonad development in both sexes. Another 13 genes were only affected by the incubation temperature by day interaction. These genes may be downstream of genes that are directly temperature-responsive. There were 193 DEGs responding to both incubation temperature and day and 123 DEGs responding to all 3 factors (incubation temperature, day and the interaction between incubation temperature and day). The functional enrichment analyses for the 193 DEGs and 123 DEGs are shown in Figure 20 and Figure 21. The enriched GO terms for the genes responsive to both incubation temperature and day were similar to those for the genes only responsive to incubation temperature (compare Figure 19 and Figure 20). However, significantly enriched functions for the genes responsive to all 3 factors were largely different from those only responsive to one or two factors. These functions included cell death, protein binding and extracellular matrix (Figure 21).

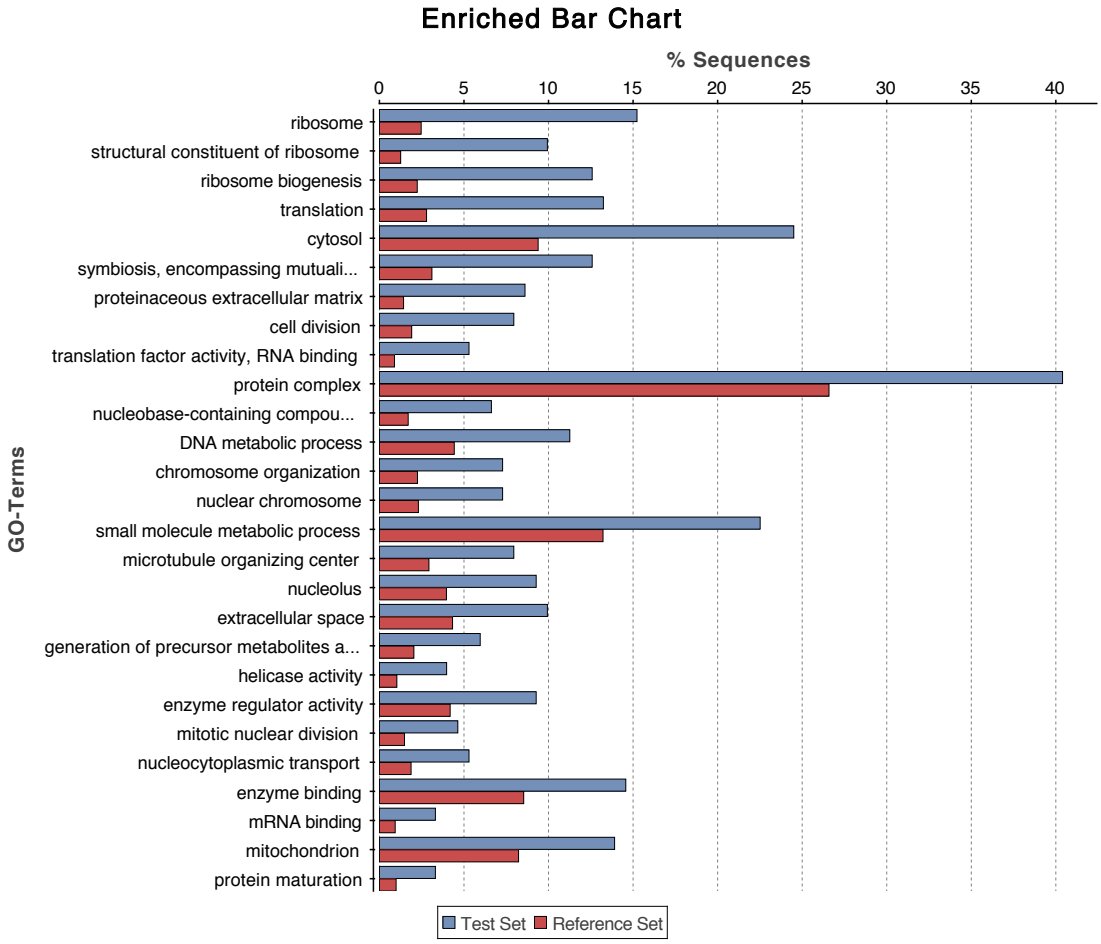


Figure 20. Gene enrichment analysis (Fisher's Exact Test) for the 193 DEGs responding to both incubation temperature and day. The annotated Chelydra serpentina transcriptome is used as the reference set and the differentially expressed genes is the test set.

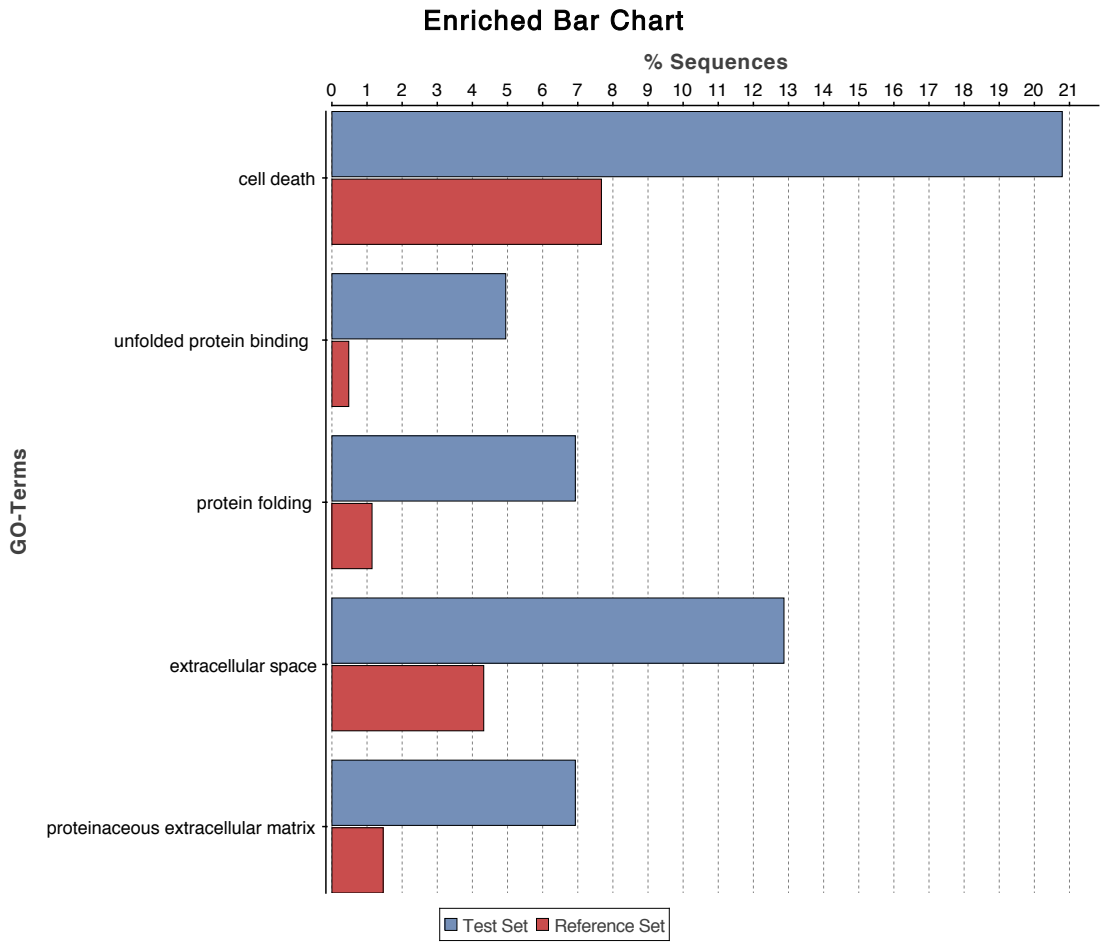
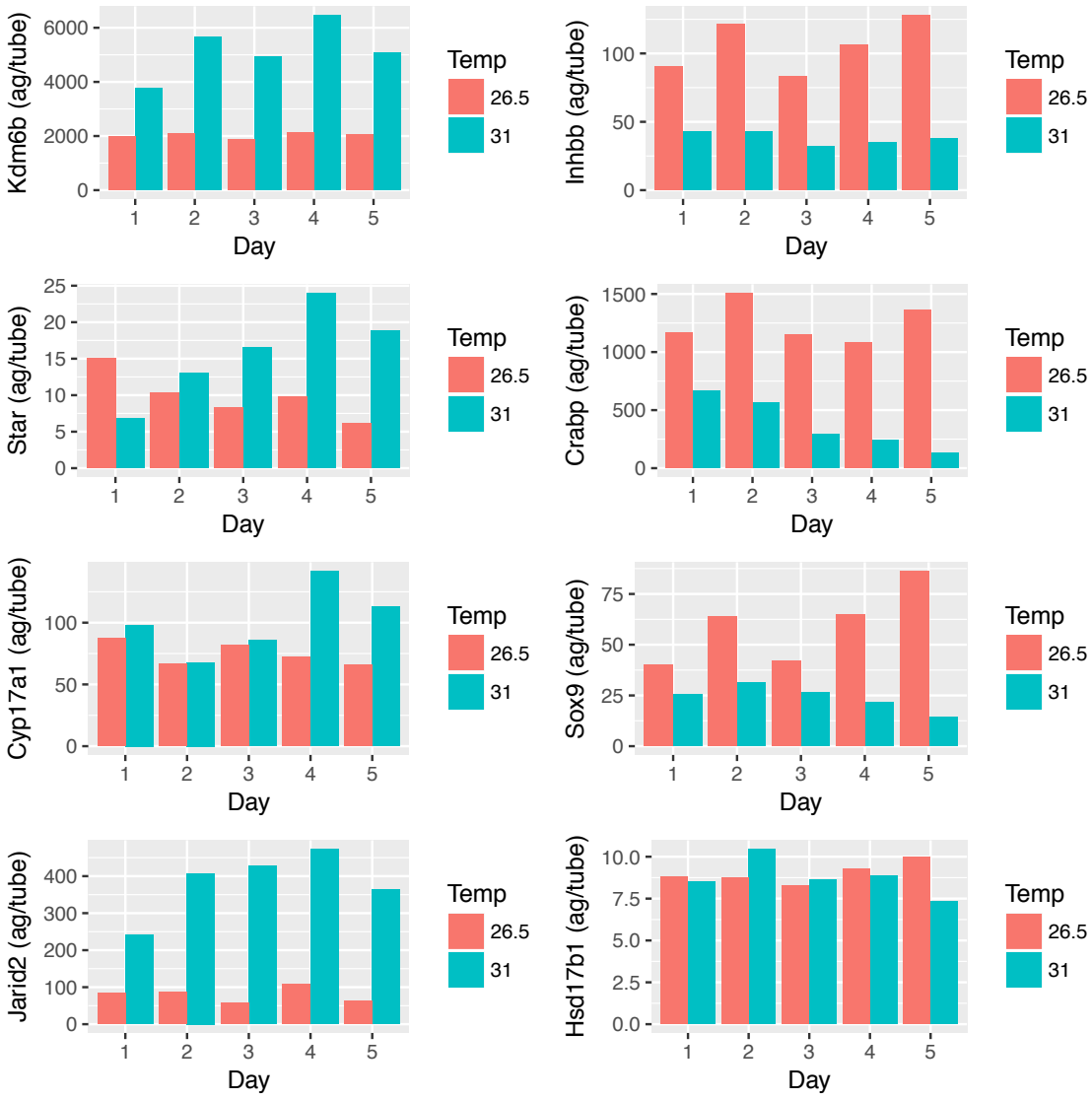


Figure 21. Gene enrichment analysis (Fisher's Exact Test) for the 123 DEGs responding to all 3 factors (incubation temperature, day and the interaction between incubation temperature and day). The annotated Chelydra serpentina transcriptome is used as the reference set and the differentially expressed genes is the test set.

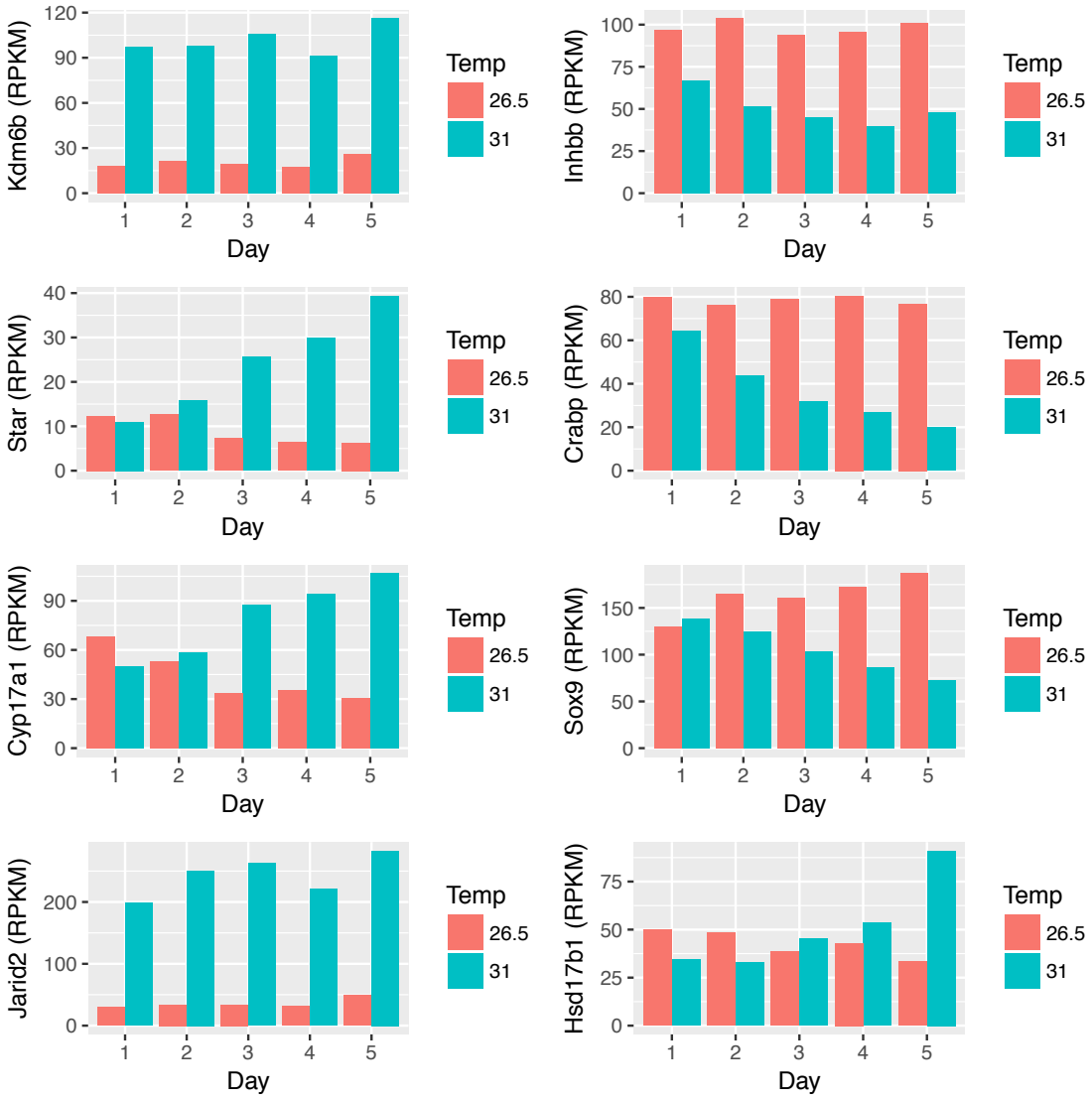
Validation Of RNA-Seq Results for Putative Sex-Determining Genes

We performed qPCR on *Kdm6b, Crabp, Star, Hsd17b1, Cyp17, Inhbb*, *Jarid2* and *Sox9* from the developing gonads at male and female temperatures to validate RNA-Seq results. Our qPCR results indicated that *Kdm6b, Star, Cyp17* and *Jarid2* were expressed at higher levels at FPT while expression of *Crabp, Inhbb* and *Sox9* was higher at MPT across time (Figure 22A). The RNA-seq analysis showed similar expression patterns for these genes (Figure 22B). The only difference between our qPCR and RNA-seq analyses

was the expression pattern of *Hsd17b1*. The RNA-seq analysis indicated the expression of *Hsd17b1* was higher in female gonads on Day 5 during the temperature shift while the qPCR showed the opposite expression pattern (Figure 22). We suspected the *Hsd17b1* gene tested in qPCR and RNA-seq may be two variances of the same gene or two different genes from the same gene family. This hypothesis will be tested in future studies using sequencing methods.



A



B

Figure 22. Expression patterns for 8 genes at MPT (26.5 ℃) or FPT (31 ℃) in the developing gonads of the snapping turtle during the 5-day temperature shift. A shows the absolute expression of mRNAs for the genes and B shows the RPKM for the genes.

Conclusion

With a total of 360.4 million single-ended reads, we assembled 270,094 non-redundant contigs. By comparing these contigs with the transcriptomes of 4 different species (human, chicken, painted turtle and Chinese softshell turtle), we generated a reference gene set comprised of 19,602 unique protein-coding genes. Functional

annotation of these genes was performed with GO analysis. Subsequent RNA-Seq analysis identified 293 temperature-responsive genes that are potentially located upstream in the gene regulatory cascade during TSD. More interestingly, our study confirmed the previously reported differential expression of two genes that are involved in epigenetic regulation, *Kdm6b* and *Jarid2*. Expression of these genes was affected in the first 24 hours of the temperature shift, suggesting epigenetic mechanisms might be involved in the earliest stages of TSD.

