# STUDY OF GENOMIC COPY NUMBER VARIATION IN EQUINE HEALTH AND DISEASE

A Dissertation

by

SHARMILA GHOSH

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

DOCTOR OF PHILOSOPHY

| | |
|---|---|
| Chair of Committee, | Terje Raudsepp |
| Committee Members, | Ernest Gus Cothran |
| | Penny K. Riggs |
| | Bhanu P. Chowdhary |
| | James Cai |
| Head of Department, | Evelyn Castiglioni |

August 2014

Major Subject: Biomedical Sciences

# ABSTRACT

This is a study of copy number variations (CNVs) in the horse genome to gain knowledge about the role of CNVs in equine biology, and their contribution to complex diseases and disorders.

We constructed a 400K whole-genome tiling array and applied it for the discovery of CNVs in 38 normal horses of 16 diverse breeds, and the Przewalski horse. Altogether, 258 CNV regions (CNVRs) were identified across all autosomes, chrX, and chrUn. The CNVRs comprised 1.3% of the horse genome with chr12 being most enriched. American Miniature Horses had the highest and American Quarter Horses the lowest number of CNVs in relation to Thoroughbred references. The Przewalski horse was similar to native ponies and draft breeds. About 20% of CNVRs were intergenic, while 80% involved 750 annotated genes with molecular functions predominantly in sensory perception, immunity, and reproduction.

The findings were integrated with previous CNV studies in the horse to generate a composite genome-wide dataset of 1476 CNVRs. Of these, 301 CNVRs were shared between studies, while 1174 were novel and require further validation. Integrated data revealed that only 41 out of over 400 breeds of the domestic horse have been analyzed for CNVs, whereas this study added 11 new breeds.

The composite CNV dataset served as a foundation for the discovery of variants contributing to Recurrent Airway Obstruction (RAO) and XY disorders of sexual development (DSDs), such as cryptorchidism and XY sex reversal. In 16 RAO affected

horses 363 CNVRs were identified, of which 31 were novel and not found in healthy horses. A deletion in *SPI2* and *SERPINA1* was studied in detail because the genes are involved in respiratory diseases in human. In horses with XY DSDs, over 50 novel CNVRs were identified including deletions of functional interest in the pseudoautosomal region and the *ATRX* gene. A potentially causative homozygous deletion in chr29 disrupting *AKR1C* genes with functions in sex hormone metabolism was shared between a cryptorchid and two sex reversal horses.

The findings effectively improved the knowledge about CNVs in horses, in health and disease, and generated resources for future studies.

# ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Raudsepp, and my committee members, Dr. Cothran, Dr. Cai, Dr. Chowdhary, and Dr. Riggs, for their guidance and support throughout the course of this research.

I want to thank my friends, Anindita Das, Nandina Paria, Anuradha Ghosh, Anrini Majumder who always helped me to go through all kind of situations. I also like to thank my colleagues, Pranab Jyoti Das, Felipe Avila, Fahad Alshanbari, Daisy Johnson, Priyanka Kachroo, Jana Caldwell, Joana Rocha, Alex Trott, Jan Janecka, and Samantha Steelman and the department faculty and staff for making my time at Texas A&M University a great experience. I also like to extent my gratitude to Ryan Doan, for his guidance on a very important part of my research. A special thanks to Dr. Jane Welsh, who helped a lot in difficult situations.

Finally, thanks to my parents and brother for their encouragement, patience, and love.

# NOMENCLATURE

| | |
|---|---|
| CNV | Copy Number Variants |
| CNVR | Copy Number Variable Region |
| aCGH | array Comparative Genomic Hybridization |
| qPCR | quantitative PCR |
| FISH | Fluorescent in situ Hybridization |
| RAO | Recurrent Airway Obstruction |
| DSD | Disorder of Sexual Development |
| CO | Cryptorchids |

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

Page

# CHAPTER I

# INTRODUCTION AND LITERATURE REVIEW

## The horse: domestication, breeds, importance

Horses were domesticated over 5500 years ago (Vila et al. 2001; Levine 2005; Outram et al. 2009; Orlando et al. 2013) and occupy a special place amongst domestic species. Unlike cattle, pigs, sheep, goats and poultry which are mostly used as a source of food (milk, meat and eggs), horses have been used as a source of power and transport. Thanks to natural athleticism and endurance for physical activities even in rough environmental conditions with scarce resources, horses were valuable for transportation in older times. Today, they are mainly used for racing and leisure industry, and as companion animals. Since domestication, humans have bred horses for several different purposes, resulting in over 400 horse and pony breeds (Hendricks 1995) which can be divided into three big groups: warmblood horses for racing and riding, coldblood horses for draft and heavy duty jobs, and native ponies, which are often adapted to local specific environmental and climatic conditions (Austen et al. 2008).

