MOLECULAR STUDIES IN HORSES WITH *SRY*-POSITIVE XY SEX REVERSAL SYNDROME

A Thesis

by

ERICA FANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2011

Major Subject: Genetics

Molecular Studies in Horses with *SRY*-positive XY Sex Reversal Syndrome

MOLECULAR STUDIES IN HORSES WITH *SRY*-POSITIVE XY SEX REVERSAL SYNDROME

A Thesis

by

ERICA FANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Co-Chairs of Committee,	Terje Raudsepp
	Bhanu Chowdhary
Committee Member,	Penny Riggs
Intercollegiate Faculty Chair,	Craig Coates

December 2011

Major Subject: Genetics

ABSTRACT

Molecular Studies in Horses with *SRY*-positive XY Sex Reversal Syndrome.

(December 2011)

Erica Fang, B.S., Texas A&M University

Co-Chairs of Advisory Committee, Dr. Terje Raudsepp
Dr. Bhanu Chowdhary

Sex determination in mammals is regulated by the sex-determining region on the Y chromosome (*SRY*); the presence of *SRY* activates the male developmental pathway and suppresses the gene network necessary for female gonad development. Mutations in sex determination genes lead to various abnormal sexual phenotypes, including sex reversal syndrome in which the genetic and phenotypic sex do not match. Sex reversal syndrome has been reported in humans, mouse, and several domestic species. In horses, *SRY*-negative XY sex reversal syndrome has been well described and is caused by deletions on the Y chromosome. However, the molecular causes of the *SRY*-positive condition in horses and other mammals are not known.

This research investigated five horses affected with *SRY*-positive XY sex reversal syndrome. Sequencing of the coding exon region of the *SRY* gene in the five cases showed 99-100% alignment with the sequences of normal males. Genotyping of two closely related individuals with 46 normal male controls on an equine SNP50 Beadchip identified two statistically significant SNPs in a ~16 Mb region on the long arm of horse chromosome 3 (ECA3q). The region was analyzed using Gene Ontology (GO) and

Gene Relationships Across Implicated Loci (GRAIL) to select functionally relevant candidate genes for sequencing. Further analysis of the entire horse genome was done through array comparative genomic hybridization (aCGH), which investigated possible structural rearrangements, such as copy number variants (CNVs). Deletions of olfactory receptor genes were detected on multiple chromosomes and confirmed through quantitative real-time PCR (qPCR). A homozygous deletion on ECA29 in a region containing genes of the aldo-keto reductase gene family, known to play a role in interconverting sex hormones between active forms and inactive forms, was discovered in two sex reversed animals. The findings were confirmed through qPCR and fluorescence *in situ* hybridization (FISH), and experiments to define the specific breakpoints of the deletion through PCR have been initiated.

This research represents the first systematic search in the horse genome for mutations and CNVs related to sex determination. The findings contribute to better understanding of the molecular mechanisms of sex determination in horses and other mammals, including humans.

DEDICATION

This work is dedicated to my family.

ACKNOWLEDGEMENTS

I would like to thank my committee co-chairs, Drs. Raudsepp and Chowdhary, and my committee member, Dr. Riggs, for their guidance and support throughout the course of this research. Also, thanks to the individuals in our lab who have contributed to my understanding of horse genomics and cytogenetic laboratory techniques.

Thanks to my friends at Texas A&M University who have made my undergraduate and graduate experiences a memorable one.

Finally, thanks to my family for their unending support, encouragement, and love to make me a better person.

NOMENCLATURE

aCGH	Array comparative genomic hybridization
AIS	Androgen insensitivity syndrome
AKR1CL1	Aldo-keto reductase family 1, member C-like 1
AR	Androgen receptor
BAC	Bacterial Artificial Chromosome
bp	base-pair
BIO	Biotin
Cbln4	Cerebellin 4 precursor
CBX2	Chromobox homolog 2
CNV	Copy Number Variation
CNVR	Copy number variation region
CSMD1	CUB and Sushi multiple domains 1
DAX1	Dosage sensitive sex-reversal
DIG	Digoxigenin
DLRSD	Derivative log ratio standard deviation
DMRT1	Double sex and mab-3 related transcription factor 1
Fgf9	Fibroblast growth factor 9
FISH	Fluorescence *in situ* Hybridization
GATA4	GATA binding protein 4
GO	Gene ontology

GRAIL	Gene relationships across implicated loci
HMG	High-mobility group
IBD	Identity-by-descent
kb	kilobase-pair
Mb	Megabase-pair
MHC	Major histocompatibility complex
MNS1	Meiosis-specific nuclear structural 1
NIPAL1	NIPA-like domain containing 1
PAR	Pseudoautosomal region
PCR	Polymerase chain reaction
qPCR	Quantitative real-time PCR
Sf1	Steroidogenic factor 1
SNP	Single nucleotide polymorphism
SOX9	*SRY*-box 9 protein
SRY	Sex determining region on the Y chromosome
TDF	Testis-determining factor
TEX9	Testis expressed 9
WNT4	Wingless type MMTV integration site family, member 4
Wt1	Wilms' tumour suppressor gene 1

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

INTRODUCTION

Sex Determination and Sex Chromosomes

The mechanism that determines whether an individual is a male, female, or in some cases, a hermaphrodite, has been a topic of interest for centuries. Environmental theories were prevalent in ancient days; for example, in 335 B.C.E., Aristotle proposed that *if the heat from the male partner overwhelmed the coldness of the female partner, the child would develop as a male*. As science progressed, the invention of the microscope and other scientific tools revolutionized many biological fields, including genetics. In 1891, German biologist Hermann Henking observed a difference in chromosome numbers while studying sperm in wasps (*Pyrrhocoris apterus*): some had 11 chromosomes, and others had 12. Noticing that the additional twelfth chromosome looked different from the others, he named this mysterious chromosome the "X element" and hypothesized that it played some role in sex determination in insects (Henking 1891). About a decade later, American zoologist C. E. McClung showed that the "X element" or "accessory chromosome" was present only in half of the cells and proposed that it might influence sexual determination (McClung 1902).

This thesis follows the style of *Animal Genetics.*

Around the same time, one of McClung's former students, Walter Sutton, was studying meiosis in sperm cells in grasshoppers. Counting and measuring the size of each chromosome led him to conclude that gametes carry half of the genetic material in a zygote and that the merging of two gametes during fertilization forms chromosome pairs (Sutton 1902). Sutton's work set the basis for the chromosome theory of inheritance.

Shortly after Sutton published his work, sex chromosomes were discovered by American researchers Nettie Stevens and Edmund Beecher Wilson, who noted differences in chromosome patterns between male and female beetles (*Tenebrio molitor*) (Stevens 1905; Wilson 1905). Wilson noticed that one chromosome pair in the males was mismatched and assumed the larger chromosome was the "X element" or the "accessory chromosome" known today as the "X chromosome", while the smaller, crumpled partner was subsequently named the "Y chromosome". His later work established chromosomal sex determination systems XX/XO and XX/XY where maleness was determined either by the lack of one X chromosome or the presence of the Y chromosome, respectively (Wilson 1905).

Additional analyses of sex chromosomes in other species over the next couple of decades revealed diversity in sex determination systems. Species of insects, reptiles and amphibians align with Aristotle's environmental hypothesis and are temperature dependent for sex determination; for example, in a certain species of moths (*Talaeporia tubulosa*), ambient temperature controls sex determination (Traut *et al.* 2007). The haplodiploidy sex determination system distinguishes the sex of insects such as

honeybees and wasps (van Wilgenburg *et al.* 2006). Many insects and nematodes utilize the XX/XO system in which the male has only one sex chromosome, as was seen in McClung's and Sutton's works. Research in birds, snakes, fish, and butterflies has revealed a ZZ/ZW system where females are heterogametic (ZW) and males are homogametic (ZZ) (Clinton 1998). Some extreme sex determination systems have multiple X and Y chromosomes or multiple Z and W chromosomes, for example, the X1-X5/Y1-Y5 system observed in monotremes (Watson 1990). Most mammals, including humans, have the XX/XY sex chromosome system where males are heterogametic (XY) and females are homogametic (XX) (Tijo & Levan 1956). There are exceptions in mammals, such as the Ryukyu spiny rats (genus *Tokudaia*) and several members of the genus *Microtidae*, in which the males lack a Y chromosome or the females and males have the same karyotype (Marchal *et al.* 2003), but generally, the XX/XY system in mammals differentiates females and males genetically.

Mammalian Sex Chromosomes

The mammalian X and Y chromosomes are thought to have evolved from a pair of identical autosomes after one chromosome acquired an allelic variation beneficial to males (Muller 1914; Lahn & Page 1999). As the Y chromosome gained other genes beneficial to males, recombination between the two chromosome pairs became increasingly suppressed, and the Y chromosome gradually degenerated in size and gene content (Lahn & Page 1999). As a result, the two chromosomes diverged in sequence homology except for the pseudoautosomal region (PAR), which allows the two

chromosomes to pair during male meiosis (Burgoyne 1982; Freije *et al.* 1992). In contrast to the smaller Y chromosome, the X chromosome is larger, represents approximately 5% of the genome, contains about 1000 genes, and is evolutionarily well-conserved across mammalian species (Ross *et al.* 2005). Due to its degeneration, small size, and low gene content, the Y chromosome was thought to be degenerate and on the road to extinction. However, the completion of the human (Skaletsky *et al.* 2003) and chimp (Hughes *et al.* 2010) Y chromosome sequencing and the Y chromosome studies in other mammalian species, such as cats (Murphy *et al.* 1999), cattle (Liu *et al.* 2002), and horses (Raudsepp *et al.* 2004; Paria *et al.* 2011) refuted these ideas, providing a clearer understanding of the structure and organization of this underestimated sex chromosome. The mammalian Y chromosome, though small and gene-poor compared to the autosomes and the X chromosome, contains a number of genes, several of which are involved in spermatogenesis, male fertility, and sex determination (Rozen *et al.* 2003).

Sex Determining Region on the Y Chromosome (SRY)

In the 1950s, human studies in females with Turner syndrome (45,XO) and males with Klinefelter's syndrome (47,XXY) established that the Y chromosome determined maleness (Jacobs & Strong 1959). The search for the testis-determining factor (*TDF* in human or *Tdy* in mouse) on the Y chromosome lasted for several decades with numerous candidate genes being studied and discounted (Silvers *et al.* 1977; Page *et al.* 1987) until the early 1990s when the sex determining region on the Y chromosome (*SRY*) was

identified as the necessary and sufficient locus for male sex determination (Gubbay *et al*. 1990; Sinclair *et al.* 1990; Koopman *et al.* 1991). Experiments in transgenic mice proved that *SRY* acts as the dominant male determinant gene. The expression of *SRY* in the early embryo triggers and coordinates testes formation from the bipotential gonad and effectively suppresses the genes involved in the female developmental pathway (Gubbay *et al.* 1990). Subsequent research on the *SRY* identified several other genes involved in mammalian sex determination, including *SRY*-box 9 protein (*SOX9*) (Morais da Silva *et al.* 1996), steroidogenic factor (*Sf1*) (Parker *et al.* 2002), fibroblast growth factor (*Fgf9*) (Schmahl *et al.* 2004), and many others (Sekido & Lovell-Badge 2008; Marshall Graves 2008). Despite these achievements, the overall knowledge about the complex genetic mechanisms underlying sex determination and sexual differentiation beyond the role of *SRY* in mammals is limited due to many contributing genes and pathways. Consequently, very little is known about the underlying genetics and molecular mechanisms of abnormal sexual development, including sex reversal syndrome.

Genetic Regulation of Sex Determination

The *SRY* gene is directly related to the production of testes in mammals. Initially, the genital ridge in the embryo has the potential to develop into either ovaries or testes (Fig. 1). The pathway for the development of testes is determined by the activation of the transiently expressed *SRY* gene and effective down-regulation of genes that regulate the ovarian pathway, specifically β-catenin (Wilhelm 2007; Sekido & Lovell-Badge 2008; Marshall Graves 2008). Regulation of the *SRY* gene is controlled by factors such as Wilms' tumour suppressor gene 1 (*Wt1*) and *Sf1*; mutations in these regulatory factors lead to a decrease in testis markers, gonad degeneration, and decreased *SRY* expression (Luo *et al.* 1994; Hammes *et al.* 2001). The normal expression of *SRY* triggers the activity of *SOX9*, another *SRY*-box containing gene, which indispensably regulates the differentiation of the pre-Sertoli cells that lead to the production of testes (Kent *et al.* 1996). Once *SOX9*, is up-regulated, *SRY* is down-regulated, and the up-regulation of *SOX9* is maintained by several genes including *Sf1* (Sekido & Lovell-Badge 2008). Interestingly, while *SRY* and *SOX9* were discovered to be the "master" genes in mammalian male sex determination nearly two decades ago, their direct target gene, cerebellin 4 precursor (*Cbln4*), was identified only recently in mouse (Bradford *et al.* 2009).

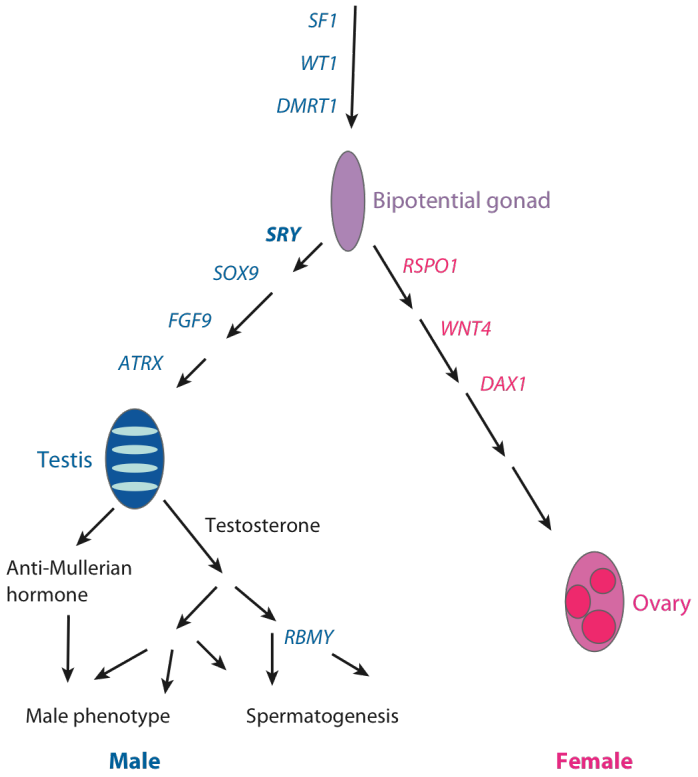


Figure 1. Mammalian gonad differentiation pathway. The bipotential gonad can differentiate into testis when *SRY* is present at the correct time and dosage. When *SRY* is not present, genes in the ovary pathway promote female development (Marshall Graves 2008).

High-mobility group (HMG) domains allow the *SRY* protein to bind and bend DNA after it has been transported from the cytoplasm into the nucleus, a dosage-dependent step critical for the activation of the testes-developing pathway (Kaur *et al*. 2010; Ross *et al.* 2008). The HMG box is the most conserved region of the *SRY* gene, but the regulation of the remaining *SRY* coding region and 5' sequences has been difficult to study due to the lack of conservation between species (Ross *et al.* 2008). A study done by Ross *et al.* (2008) used comparative genomic analysis of 5' sequences in human, bull, pig, goat, and mouse *SRY* to identify candidate regulatory regions that are homologous amongs those species and potential candidate regions to study sex reversal syndrome. Previous data show the presence of ten highly conserved DNA elements from the 5' region of the *SRY* gene in ten different mammalian species (Margarit *et al.* 1998). The regulation of transcription factors is crucial for the regulation of *SRY* expression to initiate the cascade of genes that control testes development.

Abnormal Sex Determination – Sex Reversal

Disruption of the early steps involving the bipotential gonad can result in abnormal gonadal development in both sexes. Mutations in factors that regulate *SRY* and *SOX9*, such as *Sf1* and *Wt1* have been linked to disorders, such as gonadal dysgenesis and sex reversal (Fleming and Vilain 2005; Lin *et al.* 2007). Sex reversal syndrome is a developmental disorder in which there is a discrepancy between the phenotypic and genetic sex. Both male-to-female and female-to-male sex reversal cases have been reported and studied in several species, such as mouse (Koopman *et al.* 1991),

human (German *et al.* 1978; Barbaro *et al.* 2007; Michala *et al.* 2008; Biason-Lauber *et al.* 2009; Hersmus *et al.* 2009), cattle (Ferrer *et al.* 2009), and horse (Kent *et al.* 1986). Female-to-male XX sex reversed humans are phenotypic males with a fragment of the Y chromosome in their genome, generally arising from an illegitimate recombination between the X chromosome and Y chromosome during meiosis where the *SRY* gene is translocated onto the X chromosome (de la Chapelle *et al.* 1984). Male-to-female XY individuals were first described in 1955 by G.I. Swyer; the condition, thus named Swyer syndrome, describes individuals who are phenotypically female, but the *SRY* on their Y chromosome may be mutated or deleted, although the majority of reported cases have a normal *SRY* gene (Swyer 1955). In humans, 10-20% of XY women are *SRY*-negative, and 80-90% have a normal *SRY* (Sarafoglou & Ostrer 2000; Michala *et al.* 2008). The molecular causes of this condition may be due to changes on the Y chromosome or other regions of the genome; the molecular heterogeneity of XY sex reversal make it difficult to pinpoint one cause.

Extensive studies in mice have shown that misexpression of *SRY* due to delayed onset or insufficient levels leads to partial or full sex reversal (Albrecht & Eicher 1997, Albrecht *et al.* 2003). Another study in mice showed that ablation of *M33*, which is involved in the regulation of *Sf1* and subsequently *SRY*, caused male-to-female sex reversal (Katoh *et al.* 1998). In recent years, extensive genome-wide molecular studies of sex reversal have been done in humans using Single Nucleotide Polymorphism (SNP) genotyping and Array Comparative Genomic Hybridizations (aCGH). Loss-of-function mutations have been found in the human orthologue of *M33*, chromobox homolog 2

(*CBX2*) (Biason-Lauber *et al.* 2009). A mutation in the human *SRY* gene, W70L, led to decreased *SOX9* expression, and therefore a lack of proper testis-determination coordination in development (Hersmus *et al.* 2009). Further studies in humans have also shown that gene dosage plays a critical role in sex determination; duplications of dosage sensitive sex-reversal (*DAX1*) on the X chromosome in humans result in XY individuals developing as females (Swain *et al.* 1998). Similarly, changes in autosomes can result in XY sex reversal. For example, the short arm of human chromosome 9 (HSA9p) contains genes related to gonadal development, and distal deletions in this region have been reported in 46,XY sex reversed patients (Barbaro *et al.* 2009). In order to examine the genome as a whole using a high-throughput method, other recent studies have utilized aCGH to examine copy number variations (CNVs) at a high resolution in 46,XY sex reversed patients. CNVs are structural rearrangements in the genome due to insertions, deletions, and duplications that result in a change in the effective DNA copy number. Recently, genome-wide aCGH analysis revealed structural changes in three regions that likely cause gonadal dysgenesis (White *et al.* 2011). These regions involve genes previously implicated in sex determination, *DAX1* and *SOX9*, as well as a deletion downstream of *GATA4*, which suggests the involvement of non-coding sequences in abnormal sexual development.

Taken together, molecular causes of XY sex reversal in humans are variable and involve single gene mutations, gene copy number variations, or other structural rearrangements. While relatively more information is available for male-to-female XY

sex reversal in mouse and human, little is known about the genetics of this condition in other species.

Sex Reversal in Horses

Sex reversal syndrome is of great concern in the horse (*Equus caballus*) because it affects mare and stallion fertility, an economically valuable trait in the equine industry. Additionally, male-to-female sex reversal among high-performance horses, such as trotters (http://www.nytimes.com/2009/09/19/sports/19racing.html; http://www.thehorse.com/ViewArticle.aspx?ID=14677), might give the carriers an advantage over normal females and result in disqualification.

Both female-to-male and male-to-female sex reversal conditions have been reported in several horse breeds (Milliken *et al.* 1995; Kent *et al.* 1986). Male-to-female XY sex reversal is the second most frequent sex chromosome abnormality after X chromosome monosomy; incidence of XY mares in cytogenetic analysis practice at Texas A&M University is as high as 26% among all chromosomally abnormal animals studied between 2001 to 2009. This is consistent with previous data showing that 12-30% of cytogenetic abnormalities in horses account for XY females (Bowling *et al.* 1987; Power 1990; Lear and Bailey 2008; Raudsepp *et al.* 2010) that are described as mares with the karyotype of a stallion (64,XY). Phenotypes of such animals range from feminine to highly-masculinized mares. The “feminine”-type animals have normal female external and internal genitalia and no somatic or behavioral abnormalities (Kent *et al.* 1986; Power 1986; Bowling *et al.* 1987). The “masculine”-type animals are often

categorized as male pseudohermaphrodites, showing stallion behavior and having abnormal genital tracts and gonads (Kent *et al.* 1986; Bowling *et al*. 1987; Kent *et al.* 1988; Howden 2004; Villagomez *et al.* 2009; Raudsepp *et al.* 2010). All cases thus far have been infertile except one that produced a normal 64,XX filly (Sharp *et al.* 1980). Furthermore, cases of both *SRY*-negative and *SRY*-positive have been reported in horses.

The horse *SRY* is a single copy intron-less *SRY* gene, which produces a 1,403 base-pair (bp) linear transcript (Hasegawa *et al.* 1999). The majority of mares with a stallion karyotype (64,XY) are *SRY*-negative due to various deletions on the Y chromosome. The deletions can range from 21 kilobase-pairs (kb) to a complete loss of the euchromatic region (Raudsepp *et al.* 2010). In contrast, the genetic causes of the *SRY*-positive condition remain elusive with only one previous report (Switonski *et al.* 2005).

Goals of This Study

This study builds upon the findings in humans showing that *SRY*-positive XY sex reversal is molecularly heterogeneous (German *et al.* 1978; Bernstein *et al.* 1980; Vaiman & Pailhoux 2000) and can be associated with both small (SNPs) (Kim *et al.* 2010) and large (CNVs) genomic variants (Tannour-Louet *et al.* 2010; White *et al.* 2011). These changes can affect sex determination genes on the sex chromosomes (e.g. *SRY*, *DAX1*) or autosomes (e.g. *SOX9*). In this study, we used a variety of genomic approaches to identify gene mutations and large-scale genomic rearrangements possibly associated with XY sex reversal in horses. The goal of the study was to identify putative

causes of sex reversal and facilitate the design of molecular tests for diagnostics and will contribute to better understanding of the mechanisms underlying sex determination and sexual development in horses and other mammals.

MATERIALS AND METHODS

Animals

This study involved five *SRY*-positive XY sex reversed mares referred to the Texas A&M University Molecular Cytogenetic Laboratory for cytogenetic testing due to their infertility and/or abnormal genitalia (Table 1).

Table 1. Summary of *SRY*-positive XY sex reversed females. Phenotypes of these affected horses are variable and cover a broad spectrum from "feminine"-type animals to "masculine"-type animals. Male pseudohermaphrodites are classified by abnormal genital tract development and gonads ranging from ovotestes to testicular feminization.