Our study provides a first description of the snapping turtle transcriptome and the effects of temperature on transcriptome-wide patterns of gene expression during the TSP. In this study, turtle eggs were incubated at 26.5°C, a potent masculinizing temperature, until embryos reached stage 17.5. Embryos at this stage are extremely sensitive to exposure to a female-producing temperature. A brief temperature shift to 31^{o}C at this stage will permanently change gonad fate. Previous studies have shown ovarian fate can be determined with a 5 day exposure to this female temperature (Rhen et al., 2015). The RNA-Seq analysis revealed the molecular changes underlying TSD. Several other genes changed their expression patterns within 24 hours of the MPT to FPT shift, suggesting their potential role in specification of gonad fate (Figure 18). The gonad at this phase still maintains its potential to develop into either ovaries or testes. After the first day, increasing numbers of genes were differentially expressed until ovarian fate is determined for all embryos on the 5th day. This accumulation of changes gradually tips the balance towards ovarian determination. The snapping turtle transcriptome and list of DEGs will guide future studies aimed at deciphering the molecular mechanisms of TSD at both the genetic and epigenetic level.

REFERENCES

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics 25, 25–29.

Baggerly, K.A., Deng, L., Morris, J.S., and Aldaz, C.M. (2003). Differential expression in SAGE: accounting for normal between-library variation. Bioinformatics 19, 1477–1483.

Bull, J.J. (1987). Temperature-sensitive periods of sex determination in a lizard: Similarities with turtles and crocodilians. The Journal of Experimental Zoology 241, 143–148.

Burke, R.L., and Calichio, A.M. (2014). Temperature-Dependent Sex Determination in the Diamond-backed Terrapin (Malaclemys terrapin). Journal of Herpetology 48, 466–470.

Crawford, J.E., Guelbeogo, W.M., Sanou, A., Traoré, A., Vernick, K.D., Sagnon, N., and Lazzaro, B.P. (2010). De Novo Transcriptome Sequencing in Anopheles funestus Using Illumina RNA-Seq Technology. Plos One 5, e14202.

Czerwinski, M., Natarajan, A., Barske, L., Looger, L.L., and Capel, B. (2016). A timecourse analysis of systemic and gonadal effects of temperature on sexual development of the red-eared slider turtle Trachemys scripta elegans. Developmental Biology 420, 166–177.

Deininger, P.L., and Batzer, M.A. (2002). Mammalian retroelements. Genome Research 12, 1455–1465.

Ewert, M.A., Lang, J.W., and Nelson, C.E. (1999). Geographic variation in the pattern of temperature-dependent sex determination in the American snapping turtle (Chelydra serpentina). Journal of Zoology 265, 81–95.

Eyre-Walker, A., and Hurst, L.D. (2001). OPINION: The evolution of isochores. Nature Reviews. Genetics 2, 549–555.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152.

Georges, A., Li, Q., Lian, J., O'Meally, D., Deakin, J., Wang, Z., Zhang, P., Fujita, M., Patel, H.R., Holleley, C.E., et al. (2015). High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard Pogona vitticeps. GigaScience 4.

Graves, J.A.M. (2016). Evolution of vertebrate sex chromosomes and dosage compensation. Nature Reviews. Genetics 17, 33–46.

Hudson, Q.J., Smith, C.A., and Sinclair, A.H. (2005). Aromatase inhibition reduces expression of *FOXL2* in the embryonic chicken ovary. Developmental Dynamics: an Official Publication of the American Association of Anatomists 233, 1052–1055.

Kent, J., Wheatley, S.C., Andrews, J.E., Sinclair, A.H., and Koopman, P. (1996). A male-specific role for SOX9 in vertebrate sex determination. Development (Cambridge, England) 122, 2813–2822.

Lan, F., Bayliss, P.E., Rinn, J.L., Whetstine, J.R., Wang, J.K., Chen, S., Iwase, S., Alpatov, R., Issaeva, I., Canaani, E., et al. (2007). A histone H3 lysine 27 demethylase regulates animal posterior development. Nature 449, 689–694.

Lance, V.A., Elsey, R.M., and Lang, J.W. (2000). Sex ratios of American alligators (Crocodylidae): male or female biased? Journal of Zoology 252, 71–78.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Lang, J.W., and Andrews, H.V. (1994). Temperature-dependent sex determination in crocodilians. Journal of Experimental Zoology 270, 28–44.

Loffler, K.A., Zarkower, D., and Koopman, P. (2003). Etiology of ovarian failure in blepharophimosis ptosis epicanthus inversus syndrome: *FOXL2* is a conserved, early-acting gene in vertebrate ovarian development. Endocrinology 144, 3237–3243.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics 9, 387–402.

Miller, H.C., Biggs, P.J., Voelckel, C., and Nelson, N.J. (2012). De novo sequence assembly and characterisation of a partial transcriptome for an evolutionarily distinct reptile, the tuatara (Sphenodon punctatus). BMC Genomics 13, 439.

Münsterberg, A., and Lovell-Badge, R. (1991). Expression of the mouse anti-müllerian hormone gene suggests a role in both male and female sexual differentiation. Development (Cambridge, England) 113, 613–624.

Pasini, D., Cloos, P.A.C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature 464, 306–310.

Pieau, C., Dorizzi, M., and Richard-Mercier, N. (2001). Temperature-dependent sex determination and gonadal differentiation in reptiles. Exs 117–141.

Pieau, C., and Dorizzi, M. (1981). Determination of temperature sensitive stages for sexual differentiation of the gonads in embryos of the turtle, Emys orbicularis. Journal of Morphology 170, 373–382.

Raymond, C.S., Murphy, M.W., O'Sullivan, M.G., Bardwell, V.J., and Zarkower, D. (2000). Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. Genes \& Development 14, 2587–2595.

Renaut, S., Nolte, A.W., and Bernatchez, L. (2010). Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (Coregonus spp. Salmonidae). Molecular Ecology 19, 115–131.

Rhen, T., and Schroeder, A. (2010). Molecular mechanisms of sex determination in reptiles. Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation 4, 16–28.

Rhen, T., Metzger, K., Schroeder, A., and Woodward, R. (2007). Expression of putative sex-determining genes during the thermosensitive period of gonad development in the snapping turtle, Chelydra serpentina. Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation 1, 255–270.

Rhen, T., and Lang, J.W. (1998). Among-family variation for environmental sex determination in reptiles. Evolution 52, 1514–1520.

Rhen, T., Fagerlie, R., Schroeder, A., Crossley, D.A., and Lang, J.W. (2015). Molecular and morphological differentiation of testes and ovaries in relation to the thermosensitive period of gonad development in the snapping turtle, Chelydra serpentina. Differentiation; Research in Biological Diversity 89, 31–41.

Rhen, T., Jangula, A., Schroeder, A., and Woodward-Bosh, R. (2009). The platelet-derived growth factor signaling system in snapping turtle embryos, Chelydra serpentina: potential role in temperature-dependent sex determination and testis development. General and Comparative Endocrinology 161, 335–343.

Schroeder, A.L., Metzger, K.J., Miller, A., and Rhen, T. (2016). A Novel Candidate Gene for Temperature-Dependent Sex Determination in the Common Snapping Turtle. Genetics 203, 557–571.

Sen, G.L., Webster, D.E., Barragan, D.I., Chang, H.Y., and Khavari, P.A. (2008). Control of differentiation in a self-renewing mammalian tissue by the histone demethylase JMJD3. Genes \& Development 22, 1865–1870.

Shi, Y. (2007). Histone lysine demethylases: emerging roles in development, physiology and disease. Nature Reviews. Genetics 8, 829–833.

Shoemaker, C.M., Queen, J., and Crews, D. (2007). Response of candidate sex-determining genes to changes in temperature reveals their involvement in the molecular network underlying temperature-dependent sex determination. Molecular Endocrinology (Baltimore, Md.) 21, 2750–2763.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, btv351–btv3212.

Siroski, P., Larriera, A., Piña, C., Lance, V., and Verdade, L. (2007). The temperature-sensitive period (TSP) during incubation of broad-snouted caiman (Caiman latirostris) eggs. Amphibia-Reptilia 28, 123–128.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [Et Al.] Chapter 4, Unit4.10.

Western, P.S., Harry, J.L., Graves, J.A., and Sinclair, A.H. (1999). Temperature-dependent sex determination in the American alligator: AMH precedes SOX9 expression. Developmental Dynamics: an Official Publication of the American Association of Anatomists 216, 411–419.

Wilhelm, D., Palmer, S., and Koopman, P. (2007). Sex determination and gonadal development in mammals. Physiological Reviews 87, 1–28.

WITSCHI, E. (1959). Age of sex-determining mechanisms in vertebrates. Science (New York, N.Y.) 130, 372–375.

Yatsu, R., Miyagawa, S., Kohno, S., Parrott, B.B., Yamaguchi, K., Ogino, Y., Miyakawa, H., Lowers, R.H., Shigenobu, S., Guillette, L.J., et al. (2016). RNA-seq analysis of the gonadal transcriptome during Alligator mississippiensis temperature-dependent sex determination and differentiation. BMC Genomics 17, 1–13.

Ye, L., Fan, Z., Yu, B., Chang, J., Hezaimi, Al, K., Zhou, X., Park, N.-H., and Wang, C.-Y. (2012). Histone demethylases KDM4B and KDM6B promotes osteogenic differentiation of human MSCs. Cell Stem Cell 11, 50–61.

Yntema, C.L. (1968). A series of stages in the embryonic development of Chelydra serpentina. Journal of Morphology 125, 219–251.

Yntema, C.L. (1979). Temperature levels and periods of sex determination during incubation of eggs of Chelydra serpentina. Journal of Morphology 159, 17–27.

Zeng, V., Villanueva, K.E., Ewen-Campen, B.S., Alwes, F., Browne, W.E., and Extavour, C.G. (2011). De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean Parhyale hawaiensis. BMC Genomics 12, 581.

Zhou, Y., Bizzaro, J.W., and Marx, K.A. (2004). Homopolymer tract length dependent enrichments in functional regions of 27 eukaryotes and their novel dependence on the organism DNA (G+C)% composition. BMC Genomics 5, 95

CHAPTER IV

RECONSTRUCTION OF GENE REGULATORY NETWORKS USING PUBLICLY AVAILABLE MICROARRAY DATA REVEALS NOVEL REGULATORS FOR SEX DETERMINATION

Abstract

Mammalian sexual development is a unique biological process in which a common precursor, the bipotential genital ridge, differentiates into two morphologically distinct organs, testes and ovaries. The molecular pathways that specify gonad differentiation are still poorly understood. To identify the complex interplay of cellular signals that regulates this process, this study reconstructed gene regulatory networks using a large number of gene expression profiles from public microarray experiments. We reconstructed gene regulatory networks using an entropy based network reconstructing algorithm – ARACNE. We then applied hub gene analyses and master regulator analyses to identify genes playing crucial roles in gonad fate determination in XX samples and XY samples. The functional enrichment analyses performed on 100 most connected genes in both XX and XY samples suggest the basic molecular pathways underlying gonadal development differ between sexes. The master regulator analyses identified 110 candidate sex-determining genes including both known sex-determining genes and novel candidate genes. In addition, the comparison between the inferred interaction partners for *Sox9* and *Sry* demonstrated the networks inferred in this study

were reliable. This study provides an overview of the transcriptional pathways underlying mammalian gonad determination and will guide the direction of future studies in the field of sex determination.

Introduction

Sexual reproduction is nearly universal in multicellular animals. Yet the genetic and molecular mechanisms underlying this complex process have not been fully elucidated. To allow sexual reproduction to take place, animals must be prepared both anatomically and physiologically. The sex of an individual is determined as early as during the embryonic phase (sex determination) with subsequent development of all other differences between the sexes (sexual differentiation). In mice, sex is determined by sex chromosomes, X and Y, at mid-gestation after which the bipotential gonads start to differentiate into testes or ovaries (Wilhelm, Palmer, & Koopman, 2007). The bipotential gonad, which is competent to differentiate into a testis or an ovary regardless of sex chromosomes, is the initial developmental stage of the gonad (Brennan & Capel, 2004). There are four main cell lineages that comprise the bipotential gonad, which includes supporting cells, interstitial/stromal cells, germ cells, and endothelial cells (Jameson et al., 2012b). Cells of each gonadal lineage are involved in a binary fate decision during primary sex determination (Adams & McLaren, 2002; Albrecht & Eicher, 2001). The plasticity of the biopotential gonad and rapid cell fate transitions allow us to investigate the dynamics of gene regulatory networks in developmental systems.

The transient sexual plasticity of the bipotential gonad is a result of a balanced network state established by antagonistic signals (Kim et al., 2006). A sex-determining switch can tip this balance toward one of two opposite sexual fates. In mammals, the

transcription factor *Sry* on the Y chromosome is the genetic switch responsible for directing the bipotential gonad to a testicular fate. *Sry* is probably the most well studied male-determining gene, which is only expressed in precursors of the somatic supporting cell lineage for a short period during gonadogenesis (Bullejos & Koopman, 2001). Its expression activates a cascade of signaling pathways to enable testes differentiation and repress ovary development (Hiramatsu et al., 2009). The closest counterpart for *Sry* in female determination is *Foxl2*, an antagonist of *Sox9*, which is the direct target of *Sry* (Uhlenhaut et al., 2009). Although *Foxl2* is not required for primary sex determination in mice, it is a female sex-determining gene in goat and is needed for maintaining ovarian cell identity in mice (Boulanger et al., 2014; Ottolenghi et al., 2005). Failure to activate or maintain expression of these sex-determining switches can disrupt gonadal development and sometimes cause sex reversal (Jameson, Lin, & Capel, 2012a; Uhlenhaut et al., 2009).

Besides the key sex-determining genes that have been studied in detail, there are many more genes playing important yet unknown roles in sex determination. The mouse gonad forms initially at around 10 days post coitum (dpc) or embryonic day 10 (E10.0) as a bipotential gonad capable of developing into either testes or ovaries. Expression of key sex-determining genes such as *Sry* and *Foxl2* can be detected as early as E10.5 and E11.5 respectively (Greenfield, 2015). Meanwhile, over 2,000 genes are differentially expressed between the sexes during the short sex determination window, meaning much of the mystery of sex determination and differentiation still waits to be unraveled (Beverdam & Koopman, 2006; Small, Shima, Uzumcu, Skinner, & Griswold, 2005).

Efforts have been made to identify novel genes involved in sex determination and the regulatory cascade controlling this process. Some studies used high-throughput whole-mount *in situ* hybridization to identify genes specifically expressed in the developing gonad (Wertz & Herrmann, 2000). Some studies used microarrays to determine the expression profiles of whole embryonic mouse gonads and identified candidate sex-determining genes with subsequent differential expression analysis (Munger et al., 2009; Small et al., 2005). Others went further by examining gene expression profiles in separate cell lineages in the developing gonad in a fine-tuned time course (Jameson et al., 2012b; Munger, Natarajan, Looger, Ohler, & Capel, 2013). However, none of these studies revealed how these genes are regulated specifically in gonads and even less is known about how they interact with each other. To fully understand the mechanisms of sex determination, the gene regulatory network during this biological process needs to be reconstructed.

High-throughput technologies such as microarray and RNA-Seq provide us with powerful means of identifying large numbers of differentially expressed genes among samples of interest at transcriptome/genome-wide scales. Reconstructing regulatory networks based on the gene expression profiles generated by these tools has proven to be promising approach to answering complex questions in many biological and medical fields (Cho, Kim, & Przytycka, 2012; Thompson, Regev, & Roy, 2015).

Numerous computational algorithms were developed to dissect genome-wide gene regulatory networks, which can be put into 4 categories – 1) optimization methods which maximize a scoring function over alternative network models, 2) regression techniques which fit the data to a priori models and are limited to relatively simple

models, 3) integrative bioinformatics approaches which combine data from a number of independent studies and 4) correlation methods which rely on a variety of pairwise gene expression correlation measures (Margolin, Nemenman, et al., 2006a). Though some of these methods were successfully applied to infer regulatory modules from gene expression data in simple eukaryotes, various limitations confine their application to small and less complex networks (Margolin, Nemenman, et al., 2006a). A greater challenge arises when trying to organize large number of genes into complex, functionally meaningful networks in higher-order eukaryotes (Jiang, Tang, & Zhang, 2004). A number of algorithms have been proposed in recent years, including entropy-based network modeling (Margolin, Nemenman, et al., 2006a; Villaverde, Ross, Morán, & Banga, 2014; J. Wang et al., 2013), networks based on marginal dependencies (Liu et al., 2016), network reconstruction by integrating prior biological knowledge (Yupeng Li & Jackson, 2015), or integrative predictions from multiple inference methods (Ceci, Pio, Kuzmanovski, & Džeroski, 2015).

The interactions between genes are not always linear and straightforward. They can be nonlinear, condition-dependent or time-lagged (Liu et al., 2016). Previously proposed linear models in most studies are restricted not only by the need for estimating linear high-dimensional dependency structures but also suffer from limited ability to capture nonlinear interactions (Hausser & Strimmer, 2009). To loosen the assumptions of linear models and capture nonlinear associations among genes, entropy-based network reconstructing algorithms, such as ARACNE and MRNET, were proposed (Margolin, Nemenman, et al., 2006a; Meyer, Kontos, Lafitte, & Bontempi, 2007). These methods rely on computing the mutual information (MI) between genes, a concept arising in

probability theory and information theory. Mutual information is always positive if two variables are related and zero if they are independent regardless whether their relationship is linear or nonlinear (Kraskov, Stögbauer, Andrzejak, & Grassberger, 2003). This makes MI an ideal measure for identifying genes with correlated expression patterns.

