Title
Molecular Mechanisms in Male Sex Determination

Author
Arboleda, Valerie Anne

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Molecular Mechanisms in Male Sex Determination

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

Doctor of Philosophy in Human Genetics

By

Valerie Anne Arboleda

2012

ABSTRACT OF THE DISSERTATION

Molecular Mechanisms in Male Sex Determination

By

Valerie Anne Arboleda

Doctor of Philosophy in Human Genetics

University of California, Los Angeles, 2012

Professor Eric Vilain, Chair

Disorders of Sex Development (DSD) encompass a wide range of urogenital anomalies, ranging from mild hypospadias to sex reversal with genital ambiguity that occur in approximately 0.5-1% of live births. Despite the prevalence of DSD, the molecular mechanisms behind the transformation of the bipotential gonad into a testis or an ovary are not completely understood. In this thesis, we explore molecular mechanisms in mammalian sex determination using mouse and human models of sex development.

To identify novel genomic regions and genes in sex determination, we utilize a long-standing and powerful model in mammalian sex determination, the C57BL/6J (B6)-Y^{POS}. None of the B6-Y^{POS} animals have fully normal testis during embryonic development. Our approach utilizes a congenic strain in which a region of 129S1/SvIM (129) origin on chromosome 11 protects from B6-Y^{POS} sex reversal (abbreviated as B6.129-Y^{POS}). Using traditional mouse genetics and backcrossing, we narrowed the 50 megabase (Mb) congenic region to a protective

two Mb interval containing only the promoter region of *Sox9*, a known male sex determination gene.

In order to facilitate a genetic diagnosis of patients with DSD, we developed a novel targeted diagnostic tool based on next generation sequencing technology to rapidly provide a genetic diagnosis in these patients. Using a single test, we can assess sex chromosome karyotype, copy number variation (CNV), and identify damaging single nucleotide variants (SNV) which may result in disease. We have validated our targeted sequencing approach in a pilot group of 14 patients with DSD. Currently, less than 20% of all patients with a DSD receive a definitive diagnosis, yet this type of comprehensive genetic test is critical to assigning appropriate treatment, predicting future development, and evaluating outcomes in this unique patient population.

The synergism between discoveries in mouse and human genetics has existed throughout the history of sex development research—with animal models informing human research, and vice versa. As genomic tools continue to exponentially increase the amount of data generated, this long and healthy relationship between human and mouse genetics will continue to elucidate the complex genetic regulation in sex determination.

The dissertation of Valerie Anne Arboleda is approved.

Barnett Schlinger

A. Jake Lusis

Edward McCabe

Eric Vilain, Committee Chair

University of California, Los Angeles

2012

Dedication

This thesis is dedicated to my wonderful and supportive family. In particular, my husband and my beautiful son who have forever changed my life for the better.

I would also like to dedicate this thesis to my two wonderful female mentors that I had early in my research career. Dr. Carol Troy, my first PI, who believed in my abilities and continues to encourage me to explore the world of research. Second, my organic chemistry professor, Dr. Virginia Cornish, who told me never to doubt my abilities, my intelligence, and myself and whose wise words I will always remember as I continue on this journey.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

The journey to his point in my scientific career has not been a linear path, but one in which all the branchpoints have led me to pursue a career that I truly enjoy and in which I find great reward. I have come to recognize that the path to scientific happiness lies in enjoying the process, because the fruits of the labor are elusive, yet wonderful when one finally finds them.

I would like to thank my PI, Dr. Eric Vilain for his support, academically and intellectually, throughout my PhD. I would also like to thank the esteemed members of my committee: Dr. A. Jake Lusis, Dr. Barney Schlinger, and Dr. Edward McCabe for their guidance and support of my educational endeavors. None of the projects in this dissertation would be possible without our many collaborators, Dr. Emmanuèle Délot, Dr. Hane Lee, Dr. Janet Sinsheimer, Dr. Stanley F. Nelson, and Dr. Julian Martinez-Agosto, who have provided their expertise to these research studies. Finally, I cannot forget all the wonderful people who have influenced my time in the graduate program: Dr. Alice Fleming, Dr. Saunders Ching, Dr. Francisco Sanchez, Dr. Ruth Baxter, Negar Gharamani, and Tuck Ngun for their training, mentorship, and friendship.

I also would like to thank my various sources of funding, including the HHMI Medical Students Research Training Fellowship, which started me working in the Vilain Lab and started me on a research-oriented path, UCLA Medical Scientist Training Program, and NIH NICHD F31 NRSA Predoctoral Fellowship. All of these funds have allowed me the freedom to pursue new ideas and become a better scientist.

Chapter 1 contains excerpts from the chapter published in Yen and Jaffe's Reproductive Endocrinology and from a chapter published in the Genetic Diagnosis of Endocrine disorders, with the following citations: Arboleda V.A., Vilain E. (2009). Disorders of Sex Development. In

J.F. Strauss & R.L. Barbieri (Eds.) Yen and Jaffe's Reproductive Endocrinology: Physiology, Pathophysiology, and Clinical Management (6th ed., pp. 1367-1393). Philadelphia, PA: Saunders Elsevier and Arboleda V.A., Fleming A., Vilain E.. Disorders of Sex Development. Genetic Diagnosis of Endocrine Disorders. Eds. R. Weiss and S. Refetoff. McGraw-Hill. 2010.

Chapter 2 includes data for future publication. We thank Dr. Alice Fleming for her help on Immunofluorescence experiments and Dr. Janet Sinsheimer for her statistical expertise.

Chapter 3 includes published work from Arboleda VA, Lee H, Sánchez F, Délot E, Sandberg D, Grody W, Nelson S, Vilain E. Arboleda V, Lee H, Sánchez F, Délot E, Sandberg D, Grody W, Nelson S, Vilain E. Clin Genet. 2012 Mar 21.

Chapter 4 is work that is accepted for publication: Arboleda VA, Lee H, Parnaik R, Fleming A, Banerjee A, Ferraz-de-Souza B, Délot EC , Rodriguez-Fernandez IA, Braslavsky D, Bergadá I, Dell'Angelica EC, Nelson SF, Martinez-Agosto JA, Achermann JC, Vilain E. Mutations in the PCNA-binding domain of CDKN1C cause IMAGe Syndrome. Nature Genetics. *In Press*.

Chapter 5 was published as a review with the following citation: Arboleda VA, Vilain E. The evolution of the search for novel genes in mammalian sex determination: from mice to men. Mol Genet Metab. 2011 Sep-Oct;104(1-2):67-71. 2011 Jul 6. Review.

VITA

2005	B.A., Biology with Concentration in Asian American Studies Columbia College, Columbia University New York, New York
2005-7	David Geffen School of Medicine (I & II) University of California, Los Angeles Los Angeles, California
2006	AFAR NIA & Lillian R. Gleitsman Fellowship Medical Student Training in Aging Research Fellowship
2007-8	Howard Hughes Medical Institute Medical Student Research Training Fellowship Vilain Lab and Rodriguez Lab University of California, Los Angeles Los Angeles, California
2011, April	National Research Service Award F31 National Institutes of Health National Institute of Child Health and Development
2011	61st Nobel Laureate Meetings at Lindau NIH sponsored Attendee Physiology and Medicine Lindau, Germany

PUBLICATIONS AND PRESENTATIONS

Arboleda VA, Lee H, Parnaik R, Fleming A, Banerjee A, Ferraz-de-Souza B, Délot EC , Rodriguez-Fernandez IA, Braslavsky D, Bergadá I, Dell'Angelica EC, Nelson SF, Martinez-Agosto JA, Achermann JC, Vilain E. Mutations in the PCNA-binding domain of CDKN1C cause IMAGe Syndrome. Nature Genetics. 2012. *In Press*.

Arboleda VA, Fleming A, Sinsheimer J, Vilain E. Protection from C57BL/6-Ypos sex reversal is conferred by a POSa-specific enhancer of Sox9. 6th International Symposium on Vertebrate Sex Determination. April 24, 2012. Kona Hawaii. Presentation.

Arboleda VA, Lee H, Sanchez F, Delot EC, Sandberg DE, Grody W, Nelson SF, Vilain E. Targeted Massively Parallel Sequencing Provides Comprehensive Genetic Diagnosis for Patients with Disorders of Sex Development. Clinical Genetics. 2012 Mar 15.

Arboleda VA, Vilain E. The evolution of the search for novel genes in mammalian sex determination: from mice to men. Mol Genet Metab. 2011 Sep-Oct;104(1-2):67-71. Epub 2011 Jul 6. Review.

White S, Ohnesorg T, Notini A, Smith C, Hewitt J, Daggag H, Roeszler K, Turbitt E, Gustin S, van den Bergen J, Miles D, Western P, **Arboleda, VA**, Schumacher V, Gordon L, Bell K, Bengtsson H, Speed T, Hutson J, Warne G, HarleyV, Koopman P, Vilain E, Sinclair A Copy number variation in patients with disorders of sex development due to 46,XY gonadal dysgenesis. PLOS One. 2011 Mar 7;6(3):e17793.

Sutton E, Hughes J, White S, Sekido R, Tan J, **Arboleda V**, Rogers N, Knower K, Rowley L, Eyre H, Rizzoti K, McAninch D, Goncalves J, Slee J, Turbitt E, Bruno D, Bengtsson H, Harley V, Vilain E, Sinclair A, Lovell-Badge R, Thomas P. Identification of SOX3 as an XX male sex reversal gene in mice and humans. J. Clin Invest. 2011 Jan 4;121(1):328-41.

Arboleda VA, Fleming A, Vilain E. Disorders of Sex Development *Genetic Diagnosis of Endocrine Disorders.* Eds. R Weiss and S Refetoff. McGraw-Hill 2010.

Arboleda VA, Lee H, Braslavsky D, Nelson SF, Bergada I, Vilain E. Identification of a novel locus on Chromosome 11 in a large family with IMAGe Syndrome. Abstract 2155 Presented at the annual meeting of The American Society of Human Genetics, November 3, 2010, Washington DC. Abstract.

Vilain E, **Arboleda VA**. A Congenic Approach To The Identification Of Novel Sex Determining Genes. Abstract 1401. Presented at the annual meeting of The American Society of Human Genetics, October 23, 2009, Honolulu, Hawaii. Platform Presentation.

Arboleda VA, Vilain E (2009) Disorders of Sex development In JF Strauss & RL Barbieri (Eds) *Yen and Jaffe's Reproductive Endocrinology: Physiology, Pathophysiology, and Clinical Management* (6th ed, pp 1367-1393) Philadelphia, PA: Saunders. Elsevier.

Davidson TJ, Harel S, **Arboleda VA**, Shelanski ML, Greene LA, Troy CM Highly efficient siRNA delivery to primary mammalian neurons induces microRNA-like effects before mRNA degradation. Journal of Neuroscience 24 (45):10040-6 2004.

Prunell GF, **Arboleda VA**, Troy CM. Caspase Function in Neuronal Death: Delineation of the Role of Caspases in Ischemia Current drug targets-CNS and Neurological disorders. 4(1): 51-61. 2005.

Prunell GF, **Arboleda VA**, Shelanski ML, Connolly ES, Troy CM "Neuronal SOD1 Knockdown: A Cellular Model for Ischemia." Washington, DC: Society for Neuroscience, 2004 Poster Presentation.

Chapter 1

Introduction

Introduction to Sex Development

Even before birth, one of the major defining characteristics is the phenotypic sex, visable beginning at 16 weeks gestation by ultrasound. The reproductive tracts are among the earliest sexually dimorphic structures to form within the developing fetus. The presence of either testes or ovaries portend the vast differences not only in the internal and external genitalia, but in basic physiology, behavior, size, muscle mass, and disease susceptibility. The most well studied sex differences are those resulting from hormonal secretions of the testis and ovaries, which have direct effects on external and internal genitalia differentiation and exert a strong influence over development of non-reproductive structures for the duration of an individuals' life. Underpinning these hormonally regulated anatomic sex differences are genetic factors, primarily genes which are critical to the formation of hormone-secreting tissues, such as the gonads and adrenal glands. Therefore, the root of sex differences lies in two factors: genetic factors (e.g. sex chromosome complement, autosomal genes) and hormonal factors largely derived by the presence of either ovaries or testes.

One of the major genetic factors responsible for directing the fate of the bipotential gonad towards a testis is the Y-chromosome; male embryos have a chromosome complement of 46, XY while female embryos are 46, XX. The Y-chromosome contains the testis-determining gene, *SRY*[1], whose expression initiates the formation of the testis from the bipotential gonad. Until the point of gonad differentiation, male and female embryos are anatomically indistinguishable. Before the point of gonad determination, there are few described differences between XY and XX embryos. Gene expression studies between XX and XY mouse whole brains show distinct differences in the expression profile before the presence of gonadal hormones[2]. These results

underscore the importance of the genetic sex differences that are present prior to the determination of the gonads.

The pathways regulating gonadal development require complex interaction of multiple genes that force the bipotential gonad to make a binary choice to become either a testis or an ovary. The earliest evidence of sex differences within the bipotential gonad occurs when *SRY* is up-regulated in gonadal somatic support cells around 6-8 weeks of human embryonic development. Expression of *SRY* in an autosomal dominant fashion initiates a cascade of genetic events, ultimately resulting in the formation of a testis, whose hormonal secretions work throughout the body to develop male–specific internal and external genitalia, and program the developing brain. In the absence of *SRY*, a female-specific pathway is initiated, up-regulating genes that promote the differentiation and maintenance of the ovary, the Müllerian ducts, and female external genitalia. The developmental decision to differentiate into a testis or an ovary sets the groundwork for sex differences throughout one's lifespan.

The binary decision of the bipotential gonad to differentiate into a testis or an ovary determines whether a fetus is male or female. These chromosomal, gonadal, and anatomical sex differences lies at the heart of our societal structure, in which biological sex and gender influences our interactions with people and the environment. In this thesis, I will explore the molecular and genetic mechanisms at the root of male sex determination and testis formation. A clearer understanding of these mechanisms will allow us to better understand normal sex development and disorders of sex development (DSD) arising from mutations in genes that are essential to testis and ovarian determination. Using both mouse models of sex determination and human disorders of sex development, we can elucidate the molecular mechanisms influencing

testis development and ultimately provide better diagnostic and treatment outcomes in patients with DSD.

Gonadal Development

At conception, genetic sex is determined based on whether one inherits an X or Y chromosome from one's father. *Sex determination* refers to the developmental decision that directs the bipotential, undifferentiated gonad into either a testis or an ovary. *Sex differentiation* refers to the downstream effects that are modulated by hormones secreted from the testis or ovary (Figure 1-1). Sex determination is governed by genetic factors, requiring a complex orchestration of timing, cell-specific localization, and expression levels of sex-specific gene products to force an undifferentiated bipotential gonad into becoming either a testis or an ovary.

Sex Determination: Testes or Ovary?

When the chromosomal karyotypes of patients with 47, XXY Klinefelter syndrome who are male[3], and 45, XO Turner syndrome who are female[4] were determined, it became clear that the Y-chromosome is both necessary and sufficient for male sex determination. Therefore, the Y-chromosome was deemed "sex-determining" and it was hypothesized that the gene controlling testis formation, the testis determining factor (TDF), had to be located on the Y-chromosome.

In the early 1990s, a series of elegant experiments found the testis-determining factor using patients who had 46, XX testicular or ovotesticular DSD. The karyotype of these patients identified a small portion of the Y-chromosome translocated to one of the X-chromosomes. Positional cloning located a small 35-kb fragment of the Y-chromosome translocated to the X-chromosome of these patients [5,6]. Within the 35-kb sequence, a conserved region was identified, which coded for the *SRY* gene. Gene expression analysis revealed a male-specific increase in

SRY mRNA consistent with earliest known divergence of male and female gonadal development. Furthermore, XX mice transgenic for a 14-kb fragment containing *Sry* developed testes and the full male phenotype, but were infertile[7]. These initial findings were followed up with experiments showing that deletions and mutations in *SRY*/*Sry* result in XY sex reversal in both humans and mice [8-10].

In the presence of a Y-chromosome with an intact *SRY*, the bipotential gonad diverges towards testes organogenesis[1,7,11,12] by week 8 of embryonic gestation. Testis determination is influenced by autosomal factors which prime the bipotential gonad, making it receptive to initiation of the testis determination pathway by *SRY*. Patients with mutations in *MAP3K1*[13] have 46, XY Gonadal Dysgenesis, implicating the map kinase-signaling cascade as critical to establishing the cell-lineage priming for testis differentiation. Mice with mutations in a related map kinase gene, *MAP3K4* have XY sex reversal associated with decreased levels of SRY and decreased numbers of Sertoli cells[14]. Other genes known to be important in early sex development are Chromobox-2 (*CBX2)*, Steroidogenic Factor-1 (*SF-1, NR5A1)*, Wilms-Tumor-1 (*WT1)*, *GATA4,* and *FOG2*[15,16]. Mutations in any of these genes result in XY Gonadal Dysgenesis in humans. *SRY* expression, in conjunction *SF-1*, in the Sertoli cell precursors initiates testes determination by activating downstream effectors such as *SOX9* (*SRY*-related HMG-box 9)[17]. One gene with roles in both testis and ovarian determination is *DAX1*. A single copy of the orphan nuclear receptor *DAX-1* is necessary for testes development and spermatogenesis[18], while duplication of the *DAX-1* locus in XY fetuses suppresses testes development resulting in 46, XY gonadal dysgenesis[19,20].

Once testis development is initiated, *SOX9*'s high level of expression is maintained by itself, fibroblast growth factor-9 (*FGF9*), and fibroblast growth factor receptor 2 (*FGFR2*).

Sertoli cell precursors begin to proliferate and organize around germ cells into tubular cords. Precursors to steroidogenic Leydig cells (under stimulation from placental human chorionic gonadotropin) migrate into the developing testes in the 10th week, while vascular endothelial cells from the mesonephros migrate from the mesonephros. Vascular endothelial growth factor (VEGF) and platelet derived growth factor (PDGF) signaling initiate testes-specific vasculature and expansion of the mesenchymal lineage[21]. Desert hedgehog (DHH) signaling positively regulates Leydig cell proliferation by up-regulating *SF-1*[22,23]. Mature testes are morphologically recognizable at approximately 10 weeks gestational age.

Compared to testis determination, ovarian organogenesis is poorly understood. Initial studies in mice suggest that ovary development is a relatively gene-poor process compared to testes differentiation[24,25], although this may reflect differential timing of maximal transcriptional activity between developing testes and ovaries. Nevertheless, specific X-chromosome genes as well as autosomal genes such as *WNT-4* (wingless-related MMTV integration site 4), Rspondin-1 (*RSPO1*), both acting through Beta-Catenin, and likely a downstream factor Follistatin[26] are required for proper ovarian development. Recent studies have shown that the ovarian cell-fate needs to be actively maintained throughout life. In mice, deletion of *FOXL2* in adult granulosa cells results in transdifferentiation of granulosa cells into Sertoli-like cells[27]. Therefore, a secondary role of many ovary-promoting genes may be to continually suppress *SOX9* and the male developmental pathway.

Once the developmental choice has been made, and either a testis or an ovary has formed, hormones direct the process of sex differentiation throughout the rest of the body using endocrine and paracrine signals. Critical to this dogma was work of French physiologist, Alfred Jost. In elegant microsurgery experiments, rabbit fetuses were castrated *in utero* around the time

of sex determination, day 15-20. The early removal of the testis resulted in the development of female internal and external genitalia while early removal of the ovary appeared to have little effect on the development of female reproductive structures[28,29]. Furthermore, transplanting a fetal testis unilaterally in female fetuses results in unilateral regression of the Müllerian ducts and complete external masculinization. Simply transplanting a crystal releasing testosterone unilaterally did not have any effect on the development of Müllerian structures, but resulted in external masculinization. These results heralded the discovery that the testis secreted two major "masculinizing factors", testosterone and anti-Müllerian hormone (AMH). Testosterone and its more potent form dihydrotestosterone (DHT) are responsible for masculinizing the external genitals and differentiating the Wollfian ducts. Sertoli cells secrete anti-Müllerian hormone (AMH), which causes regression of the precursors to the uterus and fallopian tubes in male fetuses. AMH expression *in situ* has been consistently detected by the 9th week of gestation[30,31]. Sex differentiation is a long process occuring throughout the embryonic development and adulthood to program the non-gonadal tissues in a sex-specific fashion.

We have only begun to understand the process by which normal gonads are differentiated at the molecular level. The molecular mechanisms required for patterning the bipotential gonad into a testis or an ovary are still largely unexplained. Our current knowledge stems from genetic diagnoses in patients with disturbances in sex determination or sex differentiation. However, these studies are limited by our inability to detect nuanced gonadal phenotypes that are not visible to the naked eye. Studies of gonad development in model organisms where gene dosage, localization, and timing can be perturbed, can dissect out complicated molecular pathways in testis development.

Disorders of Sex Development

For the first time in 2005, a consensus on the management of "intersex" disorders emerged, bringing together experts from a variety of fields (endocrinology, genetics, surgery, psychology, advocacy groups). The consensus statement provides guidelines in all aspects of management of "intersex", including genital surgery, the requirement of a multidisciplinary team and of mental health professionals, the diagnostic approach. This thesis reflects the advances made in the field as well as the recommendations of the consensus statement. One area that changed considerably was the nomenclature of intersexuality. The term *Disorders of Sex Development (DSD)* was proposed, as defined by "congenital conditions in which development of chromosomal, gonadal or anatomical sex is atypical"[32]. This broad definition replaces the word "intersex", which has social connotations and reflects a concept of sexual identity. It includes patients that do not necessarily have ambiguous external genitalia (such as complete androgen insensitivity, or Turner syndrome) and yet belong to the broad family of DSD. Here we will cover the three broad categories of DSD: Sex chromosome DSD, 46, XY DSD and 46, XX DSD.

Sex Chromosome Disorders of Sex Development

Sex Chromosome DSDs are defined by aneuploidy of the sex chromosomes, X and Y. In disorders of autosomal aneuploidy (i.e.Trisomy 21, etc), maternal errors in meiotic non-disjunction account for the majority of cases. However, paternal errors in meiosis are the etiology of the majority of sex chromosome aneuploidy[33]. Two of the most common sex chromosome DSD are Turner Syndrome and Klinefelters Syndrome.

Turner Syndrome

In 1938, a group of females with short stature, primary amenorrhea, and a lack of secondary sex characteristics were described by Turner[34]. Turner Syndrome is classified as a sex chromosome DSD, with the majority having a 45,X karyotype. The remaining Turner syndrome patients have a 46,XX karyotype with a deletion in part of one X chromosome, have various mosaics involving 45, X and 46, XX, and 46, XY cells, or more complex combinations. The universal feature in 45, X patients is short stature, with adult height less than 58 inches. The gene responsible for short stature in both Turners Syndrome and idiopathic short stature is *SHOX,* located on in the pseudoautosomal region (PAR1) of the X-chromosome[35].

In the prenatal period, fetuses with 45,X DSD and its variants frequently have intrauterine growth restriction. Suspicion for Turner syndrome should be raised if the presence of large septate cystic hygromas, nuchal thickening, short femur, total body lymphangiectasia, and/or cardiac defects is detected by ultrasound. At birth, Turner syndrome babies often have low birth weights; 30% of babies will present with lymphedema of the upper and lower extremities, an extension of the lymphangiectasias *in utero*, which disappears in the first few months of life. Additionally, the presence of a webbed neck (pterygium colli) and dysmorphic features including: high-arched palate, low set prominent ears, low posterior hairline, epicanthal folds, micrognathia, hypoplastic nail beds, and/or hypoplastic 4^{th} and 5^{th} metacarpals at birth should raise the suspicion for 45,X DSD [36]. Turner syndrome patients may have increased carrying angle of the arms (cubitus valgus), shield-like chest with wide set nipples, hearing loss, and a higher frequency of cardiovascular malformations such as co-arctation of the aorta. The incidence of renal anomalies is between 30-50%, with the most common being a horseshoe kidney followed by abnormal vasculature.

The classic Turner 45, X karyotype is thought to be one of the most common human chromosomal abnormalities and is estimated to occur in 0.8% of all zygotes. However, less than 3% of 45,X zygotes survive to term and this karyotype is commonly found in spontaneous abortions[37]. The incidence of 45, X karyotypes is approximately 1 in 2000 live newborn phenotypic females[38].

The gonads of Turner syndrome females do not develop normally and have reduced follicle formation and growth *in utero*[39]. A normal number of eggs develop in girls with Turner Syndrome, but, for unknown reasons, they disappear prematurely[40]. Interestingly, the gonads appear as streaks of white tissue next to the fallopian tubes—called "streak gonads" which on histology have primitive connective tissue stroma without primary follicles. Patients with Turner syndrome have fewer follicles and therefore less estrogen secretion from granulosa cells than normal females, resulting in delayed or failure to undergo puberty.

During adolescence, the most common presentation is short stature, amenorrhea, and lack of secondary sex characteristics. Patients may also have any of the other characteristics mentioned above but 45, X DSD patients have normal intelligence. Depending on the degree of gonadal dysgenesis, up to 30% of Turner patients undergo some degree of spontaneous puberty[41].

Klinefelter Syndrome

Klinefelter syndrome was first described in 1942, in a group of 9 men who had sparse facial and body hair, small testis, infertility due to azoospermia, and gynecomastia[42]. This particular set of characteristics was determined to be due to the presence of an extra X-chromosome in these males, making them 47, XXY.

In early life, 47, XXY boys may be diagnosed with behavioral disorders, abnormally small testes, and long legs. The presence of long lower extremities distinguishes Klinefelter from the other forms of eunuchoidism that results in equally long upper and lower extremities. Klinefelter patient's IQ typically falls in the normal range, however it tends to be below that of other siblings[43]. Most patients' present in adolescence with small firm testes and hypogonadism with varying degrees of androgen deficiency[44]. In later life, many males present at infertility centers with azoospermia.

In utero, the XXY fetal testis has the normal complement of primordial germ cells in the XXY testes. However, the germ cells degenerate during childhood, most likely due to a defect in Sertoli cells and germ cell communication during testes maturation[45]. In theory, the etiology of the non-disjunction in Klinefelter can be from maternal meiosis I or II or paternal meiosis II, and thus each situation should contribute 33% of cases. However, nearly 50% of Klinefelter patients show a paternal origin of the extra X[46]. Some studies also showing increased frequency of XY sperm in men with advanced paternal age, but the data remains inconclusive[47].

46, XY Disorders of Sex Development

46, XY Gonadal Dysgenesis

XY gonadal dysgenesis (GD) occurs when fetal testes fail to develop normally *in utero*. GD can be pure (or complete), partial, or mixed. The major criteria for the diagnosis of pure, partial, or mixed gonadal dysgenesis are the appearance and histology of both gonads. Therefore, the ultimate diagnosis of any of the forms of gonadal dysgenesis, particularly the distinction between mixed and partial GD, requires biopsy of both gonads. In many patients, this type of diagnosis is difficult to acquire because of the invasiveness of the procedure. There exist some

distinguishing phenotypic features that can determine the difference between the different types of GD, but these are not 100% accurate.

Pure GD individuals possess intraabdominal bilateral fibrous streaks instead of testes and therefore do not secrete any of the normal hormones and paracrine factors, such as testosterone and anti-Müllerian hormone, which work together to form the male external and internal genitalia. Pure XY GD individuals appear as unambiguous phenotypic females (previously known as Swyer syndrome) with hypoplastic (but occasionally normal) Müllerian structures on pelvic ultrasound. Additionally, these women are of normal or tall stature, delayed or absent puberty, and display none of the classic Turner stigmata.

Isolated 46,XY pure GD is considered in an adolescent with primary amenorrhea and sexual immaturity with a full female external phenotype, while partial or mixed GD should be considered more likely in the differential diagnosis of a 46, XY patient with ambiguous genitalia.

Patients with "partial" gonadal dysgenesis have varying degrees of Wolffian and Müllerian development and genital ambiguity correlating with the degree of testes dysgenesis. The medical literature is confusing in regards to the difference between "partial" and "mixed" GD, which are often used interchangeably. Partial GD refers to intermediate stages of dysgenic testes, between streak gonads and normal testes, and usually has a 46, XY, non-mosaic karyotype. However, mixed 46, XY GD often refers specifically to individuals with a completely dysgenic streak gonad on one side and a partially dysgenic (or even a normal-appearing) testis opposite.

Partial and mixed 46,XY GD patients present in infancy or early childhood with varying degrees of ambiguous genitalia and Wolffian and Müllerian structures development. While the streak gonad in mixed 46, XY GD is always abdominally located, dysgenic testes may be either abdominally or inguinally located, depending on the degree of testicular tissue. If one of the

testes is relatively normal, it can present in the scrotum. Patients with partial 46,XY GD usually have female external genitalia with some degree of virilization, such as clitoromegaly or a bifid scrotum. Uterus and Fallopian tubes are usually well-formed, but occasionally may be hypoplastic. In mixed GD, the development of Wolffian and Müllerian structures as well as the virilization of external genitalia correlates with the degree of development of the ipsilateral testis resulting in asymmetric virilization of the external or internal genitalia and unilateral cryptorchidism. However, in mixed GD pelvic ultrasound may show asymmetric Müllerian development, with female internal genitalia present ipsilateral to the streak gonad and male internal genitalia present ipsilateral to normal or dysgenic testes.

Given the primary importance of *SRY* in human testes determination, it is surprising that mutations in *SRY* only account for 15% of all pure XY GD [10,48-50]. SRY is a transcription factor that binds to a highly conserved 1.4kb region of the *SOX9* promoter that is required for testis specific *SOX9* transcription[17]. Early synergistic binding by SRY and SF-1 (*NR5A1*) are essential for initiating *SOX9* transcription. *SOX9* transcription is maintained by a positive feedback loop in which SOX9 binds to its own promoter and is further stimulated by FGF9 and FGFR2. The majority of *SRY* mutations causing XY pure GD disrupt the high mobility group (HMG) box, thereby either reducing DNA binding[50], interfering with DNA bending[51], or inhibiting nuclear import of the SRY protein[52]. There are over 50 known mutations of the *SRY* open reading frame (ORF). Deletions of Yp involving *SRY* have also been implicated in XY GD[48,49,53].

Steroidogenic factor 1 (SF-1/NR5A1) is an orphan nuclear receptor necessary for adrenal and testicular development. Recent studies have underlined the importance of SF-1 in testicular development, as nearly 15% of isolated XY GD can be attributed to SF-1 haploinsufficiency[54]. However, three cases have reported *SF-1* mutations with XY, GD and adrenal hypoplasia

congenita, which reflect the role of SF-1 in adrenal development. Together, mutations or deletions of *SRY* and *SF-1* account for 30% of XY GD, indicating that other genes critical to the testes determination cascade still remain to be discovered.

Mutations in genes necessary for testes organogenesis (*MAPK1*, *GATA4*, *CBX2*, *XH2, SOX9, WT-1, DHH*) and duplications of putative "anti-testis" genes (*WNT-4, DAX1*) are responsible for a very small minority of all XY GD[55-57], but are usually associated with other syndromic features.

More recently, copy number variations involving patients who are 46, XY female have involved intergenic regulatory regions of known sex determining genes[58,59]. A 1.2 Mb deletion in the promoter region of *SOX9* results in 46, XY female and a deletion in the *GATA-4* promoter leads to 46, XY Female with *adrenal hypoplasia congenita*. Overlapping deletions in the 9p region in non-related patients with 46, XY Gonadal Dysgenesis identified a small 256kb region containing 4 genes, including *DMRT1*, a homolog of *doublesex*, a sex determination gene in drosophila.

While most cases are sporadic, familial cases of 46,XY pure[50,60] and partial GD[50,61] have been reported (reviewed in[62]). Some intriguing cases involve *SRY* mutations detectable also in the normal father and male relatives[50,60]. Other reports are of fathers who are mosaics for mutant and normal *SRY* transmitting the mutation to their XY female daughters[63,64] are less puzzling. Presumably, the normal father's mosaicism reflects a post-zygotic mutation event. These cases demonstrate the influence of genetic background on sex determination, and may provide further clues for elucidating the molecular basis for gonad formation.

Mosaicism for 45,X/46,XY is the most frequent cause of mixed GD, although a minority have a 46,Xi(Yq) karyotype[65]. Individuals with 45,X/46,XY mosaicism present with quite

variable phenotypes: in one series of ten 45,X/46,XY patients, three presented as Turner females with bilateral streak gonads, three presented as mixed GD and genital ambiguity, and four as males possessing bilateral testes and varying degrees of undervirilization[66]. There is evidence that increasing proportion of gonadal Y chromosome material correlates with increasing testicular tissue and phenotypic maleness[67,68].

Since the risk of gonadoblastoma in patients with gonadal dysgenesis is so high, precise diagnosis of the type of gonadal dygenesis is determined after prophylactic or therapeutic gonadectomy. In pure 46, XY gonadal dysgenesis both gonads are streak gonads. Bilateral dysgenic testes define partial GD, while the presence of at least one streak gonad is more consistent with mixed GD. Karyotype of peripheral leukocytes shows 46,XY in pure and partial GD, and mosaic 45,X/46,XY is frequent in mixed GD. Furthermore, if gonadal histology is asymmetric (e.g. streak gonad on one side, testis on the other), karyotyping of the gonads may reveal cryptic mosaicism.

Disorders of Androgen Biosynthesis and Action

46, XY DSD patients that have normally developed testis with phenotypes ranging from undervirilized males to full sex reversal can be a result of defects in steroidogenic enzymes or in the sex steroid receptor. The biosynthesis of testosterone is essential for development of secondary sex characteristics in XY individuals, including differentiation of Wolffian structures, inguinoscrotal testes descent, and masculinization of external genitalia after conversion to DHT. Mutations in one of the steroidogenic enzymes responsible for converting progesterone to testosterone or mutations that affect the actions of testosterone, can cause incomplete virilization in XY individuals (Figure 1-2).

Congenital Lipoid Adrenal Hyperplasia

Mutations of enzymes early in the transport and synthesis of steroid hormones, such as steroidogenic acute regulatory (StAR) protein, and of cytochrome P450 side chain cleavage P450scc (*CYP11A)* [69,70], cause a total loss of adrenal and gonadal steroidogenesis. StAR protein facilitates the initial transport of cholesterol across the mitochondrial membrane of the adrenal and gonadal steroidogenic cells. P450scc catalyzes the initial reaction in all steroidogenic tissues, the conversion of cholesterol to pregnenolone (Figure 1-2)[71,72]. Perturbations in either step result in progressive cholesterol accumulation in steroidogenic organs, ultimately leading to adrenal failure[73]. Patients with mutations in StAR protein or P450scc have lipid vacuoles in adrenals on histology. However, only patients with StAR protein mutations have adrenal hyperplasia.

The lack of steroidogenesis results in a severe decrease in masculinizing sex steroids, resulting in partial to complete sex reversal only in XY fetuses. All newborns will also exhibit the features of adrenal insufficiency at birth, the severity of which depends on whether there is a partial or complete loss of function of the enzyme.

Patients with congenital lipoid adrenal hyperplasia have testes, a blind vaginal pouch, no Müllerian structures, partial or absent Wolffian derivatives and a range of external genital phenotype: from female external genitalia to undervirilized genitals. Testes can be in the abdomen, inguinal canal, or labia. The majority of affected patients have excess generalized pigmentation at birth due to intrauterine glucorticoid deficiency that results in elevated ACTH levels. Partial deficiency of StAR protein may still result in spontaneous puberty in 46, XX patients 46, XX patients are still able to have anovulatory menses because the ovaries are able to produce estrogen through StAR-independent pathways[73-75]. Additionally, because ovaries are

dormant until puberty, there is no cholesterol-induced damage to ovarian cells [74]. At puberty, XX individuals develop multiple ovarian cysts, likely from anovulation and cholesterol accumulation.

Definitive diagnosis of congenital lipoid adrenal hyperplasia is done by sequencing of STAR or P450scc. Low deoxycorticosterone levels, hyperkalemia, and absence of hypertension differentiate 46, XY males with lipoid hyperplasia from patients with 17-hydroxylase/lyase deficiency, as both disorders are characterized by low 17-OH progesterone levels. Patients should be given physiological replacement doses of glucocorticoid and mineralocorticoids. Furthermore, patients should be given sex hormones at the onset of puberty in accordance with phenotypic sex.

3β-Hydroxysteroid Dehydrogenase deficiency

Rare defects in adrenal and gonadal 3β-hydroxysteroid dehydrogenase (3βHSD2, *HSD3B2* gene) affect cortisol, aldosterone, and testosterone synthesis result in adrenal insuffienciency in XX and XY patients but genital ambiguity only in XY patients. 3βHSD2 is the enzyme responsible for conversion of all steroid precursors to those required for synthesis of androgens and cortisol in all steroidogenic tissues, including the testes, ovaries, and adrenal glands (Figure 1-2). This rare variant of CAH can have a wide range of phenotypes, and ranging from classical presentation (salt wasting, with a range of female external genitalia to ambiguous genitalia in genetic XY individuals) to non-classical (no salt-wasting and later presentation).

17α-Hydroxylase/17,20-Lyase Deficiency

Biosynthesis of adrenal and sex hormones require cytochrome P450 17α-hydroxylase/17,20-lyase (*CYP17*). *CYP17* catalyzes two distinct reactions in the adrenals and gonads: the hydroxylation of carbon-17 in C21 steroids and the lysis of the 17,20 carbon-carbon bond. This rare variant of congenital adrenal hyperplasia is characterized by decreased cortisol

synthesis with shunting of steroidogenesis towards mineralocorticoid precursor production (Figure 1-2).

Most patients present to a clinician at puberty with primary amenorrhea, as well as low-renin hypertension and hypokalemia due to elevated mineralocorticoids. This autosomal recessive disorder results in phenotypes ranging from female external genitalia with no Wolffian duct development to ambiguous genitalia with partial development of the male ductal system. The degree of ambiguous genitalia depends on the severity of the enzymatic defect. However, the ambiguous genitalia phenotype is only in 46, XY individuals. All patients show Müllerian structure regression, due to presence of a testis with functional Sertoli cells. In contrast to individuals with 3βHSD deficiency or Lipoid CAH, 17,20-lyase deficiency patients do not develop adrenal crisis[76]. Rare *CYP17* mutations cause isolated 17,20-lyase deficiency [77], leaving glucocorticoid synthesis intact. Therefore, only sex steroid synthesis is decreased, resulting in variably feminized 46, XY individuals. Patients should be treated with physiological levels of glucocorticoids and appropriate sex steroid replacements to develop secondary sex characteristics.

P450 Oxidoreductase Deficiency

Cytochrome P450 oxidoreductase (POR) deficiency is a recently characterized cause of congenital adrenal hyperplasia, which affects the enzymatic activity all microsomal P450 enzymes, including the steroidogenic enzymes P450c17, P450c21, and P450aro [78-81]. POR is a flavoprotein associated with the endoplasmic reticulum, and is an electron donor to proteins in the P450 system. Unlike other genetic defects, which only affect one enzyme in the steriodogenesis pathway (Figure 1-2), POR causes partial enzymatic activity in multiple enzymes. Therefore, POR deficiency presents with a wide range of phenotypes and is particularly difficult

to diagnose using serum hormone levels. Partial enzymatic deficiencies in 21α-hydroxylase account for the CAH phenotype and virilization in XX female, while partial 17α-hydroxylase deficiency accounts for undervirilization of XY individuals. Mothers carrying fetuses with POR deficiency are virilized due to aromatase deficiency.

The unique pathophysiology of POR deficiency results in a wide phenotypic spectrum, in which patients present during prenatal period, as newborns with ambiguous genitalia, or as XX adolescents with amenorrhea and/or Polycystic Ovarian Syndrome. During prenatal screening, this diagnosis should be considered in the differential of a pregnant woman with low estriol levels but with a normal ultrasound exam.

17β-Hydroxysteroid Dehydrogenase Deficiency

The17β-HSD3 enzyme (*HSD17B3*) is a member of the 17β-HSD gene family, which catalyzes the conversion of androstenedione to testosterone in the testes (Figure 1-2). This is a rare cause of 46,XY DSD, with only 10 reported cases in the United States[82].

HSD17B3 deficiency may be difficult to recognize in infancy since XY individuals most often appear with female external genitalia, lacking genital ambiguity, and they have no other clinical problems. Depending on the degree of residual enzyme activity, 17βHSD3 deficiency can, in few cases, cause a range of undervirilization of XY individuals at birth, from an empty bifid scrotum with perineoscrotal hypospadias to clitoromegaly and posterior labioscrotal fusion. Testes can be undescended or labial. Müllerian ducts are absent but interestingly, testosterone derived Wolffian ducts are present. During puberty, 46, XY patients may experience progressive virilization, clitoral/phallic enlargement, muscle development, breast development, testicular descent, and plasma testosterone levels within the normal range of pubertal males[82-84]. No affected 46, XX females have been reported.

5α-Reductase Deficiency

5α-reductase deficiency (*SRD5A2*) results from a defect in the 5α-reductase enzymes that converts testosterone to dihydrotestosterone (DHT). DHT is a more potent form of testosterone and is responsible for the differentiation of male external genitalia *in utero* (see Figure 2). Peripheral conversion of testosterone to dihydrotestosterone (DHT) is an irreversible reaction catalyzed by the two isoenzymes of 5α-reductase, SRD5A1 and SRD5A2. The SRD5A1 isozyme is mainly expressed in skin during puberty and genital fibroblasts. SRD5A2 isozyme is predominantly expressed in the genital skin tissue, male accessory sex organs, and prostate during embryogenesis. Defects in SRD5A2 result in undervirilization of male external genitalia, differentiated Wolffian ducts, which terminate in a blind vaginal pouch, and a hypoplastic prostate[85,86]. Testicular and adrenal androgen synthesis is unaffected.