Horses have a great impact on the economy of the U.S.A. and throughout the world; the revenue of the US equine industry in 2010 was $1.6 billion as per IBIS-World Industry Report 2013. As per 2005 American Horse Council study, in the United States, the horse industry directly produces goods and services in excess of $39 billion and has a total impact of $102 billion on U.S. GDP.

Direct employment and additional jobs generated through spending, leads to a total employment impact of 1.4 million FTE jobs. Approximately 34% of horse owners have a household annual income of less than $50,000, and over 70% of horse owners live in communities with <50,000 population. Thus the equine industry has a strong impact on the rural economy and supports a considerable proportion of people in the relatively lower income group.

Being a species with great economic importance, good health and performance of horses are of high priority to the owners, breeders and the equine industry. Therefore, the primary goal of genome analysis in horses is to identify genetic factors governing equine diseases, congenital disorders and traits of biological importance (e.g., disease resistance) and of human interest (e.g., coat color, athletic performance), so that molecular tools can be developed for the improvement of equine health and performance.

**The horse genome**

The nuclear genome of the horse is packaged into 64 chromosomes (2n=64) - 31 pairs of autosomes and one pair of sex chromosomes (XX in females and XY in males). In 1997 an International System for Chromosome Nomenclature of the Domestic Horse (ISCNH 1997) was developed. This System determined the arrangement of horse chromosomes into a standard karyotype and established a banding nomenclature for chromosome identification.

The size of the horse genome is about 2.7 billion DNA base pairs with at least 20,449 protein-coding genes as revealed by the current horse genome sequence assembly, EquCab2, which is based on the genome of a Thoroughbred mare (Wade et al. 2009).

The mitochondrial genome of the horse is 16.6 kilo base pairs (kb) in size and consists of 13 protein-coding genes, 22 tRNA genes (transfer RNA) and 2 rRNA genes (ribosomal RNA) (Xu and Arnason 1994). Similarly to other mammals, the mitochondrial genome of the horse is maternally inherited. Thus, horses of the same maternal ancestry share similar mitochondrial sequences (Bower et al. 2013). Studies of mitochondrial DNA have been widely used to explore the origins and relationships of horse breeds and populations (Hill et al. 2002; Cieslak et al. 2010; Khanshour and Cothran 2013).

**Horse genome mapping**

Until whole genome sequencing, the primary sources of information about the molecular organization of the horse genome have been various types of maps (Chowdhary et al. 2008; Chowdhary and Raudsepp 2008; Raudsepp et al. 2012; Raudsepp and Chowdhary 2013). Gene mapping in horses started with linkage analysis which is based on meiotic recombination and linkage maps show linear order and relative distance between markers.

The first linkage mapped gene was *G6PD* (Glucose 6 Phosphate Dehydrogenase) which was assigned to the X-chromosome almost 5 decades ago (Trujillo et al. 1965; Mathai et al. 1966). Thereafter linkage maps were constructed for autosomal regions (Sandberg 1974; Andersson and Sandberg 1984; Sandberg and Andersson 1984) and for the whole genome (Lindgren et al. 1998; Guerin et al. 1999; Swinburne et al. 2000; Penedo et al. 2005; Swinburne et al. 2006).

The first physical chromosome maps for the horse were constructed by synteny mapping (Lear et al. 1992; Williams et al. 1993; Bailey et al. 1995; Raney et al. 1998; Shiue et al. 1999). The method is based on somatic cell hybrid technology and assigns genes and markers into synteny groups. However, in the early years of gene mapping, neither linkage nor synteny maps were able to anchor markers to specific chromosomes or chromosomal regions. This limitation was overcome by another physical mapping technique - cytogenetic mapping, also known as *in situ* hybridization (ISH) or fluorescence *in situ* hybridization (FISH).

The first ISH mapped loci were the equine major histocompatibility complex (*ELA*) (Ansari et al. 1988; Makinen et al. 1989) and glucose phosphate isomerase (*GPI*) (Harbitz et al. 1990), while hemoglobin alpha (*HBA*) was the first gene mapped by FISH (Oakenfull et al. 1993). Currently, over 1000 genes and markers have been mapped in the horse genome by FISH (Raudsepp et al. 2008).