This study aims to elucidate the molecular mechanisms of mammalian sex determination by reconstructing a gene regulatory network in the developing mouse gonad using one of the entropy-based network reconstructing algorithms – ARACNE. This algorithm enriches for direct gene-gene interactions by applying a property of MI, known as data processing inequality (DPI) (Margolin, Wang, Lim, Kustagi, Nemenman, & Califano, 2006c). Such direct regulatory interactions may be mediated by transcription factor (TF) binding activities though many other types of regulatory interactions are also identified as ARACNE is agnostic to the molecular details of the interactions. Transcription factors are essential for the regulation of gene expression and many of them are involved in animal development, including sex determination.

To reconstruct gene regulatory networks, we applied ARACNE to microarray expression profile data of developing mouse gonads and isolated cells at different time points with separate female (XX) and male (XY) samples. The resulted regulatory network was then interrogated by means of hub gene analysis and master regulator analysis (MRA) algorithm which tests for overlap between a TF regulon (TF targets inferred by ARACNE) and genes that are differentially expressed between XX and XY samples (Carro et al., 2010). The hub gene analysis and MRA algorithm helped us identify important TFs (master regulators) and novel pathways that might play key roles in sex determination and differentiation.

In this study, we (i) reconstructed gene regulatory networks in the developing mouse gonad by applying ARACNE to the gene expression profiles of 112 female samples, 114 male samples and 226 combined samples from 10 publicly available datasets; (ii) inferred critical hub genes, master regulators (MR) and novel regulatory relationships of well-studied sex-determining genes responsible for sex determination and differentiation.

Materials and Methods

Microarray Data Processing and Cross-Platform Normalization

The gene expression data were obtained from the NCBI Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo/) using "GEOquery" Bioconductor package (Davis & Meltzer, 2007). To reconstruct transcriptional networks for the developing gonad during sex determination, it is ideal to have a large number of gene expression profiles that cover cell lineages of the developing gonad at different developing stages. In this study, we used 10 previously published gene expression datasets available from the GEO portal, which included GSE27715, GSE41948, GSE85267, GSE23908, GSE18211, GSE3463, GSE4928, GSE4818, GSE6916 and GSE5334. The mouse developmental stages covered by these datasets range from 10.5 dpc to 18 dpc and the cell lineages examined include supporting cells, interstitial/stromal cells, germ cells and endothelial cells (Table 6). To focus on the sex-determining process, we only used gene expression profiles during the critical sex determination window (10.5 dpc ~ 13.5 dpc). To avoid possible perturbation of the underlying gene regulatory networks, we also removed all transgenic samples (mutants, gene knockouts, over-

expression and so on). Among a total of 342 microarrays in the 10 studies, we used 226 microarrays that covered stage 10.5 dpc to 13.5 dpc. Among them, 112 arrays were from XX gonads and 114 were from XY gonads.

Table 6: The microarray data from 10 studies

GEO Dataset	Sample Size	Developmental Stage Covered	Tissue Type Covered
GSE27715	91	E11.5, E12.5, E13.5	Germ cells, supporting cells, interstitial cells
GSE41948	74	E11, E11.2, E11.4, E11.6, E11.8, E12	Whole gonad
GSE85267	54	E11.5, E12.5, E13.5	Supporting, interstitial/stromal, and germ cells
GSE23908	31	E12, E14, E16	Whole gonad but without germ cells
GSE18211	12	E11.5, E12.5	Somatic support cells
GSE3463	12	E10.5, E11.5	Somatic gonadal cells
GSE4928	8	E13	Whole gonad and somatic gonadal cells
GSE4818	21	E11, 12, 14, 16, 18	Whole testis
GSE6916	20	E11.5, 12.5, 14.5, 16.5, 18.5	Whole gonad
GSE5334	19	E11, 12, 14, 16, 18	Whole ovary

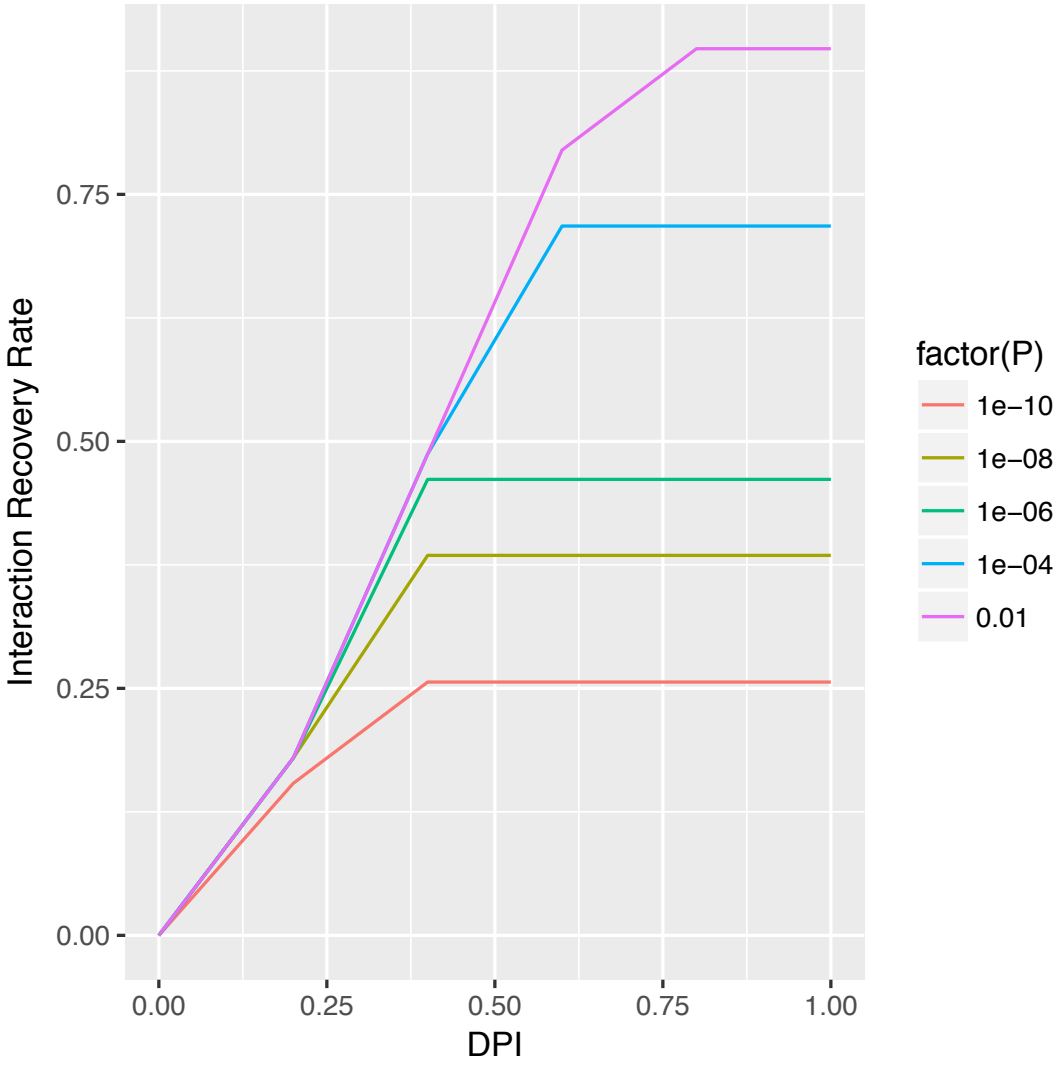
The original investigators used a variety of microarray normalization methods. We $\log_2$ transformed expression values if they were not already transformed for cross-platform normalization. We annotated each dataset with the gene symbols provided by investigators and collapsed multiple probes that represented the same gene by the median expression value. To integrate the gene expression profiles from different platforms, we generated a common gene list that contains 10,052 genes that were represented in all 10 studies. For each dataset, genes that do not belong to the common gene list were removed along with their expression values. To merge 10 gene expression studies into a single and unified dataset with minimal batch effects, we applied the cross-study normalization method Combat which (Johnson, Li, & Rabinovic, 2007). This unified dataset was then

divided into female (XX) and male (XY) subsets for further analyses. The resulting normalized datasets (the unified dataset, the female and the male subsets) were then used as input for the ARACNE algorithm.

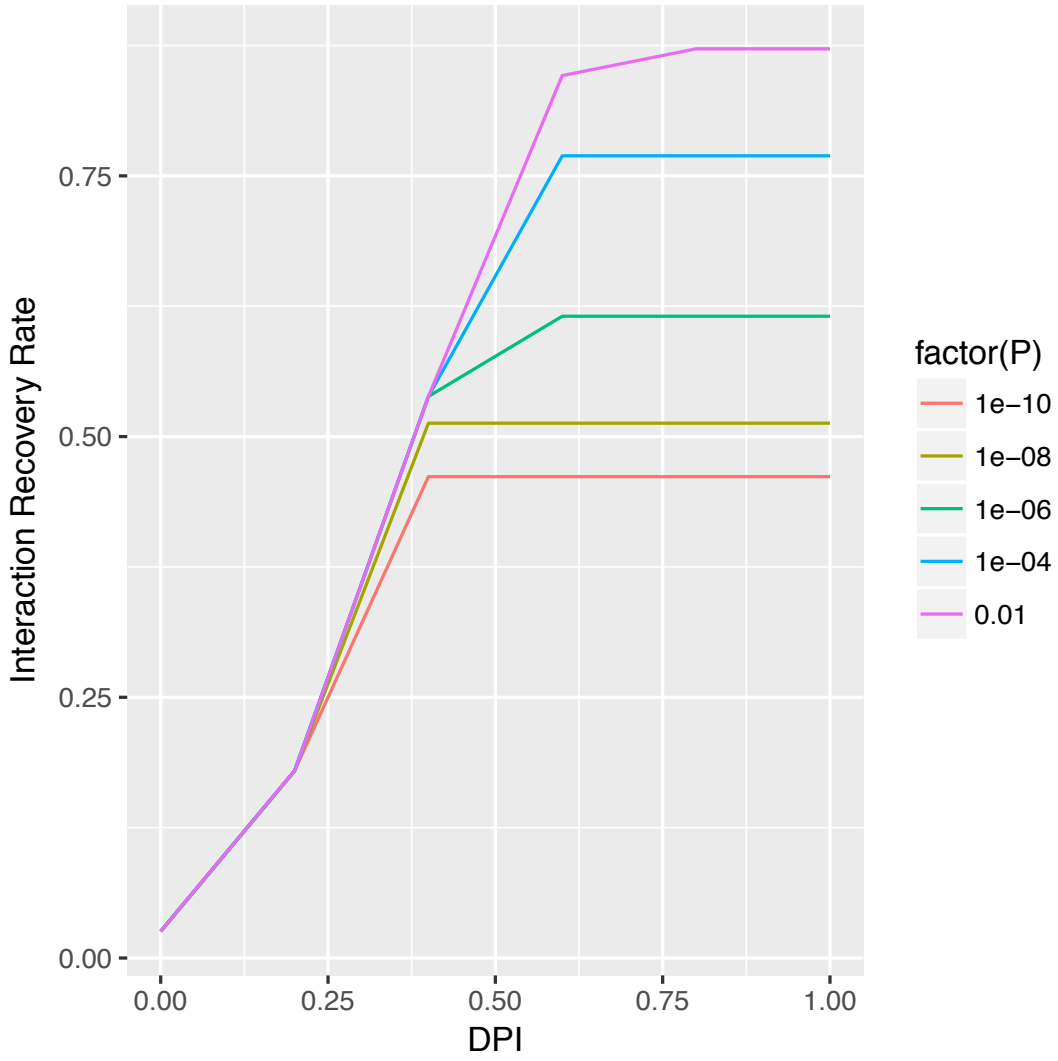
Network Inference Using ARACNE

We reconstructed the gene regulatory network for the XX gonad, the XY gonad and the entire developing gonad regardless of its sex using the ARACNE algorithm (Margolin et al., 2006a). The entropy based algorithm uses mutual information to identify regulatory interactions between genes whatever the underlying mechanism. To examine potential transcriptional interactions during the DPI process, we generated a list of TFs (737 TFs in total) by identifying all TFs in the common gene list using AmiGO2 (Carbon et al., 2009). The transcriptional regulation network for the mouse gonad during the sex determination period was reconstructed using the Linux command-line ARACNE program (http://califano.c2b2.columbia.edu/ARACNE/). The ARACNE configuration files (config_kernel.txt and config_threshold.txt) were generated individually for each dataset using the author provided Matlab scripts. To infer direct interactions between genes with high fidelity, the algorithm relies on two parameter settings – a specific p-value which is used to filter out insignificant MI values and a DPI value which is used by ARACNE to remove indirect interactions (Margolin et al., 2006a). To find out the appropriate p-value and DPI combination for each of our datasets, we performed ARACNE on each dataset with 30 different p-value and DPI combinations, which resulted in 30 networks for each dataset and 90 networks in total. The recovery rate of 39 known mechanistic interactions between sex-determining genes (Table 7) was examined for each network and the results are shown in Figure 23. To optimize the balance between

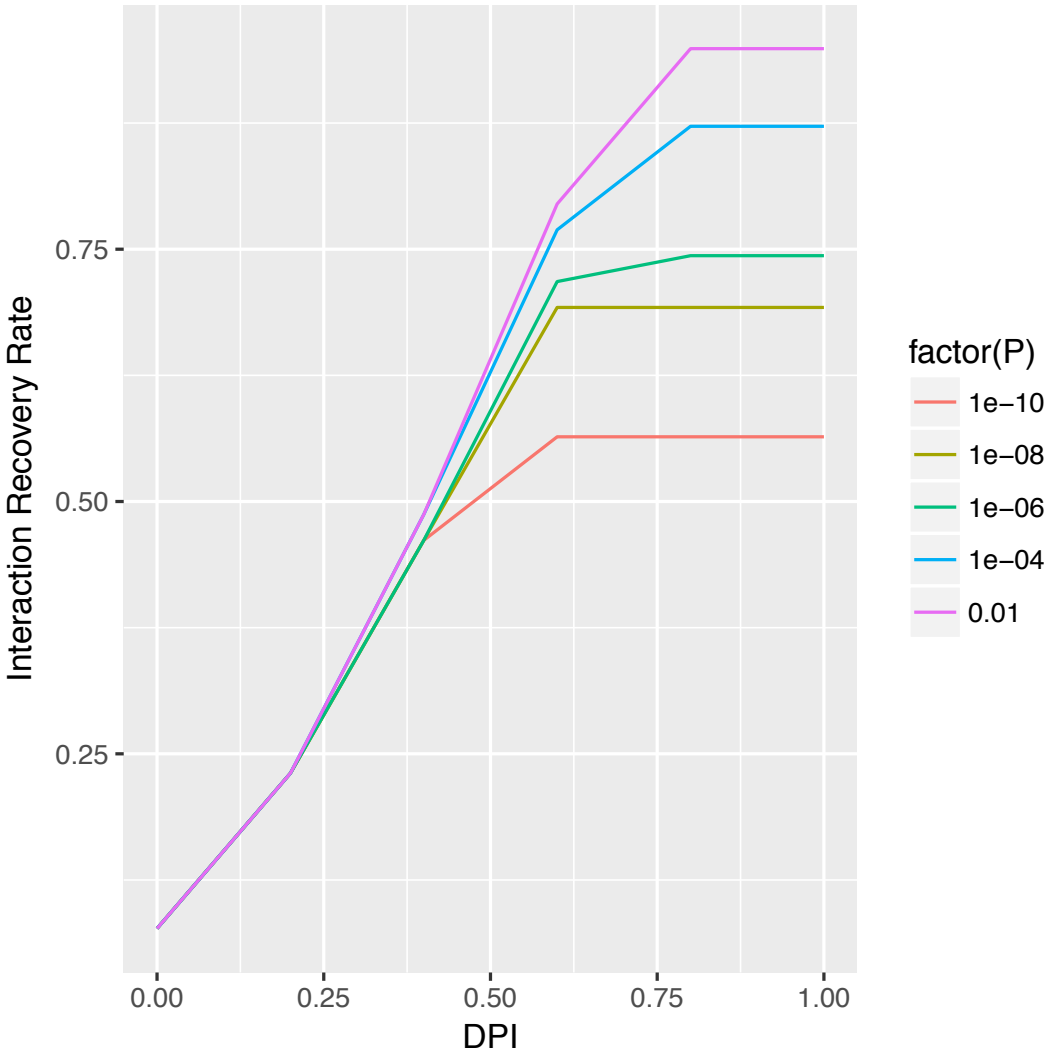
false positive and false negative error probabilities when inferring gene-gene interactions, we set the p-value and DPI to the level at which at least 50% of the 39 validated interactions were recovered (Figure 23). That is $p \leq 1e\text{-}4$ and DPI = 0.4 for male and female datasets (Figure 23A, 23B) and is $p \leq 1e\text{-}6$ and DPI = 0.6 (Figure 23C) for the combined dataset.



A


Interaction Recovery Rate
0.00
0.25
0.50
0.75
0.00
0.25
0.50
0.75
1.00
DPI
factor(P)
1e−10
1e−08
1e−06
1e−04
0.01

B



C

Figure 23. The recovery rate of 39 known mechanistic interactions between sex-determining genes in the ARACNE inferred networks with different P and DPI settings. Figure A, B and C show the recovery rates in the networks reconstructed from XX gonad samples, XY gonad samples and combined gonad samples respectively.

