Title
Identification of Genetic Etiology in Disorders of Sex Development

Author
Barseghyan, Hayk

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Identification of Genetic Etiology in Disorders of Sex Development

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Human Genetics

by

Hayk Barseghyan

2017

ABSTRACT OF THE DISSERTATION

Identification of Genetic Etiology in Disorders of Sex Development

by

Hayk Barseghyan

Doctor of Philosophy in Human Genetics

University of California, Los Angeles, 2017

Professor Eric J.N. Vilain, Chair

Disorders of Sex Development (DSD) are defined as “congenital conditions in which development of chromosomal, gonadal, or anatomic sex is atypical.” These conditions have an approximate frequency of 0.5-1% of live births and encompass a wide variety of urogenital abnormalities ranging from mild hypospadias to sex reversal. Lack of standardized anatomical/endocrine phenotyping and the limited number of known DSD genes with poor genotype/phenotype correlation have hampered the field of clinical management, leaving many patients without a definitive genetic diagnosis. Thus, the focus of this dissertation is to identify the underlying pathogenic genetic mutations that disrupt development of urogenital structures, leading to Disorders of Sex Development in humans.

The traditional trend of diagnostic approach for patients with DSD is to select candidate gene testing by searching for additional phenotypic and metabolic information

through imaging studies and endocrine tests that could explain the patient's phenotype. This approach is usually ineffective, costly and time-consuming. To address this issue and identify genetic variants leading to DSD, we utilized exome sequencing, in patients diagnosed with abnormal sex development. We show that exome sequencing has transformed the field of clinical genetic diagnosis by increasing the rate of diagnosis by approximately 30% and has become a method of choice for many clinicians.

Although exome sequencing provides much higher diagnostic yields for DSD patients than the conventional techniques, more than half of the patients tested with ES still do not possess a specific genetic diagnosis. Rather, in these patients ES identifies hundreds of variants of unknown clinical significance (VUS). To investigate the role of these variants in the 46,XY subset of DSD patients, we performed gonadal gene expression studies in C57BL/6J-Y^{POS} mice modeling the phenotype of human 46,XY individuals to identify genes important in sex development. We used these genes to filter VUS identified in 46,XY DSD exome negative cases. We identified 15 novel candidate genes with mutations in 46,XY DSD patents that may be associated with disease pathogenesis.

Due to innate limitations of exonic short read sequencing, many native variants are not identified by exome sequencing. To this end, we utilize genome sequencing in conjunction with a novel genome mapping technology in order to uncover the full spectrum of variations present in a human genomes. Collectively, these two technologies provide a physical map and a base pair-level DNA resolution allowing for identification of novel pathogenic variants that were previously inaccessible.

The dissertation of Hayk Barseghyan is approved.

Katrina Dipple

Barnett Schlinger

Stephen D. Cederbaum

Eric J. N. Vilain, Committee Chair

University of California, Los Angeles

2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Eric Vilain, laboratory principal investigator and chair of my committee, for academic, intellectual and mentoring support that he has provided throughout my PhD career. I would also like to thank the respected members of my committee: Dr. Barney Schlinger, Dr. Katrina Dipple, and Dr. Stephen Cederbaum for the provided guidance and support for my academic career. Lastly, I would like to thank my collaborators Dr. Emmanuele Delot, Dr. Alice Fleming, Dr. Ruth Baxter, Dr. Valerie Arboleda, Dr. Francisco Sanchez, Dr. Hane Lee, Dr. Negar Gharamani, Dr. Tuck Ngun, Dr. Matthew Bramble, Dr. Ascia Eskin, Dr. Stanley Nelson Dr. Vincent Harley who have provided me with outstanding training, mentorship and friendship needed for success in graduate program.

In addition, I would like to thank various funding sources that have allowed me to endure through PhD with the much needed financial and intellectual support. Laboratory of Neuroendocrinology NIH training grant entitled “Neuroendocrinology, Sex Differences, and Reproduction” (5T32HD007228); The DSD-TRN (“Disorders of Sex Development: Translational Research Network”) Eunice Kennedy Shriver NICHD research grant (HD068138); Privately Endowed Fellowships: Dr. Ursula Mandel; Mangasar M. Mangasarian; Armenian General Benevolent Union Graduate Fellowship.

Chapter 1 was published as a review and book chapter with the following citations:

Barseghyan H, Vilain E. Chapter 7: The Genetics of Ovotesticular Disorders of Sex Development (p261-263). *Genetic Steroid Disorders*. Maria I. New, Oksana Lekarev, Alan Parsa, Bert O'Malley, Gary D Hammer (2013)

Barseghyan, H., E. Delot, and E. *Vilain, New genomic technologies: an aid for diagnosis of disorders of sex development.* Horm Metab Res 2015; 47(5): 312-320.

Chapter 2 contains some data from the following publications:

Baxter, R.M., V.A. Arboleda, H. Lee, H. Barseghyan, M.P. Adam, P.Y. Fechner, R. Bargman, C. Keegan, S. Travers, S. Schelley, L. Hudgins, R.P. Mathew, H.J. Stalker, R. Zori, O.K. Gordon, L. Ramos-Platt, A. Pawlikowska-Haddal, A. Eskin, S.F. Nelson, E. Delot, and E. Vilain, *Exome sequencing for the diagnosis of 46,XY disorders of sex development*. J Clin Endocrinol Metab 2015; 100(2): E333-344.

Bashamboo, A., P.A. Donohoue, E. Vilain, S. Rojo, P. Calvel, S.N. Seneviratne, F. Buonocore, H. Barseghyan, N. Bingham, J.A. Rosenfeld, S.N. Mulukutla, M. Jain, L. Burrage, S. Dhar, A. Balasubramanyam, B. Lee, U.D.N. Members of, M.C. Dumargne, C. Eozenou, J.P. Suntharalingham, K. de Silva, L. Lin, J. Bignon-Topalovic, F. Poulat, C.F. Lagos, K. McElreavey, and J.C. Achermann, *A recurrent p.Arg92Trp variant in steroidogenic factor-1 (NR5A1) can act as a molecular switch in human sex development*. Hum Mol Genet 2016; 25(16): 3446-3453

Granados A., V.I. Alaniz, L. Mohnach, H. Barseghyan, E. Vilain, H. Ostrer, E.H. Quint, M. Chen, and C.E. Keegan. *MAP3K1-related Gonadal Dysgenesis: Six new cases and review of the literature.* (Manuscript submitted, 2017)

Chapter 3 includes data submitted for publication:

Barseghyan H, Symon, A, Zadikyan M, Almalvez M, Segura E, Eskin A, Bramble MS, Arboleda VA, Baxter R, Nelson SF, Délot E, Harley V and Vilain E. *Identification of Novel Candidate Genes for 46,XY Disorders of Sex Development (DSD) Using C57BL6J-YPOS Mouse Model*. Genome Biology (Manuscript submitted, 2017)

Chapter 4 partially includes data submitted for publication:

Barseghyan H, W. Tang, M. Almalvez, R. Wang, Douine E, L. Hane, M.S. Bramble, E. Délot, S.F. Nelson and E. Vilain. *Next Generation Mapping (NGM): A Novel Approach for Clinical Genetic Diagnosis in Duchenne Muscular Dystrophy*. Genome Medicine (Manuscript submitted, 2017)

VITA

2010	A.S., Science Chemistry Los Angeles City College (Honors)
2012	B.S., Physiological Science B.A., Russian Language and Literature University of California Los Angeles (Honors)
2014-16	Laboratory of Neuroendocrinology Trainee National Institutes of Health, T32 University of California Los Angels

PUBLICATIONS

1) Chen, X., S.M. Williams-Burris, R. McClusky, T.C. Ngun, N. Ghahramani, **H. Barseghyan**, K. Reue, E. Vilain, and A.P. Arnold, The Sex Chromosome Trisomy mouse model of XXY and XYY: metabolism and motor performance. Biol Sex Differ 2013; 4(1): 15.

2) **Barseghyan H**, Vilain E. *Chapter 7: The Genetics of Ovotesticular Disorders of Sex Development* (p261-263). Genetic Steroid Disorders. Maria I. New, Oksana Lekarev, Alan Parsa, Bert O'Malley, Gary D Hammer (2013)

3) Arboleda, V.A., A. Fleming, **H. Barseghyan**, E. Delot, J.S. Sinsheimer, and E. Vilain, Regulation of sex determination in mice by a non-coding genomic region. Genetics 2014; 197(3): 885-897.

4) Ngun, T.C., N.M. Ghahramani, M.M. Creek, S.M. Williams-Burris, **H. Barseghyan**, Y. Itoh, F.J. Sanchez, R. McClusky, J.S. Sinsheimer, A.P. Arnold, and E. Vilain, Feminized behavior and brain gene expression in a novel mouse model of Klinefelter Syndrome. Arch Sex Behav 2014; 43(6): 1043-1057

5) Baxter, R.M., V.A. Arboleda, H. Lee, **H. Barseghyan**, M.P. Adam, P.Y. Fechner, R. Bargman, C. Keegan, S. Travers, S. Schelley, L. Hudgins, R.P. Mathew, H.J. Stalker, R. Zori, O.K. Gordon, L. Ramos-Platt, A. Pawlikowska-Haddal, A. Eskin, S.F. Nelson, E. Delot, and E. Vilain, Exome sequencing for the diagnosis of 46,XY disorders of sex development. J Clin Endocrinol Metab 2015; 100(2): E333-344.

6) **Barseghyan, H.**, E. Delot, and E. Vilain, New genomic technologies: an aid for diagnosis of disorders of sex development. Horm Metab Res 2015; 47(5): 312-320.

7) Bashamboo, A., P.A. Donohoue, E. Vilain, S. Rojo, P. Calvel, S.N. Seneviratne, F. Buonocore, **H. Barseghyan**, N. Bingham, J.A. Rosenfeld, S.N. Mulukutla, M. Jain, L. Burrage, S. Dhar, A. Balasubramanyam, B. Lee, U.D.N. Members of, M.C. Dumargne, C. Eozenou, J.P. Suntharalingham, K. de Silva, L. Lin, J. Bignon-Topalovic, F. Poulat, C.F.

Lagos, K. McElreavey, and J.C. Achermann, A recurrent p.Arg92Trp variant in steroidogenic factor-1 (NR5A1) can act as a molecular switch in human sex development. Hum Mol Genet 2016; 25(16): 3446-3453.

8) Bramble, M.S., E.H. Goldstein, A. Lipson, T. Ngun, A. Eskin, J.E. Gosschalk, L. Roach, N. Vashist, **H. Barseghyan**, E. Lee, V.A. Arboleda, D. Vaiman, Z. Yuksel, M. Fellous, and E. Vilain, A novel follicle-stimulating hormone receptor mutation causing primary ovarian failure: a fertility application of whole exome sequencing. Hum Reprod 2016; 31(4): 905-914.

9) Bramble, M.S., L. Roach, A. Lipson, N. Vashist, A. Eskin, T. Ngun, J.E. Gosschalk, S. Klein, **H. Barseghyan**, V.A. Arboleda, and E. Vilain, Sex-Specific Effects of Testosterone on the Sexually Dimorphic Transcriptome and Epigenome of Embryonic Neural Stem/Progenitor Cells. Sci Rep 2016; 6: 36916.

10) **Barseghyan H**, Falk R, Vilain E. *Chapter 3: Genetics of Endocrinology* (p31-43). Manual of Endocrinology and Metabolism. (2017)

11) **Barseghyan H**, Symon, A, Zadikyan M, Almalvez M, Segura E, Eskin A, Bramble MS, Arboleda VA, Baxter R, Nelson SF, Délot E, Harley V and Vilain E. *Identification of Novel Candidate Genes for 46,XY Disorders of Sex Development (DSD) Using C57BL6J-Y^{POS} Mouse Model.* (Manuscript submitted, 2017)

12) Granados A., V.I. Alaniz, L. Mohnach, **H. Barseghyan**, E. Vilain, H. Ostrer, E.H. Quint, M. Chen, and C.E. Keegan. *MAP3K1-related Gonadal Dysgenesis: Six new cases and review of the literature.* (Manuscript submitted, 2017)

13) **Barseghyan H**, W. Tang, M. Almalvez, R. Wang, Douine E, L. Hane, M.S. Bramble, E. Délot, S.F. Nelson and E. Vilain. *Next Generation Mapping (NGM): A Novel Approach for Clinical Genetic Diagnosis.* (Submission ready for submission, 2017)

PRESENTATIONS

1) **Barseghyan H.** (2016). *Identification of Large Pathogenic Structural Variants Using Next Generation Genome Mapping Technology.* UCLA Joint Academic Retreat of Department of Human Genetics and the Intercampus Medical Genetics Training, Los Angeles, USA

2) **Barseghyan H.** (2016). *Identification of Genetic Etiology in Disorders of Sex Development.* Laboratory of Neuroendocrinology, University of California, Los Angeles, USA

3) **Barseghyan H.** (2017). *NGM in Duchene Muscular Dystrophy and Undiagnosed Genetic Disorders.* Bionano User Group Meeting, San Diego, USA

Chapter 1

Introduction

Abstract

The Chicago Consensus Conference of 2005 defined Disorders of Sex Development (DSD) as “congenital conditions in which the development of chromosomal, gonadal or anatomic sex is atypical.” DSD diagnoses are difficult to establish. A lack of standardization of anatomical and endocrine phenotyping and the limited number of known DSD genes and genotype/correlation has long hampered the field, leaving many patients without a definitive diagnosis. The resulting uncertainty may intrinsically pose a great amount of discomfort to affected individuals and their families. DSD-causative genes have historically been identified thanks to positional cloning of disease-associated variants segregating in families or chromosomal rearrangements. Recent advances of chromosomal microarray and exome sequencing technologies are allowing for higher rates of diagnostic success for DSD patients and are changing clinical practice. In this review, we discuss the application of these technologies and their findings as an upcoming model for clinical diagnosis of DSD. We show that exome sequencing is a valuable tool and we propose that it should be used as a first-tier diagnostic technique because it allows for early identification of a genetic cause that may be critical for patient management.

The Pathways of Sex Development

There are many interesting biological processes occurring during embryonic development, but perhaps the most captivating one is the development of one's sex. The significance is not only biological, as it allows deciphering of underlying mechanisms of reproduction, fertility and sex differences in physiology and medicine, but also human, allowing for better care of chronic conditions involving sex organs and their societal implications.

Sex development in humans is divided into two sequential steps: sex determination and sex differentiation (Figure 1-1). Sex determination refers to the expression of gene networks that direct the development of undifferentiated bipotential gonads into either testes or ovaries. The earliest evidence of differential gene expression within the bipotential gonads occurs when *SRY* is upregulated in male gonadal somatic cells, initiating a cascade of genetic events that lead to testicular organogenesis [1-3]. It was shown in mice that expression of *Sry* and *Sf1/Nr5a1* in Sertoli cells initiates upregulation of *Sox9* [4], which organizes Sertoli cells into testicular tubular cords[5]. In the absence of *SRY* the female-specific pathways are initiated, promoting development of the ovaries. While female sex determination has been studied to a lesser degree, several genes are known to be required for proper ovarian development including *RSPO1 [6]*, *WNT4 [7-9],* and *DAX1/NR0B1 [10].*

Once developed, testes and ovaries secrete hormones that promote further sex differentiation of the body throughout embryonic development and adulthood. Hormones secreted from testes are essential for development of male internal (epididymis, vas deferens, seminal vesicles) and external (penis, scrotum) genitalia. In males, secretion

of anti-Müllerian hormone (AMH) by Sertoli cells stimulates regression of the Müllerian structures that would develop into female internal genitalia. Secretion of testosterone and insulin-like peptide 3 by Leydig cells promotes development of male internal structures and testicular descent respectively. Development of external genitalia and prostate is mediated by conversion of testosterone to dihydrotestosterone (DHT) by 5α-reductase. In females, development of internal genitalia (uterus, Fallopian tubes, and upper third of vagina) is driven by absence of AMH and testosterone and, in the absence of DHT, virilization of external genitalia does not occur.

Anomalies in these developmental pathways lead to Disorders of Sex Development (DSDs), defined as “congenital conditions in which the development of chromosomal, gonadal or anatomic sex is atypical” [11]. Most of the genes known to drive sex development were identified as mutated in patients with DSD. Study of these disorders increases our understanding of sex differences, biology of reproduction, and plays a crucial role in understanding human chronic conditions affecting the sexual organs.

The incidence of each discrete DSD is relatively low; however, the umbrella term DSD encompasses both rare and not so rare conditions, ranging from mild hypospadias (abnormal location of the urinary meatus) to gonadal dysgenesis (and complete discrepancy between sex chromosomes and external genital phenotype), affecting a larger population than generally assumed. The birth of a child with a DSD is believed to be extremely stressful for families, bringing uncertainty in regards to child’s gender, medical treatment and future psychosexual development [11-13]. However, evidence-based management of these patients is lacking [14].

Examples of Disorders of Sex Development

Disorders of Sex Chromosomes:

Sex chromosome abnormalities in DSD are primarily due to nondisjunction errors in paternal meiosis and are defined as sex chromosome X or Y aneuploidies. This is unlike other autosomal aneuploidies that primarily arise due to nondisjunction events during maternal meiotic division.

Turner Syndrome

Monosomy X, the underlying chromosomal defect in Turner syndrome patients is estimated to be relatively common. However, in the majority of the cases (>97%) it is lethal [15] and 45,X is thought to be the most frequent karyotype of spontaneous abortuses. The karyotype of the majority of Turner cases is 45,X with an estimated frequency of 1 in 2000 female births [16]. A minority of Turner cases have 46,XX karyotype with partial deletion on one X chromosome; mosaics of 45,X and 46,XX or 45,5 and 46,XY. Patients diagnosed with Turner syndrome generally have short stature due to lack of a copy of the homeobox gene *SHOX,* located in the pseudoautosomal region of the X chromosome. Ovarian development is compromised and gonads appear as streak (fibrous tissue) with poorly developed follicles [17]. Lower estrogen secretion due to underdeveloped follicles results in delayed puberty and primary amenorrhea. In a patient with Turner-like characteristics, a genetics diagnosis can be achieved by karyotyping for either 45,X or mosaics. Fluorescent *in situ* hybridization (FISH) or chromosomal microarray (CMA) should be performed in 46,XX patients to determine if a deletion is present on a pseudoautosomal region of X chromosome.

Klinefelter Syndrome

Klinefelter syndrome affects males and the frequency estimates are difficult to make because many males with this syndrome live undiagnosed [18]. The karyotype of affected individuals is 47,XXY. Klinefelter patients are born with normal number of primordial germ cells that degenerate during childhood [19]. Intellectual abilities of Klinefelter males may be similar to that of the general population, but possibly lower than their siblings'. Testes are small and show low levels of androgen production [20] (hypergonadotropic hypogonadism). Genetic diagnosis is generally achieved by karyotyping, but can also be detected by CMA [21].

Disorders of Hormone Synthesis or Action:

Congenital Adrenal Hyperplasia

Congenital adrenal hyperplasia (CAH), an autosomal recessive disorder, can affect both 46,XX or 46,XY individuals, with an approximate frequency of CAH is 1 in 15,000 people. However, the DSD cases are 46,XX females, most frequently with 21-hydroxylase (*CYP21)* deficiency. Mutations in 21-hydroxylase block the production of aldosterone and cortisol. The excess precursors are used in testosterone biosynthesis, which increases the degree of virilization in 46,XX females. 46,XY individuals share the adrenal crisis phenotype but have a more poorly described genital phenotype. They have no external genital anomalies, but may have precocious puberty. An additional 5% of 46,XX CAH cases are explained by mutations in 11β-hydroxylase *(CYP11B1)* needed for production of cortisol [22].

More rarely, mutations in 3β-hydroxysteroid dehydrogenase 2 *(HSD3B2)* or 17α-hydroxylase *(CYP17)* are responsible for CAH in 46,XY individuals who present with ambiguous genitalia. *HSD3B2* mutations cause adrenal insufficiency in both 46,XX and 46,XY cases due to lack of production of cortisol and aldosterone and additionally genital ambiguity in 46,XY cases due to lack of production of testosterone.

The phenotypes of CAH patients are rather variable ranging from highly masculinized females to highly feminized males. This is due to multitude of possible mutations in both 46,XX and 46,XY karyotypes, their severity and resulting degree of *in utero* androgen exposure. Diagnosis is generally performed using biochemical tests for elevation of 17-OH progesterone, but endocrine profile overlap frequently yields ambiguous diagnoses[23]. The genetic diagnosis is identified by either single gene testing or targeted sequencing of several genes.

Androgen Insensitivity Syndrome:

Androgen insensitivity syndrome (AIS) affects 46,XY individuals and has an estimated frequency of 1 in 20,000 people. Individuals with AIS have normal testicular development that is capable of producing normal levels of testosterone and other testicular hormones. However, the effects of the circulating testosterone are minimized by a dysfunctional androgen receptor *(AR)*. To this day more than 300 mutations have been identified in *AR* with varying degrees of impairment of androgen binding [24]. The phenotypic spectrum of individuals with AIS is wide and can range from complete female appearance (CAIS) to minimal alteration of a typical male phenotype (MAIS). Individuals with partial AIS (PAIS) generally present with various degrees of genital ambiguity.

Typically, CAIS presentation is during puberty, when affected women present with primary amenorrhea. Further examination reveals absent uterus, testosterone dependent structures such as prostate, short vagina and palpable inguinal or labial testes. PAIS presentation is during infancy with micropenis, hypospadias, cryptorchidism or genital ambiguity. The diagnosis is achieved by sequencing of *AR* gene, and biochemical assays of AR function have been abandoned because of high variability [25].

Disorders of Gonadal Development:

46,XY DSD with Gonadal Dysgenesis:

46,XY disorders of sex development are characterized by 46,XY karyotype and intra-abdominal bilateral dysgenetic gonads. Individuals with complete gonadal dysgenesis (CGD) present with female external genitalia, hypoplastic gonads mainly composed of fibrous tissue, no production of sperm, AMH or testosterone, and presence of normal Müllerian structures such as uterus and Fallopian tubes [26]. Individuals with partial gonadal dysgenesis possess varying degrees of dysgenetic testes, ambiguous genitalia, mild to severe hypospadias, low sperm production, Mullerian structures ranging from mild to full development [26].

While actual percentages are difficult to evaluate, historical data suggest that approximately 15% of 46,XY gonadal dysgenesis cases are due to *SRY* mutations or deletions [27-29], 13% to mutations in *SF1/NR5A1* [30], and a few cases due to other rare genetic causes (*SOX9, DAX1/NR0B1,...*) leaving 70% with no known genetic

etiology. More recent data suggest that MAP3K1 variants may explain an additional 10-18% [31, 32].

46,XX Testicular DSD:

46,XX testicular disorders of sex development are characterized by a 46,XX karyotype; external genitalia ranging from ambiguous to that of a typical male; presence of varying amounts of testicular tissue; absence of Mullerian structures [33]. The incidence of 46,XX testicular DSD is approximately 1 in 20,000 individuals. Among 46,XX testicular DSD cases, 85% have a typical male phenotype after puberty with normal pubic hair and penile size, but consult physicians due to delayed puberty or infertility. Additional 15% of cases present at birth with ambiguous genitalia [33]. The vast majority (80-90%) of 46,XX testicular DSD can be explained by translocations of *SRY*. The remaining 15% who carry no detectable Y material typically have ambiguous genitalia and no known genetic etiology [33, 34]. Recently, the chromosomal rearrangements such as duplication upstream of *SOX9* and duplication involving *SOX3*, have been shown to be involved in 46,XX DSD [35].

46,XX Gonadal Dysgenesis:

46,XX gonadal dysgenesis is characterized by a 46,XX karyotype, normal female external genitalia, but impaired ovarian development. Affected individuals remain infantile and don't undergo normal puberty presenting with amenorrhea and infertility in adulthood. The gonadal histology typically shows streak gonads bilaterally. Mutations in follicle stimulating hormone receptor (*FSHR)* in an autosomal recessive form have been reported in several cases [36-38]. The molecular basis for 46,XX gonadal dysgenesis

and primary ovarian failure (POF) remain largely unknown with an estimated diagnostic rate of 25% [39]. However, genes such as transcription factors (*NR5A1, NOBOX, FIGLA, FOXL2*), folliculogenesis growth factors (*BMP15, GDF9, INHA*), LH/FSH receptors and proteins important in DA repair/replication (*STAG3, HFM1, MCM8, MCM9*) have recently been identified in part due to unbiased search using exome sequencing [40, 41].

Ovotesticular DSD:

Formerly ovotesticular DSD (OT-DSD) was known as “true hermaphroditism”, which is defined by the presence of both ovarian and testicular tissue in the same individual. OTDSD is one of the rarest forms of DSD, and one of the most poorly explained mechanistically. The genetic etiology of OTDSD remains unknown for a vast majority of patients.

The gonadal distribution of OT-DSD can be referred to as unilateral (ovotestis on one side and testis or ovary on the other), bilateral (ovotestis on each side), or lateral (testis on one side and ovary on the other) [42]. The most common form is unilateral, with ovotestis/ovary in 34%, followed by bilateral (ovotestis/ovotestis) in 29%, and lateral (ovary/testis) in 25%. The least frequent is unilateral with ovotestis/testis (12%) [42, 43] (Table 1-1).The position of the gonads in patients with OT-DSD corresponds to the amount of testicular tissue present, with a correlation between the degree of gonadal descent and the amount of testicular tissue. About half of the ovotestes are found in an abdominal position (with a quarter inguinal and a quarter labioscrotal), while almost all ovaries (85%) are abdominal and half of the testes are labioscrotal [42].

The geographical distribution of OT-DSD shows an overrepresentation in Africa, with the number of published cases at 17 per 100 million inhabitants, followed by Europe at 15.3. Asia is underrepresented at 1.2 cases per 100 million and South America at 3.6 per 100 million [42].

Fertility in both males and females diagnosed with OT-DSD is significantly reduced. Fertility in patients diagnosed with OT-DSD is significantly reduced: two cases of spermatogenesis and bout a dozen pregnancies have been reported [42]. About a dozen cases of pregnancies in females with OT-DSD have been characterized. However, only half of these females were able to give birth to a healthy child [44-47]. Malignant tumor development was calculated to occur in less than 5% of affected individuals. The susceptibility of patients to development of gonadal tumors seemed to be independent of the individual's karyotype [42, 43]. Sex assignment of OT-DSD has been historically approximately divided equally between male and female [42, 43] (Table 1-2).

Analysis of the chromosomal distribution of ovotesticular DSD shows that about 60% have an XX chromosomal complement, of whom about 10% could be explained by *SRY* translocations, leaving the majority unexplained. The remaining 40% have 46,XY, chimeric or mosaic 46XX/46XY, other sex chromosome mosaic, or partial trisomy 22 karyotypes [34], and the genetic etiology is still elusive [42, 43, 48] (Table 1-3). A Y chromosome is more often observed when there is no ovotestis (e.g. lateral). From a molecular perspective, the presence of *SRY* explains about 10% of XX OT-DSD, and in some patients, it is limited to gonadal mosaicism [49]. A small subset of XY patients with OT-DSD carry point mutations in SRY [50]. The cases of SRY-negative OT-DSD

remain, for the most part, unexplained genetically. Mutations in the *RSPO1* gene have been shown to be associated with a rare syndromic form of OT-DSD with palmoplantar keratosis [6], and there are some reports of partial 22q duplication (containing the *SOX10* gene) with OT-DSD (46,XX) sex reversal with partial duplication of chromosome arm 22q [51]. It seems clear, however, that both *SRY*-positive and *SRY*-negative OT-DSD and testicular DSD (T-DSD) are part of the same clinical spectrum, as OT-DSD and T-DSD can be observed in the same family, and with the same causative genotype, such as the presence of a translocated *SRY* gene [52, 53]. The non-random inactivation of the SRY-bearing X chromosome has been proposed as an explanation for OT-DSD in SRY-positive XX OT-DSD [54].

Animal models of OT-DSD

Since additional genetic factors explaining OT-DSD are yet to be identified, mouse models may become invaluable to this process. C57BL/6J (B6) mice containing the *Mus domesticus* poschiavinus Y chromosome, YPOS, develop ovarian tissue, whereas testicular tissue develops in DBA/2J or 129S1/SvlmJ (129) mice containing the YPOS chromosome. C57BL/6J-YPOS fetuses develop gonads varying from ovary to ovotestis, and constitute an excellent model of OT-DSD. To identify novel genes involved in sex determination, our lab used a congenic strain approach to determine which chromosomal regions from 129S1/SvlmJ provide protection against sex reversal in XYPOS mice of the C57BL/6J.129-YPOS strain. Genome scans using microsatellite and SNP markers identified a chromosome 11 region of 129 origin in C57BL/6J.129-YPOS mice [55]. To determine if this region influenced testis development in XYPOS mice, two strains of C57BL/6J-YPOS mice were produced and used in genetic

experiments. XYPOS adults homozygous for the 129 region had a lower incidence of sex reversal than XYPOS adults homozygous for the B6 region. In addition, many homozygous 129 XYPOS fetuses developed normal-appearing testis, an occurrence never observed in XYPOS mice of the C57BL/6J-YPOS strain. It was concluded that a chromosome 11 locus derived from 129Sl/SvlmJ protects against sex reversal in XYPOS mice. Further back-crossing of C57BL/6J.129-YPOS with C57BL/6J mice combined with bioinformatics approaches based on sequence conservation allowed refinement of the critical region of chromosome 11 protecting against XY sex reversal. Candidate genes generated by this mouse model, combined with other genetic approaches involving human cases of OT-DSD will help elucidate the biology of OT-DSD (See chapter 3). Other animal models of OT-DSD have been described, and could also help the identification of molecular mechanisms of this rare condition [56, 57].

OT-DSD is a rare condition affecting gonadal development and, although it has fascinated generations of clinicians and scientists, it remains mysterious in terms of molecular and genetic mechanisms. It is likely, however, that the advances in next-generation sequencing, with the possibility of performing whole exome sequencing rapidly and at a relatively low cost [58], will eventually allow investigators to understand the development of both testicular and ovarian tissue in the same individual, and therefore allow for rapid diagnosis.

Current Genetic Diagnostic Practice

The typical trend of diagnostic approach for patients with DSDs, who present in clinic with atypical features that do not match their genotypic sex, is to search for additional phenotypic and metabolic information through imaging studies and endocrine

tests. This type of approach helps clinicians select candidate gene for testing that could explain the patient's phenotype. This is usually inefficient as phenotypes frequently overlap and phenotype/genotype correlations are still poor. Thus, the current standard of genetic diagnosis of DSDs relies on sequential sequencing of genes, handpicked as likely candidates for the given phenotype, performed after imaging and endocrine exploration is exhausted, an ineffective, costly, and time-consuming practice. Genetic testing is also sometimes viewed by providers – and insurance – as dispensable. Due to these limitations, many patients do not receive a definitive clinical diagnosis.

The developing diagnostic practice focuses on using next-generation sequencing and chromosomal microarray (CMA) techniques to identify causative mutations or copy-number variants (CNVs) that would guide the diagnostic process (Figure 1-2). Genetic testing would be performed alongside urgent metabolic tests that may be needed to rule out the life-threatening condition Congenital Adrenal Hyperplasia. The benefits of using these genetic diagnostic tools are twofold. From the practitioner's perspective, these tools will allow for a greater diagnostic specificity and shorter turnaround times. They permit detection of mutations, microdeletions and microduplications for known and novel genes, which allows for better classification and more precise outcome studies. Advantages also include noninvasiveness, opposed to hormonal stimulation tests used to test the responsiveness of a specific biosynthetic pathway – which may modify the phenotype of a patient in a way that may be discordant with the chosen gender. Moreover, by knowing the genetic diagnosis, physicians will be able to monitor risks associated with a particular genotype avoiding unnecessary, costly and invasive exploratory endocrine or imaging tests. Rather, a limited number of these tests will be

specifically guided by the genetic findings. From the patient's perspective, it will serve as a shortcut to the long, potentially stressful waits for diagnosis experienced by families. It will also be critical for genetic counseling, reproductive options and associated risks.

Diagnostic Capabilities of Chromosomal Microarrays

In comparison with classical karyotyping techniques, the high-resolution scanning of the entire genome afforded by chromosomal microarrays (CMA) has advanced the field of genetic diagnosis with increased precision[59].

Two types of microarrays, Comparative Genomic Hybridization (CGH) and Single-Nucleotide Polymorphism (SNP) arrays, have been used in genetic testing to detect chromosomal defects. Array CGH uses direct comparison of the genomic DNA sample to be tested with a control. Hybridization of these two samples, labeled with different fluorescent tags, to an array of cloned genomic DNA fragments allows visualization of CNVs, micro-deletions or micro-duplications. In SNP arrays a single, fluorescently labeled DNA sample is hybridized to an array of synthetic oligonucleotides representative of the genome and the relative intensities of the test sample are compared to those of a pool of control samples available on the web. Due to uneven distribution of SNPs throughout the genome, SNP arrays initially did not have a uniform coverage; however, addition of additional oligonucleotide probes allowed for an increased genomic coverage. Recently, manufacturers combined these techniques, which now allow for detection of CNVs as well as regions of heterozygosity and homozygosity, with increased precision. Detection of CNVs in genomic regions with known DSD genes can dramatically increase chances for definitive genetic diagnosis or,

if CNVs are identified in regions with genes not known to be involved with DSDs, it could facilitate identification of novel genes involved in sex development.

For example, CNVs of *DAX1*/*NR0B1* are a known cause of DSD: deletions cause Adrenal Hypoplasia Congenita and Hypogonadotropic Hypogonadism [60] while duplications cause 46,XY DSD with complete gonadal dysgenesis [61]. CMA is now able to routinely identify cases of duplications of the X chromosome that include *DAX1*/*NR0B1* gene in patients diagnosed with the latter phenotype [62, 63].

Another example of diagnostic success of CMA is the validation of the candidate gene *SOX3* as an etiology in 46,XX DSD. Previously, *SOX3* had not been associated with DSD because mutations did not result in anomalies of sex determination either in humans or mice [64]. However, ectopic expression of *Sox3* in gonads of transgenic mice caused sex reversal of XX animals to phenotypic males, suggesting a dosage-dependent effect of this gene in sex determination [65]. CMA screening of human *SRY*-negative XX males identified *SOX3* duplications in several patients [65], a finding confirmed in another two recent clinical case reports [66, 67]. These reports strongly suggest that increased expression of *SOX3* is sufficient to drive the male sex determination pathway in the absence of *SRY*.

Routine use of CMA analysis in clinical practice can also potentially identify novel candidate genes as well as CNVs in gene regulatory regions of known DSD-associated genes, which are not scrutinized when sequencing coding regions. A recent case report identified a deletion in the *WWOX* gene in a 46,XY DSD patient with ambiguous genitalia and gonadal germ line tumors [68]. Although *WWOX* had not been previously associated with DSDs, a knockout mouse model showed abnormal gonad development

[69], which led to the genetic diagnosis. CNVs identified outside of coding regions of the genome that are close to known genes involved in DSDs could be important, previously unrecognized, regulatory sequences that affect gene expression. For example, array CGH identified deletions upstream of the *SOX9* gene open reading frame region, as the etiology of 46, XY gonadal dysgenesis [63, 70] and very small duplications in 46,XX Testicular and Ovotesticular DSD [71-73]. Increasingly refined SNP arrays are helping delineate these upstream sequences necessary for correct testis-specific regulation of *SOX9 [72]*.

To date, microarrays have yielded dozens of candidate regions whose CNV may cause DSD. In addition to *SOX9, DAX1/NROB1, SOX3, WWOX* described above, deletion of 9p24, including the *DMRT1* gene [62, 74-76] and duplication of *VAMP7* in Xq28 [77] are well documented. Other regions await validation [62, 74-76].