Presentation of 5α-reductase deficiency is unique because these patients are born of one gender, typically female, but during puberty, due to the surge in testosterone, become more phenotypically male. Affected 46,XY infants present with various degrees of undervirilization and ambiguous genitalia ranging from isolated glandular hypospadias to the most common presentation: perineal hypospadias with a blind perineal pouch, micropenis, and bifid scrotum [87]. 46, XY patients have Wolffian duct differentiation but absent Müllerian derivatives. Due to the huge surge of testosterone at puberty, patients develop male habitus, deepening of the voice, muscular growth, and penile enlargement. Due to underdevelopment of the prostate and seminal vesicles, semen is highly viscous and ejaculate volume is extremely low(<0.5-1 ml), but sperm counts may be normal. 46,XX females with *SRD5A2* deficiency have normal sexual differentiation, delayed puberty, sparse sexual hair and normal fertility.

Diagnosis of SRD5A2 deficiency during early infancy and puberty can be made based on an elevated ratio of testosterone to DHT (normal < 30:1), with or without hCG stimulation [85,88]. DHT is low in infants but can reach near-normal levels during adolescence without treatment, presumably by peripheral SRD5A1 activity. If the diagnosis is suspected during pre-pubertal childhood or during adolescence, hCG stimulation is necessary to elicit the diagnostic hormone levels since the hypothalamo-pituitary-gonadal axis is inactive. Affected females are phenotypically normal, but have the same biochemical abnormalities as affected males.

Five-alpha reductase deficiency is inherited in an autosomal recessive manner. Mutations have been reported in all five exons and over 50 mutations in *SR5DA2* have been found, ranging from point mutations to deletions, however but the genotype-phenotype correlation is poor.

Most XY patients with *SR5A2* deficiency who were raised as females revert to male gender at puberty[89-91]. It is not clear why there is such a high prevalence of change in gender identity in this disorder. However, since this disorder has been described in consanguineous families in the Dominican Republic, Turkey, New Guinea, and Saudi Arabia, cultural and societal reasons may have an important role. Topical DHT cream application above the pubic area (25 mg-50 mg daily) promotes phallic growth prior to puberty and increases the presence of facial and body hair. Fertility is possible with surgical correction of the male ductal system or through assisted reproductive technology.

Complete and Partial Androgen Insensitivity Syndrome

Androgen insensitivity syndrome (AIS) is an X-linked disorder caused by mutations affecting the androgen receptor (*AR*) gene resulting in decreased peripheral responsiveness to circulating androgens. AR is a member of the nuclear receptor superfamily of transcription factors and activated by both testosterone and DHT. Its function is essential for sex

differentiation and maintenance of normal spermatogenesis in males. Over 300 mutations have been identified worldwide with an estimated incidence of 1 per 20,400 liveborn XY individuals.

AIS is characterized by a range of undervirilization in 46, XY individuals from complete XY female phenotype, termed complete AIS (CAIS) while phenotypic XY males with the only manifestation being infertility, gynecomastia, or hypospadias [92,93], referred to as minimal AIS (MAIS). Between these two extremes exists a large grey area, with various degrees of XY ambiguity, referred to as partial AIS (PAIS). All known forms of AIS are caused by disruption of the androgen receptor activity. CAIS is generally associated with a complete absence of androgen binding and AR activation. Beyond this, there is little correlation between the level of residual AR activity and degree of AIS [94], suggesting that factors such as genetic background and interactions of AR with co-activators or promoters, influence tissue response to androgens and the resultant phenotype.

The phenotype of patients with PAIS is extremely heterogeneous. Patients present in infancy or childhood with variable degrees of virilization, ranging from micropenis to isolated cryptorchidism. Infants and children may also present with unilateral or bilateral inguinal hernias. Testes are present and functional, producing high levels of testosterone and AMH resulting in variable Wolffian derivative development and regression of Müllerian structures in most patients, respectively [95]. In patients presenting at puberty, breast development and sparse pubic hair are suggestive of PAIS, and help differentiate it from 5α-Reductase 2 deficiency.

CAIS usually presents at puberty with primary amenorrhea. Exam reveals a short, blind vagina, absent uterus, and palpable inguinal or labial testes. Testosterone-dependent Wolffian derivatives and prostate are absent or vestigial. Height, bone maturation, and breast development are normal, but pubic and axillary hair, an androgen-mediated feature, is absent or sparse.

Patients' identity and behavior is feminine[96]. Less commonly, CAIS may present in infancy with phenotypic female genitalia and inguinal or labial masses representing testes. Minimal AIS is considered in phenotypic XY males with infertility or gynecomastia, but normal masculine genitalia.

In familial PAIS, and in both familial and sporadic CAIS, mutations in *AR* exonic sequences are found in 85-90% of cases[97,98]. Even within a family, the phenotype can vary between PAIS and CAIS[97,99]. Mutations causing AIS are equally spread throughout the *AR* open reading frame with no one defect particularly prevalent, although hotspots do exist [98]. Single amino acid substitutions account for 90% of reported cases. The majority of cases are inherited, but 30% of all AIS cases are *de novo* mutations. *De novo* mutations typically originate in the mother in a single germ cell or as germ cell mosaicism, in which the recurrence risk is lower.

Normal 17-OH progesterone levels and normal androgen levels differentiates AIS from the various forms of CAH. In prepubertal children, basal LH and testosterone may be normal, but hCG stimulation elicits an exaggerated androgen response (a tripling instead of a doubling of testosterone and DHT). Pelvic exam and ultrasound reveals absence of uterus and fallopian tubes and the presence of abdominal testes.

Historically, residual AR function has been used to aid in the diagnosis of AIS. However, variations between tests done in different laboratories, and biopsies taken from different areas limits the usefulness of these assays[100]. Molecular genetic testing of the AR gene detects mutations in more than 95% of probands with CAIS. PCR-based sequencing of *AR* exons 2-10 can be routinely performed[101], as well as sequencing of the much longer exon 1, and some intronic and promoter regions. Prenatal testing by mutation analysis is available for families in which the AR-mutation has been identified in an affected family member[97].

Leydig Cell Hypoplasia/Agenesis

Leyding Cell Hypoplasia (LCH) is disorder caused by inactivating mutations of LH Receptor (LHR) and characterized by impaired Leydig cell differentiation and testosterone production. HCG and LH activate a shared G-protein-coupled hCG/LH receptor (LHR). *In utero*, placental hCG stimulates Leydig cells to produce testosterone, resulting in male internal and external genitalia. LH takes over during the third trimester of gestation and neonatal life to complete Leydig development and continue testosterone production.

Over 20 inactivating mutations of *LHR* have been identified scattered throughout the gene[102,103] which cause variable loss of receptor activity. Males with a milder phenotype, such as micropenis or hypospadias result from milder missense mutations[104]. An intriguing mutation in LHR causing absence of exon 10 has been described in an 18 year old male that presented with normal male phenotype, pubertal delay, small testicles and delayed bone age[105]. This unique phenotype highlights the differential actions of hCG and LH on the shared LHR receptor. In this patient, normal male sex differentiation occurred *in utero* confirming LHR's response to hCG but the LHR did not respond to LH, resulting in delayed puberty and bone age.

The presentation of patients with LCH depends on the degree of residual LHR activity ranging from patients presenting at birth with micropenis, hypospadias, and cryptorchidism to patients with completely female external genitalia and palpable labial or inguinal masses who are often missed in the newborn period. Patients with complete female external phenotype usually present in puberty with amenorrhea and sexual infantilism. Inactivating mutations of LHR in XX females cause hypergonadotropic hypogonadism with primary amenorrhea or oligomenorrhea, cystic ovaries, and infertility[106-109]. However, these females undergo spontaneous breast and pubic hair development during puberty.

Pelvic ultrasound shows absence of uterus and fallopian tubes. To distinguish LCH from patients with XY GD in which the entire testis fails to form, AMH is used as a marker of testicular function and is normal-high in LCH patients and low to undetectable in XY GD patients[110]. Testes are only slightly smaller and histologic examination of the testis of LCH patients shows normal Sertoli cells, hyalinization of the seminiferous tubules, absence of mature Leydig cells, and spermatogenesis. Patients with partially inactivating mutations may show some early spermatogenesis, with arrest prior to spermiation. Therefore, patients have potential for fertility using assisted reproductive technology.

Although the presentation can be similar to 17βHSD3 or 5α-reductase deficiency, LCH patients have significantly lower testosterone and DHT levels, are less responsive to hCG, and have a normal Androsteindione/Testosterone ratio (< 15:1) as well as T/DHT ratio. Definitive diagnosis requires sequencing of the LHR gene for deletions, insertions, and point mutations.

46, XX Disorders of Sex Development

46, XX Testicular and Ovotesticular Disorders of Sex Development

Testicular and Ovotesticular DSD are distinguished on purely histological grounds in which Testiscular DSD individuals have immature testis without proper germ cell formation. Ovotesticular DSD signifies the presence of both testicular (defined by seminiferous tubules) and ovarian (defined by follicles) tissues in one individual. This can take the following forms: ovotestis on one side with contralateral ovary or testis (50%); ovary and testis opposite on opposite sides (30%); or bilateral ovotestes (20%). For unknown reasons, testicular tissue is more often located on the right, and ovarian on the left [111].

The spectrum of individuals with XX testicular DSD can vary from phenotypic males with immature testes with varying degrees of genital ambiguity to patients with ovotesticular DSD (who will be addressed more thoroughly in the following section). The proportion of ovarian and testicular tissues within the gonad leads to variable development of the ipsilateral internal and external genitalia. The location of the individual gonad usually correlates with the predominant tissue type: gonads with ovarian-predominance are most often abdominal and correctly positioned, while testes are usually scrotal[111]. The degree of Müllerian and Wolffian structure correlates with the ipsilateral gonad; Wolffian structures generally develop only with a well-formed testis, confined to the same side. However, Müllerian structures develop in the presence of an ovary or an ovotestis. Gonadal hormone production and type of external genitalia depend on the predominant type of gonadal tissue.

The incidence of XX-testicular DSD is estimated at 1:20,000. Approximately 85% of patients with XX-testicular DSD have unambiguously masculine genitalia[112]. Ninety percent of these males harbor an Xp:Yp translocation containing *SRY*. One-third of all recombinations occur at the *PRKX* locus hotspot on Xp22.3 and the Y-homologue *PRKY* [113].

Of the 15% of patients with 46, XX-testicular DSD and ambiguous genitals, only a minority carries *SRY* [114,115]. Preferential inactivation of the Y-bearing X chromosome[112,114,116], and cryptic mosaicism with *SRY* expression confined to the testes[112,117] has been implicated in causing partial, rather than complete, sex reversed phenotype. XX-testicular DSD has also been associated with the overexpression of Sox family genes. For example, a 17q23-24 duplication of the SOX9 promoter, that may or may not include SOX9 has been associated with XX-Testicular DSD[118,119]. Similarly, misexpression of SOX3 in the developing gond, through duplications or deletion of SOX3 enhancer elements has also been shown to result in XX-Testicular DSD[120].

Recently, a form of syndromic ovotesticular DSD has been shown to be due to a single gene mutation in Rspondin1 [121] and associated with hyperkeratosis and squamous cell carcinoma.

Given the phenotypic variability of ovotesticular DSD, there is less genetic homogeneity in the genetic etiology of ovotesticular DSD, relative to testicular DSD. Approximately 60% of ovotesticular DSD have a 46,XX karyotype of which 10% are *SRY*-positive[49,122-125], suggesting that these patients are part of the phenotypic spectrum which includes XX-testicular DSD. Both, testicular and ovotesticular DSD may be caused by partial inactivation of the X chromosome with the Xp:Yp translocation[114,126], or by cryptic mosaicism for *SRY*[112,117,127]. Between 30-33% of ovotesticular DSD are mosaics with at least one cell line containing Yp material[111,127]. One-third of these mosaics are 46,XX/46,XY, though an unknown number of 46,XX/46,XY mosaics may actually represent chimeras [128], which has been shown in a 5 cases tetragametic chimeras[129-133]. The remaining 7-10% of ovotesticular DSD cases occur in 46,XY individuals, and 2 cases have been reported with mutations found *SRY*[49,111,127]. Over all, ovotesticular DSD is much more common in 46, XX individuals rather than 46, XY.

Familial cases of XX-testicular DSD represent predominantly *SRY*-negative kindred, whose members can present with either 46,XX partial or complete testicular DSD, or ovotesticular DSD[134].

The majority of patients with XX-testicular DSD (80%) have classic “de la Chapelle” syndrome with completely masculinized genitals. The patients exhibit some similarities with Klinefelter (XXY), such as small testes, azoospermia, gynecomastia (37%), and lack of secondary sex development [116] but unlike Klinefelter individuals, patients with XX testicular DSD are short and typically do not have learning disabilities or behavioral issues. Phallus size varies from small to normal, and sexual function and ejaculation are normal.

The degree of genital ambiguity often determines when, during childhood, one presents with testicular or ovotesticular DSD. In males without genital ambiguity, a diagnosis is often made during an investigation of infertility or delayed puberty. Patients have with XX Testicular DSD have hypergonadotropic hypogonadism with elevated FSH and LH, decreased T, DHT, and a less than two-fold increase in response to the hCG-stimulation test. Unlike individuals with GD or XX Testicular DSD, the gonads of patients with ovotesticular DSD have some degree of function and therefore normal levels of FSH, LH, E_2, T, and DHT. Most patients with ovotesticular DSD come to clinical attention at birth or early childhood due to ambiguous genitalia. The phenotype of external and internal genitalia reflects the hormonal status of the gonad. Unicornate uterus is a common finding in all patients, whereas adnexa and vagina are generally better developed in phenotypic females. Approximately half of phenotypic females menstruate [122,125]. Patients with a high degree of masculinization and male phenotype may have some uterine remnants, such as a prostatic utricle (prostatic vagina). Phenotypic males often have bilateral palpable gonads or at least one descended gonad. The undescended gonad may be an intraabdominal ovary or an ovotestis located at any point of the pathway of testicular descent.

It is important to note that the only way to obtain definitive diagnosis either ovotesticular or testicular DSD is by gonadal tissue confirmation in which both gonads are extensively examined for the presence of ovarian tissue. Complete absence of ovarian tissue can allow for definitive diagnosis of 46, XX Testicular DSD, while the slightest presence of ovarian tissue gives rise to 46, XX Ovotesticular DSD. Gonadal tissue can also be tested for karyotype, cryptic mosaicism, or *SRY* mutations since *SRY*-negative patients with ovotesticular DSD have been found to have *SRY* expression in the gonads[112,117,135]. Patients with a mosaic karyotype (most

commonly 46,XX/46,XY) and their parents should be tested with microsatellite markers to determine potential chimerism, especially in twin cases, or those conceived with IVF.

Serum AMH level greater than 75 nmol/L is the biochemical marker which gives unequivocal evidence of functioning testicular tissue, suggestive of XX testicular DSD or ovotesticular DSD[110]. Pelvic ultrasound shows absence of a uterus and semen analysis demonstrates normal semen volume with azoospermia. A combination of both male and female internal and external genitalia is suggestive of 46, XX Ovotesticular DSD. FISH for *SRY* is positive in 90% of adults with XX testicular DSD who present without genital ambiguity. However, in patients with XX testicular DSD with negative SRY, further cytogenetic testing should include Copy Number Variant analysis to identify *SOX9* (17q24) microduplications or *SOX3* duplications and deletions[136].

46, XX Gonadal Dysgenesis

The majority of cases of ovarian dysgenesis involve abnormal complements of X chromosomes in phenotypic females[137]. The most common of these is Turner's syndrome, discussed earlier as a Sex Chromosome DSD. Impaired ovary development can occur with a normal female sex chromosome complement, referred to as 46,XX GD, which can occur as either an isolated entity or as part of a syndrome. In cases of XX, GD, females possess normal female genitalia at birth, but remain sexually infantile and do not experience normal puberty. Ovarian failure with hypergonadotropic hypogonadism at puberty is highly suggestive of ovarian dysgenesis. Müllerian structures are normal, and gonadal histology reveals bilateral streaks. Patients present with amenorrhea and infertility.

Several studies in familial ovarian dysgenesis have noted the frequent reports of consanguinity, suggesting that isolated 46, XX GD is inherited as an autosomal recessive

condition[138-141]. Mutations in the gene for FSH receptor (gene *FSHR*, 2p16) have been implicated in some cases[140,142,143], but does not explain the phenotype in all populations studied[144,145]. Translocations involving the critical region Xq13.3-q26 have also been described[146]. The molecular basis for 46,XX GD remains unknown for the majority of patients.

46, XX Disorders of Sex Differentiation: Androgen Excess

High levels of intrauterine androgens and/or androgenic precursors result in masculinization of the 46, XX fetuses' internal and external genitalia. External genitalia often present as clitoromegaly, "scrotalization" of the labia, urogenital sinus, and phallic urethra. There are three major sources of excess androgens *in utero*: fetal steroidogenesis, fetoplacental steroidogenesis, and maternal causes of excess androgens (i.e. iatrogenic, androgenic tumors).

The most common causes of genital ambiguity in newborns are the virilizing forms of CAH [147]. Inactivating mutations of any of the enzymes in the steroidogenic pathway that result in an increased androgen load, shown in Figure 2, are the most common cause of virilization in XX infants. Rare causes of CAH that can result in either 46,XY DSD or 46, XX DSD, including P450scc deficiency, StAR protein mutations, 3βHSD2 deficiency, and 17-βHSD3 are described above in 46, XY DSD: Disorders in Androgen Biosynthesis.

Congenital Adrenal Hyperplasia

Depending on the enzyme implicated, Congenital Adrenal Hyperplasia (CAH) can cause either 46, XY DSD or 46, XX DSD. Mutations in 21-hydroxylase (*CYP21)* accounts for over 90% of CAH and represent the most common etiology of ambiguous genitalia in the XX newborn. The disease frequency in the general population is 1:15,000 people, but is much higher in certain ethnic groups, including Hispanics and Ashkenazi Jews[148]. The mechanisms of excess

androgen production is through blockage of both aldosterone and cortisol biosynthetic pathways. Therefore early precursors are shunted towards androgen biosynthesis, resulting in virilization of XX fetuses and, in some, maternal virilization. Mutations in 11β-hydroxylase (*CYP11B1*) account for another 5% of virilizing CAH[147,149].

Cytochrome P450 oxidoreductase (*POR*) deficiency is a recently characterized cause of CAH that affects the enzymatic activity of all microsomal P450 enzymes, including the steroidogenic enzymes *CYP17*, *CYP21*, and *CYP19* [78-81]. POR causes partial enzymatic activity in multiple enzymes, presents with a wide range of phenotypes, and is particularly difficult to diagnose using serum hormone levels. Additionally, it can cause ambiguous genitalia in both XX and XY individuals. In mothers carrying a fetus with a *POR* mutation, virilization of the mother can occur. This virilization is due to decreased fetal aromatase activity, which causes excess testosterone to be present in bloodstream of both the fetus and mother.

Conclusion

Disorders of sex development are a common congenital malformation with consequences for the child's gender assignment, medical management, surgical management, and future fertility. The approach to management of these complex cases require an interdisciplinary team of clinicians and health care specialists to assess not only the child's medical history, but the psychosocial impact of decisions that are made early in the child's life. While understanding the molecular mechanisms will allow us to better predict the progression of the condition and the risk of recurrence within the family, there is also a significant psychosocial aspect to this condition—both for the parents and the child. Therefore, careful thought must be given to irreversible decisions (e.g. surgery) that may not be in the best interest of the child but provide

some level of comfort to the parents. The greater implications of this research into the molecular mechanisms of sex determination will allow us to identify novel genes and mechanisms important in disorders of sex development and ultimately provide insight into the clinical management of DSD.

Figure 1-1: Sex Development. Sex determination is the process by which genetic sex (presence or absence of the Y chromosome) leads to gonadal sex. Sex Differentiation is the process of sex development after gonadal sex has been determined, and determined primarily by the actions of gonadal hormones on non-gonadal tissue. In XY males, sex differentiation is an active process requiring testes derived hormones, testosterone, Müllarian inhibiting substance, and Insulin-3 (Insl3). In XX females, sex differentiation is a default pathway, which occurs in the absence of testes-derived hormones.

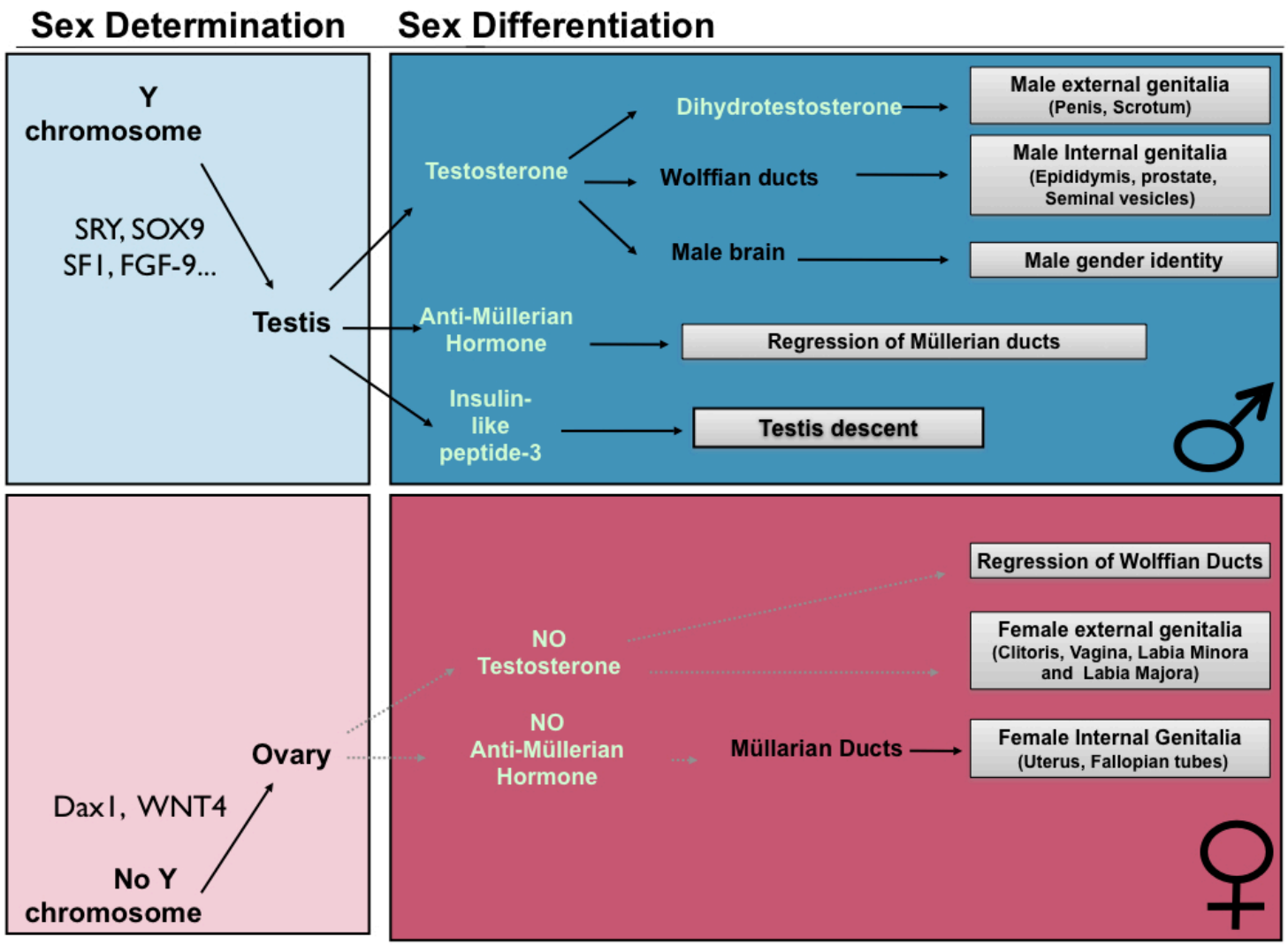
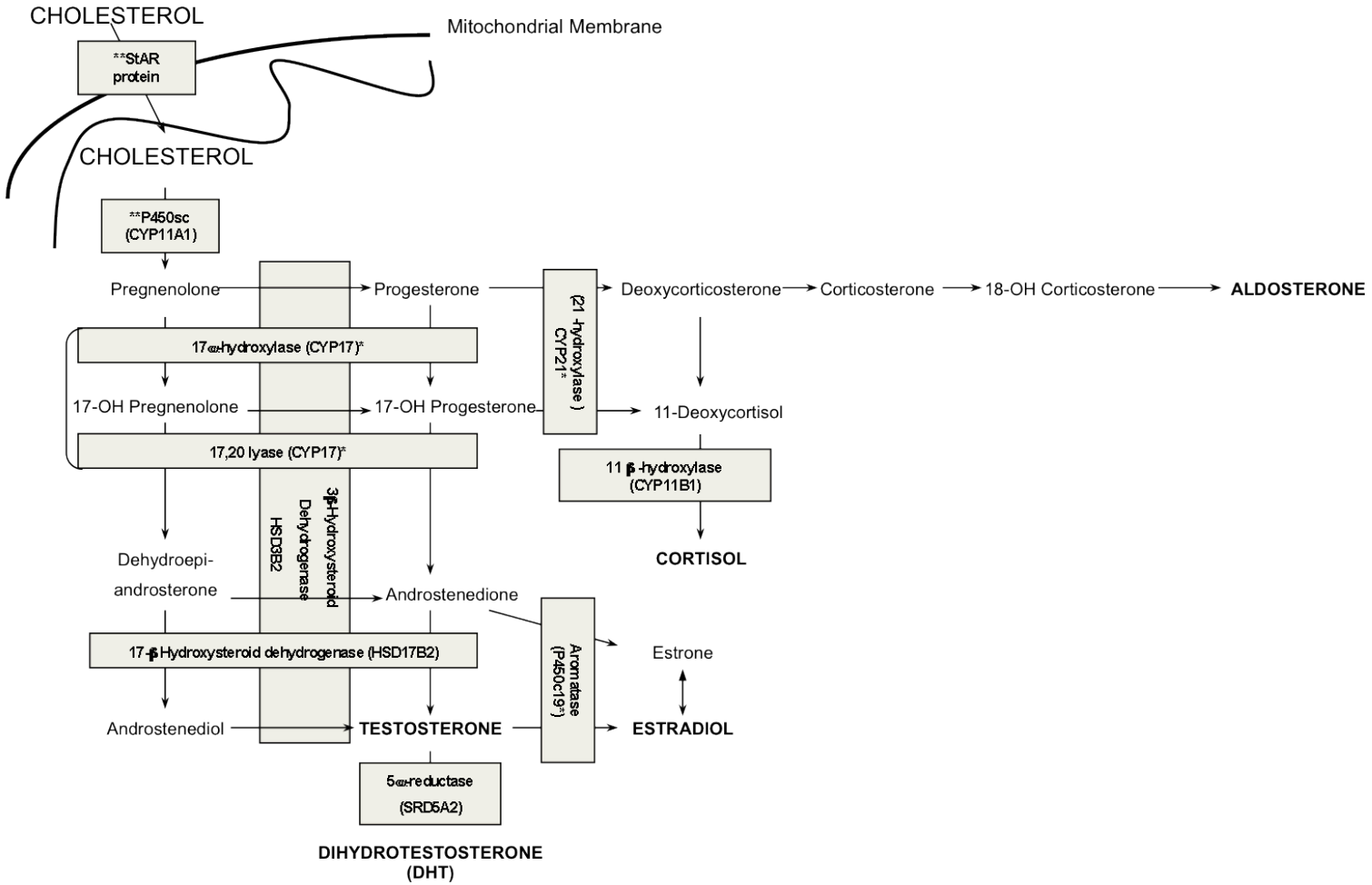

Sex Determination
Sex Differentiation
Y chromosome
SRY, SOX9
SF1, FGF-9...
Testis
Testosterone
Dihydrotestosterone
Male external genitalia
(Penis, Scrotum)
Wolffian ducts
Male Internal genitalia
(Epididymis, prostate, Seminal vesicles)
Male brain
Male gender identity
Anti-Müllerian Hormone
Regression of Müllerian ducts
Insulin-like peptide-3
Testis descent
Ovary
Dax1, WNT4
No Y chromosome
NO Testosterone
NO Anti-Müllerian Hormone
Regression of Wolffian Ducts
Female external genitalia
(Clitoris, Vagina, Labia Minora and Labia Majora)
Müllarian Ducts
Female Internal Genitalia
(Uterus, Fallopian tubes)

Figure 1-2: Steroidogenesis Pathway. Many disorders of sex differentiation are due to defects in steroidogenesis, with the majority of them occurring in the adrenal gland.



*P450 Oxidoreductase deficiency results in partial activity of this enzyme
**Both these proteins result in congenital Lipoid Adrenal Hyperplasia

Table 1-1
Proposed Revised Nomenclature for DSD

Previous	Proposed
Intersex	DSD
Male pseudohermaphrodite, undervirilization of and XY male, and undermasculinization of an XY male	46, XY DSD
Female pseudohermaphrodite, overvirilization of and XX female, and masculinization of an XX female	46, XX DSD
True hermaphrodite	Ovotesticular DSD
XX male or XX sex reversal	46, XX Testicular DSD
XY sex reversal	46, XY complete Gonadal Dysgenesis

From "Consensus Statement on Management of Intersex Disorders"

References

1 Sinclair, A. H. *et al.* A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240-244 (1990).

2 Dewing, P., Shi, T., Horvath, S. & Vilain, E. Sexually dimorphic gene expression in mouse brain precedes gonadal differentiation. *Brain Res Mol Brain Res* **118**, 82-90 (2003).

3 Jacobs, P. A. & Strong, J. A. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature* **183**, 302-303 (1959).

4 Ford, C. E., Jones, K. W., Polani, P. E., De Almeida, J. C. & Briggs, J. H. A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet* **1**, 711-713 (1959).

5 Vergnaud, G. *et al.* A deletion map of the human Y chromosome based on DNA hybridization. *Am J Hum Genet* **38**, 109-124 (1986).

6 Palmer, M. S. *et al.* Genetic evidence that ZFY is not the testis-determining factor. *Nature* **342**, 937-939,

7 Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. Male development of chromosomally female mice transgenic for Sry. *Nature* **351**, 117-121 (1991).

8 Gubbay, J. *et al.* A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**, 245-250 (1990).

9 Lovell-Badge, R. & Robertson, E. XY female mice resulting from a heritable mutation in the primary testis-determining gene, Tdy. *Development* **109**, 635-646 (1990).

10 Knower, K. C., Kelly, S. & Harley, V. R. Turning on the male--SRY, SOX9 and sex determination in mammals. *Cytogenet Genome Res* **101**, 185-198,

11 Koopman, P., Munsterberg, A., Capel, B., Vivian, N. & Lovell-Badge, R. Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* **348**, 450-452 (1990).

12 Berta, P. *et al.* Genetic evidence equating SRY and the testis-determining factor. *Nature* **348**, 448-450 (1990).

13 Pearlman, A. *et al.* Mutations in MAP3K1 cause 46,XY disorders of sex development and implicate a common signal transduction pathway in human testis determination. *American journal of human genetics* **87**, 898-904,

14 Bogani, D. *et al.* Loss of mitogen-activated protein kinase kinase kinase 4 (MAP3K4) reveals a requirement for MAPK signalling in mouse sex determination. *PLoS Biol* **7**, e1000196.

15 Biason-Lauber, A., Konrad, D., Meyer, M., DeBeaufort, C. & Schoenle, E. J. Ovaries and female phenotype in a girl with 46,XY karyotype and mutations in the CBX2 gene. *American journal of human genetics* **84**, 658-663, (2009).

16 Lourenco, D. *et al.* Loss-of-function mutation in GATA4 causes anomalies of human testicular development. *Proc Natl Acad Sci U S A* **108**, 1597-1602,

17 Sekido, R. & Lovell-Badge, R. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* **453**, 930-934

18 Achermann, J. C., Meeks, J. J. & Jameson, J. L. Phenotypic spectrum of mutations in DAX-1 and SF-1. *Mol Cell Endocrinol* **185**, 17-25 (2001).

19 Muscatelli, F. *et al.* Mutations in the DAX-1 gene give rise to both X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism. *Nature* **372**, 672-676 (1994).

20 Barbaro, M. *et al.* Isolated 46,XY gonadal dysgenesis in two sisters caused by a Xp21.2 interstitial duplication containing the DAX1 gene. *The Journal of clinical endocrinology and metabolism* **92**, 3305-3313, (2007).

21 Cool, J., DeFalco, T. J. & Capel, B. Vascular-mesenchymal cross-talk through Vegf and Pdgf drives organ patterning. *Proc Natl Acad Sci U S A* **108**, 167-172, (2011).

22 Clark, A. M., Garland, K. K. & Russell, L. D. Desert hedgehog (Dhh) gene is required in the mouse testis for formation of adult-type Leydig cells and normal development of peritubular cells and seminiferous tubules. *Biol Reprod* **63**, 1825-1838 (2000).

23 Yao, H. H., Whoriskey, W. & Capel, B. Desert Hedgehog/Patched 1 signaling specifies fetal Leydig cell fate in testis organogenesis. *Genes Dev* **16**, 1433-1440 (2002).

24 McClive, P. J., Hurley, T. M., Sarraj, M. A., van den Bergen, J. A. & Sinclair, A. H. Subtractive hybridisation screen identifies sexually dimorphic gene expression in the embryonic mouse gonad. *Genesis* **37**, 84-90 (2003).

25 Menke, D. B. & Page, D. C. Sexually dimorphic gene expression in the developing mouse gonad. *Gene Expr Patterns* **2**, 359-367 (2002).

26 Yao, H. H. *et al.* Follistatin operates downstream of Wnt4 in mammalian ovary organogenesis. *Dev Dyn* **230**, 210-215 (2004).

27 Uhlenhaut, N. H. *et al.* Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell* **139**, 1130-1142, (2009).

28 A, J. Recherches sur la differenciation sexuelle de l'embryon de lapin. III. Role des gonades foetales dans la differenciation sexuelle somatique. *Archives d'Anatomie et de Microscopie Morphologique Experimentale* **36**, 271-315 (1947).

29 Jost, A. Problems of fetal endocrinology: the gonadal and hypophyseal hormones. *Recent Prog Horm Res* **8**, 379-418 (1953).

30 Josso, N. *et al.* Anti-mullerian hormone in early human development. *Early Hum Dev* **33**, 91-99 (1993).

31 Rajpert-De Meyts, E. *et al.* Expression of anti-Mullerian hormone during normal and pathological gonadal development: association with differentiation of Sertoli and granulosa cells. *J Clin Endocrinol Metab* **84**, 3836-3844 (1999).

32 Lee, P. A., Houk, C. P., Ahmed, S. F. & Hughes, I. A. Consensus statement on management of intersex disorders. International Consensus Conference on Intersex. *Pediatrics* **118**, e488-500, (2006).

33 Martin, R. H. Meiotic errors in human oogenesis and spermatogenesis. *Reproductive biomedicine online* **16**, 523-531 (2008).

34 Turner, H. H. A syndrome of infantilism, congenital, webbed neck and cubitus valgus. *Endocrinology*, 566-574 (1938).

35 Rao, E. *et al.* Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nature genetics* **16**, 54-63, (1997).

36 Ferguson-Smith, M. A. Karyotype-Phenotype Correlations in Gonadal Dysgenesis and Their Bearing on the Pathogenesis of Malformations. *J Med Genet* **39**, 142-155 (1965).

37 Eiben, B., Borgmann, S., Schubbe, I. & Hansmann, I. A cytogenetic study directly from chorionic villi of 140 spontaneous abortions. *Hum Genet* **77**, 137-141 (1987).

38 Stochholm, K., Juul, S., Juel, K., Naeraa, R. W. & Gravholt, C. H. Prevalence, incidence, diagnostic delay, and mortality in Turner syndrome. *J Clin Endocrinol Metab* **91**, 3897-3902, (2006).

39 Reynaud, K. *et al.* Number of ovarian follicles in human fetuses with the 45,X karyotype. *Fertil Steril* **81**, 1112-1119 (2004).

40 Weiss, L. Additional evidence of gradual loss of germ cells in the pathogenesis of streak ovaries in Turner's syndrome. *J Med Genet* **8**, 540-544 (1971).

41 Pasquino, A. M., Passeri, F., Pucarelli, I., Segni, M. & Municchi, G. Spontaneous pubertal development in Turner's syndrome. Italian Study Group for Turner's Syndrome. *J Clin Endocrinol Metab* **82**, 1810-1813 (1997).

42 Harry F. Klinefelter, E. C. R., And Fuller Albright. Syndrome Characterized by Gynecomastia, Aspermatogenesis without A-Leydigism, and Increased Excretion of Follicle-Stimulating Hormone. *The Journal of Clinical Endocrinology* **2**, 615-627 (1942).

43 Robinson, A., de la Chapelle, A. in *Emery and Rimoin's Principles and Practive of Medical Genetics* (ed Emery and Rimoin's) Ch. 973, (Churchill Livingston 1996).

44 Lanfranco, F., Kamischke, A., Zitzmann, M. & Nieschlag, E. Klinefelter's syndrome. *Lancet* **364**, 273-283, (2004).

45 Lue, Y. *et al.* XXY male mice: an experimental model for Klinefelter syndrome. *Endocrinology* **142**, 1461-1470 (2001).

46 Thomas, N. S. & Hassold, T. J. Aberrant recombination and the origin of Klinefelter syndrome. *Hum Reprod Update* **9**, 309-317 (2003).

47 Eskenazi, B. *et al.* Sperm aneuploidy in fathers of children with paternally and maternally inherited Klinefelter syndrome. *Hum Reprod* **17**, 576-583 (2002).

48 McElreavey, K. *et al.* XY sex reversal associated with a deletion 5' to the SRY "HMG box" in the testis-determining region. *Proc Natl Acad Sci U S A* **89**, 11016-11020 (1992).

49 Uehara, S. *et al.* Complete XY gonadal dysgenesis and aspects of the SRY genotype and gonadal tumor formation. *J Hum Genet* **47**, 279-284 (2002).

50 Assumpcao, J. G. *et al.* Novel mutations affecting SRY DNA-binding activity: the HMG box N65H associated with 46,XY pure gonadal dysgenesis and the familial non-HMG box R30I associated with variable phenotypes. *J Mol Med* **80**, 782-790 (2002).

51 Mitchell, C. L. & Harley, V. R. Biochemical defects in eight SRY missense mutations causing XY gonadal dysgenesis. *Mol Genet Metab* **77**, 217-225 (2002).

52 Li, B. *et al.* Human Sex Reversal Due to Impaired Nuclear Localization of SRY. *J Biol Chem* **276**, 46480-46484 (2001).

53 Uehara, S. *et al.* SRY mutation and tumor formation on the gonads of XY pure gonadal dysgenesis patients. *Cancer Genet Cytogenet* **113**, 78-84 (1999).

54 Kohler, B. *et al.* Five novel mutations in steroidogenic factor 1 (SF1, NR5A1) in 46,XY patients with severe underandrogenization but without adrenal insufficiency. *Hum Mutat* **29**, 59-64,:(2008).

55 Capel, B. Sex in the 90s: SRY and the switch to the male pathway. *Annu Rev Physiol* **60**, 497-523 (1998).

56 Saylam, K. & Simon, P. WT1 gene mutation responsible for male sex reversal and renal failure: the Frasier syndrome. *Eur J Obstet Gynecol Reprod Biol* **110**, 111-113 (2003).

57 Domenice, S. *et al.* Mutations in the SRY, DAX1, SF1 and WNT4 genes in Brazilian sex-reversed patients. *Braz J Med Biol Res* **37**, 145-150 (2004).

58 White, S. *et al.* Copy Number Variation in Patients with Disorders of Sex Development Due to 46,XY Gonadal Dysgenesis. *PLoS One* **6**, e17793, (2011).

59 Tannour-Louet, M. *et al.* Identification of de novo copy number variants associated with human disorders of sexual development. *PLoS One* **5**, e15392, (2010).

60 Jordan, B. K., Jain, M., Natarajan, S., Frasier, S. D. & Vilain, E. Familial mutation in the testis-determining gene SRY shared by an XY female and her normal father. *J Clin Endocrinol Metab* **87**, 3428-3432 (2002).

61 Jawaheer, D. *et al.* Mapping a gene for 46,XY gonadal dysgenesis by linkage analysis. *Clin Genet* **63**, 530-535 (2003).

62 Sarafoglou, K. & Ostrer, H. Clinical review 111: familial sex reversal: a review. *J Clin Endocrinol Metab* **85**, 483-493 (2000).

63 Bilbao, J. R., Loridan, L. & Castano, L. A novel postzygotic nonsense mutation in SRY in familial XY gonadal dysgenesis. *Hum Genet* **97**, 537-539 (1996).