Animal	Breed	Phenotype	Karyotype	PCR with *SRY*
H169	Appaloosa	Normal external genitalia, hypoplastic uterus, underdeveloped mammary glands. Shows estrous behavior to stallions.	64,XY	Positive
H252	Mixed warmblood	Abnormal: uterus, vulva located ventrally between rear legs at the location of mammary glands.	64,XY	Positive
H348	Standardbred	Male pseudohermaphrodite.	64,XY	Positive
H369	Standardbred	Male pseudohermaphrodite.	64,XY	Positive
H423	Quarter Horse	Male pseudohermaphrodite with two testicles.	64,XY	Positive

DNA Isolation from Peripheral Blood

Peripheral blood from the cases was collected in sodium heparin and EDTA Vacutainers (VACUTAINER™; Becton Dickinson, Franklin Lakes, NJ, USA). Genomic DNA was isolated from the peripheral blood lymphocytes by chloroform-

phenol extraction (Birren *et al.* 1997). Briefly, an equal volume of 2X sucrose/Triton X-100 (pH 7.6) (Omnipur®, EMD Chemicals, Gibbstown, NJ, USA) lysis buffer was added to the peripheral blood samples to lyse the red blood cells and platelets. After a 20-minute incubation on ice, the cells were centrifuged at 2000 rpm for 20 minutes, and the supernatant was discarded. The pellet was re-suspended in 2X Sucrose/Triton X-100 with a 5-minute incubation on ice. The cells were centrifuged again at 2000 rpm for 20 minutes, and the supernatant discarded. The pellet was re-suspended in saline EDTA, made by dissolving 4.38 g of NaCl and 8.93 g of EDTA (disodium) in 1000 mL of Milli-Q water and with pH adjusted to 8.0 with 10 N NaOH, to lyse the nucleated white blood cells containing the DNA. Next, 1 mg of Proteinase K (Thermo Fisher Scientific, Inc., Waltham, MA, USA) and 250 μL of 20% SDS were was added to digest proteins in the cell and dissolve the cell membranes, respectively. Following an overnight incubation at 37ºC, an equal volume of phenol:chloroform:isoamyl alcohol (PCI; Roche Diagnostics, Mannheim, Germany) was added to separate the DNA from the carbohydrates and proteins. After centrifuging at 13,000 rpm for 5 minutes, the upper aqueous layer containing the DNA was transferred into a fresh tube, and the PCI extraction was repeated. The upper aqueous layer was mixed with an equal volume of chloroform:isoamyl alcohol (AMRESCO Inc., Solon, OH, USA) and centrifuged at 13,000 rpm for 5 minutes to remove all traces of phenol. The final aqueous layer was transferred to a fresh tube, and the DNA was precipitated with 2.5 volumes of ice cold 100% ethanol and 0.1 volumes of 5 M. NaCl. After incubating at room temperature for 15 minutes, the samples were centrifuged at 13,000 rpm for 10 minutes, and the

supernatant was removed. The pellets were washed in 70% ethanol, centrifuged at 13,000 rpm, air-dried, and re-suspended in ddH_2O. Quality of the DNA was checked on a Nanodrop spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, MA, USA) and by electrophoresis on an 1% agarose gel.

Chromosome Preparations and Karyotyping

Chromosome preparations for the five horses were obtained from short-term peripheral blood lymphocyte cultures using standard procedures (Raudsepp & Chowdhary 2008). Briefly, peripheral blood was collected in sodium-heparin vacutainer tubes, and 1 mL of plasma with buffy coat was added to 9 mL of prewarmed (37°C) cell culture media containing RPMI Medium 1640 with Glutamax and 25 mM. HEPES buffer (Gibco, Invitrogen, Gaithersburg, MD, USA), 30% fetal bovine serum (GEMINI BioProducts, West Sacramento, CA, USA), 1.4% antiobiotic-antimycotic solution (Gibco BRL, Life Technologies, Carlsbad, CA, USA), and 1% mitogen (pokeweed: lectin from *Phytolacca americana*, Sigma-Aldrich, St. Louis, MO, USA). The blood and cell culture media were mixed and incubated for 72 hours at 37°C. One hour before harvest, 100 μL of ethidium bromide (1 mg/mL) (Bio-Rad, Hercules, CA, USA) and 100 μL of demecolcine solution in Hanks' balanced salt solution (HBSS; 10 μg/mL) (Sigma-Aldrich, St. Louis, MO, USA) were added and incubated at 37°C. For the harvest, the tubes were centrifuged at 1000 rpm for 10 minutes, and the supernatant was aspirated leaving about 1 mL of medium at the bottom. The cells were re-suspended, mixed with prewarmed (37°C) hypotonic solution (0.075 M. KCl) (Rainbow Scientific, Windsor,

CT, USA) to a final volume of 10 mL, and then incubated for 30 minutes at 37°C. The tubes were centrifuged at 1000 rpm for 10 minutes, the supernatant was removed, and the cells were re-suspended in 5 mL of fresh fixative (methanol/glacial acetic acid in a ratio of 3:1) and centrifuged at 1000 rpm for 10 minutes. The fixative step was repeated three times, and after the last treatment, the supernatant was removed leaving about 200 μL of cells in the fixative, and 10 μL of the fixed suspension was spread on clean slides.

For karyotyping, the slides were stained with 5% Giemsa in Sørensen phosphate buffer (pH 6.8). The sex chromosomes were identified by C-banding (Arrighi & Hsu 1971) by treating the slides with 0.2 N HCl for 3 minutes, in 2.5% $Ba(OH)_2$ at 60ºC for 3 minutes, in 2X SSC at 60ºC for 5 minutes, followed by staining in 5% Giemsa solution for 1 hour. Microscopy was carried out using a Zeiss Axioplan2 fluorescent microscope equipped with karyotyping software Ikaros v5.2 (Metasystems GmbH, Altlussheim, Germany).

Polymerase Chain Reaction (PCR) with SRY *Primers*

The presence of the *SRY* gene was tested by PCR amplification of genomic DNA from each case with the following primers: SRY5.F 5'- TGC ATT CAT GGT GTG GTC TC-3' and SRY5.R 5'-GAA GCC GAA AAA TTG CCA T-3' (Raudsepp *et al*. 2010). The 10 μL PCR reactions contained 0.25 unit Taq polymerase (JumpStart RedTaq, Sigma-Aldrich), 50 mM KCl, 10 mM Tris-HCl (pH 8.4), 0.2 mM dATP, dCTP, dTTP, and dGTP, 0.3 μM of each primer, and 50 ng of genomic DNA as a template. The DNA was denatured at 95°C for 30 seconds, annealed at 60°C for 30 seconds, extended at

72°C seconds, followed by amplification for 30 cycles using the following program: 94°C for 30 seconds, 58°C for 30 seconds, 72°C for 30 seconds. The final extension was at 72°C for 10 minutes. The PCR products were resolved on 2% agarose gels containing ethidium bromide.

Sequencing

Primers were designed for four overlapping fragments in a 1,955 bp region containing the *SRY* exon using Primer 3 v0.4.0 software (Rozen & Skaletsky 2000) (Table 2, Fig. 2). The four fragments were amplified by PCR in all five *SRY*-positive XY sex reversed females and a normal male Thoroughbred (*Bravo – the DNA donor of the horse genomic bacterial artificial chromosome (BAC) library CHORI-241*); the DNA was denatured at 95°C for 30 seconds, annealed at 60°C for 30 seconds, extended at 72°C seconds, and then amplified for 40 cycles using the following program: 94°C for 30 seconds, 58°C for 30 seconds, 72°C for 2 minutes. The final extension was at 72°C for 10 minutes. The PCR products were purified in QIAquick Spin columns (QIAGEN, Valencia, CA, USA) and subjected for direct sequencing using BigDye chemistry (Applied Biosystems, Carlsbad, CA, USA). Briefly, a 10 μL sequencing reaction contained 2 μL Big Dye (Applied Biosystems, Carlsbad, CA, USA), 2 μL Half Big Dye (Applied Biosystems), 1 μL of 20 μM forward or reverse primer, and the template DNA (6 ng/100 bp). The initial denaturation was at 94°C for 3 minutes, followed by a sequencing program of 50 cycles: 95°C for 30 seconds, 50°C for 20 seconds, 60°C for 4 minutes. The final extension was at 60°C for 10 minutes. The samples were cleaned in

Sephadex Spin-50 mini columns (BIOMAX Inc., Rockville, MD, USA) at 3,600 rpm for 3 minutes, dried in a vacuum centrifuge (Eppendorf AG, Hamburg, Germany) at 45°C and re-suspended in formamide. Sequencing reactions were resolved on an ABI-3730 capillary sequencer. The sequences were aligned with each other and the reference sequence (GenBank AC215855.2) using Sequencher v4.7 (Gene Codes) software.

Table 2. Primer sequences for *SRY* exon sequencing.

	Forward Primer 5'-3'	Reverse Primer 5'-3'	Product Size (bp)
Fragment 1	CGT TTG TGT TTG TTA ACG GCT A	ATC ACG AGA CCA CAC CAT GA	681
Fragment 2	GCA ATC CTG GCT CAC ATTT T	CCT GTT CGG CTC TAT TTT CGT	660
Fragment 3	AAC AGG AGC GTC ACA GCA G	CGC AAG GTA GCT GAA AGA CC	510
Fragment 4	GTT GCG TTT TGA TGG GAA AC	GGG TGT CCC ACT CCC ATA AT	468

CGTTTGTGTTTGTTAACGGCTAATAGAATTAGGTCATCTGTAGGCTCTGTGGTTTTAAGATCACACTCACATATATTCA
TATTGATAAAACAATTCATTAGTATCCGTCTGTGCTCCACCTGCATCCTTTCATTTATAGACCTAATAAAATAATAATG
ACAAAGTTTGTTTTGTATTATTTTAAGGAGGGGCAGAGCCTTCAGCGCTGGGGATTAGAAGTAGGGCACAGAAACAGTG
TGTTATTACTCTGTTATCAGGTTTTGCTTGAGAGTGGATAGGCTGGTTGGGCTTTGGCTGACGGCCCAGGACACGTTTG
AGGGGAGGAGTTGGGGGCGGAGAAATAAGTATTTTATTACAAAAGTTGTAGTAACAAAGCCTCTG**CTAGAATAACC**
TTGGAAAAGTAAGATAATTTTCCCAACGCTTTATCTTCGCATTTTGCTACCACCCTCCTCTTCA
ACGGTGCCATCTTAAGCTTCTGCTATGTCCAGAGTATCCAACAGCGATAATTACAGTCTAGCAG
GACAGCAACATACCGTTCTCGGCTCTGGGAGAACCTCATCCCTACTTTGGACGAGCAATCCTGG
CTCACATTTTCGGAGTGAAACAAGAGGAAATGGTAGAGAGAACGGCCAGGACCGTGTCAAACGA
CCCATGAATGCATTCATGGTGTGGTCTCGTGATCACAGGCGCAAGGTCGCTCTAGAGAATCCCC
AACTGCAAAACTCAGAGATCAGCAAGCGGCTGGGATGCCAGTGGAAAATGCTTACGGAAGCCGA
AAAATTGCCATTCTTCGAGGAGGCACAGAGACTACGGGCTATGCATCAAGAGAAATACCCGGAC
TATAAATATCGACCTCGTCGGAAGGCCAAGATGCCACAGAAAAGTGACAAACCGCTTCCCGCAG
ACTCCTCTTCTACACTGTGCGGGCAGGCGCACGTACACGTCGACGAGTGGTTGAACCCTTTCAC
ATTCACGGACGACTGTACTGAGGCCACACAGTCACAAACGGAGGAGCGGTTAAACCATTCGCAG
CCCGCGAACACAGCTAATTCGGCGCTGCAACAGGAGCGTCACAGCAGCACCGCAACCCTGCGTG
ACAATCGGGTAACGTTGGCTACGCAGACATACGCAGACGTTCCCTTTCACTGTAATTTACCCTC
CGGACTTTCTCACGGTGATTTTCCTTGATTTCCTTACTGGTGGCGGCAGAGTCCCTGTTCGGCT
CTATTTTCGTATTTCTTCTGGAGCCCCAGCGTATCAACACCAGTGAAATTTTGAGTTCCAAGGT
CAGCTGTTTTTCTGTTAATGGAACAGTTTGTAATCTAATTTTAGTCTTCCAGAGGTTGTCCTTT
AAATAGCGCTAAGCATATATTGATACACTAATAATCGCCAGCATAGATCACAGAATTCTAACTCT
TTATTTTAAATACTATAAGTCACAAACATAGTGGAGAGAAGCATGCACAAAATTATGCTATGGAAAAATTGTTGCGTTT
TGATGGGAAACTGCGTAGCGTCTACAGTAGTCCCTTCTCGACGGTCTGTTCATTGGTCTTTCAGCTACCTTGCGCGTCT
TGGTAAACTGCACATAATTAACGCATGGAAATGTAAATTCTTTCCTGTTGGATAGTTTATGTTACCTCCCCTCCCCGCT
GTGAGAAACCAGTCTTGCGGCCCTTTGCTATTCAATTCTTTTACTCCGCACAAAACTGAGAGCTTCGATGCGCTGTGAA
CCGGTTATATTCTCTATCTGGTAAAATAAAATGTTTATTTAACTCGTGATCATTTTGATATCCTTATGAGACTCATAGT
TAGTTTTGCCTGTTTTTCCCAGAACTTTCTTTTGACAATGGCAATGTACCTGTTTCCATTGCCGCCTAACTTTTAACTT
TACGGAGGAAAAAATAACGAATGCTGAGAATATGAACTGCTGAGATTATGGGAGTGGGACACCC

Figure 2. DNA sequence of *SRY* exon and PCR primers. Contig from ECAY showing the *SRY* exon in bold lettering and the primers for the four overlapping fragments used for sequencing. The first primer's forward and reverse sequences are in yellow; the second in green; the third in blue; and the fourth in purple.

Equine SNP50 Beadchip Genotyping

Animals H348 and H369 were closely related Standardbreds sharing a common grand-dam. Their genome sharing value, $\hat{p} = 0.1905$, which is the degree of relatedness between two individuals, was calculated by pairwise clustering based on identity-by-descent (IBD) using the PLINK Whole Genome Analysis Toolset software (Purcell 2009). They were selected for Equine SNP50 Beadchip (Illumina, Inc., San Diego, CA, USA) genotyping because of their relatedness and because whole-genome genotyping data was already available for 46 normal male Standardbreds (Drs. McCue and Mickelson; University of Minnesota; personal communication) unrelated to the two cases (the highest genome sharing value from pairwise clustering based on IBD from PLINK between a case and a control was $\hat{p} = 0.1248$). The Equine SNP50 Beadchip features 54,602 SNPs that are evenly distributed across 15 different horse breeds with an average 43.2 kb space between probes for all 31 autosomes and the X chromosome. Beadchip genotyping was carried out by GeneSeek Inc. using 1 μg of genomic DNA from each case. The data was analyzed using the PLINK software (Purcell 2009). The data was analyzed using a basic association test (chi-square test for allele frequencies), and the results were visualized in Manhattan plots using Haploview v4.2 software (Barrett *et al.* 2005). Regions of interest were analyzed using GO (Gene Ontology) from DAVID Bioinformatics Resources (Huang *et al.* 2009a, 2009b) and Gene Relationships Across Implicated Loci (GRAIL) (Raychaudhuri *et al.* 2009) to find functionally relevant candidate genes.

Candidate Gene Sequencing

A candidate gene, NIPA-like domain containing 1 (*NIPAL1*), was selected for sequencing and mutation discovery. Primers were designed for the first three exons of this gene using Primer 3 v0.4.0 software (Table 3), and the exons were sequenced in all five cases and a control (*Bravo*) using BigDye chemistry as described above. The sequences were aligned with each other and the reference sequence (GenBank XM_001495264.1) using Sequencher v.4.7 software.

Table 3. Primer sequences for *NIPAL1* sequencing.

	Forward Primer	Reverse Primer	Product Size (bp)
Exon 1	5'-ACC TGT CCG AGA CTC CTC AG-3'	5'-GGT CTC ATC GCT ACC CAG AC-3'	205
Exon 2	5'-TGT TTC AGT ACC ATC TAC GCA ACT-3'	5'-GCT CAA TCC TGG AAA CTG CT-3'	536
Exon 3	5'-CTT CGA AGC TTC TTC CTC CA-3'	5'-TAC AAG CAA ACA CTA TTC TCA GCA-3'	150

Array Comparative Genomic Hybridization (aCGH)

The five cases were analyzed using the Texas-Adelaide equine whole genome oligonucleotide tiling array (Qu *et al.* 2011) The array is comprised of 417,277 60-mer oligonucleotides distributed in gene containing and non-genic sequences of the horse genome. Average distance between the tiles on the array is 7.5 kb. Array comparative genomic hybridization (aCGH) was carried out using Agilent Technologies' (Santa Clara, CA, USA) aCGH protocol. Briefly, 1 µg of the case genomic DNA and 1 µg of

the reference (male Thoroughbred *Bravo*) genomic DNA were digested with restriction enzymes, *Alu I* and *Rsa I*, at 37°C for 2 hours to fragment the DNA to about 200-500 bp fragment sizes. The enzymes were inactivated at 65º for 20 min and cooled to 4ºC. The DNA fragments were differentially labeled with fluorescent dyes using random primers and exo-Klenow enzyme. The DNA from the cases was labeled with Cy5 fluorescent dye, which excites at 650nm and emits at 670nm; and the control DNA was labeled with Cy3 fluorescent dye, which excites at 550nm and emits at 570nm. The labeling reactions used the following program: 37ºC for 2 hours, 65º for 10 min, 4ºC hold. The labeled products were cleaned by adding 430 μL 1X TE buffer to each reaction and centrifuging twice through Amicon 30kDa filters (Millipore, Billerica, MA, USA) at 13,000 rpm for 10 minutes. The cleaned samples were collected by inverting the filters into fresh tubes and centrifuging at 1000 rpm for 1 minute, resulting in approximately 20 μL of the product. An additional 20 μL of 1X TE was added to the filter and collected by centrifuging at 1000 rpm for 1 minute. DNA and dye concentrations of the labeled DNA were measured by Nanodrop spectrophotometry, and the genomic DNA yield and dye specific activity values were calculated using the following equations: genomic DNA yield = [DNA concentration (ng/μL) x sample volume (μL)]/1000 ng/μg; specific activity = (pmol per μL of dye)/(μg per μL genomic DNA). Equal concentrations of the labeled case and reference DNA were mixed with 25 μL of horse Cot-1 DNA (1 mg/mL), 26 μL of 10X blocking agent, and 130 μL of 2X Hi-RPM buffer, and the reactions were hybridized to the array for 40 hours at 65ºC in a Sure Hyb chamber (Agilent Technologies). After hybridization, the array slides were disassembled in

Agilent wash buffer 1. The slides were washed in Agilent wash buffer 1 for 5 minutes with agitation, Agilent wash buffer 2 at 37ºC for 2 minutes, acetonitrile for 10 seconds with agitation, and stabilizing/drying solution for 30 seconds with agitation. All washing steps were carried out under stringent conditions to limit exposure to light, air, and ozone degradation. The array slides were scanned on an Agilent SureScan DNA Microarray Scanner with 3μm resolution and XDR = 0.05, which takes the lightest and darkest areas of the image to more accurately represent the probe intensity levels across the array. The data was extracted using Agilent's Feature Extraction software. CNV detection and analysis was carried out using Agilent's Genomic Workbench v6.0 software. Significant gains and losses were determined by $\log_2$-threshold cutoffs of +0.5 and -0.5, respectively, containing a minimum of five consecutive probes.

Quantitative Real-Time PCR (qPCR)

Two genomic regions containing CNVs in array CGH experiments were validated by qPCR using DNA from the five cases and five normal controls. The two CNV regions were a cluster of olfactory receptor genes on ECA6 and a cluster of aldo-keto reductase genes on ECA29. Primers were designed around the two probes with the most significant "loss" scores: one region around the probe with the greatest "loss" value and two regions around the probe with the next greatest "loss" value (Table 4). Primers were also designed for *AKR1CL1* in the aldo-keto reductase gene cluster (Table 5). All primers were designed using the Primer 3 v0.4.0 software. Primer efficiency was evaluated by creating a standard curve using a DNA sample from one of the cases and

primers for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as a reference gene. The qPCR reactions were carried out in duplicates using 50 ng of DNA from the five sex reversed horses and five controls (an Arabian, an Appaloosa, two Quarter Horses, and a Thoroughbred). Each reaction contained 10 μL of Light Cycler® 480 Sybr Green 1 Master Mix (Roche Diagnostics, Mannheim, Germany), 1 μL of 10 μM forward or reverse primers, 8 μL of ddH_2O, and 2 μL of DNA (50ng). Copy numbers were evaluated using qPCR on a Light Cycler® 480 (Roche Diagnostics, Mannheim, Germany). Results were analyzed by calculating $\log_2^{-\Delta Ct}$, and the p-value was calculated by performing Student's t-test; any p-value <0.05 was considered to be significant.

Table 4. Primer sequences for an olfactory receptor gene for qPCR. The regions on ECA6 were selected based on the $\log_2$ ratio of the probes. The 60-mer probes are contained in the products.

Region	Probe Score	Forward Primer 5'-3'	Reverse Primer 5'-3'	Product Size (bp)
1	-1.694	TCT CAT CCC TAC ACC TTC TGG	ATT GGA GGT GGG GTA GGA AG	249
2	-1.576	TTC CAC ATT AAT CAG CAG ATG G	TGC GGA TGC ACA TAA AGA AG	173
3	-1.576	CGC AAT GAA TGG AAG ATG AA	GAT CCA AAC AAC AGC AGC AA	228

Table 5. Primer sequences for *AKR1CL1* for qPCR and FISH

Exon	Forward Primer 5'-3'	Reverse Primer 5'-3'	Product Size (bp)
1	CCC CAG TGA CAG TGG ATT CT	GGT CCG AGA TCA AGG ATG AA	217
2	CAG CTT CCA GTC AGG GAA AG	TTT GCA TTC AAC CCT CCA GT	187
3	CTC CTC TCC GAG TCA GAT GG	CAG CAC CTT GGT CGT GTA GA	225

Screening the Equine Bacterial Artificial Chromosome (BAC) Library and BAC DNA Isolation

A massive deletion including a cluster of an aldo-keto reductase gene family on ECA29 in two of the sex reversed cases was further verified by fluorescence *in situ* hybridization (FISH) using BACs containing the genes of interest. The corresponding BACs were identified from the equine CHORI-241 BAC library (http://bacpac.chori.org/equine241.htm) by screening the library superpools and plate pools by PCR using primers for *AKR1CL1* (Table 5). The BACs were grown overnight at 37°C in 2YT media (Invitrogen, Carlsbad, CA, USA) containing chloramphenicol (30 mg/mL) and plated on LB agar (Invitrogen, Carlsbad, CA, USA) containing chloramphenicol. Clones were grown overnight at 37°C, single colonies were picked and incubated in 2YT media containing chloramphenicol for 8 hours. The identity of the single colonies was once again verified by PCR, which was carried out on cell lysates, and the final positive clones were grown in 100 mL of 2YT containing chloramphenicol at 37°C overnight. The BAC DNA was isolated using the Plasmid Midi Kit (Qiagen,

Valencia, CA, USA) following the manufacturer's protocol. Briefly, the 100 mL cultures were divided into two 50 mL tubes and centrifuged at 8000 rpm and 4°C for 30 minutes. The supernatant was discarded, and the pellet was re-suspended in 10mL of Qiagen P1 buffer containing RNase A (100 μg/mL). Next, the samples were incubated for 5 minutes in 10 mL of Qiagen P2 buffer on ice with constant inverting. Finally, the P2 buffer was neutralized by adding 10 mL of cold Qiagen P3 buffer and incubated on ice for 15 minutes with occasional inverting. The samples were centrifuged at 8000 rpm and 4°C for 30 minutes, the supernatant was filtered through 3 layers of coffee filters into fresh 50 mL tubes, and passed through Qiagen-tip 100 columns, which were equilibrated with Qiagen QBT buffer. The columns were washed with Qiagen QC buffer twice, and the DNA was eluted with 5 mL of Qiagen QF buffer (prewarmed to 65°C and added in 1 mL aliquots) into fresh 50 mL tubes. Equal volume of isopropanol was added to the 5 mL of eluted DNA for precipitation. The samples were centrifuged at 8000 rpm and 4°C for 30 minutes, and the supernatant was discarded. The DNA pellet was washed with 70% ethanol, transferred into 1.5 mL microcentrifuge tubes, and centrifuged at 13,000 rpm for 15 minutes. The pellet was washed with 70% ethanol, air-dried, and re-suspended in ddH_2O. DNA quality was checked bye Nanodrop spectrophotometry and electrophoresis on a 1% agarose gel.

Fluorescence in situ *Hybridization (FISH)*

Probe labeling, *in situ* hybridization, detection, microscopy, and image analysis were carried out according to the detailed protocol by Raudsepp and Chowdhary (2008).