CMAs are becoming increasingly used by clinics for many patients. In a study of 23 patients diagnosed with 46, XY gonadal dysgenesis, likely causative CNVs were identified in 3 cases [63]. In a different study of a cohort of 16 *SRY*-negative XX males screened by CMAs, likely causative CNVs were identified in 3 patients [65]. Taken together, these and other studies suggest that approximately 15% of DSDs may be caused by CNVs. Higher rates may be seen in syndromic than isolated forms of 46,XY gonadal dysgenesis (25% vs 5.6%, respectively) [62]. In a cohort of 116 patients that included patients with only hypospadias and/or cryptorchidism (and no further complex genitalia), less profound differences were observed: non-polymorphic, potentially clinically significant CNVs were found in 17% of isolated and syndromic cases of hypospadias, 18% of isolated vs. 30% of syndromic cryptorchidism, and 28% isolated

vs. 16% syndromic ambiguous genitalia [76]. Actual rates are likely to be refined as methods evolve to allow identification of smaller CNVs and more genomes are analyzed, facilitating the accuracy of more polymorphic vs. pathogenic variant calls.

New Diagnostic Practice: Next-Generation Sequencing Technologies

Although CMA has proved to be a useful diagnostic technique for detecting chromosomal abnormalities with higher precision than traditional karyotype, it is ineffective for detection of small genetic variations (the current cutoff for clinical CMA reporting is 50 kb). Next-generation sequencing technologies have dramatically decreased both the cost and turnaround time for whole-genome and exome sequencing revolutionizing the diagnostic methods for complex disorders [78, 79]. Exome sequencing covers approximately 95% of the RefSeq protein-coding regions of the genome, which currently harbor 80-90% of known disease-causing variants [80]. Advantages of this approach are that all genes involved in sex development can be analyzed simultaneously and newly discovered genes can be easily included in the analysis without re-sequencing the sample.

The UCLA Clinical Genomics Center has performed exome sequencing on 90 research and 35 clinical DSD proband samples. For clinical samples, we utilize two gene lists to filter variants specifically in genes of interest. The primary gene list contains genes that have been shown to cause at least one published case of human DSD (Table 1-4). The secondary gene list contains genes that are more loosely related to sex development. For example data may only exist from animal models or the Online Mendelian Inheritance of Man (OMIM) description contains keywords related to sex development. Following American College of Medical Genetics and Genomics (ACMG)

guidelines [81], the Center reports variants when pathogenic (previously reported in humans as the recognized cause of the disorder) or likely pathogenic (previously unreported in a known human DSD gene and of a type expected to cause the disorder). In addition, variants of unknown significance (VUS) in genes of the secondary gene list are frequently reported with the expectation that these might help guide the ordering physician's further imaging and endocrine exploration.

For research samples, in addition, we search for significant variants outside the DSD gene lists. Since on average, a single exome sample generates around 20,000 variants compared to reference genomes, it is most effective to perform WES of the patient as a trio, with phenotyped parents (for example to identify inherited vs. *de novo* variants [82]), which may increase the diagnostic capabilities of WES by almost 10% when compared to proband-only cases [78]. Generally, the ability of WES to provide an accurate genetic diagnosis heavily relies on the existing literature, good clinical assessment and presence of parental samples.

At UCLA, all cases are reviewed at a weekly Genomic Data Board meeting, attended by clinical geneticists, genetic counselors, sequencing laboratory personnel, researchers and, if available, the ordering physician to discuss the relevance of exome variants in each case before clinical report [78].

Diagnostic Capabilities of Exome Sequencing

Several clinical WES laboratories around U.S. have reported greater diagnostic yields when using exome sequencing than using traditional molecular diagnostic methods such as single gene testing or panels [78, 79]. The diagnostic rate of exome

sequencing at UCLA's Clinical Genomics Center ranged from 22% (single cases) to 31% (trio cases) in a cohort of 814 consecutive cases [78]. Similar diagnostic rates were seen at BCM's Human Genome Sequencing Center, where the overall diagnostic rate in a cohort consisting of 504 patients was 25.2% [79]. The diagnostic rates for different phenotypic groups had a greater variability and ranged from 13% to as high as 48% [78, 79, 83-85]. In many cases the precise genetic diagnosis has altered the medical treatment and management.

A recent report by our laboratory illustrates the power of exome sequencing in terms of its diagnostic capabilities and ability to guide DSD patient management [31]. We studied a cohort of 40 DSD patients with a 46, XY karyotype, either clinical samples sent to the UCLA Clinical Genomics Center or research samples from our laboratory. Most of the latter had previously undergone extensive, unsuccessful, endocrine and genetic testing including single-gene and *SOX9* promoter sequencing as well as CMA.

In that cohort, we were able to identify a significant variant in the primary gene list in half of the samples, including a pathogenic variant in a quarter of them, and another 10% that received a likely pathogenic call. Exome sequencing identified rare, sometimes unsuspected genetic causes of DSD, including novel variants and diagnostic variants in genes for which individual gene testing was not available on a clinical basis in the US. For these patients, this dramatically streamlined the diagnosis process and, most importantly, in several cases, the early genetic diagnosis was critical to orient clinical management.

For example, patients presenting clinical evidence of androgen resistance or insensitivity (AIS), *i.e.* male levels of testosterone, yet female genitalia, constitute a

genetic diagnostic challenge. Those who have complete AIS typically have mutations in the androgen receptor [86]. However, those with various degrees of androgen insensitivity often don't, and genetic diagnosis is rarely achieved [87]. In the subset of patients with a working diagnosis of partial androgen insensitivity (PAIS), we found diagnostic variants in 4 different genes with very different prognoses. Only one of these suggested true androgen insensitivity, a novel, likely pathogenic variant in the androgen receptor [31].

In two women with suspected PAIS and inguinal gonads, exome sequencing generated a diagnosis for a very rare condition, often misdiagnosed as PAIS, HSD17B3 deficiency. One was compound heterozygote for known published pathogenic variants in *HSD17B3*. The other carried a homozygous deletion including exon 1. As WES is not intended to detect larger deletions, an external laboratory confirmed the presence of a 461-bp deletion, too small to be reportable with CMA on a clinical basis. In these cases, early genetic diagnosis was critical, because individuals with HSD17B3 deficiency are not actually androgen-resistant, and these young girls are at risk for virilization at puberty, when production of testosterone and DHT spikes. In these patients, the traditional diagnostic approach had either failed or been inapplicable. One patient had been tested for *SRY* presence, *AR* mutations and Sertoli cell function, none of which resulted in a diagnosis. The other patient came to the attention of a U.S. team at age 7. Response to androgen tests would not be performed to avoid masculinization in a patient with established female gender and sex of rearing. In any case, hCG stimulation testing to assess testosterone synthesis pathway function would not have been informative as she had already undergone bilateral gonadectomy.

In that cohort, exome sequencing also allowed to make a diagnosis of Leydig Cell Hypoplasia, caused by a homozygous nonsense mutation in *LHCGR*. *LHCGR* codes for the receptor for the brain-derived hormones that control the maturation of testosterone-producing Leydig cells in the testis. This very rare syndrome had not been suspected and this brought closure to a very long diagnostic odyssey for this family. In this case, years of traditional diagnostic testing had not yielded a plausible result, introducing anxiety and uncertainty to the patient's and family's lives.

Finally, we identified a missense variant in *NR5A1/SF1*, a gene associated with a wide range of DSD phenotypes [88, 89]. This particular variant had previously been reported in a male with isolated distal hypospadias, generally considered a mild form of DSD [90]. This turned out to be extraordinarily important for this patient. She had been raised as female, but did not feel comfortable in that role. The original diagnosis of AIS meant that she would be unlikely to respond to testosterone treatment but, having self-administered testosterone, she felt she had responded to it. The finding of an *NR5A1* variant previously reported in a male was very reassuring, supported her feeling that she should be male and, ultimately, her transition to a male body habitus.

The early genetic diagnosis offered by exome sequencing is an invaluable tool for such cases where the candidate gene approach is ineffective and early diagnosis is critical to optimal patient management. We would like to propose that, outside of the cases of medical emergency, exome sequencing (with chromosomal microarrays in cases of syndromic DSD) should be considered as a first-tier diagnostic or rule/out diagnosis tool for DSD, and be used to guide further imaging and endocrine exploration.

Targeted panels using next generation sequencing have also proven to be very efficient diagnostic tools [91].

These data show that exome sequencing is capable of providing DSD patients with early genetic diagnosis that may be critical for their clinical management and life-long choices. These data also show that exome sequencing can and should be used as a diagnostic tool and not only as a research tool, an argument generally used by US insurance companies to deny coverage.

At this time, insurance coverage is often denied for the clinical exome sequencing, on the grounds that it may bring information important for research or for reproductive options for the family, but not for actual management of the individual patient. Exome sequencing has now been available on a clinical basis in the US for 5 years. Hopefully, the examples above will help dispel this misconception and allow generalization of the practice of exome and, in a close future, whole-genome sequencing.

Ethical Considerations and Future Directions

When sequencing the genome of an individual with a certain disorder there is a possibility that some variants will be identified in genes not related to the condition for which the patient is referred. The current guidelines from ACMG for so-called “incidental findings” are to report pathogenic and likely pathogenic mutations in a selected list of high-penetrance, potentially lethal conditions for which other diagnostic methods and medical interventions are available to the ordering clinician regardless of patients’ age [92]. The same guidelines apply for sequencing of unaffected parents as part of a “trio”

sequencing. This practice has however been controversial (e.g. [93]) and, currently, the UCLA consent form asks families whether or not they would like to be informed of any medically actionable incidental finding. Recently, the ACMG has recognized this as a valid approach as well [94]. This practice is likely to evolve as the field continues to grapple with these ethical considerations.

It is not always that exome sequencing will yield a genetic diagnosis. A negative exome could be due to several reasons. Firstly, our ability to interpret the pathogenicity of variants identified in known or new genes not previously associated with DSD may require further, extensive functional studies (such as *in vivo*/*ex vivo* gene knock-out/knock-in animal studies or *in vitro* cell cultures to assess gene variant pathogenicity) as well as RNA sequencing from human tissues to asses differential gene expression changes that may be due to identified variants. Secondly, we are reaching the limits of understanding complex conditions by looking only at single genes, when the etiology could be multigenic (which is possible to investigate using exome data on a research basis) or multifactorial. Finally, it is likely that the genetic cause of some DSD may reside in non-coding regions of the genome. These will have to be explored by whole-genome sequencing and other emerging technologies identifying structural variants of the genome.

Next-generation sequencing technologies may also have an impact on our understanding of genotype/phenotype correlation of conditions because they can detect variants in multiple DSD-associated genes without a preconceived idea of the phenotypic spectrum associated with reported mutations in a given gene. For example, we identified the same pathogenic MAP3K1 variant previously described in a familial

case of 46,XY pure gonadal dysgenesis [32] in an individual with an ovotestis [31]. Variants in MAP3K1 are thought to act by shifting the signaling network from pro-male (Sox9/Fgf9) to pro-female (Wnt/beta-catenin) [95], but it had not previously been seen in an individual with an ovotestis.

In a sort of “virtuous cycle,” the increased prevalence of exome and genome sequencing in clinical practice will improve the diagnostic yields of clinical next-generation sequencing. As more genomes are sequenced, associated with detailed phenotyping, our understanding of phenotype/genotype correlation will progress, allowing for a greater diagnostic yield from exome sequencing for single-gene or complex diseases.

All these considerations highlight the need for deep phenotyping and better genotype/phenotype correlation in DSD. To this end, we and others (see acknowledgements) have created the DSD-Translational Research Network, an NICHD-funded network of clinics and research centers, informed by patient advocates. Its overarching goal is to generate evidence for and promote best practices in DSD care to achieve best possible health outcomes and quality of life for patients and their families. The DSD-TRN has created the first DSD registry in the US to collect standardized clinical practice data and standardized measurements in all fields involved in clinical care for DSD families: genetic, psychosocial, endocrine, anatomy/surgery, etc. Data entered into the registry from all the network’s sites will be analyzed to try to derive new genetic etiologies for DSD, refine genotype/phenotype correlations and, eventually offer evidence-based recommendations for optimal practice.

The current/upcoming genetic diagnostic practice that consists of chromosomal microarrays and exome sequencing as a first line test (after life threatening conditions have been ruled out by endocrine tests) is the most effective method to screen patients' genome for CNVs, known DSD genes, and in some cases suggest additional causative variants in genes previously not associated with DSDs. This practice, involving genetics at the heart of DSD diagnostic practice, provides a path towards acceleration of clinical diagnosis and improvement of patient care.

Figures and Tables:

Figure 1-1: Overview of sex determination and differentiation in humans

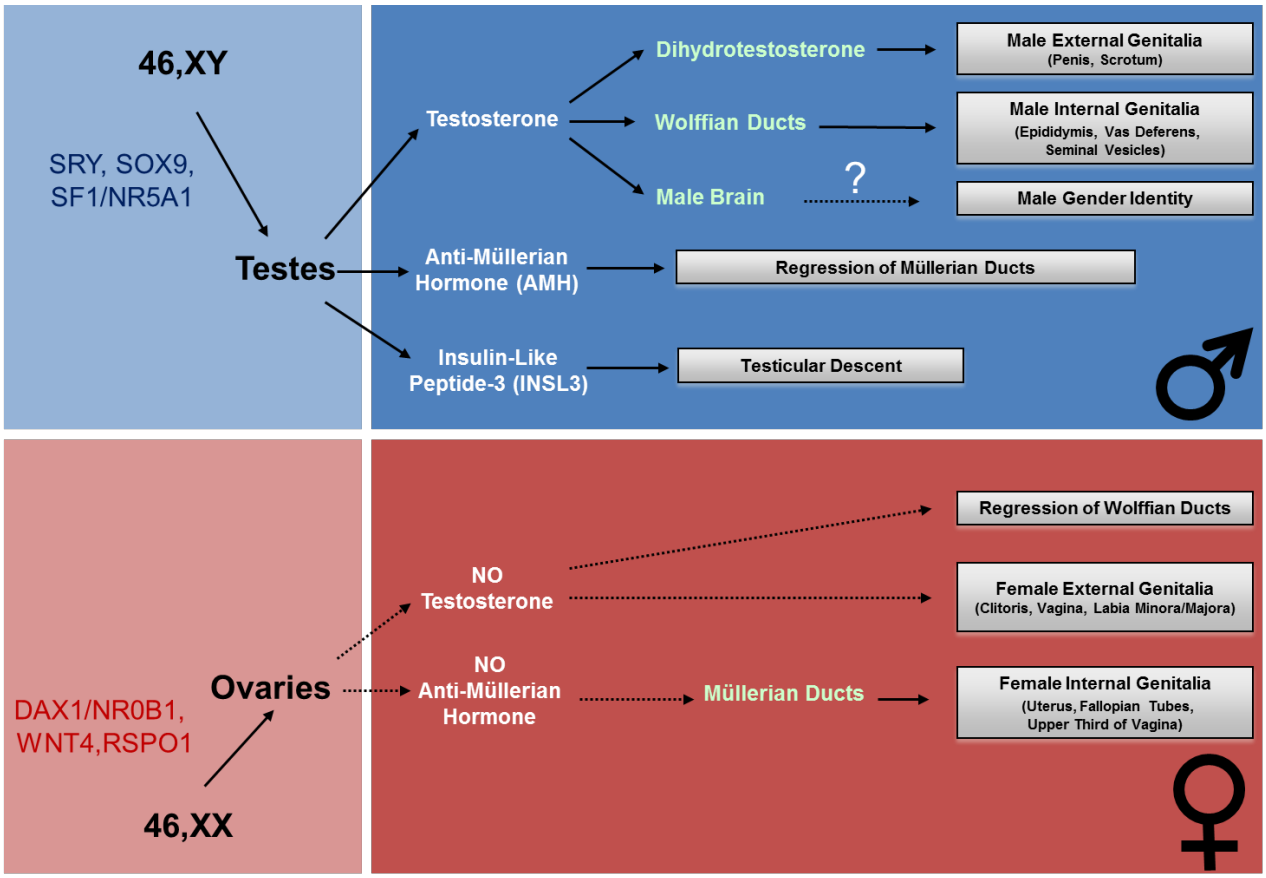


Figure 1-2: Inclusion of genomic technologies in genetic diagnostic practice for DSDs

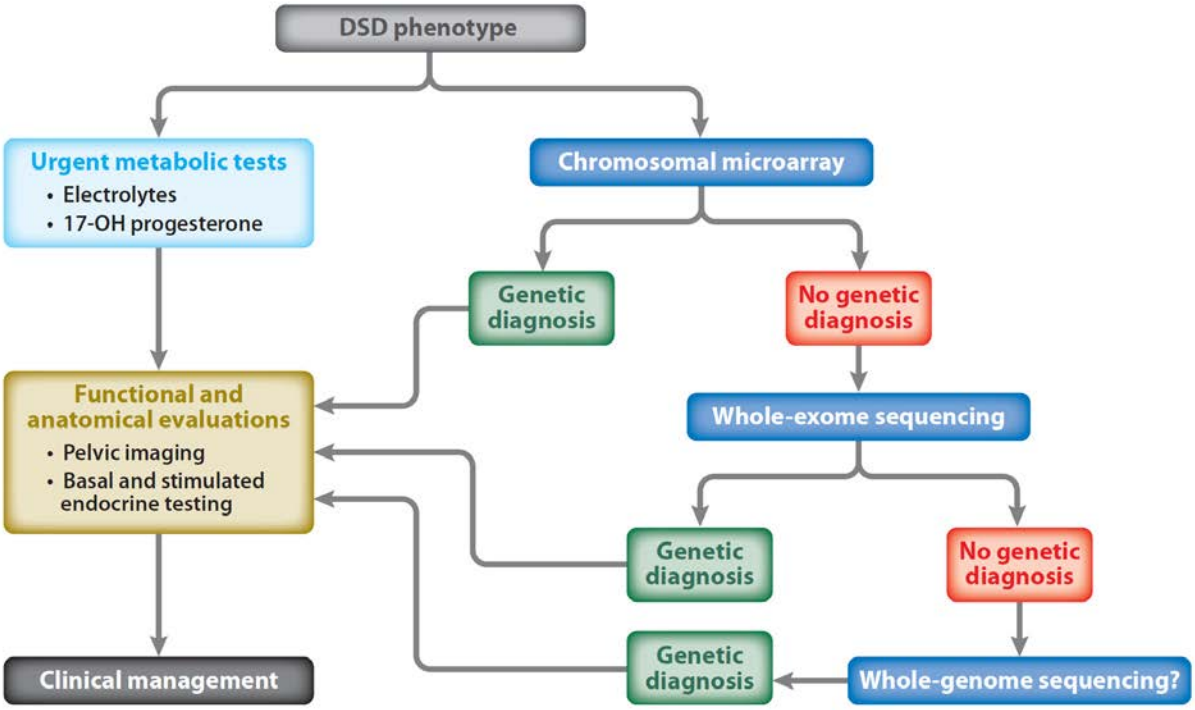


Table 1-1: Gonadal distribution of ovotesticular DSD relative to the presence or absence of Y chromosome

	Unilateral		Bilateral	Lateral
	Ovotestis/Ovary	Ovotestis/Testis	Ovotestis/Ovotestis	Ovary/Testis
Without Y	107 (27%)	29 (7%)	96 (24%)	42 (10%)
With Y	31 (8%)	19 (5%)	19 (5%)	58 (14%)
Total	138 (34%)	48 (12%)	115 (29%)	100 (25%)

Note: Table generated from combined data of [43, 96]

Table 1-2: Sex assignment of ovotesticular DSD relative to the presence or absence of Y chromosome

	Without Y	With Y	Total
Male (54%)	115	63	178
Female (46%)	110	42	152
Total	225	105	330

Note: Table generated from combined data of [43, 96]

Table 1-3: Chromosomal distribution of ovotesticular DSD

Case Reports (1970-1992)	n	%
XX	277	65%
Mosaic with Y	89	21%
XY	40	9%

Note: Table generated from combined data of [43, 96]

Table 1-4: Primary gene list used for variant filtration identified in exome

Gene	Alternative Name	Coverage	Reported Associated Phenotype
Sex Determination (gonadal dysgenesis, testicular and ovotesticular DSD)			
CBX2	CDCA6	100%	46,XY sex reversal
DHH	HHG	100%	46,XY partial or complete gonadal dysgenesis
DMRT1	DMT1	100%	46,XY gonadal dysgenesis
DMRT2		99%	46,XY gonadal dysgenesis
GATA4		82%	46, XY ambiguous genitalia
HHAT		99%	46,XY gonadal dysgenesis
MAP3K1	MEKK	98%	46,XY sex reversal
NR0B1	DAX1/AHCH	100%	46,XY sex reversal
NR5A1	SF1	100%	46,XY sex reversal; 46, XX premature ovarian failure
RSPO1	RSPONDIN	100%	46, XX sex reversal and palmoplantar hyperkeratosis
SOX3	PHP	98%	46, XX sex reversal
SOX9	SRA1	100%	46, XX sex reversal and campomelic dysplasia
SRY	TDF	100%	46,XX testicular DSD and 46, XY ovarian DSD
WNT4		100%	46,XY DSD 46,XY complete gonadal dysgenesis
WT1	AWT1/WAGR	100%	Wilms' tumor-aniridia-genital anomalies-retardation syndrome
WWOX		100%	46,XY gonadal dysgenesis
ZFPM2	FOG2	100%	46,XY gonadal dysgenesis
Sex Differentiation (e.g. steroid synthesis/receptors)			
AKR1C2	BABP/DD/DD2/HAKRD/MCDR2	91%	46,XY DSD
AKR1C4	3-a-HSD, C11/CDR/DD4/HAKRA	100%	46,XY DSD
AMH	MIS	98%	Persistent Müllerian Duct Syndrome (PMDS)
AMHR2	MISR2	100%	Persistent Müllerian Duct Syndrome (PMDS)
AR	AIS	99%	Androgen Insensitivity Syndrome (CAIS/PAIS)
ARX	CT121/EIEE1/ISSX	89%	X-linked lissencephaly with ambiguous genitalia (XLAG)
ATRX	RAD54	100%	Alpha-thalassemia X-linked intellectual disability syndrome
CYP11A1	P450SCC	100%	Congenital adrenal hyperplasia (CAH)
CYP17A1		100%	17-alpha-hydroxylase-deficient CAH
CYP19A1		100%	46,XX and 46,XY DSD
CYP21A2	CA21H/CAH1/CPS1	90%	21-hydroxylase-deficient CAH
FGFR2		100%	Apert syndrome
FOXL2	BPES	97%	Blepharophimosis, ptosis, and epicanthus inversus
HSD17B3	SDR12C2	100%	17-beta hydroxysteroid dehydrogenase III deficiency
HSD3B2	SDR11E2	100%	3-beta-hydroxysteroid dehydrogenase-deficient CAH
LHCGR	LCGR/LGR2/LHR/ULG5	100%	Leydig cell hypoplasia
MAMLD1	CG1/F18/CXORF6	100%	Hypospadias
POR		100%	Cytochrome P450 oxidoreductase deficiency
SRD5A2		100%	Steroid 5-alpha-reductase deficiency
STAR	StAR/STARD1	100%	Cholesterol desmolase-deficient CAH
VAMP7		100% (Chr X) 0% (Chr Y)	46,XY DSD
Central Causes of Hypogonadism			
ARL6	BBS3	100%	Bardet Biedl syndrome
BBS1		100%	Bardet Biedl syndrome
BBS2		100%	Bardet Biedl syndrome
BBS4		100%	Bardet Biedl syndrome
BBS5		100%	Bardet Biedl syndrome
BBS7	BBS2L1/FLJ10715	100%	Bardet Biedl syndrome
BBS9	B1/PTHB1	100%	Bardet Biedl syndrome
BBS10	FLJ23560	100%	Bardet Biedl syndrome
BBS12	FLJ35630/ FLJ41559	100%	Bardet Biedl syndrome
CHD7	FLJ20357/FLJ20361/KIAA1416	100%	Kallman syndrome; normosmic IGD; CHARGE syndrome
FGF8	AIGF	97%	IGD with anosmia (Kallman syndrome) and normosmic IGD
FGFR1	BFGFR/CD331/CEK/FLG	100%	Kallman syndrome, normosmic IGD, and Pfeiffer Syndrome
GNRH1	GNRH/ GRH/LHRH	100%	Isolated abnormality in GnRH secretion or response
GNRHR	LHRHR	100%	Isolated abnormality in GnRH secretion or response
HESX1	ANF/ RPX	100%	Combined pituitary hormone deficiency
HFE	HLA-H	100%	Hemochromatosis
KAL1	anosmin-1/KALIG-1	99%	IGD with anosmia (Kallman syndrome)
KISS1R	AXOR12/HOT7T175	100%	Isolated abnormality in GnRH secretion or response
LEP		100%	Morbid obesity
LEPR	CD295/ OBR	97%	Morbid obesity
LHX3		100%	Combined pituitary hormone deficiency

References:

1. Berta, P., et al., *Genetic evidence equating SRY and the testis-determining factor.* Nature, 1990. **348**(6300): p. 448-50.
2. Koopman, P., et al., *Male development of chromosomally female mice transgenic for Sry.* Nature, 1991. **351**(6322): p. 117-21.
3. 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.
4. Sekido, R. and R. Lovell-Badge, *Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer.* Nature, 2008. **453**(7197): p. 930-4.
5. Svingen, T. and P. Koopman, *Building the mammalian testis: origins, differentiation, and assembly of the component cell populations.* Genes Dev, 2013. **27**(22): p. 2409-26.
6. Parma, P., et al., *R-spondin1 is essential in sex determination, skin differentiation and malignancy.* Nat Genet, 2006. **38**(11): p. 1304-9.
7. Biason-Lauber, A., *WNT4, RSPO1, and FOXL2 in sex development.* Semin Reprod Med, 2012. **30**(5): p. 387-95.
8. Biason-Lauber, A., et al., *A WNT4 mutation associated with Mullerian-duct regression and virilization in a 46,XX woman.* N Engl J Med, 2004. **351**(8): p. 792-8.
9. Jordan, B.K., et al., *Up-regulation of WNT-4 signaling and dosage-sensitive sex reversal in humans.* Am J Hum Genet, 2001. **68**(5): p. 1102-9.

10. Barbaro, M., et al., *Isolated 46,XY gonadal dysgenesis in two sisters caused by a Xp21.2 interstitial duplication containing the DAX1 gene.* J Clin Endocrinol Metab, 2007. **92**(8): p. 3305-13.

11. Lee, P.A., et al., *Consensus statement on management of intersex disorders. International Consensus Conference on Intersex.* Pediatrics, 2006. **118**(2): p. e488-500.

12. Lux, A., et al., *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, 2009. **9**: p. 110.

13. Sandberg, D.E., M. Gardner, and P.T. Cohen-Kettenis, *Psychological aspects of the treatment of patients with disorders of sex development.* Semin Reprod Med, 2012. **30**(5): p. 443-52.

14. Warne, G.L., *Long-term outcome of disorders of sex development.* Sex Dev, 2008. **2**(4-5): p. 268-77.

15. Eiben, B., et al., *A cytogenetic study directly from chorionic villi of 140 spontaneous abortions.* Hum Genet, 1987. **77**(2): p. 137-41.

16. Stochholm, K., et al., *Prevalence, incidence, diagnostic delay, and mortality in Turner syndrome.* J Clin Endocrinol Metab, 2006. **91**(10): p. 3897-902.

17. Reynaud, K., et al., *Number of ovarian follicles in human fetuses with the 45,X karyotype.* Fertil Steril, 2004. **81**(4): p. 1112-9.

18. Bojesen, A., S. Juul, and C.H. Gravholt, *Prenatal and postnatal prevalence of Klinefelter syndrome: a national registry study.* J Clin Endocrinol Metab, 2003. **88**(2): p. 622-6.

19. Lue, Y., et al., *XXY male mice: an experimental model for Klinefelter syndrome.* Endocrinology, 2001. **142**(4): p. 1461-70.

20. Lanfranco, F., et al., *Klinefelter's syndrome.* Lancet, 2004. **364**(9430): p. 273-83.

21. Emmanuèle C. Délot*, J.C.P., the DSD-TRN Genetics Workgroup, David E. Sandberg, Eric Vilain, *Disorders of Sex Development.* Endocrinol Metab Clin N Am, 2017. **In Press**.

22. Menabo, S., et al., *Congenital adrenal hyperplasia due to 11-beta-hydroxylase deficiency: functional consequences of four CYP11B1 mutations.* Eur J Hum Genet, 2014. **22**(5): p. 610-6.

23. Grimbly, C., et al., *46,XY disorder of sex development due to 17-beta hydroxysteroid dehydrogenase type 3 deficiency: a plea for timely genetic testing.* Int J Pediatr Endocrinol, 2016. **2016**: p. 12.

24. Gottlieb, B., et al., *The androgen receptor gene mutations database (ARDB): 2004 update.* Hum Mutat, 2004. **23**(6): p. 527-33.

25. Brown, T.R. and C.J. Migeon, *Cultured human skin fibroblasts: a model for the study of androgen action.* Mol Cell Biochem, 1981. **36**(1): p. 3-22.

26. Mohnach, L., P.Y. Fechner, and C.E. Keegan, *Nonsyndromic Disorders of Testicular Development*, in *GeneReviews(R)*, R.A. Pagon, et al., Editors. 1993: Seattle (WA).

27. Cameron, F.J. and A.H. Sinclair, *Mutations in SRY and SOX9: testis-determining genes.* Hum Mutat, 1997. **9**(5): p. 388-95.

28. Hawkins, J.R., *Mutational analysis of SRY in XY females.* Hum Mutat, 1993. **2**(5): p. 347-50.

29. Veitia, R., et al., *Mutations and sequence variants in the testis-determining region of the Y chromosome in individuals with a 46,XY female phenotype.* Hum Genet, 1997. **99**(5): p. 648-52.

30. Ostrer, H., *46,XY Disorder of Sex Development and 46,XY Complete Gonadal Dysgenesis*, in *GeneReviews(R)*, R.A. Pagon, et al., Editors. 1993: Seattle (WA).

31. Baxter, R.M., et al., *Exome sequencing for the diagnosis of 46,XY disorders of sex development.* J Clin Endocrinol Metab, 2015. **100**(2): p. E333-44.

32. 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.* Am J Hum Genet, 2010. **87**(6): p. 898-904.

33. Delot, E.C. and E.J. Vilain, *Nonsyndromic 46,XX Testicular Disorders of Sex Development*, in *GeneReviews(R)*, R.A. Pagon, et al., Editors. 1993: Seattle (WA).

34. Zenteno-Ruiz, J.C., S. Kofman-Alfaro, and J.P. Mendez, *46,XX sex reversal.* Arch Med Res, 2001. **32**(6): p. 559-66.

35. Vetro, A., et al., *Testis development in the absence of SRY: chromosomal rearrangements at SOX9 and SOX3.* Eur J Hum Genet, 2015. **23**(8): p. 1025-32.

36. Katari, S., et al., *Novel Inactivating Mutation of the FSH Receptor in Two Siblings of Indian Origin With Premature Ovarian Failure.* J Clin Endocrinol Metab, 2015. **100**(6): p. 2154-7.

37. Kuechler, A., et al., *An unbalanced translocation unmasks a recessive mutation in the follicle-stimulating hormone receptor (FSHR) gene and causes FSH resistance.* Eur J Hum Genet, 2010. **18**(6): p. 656-61.

38. Bramble, M.S., et al., *A novel follicle-stimulating hormone receptor mutation causing primary ovarian failure: a fertility application of whole exome sequencing.* Hum Reprod, 2016. **31**(4): p. 905-14.

39. Persani, L., R. Rossetti, and C. Cacciatore, *Genes involved in human premature ovarian failure.* J Mol Endocrinol, 2010. **45**(5): p. 257-79.

40. Caburet, S., et al., *Mutant cohesin in premature ovarian failure.* N Engl J Med, 2014. **370**(10): p. 943-9.

41. Chapman, C., L. Cree, and A.N. Shelling, *The genetics of premature ovarian failure: current perspectives.* Int J Womens Health, 2015. **7**: p. 799-810.

42. King, N., *The reproductive tract. In: Pathology of Laboratory Animals*, ed. G.F. Benirschke K, Jones TC Vol. I. 1978, Springer, New York, 509-580.

43. van Niekerk, W.A. and A.E. Retief, *The gonads of human true hermaphrodites.* Hum Genet, 1981. **58**(1): p. 117-22.

44. Narita, O., et al., *Pregnancy and childbirth in a true hermaphrodite.* Obstet Gynecol, 1975. **45**(5): p. 593-5.

45. Starceski PJ, S.W., Lee PA, *Fertility in true hermaphroditism.* Adolesc Pediatr Gynecol, 1988. **1**: p. 55-56.

46. Verp, M.S., et al., *Chimerism as the etiology of a 46,XX/46,XY fertile true hermaphrodite.* Fertil Steril, 1992. **57**(2): p. 346-9.

47. Williamson, H.O., S.A. Phansey, and R.S. Mathur, *True hermaphroditism with term vaginal delivery and a review.* Am J Obstet Gynecol, 1981. **141**(3): p. 262-5.

48. McElreavey, K., et al., *A minority of 46,XX true hermaphrodites are positive for the Y-DNA sequence including SRY.* Hum Genet, 1992. **90**(1-2): p. 121-5.

49. 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, 2002. **111**(3): p. 278-83.

50. Hiort, O., B. Gramss, and G.T. Klauber, *True hermaphroditism with 46,XY karyotype and a point mutation in the SRY gene.* J Pediatr, 1995. **126**(6): p. 1022.

51. Seeherunvong, T., et al., *46,XX sex reversal with partial duplication of chromosome arm 22q.* Am J Med Genet A, 2004. **127A**(2): p. 149-51.

52. 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, 1993. **316**(4): p. 375-83.

53. Ramos, E.S., et al., *SRY-negative true hermaphrodites and an XX male in two generations of the same family.* Hum Genet, 1996. **97**(5): p. 596-8.

54. Fechner, P.Y., et al., *Nonrandom inactivation of the Y-bearing X chromosome in a 46,XX individual: evidence for the etiology of 46,XX true hermaphroditism.* Cytogenet Cell Genet, 1994. **66**(1): p. 22-6.

55. Nikolova, G., et al., *The chromosome 11 region from strain 129 provides protection from sex reversal in XYPOS mice.* Genetics, 2008. **179**(1): p. 419-27.

56. Tarkowski, A.K., *True Hermaphroditism in Chimaeric Mice.* J Embryol Exp Morphol, 1964. **12**: p. 735-57.

57. Ward, H.B., T.G. Baker, and A. McLaren, *A histological study of the gonads of T16H/XSxr hermaphrodite mice.* J Anat, 1988. **158**: p. 65-75.

58. Ng, S.B., et al., *Exome sequencing identifies the cause of a mendelian disorder.* Nat Genet, 2010. **42**(1): p. 30-5.

59. Vinci, G., et al., *Association of deletion 9p, 46,XY gonadal dysgenesis and autistic spectrum disorder.* Mol Hum Reprod, 2007. **13**(9): p. 685-9.

60. Muscatelli, F., et al., *Mutations in the DAX-1 gene give rise to both X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism.* Nature, 1994. **372**(6507): p. 672-6.

61. Bardoni, B., et al., *A dosage sensitive locus at chromosome Xp21 is involved in male to female sex reversal.* Nat Genet, 1994. **7**(4): p. 497-501.