64 Hines, R. S. *et al.* Paternal somatic and germ-line mosaicism for a sex-determining region on Y (SRY) missense mutation leading to recurrent 46,XY sex reversal. *Fertil Steril* **67**, 675-679 (1997).

65 Migeon, C. J. & Wisniewski, A. B. Human sex differentiation and its abnormalities. *Best Pract Res Clin Obstet Gynaecol* **17**, 1-18 (2003).

66 Knudtzon, J. & Aarskog, D. 45,X/46,XY mosaicism. A clinical review and report of ten cases. *Eur J Pediatr* **146**, 266-271 (1987).

67 Reddy, K. S. & Sulcova, V. Pathogenetics of 45,X/46,XY gonadal mosaicism. *Cytogenet Cell Genet* **82**, 52-57 (1998).

68 Rosenberg, C., Frota-Pessoa, O., Vianna-Morgante, A. M. & Chu, T. H. Phenotypic spectrum of 45,X/46,XY individuals. *Am J Med Genet* **27**, 553-559 (1987).

69 Tajima, T., Fujieda, K., Kouda, N., Nakae, J. & Miller, W. L. Heterozygous mutation in the cholesterol side chain cleavage enzyme (p450scc) gene in a patient with 46,XY sex reversal and adrenal insufficiency. *J Clin Endocrinol Metab* **86**, 3820-3825 (2001).

70 Katsumata, N. *et al.* Compound heterozygous mutations in the cholesterol side-chain cleavage enzyme gene (CYP11A) cause congenital adrenal insufficiency in humans. *J Clin Endocrinol Metab* **87**, 3808-3813 (2002).

71 Degenhart, H. J., Visser, H. K., Boon, H. & O'Doherty, N. J. Evidence for deficient 20 -cholesterol-hydroxylase activity in adrenal tissue of a patient with lipoid adrenal hyperplasia. *Acta Endocrinol (Copenh)* **71**, 512-518 (1972).

72 Koizumi, S., Kyoya, S., Miyawaki, T. M., Kidani, H. & Funabashi, T. Cholesterol side-chain cleavage enzyme activity and cytochrome P-450 content in adrenal mitochondria of a patient with congenital lipoid adrenal hyperplasia (Prader disease). *Clin Chim Acta* **77**, 301-306 (1977).

73 Stocco, D. M. Clinical disorders associated with abnormal cholesterol transport: mutations in the steroidogenic acute regulatory protein. *Mol Cell Endocrinol* **191**, 19-25 (2002).

74 Fujieda, K. *et al.* Spontaneous puberty in 46,XX subjects with congenital lipoid adrenal hyperplasia. Ovarian steroidogenesis is spared to some extent despite inactivating

mutations in the steroidogenic acute regulatory protein (StAR) gene. *J Clin Invest* **99**, 1265-1271 (1997).

75 Fujieda, K. *et al.* Molecular pathogenesis of lipoid adrenal hyperplasia and adrenal hypoplasia congenita. *J Steroid Biochem Mol Biol* **85**, 483-489 (2003).

76 Scaroni, C., Biason, A., Carpene, G., Opocher, G. & Mantero, F. 17-alpha-hydroxylase deficiency in three siblings: short- and long-term studies. *J Endocrinol Invest* **14**, 99-108 (1991).

77 Geller, D. H., Auchus, R. J., Mendonca, B. B. & Miller, W. L. The genetic and functional basis of isolated 17,20-lyase deficiency. *Nat Genet* **17**, 201-205 (1997).

78 Peterson, R. E., Imperato-McGinley, J., Gautier, T. & Shackleton, C. Male pseudohermaphroditism due to multiple defects in steroid-biosynthetic microsomal mixed-function oxidases. A new variant of congenital adrenal hyperplasia. *N Engl J Med* **313**, 1182-1191 (1985).

79 Shackleton, C., Marcos, J., Arlt, W. & Hauffa, B. P. Prenatal diagnosis of P450 oxidoreductase deficiency (ORD): a disorder causing low pregnancy estriol, maternal and fetal virilization, and the Antley-Bixler syndrome phenotype. *Am J Med Genet* **129A**, 105-112 (2004).

80 Arlt, W. *et al.* Congenital adrenal hyperplasia caused by mutant P450 oxidoreductase and human androgen synthesis: analytical study. *Lancet* **363**, 2128-2135 (2004).

81 Scott, R. R. & Miller, W. L. Genetic and Clinical Features of P450 Oxidoreductase Deficiency. *Horm Res* **69**, 266-275, (2008).

82 Mains, L. M. *et al.* 17beta-hydroxysteroid dehydrogenase 3 deficiency in a male pseudohermaphrodite. *Fertil Steril* **89**, 228 e213-227, (2008).

83 Imperato-McGinley, J., Peterson, R. E., Stoller, R. & Goodwin, W. E. Male pseudohermaphroditism secondary to 17 beta-hydroxysteroid dehydrogenase deficiency: gender role change with puberty. *J Clin Endocrinol Metab* **49**, 391-395 (1979).

84 Wilson, J. D. Androgens, androgen receptors, and male gender role behavior. *Horm Behav* **40**, 358-366 (2001).

85 Peterson, R. E., Imperato-McGinley, J., Gautier, T. & Sturla, E. Male pseudohermaphroditism due to steroid 5-alpha-reductase deficiency. *Am J Med* **62**, 170-191 (1977).

86 Imperato-McGinley, J. *et al.* Prostate visualization studies in males homozygous and heterozygous for 5 alpha-reductase deficiency. *J Clin Endocrinol Metab* **75**, 1022-1026 (1992).

87 Imperato-McGinley, J. 5alpha-reductase-2 deficiency and complete androgen insensitivity: lessons from nature. *Adv Exp Med Biol* **511**, 121-131; discussion 131-124 (2002).

88 Saenger, P. *et al.* Prepubertal diagnosis of steroid 5 alpha-reductase deficiency. *J Clin Endocrinol Metab* **46**, 627-634 (1978).

89 Canto, P. *et al.* Mutations of the 5 alpha-reductase type 2 gene in eight Mexican patients from six different pedigrees with 5 alpha-reductase-2 deficiency. *Clin Endocrinol (Oxf)* **46**, 155-160 (1997).

90 Imperato-McGinley, J., Peterson, R. E., Gautier, T. & Sturla, E. Androgens and the evolution of male-gender identity among male pseudohermaphrodites with 5alpha-reductase deficiency. *N Engl J Med* **300**, 1233-1237 (1979).

91 Mendonca, B. B. *et al.* Male pseudohermaphroditism due to steroid 5alpha-reductase 2 deficiency. Diagnosis, psychological evaluation, and management. *Medicine (Baltimore)* **75**, 64-76 (1996).

92 Galli-Tsinopoulou, A., Hiort, O., Schuster, T., Messer, G. & Kuhnle, U. A novel point mutation in the hormone binding domain of the androgen receptor associated with partial and minimal androgen insensitivity syndrome. *J Pediatr Endocrinol Metab* **16**, 149-154 (2003).

93 Pinsky, L. *et al.* Human minimal androgen insensitivity with normal dihydrotestosterone-binding capacity in cultured genital skin fibroblasts: evidence for an androgen-selective qualitative abnormality of the receptor. *Am J Hum Genet* **36**, 965-978 (1984).

94 Bevan, C. L., Hughes, I. A. & Patterson, M. N. Wide variation in androgen receptor dysfunction in complete androgen insensitivity syndrome. *J Steroid Biochem Mol Biol* **61**, 19-26 (1997).

95 Rutgers, J. L. & Scully, R. E. The androgen insensitivity syndrome (testicular feminization): a clinicopathologic study of 43 cases. *Int J Gynecol Pathol* **10**, 126-144 (1991).

96 Mazur, T. Gender dysphoria and gender change in androgen insensitivity or micropenis. *Arch Sex Behav* **34**, 411-421, (2005).

97 Sultan, C. *et al.* Disorders of androgen action. *Semin Reprod Med* **20**, 217-228 (2002).

98 Gottlieb, B., Beitel, L. K., Wu, J. H. & Trifiro, M. The androgen receptor gene mutations database (ARDB): 2004 update. *Hum Mutat* **23**, 527-533 (2004).

99 Brinkmann, A. O. Molecular basis of androgen insensitivity. *Mol Cell Endocrinol* **179**, 105-109 (2001).

100 Brown, T. R. & Migeon, C. J. Cultured human skin fibroblasts: a model for the study of androgen action. *Mol Cell Biochem* **36**, 3-22 (1981).

101 Ris-Stalpers, C. *et al.* A practical approach to the detection of androgen receptor gene mutations and pedigree analysis in families with x-linked androgen insensitivity. *Pediatr Res* **36**, 227-234 (1994).

102 Richter-Unruh, A. *et al.* Leydig cell hypoplasia: cases with new mutations, new polymorphisms and cases without mutations in the luteinizing hormone receptor gene. *Clin Endocrinol (Oxf)* **56**, 103-112 (2002).

103 Wu, S. M. & Chan, W. Y. Male pseudohermaphroditism due to inactivating luteinizing hormone receptor mutations. *Arch Med Res* **30**, 495-500 (1999).

104 Martens, J. W. *et al.* A homozygous mutation in the luteinizing hormone receptor causes partial Leydig cell hypoplasia: correlation between receptor activity and phenotype. *Mol Endocrinol* **12**, 775-784 (1998).

105 Gromoll, J., Eiholzer, U., Nieschlag, E. & Simoni, M. Male hypogonadism caused by homozygous deletion of exon 10 of the luteinizing hormone (LH) receptor: differential action of human chorionic gonadotropin and LH. *J Clin Endocrinol Metab* **85**, 2281-2286 (2000).

106 Latronico, A. C. *et al.* Brief report: testicular and ovarian resistance to luteinizing hormone caused by inactivating mutations of the luteinizing hormone-receptor gene. *N Engl J Med* **334**, 507-512 (1996).

107 Toledo, S. P. *et al.* An inactivating mutation of the luteinizing hormone receptor causes amenorrhea in a 46,XX female. *J Clin Endocrinol Metab* **81**, 3850-3854 (1996).

108 Latronico, A. C. *et al.* A homozygous microdeletion in helix 7 of the luteinizing hormone receptor associated with familial testicular and ovarian resistance is due to both decreased cell surface expression and impaired effector activation by the cell surface receptor. *Mol Endocrinol* **12**, 442-450 (1998).

109 Stavrou, S. S. *et al.* A novel mutation of the human luteinizing hormone receptor in 46XY and 46XX sisters. *J Clin Endocrinol Metab* **83**, 2091-2098 (1998).

110 Rey, R. A. *et al.* Evaluation of gonadal function in 107 intersex patients by means of serum antimullerian hormone measurement. *J Clin Endocrinol Metab* **84**, 627-631 (1999).

111 Krob, G., Braun, A. & Kuhnle, U. True hermaphroditism: geographical distribution, clinical findings, chromosomes and gonadal histology. *Eur J Pediatr* **153**, 2-10 (1994).

112 Queipo, G. *et al.* Molecular analysis in true hermaphroditism: demonstration of low-level hidden mosaicism for Y-derived sequences in 46,XX cases. *Hum Genet* **111**, 278-283 (2002).

113 Schiebel, K. *et al.* Abnormal XY interchange between a novel isolated protein kinase gene, PRKY, and its homologue, PRKX, accounts for one third of all (Y+)XX males and (Y-)XY females. *Hum Mol Genet* **6**, 1985-1989 (1997).

114 Kusz, K. *et al.* Incomplete masculinisation of XX subjects carrying the SRY gene on an inactive X chromosome. *J Med Genet* **36**, 452-456 (1999).

115 Zenteno-Ruiz, J. C., Kofman-Alfaro, S. & Mendez, J. P. 46,XX sex reversal. *Arch Med Res* **32**, 559-566 (2001).

116 Bouayed Abdelmoula, N. *et al.* Skewed X-chromosome inactivation pattern in SRY positive XX maleness: a case report and review of literature. *Ann Genet* **46**, 11-18 (2003).

117 Jimenez, A. L. *et al.* Partially deleted SRY gene confined to testicular tissue in a 46,XX true hermaphrodite without SRY in leukocytic DNA. *Am J Med Genet* **93**, 417-420 (2000).

118 Huang, B., Wang, S., Ning, Y., Lamb, A. N. & Bartley, J. Autosomal XX sex reversal caused by duplication of SOX9. *Am J Med Genet* **87**, 349-353 (1999).

119 Cox, J. J., Willatt, L., Homfray, T. & Woods, C. G. A SOX9 duplication and familial 46,XX developmental testicular disorder. *The New England journal of medicine* **364**, 91-93, (2011).

120 Sutton, E. *et al.* Identification of SOX3 as an XX male sex reversal gene in mice and humans. *J Clin Invest* **121**, 328-341 (2011).

121 Tomaselli, S. *et al.* Syndromic true hermaphroditism due to an R-spondin1 (RSPO1) homozygous mutation. *Hum Mutat* **29**, 220-226, (2008).

122 Boucekkine, C. *et al.* Clinical and anatomical spectrum in XX sex reversed patients. Relationship to the presence of Y specific DNA-sequences. *Clin Endocrinol (Oxf)* **40**, 733-742 (1994).

123 McElreavey, K. *et al.* A minority of 46,XX true hermaphrodites are positive for the Y-DNA sequence including SRY. *Hum Genet* **90**, 121-125 (1992).

124 Berkovitz, G. D. *et al.* The role of the sex-determining region of the Y chromosome (SRY) in the etiology of 46,XX true hermaphroditism. *Hum Genet* **88**, 411-416 (1992).

125 Damiani, D. *et al.* True hermaphroditism: clinical aspects and molecular studies in 16 cases. *Eur J Endocrinol* **136**, 201-204 (1997).

126 Abbas, N. *et al.* Familial case of 46,XX male and 46,XX true hermaphrodite associated with a paternal-derived SRY-bearing X chromosome. *C R Acad Sci III* **316**, 375-383 (1993).

127 Hadjiathanasiou, C. G. *et al.* True hermaphroditism: genetic variants and clinical management. *J Pediatr* **125**, 738-744 (1994).

128 Niu, D. M., Pan, C. C., Lin, C. Y., Hwang, B. & Chung, M. Y. Mosaic or chimera? Revisiting an old hypothesis about the cause of the 46,XX/46,XY hermaphrodite. *J Pediatr* **140**, 732-735 (2002).

129 Cui, Y. *et al.* [The mechanism of tetragametic chimerism in a true hermaphroditism with 46, XX/46 ,XY]. *Zhonghua Nan Ke Xue* **10**, 107-112 (2004).

130 Repas-Humpe, L. M. *et al.* A dispermic chimerism in a 2-year-old Caucasian boy. *Ann Hematol* **78**, 431-434 (1999).

131 Strain, L., Dean, J. C., Hamilton, M. P. & Bonthron, D. T. A true hermaphrodite chimera resulting from embryo amalgamation after in vitro fertilization. *N Engl J Med* **338**, 166-169 (1998).

132 Uehara, S. *et al.* Molecular biologic analyses of tetragametic chimerism in a true hermaphrodite with 46,XX/46,XY. *Fertil Steril* **63**, 189-192 (1995).

133 Verp, M. S. *et al.* Chimerism as the etiology of a 46,XX/46,XY fertile true hermaphrodite. *Fertil Steril* **57**, 346-349 (1992).

134 Jarrah, N. *et al.* Familial disorder of sex determination in seven individuals from three related sibships. *Eur J Pediatr* **159**, 912-918 (2000).

135 Ortenberg, J. *et al.* SRY gene expression in the ovotestes of XX true hermaphrodites. *J Urol* **167**, 1828-1831 (2002).

136 Hughes, I. A., Houk, C., Ahmed, S. F. & Lee, P. A. Consensus statement on management of intersex disorders. *J Pediatr Urol* **2**, 148-162, (2006).

137 Simpson, J. L. & Rajkovic, A. Ovarian differentiation and gonadal failure. *Am J Med Genet* **89**, 186-200 (1999).

138 Boczkowski, K. Pure gonadal dysgenesis and ovarian dysplasia in sisters. *Am J Obstet Gynecol* **106**, 626-628 (1970).

139 Portuondo, J. A., Neyro, J. L., Benito, J. A., de los Rios, A. & Barral, A. Familial 46,XX gonadal dysgenesis. *Int J Fertil* **32**, 56-58 (1987).

140 Aittomaki, K. The genetics of XX gonadal dysgenesis. *Am J Hum Genet* **54**, 844-851 (1994).

141 Namavar-Jahromi, B., Mohit, M. & Kumar, P. V. Familial dysgerminoma associated with 46, XX pure gonadal dysgenesis. *Saudi Med J* **26**, 872-874 (2005).

142 Aittomaki, K. *et al.* Mutation in the follicle-stimulating hormone receptor gene causes hereditary hypergonadotropic ovarian failure. *Cell* **82**, 959-968 (1995).

143 Lacombe, A. *et al.* Disruption of POF1B binding to nonmuscle actin filaments is associated with premature ovarian failure. *Am J Hum Genet* **79**, 113-119, (2006).

144 de la Chesnaye, E., Canto, P., Ulloa-Aguirre, A. & Mendez, J. P. No evidence of mutations in the follicle-stimulating hormone receptor gene in Mexican women with 46,XX pure gonadal dysgenesis. *Am J Med Genet* **98**, 125-128 (2001).

145 Layman, L. C., Amde, S., Cohen, D. P., Jin, M. & Xie, J. The Finnish follicle-stimulating hormone receptor gene mutation is rare in North American women with 46,XX ovarian failure. *Fertil Steril* **69**, 300-302 (1998).

146 Therman, E., Laxova, R. & Susman, B. The critical region on the human Xq. *Hum Genet* **85**, 455-461 (1990).

147 New, M. I. Inborn errors of adrenal steroidogenesis. *Mol Cell Endocrinol* **211**, 75-83 (2003).

148 New, M. I., Rapaport R. in *Pediatric Endocrinology* (ed M.A. Sperling) Ch. 12, 282-314 (W.B. Saunders Co, 1996).

149 Peter, M. Congenital adrenal hyperplasia: 11beta-hydroxylase deficiency. *Semin Reprod Med* **20**, 249-254 (2002).

Chapter 2:

Protection from C57BL/6-Y^{POS} Sex Reversal is conferred by a POSA specific enhancer of SOX9.

Introduction

Since the discovery of the testis-determining gene *SRY* in 1991[1,2], the field of genetics has rapidly evolved, allowing for identification many genes involved in testis and/or ovarian determination in mice and in humans. In 15% of 46, XY Gonadal Dysgenesis (GD) patients, point mutations in *SRY* result in decreased binding affinity of SRY to its' downstream targets, including *SOX9*[3], ultimately resulting in 46, XY female phenotype. Heterozygous mutations in one of *SRY's* downstream targets, *SOX9*, results in a severe developmental condition, campomelic dysplasia with autosomal sex reversal, in which all patients have a severe cartilage defect but only XY patients are sex reversed. Mouse models in which SOX gene family members are overexpressed in the developing gonad result in XX-males[4-7], implying that aberrant SOX gene expression in the bipotential gonad during sex determination can ultimately shift the delicate balance between ovarian and testis development towards testis development [4]. Despite our extensive knowledge about SRY and SOX9 actions, we still do not fully understand their complex regulation in mammalian sex determination.

A powerful model in mammalian sex determination is the C57BL/6J (B6) mouse with a Y-chromosome from one of the *M. domesticus* (*DOM*) substrains. Consomic mice with autosomes and X-chromosome of B6 origin and the Y^{DOM} show a range of sex determination phenotypes from full sex reversal to delayed testicular development and partially masculinized males. Interestingly, B6 males with their native Y-chromosome, Y^{B6} (*M. musculus*) develop normally as males. Specifically, the Y-chromosome from *M. domesticus poschiavinus (*Y^{POS}), a wild derived strain from the Poschiavo valley in Italy, does not result in normal testicular development in B6-XY^{POS} animals. Y^{POS} sex reversal is specific to the B6 background, as all

other tested strains, 129S1/SvIMJ(129), DBA/2J(DBA), C58/J(C58) and BALB/cBy(BALBc)[8] do not sex reverse in the presence of Y^{POS}.

This observation indicates that a complex interaction between a genetic background (B6) and a specific Y-chromosome (Y^{DOM}) perturbs normal male development. Sequencing of the male sex determination gene, *Sry*, from Y^{DOM} and Y^{MUS} demonstrated that the *M. musculus* isoform was nearly twice as large (395 amino acids) as the *M. domesticus Sry* (230aa). This isoform difference resulted from a premature termination codon located in the glutamine repeat region of the *M. domesticus Sry* [9]. However, these isoform differences did not correlate with sex-reversal when the respective Y-chromosome was placed on a B6 background. Functional studies in which mouse embryos were injected with transgenic constructs of truncating SRY mutations that eliminated the entire glutamine repeat region, showed that these mutated constructs could not induce XX-males while the full-length wildtype *Sry* causes XX-Maleness[10]. These data suggest that differences in the length of the glutamine repeat region might account for some of the speciation among mouse strains, such that a specific SRY requires a specific genetic background in order to induce normal testis development.

The interaction between genetic background and *Sry* expression is critical to initiation of testis development. The role of the genetic background likely regulates the environmental receptiveness of the bipotential gonad to organ differentiation, through influence of timing, level, and localization of sex-specific genes. Studies comparing B6-Y^{POS} and B6-Y^{B6} gonads indicate that *Sry* expression levels are delayed by two tail somites, or approximately four hours, in B6-Y^{POS} embryos[11]. Further evidence supporting this theory that *Sry*'s temporal expression is a critical step in initiating testis differentiation were studies done in a transgenic mouse with a heat-shock promoter inducible SRY. SRY's ability to up-regulate testis determination was

limited to a short six-hour window from embryonic days 11.0-11.25. Induction of *Sry* expression after this time could not activate the male pathway and resulted in ovarian development[12]. This suggests that testis determination in the B6 background is primed towards male testis determination only during a short window in development. Thus, SRY^{POS} expression does not occur in the correct time frame on the B6 background to initiate testis development. While it is clear that temporal expression patterns are important in testis determination, it is only one of many genetic factors that contribute to sex determination.

To elucidate autosomal differences between B6 and other protected strains, such as DBA, a linkage study was performed on F2 intercross mice between B6 and DBA to identify regions that contribute to B6-Y^{POS} sex reversal. This study yielded three autosomal testis-determining loci on chromosomes 2, 4, and 5[13]. This early study demonstrated that homozygosity for the B6 allele is one of the major contributing factors in B6-Y^{POS} sex reversal.

Recently, a comparative study of gene expression profiles was performed between developing testis of the B6 and 129S1 strains to identify strain-specific differences in gene expression. Their data showed that compared to 129-XY^{129} gonads, B6-XY^{B6} gonads had higher levels of expression of "female" related genes and lower expression of "male" genes[14]. Additionally, in B6-Y^{B6} mice, *Sox9* expression was elevated by nearly 2-fold, presumably as a compensation for their "pro-female" gene expression profile. This data supports the observations that XX-male transgenic mice, such as *OdSex*, which were created on the 129 or mixed backgrounds, lose their XX-male phenotype when backcrossed onto the "pro-female" B6-background. B6 has always been known as a background that is exquisitely sensitive to XY-sex reversal due to increased expression of female related genes. For the same reason B6 is protected from XX-sex reversal.

In this study, we utilize a novel congenic strain in which homozygosity for a non-B6 region confers 90% protection from B6-Y^{POS} sex reversal. The initial discovery of B6-Y^{POS} sex reversal occurred after backcrossing semi-inbred *POSA* males to the B6 background. After 5 generations, the male-to-female sex ratio was skewed in favor of females, and it was discovered that half of the female animals had an XY karyotype[15].

In order to generate our protected congenic strain, B6-Y^{POS} was used to generate a 129-Y^{POS} strain by backcrossing to 129 females and selecting for the Y^{POS}. After 13 generations, the male to female ratio was 1:1. One 129-Y^{POS} was backcrossed to B6-XX to initiate the congenic strain used in this paper. At each generation, only the most masculinized males were mated to B6 females, selecting for autosomal regions that protect from B6-Y^{POS} sex reversal. After 13 generations of backcrossing to the B6 strain, congenic strains were created with varying degrees of protection from B6-Y^{POS} sex reversal.

In a previous publication, we genetically characterized B6.129-Y^{POS} and identified two large regions on chromosome 11 of 129 origin that resulted in dose-dependent protection from B6-Y^{POS} sex reversal. The identification of this region on chromosome 11 relied 300 single nucleotide polymorphisms (SNPs) chosen because they are polymorphic between the B6 and 129 strains. Of these 300 SNPs, 38 were located on chromosome 11[16].

To continue narrowing down the genomic element important in protection from B6-Y^{POS} sex reversal, we backcrossed heterozygous B6.129-Y^{POS} males to B6 females and at each generation performed SNP genotyping to identify recombinants within the congenic region. In this study, we identified a 1.5 megabase (Mb) region that confers protection from B6-Y^{POS} sex reversal. The origin of this protective region is not from the 129 inbred strain but likely derived

from the original *POSA* strain. These results about the B6.129-Y^{POS} protection from sex reversal highlight the importance of the promoter of *Sox9* in the protection from B6-Y^{POS} sex reversal.

Materials and Methods

Mouse strains

Animals were housed according to the guidelines of UCLA's Division of Laboratory Animal Medicine. All experiments were approved by the Institutional Animal Care and Use Committees of UCLA. C57BL/6J (B6) (#000664) females and males used for subcongenic generation and for breeding were obtained from the Jackson Laboratories. The Jackson Laboratory is fully accredited by the American Association for Accreditation of Laboratory Animal Care.

Construction of the B6.129-Y^{POS} subcongenic strain:

The original B6.129-Y^{POS} strain was produced as described previously [16,17]. To generate the subcongenic strains, heterozygous B6.129-Y^{POS} protected males were crossed to B6 females (Jackson Laboratory, Maine). At each generation, offspring were genotyped with polymorphic SNP markers from the 2008 SNP scan as well as those from the Mouse Diversity Array (Table 2-1, Fig. 2-1). B6-Y^{POS} fertile animals in which a recombination occurred within the congenic region were used to generate new subcongenic lines, and the first subcongenic B6.129-Y^{POS} mouse bred to homozygosity was considered the "founder" of the line. Subcongenic founder B6.129-Y^{POS} males were bred to B6 females and all N1 offspring are genotyped to determine presence of the subcongenic region. Heterozygous subcongenic brothers and sisters were mated together to generate animals that are homozygous for the subcongenic region. All N2 offspring

were genotyped to confirm the presence of the entire congenic region and confirm that there were no recombinations within the subcongenic regions.

Classification of Adults

A mouse was be classified as 1) a **female**, if female external genitalia, yellow mammary-associated hair pigmentation, bilateral uterine horns, and normal-appearing ovaries were present; 2) a **hermaphrodite**, if ambiguous genitalia, some yellow mammary-associated hair pigmentation, and an ovary and a contralateral ovotestis or testis were present; or 3) a **male**, if normal appearing male genitalia, no yellow mammary-associated hair pigmentation, and two testes were present.

Classification of Fetuses

Midday of the day a vaginal plug is detected will be considered E0.5 and the fetal stage of development will be verified by limb morphology [18]. Individual gonads from E14.5-16 fetuses will be classified as an ovary (O), an ovotestis (OT) or a testis (T) [15]. Finally, the approximate amount of testicular tissue in each ovotestis will be estimated at the time of dissection and confirmed using a captured image. If 50% or greater testicular tissue is present, the ovotestis is designated as oT. If less than 50% testicular tissue is present, the ovotestis is designated as Ot.

Genotyping

Chromosomal sex was determined with a PCR-based assay using a 2x ImmoMix Red (Bioline, London, UK) and single primer pair, which detects the X-linked *Smc-x* gene (330 bp) and the Y-linked *Smc-y* gene (301 bp)[11]. The initial denaturation step at 98° for 10min was followed by 35 cycles of 98° for 30 sec, 57° for 30 sec, 72° for 30 sec, and the process completed by a final extension at 72° for 7 min. XX individuals thus have a single PCR band, while XY individuals have 2 bands when resolved on 2% agarose gel by electrophoresis.

Genotyping along chromosome 11 was done by PCR with 2x ImmoMix Red (Bioline, London, UK) and primers listed in Table 2.1. PCR products were purified using ExoSAP-IT (Affymetrix, USA) and Sanger sequenced at Laragen (Culver City, CA). Trace files were examined to identify the presence of homozygosity for the B6 or alternate allele, or heterozygosity for either B6 or alternate allele.

Statistical Analysis

Categorical and ordinal data analyses for the genotype-phenotype correlations was assessed using Fisher's exact tests or Likelihood ratio tests for the ordinal treatment of the data. The dose effect hypothesis will be tested using Stata's ologit and mlogit options.

Mouse Diversity Genotyping Array (Affymetrix)

Genomic DNA from 10 founder homozygous males from each subcongenic line was extracted using the Gentra Puregene Mouse Tail Kit (Qiagen) and analyzed at JAX® Mouse Diversity Genotyping Array Service (Bar Harbor, ME) according to manufacturers' protocol. SNP analysis was performed (PMC2735580) using the R MouseDivGeno software package (http://cgd.jax.org/tools/MouseDivGeno/). The MouseDivGeno package assumes three possible genotypes: allele A (B6), allele B and heterozygous, and can also declare missing (N) calls. This array contains 623,000 SNPs and 916,000 copy number variant probes. On chromosome 11, there are 28,000 SNPs in total of which 5,014 SNPs are known to be polymorphic between the B6 and 129 strains. The mean distance between informative SNPs is 24kb [19].

CNV analysis was performed using the SimpleCNV function within the R MouseDivGeno package. The SimpleCNV function requires genotyping data derived from the SNP genotyping process; therefore, a subset of 249 high-quality arrays generated with the Diversity Array and .CEL files generated from this experiment were used for normalization of

our data. The quality of the CNV output is improved with analysis of additional high-quality array data.

Immunofluorescence

The time course of *Sox9* expression was performed by whole mount immunofluorescence (WM-IF) following the experimental design in [20] and using standard WM-IF protocols. For these experiments, the developmental stage of embryos will be determined by tail somites [21,22]. We examined B6-Y^{B6}, B6-Y^{POS} and B6.129-Y^{POS} XY gonads at 17/18, 19/20, 21/22, and 23/24 tail somites, roughly corresponding to E11.75 to 12.5. IHCs against, Sox9 (H-90, Santa Cruz Biotechnology, sc-20095), PE-CAM (CD-31, BD Pharminigin, #550274), and will be performed according to standard protocols.

RNA Expression Analysis

We examined B6-XX, B6-Y^{B6}, B6-XY^{POS}, and B6.129-Y^{POS} gonads at 17/18, 19/20, and 21/22 tail somites, roughly corresponding to E11.75 to 12.5. All gonads will be preserved in Trizol and similar sex and age gonads will be pooled and extracted for mRNA. Real-time PCR will be performed to assess the presence of these gene products and splice isoforms within the developing gonads.

Database use

The NCBI http:/www.ncbi.nlm.nih.gov/, UCSC Genome Bioinformatics http:/genome.ucsc.edu/ and the Mouse Genome Informatics http:/www.informatics.jax.org/ sites were used in retrieving gene, marker, and sequence information. The genomic location information used is based on Mouse Build 37, July 2007.

Genomic DNA Capture and Sequencing of the 110 Region

Capture of the 110 region was performed using Agilent custom designed 244K CGH array. Oligonucleotide baits were tiled every 27 base pairs on all regions spanning 110-115 Mb on chromosome 11, except for repeat regions. In total, 3.2 Mb was covered by our array. BED files of genomic locations of all cRNA oligonucleotide probes are available upon email request to the authors. A genomic DNA library was created for each sample following manufacturers' protocol (Illumina protocol Preparing Samples for Sequencing Genomic DNA (p/n 11251892 Rev. A) and Agilent Technologies SureSelect Target Enrichment System for Illumina Single-End Sequencing Library) and ligated to custom bar-coded adaptors internally validated. After the PCR amplification, up to 7 bar-coded libraries were pooled to be captured together by the in-solution hybridization custom-designed 120-bp cRNA oligonucleotide baits [4] according to manufacturer's protocol (Agilent Technologies). After the capture, the baits were washed, eluted, amplified and diluted to 10nM final concentration based on the Qubit concentration measure and Agilent Bioanalyzer size measure. Single end sequencing was performed at the UCLA Genomic Sequencing Center (GSC) on the Illumina Genome Analyzer II for 76x cycles following manufacturer's protocol. Each pool of captured library was sequenced on one lane of a flowcell. The base-calling was performed by the real time analysis (RTA) software provided by Illumina.

Sequence Data Analysis

Alignment

The Illumina output files (.qseq) were converted to fastq formats using BFAST (http://bfast.sourceforge.net) [5, 6] script ill2fastq.pl that strips the barcodes in the beginning of the read and add to the read name. The fastq files were then parsed into multiple fastq files, each for one unique barcode. Only the reads with bar-code sequences matching the original bar-codes 100% were carried over to the second set of fastq files. The sequence reads in each fastq file

were aligned to the mouse reference genome (mm9, July 2007, build 37) using Novoalign from Novocraft Short Read Alignment Package (http://www.novocraft.com/index.html). Mouse Genome reference sequence (mm9), downloaded from the UCSC genome database located at http://genome.ucsc.edu and mirrored locally, was indexed using novoindex program (–k 14 –s 3). The output format was set to SAM and default settings were used for all options. Using SAMtools (http://samtools.sourceforge.net/), the SAM file of each sample was converted to BAM file, sorted and merged and potential PCR duplicates were removed using Picard (http://picard.sourceforge.net/). [7]

Variant Calling

The variants, both single nucleotide variants (SNVs) and small INDELs (insertions and deletions), within the captured coding exonic intervals were called using SAMtools pileup tool. Here, we used processed BAM files and called the SNVs and indels separately so that the likelihood of using spuriously mismapped reads to call the variants was reduced: for SNV calling, the last 5 bases were trimmed and only the reads lacking indels were retained and for INDEL calling, only the reads that contained one contiguous indel, not occurring on either end of the read, were retained. [8] The variants were further annotated using the SeqWare project (http://seqware.sourceforge.net) and loaded into the SeqWare QueryEngine database [9]. Variants from each sample with the following criteria were identified: (i) variant base or in/del observed at least twice and at ≥5% of the total coverage per base, (ii) variant observed at least once on both forward and reverse strands, (iii) SNP quality score ≥10. To retrieve the depth of coverage information of each base, we generated a PILEUP file for each sample using SAMtools and calculated the average coverage per capture interval using a custom script.

Results

Generation and Analysis of Subcongenic Lines

The original B6.129-Y^{POS} congenic strain identified three major congenic regions on chromosome 11: the first spanning 1-14,267,131 base pairs (bp), the second 58,783,266-71,981,592 bp, and the third region spanning 76,380,940-94,191,519 bp[16]. Identification of these three regions relied on a small set of polymorphic SNPs located within and around the boundaries (see Table 2-1 and Fig. 2-1). We generated seven subcongenic lines with different sub-regions of the original congenic region identified in [16] (Fig. 2-1).

Previous studies had demonstrated our B6-Y^{POS} animals developed as either XY-females or XY-hermaphrodites with sufficient testicular tissue to continue breeding. Fifty percent of the B6-Y^{POS} animals were phenotypically female or hermaphrodites, while the other half developed as phenotypic males. However, in our genotype and phenotype studies, we observed several interesting occurrences within the subcongenic lines. First, several lines were significantly protected compared to the B6-Y^{POS} control group initially, but appeared to drift over time. Second, in some males in which the 2008 SNP genotyping indicated that the congenic regions were identical, we found that offspring from one male were protected from sex reversal, while offspring from the other male were not protected from sex reversal. These results implied that although two B6.129-Y^{POS} males appeared to have the same genotype, they were passing on these "protective" alleles to their offspring at different rates, therefore causing different rates of protection from B6-Y^{POS} sex reversal.

As we continued to phenotype animals from each of the subcongenic lines, we hypothesized several scenarios that could account for these observances.

1.) A second region outside of chromosome 11 that is inherited independently of our subcongenic regions, which are conferring this "change" in protection from sex reversal.

2.) There was a *de novo* event, which occurred, undetected in one of our founder subcongenic males and resulted in an increased or decreased protection from XY sex reversal in all this particular male's offspring.

3.) The resolution of our SNP genotyping was insufficient to detect all recombination's and therefore the differences within the subcongenic regions may not be distinguished by our SNP genotyping methods.

4.) There was another region on chromosome 11 that was missed in the original SNP scan. Deciphering between these possibilities requires a high-density SNP array that included SNP and CNV probes that were not biased towards SNPs polymorphic between B6 and 129.

Genome scans

Ten Y^{POS} males from a the subcongenic and B6-Y^{POS} lines were analyzed using the Affymetrix Mouse SNP Diversity Array developed at the Jackson Labs[23]. For these arrays, the data was of high quality, with genotype call rates of 97% and higher. For each of the subcongenic B6.129-Y^{POS} lines and in two B6-Y^{POS} animals, DNA was analyzed to determine which regions had SNPs of B6 origin and which regions had SNPs of non-B6 origin. This would allow us to identify congenic regions that may not have been identified in the original SNP scan due to a paucity of appropriate SNP markers.

In the whole genome scan we did not find any other regions with significant numbers of non-B6 SNPs (data not shown). The congenic regions on chromosome 11 had a large

concentration of non-B6-SNPs/Mb (>100), while all other regions had significantly smaller concentrations (< 20) of non-B6 SNPs.

Upon inspection of chromosome 11 we established the boundaries of our congenic regions. For the majority of the lines, the boundaries established by our previous genotyping methods matched those identified in the dense genotyping array (Fig. 2-2). All expected congenic regions were present in all the lines. Within line 5, there appeared to be a recombination at the telomeric end of the congenic region leaving 1 Mb heterozygous while the rest of the congenic region as homozygous.

Additionally, our analysis identified two previously unknown regions of that were not of B6 origin, on chromosome 11 from 20 to 24Mb and from 110 to 114.5Mb (110 region). The 110 region was present in all seven subcongenic lines, either in a heterozygous or homozygous state. Among these animals, there were two animals, Line 6A and Line 7, that exhibited further recombination within the 110 congenic region. Most interestingly, we also found B6-Y^{POS} animals were heterozygous for the 110 congenic region. The congenic region spanning 20-24Mb was identified in five of the seven founder males in the heterozygous or homozygous state but this region was not present in the B6-Y^{POS} males (Fig. 2-2).

This new data allowed for the formation of a new hypothesis, in which heterozygosity for the 110 region was required for the protection from B6-Y^{POS} sex reversal. Given that our rates of B6-Y^{POS} females and hermaphrodites were lower than previously reported in the literature[15] it was likely that the original group classified as B6-Y^{POS} animals in[16], contains two genotypes: homozygous B6-Y^{POS} or heterozygous B6/110h-Y^{POS}.

These two new congenic regions forced us to revisit our genotypic data, and new SNPs (Table 2-1) that encompassed the boundaries of these new regions were genotyped. This

allowed us to determine whether the presence of absence of these new regions had a direct effect on protection from B6-Y^{POS} sex reversal. Upon further genotyping, we determined that we actually possessed a total of 14 lines when the presence of the 110 region was taken into account. A revised list of subcongenic lines, incorporating the 110 region is shown in Fig. 2-3. All future references to lines in this paper will be from this Fig. 2-3.

Copy Number Variation (CNV) Analysis

CNV analysis was performed on the genotyped founders from Lines 3, 4, 6B and 7. Results from the CNV analysis identified between 5 and 9 CNVs in each of these samples. These CNVs were classified as either a gain or a loss, but the number of copies was not quantitatively determined. Four of the CNVs recurred in most of the samples and had the same boundaries. These consisted of CNVs encompassing the variable antibody region on chromosome 12, the histocompatibility locus, erythropoietin-4 immediate early response gene, the intron of Neuroligin-1, and Lrmp. The locations of other CNVs identified are listed in Table 2-2. Only two CNVs were identified on chromosome 11, in the Line 5 homozygous male. None of the CNVs were associated with protection from B6-Y^{POS} sex reversal.

Association of the Subcongenic Lines and Sex Reversal in Adults

The previous study showed that the original congenic region conferred a dose dependent increase in protection from B6-Y^{POS} sex reversal. We then constructed subcongenic lines to isolate the genomic region(s) responsible for protection from B6-XY^{POS} sex reversal. Genotype-phenotype correlations were performed for each of the subcongenic lines.

We analyzed the congenic regions spanning 20-24Mb in each of the lines, and the presence or absence of this region did not have a significant effect on protection from sex reversal (data not shown). The association between each line's genotype (heterozygous or homozygous) and phenotypes: XY^{POS}-female, XY^{POS}-hermaphrodite, and XY^{POS}-male are illustrated in Fig. 2-4. Each line is also analyzed based on the presence or absence of the 110 region.