Briefly, 1 μg of the BAC DNA was labeled with either biotin (BIO) or digoxigenin (DIG) using BIO- or DIG-nick translation mix (Roche Diagnostics), respectively. Following incubation at 15°C (1 hour 10 minutes for BIO; 1 hour 30 minutes for DIG), the labeling reactions were cleaned using Sephadex Spin-50 mini columns and checked on a 1% agarose gel. In order to bind non-specific repetitive sequences, 10 μL of unlabeled horse genomic DNA (1 μg/μL) was added to the labeled probe, concentrated in a vacuum centrifuge, and re-suspended in 3:7 water:hybridization master mix. The latter contained 70% deionized formamide (molecular biology grade, Sigma Aldrich, St. Louis, MO, USA), 14% dextran sulphate sodium salt (Sigma Aldrich), and 3X SSC. Slides with metaphase chromosome spreads were treated with RNase (100 μg/mL in 2X SSC) (Fisher Scientific) at 37°C for 1 hour, denatured in 70% formamide; 2X SSC (pH 7.0) at 70°C for 2 minutes, dehydrated in an ascending series of 70%, 80%, 90%, and 100% ethanol, and air-dried. The probes were denatured at 90°C for 12 minutes, pre-annealed at 37°C for 20 minutes, and applied to the slides under coverslips. Hybridizations were carried out in a moist chamber at 37°C overnight. The next day, the slides were washed three times in 50% formamide; 2X SSC at 40°C for 5 minutes each, three times in 4X SSC; 0.05% Tween-20 at room temperature for 2 minutes each with gentle shaking, and in 4X SSC at room temperature for 2 minutes with gentle shaking. Antibodies were diluted in 1X blocking solution (Vector Laboratories) and pooled together because biotin- and dig-labeled probes were co-hybridized on the same slide. For the first antibody layer, fluorescein-avidin D (avidin-FITC; Vector Laboratories) was used to tag biotin, and mouse antidig (Roche Biochemicals) was used to tag DIG.

Antibodies were incubated at 37°C for 30 minutes. The slides were washed three times in 4X SSC; Tween20 and once in 4X SSC as described above. For the second antibody layer, biotinylated antiavidin D (Vector Laboratories) was used to tag the avidin-FITC on BIO, and antimouse Ig-dig (Chemicon International, Billerica, MA, USA) was used to tag the antidig on DIG. After 30 minutes of incubation at 37°C, the slides were washed as described above. For the third layer, avidin-FITC was used to tag the biotinylated antiavidin D and add additional fluorochromes for better biotin signal detection, and antidig-rhodamine (Roche Biochemicals) was used to tag the antimouse Ig-dig for DIG signal detection. Lastly, the slides were mounted with DAPI-antifade. The results were analyzed with a Zeiss Axioplan2 fluorescent microscope using DAPI, FITC, and Cy3 fluorescent filters, and the images were captured using Isis v5.2 (MetaSystems GmbH) software. For each hybridization, 10-15 cells were analyzed.

PCR Refinement to Define the ECA29 Deletion Breakpoints

Defining the breakpoints of the massive deletion on ECA29 in H348 and H369 was initiated by designing primers flanking the deletion directly outside of the deletion and immediately inside the deletion start site and inside the deletion end site (Table 6, Fig. 3) using the Primer 3 v0.4.0 software. Repetitive sequences in the region present in other regions of the genome were masked using RepeatMasker (Smit *et al.* 1996-2010; http://www.repeatmasker.org/) to ensure unique primers that only amplify the region on ECA29. Genomic DNA from all five cases, a male control (*Bravo*), and a female control (*Twilight*) was amplified by PCR. The 10 μL PCR reactions contained 0.25 unit

Taq polymerase, 50 mM KCl, 10 mM Tris-HCl (pH 8.4), 0.2 mM dATP, dCTP, dTTP, and dGTP, 0.3 μM of each primer, and 50 ng of genomic DNA as a template. The DNA was denatured at 95°C for 30 seconds, annealed at 60°C for 30 seconds, extended at 72°C seconds, followed by amplification for 30 cycles using the following program: 94°C for 30 seconds, 58°C for 30 seconds, 72°C for 30 seconds. The final extension was at 72°C for 10 minutes. The PCR products were resolved on 1% agarose gels containing ethidium bromide.

Table 6. Primer sequences for defining the ECA29 deletion breakpoints. Primers were designed for the region immediately outside of the deletion and the regions immediately inside the deletion start and end site.

Region	Forward Primer 5'-3'	Reverse Primer 5'-3'	Product Size (bp)
OUT (Region outside of deletion)	ACC CCA CCC ACC CTA TAT GT	CAT ATG GGG GTG ACA GGA AT	226,069
DEL.IN1 (Region inside deletion start site)	CCC ATG GCT GTG GAA CTA GA	ATC ACC TGC CCT TCA CAC TC	180
DEL.IN2 (region inside deletion end site)	TGA TTT GCT TGG ATT TTC TGG	TGT GCA GAA AAA TAC CGA AGA TT	154

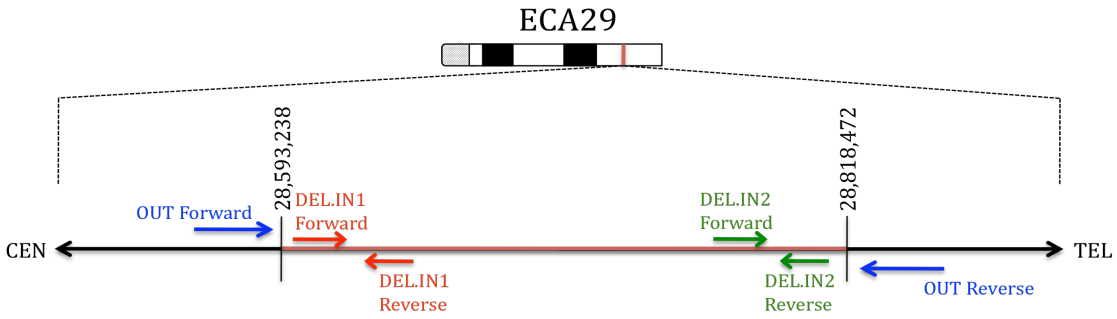


Figure 3. Primers designed around the ECA29 deletion site to define the breakpoints.

RESULTS

Cytogenetic Analysis and SRY *Testing*

Cytogenetic analysis of Giemsa stained metaphase spreads showed that the five cases had normal diploid chromosome number (2n=64), but the XY chromosome complement, which is typical of males (Fig. 4a, b). The results were confirmed by C-banding, which clearly showed the presence of one X chromosome and one Y chromosome (Fig. 4c). Amplification by PCR of the *SRY* gene was observed in all five animals, thus confirming the presence of the Y chromosome and the male sex determination gene (Fig. 5).

a)

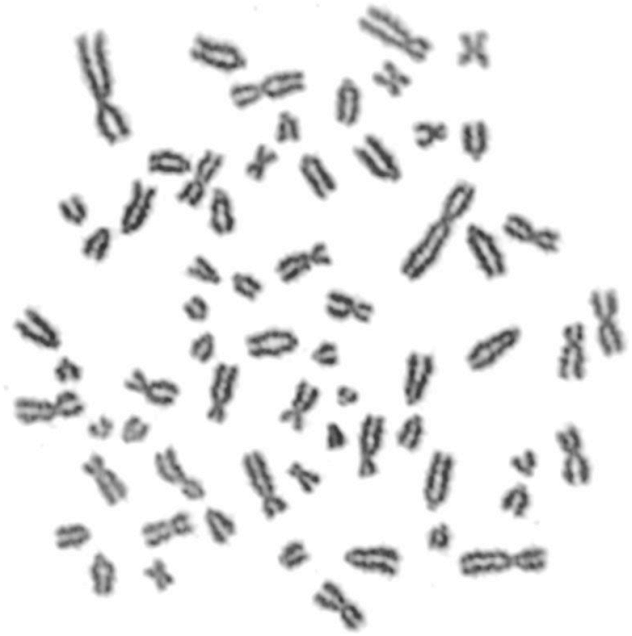
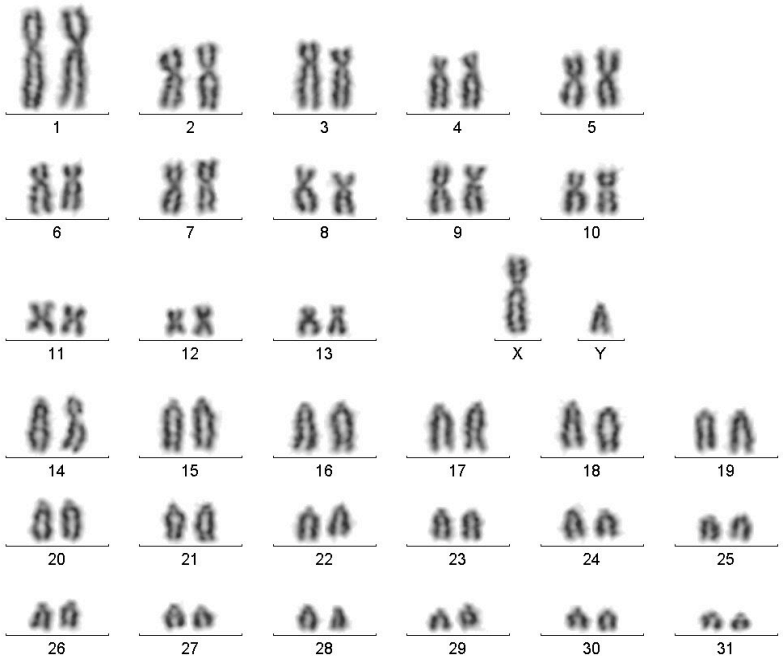


Figure 4. Cytogenetic analyses. A Giemsa stained metaphase spread (a), a karyotype (b), and a C-banded metaphase spread (c) of an *SRY*-positive XY sex reversed female horse.

b)



c)

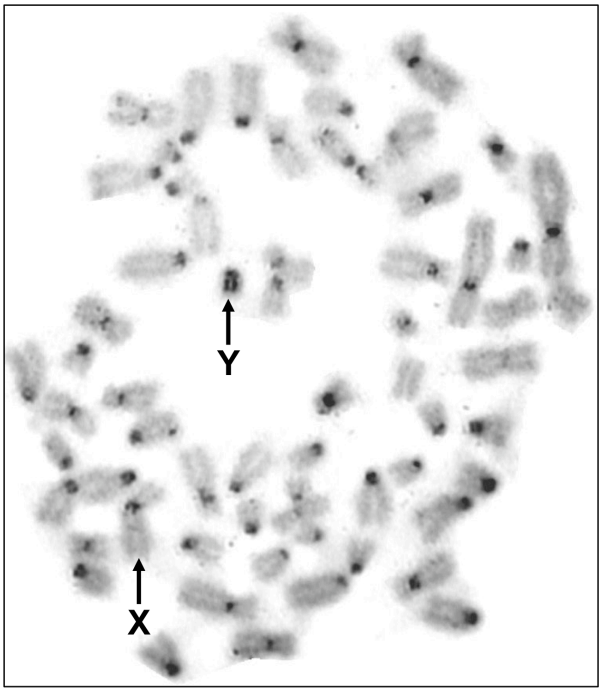
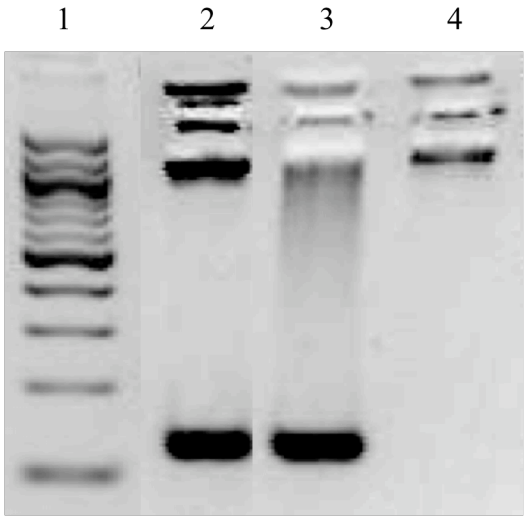


Figure 4. Continued



1. 100 bp ladder
2. XY female blood DNA
3. Male control
4. Female control

Figure 5. PCR gel image for *SRY* product. Gel image showing the presence of *SRY* PCR product (131 bp) in an XY female and male control but not in normal female control.

SRY *Sequencing*

Sequence analysis of the *SRY* exon and predicted promoter region revealed 99-100% similarity between the five cases, the normal male control, and the reference sequence (Fig. 6). These results effectively excluded mutations in the *SRY* gene as a possible cause of the sex reversal condition.

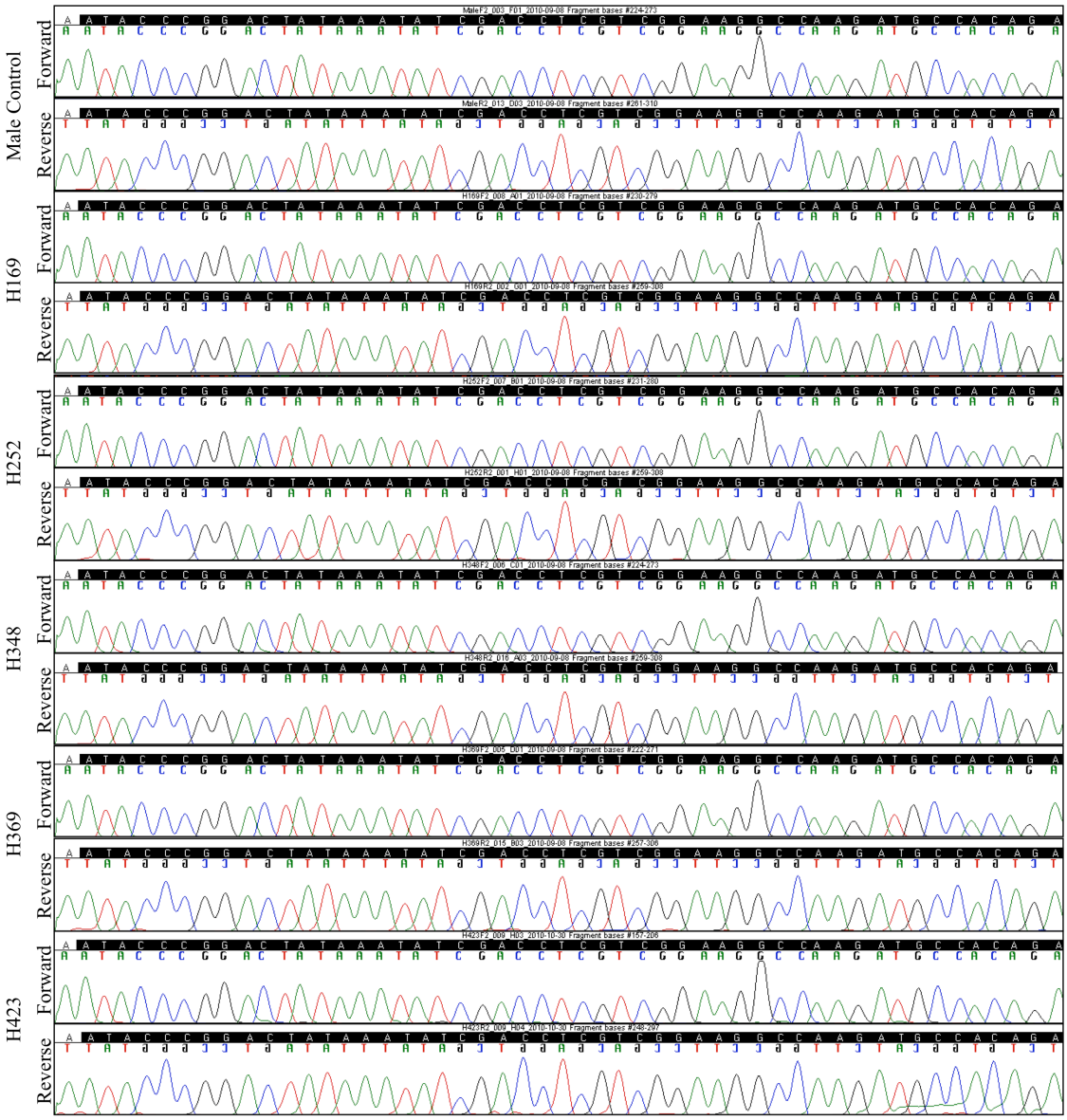
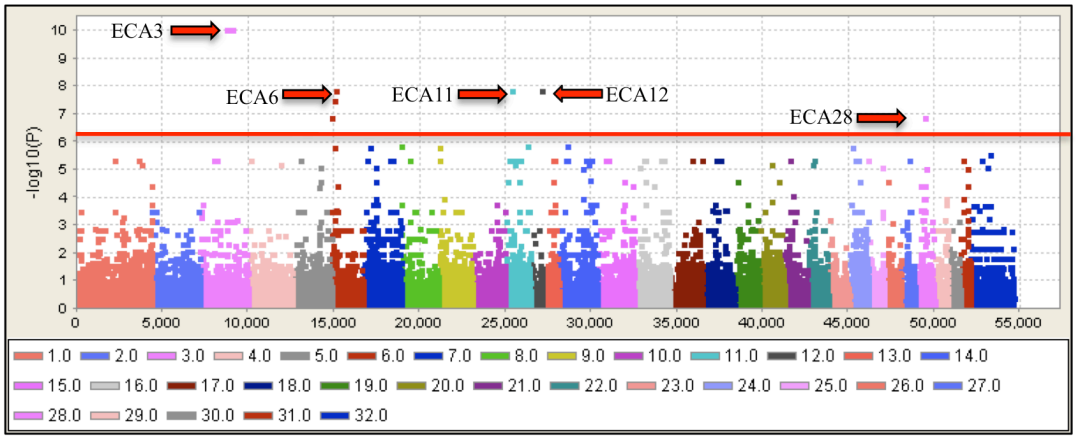


Figure 6. DNA sequence alignment of *SRY*. The alignments show identical sequences between a male control and the five cases. The region corresponds to Fragment 2 in Table 2 including the predicted promoter sequence and part of the *SRY* exonic sequence.

SNP50 Beadchip Genotyping and Data Analysis

Good quality (97-98%) equine SNP50 Beadchip genotypes were obtained for H348 and H369 and the 46 normal male Standardbred horses. Allele frequencies between the two cases and controls were analyzed using basic association (chi-square) tests and PLINK software. For the initial, non-permuted data, SNPs above the threshold cutoff p-value of $-\log_{10}>6$ were considered statistically significant and were observed in five horse chromosomes: ECA3, 6, 11, 12, and 28 (Fig. 7a). The initial data were corrected by 10,000 permutations, and two statistically significant SNPs were retained above the threshold cutoff p-value of $-\log_{10}>1.5$ on ECA3 (Fig. 7b). Both retained SNPs were located on the long arm (ECA3q) at base pair positions 64,917,308 and 80,911,978 both with p-value = 9.13E-11.

a)



b)

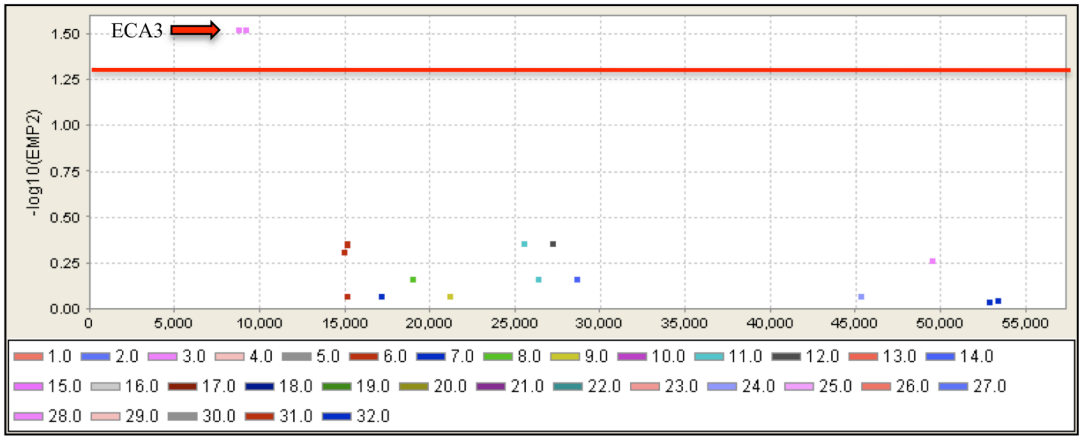
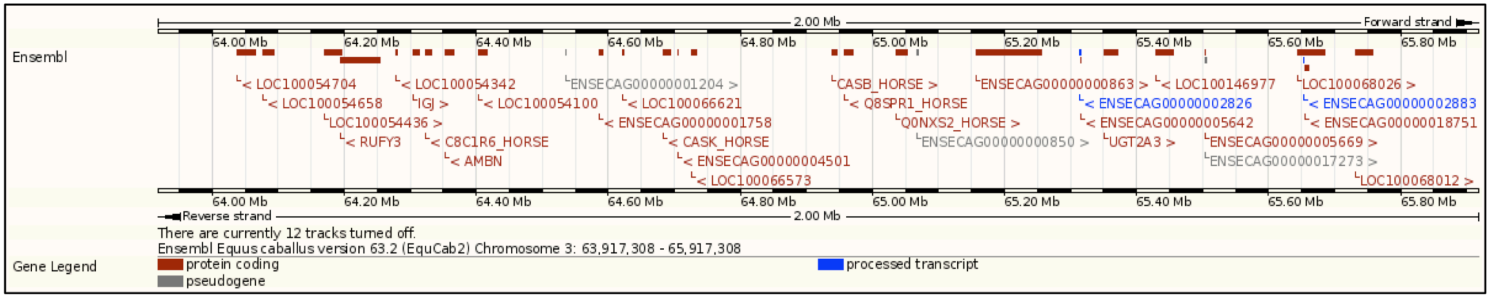


Figure 7. Manhattan plots of genotyping results. The plots show significant SNPs using non-permuted data (a) and permuted data (b).

Candidate Gene Discovery and Sequencing

The distance between the two SNPs on ECA3q was 16 megabase-pairs (Mb), which is very large for targeting candidate genes. Candidate genes were searched for in 1 Mb regions up- and downstream from the SNPs using Ensembl Genome Browser (http://useast.ensembl.org/index.html; Fig. 8a, b) and GRAIL (Raychaudhuri *et al.* 2009). Searching 1 Mb up- and downstream of the SNPs in Ensembl revealed the two ~2 Mb regions were relatively gene poor, harboring 25 and 18 protein coding genes, respectively (Fig. 8). Analysis using GRAIL mined 26 genes from the first SNP and 15 from the second SNP (Table 7). The entire 16 Mb region was also analyzed using GRAIL and Gene Ontology (GO) (Huang *et al.* 2009a, 2009b) to cluster the genes based on function. The combined analysis identified 52 candidate genes (Table 8). Of these, NIPA-like domain containing 1 (*NIPAL1*), a membrane-related magnesium transporter, was selected for exon sequencing and mutation discovery. Sequencing the first three exons and exon-intron boundaries identified two SNPs in non-coding regions: a G to A transition at base pair position 80,901,305 preceding the start of exon 1 (Fig. 9) and a C/A polymorphism at base pair position 80,892,949 in intron 1, near the intron 1/exon 2 boundary (Fig. 10). Notably, in the first SNP site, H348 and H369 were heterozygous for the G/A allele while the remaining cases were homozygous for the G allele (Fig. 9). All five cases and the male control were heterozygous for the second SNP. Neither SNP was recorded in the Broad Institute database for Horse SNPs (http://www.broadinstitute .org/mammals/horse/snp). No SNPs were found in the coding exons.

a)



b)

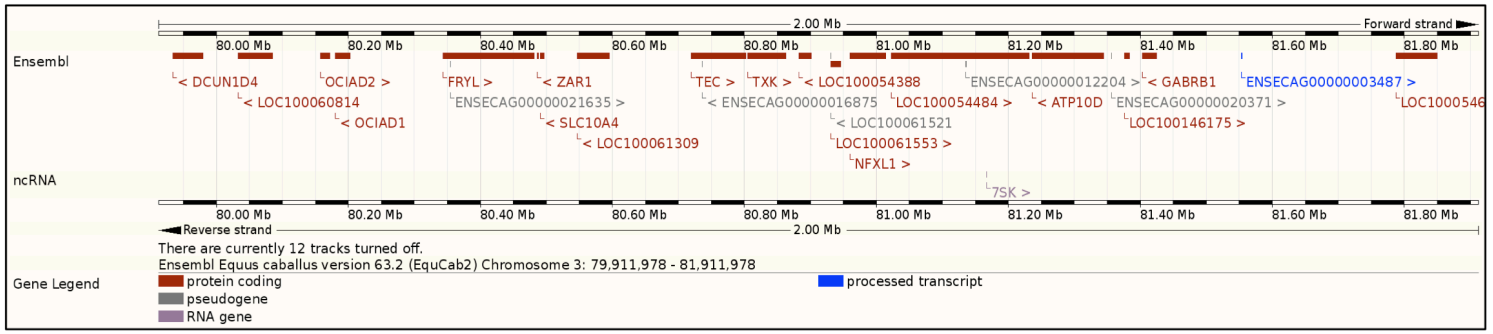


Figure 8. Ensembl Genome Browser depicting genes 1 MB up- and downstream of the SNPs on ECA3. Region 1 Mb up- and downstream of the SNP at base pair position 64,917,308 on ECA3 contains 25 protein coding genes, 3 pseudogenes, and 2 processed transcripts (a), and region 1 Mb up- and downstream of the SNP at base pair position 80,911,978 on ECA3 contains 18 protein coding genes, 5 pseudogenes, 1 RNA gene, and 1 processed transcript.