62. Ledig, S., et al., *Array-CGH analysis in patients with syndromic and non-syndromic XY gonadal dysgenesis: evaluation of array CGH as diagnostic tool and search for new candidate loci.* Hum Reprod, 2010. **25**(10): p. 2637-46.

63. 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.

64. Weiss, J., et al., *Sox3 is required for gonadal function, but not sex determination, in males and females.* Mol Cell Biol, 2003. **23**(22): p. 8084-91.

65. Sutton, E., et al., *Identification of SOX3 as an XX male sex reversal gene in mice and humans.* J Clin Invest, 2011. **121**(1): p. 328-41.

66. Moalem, S., et al., *XX male sex reversal with genital abnormalities associated with a de novo SOX3 gene duplication.* Am J Med Genet A, 2012. **158A**(7): p. 1759-64.

67. Vetro, A., et al., *Testis development in the absence of SRY: chromosomal rearrangements at SOX9 and SOX3.* Eur J Hum Genet, 2014.

68. White, S., et al., *A multi-exon deletion within WWOX is associated with a 46,XY disorder of sex development.* Eur J Hum Genet, 2012. **20**(3): p. 348-51.

69. Ludes-Meyers, J.H., et al., *WWOX hypomorphic mice display a higher incidence of B-cell lymphomas and develop testicular atrophy.* Genes Chromosomes Cancer, 2007. **46**(12): p. 1129-36.

70. Lecointre, C., et al., *Familial acampomelic form of campomelic dysplasia caused by a 960 kb deletion upstream of SOX9.* Am J Med Genet A, 2009. **149A**(6): p. 1183-9.

71. Benko, S., et al., *Disruption of a long distance regulatory region upstream of SOX9 in isolated disorders of sex development.* J Med Genet, 2011. **48**(12): p. 825-30.

72. Kim, G.J., et al., *Copy number variation of two separate regulatory regions upstream of SOX9 causes isolated 46,XY or 46,XX disorder of sex development.* J Med Genet, 2015.

73. Vetro, A., et al., *XX males SRY negative: a confirmed cause of infertility.* J Med Genet, 2011. **48**(10): p. 710-2.

74. Igarashi, M., et al., *Cryptic genomic rearrangements in three patients with 46,XY disorders of sex development.* PLoS One, 2013. **8**(7): p. e68194.

75. Quinonez, S.C., et al., *9p partial monosomy and disorders of sex development: review and postulation of a pathogenetic mechanism.* Am J Med Genet A, 2013. **161A**(8): p. 1882-96.

76. 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.

77. Tannour-Louet, M., et al., *Increased gene copy number of VAMP7 disrupts human male urogenital development through altered estrogen action.* Nat Med, 2014. **20**(7): p. 715-24.

78. Lee, H., et al., *Clinical exome sequencing for genetic identification of rare Mendelian disorders.* JAMA, 2014. **312**(18): p. 1880-7.

79. Yang, Y., et al., *Molecular findings among patients referred for clinical whole-exome sequencing.* JAMA, 2014. **312**(18): p. 1870-9.

80. Genomes Project, C., et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.

81. Richards, C.S., et al., *ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007.* Genet Med, 2008. **10**(4): p. 294-300.

82. Biesecker, L.G. and R.C. Green, *Diagnostic clinical genome and exome sequencing.* N Engl J Med, 2014. **370**(25): p. 2418-25.

83. Fogel, B.L., et al., *Exome sequencing in the clinical diagnosis of sporadic or familial cerebellar ataxia.* JAMA Neurol, 2014. **71**(10): p. 1237-46.

84. Srivastava, S., et al., *Clinical whole exome sequencing in child neurology practice.* Ann Neurol, 2014. **76**(4): p. 473-83.

85. Yang, Y., et al., *Clinical whole-exome sequencing for the diagnosis of mendelian disorders.* N Engl J Med, 2013. **369**(16): p. 1502-11.

86. Gottlieb, B., L.K. Beitel, and M.A. Trifiro, *Androgen Insensitivity Syndrome*, in *GeneReviews(R)*, R.A. Pagon, et al., Editors. 1993: Seattle (WA).

87. Lek, N., et al., *Low frequency of androgen receptor gene mutations in 46 XY DSD, and fetal growth restriction.* Arch Dis Child, 2014. **99**(4): p. 358-61.

88. El-Khairi, R. and J.C. Achermann, *Steroidogenic factor-1 and human disease.* Semin Reprod Med, 2012. **30**(5): p. 374-81.

89. Philibert, P., et al., *Phenotypic variation of SF1 gene mutations.* Adv Exp Med Biol, 2011. **707**: p. 67-72.

90. Allali, S., et al., *Mutation analysis of NR5A1 encoding steroidogenic factor 1 in 77 patients with 46, XY disorders of sex development (DSD) including hypospadias.* PLoS One, 2011. **6**(10): p. e24117.

91. Eggers, S., et al., *Disorders of sex development: insights from targeted gene sequencing of a large international patient cohort.* Genome Biol, 2016. **17**(1): p. 243.

92. Green, R.C., et al., *ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing.* Genet Med, 2013. **15**(7): p. 565-74.

93. Allyse, M. and M. Michie, *Not-so-incidental findings: the ACMG recommendations on the reporting of incidental findings in clinical whole genome and whole exome sequencing.* Trends Biotechnol, 2013. **31**(8): p. 439-41.

94. Relations, A.M. *ACMG Updates Recommendation on "Opt Out" for Genome Sequencing Return of Results*. 2014.

95. Loke, J., et al., *Mutations in MAP3K1 tilt the balance from SOX9/FGF9 to WNT/beta-catenin signaling.* Hum Mol Genet, 2014. **23**(4): p. 1073-83.

96. Krob, G., A. Braun, and U. Kuhnle, *True hermaphroditism: geographical distribution, clinical findings, chromosomes and gonadal histology.* Eur J Pediatr, 1994. **153**(1): p. 2-10.

Chapter 2

Identification of Causative Genetic Variants in Disorders of Sex Development via Exome Sequencing

Abstract

The Chicago Consensus Conference of 2005 defined Disorders of Sex Development (DSD) as “congenital conditions in which the development of chromosomal, gonadal or anatomic sex is atypical.” DSD diagnoses are difficult to establish. A lack of standardization of anatomical and endocrine phenotyping and the limited number of known DSD genes and genotype/correlation has long hampered the field, leaving many patients without a definitive diagnosis. The resulting uncertainty may intrinsically pose a great amount of discomfort to affected individuals and their families. DSD-causative genes have historically been identified by positional cloning of disease-associated variants segregating in families or chromosomal rearrangements. Recent advances of chromosomal microarray and exome sequencing technologies have allowed for higher rates of diagnostic success for DSD patients and are changing clinical practice. Here, we use exome sequencing to identify the pathogenic variants in DSD cases. We show that a much higher diagnostic yield is reached when using exome sequencing rather than the conventional single gene testing.

Introduction

Disorders of Sex Development (DSD) are defined as “congenital conditions in which development of chromosomal, gonadal, or anatomic sex is atypical” [1, 2]. Individuals with DSD have a discordance between their phenotypic and genotypic sex. Their gonads are usually abnormally developed and are described as dysgenic. The incidence of each discrete DSD is relatively rare; however, the umbrella term DSD encompasses both rare and not so rare conditions, ranging from mild hypospadias to sex reversal with genital ambiguity affecting a larger population than generally assumed. The birth of a child with a DSD is believed to be extraordinarily stressful for families, bringing uncertainty in regards to child’s gender and future psychosexual development [1, 3, 4].

These disorders are difficult to diagnose because little is known about their pathogenesis. When detected at birth by malformation of the genitalia, they are typically considered to be a medical and social emergency, since in many cases decisions about gender assignment, medical treatment and surgery have to be made. Moreover, evidence based management of these patients is lacking [5]. At present, a specific molecular diagnosis is identified in only a minority of DSD patients, leaving the majority of DSD cases unexplained [6-8]. This suggests the existence of a number of unknown sex determining genes.

When a patient presents in clinic with a genital malformation or phenotype that does not match the chromosomal sex, the current trend of care (although not universal practice yet) calls for investigation of a potential chromosomal defect, using array CGH

(Comparative Genomic Hybridization). This technique uses two DNA samples, the patient DNA and control genomic DNA that are hybridized to the array, and the relative fluorescence signals of labeled DNA show regions of gain or loss of genetic material of the patient sample in direct comparison with the control sample. This array is considered the best for detecting gain or loss of genetic material and helps significantly with diagnosis by revealing abnormal copy numbers in regions with known DSD genes. However, this method does not offer clear diagnostic help when deletions/insertions are found in regions of unknown involvement in DSD. Additionally, findings that are smaller than 50kb are unlikely to be included in a clinical report due to inability to interpret the pathogenic significance. The cost for such test is approximately $1,500, but may vary widely based on a test type and institution. Based on the array results, clinician may suspect involvement of a particular gene associated with sex determination. Although our knowledge of sex determining genes has increased, only limited number of them are available for clinical testing, and a single gene sequence can cost up to $1,500, depending on the length and complexity of the target [9]. Thus, the current standards for genetic diagnosis of DSDs are limited to genotyping of one or two genes, chosen as likely candidates based on disease phenotype. Because of this narrow scope and exclusion of unknown genes, large numbers of patients do not receive a molecular diagnosis.

Statistical models for identifying new mutations in unknown genes require collection of either a sample of dozens of families with an identical phenotype and at least two affected members (linkage analysis), or thousands of unrelated patients and controls (genome-wide association studies). Due to the nature and frequency of DSD,

neither approach has proven feasible. However, Next Generation Sequencing (NGS) technologies have a higher potential for diagnostic capabilities in genetics. NGS is capable of sequencing the entire human genome in a matter of days and also allows for a rapid sequencing of smaller selections of DNA by targeted capture of specific set of genes or the entire subset of the genome that is expressed (exome). Here, we use exome sequencing to identify pathogenic variants in patients diagnosed with DSD. This approach improves significantly the limited traditional medical diagnosis, while offering a realistic program for identifying new mutations or disease genes.

Methods

DNA isolation and sequencing

DNA was isolated from peripheral blood Gentra Puregene Blood Kit (Qiagen, Germantown, MD, USA) or saliva collected using ORAgene Discover ORG-500 (DNAgenoteck Ottawa, ON, Canada). Sequencing libraries were created for each individual sample following manufacturer's protocol (Illumina protocol *Preparing Samples for Sequencing Genomic DNA*, p/n 11251892 Rev. A). Exomes were captured using Agilent SureSelect All Exon 50Mb capture kit and and Nextera Rapid Capture (Illumina, San Diego, CA USA) Sequencing was performed on an Illumina HiSeq2000 or HiSeq2500 as 50bp or 100bp paired-end run at the UCLA Clinical Genomics Center. The base-calling will be performed using the real-time analysis (RTA) software provided by Illumina.

Exome analysis

The sequence reads (QSEQ or FASTQ files) were aligned to the human reference genome (hg_g1k_b37 assembly) using Novoalign V2.07.13 from the Novocraft Short Read Alignment Package (http://www.novocraft.com/index.html) or using BWA (Burrows-Wheeler Alignment Tool) [10]. The output BAM files were sorted merged and PCR duplicates were removed using Picard (http://picard.sourceforge.net/). INDEL (insertions and deletions) realignment and recalibration were performed using Genome Analysis Took Kit (GATK) from the Broad Institute (http://www.broadinstitute.org/gatk/). In the past we have reached a mean coverage of over 80x for each sample and approximately 93% of the RefSeq gene coding regions +/-2bp was covered at 10x or greater. Both single-nucleotide variants (SNVs) and small INDELs were called within the Ensembl coding exonic intervals +/-2bp using GATK's Unified Genotyper and recalibrated and filtered using GATK variant-quality score recalibration and variant filtration tools. Consanguinity analysis was performed to identify regions > 100kb of homozygosity using Plink software (http://pngu.mgh.harvard.edu/~purcell/plink/). All high-quality variants were reported and annotated using Variant Annotator X (VAX), a custom-designed using Ensembl variant effect predictor [11] or VarSeq Variant Annotation, Filtering and Interpretation Software (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). All variants were filtered by minor allele frequency (MAF) of <1% and intersected with the DSD gene list to identify mutations in known DSD genes.

.

To assess previously unreported missense variants we used three *in silico* algorithms SIFT [12], PolyPhen [13] and Condel [14]. These algorithms predict the likelihood that a given missense variant is pathogenic, based on conservation of the amino acid across species, the physical characteristics of the altered amino acid and the possible impact on protein structure and function.

Results

Exome sequencing was able to cover approximately 95% of the RefSeq coding regions of the genome, including +/-2bp on each side of each exon. This enabled us to identify the majority of variants in protein-coding regions, which currently are known to harbor 80-90% of known disease-causing variants [15]. To identify potential variants involved in DSD pathogenesis we used a DSD specific gene list to filter out variants identified by exome. For that, we developed a list of well annotated genes involved in sex determination and differentiation, as well as a secondary list of genes that are more loosely associated with sex development e.g. the OMIM (Online Medelian Inheritance of Man) description contains sex specific key words or data indicating involvement in sex development is available only from animal models (Table 1-4, 2-1).

The variants identified by exome sequencing were classified into three categories: causative, likely causative and variants of unknown significance (VUS) following the recommendations of the American College of Medical Genetics [16]. Causative variants are defined as variants that have been previously reported with strong evidence of being pathogenic for a particular DSD phenotype. Likely causative variants are variants that have not been previously reported, but are identified as damaging in DSD genes directly related to the clinical phenotype. Variants of unknown

significance are variants with minor allele frequency of less than 1% that may or may not be the causative. This category excludes benign and likely benign variants. Summary depicted in Figure 2-1.

We performed exome sequencing on a number of individuals diagnosed with DSD both with 46,XY and 46,XX karyotypes, a wide variety of clinical features and phenotypes of external genitalia (outlined in Tables 2-2; 2-3). In the 46,XY DSD subset of cases we were able to identify the genetic diagnosis in 35% of cases most of which had been worked up by candidate gene testing or chromosomal microarrays without a success (Table 2-4). The rate of diagnosis in 46,XX DSD subset of cases was much lower possibly due to smaller number of cases not explained in comparison with 46,XY DSD and smaller number of genes known to be associated with 46,XX DSD. The historic rate of undiagnosed cases for 46,XX DSD is only 10% whereas for 46,XY DSD it is 70%, indicating that more pathways exist in 46,XY DSD pathogenesis.

We have identified a c.2522G>A (p.Arg841His) variant in the androgen receptor (*AR*) in the DSDEX94 patient with a working diagnosis of Complete Androgen Insensitivity Syndrome (CAIS). This variant was predicted to be damaging/pathogenic by all three *in silico* algorithms SIFT, PolyPhen and Condel. Using the androgen receptor database we were able to find a previous report of the same mutation indicating the pathogenicity of the variant [17]. The external appearance of this case was that of a normal female; however, testes were present inguinally producing enough testosterone to be in the male range. Identification of this variant had finally put to an end the genetic odyssey for this patient.

We identified several mutations in the mitogen activated protein kinase x3 1, E3 ubiquitin protein ligase (*MAP3K1)* gene that has been emerging as one of the more common genes mutations in which result in 46,XY DSD. A previously reported c.1846G>A (p.Gly616Arg) pathogenic variant was identified in two of our cases who presented with two different phenotypes (a female with complete gonadal dysgenesis and a male with ovotesticular DSD and ambiguous genitalia respectively) [18]. These findings extend the spectrum of phenotypes associated with this variant. Additionally, we identified another three previously unreported mutations in *MAP3K1* gene c.1016G>A (p.Arg339Gln); c.770C>T (p.Pro257Leu); c.1760T>A (p.Leu587His) in 3 distinct patients DSDEX73; SN00458 and DSDEX118 respectively. All three variants were predicted to be damaging by *in silico* tools. DSDEX73 presented as a 46,XY female with complete gonadal dysgenesis; hypoplastic uterus, fallopian tubes, bilateral streak gonads, primary amenorrhea and tall stature. SN00458 presented as a 46,XY individual with ambiguous genitalia; 2cm phallus, hypospadias, bifid scrotum, bilateral descended testes, no mullerian structures, no gonadal dysgenesis. These two mutations along with previously reported p.Gly616Arg variant indicate that the phenotypic spectrum of mutations in *MAP3K1* is very wide.

The DSDEX118 proband presented as a 46,XY female with complete gonadal dysgenesis and failure to enter puberty. The external genitalia were of a normal female. Internally this patient had small uterus, bilateral fallopian tubes and bilateral streak gonads. The proband also had one affected sister who had undergone bilateral inguinal hernia repair at an early age and one affected cousin with similar evaluations. The identified heterozygous variant c.1760T>A (p.Leu587His), was present in all affected

sisters with 46,XY karyotype and in unaffected aunt who had 46,XX karyotype indicating the 46,XY karyotype necessity for disease progression (Figures 2-2 and 2-3).

We have generally been unlucky at finding pathogenic variants in 46,XX DSD cases with exome sequencing. However, we found a heterozygous c.274C>T (p.Arg92Trp) mutation in *NR5A1* gene in one 46,XX male who did not have a copy of *SRY* gene. This variant was highly conserved, predicted damaging and was not seen in EVS, 1000G or ExAC. Mutations in *NR5A1* have historically be associated with 46,XY sex reversal that is why at first this variant seemed unconvincing. Nevertheless, we were able to find two other families in collaboration with colleagues from France and United States that had an affected family member with similar phenotype as our case and who carried the exact p.Arg92Trp variant [19]. Functional studies show that this variant switches organ fate from ovary to testis through disruption of ovary-specific pathways that normally oppose testis development [19]. Several other groups have also identified p.Arg92Trp to be associated with 46,XX testicular and ovotesticular DSD [20-22].

Discussion

Exome sequencing allows identification of genetic cause in rare cases at much higher success rate when compared with other diagnostic techniques of comparable pricing. The use of gene lists as bioinformatics filters for exome sequencing allows for rapid identification of variants in known DSD genes in an individual sample. Moreover, when no causative variants are identified from one of these gene lists, one is able to analyze the complete exome sequencing data set to look for additional variants that might be disease causing. We prefer to use a trio analysis in which we sequence the

patient and both unaffected parents because each individual exome on average has 20,000 variants compared to the reference sequence [15]. The majority of these variants are inherited, and by including parent's sequences we can define de novo, heterozygous variants in addition to any homozygous or compound heterozygous variants.

Exome sequencing allows for identification of variants in genes that have not been associated with DSD previously, such as the p.Arg92Trp mutation that we identified in *NR5A1* gene in 46,XX males that is now associated with testicular DSD. One of the most important advantages of this approach is that all genes with any involvement in sex development can be analyzed concurrently, and new genes can be readily included in the analysis without needing to reconfigure the sequencing pipeline or re-sequence the samples.

Conclusion

Exome sequencing provides high throughput genetic diagnostic capability that has become the core of modern clinical genetics. In the field of disorders of sex development it has revolutionized the rate of successive genetic diagnosis. Nevertheless, exome sequencing is limited to only 1% of the entire genome which is a big limitation when trying to achieve a high diagnostic rate. Moreover, clinical exome sequencing relies on the information available about the gene such as function and molecular mechanism of action in order to be effective. Much of this information is currently unavailable. Thus, it is necessary to investigate molecular pathways, identify gene functions in order to be able to interpret variants identified by exome sequencing that are classified as "variants of unknown significance".

Figures and Tables

Table 2-1: Secondary gene list used for variant filtration in exome sequencing

Gene ID	Search Term	Gene ID	Search Term	Gene ID	Search Term
AKR1C2	'sexual'	TRIM37	'testicular'	DHCR7	hypospadias
AKR1D1	'steroid'	TRPC6	'steroid'	DISC1	bardet_biedl_syndrome
ATF3	hypospadias'	TTC21B	'+bardet +biedl'	DNMT3L	hypogonadism
AZF1	'+gonadal +dysgenesis'	UGT2B17	'steroid'	EFNB1	sex_reversal
BMP4	'hypospadias'	UGT2B28	'steroid'	EHD1	bardet_biedl_syndrome
BMP7	'hypospadias'	WDPCP	'+bardet +biedl'	ELAVL4	sex_determination
BNC2	'hypospadias'	WDR11	'hypogonadism	EPHX1	disorder_of_sex_development
CCDC28B	'+bardet +biedl'	WWOX	'disorder of +sex +development'	EPHX1	ovotesticular_dsd
CEP290	'+bardet +biedl'	ACP5	steroid_synthesis	ERCC3	sexual_development
CUL4B	'hypogonadism	AHR	steroid_receptors	ERCC3	sexual_disorder
CYB5A	'disorder of +sex +development'	ALMS1	hypogonadism	ERCC5	hypogonadism
CYP11B1	'steroid'	ARL13B	bardet_biedl_syndrome	ESR1	sexual_disorder
CYP11B1	'+Adrenal +Hyperplasia'	ATM	gonadal_dysgenesis	ESR1	steroid_receptors
CYP11B2	'steroid'	ATM	sexual_disorder	ESR2	hypospadias
DND1	'testicular'	AUTS2	hypospadias	ESR2	sexual_differentiation
ESR2	'hypospadias'	BKMA1	hypogonadism	ESR2	sexual_disorder
FOXF2	'disorder of +sex +development'	BKMA1	sexual_differentiation	ESR2	steroid_receptors
FSHR	'hypogonadism	BMPR1A	sexual_development	ESRRA	steroid_synthesis
GHRHR	'+pituitary +hormone deficiency'	BMPR1A	sexual_differentiation	F9	sexual_disorder
HOXA4	'hypospadias'	BRCC3	hypogonadism	F9	steroid_synthesis
HOXB6	'hypospadias'	CAPN5	sex_determination	FANCA	hypogonadism
HS6ST1	'hypogonadism	CCDC28B	bardet_biedl_syndrome	FANCG	hypogonadism
KCNJ5	'+Adrenal +Hyperplasia'	CEP290	bardet_biedl_syndrome	FANCM	hypogonadism
KIF7	'+bardet +biedl'	CEP41	ambiguous_genitalia	FEM1C	sex_determination
KISS1	'hypogonadism	CGNL1	hypogonadism	FEM1C	sexual_development
LHB	'hypogonadism	CITED2	sex_determination	FGF10	hypospadias
LHFPL5	'hypospadias'	CITED2	sex_reversal	FGFR3	disorder_of_sex_development
LHX4	'+pituitary +hormone deficiency'	CLCN4	sexual_differentiation	FGFR3	ovotesticular_dsd
LZTFL1	'+bardet +biedl'	CLTCL1	hypospadias	FKBP4	hypospadias
MID1	'hypospadias'	COL2A1	disorder_of_sex_development	FLNA	sexual_differentiation
MKS1	'+bardet +biedl'	COL2A1	ovotesticular_dsd	FLNA	sexual_disorder
MYO1E	'steroid'	COQ2	steroid_synthesis	FREM2	ambiguous_genitalia
NELF	'hypogonadism	COX14	Adrenal_Hyperplasia	FSHB	hypogonadism
NMT2	'hypogonadism	CRH	sexual_disorder	FSHR	hypogonadism
NPHS2	'steroid'	CSF2RA	sexual_disorder	G6PC2	bardet_biedl_syndrome
OTX2	'+pituitary +hormone deficiency'	CUL4B	hypogonadism	GAL	hypogonadism
PDE11A	'testicular'	CUL4B	hypospadias	GATA5	hypospadias
PDE8B	'+Adrenal +Hyperplasia'	CUL4B	steroid_receptors	GHRHR	pituitary_hormone_deficiency
POLR3A	'hypogonadism	CYP11B1	Adrenal_Hyperplasia	GLYCTK	pituitary_hormone_deficiency
POU1F1	'+pituitary +hormone deficiency'	CYP11B2	steroid_synthesis	GNAS	Adrenal_Hyperplasia
PSMC3IP	'+gonadal +dysgenesis'	CYP19A1	disorder_of_sex_development	GPC3	ambiguous_genitalia
RET	'+bardet +biedl'	CYP19A1	hypogonadism	HAMP	hypogonadism
RPGRIP1L	'+bardet +biedl'	CYP19A1	ovotesticular_disorder_of_sex_deve	HBA1	hypospadias
SDCCAG8	'+bardet +biedl'	CYP19A1	sexual_disorder	HBA2	ambiguous_genitalia
SDHB	'+Adrenal +Hyperplasia'	CYP19A1	testicular_disorder_of_sex_developr	HBA2	hypospadias
SNORD116	'hypogonadism	CYP7B1	steroid_receptors	HBB	disorder_of_sex_development
SOX2	'hypogonadism	DCAF17	hypogonadism	HBB	ovotesticular_dsd
TDRD7	'hypospadias'	DGKK	hypospadias	HBB	sexual_differentiation
TMEM67	'+bardet +biedl'	DHCR24	steroid_synthesis	HES1	sex_determination
TRIM37	'testicular'	DHCR7	ambiguous_genitalia	HEXB	sexual_disorder

Table 2-1: continued

Gene ID	Search Term	Gene ID	Search Term	Gene ID	Search Term
HFE2	hypogonadism	MSC	sex_determination	SLC6A4	ovotesticular_dsd
HOXA13	hypospadias	MTCP1	hypogonadism	SLC9A3R2	sex_determination
HOXD10	bardet_biedl_syndrome	MTHFR	disorder_of_sex_development	SMPD2	pituitary_hormone_deficiency
HSD11B2	Adrenal_Hyperplasia	MTHFR	ovotesticular_dsd	SMPD3	pituitary_hormone_deficiency
HTR2A	sexual_disorder	MTMR1	sexual_development	SNORD116-1	hypogonadism
HTR3A	sexual_disorder	MTMR1	sexual_disorder	SOD2	hypogonadism
IGF1R	sex_determination	NCOA2	steroid_receptors	SOX10	sex_reversal
IGF1R	sexual_differentiation	NCOA4	steroid_receptors	SOX10	sexual_development
IGFALS	sexual_development	NDN	hypogonadism	SOX10	sexual_disorder
INHBA	hypogonadism	NDUFS4	hypospadias	SOX2	hypogonadism
INHBB	hypogonadism	NEDD4	steroid_receptors	SOX2	hypospadias
INSR	sex_determination	NELF	hypogonadism	SPO11	gonadal_dysgenesis
INSR	sexual_differentiation	NKAIN2	hypogonadism	SRD5A1	hypospadias
INSRR	sex_determination	NMT2	hypogonadism	SRD5A3	sexual_disorder
INSRR	sexual_differentiation	NOS1	sexual_disorder	STAT5B	sexual_development
IRF6	hypospadias	NPC1	steroid_synthesis	STS	hypogonadism
ITGB3	disorder_of_sex_development	NR2C1	steroid_receptors	TAB2	steroid_receptors
ITGB3	ovotesticular_dsd	NR2E3	bardet_biedl_syndrome	TDRD7	hypospadias
KCNJ5	Adrenal_Hyperplasia	NR3C1	Adrenal_Hyperplasia	TFR2	hypogonadism
KDM5D	sex_determination	NR3C1	ambiguous_genitalia	THRB	disorder_of_sex_development
KDM5D	sexual_differentiation	NR3C1	steroid_receptors	THRB	ovotesticular_dsd
KISS1	hypogonadism	NR4A1	sexual_differentiation	TMEM67	bardet_biedl_syndrome
KRT86	sexual_disorder	NR4A1	steroid_receptors	TMEM70	hypospadias
LATS1	hypogonadism	OTX2	pituitary_hormone_deficiency	TNXB	Adrenal_Hyperplasia
LHB	disorder_of_sex_development	PEX2	sexual_disorder	TP63	hypogonadism
LHB	hypogonadism	PGR	steroid_receptors	TPH2	sexual_disorder
LHB	ovotesticular_dsd	PHF6	hypogonadism	TPI1	sexual_disorder
LHB	sexual_development	PKD1	Adrenal_Hyperplasia	TRA2A	sex_determination
LHB	testicular_dsd	PLCB3	bardet_biedl_syndrome	TRA2A	sexual_differentiation
LHX4	pituitary_hormone_deficiency	PLLP	bardet_biedl_syndrome	TRNL1	hypogonadism
LHX9	gonadal_dysgenesis	POLG	hypogonadism	TTR	sexual_disorder
LHX9	sex_determination	POLR3A	hypogonadism	TUB	bardet_biedl_syndrome
LIG4	hypogonadism	POLR3B	hypogonadism	VANGL2	bardet_biedl_syndrome
LIPE	hypogonadism	POU1F1	pituitary_hormone_deficiency	WDPCP	bardet_biedl_syndrome
LMNA	hypogonadism	RAB3GAP2	hypogonadism	XG	sex_reversal
MAGEB1	sex_determination	RCN2	bardet_biedl_syndrome	XG	sexual_development
MAGEB2	sex_reversal	RET	disorder_of_sex_development	ZDHHC21	steroid_receptors
MAGEL2	hypogonadism	RET	ovotesticular_dsd	ZDHHC7	steroid_receptors
MAOA	sexual_disorder	RFXAP	pituitary_hormone_deficiency	ZEB2	hypospadias
MBTPS2	hypospadias	RPL35A	hypospadias	ZFX	gonadal_dysgenesis
MC4R	hypogonadism	RPS4X	gonadal_dysgenesis	ZFX	hypogonadism
MC4R	sexual_development	RPS4Y1	gonadal_dysgenesis	ZFX	sex_determination
MC4R	sexual_disorder	SALL1	hypospadias	ZFY	sex_determination
MEF2B	steroid_receptors	SAT1	sex_reversal		
MEN1	Adrenal_Hyperplasia	SDCCAG8	bardet_biedl_syndrome		
MEN1	disorder_of_sex_development	SHOX	gonadal_dysgenesis		
MEN1	ovotesticular_dsd	SIL1	hypogonadism		
MID1	hypospadias	SLC29A3	hypogonadism		
MKS1	bardet_biedl_syndrome	SLC6A4	disorder_of_sex_development		

Table 2-2: List of DSD patients with corresponding phenotypes and clinical features that have undergone exome sequencing.

Patient ID	Diagnosis	Clinical Features
DSDEX53	46,XY Female, Complete Gonadal Dysgenesis (CGD)	Normal uterus and Fallopian tubes, streak gonads removed
DSDEX02	46,XY Female, CGD	-
DSDEX01	46,XY Female, CGD	No gonads found, cerebellar hypoplasia, optic nerve hypoplasia, ear pits, blindness
DSDEX21	46,XY Female, CGD	Streak gonads, rete testis and seminiferous tubules
DSDEX54	46, XY Female, CGD	-
DSDEX73	46, XY Female, CGD	Hypoplastic partial bicornuate uterus with cervix, Fallopian tubes; bilateral streak gonads; tall stature, primary amenorrhea
DSDEX118	46,XY Female, CGD	3 affected (2 siblings, 1 maternal aunt); Failure to enter puberty; underwent gonadectomy; bilateral streak gonads; small uterus; bilateral fallopian tubes; occasional sertoli-like tubules and ovariant stroma present in gonads; Wolffian ducts also present
DSDEX16	46,XY Female, Partial Gonadal Dysgenesis (PGD)	No uterus, Fallopian tubes present, short vagina, very low T and undetectable Estradiol, gonads not found
DSDEX17	46,XY Female, Gonadal Dysgenesis (GD)	-
DSDEX25	46,XY Female, GD	Adrenal rests
DSDEX111	46,XY Female, GD	-
SN00275-P	46,XY Ambiguous Genitalia, CGD	No oocytes, no seminiferous tubules, no clitoromegaly
DSDEX37	46,XY Female	Clitoromegaly
DSDEX38	46,XY Female	Short stature
DSDEX51	46,XY Female	Kidney disease, possible Denys-Drash syndrome
SN00366-P	46,XY Female	Inguinal testes w/ immature seminiferous tubules, no uterus and Fallopian tubes, deafness, impaired cognition
SN00212-P	46,XY Female	Large clitoris, no uterus and vaginal opening, inguinal testes
DSDEX61	46 XY Female	Complete Androgen Insensitivity Syndrome (CAIS)
DSDEX94	46, XY Female	CAIS
DSDEX36	46,XY Female	No uterus; blind vagina; typical labia
DSDEX62	46, XY Female	No clitoromegaly, non rugated labia, vagina; small bilat. Inguinal gonads; fibrous tissue with Sertoli cells
DSDEX72	46, XY Female	Enlarged clitoris; no uterus; two testes; Androgen Insensitivity Syndrome (AIS)
SN00113-P	46, XY Female	No uterus, no adnexa (U/S); inguinal
SN00375-P	46, XY Female	Bilateral hernias with palpable gonads, normal appearing testes descended into labial folds after orchiopexy
DSDEX32	46, XY Female	No uterus; adrenal insufficiency; small adrenals
SN02421-P	46, XY Female	Growth delay, short stature
DSDEX125	46,XY Female	Multiple congenital anomalies; no uterus; internal gonads - testes
DSDEX23	46,XY Female	Amelia (missing limbs)

Table 2-2: continued

Patient ID	Diagnosis	Clinical Features
DSDEX04	46,XY Ambiguous Genitalia	Partial fusion of labioscrotal folds, small phallus, penoscrotal hypospadias
DSDEX50	46,XY Ambiguous Genitalia	Developmental delay, agenesis of corpus callosum
DSDEX52	46,XY Ambiguous Genitalia	Adrenal Hypoplasia Congenita, other dysmorphic features
DSDEX43	46,XY Ambiguous Genitalia	Microcephaly, intestinal dysmotility, optic nerve hypoplasia
SN00188-P	46,XY Ambiguous Genitalia	Undescended testes, bifid scrotum, hypospadias
SN00246-P	46,XY Ambiguous Genitalia	Bilateral descended testes, midshaft hypospadias, chordee
SN00345-P	46,XY Ambiguous Genitalia	Bilateral descended testes, hypospadias, micropenis, chordee, congenital hypothyroidism
DSDEX05	46,XY Ambiguous Genitalia	Uterus, cervix, Fallopian tube on L. long urethra; disorganized testicular tissue on L; normal testis on R
DSDEX47	46,XY Ambiguous Genitalia	UG sinus, cryptorchidy, "mild phallus", "unfold aread of scrotum"; Left: immature testis, seminiferous tubules, epididymis; Right: fibro-fatty connective tissue, ductal structures consistent with mesonephric ducts, possible vas. Denys Drash syndrome; end-stage renal disease; bilateral nephrectomy before age 2; no Wilms tumor.
DSDEX65	46, XY Ambiguous Genitalia	Microphallus, chordee, hypospadias; incompletely fused scrotum; ovotestis with Fallopian tube & partial uterus; contralateral normal testis
SN00458-P	46, XY Ambiguous Genitalia	Perineal hypospadias, 2 cm phallus, bifid scrotum, penoscrotal transposition, no Müllerian structures; bilateral descended testes; no gonadal dysgenesis; premature birth, IUGR
DSDEX120	46,XY Male, Anorchia	Congenital Bilateral Anorchia; responsive to testosterone; definite penis (mildly shortened); no hypospadias, fully formed scrotum
SN00341-P	46,XY Male, Micropenis	Vanishing testes
SN00441-P	46,XY Male, Micropenis	No uterus or ovaries per ultrasound, ambiguous genitalia, under virilization
DSDEX131	46,XY Male, Micropenis	-
DSDEX29	46,XY Micropenis/Cryptochidism	Severe growth and developmental retardation, testes not seen by ultrasound
DSDEX127	46,XY Male, Micropenis, Hypospadias	Hypogonadism
DSDEX124	46,Male, Hypospadias/Cryptorchidism	Azoospermia, high T levels, PAIS
SN00134-P	46,XY Male, Hypospadias	Bilateral descended gonads (likely testes), bifid scrotum
DSDEX108	46,XY Male, Hypospadias	-
SN02417-P	46,XY Male, Hypospadias	Penoscrotal Hypospadias, chordee, bifid scrotum, cryptorchidism
DSDEX132	46,XY Male, Hypospadias	-
SN00198-P	46,XY Male	Undescended testes, neuropathy, hypotonia
DSDEX57	46, XY Male, PMDS	Hypoplastic vas, epididymis, rete testis, microcalcifications, portion of seminal vesicle & prostatic tissue, primitive Fallopian tube, vagina, endocervix, uterine structure. No gonad on Right; Left: inguinal fibrotic & atrophic testis, no Sertoli; exuberant Leydig cell proliferation, in complex specimen
DSDEX60	46, XY Male, PMDS	Fallopian tubes and small uterus; two abdominal testes, normal testicular tissue