In the early 2000s, another somatic cell hybrid based technique - radiation hybrid (RH) mapping - was adopted for the horse (Chowdhary et al. 2002) resulting in the development of high-resolution physical maps for individual chromosomes (Lee et al.

2004; Raudsepp et al. 2004a; Raudsepp et al. 2004b; Gustafson-Seabury et al. 2005; Goh et al. 2007), as well as for the whole genome (Chowdhary et al. 2003; Raudsepp et al. 2008). The most comprehensive of these is one 4000 marker containing 2$^{nd}$ generation RH map (Raudsepp et al. 2008) which integrates all linkage and cytogenetic maps, and aligns about 2000 loci with the human genome.

Another type of high-resolution maps were based on contigs of Bacterial Artificial Chromosome (BAC) clones and were constructed for selected genomic regions, such as the major histocompatibility complex, MHC (Gustafson et al. 2003), pseudoautosomal region, PAR (Raudsepp and Chowdhary 2008b), and a small region in chromosome 21 (Brinkmeyer-Langford et al. 2008). Attempts were also made to construct a WG BAC fingerprint map (Leeb et al. 2006) with the aim to guide and validate the horse WG sequence assembly (Wade et al. 2009).

An integral part of constructing various maps was the development of mapping resources, such as pedigree material (Lindgren et al. 1998; Guerin et al. 1999; Penedo et al. 2005), polymorphic microsatellite markers for linkage mapping (Lindgren et al. 1998; Guerin et al. 1999; Mittmann et al. 2010a), markers for PCR-based synteny and RH mapping (Chowdhary et al. 2003; Swinburne et al. 2006; Raudsepp et al. 2008), and large insert clone libraries for cytogenetic and BAC-based mapping. The latter include one cosmid library (Breen et al. 1997) and three male genomic BAC libraries: INRA (Milenkovic et al. 2002), CHORI-241 BAC library, and TAMU BAC library (Rubes et al. 2009).

All these maps have served as valuable tools for the isolation of genes and markers associated with equine diseases and economically important traits. Besides, the linkage (Penedo et al. 2005; Swinburne et al. 2006), radiation hybrid (RH) and cytogenetic maps (Raudsepp et al. 2008) have been instrumental for anchoring and orienting WG sequence contigs and scaffolds on horse chromosomes (Wade et al. 2009). The need for good maps will be even more pronounced in re-sequencing projects to assemble millions of short reads produced by next generation sequencing technologies (NGS) (Mardis 2008; Huddleston et al. 2014). The whole genome maps remain to offer a well-tested resource for the analysis of genome architecture, function, and evolution.

**The genome sequence of the domestic horse**

Sequencing and assembly of the genome of a Thoroughbred mare *Twilight* in 2009 (Wade et al. 2009) marked the start of a new era in equine genomics. The assembly of 2.4 billion DNA base-pairs provided the first detailed characterization of the horse genome and serves currently as a reference for all other horse genome sequencing projects.

The horse genome was sequenced with coverage of 6.97X. Although the total length of the assembly is 2.43 billion base pairs, with gaps the size of the horse genome is estimated to be 2.7 billion base pairs (Wade et al. 2009). Using gene predictions from mammalian and non-mammalian vertebrates, 20,449 protein-coding genes, 2142 non-

coding genes, and 4400 pseudogenes and retrotransposed genes were annotated. Gene annotation showed that among 20,322 equine genes, 16,617 have orthologs in humans, 17,106 in mouse, and 16,159 in dogs. Besides determining the DNA sequence for the horse, the project provided data regarding the functions and evolution of the genome, and the genomics of equine populations (Wade et al. 2009). Comparison of the horse genome with the human showed that horse chromosomes have undergone few evolutionary rearrangements: 53% of equine chromosomes show conserved synteny to a single human chromosome. There are a total of 403 syntenic segments between human and horse (Wade et al. 2009).

About 95% of the genome sequences are assigned to the 31 autosomes and the X chromosome. The remaining 5% are not assigned to any chromosomes and considered as unassigned chromosome, "chrUn". The horse genome is relatively repetitive but has little segmental duplications. About 46% of the horse genome comprises repetitive elements of which 19% are long interspersed elements (LINEs), 7% are short interspersed elements (SINEs) and the remaining are novel equine repetitive elements. Detailed analysis of equine repetitive elements showed that besides typical eutherian mammalian repeats, the horse genome also contains a significant number of hybrid repeats and clade-specific LINEs (Adelson et al. 2010).