Table 7. Genes mined by GRAIL around the two SNPs on ECA3. Regions 1 Mb up- and downstream of the two SNPs on ECA 3 at base pair positions 64,917,308 (SNP1) and 80,911,978 (SNP2) were analyzed. This returned 26 GRAIL and 25 Ensembl genes for SNP1, and 15 and 18 genes for SNP2. Of these, 16 and 14 genes were common between the two analyses for SNP1 and SNP2, respectively.

GRAIL Results for SNP at 64,917,308	**Ensembl Results for SNP at 64,917,308**	**Both**
AMBN	AMBN	AMBN
AMTN	AMTN	AMTN
C4orf7	CABS1	CSN1S1
C4orf35	CSN1S1	CSN2
C4orf40	CSN2	CSN3
CSN1S1	CSN3	ENAM
CSN2	DCK	ODAM
CSN3	ENAM	SMR3A
ENAM	GRSF1	SULT1E1
HTN1	IGJ	UGT2A1
HTN3	MOBKL1A	UGT2A3
MUC7	ODAM	UGT2B10
ODAM	RUFY3	UGT2B11
PROL1	SMR3A	UGT2B28
SMR3A	SULT1E1	UGT2B4
SMR3B	UGT2A1	UGT2B7
STATH	UGT2A3	
SULT1B1	UGT2B10	
SULT1E1	UGT2B11	
UGT2A1	UGT2B15	
UGT2A3	UGT2B17	
UGT2B10	UGT2B28	
UGT2B11	UGT2B4	
UGT2B28	UGT2B7	
UGT2B4	UTP3	
UGT2B7		
GRAIL Results for SNP at 80,911,978	**Ensembl Results for SNP at 80,911,978**	**Both**
ATP10D	ATP10D	ATP10D
CNGA1	CNGA1	CNGA1
COMMD8	COMMD8	COMMD8
CORIN	CORIN	CORIN
FLJ21511	CWH43	GABRB1
GABRB1	DCUN1D4	NFXL1

Table 7. Continued

GRAIL Results for SNP at 80,911,978	Ensembl Results for SNP at 80,911,978	Both
NFXL1	FRYL	NIPAL1
NIPAL1	GABRA4	OCIAD1
OCIAD1	GABRB1	OCIAD2
OCIAD2	NFXL1	SLAIN2
SLAIN2	NIPAL1	SLC10A4
SLC10A4	OCIAD1	TEC
TEC	OCIAD2	TXK
TXK	SLAIN2	ZAR1
ZAR1	SLC10A4	
	TEC	
	TXK	
	ZAR1	

Table 8. Genes mined by GRAIL between the two SNPs on ECA3. The 16 Mb region was submitted to GRAIL, and the region was also manually analyzed from the Ensembl Genome Browser. Genes present in both lists were analyzed by function using Gene Ontology (GO). Functionally relevant genes from GO are denoted by an asterisk

GRAIL	Ensembl	Both
AASDH	1700023E05Rik	AASDH
AMBN	AASDH	ARL9
AMTN	AC068620.2	C4orf14
ARL9	ARL9	CENPC1
C4orf14	C4orf14	CEP135
C4orf35	CENPC1	CHIC2
C4orf40	CEP135	CLOCK
C4orf7	CHIC2	CSN2
CENPC1	CLOCK	DCUN1D4
CEP135	CSN2	EPHA5*
CHIC2	CWH43	EXOC1
CLOCK	DCUN1D4	FIP1L1
CNGA1	EPHA5	FRYL
CORIN	EXOC1	GNRHR*
CSN1S1	FIP1L1	HOPX
CSN2	FOLR1	IGFBP7
CSN3	FRYL	KDR*
DCUN1D4	GNRHR	KIAA1211
ENAM	HOPX	KIT*
EPHA5	IGFBP7	LNX1
EXOC1	KDR	LPHN3
FIP1L1	KIAA1211	NMU
FLJ21511	KIT	NIPAL1*
FRYL	LNX1	OCIAD1
GNRHR	LOC518437	OCIAD2
GSX2	LPHN3	PAICS
HOPX	LRRC66	PDCL2
HTN1	NIPAL1	PDGFRA*
HTN3	NMU	POLR2B
IGFBP7	OCIAD1	PPAT
IGJ	OCIAD2	RASL11B
KDR	PAICS	REST
KIAA1211	PDCL2	SCFD2
KIT	PDGFRA	SGCB*
LNX1	POLR2B	SCL10A4
LPHN3	PPAT	SPATA18

Table 8. Continued

GRAIL	Ensembl	Both
MUC7	Q1L836_BOVIN	SPINK2
NFXL1	RASL11B	SRD5A3
NMU	REST	SRP72
NIPAL1	SCFD2	STAP1
OCIAD1	SGCB	SULT1E1
OCIAD2	SLC10A4	TMEM165
ODAM	SN1S1	TMPRSS11A*
PAICS	SNORA31	TMPRSS11B*
PDCL2	SPATA18	TMPRSS11D*
PDGFRA	SPINK2	TMPRSS11E2
POLR2B	SRD5A3	TXK
PPAT	SRP72	UBA6
PROL1	STAP1	UGT2A3
RASL11B	SULT1E1	UGT2B10
REST	TECRL	YTHDC1
SCFD2	TMEM165	ZAR1
SGCB	TMPRSS11A	
SLAIN2	TMPRSS11B	
SLC10A4	TMPRSS11D	
SMR3A	TMPRSS11E2	
SMR3B	TMPRSS11F	
SPATA18	TXK	
SPINK2	UBA6	
SRD5A2L2	UGT1A6	
SRD5A3	UGT2A3	
SRP72	UGT2B10	
STAP1	YTHDC1	
STATH	ZAR1	
SULT1B1		
SULT1E1		
SYT14L		
TEC		
TMEM165		
TMPRSS11A		
TMPRSS11B		
TMPRSS11D		
TMPRSS11E		
TMPRSS11E2		
TXK		

Table 8. Continued

GRAIL	Ensembl	Both
UBA6		
UGT2A1		
UGT2A3		
UGT2B10		
UGT2B11		
UGT2B15		
UGT2B17		
UGT2B28		
UGT2B4		
UGT2B7		
USP46		
YTHDC1		
ZAR1		

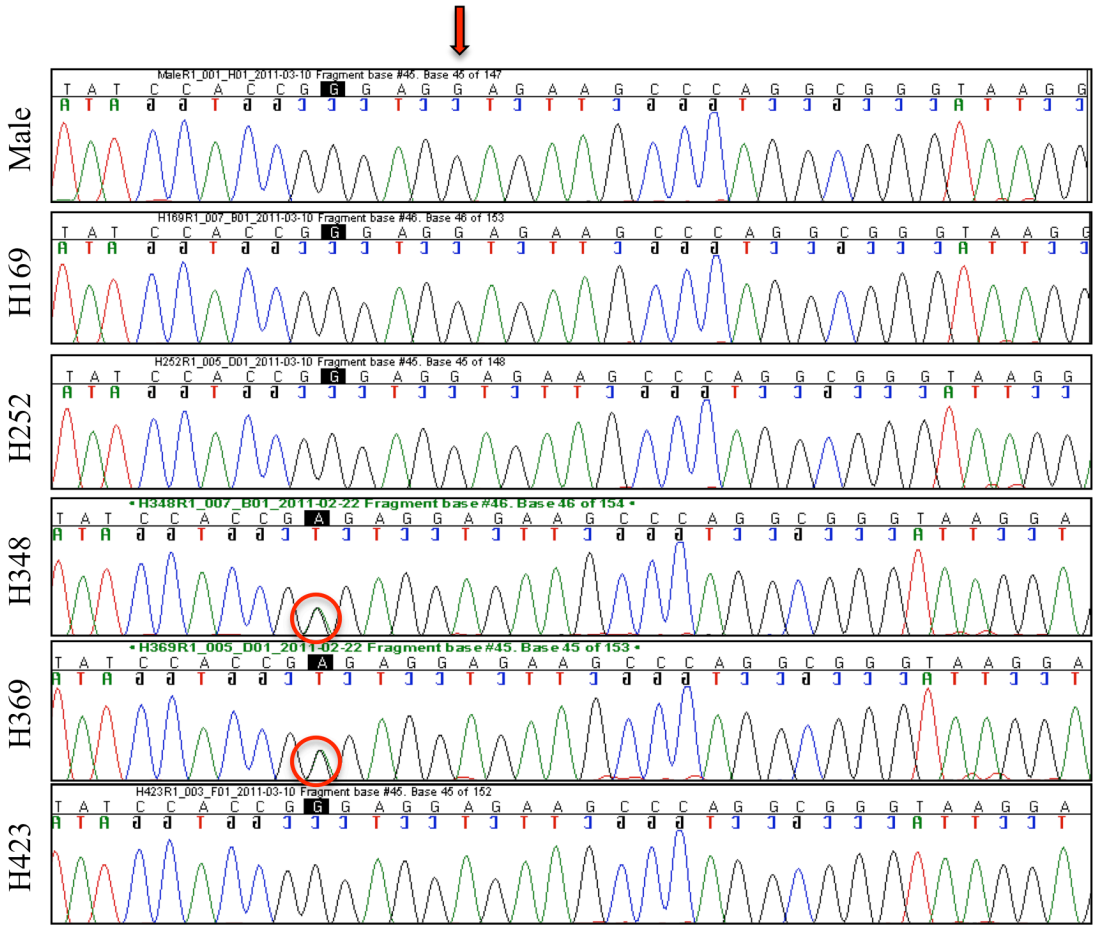


Figure 9. DNA sequence alignment of part of the first exon of *NIPAL1*. The start of the first exon is indicated by the red arrow. The sequence alignment shows a SNP in H348 and H369 in the non-coding region immediately preceding the first exon, indicated by the red circles.

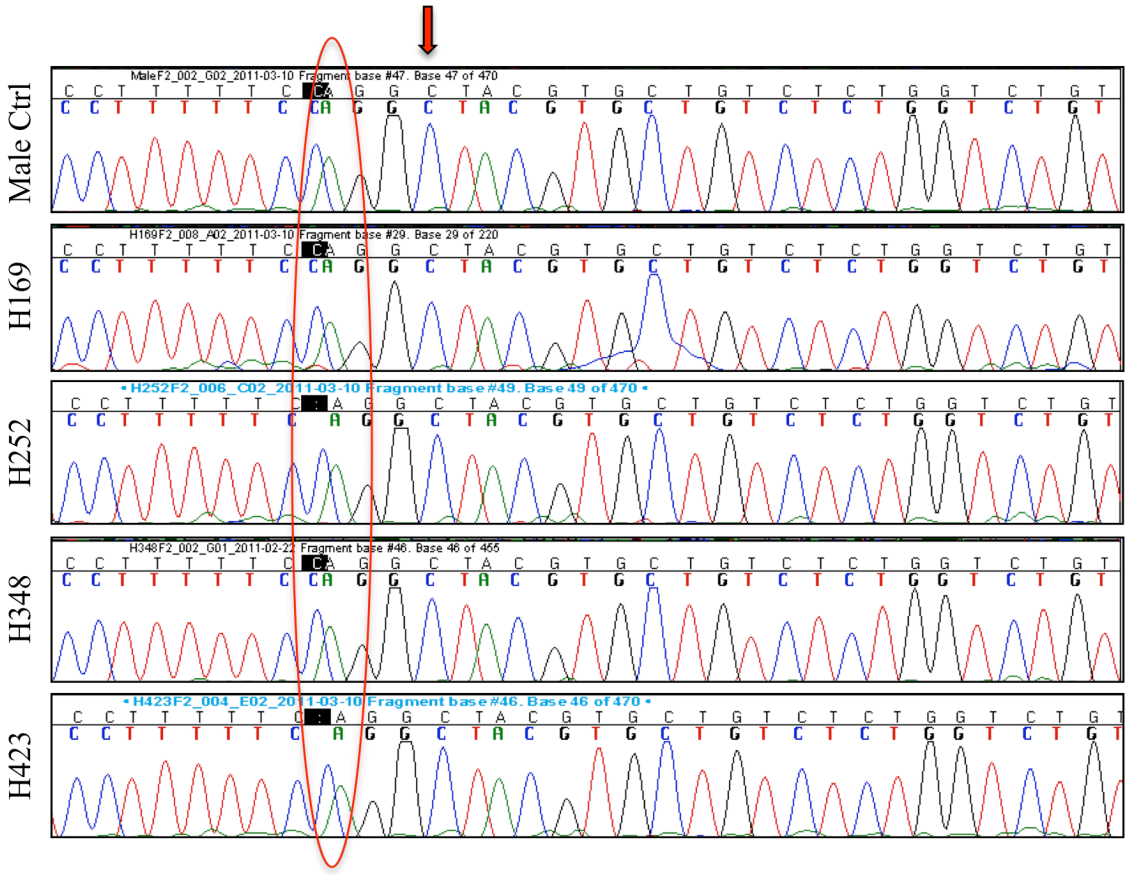


Figure 10. DNA sequence alignment of part of the second exon of *NIPAL1*. The start of exon 2 is indicated by the red arrow. This sequence alignment shows a C/A heterozygous position in all of the cases and male control indicated by the red circle.

The Discovery of Structural Rearrangements Using Array Comparative Genomic Hybridization (aCGH)

The quality of aCGH raw data was evaluated based on the derivative log ratio standard deviation (DLRSD) score, which calculates the noise level by measuring the distance between each spot and taking the standard deviation to acquire a baseline for detecting gains and losses. A lower DLRSD value equates to less noise, thus only results with DLRSD <0.30 were used for further analysis (Table 9). The DLRSD scores were acceptable for all but one of the *SRY*-positive XY sex reversed females. The mixed warmblood, H252, consistently produced poor results in three separate aCGH experiments and was thus excluded from the array data set.

Altogether, 187 CNVs were detected in the genomes of the four sex reversed animals (H169, H348, H369, and H423) in relation to the normal male Thoroughbred reference genome. Of these, 126 were gains, and 61 were losses. The CNVs were distributed over all autosomes and the X chromosome, with the exception of ECA22 and the Y chromosome, which showed no gains and losses in the studied individuals (Figs. 11 & 12). Distribution of CNVs along individual chromosomes (Fig. 11) shows that genomic gains and losses occur at both G-positive and G-negative bands, intercalary, and at the pericentromeric and subtelomeric regions. The number of CNVs per chromosome (Fig. 12) ranges from 14 on ECA1 to 2 on ECA13, 14, 18, 21, and 28-31, thus being overall correlated with the chromosome size. However, chromosomes 7, 10, 12, 16, and 20 showed a disproportionately high number of CNVs compared to the chromosome size (Fig. 12). Structural variants (n=5) were also found in the unassigned

sequences (chromosome Un). The size of the detected CNVs ranged from 1 kb to almost 3,500 kb, while approximately half of all the CNVs were in the size range of 10-300 kb (Fig. 13).

Analysis of aCGH results for individual animals revealed differences in the number and distribution of CNVs in the four sex reversed horses. The highest number of CNVs was detected in one of the Standardbreds (H348) and the least in the Appaloosa horse (H169) (Fig. 14). However, only a few CNVs were shared between the cases (Fig. 15). As expected, the two related Standardbreds, H348 and H369, had similar array results showing 25 common CNVs: 23 duplications and 2 deletions (Fig. 15). In contrast, there were only four common structural variants between the Appaloosa (H169) and one or the other Standardbred. Of these, a loss containing CUB and Sushi multiple domains 1 gene (*CSMD1*) on ECA27 was shared between the three animals. Further, the Appaloosa (H169) shared three losses and one gain with the Quarter Horse (H423), while only one common CNVs was detected between the latter and the two Standardbreds despite the fact that all three were classified as male pseudohermaphrodites. The only region containing a CNV commonly shared between all four cases was on ECA1, although the CNV was a deletion in H169, H348, and H369 and a duplication in H423.

Many of the largest CNVs contained clusters of olfactory receptor genes; other types of genes found in the CNVs included major histocompatibility complex (MHC) genes, immunity-related genes, muscle and skeletal genes, solute carrier genes, keratins, LINE1 associated reverse transcriptases, and a variety of others. Genes that appeared

more functionally relevant included a meiosis-specific nuclear structural 1 (*MNS1*) gene and a testis expressed 9 (*TEX9*) gene both in the same CNV region on ECA1. Overall, 41% of the CNVs contained coding genes, and the remaining 59% contained non-genic regions (Fig. 16).

Out of all of the CNVs, the functionally most interesting and relevant CNV was a massive (>225 kb) deletion on ECA 29 at 28,593,238-28,818,472 Mb found in H348 and H369, the two related Standardbreds (Fig. 17). The CNV contained a cluster of genes from the aldo-keto reductase gene family, which is known to catalyze the reactions to inter-convert sex hormones (testosterone and progesterone) into their active and inactive forms (Jin *et al.* 2011). However, the deletion was not observed in H169 or H423.

Table 9. Derivative log ratio standard deviation scores (DLRSD). The DLRSD scores for aCGH experiments were calculated by Agilent's Genome Workbench for the five sex reversed female horses.

Case	DLRSD Score
H169	0.190
H252	0.467 (average for three experiments)
H348	0.223
H369	0.189
H423	0.154

a) CNVs across all four cases

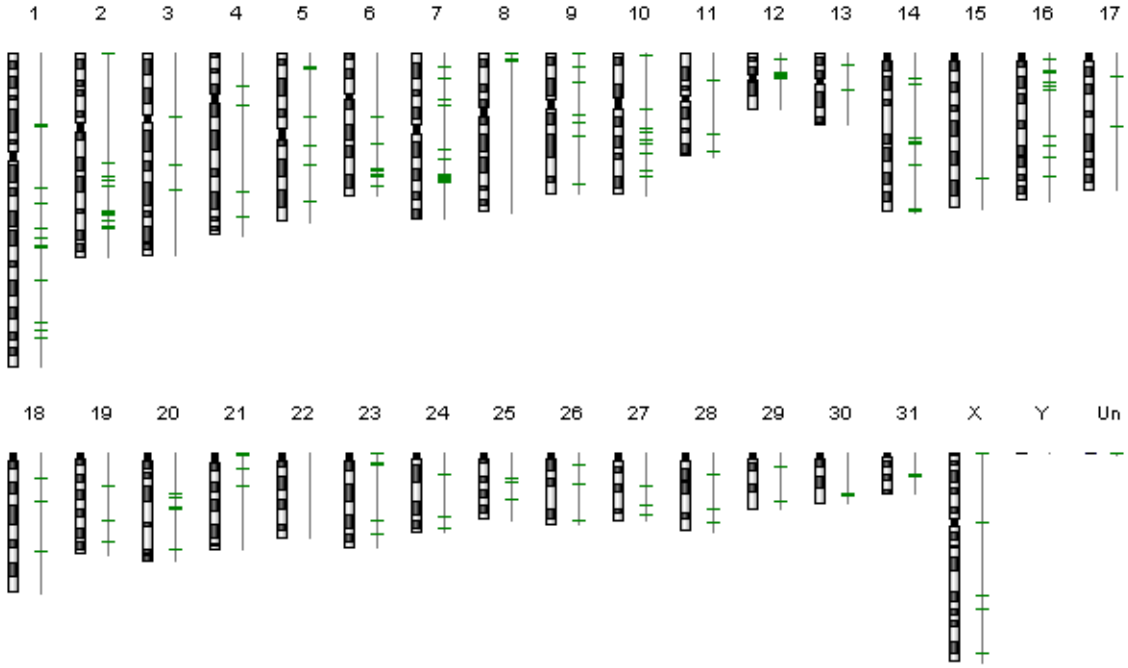


Figure 11. G-banded ideograms of CNV distribution. Ideograms include the horse autosomes, the sex chromosomes, and unassigned (Un) regions. Chromosome numbers are shown at the top of each ideogram. CNV distribution is shown across all four cases (a), as well each individual animal, H169 (b), H348 (c), H369 (d), and H423 (e).