Table 2-3: List of 46,XX DSD patients with corresponding phenotypes and clinical features that have undergone exome sequencing

Patient ID	Diagnosis	Clinical Features
DSDEX105	46,XX Male	SRY-
DSDEX115	46,XX Male	SRY-
DSDEX08	46,XX Male	SRY-
DSDEX09	46,XX Male	SRY-
DSDEX11	46,XX Male	SRY-
DSDEX14	46,XX Male	SRY-
DSDEX55	46,XX Male	SRY-
DSDEX56	46,XX Male	SRY-
DSDEX10	46,XX Male	SRY-, penoscrotal hypospadias, 2 testicles
DSDEX81	46,XX Male	SRY-; severe hypospadias
DSDEX13	46,XX Male	SRY-; abnormal bones
DSDEX26	46,XX Male	SRY-, bilateral cryptorchidism, no uterus
SN00048-P	46,XX Male	-
DSDEX06	46,XX Male	-
DSDEX76	46,XX Male	Hypospadias
DSDEX41	45X, Male	Hypospadias
DSDEX12	46,XX Ambiguous Genitalia	SRY-
DSDEX6	46,XX Ambiguous Genitalia	SRY-
DSDEX78	46,XX Ambiguous Genitalia	SRY-; complex phenotype
DSDEX18	46,XX Ambiguous Genitalia	-
DSDEX07	46,XX Ambiguous Genitalia	-
DSDEX34_1	46,XX Ambiguous Genitalia	-
DSDEX30	46,XX Female MRKH	-
DSDEX03	46,XX Female MRKH	Mullerian duct abnormalities
DSDEX85	46,XX Female POF	-
DSDEX91	46,XX Female POF	-
DSDEX46?	46,XX Female	Duplication of: vagina, uterus, ureter, bladder, colon (other)
DSDEX48	46,XX TH	-
DSDEX101	46,XX Female GD	-

Table 2-4: DSD patients with 46,XY karyotype with corresponding exome result

Patient ID	Phenotype	Diagnosis	Patient ID	Phenotype	Diagnosis
DSDEX54	46,XY Female, CGD	MAP3K1	DSDEX47	46,XY Ambiguous Genitalia	WT1
DSDEX73	46,XY Female, CGD	MAP3K1	DSDEX65	46,XY Ambiguous Genitalia	MAP3K1
DSDEX118	46,XY Female, CGD	MAP3K1	DSDEX05	46,XY Ambiguous Genitalia	WT1
DSDEX01	46,XY Female, CGD	CHD7	SN00458-P	46,XY Ambiguous Genitalia	MAP3K1
DSDEX21	46,XY Female, CGD	DHH	SN00563-P	46,XY Ambiguous Genitalia	DHX37
DSDEX53	46,XY Female, CGD		SN00345-P	46,XY Ambiguous Genitalia	BNC2; FGFR1
DSDEX02	46,XY Female, CGD		DSDEX04	46,XY Ambiguous Genitalia	
DSDEX16	46,XY Female, PGD		DSDEX50	46,XY Ambiguous Genitalia	
DSDEX17	46,XY Female, GD		DSDEX52	46,XY Ambiguous Genitalia	
DSDEX25	46,XY Female, GD		DSDEX43	46,XY Ambiguous Genitalia	
DSDEX111	46,XY Female, GD		SN00188-P	46,XY Ambiguous Genitalia	
SN00275-P	46,XY Ambiguous Genitalia, CGD		SN00246-P	46,XY Ambiguous Genitalia	
SN00815-P	46,XY Female	HSD17B3	DSDEX57	46,XY Male, PMDS	AMHR2
DSDEX62	46,XY Female	LHCGR	DSDEX60	46,XY Male, PMDS	AMHR2
SN00113-P	46,XY Female	HSD17B3	SN00178-P	46,XY Cryptorchidism	PTPN11
SN00375-P	46,XY Female	HSD17B3	SN03225-P	46,XY Male, Hypospadias	MAP3K1
DSDEX32	46,XY Female	STAR	SN00198-P	46,XY Male	MAMLD1
DSDEX94	46,XY Female	AR	SN00341-P	46,XY Male, Micropenis	CHD7
SN01691-P	46,XY Female	FGFR2	SN00441-P	46,XY Male, Micropenis	
DSDEX36	46,XY Female	AR	DSDEX131	46,XY Male, Micropenis	
DSDEX72	46,XY Female	NR5A1	DSDEX29	46,XY Micropenis/Cryptochidism	
SN01607-P	46,XY Female	NR5A1	DSDEX127	46,XY Male, Micropenis, Hypospadias	
SN00366-P	46,XY Female	NRP1	DSDEX124	46,Male, Hypospadias/Cryptorchidism	
SN00212-P	46,XY Female		SN02417-P	46,XY Male, Hypospadias	
DSDEX37	46,XY Female		SN00134-P	46,XY Male, Hypospadias	
DSDEX38	46,XY Female		DSDEX108	46,XY Male, Hypospadias	
DSDEX51	46,XY Female		DSDEX132	46,XY Male, Hypospadias	
DSDEX23	46,XY Female		DSDEX120	46,XY Male, Anorchia	
DSDEX61	46,XY Female				
DSDEX125	46,XY Female				
SN02421-P	46,XY Female				

Green – pathogenic variant ~**22%** cases; Orange – likely pathogenic variant ~**13**% cases;
Yellow – possible genetic diagnosis ~**10%** cases; Red – no genetic diagnosis in >**50%** of cases.

Table 2-5: DSD patients with 46,XX karyotype with corresponding exome result.

Patient ID	Diagnosis	Diagnosis
DSDEX105	46,XX Male	
DSDEX115	46,XX Male	NR5A1
DSDEX08	46,XX Male	
DSDEX09	46,XX Male	
DSDEX11	46,XX Male	
DSDEX14	46,XX Male	
DSDEX55	46,XX Male	
DSDEX56	46,XX Male	
DSDEX10	46,XX Male	
DSDEX81	46,XX Male	
DSDEX13	46,XX Male	
DSDEX26	46,XX Male	
SN00048-P	46,XX Male	
DSDEX06	46,XX Male	
DSDEX76	46,XX Male	
DSDEX41	45X, Male	
DSDEX12	46,XX Ambiguous Genitalia	
DSDEX6	46,XX Ambiguous Genitalia	
DSDEX78	46,XX Ambiguous Genitalia	
DSDEX18	46,XX Ambiguous Genitalia	
DSDEX07	46,XX Ambiguous Genitalia	
DSDEX34_1	46,XX Ambiguous Genitalia	
DSDEX30	46,XX Female MRKH	
DSDEX03	46,XX Female MRKH	
DSDEX85	46,XX Female POF	
DSDEX91	46,XX Female POF	FSHR
DSDEX46?	46,XX Female	
DSDEX48	46,XX TH	
DSDEX101	46,XX Female GD	

Figure 2-1: Classification of variants identified by exome sequencing into categories.

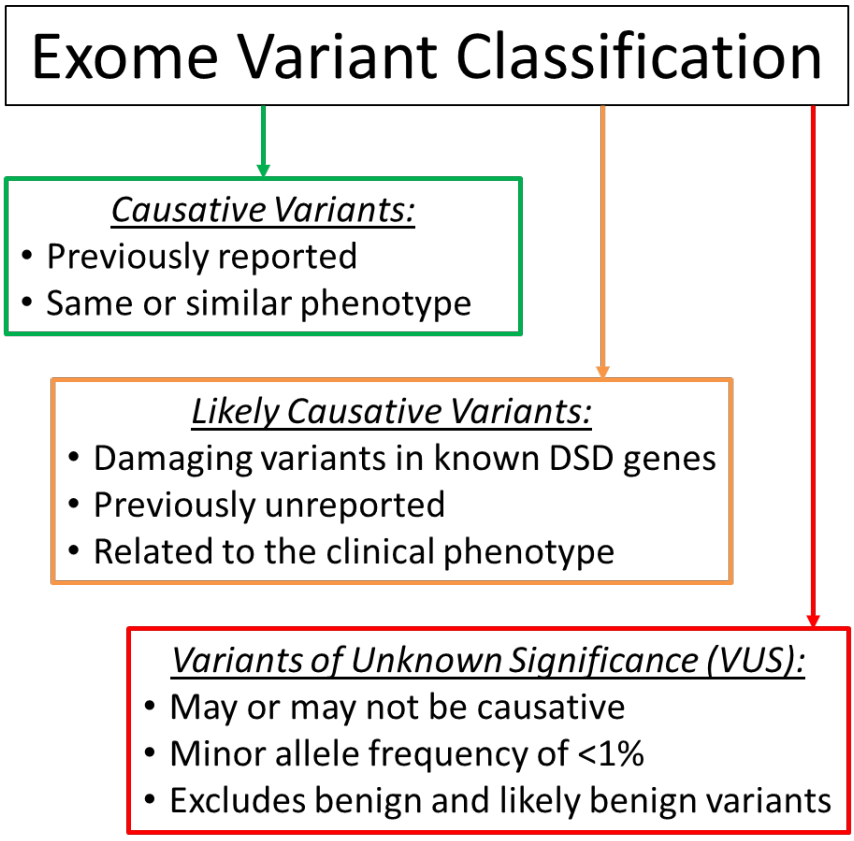
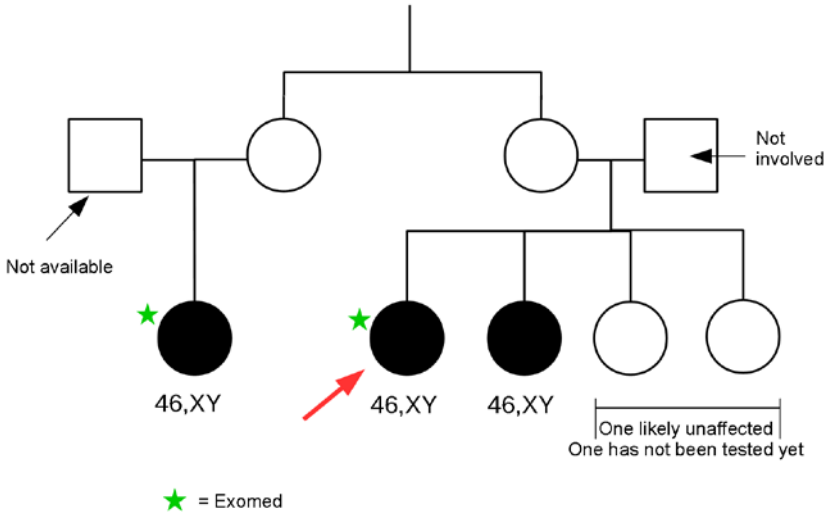


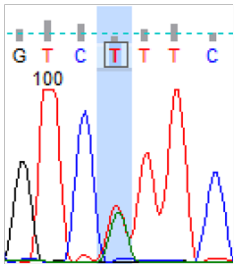
Figure 2-2: Pedigree of a familial case of 46,XY complete gonadal dysgenesis

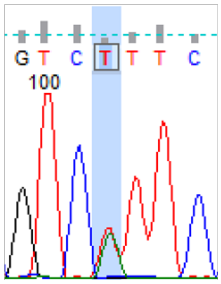


Arrow indicates the proband.

Figure 2-3: Sanger sequencing of family members showed that all affected individuals with 46,XY karyotype were heterozygous for the identified mutation.

Affected	Unaffected
46,XY Females, proband, cousin, sister	**46,XX Female** Aunt





References

1. Lee, P.A., et al., *Consensus statement on management of intersex disorders. International Consensus Conference on Intersex.* Pediatrics, 2006. **118**(2): p. e488-500.
2. Vilain, E., et al., *We used to call them hermaphrodites.* Genet Med, 2007. **9**(2): p. 65-6.
3. Lux, A., et al., *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, 2009. **9**: p. 110.
4. Stein, M.T., et al., *A newborn infant with a disorder of sexual differentiation.* J Dev Behav Pediatr, 2003. **24**(2): p. 115-9.
5. Warne, G.L., *Long-term outcome of disorders of sex development.* Sex Dev, 2008. **2**(4-5): p. 268-77.
6. Fleming, A. and E. Vilain, *The endless quest for sex determination genes.* Clin Genet, 2005. **67**(1): p. 15-25.
7. Ohnesorg, T., E. Vilain, and A.H. Sinclair, *The Genetics of Disorders of Sex Development in Humans.* Sex Dev, 2014.
8. Bashamboo, A. and K. McElreavey, *Gene mutations associated with anomalies of human gonad formation.* Sex Dev, 2013. **7**(1-3): p. 126-46.
9. Baxter, R.M. and E. Vilain, *Translational genetics for diagnosis of human disorders of sex development.* Annu Rev Genomics Hum Genet, 2013. **14**: p. 371-92.

10. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform.* Bioinformatics, 2010. **26**(5): p. 589-95.
11. Yourshaw, M., et al., *Rich annotation of DNA sequencing variants by leveraging the Ensembl Variant Effect Predictor with plugins.* Brief Bioinform, 2014.
12. Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.* Nat Protoc, 2009. **4**(7): p. 1073-81.
13. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.
14. Gonzalez, S.A., *Consensus interferon: tailored therapy and the impact of adherence.* Dig Dis Sci, 2011. **56**(3): p. 631-4.
15. Genomes Project, C., et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
16. Rehm, H.L., et al., *ACMG clinical laboratory standards for next-generation sequencing.* Genet Med, 2013. **15**(9): p. 733-47.
17. Kohler, B., et al., *Androgen insensitivity syndrome: somatic mosaicism of the androgen receptor in seven families and consequences for sex assignment and genetic counseling.* J Clin Endocrinol Metab, 2005. **90**(1): p. 106-11.
18. 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.* Am J Hum Genet, 2010. **87**(6): p. 898-904.

19. Bashamboo, A., et al., *A recurrent p.Arg92Trp variant in steroidogenic factor-1 (NR5A1) can act as a molecular switch in human sex development.* Hum Mol Genet, 2016.

20. Baetens, D., et al., *NR5A1 is a novel disease gene for 46,XX testicular and ovotesticular disorders of sex development.* Genet Med, 2016.

21. Igarashi, M., et al., *Identical NR5A1 Missense Mutations in Two Unrelated 46,XX Individuals with Testicular Tissues.* Hum Mutat, 2017. **38**(1): p. 39-42.

22. Swartz, J.M., et al., *A 46,XX Ovotesticular Disorder of Sex Development Likely Caused by a Steroidogenic Factor-1 (NR5A1) Variant.* Horm Res Paediatr, 2016.

Chapter 3

Identification of Novel Candidate Genes for 46,XY Disorders of Sex Development (DSD) using C57BL/6J-YPOS Mouse Model

Abstract

Disorders of Sex Development (DSD) have an estimated frequency of 0.5-1% of live births encompassing a variety of urogenital abnormalities ranging from mild hypospadias to complete sex reversal. In order to identify the underlying genetic etiology in a subset of DSD cases with 46,XY karyotype, we performed exome sequencing and were able to identify the causative genetic variant in 35% of cases. While the genetic etiology was not ascertained in more than half of the cases, a large number of variants of unknown clinical significance (VUS) were identified in those exomes. To investigate the relevance of these VUS in regards to the patient's phenotype, we utilized a mouse model in which the presence of a Y chromosome from the poschiavinus strain (YPOS) on a C57BL/6J (B6) background results in XY undervirilization and sex reversal, a phenotype characteristic to 46,XY DSD cases. We hypothesized that abnormal gonadal expression of specific genes in B6-YPOS males during embryonic development would correlate with genes in which VUS were identified in the exomes of 46,XY DSD patients. We isolated gonadal tissue from wild type (WT) B6 and undervirilized B6-YPOS males at embryonic day 11.5 and performed RNA sequencing to assess differential gene expression. The list of differentially expressed genes from the mouse model was filtered using the list of human genes containing VUS in 46,XY DSD cases to identify overlap. This identified 15 novel candidate genes with mutations in 46,XY DSD cases that may be associated with disease pathogenesis. We also show that all of the candidate genes were under direct regulation of the well-known sex determination gene Sox9.

Introduction

Human sex development is dictated by the inheritance of either an X or Y chromosome from parents to offspring. The male sex determination starts with the expression of a transcription factor *SRY* on Y chromosome in bipotential gonads initiating testicular organogenesis [1]. In the absence of the Y chromosome, female specific pathways are initiated for proper ovarian development. Secretion of testicular or ovarian hormones further differentiate body into typical male or female structures, including both internal and external genitalia. Anomalies in hormonal exposure and/or gene mutations disrupting sex development pathways lead to Disorders of Sex Development (DSD), defined as “congenital conditions in which development of chromosomal, gonadal, or anatomic sex is atypical” [2]. The umbrella term DSD encompasses conditions ranging from mild hypospadias (abnormal location of the meatus) to discrepancy between sex chromosomes and external genital phenotype (formerly known as sex reversal, either complete or with ambiguous genitalia). DSDs are estimated to affect 0.5-1% of the population. The birth of a child with a DSD is potentially quite stressful for families, bringing uncertainty in regards to future psychosexual development and clinical management [2-4].

At present a specific molecular diagnosis is identified at variable rates in different DSD subtypes. The majority (80-90%) of isolated 46,XX testicular DSD are explained by *SRY* translocations [5]. More recently a single nucleotide variant in *NR5A1* gene resulting in p.Arg92Trp amino acid change has been associated with 46,XX testicular (and ovotesticular) DSD [6, 7]. Approximately 50% of ovotesticular DSD are explained by *SRY* translocations, sex chromosome mosaicism and partial translocations [8].

Among 46,XY DSD cases 15% are due to *SRY* defects, 13% due to *NR5A1* defects and several due to other rare mutations (*SOX9, NR0B1, FGFR,…*). More recent data suggests that an additional 10-18% of 46,XY DSD cases may be explained by mutations in *MAP3K1* gene. Nevertheless, collectively the genetic etiology is still not identified in greater than 50% of DSD patients, suggesting the existence of a number of unknown sex determining genes.

Next-generation sequencing has now become instrumental in DSD diagnosis, including clinical exome sequencing and gene panels [9-11], with high diagnostic rates reported for known DSD genes. In a cohort of 46,XY DSD patients, we were able to establish a diagnosis in approximately 1/3 of cases [12], similar to rates for other rare disorders [13, 14]. Another 15% of the exomes in the cohort contained variants in known DSD genes that couldn't be validated but were reported to the referring clinicians to orient further endocrine or imaging testing toward a definitive diagnosis. Half of the cases from our cohort remained undiagnosed but contained hundreds of variants of unknown significance (VUS) that provide opportunity for research into new etiology of DSD. Here, we utilize an animal model resembling the phenotype of human 46,XY DSD patients [15, 16] to identify the genetic diagnosis in exome-negative cases (no pathogenic variant identified). We present a novel method that allows for identification of candidate genes involved in 46,XY DSD pathogenesis.

Results

Exome-negative 46,XY DSD cases

As previously described, we have performed exome sequencing on a cohort of 40 individuals diagnosed with 46,XY DSD [12]. To identify the disease-causing mutations a DSD-specific gene list, published elsewhere [9], was used for variant filtration. Exome sequencing was not able to identify the genetic diagnosis in 50% of cases. To address this issue, we compiled a cohort of 32 exome-negative DSD cases with 46,XY karyotype for further investigation (Table 3-1) (this new cohort includes 21 unresolved cases from [12] and additional 11 exome-negative cases enrolled since). As evident from Table 3-1, the range of associated clinical features was wide, which is a typical characteristic of DSD presentation. Patients could be grouped into four categories based on the appearance of the external genitalia and gonadal development: 46,XY women with complete gonadal dysgenesis (CGD), when gonadal phenotype had been ascertained by the clinical team; 46,XY Females; 46,XY with ambiguous genitalia (and unknown sex of rearing at the time of enrollment); 46,XY males, with hypogonadism.

C57BL/6J-*Y poschiavinus* mice as a model for 46,XY DSD

In exome-negative cases, where no pathogenic variant was found, we identified many variants of unknown significance outside of the DSD clinical gene list. To investigate the relevance of these VUS in regards to patients' phenotype, we utilized a powerful mouse model used for studying undervirilization in human 46,XY individuals. In this model, the presence of a Y chromosome originating from a *M. domesticus poschiavinus* strain (Y^{POS}) on a C57BL/6J (B6) background (B6-Y^{POS}), an inbred

laboratory strain that normally carries a *M. musculus* Y chromosome, results in disrupted testicular development and sex reversal [17].

This inherited phenomenon has been extensively studied in the B6-Y^{POS} males. The failure to develop testes stems from the inability of the Sry^{POS} gene to initiate normal testicular development when B6 autosomal and/or X-linked factors are present. Virtually all B6-Y^{POS} animals develop some ovarian tissue; half develop exclusively ovarian tissue, classified as completely sex-reversed, and the remainder develop both ovarian and testicular tissue, classified as partially sex-reversed (gonad morphology shown in Figure 3-1A). We have previously identified a 1.5-Mb congenic region on chromosome 11 that confers 80% protection from B6-Y^{POS} sex reversal in the heterozygous state (B6-110h-Y^{POS}) and complete protection in the homozygous state (B6-110H-Y^{POS}) [18]. This protective region allows for continual maintenance of subfertile *Poschiavinus* male mice as a breeding colony, with an option of generating unprotected B6-Y^{POS} males by mating heterozygous B6-110h-Y^{POS} males with wild type (WT) B6 females.

In the embryonic mouse gonad, *Sry* is normally expressed in a dynamic wave (central to distal) between E10.5 and E12.5 in the XY genital ridge with peak *Sry* expression occurring in normal XY B6 genital ridges at ~E11.5, i.e., at the 16-18 tail somites stage of development [19]. In contrast, expression of the Sry^{POS} gene peaks 10 to 14 hours later in the genital ridges of B6-Y^{POS} fetuses [20]. We hypothesized that abnormal gonadal expression of specific genes in B6-Y^{POS} males, after the surge of *Sry* during gonadal development, would correlate with VUS in genes of 46,XY DSD patients identified through exome sequencing.

Gene expression differences between B6-Y^{B6} and of B6-Y^{POS} males

Since all of the 46,XY DSD patients in the exome-negative cohort carried a functional *SRY* gene it was important to perform gene expression analysis in the animal model after the peak of *Sry* expression for optimal comparability. To achieve this, gonadal tissue from WT B6-Y^{B6} and undervirilized B6-Y^{POS} males at embryonic day E11.5, specifically at 21 tail somites (ts) (a time point when the surge of *Sry* gene was complete in both B6-Y^{B6}/B6-Y^{POS} males) was collected to perform RNA sequencing for assessment of differential gene expression. To obtain sufficient RNA for sequencing and to minimize the impact of gonad-to-gonad variation, dissected embryonic gonads of identical genotypes were pooled together for a single RNA extraction and subsequent sequencing.

Using this method, we identified 515 genes that were differentially expressed between B6-Y^{B6}/B6-Y^{POS} males with a fold change greater than 1.5. Out of these 515 genes, 308 were underexpressed and 207 were overexpressed in B6-Y^{POS} males (Figure 3-2) (Supplemental Table 3-S1). To validate the integrity of the method, we verified which of the genes in our DSD gene list used for exome variant filtering were present in the B6-Y^{B6}/B6-Y^{POS} differentially expressed gene list. The comparison of the two lists revealed that 21 genes were in common - 15 underexpressed and 6 overexpressed (Supplemental Table 3-S1). In our previous cohort [12], out of these 21 genes, 3 (*HSD17B3, STAR, FGFR2)* contained a previously identified pathogenic variant by exome sequencing explaining a total of 5 cases and 2 (*DHH, MAMLD1)* contained a convincing VUS. Cumulatively, these findings indicate the ability of the method to identify genes important in sex development.

Filtering of VUS in 46,XY DSD cases using the B6-Y^{POS} gene list

On average exome sequencing identifies ~ 21,000 variants per single case [13]. Since DSDs are rare conditions, all variants identified in exome with a minor allele frequency (MAF) of more than 1% in the population were excluded. The variants remaining after the MAF cutoff, were classified as variants of unknown significance. The gene list generated via expression studies in B6-Y^{B6}/B6-Y^{POS} mice, consisting of 515 genes, was filtered against the list of VUS-containing human genes identified by exome sequencing. The comparison of two lists identified 305 (189 underexpressed and 116 overexpressed in B6-Y^{POS}) genes that were both differentially expressed in B6-Y^{POS} males and contained a VUS in our exome-negative 46,XY DSD cohort (Figure 3-2).

All these genes are known to be expressed in the developing gonad at the time of sex determination. In order to increase the probability of identifying relevant candidate genes involved in 46,XY DSD pathogenesis, we queried if the genes from the mouse model were involved in any known biological processes. Gene Ontology Consortium (GOC) [21] enrichment analysis confirmed that genes underexpressed in B6-Y^{POS} males were indeed enriched in biological processes known to control multicellular organism and anatomical structure development, including male reproductive development (Table 3-2). Understanding the relevance of the genes that were overexpressed in B6-Y^{POS} males was less straightforward. Those were enriched in only two biological processes: response to extracellular stimulus and epithelial cell differentiation. Both of these categories had a high P-value indicating that many genes in the overexpressed category are not associated with any known biological processes

at this time. In addition, all of the pathogenic variants identified in our previous 46,XY DSD cohort [12] were in the underexpressed category of genes, indicating that they need to be expressed at higher levels in the developing gonad for proper testicular formation.

Based on these findings, we decided to focus on variants identified in the 189 genes underexpressed in B6-Y^{POS} males whose higher expression in WT males correlates with normal male sex development. To choose a fold change cutoff, we looked at fold change differences in expression between B6-Y^{B6} and B6-Y^{POS} males for genes present in our primary clinical gene list (Supplemental Figure 3-S1). We found that more genes have an expression that is 2 fold higher in WT males compared to B6-Y^{POS} males. To make the analysis more stringent and increase the confidence of identifying true candidate genes involved in male sex development we therefore increased the fold change cutoff in expression between the B6-Y^{B6} and B6-Y^{POS} males from 1.5 to 2. This change decreased the number of genes from 189 to 53. Additional, filtering was performed based on variant frequency, amino acid conservation, number of variants, *in silico* predictions for pathogenicity, availability of the literature, and gonadal cell-specific expressivity using GenitoUrinary Developmental Molecular Anatomy Project data (GUDMAP) [22].

Using the above mentioned filtering criteria, we identified 15 novel candidate sex developmental genes variants that may be involved in 46,XY DSD pathogenesis. The list of VUS identified in the 46,XY cohort is shown in Table 3-3. The relative expressions of these genes in B6-Y^{B6}, B6-Y^{POS} and WT female are shown in Figure 3-3. The

expression changes of all 15 genes were confirmed by quantitative real time PCR (Supplemental Figure 3-S2).

Novel candidate genes are under *Sox9* regulation

Although the time point chosen for our gene expression analysis was such that the *Sry* gene expression was similar between in B6-Y^{B6} and B6-Y^{POS} males, the downstream target of *Sry, Sox9,* was significantly higher in WT B6-Y^{B6} males (Supplemental Figure 3-S3). In order to identify if any of our candidate genes fell under the direct regulation of *Sox9,* we used a mouse model where *Sox9* expression is suppressed in Sertoli cells using *Amh-Cre Sox9flox/flox* mice [23]. By E13.5, *Sox9* protein is completely absent (Figure 3-1B) and these mice show postnatal fertility defects. This model allows the examination of *Sox9* loss in an intact Sertoli cell environment. Earlier *Sox9* knock out models result in complete sex reversal (XY ovaries) [24, 25] or embryonic lethality [26]; neither situation sheds light on *Sox9* target genes during sex determination.

The profiles of gonadal gene expression from GUDMAP reveal that almost all cases, the patterns of gene expression are similar to bona fide target genes such as *Amh* and *Ptgds* [27, 28]. The target gene expression is higher in the male supporting cells (Sertoli) than in the female supporting cells (granulosa) (Supplementary figure 3-S4).

We identified that all 15 genes were also underexpressed in mice with suppressed *Sox9* expression, suggesting that the novel candidate genes for 46,XY DSD are downstream targets of *Sox9, in the sex-determining pathway*. This in turn strengthens our findings that mutations in these candidate genes such as the ones

we've identified in our 46,XY DSD cases (Table 3-3) may be responsible for the patient's phenotype.

Discussion

The use of the undervirilized B6-Y^{POS} mice as a model for 46,XY DSD in humans provides valuable screening information towards the identification of novel genes involved in male sex development, mutations in which could lead to anomalies of genital development in 46,XY patients with DSD. All these genes are expressed in the developing mouse gonad at the relevant time for sex determination and are all under the control of major male sex determination *Sox9*.

Mutations in the novel candidate genes identified via the Y^{POS} mouse model are likely to be causative. For example, one of the candidate genes *Adamts16* (A Disintegrin And Metalloproteinase with Thrombospondin Type 1 Motif, 16) has been shown to be co-expressed with the known DSD gene *Wt1* in embryonic gonads, adult testes and spermatids [17]. Moreover, targeted disruption of *Adamts16* in rats results in cryptorchidism and sterility [29]. In our exome-negative DSD cohort, we identified three heterozygous variants in this gene (Table 3-3). Patients DSDEX53 and DSDEX02, both 46,XY females with complete gonadal dysgenesis, had a missense variant leading to amino acid change at positions p.Val734Ile and p.Arg100Trp respectively. These changes were located in propeptide and Cysteine-rich domains of the *Adamts16* protein and may prevent expression or proper folding of the protein. The third missense variant (p.Phe469Val) in patient DSDEX43 (46,XY, with ambiguous genitalia) was located in

the peptidase domain of the protein and predicted damaging by *in silico* tools suggesting a possible impairment of the enzymatic function of Adamts16.

We have also identified two rare variants (p.Ala496Thr; p.Ala1202Gly) in the *FBLN2* gene in two exome-negative cases with different phenotypes: 46,XY Female with inguinal testes/enlarged clitoris and 46,XY Male with hypospadias. Additional rare *FBLN2* variants were present in six other unrelated cases with previously identified genetic diagnosis (with a pathogenic variants identified in known DSD genes).This suggest that variants in *FBLN2* are overrepresented in DSD population and may act as modifiers of the phenotype. We show that the *Fbln2* protein is expressed in a sexually dimorphic pattern in the developing gonad by immunohistochemical staining (Supplemental Figure 3-S5): at E12.5 WT B6 females have virtually no expression in the developing ovaries, whereas WT B6 males have very high expression in developing testes suggesting an important role of *Fbln2* in gonad development. *FBLN2* has been proposed as a candidate gene for 46, XY DSD in an unpublished meeting abstract (K. MacElreavey, personal communication).

We identified a single variant, predicted to be damaging by *in silico* tools, in the *SPRY4* gene in a 46,XY male patient (SN00441-P) with hypogonadism. *SPRY4* is one of the genes in which variants have been found in a cohort of patients presenting with Hypogonadotropic Hypogonadism with or without anosmia (HH17, OMIM #615266) [30]. These genes are believed to be functioning in an oligogenic model, with variants in several genes possibly needed for phenotypic expression. Variants in *SPRY4* have been found in association with variants in *FGFR1* (HH2, OMIM #147950) and *DUSP6* (HH19, OMIM #615269). We didn't identify *FGFR1* or *DUSP6* variants in the exome of

patient SN00441-P (which would have been diagnostic for this patient). However *DUSP6* is present in the differentially expressed gene list (underexpressed in B6-Y^{POS} with a fold-change of 1.7) and another gene coding for a dual-specificity phosphatase, DUSP15, is in our final candidate gene list, with underexpression in YPOS (fold change >2) and a VUS in one patient.

Conclusion:

Exome sequencing provides high throughput genetic diagnostic capability that has become the core of modern clinical genetics. However, many variants identified by whole exome sequencing are uninterpretable clinically. The C57/BL6J-YPos model allows to improve the interpretive gap by correlating human sequence variants and RNA sequencing variations. This approach allowed the identification of 15 novel candidate genes for human DSD.

Materials and Methods

Exome sequencing and analysis

Exome sequencing (ES) covers approximately 95% of the RefSeq coding regions of the genome, including +/-2bp on each side of each exon, and thus identifies the majority of variants in protein-coding regions, which currently harbor 80-90% of known disease-causing variants [31].

DNA was isolated from peripheral blood using Gentra Puregene Blood Kit (Qiagen, Germantown, MD, USA) or saliva collected using ORAgene Discover ORG-500 (DNAgenoteck Ottawa, ON, Canada). Sequencing libraries and exome capture was

done for each sample following manufacturer's protocols for SureSelect All Exon 50 Mb capture kit (Agilent Technologies) and Nextera Rapid Capture (Illumina, San Diego, CA USA). Sequencing was performed on an Illumina HiSeq 2000 or HiSeq2500 as 50bp or 100bp paired-end run at the UCLA Clinical Genomics Center.

The sequence reads FASTQ files were aligned to the human reference genome (GRCh37/hg19 Feb. 2009 assembly) using BWA (Burrows-Wheeler Alignment Tool) [32] . The output BAM file were sorted, merged, and PCR duplicates removed using Picard. INDEL (insertions and deletions) realignment and recalibration was performed using Genome Analysis Tool Kit (GATK) from the Broad Institute. Both single-nucleotide variants (SNVs) and small INDELs were called within the Ensembl coding exonic intervals +/-2bp using GATK's Unified Genotyper and recalibrated and filtered using GATK variant-quality score recalibration and variant filtration tools. All high-quality variants were annotated using VarSeq Variant Annotation, Filtering and Interpretation Software (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). All variants were filtered by minor allele frequency (MAF) of <1% and intersected with the DSD gene list to identify mutations in known DSD genes. The list is comprised of a primary gene list of well annotated genes involved in sex determination and differentiation [9], as well as a secondary list of genes that are more loosely associated with sex development: *e.g.* their OMIM (Online Mendelian Inheritance of Man) description contains sex development key words or data is available only from animal models.

The variants identified by exome sequencing were classified into three categories: causative, likely causative and variants of unknown significance (VUS) following the recommendations of the American College of Medical Genetics [33]. To

assess previously unreported missense variants we used two *in silico* algorithms SIFT [34] and PolyPhen [35]. These algorithms predict the likelihood that a given missense variant is pathogenic, based on conservation of the amino acid across species, the physical characteristics of the altered amino acid and the possible impact on protein structure and function. All variants with low quality scores were validated by Sanger sequencing.

Animal care and dissections

All animals were housed at the UCLA Animal Care Facility following the guidelines of the University of California, Los Angeles (UCLA) Division of Laboratory Animal Medicine. All experiments were approved by the Institutional Animal Care and Research Committees of UCLA. Wild type C57BL/6J males and females used for breeding and experiments were purchased from the Jackson Laboratory (Bar Harbor, ME, USA) that is fully accredited by the American Association for Accreditation of Laboratory Animal Care.