Table 7: Known mechanistic interactions between sex-determining genes

Sex	Gene	Gene	Interaction	Experimental Evidence	Citation
Male	*Sox9*	*Amh*	+	KO/EMSA	Barrionuevo et al., 2006/De Santa Barbara et al., 1998
Male	*Sox9*	*Cyp26b1*	+	overexpression&KO/chip-chip	Kashimada et al., 2011; Li et al., 2014
Male	*Sox9*	*Dhh*	+	KO/chip-chip	Barrionuevo et al., 2009; Li et al., 2014
Male	*Sox9*	*Vnn1*	+	Reporter gene assay & EMSA	Wilson et al., 2005
Male	*Sox9*	*Cbln4*	+	knockdown & overexpression & EMSA	Bradford et al., 2009
Male	*Sox9*	*Ptgds*	+	KO/chip-chip	Moniot et al., 2009; Li et al., 2014
Male	*Sox9*	*Etv5*	+	ChIP, knockdown & overexpression	Alankarage et al., 2016
Male	*Nr5a1*	*Sox9*	+	two-hybrid assay, reporter gene assay, ChIP	De Santa Barbara et al., 1998; Münsterberg and Lovell-Badge, 1991
Male	*Nr5a1*	*Amh*	+	EMSA	Natchigal et al., 1998; De Santa Barbara et al., 1998
Male	*Wt1*	*Sox9*	+	KO	Hammes et al., 2001
Male	*WT1*	*Amhr2*	+	ChIP, knockdown & overexpression, reporter gene assay, EMSA	Klattig et al., 2007
Male	*WT1*	*Amh*	+	Reporter gene assay	Natchigal et al., 1998
Male	*Ovol1*	*Id2*	-	promoter-luciferase reporter gene assay	Li et al., 2005
Male	*Rnf2*	*Stra8*	-	KO	Yokobayashi et al., 2013
Male	*Ring1*	*Stra8*	-	KO	Yokobayashi et al., 2013
Male	*Nr5a1*	*Cyp11a1*	+	overexpression & knockdown	Bashamboo et al., 2010

Sex	Gene	Gene	Interaction	Experimental Evidence	Citation
Male	*Nr5a1*	*Cyp17*	+	Reporter gene assay	Li et al., 2007
Male	*Nr5a1*	*Vnn1*	+	promoter-luciferase reporter gene assay	Wilson et al., 2005
Male	*Notch2*	*Hes1*	+	KO	Liu et al., 2016
Male	*Nr5a1*	*Vcam1*	-	ChIP-seq & RNA-seq & knockdown	Baba et al., 2014
Male	*Nr5a1*	*Bmp2*	+	ChIP-seq & RNA-seq & knockdown	Baba et al., 2014
Male	*Nr5a1*	*Cyp26b1*	+	overexpression&KO	Kashimada et al., 2011
Male	*Dhh*	*Nr5a1*	+	KO	Yao et al., 2002
Male	*Dhh*	*Cyp11a1*	+	KO	Yao et al., 2002
Male	*Nr5a1*	*Aldoa*	+	ChIP-seq & RNA-seq & knockdown	Baba et al., 2014
Female	*Wnt4*	*Fst*	+	KO	Yao et al., 2004
Female	*Wnt4*	*Runx1*	+	KO	Naillat et al., 2015
Female	*Wnt4*	*Star*	-	ChIP	Jordan et al., 2003
Female	*Ctnnb1*	*Fst*	+	overexpression	Boyer et al., 2012
Female	*Ctnnb1*	*Axin2*	+	Reporter gene assay	Jho et al., 2002
Female	*Wnt4*	*Bmp2*	+	KO	Yao et al., 2004
Female	*Ctnnb1*	*Sox9*	-	various assays	Bernard et al., 2012
Female	*Rarb*	*Stra8*	+	EMSA/overexpression	Taneja et al., 1995; Koubova et al., 2006
Female	*Cyp26b1*	*Stra8*	-	overexpression	Koubova et al., 2006
Female	*Msx1*	*Stra8*	+	KO	Le Bouffant et al., 2011
Female	*Dmrt1*	*Stra8*	-	KO, ChIP	Krentz et al. 2011
Female	*Stra8*	*Sycp3*	+	Knockdown	Soh et al., 2015
Female	*Stra8*	*Msh5*	+	Knockdown	Soh et al., 2015
Female	*Dazl*	*Stra8*	+	Knockdown	Lin et al., 2008

Hub Gene Analysis and Known Sex-Determining Gene Analysis

In network science, hubs refer to nodes with a number of links that greatly exceeds the average. In a gene regulatory network, hub genes have a large number of interaction partners and often have significantly different biological properties than non-hub genes (Almaas, 2007). To identify hub genes in the XX network and the XY network, and potential molecular mechanisms underlying ovary and testis determination and differentiation, we sorted all genes in each network based on the number of interaction partners. Frequencies of genes with and the numbers of interactions they had in both female and male networks were shown in the histograms in Figure 24. Genes with 1,000 to 1,500 interactions were most frequent in both female and male networks. Both histograms showed a declining frequency pattern of genes as interaction numbers increased. However, the male network tends to have more genes with 2,000 ~ 3,000 interactions compared to the female network. The top 100 genes in the above list, which had about more than 4,000 interaction partners, were selected for gene ontology (GO) term enrichment analysis using BLAST2GO (version 4.0.7). We also compared the ARACNE inferred interaction partners of *Sry* and *Sox9* to experimentally determined target genes.

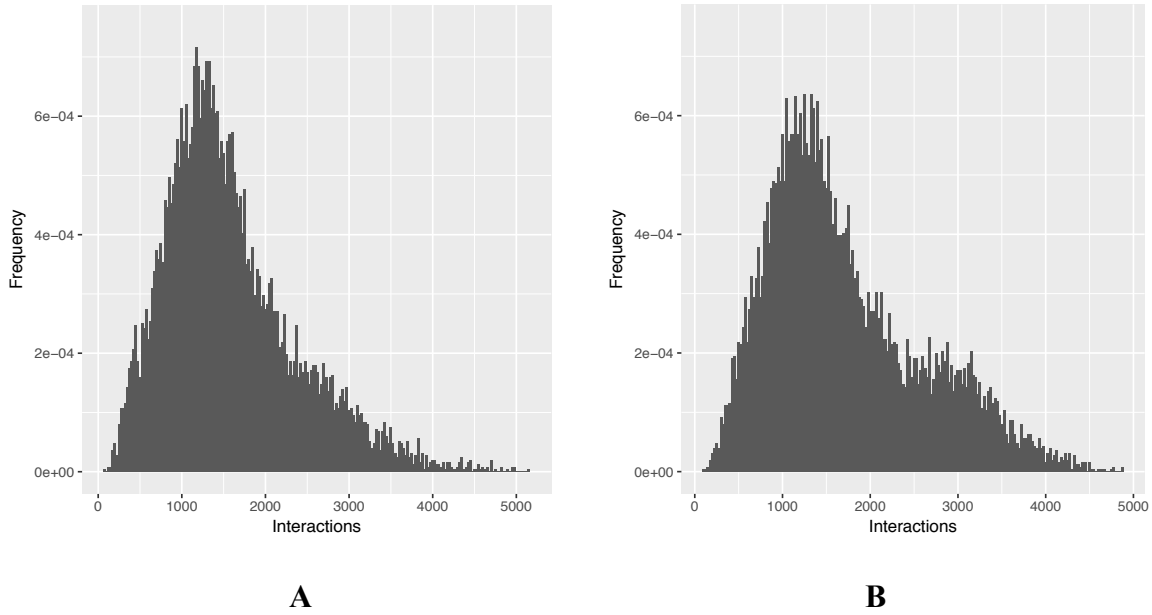


Figure 24: The distribution of genes in the ARACNE inferred gene regulatory networks based on their number of interactions. Figure A shows the gene distribution frequency in the ARACNE inferred female network while Figure B shows the gene distribution frequency in the ARACNE inferred male network.

Master Regulator Analysis

To identify transcription factors that play key roles in sex determination and differentiation (master regulators), we performed MRA analysis on the combined dataset. The master regulator analysis (MRA) algorithm tests whether putative TF targets (i.e., regulons) are enriched within a set of differentially expressed genes (Carro et al., 2010). There are two different algorithms for evaluating the significance of the enrichment, Fisher's exact test (FET) and gene set enrichment analysis (GSEA) (Subramanian et al., 2005). In this work, we used MRA-FET method with Bonferroni correction for multiple-testing for the enrichment analysis. The MRA algorithm requires a list of transcription factors, an interaction network for those transcription factors, and a list of differentially expressed genes. In our case, the interaction network was inferred by ARACNE, the

candidate master regulators were the 737 transcription factors identified above, and the differentially expressed genes (DEGs) generated by Student's *t*-test with Bonferroni correction and a p-value less than 0.01 between XX and XY samples. The MRA-FET was performed on an open-source software platform – geworkbench 2.6.0 (Floratos, Smith, Ji, Watkinson, & Califano, 2010).

Results and Discussion

Reconstruction of Mouse Gonad Gene Regulatory Network

Microarrays have been widely used to simultaneously measure the expression of thousands of genes. Only a small subset of DEGs, also known as "a gene signature", have a collective expression pattern that is unique to a trait of interest (Chang et al., 2011). Efforts have been made to identify gene signatures for traits of interest in different research fields, such as basic biology and medical science. However, it has often been found that gene signatures derived from different microarray studies for the same trait show little overlap (Shen, Chinnaiyan, & Ghosh, 2008). Low reproducibility may be caused by differences in sample collection methods, processing protocols, microarray platforms, normalization methods and small sample sizes (Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma et al., 2008). Integration of multiple microarray datasets has been shown to improve detection of gene signatures by increasing sample sizes, attenuating data heterogeneity and reducing study-specific biases (Hamid et al., 2009; Hu, Greenwood, & Beyene, 2005; Shabalin, Tjelmeland, Fan, Perou, & Nobel, 2008; Taminau, Lazar, Meganck, & Nowé, 2014). Information from multiple independent microarray studies performed on different

platforms can be combined either at an early stage (cross-platform normalization/merging) or at a late stage (meta-analysis/integrative analysis) (Walsh et al., 2015). With meta-analysis, analyses are performed for each experiment first and their results are subsequently combined. With cross-platform normalization, also known as data merging, datasets from different studies are first merged into a single dataset and then analyzed (Walsh et al., 2015).

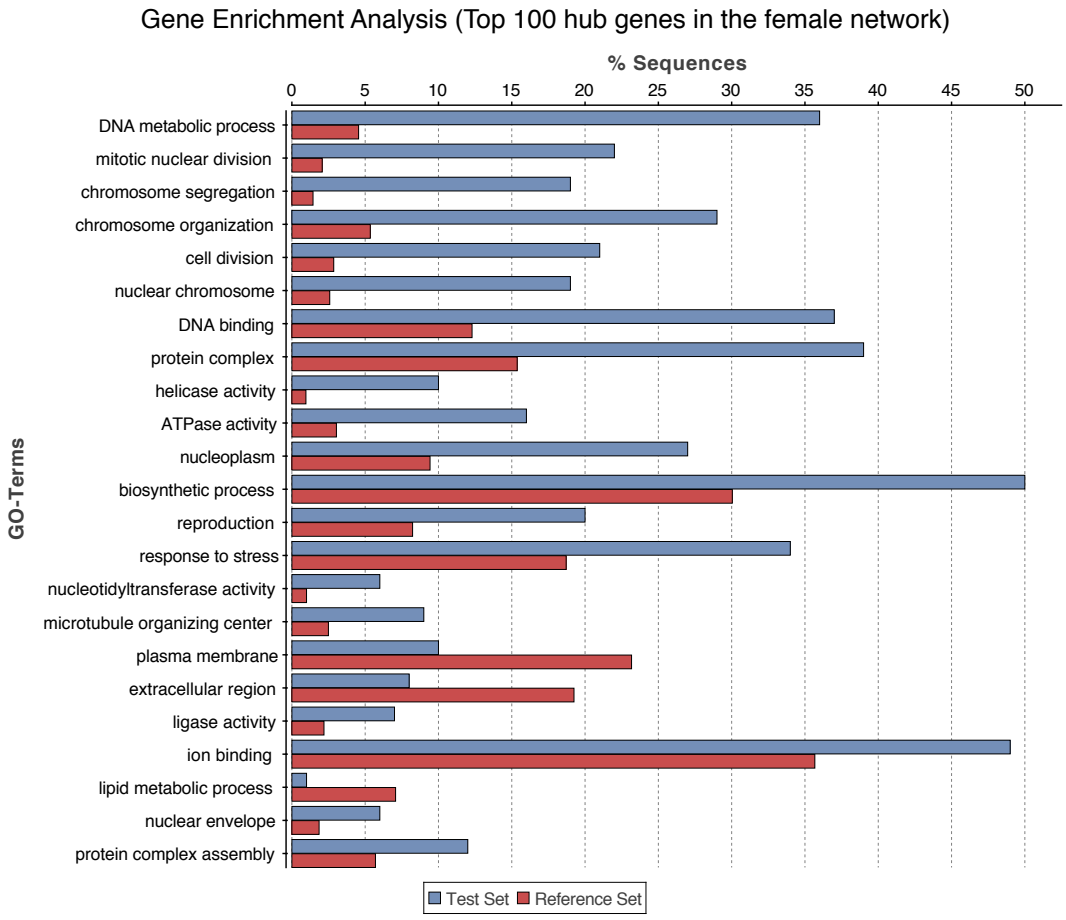
With the normalized and unified microarray expression profiles, ARACNE was able to reliably estimate the MI, a measure of the statistical dependence between expression levels of two genes. Individual studies of mouse gonad development during 10.5 dpc ~ 13.5 dpc usually have small sample sizes. Besides, microarray expression profiles from different platforms are often heterogeneous in genes and normalization methods. It is also necessary to accurately identify the gene signature for subsequent MRA analysis. Therefore, merging microarray datasets from different studies into a unified single dataset is arguably the best approach for running ARACNE and network analyses. Different cross-platform normalization methods have been compared to determine which method is most effective in reducing batch effects (Rudy & Valafar, 2011; Turnbull et al., 2012). Four cross-platform normalization methods, Combat, XPN, DWD and GQ, stand out in their ability to substantially improve inter-platform concordance (Walsh et al., 2015). In this study, we used the Empirical Bayes (EB) method, known as Combat, to merge expression profiles from 10 mouse gonad studies. The unified gene expression profile contained 226 samples and 10,052 genes and was divided into a female subset, which contained 112 samples, and a male subset, which

contained 114 samples. All 3 datasets were subjected to ARACNE for gene regulatory network reconstruction.

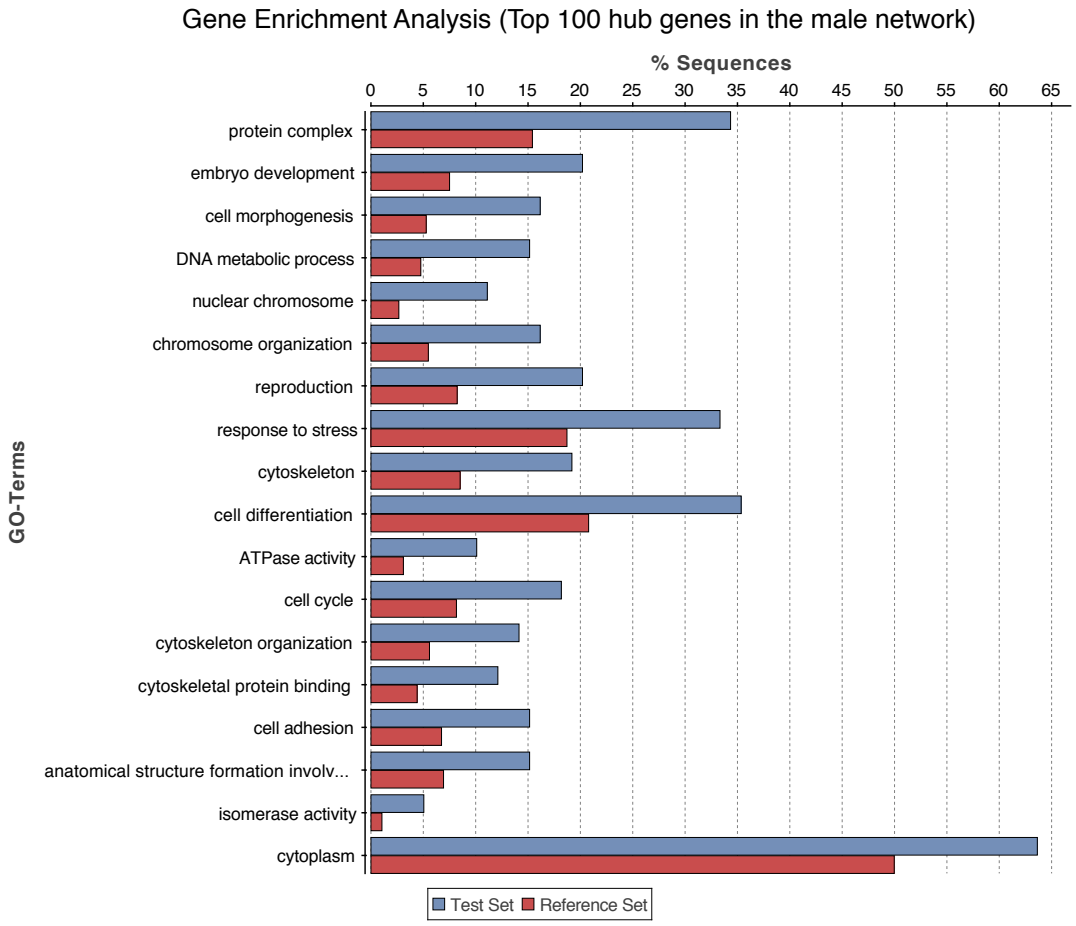
ARACNE uses a significance threshold for the DPI value to eliminate interactions that are likely to be indirect (Margolin, Wang, Lim, Kustagi, Nemenman, & Califano, 2006b). We selected 39 experimentally validated gene-gene interactions (Table 7) and examined the recovery rate of these interactions in the networks inferred by ARACNE with different p-value and DPI settings to balance false positive and false negative errors. We examined the recovery rate of known interactions (interaction detected = 1 or interaction not detected = 0) as a function of p-value and DPI for each dataset (5 p-values x 6 DPIs = 30 total combinations) (Figure 23). Recovery rates were calculated and least square means plotted as a function of p-value and DPI (Figure 23). As expected, recovery rate for known interactions increased as the stringency of the test parameters was lowered (i.e., increasing p-value and increasing DPI value) (Figure 23). For both male and female networks, the recovery rate stopped increasing when DPI hit 0.4 for all p-values. The recovery rate was slightly higher in male networks than in female networks for each p-value and DPI combination. In the combined dataset, the recovery rate leveled off at DPI = 0.6. The recovery rate in the combined dataset was higher at a given p-value, which was probably due to larger sample size. We decided to use the p-value and DPI combination that recovered more than 70% of known gene interactions. For male and female subsets, the threshold was set to $p \leq 1e\text{-}4$ (Figure 23A) and DPI = 0.4 (Figure 23B) and for the combined dataset it was set to $p \leq 1e\text{-}6$ and DPI = 0.6 (Figure 23C).