b) H169

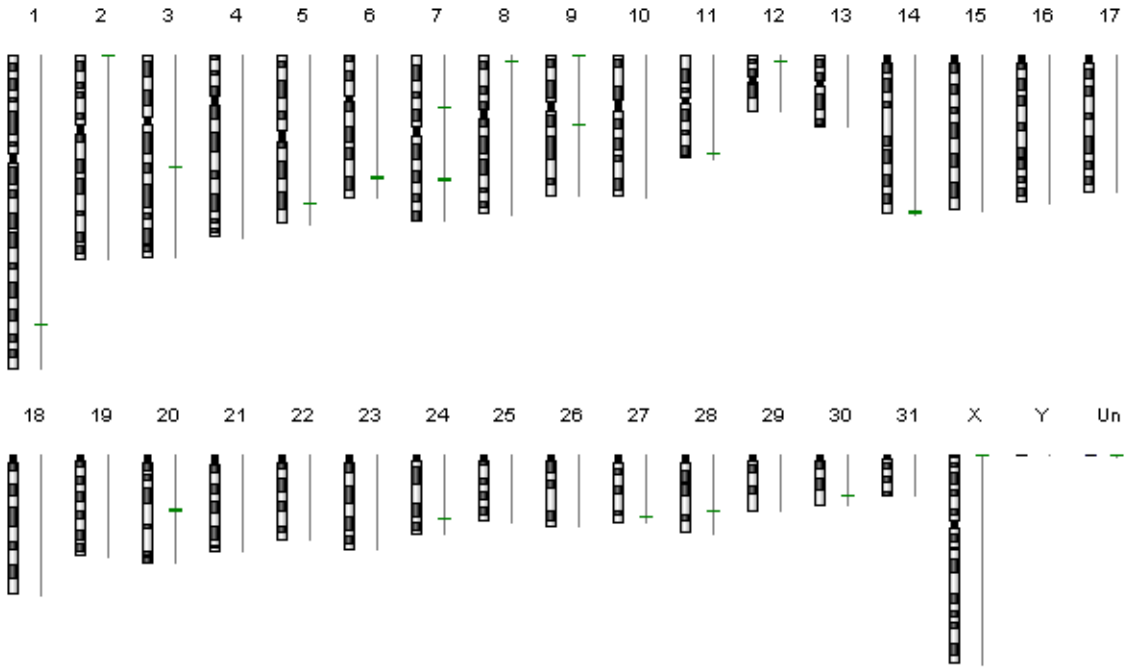


Figure 11. Continued

c) H348

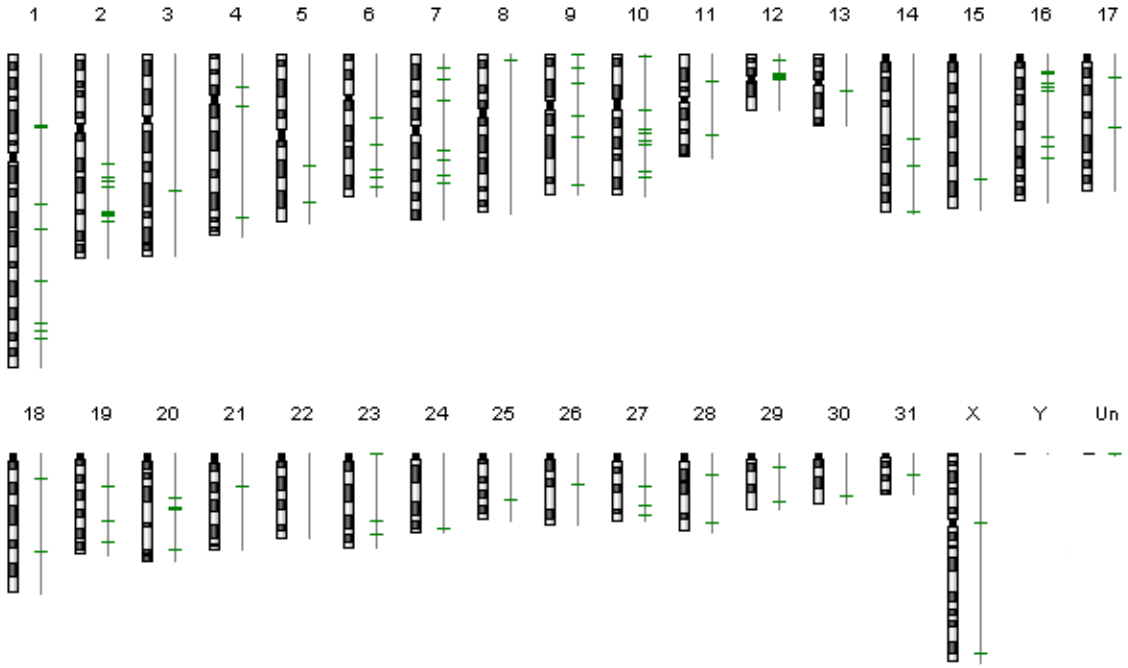


Figure 11. Continued

d) H369

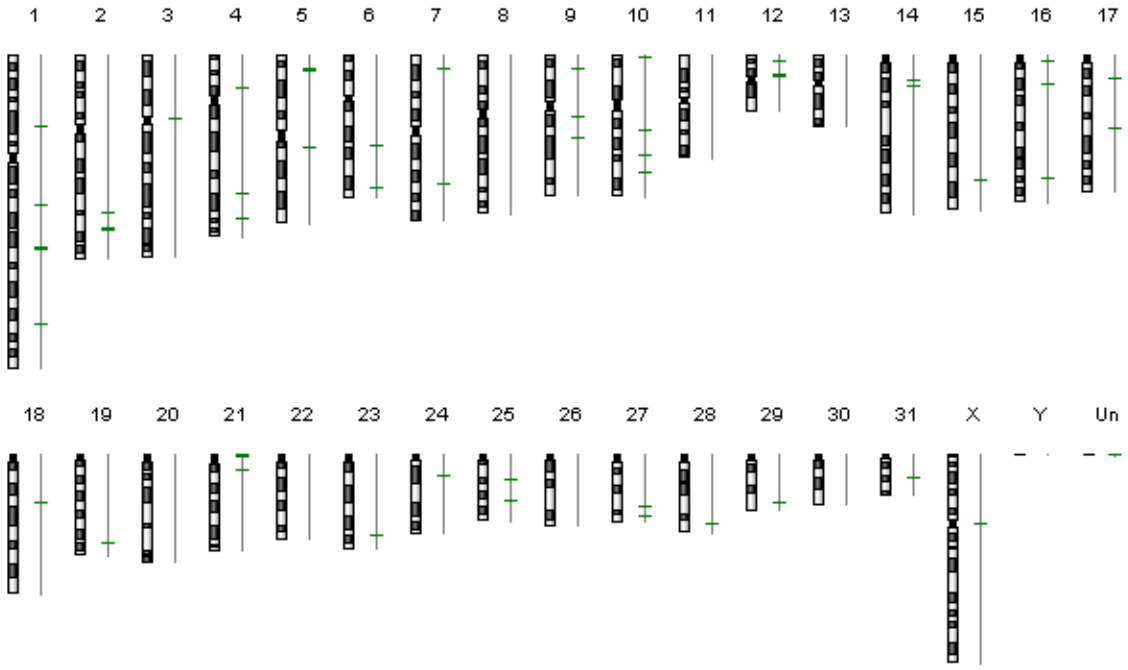


Figure 11. Continued

e) H423

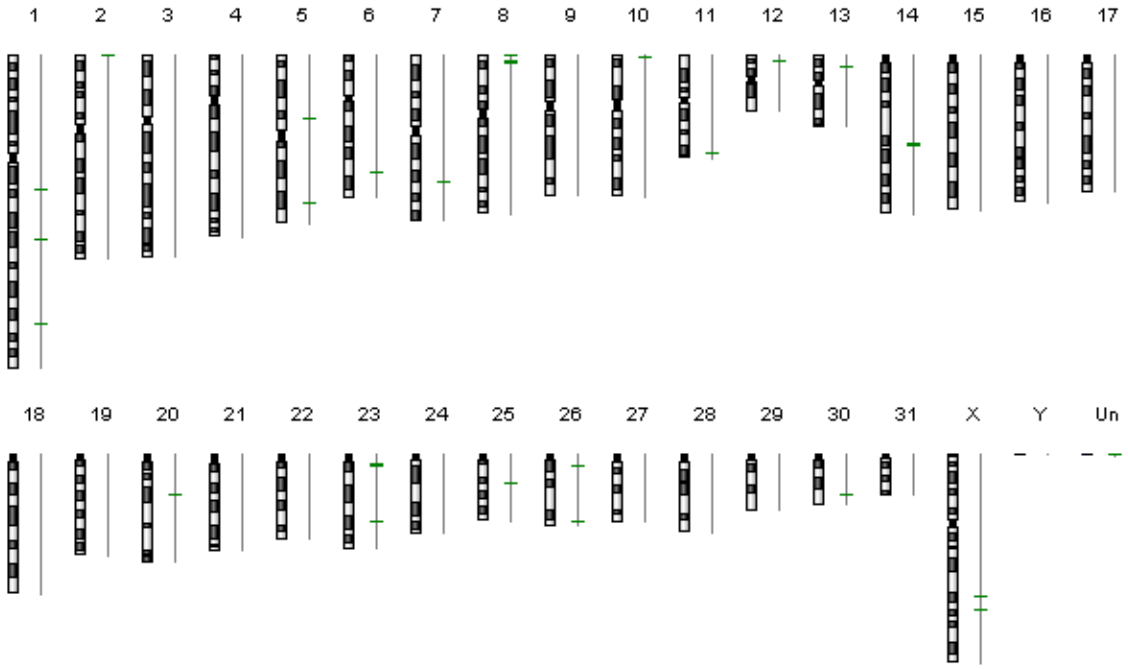


Figure 11. Continued

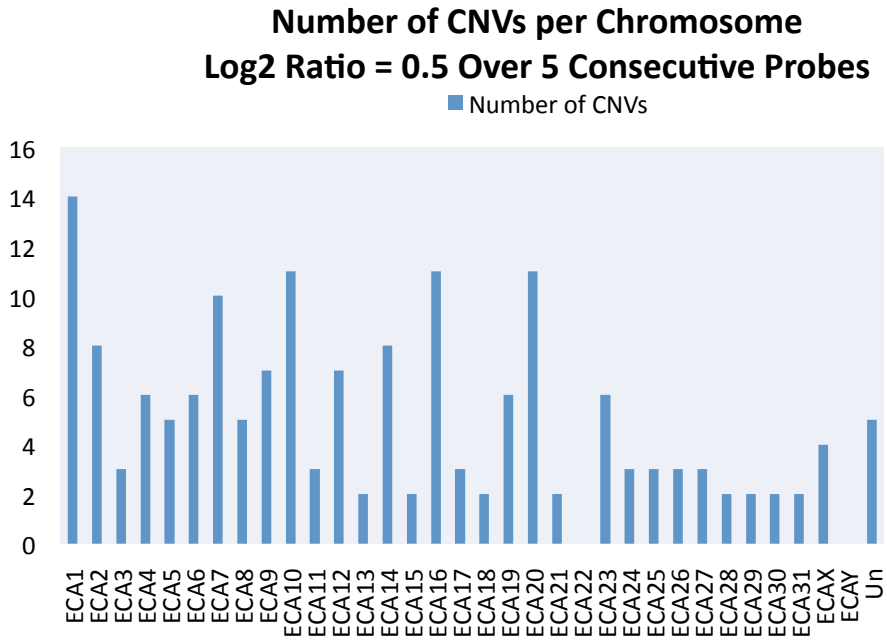


Figure 12. Bar graph of CNVs per chromosome.

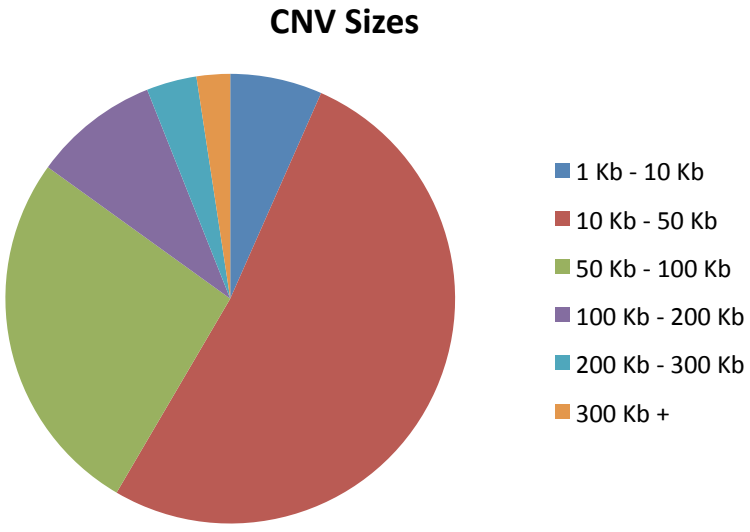


Figure 13. Distribution of CNVs by size.

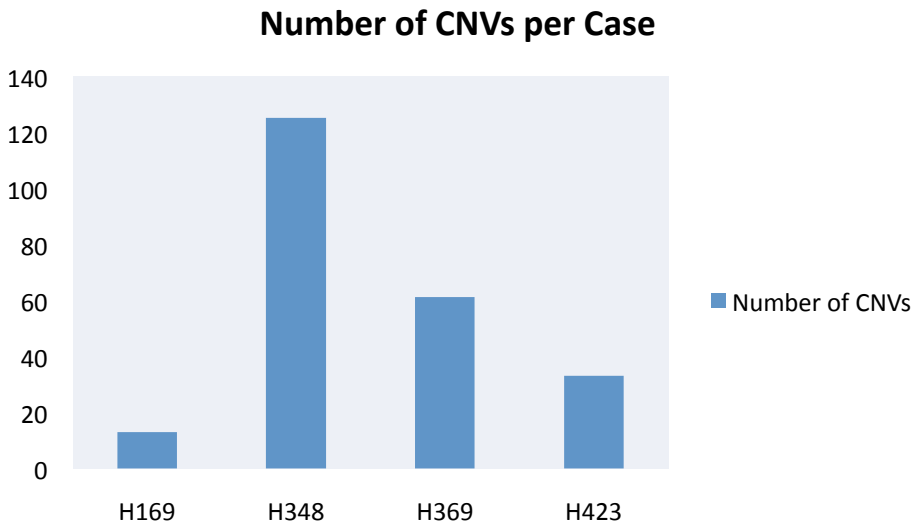


Figure 14. Number of CNVs per case.

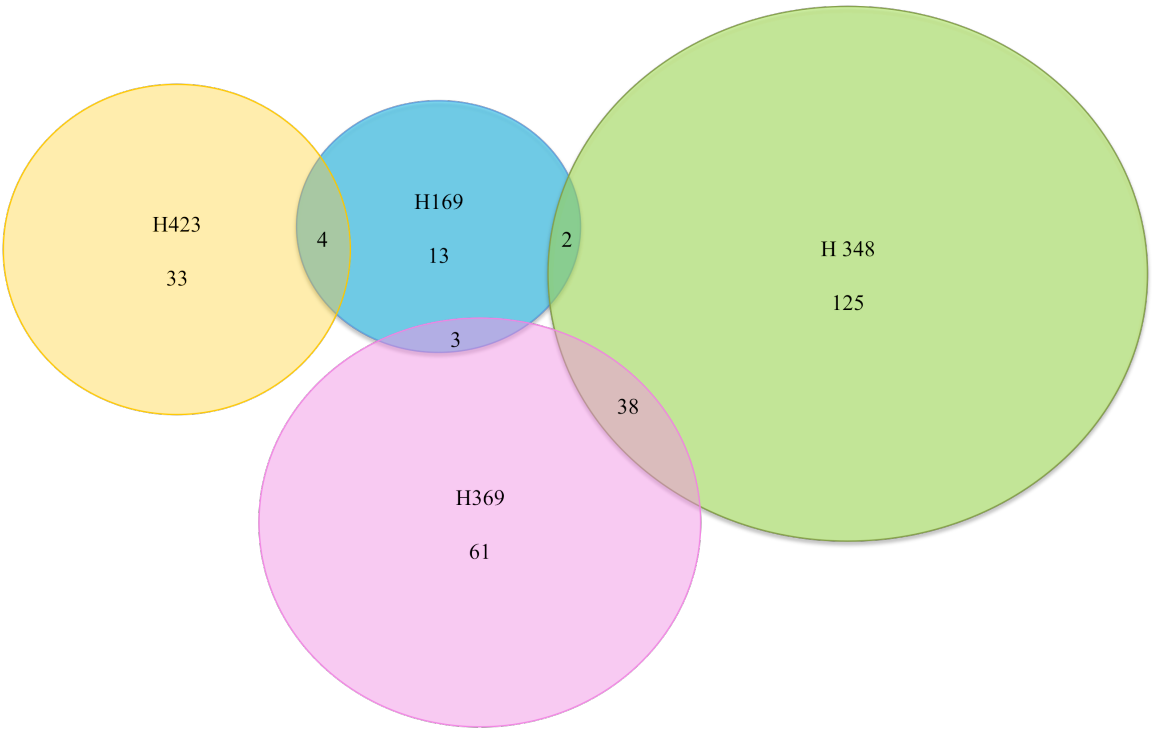


Figure 15. Shared CNVs between the cases.

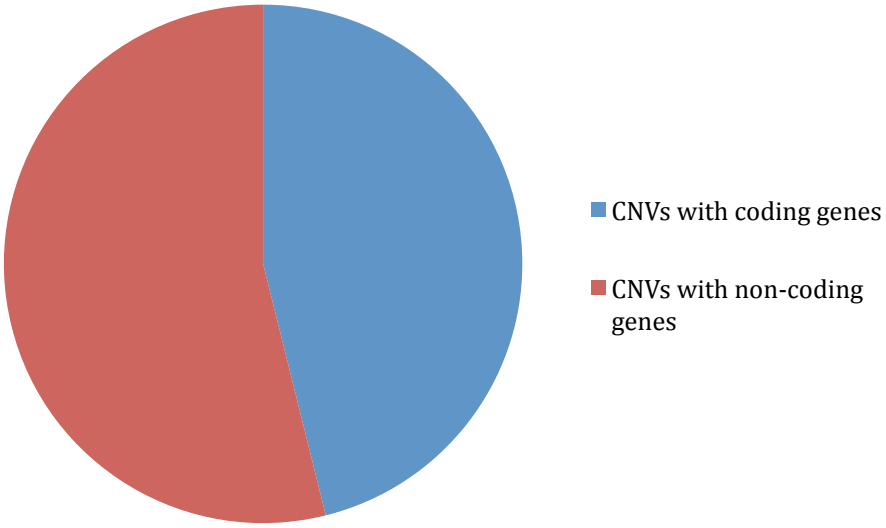


Figure 16. Distribution of CNVs containing coding genes and non-genic regions.

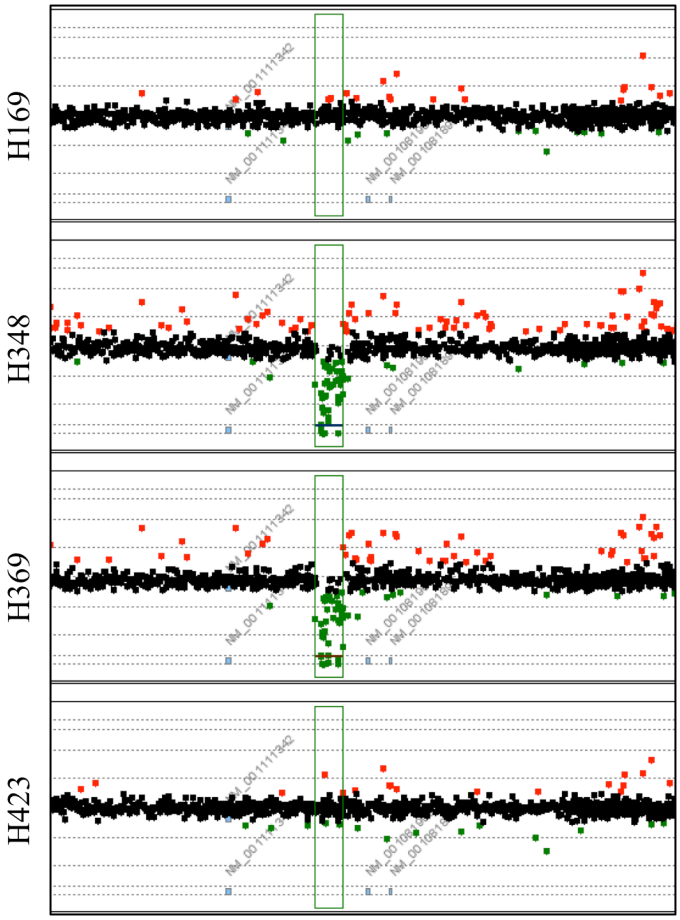


Figure 17. Array CGH image of deleted region on ECA29. The deletion was observed in H348 and H369, indicated by the box with the green outline. Each dot in the figure corresponds to a 60-mer oligonucleotide on the array. Normal copy number values are represented by the black dots along the middle, and deviations from the normal values are represented by either the red dots (gains) or the green dots (losses). Genes within the region are indicated by the gray text.

Array CGH Data Validation by Quantitative Real-Time PCR (qPCR)

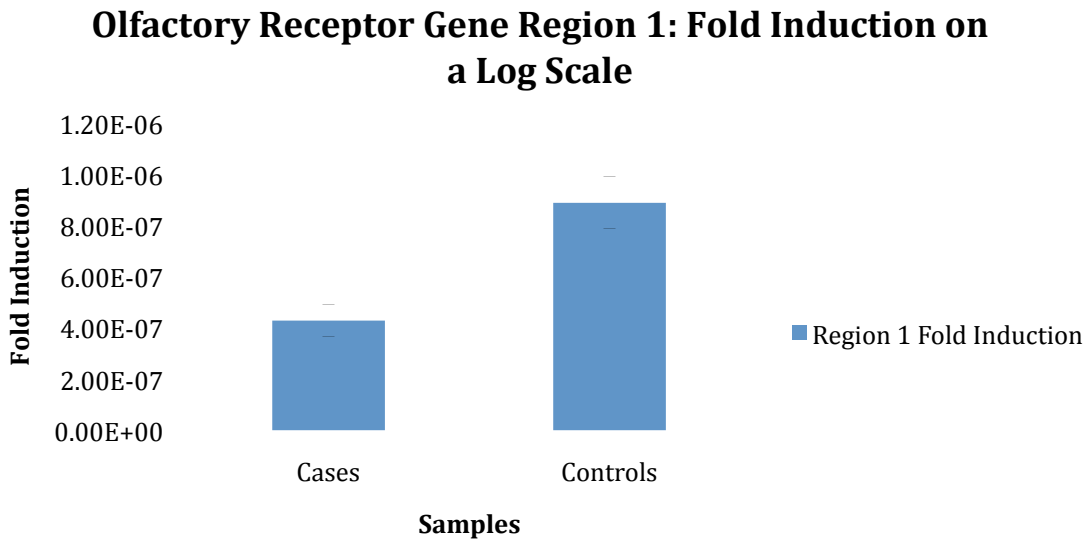
Quantitative Realtime PCR (qPCR) was used to confirm and measure the loss of DNA copy number levels in all five sex reversed cases compared to five normal male controls. Two regions from the aCGH results were selected for qPCR testing: an olfactory receptor gene cluster on ECA6 and a massive deletion on ECA29 containing a family of aldo-keto reductase genes. Although aCGH results were not available for the mixed warmblood, H252, genomic DNA was available for qPCR, thus it was included in this part of the experiment. The normal male controls were selected based on their breeds in order to compare them to the cases, thus two Quarter Horses and one Appaloosa were used, plus an Arabian and a Thoroughbred (*Bravo*) because a normal male Standardbred and a normal male mixed warmblood were not available. Standard curves were successfully calculated from the technical replicates from the first two primer sets but were not usable for the third primer. Therefore, the results were only available for the first two regions in the olfactory receptor gene cluster on ECA6 (see Table 4). As expected, DNA copy numbers were lower in the five cases compared to the controls for the olfactory receptor region on ECA6 (Fig. 18) thus validating the deletion observed by aCGH for H169, H348, H369, and H423 and also expanding this to H252. The calculated p-value for the first region was 3%, and the calculated p-value for the second region was 8% (Table 10); p-values <5% were considered highly significant.

Quantification of gene copy numbers by qPCR in the second region of interest, the aldo-keto reductase gene cluster on ECA29, gave no results because *AKR1CL1* was completely deleted in H348 and H369. Likewise, regular PCR with *AKR1CL1* primers

produced no amplification product in H348 and H369, while the sequence was successfully amplified in H169, H252, and H423, and in a male and a female control (Fig. 19). Altogether, the qPCR and PCR results were in agreement with the aCGH data and confirmed the findings.

Table 10. P-value scores for olfactory receptor gene regions in qPCR. Values <0.05 were considered highly significant.

Sample	P-value
Region 1	0.031195
Region 2	0.081011



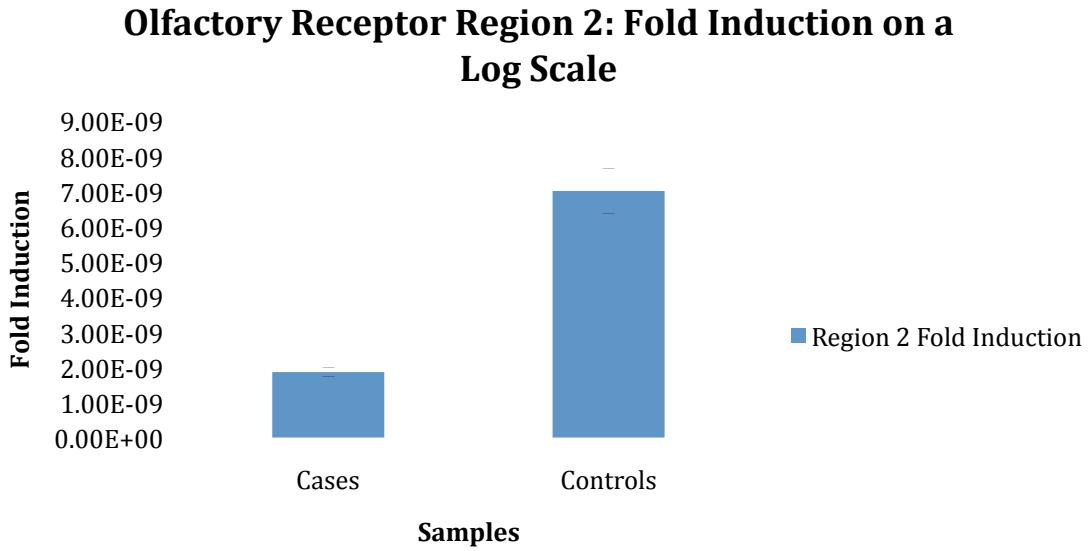


Figure 18. Fold induction between cases and controls from qPCR. Bar graphs are on a log scale.