Congenic mice protected from sex reversal were used to generate B6-Y^{POS} males [15]. Overnight mating was performed using either wild type (WT) B6 or B6-110h-Y^{POS} (protected from sex reversal) males and three 8-week-old WT B6 female mice each. The morning after the overnight mating the WT B6 females were checked for presence of a vaginal plug and separated into a new cage. At E11.5 of gestation, pregnant females were sacrificed. Embryos were removed from and placed in PBS for the duration of dissections that were performed under light microscope. Both right and left embryonic gonads were separated from the mesonephros and placed in RNA

stabilizing solution (RNA*later*, Ambion, Life Technologies). DNA was extracted from the rest of the embryos' bodies for genotyping PCRs. Chromosomal sex was determined using a single primer pair that detects the X-linked Smc-x gene (330 bp) and the Y-linked Smc-y gene (301 bp) (Forward: 5´CCGCTGCCAAATTCTTTGG3´; Reverse: 5´TGAAGCTTTTGGCTTTGAG3´). The presence of the Y^{POS} chromosome was determined by a single nucleotide polymorphism (SNP) between Y^{B6} and Y^{POS} Sry gene using the primer sets 5´TGAATGCATTTATGGTGTGGTC3´; 5´AGCTTTGCTGGTTTTTGGAGTA3´. Immomix Red 2x (Bioline, London) was used for PCR amplification following manufacturer's guidelines with annealing temperature at 60˚C and resolved by electrophoresis on a 2% agarose gel. Presence or absence of the 110h protective region in B6-Y^{POS} males was checked by Sanger sequencing of two regions 11-10 and 11-11 containing SNP rs27019103 (5´AAAGTGTGCTTCCCAGGAGA3´; 5´CCTCTCCCTCAACCCCTAAG3´) and SNP rs28240850 (5´CCACAGCTGGAGGTAGGGTA3´; 5´CCTAAGATGCCATGGGAAGA3´) respectively [15].

Total RNA was isolated from combined embryonic gonadal tissue (50-70 gonads per group) using Qiagen RNeasy Plus Mini Kit (Qiagen) following manufacturer's guidelines. RNA quality was assessed by Agilent 2100 Bioanalyzer (Agilent Technologies) following standard protocols at the UCLA Genotyping and Sequencing Core. All samples were required to have RNA integrity scores (RIN) greater than 7.

RNA sequencing and expression analysis

Good quality RNA from each sample was submitted to UCLA Neuroscience Genomic Core (UNGC) for library preparation and sequencing. Library preparation was

performed using TruSeq Stranded Total RNA (Illumina) kit with Poly-A selection following manufactures guidelines. Libraries were multiplexed and sequenced across 2 lanes using HiSeq 2500 (Illumina) with 69 bp paired-end run on a rapid flowcell generating 150M reads per lane. Four samples were multiplexed and sequenced over two rapid lanes with each sample receiving approximately 75 million reads with >85% map rate.

The generated sequencing reads were aligned to the mouse genome, versions mm9 and mm10 with TopHat v2.0.2 [36] and STAR [37] aligner respectively. Transcript abundance was assessed by Cufflinks (v2.1.1) [38], using a GTF file based on Ensembl mouse NCBI37. Differential expression analysis was based on fold change differences greater than 1.5 between the groups being compared.

Differentially expressed genes were split into two categories: underexpressed and overexpressed in B6-Y^{POS} males. Both categories were separately subjected to pathway enrichment analysis using Gene Ontology Consortium [39] and PANTHER classification system [40].

Quantitative PCR

Reverse transcription of RNA to cDNA was performed using Tetro cDNA Synthesis Kit (Bioline, Taunton, MA USA) following manufacturer's protocol with 1µg of total RNA as template. The RNA samples used for validation were from the same batch of RNA used for RNA sequencing above. The primer sequences used are detailed in (Supplemental Table 3-S2). Primers were designed using autoprime software (www.autoprime.de) and spanned exon-exon junctions for optimal RNA quantification.

cDNA was quantified using QuBit HS (Invitrogen) for double stranded DNA and a total of 3ng of cDNA was used per sample for amplification. qPCR was carried out in duplicate utilizing Syber Green-based SensiMix SYBR No-Rox Kit (Bioline). Reaction conditions were as follows: 95ºC for 10 min, then 40 cycles of 95ºC for 15 sec, 60-64ºC (see Table S2), and 72ºC for 15 sec.

Immunohistochemistry

Fbln2 expression in embryonic gonads at E12.5 was assessed using immunohistochemistry (IHC) following the experimental design of Wilhelm et al. [41], using the *Fbln2* rabbit polyclonal, sc-30176 (Santa Cruz Biotechnology) antibody. Topro (Invitrogen) was used to counterstain nuclei. All images were taken on a Zeiss LSM 510 Meta confocal microscope.

Figures and Tables:

Table 3-1: Exome-negative cohort of 46,XY DSD cases

Patient ID	DSD Category	Clinical Features
RDSD002	46,XY Female, CGD	–
RDSD003	46,XY Female, PGD	No uterus, Fallopian tubes present, short vagina, very low T and undetectable Estradiol, gonads not found
RDSD006	46,XY Female	Amelia (missing limbs)
RDSD007	46,XY Female, GD	Adrenal rests
RDSD010	46,XY Female	Clitoromegaly
RDSD011	46,XY Female	Short stature
RDSD012	46,XY Female	Kidney disease, possible Denys-Drash syndrome
RDSD013	46,XY Female, CGD	Normal uterus and Fallopian tubes, streak gonads
RDSD017	46,XY Female, GD	–
RDSD018	46,XY Ambiguous Genitalia	Partial fusion of labioscrotal folds, small phallus, penoscrotal hypospadias
RDSD020	46,XY Ambiguous Genitalia	Developmental delay, agenesis of corpus callosum
RDSD021	46,XY Ambiguous Genitalia	Adrenal Hypoplasia Congenita, other dysmorphic features
RDSD022	46,XY Ambiguous Genitalia	Microcephaly, intestinal dysmotility, optic nerve hypoplasia
RDSD025	46,XY Male, Micropenis/Cryptochidism	Severe growth and developmental retardation, testes not seen by ultrasound
CDSD029	46,XY Male, Hypospadias	–
CDSD030	46,XY Female	Large clitoris; no uterus or vaginal opening; inguinal testes
CDSD031	46,XY Ambiguous Genitalia, CGD	Abdominal gonads with no oocytes, no seminiferous tubules; no clitoromegaly; posterior fusion of labia; urogenital sinus
CDSD032	46,XY Female	Inguinal testes w/ immature seminiferous tubules; no uterus or Fallopian tubes; deafness; impaired cognition
CDSD034	46,XY Ambiguous Genitalia	Undescended testes, bifid scrotum, hypospadias
CDSD036	46,XY Ambiguous Genitalia	Bilateral descended testes, midshaft hypospadias, chordee
CDSD039	46,XY Male, Micropenis	No uterus or ovaries per ultrasound, ambiguous genitalia, under virilization

RDSD041	46,XY Female	Complete Androgen Insensitivity Syndrome (CAIS)
RDSD042	46,XY Male, Hypospadias	–
RDSD043	46,XY Female, GD	–
RDSD044	46,XY Male, Anorchia	Congenital Bilateral Anorchia; responsive to testosterone; definite penis (mildly shortened); no hypospadias, fully formed scrotum
RDSD045	46,XY Male, Hypospadias/Cryptorchidism	Azoospermia, high T levels
RDSD046	46,XY Female	Multiple congenital anomalies; no uterus; internal gonads - testes
RDSD047	46,XY Male, Microphallus	Hypogonadism, hypospadias
RDSD048	46,XY Male, Micropenis	–
RDSD049	46,XY Male, Hypospadias	–
CDSD050	46,XY Male, Hypospadias	Chordee, bifid scrotum, cryptorchidism
CDSD051	46,XY Female	Growth delay, short stature

Abbreviations: CGD – complete gonadal dysgenesis; GD – gonadal dysgenesis.

Anatomical description follows the standardized nomenclature in Hennekam et al., 2013 [42], except when only historical description was available in patient's file.

Table 3-2: Biological Processes in which differentially expressed mouse gonad genes are involved

Gene Ontology Biological Process	Expression in B6-Y^{POS} vs. B6-Y^{B6} Males	# of genes in the GO Term Reference Gene List	# of genes Test Gene List	P-value
Single-multicellular organism process (GO:0044707)	Underexpressed	5417	151	4.49E-15
System development (GO:0048731)	Underexpressed	4042	125	1.45E-14
Multicellular organism development (GO:0007275)	Underexpressed	4640	136	1.61E-14
Anatomical structure development (GO:0048856)	Underexpressed	4986	139	4.32E-13
Multicellular organismal process (GO:0032501)	Underexpressed	6482	163	1.58E-12
Reproductive system development (GO:0061458)	Underexpressed	429	26	1.21E-05
Male sex differentiation (GO:0046661)	Underexpressed	161	15	1.86E-04
Reproduction (GO:0000003)	Underexpressed	1373	45	2.52E-03
Male gonad development (GO:0008584)	Underexpressed	139	12	1.05E-02
Response to extracellular stimulus (GO:0009991)	Overexpressed	488	18	3.08E-02
Epithelial cell differentiation (GO:0030855)	Overexpressed	503	18	4.65E-02

The list of 515 genes found to be differentially expressed between B6-Y^{POS} and WT male embryonic gonads was analyzed using the Gene Ontology Consortium functional annotation software. The categories of Gene Ontology (GO) biological processes are shown in column 1. P-value – probability of seeing the indicated number of genes from custom list (column 4) in GO term gene list (column 3), given the total number of annotated genes in the whole genome.

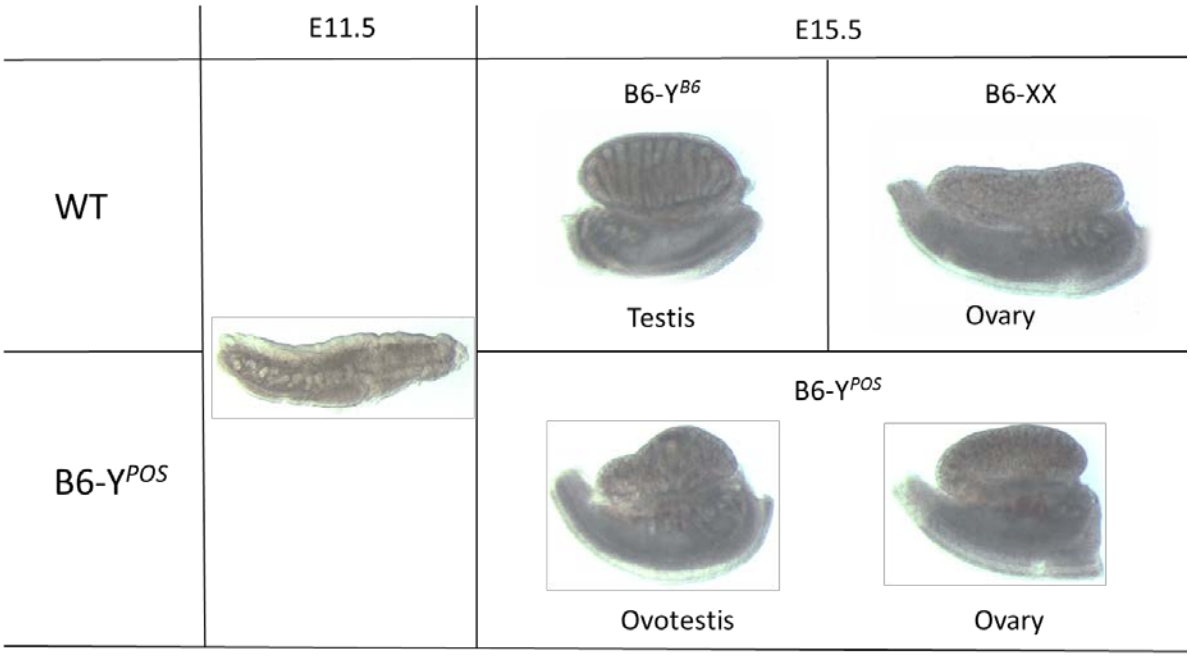
Table 3-3: List of VUS in candidate genes

Gene	DSD Case ID	Zygosity	HGVSc	HGVSp	MAF gnomAD (%)
TOX2	RDSD021	Het	c.319G>A	p.Gly107Ser	0
	CDSD036	Cmpd Het	c.448A>G	p.Met150Val	0
	CDSD036	Cmpd Het	c.1201C>G	p.Pro401Ala	0
	CDSD036	Cmpd Het	c.1122_1124dupGCC	p.Pro376dup	0
DUSP15	RDSD020	Het	c.563G>C	p.Arg188Pro	0.002
NKD2	RDSD003	Het	c.1151G>A	p.Arg384Gln	0.001
CNGA1	CDSD030	Het	c.1478G>A	p.Arg493Gln	0.09
	RDSD022	Het	c.398G>T	p.Gly133Val	0.03
PTK2B	RDSD011	Het	c.1799G>A	p.Arg600Gln	0.0008
ESPN	RDSD044	Het	c.2230G>A	p.Asp744Asn	0.02
SMOC2	CDSD030	Het	c.1276G>A	p.Val426Met	0.3
ADAMTS16	RDSD013	Het	c.2200G>A	p.Val734Ile	0.8
	RDSD002	Het	c.298C>T	p.Arg100Trp	0.1
	RDSD022	Het	c.1405T>G	p.Phe469Val	0.02
FBLN2	CDSD030	Het	c.1486G>A	p.Ala496Thr	0.033
	CDSD029	Het	c.3605C>G	p.Ala1202Gly	0.004
NIPAL1	RDSD003	Het	c.1207A>G	p.Thr403Ala	0.1
	CDSD031	Het	c.31G>A	p.Glu11Lys	0
CYP26B1	CDSD032	Het	c.805C>G	p.Leu269Val	0.008
SPRY4	CDSD039	Het	c.446C>G	p.Pro149Arg	0.0004
MYBL1	RDSD004	Het	c.754T>A	p.Phe252Ile	0.05
	CDSD029	Het	c.1832G>C	p.Ser611Thr	0.0008
	RDSD049	Het	c.936T>A	p.Asn312Lys	0.03
ETV4	RDSD006	Het	c.523C>A	p.His175Asn	0.1
LGR5	RDSD007	Het	c.1834G>A	p.Val612Met	0.004
	RDSD020	Het	c.2341C>G	p.Pro781Ala	0.8
	RDSD048	Het	c.2537C>A	p.Thr846Asn	0

Abbreviations: Het - heterozygous; Cmpd het – compound heterozygous; HGVSc – Human Genome Variation Society coding sequence location; HGVSp – Human Genome Variation Society protein sequence location; MAF – minor allele frequency; gnomAD – genome Aggregation Database.

Figure 3-1: A. Abnormal gonadal development in B6-Y^{POS} fetuses (E15.5)

The morphology of gonad development in mice shown at embryonic day E11.5 (middle: capable of giving rise to both testes and ovaries) and E15.5. Top left: B6-Y^{B6} – testicular development in WT male. Top right: B6-XX – ovarian development in WT female. Bottom two images show development of ovotestis (left) and ovary (right) in B6-Y^{POS} males.



B. Immuno fluorescence of wild type and Sox9 knockout gonad at E13.5

At E13.5 Sox9 protein is lost from the testis cords (white arrows) in the *Amh-Cre, Sox9floxflox* mice yet the testis cords remain intact, as shown by the laminin stain. Sox9/Laminin is shown in green and Dapi is shown in blue. Scale bar represents 100um

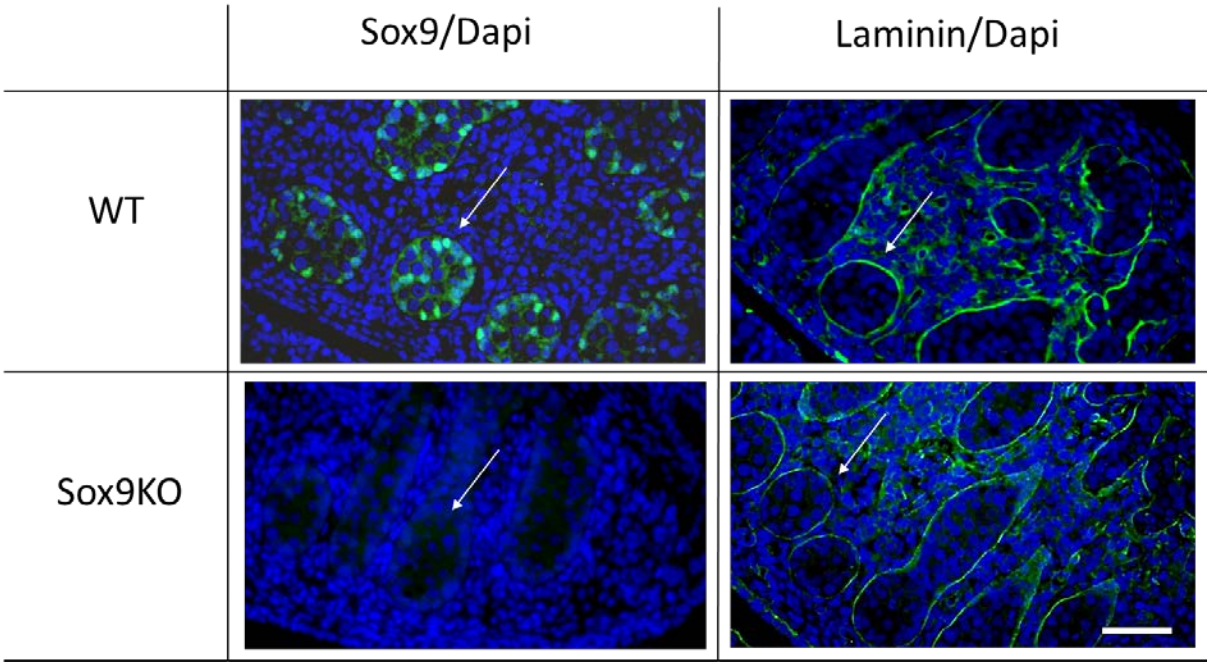


Figure 3-2: 515 Genes are differentially expressed in B6-Y^{B6} vs B6-Y^{POS} gonads

Pie chart representing the number of differentially expressed genes between WT B6-Y^{B6} and undervirilized B6-Y^{POS} males. A total of 308 and 207 genes were underexpressed (red) and overexpressed (green) in B6-Y^{POS} males, respectively. VUS were identified in 189 and 116 of the human syntenic genes in underexpressed and overexpressed categories respectively.

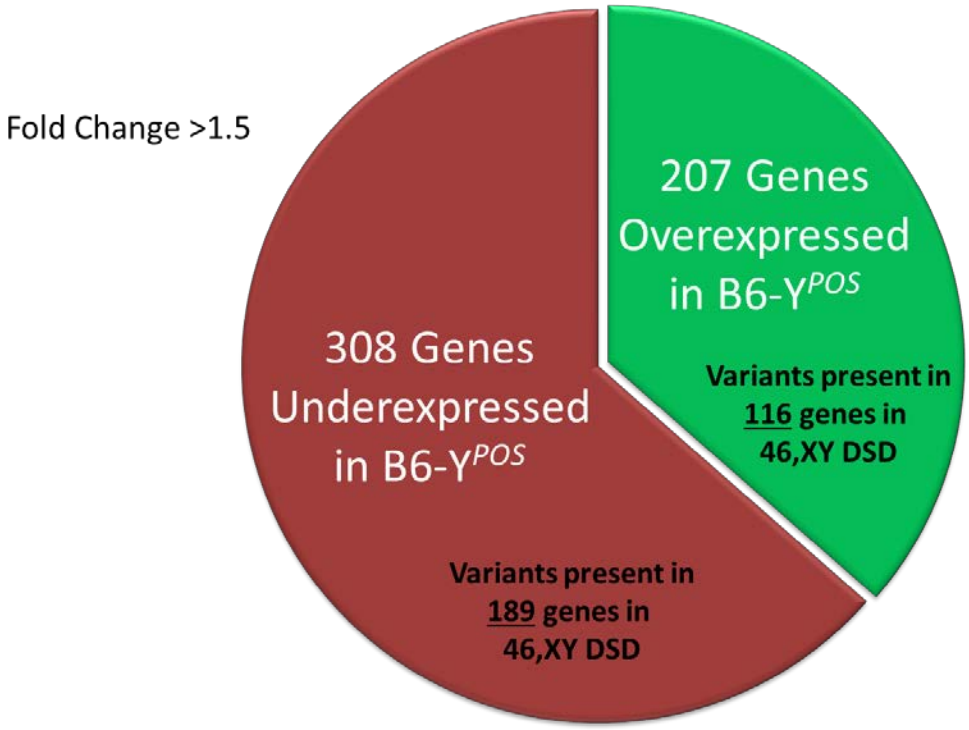
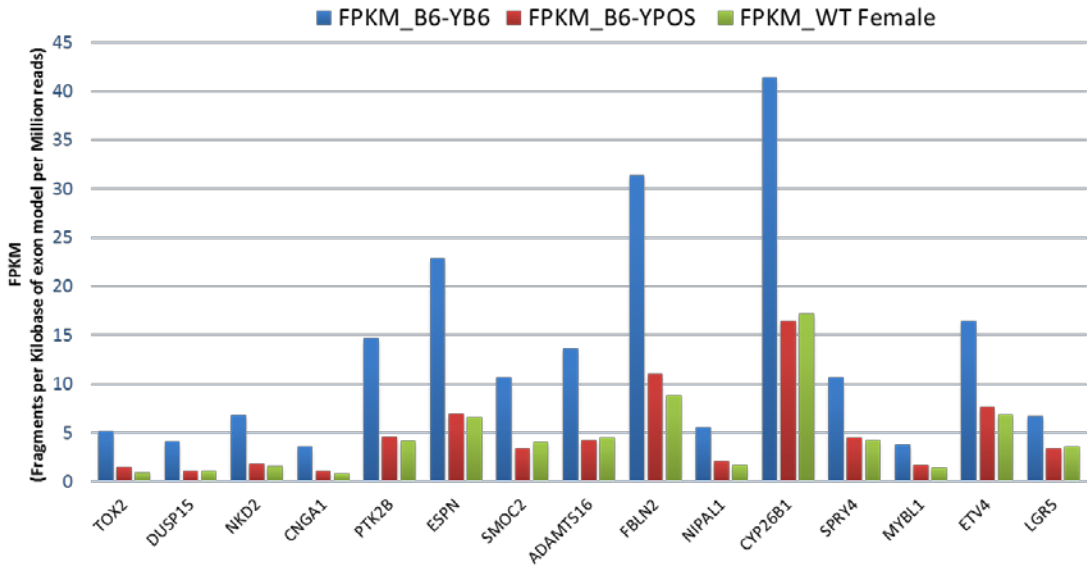


Figure 3-3: Relative expression of novel candidate genes in the B6-Y^{POS} mouse model Gene expression differences in candidate genes between B6-Y^{B6} (blue) males, B6-Y^{POS} (red) males, and WT B6 females (green). The expression values are shown in FPKM values (fragments per kilobase of exon model per million reads).



Supplemental Table 3-S1: Genes differentially expressed between B6-Y^{B6} and B6-Y^{POS} males

List of all differentially expressed genes either underexpressed or overexpressed in B6-Y^{POS} males is shown. Additionally, the table contains DSD gene list used for exome variant filtration. Genes in common between the two list are also present.

Differentially expressed genes	Fold Change (WTB6 Male/Ypos Male)	Expression in B6-Ypos	DSD Gene list used to filter exome variants	Genes in common
MIR6516	Not Expressed in Ypos Males	Underexpressed	ACP5	AMH
SNORA28	Not Expressed in Ypos Males	Underexpressed	AHR	DHH
SNORA44	Not Expressed in Ypos Males	Underexpressed	AKR1C2	SOX9
SNORA47	Not Expressed in Ypos Males	Underexpressed	AKR1C4	SOX10
SNORA7A	Not Expressed in Ypos Males	Underexpressed	AKR1D1	FGF9
SNORD15B	Not Expressed in Ypos Males	Underexpressed	ALMS1	PROKR2
CERS1	1266.999808	Underexpressed	AMH	SLC6A4
AMD1	1156.644062	Underexpressed	AMHR2	COL2A1
AMH	41.14940201	Underexpressed	AR	HSD17B3
NR4A3	25.34241571	Underexpressed	ARL13B	STAR
MT3	21.34265152	Underexpressed	ARL6	MAMLD1
DHH	12.92685038	Underexpressed	ARX	NR4A1
PTGDS	11.72633858	Underexpressed	ATF3	TCF21
SIGLEC12	11.34107437	Underexpressed	ATM	HSD11B2
TESC	10.18455429	Underexpressed	ATRX	AUTS2
COL9A3	9.673400875	Underexpressed	AUTS2	FOXL2
KIF19	9.120920405	Underexpressed	AZF1	TPH2
MMD2	8.62999358	Underexpressed	BBS1	CYP11A1
GJB1	7.303379673	Underexpressed	BBS10	WNT4
RNASE1	6.981688543	Underexpressed	BBS12	RSPO1
GDNF	6.678281727	Underexpressed	BBS2	FGFR2
SOX9	6.124617642	Underexpressed	BBS4	
ISLR2	5.981043326	Underexpressed	BBS5	
CBLN4	5.972338095	Underexpressed	BBS7	
SOX10	5.95121736	Underexpressed	BBS9	
ANKRD63	5.830259494	Underexpressed	BKMA1	
DKK4	5.287193208	Underexpressed	BMP4	
FGF9	5.197581132	Underexpressed	BMP7	
PROKR2	5.015227646	Underexpressed	BMPR1A	
MALL	4.862568301	Underexpressed	BNC2	
FZD9	4.846703066	Underexpressed	BRCC3	
SOX8	4.589458322	Underexpressed	CAPN5	
NIPAL4	4.390635167	Underexpressed	CBX2	
PRSS46	4.318670654	Underexpressed	CCDC28B	
BRICD5	4.196793202	Underexpressed	CEP290	
AARD	4.137640679	Underexpressed	CEP41	
DIRAS2	3.944633523	Underexpressed	CGNL1	
GLDN	3.920223178	Underexpressed	CHD7	
CST9	3.840364214	Underexpressed	CITED2	
NKD2	3.798796974	Underexpressed	CLCN4	
DUSP15	3.786206266	Underexpressed	CLTCL1	
CITED4	3.722147786	Underexpressed	COL2A1	
TMEM200B	3.600024105	Underexpressed	COQ2	
EFHD1	3.475004837	Underexpressed	COX14	
TOX2	3.444991647	Underexpressed	CRH	
FSTL4	3.434470546	Underexpressed	CSF2RA	
SLC26A7	3.416325532	Underexpressed	CUL4B	
FAM46C	3.404545742	Underexpressed	CYB5A	
KIAA1804	3.364221265	Underexpressed	CYP11A1	

ESPN	3.304105135	Underexpressed	CYP11B1	
CNGA1	3.289376992	Underexpressed	CYP11B2	
CFTR	3.252877447	Underexpressed	CYP17A1	
PTK2B	3.198714887	Underexpressed	CYP19A1	
ADAMTS16	3.190049216	Underexpressed	CYP21A2	
SMOC2	3.18133361	Underexpressed	CYP7B1	
CSPG5	3.151143194	Underexpressed	DCAF17	
PPARGC1A	3.14794475	Underexpressed	DGKK	
CBLN1	3.143573222	Underexpressed	DHCR24	
AHSG	3.047744547	Underexpressed	DHCR7	
DTNA	3.025463286	Underexpressed	DHH	
JAML	3.012464658	Underexpressed	DISC1	
MRO	2.980612041	Underexpressed	DMRT1	
LPCAT2	2.919351	Underexpressed	DMRT2	
LRP4	2.906787146	Underexpressed	DND1	
COL27A1	2.906445023	Underexpressed	DNMT3L	
POPDC3	2.90460229	Underexpressed	EFNB1	
SPP1	2.877270188	Underexpressed	EHD1	
FBLN2	2.839900254	Underexpressed	ELAVL4	
SLC6A4	2.837312993	Underexpressed	EPHX1	
MFSD2A	2.817328515	Underexpressed	ERCC3	
ST3GAL4	2.8149864	Underexpressed	ERCC5	
GATM	2.812379197	Underexpressed	ESR1	
SMTNL2	2.798053878	Underexpressed	ESR2	
CEMIP	2.782810155	Underexpressed	ESRRA	
ADHFE1	2.770123692	Underexpressed	F9	
ACAP1	2.753425427	Underexpressed	FANCA	
NIPAL1	2.711846818	Underexpressed	FANCG	
MYBPHL	2.703880995	Underexpressed	FANCM	
COCH	2.687715023	Underexpressed	FEM1C	
PLA2G2F	2.657842557	Underexpressed	FGF10	
SERPINE2	2.655779698	Underexpressed	FGF8	
COL2A1	2.622345814	Underexpressed	FGF9	
TRIM47	2.61588044	Underexpressed	FGFR1	
PAK3	2.609585234	Underexpressed	FGFR2	
DCT	2.572618263	Underexpressed	FGFR3	
KCNK2	2.559737823	Underexpressed	FKBP4	
ATG9B	2.535018285	Underexpressed	FLNA	
EIF3J	2.513811934	Underexpressed	FOXF2	
CYP26B1	2.508865967	Underexpressed	FOXL2	
GAS7	2.469701009	Underexpressed	FREM2	
BARX2	2.469675205	Underexpressed	FSHB	
PPP1R3C	2.446262391	Underexpressed	FSHR	
RELT	2.440486625	Underexpressed	G6PC2	
GAL3ST1	2.438966909	Underexpressed	GAL	
RTN4RL1	2.392479459	Underexpressed	GATA4	
SPRY4	2.390869506	Underexpressed	GATA5	
RERG	2.384200023	Underexpressed	GHRHR	
REN	2.381279951	Underexpressed	GLYCTK	
COL9A2	2.380238517	Underexpressed	GNAS	
HSD17B3	2.361614016	Underexpressed	GNRH1	
SPATA2L	2.351927188	Underexpressed	GNRHR	
MAPK13	2.348581405	Underexpressed	GPC3	
PPP1R16B	2.31830908	Underexpressed	HAMP	
ADAMTSL2	2.314326038	Underexpressed	HBA1	
TMEM184A	2.297886065	Underexpressed	HBA2	
MYBL1	2.288200893	Underexpressed	HBB	
NGEF	2.271001818	Underexpressed	HES1	
DEFB119	2.251651252	Underexpressed	HESX1	
SCRN1	2.199678762	Underexpressed	HEXB	
RAB20	2.197794982	Underexpressed	HFE	
ZAR1	2.191764202	Underexpressed	HFE2	
GSTM1	2.180067453	Underexpressed	HOXA13	
HCN4	2.17783067	Underexpressed	HOXA4	
WIPF3	2.153849115	Underexpressed	HOXB6	
COL9A1	2.145411571	Underexpressed	HOXD10	
ETV4	2.141425931	Underexpressed	HS6ST1	
GFRA1	2.137688849	Underexpressed	HSD11B2	
RHPN1	2.133648616	Underexpressed	HSD17B3	
EMILIN3	2.133309197	Underexpressed	HSD3B2	

TNFRSF12A	2.12736642	Underexpressed	HTR2A	
GPR157	2.120702551	Underexpressed	HTR3A	
MYL4	2.114417637	Underexpressed	IGF1R	
BHLHE40	2.089660533	Underexpressed	IGFALS	
FOXS1	2.071701277	Underexpressed	INHBA	
TPD52L1	2.070663088	Underexpressed	INHBB	
KCTD12	2.063090789	Underexpressed	INSL3	
BVES	2.061582638	Underexpressed	INSR	
SNORA78	2.055478272	Underexpressed	INSRR	
GSTM2	2.054969791	Underexpressed	IRF6	
EGR1	2.03247483	Underexpressed	ITGB3	
PITPNM3	2.013541789	Underexpressed	KAL1	
LGR5	2.009966014	Underexpressed	KCNJ5	
TMEM59L	2.008618477	Underexpressed	KDM5D	
MAFF	1.989893456	Underexpressed	KIF7	
HIST1H2AJ	1.987322783	Underexpressed	KISS1	
LIPG	1.965736363	Underexpressed	KISS1R	
ABTB2	1.964094251	Underexpressed	KRT86	
LEFTY2	1.955667499	Underexpressed	LATS1	
CTGF	1.953865199	Underexpressed	LEP	
ABCB1	1.953209539	Underexpressed	LEPR	
PITX2	1.951544627	Underexpressed	LHB	
GALR3	1.927555487	Underexpressed	LHCGR	
CITED1	1.927251738	Underexpressed	LHFPL5	
GSTM2	1.923317345	Underexpressed	LHX3	
RGN	1.921622348	Underexpressed	LHX4	
PLA2G5	1.91384394	Underexpressed	LHX9	
TNFAIP6	1.905623944	Underexpressed	LIG4	
C16orf89	1.902021758	Underexpressed	LIPE	
AP1M2	1.90097902	Underexpressed	LMNA	
RTN1	1.893716951	Underexpressed	LZTFL1	
FAM46B	1.884866914	Underexpressed	MAGEB1	
RSG1	1.884228409	Underexpressed	MAGEB2	
STAR	1.883137846	Underexpressed	MAGEL2	
AGPAT4	1.881669037	Underexpressed	MAMLD1	
CHST1	1.874657235	Underexpressed	MAOA	
APBA1	1.874487801	Underexpressed	MAP3K1	
ETV5	1.86789436	Underexpressed	MAP3K4	
EYA1	1.862463216	Underexpressed	MBTPS2	
VDR	1.861342891	Underexpressed	MC4R	
RNF223	1.856798156	Underexpressed	MEF2B	
NTF3	1.842122036	Underexpressed	MEN1	
LHX1	1.840754706	Underexpressed	MID1	
COL18A1	1.839971455	Underexpressed	MKKS	
SHBG	1.837639646	Underexpressed	MKS1	
MYLK3	1.836429476	Underexpressed	MNX1	
WSCD1	1.824120006	Underexpressed	MSC	
ALDH1A1	1.818860004	Underexpressed	MTCP1	
CA12	1.809182273	Underexpressed	MTHFR	
FAM189A2	1.80784961	Underexpressed	MTMR1	
RASL12	1.803755689	Underexpressed	MYO1E	
RCAN1	1.799798957	Underexpressed	NCOA2	
SOD3	1.792261668	Underexpressed	NCOA4	
CLDN4	1.791679726	Underexpressed	NDN	
MICAL2	1.790454985	Underexpressed	NDUFS4	
SLCO3A1	1.790215506	Underexpressed	NEDD4	
GRAMD2	1.787934596	Underexpressed	NELF	
APH1B	1.787322077	Underexpressed	NKAIN2	
SEMA5A	1.786794482	Underexpressed	NMT2	
ACVR1C	1.777613166	Underexpressed	NOS1	
NUP62CL	1.774050708	Underexpressed	NPC1	
AKR1B10	1.773597347	Underexpressed	NPHS2	
DYNLT1	1.773091673	Underexpressed	NR0B1	
LRRC25	1.772346133	Underexpressed	NR2C1	
LMO4	1.768978609	Underexpressed	NR2E3	
PDZD2	1.765992318	Underexpressed	NR3C1	
DNM3OS	1.764734544	Underexpressed	NR4A1	
CARMN	1.764534059	Underexpressed	NR5A1	
HAS3	1.759283451	Underexpressed	OTX2	
S100A8	1.753832225	Underexpressed	PCSK1	