A novel feature of the horse genome was the discovery of an evolutionarily new centromere in ECA11. As this functional centromere did not contain any major horse satellite sequences, it was concluded that centromeric satellite DNA is not obligatory for centromere functions.

Importantly, the horse genome sequencing project also determined DNA sequence variation across horses by identifying 1,154,177 short variants - single nucleotide polymorphisms (SNPs) and indels. Single nucleotide polymorphisms are the most common types of genetic variations in mammalian populations and typically involve an alteration in a single nucleotide. For example, in humans, SNPs occur once in every 300 nucleotides suggesting that there are approximately 10 million SNPs in the human genome (NCBI). Most SNPs have no effect on health or development, however, when SNPs occur within a gene or in a regulatory region, they may affect phenotype more directly. The discovery of SNPs in horses was based on 7 horses of diverse breeds and "Twilight". These 7 horses are breeds Akhal-Teke, Andulasian, Arabian, Icelandic horse, Quarter Horse, Standardbred, and a second Thoroughbred. On average, there is one SNP per 2000 nucleotides in horses, suggesting the presence of over one million SNPs in the genome. Indeed, since then, over 3 million SNPs have been discovered by sequencing additional genomes of individual horses (Orlando et al. 2011; Doan et al. 2012b; Orlando et al. 2013).

Besides SNPs, the horse genome sequence draft assembly showed that segmental duplications make up about 0.5 % of the equine genome and are most abundant in chromosome 25 (Wade et al. 2009). Segmentally duplicated areas are known as hotspots for non-allelic homologous recombination (NAHR), which is one of the main causes for the formation of copy number variable (CNV) regions (Hurles and Lupski 2006). These are stretches of DNA larger than 1 kb, although smaller CNVs are also common (Mills et al. 2011; Xu et al. 2011) that are involved in large scale genomic rearrangements, such

as deletions, insertions, duplications, inversions, translocations and mobile element transposition.

These rearrangements result in structural differences between genomes generating the genetic basis for normal phenotypic variations. At the time of the draft assembly of the horse genome in 2009, it was known that gray coat color is caused by a CNV in *STX17* (syntaxin-17) gene (Rosengren Pielberg et al. 2008). Thus, in the study by Wade and colleagues (2009) a breed wise CNV analysis was performed in horses of different coat colors as a proof of the principle. Since CNVs and their biological importance is the central theme of this dissertation, detailed description of CNV research in horses and other mammals will be provided later in "Genomic Copy Number Variation (CNV)" and "Present status of CNV research in the horse".

The current horse reference genome sequence assembly draft was produced by shotgun sequencing of the genome of a Thoroughbred mare "*Twilight*" using Sanger technology. During the recent years, taking advantage of various next generation sequencing (NGS) platforms, whole genomes of several more horses have been sequenced. The first was sequencing a Quarter Horse mare using Illumina platform (Doan et al. 2012b). Very recently single molecule sequencing technology (Helicos HeliScope and Illumina GAIIX) was applied of two ancient horses - a 700,000 years old early middle Pleistocene horse and a 43,000 years old pre-domestication late-Pleistocene horse, 5 modern horses (Arabian, Icelandic, Norwegian Fjord, Standardbred and Thoroughbred), a Przewalski's horse and a donkey (Orlando et al. 2011; Doan et al. 2012b; Orlando et al. 2013). Analysis of WG sequences of additional horses, by PacBio

and Illumina, is in progress. Altogether, these data are aimed to improve the current EcuCab2 assembly and upgrade it into version EcuCab3 (Kalbfleisch et al. 2013).

## The development and application of advanced genome analysis tools

Achievements in horse genomics, in particular the availability of the reference sequence EcuCab2 and additional WG or partial sequences from multiple individuals, lay the foundation for the development of advanced tools for further analysis of the organization and function of the horse genome. These include a variety of array-based platforms and improved panels of polymorphic markers for WG scan.