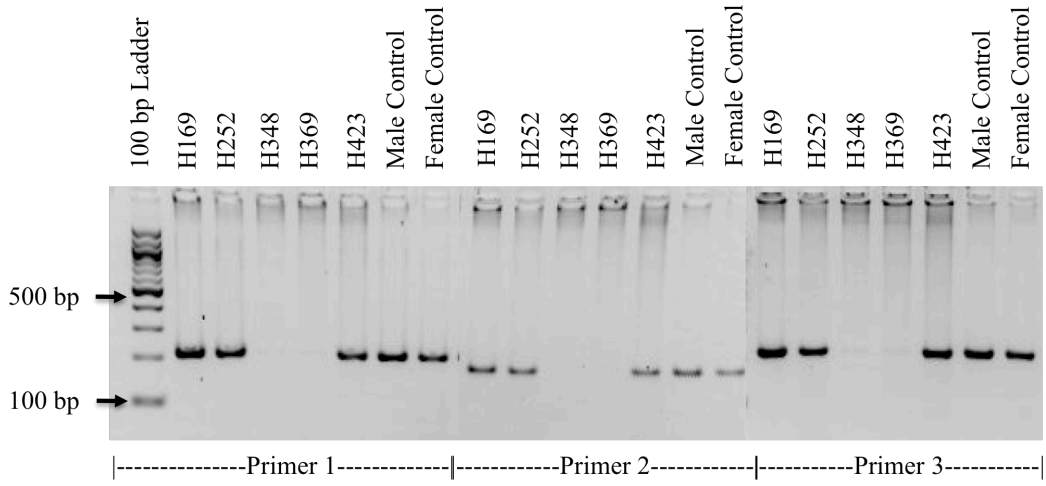


Figure 19. PCR products using *AKR1CL1* primers. The conventional PCR products are seen in all five cases and male and female control for the primers from *AKR1CL1* validating the deletion in H348 and H369.

Array CGH Data Validation by Fluorescence in situ *Hybridization (FISH)*

FISH was used to confirm and refine the massive deletion of aldo-keto reductase genes on ECA29 in animals H348 and H369. Metaphase chromosome preparations were obtained from all sex reversed horses, except H252. Metaphase spreads from a normal male horse were used as a control. Dual color FISH was carried out with the BAC clones (23N13, 67D20, 161K5, and 177J11) for the *AKR1CL1* gene and a control gene, *CREM*. The latter maps to a different region on ECA29 (q13, Raudsepp *et al.* 2008) and is not involved in any CNVs. Good quality FISH signals on both homologues were obtained with *CREM* and all four *AKR1CL1* BACs in H169, H423, and the control horse (Fig. 20). In cases H348 and H369, good signals on both homologues were only seen with *CREM* while signals with *AKR1CL1* clones 67D20 and 161K5 were consistently weaker than the *CREM* signal (Fig. 21) or the signals produced by the two *AKR1CL1* BACs in other animals. Notably, no FISH signals on the metaphase spreads of animals H348 and H369 were obtained with BACs 23N13 and 177J11 (Fig. 22). Alignment of the ends of the four *AKR1CL1* BAC clones with EquCab2 reference sequence showed that clones 23N13 and 177J11 only contain sequences within the deletion, while the sequences in clones 67D20 and 161K5 also extend out of the deletion (Fig. 23). This explains why clones 67D20 and 161K5 produced weak FISH signals and why no signals were observed with clones 23N13 and 177J11 in the two Standardbred horses. The FISH analysis thus further confirmed the aCGH and PCR results, but additionally refined the results showing that the deletion in the two Standardbreds is homozygous and that the deletion breakpoints are contained in clones 67D20 and 161K5.

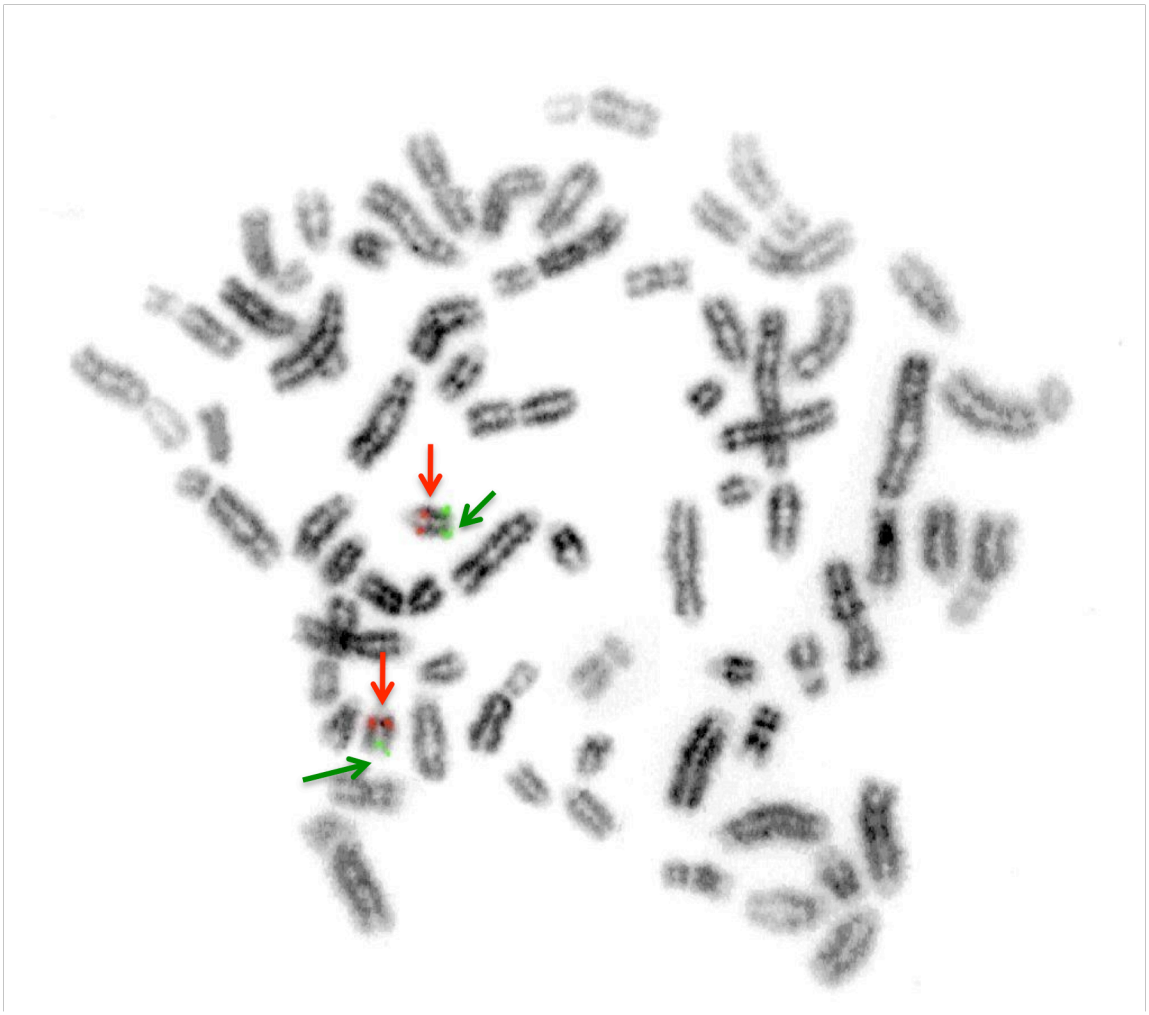


Figure 20. FISH on a normal male control using *CREM* and BAC 23N13. The FISH results show strong signals from *CREM* (red signal, red arrows), a control gene known to be mapped to the distal end of ECA29, and BAC 23N13 (green signal, green arrows), which is located within the deletion.

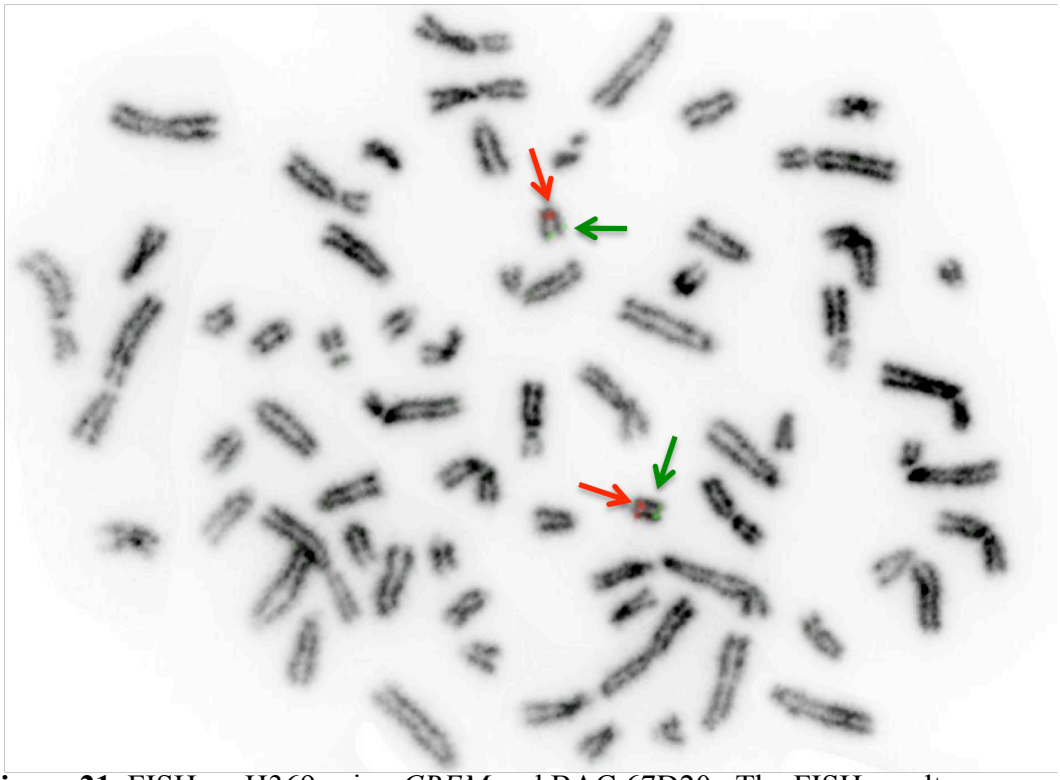


Figure 21. FISH on H369 using *CREM* and BAC 67D20. The FISH results on a metaphase spread of H369 show a strong signal for *CREM* (red signal, red arrows) on ECA29 and a weaker signal for BAC 67D20 (green signal, green arrows), which is located only partially in the deletion on ECA29.

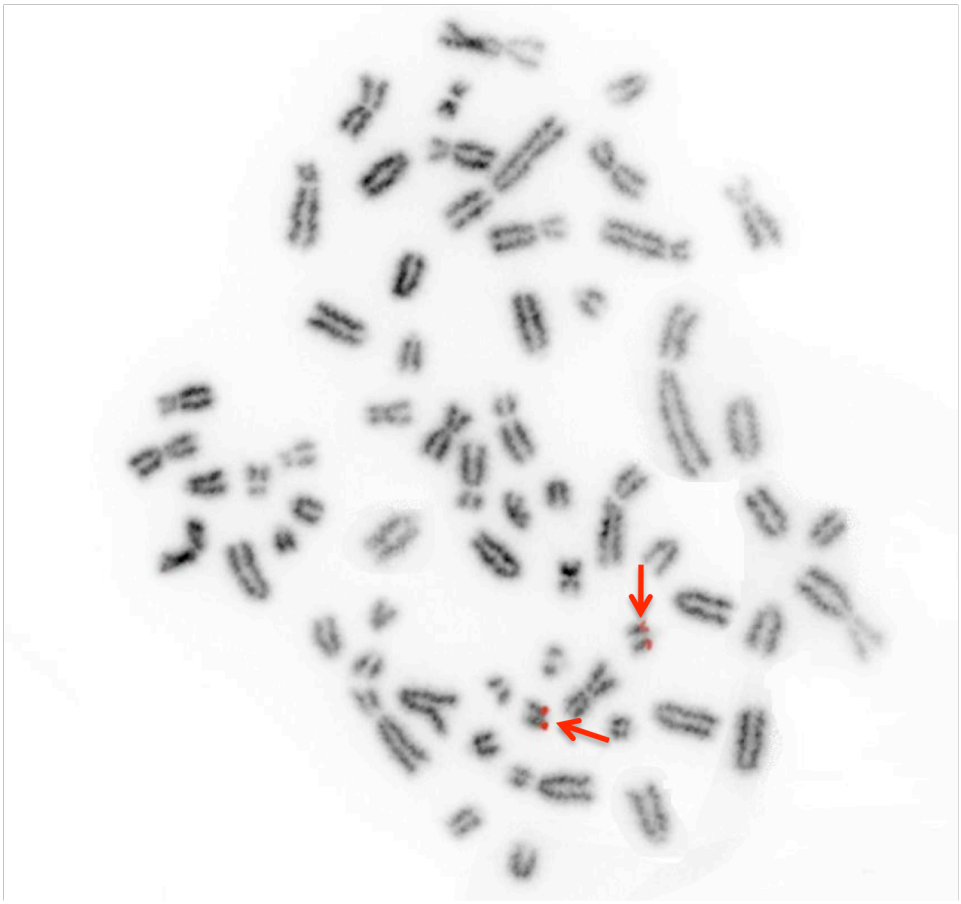


Figure 22. FISH on H369 using *CREM* and BAC 23N13. The FISH results on a metaphase spread of H369 show a strong signal for *CREM* (red signal, red arrows) on ECA29 and an absence of green signal for BAC 23N13, which is located inside the deleted region.

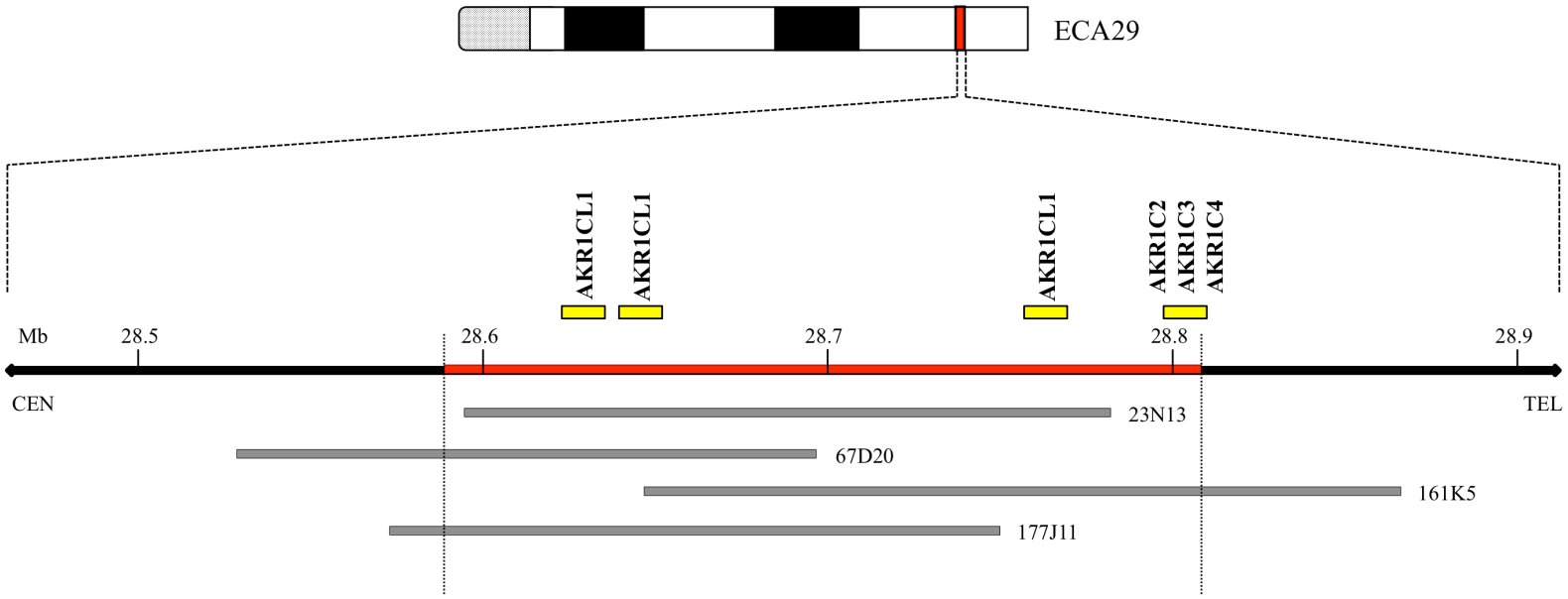


Figure 23. Schematic showing the deletion on ECA29 and the relative positions of the BACs. The location of the deletion is shown on an ideogram of ECA29 (red bar), the genes present within the deletion are indicated by the yellow bars, and the positions of the BAC clones respective to the deleted region are indicated by the gray bars.

PCR Refinement to Define the ECA29 Deletion Breakpoints

PCR was used to amplify regions inside and around the massive deletion on ECA29 using genomic DNA from H348, H369, a male control (*Bravo*), and a female control (*Twilight*). Primers immediately after the deletion start site ("DEL.IN1", Figs. 3 & 24a) and immediately before the deletion end site ("DEL.IN2", Figs. 3 & 24b) did not amplify in H348 and H369 but were successfully amplified in the two controls. A forward primer immediately preceding the deletion start site ("OUT") was combined with the reverse primer of "DEL.IN1," and no PCR product was observed in H348 and H369 while PCR product was present for the two controls (Fig. 24c). Similarly, the forward primer from "DEL.IN2" was combined with a reverse primer immediately following the deletion end site ("OUT"), and no PCR product was observed in H348 and H369 while PCR product was present for the two controls (Fig. 24d). Notably, no PCR product was obtained with the "OUT" forward and reverse primers in any of the horses analyzed.

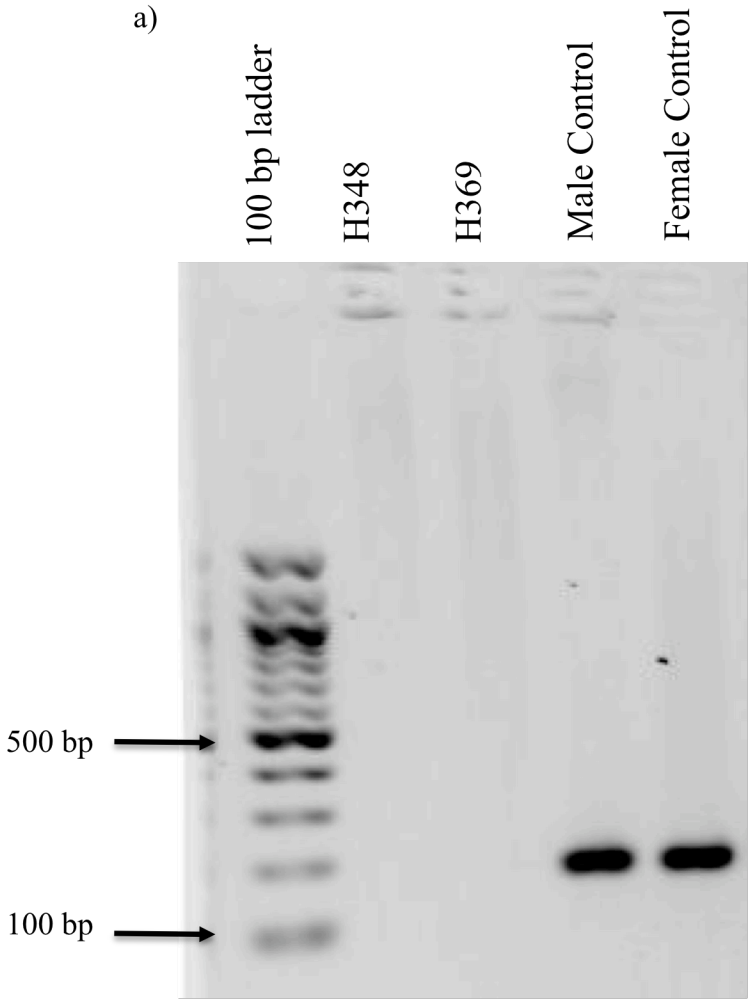


Figure 24. PCR results for primers designed in and around the ECA29 deletion. "DEL.IN1" did not amplify in H348 and H369 immediately after the deletion start site but did amplify in the two controls (a); "DEL.IN2" did not amplify in H348 and H369 immediately preceding the deletion end site but did amplify in the two controls (b); a forward primer immediately preceding the deletion start site ("OUT") and reverse primer from "DEL.IN1" did not amplify in H348 and H369 but did amplify in the two controls (c); forward primer from "DEL.IN2" and a reverse primer immediately following the deletion end site ("OUT") did not amplify in H348 and H369 but did amplify in the two controls (d).

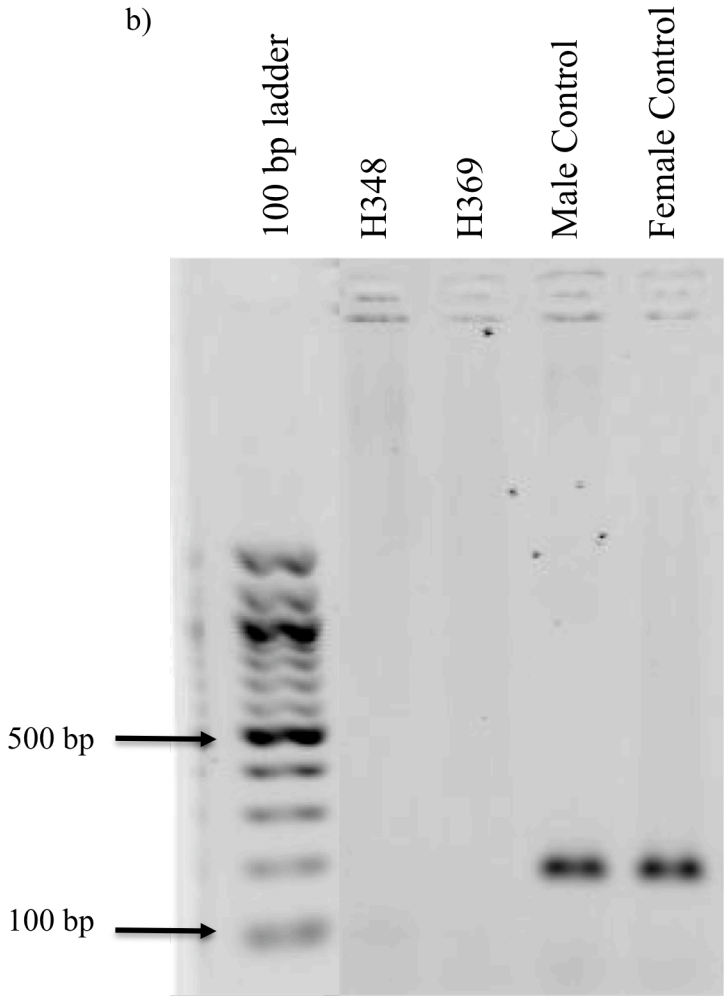


Figure 24. Continued

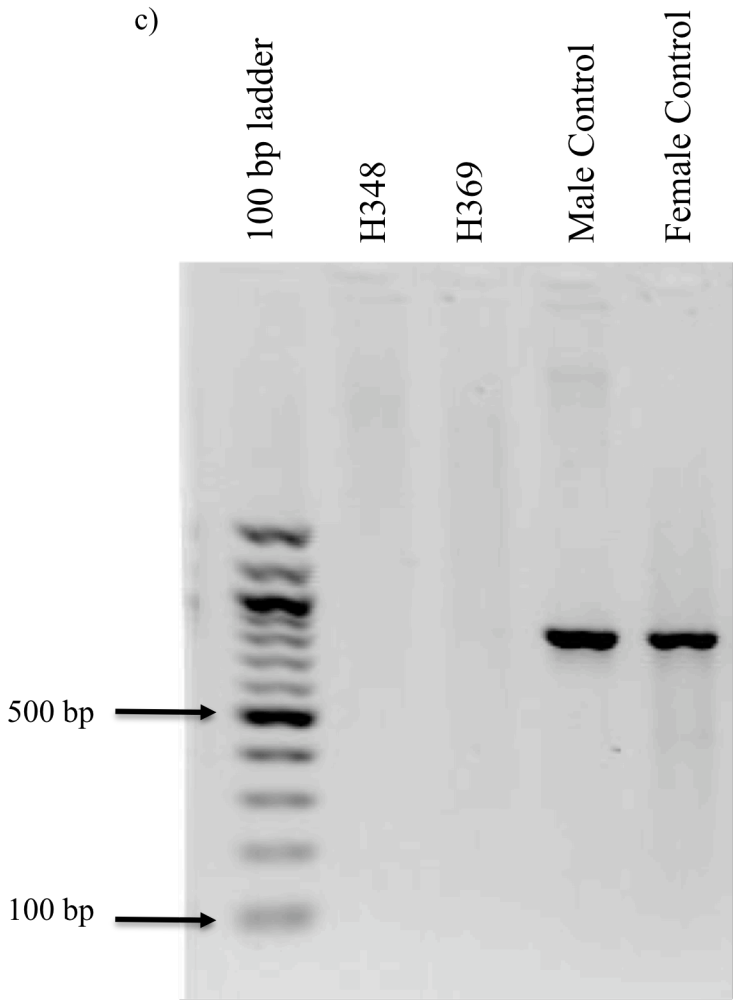


Figure 24. Continued

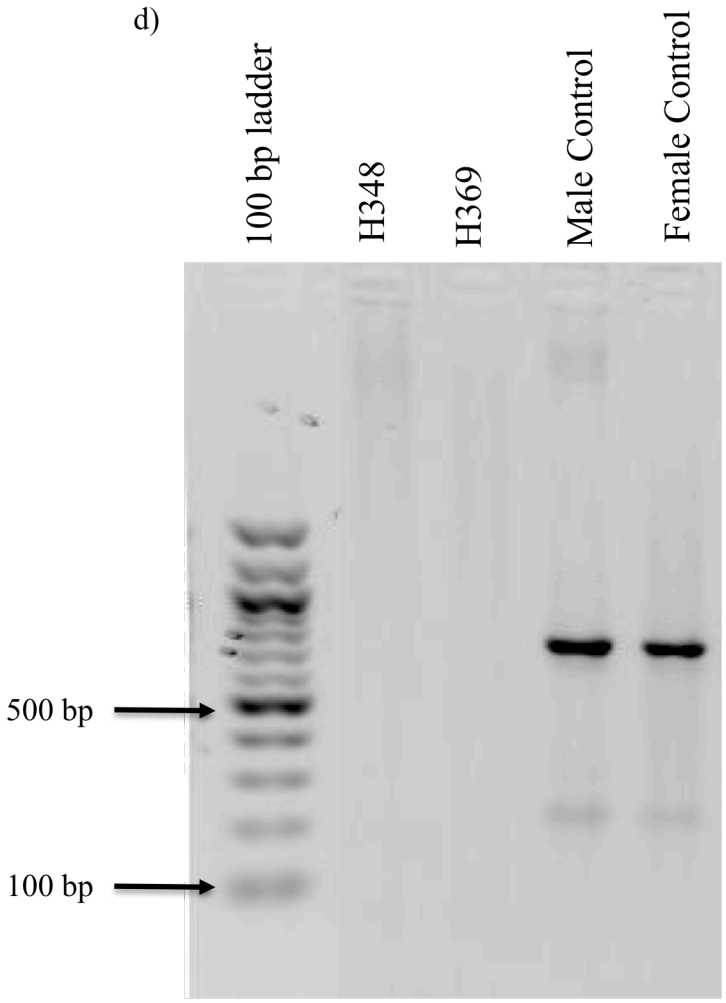


Figure 24. Continued

DISCUSSION

The search for genetic factors underlying mammalian sex determination has lasted for decades. It is well established that sex in mammals is determined in the fetal gonad by the presence or absence of the Y chromosome gene, *SRY* (Gubbay *et al*. 1990; Sinclair *et al.* 1990; Koopman *et al.* 1991). This initial trigger activates a complex regulatory gene network, which involves autosomal and X-linked genes, bending the balance either towards male or female sexual development (Matson *et al*. 2011; Migeon & Wisniewski 2003; Bagheri-Fam *et al.* 2010; Piprek 2009). Though the molecular details of these processes are not fully understood, it is generally agreed that mammalian sex determination is a dynamic and multidimensional process making identification of genes and non-genic factors involved challenging. It is therefore of no surprise that very little is currently known about the underlying genetics of abnormal sexual development, for example, the male-to-female XY sex reversal syndrome.