Gene Ontology Term Enrichment Analysis of The Top 100 Gonad Development-Related Hub Genes

We carried out a functional enrichment analysis (Fisher's Exact Test) for the top 100 most highly connected genes (hub genes) in both female and male networks. Enriched GO terms in both sexes included protein complex, chromosome organization, response to stress, nuclear chromosome, reproduction, DNA metabolic process, ATPase activity. A large number of GO terms for hub genes differed between the sexes, including anatomical structure formation involvement, biosynthetic process, lipid metabolic process, nucleoplasm, cytoplasm, mitotic nuclear division, chromosome segregation, protein complex assembly, nucleotidyltransferase activity, ligase activity, cell cycle, embryo development, helicase activity, cytoskeleton, plasma membrane, nuclear envelope, extracellular region, DNA binding, microtubule organizing center, cell division, isomerase activity, cytoskeletal protein binding, ion binding, cell differentiation, cell adhesion, cytoskeleton organization, cell morphogenesis (Figure 25A and 25B).


Gene Enrichment Analysis (Top 100 hub genes in the female network)
% Sequences
0
5
10
15
20
25
30
35
40
45
50
GO-Terms
DNA metabolic process
mitotic nuclear division
chromosome segregation
chromosome organization
cell division
nuclear chromosome
DNA binding
protein complex
helicase activity
ATPase activity
nucleoplasm
biosynthetic process
reproduction
response to stress
nucleotidyltransferase activity
microtubule organizing center
plasma membrane
extracellular region
ligase activity
ion binding
lipid metabolic process
nuclear envelope
protein complex assembly
Test Set
Reference Set

A



B

Figure 25. GO term enrichment analyses of the top 100 gonad development-related hub genes. Figure A shows the enriched GO terms for the top 100 hub genes in the networks reconstructed from XX gonad samples and Figure B shows the enriched GO terms for the top 100 hub genes in the networks reconstructed from XY gonad samples. The top 100 hub genes were used as the test set (blue) and all genes in the network were used as the reference set (red).

Once the bipotential gonad forms, its fate is determined by complex signaling networks that regulate a series of biological processes such as cell proliferation, cell differentiation, cell migration, apoptosis, and morphological changes. The GO term distribution of the 100 hub genes in the XX gonad and XY gonad reflected these biological events. The most abundant GO terms for the hub genes in both networks were

those involved in cell division and DNA replication, suggesting the roles of the hub genes in basic development of gonad. Differences in enrichment of GO terms in the female and male networks reflect the different biological processes of ovary and testis determination and differentiation. For example, enriched GO terms cytoskeleton, cell adhesion and anatomical structure formation for the top hub genes in the XY gonad are related to key sex specific events in testis development, including cell migration, testis cord formation and testis-specific vasculature development (Brennan & Capel, 2004).

We also noticed that *Emx2* and *Runx1,* two genes known to be involved in sexual development, were among the top 100 hub genes in the reconstructed female network. *Emx2* is a homolog of the Drosophila head gap gene *empty spiracles (ems)*. *Emx2* mutant mice display defects in the kidneys, ureters, gonads and genital tracts (Miyamoto, Yoshida, Kuratani, Matsuo, & Aizawa, 1997). *Runx1* is a transcription factor involved in cell proliferation and differentiation. Its expression levels are similar in male and female genital ridges initially but are restricted to the ovaries and mesonephric ducts in later stages (Nef et al., 2005). Although these two genes are widely recognized as markers for gonadal development, the function and regulatory mechanisms of both genes are poorly understood. This study provided the first insight into the molecular functions of these genes.

Comparison of ARACNE Inferred and Experimentally Validated Target Genes of Sry And Sox9

Sex-determining region Y (*Sry*) gene is the most important gene for testis determination in mammals because it initiates differentiation of Sertoli cells. These cells, in turn, are involved in morphogenesis of seminiferous tubules (McLaren, 1991). Studies have shown that *Sry* synergizes with steroidogenic factor 1 (*Sf1*) to regulate the

expression of *SRY*-box 9 (*Sox9*), a gene critical for testis determination, by directly binding to the core domain of the testis enhancer of *Sox9* (*TESCO*) (Sekido & Lovell-Badge, 2008). It has been suggested that the primary function of *Sry* in testis determination is to activate *Sox9* expression (Sekido & Lovell-Badge, 2009). However, recent studies have demonstrated that in addition to *Sox9, Sry* activates a large number of genes important to sex determination (Yunmin Li, Zheng, & Lau, 2014). With ChIP-chip, a recent study identified 3,083 direct *Sry* target genes and 1,903 direct *Sox9* target genes in developing mouse gonads (Yunmin Li et al., 2014). A total of 707 common target genes were found to be regulated by both transcriptional factors (Figure 26) (Yunmin Li et al., 2014).

In this study, interaction partners for *Sry* and *Sox9* were inferred computationally with ARACNE and compared to those derived from the ChIP-chip experiment mentioned above (Yunmin Li et al., 2014). In total, we inferred 1,262 interaction partners for *Sry* and 3,981 interaction partners for *Sox9*. The inferred common interaction partners for *Sry* and *Sox9* are shown in Figure 26. The comparisons suggested ARACNE can accurately infer large numbers of transcription factor targets from gene expression profiles. However, there were also significant differences between computationally inferred gene-gene interactions and interactions derived from ChIP-chip. Such divergence is expected because ARACNE and ChIP-chip identify different types of interactions. ARACNE should detect any type of regulatory interaction, transcriptional or otherwise, producing undirected edges (e.g., upstream genes that regulate expression of *Sry* and *Sox9* as well as their downstream targets). In contrast, ChIP-chip only identifies DNA sequences that are bound by TFs. Those sequences are putative *cis* regulatory sequences that may or may

not influence transcription of adjacent genes. Li et al. (2014) aimed to identify transcriptional targets of *Sry* and *Sox9* during testis determination and differentiation via ChIP-chip. Our study aims to reveal the overall molecular interactions underlying sex determination and gonad differentiation (i.e., all types of genetic interactions, direct and indirect, upstream and downstream).

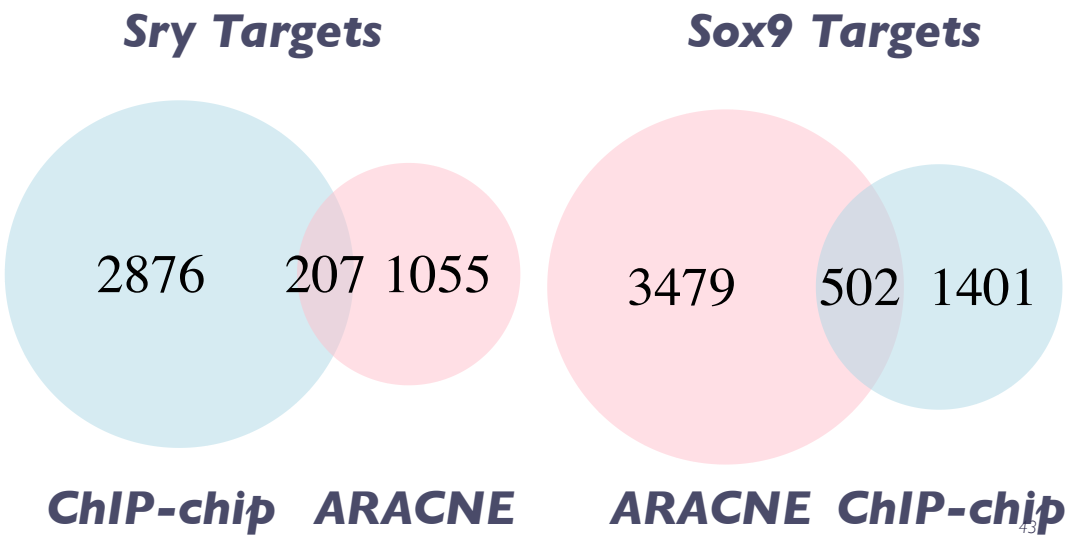


Figure 26. Comparison of ARACNE inferred and ChIP-chip derived target genes of Sry *and* Sox9. *Overlaps represent the common interaction partners derived from ARACNE and ChIP-chip for* Sry and Sox9 *(ChIP-chip data were from Li et al. 2014)*

With our optimized p-value and DPI settings, which balanced the false positive and false negative errors when inferring networks, we were able to infer gene regulatory networks in both XX gonads and XY gonads during the gonad differentiation period. This study identified many well-known interactions that the ChIP-chip study missed. For example, interactions partners for *Sox9,* such as *Amh, Vnn1, Cbln4, Etv5, Nr5a1* and *Foxl2,* were correctly identified in our reconstructed network but were missing in the

ChIP-chip study. These interactions have been experimentally validated in studies using various molecular techniques, such as KO, knockdown, over-expression, EMSA, reporter gene assay, and ChIP qPCR (Alankarage et al., 2016; Barrionuevo et al., 2006; Bradford et al., 2009; De Santa Barbara et al., 1998; Uhlenhaut et al., 2009; Wilson, Jeyasuria, Parker, & Koopman, 2005). In addition, the ARACNE inferred network indicated that *Sox9* may be directly or indirectly involved in the regulation of some *Foxl2* targets such as *Cyp17a1, Star, Kitl, Smad3, Serpine2, Ptger2* and *Ednra* (Georges et al., 2014). These findings indicate the reconstructed gene regulatory network in this study has the potential to reveal new interactions for well-studied sex-determining genes and guide future studies of novel genes or pathways for sex determination.

Master Regulator Analysis (MRA) Revealed Novel Candidate Sex-Determining Genes

The master regulator analysis compares putative TF targets (inferred by ARACNE, ChIP-chip, ChIP-Seq, or another method) to a list of differentially expressed genes to test whether TF targets are enriched in a gene signature (Carro et al., 2010). To perform MRA, we first generated an interaction network using ARACNE. Of 10,052 genes in our unified gene expression dataset, 737 were transcription factors. To identify all possible targets for these transcription factors, we reconstructed a gene regulatory network for the combined dataset (n = 226). The p-value and DPI threshold were balanced in the same way when ARACNE was applied to female and male datasets (Figure 23). ARACNE inferred a gene regulatory network with 469,357 interactions for 737 transcription factors.

We then identified differentially expressed genes between 112 female gonad samples and 114 male gonad samples. The Bonferroni adjusted *t*-tests generated 503

differentially expressed genes with a significance threshold of 0.01, including 213 genes expressed at a higher level in female gonads and 290 genes expressed at a higher level in male gonads. We performed a functional enrichment analysis with BLAST2GO on these DEGs in an attempt to interpret their biological functions during gonadal development. The most significantly enriched GO terms were related to steroidogenesis (Figure 27), indicating the important roles of steroid hormones in sexual differentiation.

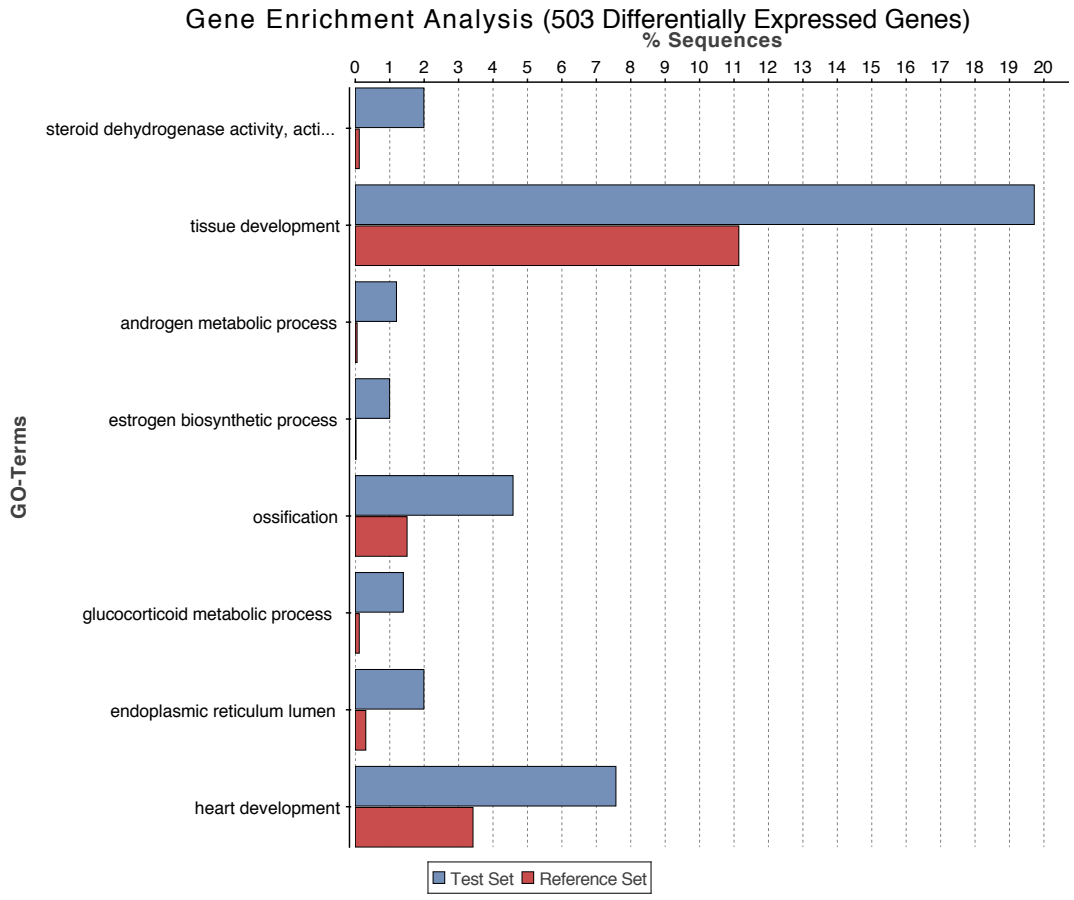


Figure 27. GO term enrichment analyses of the differentially expressed genes between sexes (Differentially expressed genes were identified from the normalized microarray data with Student's t-test with Bonferroni correction, $p < 0.01$).

With all three inputs prepared, we then performed MRA to identify the candidate master regulators that may control the trajectory of gonad differentiation. A total of 110 candidate master regulators were identified with 54 master regulators significantly up-regulated in XX gonads and 56 up-regulated in XY gonads (Table 8). A large proportion of the 110 inferred master regulators have been empirically demonstrated to play crucial roles in sex determination and gonad differentiation. These genes included the *Sox* family (*Sox8* and *Sox9*), *Dmrt1*, *Etv5* and *Lmo4* for male gonadal development and *Irx3, Msx1, Runx1, Zfp277* and *Foxl2* for female gonadal development (Alankarage et al., 2016; Garcia-Ortiz et al., 2009; Jorgensen & Gao, 2005; Koopman, 2005; Menke & Page, 2002; Minkina et al., 2014; Munger et al., 2013). The prediction of these experimentally validated master regulators indicated the MRA we performed in this study was reliable for identifying critical TFs that control the gene regulatory network underlying sex determination.