### Studies of genome function

The first gene expression arrays for the horse were constructed even before the whole genome was sequenced and were based on the sequence information of cDNA libraries and ESTs (Gu and Bertone 2004; Smith et al. 2006; Huang et al. 2008). The availability of the reference genome with over 20,000 annotated protein coding genes (Wade et al. 2009), essentially improved the production of whole genome cDNA and oligoarrays (Bright et al. 2009; Glaser et al. 2009; Mienaltowski et al. 2009; Noschka et al. 2009). These platforms have been used to study the dynamics of gene expression in normal equine cells and tissues: synoviocytes (Gu and Bertone 2004), cartilage (Huang et al. 2008), testes and sperm (Das et al. 2013a), and endometrium (Gebhardt et al. 2012;

Kamm et al. 2013; Moyo et al. 2013). Also, the arrays have been critical for the discovery of gene expression changes in pathologies: osteoarthritis (Smith et al. 2006), chondrocyte surface lesions (Mienaltowski et al. 2009), early stages of laminitis (Noschka et al. 2009), and expression of *Rhodococcus equi* in neonatal foals (Kachroo et al. 2013). However today, gene expression arrays have been largely replaced by RNA-seq technology which is cost efficient and provides most comprehensive information about the expression profile of any tissue or cell type (Coleman et al. 2013). Most importantly, RNA-seq data from adult and embryonic equine tissues is a valuable resource for improving functional annotation of the horse genome and for the discovery of novel protein coding and regulatory RNA genes.

Studies of the genome structure

Even although there is just one reference genome for the horse, genomic DNA sequences of individual horses vary in multifarious ways. These variants include polymorphisms of single nucleotides (SNPs and indels) and short-tandem repeats (microsatellites), segmental duplications and copy number variants. Such inter-individual DNA sequence differences can be used for genome wide association studies (GWAS) for detecting genotype-phenotype associations and discovering mutations underlying genetic diseases/disorders and traits of interest. Thus, development of cutting edge tools for the discovery of these associations has been an important part of equine genomics during the past years.

***SNP arrays***

Single nucleotide polymorphisms in the horse genome are considered as the next generation markers for parentage testing, breed diversity analysis, and GWAS of Mendelian and complex traits (Swinburne et al. 2009; Wade et al. 2009). In the past, candidate genes for equine genetic disorders (Tryon et al. 2007; Young et al. 2007) or traits of interest, like fertility (Hamann et al. 2007) or athletic performance (Schroder et al. 2011; McGivney et al. 2012) were directly analyzed for SNPs to discover causative associations. However, this approach requires previous knowledge about the genes and is laborious. Therefore, one of the most awaited by-products of the genome sequencing was the discovery of over million SNPs (Wade et al. 2009) and the construction of the first generation SNP chip, known as Illumina Equine SNP50 BeadChip (McCue 2009; McCue et al. 2012). The chip comprises 54,602 highly informative (polymorphic in many breeds) SNPs uniformly distributed across the equine genome. The mean inter-SNP distance across the genome is about 43 kb, and 95% of informative SNPs are apart less than 150 kb. The SNP50 chip was successfully used for the proof of principle mapping of chestnut (*MC1R*) and black (*ASIP*) coat color loci detecting a ~350 kb and ~200 kb conserved across-breeds haplotypes, respectively (McCue 2009). Since then, the SNP50 Beadchip has been extensively used for a variety of studies that can be broadly grouped into two: i) populations studies searching for similarities, differences and relationships between different horse breeds, and for signatures of domestication (Petersen et al. 2013a; Petersen et al. 2013b), and ii) studies looking for candidate genes and mutations underlying equine diseases/disorders and traits of interest (see Table 1

below). For example, a population study with 814 horses of 36 different breeds discovered candidate regions for selection during horse domestication. These included a 5.5Mb region in ECA18 with the myostatin (*MSTN)* gene showing highest selection in Paint and Quarter Horse breeds; a shared, 186 kb haplotype in ECA23with *DMRT2* and *DMRT3* genes; and a locus at ECA11 possibly responsible for size in Draft breeds and Miniature horses (Petersen et al. 2013b).

Despite of this success, it was soon recognized that the number and representation of SNPs on the chip needed improvement. Thus, in 2011 the SNP50 Beadchip was modified by adding new SNPs, removing less informative ones and enhancing the coverage of ECAX, regions involved in coat colors, and the MHC in ECA20 (Brooks and Bellone 2013; McCue and Mickelson 2013; McCue 2013). The SNP70 Beadchip was produced by Illumina and is currently available for researchers. Further, work is in progress to develop a high density SNP chip with about 600,000 probes. This is an international effort and started with identifying 20-30 million SNPS from WG sequences of a diverse set of horses.