In this study, the molecular genetic causes of XY sex reversal syndrome in horses were investigated. Although *SRY* is known to be the main trigger for male sex determination, the XY sex reversal condition in horses (Kent *et al.* 1986; Makinen *et al*. 1999; Bugno *et al.* 2003; Switonski *et al.* 2005; Raudsepp *et al.* 2010) and other mammalian species (Michala *et al.* 2008; Ferrer *et al.* 2009; Villagomez *et al.* 2009) involves both *SRY*-negative and *SRY*-positive cases. While Y chromosome deletions explain the molecular background of *SRY*-negative cases in horses (Raudsepp *et al.* 2010), the genetic causes of the *SRY*-positive sex reversal remain as of yet undetermined.

Given the complexity of genetic regulation of mammalian sex determination, successful research in this field needs the use of global (genome-wide) analysis tools, which very recently have become available for domestic species, including the horse. Taking advantage of the draft genome sequence assembly (Wade *et al.* 2009), the SNP50 Beadchip (McCue *et al.* 2009), and the 400K whole genome tiling array (Qu *et al.* 2011), the *SRY*-positive XY sex reversal condition in the horse was investigated for the first time by combining the traditional cytogenetic and candidate gene analyses with whole-genome studies.

Cytogenetics of Sex Reversal

Cytogenetic analysis has been and still is the "gold standard" for the initial detection of the sex reversal condition in horses or any other mammalian species. This is because identifying the sex chromosome complement in the karyotype will determine whether or not there is a discrepancy between the phenotypic and genetic sex. Likewise in this study, the initial identification of the five horses as sex reversals was done through chromosome analysis showing that these five female or female-like animals had 64,XY karyotypes.

The Y Chromosome and SRY

Because the sex determining region on Y (*SRY*) is the primary trigger for male sexual development, the next logical step was to determine the presence or absence of the *SRY* gene on the Y chromosome of the five abnormal horses by PCR. Such tests

have been used in horses since the *SRY* sequence became available (Hasegawa *et al.* 1999) and show that the female-like phenotype or the lack of male phenotype in the majority of sex reversed animals is primarily due to the loss of *SRY* (Pailhoux *et al.* 1995; Makinen *et al*. 1999; Raudsepp *et al.* 2010). In this light, the findings of the present study were of particular interest because all five animals were positive for the *SRY* gene. Previously, the *SRY*-positive sex reversal in horses had been reported in only two studies (Switonski *et al.* 2005; Raudsepp *et al.* 2010), in which no mutations in the *SRY* sequence were found, suggesting that the condition is probably not Y-linked. Likewise, almost 80% of 46,XY women have no mutations in *SRY* or other Y-linked genes (Barbaro *et al.* 2009; Biason-Lauber *et al.* 2009; Schimmer & White 2010). However, in the remaining 20% of sex reversed women, different missense and frameshift mutations in the *SRY* coding region have been detected (Salehi *et al.* 2006; Shahid *et al.* 2008; Marchina *et al.* 2009), thus furthering the need to carry out *SRY* sequence analysis in every individual case of *SRY*-positive sex reversal.

Sequencing and sequence analysis of the *SRY* single exon in the horses involved in this study did not reveal any SNPs, mutations, or any other type of sequence differences between the cases and the normal male control. Therefore, the results validated the molecular integrity of the *SRY* on the Y chromosome in the five sex reversed horses. These findings are in agreement with previous studies in horses (Switonski *et al.* 2005; Raudsepp *et al.* 2010) and humans (Salehi *et al.* 2006; Shahid *et al.* 2008; Marchina *et al.* 2009), suggesting that the genetic causes of *SRY*-positive sex

reversal are not necessarily Y-linked and may involve genetic factors located on autosomes and the X chromosome.

Autosomal and X-linked Genetic Factors

As already reviewed in *Introduction*, mammalian sex determination involves synergistic action of multiple genes, and the genetic causes of abnormal sexual development are not necessarily defined by a single gene mutation. For example, in human sex reversed patients, mutations have been found in *SF1* (*NR5A1*) and *SOX9* (Sekido & Lovell-Badge 2008; Domenice *et al.* 2004) – genes directly interacting with *SRY*. Additionally, XY sex reversal has been associated with mutations in autosomal genes involved in testis determination and gonadal development, such as *DMRT1* (Barbaro *et al.* 2009), *CBX2* (Biason-Lauber *et al.* 2009), *WNT4* (Domenice *et al.* 2004), and two X-linked genes – *DAX1* (*NR0B1*) (Lardone *et al.* 2011) and *AR* (Galani *et al.* 2008), just to name a few. Due to mutations in the androgen receptor (*AR*) gene, the human 46,XY sex reversal condition is sometimes referred to as *Androgen Insensitivity Syndrome* (AIS) (Hughes & Deeb 2006). It has been proposed that testicular feminization and male pseudohermaphroditism in *SRY*-positive mares might also be caused by a mutation in the *AR* (Crabbe *et al.* 1992; Howden 2004; Switonski *et al.* 2005), although, to date, there is no experimental evidence to support this. Likewise, previous sequence analysis of *SOX9* on ECA11 and *DAX1* on ECAX in our laboratory did not reveal any differences between the *SRY*-positive sex reversed and normal animals (unpublished data).

Taken together, the genetic causes of abnormal sexual development, including the *SRY*-positive sex reversal condition can involve one or more genes on autosomes or sex chromosomes and shows heterogeneity between individuals and species. Due to this, application of comparative candidate gene approach for the discovery of causative mutations underlying the *SRY*-positive sex reversal syndrome is tedious, time-consuming, and might not lead to the expected results. Therefore, in this study, attempts were made to study the equine condition using genome-wide approaches.

Single Nucleotide Polymorphism Genotyping

Single nucleotide polymorphisms (SNPs) in human and animal genomes are widely used and valuable markers to conduct genome-wide association studies of Mendelian and complex traits by comparing allele and genotype frequencies across a large number of individuals (Shriver *et al.* 2005; Hoggart *et al.* 2008). Therefore, one of the most awaited byproducts of the horse genome sequencing project was the discovery of over a million SNPs (Wade *et al.* 2009). The SNP collection (http://www.broadinstitute.org/ftp/distribution/ horse_snp_release/v2/) has been used to generate the Equine SNP50 and SNP70 Beadchips, which consist of 54,602 and ~74,000 highly informative (polymorphic in many breeds) SNPs, respectively, uniformly distributed across the equine genome (McCue *et al.* 2009; Wade *et al.* 2009). The Equine SNP50 Beadchips have been successfully used for the discovery of single gene mutations (Brooks *et al.* 2010) and candidate regions associated with complex traits, such as osteochondrosis (Teyssèdre *et al.* 2011; Lykkjen *et al.* 2010) and laryngeal

neuropathy (Dupuis *et al.* 2011). Importantly, in order to grant sufficient statistical power for the analysis, such studies typically use DNA samples from hundreds of phenotypically well-defined cases and controls.

In the present study with just five cases identified, the use of SNP Beadchip genotyping for the detection of likely candidate genomic regions for *SRY*-positive sex reversal was initially not even a consideration. This was because the phenotypes (clinical descriptions) of the five animals were not identical, the animals were not of the same breed, and the number of cases was too low for any meaningful statistical analysis. However, two sex reversed animals, Standardbreds H348 and H369, had very similar phenotypes and were closely related through the maternal lineage (dam of H348 was the grand-dam of H369). Therefore, it was considered worth trying to genotype these two animals on the Equine SNP50 Beadchip and analyze the data using genotyping information from 46 normal unrelated male Standardbreds (part of an unrelated project at the University of Minnesota). The results of the basic association analysis showing two statistically significant SNPs on ECA3 came as a surprise but had to be taken with great caution. It is very likely that the observed deviation of the allele frequencies in the two sex reversed animals was solely due to their close relatedness to each other and distant relationship to the control Standardbreds. On the other hand, it cannot be ruled out that the two regions on ECA3 are indeed somehow associated with sex reversal. Therefore a 16 Mb region demarcated by two SNPs on ECA3 was interrogated in more detail using GRAIL (Raychaudhuri *et al.* 2009) and Ensembl to search for genes possibly relevant to sex determination. Altogether, fifty-two genes were identified that

showed various degrees of functional relevance to reproduction and sexual development (Table 8) by GRAIL, Ensembl, or GO analyses. None of these genes, however, have been previously associated with sex reversal in humans or other mammalian species. The genes that appeared the most number of times in the gene lists belonged to the UGT gene family, which encodes for a protein involved in olfactory neuroepithelium. Genes involved in olfactory receptor functional pathways have not been shown to play a role in sex reversal. However olfactory receptors are known to show copy number variation between horses (Wade *et al.* 2009), which is probably the reason why the UGT gene family was detected by the SNP analysis. Indeed, the same gene cluster also showed significant copy number variations by array CGH analysis (see below).

In the search for candidate genes, a magnesium transporter, NIPA-like domain containing 1 (*NIPAL1*) was selected for sequencing. The gene was independently identified in GRAIL analysis, in manual analysis through Ensembl, and had a GO functional gene enrichment score of 1.15. Although magnesium binds to enzymes and nucleic acids and is essential as a signaling molecule (Wu & Veillette 2011), it is not clear how magnesium transport is involved in the molecular pathways of sex determination. Also, the sequencing did not reveal significant differences in *NIPAL1* exons between the cases and controls.

Nevertheless, the SNP Beadchip analysis, data mining, and candidate gene research were useful exercises as well as the first attempt to investigate the genetic causes of equine sex reversal using genome-wide approaches. Despite the limited

results, the 16 Mb region on ECA3 and the 52 identified candidate genes set an important starting point for future studies.

Copy Number Variation Analysis Using aCGH

In addition to SNPs, genomes of individuals within a species can vary by submicroscopic insertions, deletions, and segmental duplications ranging from a few kilobases (kb) to several megabase pairs (Mb) in size (Redon *et al.* 2006; Feuk *et al.* 2006). Such copy number variants (CNVs) form a part of normal variation between genomes but can also be associated with genome instability, genetic diseases, and congenital disorders (Beckmann *et al*. 2008). Very recently, CNVs have been associated with disorders of sexual development in humans, including the XY sex reversal disorder (van Silfhout *et al* 2009; Tannour-Louet *et al.* 2010; White *et al.* 2011), strongly encouraging and justifying initiation of array CGH studies for equine sex reversal syndrome in this study.

Indeed, despite the limited number of cases and the fact that no aCGH data was obtained for H252, the most interesting results of this study came from CNV analysis.

In general, the number of CNVs detected in the four individuals was proportional with the size of the chromosomes, with the exceptions of ECA7, 10, 12, 16, and 20 where the number of CNVs was higher. Many of the CNVs detected on ECA7, 12, and 20 included large regions of olfactory receptor gene clusters, although olfactory receptor gene clusters were in CNVs detected on other chromosomes as well. Olfactory receptor genes are widespread throughout the genomes of many species and show high degree of

individual variation (Niimura & Nei 2007; Zozulya *et al.* 2001; Zhang & Firestein 2002; Quignon *et al.* 2005). Thus, the CNVs containing olfactory receptor genes most likely reflect individual differences and are not related to the sex reversal condition. Furthermore, the CNVs containing the olfactory receptor gene clusters were present sporadically showing no consistent patterns across the individual cases. As expected, the major histocompatibility complex (MHC) on ECA20 was another genomic region showing considerable variation between all four cases and the reference genome. Like CNVs in olfactory receptor regions, the CNVs in MHC did not show any common pattern between the cases and were considered a part of normal genomic variants. MHC genes play an important role in the immune system and autoimmunity and are known to be highly polymorphic across individual horses and equine populations (Brinkemeyer-Langford *et al.* 2010).

Another important observation from aCGH analysis was that many CNVs (>50%) were detected in non-genic regions. This does not reduce their potential functional relevance because duplications or deletions in promoters, enhancers, silencers, or other regulatory regions may affect the genes that have a role in sex determination. These results also underlie the need to study equine CNVs using a whole-genome tiling array and not restricting the search to coding sequences only.

Interestingly, ECA22 and ECAY did not show any CNVs in the four cases. The absence of CNVs on ECAY can be explained by the known low degree of Y sequence variation in horses (Wallner *et al.* 2003; Lindgren *et al.* 2004) and by the limited number (n=512) of ECAY sequences present on the array. Also, as shown in this and previous

studies (Raudsepp *et al.* 2010), the Y chromosome of *SRY*-positive XY females has retained its integrity and is molecularly the same as in normal males. In contrast, ECA22 is a medium sized (49.9 Mb) chromosome containing 403 protein coding genes, thus the lack of CNVs on ECA22 came as a surprise. This might be because there were no structural rearrangements on ECA22 in the four cases, and the chromosome plays no role in sex reversal; or this might be because the design of the array with a 7.5 kb resolution was not sufficient to detect variations. Some answers are expected to come from our ongoing studies of CNVs in normal individuals representing different horse breeds.

The primary goal of the aCGH experiments was to identify CNVs that are uniquely present in the cases and not in the reference genome, thus attempting to detect structural variants that might be associated with the sex reversal condition. Interestingly, there was only one CNV region (CNVR) that was shared between all four cases. This was a region on ECA1 at 158.8-159.1 Mb containing a cluster of genes encoding uncharacterized human proteins, and the orthologues have not been described in other mammalian species. Furthermore, the region showed a loss in H169, H348, and H369 and a gain in H423. Due to this, the CNVR was not further analyzed in this study but might be of interest in future research when annotation of horse genes has improved.

The remaining CNVs were present in one, two, or three animals, but not in all four. The highest number of CNVs was found in H348 and the lowest in the Appaloosa, H169. It is possible that this was partially due to technical reasons, although aCGH conditions have been thoroughly optimized and both DNA samples and hybridization

raw data must pass strict quality control before accepting the data. Therefore, it is more likely that the differences in the numbers and locations of CNVs reflect true individual variation. Not coincidentally, the animal H169 with the lowest number of CNVs also had the least severe sex reversal phenotype. In contrast, the two male hermaphrodite Standardbreds, H348 and H369, had the highest numbers of CNVs. Many of these CNVs were shared between the two animals reflecting their close relatedness. At the same time, it was surprising that almost no common CNVs were found between the Standardbreds and the Quarter Horse, H423, though all three were described as male pseudohermaphrodites. Instead, H423 shared several CNVs with the Appaloosa, H169, suggesting that *SRY*-positive sex reversal in horses is a genetically heterogeneous condition with varying degrees of genetic and phenotypic differences between individuals. Certainly, in order to prove or refute this statement, many more cases from the same and different breeds must be investigated.

A key for identifying at least some of the genetic causes of this complex condition is to investigate phenotypically similar and closely related animals, such as the two Standardbreds, H348 and H369, in this study. The two horses not only had the highest numbers of CNVs but also had many in common. While the majority of shared CNVs likely reflect similar genetic structure of the two genomes, a few might pinpoint the key regions involved in the abnormal sexual development. For example, both Standardbreds showed a significant gain in the region on ECA1 containing a meiosis-specific gene, *MNS1*, and a testis-expressed gene, *TEX9*. Notably, this CNV region is located just 2 Mb upstream from the only CNVR that was shared among all four cases.

Further investigations are needed to determine the connection, if any, of this CNVR to male hermaphroditism and sex reversal.

The most significant and functionally interesting CNVR in H348 and H369 was a massive deletion on ECA29 in a region containing genes belonging to the aldo-keto reductase family, including *AKR1CL1, AKR1C1, AKR1C2, AKR1C3,* and *AKR1C4*. Aldo-keto reductases convert aldehydes and ketones into their corresponding alcohols, carry out bile transport, and most relevantly, convert androgens, estrogens, and progestins into their less active metabolite form and back into their active form (Jin & Penning 2007; Takahashi *et al*. 2009). Androgen is a key factor that stimulates and controls the development and maintenance of male characteristics. It is possible that the loss of *AKR1CL1* and other aldo-keto reductase genes on ECA29 in the two Standardbreds has affected sex hormone regulation and might be one of the direct genetic causes of the abnormal phenotype. With so limited information available, it is too early to say whether the deletion represents an isolated case in one Standardbred family or if it is a common mutation underlying similar phenotypes in other horses. Regardless, this is the first report about a likely genetic cause of *SRY*-positive sex reversal in horses. Furthermore, the deletion was effectively validated both by PCR and FISH, providing useful tools for testing suspected XY females for the deletion in future studies. Given that CGH is a costly and elaborate analysis method, such tests will be of practical value.

Experiments have been initiated to precisely define the ECA29 deletion breakpoints so that a simple and reliable PCR test for *SRY*-positive XY sex reversed

mares could be developed. The currently available PCR primers and amplification results have narrowed down the breakpoints to be within 500 bp up- and downstream of the deletion start and end sites. We designed primers, "OUT", flanking the entire 300 kb deletion (Fig. 3) so that the PCR products could be obtained only in animals carrying the deletion and not from the normal sequence. However, no products were obtained in cases or controls. It is possible that due to the resolution of the array in which the 60-mer oligonucleotides are separated, on average, by 7.5 kb, the deletion start and end points were not demarcated precisely, and the primers were designed in the deleted sequences. It is also possible that the primers were designed in regions with incorrect sequence assembly. Nevertheless, demarcation and characterization of deletion breakpoints in ECA9 need further experimental work and analyses.

Finally, it was of interest to compare the results of aCGH with those of SNP genotyping. Overall, no overlaps between the two datasets were observed. The regions on ECA3, 6, 11, 12, and 28 that showed statistically significant SNPs contained no significant CNVs. This confirms our doubts and caution while interpreting the SNP data and supports the idea that the few statistically significant SNPs were detected solely because of the close relatedness in H348 and H369. However, in animal H348, a significant gain was found on ECA3 approximately 1 Mb downstream from one of the SNPs at 81,332,148-81,380,244 Mb. The region included a gene in the ATPase gene family and a COMM gene, which were also both present when the SNP data was mined by GRAIL and Ensembl. This proves that there is a structural variation at that position on ECA3 in H348. Additional experiments are needed to validate this CNV by qPCR.

Also, functional relevance of ATPase and COMM-domain genes to sexual development needs to be ascertained.

This is the first time aCGH has been used to investigate genetic causes of equine congenital disorders. Though the results are preliminary, need further validation by PCR and FISH, and require comparison with CNVs in normal horse populations, the findings clearly show the involvement of structural rearrangements in disorders of sexual development in horses. Additionally, the experiments carried out in this study validate the utility of the equine whole-genome tiling array for the study of variation in the horse genome in health and disease.

CONCLUSIONS

- The five horses involved in this study were genetically males having a 64,XY karyotype but phenotypically females or female-like;
- The Y chromosome of all studied animals carried the *SRY* gene, and the sequence was the same as in normal male horses;
- A ~300 kb deletion in aldo-keto reductase gene family on ECA29 was found to be a likely cause of the sex reversal condition in two out of the five studied animals;
- The *SRY*-positive male-to female sex reversal condition in horses is a genetically heterogeneous condition.

REFERENCES

Albrecht, K.H. & Eicher E.M. (1997) DNA sequence analysis of Sry alleles (subgenus *Mus*) implicates misregulation as the cause of C57BL/6J-Y(POS) sex reversal and defines the SRY functional unit. *Genetics* **147,** 1267-77.

Albrecht K.H., Young M., Washburn L.L. & Eicher E.M. (2003) Sry expression level and protein isoform differences play a role in abnormal testis development in C57BL/6J mice carrying certain Sry alleles. *Genetics* **164,** 277-88.

Arrighi F.E. & Hsu, T.C. (1971) Localization of heterochromatin in human chromosomes. *Cyto* **10,** 81-6.

Bagheri-Fam S., Sinclair A.H., Koopman P. & Harley V.R. (2010) Conserved regulatory modules in the Sox9 testis-specific enhancer predict roles for SOX, TCF/LEF, Forkhead, DMRT, and GATA proteins in vertebrate sex determination. *Int J Biochem Cell Biol* **42,** 472-7.

Barbaro M., Oscarson M., Schoumans J., Staaf J., Ivarsson S.A. & Wedell A. (2007) Isolated 46,XY gonadal dysgenesis in two sisters caused by a Xp21.2 interstitial duplication containing the *Dax1* gene. *J Clin Endocrin Metab* **92,** 3305-13.

Barbaro M., Balsamo A., Anderlid B.M., Myhre A.G., Gennari M., Nicoletti A., Pittalis M.C., Oscarson M. & Wedell A. (2009) Characterization of deletions at 9p affecting the candidate regions for sex reversal and deletion 9p syndrome by MLPA. *Euro J Human Genet* **17,** 1439-47.

Barrett J.C., Fry B., Maller J. & Daly M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263-5.

Beckmann J.S., Sharp A.J. & Antonarakis S.E. (2008) CNVs and genetic medicine (excitement and consequences of a rediscovery). *Cytogenet Genome Res* **123,** 7-16.

Bernstein R., Koo G.C. & Wachtel S.S. (1980) Abnormality of the X chromosome in human 46,XY female siblings with dysgenetic ovaries. *Science* **207,** 768-9.