Compared to the B6-Y^{POS} animals, animals that are either heterozygous or homozygous for the 110 region (regardless of other congenic regions) are significantly protected from B6-Y^{POS} in a dose dependent fashion. Heterozygosity(h) for the 110 region results in 80% protection from B6-Y^{POS} sex reversal and homozygosity(H) for the 110 region results in 100% protection from sex reversal. These levels of protection were not substantially changed in the presence of other sub-congenic regions (see lines 1-110h, 2-110h, 2-110H, 3-110h, 3-110H, 4-110H, 5-110h, 5-110H, 6-110h, 6-110H, 7-110h, and 7-110H in Fig. 2-4). Subcongenic Lines 1h, 1H, 3h, 3H, 4h, 6h, 7h, and 7H, did not contain the 110 region and were not significantly protected from B6-Y^{POS} sex reversal (Fig. 2-4). This lack of protection indicates that the originally identified congenic regions (as depicted in Fig. 2-1) were not contributing to the protection from sex reversal phenotype. The 110 region is the smallest subcongenic region which confers full protection from B6-Y^{POS} sex reversal. Only lines containing at least one copy of the 110 region had a significant number of XY^{POS}-males, which demonstrates the large effect this genomic element plays in normal testis development and fertility.

Association of the Subcongenic Lines and Sex Reversal in Embryos

Embryonic gonads were classified into four phenotypic categories and analyzed between embryonic day 14 and 15.5. The four phenotypic categories are: normal ovary(O), normal testis(T), and a range of ovotestis. The distributions of embryonic gonadal phenotype were compared based on the genotypes in seven lines in Fig. 2-5. The embryonic timepoint provides a more accurate picture as to the rates of sex reversal in our animals, allowing us to assess the degree of testis formation in early testis development. This data showed that B6-Y^{POS} animals had over 80% XY^{POS} ovaries during embryonic development. Heterozygosity for the 110 region showed a dramatic decrease in the number of XY^{POS} ovaries, with 98% of all animals displaying some degree of testis formation. Furthermore, 110 region homozygosity results visually normal testes in a subset of XY^{POS} embryos.

Lines that did not contain the 110 region, Lines 1h, 1H, 3h, 3H, 4h, 6h, and 7h did not show any significant development of testis tissue. In each of the above lines, only 0-20% of XY^{POS} gonads developed even some degree of testicular tissue, which never amounted to more than 50% of the gonad. There were no significant differences between heterozygosity and homozygosity for these non-110 lines, confirming that these regions do not confer any significant protection from B6-Y^{POS} sex reversal.

The embryonic and adult genotype and phenotype correlation narrowed the critical region for protection from B6-Y^{POS} sex reversal to the novel region spanning 110-114.5Mb on chromosome 11.

Narrowing the protective 110 region

The 110 region consists of 4.5 Mb, which encompasses 15 genes. However, this region contains a well-described male sex determination gene, *Sox9* and the entirety of its 1.5 Mb promoter region. In addition to *Sox9*, another putative gene that may play a role in sex development is *Map2k6*, which is at the centromeric end of the congenic region. In order to further narrow down the region we identified a XY^{POS} mouse that had a recombination within the *Sox9* region. This recombination created a narrowed congenic region, which we named sub-110 #1 (s110-1), spanning 110-112.5 Mb, with a telomeric boundary 120kb upstream of *Sox9* (Fig. 2-2, Line 6A). A second line6-110h-XY^{POS} mouse was identified in our genotyping s to also have a recombination within the 110 region, giving rise to a second sub-110 region #2 (s110-2) region spanning 111Mb-114Mb.

To determine if either sub-110 region was sufficient to confer protection we performed a phenotypic embryonic analysis on matings from the s110-1 or s110-2 XY^{POS} animals. The offspring had two distinct genotypes: heterozygosity for the Line 6 region only or heterozygous for Line 6 and one of the sub-110 regions. In the absence of 110 region, the Line 6 congenic region does not protect from B6-Y^{POS} sex reversal (Figs. 2-4, 2-5, 2-6). Therefore, any protection phenotype identified in the 6-s110-1h or 6-s110-2h heterozygous animals is due to the sub-110h region, and not the attached subcongenic region. Embryonic phenotype analysis showed that both of the sub-110 regions could drive formation of testicular tissue *in* XY^{POS} animals resulting in fertile XY^{POS} males (Fig. 2-6). This data further narrowed the smallest congenic region required for protection from sex reversal in half, to 1.5 Mb region, including the *Sox9* promoter and enhancer region, but not *Sox9* or the Testis specific enhancer of $Sox9^{3}$.

Next Generation Sequencing of sub-110 region

We confirmed the boundaries of the s110-1 region using targeted next generation sequencing for the entire 110 region. Genomic DNA from a 129S1/SvimJ mouse and the founder Line 6A-Y^{POS} were used to create single-end genomic DNA libraries with custom bar-coded adaptors. The genomic libraries were pooled and hybridized against a custom-array with probes tiled every 27bp between 110 and 115 Mb on chromosome 11. Since we were targeting a large region, including the intronic and promoter regions, the capture specificity and coverage was low due to the large number of repetitive and difficult to map regions. However, we were still able to obtain an average of 10x coverage for the 129S1 strain and 8x coverage for Line 6A founder. Only SNPs that had at least 10x coverage were considered real, in order to decrease our risk of false positives calls. To ensure that our sequencing was accurate, we compared our sequencing of the 129S1/SvIM strain to that generated by the Sanger Sequencing database (www.sanger.ac.uk/resources/mouse/genomes/). The pattern of SNP calling indicated that the high quality SNPs called in our sequencing were true sequence polymorphisms and not sequencing errors.

For the founder of the sub-110 region, we sequenced the region from 110-115Mb. We were able to call 469 sequence variants within this region, with a coverage of 10x or greater. Using this approach, we confirmed that the boundaries for the sub-110 region identified by the SNP analysis were accurate. The boundaries identified based on the sequencing a portion of Chromosome 11, from 110,020,917 – 112,515,300. This proved that regulation at the TESCO enhancer region was not responsible for the protective phenotype in our congenic strain and that a sequence variant(s) located within the sub-110 region was responsible for modulating *Sox9* expression and protecting from B6-Y^{POS} sex reversal.

Visual analysis of the SNP distribution pattern showed that the 110 region was drastically different in the 129S1/SviM strain as compared to the Line 6A founder. We would have expected the regions to be nearly identical if the protection from sex reversal phenotype originated from the 129 strain. The only other potential origin for the protected strain's genomic elements is the original POSA semi-inbred strain, from which the Y^{POS} chromosome was originally derived. Unfortunately, this strain is no longer in existence and cannot be genotyped or sequenced to confirm this hypothesis directly.

Sox9 Protein Expression and Localization in B6-Y^{POS} and B6-Y^{B6}

Previous studies have demonstrated that *Sry* mRNA expression is delayed beyond the critical window to initiate testis sex determination in B6-Y^{POS} [11,12]. It has been hypothesized that the interaction between autosomal factors and *Sry* is responsible for B6-Y^{POS} sex reversal, and it is intriguing that the protection from B6-Y^{POS} sex reversal is in a genetic rescue of the expression of a downstream step, *Sox9*.

The importance of *Sox9* in sex determination led us to further explore the protein expression levels of Sox9 in B6-Y^{B6}, B6-Y^{POS}, 110h-Y^{POS}, and 110H-Y^{POS} embryonic gonads at 20 and 24 ts. Whole mount immunofluorescence was performed with antibodies to SOX9 and PE-CAM (CD-31) were done to assess if protein expression was delayed or altered in the B6-Y^{POS} gonads. Furthermore, we wanted to confirm that Sox9 expression was rescued in animals that were heterozygous or homozygous for the 110 region.

Compared to the wildtype B6-Y^{B6} gonads, B6-Y^{POS} gonads show decreased levels of Sox9 expression and fewer somatic cells expressing Sox9 at 20ts. By 24ts, expression of Sox9 in B6-Y^{POS} is limited to the central portion of the gonad, and there is a failure to up-regulate Sox9

at the anterior and posterior poles of the gonad. In contrast to the B6-Y^{POS}, gonads that were either heterozygous or homozygous for the 110 region displayed high number of cells expressing Sox9 throughout the gonad, extending to both the anterior and posterior poles. The expression localization and levels of Sox9 and Pe-cam in heterozygous and homozygous 110-Y^{POS} gonads are indistinguishable from those of the B6-Y^{B6} gonads (Fig. 2-8).

While the primary cause of B6-Y^{POS} sex reversal is due to a delayed activation of Sry^{POS}, the protection conferred by the 110 region rescues the sex reversal phenotype by up-regulating Sox9 expresion. Overexpression of Sox9 has been shown in both humans and mice to be both necessary and sufficient for testis formation[24,25].

RT-PCR of genes in s110h_regions

In order to confirm that the effect we were seeing was due to SOX9 and not due to one of the 5 other genes located in the sub-110 region, we isolated cDNA from E12.5 gonads B6-Y^{B6}, B6-Y^{POS}, and 110h-Y^{POS}. We did not find any significant expression of KCNJ2, Abca9, Abca5, or Abca6 (data not shown). Both *Map2k6* and *Sox9* were expressed within both the developing male and female gonad in a sexually dimorphic fashion.

Human SOX9 duplications and deletions

The SOX gene family plays a major role in male sex determination. SRY and SOX9 are both known components of the initial cascade that pushes the bipotential gonad into forming a testis. A loss of either component ultimately results in XY-Sex reversal in humans and mice. While transgenic overexpression of either SRY or SOX9 causes 46, XX Testicular DSD. More recently, the importance of the regulation of the SOX gene family has shown that aberrant

gonadal expression of sox genes at the appropriate time window and cell type is sufficient to initiate testis development[4,6,7]. Both small duplications or deletions in the promoter and enhancer regions of SOX3 result in 46, XX Testicular DSD, indicating that the regulation of developmental gene expression is not a linear process and instead requires a dynamic and complex interaction of both enhancer and suppressor elements (Fig. 2-9).

In humans, small duplications or deletions encompassing only regulatory elements of the SOX9, and not the SOX9 gene itself, cause 46, XX Testicular DSD and 46, XY Gonadal Dysgenesis. The developmental regulation of SOX9 is controlled by the large 2 Mb promoter region contains cell-type and timepoint specific enhancer elements that bind specifically to transcription factors that activate SOX9 expression. Recent studies in humans have identified deletions and duplications encompassing genetic elements within the SOX9 promoter[26,27]. These copy number variants were mapped to the syntenic region in the mouse genome and plotted against the smallest protective congenic region, spanning 111-112.5 Mb. In addition to this, we also mapped the regions for the OdSex insertion [5] and the known location for the Testis specific enhancer of SOX9.

Discussion

Using an unbiased genetic approach, we have elucidated the genetic mechanisms underlying protection from B6-Y^{POS} sex reversal in a long-standing congenic mouse model. After continued backcrossing to decrease the size of the congenic region, a high-density SNP array was used to identify all non-B6 SNPs throughout the genome. With this new data, we discovered a new 1.5 Mb region, that includes only the promoter of *Sox9,* is sufficient to confer protection from B6-Y^{POS} sex reversal. Finally, we discovered that the 110 region, was not of

129-origin as expected, but based on the congenic line's well documented history, we hypothesize that the origins of this 110 region must be from the original POSA strain.

Our finding calls shows the importance of taking an unbiased mapping approach, using relatively dense SNP mapping in order to identify all the regions contributing to the phenotype. In studies where the phenotype of interest, sex reversal, is intimately tied with fertility, maintaining a the consomic strain, B6-Y^{POS} is very difficult. Our B6-Y^{POS} line was derived from the original congenic B6.129-Y^{POS} stock and verified based on genotyping of the B6-129 SNP markers used in [16]. Although there were markers within the 110 region, none were capable of identifying the protective region because it was not of 129 genetic origin. The originally identified 129 congenic regions on chromosome 11 linked were in linkage disequilibrium with the POSA region, allowing us to identify chromosome 11 as containing the protective allele.

Heterozygosity for the 110 region has a very strong protective effect from sex reversal. Therefore, it is not surprising that in order to generate a fertile B6-Y^{POS} male that the 110 region was surreptitiously maintained through 30 years of breeding. After re-genotyping for the 110 region in our B6-Y^{POS} line, we discovered that the rates of sex reversal in our true B6-Y^{POS} lines are comparable to data from other strains who keep colonies of B6-Y^{POS}, which sees rates of embryonic sex reversal similar to ours, with 75% ovaries and 25% ovotestis during embryonic development. All animals of a pure B6-Y^{POS} background are XY^{POS} females or hermaphrodites, which, in our experience, are not fertile.

A close examination of early studies of the B6-Y^{POS} strain, indicate that the origin of the 110 region was likely present in the initial backcrosses for the generation of the B6.129 strain. Since heterozygous genomic region confers significant protection from B6-Y^{POS} sex reversal, then only when that region is present will the resulting male be fertile and able to pass on this

particular genetic element. However, the initial backcrossing described in Eicher et al 1982 only had backcrossing for 7 generations and animals from the N4-N7 backcross generation were analyzed. Of the 309 XY mice, 117 were XY females with two ovaries, and 88 were externally normal males with 2 testis, which were smaller than normal adult testis, and 103 mice with variable external genitalia (ranging from normal male to ambiguous) all of which were overt hermaphrodites with 1 or more ovotestis. However, since the backcross generations were between N4 and N7, it meant that between 1.6-12.5% of its genome was still unfixed, or heterozygous, from the original donor strain, POSA. Therefore, in the 309 mice analyzed, it's highly likely that the 110 region was present and conferring protection in many of the 28 percent of XY mice that were externally male with testis development. Since this subset of mice were externally male and had functioning testis, they were the ones used for further breeding and later studies—therefore passing along the protective genomic element, the 110 region to the B6.129 congenic strain.

In the absence of a selecting factor, fertility, the 110 region should have been removed from the congenic line after 10 generations of backcrossing. The underlying assumption that the backcrossing of Y^{POS} to the B6 background was not subject to some selection factor was not true. In all consomic and congenic lines, a major selection factor is fertility, because without it, the next generation would not exist. This inadvertent selection enabled the B6-Y^{POS} strain to maintain heterozygosity for regions that promoted male sex development and fertility.

The early use of congenic and consomic strains to understand male sex determination was initially hindered by a lack of genetic markers that could distinguish between the genetic background of multiple inbred strains. Based on our findings, other inbred "protected" strains need to be reassessed to ensure that their genomic backgrounds are close to the theoretical values.

Only in the past two years have we had the tools and power to do large scale SNP analysis on our various subcongenic lines. Both the number of SNPs and the strain derivation of the SNPS are critically important in identifying novel congenic regions that affect a phenotype. This is especially true when there is a potential for multiple genetic backgrounds, as happened in our congenic strain. Finally, this is also important when the phenotypic variability of the analyzed animals appears larger than expected, implying the existence of a region that is present in a heterozygous state. The new genomic technologies have elucidated and corrected long-held assumptions in models of mouse sex determination.

A study to identify autosomal sex determining genes was published in 1996 and exploited the known genomic differences between the B6 and DBA strains in protection from sex reversal in the presence of Y^{POS}. The F1 hybrid offspring were intercrossed to generate a segregating F2 population. However, this study had two major confounders. First, the SNP genotyping of the F2 population was performed with 241 markers, chosen because they were polymorphic for B6 and DBA. However, if there was a possibility for a third strain's genomic elements accounting for the ovo-testis formation it might not be identified due to the scarcity of markers, because these same markers were not necessarily polymorphic for B6 and POSA, or the phenotype examined was the XY-ovotestis which remains in the middle of the phenotypic spectrum between testis and ovaries and likely due to many small effect sizes rather than a few loci of large effect size. The use of hermaphrodites to perform the linkage analysis might have made the genotype-phenotype correlation more difficult to detect using a linkage analysis. Extreme phenotypes are often easier to identify using small numbers rather than intermediate phenotypes.

We had been perplexed as to why a recent study of gonad development did not find major regions within our congenic region associated with male sex determination. This study was

performed on B6 and 129S1 backgrounds, looking at gonads in early sex determination to identify novel genes and gene pathways. Their findings from this study indicate that other there are many other autosomal elements from the B6 and 129 strains are responsible for the susceptibility and protection, respectively, to B6-Y^{POS} sex reversal.

SOX9

Our study adds to the continually expanding body of literature demonstrating the critical importance of *SOX9* in mammalian sex determination. SOX9 is a transcription factor required for normal development of multiple organ systems. During development *SOX9* is actively expressed in the neural tube, cartilage, brachial arches, and testis. In the testis, the control of SOX9 expression is thought to have three major components that are important in maintaining its sexually dimorphic expression: early inducers, maintenance of *SOX9* expression, and negative regulators[28]. Early regulation of *SOX9* is primarily through *SF1* and *SRY*. Within XY *Sf1*$^{-/-}$ gonads, *SOX9* expression is abolished. *In vitro* work confirms that SF1 directly interacts with the *SOX9* TESCO enhancer. SRY is only active during a short time period of development from embryonic day 10.75-12.5 in the mouse. Within this short and critical time window *SRY* acts as a molecular switch to activate *SOX9* within Sertoli cells. Heat-shock-inducible SRY transgene have illustrated the importance of this developmental window, and if SRY is activated beyond the critical 6-hour window, SOX9 levels remain low and ovary development is initiated[12].

Once *SOX9* expression is turned on, it is maintained by a separate set of factors, including SOX9, Fibroblast growth factor 9 (FGF9), Fibroblast growth factor receptor-2 (FGFR2), and Prostaglandin D2 (PGD2). These factors maintain an auto-regulatory loop which promote Sertoli Cell development. In XX gonads, multiple genes are up-regulated that suppress

SOX9 expression in the gonad throughout embryonic development and adulthood. Interestingly, conditional knockout of FOXL2 in adult Granulosa cells results in somatic cell sex reversal [29]. In the absence of FOXL2, the Granulosa cells trans-differentiate into Sertoli cells, which creates an environment that is hostile to normal oocyte maturation. Testis and ovarian differentiation is not fixed after differentiation, but dynamic and actively maintained in its differentiated state using complex genetic and epigenetic modifications.

Recent case reports of patients with 46, XY Gonadal Dysgenesis or 46, XX Testicular DSD were found to have deletions or duplications that encompass only the SOX9 promoter and not the gene itself, implying the regulation of SOX9 expression is under precise control. These rare cases provide insight into the complex regulation of SOX9 expression within the developing embryos.

Another model of long-range regulation of SOX9 is OddSex (*Ods*) in which a dominant transgenic insertion 1 Mb upstream of SOX9 of a tyrosinase transgene resulted in micropthalmia and XX-males[5]. The mechanism was inappropriate activation of SOX9 in the XX gonad. Sox9 did not act in place of SRY, but its expression pattern and localization mirrored that of a wild-type male. The insertion of the minigene was accompanied by a 134 kb deletion, but this deletion alone did not have a sex reversal phenotype[30]. When this particular transgene was crossed onto the A/J background, there was a surprising suppression of XX sex reversal that was mapped to chromosome 18 of A/J[31]. However, the suppressive locus displayed parent-of-origin effects, such that only when the transgene was inherited from the father and the suppressive factor was inherited from the mother, did they get suppression of XX-Males[32].

The spatiotemporal regulation of expression of developmental genes is controlled by several factors: chromatin conformation, epigenetic modifications, and transcription factors that

are bound to hundreds of enhancer and repressor elements along the promoter. This regulation can occur even at very long-range enhancer elements. In a subset of patients with Pierre-Robin Sequence (PRS), which consists of micrognathia, glossoptosis, and cleft palate, microdeletions occurring 1.5Mb centromeric to the *SOX9* start site were identified as one of the genetic causes of this unique syndrome. Furthermore, a single mutation within the smallest 75-kb microdeletion, was identified in a family with PRS [33,34].

The only known testis specific enhancer of SOX9 is TESCO[3], however mutation analysis of TESCO in 16 patients with 46, XY gonadal Dysgenesis[35] did not find any mutations. This data along with finding of deletions and duplications within the promoter region leading to 46, XY Gonadal Dysgenesis or 46, XX ovotesticular or testicular disorders of sex development suggests that there are other long-range testis-specific enhancers which regulate *SOX9* testis expression. Understanding the genetic regulation of SOX9 might identify novel molecular mechanisms in sex determination. Long-range enhancer regions with large effect sizes are only recently becoming detectable by current genetic technologies and bioinformatic analyses. The implications of this new data are that subsets of patients with 46, XY Gonadal Dysgenesis or 46, XX Testicular DSD may fact have rare mutations in unidentified testis specific enhancers of SOX9.

The role of non-coding variation in the development

Non-coding regions of DNA make up 99% of the genome, implying the genetic regulation is one of the primary determinants of phenotypic diversity among humans. However, unlike coding regions of the genome, the functional effects of variation in non-coding regions are much more difficult to assess. Tools to identify functional non-coding genomic elements are

limited and typically rely on conservation of sequence between human and distantly related species. These tools rely on the assumption that a high rate of sequence conservation is an indicator of functional importance, which is not always true. Functional testing of highly conserved genetic elements shows that only 50% of them have a functional role on gene expression. Furthermore, in addition to non-coding variation, epigenetic modifications such as methylation and chromatin conformation can modify the accessibility of these genomic factors to other factors, ultimately resulting in disease specific susceptibility[36]. These epigenetic effects vary based on time, cell-type, tissue type, and are subject to environmental effects, allowing for infinite variation in expression profiles.

Genes can be grouped based on their different methods of regulation into three general categories: housekeeping, tissue-specific, and developmental-regulator genes. Of these groups, developmental genes have strict parameters that ensure their expression is limited to specific tissues and completely inactive in all other tissues. Genes that are important in developmental regulation often have very large and complicated regulatory regions, which can extend within introns of or even beyond neighboring genes. They require that levels reach a specific threshold in certain tissues and are not expressed in any other tissue. Therefore developmentally important genes such as SOX9 require multiple cis-enhancer and repressor elements to fit around the gene.

The evolution of these very specific controls of developmental gene expression has evolved through slight genomic rearrangements and mutations that confer an evolutionary advantage and are therefore maintained within the genome. The most successful developmental genes have recurrent roles in organ determination, patterning, and differentiation. Some genes may even play distinct roles in the same organ, but in specific cell types or at different timepoints during the organs' development. Given that the regulatory regions of developmental genes are

so complex, nearly all disorders involving gene misregulation are developmental disorders in which the genes involved have a very strict tissue and temporal specificity. The role of enhancer regions in developmental disorders is highlighted by the developmental disorder, pre-axial polydactyly, in which mutations in an enhancer located more than 1.5Mb upstream of the Sonic Hedgehog gene, and within an intron of a neighboring gene, can result in misregulation of sonic hedgehog expression *in utero*[37].

In this study, we have begun to unravel complex interactions between SRY, SOX9, and autosomal factors that protect from B6-Y^{POS} sex reversal in an *in vivo* model system. The role of SOX9 in sex determination is complex and actively regulated throughout life. However, even more than the important role of SOX9 regulation, this study suggests that complex interactions between two "weak" alleles, SRY^{POS} and the B6 *Sox9* promoter region could account for rare cases of sex reversal. In this scenario, it would be difficult to diagnose a single genetic variant as causative in DSD, as it would be two independent mutations, which by current diagnostic methods could not be diagnosed. However, this posits the potential for interaction between "common" variants that do not have a decipherable phenotype unless coupled with a deleterious "rare" variant. If these types of patients exists, they would be incredibly difficult, if not impossible, to provide a genetic diagnosis based on current technology and knowledge about clinical causes of disease .

In patients in whom we cannot provide a genetic diagnosis, we attribute this to novel and yet-undiscovered genes. While there are still undiscovered disease genes, the alternative hypothesis is that we have multiple genes of small effect size, or "common variants" that predispose a person to a condition that interact (or fail to interact) with a "rare" variants, therefore leading to a disease phenotype. The tools to identify these type of interactions are still

in their infancy, but will require sophisticated molecular biology and bioinformatic approaches to identify real and deleterious interactions. As the coding regions of the genome are deciphered, the importance of the non-coding region in the spectrum of phenotypic traits to disease will continue to grow. As we move into the post-genomic era, understanding the complex interaction among genes and between genes and regulatory regions will help us unravel the subtle nuances of human disease and development.

FIGURES AND TABLES

Figure 2-1: Subcongenic Lines Generated Through Breeding. In the first row we have an ideogram of Chromosome 11 and the physical distances along the chromosomes. The physical location of the SNPs (Table 1) are spread throughout the chromosome. The physical location of original congenic regions (blue) are shown compared to the B6 regions (red). The various subcongenic lines are identified below, which all have recombination within the original congenic region.

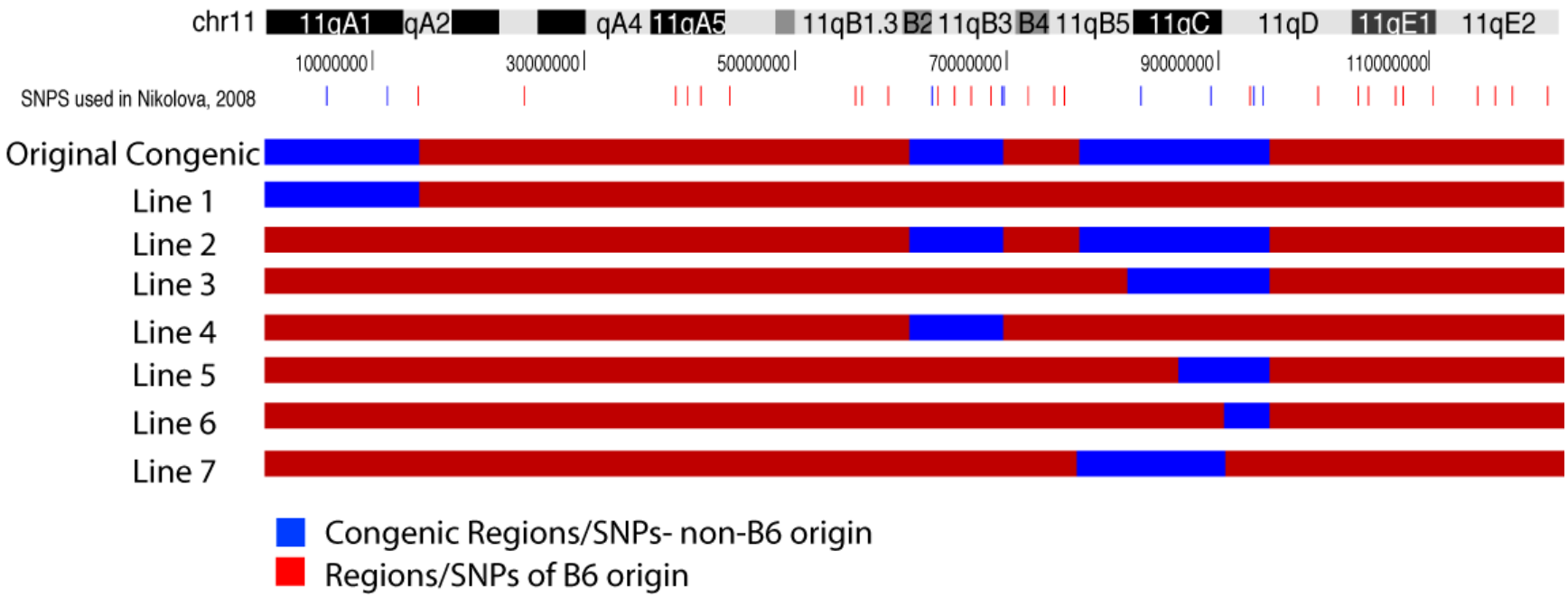

chr11
11qA1
qA2
qA4
11qA5
11qB1.3
B2
11qB3
B4
11qB5
11qC
11qD
11qE1
11qE2
10000000
30000000
50000000
70000000
90000000
110000000
SNPS used in Nikolova, 2008
Original Congenic
Line 1
Line 2
Line 3
Line 4
Line 5
Line 6
Line 7
Congenic Regions/SNPs- non-B6 origin
Regions/SNPs of B6 origin

Figure 2-2: Diversity Array identifies novel regions on chromosome 11 not identified by original SNP genotyping scan. Founder homozygous males generated for each the lines were run through a SNP Diversity Array to identify regions of non-B6 origin. For each line, the regions that were previously genotyped as homozygous for non-B6 SNPs were fully identified in our array. However, two novel regions were identified between 20-24 Mb was present in lines 1 2, 4, and 6A and 6B as heterogyzous or homozygous. A second region spanning 110-114.5 Mb, was present in all the founders and in the B6-Y^{POS} animals.

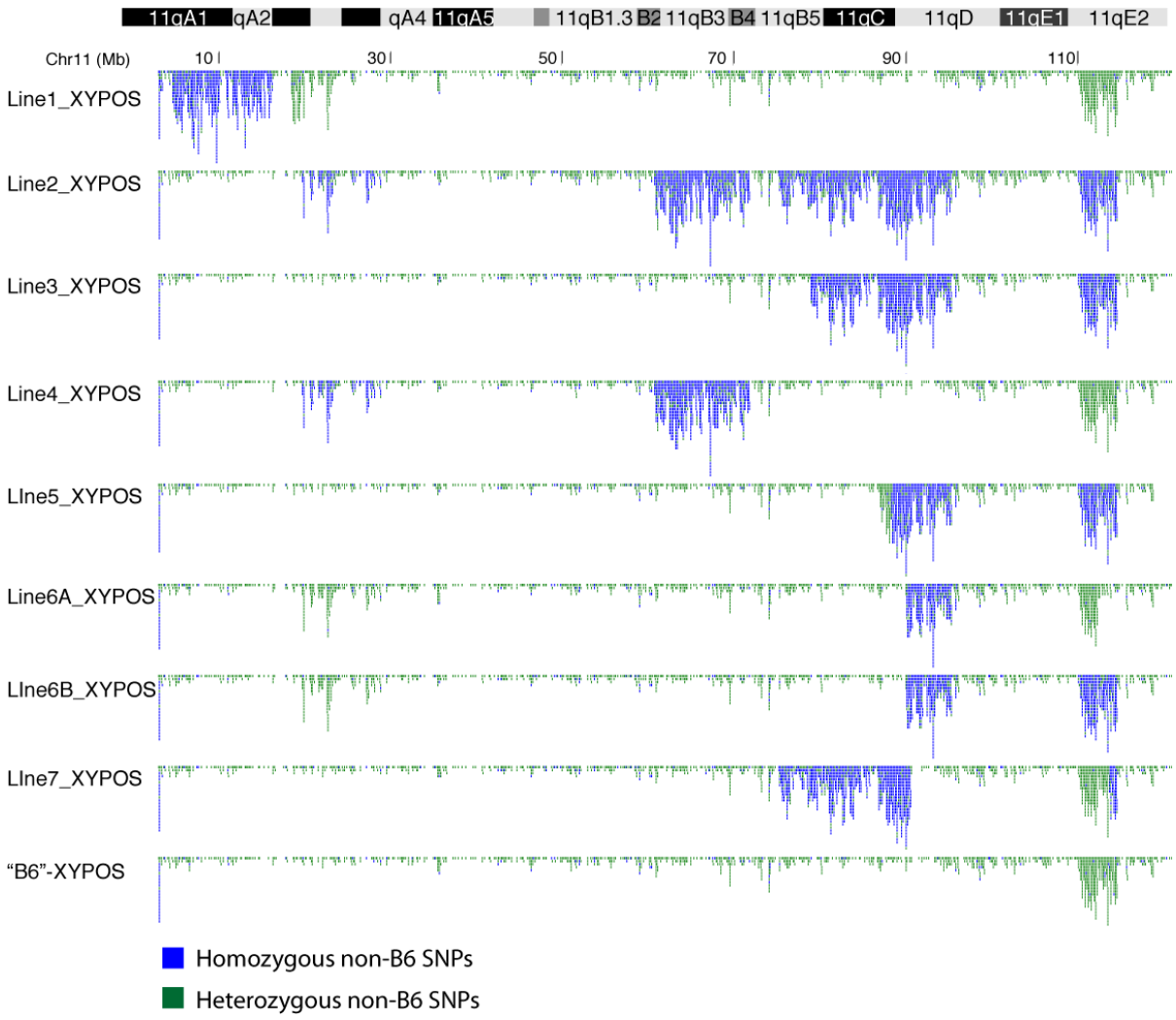

11qA1
qA2
qA4
11qA5
11qB1.3
B2
11qB3
B4
11qB5
11qC
11qD
11qE1
11qE2
Chr11 (Mb)
10
30
50
70
90
110
Line1_XYPOS
Line2_XYPOS
Line3_XYPOS
Line4_XYPOS
LIne5_XYPOS
Line6A_XYPOS
LIne6B_XYPOS
LIne7_XYPOS
“B6”-XYPOS
Homozygous non-B6 SNPs
Heterozygous non-B6 SNPs

Figure 2-3: New Set of Subcongenic Lines based on Diversity Array Results. Based on the results from the SNP Diversity Array, all analyzed males were re-genotyped to assess the presence of the two new regions (Figure 2). From these results, we had an expanded set of congenic lines in which the region from 110-115 Mb is present of absent. The Line number has remained the same, but the presence of the 110 region is indicated by -110. Furthermore, we now have a new line that only has the 110 region and a true B6-Y^{POS} animals for analysis.

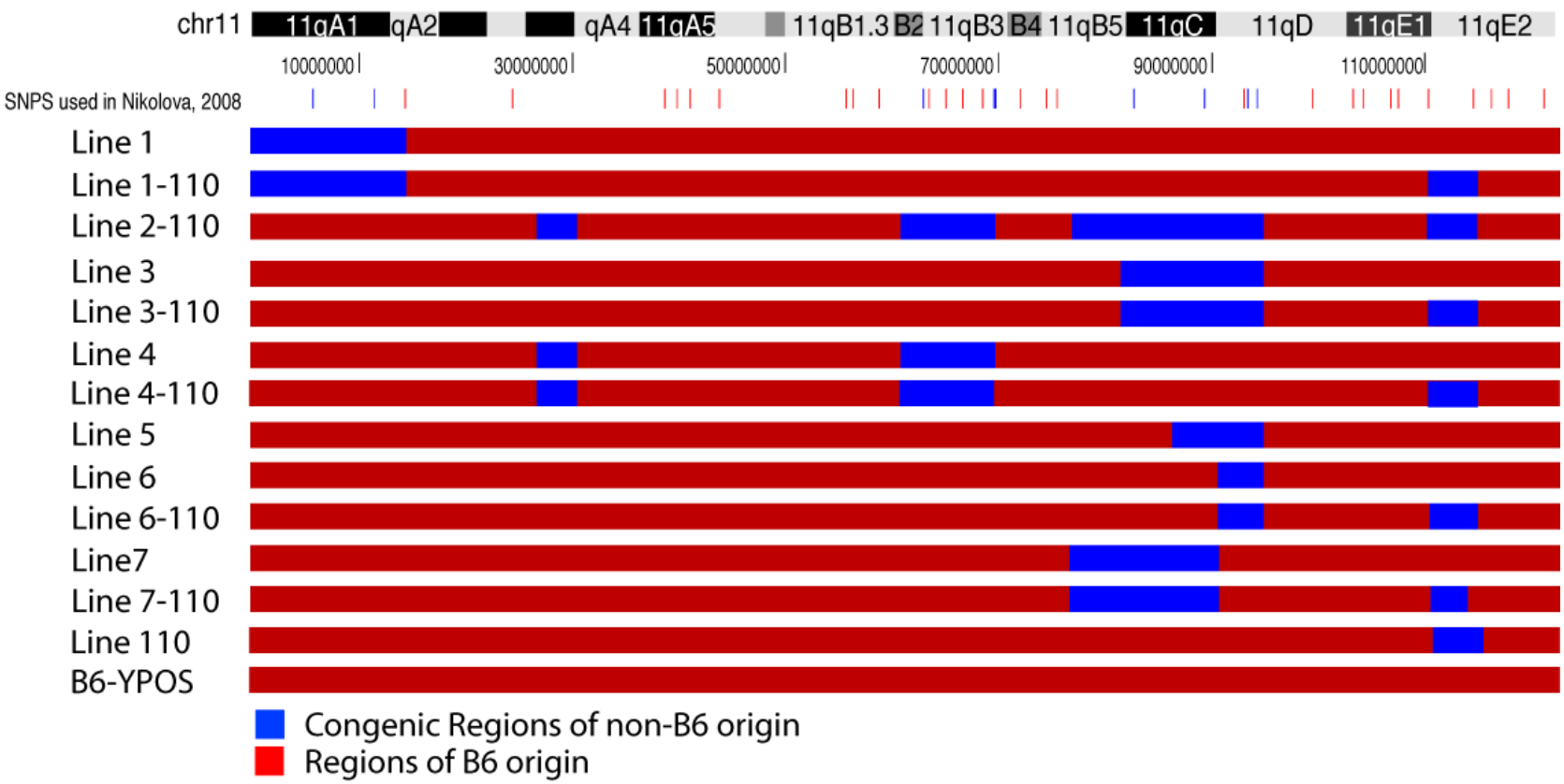

chr11
11qA1
qA2
qA4
11qA5
11qB1.3
B2
11qB3
B4
11qB5
11qC
11qD
11qE1
11qE2
10000000
30000000
50000000
70000000
90000000
110000000
SNPS used in Nikolova, 2008
Line 1
Line 1-110
Line 2-110
Line 3
Line 3-110
Line 4
Line 4-110
Line 5
Line 6
Line 6-110
Line7
Line 7-110
Line 110
B6-YPOS
Congenic Regions of non-B6 origin
Regions of B6 origin

Figure 2-4: Association of Protection from B6-Y^{POS} Sex Reversal in Adults is Dependent on the Presence of the 110 Region. All subcongenic lines were analyzed for the rates of XY-Females for each genotype. An h indicates heterozygosity for the line, while H indicated homozygosity for the line. All references are to lines in Figure 3. All lines that did not have the 110 region were not protected from B6-Y^{POS} sex reversal. However, all lines heterozygous or homozygous for the 110 region conferred significant protection ($p < 0.00001$).

Genotype-Phenotype Correlations in Adult XYPOS Mice
Percent of XYPOS animals
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
B6/B6
110h
110H
Line1h
Line1H
Line1-110h
Line2-110h
Line3h
Line3H
Line3-110h
Line4h
Line4-110h
Line5-110h
Line6h
Line6-110h
Line7h
Line7H
Line7-110h
XY Male
XY Hermaphrodite
XY Female

F**igure 2-5: Association of Protection from B6-Y^{POS} Sex Reversal in Embryonic Gonads is Dependent on the Presence of the 110 Region**. All gonads were analyzed from XY^{POS} embryos. Compared to the B6-Y^{POS} embryos, there was no protection conferred by the congenic strains in the absence of the 110 region. If the 110 region was present, either in the heterozygous or homozygous form, there was a significant increase in the formation of testicular tissue, as evidenced by more gonads classified as T and oT. h: heterozygous, H: homozygous

Genotype-Phenotype Correlations in Embryonic Gonads of XYPOS Mice
Percent of XYPOS gonads
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
B6/B6
s110h
s110h
line1
line1
line1-110h
line2-110h
line2-110h
line3
line3
line3-110h
line3-110h
line4
line4-110h
lines-110h
lines-110h
line6
line6-110h
line6-110h
line7
line7-110h
Testis
ovoTestis T>O
Ovotestis T<O
Ovaries

Figure 2-6: Genotype-Phenotype correlation with the narrowed sub-110 region.