BCAR3	1.753076567	Underexpressed	PDE11A	
SPACA6	1.743499404	Underexpressed	PDE8B	
SLC16A7	1.742868522	Underexpressed	PEX2	
MAMLD1	1.73911045	Underexpressed	PGR	
CGA	1.738851327	Underexpressed	PHF6	
DKK1	1.734203792	Underexpressed	PKD1	
C8orf88	1.733425702	Underexpressed	PLCB3	
ENPP1	1.732472245	Underexpressed	PLLP	
LAMA4	1.731224309	Underexpressed	POLG	
CAMK2N1	1.730778472	Underexpressed	POLR3A	
CEND1	1.729012687	Underexpressed	POLR3B	
MFSD13A	1.725020768	Underexpressed	POR	
FJX1	1.724142221	Underexpressed	POU1F1	
SH2B2	1.723018218	Underexpressed	PROK2	
FGF1	1.720638406	Underexpressed	PROKR2	
HIST1H1E	1.71772735	Underexpressed	PROP1	
ATP1A2	1.71499176	Underexpressed	PSMC3IP	
DUSP14	1.711981287	Underexpressed	RAB3GAP2	
SCX	1.711367324	Underexpressed	RCN2	
TRANK1	1.698087228	Underexpressed	RET	
AQP1	1.696429719	Underexpressed	RFXAP	
REL	1.692285078	Underexpressed	RPGRIP1L	
DUSP6	1.689337851	Underexpressed	RPL35A	
TLX2	1.68136116	Underexpressed	RPS4X	
PDE2A	1.680291408	Underexpressed	RPS4Y1	
TINAGL1	1.679650642	Underexpressed	RSPO1	
TMEM114	1.668182154	Underexpressed	SALL1	
CMKLR1	1.665284304	Underexpressed	SAT1	
SYNJ2	1.66348239	Underexpressed	SDCCAG8	
EPS8	1.663175277	Underexpressed	SDHB	
RPRM	1.660709082	Underexpressed	SHOX	
HSD3B1	1.659951155	Underexpressed	SIL1	
IRS2	1.658844335	Underexpressed	SLC29A3	
JUNB	1.657983445	Underexpressed	SLC6A4	
TYRO3	1.655633933	Underexpressed	SLC9A3R2	
CTRB1	1.654668669	Underexpressed	SMPD2	
NR4A1	1.653278706	Underexpressed	SMPD3	
EPHB6	1.651455778	Underexpressed	SNORD116-1	
PTPRE	1.645823108	Underexpressed	SNORD116@	
CLDN8	1.643600868	Underexpressed	SOD2	
SOCS2	1.641928301	Underexpressed	SOX10	
EPHA4	1.639652309	Underexpressed	SOX2	
FBXL22	1.638852818	Underexpressed	SOX3	
FAM84A	1.636644604	Underexpressed	SOX9	
SPRY2	1.633819365	Underexpressed	SPO11	
PLEKHA4	1.631924426	Underexpressed	SRD5A1	
AFAP1L2	1.631016164	Underexpressed	SRD5A2	
TUBB3	1.63028545	Underexpressed	SRD5A3	
CDO1	1.626459314	Underexpressed	SRY	
SEMA3D	1.626093996	Underexpressed	STAR	
CCL25	1.625684728	Underexpressed	STAT5B	
SLC45A4	1.624164382	Underexpressed	STS	
CHCHD10	1.621336026	Underexpressed	TAB2	
PRSS53	1.619839241	Underexpressed	TAC3	
PMEPA1	1.614384903	Underexpressed	TACR3	
ERBB3	1.611556682	Underexpressed	TCF21	
C4orf47	1.610416521	Underexpressed	TDRD7	
RASGEF1B	1.609758174	Underexpressed	TFR2	
VAMP5	1.609647744	Underexpressed	THRB	
ART3	1.607507562	Underexpressed	TMEM67	
HS3ST1	1.606398673	Underexpressed	TMEM70	
GPR20	1.605300627	Underexpressed	TNXB	
NGFR	1.604264065	Underexpressed	TP63	
GALNT14	1.603235988	Underexpressed	TPH2	
IL17RD	1.600321585	Underexpressed	TPI1	
WIF1	1.599531822	Underexpressed	TRA2A	
DUSP4	1.59694461	Underexpressed	TRIM32	
ITGB8	1.596143315	Underexpressed	TRIM37	
TMEM150C	1.593567296	Underexpressed	TRNL1	
ANKH	1.588046739	Underexpressed	TRPC6	

WNT11	1.58549257	Underexpressed	TSPYL1	
CYP27A1	1.584928824	Underexpressed	TTC21B	
LDLRAD4	1.582911629	Underexpressed	TTC8	
HAS2	1.579712039	Underexpressed	TTR	
OPTC	1.577373246	Underexpressed	TUB	
SOX6	1.575725432	Underexpressed	UGT2B17	
JUN	1.574952164	Underexpressed	UGT2B28	
MAL2	1.570776774	Underexpressed	VANGL2	
TBX20	1.57031401	Underexpressed	WDPCP	
INF2	1.568763762	Underexpressed	WDR11	
MASP1	1.567569942	Underexpressed	WNT4	
NEDD9	1.565066197	Underexpressed	WT1	
TSPAN15	1.56362657	Underexpressed	WWOX	
TGOLN2	1.562371179	Underexpressed	XG	
CALY	1.56181677	Underexpressed	ZDHHC21	
WNT9B	1.561582466	Underexpressed	ZDHHC7	
NODAL	1.557602118	Underexpressed	ZEB2	
ADCYAP1R1	1.557479233	Underexpressed	ZFPM2	
ZFHX2	1.557323508	Underexpressed	ZFX	
GSG1L	1.557184237	Underexpressed	ZFY	
PLCB2	1.553184034	Underexpressed		
FAM19A5	1.550824214	Underexpressed		
LRRC75B	1.545668424	Underexpressed		
SEMA4A	1.544618596	Underexpressed		
CSPG4	1.540188486	Underexpressed		
TCF21	1.539581849	Underexpressed		
SLC16A4	1.533941193	Underexpressed		
PAQR8	1.533883211	Underexpressed		
AQP5	1.532708087	Underexpressed		
SPSB1	1.532008193	Underexpressed		
RND2	1.531732478	Underexpressed		
VSTM4	1.5255081	Underexpressed		
PPP1R36	1.524634844	Underexpressed		
MMP2	1.521252375	Underexpressed		
FHDC1	1.520483207	Underexpressed		
PITPNM2	1.516215956	Underexpressed		
CNTNAP1	1.515016598	Underexpressed		
C7orf31	1.514247829	Underexpressed		
PARM1	1.513625802	Underexpressed		
PGP	1.513071922	Underexpressed		
HSD11B2	1.511952732	Underexpressed		
PTPN7	1.508987116	Underexpressed		
PAG1	1.507527242	Underexpressed		
AUTS2	1.506979311	Underexpressed		
CHRNB4	1.504837673	Underexpressed		
EFHD2	1.50472639	Underexpressed		
CTTNBP2	1.500523263	Underexpressed		
CPM	1.500148848	Underexpressed		
SLMAP	1.500086017	Underexpressed		
SNORA17	Not Expressed in WT B6 Males	Overexpressed		
SNORA52	Not Expressed in WT B6 Males	Overexpressed		
RPS27	11.32170688	Overexpressed		
FST	5.015775937	Overexpressed		
SNORA81	4.354569893	Overexpressed		
FOXL2	4.221084368	Overexpressed		
WFDC1	3.703793901	Overexpressed		
KCNG4	3.448244778	Overexpressed		
RNF152	3.446044859	Overexpressed		
SLC7A11	3.297906051	Overexpressed		
ONECUT1	3.218813544	Overexpressed		
STUM	3.056932247	Overexpressed		
MIXL1	3.045712849	Overexpressed		
FAM196B	3.016041036	Overexpressed		
EIF3J	3.002723589	Overexpressed		
NMUR2	2.92225111	Overexpressed		
TSHR	2.807169723	Overexpressed		
VGLL2	2.776733916	Overexpressed		
C17orf97	2.681585036	Overexpressed		
CDH9	2.671882459	Overexpressed		
HSD17B13	2.638580788	Overexpressed		

SIGLEC10	2.599807364	Overexpressed		
HIST1H2BB	2.596494891	Overexpressed		
SPRR2D	2.575460231	Overexpressed		
PDZD7	2.522809788	Overexpressed		
SSTR3	2.495861846	Overexpressed		
SLITRK1	2.412227995	Overexpressed		
AGTR1	2.40639021	Overexpressed		
KLHDC7A	2.334831452	Overexpressed		
GPR55	2.331359995	Overexpressed		
KCNJ12	2.330557327	Overexpressed		
GBP3	2.277363542	Overexpressed		
CSGALNACT1	2.275820796	Overexpressed		
DCAF12L1	2.265471897	Overexpressed		
HIST1H4F	2.222366198	Overexpressed		
QPRT	2.211066075	Overexpressed		
ELFN2	2.203547914	Overexpressed		
BMP2	2.19205753	Overexpressed		
IRF7	2.173973271	Overexpressed		
TPH2	2.165042799	Overexpressed		
GBP2	2.163510119	Overexpressed		
LBX2	2.153010266	Overexpressed		
CTXN3	2.148315454	Overexpressed		
TNS4	2.136426679	Overexpressed		
SCRT2	2.133960795	Overexpressed		
CHRM4	2.12644749	Overexpressed		
WNT9A	2.090359813	Overexpressed		
C9orf152	2.089763669	Overexpressed		
EDN2	2.061351233	Overexpressed		
AKR1C3	2.034186876	Overexpressed		
PTCHD1	2.014200415	Overexpressed		
CGN	2.011798532	Overexpressed		
STX11	1.997671242	Overexpressed		
IRX3	1.970579587	Overexpressed		
NRN1	1.966620624	Overexpressed		
TAF7L	1.953627526	Overexpressed		
CDKN1B	1.948892084	Overexpressed		
PPARA	1.94436905	Overexpressed		
CPT1B	1.943004401	Overexpressed		
MGARP	1.92895282	Overexpressed		
PDK4	1.927050628	Overexpressed		
SYT13	1.923379289	Overexpressed		
ADTRP	1.921132122	Overexpressed		
KCNIP1	1.913031151	Overexpressed		
PLBD1	1.895975731	Overexpressed		
KRT20	1.893712982	Overexpressed		
ACAT2	1.889804397	Overexpressed		
EREG	1.88709216	Overexpressed		
KLHL5	1.877198817	Overexpressed		
ADM	1.874204743	Overexpressed		
BHLHA15	1.870258393	Overexpressed		
PLIN1	1.857005825	Overexpressed		
SP5	1.855243798	Overexpressed		
CYP11A1	1.852878122	Overexpressed		
IL15	1.836812458	Overexpressed		
KYAT3	1.828685522	Overexpressed		
CCDC182	1.82230537	Overexpressed		
PANK1	1.809516395	Overexpressed		
ISG15	1.804757885	Overexpressed		
CDKN1A	1.803095995	Overexpressed		
TXNDC2	1.798272949	Overexpressed		
FAM178B	1.794897531	Overexpressed		
FOXR1	1.780932866	Overexpressed		
LYPD6B	1.779994878	Overexpressed		
PRRT4	1.779099395	Overexpressed		
NPY1R	1.77835442	Overexpressed		
COL12A1	1.771471122	Overexpressed		
SPINK4	1.771210239	Overexpressed		
CFAP161	1.770392383	Overexpressed		
CCL2	1.770064146	Overexpressed		
HIST1H1A	1.769393464	Overexpressed		

TMEM171	1.765579488	Overexpressed		
PPP2R2B	1.765193267	Overexpressed		
GPRC5A	1.762134052	Overexpressed		
GNG13	1.761490003	Overexpressed		
RBKS	1.759679541	Overexpressed		
SH3BGR	1.759456337	Overexpressed		
VAV3	1.758471534	Overexpressed		
AIF1	1.756964897	Overexpressed		
EFCAB10	1.75594238	Overexpressed		
MCTP1	1.748212112	Overexpressed		
ISX	1.744613287	Overexpressed		
WNT4	1.742707143	Overexpressed		
PIPOX	1.741384502	Overexpressed		
WFDC13	1.734425578	Overexpressed		
C16orf45	1.733405527	Overexpressed		
ADAM8	1.730807823	Overexpressed		
SPRR2E	1.72918969	Overexpressed		
RORC	1.722362727	Overexpressed		
TRNP1	1.71335502	Overexpressed		
SLC37A2	1.710394409	Overexpressed		
SMIM10L2A	1.71011236	Overexpressed		
HIST1H3I	1.708885931	Overexpressed		
RNASET2	1.708081348	Overexpressed		
EPX	1.702774666	Overexpressed		
PGLYRP1	1.699234872	Overexpressed		
TMEM40	1.697091467	Overexpressed		
PET117	1.696923289	Overexpressed		
OAS2	1.695282499	Overexpressed		
ACTR6	1.683190047	Overexpressed		
CYBRD1	1.681156707	Overexpressed		
TMEM174	1.680620355	Overexpressed		
AHNAK	1.677631618	Overexpressed		
PLD1	1.676163644	Overexpressed		
GABRA4	1.675159704	Overexpressed		
SLC1A6	1.673739143	Overexpressed		
TGM3	1.671730237	Overexpressed		
PLAC1	1.665417966	Overexpressed		
SIGLEC5	1.663814485	Overexpressed		
MSX1	1.661100614	Overexpressed		
IDI1	1.660112402	Overexpressed		
NPB	1.659384896	Overexpressed		
MSX2	1.657707699	Overexpressed		
ELFN1	1.656938083	Overexpressed		
ORC1	1.656137519	Overexpressed		
LZTS1	1.654276926	Overexpressed		
TSPAN33	1.652136543	Overexpressed		
ASNS	1.651053646	Overexpressed		
PHLDA2	1.647210926	Overexpressed		
SPRR1A	1.643740569	Overexpressed		
GXYLT2	1.640503814	Overexpressed		
PNLIPRP1	1.633228932	Overexpressed		
DHDH	1.62881779	Overexpressed		
SMPX	1.627926953	Overexpressed		
FAM83G	1.624749915	Overexpressed		
DCAF12L2	1.624211209	Overexpressed		
FRMD6	1.623913577	Overexpressed		
EGFL6	1.617778676	Overexpressed		
SLC7A2	1.615677244	Overexpressed		
TRIM50	1.611484558	Overexpressed		
C11orf65	1.604872802	Overexpressed		
RSPO1	1.602248498	Overexpressed		
HPSE	1.599898053	Overexpressed		
ANGEL1	1.59674608	Overexpressed		
COQ8A	1.596129176	Overexpressed		
GK5	1.590209768	Overexpressed		
SH2D4A	1.58545061	Overexpressed		
SFXN3	1.584349425	Overexpressed		
FOXJ1	1.584320251	Overexpressed		
CPNE8	1.577968063	Overexpressed		
CDKN1C	1.576690526	Overexpressed		

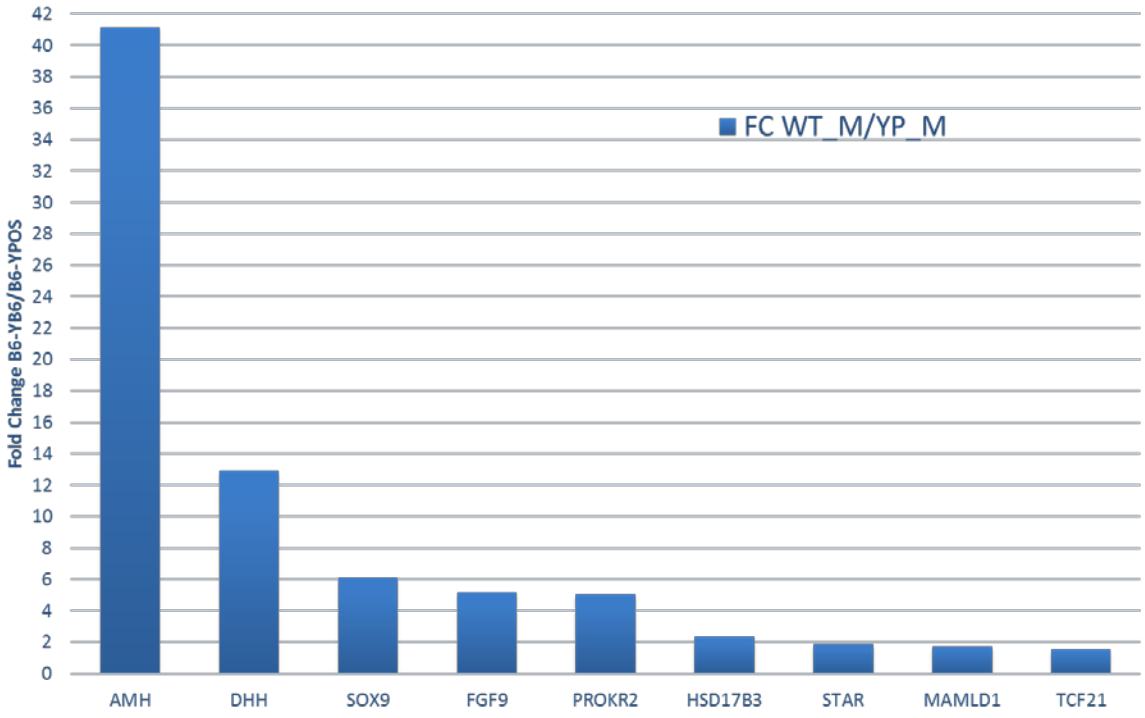
TCEA3	1.576256699	Overexpressed		
TGFA	1.574456597	Overexpressed		
RPL39L	1.574046014	Overexpressed		
GTF2A1	1.570612911	Overexpressed		
HPGD	1.569853795	Overexpressed		
RASGRP1	1.564330869	Overexpressed		
DBNDD1	1.562684113	Overexpressed		
GRIN2C	1.561425831	Overexpressed		
FGFR2	1.558340878	Overexpressed		
KCNIP2	1.55752982	Overexpressed		
NECAB3	1.555544383	Overexpressed		
GABRB1	1.554939015	Overexpressed		
RNF128	1.553661667	Overexpressed		
PLCH1	1.551738416	Overexpressed		
PCBD1	1.549787805	Overexpressed		
ZNF703	1.548053594	Overexpressed		
MTERF1	1.547233678	Overexpressed		
IFI27	1.545121063	Overexpressed		
CRYL1	1.544270434	Overexpressed		
TMEM220	1.543977958	Overexpressed		
TRIM14	1.541887737	Overexpressed		
TSPYL5	1.541854931	Overexpressed		
TMEM141	1.53932226	Overexpressed		
CHST4	1.538911709	Overexpressed		
ZNF277	1.534721179	Overexpressed		
HIST1H2BJ	1.531751971	Overexpressed		
IGF1	1.531496286	Overexpressed		
SPOCK2	1.529845464	Overexpressed		
MORN1	1.528812451	Overexpressed		
TTC38	1.528809022	Overexpressed		
SLC30A5	1.523289108	Overexpressed		
HSD17B11	1.523104472	Overexpressed		
MOAP1	1.521789255	Overexpressed		
SALL3	1.520033582	Overexpressed		
MID1IP1	1.516664671	Overexpressed		
IFITM3	1.514798276	Overexpressed		
KCND2	1.512721511	Overexpressed		
IBSP	1.511966492	Overexpressed		
SOHLH2	1.508940653	Overexpressed		
NLRP4	1.50774366	Overexpressed		
SERPINB6	1.507622765	Overexpressed		
CBS	1.507059254	Overexpressed		
SPRR2G	1.50579062	Overexpressed		
NEU3	1.505199617	Overexpressed		
REC8	1.50127842	Overexpressed		
NIM1K	1.500014461	Overexpressed		

Supplemental Table 3-S2: Primer sets used for qPCR validation

Gene	Forward primer	Reverse primer	Annealing temperature (in Celsius)
Tox2	GGCCTACGCTCTCTTCTTC	CCTCTTATACGCCTGTTTCTG	60
Dusp15	AACTTCATTGATGCCAAAGAC	GGTGATTCGTGGATAGAGATG	61
Nkd2	GTGGCAGAACAGAGATTGAG	TTGGGCTTCCTGCTGTAG	60
Cnga1	TCAACAACAGCAGCAACAAAG	TATCATCGGCCTTGCTCTTC	64
Ptk2b	TGGATGTGGAGAAGGAAGAC	TGATGATCTCCTGGATCTCTG	61
Espn	ATTACCCTGAGGGAGTGAATG	AAGGTACTTCGTCACTTCCAG	60
Smoc2	CAAATGGAAGACCCATCAG	CAGCATCATCTGCTTTCC	60
Adamts16	AACATGGTGTCTGCCTTATTC	CCCTGGCTGTTCATCTTC	63
Fbln2	GTGATCTTGATGGCTCCAC	CTGGGCTATCCTACAGATGTC	60
Nipal1	GGGTCAACTGTGATGGTTATC	AACGAACCCTGGATCTCTC	61
Cyp26b1	GTACCCAGGGCAAAGACTAC	GGTTCCATCCTTCAGCTC	60
Spry4	TGCAGCTCCTCAAAGACC	ATGACTGAGCTGGGATTCAC	60
Mybl1	GTCAGCCGAGAATGAAGTTAG	AGCTTCCAGGTTGAGGTG	61
Etv4	CAGGACCTCAGTCACTTCC	CGGTACCTGAGCTTCTGC	62
Lgr5	AACCTCCGATCTCTGAACTTAG	CGACAGGAGATTGGATGATAG	63

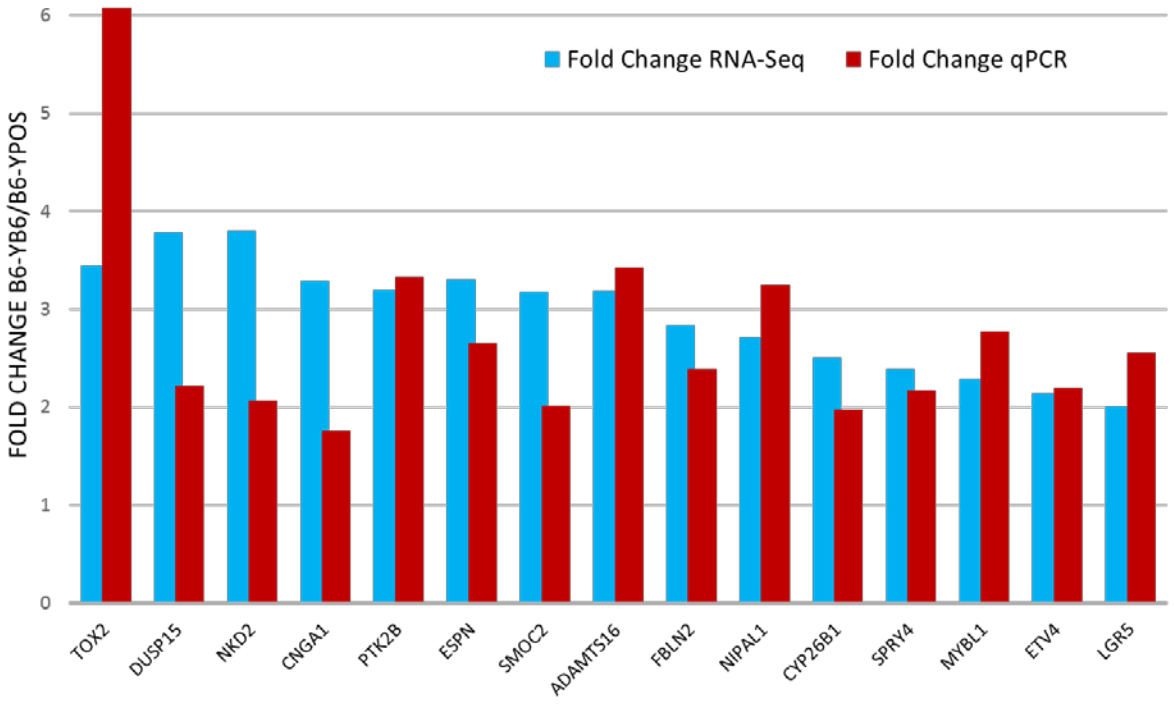
Supplemental Figure 3-S1: Fold expression changes in known DSD genes underexpressed in B6-Y^{POS} males

Expression values are shown as fold change differences between B6-Y^{B6} and B6-Y^{POS} males for genes present in primary gene list used for exome variant filtration.

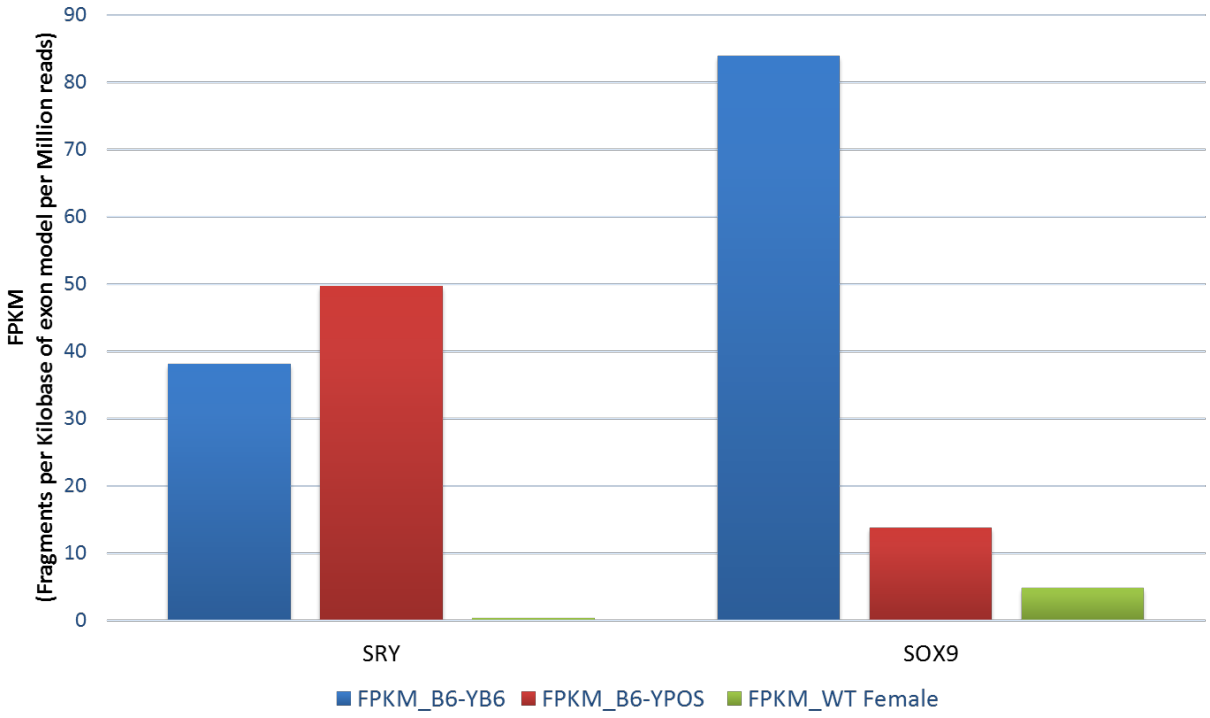


Supplemental Figure 3-S2: Validation of candidate gene expression differences between B6-Y^{B6} and B6-Y^{POS} males via qPCR

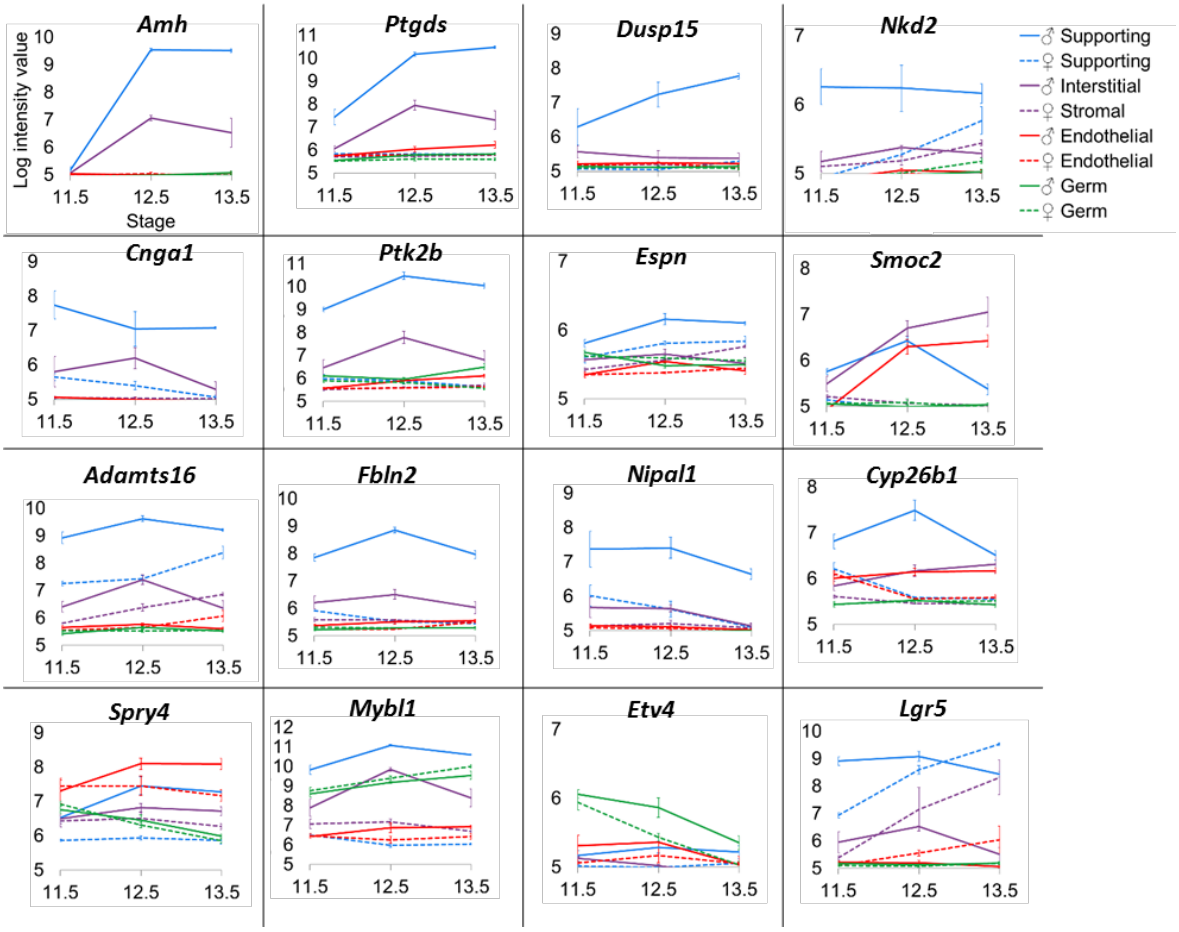
Expression values are shown as fold change differences between B6-Y^{B6} and B6-Y^{POS} males using RNA-Seq data (blue) and qPCR data (red).



Supplemental Figure 3-S3: *Sry* and *Sox9* expression in the E11.5 embryonic gonad Expression values shown in FPKM (fragments per kilobase of exon model per million reads) measured by RNA sequencing.



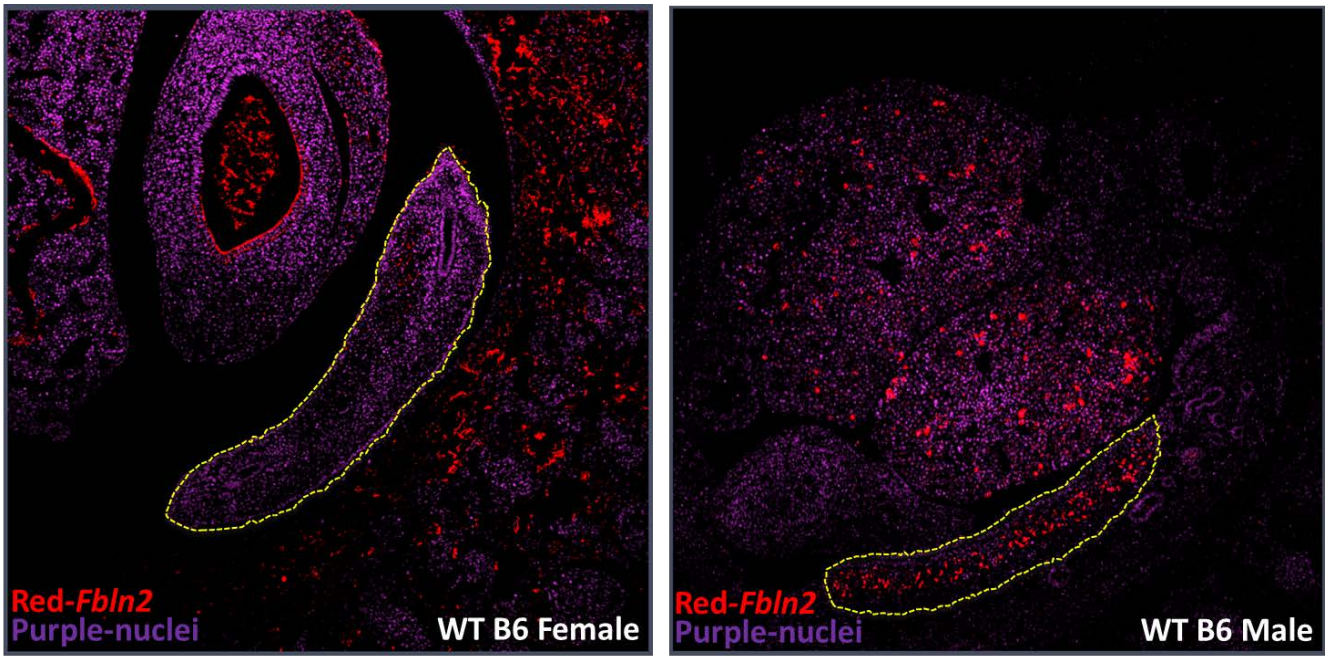
Supplemental Figure 3-S4: Profiles of candidate gene expression in the gonad across sex determination



Candidate gene profile graphs were generated from the microarray performed by Jameson et al. 2012 [27] where gene expression was profiled in each cell population of the gonad at E11.5, E12.5 and E13.5. Similar to the *Sox9* target genes, *Amh* and *Ptgds* the candidate genes show strong expression in the male supporting lineage (blue dotted line) compared to the female (blue solid line). There was no information available in the microarray data for *Tox2*.

Supplemental Figure 3-S5: *Fbln2* protein expression in WT B6 females and males at E12.5 by immunohistochemistry

Embryonic section of WT B6 females and males at E12.5 stained for Fbln2 (red) and cell nuclei (purple). Fbln2 is expressed in a sexually dimorphic pattern as no expression is present in WT B6 female (left), whereas the expression in WT B6 male is high. The gonads are encircled by yellow dashed lines.



References:

1. Berta, P., J.R. Hawkins, A.H. Sinclair, A. Taylor, B.L. Griffiths, P.N. Goodfellow, et al., *Genetic evidence equating SRY and the testis-determining factor.* Nature 1990; **348**(6300): 448-450.
2. Lee, P.A., C.P. Houk, S.F. Ahmed, and I.A. Hughes, *Consensus statement on management of intersex disorders. International Consensus Conference on Intersex.* Pediatrics 2006; **118**(2): e488-500.
3. Sandberg, D.E., M. Gardner, and P.T. Cohen-Kettenis, *Psychological aspects of the treatment of patients with disorders of sex development.* Semin Reprod Med 2012; **30**(5): 443-452.
4. Warne, G.L., *Long-term outcome of disorders of sex development.* Sex Dev 2008; **2**(4-5): 268-277.
5. Delot, E.C. and E.J. Vilain, *Nonsyndromic 46,XX Testicular Disorders of Sex Development*, in *GeneReviews(R)*, R.A. Pagon, et al., Editors. 1993: Seattle (WA).
6. Baetens, D., H. Stoop, F. Peelman, A.L. Todeschini, T. Rosseel, F. Coppieters, et al., *NR5A1 is a novel disease gene for 46,XX testicular and ovotesticular disorders of sex development.* Genet Med 2016.
7. Bashamboo, A., P.A. Donohoue, E. Vilain, S. Rojo, P. Calvel, S.N. Seneviratne, et al., *A recurrent p.Arg92Trp variant in steroidogenic factor-1 (NR5A1) can act as a molecular switch in human sex development.* Hum Mol Genet 2016.