Biason-Lauber A., Konrad D., Meyer M., deBeaufort C. & Schoenle E.J. (2009) Ovaries and female phenotype in a girl with 46,XY karyotype and mutations in the *CBX2* gene. *Am J Human Genet* **84,** 658-63.

Birren B., Green E.D., Klapholz S., Myers R.M. & Roskams J. (1997) *Genome Analysis: Analyzing DNA, A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Bowling A.T., Millon L. & Hughes J.P. (1987) An update of chromosomal abnormalities in mares. *J Reprod Fertil Suppl* **35,** 149-55.

Bradford S.T., Hiramatsu R., Maddugoda M.P., Bernard P., Chaboissier M., Sinclair A., Schedl A., Harley V., Kanai Y., Koopman P., Wilhelm D. (2009) The cerebellin 4 precursor gene is a direct target of *SRY* and *SOX9* in mice. *Biol Reprod* **80,** 1178-88.

Brinkmeyer-Langford C.L., Murphy W.J., Childers C.P. & Skow L.C. (2010) A conserved segmental duplication within ELA. *Anim Genet* **41,** 186-95.

Brooks S.A., Gabreski N., Miller D., Brisbin A., Brown H.E., Streeter C., Mezey J., Cook D. & Antczak D.F. (2010) Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for lavender foal syndrome. *PLoS Genet* **6(4),** e1000909. doi:10.1371/journal.pgen.1000909.

Bugno M., Klukowska J., Slota E., Tischner M. & Switonski M. (2003) A sporadic case of the sex-reversed mare (64,XY; *SRY*-negative): molecular and cytogenetic studies of the Y chromosome. *Theriogenology* **59,** 1597-603.

Burgoyne P.S. (1982) Genetic homology and crossing over in the X and Y chromosomes of mammals. *Hum Genet* **61,** 85-90.

Clinton M. (1998) Sex determination and gonadal development: a bird's eye view. *J Exp Zool* **281,** 188-92.

Crabbe B.G., Freeman D.A., Grant B.D., Kennedy P., Whitlatch L. & MacRae K. (1992) Testicular feminization syndrome in a mare. *J Am Vet Med Assoc* **200,** 1689-91.

De la Chapelle A., Tippett P.A., Wetterstrand G. & Page, D. (1984) Genetic evidence of X-Y interchange in a human XX male. *Nature* **307,** 170-1.

Domenice S., Correa R.V., Costa E.M., Nishi M.Y., Vilain E., Arnhold I.J. & Mendonca B.B. (2004) Mutations in the *SRY, DAX1, SF1* and *WNT4* genes in Brazilian sex-reversed patients. *Braz J Med Biol Res* **37,** 145-50.

Dupuis M., Zhang Z., Druet T., Denoix J., Charlier C., Lekeux P. & Georges M. (2011) Results of a haplotype-based GWAS for recurrent laryngeal neuropathy in the horse. *Mamm Genome* doi:10.1007/s00335-011-9337-3.

Ferrer L.M., Monteagudo L.V., Garcia de Jalon J.A., Tejedor M.T., Ramos J.J. & Lacasta, D. (2009) A case of ovine female XY sex reversal syndrome not related to anomalies in the sex-determining region Y (*SRY*). *Cyto Genome Res* **126,** 329-32.

Feuk L., Carson A.R. & Scherer S.W. (2006) Structural variation in the human genome. *Nature Rev Genet* **7,** 85-97.

Finley, B. (2009) Another sex dispute, but this athlete has four legs. *The New York Times* http://www.nytimes.com/2009/09/19/sports/19racing.html.

Fleming A. & Vilain E. (2005) The endless quest for sex determination genes. *Clin Genet* **67,** 15-25.

Freije D., Helms C., Watson M.S. & Donis-Keller H. (1992) Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science* **258,** 1784-7.

Galani A., Kitsiou-Tzeli S., Sofokleous C., Kanavakis E. & Kalpini-Mavrou A. (2008) Androgen insensitivity syndrome: clinical features and molecular defects. *Hormones* **7,** 217-29.

German J., Simpson J.L. & Chaganti, R.S.K. (1978) Genetically determined sex reversal in 46,XY humans. *Science* **202,** 53-6.

Gubbay J., Collignon J., Koopman P., Capel B., Economou A., Munsterberg A., Vivian N., Goodfellow P. & Lovell-Badge R. (1990) A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346,** 245-50.

Hammes A., Guo J.K., Lutsch G., Leheste J.R., Landrock D., Ziegler U., Gubler M. & Schedl A. (2001) Two splice variants of Wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. *Cell* **106,** 319-29.

Hasegawa T., Ishida M., Harigaya T., Sato F., Ishida N. & Mukoyama H. (1999) Linear *SRY* transcript in equine testis. *J Vet Med Sci* **61,** 97-100.

Henking H. (1891) Über Spermatogenese und deren Beziehung zur Entwicklung bei *Pyrrhocoris apterus. L Zeitschrift für wissenschaftliche Zoologie* **51,** 685-736.

Hersmus R., HCGM de Leeuw B., Stoop H., Bernard P., C van Doorn H., Bruggenwirth H.T., LS Drop S., Oosterhuis J.W., Harley V.R. & Looijenga L.H.J. (2009) A novel *SRY* missense mutation affecting nuclear import in a 46,XY female patient with bilateral gonadoblastoma. *Euro J Human Genet* **17,** 1642-9.

Hoggart C.J., Whittaker J.C., De Iorio M. & Balding D.J. (2008) Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. *PLoS Genet* 4(7): e1000120. doi:10.1371/journal.pgen.1000130.

Howden K.J. (2004) Androgen insensitivity syndrome in a thoroughbred mare (64, XY-testicular feminization). *Canadian Veterinary Journal* **45,** 501-3.

Huang D.W., Sherman B.T. & Lempicki R.A. (2009a) Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc* **4,** 44-57.

Huang D.W., Sherman B.T. & Lempicki R.A. (2009b) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37,** 1-13.

Hughes J.F., Skaletsky H., Pyntikova T., Graves T.A., van Daalen S.K.M., Minx P.J., Fulton R.S., McGrath S.D., Locke D.P., Friedman C., Trask B.J., Mardis E.R., Warren W.C., Repping S., Rozen S., Wilson R.K. & Page D.C. (2010) Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463,** 536-9.

Hughes I.A. & Deeb A. (2006) Androgen resistance. *Best Pract Res Clin Endocrin Metab* **20,** 577-98.

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

Jin Y., Mesaros A.C., Blair I.A. & Penning T.M. (2011) Stereospecific reduction of 5β-reduced steroids by human ketosteroid reductases of the AKR (aldo-keto reductase) superfamily: role of *AKR1C1-AKR1C4* in the metabolism of testosterone and progesterone via the 5β-reductase pathway. *J Biochem* **437,** 53-61.

Jin Y. & Penning T.M. (2007) Aldo-keto reductases and bioactivation/detoxication. *Annu Rev Pharmacol Toxicol* **47,** 263-92.

Katoh-Fukui Y., Tsuchiya R., Shiroishi T., Nakahara Y., Hashimoto N., Noguchi K. & Higashinakagawa T. (1998) Male-to-female sex reversal in M33 mutant mice. *Nature* **393,** 688-92.

Kaur G., Delluc-Clavieres A., Poon I.K.H., Forwood J.K., Glover D.J. & Jans D.A. (2010) Calmodulin-dependent nuclear import of HMG-box family nuclear factors: importance of the role of *SRY* in sex reversal. *J Biochem* **430,** 39-48.

Kent M.G., Shoffner R.N., Buoen L. & Weber A.F. (1986) XY sex-reversal syndrome in the domestic horse. *Cyto and Cell Genet* **42,** 8-18.

Kent M.G., Schneller H.E., Hegsted R.L., Johnston S.D. & Wachetl S.S. (1988) Concentration of serum testosterone in XY sex reversed horses. *J Endocrin Invest* **11,** 609-13.

Kent J., Wheatley S.C., Andrews J.E., Sinclair A.H. & Koopman P. (1996) A male-specific role for *SOX9* in vertebrate sex determination. *Dev* **122,** 2813-22.

Kim J., Han B., Lee H., Yoo H. & Lee J. (2010) Development of SNP-based human identification system. *Int J Legal Med* **124,** 125-31.

Koopman P., Gubbay J., Vivian N., Goodfellow P. & Lovell-Badge R. (1991) Male development of chromosomally female mice transgenic for *Sry*. *Nature* **351,** 117-21.

Lahn B. & Page D. (1999) Four evolutionary strata on the human X chromosome. *Science* **286,** 964-7.

Lardone M.C., Parada-Bustamante A., Ebensperger M., Valdevenito R., Kakarieka E., Martinez D., Pommer R., Piottante A. & Castro A. (2011) *DAX-1* and *DAX-1A* expression in human testicular tissues with primary spermatogenic failure. *Mol Hum Reprod* doi: 10.1093/molehr/gar051.

Lear T.L. & Bailey E. (2008) Equine clinical cytogenetics: the past and the future. *Cyto Genome Res* **120,** 42-9.

Lin Y., Philibert P., Ferraz-de-Souza B., Kelberman D., Homfray T., Albanaese A., Molini V. Sebire N.J., Einaudi S., Conway G.S., Hughes I.A., Jameson J.L.,

Sultan C., Dattani M.T. & Achermann J.C. (2007) Heterozygous missense mutations in steroidogenic factor 1 (*SF1/Ad4BP, NR5A1*) are associated with 46,XY disorders of sex development with normal adrenal function. *J Clin Endocrin Metab* **92,** 991-9.

Lindgren G., Backström N., Swinburne J., Hellborg L., Einarsson A., Sandberg K., Cothran G., Vilà C., Binns M. & Ellegren H. (2004) Limited number of patrilines in horse domestication. *Nat Genet* **36,** 335-6.

Liu W.S., Mariani P., Beattie C.W., Alexander L.J. & Ponce De Leon F.A. (2002) A radiation hybrid map for the bovine Y Chromosome. *Mamm Genome* **13,** 320-6.

Luo X., Ikeda Y. & Parker K. (1994) A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell* **77,** 481-90.

Lykkjen S., Dolvik N.I., McCue M.E., Rendahl A.K., Mickelson J.R. & Roed K.H. (2010) Genome-wide association analysis of osteochondrosis of the tibiotarsal joint in Norwegian Standardbred trotters. *Anim Genet* **41,** 111-20.

Makinen A., Hasegawa T., Makila M. & Katila T. (1999) Infertility in two mares with XY and XXX sex chromosomes. *Equine Veterinary Journal* **31,** 346-9.

Marchal J.A., Acosta M.J., Bullejos M., Diaz de la Guardia R. & Sanchez A. (2003) Sex chromosomes, sex determination, and sex-linked sequences in Microtidae. *Cytogenet Genome Res* **101,** 266-73.

Marchina E., Gambera A., Spinelli E., Clerici P., Scagliola P., Sartori E. & Barlati S. (2009) Identification of a new mutation in the *SRY* gene in a 46,XY woman with Swyer syndrome. *Fertility and Sterility* **91,** 932.e937-11.

Margarit E., Guillen A., Rebordosa C., Vidal-Toboada J., Sanchez M., Ballesta F. & Oliva R. (1998) Identification of conserved potentially regulatory sequences of the *SRY* gene from 10 different species of mammals. *Biochem Biophys Res Commun* **245,** 370-5.

Marshall Graves J.A. (2008) Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annu Rev Genet* **42,** 565-86.

Matson C.K., Murphy M.W., Sarver A.L., Griswold M.D., Bardwell V.J. & Zarkower D. (2011) *DMRT1* prevents female reprogramming in the postnatal mammalian testis. *Nature* **476,** *101-4.*

McClung C.E. (1902) The accessory chromosome-sex determinant? *Biological Bulletin* **3,** 43-84.

McCue M., Mickelson J., Bannasch D., Penedo C., Bailey E., Binns M., Distl O., Guerin G., Hasegawa T., Hill E., Leeb T., Lindgren G., Roed K., Swinburne J., Tozaki T., Vaudin M. & Wade C.M. (2009a) The horse gentrain project: initial evaluation of the Equine SNP50 BeadChips. 8th Dorothy Russell Havemeyer Foundation International Equine Genome Mapping Workshop, July 22-25, The West Wing at Ickworth, Suffolk, UK, pg. 11.

Melsky, R. (2009) Pacer Arizona Helen reclassified as "horse." *theHORSE.com* http://www.thehorse.com/ViewArticle.aspx?ID=14677.

Michala L., Goswami D., Creighton S.M., & Conway G.S. (2008) Swyer syndrome: presentation and outcomes. *Bio J Genet* **115,** 737-41.

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

Milliken J.E., Paccamonti D.L., Shoemaker S. & Green W.H. (1995) XX male pseudohermaphroditism in a horse. *J Am Vet Med Assoc* **207,** 77-9.

Morais da Silva S., Hacker A., Harley V., Goodfellow P., Swain A. & Lovell-Badge R. (1996) *Sox9* expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. *Nature* **14,** 62-8.

Muller H.J. (1914) A gene for the fourth chromosome of Drosophila. *J Exp Zoology* **17,** 325-36.

Murphy W.J., Sun S., Chen Z., Pecon-Slattery J. & O'Brien S.J. (1999) Extensive conservation of sex chromosome organization between cat and human revealed by parallel radiation hybrid mapping. *Genome Res* **9,** 1223-30.

Niimura Y. & Nei M. (2007) Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* **2,** e708. doi:10.1371/journal.pone.0000708.

Page D.C., Mosher R., Simpson E.M., Fisher E.M.C., Mardon G., Pollack J., McGillivray B., de la Chapelle A. & Brown L.G. (1987) The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* **51,** 1091-104.

Pailhoux E., Cribiu E.P., Parma P. & Cotinot C. (1995) Molecular analysis of an XY mare with gonadal dysgenesis. *Hereditas* **122,** 109-12.

Paria N., Raudsepp T., Wilkerson A.J.P., O'Brien P.C.M., Ferguson-Smith M.A., Love C.C., Arnold C., Rakestraw P., Murphy W.J. & Chowdhary B.P. (2011) A gene

catalogue of the euchromatic male-specific region of the horse Y chromosome: comparison with human and other mammals. *PLoS ONE* **6,** e21374.

Parker K.L., Rice D.A., Lala D.S., Ikeda Y., Luo X., Wong M., Bakke M., Zhao L., Frigeri C., Hanley N.A., Stallings N. & Schimmer B.P. (2002) Steroidogenic factor 1: an essential mediator of endocrine development. *The Endocrin Soc* **57,** 19-36.

Piprek R.P. (2009) Genetic mechanisms underlying male sex determination in mammals. *J Appl Genet* **50,** 347-60.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J. & Sham P.C. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Human Genet* **81**.

Qu Z., Chowdhary B.P., Raudsepp T., & Adelson D.L. (2011) Design of an equine whole genome tiling array. *Plant and Animal Genome XIX*, January 15-19, San Diego, CA, USA, W214.

Quignon P., Giraud M., Rimbault M., Lavigne P., Tacher S., Morin E., Retout E., Valin A., Lindblad-Toh K., Nicolas J. & Galibert F. (2005) The dog and rat olfactory receptor repetoires. *Genome Biol* **6,** R83.

Raudsepp T. & Chowdhary B.P. (2008) FISH for mapping single copy genes. *Methods in Molec Biol: Phylogenomics* **422,** 31-49.

Raudsepp T., Durkin K., Lear T.L., Das P.J., Avila F., Kachroo P. & Chowdhary B.P. (2010) Molecular heterogeneity of XY sex reversal in horses. *Animal Genet* **41,** 41-52.

Raudsepp T., Santani A., Wallner B., Srinivas R.K., Ren C., Zhang H., Womack J.E., Skow L.C. & Chowdhary B.P. (2004) A detailed physical map of the horse Y chromosome. *PNAS* **101,** 9321-6.

Raychaudhuri S., Plenge R.M., Rossin E.J., Ng A.C.Y., International Schizophrenia Consortium, Purcell S.M., Sklar P., Scolnick E.M., Xavier R.J., Altshuler D. & Daly M.J. (2009) Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLOS Genetics* **5,** 1-15.

Redon R., Ishikawa S., Fitch K.R., Feuk L., Perry G.H. *et al.* (2006) Global variation in copy number in the human genome. *Nature* **444,** 444-54.

Ross M.T., Graham D.V., Coffey A.J., Scherer S., McLay K. *et al.* (2005) The DNA sequence of the human X chromosome. *Nature* **434,** 325-37.

Ross D.G.F., Bowles J., Koopman P. & Lehnert S. (2008) New insights into *SRY* regulation through identification of 5' conserved sequences. *BMC Molec Biol* **9,** 85.

Rozen S. & Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols: Methods in Molecular Biology* **132,** 365-86.

Rozen S., Skaletsky H., Marszalek J., Minx P., Cordum H., Waterston R., Wilson R. & Page D. (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423,** 873-6.

Salehi L.B., Scarciolla O., Vanni G.F., Nardone A.M., Frajese G., Novelli G. & Stuppia L. (2006) Identification of a novel mutation in the *SRY* gene in a 46,XY female patient. *Eur J Med Genet* **49,** 494-8.

Sarafoglou K. & Ostrer H. (2000) Familial sex reversal: a review. *J Clin Endo Metab* **85,** 483-93.

Schimmer B.P. & White P.C. (2010) Minireview: steroidogenic factor 1: its roles in differentiation, development, and disease. *Molec Endocrin* **24,** 1322-37.

Schmahl J., Kim Y., Colvin J.S., Ornitz D.M. & Capel B. (2004) *Fgf9* induces proliferation and nuclear localization of *FGFR2* in Sertoli precursors during male sex determination. *Dev* **131,** 3627-36.

Sekido R. & Lovell-Badge R. (2008) Sex determination involves synergistic action of *SRY* and *SF1* on a specific *Sox9* enhancer. *Nature* **453,** 930-4.

Shahid M., Dhillon V.S., Hussain Z. *et al.* (2008) Analysis of the *SRY* gene in two sex-reversed XY sisters identifies two new novel point mutations in the high mobility group box domain. *Fertility and Sterility* **90,** 1191-8.

Sharp A.J., Wachtel S.S. & Benirschke K. (1980) H-Y antigen is a fertile XY female horse. *J Reproduction and Fertility*. **58,** 157-60.

Shriver M.D., Mei R., Parra E.J., Sonpar V., Halder I., Tishkoff S.A., Schurr T.G., Zhadanov S.I., Osipova L.P., Brutsaert T.D., Friedlaender J., Jorde L.B., Watkins

W.S., Bamshad M.J., Gutierrez G., Loi H., Matsuzaki H., Kittles R.A., Argyropolous G., Fernandez J.R., Akey J.M. & Jones K.W. (2005) Large-scale SNP analysis reveals clustered and continuous patterns of human genetic variation. *Human Genomics* **2,** 81-9.

Silvers W.K. & Wachtel S.S. (1977) H-Y antigen: behavior and function. *Science* **195,** 956-60.

Sinclair A.H., Berta P., Palmer M.S., Hawkins J.R., Griffiths B.L., Smith M.J., Foster J.W., Frischauf A.M., Lovell-Badge R. & Goodfellow P. (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346,** 240-4.

Skaletsky H., Kuroda-Kawaguchi T., Minx P.J., Cordum H.S., Hillier L., Brown L.G., Repping S., Pyntikova T., Ali J., Bieri T., Chinwalla A., Delehaunty A., Delehaunty K., Du H., Fewell G., Fulton L., Fulton R., Graves T., Hou S., Latrielle P., Leonard S., Mardis E., Maupin R., McPherson J., Miner T., Nash W., Nguyen C., Ozersky P., Pepin K., Rock S., Rohlfing T., Scott K., Schultz B., Strong C., Tin-Wollam A., Yang S., Waterston R.H., Wilson R.K., Rozen S. & Page D.C. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423,** 825-37.

Smit A.F.A., Hubley R. & Green P. (1996-2010) *RepeatMasker Open-3.0* <http://www.repeatmasker.org>.

Stevens N.M. (1905) Studies in spermatogenesis with especial reference to the "accessory chromosome." *Carnegie Institute of Washington Publication* **36,** 1-33.

Sutton W.S. (1902) On the morphology of the chromosome group in *Brachystola magna*. *Biological Bulletin* **4,** 24-39.

Swain A., Narvaez V., Burgoyne P., Camerino G. & Lovell-Badge R. (1998) *DAX1* antagonizes *SRY* action in mammalian sex determination. *Nature* **391,** 761-7.

Switonski M., Chmurzynska A., Szczerbal I., Lipczynski A., Yang F. & Nowicka-Posluszna A. (2005) Sex reversal syndrome (64,XY; SRY-positive) in a mare demonstrating masculine behavior. *J Anim Breed Genet* **122,** 60-3.

Swyer G.I. (1955) Male pseudohermaphroditism: a hitherto undescribed form. *Br Med J* **2,** 709-12.

Takahashi R.H., Grigliatti T.A., Reid R.E. & Riggs K.W. (2009) The effect of allelic variation in aldo-keto reductase 1C2 on the *in vitro* metabolism of dihydrotestosterone. *J Pharmacol and Exp Therapeutics* **329,** 1032-9.

Tannour-Louet M., Han S., Corbett S.T., Louet J., Yatsenko S., Meyers L., Shaw C.A., Kang S.L., Cheung S.W. & Lamb D.J. (2010) Identification of *de novo*, copy number variants associated with human disorders of sexual development. *PLoS ONE* **5,** 1-13.

Teyssèdre S., Dupuis M.C., Guérin G., Schibler L., Denoix J.M., Elsen J.M. & Ricard A. (2011) Genome-wide association studies for osteochondrosis in French trotters. *J Anim Sci* doi:10.2527/jas.2011-4031.

Tijo H. & Levan A. (1956) The chromosome number of man. *Hereditas.* **42,** 1-6.

Traut W., Sahara K. & Marec F. (2007) Sex chromosomes and sex determination in Lepidoptera. *Sex Dev* **1,** 332-46.

Vaiman D. & Pailhoux E. (2000) Mammalian sex reversal and intersexuality: deciphering the sex-determination cascade. *Trends in Genetics* **16,** 488-94.

Van Silfhout A., Boot A.M., Dijkhulzen T., Hoek A., Nijman R., Sikkema-Raddatz B. & van Ravenswaaij-Arts C.MA. (2009) A unique 970 kb microdeletion in 9q33.3, including the *NR5A1* gene in a 46,XY female. *Euro J Med Genet* **52,** 157-60.

Van Wilgenburg E., Driessen G. & Beukeboom L.W. (2006) Single locus complementary sex determination in Hymenoptera: an "unintelligent" design? *Frontiers in Zoology* **3,** 1-15.

Wade C.M., Giulotto E., Sigurdsson S., Zoli, M., Gnerre S. *et al.* (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326,** 865-7.

Wallner B., Brem G., Müller M. & Achmann R. (2003) Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii* and *Equus caballus*. *Anim Genet* **34,** 453-6.

Watson J.M. (1990) Monotreme genetics and cytology and a model for sex-chromosome evolution. *Aus J Zoology* **37,** 385-406.

White S., Ohnesorg T., Notini A., Roeszler K., Hewitt J., Daggag H., Smith C., Turbitt E., Gustin S., van den Bergen J., Miles D., Western P., Arboleda V., Schumacher V., Gordon L., Bell K., Bengtsson H., Speed T., Hutson J., Warne G., Harley V.,

Koopman P., Vilain E. & Sinclair A. (2011) Copy number variation in patients with disorders of sex development due to 46,XY gonadal dysgenesis. *PLoS ONE* **6,** 1-10.

Wilhelm D. (2007) R-spondin1 – discovery of the long-missing mammalian female-determining gene? *BioEssays* **29,** 314-8.

Wilson E.B. (1905) The chromosomes in relation to determination of sex in insects. *Science* **22,** 500-2.

Wu N. & Veillette A. (2011) Immunology: magnesium in a signaling role. *Nature* **475,** 462-3.

Zhang X. & Firestein S. (2002) The olfactory receptor gene superfamily of the mouse. *Nat Neurosci* **5,** 124-33.

Zozulya S., Echevveri F. & Nguyen T. (2001) The human olfactory receptor repertoire. *Genome Biol* **2,** RESEARCH0018.

VITA

Name: Erica Fang

Address: 5503 Caversham Dr. Houston, TX 77096

Email Address: efang387@gmail.com

Education: B.S., Genetics, Texas A&M University, 2009