Heterozygosity for a subset of the 110 region (s110h), spanning 2.5 Mb from Chr11: 110-112.5 Mb, confers significant protection from B6-Y^{POS} sex reversal. This small region allows for the formation of normal XY^{POS} testis and none of the gonads form XY^{POS}-Ovaries. Since Line 6h alone cannot confer protection, the s110h1 and s110h2 region contain the smallest region required for protection against B6-Y^{POS} sex reversal. h: heterozygous, H: homozygous

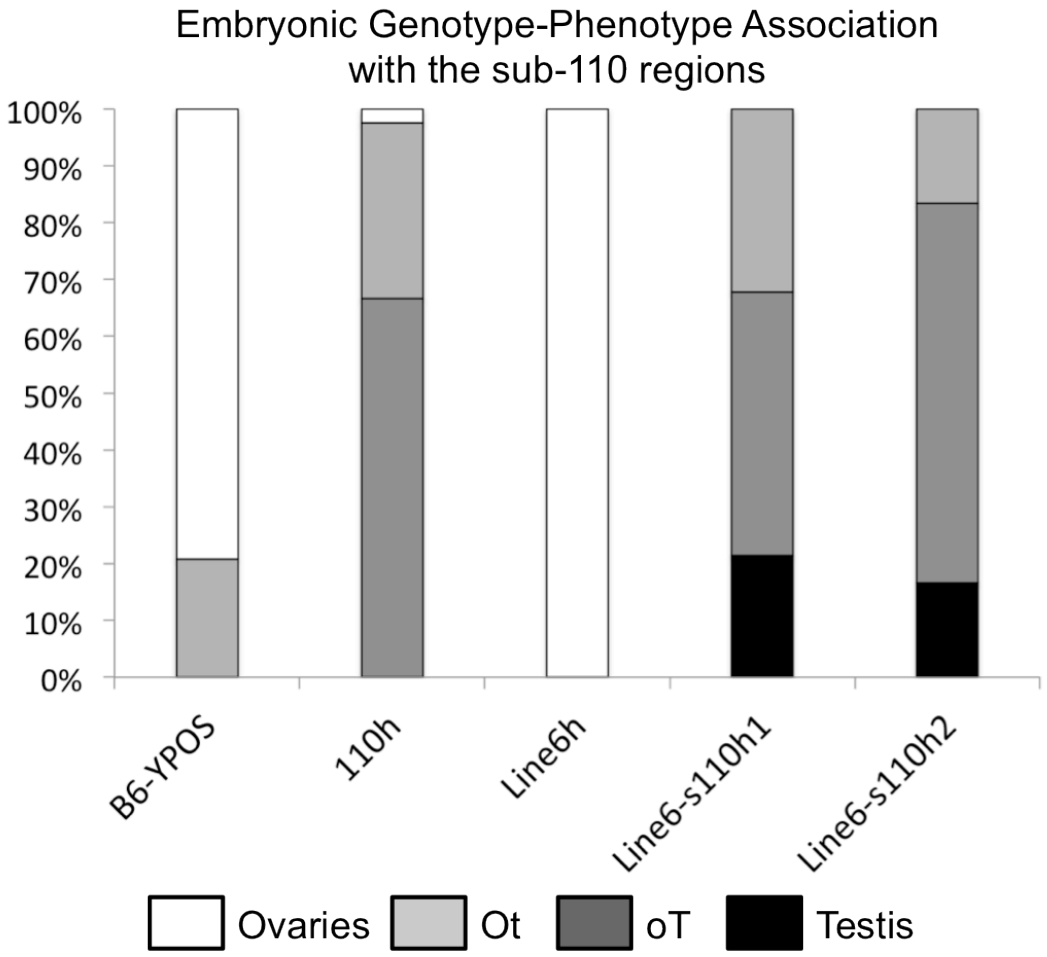

Embryonic Genotype-Phenotype Association
with the sub-110 regions
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
B6-YPOS
110h
Line6h
Line6-s110h1
Line6-s110h2
Ovaries
Ot
oT
Testis

Figure 2-7: The sub-110 Region. Targeted next generation sequencing of the 110 region was performed in a 129S1SviM/J (129) animal and the Line6A-sub110h founder animal. A comparison of the region in the 129 strain and the sub110 region in our protected congenic line illustrates that protective 110 region is not derived from a congenic region of 129 origin. The sub-110 region encompasses nearly the entire *Sox9* promoter and 6 genes, but not *Sox9* itself. h: heterozygous, H: homozygous

chr11(Mb):
110.5
111
111.5
112
112.5
Line 6A-sub-110
129S1/SviMJ
Abca9
Abca5
Abca6
Map2k6
Kcnj16
Kcnj2
Sox9

Figure 2-8: *Sox9* Expression in rescued in the B6-Y^{POS} animals heterozygous or homozygous for the 110 region. Whole mount immune fluorescence of *Sox9* expression in the developing gonad at the point of sex determination in B6-110 region Y^{POS} is significantly higher than in the unprotected B6-Y^{POS} gonad. Expression in protected gonads is comparable to B6-Y^{B6} animals. PE-CAM is present to show the development of male-specific vasculature and germ cells. h: heterozygous, H: homozygous

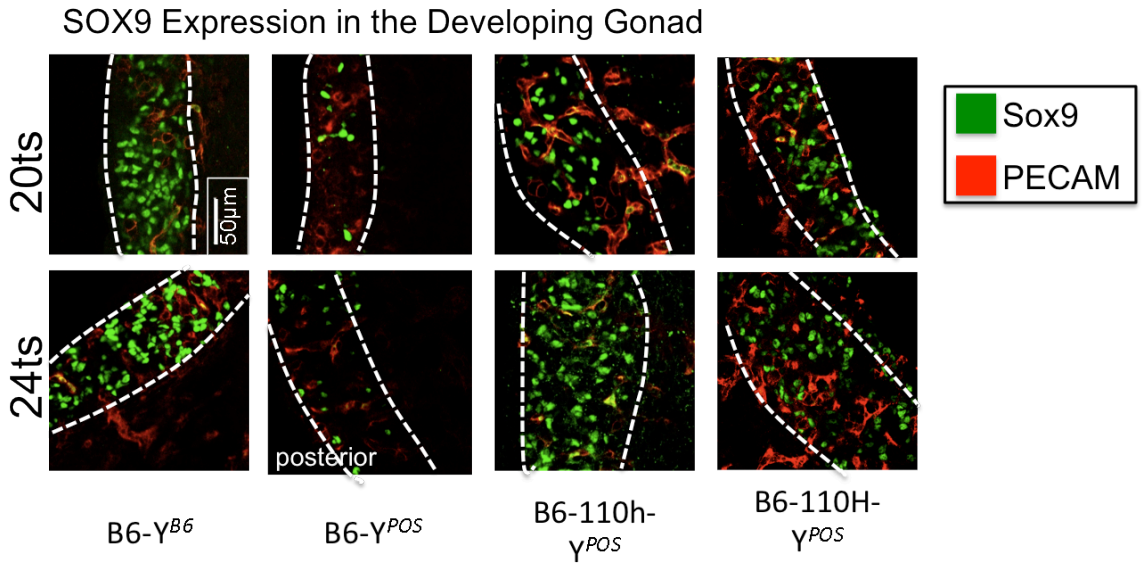

SOX9 Expression in the Developing Gonad
20ts
24ts
50μm
posterior
Sox9
PECAM
B6-Y^B6
B6-Y^POS
B6-110h-Y^POS
B6-110H-Y^POS

Figure 2-9: Human and Mouse Duplications and Deletions involving the SOX9 promoter.

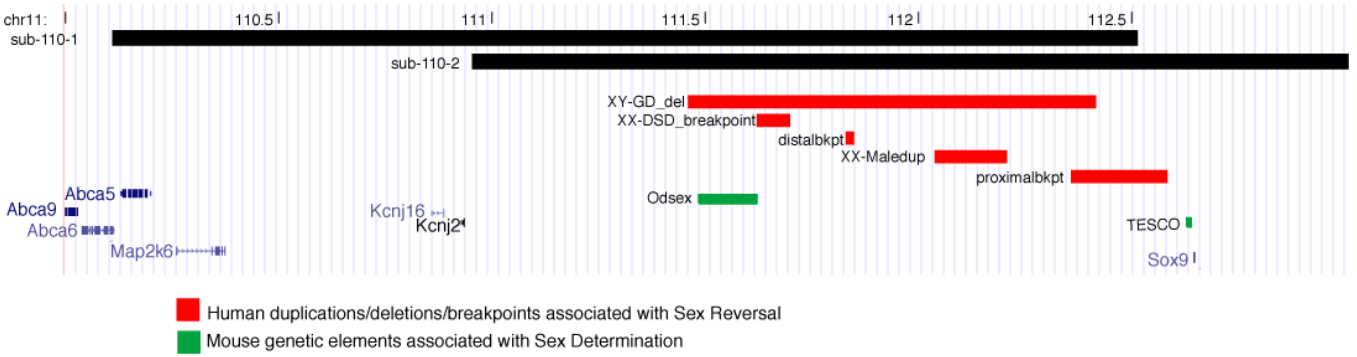

chr11:
110.5
111
111.5
112
112.5
sub-110-1
sub-110-2
XY-GD_del
XX-DSD_breakpoint
distalbkpt
XX-Maledup
proximalbkpt
Odsex
Abca5
Abca9
Abca6
Map2k6
Kcnj16
Kcnj2
TESCO
Sox9
Human duplications/deletions/breakpoints associated with Sex Reversal
Mouse genetic elements associated with Sex Determination

Table 2-1: SNPs Polymorphic for B6 and 129 used for Chromosome 11 Genotyping

Recombinant SNPs used in subcongenic line generation

dbSNP ID	B6	Other	Search string	Forward primer	Reverse primer	Size
rs13480100	T	C	GGAGGG	CCAAGTATGGAGAAGACTTC	CCACAGTCAAAGACCAGTCA	203
rs4222040	T	A	TTCCAG	TCACTTCTGAAGGAATGACCA	AAGGGAGATTCGGGTTATG	278
rs3723733	G	A	GGTTCA	GGAACATACTATAGAGGACA	GCTGACTTAGGTTTCTGTC	167
rs3024185	G	A	AGCTC	CATGGGGTCCCAGACAGT	GCAGAAGGAAGAGTTTCCTAGC	~180
rs3668244	T	C	CACAGG	AAGACCTACTCACAGAATCTACAAGC	AAAAGAAAGAAAGTGTCTGGTGTG	~200
rs3658198	G	A	CTCAGG	TGATTCCATAAGATAGGGGTCTTG	GAGGCACAGGGCTACAGTTC	~200
rs3682081	T	C	CTGAGC	TCTGAAGGGTACAACTAAGGGAAG	CTTAGGTGTCTCTGCCAGCTC	~200
rs3088501	T	C	GAACTA	ACAAAGCTTGGCACTGTCCT	AGAATCCAGGCTGTGCTGTT	386

SNPs from Diversity Array used in redefinition of existing lines

dbSNP ID	B6	Other	Search string	Forward primer	Reverse primer	Size
rs26831216	G	C	TCTCTCA	CTTGGTCAAAGAGCCCTCAG	GTTCCAAAGGTGTGCTGGTT	234
rs27019103	C	T	TGGAGG	AAAGTGTGCTTCCCAGGAGA	CCTCTCCCTCAACCCCTAAG	252
rs28240850	G	A	AGGGAAG	CCACAGCTGGAGGTAGGGTA	CCTAAGATGCCATGGGAAGA	176
rs27029785	C	A	aaggct	AGGTGGTGCTGGTTTTGAAG	TTTGCACAACCCCCTTAGTC	139

SNP Name	Forward primer	Reverse primer	Size
Sexing	CCGCTGCCAAATTCTTTGG	TGAAGCTTTTGGCTTTGAG	~300

Table 2-2: Whole Genome Copy Number Variations Analysis. For the founders of lines 3, 4, 5 and 6B we performed Copy Number Variation Analysis to identify regions that were duplicated or deleted within the genome.

Identified in Line	Status	Chromosome	StartPosition	EndPosition	Number of Probe sets	Size	Gene involved
3	gain	2	83493534	83502417	22	8884	erythropoietin 4 immediate early response
3	loss	3	25426121	25445541	11	19421	neuroligin 1-intron
3	gain	5	29863905	29864556	7	652	
3	gain	6	139604484	139664403	15	59920	heat shock protein 1
3	gain	6	145096182	145096552	9	371	Lrmp
3	gain	12	114683338	114848428	51	165091	Antibody variable region
3	gain	16	14461082	14461464	6	383	ATP-binding cassette, sub-family C, member 1
4	gain	2	83493534	83502417	22	8884	erythropoietin 4 immediate early response
4	loss	3	25426121	25445541	11	19421	neuroligin 1-intron
4	loss	5	110372511	110415466	9	42956	
4	gain	6	139605739	139605813	4	75	heat shock protein 1
4	gain	6	145096182	145096552	9	371	Lrmp
4	loss	11	65644695	65645861	8	1167	dynein, axonemal, heavy chain 9
4	loss	11	71038160	71137579	74	99420	NLR family, pyrin domain containing 1B
4	gain	12	114668031	114827120	44	159090	Antibody variable region
4	loss	14	69880932	70064719	133	183788	ectonucleoside triphosphate diphosphohydrolase
5	gain	2	83493534	83502417	22	8884	erythropoietin 4 immediate early response
5	loss	3	25414824	25445541	12	30718	neuroligin 1-intron
5	gain	6	145096182	145096552	9	371	Lrmp
5	gain	12	114683338	114848428	51	165091	Antibody variable region
5	loss	14	69880932	69970736	58	89805	ectonucleoside triphosphate diphosphohydrolase
6B	gain	2	83493534	83502417	22	8884	erythropoietin 4 immediate early response
6B	loss	3	25414824	25445541	12	30718	neuroligin 1-intron
6B	gain	12	114683338	114848428	51	165091	Antibody variable region

REFERENCES

1 Sinclair, A. H. *et al.* A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240-244 (1990).

2 Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. Male development of chromosomally female mice transgenic for Sry. *Nature* **351**, 117-121 (1991).

3 Sekido, R. & Lovell-Badge, R. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* **453**, 930-934, (2008).

4 Sutton E, H. J., White S, Rogers N, Tan J, RowleyL, Eyre H, Knower K, Sekido R, Rizzoti K, GoncalvesJ, Arboleda VA, Harley V, Vilain E, Sinclair A, Lovell-BadgeR & Thomas P. . A novel role for SOX3 in mouse and human XX male sex reversal. *Submitted* (2009).

5 Bishop, C. E. *et al.* A transgenic insertion upstream of sox9 is associated with dominant XX sex reversal in the mouse. *Nature genetics* **26**, 490-494, (2000).

6 O'Bryan, M. K. *et al.* Sox8 is a critical regulator of adult Sertoli cell function and male fertility. *Developmental biology* **316**, 359-370, (2008).

7 Polanco, J. C., Wilhelm, D., Davidson, T. L., Knight, D. & Koopman, P. Sox10 gain-of-function causes XX sex reversal in mice: implications for human 22q-linked disorders of sex development. *Human molecular genetics* **19**, 506-516, (2010).

8 Eicher, E. M. Autosomal genes involved in mammalian primary sex determination. *Philos Trans R Soc Lond B Biol Sci* **322**, 109-118 (1988).

9 Albrecht, K. H. & Eicher, E. M. DNA sequence analysis of Sry alleles (subgenus Mus) implicates misregulation as the cause of C57BL/6J-Y(POS) sex reversal and defines the SRY functional unit. *Genetics* **147**, 1267-1277 (1997).

10 Bowles, J., Cooper, L., Berkman, J. & Koopman, P. Sry requires a CAG repeat domain for male sex determination in Mus musculus. *Nat Genet* **22**, 405-408, (1999).

11 Bullejos, M. & Koopman, P. Delayed Sry and Sox9 expression in developing mouse gonads underlies B6-Y(DOM) sex reversal. *Dev Biol* **278**, 473-481, (2005).

12 Hiramatsu, R. *et al.* A critical time window of Sry action in gonadal sex determination in mice. *Development* **136**, 129-138, doi:10.1242/dev.029587 (2009).

13 Eicher, E. M. *et al.* Sex-determining genes on mouse autosomes identified by linkage analysis of C57BL/6J-YPOS sex reversal. *Nat Genet* **14**, 206-209, doi:10.1038/ng1096-206 (1996).

14 Munger S, A., D, Haider A, Threadgill D, Capel B. in *Fifth International Symposium on the Biology of Vertebrate Sex Determination* (Kona, Hawaii, 2009).

15 Eicher, E. M., Washburn, L. L., Whitney, J. B., 3rd & Morrow, K. E. Mus poschiavinus Y chromosome in the C57BL/6J murine genome causes sex reversal. *Science* **217**, 535-537 (1982).

16 Nikolova, G., Sinsheimer, J. S., Eicher, E. M. & Vilain, E. The chromosome 11 region from strain 129 provides protection from sex reversal in XYPOS mice. *Genetics* **179**, 419-427, (2008).

17 Whitney, J. B., Mills, T. M., Lewis, R. W., Wartell, R. & Abney, T. O. A single genetic determinant that prevents sex reversal in C57BL-YPOS congenic mice. *Biochem Genet* **38**, 119-137 (2000).

18 Theiler, K. The House Mouse: Atlas Of Embryonic Development. *Springer, N.Y.* (1989).

19 Yang, H. *et al.* A customized and versatile high-density genotyping array for the mouse. *Nat Methods* **6**, 663-666,(2009).

20 Wilhelm, D. *et al.* Antagonism of the testis- and ovary-determining pathways during ovotestis development in mice. *Mech Dev* **126**, 324-336, (2009).

21 Bullejos, M. & Koopman, P. Spatially dynamic expression of Sry in mouse genital ridges. *Dev Dyn* **221**, 201-205, (2001).

22 Hacker, A., Capel, B., Goodfellow, P. & Lovell-Badge, R. Expression of Sry, the mouse sex determining gene. *Development (Cambridge, England)* **121**, 1603-1614 (1995).

23 Yang, H. *et al.* A customized and versatile high-density genotyping array for the mouse. *Nature methods* **6**, 663-666, doi:10.1038/nmeth.1359 (2009).

24 Huang, B., Wang, S., Ning, Y., Lamb, A. N. & Bartley, J. Autosomal XX sex reversal caused by duplication of SOX9. *Am J Med Genet* **87**, 349-353 (1999).

25 Vidal, V. P., Chaboissier, M. C., de Rooij, D. G. & Schedl, A. Sox9 induces testis development in XX transgenic mice. *Nat Genet* **28**, 216-217, (2001).

26 Cox, J. J., Willatt, L., Homfray, T. & Woods, C. G. A SOX9 duplication and familial 46,XX developmental testicular disorder. *The New England journal of medicine* **364**, 91-93, doi:10.1056/NEJMc1010311 (2011).

27 White, S. *et al.* Copy Number Variation in Patients with Disorders of Sex Development Due to 46,XY Gonadal Dysgenesis. *PLoS One* **6**, e17793, doi:10.1371/journal.pone.0017793 (2011).

28 Jakob, S. & Lovell-Badge, R. Sex determination and the control of Sox9 expression in mammals. *Febs J* **278**, 1002-1009, (2011).

29 Uhlenhaut, N. H. *et al.* Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell* **139**, 1130-1142, (2009).

30 Qin, Y. *et al.* Long-range activation of Sox9 in Odd Sex (Ods) mice. *Human molecular genetics* **13**, 1213-1218, (2004).

31 Qin, Y. *et al.* A major locus on mouse chromosome 18 controls XX sex reversal in Odd Sex (Ods) mice. *Human molecular genetics* **12**, 509-515 (2003).

32 Poirier, C. *et al.* A complex interaction of imprinted and maternal-effect genes modifies sex determination in Odd Sex (Ods) mice. *Genetics* **168**, 1557-1562, (2004).

33 Benko, S. *et al.* Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nature genetics* **41**, 359-364, doi:10.1038/ng.329 (2009).

34 Epstein, D. J. Cis-regulatory mutations in human disease. *Brief Funct Genomic Proteomic* **8**, 310-316, (2009).

35 Georg, I. *et al.* Mutations of the SRY-responsive enhancer of SOX9 are uncommon in XY gonadal dysgenesis. *Sex Dev* **4**, 321-325, (2010).

36 Birney, E., Lieb, J. D., Furey, T. S., Crawford, G. E. & Iyer, V. R. Allele-specific and heritable chromatin signatures in humans. *Human molecular genetics* **19**, R204-209, (2010).

37 Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human molecular genetics* **12**, 1725-1735 (2003).

Chapter 3:

Targeted Massively Parallel Sequencing Provides Comprehensive Genetic Diagnosis for Patients with Disorders of Sex Development

ABSTRACT

Disorders of Sex Development (DSD) are rare disorders in which there is discordance between chromosomal, gonadal, and phenotypic sex. Only a minority of patients clinically diagnosed with DSD obtains a molecular diagnosis, leaving a large gap in our understanding of the prevalence, management, and outcomes in affected patients. We created a novel DSD-genetic diagnostic tool, in which sex development genes are captured using RNA probes and undergo massively parallel sequencing. In the pilot group of 14 patients, we determined sex chromosome dosage, copy number variation, and gene mutations. In the patients with a known genetic diagnosis (obtained either on a clinical or research basis), this test identified the molecular cause in 100% (7/7) of patients. In patients in whom no molecular diagnosis had been made, this tool identified a genetic diagnosis in 2 of 7 patients. Targeted sequencing of genes representing a specific spectrum of disorders can result in a higher rate of genetic diagnoses than current diagnostic approaches. Our DSD diagnostic tool provides for first time, in a single blood test, a comprehensive genetic diagnosis in patients presenting with a wide range of urogenital anomalies.

INTRODUCTION

Disorders of Sex Development (DSD) are defined as a rare set of conditions in which the chromosomal, gonadal, and phenotypic sex is atypical. DSD has a prevalence of 0.1%-0.5% of live births, yet only 13% of patients will ever receive a definitive genetic diagnosis[1]. [1-3] The uncertainty regarding the child's gender and future psychosocial and psychosexual development is extraordinarily stressful for the child's family.[1,4,5] From the time of initial presentation, patients with DSD undergo a wide spectrum of clinical and endocrine tests, from which life-altering decisions are made about gender assignment, medical treatments, and surgery. Yet, to date, evidence is lacking to justify support for specific management strategies of these patients.[6]

The promise of next generation sequencing in the clinical arena is hindered by the difficulty in differentiating between an inconsequential sequence polymorphism and a disease-causing mutation. Although the first predictive test for *BRCA1* and *BRCA2* mutations had numerous detractors, genetic testing for *BRCA1* and *BRCA2* has transformed the management of high-risk patients and in the process, researchers have discovered a vast number of gene variants, which are now classified based on their cancer risk.[7,8]

Unlike traditional genetic diagnostic tests that at most sequence a handful of genes or target a panel of known mutations, we have combined multiple genetic tests and put forward a novel and integrated role for comprehensive molecular genetic diagnostics in the clinical realm. Our test combines multiple genetic testing modalities routinely ordered in DSD patients, including sex chromosome complement determination, copy number variant (CNV) analysis, and gene sequencing. Currently, gene sequencing is done on a gene-by-gene basis. Many genes, particular those for rare or complex disorders, are only offered on a research basis, further

[1] This percentage is based on a systematic electronic medical chart review targeting patients categorized as DSD at one major mid-western academic medical center.

complicating the genetic diagnostic process. This strategy replaces multiple single-gene sequencing tests with a unified test, thereby drastically improving the odds of identifying a high-risk variant and of assigning the appropriate management based on the individual's genetic risk.

We propose a novel diagnostic process allowing clinicians to initially identify a genetic mutation, which would be followed by relevant metabolic, endocrine, and imaging tests for functional assessment of the gene mutation. This diagnostic approach can eliminate non-indicated clinical tests, sparing the patient unnecessary stress and saving healthcare system's resources. Finally, by pinpointing the genetic diagnosis at the beginning of the diagnostic process, we can more accurately analyze and predict both future developmental issues in the child and the risk of recurrence within the family. In our pilot group of patients, we have shown that this novel targeted diagnostic approach can accurately diagnose the genetic basis of DSD in the majority of patients. This new test shifts the paradigm of the diagnostic processes and ultimately has the potential to increase the rate of genetic diagnosis, provide more cost-effective care, and allow for more informed clinical management in patients with DSD.

MATERIALS AND METHODS

Clinical diagnosis of all DSD patients is outlined in Table 1 and in Study Subjects below. Patients with known genetic diagnoses were diagnosed in either clinical laboratories or on a research basis. Patients 45, X and 47, XXY, and DSDPt7 were diagnosed in a clinical laboratory using karyotype and/or Sanger Sequencing. DSDPts 2, 3, 8 and 9 were genetically diagnosed on a research basis only after extensive endocrine work-up. All other patients did not have a genetic diagnosis. The clinical and genetic diagnoses were blinded to the investigators. This study was approved by the Institutional Review Board at the University of California, Los Angeles.

Study Subjects

All participants were patients diagnosed clinically with a DSD. We also included healthy 46,XX and a 46,XY controls, for normalization of sex chromosome dosage and copy number variation analysis. All DSD patients had clinical record of karyotype analysis. For 7/14 of the DSD patients, the causative genetic mutation had already been determined, by clinical or research laboratories using traditional molecular diagnostics. The remaining 7 patients had been further assessed by FISH and gene sequencing of SRY and underwent an endocrine evaluation but all findings were negative and received no definitive genetic diagnosis. This study was approved by the Institutional Review Board at the University of California, Los Angeles. All participants provided informed consent and DNA samples were obtained from saliva (Oragene) or whole blood samples purified with the Puregene Blood DNA Extraction Kit (Qiagen).

DSDPtTS: This patient was diagnosed at birth with classic Turner Syndrome and a non-mosaic 45,X karyotype.

DSDPtKS: Patient was diagnosed with Klinefelter Syndrome and a non-mosaic 47,XXY karyotype.

DSDPt1: This patient presented early in life with ambiguous genitals but was never clinically worked up until adulthood. Clinical testing revealed an abnormally low DHT/T ratio.

DSDPt2: Initially presented with severe combined adrenal and gonadal deficiency. This patient was previously reported in[9].

DSDPt3: This patient presented with 46,XY Gonadal Dysgenesis and no genital ambiguity. Karyotype was 46,XY and SRY was present. Sequencing of the SRY gene at the research level revealed a mutation in the SRY gene by Sanger sequencing, which was independently confirmed in a CLIA approved laboratory.

DSDPt4: This patient was clinically diagnosed with Campomelic Dysplasia with 46, XY autosomal sex reversal. However, no mutation was identified after Sanger sequencing of the SOX9 gene.

DSDPt5: This patient presented in the newborn period with 46, XY Gonadal Dysgenesis, a female phenotype, and galactosemia.

DSDPt6: This patient presented with 46, XX Testicular DSD with no SRY on FISH. CNV analysis revealed no deletions or duplications that might explain the phenotype.

DSDPt7: Patient was diagnosed clinically with complete Androgen Receptor insensitivity which was confirmed by mutation analysis in a clinical laboratory.

DSDPt8: Patient was diagnosed with adrenal hypoplasia congenita with XY sex reversal. A hemizygous DAX1 mutation was identified.

DSDPt9: This patient was clinically described as 46, XY Gonadal Dysgenesis with a maternal aunt who was also 46, XY Female. Genetic diagnosis by CNV microarray was performed in [10].

DSDPt10: This patient initially presented as 46, XX DSD with ambiguous genitalia. Patient has a half-sister related through the father with 46, XX DSD with ambiguous genitalia. No genetic diagnosis has been made in this family.

DSDPt11: This patient was clinically diagnosed with Swyer syndrome and is 46, XY. No genetic diagnosis has been made in this case.

DSDPt12: This patient presented in the newborn period with full posterior labial fusion, underdeveloped scrotum, no testes palpable bilaterally, an enlarged clitoris (1cm in length), and no hyperpigmentation. No other anomalies were detected. Sex chromosome analysis revealed a 46, XY karyotype. Serum 17-OH-Progesterone, cortisol, and pituitary function were in the normal range. Testosterone and Anti-Mullerian Hormone levels were below the normal range,

indicating decreased testicular tissue. On ultrasound of the pelvis, no Müllerian structures or testis were found. At 2 years of age, there was no apparent developmental delay.

Targeted Capture was performed using custom Agilent SureSelect Target Enrichment System Kit[11] (Fig.1). We designed oligonucleotide baits tiled against exonic and intronic regions of 35 known genes of sex development, up to 10kb regions upstream and downstream of all known genes in sex development, and 3-10kb spaced every 10Mb along the X- and Y-chromosomes (Supplemental Data Table 1A and 1B). All clinically-associated genes reported in the literature (as of December, 2009) in both sex determination and sex differentiation were included. We also included a subset of genes for ovarian insufficiency. Up to six custom bar-coded samples were pooled, captured with the baits designed for one reaction and sequenced on a single lane of Illumina GAIIx for 76 cycles or HiSeq2000 for 50 cycles. The reads were aligned to the whole genome using Novoalign (http://www.novocraft.com/index.html) and the aligned reads were processed using SAMtools (http://samtools.sourceforge.net/) and Picard (http://picard.sourceforge.net/)[12] to remove potential PCR duplicates. Both single nucleotide variants (SNVs) and small insertions and deletions (INDELs) within the captured coding exonic and splice-site intervals of the DSD genes were called using SAMtools pileup tool and annotated using the SeqWare project (http://seqware.sourceforge.net).[13] The SNVs and INDELs were further filtered to include only those resulting in non-synonymous nucleotide substitution, frameshift, in-frame INDELs, splice-site, or early-termination mutations. Finally, in order to minimize the risk of false-positive SNV findings, only the variants called with SNV Phred score ≥ 30, total coverage ≥10X and % of non-reference call ≥15 were further analyzed (Fig. 3-1). All variants with coding consequences were analyzed against the public Human Gene Mutation Database (HGMD), dbSNP132 (common variants present at ≥1% frequency), and SNVs were

run through three independent protein pathogenicity predictors: PolyPhen-2, SIFT, and MutationAssessor in order to determine whether they were likely to be disease-causing.[14-16] When used together, the three independent *in silico* pathogenicity predictors have a higher positive predictive value and any of the predictors alone[17]. If two of the three algorithms predicted a tolerable/benign effect of the mutation this was considered "likely benign". All bioinformatic analysis was performed blinded to the patient's chromosomal sex, phenotype, and diagnosis.

SNV Validation

To determine the accuracy of our sequencing pipeline, we compared DSDPt12 to previously acquired Illumina IM SNP Genotyping data. This array was run and analyzed as per manufacturer's protocols at the Southern California Genotyping Consortium and had a call rate of 0.9963. 691 SNPs within the targeted region were also genotyped on this array and compared to our SNP and INDEL variant callers. 4/691 SNPs were discordant between the genotyping data and the Illumina Sequencing data, giving a false negative rate of 0.57%. These false negatives were 4 SNVs that the genotyping data called heterozygous while the sequencing data did not. Conversely, the sequencing did not identify any SNVs at base positions where the genotyping data did not call an SNV.

Sex Chromosome Dosage and CNV analysis

Our sex chromosome complement analysis comprised of two normalization steps. First, we normalize for differences in coverage levels because starkly different coverage levels between samples is a common issue, known to hinder CNV analysis of targeted sequencing data[18]. Second, we normalize the X- and Y- chromosome coverage to the sample's autosomes in order to perform inter-sample comparisons.

More precisely, any sample whose DOC coverage was 0.5 standard deviation higher than the mean DOC of all the samples, was subjected to normalization of the DOC by reanalyzing a randomly selected subset of reads. Once the DOC was similar among all samples, we further normalized the samples, by dividing the mean depth of coverage (DOC) on the X and the Y chromosomes, Ci(chrX) and Ci(chrY) respectively, with the mean DOC of the patient's autosomes, Ci(chrAut). Since there are 2 copies of every autosome, and 0, 1, or 2 copies of the X- and Y-chromosomes, the ratio derived (0, 0.5, or 1) allows us to estimate the number of X- and Y- chromosomes per sample.

After the relative ratio of the sex chromosome to autosome for each sample was calculated, they were grouped by their estimated karyotype: XX, XY, XXY and XO. Two-sample t-test was used to assess the significance of the separation between different copy number groups: 1 versus 2 X-chromosomes and 0 vs 1 Y-chromosome. The X-chromosome and the Y-chromosome were tested separately as their copy number states can be considered to be independent of each other.

The same approach was taken for the copy number assessment of the DSD genes, except that instead of taking the mean DOC of the chromosomes, the mean DOC of each gene (G) was calculated (Ci(Gj)) and normalized by dividing with the mean DOC of the autosomes (Ci(chrAut)).

To detect copy number variants at genic or exonic level for sample *i*, gene G_j or exon E_j, the normalized DOC $C_i(G_j)/C_i(\text{chrAut})$ or $C_i(E_j)/C_i(\text{chrAut})$, was compared to those of the rest of the samples to determine if it was significantly greater or less. For the CNV analysis of the genes or the exons, outlier sample(s) for each gene was determined by assessing how the normalized

coverage of a sample is significantly different from the rest of the samples. The Z-score of a known duplication was used to determine the lower bound (LB) and upper bound (UB) for each gene or exon and the gene(s) or exons(s) falling outside the LB or UB were considered outliers.

RESULTS

Targeted sequencing achieved a capture efficiency of ~51.5% and a mean coverage of 48.3X per sample with 92.6% of the targeted base positions being sequenced at ≥10X. A total of 16 individuals were sequenced in our targeted approach, 2 unaffected individuals and 14 patients clinically diagnosed with DSD (Table 3-1). Some of the targeted regions were not covered due to the presence of repetitive regions, which comprised 2.3% of the total regions covered. Repetitive regions are historically difficult to sequence and map back to the reference genome and these findings are consistent with previously published results of targeted sequencing.[11,19] However, none of the genes sequenced are known to have mutations dependent on the size of these repetitive regions. To estimate the rates of false positive and false negative SNV calls, we compared SNP genotyping data to our SNV and INDEL calls in DSDPt12 and calculated a false negative call rate of <1% and found 0 false positives.

In order to analyze sex chromosome dosage, the normalized depth of coverage (DOC) for both chromosomes, C_i(chrX)/C_i(chrAut) and C_i(chrY)/C_i(chrAut) were calculated and independently examined (Fig. 3-2). First, we reliably distinguish between samples with one or two X-chromosomes (p-value<0.001). Calling the number of Y-chromosome was also clear as the four samples with no Y-chromosome had nearly null coverage while the ten samples with Y-chromosome had close to half the coverage compared to the (diploid) autosomes (p-value<0.001). Sample 47, XXY was properly clustered with the XX samples on the X-axis and the XY samples

on the Y-axis. Sample 45, X was properly clustered with the XY samples on the X-axis and had null coverage on the Y-chromosome. All of our called sex chromosome dosage matched clinically performed cytogenetic karyotype tests.

To identify both rare and common variants that might result in DSD, we employed a number of filters. First, to identify variants previously identified in the literature as causative of DSD we compared all coding SNVs and INDELs against Human Gene Mutation Database public (HGMD), which includes both rare and more "common" causes of DSD. We then filtered out common variants using dbSNP132 (≥1% frequency) and then ran all novel variants through *in silico* protein pathogenicity predictors to determine if they were likely benign and causative (see Fig. 3-1).

An average of 30 SNVs and INDELs were called along all coding exons +/- 3bp and in the testis-specific SOX9 enhancer in each patient.[20] Few variants led to protein-level changes (frameshift, inframe-INDELs, early-termination, missense, splice-site) in each sample. Of 19 high-quality protein-changing variants in all patients, five were reported to be causal variants for a similar phenotype in the HGMD public version (Table 2).[21] In the DSD patients without sex chromosome abnormalities, four patients (DSDPts3, 7, 8, and 9) had a previously identified genetic diagnosis, all of which were identified through screening of HGMD. This approach also identified a genetic diagnosis of five-alpha reductase deficiency in DSDPt1. None of the mutations identified in HGMD were present in dbSNP132 (≥1% frequency), indicating that these are rare variants. Additionally, we also identified a genetic mutation not present in HGMD public in DSDPt12, which is described below.

Of the remaining 14 variants, six were found to be common polymorphisms recorded in dbSNP132 (≥1% frequency), and therefore classified as likely benign mutations. The remaining

eight variants were not present in dbSNP132 (≥1% frequency) indicating they were rare variants, thus potentially pathogenic. Discerning between benign and potentially pathogenic rare variants required a multistep approach.

INDELs that result in out-of-frame coding consequences or lie in canonical splice junctions are automatically classified as "likely pathogenic".[22] A single insertion disrupting a canonical splice-site found in *CYP11A1* for DSDPt9 was not identified in HGMD or dbSNP132 (≥1% frequency). This was classified as likely pathogenic and is concordant with the known compound heterozygous genetic diagnosis in *CYP11A1*.

The remaining seven rare SNV variants were analyzed using two independent methods: 1) SNVs were run through three *in silico* protein pathogenicity prediction algorithms and 2) SNVs were manually evaluated based on inheritance of the disease (ie. sex-limited, recessive, dominant) and patient phenotype. These two analyses ensured that we were not excessively filtering out rare variants solely based on *in silico* pathogenicity predictors. We opted to be more conservative in our calling of benign variants, requiring that two of three pathogenicity predictors predict a tolerable effect and that the manual analysis did not find the SNV as likely to be causative.

In DSDPt12, diagnosed with 46,XY gonadal dysgenesis, we identified one hemizygous K1045E mutation in *ATRX*[23]. One of the three pathogenicity predictors called the missense mutation 'probably damaging' and manual inspection identified SNV in *ATRX* as causative in DSDPt12 with 46, XY gonadal dysgenesis. Therefore we called this a likely pathogenic mutation in the patient. All other SNVs were called as benign or tolerable by two of the three predictors. When manually evaluated, these same SNVs were either not pathogenic because of inheritance (for example, the gene typically requires pathogenic mutations on both alleles to demonstrate a

phenotype) and/or the phenotype was sex-limited and only displayed in either XY or XX individuals. In combining the data from *in silico* protein pathogenicity predictors and manual evaluation, the remaining six SNVs were classified as likely benign.

As duplication and deletions can contribute to DSD, we screened all of our patients for causative DSD duplications and deletions.[10,24] In DSDPt2, who has an XY karyotype, the mean DOC of *NROB1 (DAX1)*, C_{DSDpt2}(*NROB1*), was elevated to the level of the mean coverage achieved by the samples with XX karyotype (Fig.3-3), indicating copy number increase at the locus.[10] The normalized coverage of *NROB1* gene in DSDPt2 was 2.75 standard deviations away (Z-score = 3.04) from the mean of the normalized coverage of *NROB1* of all XY samples (p-value = 0.002). To call CNVs, we have chosen a significance threshold of Z-score 3.04 away from the mean to call a deletion or duplication involving an entire gene based on the results from DSDPt2, which generates a false positive rate for CNVs of 0.1%. All other DSD genes were tested in the same manner, with no additional duplications or deletions identified (Fig. 3-3).

DISCUSSION

The proposed method of broad-scale sequencing of all known DSD genes offers significant advantages over current diagnostic procedures for the assessment of DSD. The vast majority of disease-causing mutations can be attributed to sequence and copy number variations affecting the coding regions of genes. We limited the targeted genes to those with known roles in sex development, as pathogenic mutations in these genes can be confidently reported back to the clinician and patient. While we cannot identify novel genes in sex development using this approach, limiting the number of genes sequenced streamlines the bioinformatic analysis and restricts the pathogenic variants to those genes relevant to phenotype. Another major advantage

of this DSD-specific approach, rather than whole-exome or whole-genome sequencing, is that we decrease the chance of incidental findings unrelated to DSD. For instance, we eliminate the possibility of diagnosing minors with adult onset-diseases unrelated to the reason for genetic testing. Genetic testing for adult-onset diseases is ethically questionable in children, and under current guidelines is only performed in exceptional circumstances.[25] However, because the targeted method is readily scalable, inclusion of novel sex development genes or expansion of the targeted region to also include all genes resulting in ovarian insufficiency or male infertility can be easily updated in future capture designs.

Intensive study of important disease genes such as *CFTR* and *BRCA1/BRCA2* has taught us the complexity of single-gene disorders. The vast majority of disease genes demonstrate remarkable mutational heterogeneity in the general population (as opposed to the more restricted mutation sets found in certain ethnic groups), with mutations scattered across all exons of the genes with no particular "hot-spots". In such situations, there is really no alternative to whole-gene sequencing (at least of the exons and intron-exon junctions) if one is to entertain any hope of identifying the majority of causative mutations in affected individuals. The current price for full-gene sequencing of *BRCA1* and *BRCA2*, for example, is, at time of this writing, about $3600. The price for other individual genes offered in the clinical setting ranges from about $1500 to $3000, depending on the exonic size and other factors such as overall demand and test exclusivity. For those genetic disorders caused by many different genes, such as DSD, the cost of sequencing them all is prohibitive and is rarely done.

The majority of the patients with congenital adrenal hyperplasia, for example, have mutations in *CYP21A2*, while a smaller proportion of patients with the same condition have mutations in one of the four other genes that give rise to similar phenotypes[26] (*POR*, *STAR*,

HSD3B2, *CYP11B1*, *CYP17A1*). Sequencing all 6 genes by current methods would cost over $10,000. Assuming that we pool a minimum of seven bar-coded samples for one reaction worth of targeted baits and sequence on one lane of an Illumina HiSeq2000 flowcell with 50bp paired-end reads, our comprehensive sequencing approach would cost less than $1,000 per sample. Included in this cost per sample is the labor and reagents for library preparation ($350), cost of targeted baits ($150), bioinformatic analysis ($200), and the full sequencing service ($300). Even with the additional costs that need to be factored in if performed within a clinical setting, such as hospital overhead, maintenance of CLIA, CAP and state certifications, and the higher labor costs of licensed medical technologists, the proposed method would come out to be significantly cheaper while providing more information than current one-gene-at-a-time Sanger sequencing approaches. While endocrine testing and radiological imaging can help to prioritize the order in which genes are sequenced, these tests may lengthen the diagnostic process, increase costs, and are sometimes invasive. Sequencing all the genes upfront requires only a blood draw, and the entire process can be completed in less than three weeks, which is shorter than the turnaround time for clinical molecular genetic sequencing tests for single genes.

Several groups have explored similar targeted sequencing approaches to encompass all known genes conferring an inherited risk of breast/ovarian cancer, congenital ocular disorders, or hereditary hearing loss.[19],[27], [28] Both commercial and academic reference laboratories have realized the utility of such panels, and now offer sequencing services for all genes causing various cardiomyopathies, and other phenotypic traits associated with 20 or more genes.[29] All of these approaches simply replace the current one-gene-at-a-time sequencing tests traditionally offered, reducing the cost while increasing the diagnostic yield. However, we have yet to see a systematic reevaluation of the powerful role that such genetic diagnosis can play in early

diagnosis and management. By integrating these diagnostic tools into the clinical framework, we might eliminate unnecessary tests and the risks, costs, and diagnostic delays associated with them. Furthermore, with the cost of whole-genome sequencing soon falling below the aggregate cost of performing standard Sanger sequencing on two or three single genes, it is likely that health care systems will be reluctant to cover the latter, when much more comprehensive diagnostic information can be gained for the same price by massively parallel sequencing.