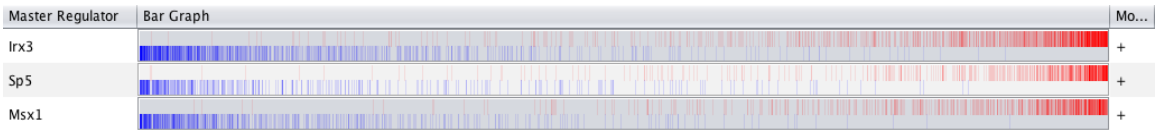
Table 8: The Results of Master Regulator Analysis

Master Regulator	FET P-Value	Genes in regulon	Genes in intersection set	Mode
Uty	2.42E-273	1227	390	-
Irx3	5.15E-190	1958	394	+
Sp5	1.07E-149	951	270	+
Sox13	1.43E-148	767	245	-
Msx1	4.11E-145	1249	295	+
Sox8	2.17E-138	1261	290	-
Spry4	2.98E-137	1552	315	-
Mixl1	5.47E-135	914	251	+
Nfe2	8.75E-124	862	235	-
Taf7l	4.12E-109	1743	305	+
Spic	4.91E-104	1430	272	-
Irx5	1.22E-103	1298	261	+
Scmh1	2.03E-103	2114	348	+
Polr2g	5.41E-98	2509	346	+
Scx	3.13E-95	715	191	-
Foxm1	2.68E-88	1699	278	-

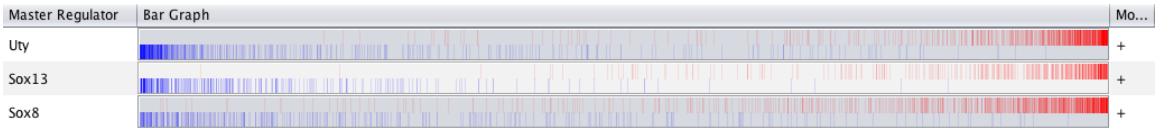
Master Regulator	FET P-Value	Genes in regulon	Genes in intersection set	Mode
Mafk	8.56E-85	1463	252	-
Bcl11b	4.20E-82	459	153	-
Mybl1	1.76E-78	2938	347	-
Pbx3	9.78E-73	2007	282	+
Foxa3	2.61E-69	930	187	-
Nkx3-1	4.16E-69	1810	262	-
Pax8	7.92E-68	980	193	+
Etv5	4.52E-62	3179	343	-
Mllt3	1.55E-61	3122	347	+
Msx2	3.24E-59	2589	307	+
Lmo4	5.71E-58	2523	308	-
Smad7	3.52E-52	1387	205	-
Pura	5.56E-52	631	135	-
Pdlim1	9.22E-51	1558	218	-
Srebf1	5.70E-50	734	143	-
Sall3	2.66E-49	869	155	+
Sin3b	4.70E-47	2830	301	-
Hoxb9	2.96E-46	1284	193	-
Npas3	1.11E-45	699	139	+
Vgll2	7.61E-44	809	143	+
Zfp553	1.81E-42	1741	216	+
Mbtd1	4.13E-40	582	116	+
Creb3l4	5.97E-40	1219	181	+
Runx1	3.18E-38	3504	350	+
Lmo1	1.22E-37	1017	155	+
Zfp532	1.55E-37	776	130	-
Zfp292	6.96E-37	1810	218	+
Zfp277	8.80E-36	2997	302	+
Sbds	1.07E-35	1155	162	-
Lmo3	1.35E-35	687	122	+
Sox9	5.28E-35	3510	357	-
Morf4l2	1.13E-33	1059	149	-
Hmgb3	3.21E-32	1851	211	-
Gne	1.28E-29	3064	309	-
E2f7	7.43E-28	2383	253	-
Foxd1	2.27E-27	1368	182	+
Epas1	5.71E-27	2448	248	-
Foxq1	1.12E-26	305	71	-
Tcea3	1.32E-26	1737	188	+
Crip3	5.03E-26	2016	242	+
Etv6	1.10E-25	1387	164	+
Irf5	1.31E-24	1141	143	+
Hivep3	5.56E-23	856	113	-
Cebpa	4.61E-21	2217	221	-
Gata1	7.28E-21	221	51	+
Bcl6	1.71E-20	1778	182	+
Rel	2.10E-20	1553	172	-
Mkl1	7.81E-20	2268	219	+
Foxc1	5.68E-19	1650	179	-

Master Regulator	FET P-Value	Genes in regulon	Genes in intersection set	Mode
Fhl1	8.33E-19	923	141	-
E2f1	1.43E-18	2273	225	-
Emx2	1.59E-17	3654	338	+
Rreb1	4.20E-17	1168	138	-
Yeats2	1.73E-16	2591	258	+
Cdca4	5.47E-16	1628	157	-
Pbx2	5.70E-16	790	97	+
Phf14	9.61E-16	2751	240	+
Pdlim3	2.46E-15	1439	147	-
Dmrt1	4.69E-15	2087	190	-
Bnc1	8.98E-15	2051	199	+
Foxl2	1.14E-14	801	94	+
Nfat5	9.30E-14	3839	356	+
Gata3	7.42E-13	655	78	+
Sox6	2.00E-12	3342	253	-
Med10	3.06E-12	1815	163	+
Spry1	5.59E-12	1412	134	-
Dmp1	7.87E-11	345	48	-
Ankra2	1.06E-10	313	45	+
Sox18	1.25E-10	2036	177	-
Barx1	1.65E-10	843	85	-
Rai1	2.34E-10	1055	104	-
Pdlim2	2.89E-10	1227	119	-
Foxp1	3.35E-10	2725	222	+
Akna	3.50E-10	870	87	-
Zfp93	9.67E-10	687	74	+
Foxj3	1.59E-09	162	31	+
Meox2	2.56E-09	1595	137	-
Dbp	7.61E-09	1250	114	+
Sertad1	1.12E-08	1243	114	+
Mxd3	1.30E-08	401	48	-
Tcf12	2.12E-08	3599	300	+
Etv4	3.03E-08	2524	201	-
Zfp462	5.79E-08	1444	125	+
Stat2	6.15E-08	529	56	+
Ahr	7.49E-08	1973	155	-
Taf7	4.72E-07	464	49	+
Spry2	8.00E-07	733	68	-
Hlf	9.14E-07	433	47	+
Zbtb17	1.07E-06	228	30	+
Maff	1.65E-06	2429	177	-
Hoxa3	1.70E-06	1726	146	+
Bud31	1.70E-06	556	55	-
Dach1	2.06E-06	2783	218	+
Foxo1	9.59E-06	2298	185	-

We further investigated the networks of the top 3 master regulators in the developing XX and XY gonads. The top master regulators in male gonads included *Uty* (FET p-value = 2.42E-273), *Sox13* (FET p-value = 1.43E-148) and *Sox8* (FET p-value = 2.17E-138). The top female counterparts were *Irx3* (FET p-value = 5.15E-190), *Sp5* (FET p-value = 1.07E-149) and *Msx1* (FET p-value = 4.11E-145). The ranked differential expression results for the top 3 master regulators in both female and male samples were shown in Figure 28A and 28B. Comparison of the target genes revealed substantial overlap among the top 3 master regulators within each sex (Figure 29), which suggesting master regulators may function cooperatively to regulate large sets of genes involved in determining gonad fate.



A



B

Figure 28. The ranked differential expression results for the top 3 master regulators in XX gonad samples (A) and XY gonad samples (B). The vertical bars represent the targets belonging to each TF's regulon. Bar positions on horizontal axis represent the expression level of each target. The expression levels are higher on the right of the horizontal axis than those on the left. The color of each bar indicates the sign of the Spearman's Correlation between the expression profile of the TF and its targets. Red means the target and the master regulator are positively correlated and blue means the target and the master regulator are negatively correlated. The color intensity of each bar is scaled to represent the number of overlapping bars at any given point in the graph.

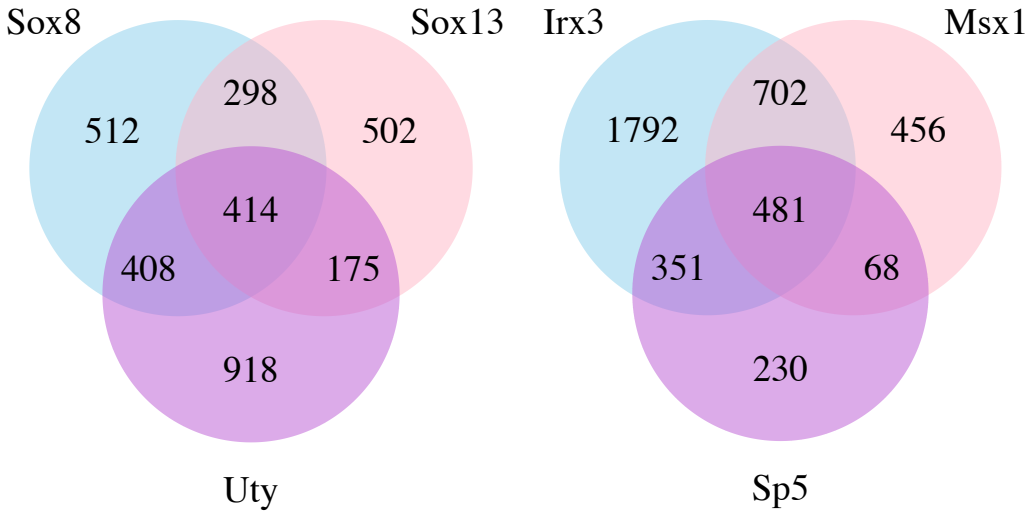


Figure 29. Comparison of the regulons of the top 3 master regulators in XX gonad samples (A) and XY gonad samples (B). Circles represent the ARACNE predicted regulons for the master regulators. Numbers in circles are gene numbers contained in that region.

Among our top 3 MRs for testis determination, *Sox* genes may be the best-known sex-determining genes. All 20 members of the *Sox* family play important roles in embryonic development (Bowles, Schepers, & Koopman, 2000). *Sox9,* a crucial downstream target of *Sry*, has been proved to be sufficient for male sex determination (Sekido & Lovell-Badge, 2009). *Sox8,* the closest paralog of *Sox9,* resembles *Sox9* in biochemical properties and expression patterns in the developing gonad (Schepers et al., 2003). Its expression during sex determination is directly regulated by *Sox9* (Chaboissier et al., 2004). Although *Sox8* does not play a decisive role in testis determination and differentiation, it functionally complements *Sox9* function in testis differentiation (Chaboissier et al., 2004). The MRA performed in this study accurately inferred the role of *Sox8* and its direct regulatory relationship with *Sox9* in gonad differentiation. Recent studies have shown *Sox13* is expressed in postnatal testis and may regulate

steroidogenesis and spermatogenesis (Daigle, Roumaud, & Martin, 2015). However, little is known about the function of *Sox13* in sex determination or gonad differentiation.

Uty (Ubiquitously Transcribed Tetratricopeptide Repeat Containing, Y-Linked), also known as *Kdm6c*, is located on Y chromosome and may contribute to gender differences in brain function (Vawter et al., 2004). One of the *Uty* related pathways is chromatin organization and its related GO annotation includes histone H3-K27 specific demethylase activity (Belinky et al., 2015; Gene Ontology Consortium, 2015). Interestingly, our studies in a species with temperature-dependent sex determination revealed its paralog *Kdm6b* is differentially expressed between incipient testes and ovaries (Chapter 3). *Uty* may participate in mammalian sex determination by epigenetically regulating expression of other genes.

Irx3, one of our inferred top 3 MRs for ovary determination and differentiation, belongs to the Iroquois homeobox gene family. Its expression was found to be restricted to somatic cells of XX gonads during gonadal development, suggesting its potential role in ovary determination (Jorgensen & Gao, 2005). Similarly, *Msx1* is highly expressed in XX gonads during sex determination and is repressed in XY gonads (Munger et al., 2013). *Sp5* is a transcription factor that shows dynamic expression pattern during mouse embryogenesis in different tissues (Treichel, Becker, & Gruss, 2001). One of the pathways related to this gene is *Wnt*-mediated beta-catenin signaling and target gene transcription (Belinky et al., 2015). Although the role of *Sp5* in sex determination is still largely unknown, its link to *Wnt* signaling suggests it may be involved in regulating ovary development.

We have identified candidate MRs in both developing testes and ovaries. Four out of six MRs (*Sox8, Sox13, Irx3 and Msx1*) listed above have been shown to play important roles in sexual development. Although there is no independent evidence showing *Uty* and *Sp5* participate in sex determination directly, our GO and pathway analyses suggested the potential role of these genes in gonad development. In addition, well studied sex-determining genes such as *Dmrt1*, *Lmo4*, *Emx2*, *Sox8*, *Sox9* and *Foxl2* were also identified as MRs in this study, indicating the reliability of our analyses (Tanaka & Nishinakamura, 2014).

Conclusion

In this study, we inferred gene regulatory networks in the developing mouse gonad by merging multiple carefully selected microarray datasets, which allows us to overcome the limitation of small sample sizes in individual studies and increase statistical power. For network reconstruction, we used ARACNE, an algorithm that has been shown to outperform other network-reconstructing algorithms in both sensitivity and precision (Basso et al., 2005; Margolin, Nemenman, et al., 2006a). We examined the recovery rate of 39 previously validated gene-gene interactions in three networks inferred with different p-value and DPI settings (combined, female and male data sets). We selected values to maximize sensitivity (recovery of true positives), while maintaining a high specificity.

Our first contribution to the understanding of sex determination was identification of hub genes in gonadal development by investigating the nodes with the highest degree in the network. Networks in both prokaryotic and eukaryotic exhibit a hierarchical scale-free nature, characterized by vertices with a degree that greatly exceeds the average,

which are known as hub genes (Albert, 2005). The stability of such networks relies on these highly-connected hub genes. Because of the importance of hub genes, one can hypothesize they are subject to severe selective and evolutionary constraints (Albert, 2005). In this study, we investigated the functions of the top 100 hub genes in the reconstructed networks for the developing mouse gonads. The functional enrichment analyses on the top 100 hub genes indicated the most enriched functions for both XX and XY gonads were related to cell cycle and DNA replication. This result suggests that processes involved in cell proliferation are important in directing gonad development. Our hub gene analyses also reflected major differences in the molecular mechanisms governing development of ovary and testis.

Our second contribution is the identification of new interaction partners for genes known to play critical roles in sex determination and differentiation. We compared our computationally inferred interaction partners for two well studied sex-determining genes to results from an independent ChIP-chip study. ChIP-chip studies focus on identifying the direct targets of TFs, while ARACNE identifies all potential interaction partners for these genes. Although there should be some overlap, we do not expect perfect concordance between lists generated with these methods. Our computationally inferred interactions for known sex-determining genes contained known interactions, which suggests that the method is accurate. The novel connections in our gene regulatory network will provide direction for future studies of known sex-determining genes and new genes to study for their role in sex determination.

Lastly, we identified 110 candidate master regulators, which may play key roles in gonad fate determination. The functional enrichment analyses and pathway analyses of

the top 3 MRs suggested they might interact with each other to regulate many of the same target genes. Although further experiments are needed to validate these results, our computational study may help guide the direction of future studies of sex determination.

REFERENCES

Adams, I. R., & McLaren, A. (2002). Sexually dimorphic development of mouse primordial germ cells: switching from oogenesis to spermatogenesis. Development (Cambridge, England), 129(5), 1155–1164.

Alankarage, D., Lavery, R., Svingen, T., Kelly, S., Ludbrook, L., Bagheri-Fam, S., et al. (2016). SOX9 regulates expression of the male fertility gene Ets variant factor 5 (ETV5) during mammalian sex development. The International Journal of Biochemistry \& Cell Biology, 79, 41–51. http://doi.org/10.1016/j.biocel.2016.08.005

Albert, R. (2005). Scale-free networks in cell biology. Journal of Cell Science, 118(Pt 21), 4947–4957. http://doi.org/10.1242/jcs.02714

Albrecht, K. H., & Eicher, E. M. (2001). Evidence That Sry Is Expressed in Pre-Sertoli Cells and Sertoli and Granulosa Cells Have a Common Precursor. Developmental Biology, 240(1), 92–107. http://doi.org/10.1006/dbio.2001.0438

Almaas, E. (2007). Biological impacts and context of network theory. The Journal of Experimental Biology, 210(Pt 9), 1548–1558. http://doi.org/10.1242/jeb.003731

Baba, T., Otake, H., Sato, T., Miyabayashi, K., Shishido, Y., Wang, C.-Y., Shima, Y., Kimura, H., Yagi, M., Ishihara, Y., et al. (2014). Glycolytic genes are targets of the nuclear receptor Ad4BP/SF-1. Nat Commun 5, 3634.

Barrionuevo, F., Bagheri-Fam, S., Klattig, J., Kist, R., Taketo, M. M., Englert, C., & Scherer, G. (2006). Homozygous inactivation of Sox9 causes complete XY sex reversal in mice. Biology of Reproduction, 74(1), 195–201. http://doi.org/10.1095/biolreprod.105.045930

Barrionuevo, F., Georg, I., Scherthan, H., Lécureuil, C., Guillou, F., Wegner, M., and Scherer, G. (2009). Testis cord differentiation after the sex determination stage is independent of Sox9 but fails in the combined absence of Sox9 and Sox8. Developmental Biology 327, 301–312.

Bashamboo, A., Ferraz-de-Souza, B., Lourenço, D., Lin, L., Sebire, N.J., Montjean, D., Bignon-Topalovic, J., Mandelbaum, J., Siffroi, J.-P., Christin-Maitre, S., et al. (2010). Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor 1. Am. J. Hum. Genet. 87, 505–512.

Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., & Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nature Genetics, 37(4), 382–390. http://doi.org/10.1038/ng1532

Belinky, F., Nativ, N., Stelzer, G., Zimmerman, S., Iny Stein, T., Safran, M., & Lancet, D. (2015). PathCards: multi-source consolidation of human biological pathways. Database : the Journal of Biological Databases and Curation, 2015(0), bav006–bav006. http://doi.org/10.1093/database/bav006

Bernard, P., Gervais, S.J., Allen, J., Campomizzi, S., and Klein, O. (2012). Integrating sexual objectification with object versus person recognition: the sexualized-body-inversion hypothesis. Psychol Sci 23, 469–471.