8. Ledig, S., O. Hiort, L. Wunsch, and P. Wieacker, *Partial deletion of DMRT1 causes 46,XY ovotesticular disorder of sexual development.* Eur J Endocrinol 2012; **167**(1): 119-124.

9. Barseghyan, H., E. Delot, and E. Vilain, *New genomic technologies: an aid for diagnosis of disorders of sex development.* Horm Metab Res 2015; **47**(5): 312-320.

10. Eggers, S., S. Sadedin, J.A. van den Bergen, G. Robevska, T. Ohnesorg, J. Hewitt, et al., *Disorders of sex development: insights from targeted gene sequencing of a large international patient cohort.* Genome Biol 2016; **17**(1): 243.

11. Kim, J.H., E. Kang, S.H. Heo, G.H. Kim, J.H. Jang, E.H. Cho, et al., *Diagnostic yield of targeted gene panel sequencing to identify the genetic etiology of disorders of sex development.* Mol Cell Endocrinol 2017; **444**: 19-25.

12. Baxter, R.M., V.A. Arboleda, H. Lee, H. Barseghyan, M.P. Adam, P.Y. Fechner, et al., *Exome sequencing for the diagnosis of 46,XY disorders of sex development.* J Clin Endocrinol Metab 2015; **100**(2): E333-344.

13. Lee, H., J.L. Deignan, N. Dorrani, S.P. Strom, S. Kantarci, F. Quintero-Rivera, et al., *Clinical exome sequencing for genetic identification of rare Mendelian disorders.* JAMA 2014; **312**(18): 1880-1887.

14. Yang, Y., D.M. Muzny, J.G. Reid, M.N. Bainbridge, A. Willis, P.A. Ward, et al., *Clinical whole-exome sequencing for the diagnosis of mendelian disorders.* N Engl J Med 2013; **369**(16): 1502-1511.

15. Arboleda, V.A., A. Fleming, H. Barseghyan, E. Delot, J.S. Sinsheimer, and E. Vilain, *Regulation of sex determination in mice by a non-coding genomic region.* Genetics 2014; **197**(3): 885-897.

16. Umemura, Y., R. Miyamoto, R. Hashimoto, K. Kinoshita, T. Omotehara, D. Nagahara, et al., *Ontogenic and morphological study of gonadal formation in genetically-modified sex reversal XY(POS) mice.* J Vet Med Sci 2016; **77**(12): 1587-1598.

17. Eicher, E.M., L.L. Washburn, J.B. Whitney, 3rd, and K.E. Morrow, *Mus poschiavinus Y chromosome in the C57BL/6J murine genome causes sex reversal.* Science 1982; **217**(4559): 535-537.

18. Arboleda, V.A., A. Fleming, H. Barseghyan, E. Delot, J.S. Sinsheimer, and E. Vilain, *Regulation of Sex Determination in Mice by a Non-Coding Genomic Region.* Genetics 2014.

19. Bullejos, M. and P. Koopman, *Spatially dynamic expression of Sry in mouse genital ridges.* Dev Dyn 2001; **221**(2): 201-205.

20. Bullejos, M. and P. Koopman, *Delayed Sry and Sox9 expression in developing mouse gonads underlies B6-Y(DOM) sex reversal.* Dev Biol 2005; **278**(2): 473-481.

21. The Gene Ontology, C., *Expansion of the Gene Ontology knowledgebase and resources.* Nucleic Acids Res 2016.

22. Harding, S.D., C. Armit, J. Armstrong, J. Brennan, Y. Cheng, B. Haggarty, et al., *The GUDMAP database--an online resource for genitourinary research.* Development 2011; **138**(13): 2845-2853.

23. Barrionuevo, F., I. Georg, H. Scherthan, C. Lecureuil, F. Guillou, M. Wegner, et al., *Testis cord differentiation after the sex determination stage is independent of Sox9 but fails in the combined absence of Sox9 and Sox8.* Dev Biol 2009; **327**(2): 301-312.

24. Barrionuevo, F., S. Bagheri-Fam, J. Klattig, R. Kist, M.M. Taketo, C. Englert, et al., *Homozygous inactivation of Sox9 causes complete XY sex reversal in mice.* Biol Reprod 2006; **74**(1): 195-201.

25. Lavery, R., A. Lardenois, F. Ranc-Jianmotamedi, E. Pauper, E.P. Gregoire, C. Vigier, et al., *XY Sox9 embryonic loss-of-function mouse mutants show complete sex reversal and produce partially fertile XY oocytes.* Dev Biol 2011; **354**(1): 111-122.

26. Bi, W., W. Huang, D.J. Whitworth, J.M. Deng, Z. Zhang, R.R. Behringer, et al., *Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization.* Proc Natl Acad Sci U S A 2001; **98**(12): 6698-6703.

27. Jameson, S.A., A. Natarajan, J. Cool, T. DeFalco, D.M. Maatouk, L. Mork, et al., *Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad.* PLoS Genet 2012; **8**(3): e1002575.

28. Behringer, R.R., M.J. Finegold, and R.L. Cate, *Mullerian-inhibiting substance function during mammalian sexual development.* Cell 1994; **79**(3): 415-425.

29. Abdul-Majeed, S., B. Mell, S.M. Nauli, and B. Joe, *Cryptorchidism and infertility in rats with targeted disruption of the Adamts16 locus.* PLoS One 2014; **9**(7): e100967.

30. Miraoui, H., A.A. Dwyer, G.P. Sykiotis, L. Plummer, W. Chung, B. Feng, et al., *Mutations in FGF17, IL17RD, DUSP6, SPRY4, and FLRT3 are identified in individuals with congenital hypogonadotropic hypogonadism.* Am J Hum Genet 2013; **92**(5): 725-743.

31. Genomes Project, C., G.R. Abecasis, D. Altshuler, A. Auton, L.D. Brooks, R.M. Durbin, et al., *A map of human genome variation from population-scale sequencing.* Nature 2010; **467**(7319): 1061-1073.

32. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform.* Bioinformatics 2010; **26**(5): 589-595.

33. Rehm, H.L., S.J. Bale, P. Bayrak-Toydemir, J.S. Berg, K.K. Brown, J.L. Deignan, et al., *ACMG clinical laboratory standards for next-generation sequencing.* Genet Med 2013; **15**(9): 733-747.

34. Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.* Nat Protoc 2009; **4**(7): 1073-1081.

35. Adzhubei, I.A., S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, et al., *A method and server for predicting damaging missense mutations.* Nature methods 2010; **7**(4): 248-249.

36. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics 2009; **25**(9): 1105-1111.

37. Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics 2013; **29**(1): 15-21.

38. Trapnell, C., B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nat Biotechnol 2010; **28**(5): 511-515.

39. The Gene Ontology, C., *Expansion of the Gene Ontology knowledgebase and resources.* Nucleic Acids Res 2017; **45**(D1): D331-D338.

40. Mi, H., A. Muruganujan, J.T. Casagrande, and P.D. Thomas, *Large-scale gene function analysis with the PANTHER classification system.* Nat Protoc 2013; **8**(8): 1551-1566.

41. Wilhelm, D., L.L. Washburn, V. Truong, M. Fellous, E.M. Eicher, and P. Koopman, *Antagonism of the testis- and ovary-determining pathways during ovotestis development in mice.* Mech Dev 2009; **126**(5-6): 324-336.

42. Hennekam, R.C., J.E. Allanson, L.G. Biesecker, J.C. Carey, J.M. Opitz, and E. Vilain, *Elements of morphology: standard terminology for the external genitalia.* Am J Med Genet A 2013; **161A**(6): 1238-1263.

Chapter 4

Identification of Large Causative Genetic Variants via Next-Generation Genome Mapping and Sequencing

Abstract

DNA sequencing has become a routine procedure in clinics across the United States and around the globe where physicians order either exome or single gene sequencing to identify pathogenic variants leading to patient's phenotype and provide the appropriate care. Massively parallel DNA sequencing has the capability of reliably identifying single nucleotide variants (SNVs) and small insertions and deletions (INDELs); however, due to its innate methodology that relies on generation of short reads, this platform fails to identifiy large structural variants (SVs) such as insertions, deletions, inversions and translocations. In order to overcome these limitations and provide genetic diagnosis for patients with previous negative exome sequencing, we used **Irys genome mapping technology** that relies on imaging of fluorescently labeled native-state DNA molecules (up to 1Mb in size) in nanochannel arrays for genome assembly and SV detection. First, to ascertain the SV detection capability of the Irys system, we investigated a series of patients diagnosed with Duchenne Muscular Dystrophy (DMD) who were known to carry a large deletion, insertion or inversion in the Dystrophin gene by PCR, but with imprecise DNA break points. Second, we performed genome sequencing and genome mapping on patients diagnosed with disorders of sex development (DSD) to discover pathogenic SVs, SNVs or INDELs. Using this strategy we have successfully identified the intronic breakpoints within the *Dystrophin* gene where our cohort of DMD patients carried either a deletion, a duplication or an inversion. The sizes of identified deletions and insertions ranged from 30kb – 200kb and 10kb – 150kb respectively encompassing several exons and introns. We also were able to identify heterozygous SVs in carrier mothers of DMD patients indicating the ability of the

method to distinguish between homozygous and heterozygous SVs, and therefore distinguish between carrier and affected status of an individual. On average, in our cohort of both DMD and DSD cases, the Irys system identified around 1300 insertions, 700 deletions, 50 inversions and 20 translocations of variable sizes per genome. To filter out possibly benign variants, we used a database containing SVs from 144 healthy individuals and were able to narrow the list of SVs down to several that could potentially play a role in patient's phenotype. Although we see that SVs are less common than SNVs and INDELs, they account for a significant fraction of genetic variation. **Next-Generation Mapping (NGM)** is poised to identify potential pathogenic SVs in affected individuals and soon be incorporated in the clinical diagnostic strategy.

Introduction

Although Sanger sequencing is still widely used for reads around 1kb in length, new developments of massively parallel sequencing reactions have taken over the global market share of sequencing due to cheap price and fast turnaround times. This is in part due to the large improvements made in both imaging technology and microengineering. Next-generation sequencing technology has allowed for the discovery of many pathogenic single-nucleotide mutations. Next-generation sequencing from Illumina utilizes flowcells covered with millions of surface-bound oligonucleotides that allow parallel sequencing of millions of short reads. Resulting reads are then aligned to a reference genome for the genome assembly. This method misses on deletions, insertions, or duplications larger than the produced reads. The most common method for detecting large insertions or deletions is through the use of chromosomal microarrays. For diseases in which the suspected mutations are copy number variations, array-based DNA hybridization techniques can be used. However, CMAs cannot detect balanced translocations or inversions.

Similarly to exome sequencing, genome sequencing uses next-generation sequencing, but is capable of identifying single nucleotide variants, insertion/deletions and copy number variants not only in exons but also in non-coding regions of the genome. This allows for identification of variants affecting gene regulation, but currently has limited use in the clinical realm due to our inability to interpret most variants. A method that is useful in conjunction with genome sequencing is genome mapping, which utilizes labelled high molecular weight DNA to preserve large elements of structure of the genome. This allows the examination of the larger genome and

detection of structural variants. Genome mapping also can be used to create scaffolds for sequence assembly. Genome mapping enables a more complete picture of the genome, and allows us to view breakpoints, large insertions, and deletions in the genome, which would not be possible with only genome sequence.

To facilitate the identification of genetic diagnosis in DSD patients with negative exome, we will utilize genome sequencing and new genome mapping technology capable of identifying large insertions, deletions, copy number variants, inversions and balanced translocations. Combined these two technologies will provide base level resolution and physical map of the human genome allowing for identification of the majority of known genetic variation.

Here, to validate the new, still under development method, we first used nanochannel-based Next Generation Mapping (NGM) technology to detect large insertions, deletions, copy number variations, as well as inversions and translocations in Duchenne muscular dystrophy (DMD). DMD is an X-linked recessive form of muscular dystrophy which affects 1 in 3500 male newborns. It is a disease that is characterized by symptoms of delayed motor reactions and limb weakness, with many affected patients becoming wheelchair-bound by 11-12 years old. The disease is caused by mutations on the dystrophin gene, which is on position Xp21. The dystrophin gene encodes one of the largest proteins in the human genome, with 79 exons, which encode for a 14kb mRNA strand. A study of a database with over 7000 recorded mutations on the dystrophin gene showed that 86% of total mutations were large deletions of 1 exon or greater.

We have also performed genome sequencing on 95 trios diagnosed with DSD and perform physical map assembly for an additional 17 families.

Methods

Genome mapping:

To assemble a genomic physical map for identification of large deletions, insertions, translocation and inversions, we used Bionano Genomics Irys technology that relies on labeling of restriction enzyme sites throughout long strands of DNA molecules and then reverse-assemble these molecules with the reference genome in order to identify structural variants.

DNA isolation for genome mapping:

Irys technology requires long strands of DNA molecules (>150kb in size) encompassing multiple restriction enzyme sites for proper genome assembly. To achieve this requirement, DNA is extracted from fresh peripheral blood following manufacturers PlugLysis protocol (Bionano Genomics, USA). Briefly, red blood cells (RBC) are lysed using RBC lysis solution (Qiagen) and white blood cells (WBC) are centrifuged and re-suspended in cell suspension buffer (Bio-Rad). The WBC are embedded into agarose plugs (CHEF Genomic DNA Plug Kit, Bio-Rad) for overnight lysis at 50°C using 16:1 ratio of lysis buffer (Bionano Genomics, USA) and Proteinase K (Qiagen). The plugs are washed with TE Buffer and digested at 43°C with GELase (Epicentre). Extracted high molecular weight DNA is purified from digested materials/enzymes via drop dialysis using Millipore membrane filters (EMD Millipore,

USA) placed on a Tris-EDTA buffer. DNA quantifications were carried out using Qubit dsDNA assay kits with Qubit 3.0 Fluorometer (Thermo Fisher).

DNA labeling and chip loading for genome mapping:

DNA labeling consists of 4 sequential steps (Figure 4-1) and was performed using the IrysPrep Reagent Kit (BioNano Genomics). Briefly, 900ng of purified high molecular weight DNA was nicked with nicking endonucleases Nt.BspQI and Nt.BssSI (New England BioLabs) in 10X Buffer 3 at 37°C for 2 hours. The nicked DNA was then labeled with 10X Labeling Mix containing fluorophore-labeled nucleotide using Taq polymerase (NEB) at 72°C for 1 hr before being repaired with Taq ligase (NEB) and IrysPrep Repair Mix, NAD+, and 10X Thermopol buffer at 37°C for 30 minutes. The backbone of DNA was then stained with IrysPrep DNA stain, 5X DTT, and 4X flow buffer overnight at 4°C. Labeled DNA was loaded on Irys chip and ran for 24 hours (Figure 4-2). In the chip, the sample is run through a low-voltage electric field. DNA is first concentrated in a gradient region (lip) before being pushed through a pillar region, need for DNA linearization to enter nanochannel arrays. The fluorescently labeled DNA molecules are imaged sequentially across nanochannels by Irys instrument producing thousands of high-resolution images of single DNA molecules that are then used for genome assembly and variant calling. To achieve the necessary coverage of 100x necessary for accurate identification of structural variants 4-6 Irys chips were run, each at 30 cycles.

Data analysis for genome mapping:

Genome assembly was performed using IrysView/IrysSolve software solutions provided by Bionano Genomics. The raw TIFF images of labeled long DNA molecules were converted to BNX files containing DNA backbone, nicked site, and quality score information for each molecule/label. The conversion was accomplished via AutoDetect software (Bionano Genomics). Due to large size of the raw data that is acquired in the form of TIFF images, it was not economical to store and we opted to store only BNX files. This decision has no consequences on the quality of the data, downstream processing or analysis. Assembly of the genome using BNX files and further structural variation detection was performed using pipelines generated by Bionano Genomics [1]. Briefly, the human genome reference hg19 was digested *in silico* using either Nt.BspQI and Nt.BssSI enzymes to generate a reference consensus map. The reference consensus map was used for initial molecule alignment and generation of contigs. Later, the molecule information was used to fine-tune the initially generated contings to accurately represent the molecule data. Comparison of the two maps (sample vs reference) was used to identify structural variants. Identified structural variants were filtered against a healthy control dataset containing 144 individuals (Bionano Genomics). Data visualization was accomplished via IrysView 2.5.1 software (Bionano Genomics).

Validation of SVs via PCR and quantitative PCR

Validation of identified SVs was performed using PCR and qPCR. The primer sequences used are detailed in (Table 4-1). Primers were designed using primer design software Primer3 [2]. DNA was quantified using QuBit HS (Invitrogen) for double

stranded DNA and a total of 2ng of DNA was used per sample for qPCR reaction. qPCR was carried out in quadruplicates and duplicates utilizing Syber Green-based SensiFAST™ SYBR No-ROX Kit (Bioline, UK) by DNA Engine Opticon® 2 real-time PCR detection system from Bio-Rad Laboratories (BioRad, USA). Reaction conditions were as follows: 95ºC for 10 min, then 40 cycles of 95ºC for 15 sec, 60ºC for 10 sec (see Table S2), and 72ºC for 15 sec. A total of 100ng of DNA per sample was used for regular PCR. PCR was carried out utilizing MyTaq™HS Red Mix by BioRad Thermal Cycling Platform,1000 Series (BioRad USA). Reaction conditions were as follows: 95° for 1min, and then 35 cycles for 95°C for 15 sec, 95°C for 15 sec, 60°C for 15 sec, and 72°C for 10sec.

DNA isolation and sequencing for WGS:

DNA was isolated from peripheral blood Gentra Puregene Blood Kit (Qiagen, Germantown, MD, USA) or saliva collected using ORAgene Discover ORG-500 (DNAgenoteck Ottawa, ON, Canada). Sequencing libraries were created for each individual sample following manufacturer's protocol (Illumina protocol *Preparing Samples for Sequencing Genomic DNA*, p/n 11251892 Rev. A). Sequencing was performed on an Illumina HiSeq X with 100bp paired-end reads at the Baylor Miraca Genetics Laboratories. The base-calling was performed using the real-time analysis (RTA) software provided by Illumina.

Data analysis for WGS:

The sequence reads FASTQ files were aligned to the human reference genome (GRCh37/hg19 Feb. 2009 assembly) using BWA (Burrows-Wheeler Alignment Tool) [3]. PCR duplicates were marked by Picard. GATK (Genome Analysis ToolKit) was used to realign INDELs, recalibrate quality scores, call, filter, recalibrate and evaluate variants (single-nucleotide variants (SNVs) and INDELs). SPLITREAD [4] and CoNIFER [5] were used to call structural variants (SVs).

Quality control:

For each sample, SNVs and small INDELs were evaluated to determine if they fall within normal human genomic variation quality parameters (total number of variants, concordance rate, het/homo ratio, Ti/Tv ratio). Coverage of the sex chromosomes were assessed for each sample to confirm match with the reported karyotype. Mendelian errors (errors that lead to genotypes incompatible with Mendelian inheritance) were checked to confirm the relationship between family members.

Sanger validations*:*

It was previously demonstrated that variants with a QUAL score (a scaled probability of a variant existing at a given site based on the sequencing data, calculated using GATK variant caller) of 500 or higher are highly accurate [5]. Variants of interest with a QUAL score lower than 500 and all INDELs were confirmed by Sanger sequencing.

Variant annotation for SNVs and INDELs:

Variants were annotated using VarSeq (GoldenHelix, USA). Each variant was annotated with information including gene names and accession numbers, reference variant, variant consequences (*e.g.*, missense, frameshift), amino-acid changes, conservation scores, MAF, expression pattern, CADD scores [6], Polyphen [7] and SIFT [8] used for predicting the deleteriousness of single nucleotide and small indel variants.

Genome sequencing variant filtering:

We first searched for rare causal risk alleles that are fully penetrant with a high genetic relative risk. A recessive model was considered for homozygous, hemizygous and compound heterozygous variants and a dominant model was considered for *de novo* variants and inherited heterozygous variants from the opposite sex parent (*e.g*, when an affected 46,XX male inherited a heterozygous variant from his 46,XX mother). *Step 1:* Common variants with MAF > 1% were filtered out. Population frequency data was retrieved from dbSNP and NHLBI Exome Sequencing Project (ESP) Exome Variant Server (EVS) for SNVs and INDELs and the DGV database of genomic variants for structural variants. Variants that are repeatedly called within the sample set of the 97 trios at >10% were also filtered out (unless the variant was only observed in the patient group) to remove variants that are most likely technical false positives. *Step2:* Within the exome, synonymous variants were filtered out. *Step 3:* Since ExAC (Exome Aggregation Consortium) database is reported to not have included pediatric cases with significant birth defect, *de novo*, homozygous, hemizygous, heterozygous or compound heterozygous variants observed in >4 individuals in ExAC (Exome Aggregation

Consortium) as heterozygous were filtered out. The identified variants were segregated into 6 types (Table 4-2).

Results

NGM in Duchenne Muscular Dystrophy

Next generation mapping is a new emerging tool in the market of genetics that still needs to prove its worth within the diagnostic arena. Previously, Bionano's Irys technology has been mostly used for *de novo* assemblies of genomes with absent or incomplete reference. More recently, we and other laboratories saw the potential of this technology to identify large structural variants (SVs) such as insertions, deletions, inversions, balanced translocations and repeats on a clinical diagnostic level. However, due to its novelty and unproven track record in the clinics, we sought to validate the method of identifying large SVs in a cohort of patients diagnosed with Duchenne muscular dystrophy (DMD) (Table 4-3). These patients were known to carry multiple exon deletions or insertions in the *Dystrophin* gene as well as a large inversion that disrupted the *DMD* gene function. We used Bionano's Irys next generation mapping system to validate its capability of identifying large, previously known SVs in DMD cases.

We performed next-generation mapping on cases diagnosed with DMD as well as on cases where carrier status for the mutation was a possibility, such as in mothers of affected males. Our validation cohort size comprised of a total of 9 affected DMD individuals (6 with deletions, 2 with insertions and 1 with an inversion) and 4 biological mothers, some of whom were known carriers of pathogenic *DMD* variants (Table 4-3).

Long DNA molecules were represented well throughout the genome, except in locations with few unique sequences such as centromeres, acromeres and long arm of Y chromosome (Figure 4-4). The size of the insertions and deletions that can be identified is dependent upon the length of the labeled DNA molecules and content of fragile sites, where endonuclease nicking sites are close to each other on both strands of DNA resulting in double stranded DNA break. Both of these issues will be addressed in a second version of the optiDNA preparation kits that are going to be released during the second half of 2017. In the new protocol DNA is stabilized in agarose, allowing for less manual manipulation, and a second endonuclease is used to allow for molecular representation of the regions where the first enzyme had double strand breaks due to fragile sites allowing for a better breakpoint resolution.

On average we were able to identify 1500 insertions, 750 deletions, 40 inversions and 20 translocations per single case in the DMD cohort (Table 4-4). A distribution of the lengths of identified SVs is shown in Figure 4-4. Most of the identified insertions and deletions in the DMD cohort fell under 50kb in size; however, when considering SVs greater than 50kb in size, more deletions were identified in comparison to insertions. To interpret those variants and identify rare pathogenic variants, we needed a reference database to filter out the common benign variants. For that we used Bionano's control database containing structural variant information from 144 healthy individuals. All variants seen in more than one individual from the control database, with at least 50% match in size of the structural variant, were filtered out. This filtration step allowed us to reduce the number of SVs to numbers allowing manual curation (Table 4-5). A control database becomes extremely important for providing a genetic diagnosis in

patients where one doesn't know the precise location on the genome where variants are associated with the disease of interest. Here, we showed that most of the structural variants are actually common and can be filtered out.

All previously known deletions in the DMD cohort were successfully identified using NGM (Figure 4-5). These cases had been clinically diagnosed using multiplex ligation dependent probe amplification (MLPA) that sequences all existing exons of *DMD* gene and identifies which exons are deleted. Unlike MLPA bionano's NGM technology is capable of identifying the intronic breakpoints in the gene providing a more accurate representation of which part of the genome is missing in DMD cases. Also from Figure 4-5 we can see the ability of the instrument to identify a range of exon deletions, from a single exon deletion to multiple. The resolution of the breakpoints is limited to endonuclease nicking site density in a given location; higher density provides more accurate estimates. With a single enzyme the resolution of breakpoints is between 3-5kb in size; however, it possible to gain higher accuracy with the use of a second endonuclease from 1.5 to 3kb.

We then tested whether NGM was capable of identifying not only hemizygous/homozygous deletions but also to determine carrier status of an individual by detecting heterozygous variants. We performed NGM using Irys in *DMD* duos (proband, mother) to determine if the mothers of the affected individuals were also carriers of the same variant. The first duo we examined (CDMD-1131-proband and CDMD-1132-mother) contained a large deletion in the *DMD* gene first identified in the proband by PCR and later confirmed by MLPA. The mother of this patient had been tested using array CGH to identify the possibility of the next child being affected with the

same disorder. The array CGH identified the same deletion in the mother that was present in the proband, which was also confirmed by MLPA. Our NGM testing of this duo confirmed the previous diagnosis of the proband, which we also confirmed by PCR (Figure 4-6A), and the carrier status of the mother (Figure 4-6B). We did not find a deletion in the mother of the second duo (who hadn't been tested clinically), indicating that the mother is not a carrier of the deletion (Figure 4-7).

In addition, two duos where exon 3-4 duplication had been identified by array CGH were tested. We identified this duplication in the both probands and it was also present in a heterozygous state in the mothers of affected males (Figure 4-8A,B). Since the carrier status of the mother hadn't been determined clinically, we validated the NGM finding by quantitative PCR (Figure 4-9).

One of the defining features of Bionano's NGM system is its capability to identify inversions, which cannot be detected with chromosomal microarrays. We had one patient diagnosed with DMD who carried a large 5.1Mb inversion in *DMD* gene that disrupted the protein starting from exon 38. The patient had undergone numerous tests in order to identify this 5.1Mb inversion. Standard tests for diagnosing DMD such as PCR sequencing for point mutations and MLPA for deletion/duplication analysis had been performed, both of which were negative. A muscle biopsy had been excised to perform RNA sequencing to identify if the mRNA had splicing issues. It was identified that mRNA was abnormally spliced, which suggested an inversion in the gene. Genome sequencing had then been performed to validate the RNA sequencing findings. Because genome sequencing is not a clinically validated test, a PCR was performed on a clinical basis and results reported to the patient and parents. We performed NGM on

this sample and were able to identify the 5.1Mb inversion in a fraction of the time and cost of previous tests (Figure 4-10).

Here, we have shown the capabilities of the NGM in identification of large pathogenic structural variants in a cohort of DMD patients with a clinical diagnosis. Next steps are to use this technology in cases where the location of the pathogenic variant is not known and try to find it. For that we have started to use this method on a number of disorders of sex development cases.

NGM in Disorders of Sex Development

We performed Bionano NGM on 9 families diagnosed with a variety of undiagnosed DSD conditions (Table 4-6) to identify the underlying genetic diagnosis in conjunction with genome sequencing.

Unlike in Duchenne Muscular Dystrophy, where the genetic etiology is well defined and limited to the DMD gene on the X chromosome, CNVs in known DSD genes explain only a very small minority of cases. Therefore, in DSD cases, we do not know where to look for a pathogenic structural variant. To perform the analysis we first filtered out all common variants using the healthy control database consisting of 144 individuals. As above, common variants were classified as being seen greater than once in the control database, with at least 50% overlap in the structural variant size or genome location breakpoint. Variant filtering based on frequency allowed us to significantly reduce the number of variants for manual curation (Table 4-7). The remaining variants were classified as rare or *de novo* based on the information available from the parents. To identify the likely causative variants, we used a DSD specific gene

list to filter structural variants based on gene overlap and proximity of the gene of interest to the structural variant. We were unable to find a structural variant with an effect on genes in the primary gene list that would explain the patients' phenotypes as a first-pass clinical test. However, because for most of these families we have both genome mapping and genome sequencing we will be looking at the combination of the two to possibly identify compound heterozygous variants that may be pathogenic. These result were expected because not many structural variants in the genome are known to be associated with DSD. Using this method we hope to identify novel regions in the genome that may be associated with DSD.

Discussion:

For many years in modern genetic diagnostic the primary focus has been on single nucleotide variants and exome sequencing. However, a large subset of the human genome is missed when performing only exome sequencing – all the protein non-coding regions are not sequenced. This leaves a large gap in data collection and interpretation. More recently, due to continuous drop in sequencing price more institutions around the world choose to perform genome sequencing that is geared towards sequencing most of the human genome. Although this is a major step forward over exome sequencing it still has limitations, in particular because of the innate limitations of next-generation sequencing in terms of fragment lengths and unavailability of tools we are not able to identify of large structural variants. To overcome these limitations we used Bionano's next generation mapping tools to look for larger structural variants in human genomes.

First, we used a cohort of patients diagnosed with Duchenne muscular dystrophy with a known structural variant in the *DMD* gene to validate the capability of the Irys system to accurately identify large deletions, insertions and inversions. We have had a 100% success rate in identifying the correct clinically diagnosed structural variants using NGM, indicating the clinical utility of the method. Second, we used NGM to identify large pathogenic variants in patients with disorders of sex development of unknown genetic etiology. Although in the case of DSD we were not able to provide a genetic diagnosis, additional research with combination of genome sequencing might reveal variants causing the patient's phenotype. NGM promises to help further our understanding of gene regulatory elements in the genome and how SNVs and SVs in these regions may affect gene regulation. For this reason we will in the future perform RNA sequencing on blood or skin-derived fibroblasts to see if any of the variants that we have identified outside of the coding regions of the genome, such as splice variants, intronic, intergenic and promoter variants have a deregulating effect of effect on gene expression.

With the implementation of the exome sequencing as a clinical test for genetic diagnosis we were able to significantly increase the rate of diagnosis. However, as technologies have evolved over the past several years we can now get much longer read lengths of DNA molecules and in turn have a more complete picture of the human genome both on a single base pair resolution and globally for larger structural variants with comparable prices. It is important to look at the locations of the genome where we have much less understanding than the coding parts because many of the disease causing variants may be located outside of exons. The interpretation of pathogenicity of

non-coding variants will present challenges; however, with larger databases and gene expression studies the scientific community will be able to solve many more previously genetically undiagnosed cases. Although data not shown here, we will utilize genome sequencing data available on 97 trios to complement the analysis of NGM as it crucial to have the most complete sequence of the genome to provide genetic diagnosis.

Figures and Tables:

Table 4-1: Primer sequences used for PCR and quantitative PCR for validation of structural variants identified in DMD cases

DMD Exon #	Forward primer	Reverse primer	Product Size	Annealing temperature
E2	tgcattttagATGAAAGAGAAGATG	aaaacggatttttaagatacacagg	162bp	60°C
E3	ttcaaaaggggataatcgtga	GCCTTCGAGGAGGTCTAGGA	152bp	60°C
E4	atgcctcacaggctctgttc	GCAGTGCCTTGTTGACATTG	182bp	60°C
E5	ccccttttctttaacagGTTGATT	catttgtttcacacgtcaagg	294bp	60°C
E44	ctttTACCTGCAGGCGATTT	caccccttcagaacctgatcttt	349bp	60°C
E45	catggggcttcatttttgtt	TGACAGCTGTTTGCAGACCT	377bp	60°C
E46	gccatgtttgtgtcccagtt	acCTTGACTTGCTCAAGCTTTT	349bp	60°C
E47	GCCAGGGAATTCTCAAACAA	gaagcacccaggaaacaaaa	427bp	60°C
E48	tCCTTTCAGGTTTCCAGAGC	CGTCAAATGGTCCTTCTTGG	387bp	60°C
E49	ttttccccagGAAACTGAAA	tagtccacgtcaatggcaaa	303bp	60°C
E50	GATCTGAGCTCTGAGTGGAAGG	ccaaagagaatgggatccag	126bp	60°C
E51	TCTGGTGACACAACCTGTGG	CACCATCACCCTCTGTGATTT	434bp	60°C
E52	tacagGCAACAATGCAGGAT	GCCTCTTGATTGCTGGTCTT	319bp	60°C
E53	GAATTCAGAATCAGTGGGATGA	CCTCCTTCCATGACTCAAGC	413bp	60°C
E54	tagCAGTTGGCCAAAGACCT	tcatggtccatccagtttca	356bp	60°C
E55	GAGAGGCTGCTTTGGAAGAA	aagcggaaatgcctgactta	200bp	60°C

Table 4-2: Genome sequence variant analysis steps

1	SNVs+INDELs within protein-coding region of known DSD genes
2	SNVs+INDELs outside protein-coding region (introns, promoters, UTRs) of known DSD genes
3	SVs involving known DSD genes
4	SNVs+INDELs within protein-coding region of genes not associated with DSD
5	SVs involving 1 or more genes not associated with DSD
6	*De novo* SNVs+INDELs outside the protein-coding region of genes not associated with DSD

Table 4-3: Cohort of patients diagnosed with Duchenne muscular dystrophy

Cases with deletions in *DMD* gene are highlighted with red, insertions with green, and inversion with purple. Bolded cases indicate unaffected mothers of DMD probands. The coverage is defined as the total amount of the data produced in base pairs divided by the genome size (3.2 in case of humans).

Sample ID	Identifier	Variant Type	Affected Exons	Coverage
CDMD-1003	Proband	Hemizygous deletion	Exons 46-51	75x
CDMD-1155	Proband	Hemizygous deletion	Exons 48-54	112x
CDMD-1156	Proband	Hemizygous deletion	Exons 49-50	89x
CDMD-1159	Proband	Hemizygous deletion	Exon 52	91x
CDMD-1131	Proband	Hemizygous deletion	Exons 45-51	123x
CDMD-1132	**Mother**	**Heterozygous deletion**	**Exons 45-51 (carrier)**	98x
CDMD-1157	Proband	Hemizygous deletion	Exons 46-51	85x
CDMD-1158	**Mother**	**Unknown**	**Unknown**	74x
CDMD-1161	Proband	Hemizygous duplication	Exons 3-4	73x
CDMD-1162	**Mother**	**Heterozygous duplication**	**Exons 3-4 (carrier)**	113x
CDMD-1163	Proband	Hemizygous duplication	Exons 3-4	91x
CDMD-1164	**Mother**	**Heterozygous duplication**	**Exons 3-4 (carrier)**	164x
CDMD-1187	Proband	Hemizygous inversion	Exons 38-end	89x

Table 4-4: Number of structural variants identified in DMD cases

Duchenne Muscular Dystrophy					
Sample ID	**Insertions**	**Deletions**	**Inversions**	**Translocations**	**Coverage**
CDMD-1003	1342	671	46	17	75x
CDMD-1131	1357	662	28	22	123x
CDMD-1132	1263	608	24	13	98x
CDMD-1155	1876	822	58	23	112x
CDMD-1156	1784	863	38	17	89x
CDMD-1157	1444	690	30	21	85x
CDMD-1158	1315	646	22	10	74x
CDMD-1159	1765	778	32	16	91x
CDMD-1161	1759	814	42	20	73x
CDMD-1162	1796	826	40	24	113x
CDMD-1163	1892	891	54	23	91x
CDMD-1164	1839	845	48	30	164x
CDMD-1187	1783	790	49	29	89x
Average	1586	749	40	20	104

Table 4-5: The remaining number of structural variants identified in DMD cases after filtering of common structural variants using a control database

Duchenne Muscular Dystrophy					
Sample ID	**Insertions**	**Deletions**	**Inversions**	**Translocations**	**Coverage**
CDMD-1003	21	45	2	3	75x
CDMD-1131	15	38	3	3	123x
CDMD-1132	20	25	1	1	98x
CDMD-1155	26	40	1	3	112x
CDMD-1156	23	58	0	0	89x
CDMD-1157	20	29	0	0	85x
CDMD-1158	36	64	0	0	74x
CDMD-1159	23	49	0	1	91x
CDMD-1161	18	45	0	0	73x
CDMD-1162	32	33	1	1	113x
CDMD-1163	30	55	2	7	91x
CDMD-1164	17	37	1	5	164x
CDMD-1187	20	38	1	3	89x
Average	24	43	1	2	104

Table 4-6: The cohort of DSD patients undergone NGM analysis

Family ID	Identifier	Diagnosis
IX-1220	Trio	46,XY Gonadal Dysgenesis
SEA-3073	Trio	46,XX Female, Ovotestis
ORD-9004	Trio	46,XY Female
ARB-2055	Trio	46,XY Mixed Gonadal Dysgenesis
ARB-2060	Trio	46,XY Anorchia
ARB-2064	Trio	46,XY AIS-working
CVG-6033	Trio	46,XY Hypospadias, UDT
CVG-6034	Duo	46,XX DSD
CVG-6035	Trio	46,XY DSD

Table 4-7: Number of structural variants identified by NGM in the DSD cohort after filtration of common variants from a control database containing 144 healthy individuals Data missing in cases in which structural variation has not been called yet.