The ability to provide a single test that produces such a variety of genetic information (copy number variation, sex chromosome complement, sequence variants) has the potential to significantly alter clinical practice. In our cohort of fourteen patients, a diagnosis was identified in nine of fourteen (64%). Parents with children afflicted with a genetic disease place a high value on obtaining a genetic diagnosis, even with the knowledge that a diagnosis is not reached in all cases and that identification of the genetic lesion will not necessarily affect medical management.[30] By establishing a primary genetic diagnosis, the patient is spared a long and difficult diagnostic process including numerous costly and sometimes invasive tests. For parents dealing with the appreciably greater stress of a child presenting with a DSD, a genetic test may provide a better understanding of the condition's etiology and outcomes. The next step is to evaluate the impact of genomic sequencing on quality of life in patients with rare genetic disorders such as DSD.

Our novel results show the potential of using next generation sequencing to reframe the typical diagnosis pipeline within clinical medicine. This is especially true for those patients with rare disorders who have variable phenotypes and multiple genes associated with the phenotype, such as DSD. The development of clinical diagnostic tools targeted towards broader phenotypes

will catapult molecular diagnostics from a confirmatory test to a primary diagnostic tool that can diagnose and triage the patient earlier into appropriate management.

Traditional clinical diagnosis for newborns presenting with atypical phenotypic sex requires karyotype tests, electrolyte measurements, hormone challenges and stimulation tests, and imaging studies to visualize the gonads and internal reproductive structures. For newborns presenting with ambiguous genitalia, the major life-threatening concern is salt-wasting adrenal crisis. Therefore, we propose that all newborns presenting with ambiguous genitalia should be monitored until it can be safely determined that there is no risk of adrenal crisis. At the same time, in lieu of performing all the other subsequent clinical tests enumerated above, we propose that a blood specimen be sent for targeted DSD sequencing to identify the causative gene mutation. Once the involved gene is identified, follow-up functional tests can be performed to direct clinical management (Fig. 3-4). Establishing a precise genetic etiology early on allows one to predict the likelihood of developmental delay as well as conditions that might not be apparent in the newborn period.

The targeted approach is ideal for disorders that have similar phenotypes, typically affecting a single organ system, which can be the result of mutations in many different genes. Rare cases that are not genetically diagnosed by the targeted approach will require a more comprehensive work up to identify novel gene variants, non-coding variants, or copy number changes. However, these cases are the exception and not the rule and we believe the targeted approach will provide a diagnosis in the majority of DSD patients. Patients in whom a targeted approach is unable to identify a genetic cause of the DSD or patients with rare cases of DSD that have not been associated with a gene, such as agonadism, would be excellent candidates for whole exome and whole genome approaches. Because non-targeted approaches identify more

novel variants, both within coding and the non-coding regions, these technologies can be more difficult to interpret clinically.

Next generation sequencing has only begun to illuminate the genetic variants responsible for rare Mendelian diseases. As targeted sequencing approaches become cheaper and generate more data, it is up to the medical community to create sophisticated tests to utilize the technology such that physicians and patients can benefit from this revolutionary technology.

Figure 3-1: Targeted Sequencing Analysis Pipeline.

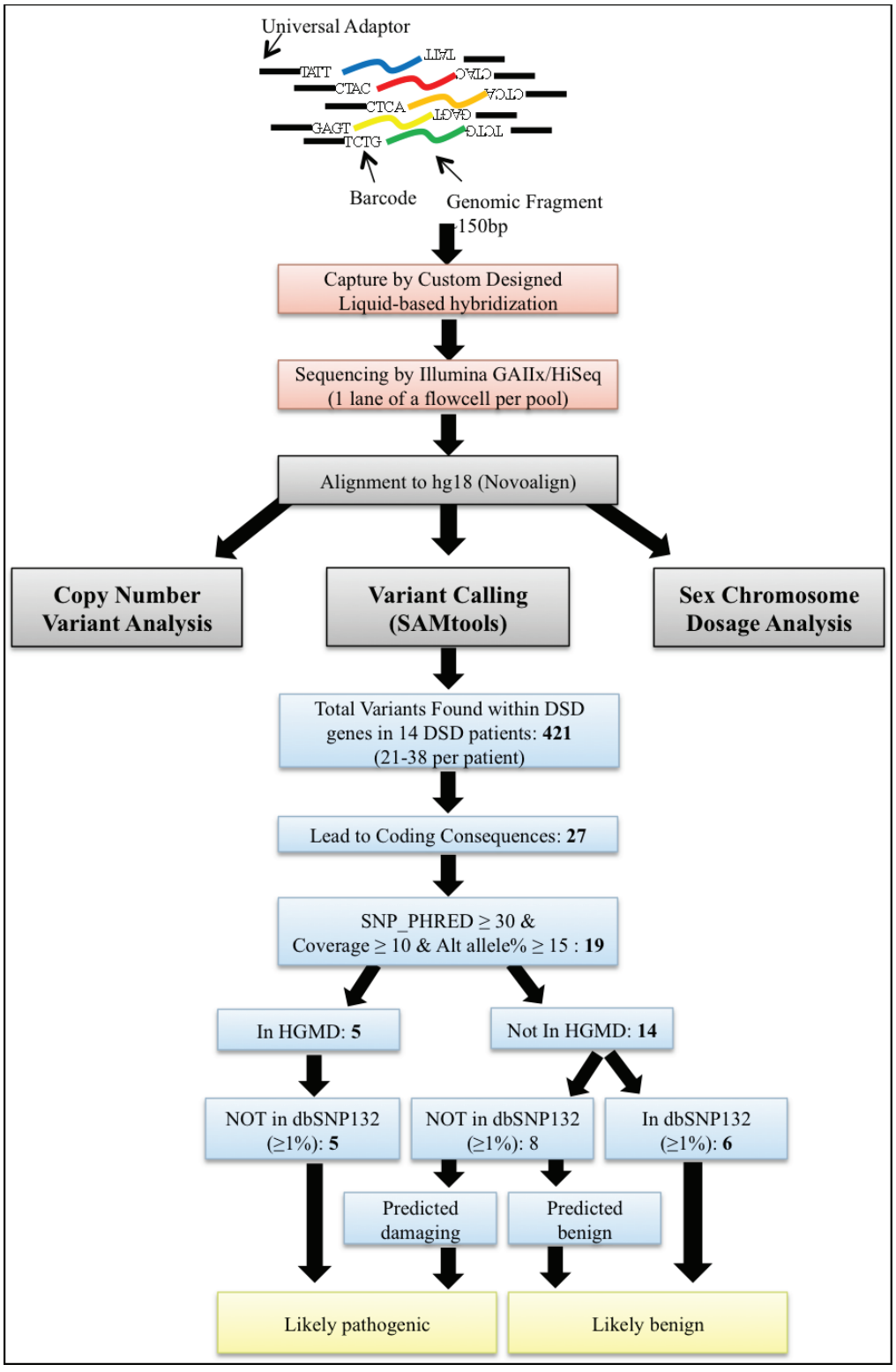

Universal Adaptor
TATT
CTAC
CTCA
GAGT
TCTG
Barcode
Genomic Fragment
150bp
Capture by Custom Designed
Liquid-based hybridization
Sequencing by Illumina GAIIx/HiSeq
(1 lane of a flowcell per pool)
Alignment to hg18 (Novoalign)
Copy Number
Variant Analysis
Variant Calling
(SAMtools)
Sex Chromosome
Dosage Analysis
Total Variants Found within DSD
genes in 14 DSD patients: 421
(21-38 per patient)
Lead to Coding Consequences: 27
SNP_PHRED ≥ 30 &
Coverage ≥ 10 & Alt allele% ≥ 15 : 19
In HGMD: 5
Not In HGMD: 14
NOT in dbSNP132
(≥1%): 5
NOT in dbSNP132
(≥1%): 8
In dbSNP132
(≥1%): 6
Predicted
damaging
Predicted
benign
Likely pathogenic
Likely benign

Figure 3-2: Sex Chromosome Complement A) Depth of coverage (DOC) along the X-chromosome and Y-chromosome was normalized to the autosomal DOC and plotted to determine sex chromosome complement. All XY samples (upper left) clustered together and had normalized coverage ranging from 0.43-0.63 for the X-chromosome and 0.58-0.72 for the Y-chromosome. All XX samples (lower right) clustered together, had null Y-chromosome coverage, and had close to 1 for normalized X-chromosome coverage, indicating an absence of the Y-chromosome and two copies of the X-chromosome. An X marks the mean coverage for the 46,XY cluster or the 46, XX cluster. The p-values for separating the two clusters were <0.001 for both directions and DSDpt3 that had the highest DOC (0.63) along the X-chromosome was separated from the 46, XX cluster with a p-value of <0.001.

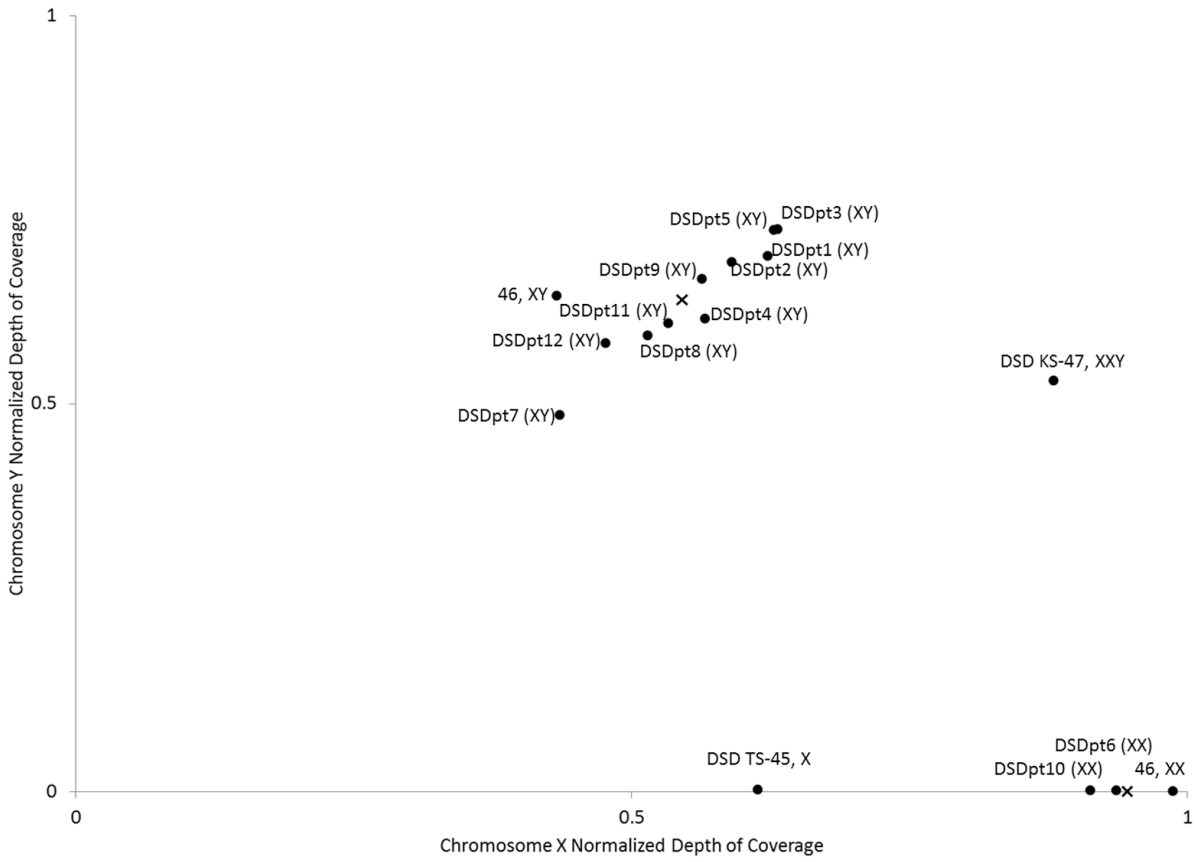
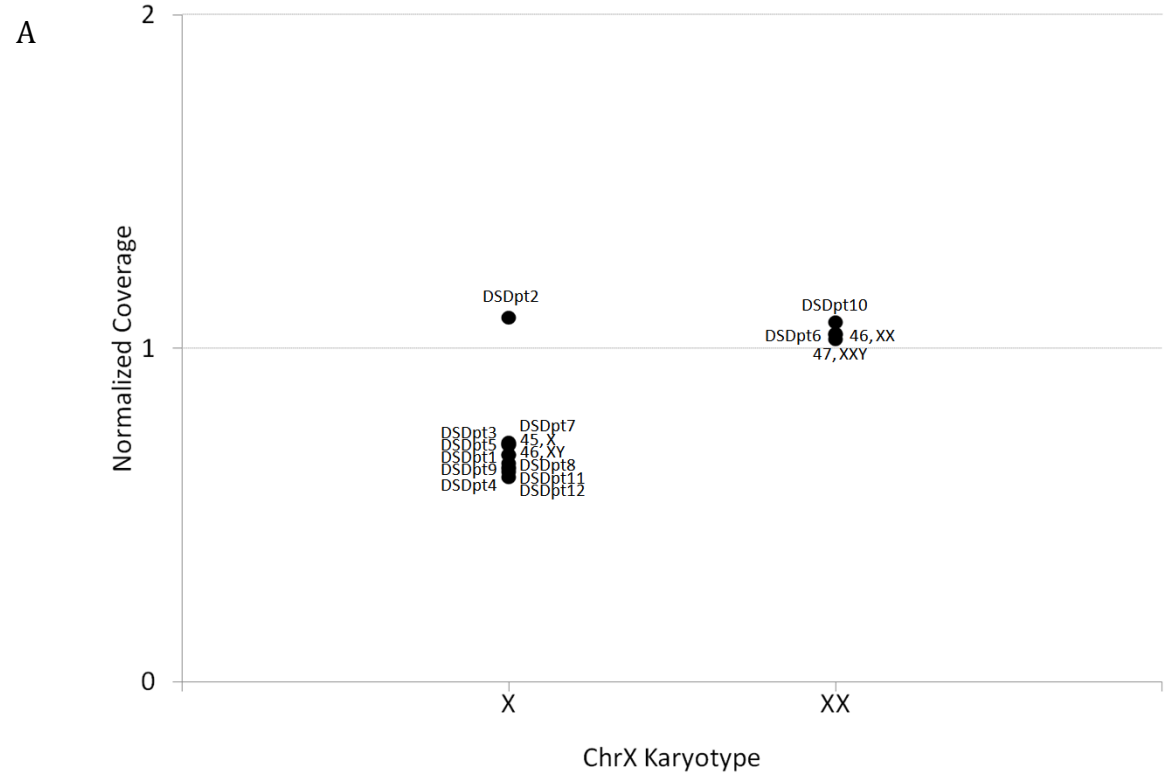
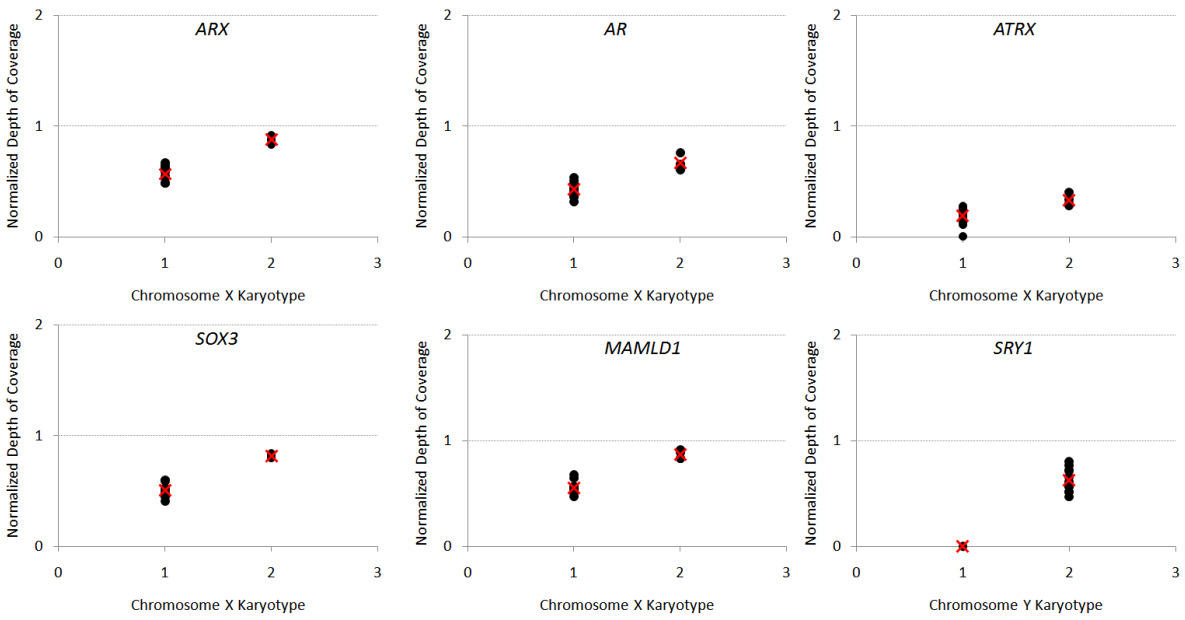

1
0.5
0
0
0.5
1
Chromosome Y Normalized Depth of Coverage
Chromosome X Normalized Depth of Coverage
DSDpt5 (XY)
DSDpt3 (XY)
DSDpt1 (XY)
DSDpt9 (XY)
DSDpt2 (XY)
46, XY
DSDpt11 (XY)
DSDpt4 (XY)
DSDpt12 (XY)
DSDpt8 (XY)
DSD KS-47, XXY
DSDpt7 (XY)
DSD TS-45, X
DSDpt6 (XX)
DSDpt10 (XX)
46, XX

Figure 3-3: Copy Number Variant Analysis

A) To determine CNV status for *NROB1* (*DAX1*), **c**overage for each gene was normalized to autosomal coverage of each sample and all samples plotted based on normalized coverage. Since *NROB1* is an X-chromosome gene, normalized coverage was plotted separately based on the X-chromosome karyotype. DSDpt2, with 46, XY GD, had significantly higher coverage than other XY individuals, indicating a duplication of the gene (p-value= 0.002). An X indicates the mean coverage for the gene. B) Genes on the X and Y chromosomes were analyzed separately for copy number dosage based on the number of sex chromosomes. No other copy number variations were identified in genes on the X or Y chromosome..C) Autosomal genes were also analyzed to determine if copy number variations were present and non were found.



B


ARX
AR
ATRX
SOX3
MAMLD1
SRY1
Normalized Depth of Coverage
Chromosome X Karyotype
Chromosome Y Karyotype

C

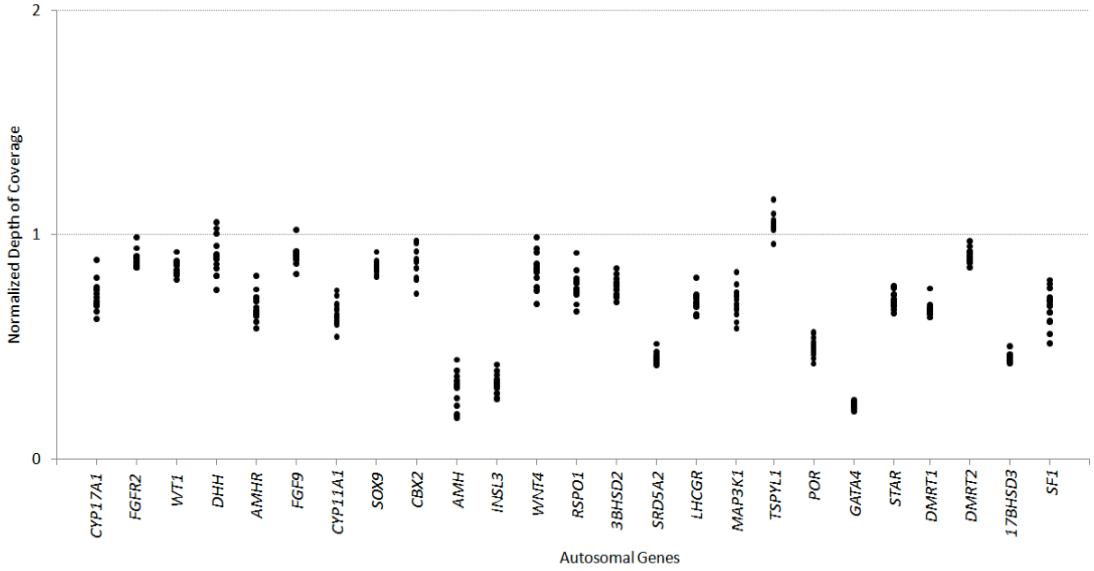

Normalized Depth of Coverage
CYP17A1
FGFR2
WT1
DHH
AMHR
FGF9
CYP11A1
SOX9
CBX2
AMH
INSL3
WNT4
RSPO1
3BHSD2
SRD5A2
LHCGR
MAP3K1
TSPYL1
POR
GATA4
STAR
DMRT1
DMRT2
17BHSD3
SF1
Autosomal Genes

Figure 3-4: Proposed Integration of Targeted Sequencing Approach to Clinical Management of Suspected DSD. Current clinical management begins with the identification of an abnormal phenotype and is followed by multiple metabolic and endocrine tests, genetic tests, and imaging studies in order to identify the mostly likely candidate for sequencing. Targeted sequencing approach would prioritize a genetic diagnosis, which would be functionally assayed and confirmed with endocrine and imaging studies of the patient.

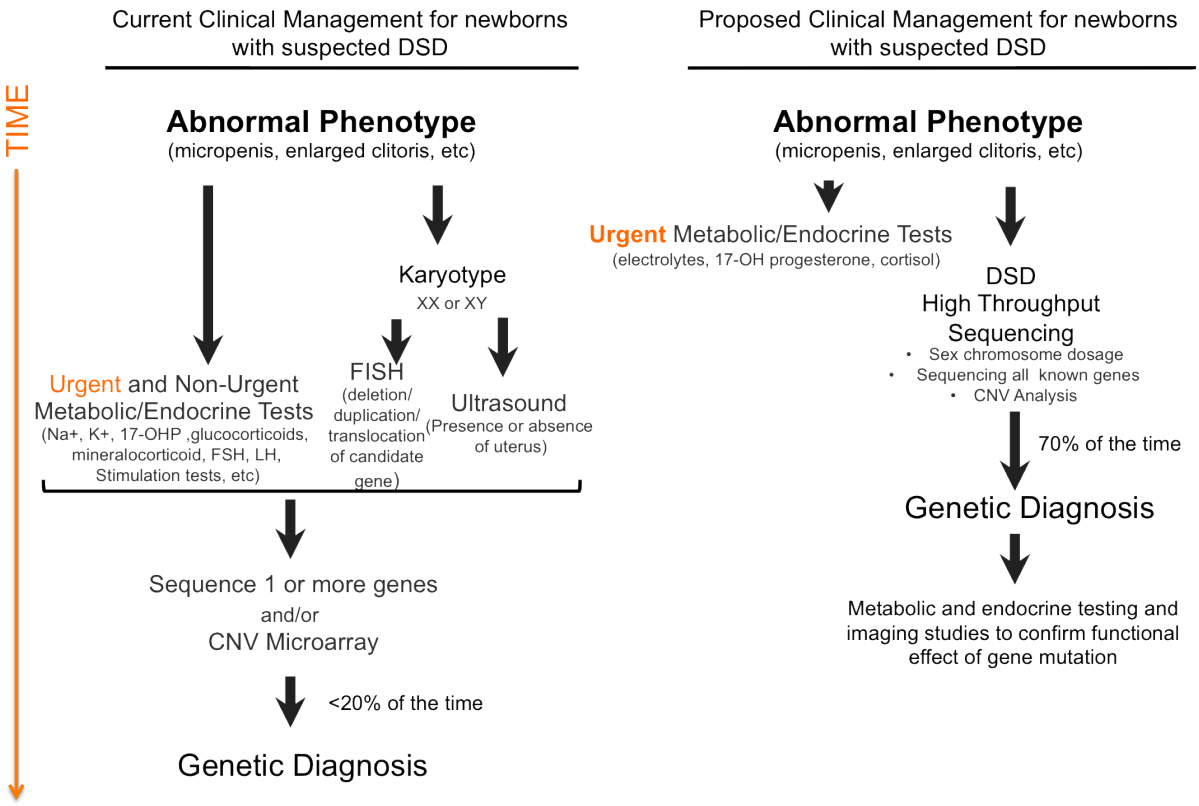

Current Clinical Management for newborns with suspected DSD
Proposed Clinical Management for newborns with suspected DSD
TIME
Abnormal Phenotype
(micropenis, enlarged clitoris, etc)
Urgent and Non-Urgent Metabolic/Endocrine Tests
(Na+, K+, 17-OHP ,glucocorticoids, mineralocorticoid, FSH, LH, Stimulation tests, etc)
Karyotype
XX or XY
FISH
(deletion/ duplication/ translocation of candidate gene)
Ultrasound
(Presence or absence of uterus)
Sequence 1 or more genes
and/or
CNV Microarray
<20% of the time
Genetic Diagnosis
Abnormal Phenotype
(micropenis, enlarged clitoris, etc)
Urgent Metabolic/Endocrine Tests
(electrolytes, 17-OH progesterone, cortisol)
DSD
High Throughput
Sequencing
• Sex chromosome dosage
• Sequencing all known genes
• CNV Analysis
70% of the time
Genetic Diagnosis
Metabolic and endocrine testing and imaging studies to confirm functional effect of gene mutation

Table 3-1: Clinical Diagnosis of Patients with DSD. All patients in this study were clinically diagnosed with a DSD.

Identification	Clinical Diagnosis	Genetic Diagnosis known (Y/N)	genetic diagnosis identified by targeted sequencing (Y/N)
46, XY Male	Control XY Male	-	-
46, XX Female	Control XX Female	-	-
47, XXY KS	Klinefelter Syndrome	Y	Y
45, XO TS	Turner Syndrome	Y	Y
DSDPt1	5-alpha reductase Deficiency	N	Y (SRD5A2, E200K))
DSDPt2	46, XY Gonadal Dysgenesis	Y (NROB1 (DAX1) duplication)	Y
DSDPt3	46, XY Gonadal Dysgenesis	Y (SRY, Y127C)	Y
DSDPt4	46, XY Gonadal Dysgenesis + Campomelic Dysplasia	N	N
DSDPt5	46, XY Gonadal Dysgenesis+ Galactosemia	N	N
DSDPt6	46, XX Testicular DSD	N	N
DSDPt7	46, XY DSD	Y (AR, M788T)	Y
DSDPt8	46, XY Female + AHC	Y (DAX1, Y121*)	Y
DSDPt9	46, XY DSD Severe Combined Adrenal and Gonadal Deficiency	Y (CYP11A1)	Y
DSDPt10	46, XX Testicular DSD	N	N
DSDPt11	46, XY Gonadal Dysgenesis	N	N
DSDPt12	46, XY Gonadal Dysgenesis	N	Y (ATRX, K1045E)

Table 3-2A: Captured Genomic Regions of DSD Genes

	Captured Genomic Region (hg 18)			
Gene	**Chromosome**	**Start**	**End**	**Size of Region**
WNT4	1	22,306,385	22,346,106	39,721
RSPO1	1	37,843,538	37,878,078	34,540
3BHSD2	1	119,758,459	119,791,480	33,021
SRD5A2	2	31,593,160	31,669,544	76,384
LHCGR	2	48,757,417	48,841,384	83,967
FOXL2	3	140,145,300	140,148,947	3,647
MAP3K1	5	56,141,657	56,230,735	89,078
CYP21A2	6	32,114,061	32,117,398	3,337
TSPYL1	6	116,703,555	116,708,857	5,302
POR	7	75,372,356	75,456,109	83,753
GATA4	8	11,561,877	11,664,918	103,041
STAR	8	38,116,375	38,132,757	16,382
ZFPM2 (FOG2)	8	106,390,323	106,895,943	505,620
DMRT1	9	821,690	969,090	147,400
DMRT2	9	1,035,858	1,050,552	14,694
17BHSD3	9	98,027,410	98,114,255	86,845
NR5A1 (SF1)	9	126,273,336	126,319,520	46,184
CYP17A1	10	104,578,278	104,589,280	11,002
FGFR2	10	123,217,834	123,357,962	140,128
WT1	11	32,355,901	32,423,663	67,762
DHH	12	47,767,473	47,775,869	8,396
AMHR	12	52,101,908	52,113,570	11,662
FGF9	13	21,140,875	21,182,586	41,711
CYP11A1	15	72,407,156	72,457,134	49,978
SOX9	17	67,618,756	67,644,155	25,399
CBX2	17	75,364,588	75,380,974	16,386
AMH	19	2,200,113	2,203,072	2,959
INSL3	19	17,788,322	17,793,320	4,998
ARX	X	24,926,732	24,953,986	27,254
NROB1 (DAX1)	X	30,230,460	30,240,416	9,956
AR	X	66,670,599	66,870,844	200,245
ATRX	X	76,637,012	76,938,375	301,363
SOX3	X	139,407,818	139,423,891	16,073
MAMLD1	X	149,359,378	149,443,104	83,726
SRY	Y	2,713,896	2,716,792	2,896
			Total Targeted Sequence for genes (bp)	2,394,810

Table 3-2B: Captured Genomic Regions for Sex Chromosome Dosage Analysis

Gene	Chromosome	Start	End	Size of Region
ZFY	Y	2,904,643	2,908,225	3,582
Jared1D	Y	20,326,764	20,331,132	4,368
RPS4Y2	Y	21,349,988	21,354,560	4,572
protein kinase, X-linked	X	3,532,411	3,541,661	9,250
amelogenin	X	11,221,454	11,228,802	7,348
ACOT9	X	23,631,698	23,635,328	3,630
hypothetical protein LOC645090	X	36,163,711	36,170,577	6,866
MAOB	X	43,510,801	43,516,665	5,864
jumonji, AT rich interactive domain 1C	X	53,237,378	53,247,329	9,951
KLHL4	X	86,807,227	86,811,790	4,563
GPRASP2	X	101,854,248	101,858,820	4,572
PGRMC1	X	118,254,395	118,261,254	6,859
ODZ1	X	123,339,463	123,344,035	4,572
			Total Targeted Sequence (bp)	75,997

Table 3-3: Sequencing Quality Statistics

Sample	Specificity	Average Coverage	% targeted bases >=10x	Autosome coverage	ChrX coverage	ChrY coverage
45, XO TS	26.70%	67.74	96.0%	77.42	50.31	0.12
47, XXY KS	24.90%	22.61	89.0%	23.29	21.45	14.09
DSDPt1	25.63%	74.07	97.1%	81.09	59.79	65.92
DSDPt2	26.60%	91.78	98.4%	99.51	75.86	87.14
DSDPt3	31.22%	68.89	97.3%	77.50	51.41	60.19
DSDPt4	18.04%	79.00	97.7%	86.86	63.08	68.60
DSDPt5	28.12%	96.60	98.3%	103.44	82.38	94.27
DSDPt6	27.40%	96.80	97.1%	98.19	97.21	0.03
DSDPt7	50.87%	33.49	88.4%	40.66	19.22	21.33
DSDPt8	51.58%	71.48	97.1%	84.12	46.08	55.29
DSDPt9	54.68%	26.08	87.1%	30.14	17.91	21.00
DSDPt10	56.67%	64.75	96.0%	66.48	63.40	0.06
DSDPt11	53.71%	46.13	95.3%	53.99	30.36	35.01
DSDPt12	41.42%	47.66	91.4%	56.88	29.11	37.29

Table 3-4: Filtered SNVs and INDELs identified in DSD patients. High quality SNVs with coding consequences identified in the 16 patients were run through a series of filters such as Human Gene Mutation Database (HGMD) to identify SNVs and INDELs previously reported in the literature. All INDELs were reported as causative. Remaining SNVs were filtered using dbSNP132 (≥1% frequency) to identify rare variants. The identified rare SNV variants were then run through multiple *in silico* protein pathogenicity predictors to determining whether the coding variation was likely to be causative of the disease. Two of three pathogenicity predictors were required to indicate the mutation was "damaging" in order to classified as causative. [1]This mutation lies in the coding region of an isoform of DMRT1. *In silico* protein predictors do not provide pathogenicity for INDELs.

Sample	Chr	Start	End	Gene	SNV/ INDEL	SNP Quality	Protein Change	Causative Variant?	in HGMD PUBLIC?	in dbSNP132 (1%)?	polyphen-2 prediction	SIFT	Mutation Assessor
47, XXY KS	chr10	104586872	104586873	CYP17A1	T->G	34	H79P		N	N	possibly damaging	tolerated	low
47, XXY KS	chr8	106870217	106870218	ZFPM2	G->C	114	S210T		N	N	benign	tolerated	neutral
DSDPt1	chr2	31607980	31607981	SRD5A2	C->T	225	E200K	Y	Y	N			medium
DSDPt3	chr8	11651992	11651993	GATA4	G->A	228	V380M		N	rs114868912	benign	tolerated	neutral
DSDPt3	chr9	1046731	1046732	DMRT2	G->A	194	R382Q		N	rs72703237	benign	tolerated	neutral
DSDPt3	chr17	75373191	75373192	CBX2	C->T	228	A452V		N	rs76915888	benign	tolerated	neutral
DSDPt3	chr9	1046397	1046398	DMRT2	C->G	100	P271A		N	rs72703236	probably damaging	damaging	medium
DSDPt3	chrY	2715264	2715265	SRY	T->C	225	Y127C	Y	Y	N	probably damaging		
DSDPt5	chr12	52104848	52104849	AMHR2	A->C	136	T108P		N	N	possibly damaging	tolerated	neutral
DSDPt5	chr8	106870217	106870218	ZFPM2	G->C	228	S210T		N	N	benign	tolerated	neutral
DSDPt6	chr17	75373191	75373192	CBX2	C->T	228	A452V		N	rs76915888	benign	tolerated	neutral
DSDPt6	chrX	139414764	139414765	SOX3	C->T	228	A43T		N	N	benign		neutral
DSDPt7	chrX	66858443	66858444	AR	T->C	225	M788T	Y	Y	N	probably damaging	damaging	medium
DSDPt8	chrX	30237128	30237129	NR0B1	G->T	212	Y121*	Y	Y	N	termination	termination	termination
DSDPt9	chr9	884196	884197	DMRT1	T->G	46	V->G[1]		N	rs116766038	non-coding by CCDS	non-coding by CCDS	non-coding by CCDS
DSDPt9	chr15	72424436	72424437	CYP11A1	INS:-->A	208	Intronic splice-sitemutation	Y	N	N	Indel[2]	indel	indel
DSDPt9	chr15	72422525	72422526	CYP11A1	DEL:T->-	726	frameshift, early-termination	Y	Y	N	Indel	indel	indel
DSDPt11	chr15	72446744	72446745	CYP11A1	C->T	204	V79I		N	N	benign	tolerated	low
DSDPt12	chrX	76824270	76824271	ATRX	T->C	139	K1045E	Y	N	N	probably damaging	tolerated	low

REFERENCES

1 Lux, A., Kropf, S., Kleinemeier, E., Jurgensen, M. & Thyen, U. Clinical evaluation study of the German network of disorders of sex development (DSD)/intersexuality: study design, description of the study population, and data quality. *BMC Public Health* **9**, 110, (2009).

2 Hughes, I. A., Houk, C., Ahmed, S. F. & Lee, P. A. Consensus statement on management of intersex disorders. *J Pediatr Urol* **2**, 148-162, (2006).

3 This percentage is based on a systematic electronic medical chart review targeting patients, n. t. y., categorized as DSD at one major mid-western academic medical center.

4 Stein MT, S. D., Mazur T, Eugster E, Daaboul J. A newborn infant with a disorder of sexual differentiation. *Journal of Developmental & Behavioral Pediatrics*, 115-119 (2003).

5 Meyer-Bahlburg HFL, M. L. in *Encyclopedia of Endocrine Diseases* 125-134 (Elsevier, 2004).

6 Warne, G. L. Long-term outcome of disorders of sex development. *Sex Dev* **2**, 268-277, (2008).

7 Easton, D. F. *et al.* Cancer risks in two large breast cancer families linked to BRCA2 on chromosome 13q12-13. *Am J Hum Genet* **61**, 120-128 (1997).

8 Goldgar, D. E. *et al.* Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Genet* **75**, 535-544, (2004).

9 Kim, C. J. *et al.* Severe combined adrenal and gonadal deficiency caused by novel mutations in the cholesterol side chain cleavage enzyme, P450scc. *J Clin Endocrinol Metab* **93**, 696-702, (2008).

10 White, S. *et al.* Copy Number Variation in Patients with Disorders of Sex Development Due to 46,XY Gonadal Dysgenesis. *PLoS One* **6**, e17793, (2011).

11 Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27**, 182-189, (2009).

12 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, (2009).

13 O'Connor, B. D., Merriman, B. & Nelson, S. F. SeqWare Query Engine: storing and searching sequence data in the cloud. *BMC Bioinformatics* **11 Suppl 12**, S2, (2010).

14 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249, (2010).

15 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-1081, (2009).

16 Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*, (2011).

17 Chan, P. A. *et al.* Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). *Hum Mutat* **28**, 683-693, (2007).

18 Sathirapongsasuti, J. F. *et al.* Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics* **27**, 2648-2654, (2011).

19 Shearer, A. E. *et al.* Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21104-21109, (2010).

20 Sekido, R. & Lovell-Badge, R. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* **453**, 930-934, (2008).

21 Finkielstain, G. P. *et al.* Comprehensive genetic analysis of 182 unrelated families with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab* **96**, E161-172, (2011).

22 Tavtigian, S. in *The Role of Genetics in Breast and Reproductive Cancers Cancer Genetics* (ed P. Welcsh) Ch. 3, 49-73 (Spinger, 2009).

23 Ion, A. *et al.* A novel mutation in the putative DNA helicase XH2 is responsible for male-to-female sex reversal associated with an atypical form of the ATR-X syndrome. *Am J Hum Genet* **58**, 1185-1191 (1996).

24 Tannour-Louet, M. *et al.* Identification of de novo copy number variants associated with human disorders of sexual development. *PLoS One* **5**, e15392, (2010).

25 Holtzman, N. A., Murphy, P. D., Watson, M. S. & Barr, P. A. Predictive genetic testing: from basic research to clinical practice. *Science* **278**, 602-605 (1997).

26 Krone, N. & Arlt, W. Genetics of congenital adrenal hyperplasia. *Best Pract Res Clin Endocrinol Metab* **23**, 181-192, (2009).

27 Walsh, T. *et al.* Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci U S A* **107**, 12629-12633, (2010).

28 Raca, G., Jackson, C., Warman, B., Bair, T. & Schimmenti, L. A. Next generation sequencing in research and diagnostics of ocular birth defects. *Mol Genet Metab* **100**, 184-192 (2010).

29 Wheeler, M. *et al.* A new era in clinical genetic testing for hypertrophic cardiomyopathy. *J Cardiovasc Transl Res* **2**, 381-391, (2009).

30 Geelhoed, E. A., Harrison, K., Davey, A. & Walpole, I. R. Parental perspective of the benefits of genetic testing in children with congenital deafness. *Public Health Genomics* **12**, 245-250, (2009).

Chapter 4

Mutations in the PCNA-binding domain of *CDKN1C* cause IMAGe Syndrome

IMAGe Syndrome (Intrauterine growth restriction, Metaphyseal dysplasia, Adrenal hypoplasia congenita, and Genital anomalies) is an undergrowth developmental disorder with life-threatening consequences[1]. Identity-by-descent analysis in a family with IMAGe syndrome[2] identified a 17.2 megabase (Mb) locus on 11p15 that segregated in affected family members. Targeted exon array capture of the disease locus, followed by high-throughput genomic sequencing and validated by dideoxysequencing, identified missense mutations in imprinted gene *CDKN1C (P57KIP2)* in two familial and four unrelated patients. Familial analysis demonstrated an imprinted mode of inheritance where only maternal transmission of the mutation resulted in IMAGe syndrome. *CDKN1C* inhibits cell-cycle progression[3] and targeted expression of IMAGe-associated *CDKN1C* mutations in *Drosophila* caused severe eye growth defects compared to wild type *CDKN1C*, suggesting a gain-of-function mechanism. All IMAGe-associated mutations clustered in the PCNA-binding domain of *CDKN1C* and resulted in loss of PCNA binding, distinguishing them from *CDKN1C* mutations that cause Beckwith-Wiedemann Syndrome, an overgrowth syndrome[4].

Materials and Methods

Study Subjects. All participants were patients diagnosed clinically with IMAGe syndrome. This study was approved by the Institutional Review Board (IRB) at the University of California, Los Angeles, the Hospital de Niños Ricardo Gutierrez in Argentina, or the Institute of Child Health at University College London. All participants provided informed consent. Phenotypes are summarized in Table 1.

Identity-by-descent analysis. Genomic DNA from 7 affected individuals (IV-10, V-1, V-2, V-5, V6, V-7, V-12) and 1 non-affected individual (V-13) from a large Argentine family (Family A) were genotyped on the Affymetrix 250K *Nsp1* SNP arrays, as per manufacturers' protocol. Familial relationships were confirmed through checking the sharing statistics in all pairs of samples and ancestral identity-by-descent (IBD) analysis was performed using a custom script (B. Merriman, available on request). The IBD analysis script was designed to search for long continuous intervals compatible with a common extended haplotype among all the affected individuals but not with the unaffected individuals.[29] A conservative error rate of 1% was used to allow the algorithm to tolerate possible genotyping errors. A rare dominant model of inheritance was assumed with rare frequency of this haplotype of 0.1% and penetrance of 100% under the assumption that the carriers (parents of the affected individuals) were not showing the phenotype for a different reason (i.e. imprinting) than low penetrance of the disease phenotype.