Beverdam, A., & Koopman, P. (2006). Expression profiling of purified mouse gonadal somatic cells during the critical time window of sex determination reveals novel candidate genes for human sexual dysgenesis syndromes. Human Molecular Genetics, 15(3), 417–431. http://doi.org/10.1093/hmg/ddi463

Boulanger, L., Pannetier, M., Gall, L., Allais-Bonnet, A., Elzaiat, M., Le Bourhis, D., et al. (2014). *FOXL2* Is a Female Sex-Determining Gene in the Goat. Current Biology, 24(4), 404–408. http://doi.org/10.1016/j.cub.2013.12.039

Bowles, J., Schepers, G., & Koopman, P. (2000). Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Developmental Biology, 227(2), 239–255. http://doi.org/10.1006/dbio.2000.9883

Boyer, A., Yeh, J.R., Zhang, X., Paquet, M., Gaudin, A., Nagano, M.C., and Boerboom, D. (2012). CTNNB1 signaling in sertoli cells downregulates spermatogonial stem cell activity via WNT4. Plos One 7, e29764.

Bradford, S. T., Hiramatsu, R., Maddugoda, M. P., Bernard, P., Chaboissier, M.-C., Sinclair, A., et al. (2009). The cerebellin 4 precursor gene is a direct target of SRY and SOX9 in mice. Biology of Reproduction, 80(6), 1178–1188. http://doi.org/10.1095/biolreprod.108.071480

Brennan, J., & Capel, B. (2004). One tissue, two fates: molecular genetic events that underlie testis versus ovary development. Nature Reviews. Genetics, 5(7), 509–521. http://doi.org/10.1038/nrg1381

Bullejos, M., & Koopman, P. (2001). Spatially dynamic expression of Sry in mouse genital ridges. Developmental Dynamics, 221(2), 201–205. http://doi.org/10.1002/dvdy.1134

Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., et al. (2009). AmiGO: online access to ontology and annotation data. Bioinformatics, 25(2), 288–289. http://doi.org/10.1093/bioinformatics/btn615

Carro, M. S., Lim, W. K., Alvarez, M. J., Bollo, R. J., Zhao, X., Snyder, E. Y., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature, 463(7279), 318–325. http://doi.org/10.1038/nature08712

Ceci, M., Pio, G., Kuzmanovski, V., & Džeroski, S. (2015). Semi-Supervised Multi-View Learning for Gene Network Reconstruction. Plos One, 10(12), e0144031. http://doi.org/10.1371/journal.pone.0144031

Chaboissier, M.-C., Kobayashi, A., Vidal, V. I. P., Lützkendorf, S., van de Kant, H. J. G., Wegner, M., et al. (2004). Functional analysis of Sox8 and Sox9 during sex determination in the mouse. Development (Cambridge, England), 131(9), 1891–1901. http://doi.org/10.1242/dev.01087

Chang, H., Gao, F., Guillou, F., Taketo, M.M., Huff, V., and Behringer, R.R. (2008). Wt1 negatively regulates beta-catenin signaling during testis development. Development (Cambridge, England) 135, 1875–1885.

Chang, C., Wang, J., Zhao, C., Fostel, J., Tong, W., Bushel, P. R., et al. (2011). Maximizing biomarker discovery by minimizing gene signatures. BMC Genomics, 12 Suppl 5(Suppl 5), S6. http://doi.org/10.1186/1471-2164-12-S5-S6

Chassot, A.-A., Ranc, F., Gregoire, E.P., Roepers-Gajadien, H.L., Taketo, M.M., Camerino, G., de Rooij, D.G., Schedl, A., and Chaboissier, M.-C. (2008). Activation of beta-catenin signaling by Rspo1 controls differentiation of the mammalian ovary. Human Molecular Genetics 17, 1264–1277.

Cho, D.-Y., Kim, Y.-A., & Przytycka, T. M. (2012). Chapter 5: Network Biology Approach to Complex Diseases. PLoS Computational Biology, 8(12), e1002820. http://doi.org/10.1371/journal.pcbi.1002820

Cosentino, C., Grieco, D., and Costanzo, V. (2011). ATM activates the pentose phosphate pathway promoting anti-oxidant defence and DNA repair. The EMBO Journal 30, 546–555.

Couse, J.F., Hewitt, S.C., Bunch, D.O., Sar, M., Walker, V.R., Davis, B.J., and Korach, K.S. (1999). Postnatal sex reversal of the ovaries in mice lacking estrogen receptors alpha and beta. Science (New York, N.Y.) 286, 2328–2331.

Daigle, M., Roumaud, P., & Martin, L. J. (2015). Expressions of Sox9, Sox5, and Sox13 transcription factors in mice testis during postnatal development. Molecular and Cellular Biochemistry, 407(1-2), 209–221. http://doi.org/10.1007/s11010-015-2470-7

Davis, S., & Meltzer, P. S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics, 23(14), 1846–1847. http://doi.org/10.1093/bioinformatics/btm254

De Santa Barbara, P., Bonneaud, N., Boizet, B., Desclozeaux, M., Moniot, B., Sudbeck, P., et al. (1998). Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Müllerian hormone gene. Molecular and Cellular Biology, 18(11), 6653–6665.

Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma, Shedden, K., Taylor, J. M. G., Enkemann, S. A., Tsao, M.-S., Yeatman, T. J., et al. (2008). Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nature Medicine, 14(8), 822–827. http://doi.org/10.1038/nm.1790

Dokshin, G.A., Baltus, A.E., Eppig, J.J., and Page, D.C. (2013). Oocyte differentiation is genetically dissociable from meiosis in mice. Nature Genetics 45, 877–883.

Floratos, A., Smith, K., Ji, Z., Watkinson, J., & Califano, A. (2010). geWorkbench: an open source platform for integrative genomics. Bioinformatics, 26(14), 1779–1780. http://doi.org/10.1093/bioinformatics/btq282

Garcia-Ortiz, J., Pelosi, E., Omari, S., Nedorezov, T., Piao, Y., Karmazin, J., et al. (2009). *Foxl2* functions in sex determination and histogenesis throughout mouse ovary development. BMC Developmental Biology, 9(1), 36. http://doi.org/10.1186/1471-213X-9-36

Gene Ontology Consortium. (2015). Gene Ontology Consortium: going forward. Nucleic Acids Research, 43(Database issue), D1049–56. http://doi.org/10.1093/nar/gku1179

Georges, A., Auguste, A., Bessière, L., Vanet, A., Todeschini, A.-L., & Veitia, R. A. (2014). *FOXL2*: a central transcription factor of the ovary. Journal of Molecular Endocrinology, 52(1), R17–33. http://doi.org/10.1530/JME-13-0159

Glister, C., Richards, S.L., and Knight, P.G. (2005). Bone Morphogenetic Proteins (BMP) -4, -6, and -7 Potently Suppress Basal and Luteinizing Hormone-Induced Androgen Production by Bovine Theca Interna Cells in Primary Culture: Could Ovarian Hyperandrogenic Dysfunction Be Caused by a Defect in Thecal BMP Signaling? Endocrinology 146, 1883–1892.

Godschalk, R.W.L., Vanhees, K., Maas, L., Drittij, M.-J., Pachen, D., van Doorn-Khosrovani, S.V.W., van Schooten, F.J., and Haenen, G.R.M.M. (2016). Does Ataxia Telangiectasia Mutated (ATM) protect testicular and germ cell DNA integrity by regulating the redox status? Reprod. Toxicol. 63, 169–173.

Greenfield, A. (2015). Understanding sex determination in the mouse: genetics, epigenetics and the story of mutual antagonisms. Journal of Genetics, 94(4), 585–590. http://doi.org/10.1007/s12041-015-0565-2

Hamid, J. S., Hu, P., Roslin, N. M., Ling, V., Greenwood, C. M. T., & Beyene, J. (2009). Data Integration in Genetics and Genomics: Methods and Challenges. Human Genomics and Proteomics, 2009(2), 1–13. http://doi.org/10.4061/2009/869093

Hammes, A., Guo, J.K., Lutsch, G., Leheste, J.R., Landrock, D., Ziegler, U., Gubler, M.C., and Schedl, A. (2001). Two splice variants of the Wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. Cell 106, 319–329.

Hausser, J., & Strimmer, K. (2009). Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks. Journal of Machine Learning Research, 10(Jul), 1469–1484.

Hiramatsu, R., Matoba, S., Kanai-Azuma, M., Tsunekawa, N., Katoh-Fukui, Y., Kurohmaru, M., et al. (2009). A critical time window of Sry action in gonadal sex determination in mice. Development (Cambridge, England), 136(1), 129–138. http://doi.org/10.1242/dev.029587

Hu, P., Greenwood, C. M. T., & Beyene, J. (2005). Integrative analysis of multiple gene expression profiles with quality-adjusted effect size models. BMC Bioinformatics, 6(1), 128. http://doi.org/10.1186/1471-2105-6-128

Jameson, S. A., Lin, Y.-T., & Capel, B. (2012a). Testis development requires the repression of Wnt4 by Fgf signaling. Developmental Biology, 370(1), 24–32. http://doi.org/10.1016/j.ydbio.2012.06.009

Jameson, S. A., Natarajan, A., Cool, J., DeFalco, T., Maatouk, D. M., Mork, L., et al. (2012b). Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad. PLOS Genet, 8(3), e1002575. http://doi.org/10.1371/journal.pgen.1002575

Jho, E.-H., Zhang, T., Domon, C., Joo, C.-K., Freund, J.-N., and Costantini, F. (2002). Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. Mol. Cell. Biol. 22, 1172–1183.

Jiang, D., Tang, C., & Zhang, A. (2004). Cluster Analysis for Gene Expression Data: A Survey. IEEE Trans. on Knowl. and Data Eng., 16(11), 1370–1386. http://doi.org/10.1109/TKDE.2004.68

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics (Oxford, England), 8(1), 118–127. http://doi.org/10.1093/biostatistics/kxj037

Jordan, B.K., Shen, J.H.-C., Olaso, R., Ingraham, H.A., and Vilain, E. (2003). Wnt4 overexpression disrupts normal testicular vasculature and inhibits testosterone synthesis by repressing steroidogenic factor 1/beta-catenin synergy. Proceedings of the National Academy of Sciences of the United States of America 100, 10866–10871.

Jorgensen, J. S., & Gao, L. (2005). Irx3 is differentially up-regulated in female gonads during sex determination. Gene Expression Patterns: GEP, 5(6), 756–762. http://doi.org/10.1016/j.modgep.2005.04.011

Kashimada, K., Svingen, T., Feng, C.-W., Pelosi, E., Bagheri-Fam, S., Harley, V.R., Schlessinger, D., Bowles, J., and Koopman, P. (2011). Antagonistic regulation of Cyp26b1 by transcription factors SOX9/SF1 and *FOXL2* during gonadal development in mice. Faseb J. 25, 3561–3569.

Kim, Y., Kobayashi, A., Sekido, R., DiNapoli, L., Brennan, J., Chaboissier, M.-C., et al. (2006). Fgf9 and Wnt4 act as antagonistic signals to regulate mammalian sex determination. PLoS Biology, 4(6), e187. http://doi.org/10.1371/journal.pbio.0040187

Klattig, J., Sierig, R., Kruspe, D., Besenbeck, B., and Englert, C. (2007). Wilms' tumor protein Wt1 is an activator of the anti-Müllerian hormone receptor gene Amhr2. Mol. Cell. Biol. 27, 4355–4364.

Knudson, C.M., Tung, K.S., Tourtellotte, W.G., Brown, G.A., and Korsmeyer, S.J. (1995). Bax-deficient mice with lymphoid hyperplasia and male germ cell death. Science (New York, N.Y.) 270, 96–99.

Koopman, P. (2005). Sex determination: a tale of two Sox genes. Trends in Genetics : TIG, 21(7), 367–370. http://doi.org/10.1016/j.tig.2005.05.006

Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates sex-specific timing of meiotic initiation in mice. Proceedings of the National Academy of Sciences of the United States of America 103, 2474–2479.

Kraskov, A., Stögbauer, H., Andrzejak, R. G., & Grassberger, P. (2003). Hierarchical Clustering Based on Mutual Information. arXiv:Q-Bio/0311039.

Krentz, A.D., Murphy, M.W., Sarver, A.L., Griswold, M.D., Bardwell, V.J., and Zarkower, D. (2011). DMRT1 promotes oogenesis by transcriptional activation of Stra8 in the mammalian fetal ovary. Developmental Biology 356, 63–70.

Le Bouffant, R., Souquet, B., Duval, N., Duquenne, C., Hervé, R., Frydman, N., Robert, B., Habert, R., and Livera, G. (2011). Msx1 and Msx2 promote meiosis initiation. Development 138, 5393–5402.

Li, B., Nair, M., Mackay, D.R., Bilanchone, V., Hu, M., Fallahi, M., Song, H., Dai, Q., Cohen, P.E., and Dai, X. (2005). Ovol1 regulates meiotic pachytene progression during spermatogenesis by repressing Id2 expression. Development (Cambridge, England) 132, 1463–1473.

Li, D., Urs, A.N., Allegood, J., Leon, A., Merrill, A.H., and Sewer, M.B. (2007). Cyclic AMP-stimulated interaction between steroidogenic factor 1 and diacylglycerol kinase theta facilitates induction of CYP17. Mol. Cell. Biol. 27, 6669–6685.

Li, Yunmin, Zheng, M., & Lau, Y.-F. C. (2014). The sex-determining factors SRY and SOX9 regulate similar target genes and promote testis cord formation during testicular differentiation. Cell Reports, 8(3), 723–733. http://doi.org/10.1016/j.celrep.2014.06.055

Li, Yupeng, & Jackson, S. A. (2015). Gene Network Reconstruction by Integration of Prior Biological Knowledge. G3 (Bethesda, Md.), 5(6), 1075–1079. http://doi.org/10.1534/g3.115.018127

Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ cell-intrinsic and -extrinsic factors govern meiotic initiation in mouse embryos. Science (New York, N.Y.) 322, 1685–1687.

Liu, H., Li, P., Zhu, M., Wang, X., Lu, J., & Yu, T. (2016). Nonlinear Network Reconstruction from Gene Expression Data Using Marginal Dependencies Measured by DCOL. Plos One, 11(7), e0158247. http://doi.org/10.1371/journal.pone.0158247

Liu, C., Rodriguez, K., and Yao, H.H.-C. (2016). Mapping lineage progression of somatic progenitor cells in the mouse fetal testis. Development 143, 3700–3710.

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla-Favera, R., & Califano, A. (2006a). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics, 7 Suppl 1(Suppl 1), S7. http://doi.org/10.1186/1471-2105-7-S1-S7

Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., Nemenman, I., & Califano, A. (2006b). Reverse engineering cellular networks. Nature Protocols, 1(2), 662–671. http://doi.org/10.1038/nprot.2006.106

Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., Nemenman, I., & Califano, A. (2006c). Reverse engineering cellular networks. Nature Protocols, 1(2), 662–671. http://doi.org/10.1038/nprot.2006.106

McClelland, K.S., Bell, K., Larney, C., Harley, V.R., Sinclair, A.H., Oshlack, A., Koopman, P., and Bowles, J. (2015). Purification and Transcriptomic Analysis of Mouse Fetal Leydig Cells Reveals Candidate Genes for Specification of Gonadal Steroidogenic Cells. Biology of Reproduction 92, 145–145.

McLaren, A. (1991). Development of the mammalian gonad: the fate of the supporting cell lineage. BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology, 13(4), 151–156. http://doi.org/10.1002/bies.950130402

Menke, D. B., & Page, D. C. (2002). Sexually dimorphic gene expression in the developing mouse gonad. Gene Expression Patterns: GEP, 2(3-4), 359–367. http://doi.org/10.1016/S1567-133X(02)00022-4

Meyer, P. E., Kontos, K., Lafitte, F., & Bontempi, G. (2007). Information-theoretic inference of large transcriptional regulatory networks. EURASIP Journal on Bioinformatics \& Systems Biology, 2007(1), 79879–9. http://doi.org/10.1155/2007/79879

Minkina, A., Matson, C. K., Lindeman, R. E., Ghyselinck, N. B., Bardwell, V. J., & Zarkower, D. (2014). DMRT1 protects male gonadal cells from retinoid-dependent sexual transdifferentiation. Developmental Cell, 29(5), 511–520. http://doi.org/10.1016/j.devcel.2014.04.017

Miyamoto, N., Yoshida, M., Kuratani, S., Matsuo, I., & Aizawa, S. (1997). Defects of urogenital development in mice lacking Emx2. Development (Cambridge, England), 124(9), 1653–1664.

Moniot, B., Declosmenil, F., Barrionuevo, F., Scherer, G., Aritake, K., Malki, S., Marzi, L., Cohen-Solal, A., Georg, I., Klattig, J., et al. (2009). The PGD2 pathway, independently of FGF9, amplifies SOX9 activity in Sertoli cells during male sexual differentiation. Development (Cambridge, England) 136, 1813–1821.

Munger, S. C., Aylor, D. L., Syed, H. A., Magwene, P. M., Threadgill, D. W., & Capel, B. (2009). Elucidation of the transcription network governing mammalian sex determination by exploiting strain-specific susceptibility to sex reversal. Genes \& Development, 23(21), 2521–2536. http://doi.org/10.1101/gad.1835809

Munger, S. C., Natarajan, A., Looger, L. L., Ohler, U., & Capel, B. (2013). Fine time course expression analysis identifies cascades of activation and repression and maps a putative regulator of mammalian sex determination. PLOS Genet, 9(7), e1003630. http://doi.org/10.1371/journal.pgen.1003630

Münsterberg, A., and Lovell-Badge, R. (1991). Expression of the mouse anti-müllerian hormone gene suggests a role in both male and female sexual differentiation. Development (Cambridge, England) 113, 613–624.