Family ID	Deletions	Insertions	Inversions	Translocations	De Novo
IX-1220	81	64	3	15	1
SEA-3073	60	54	3	12	1
ORD-9004	72	62	5	17	3
ARB-2055					
ARB-2060	85	62	2	10	2
ARB-2064					
CVG-6033					
CVG-6034					
CVG-6035					

Figure 4-1: DNA Nicking, Labeling, Repairing and Staining

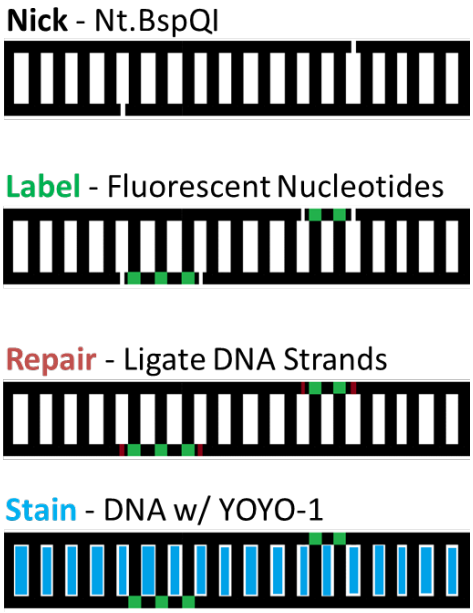


Figure 4-2: Irys Chip Nanochannel Structure and DNA Loading

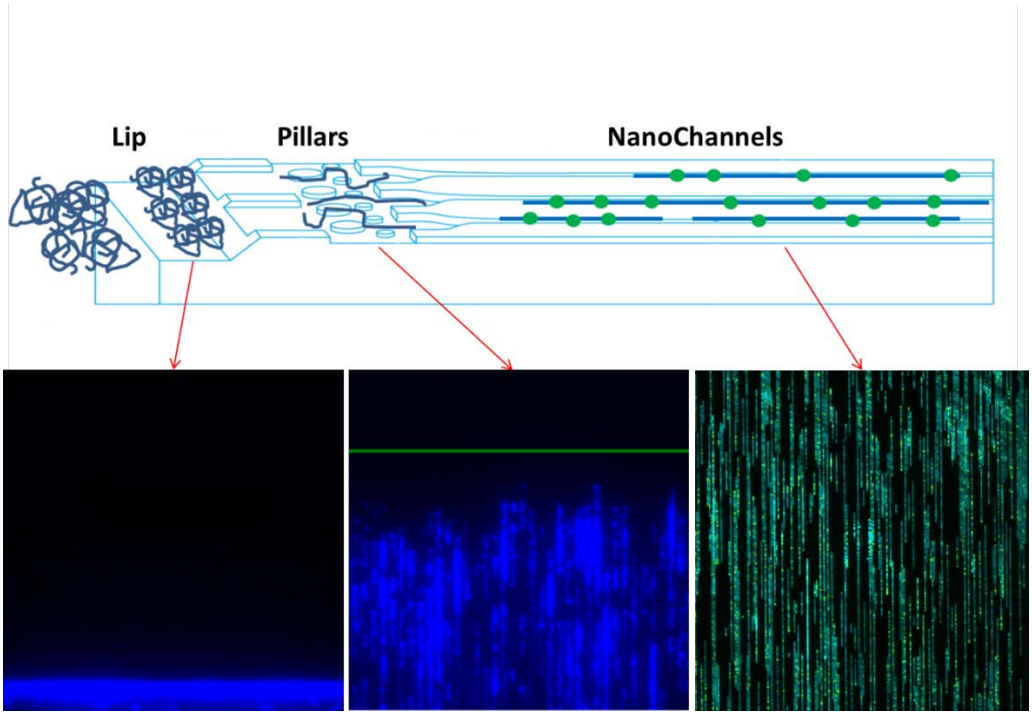


Figure 4-3: Next generation mapping genome coverage.

Vertical blue lines represent areas of the chromosome where long DNA molecules have been aligned. Below the blue lines are shown where traditional chromosome banding align.

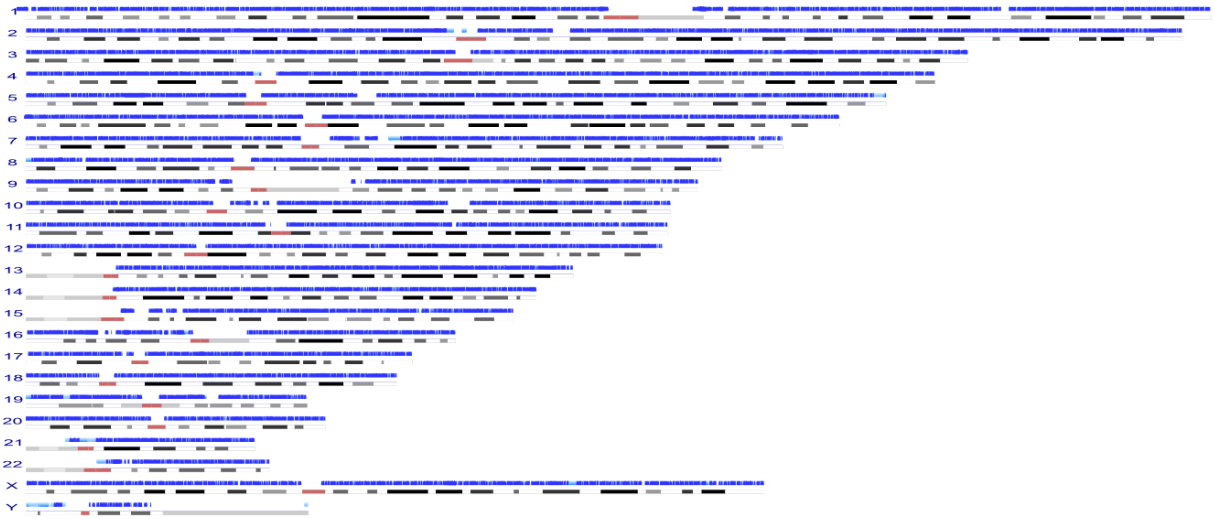


Figure 4-4: Genome-wide distribution of insertions and deletion in the DMD cohort (probands and parents)

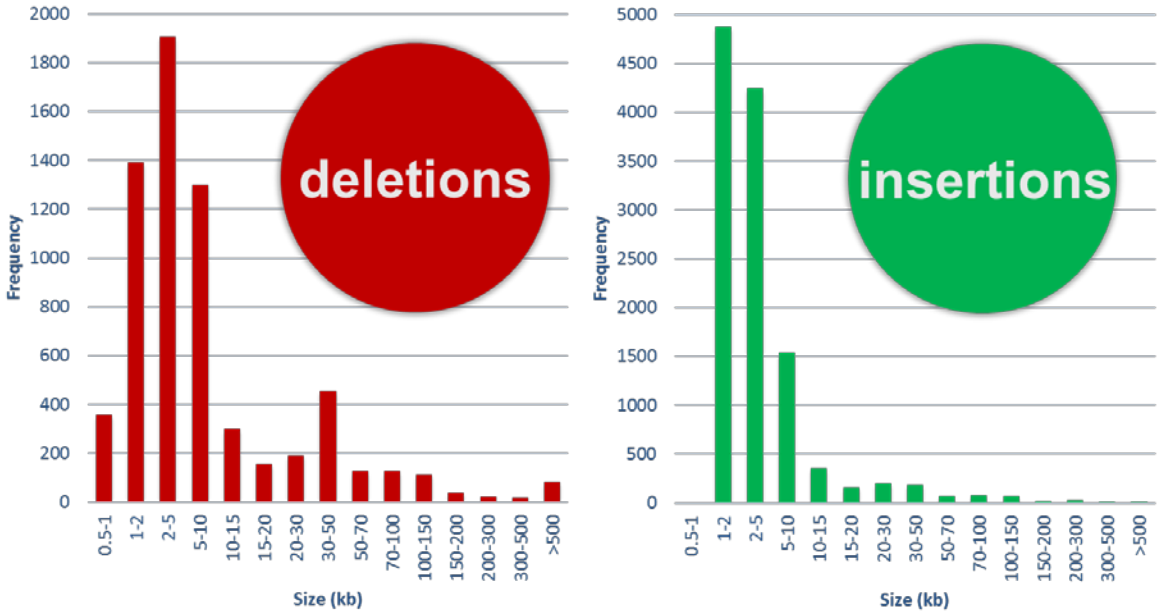


Figure 4-5: Deletions identified in DMD singletons

For each case, the green bar represents the reference X chromosome. The blue bar represents the contig map generated based on long molecule assembly of the patient's genome. The black vertical lines indicate endonuclease cut site. The lines between reference and assembled map show alignment of the two maps. The red area indicates deletion where reference (green) endonuclease sites are missing from the assembled map (blue).

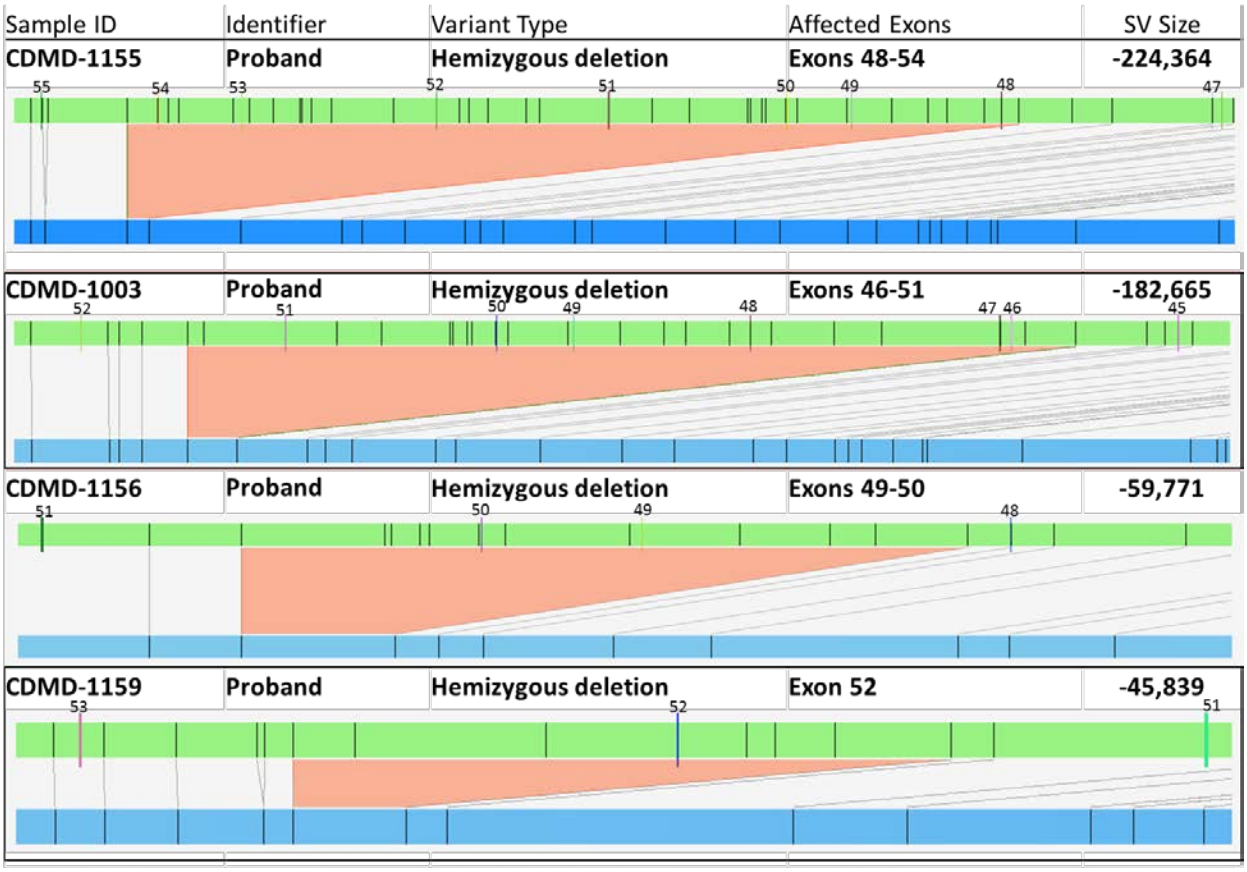


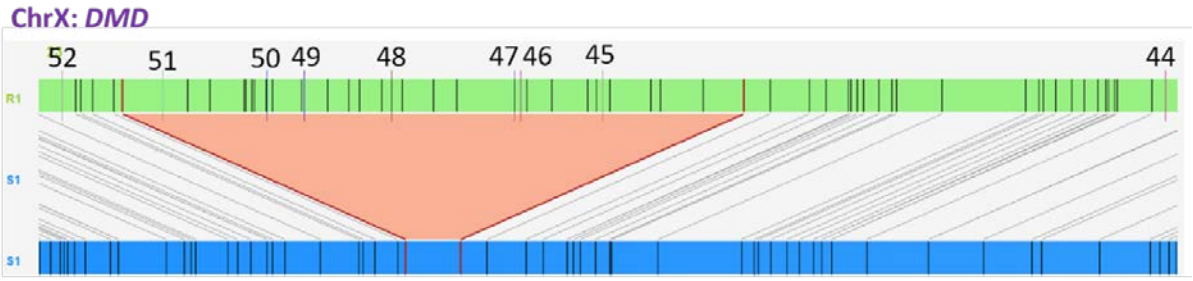
Figure 4-6: A. *DMD* deletion identified in an 46,XY muscular dystrophy patient

Top: Ref-seq locations on the X chromosome indicating possible size of the deletion based on MPLA and size identified by Irys

Middle: Visual representation of the deletion

Bottom: Confirmation of the deletion of exos 45-51 by PCR

Sample ID	DMD Deletion	Ref-Seq Location ChrX	Size (bp)
CDMD-1131	Exon 51-45	31,792,076-31,986,631	-194,555 to -487,167
Irys-1131	Exon 51-45	31,774,324-32,049,008	-250,092



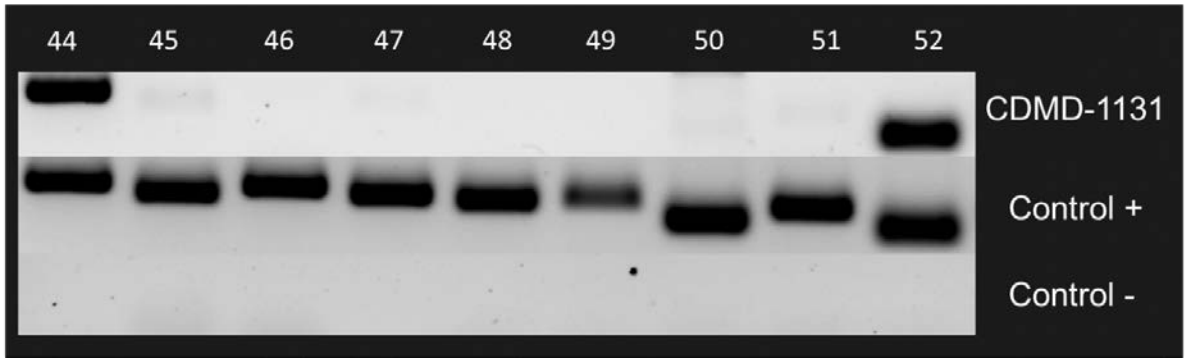


Figure 4-6: B. *DMD* heterozygous deletion identified in a 46,XX mother of a DMD patient

Top: Ref-seq locations on the X chromosome indicating possible size of the deletion based on MPLA and size identified by Irys

Bottom: Visual representation of the deletion, showing one allele is wild type (top blue line) and one of the alleles has the deletion present in the proband (bottom blue line). Individual molecules represented by single orange lines on the bottom with black dots indicate nicking sites.

Sample ID	DMD Deletion	Ref-Seq Location ChrX	Size (bp)
CDMD-1132	Carrier of E51-45	31,792,076-31,986,631	-194,555 to -487,167
Irys-1132	Exon 51-45	31,774,324-32,049,008	-249,994

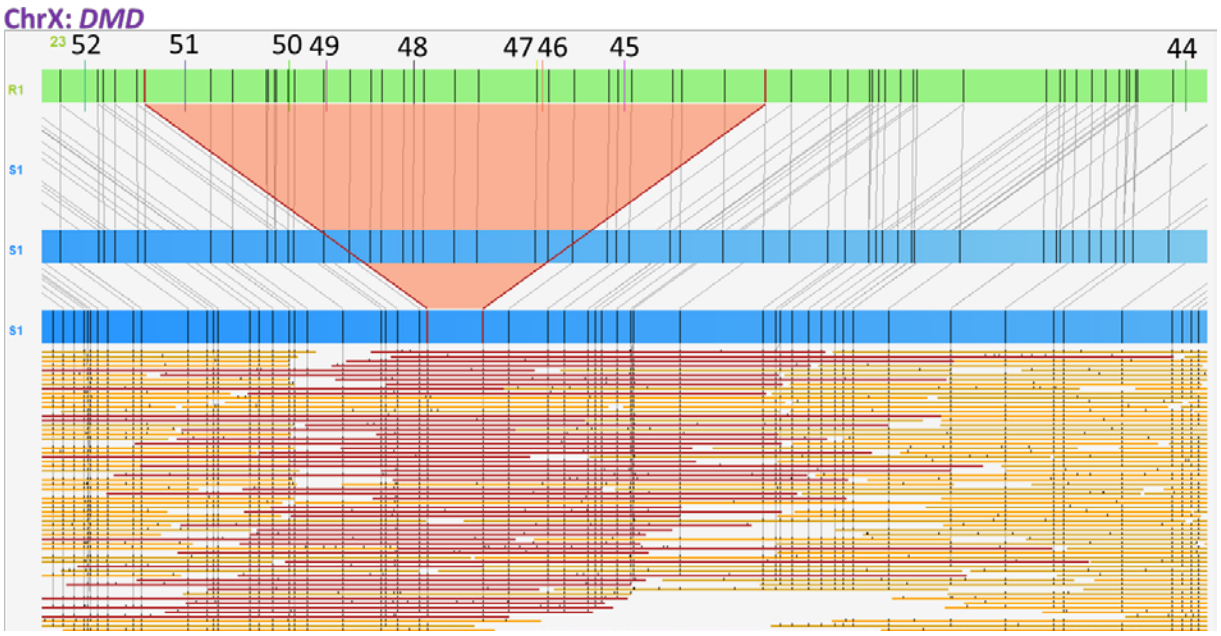


Figure 4-7: No *DMD* gene deletions identified in a mother (CDMD1158, blue) with an affected son. In this particular case several contigs (blue) are needed to cover the entirety of *DMD* gene, hence several sample maps (blue) shown in the image.

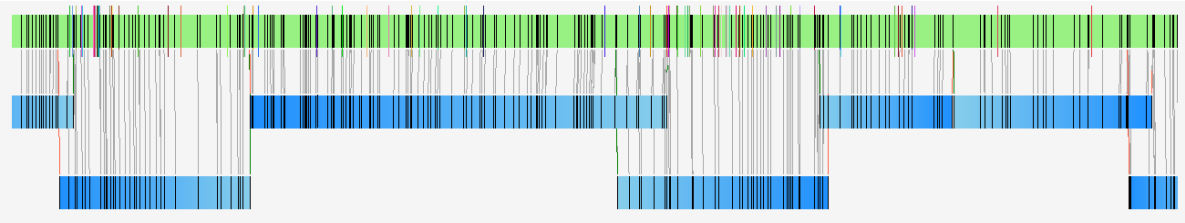


Figure 4-8: A. Insertion identified in a DMD patient. (Color conventions are the same as in Fig. 4-6)

Sample ID	DMD Duplication	Ref-Seq Location ChrX	Size (bp)
CDMD-1163	Exon 4-3	32,862,899-32,867,937	+5,038 to +196,751
Irys-1163	Exon 4-3	32,862,504-32,908,380	+12,968

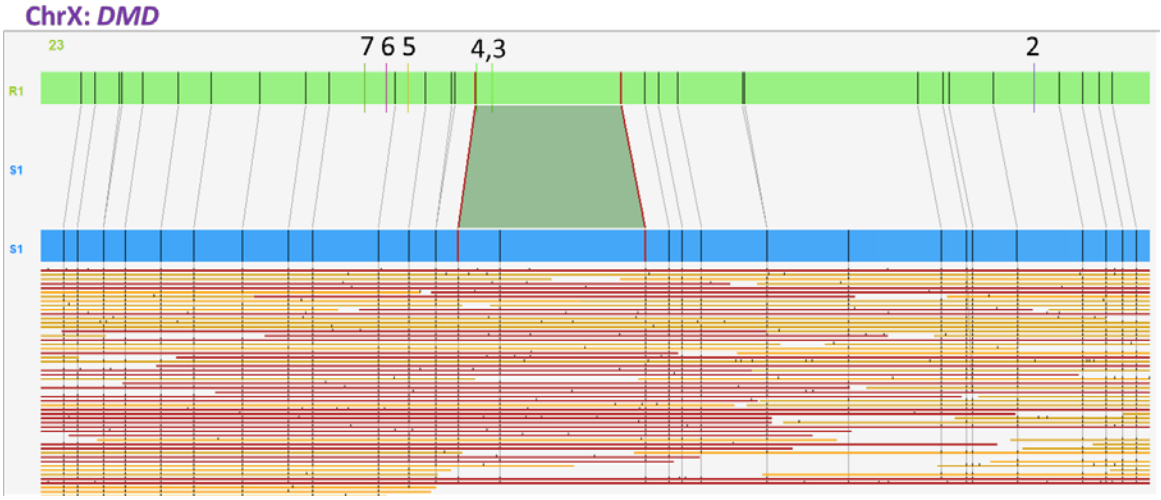


Figure 4-8: B. A heterozygous insertion identified in a mother of a DMD patient. (Color conventions are the same as in Fig. 4-6)

Sample ID	DMD Duplication	Ref-Seq Location ChrX	Size (bp)
CDMD-1164	Carrier of E4-3	32,862,899-32,867,937	+5,038 to +196,751
Irys-1164	Exon E4-3	32,862,504-32,908,380	+12,857

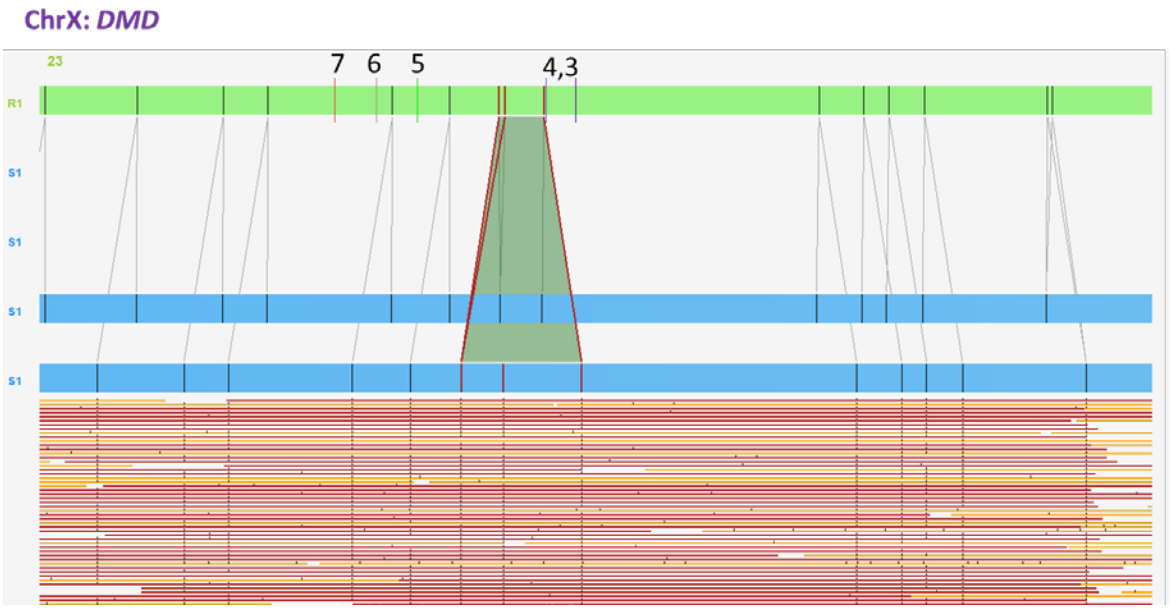


Figure 4-9: Quantitative PCR validation of the *DMD* exon 3-4 duplication in the mother of patient CDMD-1163. Exons 2 & 4 are present in single copies on X chromosome, amounts in the proband (red). The mother (CDMD-1164), who has 2 X chromosomes, has 2x more DNA (green). Exons 3 & 4 are duplicated in the proband and mother DNA amount in proband (2 copies) is approximately 1.5x lower than in mother (3 copies) indicating duplication is present in both.

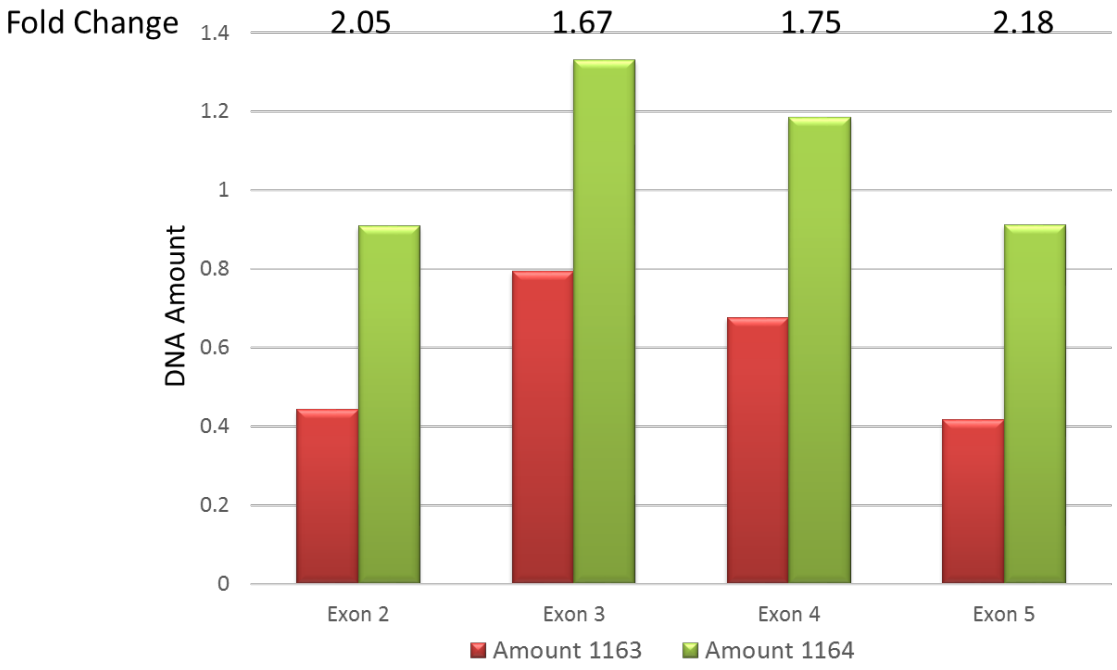
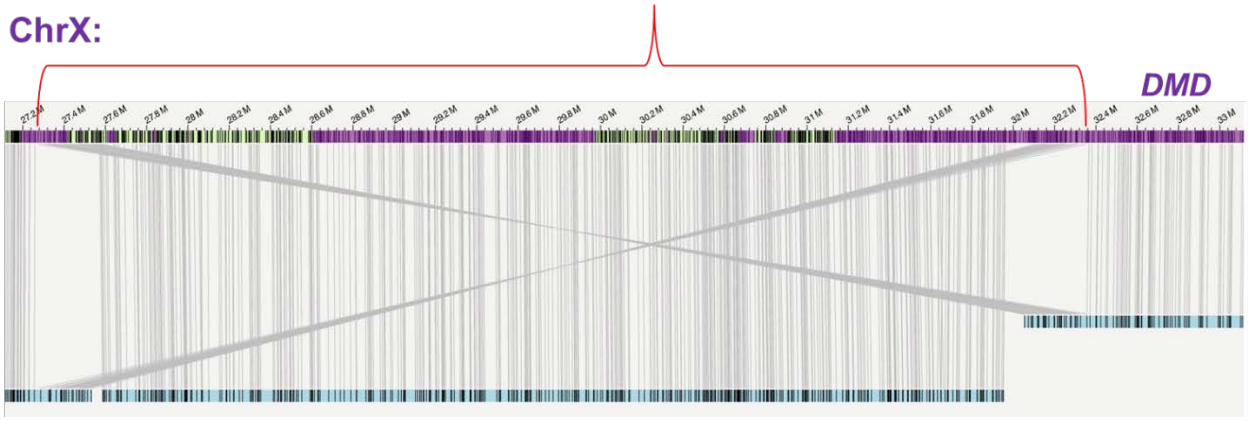


Figure 4-10: 5.1Mb inversion disrupting the protein function identified in DMD patient. The top row of the table shows the Ref-Seq locations of the inversion identified by genome sequencing. The bottom tow shows the location and size identified by Bionano NGM. Purple highlights on the reference show the locations of the genes, including *DMD.* The two lines across show that part of the sample map (blue) is aligning to the references in the opposite orientation, indicating and inversion of the entire middle section.

Sample ID	DMD Duplication	Ref-Seq Location ChrX	Size
CDMD-1187	Inversion	27,265,833-32,374,557	5.1Mb
Irys-1187	Inversion	27,260,000-32,360,000	5.1Mb



References:

1. Alex R Hastie, E.T.L., Andy Wing Chun Pang, Luna Xinyue Zhang, Warren Andrews, Joyce Lee, Tiffany Y Liang, Jian Wang, Xiang Zhou, Zhanyang Zhu, Thomas Anantharaman, Zeljko Zdzakula, Sven Bocklandt, Urvashi Surti, Michael Saghbini, Mike Austin, Mark Borodkin, R. Erik Holmlin, Han Cao, *Rapid Automated Large Structural Variation Detection in a Diploid Genome by NanoChannel Based Next-Generation Mapping.* preprint, bioRxiv, 2017.
2. Untergasser, A., et al., *Primer3--new capabilities and interfaces.* Nucleic Acids Res, 2012. **40**(15): p. e115.
3. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform.* Bioinformatics, 2010. **26**(5): p. 589-95.
4. Karakoc, E., et al., *Detection of structural variants and indels within exome data.* Nat Methods, 2012. **9**(2): p. 176-8.
5. Krumm, N., et al., *Copy number variation detection and genotyping from exome sequence data.* Genome Res, 2012. **22**(8): p. 1525-32.
6. Kircher, M., et al., *A general framework for estimating the relative pathogenicity of human genetic variants.* Nat Genet, 2014. **46**(3): p. 310-5.
7. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.
8. Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.* Nat Protoc, 2009. **4**(7): p. 1073-81.

Chapter 5

Conclusions

Exome/Genome Sequencing

DNA sequencing has become a routine procedure in clinics across the United States and around the globe where physicians order either exome or single gene sequencing to identify pathogenic variants leading to patient's phenotype and provide the appropriate care. The increased prevalence of exome and genome sequencing in clinical practice will improve the diagnostic yields of clinical next-generation sequencing. As more genomes are sequenced, associated with detailed phenotyping, our understanding of phenotype/genotype correlation will progress, allowing for a greater diagnostic yield from exome sequencing for single-gene or complex diseases.

These considerations highlight the need for deep phenotyping and better genotype/phenotype correlation in DSD and other disorders. To this end, translational research networks will be crucial to generate evidence for and promote best practices in clinical care to achieve best possible health outcomes and quality of life for patients and their families. In particular the DSD-TRN has created the first DSD registry in the US to collect standardized clinical practice data and standardized measurements in all fields involved in clinical care for DSD families: genetic, psychosocial, endocrine, anatomy/surgery, etc. Data entered into the registry from all the network's sites will be analyzed to try to derive new genetic etiologies for DSD, refine genotype/phenotype correlations and, eventually offer evidence-based recommendations for optimal practice.

Genome Mapping

For many years in modern genetic diagnostic the primary focus has been on single nucleotide variants exome/genome sequencing. However, a large subset of the human genome is missed when performing only genome/exome sequencing – all the protein non-coding regions are not sequenced. This leaves a large gap in data collection and interpretation. More recently, due to continuous drop in sequencing price more institutions around the world choose to perform genome sequencing that is geared towards sequencing most of the human genome. Although this is a major step forward over exome sequencing it still has limitations, in particular because of the innate limitations of next-generation sequencing in terms of fragment lengths and unavailability of tools we are not able to identify of large structural variants. To overcome these limitations laboratories around the world will need to use next generation mapping tools to look for larger structural variants in human genomes.

As DNA sequencing technologies have evolved and will continue to evolve the DNA sequence read lengths will become longer with higher quality enabling researchers and clinicians to see the more complete picture of the human genome both on a single base pair resolution and globally for larger structural variants. It is important to look at the locations of the genome where we have much less understanding than the protein coding regions because many of the unknown disease causing variants may be located outside of exons. The interpretation of pathogenicity of non-coding variants will present challenges; however, with larger databases and gene expression studies the scientific community will be able to solve many more previously genetically undiagnosed cases.

The upcoming clinical genetic diagnostic practice should consists of next generation mapping and sequencing technologies as a first line test (after life threatening conditions have been ruled out by other tests). The turnaround times will continue to improve and data analysis highly automated. This method will prove to be the most effective in identifying genetic diagnosis and providing a path towards acceleration of improvement of patient care.

Identification of Novel Disease Genes

There is no doubt that both genome sequencing and genome mapping have a potential for high diagnostic ability. These techniques are going to become the core of modern clinical genetics. However, many variants identified by either of the techniques will initially be uninterpretable clinically and be known as variants of unknown significance. The utilization of animal models will become of outmost importance to study VUS to improve the interpretive gap between identified variants and allowing the identification of novel genes involved in human disease pathogenesis. We have demonstrated this method in the case of 46,XY DSD. Were by utilizing undervirilized B6-YPOS mice as a model for 46,XY DSD allowed the identification of 15 novel candidate genes involved in male sex development.