Capture and Sequencing of Genomic DNA. For the capture of the genes within the IBD interval, we used Agilent Custom 244K comparative genomic hybridization Array. The 60-bp oligonucleotide probes were tiled every 20-bp against all exonic regions on chromosome 11 between 2.45-20.15Mb (hg18, March 2006, build 36.1) and every 30-bp in the flanking 5' and 3' regions spanning 0-2.45 Mb and 20.15-22.6 Mb. We included all gene models identified in RefSeq, Genbank, CCDS and UniProt. Location and sequence of all probes are available upon request.

Genomic DNA libraries were created for patients V-1 an V-6 from Family A and 3 isolated cases of IMAGe[1] (Patient1, 2, and 3) following manufacturer's protocol (Illumina protocol Preparing Samples for Sequencing Genomic DNA, p/n 11251892 Rev. A) except for the adaptor ligation step where we used custom-made, internally validated bar-coded adaptors. After

the PCR amplification, five bar-coded libraries were pooled together and captured by hybridization to custom-designed CGH array, as described in [30]. After the capture, the array was washed and the captured DNA was eluted, amplified and diluted to 10nM final concentration based on the Qubit concentration measure and Agilent Bioanalyzer. One flowcell lane of single-end sequencing was performed at the UCLA Genomic Sequencing Center on the Illumina Genome Analyzer II for 76x cycles following manufacturer's protocol. The base-calling was performed by the real time analysis software (Illumina).

Sequence Data Analysis. The Illumina output files (.qseq) were converted to fastq formats using BFAST[31,32] script ill2fastq.pl and then parsed into multiple fastq files, each for one unique barcode. Only the reads with 100% matching barcode sequences were carried over to the second set of fastq files. The sequence reads in each fastq file were aligned to the human reference genome (hg18, March 2006, build 36.1) using Novoalign from Novocraft Short Read Alignment Package. The output format was set to SAM and default settings were used for all options. Using SAMtools, the SAM file of each sample was converted to a BAM file, sorted and merged and potential PCR duplicates were removed using Picard[33]. The variants, both single nucleotide variants (SNVs) and small INDELs (insertions and deletions), within the captured coding exonic intervals were called using the SAMtools pileup tool. For SNV calling, the last 5 bases were trimmed and only the reads lacking INDELs were retained. For INDEL calling, only the reads that contained one contiguous INDEL, not occurring on either end of the read, were retained[34]. The variants were further annotated using the SeqWare project and loaded into the SeqWare QueryEngine database[35]. Variants from each sample with the following criteria were identified: (i) variant base or INDEL observed at least twice and at ≥5% of the total coverage per base, (ii) variant observed at least once on both forward and reverse strands, (iii) SNV quality score ≥10.

As IMAGe syndrome is a rare condition, only variants with coding consequences not present in dbSNP129 were further analyzed.

***CDKN1C* Sequencing.** All *CDKN1C* (RefSeq: NM_000076.2) mutations were PCRed using Phusion HF polymerase (NEB) with 5% DMSO and 0.1M Betaine. PCR products were sent for dideoxysequencing at Laragen, Inc. PCR primers used are listed in Supplementary Table 1. All mutation locations were reported using CCDS7738.1 and P49918 as the normal transcript and protein sequences, respectively.

Plasmid Constructs. The cDNA of CDKN1C cloned into pBluescript was purchased from ATCC(#99411)[36]. Mutations were generated using site-directed mutagenesis (Stratagene, primers in Supplementary Table 1). 4 mutants were created, corresponding to BWS mutations (p.L42P and c.826delT) and IMAGe syndrome mutations (p.F276V and p.K278E). Wild type and mutant versions of *CDKN1C* were subcloned into pCDNA3.1(+) and pEGFP-C2 for mammalian cell culture experiments, into pFLAG-pcDNA3.1 for the immunoprecipitation, and into pUAST for the generation of transgenic flies.

Quantitative Reverse Transcription PCR. Human tissue from 7-8 weeks post-conception (wpc) was provided by the Medical Research Council/Wellcome Trust-funded Human Developmental Biology Resource with Research Ethics Committee approval and informed consent. RNA was extracted from adrenal, brain and muscle samples using the TRIzol reagent and first-strand cDNA was generated using SuperScript® II reverse transcriptase (Invitrogen). Expression of *CDKN1C* transcript was assessed by quantitative PCR using the StepOnePlus™ Real-time PCR System, TaqMan® Gene Expression Assays for human *CDKN1C* (Hs00175938_m1) and human *GAPDH* as endogenous control (4333764T; all Applied Biosystems). Data were analyzed with StepOne software v2.1 according to the $2^{-\Delta\Delta CT}$ method.

Immunofluorescence. Fourteen-micron sections of human fetal adrenal tissue (8 weeks post conception) were fixed briefly in 4% PFA and blocked in 1% BSA before incubating overnight with antibody to CDKN1C (Fisher AFMA121866) and CYP11A1 (Sigma HPA016436). Primary antibodies were detected using Alexa647 goat anti-mouse (Invitrogen, A21235) and Alexa555 (Invitrogen, A21429) goat anti-rabbit conjugates. Nuclei were counterstained with DAPI. IMAGes were collected on a Zeiss 710 confocal microscope (Carl Zeiss).

Immunoprecipitation. HEK293T cells were transfected with constructs encoding Flag-CDKN1C with Lipofectamine 2000 (Invitrogen). For ubiquitination assays, cells were co-transfected with pCI-neo-$(HA)_3$-human Ub[37] construct and treated 3 hours in 10μM MG-132 (Millipore) prior to cell lysis. Flag-CDKN1C was immunoprecipitated from cell lysates using the ANTI-FLAG M2-Agarose Affinity Gel (Sigma-Aldrich). Western blot was performed on immunoprecipitated samples and cell lysates with primary antibody to HA (Covance MMS101R), Flag (Abcam, ab1162), PCNA (Abcam ab29), and CDKN1C (Santa Cruz biotechnology, sc-1040). Secondary antibodies used were goat-anti-mouse-HRP (Biorad, 1:20,000) and goat-anti-rabbit-HRP (Santa Cruz, 1:5,000).

Drosophila **experiments.** Five independent constructs, $CDKN1C^{WT}$, $CDKN1C^{L42P}$, $CKDN1C^{826delT}$, $CDKN1C^{F276V}$, and $CDKN1C^{K278E}$ were injected into embryos and each construct generated multiple independent transgenic lines. Overexpression was achieved by using the GAL4-UAS system[17] and the following drivers: *Ubi-gal4* (ubiquitous expression), *ey-gal4* (eye-specific expression), *MS1096-gal4* (wing-specific expression), and sal^{PE}*-gal4* (wing pouch-specific expression). *MS1096-gal4* is expressed in the entire wing imaginal disc but at higher levels in the dorsal compartment. This higher level of expression in the dorsal compartment causes enhanced phenotypic effects in the dorsal versus the ventral side of the wing.

sal^{PE}*-gal4* is specifically expressed in the pouch of the wing disc, which only gives rise to the wing proper. All UAS-*CDKN1C* constructs were larval lethal when expressed with *Ubi-gal4*, confirming their expression efficiency. At least two independent lines were used for each experiment and all yielded similar results. All eye images were taken on a Leica Z16 AP0 Camera. Wing IMAGes were taken on Leica DFC 300 FX R2 Camera.

Flow Cytometry. Wild-type or mutant versions of *CDKN1C* were co-transfected with pCMV-GFP into serum starved HEK293T cells using Lipofectamine 2000 (Invitrogen). After 24 hours, cells were grown in media containing 5% serum for 48-hours. Cells were resuspended in a hypotonic buffer with propidium iodide and GFP-positive events were analyzed on a Becton Dickinson FACScan Analytic Flow Cytometer. All experiments were performed in 2 biological replicates. Statistical significance was assessed using a 2-proportion z-test. Flow cytometry was performed in the UCLA Johansson Comprehensive Cancer Center (JCCC) and Center for AIDS Research Flow Cytometry Core Facility.

Analysis of nuclear/cytoplasmic distribution of GFP-fusion proteins. H295R cells were transfected using Lipofectamine 2000. After incubation for 24 hours, cells were fixed in 4% paraformaldehyde and IHC was performed using an anti-GFP antibody (Invitrogen), a FITC-labeled secondary antibody (Jackson Lab), and a mounting medium with DAPI (VectorLabs). Cells were imaged on an Olympus AX70 microscope.

M1 fibroblasts were transfected with 2ug/well of purified expression plasmid using the X-tremeGENE HP DNA transfection reagent (Roche). 24-hours after transfection, cells were fixed in 2% formaldehyde and analyzed using fluorescence microscopy. Fluorescent and bright-field images of randomly selected fields containing transfected M1 cells were saved electronically using a 'blind-code' file nomenclature. The distribution of GFP-fusion proteins in each

transfected cell was annotated by an experienced observer (who was unaware of the blind code) as one of the following: (1) nuclear only, (2) nuclear and cytoplasmic, and (3) cytoplasmic only.

Results

Since the initial description of IMAGe syndrome (OMIM #300290)[1], a number of isolated and familial cases have been reported[1,5-11]. In order to identify a causative gene for IMAGe syndrome, we performed a 250K *Nsp* Affymetrix SNP Array on seven affected and one unaffected sibling from Family A, a five-generation family from Argentina (Fig. 4-1a). Further analysis using a custom script detected a 17.2 Mb identical-by-descent (IBD) region on chromosome 11, shared by seven affected family members but not an unaffected sibling (Fig. 4-1b), with a LOD score of 5.4. Despite the multisystem involvement of IMAGe syndrome, we did not identify a contiguous gene deletion or duplication in the affected individuals (Fig. 4-2).

To determine the causative mutation, we performed targeted high-throughput genomic sequencing of all the exons within a conservative IBD region. An Agilent 244K custom CGH array was designed to capture all exons and splice sites within the region spanning 0-22.6 Mb on chromosome 11. In total, five custom bar-coded genomic DNA libraries (Family A, V-1 and V-6; unrelated Patients 1, 2, and 3) were prepared, pooled, and captured on a single custom array, and the DNA enriched for the IBD region was sequenced on one lane of the Illumina Genome Analyzer II. Our targeted approach yielded an average coverage of 32x. Patient 3 had a significantly lower coverage of 9x across the targeted intervals and was not used in the initial bioinformatics analysis. In the remaining four samples analyzed, ~85% of all targeted regions were covered at ≥10x.

The pedigree and IBD analysis led us to hypothesize that IMAGe syndrome was inherited

as a rare autosomal dominant disorder. Our bioinformatics analysis required that both V-1 and V-6 share the same rare gene variant, and at least one of the non-related IMAGe syndrome patients harbor a rare variant (defined as a variant not present in dbSNP129) in the same gene. This approach identified a single gene, *CDKN1C*.

Upon further examination, we noted that *CDKN1C* was captured and sequenced at a much lower rate compared to other targeted genes due to a high GC-content of ≥80%. To compound this low gene coverage, we re-sequenced *CDKN1C* by dideoxysequencing (primers listed in Supplementary Table 1) in all five individuals sequenced by high-throughput sequencing and in an additional sporadic case (Patient 4). Affected individuals from Family A carried a c.825T>G change resulting in a F276V missense mutation. The four unrelated patients with IMAGe syndrome harbored four mutations in *CDKN1C*: F276S, R279P, D274N, and K278E (Table 1). In total, we identified five rare heterozygous missense mutations in *CDKN1C* that cluster within six amino acids of the PCNA-binding domain[12] (Fig. 4-3a). All variants localized to a highly conserved region[13] (Fig. 4-3b) and are predicted to be "damaging" to the structure and function of *CDKN1C* by Polyphen analysis[14].

CDKN1C is located on chromosome 11 and encodes a protein known to play a key role in inhibiting cell cycle progression. In most tissues, the paternal allele is repressed by distant imprinting control regions, such that expression is primarily from the maternal allele[15,16]. Inheritance of IMAGe syndrome in Family A was only through maternal transmission of the *CDKN1C* mutation (Fig. 1a). Sequencing for the c.825T>G mutation in 24 members from Family A confirmed that only individuals who inherited the mutation on the maternal allele are affected. A c.825T>G mutation inherited on the paternal allele was not expressed, presumably because of epigenetic silencing of the mutated allele.

To confirm the pathogenicity of these mutations, we used an *in vivo* functional model in which IMAGe-associated human *CDKN1C* mutants were expressed in *Drosophila melanogaster* using the GAL4 UAS system[17]. Ubiquitous overexpression of wild type or mutant *CDKN1C* resulted in early larval lethality. Expression of wild type *CDKN1C* restricted to the eye did not have any effects on adult eye size, while expression of IMAGe-associated *CDKN1C* mutants showed a moderate to severe reduction in eye size (Fig. 4-4). Targeted expression of the IMAGe-associated *CDKN1C* mutants resulted in altered wing vein patterning and decreased wing size (see Fig. 4-5 & 4-6).

Overexpression of IMAGe-associated *CDKN1C* mutants in HEK293T cells did not interfere with the ability of CDKN1C to inhibit the cell cycle in G0/G1 through binding of the CDK-domain (Fig. 4-7). These data suggest that IMAGe-mutations within the PCNA-binding domain do not inhibit cell-cycle progression and likely act through a different mechanism resulting in IMAGe syndrome.

CDKN1C is located within an imprinted cluster of genes that regulate prenatal and postnatal growth and development. Genetic alterations in *CDKN1C* have been shown to give rise to Beckwith-Wiedemann Syndrome (BWS; OMIM #130650), an overgrowth disorder[18,19] Here, we show that an undergrowth condition, IMAGe syndrome, is caused by domain-specific mutations in the maternally inherited allele of *CDKN1C*.

Both BWS human patients and *CDKN1C*$^{-/-}$ knockout mice exhibit adrenal hyperplasia[20,21] in contrast to the adrenal hypoplasia in IMAGe patients. We therefore verified that *CDKN1C* mRNA and protein are expressed in the developing human adrenal gland (Fig. 4-8). Quantitative RT-PCR demonstrated that the expression of *CDKN1C* is greater in adrenal tissue during early human development than in brain or muscle (Fig. 4-8a). Immunohistochemistry showed

strongest expression of *CDKN1C* within a subset of cells in the subcapsular or developing definitive zone of the adrenal gland (Fig. 4-8b).

To determine if BWS and IMAGe mutations work through the same mechanism, we repeated the above functional studies with BWS-specific mutants. In the cell cycle analysis, transfection of the BWS-mutant *CDKN1C-L42P* resulted in a loss of cell cycle inhibition at G0/G1. Ubiquitous expression of BWS-associated *CDKN1C* mutations in *Drosophila melanogaster* were early larval lethal, however targeted expression had no effect on eye size, wing size, or wing vein patterning (Fig. 4-4g,h, 4-5e,5f,4-6e,6f). Thus, *in vitro* and *in vivo*, BWS mutants have different effects relative to the IMAGe mutants suggesting that domain-specific mutations have differential effects on cell cycle progression and developmental processes.

BWS-associated *CDKN1C* mutations are either missense mutations localized to the cyclin-dependent kinase binding domain or nonsense mutations, both of which result in protein loss-of-function, over proliferation, and predisposition to cancer[22] due to loss of cell-cycle inhibition[3]. In contrast, we show that missense mutations localized to a highly conserved region of the PCNA-binding domain of *CDKN1C* in IMAGe syndrome resulted in excess inhibition of growth and differentiation—a gain of function (Fig. 4-9).

Since two of the five mutations fall into a putative nuclear localization signal, we determined whether IMAGe mutants interfere with active nuclear transport of CDKN1C. H295R and M1 cells transfected with GFP-CDKN1C fusion constructs (Fig. 4-10), showed that none of the tested IMAGe mutants interfered with active transport mechanisms or with binding affinity to α-importin.

Since IMAGe mutations cluster in a domain known to bind PCNA, we performed co-immunoprecipitation experiments to test the effect of IMAGe mutations on PCNA binding.

HEK293T cells were transfected with flag-tagged CDKN1C constructs bearing the wild type or IMAGe alleles (F276V and K278E). Endogenous PCNA was recovered from the wild type but not IMAGe immunoprecipitates (Fig. 4-8c), suggesting that PCNA binding is disrupted in mutants.

Since one of the roles of *PCNA* is to facilitate ubiquitination of cell cycle proteins[23], we investigated the role of IMAGe mutations in PCNA-dependent ubiquitination of CDKN1C. HEK293T cells were co-transfected with flag-tagged wild type or IMAGe-mutant F276V CDKN1C and with HA-tagged Ubiquitin (12 kDa) and subjected to co-immunoprecipitation. CDKN1C migrates at ~50 kDa, and therefore we expect the mono-, di-, and poly-ubiquitinated CDKN1C to migrate at ~62 kDa or higher depending on the number and the branching of the ubiquitin moieties. Here we show that a band at 63 kDa, the approximate size of a monoubiquitinated CDKN1C protein, is present in the wild type CDKN1C but absent in the IMAGe mutant sample (Fig.4-8d).

Our data reveal a role for PCNA binding in a specific ubiquitination modification of *CDKN1C*. Many cell cycle proteins are subject to PCNA-dependent ubiquitination[24], which has pleiotropic effects. Monoubiquitination, as observed in our data, may have a number of functional consequences, such as modulation of protein localization, of protein interactions, and of proteosomal degradation[25,26,27]. The latter is less likely because it typically requires, at minimum, tetra-ubiquitination[28], but cannot be ruled out without information on CDKN1C protein stability**.**

Next generation sequencing has emerged as a powerful tool in identifying rare Mendelian disease genes, utilizing existing linkage analysis data to identify a disease gene in an unbiased fashion. Our findings show that missense mutations in the PCNA-binding domain have an

inhibitory effect on growth, via loss of binding of PCNA to CDKN1C, thereby altering ubiquitination of CDKN1C and presumably promoting its function. The contrast between BWS and IMAGe mutations in *CDKN1C* highlights the dual and opposing effects of specific *CDKN1C* mutations. Mutations within the PCNA-binding site of *CDKN1C* blocked *in vivo* growth and differentiation and may illuminate novel mechanisms regulating cell transformation[12], tumor growth, and cell cycle progression.

FIGURES

Figure 4-1. Identity-by-descent analysis in a family with IMAGe syndrome. (a) In this large family (Family A) with IMAGe syndrome[2], 24 individuals were tested for genetic mutations in *CDKN1C*. All affected individuals we tested carried the c.825T>G mutation on the maternally inherited allele resulting in a F276V amino acid change. Unaffected carriers all inherited the mutations on the paternal alleles. (b) IBD analysis identified a region on chromosome 11 spanning from base pairs 2,685,916 to 19,809,755 (according to hg19) that was shared by affected family members (IV-10, V-1, V-2, V-5, V-6, V-7, V-12) and different from an unaffected sibling (V-13).

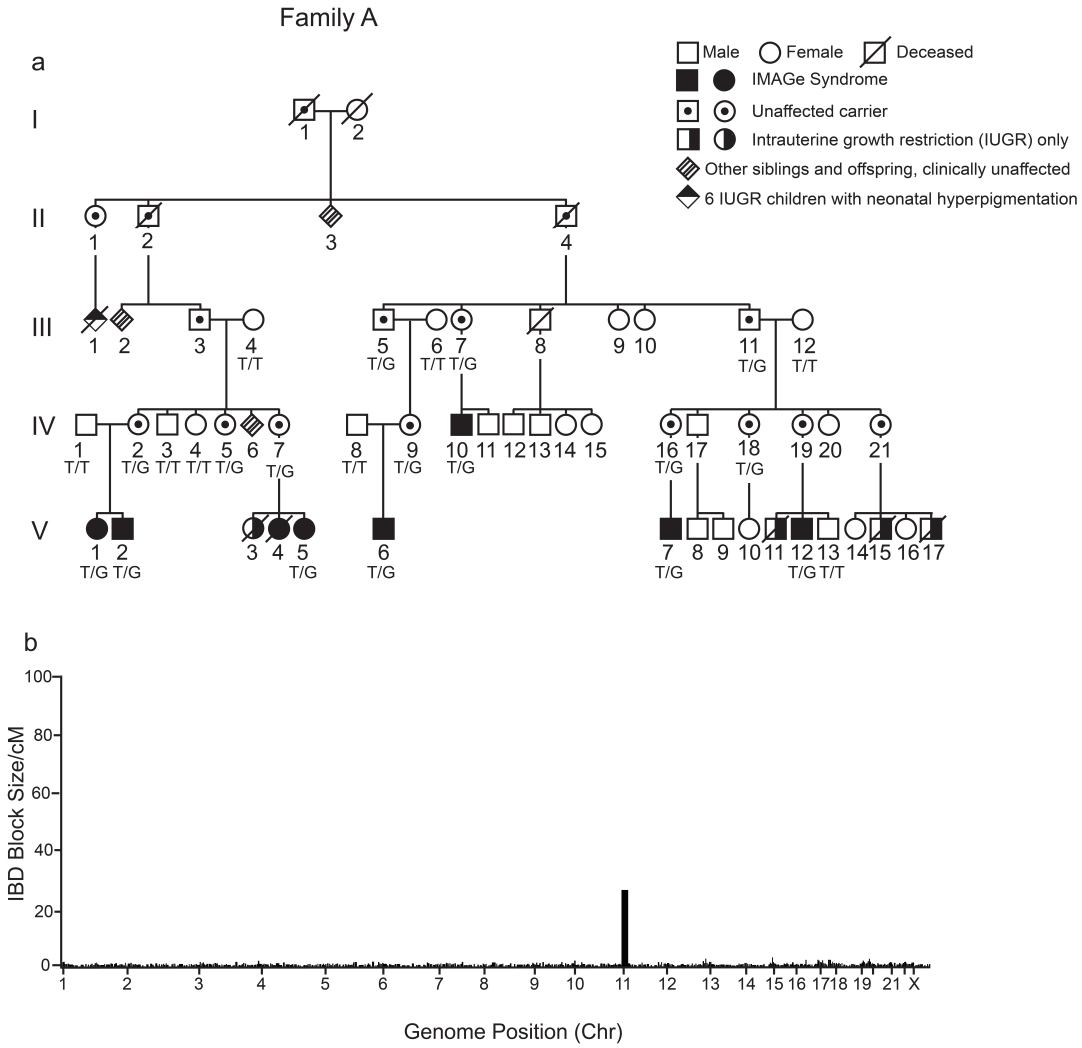


Figure 4-2. No shared contiguous gene deletions or duplications in affected IMAGe syndrome individuals. Using Affymetrix CNAT (copy number analysis tool)[1], the relative copy number was calculated to that of unaffected individual V-13 for each sample across the genome. CNAT calls the copy number variant regions by scanning the relative ratio of the normalized intensity of each probe in two samples, then smoothing and segmenting the signals. Sample V-7 was not included in the analysis due to a low call rate. The identical-by-descent region (chr11: 2,685,916-19,809,755, in human reference genome hg19) is shown above and no contiguous duplications or deletions were shared among affected family members.

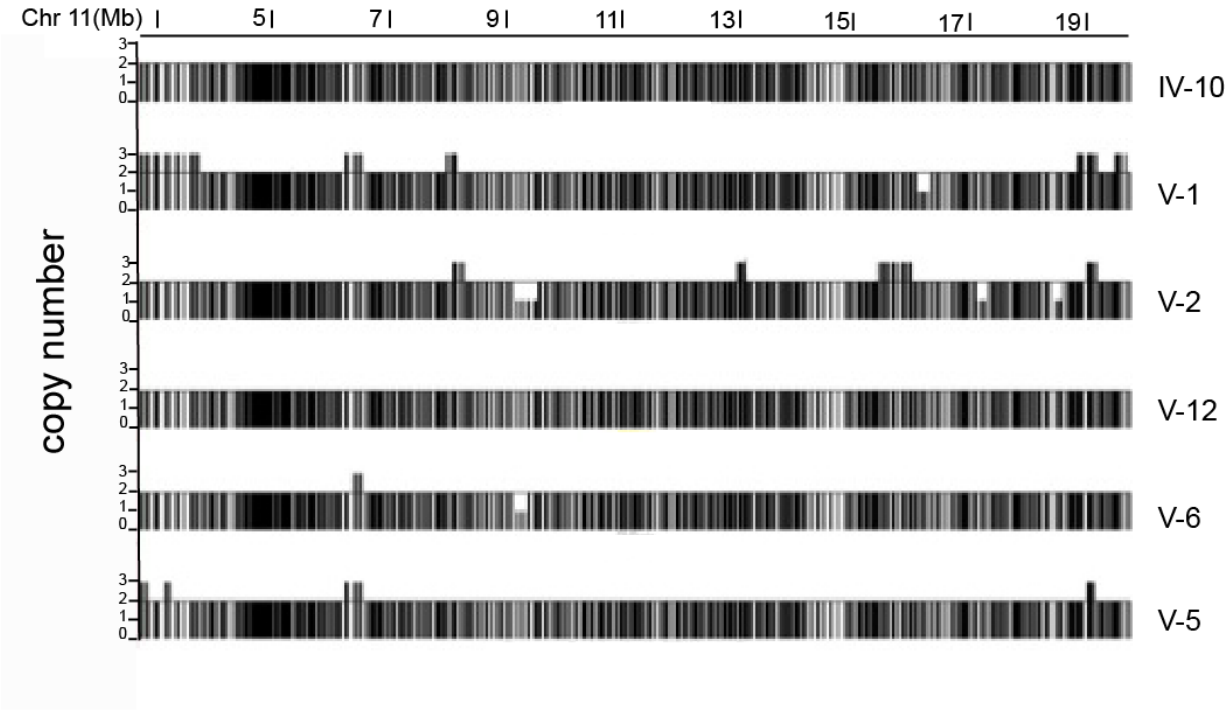


Figure 4-3. Localization of IMAGe syndrome mutations in *CDKN1C*. (a) All IMAGe syndrome mutations are localized to the region of the gene encoding the PCNA-binding domain. (b) Highly conserved amino acid residues (indicated by *) in the PCNA-binding domain of *CDKN1C* are conserved down to *D. rerio*. All IMAGe mutations affect highly conserved residues. A colon (:), indicates conservation between groups of strongly similar properties; a period indicates conservation between groups of weakly similar properties.

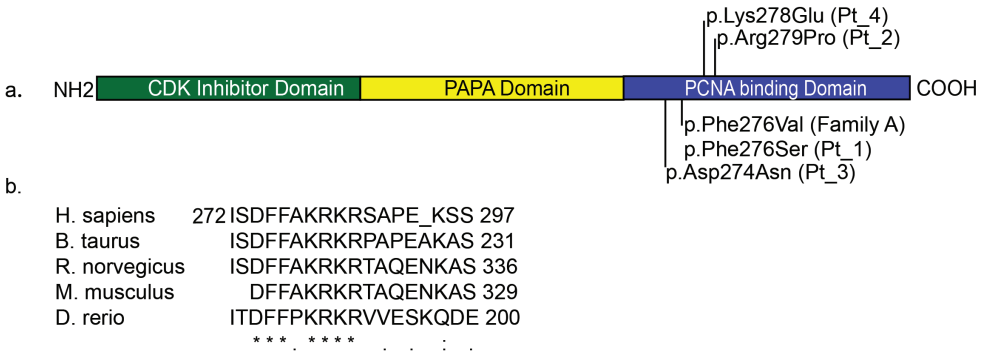


Figure 4-4. Phenotypic validation of IMAGe syndrome-associated mutations in *Drosophila melanogaster*. (**a**) Light microscope IMAGe of an adult wild type eye (*ey*>+). Scale bar, 25μm. (**b**) Eye-specific expression of human *CDKN1C* (*ey*>*CDKN1C*^{WT}) did not affect final size. (**c-g**) Overexpression of IMAGe syndrome-associated human *CDKN1C* mutations F276V (**c-d,** *ey*>*CDKN1C*^{F276V}) or K278E (**e-f,** *ey*>*CDKN1C*^{K278E}) in the developing eye led to a gain-of-function phenotype of moderate (**c**, **e**) to severely (**d**, **f**) restricted eye growth. (**g-h**) Overexpression of human *CDKN1C* carrying Beckwith-Wiedemann syndrome-associated mutations, c.826delT (**g,** *ey*>*CDKN1C*^{826delT}) or L42P (**h,** *ey*>*CDKN1C*^{L42P}), did not affect eye growth.

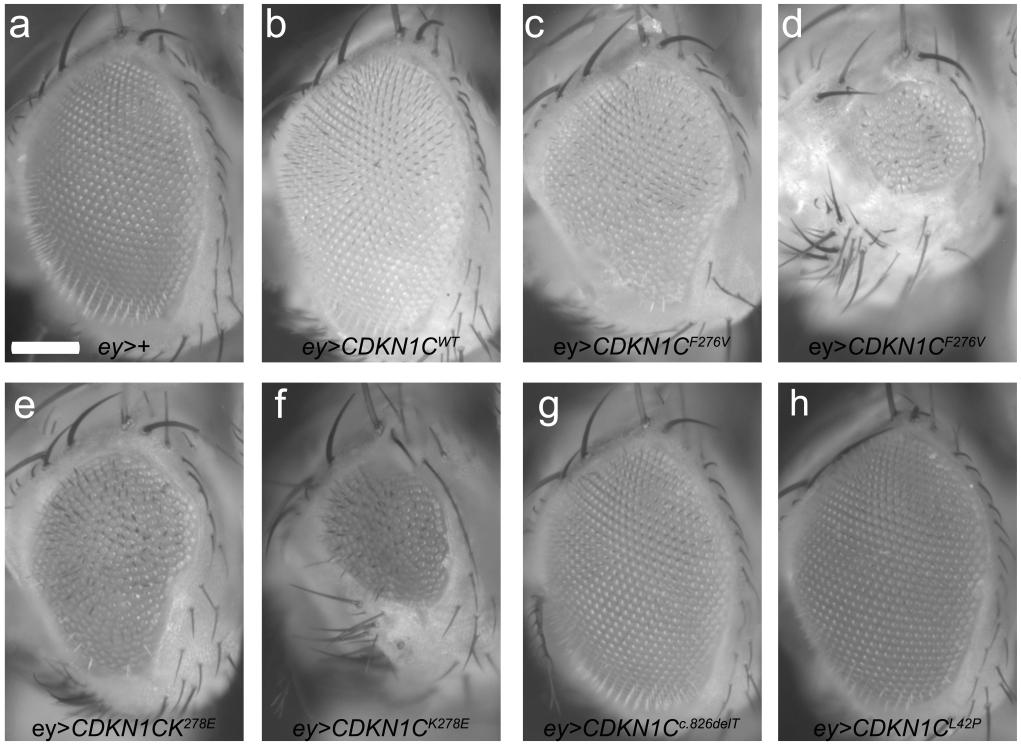


Figure 4-5. IMAGe syndrome-associated *CDKN1C* mutations alter wing venation in *Drosophila melanogaster*. sal^{PE}*-gal4* drives wing pouch-specific expression (**a**) Wild-type adult wing (sal^{PE}>+). Scale bar= 0.5 mm. (**b**) Expression of wild-type human *CDKN1C* in the developing wing pouch (sal^{PE}>$CDKN1C^{WT}$) did not affect vein patterning. Overexpression of IMAGe syndrome-associated mutations F276V (**c,** sal^{PE}>$CDKN1C^{F276V}$) or K278E (**d,** sal^{PE}> $CDKN1C^{K278E}$) in the developing wing pouch caused similar phenotypes of L2 vein truncation (white arrowhead) and anterior cross vein elimination (black arrowhead) while reducing the overall spaces between veins of the wing anterior compartment (top of each picture). In contrast, expression of mutant human *CDKN1C* with Beckwith-Wiedemann syndrome-associated mutations c.826delT (**e,** sal^{PE}>$CDKN1C^{826delT}$) or L42P (**f,** sal^{PE}>$CDKN1C^{L42P}$) did not alter vein patterning.

Figure 4-6. Overexpression of IMAGe syndrome-associated *CDKN1C* mutations restricts *Drosophila* wing growth. Side (a-f), top (a'-f'), and whole wing (a"-f") views are displayed. *MS1096>* drives wing-specific expression. (**a-a"**) Wild-type adult wing (*MS1096>+*). Scale bar= 0.5 mm. (**b-b"**) Expression of wild-type human *CDKN1C* in the developing wing (*MS1096>CDKN1C*^{WT}) did not affect final size. Overexpression of IMAGe syndrome-associated mutations F276V (**c-c",** *MS1096>CDKN1C*^{F276V}) or K278E (**d-d",** *MS1096>CDKN1C*^{K278E}) in the developing wing caused similar phenotypes of severely reduced wing size. In contrast, expression of mutant human *CDKN1C* with Beckwith-Wiedemann syndrome-associated mutations 826delT (**e-e",** *MS1096>CDKN1C*^{826delT}) or L42P (**f-f",** *MS1096>CDKN1C*^{L42P}) did not affect wing size.

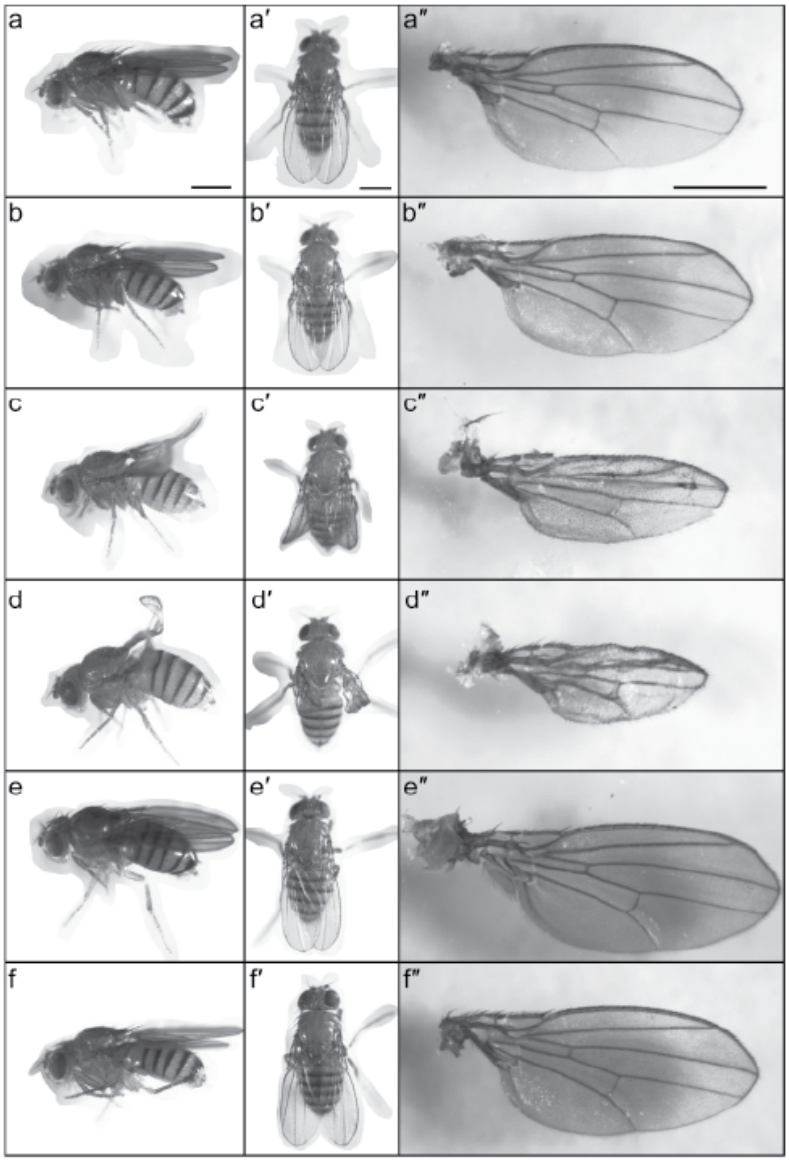
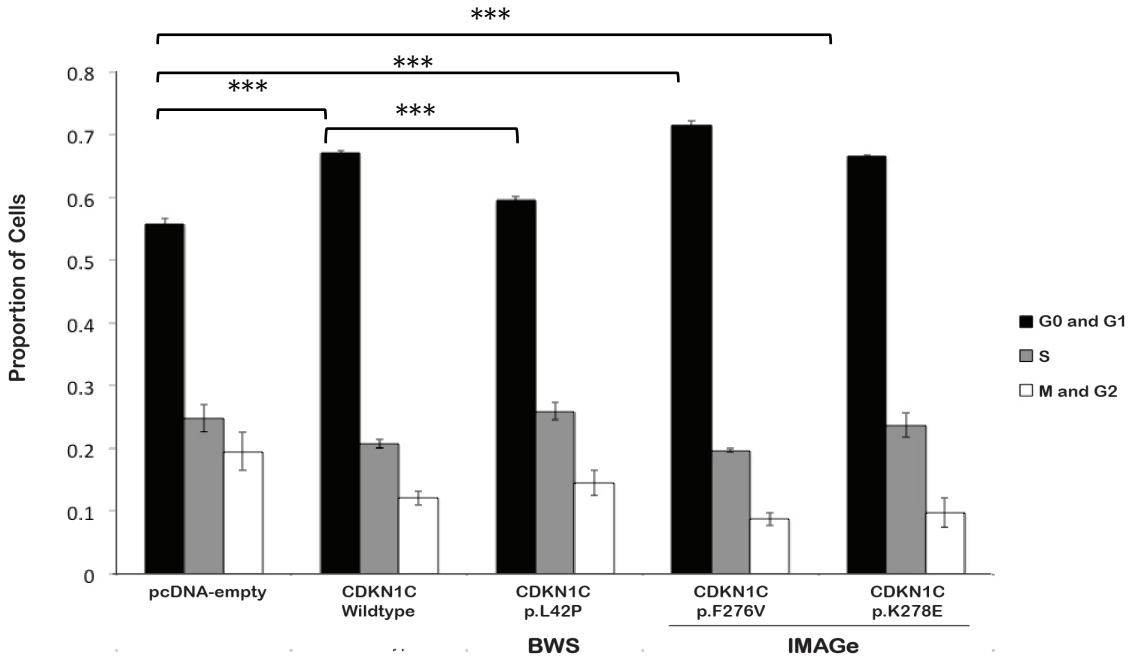

a
a′
a″
b
b′
b″
c
c′
c″
d
d′
d″
e
e′
e″
f
f′
f″

Figure 4-7. HEK293T cells transfected with wild-type and IMAGe *CDKN1C* mutants are stalled in G1. (a)We transiently transfected HEK293T cells with pcDNA-empty, $CDKN1C^{wildtype}$, the BWS mutant $CDKN1C^{L42P}$, or the IMAGe mutants $CDKN1C^{F276V}$ and $CDKN1C^{K278E}$ and performed cell cycle analysis through measurement of DNA content with propidium iodide. The representative distribution of 10,000 cells to stages of the cell cycle is shown in a. Wild-type *CDKN1C* overexpression resulted in an increased proportion of cells stalled in the G0/G1 phase compared to the empty control vector. Compared to the wild-type *CDKN1C,* the L42P mutant displayed a loss in the ability to inhibit the cells in G0/G1, consistent with loss-of-function of the cyclin-dependent kinase (CDK) binding domain[3]. Interestingly, neither IMAGe-associated mutant (F276V, K278E) showed different levels of cell cycle inhibition compared to wild-type *CDKN1C.* Both IMAGe mutants exhibited increased proportion of cells in the G0/G1 phase compared to the control empty vector. *** = p-value <0.001; (b) Differences between the sample and control proportion of cells (sample minus control) in G0 and G1, S, and M and G2 cell-cycle phases.

a



b. Difference in proportion of cells in cell-cycle phases (Sample - Control)

		G0 and G1	S	M and G2
Control	pcDNA-empty	0	0	0
Wild type	CDKN1c	0.114	-0.040	-0.074
BWS Mutant	L42P	0.038	0.011	-0.050
IMAGe Mutant	F276V	0.158	-0.051	-0.108
IMAGe Mutant	K278E	0.108	-0.010	-0.098

Figure 4-8. *CDKN1C* is expressed in the developing human adrenal gland and IMAGe mutants lose PCNA-binding, altering ubiquitination of CDKN1C. (a) Taqman RT-PCR showed higher expression of *CDKN1C* in the adrenal gland at 7-8 weeks post conception compared to control samples (muscle, brain). Error bars represent 1 s.d. (b) Immunohistochemistry showed *CDKN1C* protein expression in the developing adrenal gland at 8 weeks post conception. Nuclear expression of *CDKN1C* was observed in a subset of cells in the subcapsular region (arrowhead) and developing definitive cortex (arrow). Expression of the steroidogenic enzyme CYP11A1 is shown, predominantly in the fetal zone of the developing gland. Nuclear counterstaining was performed with DAPI. Scale bar, 200 μm. (c) HEK293T cells transfected with flag-tagged wild type or IMAGe mutant (F276V or K278E) CDKN1C were immunoprecipitated with Flag antibody. Left panel: whole lysates for wild type and non-transfected cells. Right panel: Flag immunoprecipitated fraction binds with endogenous PCNA in wild type but not in IMAGe mutants F276V and K278E. Flag-BAP: Bacterial Alkaline Phosphatase. (d) Immunoprecipitation of HEK293T cells co-transfected with Flag-CDKN1C (wild type or F276V) and HA-Ubiquitin constructs. Top panel: Immunoblot with HA antibody identifies a band at ~63kDa (indicated by *) in the CDKN1C wild type that is not present in the

276V IMAGe mutant. Bottom panel: Immunoblot of Flag protein.