Nachtigal, M.W., Hirokawa, Y., Enyeart-VanHouten, D.L., Flanagan, J.N., Hammer, G.D., and Ingraham, H.A. (1998). Wilms' tumor 1 and Dax-1 modulate the orphan nuclear receptor SF-1 in sex-specific gene expression. Cell 93, 445–454.

Naillat, F., Yan, W., Karjalainen, R., Liakhovitskaia, A., Samoylenko, A., Xu, Q., Sun, Z., Shen, B., Medvinsky, A., Quaggin, S., et al. (2015). Identification of the genes regulated by Wnt-4, a critical signal for commitment of the ovary. Experimental Cell Research 332, 163–178.

Nef, S., Schaad, O., Stallings, N. R., Cederroth, C. R., Pitetti, J.-L., Schaer, G., et al. (2005). Gene expression during sex determination reveals a robust female genetic program at the onset of ovarian development. Developmental Biology, 287(2), 361–377. http://doi.org/10.1016/j.ydbio.2005.09.008

Ottolenghi, C., Omari, S., Garcia-Ortiz, J. E., Uda, M., Crisponi, L., Forabosco, A., et al. (2005). *Foxl2* is required for commitment to ovary differentiation. Human Molecular Genetics, 14(14), 2053–2062. http://doi.org/10.1093/hmg/ddi210

Rao, M.K., Pham, J., Imam, J.S., MacLean, J.A., Murali, D., Furuta, Y., Sinha-Hikim, A.P., and Wilkinson, M.F. (2006). Tissue-specific RNAi reveals that WT1 expression in nurse cells controls germ cell survival and spermatogenesis. Genes \& Development 20, 147–152.

Rege, J., Nishimoto, H.K., Nishimoto, K., Rodgers, R.J., Auchus, R.J., and Rainey, W.E. (2015). Bone Morphogenetic Protein-4 (BMP4): A Paracrine Regulator of Human Adrenal C19 Steroid Synthesis. Endocrinology 156, 2530–2540.

Rudy, J., & Valafar, F. (2011). Empirical comparison of cross-platform normalization methods for gene expression data. BMC Bioinformatics, 12(1), 467. http://doi.org/10.1186/1471-2105-12-467

Schepers, G., Wilson, M., Wilhelm, D., & Koopman, P. (2003). SOX8 Is Expressed during Testis Differentiation in Mice and Synergizes with SF1 to Activate the Amh Promoter in Vitro. Journal of Biological Chemistry, 278(30), 28101–28108. http://doi.org/10.1074/jbc.M304067200

Sedlak, T.W., Oltvai, Z.N., Yang, E., Wang, K., Boise, L.H., Thompson, C.B., and Korsmeyer, S.J. (1995). Multiple Bcl-2 family members demonstrate selective dimerizations with Bax. Proceedings of the National Academy of Sciences of the United States of America 92, 7834–7838.

Sekido, R., & Lovell-Badge, R. (2008). Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. Nature, 456(7223), 824–824. http://doi.org/10.1038/nature07622

Sekido, R., & Lovell-Badge, R. (2009). Sex determination and SRY: down to a wink and a nudge? Trends in Genetics : TIG, 25(1), 19–29. http://doi.org/10.1016/j.tig.2008.10.008

Shabalin, A. A., Tjelmeland, H., Fan, C., Perou, C. M., & Nobel, A. B. (2008). Merging two gene-expression studies via cross-platform normalization. Bioinformatics, 24(9), 1154–1160. http://doi.org/10.1093/bioinformatics/btn083

Shen, R., Chinnaiyan, A. M., & Ghosh, D. (2008). Pathway analysis reveals functional convergence of gene expression profiles in breast cancer. BMC Medical Genomics, 1(1), 28. http://doi.org/10.1186/1755-8794-1-28

Small, C. L., Shima, J. E., Uzumcu, M., Skinner, M. K., & Griswold, M. D. (2005). Profiling gene expression during the differentiation and development of the murine embryonic gonad. Biology of Reproduction, 72(2), 492–501. http://doi.org/10.1095/biolreprod.104.033696

Soh, Y.Q.S., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. PLOS Genet 11, e1005531.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 102(43), 15545–15550. http://doi.org/10.1073/pnas.0506580102

Taminau, J., Lazar, C., Meganck, S., & Nowé, A. (2014). Comparison of merging and meta-analysis as alternative approaches for integrative gene expression analysis. ISRN Bioinformatics, 2014(1, article 3), 345106–7. http://doi.org/10.1155/2014/345106

Tanaka, S. S., & Nishinakamura, R. (2014). Regulation of male sex determination: genital ridge formation and Sry activation in mice. Cellular and Molecular Life Sciences : CMLS, 71(24), 4781–4802. http://doi.org/10.1007/s00018-014-1703-3

Taneja, R., Bouillet, P., Boylan, J.F., Gaub, M.P., Roy, B., Gudas, L.J., and Chambon, P. (1995). Reexpression of retinoic acid receptor (RAR) gamma or overexpression of RAR alpha or RAR beta in RAR gamma-null F9 cells reveals a partial functional redundancy between the three RAR types. Proceedings of the National Academy of Sciences of the United States of America 92, 7854–7858.

Thompson, D., Regev, A., & Roy, S. (2015). Comparative Analysis of Gene Regulatory Networks: From Network Reconstruction to Evolution. Annual Review of Cell and Developmental Biology, 31(1), 399–428. http://doi.org/10.1146/annurev-cellbio-100913-012908

Tishkoff, D.X., Boerger, A.L., Bertrand, P., Filosi, N., Gaida, G.M., Kane, M.F., and Kolodner, R.D. (1997). Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2. Proceedings of the National Academy of Sciences of the United States of America 94, 7487–7492.

Treichel, D., Becker, M. B., & Gruss, P. (2001). The novel transcription factor gene Sp5 exhibits a dynamic and highly restricted expression pattern during mouse embryogenesis. Mechanisms of Development, 101(1-2), 175–179.

Turnbull, A. K., Kitchen, R. R., Larionov, A. A., Renshaw, L., Dixon, J. M., & Sims, A. H. (2012). Direct integration of intensity-level data from Affymetrix and Illumina microarrays improves statistical power for robust reanalysis. BMC Medical Genomics, 5(1), 35. http://doi.org/10.1186/1755-8794-5-35

Uhlenhaut, N. H., Jakob, S., Anlag, K., Eisenberger, T., Sekido, R., Kress, J., et al. (2009). Somatic Sex Reprogramming of Adult Ovaries to Testes by *FOXL2* Ablation. Cell, 139(6), 1130–1142. http://doi.org/10.1016/j.cell.2009.11.021

Vawter, M. P., Evans, S., Choudary, P., Tomita, H., Meador-Woodruff, J., Molnar, M., et al. (2004). Gender-specific gene expression in post-mortem human brain: localization to sex chromosomes. Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology, 29(2), 373–384. http://doi.org/10.1038/sj.npp.1300337

Villaverde, A. F., Ross, J., Morán, F., & Banga, J. R. (2014). MIDER: network inference with mutual information distance and entropy reduction. Plos One, 9(5), e96732. http://doi.org/10.1371/journal.pone.0096732

Walsh, C., Hu, P., Batt, J., & Santos, C. (2015). Microarray Meta-Analysis and Cross-Platform Normalization: Integrative Genomics for Robust Biomarker Discovery. Microarrays, 4(3), 389–406. http://doi.org/10.3390/microarrays4030389

Wang, J., Chen, B., Wang, Y., Wang, N., Garbey, M., Tran-Son-Tay, R., et al. (2013). Reconstructing regulatory networks from the dynamic plasticity of gene expression by mutual information. Nucleic Acids Research, 41(8), e97–e97. http://doi.org/10.1093/nar/gkt147

Wertz, K., & Herrmann, B. G. (2000). Large-scale screen for genes involved in gonad development. Mechanisms of Development, 98(1–2), 51–70. http://doi.org/10.1016/S0925-4773(00)00452-4

Wilhelm, D., Palmer, S., & Koopman, P. (2007). Sex determination and gonadal development in mammals. Physiological Reviews, 87(1), 1–28. http://doi.org/10.1152/physrev.00009.2006

Wilson, M. J., Jeyasuria, P., Parker, K. L., & Koopman, P. (2005). The transcription factors steroidogenic factor-1 and SOX9 regulate expression of Vanin-1 during mouse testis development. Journal of Biological Chemistry, 280(7), 5917–5923. http://doi.org/10.1074/jbc.M412806200

Yao, H.H.-C., Matzuk, M.M., Jorgez, C.J., Menke, D.B., Page, D.C., Swain, A., and Capel, B. (2004). Follistatin operates downstream of Wnt4 in mammalian ovary organogenesis. Developmental Dynamics: an Official Publication of the American Association of Anatomists 230, 210–215.

Yao, H.H.-C., Whoriskey, W., and Capel, B. (2002). Desert Hedgehog/Patched 1 signaling specifies fetal Leydig cell fate in testis organogenesis. Genes \& Development 16, 1433–1440.

Yokobayashi, S., Liang, C.-Y., Kohler, H., Nestorov, P., Liu, Z., Vidal, M., van Lohuizen, M., Roloff, T.C., and Peters, A.H.F.M. (2013). PRC1 coordinates timing of sexual differentiation of female primordial germ cells. Nature 495, 236–240.

Zhang, H., Klausen, C., Zhu, H., Chang, H.-M., and Leung, P.C.K. (2015). BMP4 and BMP7 Suppress StAR and Progesterone Production via ALK3 and SMAD1/5/8-SMAD4 in Human Granulosa-Lutein Cells. Endocrinology 156, 4269–4280.

CHAPTER V

EPILOGUE

Contribution to The Field

The molecular mechanisms underlying temperature-dependent sex determination (TSD) have been intensively studied during the past decades. However, rather than identifying sex-determining genes unique to TSD, most of these studies have emphasized functions of orthologous genes to mammalian sex-determining genes. Also, steroid hormones are known to play critical roles in sex determination in reptiles. In TSD species, steroid hormones interact with temperature to determine the primary sex of the animal and sometimes can override the effect of temperature (Crews 1996). Estrogen or aromatizable androgen treatments during gonadal development cause permanent male-to-female sex reversal while non-aromatizable androgen treatments have the opposite effect (Crews 1996). However, effects of steroid hormones on sex determination vary dramatically from species to species and sometimes can have opposite effects. The mechanisms behind this phenomenon are still poorly understood. In addition, studies of sex determination often focus on functions of single genes or the identification of differentially expressed genes between the sexes while overlooking broader gene regulatory networks. The purpose of this dissertation was to address these questions and overcome the disadvantages of traditional studies in this field. The main objects of this

dissertation were to 1) investigate the role of androgen in sex determination and differentiation in the snapping turtle, 2) assemble and annotate a reference transcriptome for the snapping turtle and identify novel sex-determining genes with high throughput next generation sequencing (NGS) technology and 3) reconstruct gene regulatory networks in developing mouse gonads using publicly available microarray data and developed a valid workflow transferrable to other comparisons.

Previous studies have demonstrated the feminizing effect of dihydrotestosterone (DHT), a non-aromatizable androgen, in the snapping turtle (Rhen and Lang, 1994; Rhen and Schroeder, 2010). The DHT effect in snapping turtles is the opposite of its effect in the red-eared slider turtle (Crews 1996). Chapter II aimed to test the hypothesis that androgens regulate *Foxl2* expression using a reporter gene assay. Due to the lack of a turtle granulosa cell line, this study was performed in a mouse granulosa cell line (KK1 cells). Although the transfected *Foxl2*-mCherry construct was not affected by DHT treatments in our experiment, the expression of endogenous mouse *Foxl2* was significantly suppressed by DHT, which was the opposite of previous findings in the snapping turtle. This suggested that androgen effects on this gene require regulatory sequences outside the proximal promoter/coding sequence, 2) depend on genomic context, and/or 3) differ between species. In addition, we found that transfected turtle *Foxl2* influenced expression of *FshR, Gnrhr, Star* and aromatase in KK1 cells, which confirmed the effectiveness of our construct and its transfection. Lastly, we found newborn calf serum (NCS) significantly influenced expression of all genes studied in this chapter, which provided insights about the complex effects of NCS on steroid mediated gene regulation.

In Chapter III, a total of 363.2 million read from Roche/454 and Illumina sequencing were assembled into 421,738 contigs. Further analysis identified 19,602 unique protein-coding transcripts, which were then subjected to functional annotation and differential expression analyses. Among the assembled sequences, 16,966 sequences were found to have one or more gene ontology (GO) terms associated with them and 725 differentially expressed genes (DEGs) were identified. With the two-way ANOVA analyses, we identified 293 temperature-responsive genes among all identified DEGs, many of which were also investigated in other studies as sex-determining genes. To further validate our findings, we performed qPCR on 9 DEGs and compared the expression patterns between the two methods. Results were in agreement with our RNA-seq analyses. The comparison of the DEG patterns between the RNA-seq analyses and qPCR analyses indicated the RNA-seq analyses was reliable. Unlike traditional studies in TSD, which focused on orthologs of sex-determining genes in mammals, this study provided novel insights by presenting a reference transcriptome for a TSD species and a set of temperature-responsive genes.

In the final chapter, publicly available microarray data from mouse gonads was used for reverse engineering of gene regulatory networks in the developing gonads. Microarray data between the developmental stage 10.5 dpc to 13.5 dpc, the time window when sex is determined, were selected from 10 studies for the network reconstruction. A total of 226 gene expression profiles (112 profiles for XX gonads and 114 profiles for XY gonads), which contained 10,052 common genes, were merged into a unified dataset with minimized batch effects. We then used ARACNE to reconstruct gene regulatory networks for each sex, i.e. the XX dataset, the XY dataset as well as the combined

dataset. Networks for XX and XY category were subjected to hub gene analysis and for known sex-determining genes. The combined dataset was subjected to the master regulator analysis (MRA). Functional enrichment analyses of hub genes for both XX and XY category indicated genes with most interactions were involved in cell cycle regulation and DNA replication, which reflected the major biological events during sex determination and differentiation. We also compared the computationally inferred networks for key male-determining genes *Sry* and *Sox9* to ChIP-chip derived targets of these two genes. There was overlap between the computationally inferred networks and experimentally derived targets of *Sry* and *Sox9* but there were also many differences. As we explained in Chapter IV, this gap is likely due to differences between the types of interactions detected by the two methods. ARACNE theoretically detects all types of regulatory interactions, both upstream and downstream, while ChIP-chip only detects TF targets. The master regulator analysis in this study identified 503 DEGs and 110 candidate master regulators. We interrogated the network of the top 3 master regulators and found that master regulators may function through collaborating with each other and cross-regulating their targets. The results also suggest novel master regulators such as *Uty1* and *Sp5,* whose functions have never been examined in sex determination, may play critical roles in sex determination.

Future studies

The study in Chapter II did not detect direct androgen regulation of the turtle *Foxl2* reporter construct but did reveal an androgen effect on endogenous *Foxl2* in the murine system. The contradictory regulatory relationship observed in turtle versus murine systems raises new questions about androgen mediated gene regulation. These questions

include: 1) Does *AR* regulate the expression of *Foxl2* directly by binding to androgen response elements (ARE) or indirectly through other mechanisms? 2) Does androgen regulate female-specific genes independently or synergize with genes like *Foxl2?* 3) How does *AR* function in different genomic contexts? More experiments will be needed to further investigate the hypothesized *AR*-*Foxl2* regulatory relationship and answer questions derived from this study.

The computational work in Chapter III and Chapter IV provided large amount of information on sex determination at the transcriptional level. Further experimental studies are needed to parse and validate the results. For example, for the RNA-seq study, the function of 293 temperature-responsive genes and their regulatory network in TSD need to be clarified. For the regulatory network study, the inferred signaling pathways and newly identified master regulators and their relationships with the well-studied master regulators need to be validated.

Conclusions

In vertebrates, females and males exhibit divergent phenotypes and behaviors and sometimes this divergence even extends to diseases. A better understanding of sexual dimorphism helps elucidate selection pressures and differential life histories in animals. More importantly, a good understanding of sexual development will help improve reproductive health and promote the development of new treatments for diseases in humans. Because of the highly conserved gonadal development pattern among vertebrates, studies in TSD shed light on the molecular mechanisms underlying sex determination and differentiation and gene-environment interaction in both TSD species and species with genotypic sex determination (GSD).

The studies in this dissertation used approaches from different areas of biology, including endocrinology, bioinformatics and systems biology, to identify genes and gene networks that may be involved in sex determination. Including both TSD and GSD species in these studies not only improves our understanding of the molecular basis underlying sexual development in vertebrates but also opens up new areas for the study of sex determination across multicellular organisms, in general. In addition, it is interesting to see how might these gene regulatory networks shift in organisms that undergo sex transition within their life cycle (i.e. wrasse). I also think it would be interesting to look through invertebrate species that shift to sexual reproduction with environmental pressure, such as *Daphnia*, to see what their regulatory networks are and how they shift during stress scenarios.

REFERENCES

Crews, D. 1996. “Temperature-Dependent Sex Determination: The Interplay of Steroid Hormones and Temperature.” Zoological Science 13 (1). Zoological Society of Japan: 1–13. doi:10.2108/zsj.13.1.

Rhen, T, K Metzger, A Schroeder, and R Woodward. 2007. “Expression of Putative Sex-Determining Genes During the Thermosensitive Period of Gonad Development in the Snapping Turtle, Chelydra Serpentina.” Sexual Development: Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation 1 (4): 255–70. doi:10.1159/000104775.