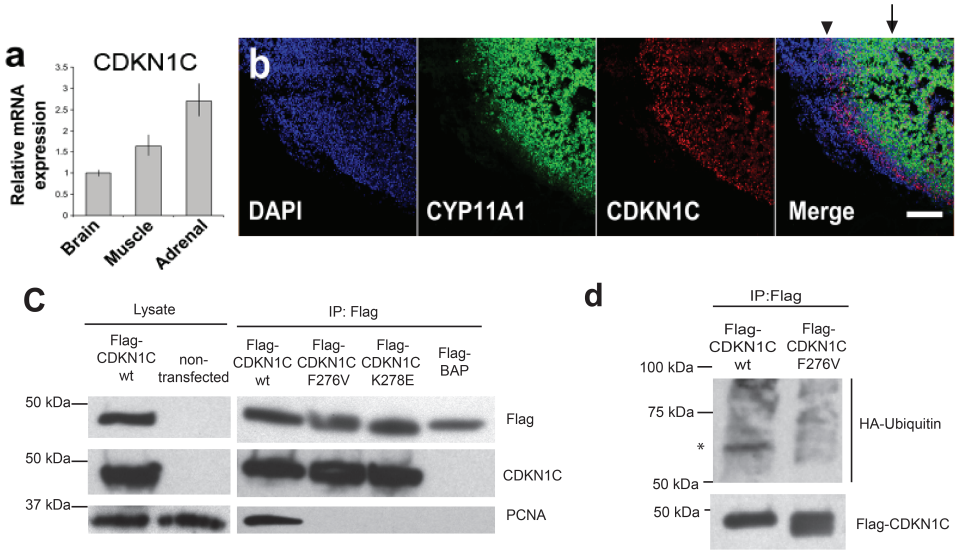


Figure 4-9. Missense mutations in CDK-binding domain and truncating mutations in CDKN1C cause Beckwith-Wiedemann Syndrome while missense mutations localized to the PCNA-binding domain result in IMAGe syndrome. Comparison between *CDKN1C* mutations resulting in IMAGe Syndrome (black arrowheads located above gene) or in Beckwith-Wiedemann Syndrome (below; red arrowheads = missense mutations; yellow arrowheads = truncating mutations). BWS mutations are either missense mutations primarily located in the cyclin-binding domain or truncating mutations, while IMAGe syndrome mutations are all missense mutations localized to a highly conserved region of the PCNA-binding domain.

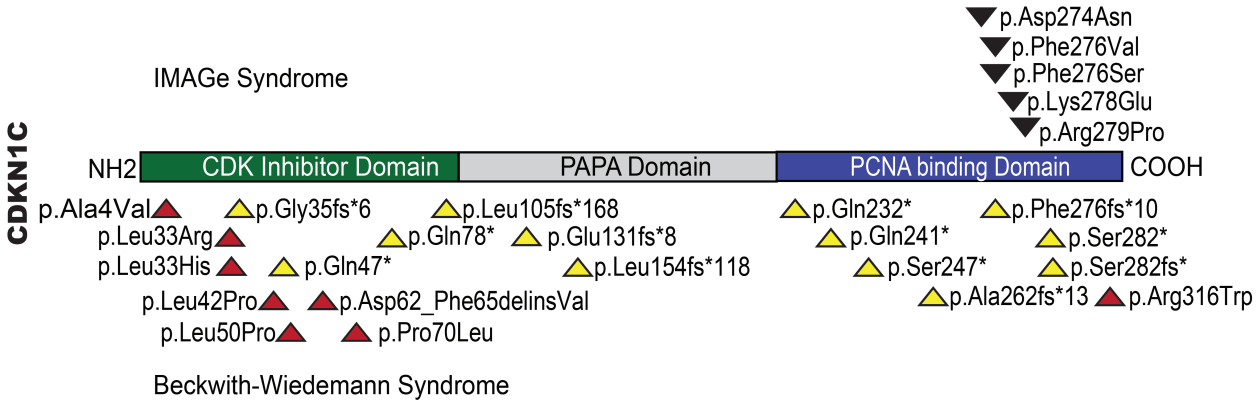

CDKN1C
IMAGe Syndrome
p.Asp274Asn
p.Phe276Val
p.Phe276Ser
p.Lys278Glu
p.Arg279Pro
NH2
CDK Inhibitor Domain
PAPA Domain
PCNA binding Domain
COOH
p.Ala4Val
p.Gly35fs*6
p.Leu105fs*168
p.Gln232*
p.Phe276fs*10
p.Leu33Arg
p.Gln78*
p.Glu131fs*8
p.Gln241*
p.Ser282*
p.Leu33His
p.Gln47*
p.Leu154fs*118
p.Ser247*
p.Ser282fs*
p.Leu42Pro
p.Asp62_Phe65delinsVal
p.Ala262fs*13
p.Arg316Trp
p.Leu50Pro
p.Pro70Leu
Beckwith-Wiedemann Syndrome

Figure 4-10. Mutations in the putative nuclear localization site do not result in disruption of active nuclear transport of CDKN1C. (a) H295R adrenocortical cells were transfected with pcDNA-CDKN1C wild-type (wt) or IMAGe mutants (F276V and K278E) which are small enough to passively diffuse in and out of the nucleus[2]. The transfected proteins show no difference in nuclear localization of CDKN1C. (b) H295R (top row) and M1 (bottom row) cells were transfected with fusion pcEGFP-CDKN1C wild type (wt) or IMAGe mutants (F276V and K278E) with a molecular weight of 61kDa, which requires active transport into the nucleus due to its larger size. In both H295R and M1 cells, nuclear transport was not disturbed by either of the tested IMAGe mutants. We quantified this effect in M1 fibroblasts and found that 80-90% of the M1 fibroblasts transfected with the GFP-CDKN1C constructs (wt, F276V and K278E) scored as nuclear only. By comparison, almost 92% of the cells expressing GFP alone(empty vector) scored as nuclear *and* cytoplasmic. (c) HEK293T cells were transfected with pFLAG-CDKN1C constructs (wt, F276V and K278E) and immunoprecipitated with Flag. Right panel: whole cell lysate of transfected CDKN1C-wt. Left panel: Immunoprecipitated CDKN1C-wt and IMAGe mutants did not differ in their ability to bind to alpha-importin, a critical mediator in active nuclear transport.

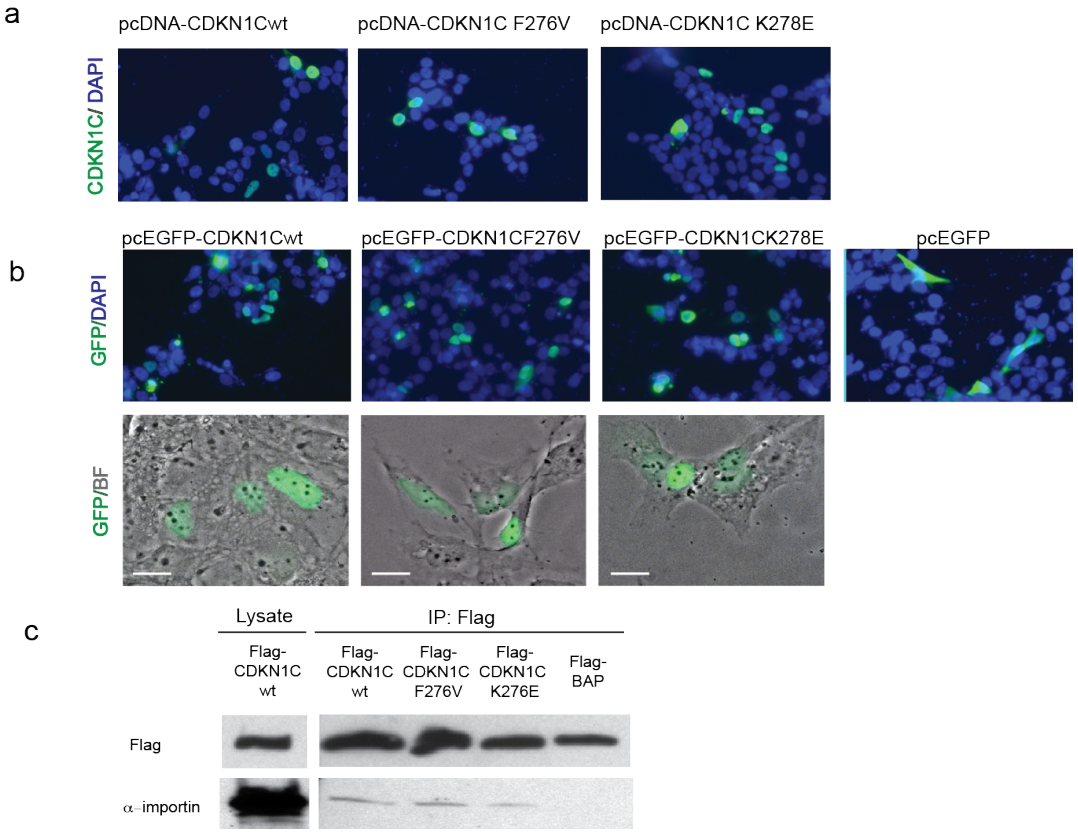

a
pcDNA-CDKN1Cwt
pcDNA-CDKN1C F276V
pcDNA-CDKN1C K278E
CDKN1C/ DAPI
b
pcEGFP-CDKN1Cwt
pcEGFP-CDKN1CF276V
pcEGFP-CDKN1CK278E
pcEGFP
GFP/DAPI
GFP/BF
c
Lysate
IP: Flag
Flag-CDKN1C wt
Flag-CDKN1C wt
Flag-CDKN1C F276V
Flag-CDKN1C K276E
Flag-BAP
Flag
α-importin

Table 4-1. Clinical characteristics of IMAGe Syndrome Patients

	Pt 1	Pt 2	Pt 3	Pt 4	IV-10	V-12	V-7	V-6	V-5	V-1	V-2
Chromosomal sex	46, XY	46, XY	46, XY	46,XY	46, XY	46, XY	46, XY	46, XY	46, XX	46, XX	46, XY
Nucleotide Change	826T>C	835G>C	819G>A	831A>G	825T>G	825T>G	825T>G	825T>G	825T>G	825T>G	825T>G
Amino Acid Change	Phe276Ser	Arg279Pro	Asp274Asn	Lys278Glu	Phe276Val	Phe276Val	Phe276Val	Phe276Val	Phe276Val	Phe276Val	Phe276Val
Intrauterine Growth Restriction	+	+	+	+	+	+	+	+	+	+	+
Adrenal crisis	+ (day 14)	+ (day 4)	+ (day 7)	+ (day 3)	+ (day 11)	+ (1^{st} week)	+ (day11)	+ (day19)	+ (day20)	+ (day5)	+ (day21)
Adrenal hypoplasia	probable	+	+	+	+	+	+	+	+	+	+
Hypercalciuria	+	+	probable	+	NE	NE	NE	NE	-	NE	NE
Bifrontal bossing, abnormal ears and nose	+	+	+	+	+	+	+	+	+	+	+
Short arms and legs	+	+	+	+	-	-	-	-	+	-	-
Craniosynostosis	+	+	+	-	-	-	-	-	-	-	-
Genital Anomalies	+	+	small penis	+	+	+	+	-	NA	NA	-
Cryptorchidism Unilateral or bilateral	+	+	+	+	+	-	-	+	NA	NA	+
Osteopenia	+	+	+	NE	NE	NE	NE	NE	NE	NE	NE
Delayed bone age	+	+	+	+	+	+	+	+	+	-	+
Small epiphyses	+	-	+	+	+	+	-	+	NE	-	+
Striated irregular metaphyses	+	+	+	+*	+	+	-	+	NE	-	+

NE: Not Evaluated; NA: Not Applicable; day refers to the postnatal day of adrenal crisis.*broad and flattened, but not striated

REFERENCES

1 Vilain, E. *et al.* IMAGe, a new clinical association of intrauterine growth retardation, metaphyseal dysplasia, adrenal hypoplasia congenita, and genital anomalies. *J Clin Endocrinol Metab* **84**, 4335-4340 (1999).

2 Bergada, I. *et al.* Familial occurrence of the IMAGe association: additional clinical variants and a proposed mode of inheritance. *J Clin Endocrinol Metab* **90**, 3186-3190, (2005).

3 Lee, M. H., Reynisdottir, I. & Massague, J. Cloning of p57KIP2, a cyclin-dependent kinase inhibitor with unique domain structure and tissue distribution. *Genes Dev* **9**, 639-649 (1995).

4 Romanelli, V. *et al.* CDKN1C (p57(Kip2)) analysis in Beckwith-Wiedemann syndrome (BWS) patients: Genotype-phenotype correlations, novel mutations, and polymorphisms. *Am J Med Genet A* **152A**, 1390-1397, (2010).

5 Hutz, J. E. *et al.* IMAGe association and congenital adrenal hypoplasia: no disease-causing mutations found in the ACD gene. *Mol Genet Metab* **88**, 66-70, (2006).

6 Ko, J. M., Lee, J. H., Kim, G. H., Kim, A. R. & Yoo, H. W. A case of a Korean newborn with IMAGe association presenting with hyperpigmented skin at birth. *Eur J Pediatr* **166**, 879-880, (2007).

7 Lienhardt, A., Mas, J. C., Kalifa, G., Chaussain, J. L. & Tauber, M. IMAGe association: additional clinical features and evidence for recessive autosomal inheritance. *Horm Res* **57 Suppl 2**, 71-78, (2002).

8 Pedreira, C. C., Savarirayan, R. & Zacharin, M. R. IMAGe syndrome: a complex disorder affecting growth, adrenal and gonadal function, and skeletal development. *J Pediatr* **144**, 274-277,(2004).

9 Tan, T. Y. *et al.* Two sisters with IMAGe syndrome: cytomegalic adrenal histopathology, support for autosomal recessive inheritance and literature review. *Am J Med Genet A* **140**, (2006).

10 Amano, N. *et al.* Radiological evolution in IMAGe association: a case report. *Am J Med Genet A* **146A**, 2130-2133, (2008).

11 Balasubramanian, M., Sprigg, A. & Johnson, D. S. IMAGe syndrome: Case report with a previously unreported feature and review of published literature. *Am J Med Genet A* **152A**, (2010).

12 Watanabe, H. *et al.* Suppression of cell transformation by the cyclin-dependent kinase inhibitor p57KIP2 requires binding to proliferating cell nuclear antigen. *Proc Natl Acad Sci U S A* **95**, 1392-1397 (1998).

13 Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* **38**, W695-699, (2010).

14 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249, (2010).

15 Diaz-Meyer, N. *et al.* Silencing of CDKN1C (p57KIP2) is associated with hypomethylation at KvDMR1 in Beckwith-Wiedemann syndrome. *Journal of medical genetics* **40**, 797-801 (2003).

16 Shin, J. Y., Fitzpatrick, G. V. & Higgins, M. J. Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. *Embo J* **27**, 168-178, (2008).

17 Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-415 (1993).

18 Wiedemann, H. R. [The EMG-syndrome: exomphalos, macroglossia, gigantism and disturbed carbohydrate metabolism]. *Z Kinderheilkd* **106**, 171-185 (1969).

19 Beckwith, J. B. Macroglossia, omphalocele, adrenal cytomegaly, gigantism, and hyperplastic visceromegaly. *Birth Defects Orig. Art. Ser.* **V(2)**, 188-196 (1969).

20 Zhang, P. *et al.* Altered cell differentiation and proliferation in mice lacking p57KIP2 indicates a role in Beckwith-Wiedemann syndrome. *Nature* **387**, 151-158, (1997).

21 Hatada, I. *et al.* An imprinted gene p57KIP2 is mutated in Beckwith-Wiedemann syndrome. *Nat Genet* **14**, 171-173, (1996).

22 Bourcigaux, N. *et al.* High expression of cyclin E and G1 CDK and loss of function of p57KIP2 are involved in proliferation of malignant sporadic adrenocortical tumors. *J Clin Endocrinol Metab* **85**, 322-330 (2000).

23 Havens, C. G. & Walter, J. C. Mechanism of CRL4(Cdt2), a PCNA-dependent E3 ubiquitin ligase. *Genes Dev* **25**, 1568-1582, (2011).

24 Kirchmaier, A. L. Ub-family modifications at the replication fork: Regulating PCNA-interacting components. *FEBS Lett* **585**, 2920-2928, (2011).

25 Ye, Y. & Rape, M. Building ubiquitin chains: E2 enzymes at work. *Nat Rev Mol Cell Biol* **10**, 755-764, (2009).

26 Mukhopadhyay, D. & Riezman, H. Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* **315**, 201-205, doi:10.1126/science.1127085 (2007).

27 Li, W. & Ye, Y. Polyubiquitin chains: functions, structures, and mechanisms. *Cell Mol Life Sci* **65**, 2397-2406, (2008).

28 Thrower, J. S., Hoffman, L., Rechsteiner, M. & Pickart, C. M. Recognition of the polyubiquitin proteolytic signal. *Embo J* **19**, 94-102, doi:10.1093/emboj/19.1.94 (2000).

29 Lee, H., Jen, J. C., Cha, Y. H., Nelson, S. F. & Baloh, R. W. Phenotypic and genetic analysis of a large family with migraine-associated vertigo. *Headache* **48**, 1460-1467, (2008).

30 Lee, H. *et al.* Improving the efficiency of genomic loci capture using oligonucleotide arrays for high throughput resequencing. *BMC Genomics* **10**, 646, (2009).

31 Homer, N., Merriman, B. & Nelson, S. F. BFAST: an alignment tool for large scale genome resequencing. *PLoS One* **4**, e7767, (2009).

32 Homer, N., Merriman, B. & Nelson, S. F. Local alignment of two-base encoded DNA sequence. *BMC Bioinformatics* **10**, 175, (2009).

33 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, (2009).

34 Clark, M. J. *et al.* U87MG Decoded: The Genomic Sequence of a Cytogenetically Aberrant Human Cancer Cell Line. *PloS Genetics* **6** (2009).

35 O'Connor, B. D., Merriman, B. & Nelson, S. F. SeqWare Query Engine: storing and searching sequence data in the cloud. *BMC Bioinformatics* **11 Suppl 12**, S2,

36 Matsuoka, S. *et al.* p57KIP2, a structurally distinct member of the p21CIP1 Cdk inhibitor family, is a candidate tumor suppressor gene. *Genes Dev* **9**, 650-662 (1995).

37 Mattera, R., Tsai, Y. C., Weissman, A. M. & Bonifacino, J. S. The Rab5 guanine nucleotide exchange factor Rabex-5 binds ubiquitin (Ub) and functions as a Ub ligase through an atypical Ub-interacting motif and a zinc finger domain. *The Journal of biological chemistry* **281**, (2006).

Chapter 5:

Conclusion

The Evolution of the Search for Novel Genes in Mammalian Sex Determination: From Mice to Men

Identifying the genetic origins of testis and ovarian determination is no easy task. Traditional linkage approaches for mapping disease genes required large families with multiple affected members. However, patients with disorders of sex development are sub- or infertile and therefore most families are too small to perform genetic linkage analysis. Some of the major breakthroughs relied on identification of karyotype abnormalities followed by positional cloning to identify the disrupted gene. Using this approach, SRY was identified as the gene responsible for initiating male sex determination in humans.

Mouse Models of Mammalian Sex Determination

Beyond the initial discovery of the testis- determination gene, SRY, in humans, the discovery of other novel sex determination genes was still limited by the size of the families, the rareness of these conditions, and access to fetal gonad tissue. To elucidate the complex interactions that result in the formation of a testis or an ovary, much of the work in identifying novel genes and understanding their interactions was done in mouse inbred mouse strains. In the developing mouse there is easy access to the tissue of interest, in this case, the bipotential gonad, at the timepoint of divergence into either a testis or ovary. Studies performed in mouse fetuses would be not only impractical but unethical in human fetuses. Furthermore, we have the technology to genetically manipulate mice, creating transgenic mice to directly assess the effect of overexpression or knockout of specific genes. All of these things have catapulted mouse inbred strains as one of the primary model organisms for studying mammalian sex determination.

The modern laboratory mouse has its origins in the 20th century around the time when the laws of Mendelian genetics were “re-discovered” and integrated into classical genetics. Fancy mouse breeders realized that valued traits, such as coat color and fur length, were also likely to

be genetically determined. Ms. Abbie Lathrop developed the first fancy mouse strains for laboratory use on her farm in Granby Massachusetts. In order to ensure the genetic homogeneity of the mouse model and solidify a place for mouse models as paramount in the study of human health and disease, Clarence Little recognized the importance of generating homogenous inbred mouse strains to remove the genetic heterogeneity from experiments therefore allowing experiments conducted in different locations to be more easily compared. After moving to Cold Spring Harbor in 1918, Little began working on other strains of inbred mice, including C57BL/6, C3H, and BALB/c [1]. Little ultimately went on to co-found the Jackson Laboratory in Bar Harbor, Maine, which today remains a non-profit institution dedicated to using mouse models to better understand and treat a wide variety of human disease.

Mouse models are ideal for studying human diseases. First, mice have the same internal organs as humans, they can be housed in small cages, easily fed, and with age and diet mice develop many of the same diseases as humans, including cancer and diabetes. Despite the evolutionary distance between mice and humans, the mechanisms of sex determination and sex development are highly conserved. The trigger for placental mammals' testis-determining pathways is the sexually dimorphic expression of a single gene, SRY, on the Y chromosome. SRY is believed to have evolved 150 million years ago, after the divergence of placental and marsupial mammals from monotremes[2]. In lower organisms, the sex determination trigger can be either genetic, environmental, or a combination of both. While trigger mechanisms to differentiate between two sexes may be different, many of the downstream actors remain conserved.

Prior to the discovery of SRY, careful investigation of sex development in mouse models also identified specific strains as susceptible to various types of perturbations in XY sex reversal.

A leader in the field of mouse genetics and sex development, Eva Eicher, pioneered many of the crosses and genetically mutated strains in which there is abnormal sex development[3]. Consomic crosses, in which a single chromosome is derived from one strain, while the remaining chromosomes come from a different strain, further illuminated the role of the Y-chromosome in genetic sex determination. Using this approach, the Y-chromosome from wild derived mouse strains was backcrossed onto inbred strains to determine if there was a sex development phenotype. The Y-chromosome of POSA, a semi-inbred strain generated from the mating of a wild-derived *mus musculus domesticus poschiavinaus* male and a Naval Medical Research Institute (NMRI) Swiss female, when backcrossed onto the C57BL/6 (B6) background resulted in a large proportion of XY-female mice. When mice were examined during embryonic development at E14.5, none of the B6-Y^{POS} embryos developed fully normal testis. All of the XX individuals had normal appearing ovaries, indicating that this trait resulted from an interaction between the Y^{POS} and one or more autosomal determinants on the B6 background[4]. This phenomenon is not particular to the Y^{POS} chromosome. On its' original genetic background, Y^{POS} initiates normal male sex determination. Backcrossing the Y^{POS} onto other inbred backgrounds, such as BALB/c, C58/J, and DBA/2J did not result in any XY sex reversal [5] implying that that the aberrant interaction was specific to the B6 inbred strain and the Y^{POS}. Placing other wild derived Y-chromosomes on the B6 background confirmed the propensity of the B6 strain to XY sex reversal. Y-chromosomes from *mus musculus domesticus AKR (Y^{AKR}), and mus musculus domesticus tirano (Y^{TIR})* showed varying degrees of sex reversal when placed onto a B6 background. Interestingly, Y^{AKR} merely showed delayed testis determination but by adulthood all XY mice were phenotypically male and fertile. These studies confirmed that the B6 genetic background is exquisitely sensitive to XY sex reversal.

Studies utilizing recombinant inbred strains focused on exploiting the genetic differences between the sex reversing B6 strain and the non-sex reversing DBA2 strain. B6-Y^{POS} males were crossed to DBA females and F1 offspring were intercrossed to create a genetically heterogeneous population with varying phenotypes. The genotype-phenotype correlation can be used to map genetic regions of B6 origin that contribute to XY hermaphroditism. However, these studies did not have the power or the resolution to definitively identify the genes or mutations responsible for B6 sensitivity to sex reversal[6].

Complementing the increased sensitivity to XY sex reversal, the B6 strain is also protected from XX-sex reversal. In genetic models of XX-males, when mutations are transferred to the B6 inbred strain, the XX-Sex reversal phenotype is lost. In both the *Odsex* mutant and the Tg-SOX3 mutant, the resulting XX-Male phenotype is lost when the genetic mutation is transferred to a B6 background. Together, the sensitivity of B6 background to XY sex reversal and protection from XX male sex reversal implies that at the genomic and transcriptomic level, the balance of the bipotential gonad is tipped in favor of ovarian formation.

The B6 strain's protection from XX-sex reversal and promotion of XY-sex reversal can be partially explained by the differing expression levels of "female" or "male" promoting genes within the developing gonad of B6 mice. B6 animals have a higher expression of "female-promoting" genes, compared to 129 animals. Therefore, B6 animals are already tipped in in favor of ovarian sex determination. To compensate for a female-tipped balance, there is an increased level of *SOX9* expression that is sufficient to for male sex determination[7].

The use of mouse models of sex development has elucidated the function of genes of large effect size (i.e. *SRY*, *SOX9*), but also has shed light on the intricate balance of genes that take place within the differentiating gonad. The developmental choice between an ovary or a

testis is not a passive process, but one that is actively battling to repress the other throughout gonadal development, and even long into adulthood.

The Role of the Human Genome Project on Disorders of Sex Development

The early history of novel gene identification in sex development relied heavily on karyotype and *in situ* hybridization to identify regions of interest within the genome from which a gene could be positionally cloned. Through positional cloning of Yp translocations to the X-chromosome, the testis-determining gene on the Y-chromosome was identified as *SRY*, a homeobox gene [8]. This initial discovery paved the way for the discovery of a second gene in the SOX gene family, *SOX9*, in patients with Campomelic Dysplasia (CD) and XY autosomal sex reversal. Translocation breakpoints identified through karyotyping were followed by *in situ* hybridization to identify the gene, which was followed up by sequencing for deleterious mutations in *SOX9*[9, 10].

An alternative source of genome variation lies in duplications and deletions of larger chromosomal regions known as copy number variation (CNV) and can give rise to a number of polymorphic traits. Karyotype and chromosome painting approaches can detect large duplications or deletions, on the order of megabases, while SNP arrays can detect copy number variations (CNV) that are as small as several kilobases. Most CNVs are benign, representing the range of genomic variation that exists within the population. However, *de novo* copy number variations are responsible for many developmental conditions. CNVs in patients with disorders of sex development (DSD) can be inherited from either the mother or the father, since the phenotype is sex specific [11, 12]. Large scale CNV analysis in a wide range of DSD patients has identified multiple candidate genes for XY sex reversal and hypospadias[13, 14]. Even more

interesting are recent reports that highlight the importance of regulatory regions in the proper expression of genes within the developing gonad. Deletions and duplications in the regulatory region of *SOX3*, *SOX9*, *DAX1*, and *GATA4* can result in misregulation (either mislocalized, over- or underexpressed) of a gene within the developing gonad resulting in a sex reversal phenotype[13-16].

The road to sequencing the entire human genome began long before the idea of sequencing all three billion base pairs even seemed to be feasible. The Human Genome project had its roots in an ambitious project to generate a complete linkage map of the Human Genome based on restriction fragment length polymorphisms (RFLP) in the early 1980s. The genetic map of the human genome paved the way for linkage analysis in large families to follow specific disease phenotypes looking for a gene with a large effect size. The first linkage study was limited to the X chromosome and identified Factor IX for the X-linked trait, hemophilia B [17] .

Identifying disease linkage peaks and/or genes in the 1980's was no small task, in which mapping a single gene comprised a graduate student's thesis. Even in the mid-1980s, the sequencing technology had not advanced to the point where sequencing model organisms was a viable option, as sequencing reactions were run on large pulse field gels and manually read. This technique was both labor-intensive and prone to human error. Upon the development of fluorescent dideoxysequencing and miniaturization of the scale of sequencing reads, on capillary electrophoresis with automated the nucleotide calling. The Human Genome Project (HGP) was officially born, with the goal to generate a dense physical map of the human genome and to sequence 20Mb of model organisms in 15 years, a goal that was ultimately dwarfed by the actual accomplishments of the HGP.

The availability of a dense physical map of the human genome allowed for linkage analysis and positional cloning efforts to identify gene in which there were large families or multiple families with a well-phenotyped disorder of sex development. In 2001, *FOXL2* was identified as the causative gene in blepharophimosis, ptosis, epicanthus inversus (BPES) syndrome after linkage analysis and positional cloning [18, 19]. The conservation of sex development pathways is highlighted by the fact that this same gene is a candidate for a nearly identical condition in XX male goats known as *polled/intersex* condition. This same approach has been used to identify novel genes in sex development, including Rspondin-1 (*RSPO1*) in XX-Males with palmoplantar keratoderma with squamous cell carcinoma of skin and sex reversal [20] and in a family with 46, XY gonadal dysgenesis revealed *MAP3K1*[21, 22] as an important signaling pathway in early sex determination. Interestingly, this same map kinase pathway was discovered using an analogous method in the mouse models, in which ENU mutagenesis generated a spontaneous mutant, which was mapped back to chromosome 17. Positional cloning identified *Map3k4* as the mutated gene and illuminated the genetic cause of an unexplained mouse model of sex determination, T-associated sex reversal (Tas) originally described in 1983 by Eva Eicher[23-25].

The sequencing of the human genome was completed in 2001, 10 years after the official start of the HGP, by two independent groups, the public international venture led by Eric Lander [26]and a private venture, led by Craig Venter[27]. These simultaneously published drafts of the human genome shed light on the true amount of genetic variation and paved the way for the advent of next generation sequencing technologies that are, as of 2011, aiming to sequence an entire genome for $1000. The ability to sequence an entire genome for $1000 is the theoretical price by which it would be feasible for use in clinical diagnostics and in large-scale research

projects for novel disease gene identification. The human genome project has revolutionized our knowledge about the variation within the human genome. The immense amount of data provided by genome sequencing has shed light on the huge number of common variants' but also on the large number of rare, personal, and deleterious variants present in each of us[28].

The traditional linkage analysis approach and disease gene identification has become even more powerful as next generation sequencing can sequence an entire linkage peak that contains hundreds of genes. Instead of using *a priori* evidence to pick a gene out of hundred to sequence, one can take an unbiased approached to identify the disease causing gene[29]. These unbiased approaches rely heavily on bioinformatic expertise to reliably identify causative gene variants and eliminate false positive signals. However, they will identify novel pathways and genes not previously described in the sex development literature. These novel discoveries will ultimately lead to a systems biology approach to understanding the genetic regulation behind the developmental choice of the bipotential gonad to become either a testis or an ovary.

The Future of Sex Determination Research

Strategies to identify sex development genes have evolved along side genomic technologies. Next generation sequencing is just the beginning of a new chapter in personalized medicine and sex development research. Today, sequencing the entire exome, and even the whole genome, in unexplained cases of DSD will continue to enlighten and broaden our understanding about both normal testis and ovarian development and patients who have disorders of sex development. However, the future of sex determination research will not end in the identification of disease genes through sequencing of genomic DNA. Based on the hundreds of variants identified of "undetermined significance", identification of the disease-causing variant is

difficult and reliant on functional data generated in *in vitro* or *in vivo* model systems. This technology will not only revolutionize how we identify novel genes, but also how we diagnose and treat patients in the clinical setting. From the research perspective, these tools provide an unprecedented view into the various layers of gene regulation in a highly specific and quantitative way that ultimately may provide better long-term management and treatments for patients with DSD.

Currently, targeted capture approaches can sequence all known genes for a specific condition, such as breast cancer [30] and hereditary hearing loss [31]. Providing a genetic diagnosis that explains the molecular defect in every patient presenting with DSD would be an invaluable tool to patients, clinicians, and to the research community. A personalized diagnosis would allow for clinicians and patients to create standardized guidelines and criteria for the complex medical, surgical, and psychological management of these patients.

As the costs of whole genome sequencing continue to drop, the cost-savings from targeting specific exons for sequencing will ultimately disappear. In such a scenario, whole genome or exome data could be targeted at the bioinformatics level, and one would analyzing only a subset of genes relevant to the disease phenotype. The major benefit to only sequencing or analyzing a subset of the genomic material is that we limit our findings to genetic variants that are known to cause disease and can be functionally validated. This targeted approach removes the ethical quandary in which whole genome or exome sequencing may identify a genetic finding completely unrelated to the reason the patient came in for testing (i.e. identification of a *BRCA1* mutation in a newborn presenting with a 46, XX DSD). Furthermore, current ethical guidelines do not support genetic testing of children for adult onset diseases, yet there exist no guidelines on inadvertent genetic diagnoses that might arise in reporting results from the whole exome analysis.

To date, there is no research on whether releasing large number of semi-predictive genetic data is harmful to the patient, and therefore we caution overzealous use of whole genome or whole-exome analysis on a clinical and predictive basis. In absence of studies that examine the best methods to provide this information and the effects of receiving large-scale genetic data, we recommend a more conservative approach, to limit the genes analyzed to those specific for the disease process in question. Despite these dilemmas, there is no question that next generation sequencing will revolutionize the diagnostic approach for patients presenting with genetic disorders. For patients with DSD, this molecular phenotyping will ultimately give rise to a greater understanding about the clinical outcomes in this groups of patients and inform our understanding of the genetics of gonadal determination.

The sequencing technologies reach far beyond variant discovery and diagnostics, as these technologies can query the entire transcriptome, epigenetic markers, chromosomal conformation, to name a few. Exploring the epigenome, will unravel how methylation changes throughout development can affect differentiation and function of cells and tissues. Combining genomic and transcriptomic studies can identify genes that are subject to allele-specific cis- or trans- effects[32]. In the aftermath of the Human Genome Project, we are still discovering new exons, some of which are tissue and time specific, and introduce new domains into known genes[33]. Understanding the genetic and transcriptomic regulation within the developing gonad is a lofty goal, requiring well-conceived experiments that ultimately require intensive bioinformatic analysis to identify the relationships between genome, transcriptome, epigenome, or as some would say: the “interactome”. These are only a few of the ways that sequencing technology has given us tools that early geneticists could only dream of.

While most novel sex determination genes are identified in human patients with DSD, much of our understanding of how sex determination genes interact within the developing gonad comes from well- designed studies in mouse models. Our understanding of genes and pathways important in gonadal determination will be elucidated from overexpression and knockout studies in inbred mice and other model organisms. The mouse model has the added benefit that it is a controlled environment in which each cell type within the heterogeneous gonad can be dissected and analyzed for their specific gene expression profile.

Several genes that when mutated can result in 46, XY GD have roles in sex determination that remain to be explained. One of these is Chromobox-2 (*CBX2*), which was first described in a knockout mouse model M33 and more recently found in a patient with 46, XY GD[34]. Interestingly, the chromobox gene family is a known regulator of gene expression, through chromatin and histone modification, and likely plays a role in programming and regulating the environmental niche that allows for proper sex development. The interaction of *CBX2* and other epigenetic regulators are likely to play a large role in controlling timing and level of expression of sex determination genes and repressing down the expression of the genes for the opposing sex. Ultimately, understanding the genome, transcriptome, and epigenome within the developing gonad will allow us to paint a comprehensive portrait of the interactions within the developing gonad. As we can better understand the complexity that has been built into the regulatory regions of the genome, we might be able to decipher how genetic elements of small effect sizes can contribute to fertility and reproductive health.

The synergistic use of mouse and human models has contributed greatly to our understanding of early sex determination. As we move forward, there are still remarkable

discoveries to be made within the field of sex development, which will help us unravel the age-old question: why are men and women different?

References

1. Silver, L.M., *Mouse Genetics: Concepts and Applications*1995, Oxford, UK: Oxford University Press.
2. Wallis, M.C., P.D. Waters, and J.A. Graves, *Sex determination in mammals--before and after the evolution of SRY.* Cellular and molecular life sciences : CMLS, 2008. **65**(20): p. 3182-95.
3. Eicher, E.M. and L.L. Washburn, *Genetic control of primary sex determination in mice.* Annual review of genetics, 1986. **20**: p. 327-60.
4. Eicher, E.M., et al., *Mus poschiavinus Y chromosome in the C57BL/6J murine genome causes sex reversal.* Science, 1982. **217**(4559): p. 535-7.
5. Eicher, E.M., *Autosomal genes involved in mammalian primary sex determination.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 1988. **322**(1208): p. 109-18.
6. Eicher, E.M., et al., *Sex-determining genes on mouse autosomes identified by linkage analysis of C57BL/6J-YPOS sex reversal.* Nat Genet, 1996. **14**(2): p. 206-9.
7. Munger S, A., D, Haider A, Threadgill D, Capel B, *Integrating Genetics and genomics to elucidate the transcription networks governing sex determination and testis organogenesis in mice.*, in *Fifth International Symposium on the Biology of Vertebrate Sex Determination*2009: Kona, Hawaii.
8. Sinclair, A.H., et al., *A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif.* Nature, 1990. **346**(6281): p. 240-4.
9. Foster, J.W., et al., *Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene.* Nature, 1994. **372**(6506): p. 525-30.

10. Wagner, T., et al., *Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9.* Cell, 1994. **79**(6): p. 1111-20.

11. Cox, J.J., et al., *A SOX9 duplication and familial 46,XX developmental testicular disorder.* The New England journal of medicine, 2011. **364**(1): p. 91-3.

12. Barbaro, M., et al., *Isolated 46,XY gonadal dysgenesis in two sisters caused by a Xp21.2 interstitial duplication containing the DAX1 gene.* The Journal of clinical endocrinology and metabolism, 2007. **92**(8): p. 3305-13.

13. Tannour-Louet, M., et al., *Identification of de novo copy number variants associated with human disorders of sexual development.* PloS one, 2010. **5**(10): p. e15392.

14. White, S., et al., *Copy Number Variation in Patients with Disorders of Sex Development Due to 46,XY Gonadal Dysgenesis.* PloS one, 2011. **6**(3): p. e17793.

15. Sutton, E., et al., *Identification of SOX3 as an XX male sex reversal gene in mice and humans.* The Journal of clinical investigation, 2011. **121**(1): p. 328-41.

16. Smyk, M., et al., *Male-to-female sex reversal associated with an approximately 250 kb deletion upstream of NR0B1 (DAX1).* Human genetics, 2007. **122**(1): p. 63-70.

17. Camerino, G., et al., *Regional localization on the human X chromosome and polymorphism of the coagulation factor IX gene (hemophilia B locus).* Proceedings of the National Academy of Sciences of the United States of America, 1984. **81**(2): p. 498-502.

18. Jost, A., et al., *Studies on sex differentiation in mammals.* Recent Prog Horm Res, 1973. **29**: p. 1-41.

19. Crisponi, L., et al., *The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome.* Nat Genet, 2001. **27**(2): p. 159-66.

20. Tomaselli, S., et al., *Syndromic true hermaphroditism due to an R-spondin1 (RSPO1) homozygous mutation.* Hum Mutat, 2008. **29**(2): p. 220-6.

21. Lois, C., et al., *Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors.* Science, 2002. **295**(5556): p. 868-72.

22. Pearlman, A., et al., *Mutations in MAP3K1 cause 46,XY disorders of sex development and implicate a common signal transduction pathway in human testis determination.* American journal of human genetics, 2010. **87**(6): p. 898-904.

23. Bergstrom, D.E., et al., *Related function of mouse SOX3, SOX9, and SRY HMG domains assayed by male sex determination.* Genesis, 2000. **28**(3-4): p. 111-24.

24. Washburn, L.L. and E.M. Eicher, *Sex reversal in XY mice caused by dominant mutation on chromosome 17.* Nature, 1983. **303**(5915): p. 338-40.

25. Bogani, D., et al., *Loss of mitogen-activated protein kinase kinase kinase 4 (MAP3K4) reveals a requirement for MAPK signalling in mouse sex determination.* PLoS biology, 2009. **7**(9): p. e1000196.

26. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

27. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.

28. Kinde, I., et al., *Detection and quantification of rare mutations with massively parallel sequencing.* Proceedings of the National Academy of Sciences of the United States of America, 2011.

29. Goudie, D.R., et al., *Multiple self-healing squamous epithelioma is caused by a disease-specific spectrum of mutations in TGFBR1.* Nature genetics, 2011. **43**(4): p. 365-9.

30. Walsh, T., et al., *Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing.* Proc Natl Acad Sci U S A, 2010. **107**(28): p. 12629-33.

31. Shearer, A.E., et al., *Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(49): p. 21104-9.

32. Pastinen, T., *Genome-wide allele-specific analysis: insights into regulatory variation.* Nature reviews. Genetics, 2010. **11**(8): p. 533-8.

33. Graveley, B.R., et al., *The developmental transcriptome of Drosophila melanogaster.* Nature, 2011. **471**(7339): p. 473-9.

34. Biason-Lauber, A., et al., *Ovaries and female phenotype in a girl with 46,XY karyotype and mutations in the CBX2 gene.* American journal of human genetics, 2009. **84**(5): p. 658-63.