Based on WG sequence data of 166 horses representing 32 diverse breeds, 5 million SNPs have been selected and sent to Affymetrix for quality control. It is expected that there will be 2 million SNPs suitable for the production of a test array. The final goal is to construct a chip with 640,000 SNPs for research. With such a SNP array in place, horse genomics will have high throughput genome analysis tools comparable to those available for humans and cattle.

## Table 1: Summary of GWAS studies using Equine SNP50 Beadchip.

| Condition | Breeds | Genes/mutations/associations | Reference |
|---|---|---|---|
| Negative effects on fitness | Thoroughbred | Inbreeding increased post-1996 and coincides with the introduction of stallions covering larger numbers of mares. | (Binns et al. 2012) |
| Athletic performance (racing distance) | Thoroughbred | *MSTN* in ECA18 | (Binns et al. 2010; Hill et al. 2010) |
| Body size | Many breeds | ECA1, 8, 9 | (Metzger et al. 2013a) |
| Dwarfism | Friesian | 2 Mb in ECA14 | (Orr et al. 2010) |
| Foal Immunodeficiency Syndrome; Fell Pony Syndrome | Fell Pony | *SLC5A3* in ECA26, causative mutation | (Fox-Clipsham et al. 2011a; Fox-Clipsham et al. 2011b) |
| Fracture risk | Thoroughbred | ECA1, 18 | (Blott et al. 2014) |
| Guttural pouch tympany | Arabian,<br>German Warmblood | ECA15 – Arabian;<br>ECA3- German Warmblood | (Metzger et al. 2012) |
| Height | Hanoverian, 42 other breeds | *LCORL* – ECA3 a candidate locus | (Metzger et al. 2013b) |
| Impaired acrosome reaction | Thoroughbred | *FKBP6* - ECA13 a susceptibility locus | (Raudsepp et al. 2012) |
| Insect bite hypersensitivity | Icelandic horse<br>Exmoor pony | ELA class II | (Andersson et al. 2012b) |
| Lavender Foal Syndrome, LFS | Arabian | Causative mutation in *MYO5A* ECA1 | (Brooks et al. 2010) |
| Locomotion, gait | Icelandic horse,<br>Trotters,<br>Gaited breeds | Causative mutation in *DMRT3* – ECA23 | (Andersson et al. 2012a) |
| Maxillary Prognathism | Various breeds | ECA13. | (Signer-Hasler et al. 2014) |
| Osteochondritis dissecans | Thoroughbred | ECA3 | (Corbin et al. 2012) |
| Osteochondrosis | Standardbred | ECA5, 10, 27, 28 | (Lykkjen et al. 2010) |
| Osteochondrosis | French trotter | ECA1, 2, 3, 13, 14, 15 | (Teyssedre et al. 2012) |
| Plantar osteochondral fragments | Standardbred | ECA1, 2, 7, 9, 11,27, 31, X | (Lykkjen et al. 2013) |
| Recurrent exertional rhabdomyolysis, RER | Thoroughbred | 13 Mb in ECA16 | (Fritz et al. 2012) |
| Recurrent laryngeal neuropathy | Warmblood,<br>Trotter,<br>Thoroughbred,<br>Draft | ECA21 and 31 | (Dupuis et al. 2011) |
| Recurrent laryngeal neuropathy | Various breeds | Duplication in ECA10 | (Dupuis et al. 2013) |
| Recurrent uveitis | German Warmblood | ECA20,18 | (Kulbrock et al. 2013) |
| Recurrent uveitis | Appaloosa | ECA1 and *ELA* (ECA20) | (Fritz et al. 2014) |
| Show-jumping | Hanoverian | ECA1, 3, 8, 9, 11, 17, 21 | (Schroder et al. 2012) |
| Susceptibility to equine arteritis virus | Thoroughbred,<br>American Saddlebred,<br>Standardbred, Quarter Horse | ECA11 | (Go et al. 2011) |
| Trisomies | Friesian<br>Welsh Pony | ECA27, 30 | (Holl et al. 2013) |

### *WG scanning panels*

Polymorphic microsatellites have, for years, been the markers of choice for linkage studies and whole genome scans. The success of these studies depends largely on the quality of panels made of informative microsatellite markers, strategically positioned over the genome. The available sequence assembly significantly improved these panels. For example, *in silico* analysis of EcuCab2.0 identified 21,781 microsatellite sequences (Mittmann et al. 2010c). Of these 72% are di-, 3.8 % tri-, 17.6 % tetra-, and 1.6% pentanucleotide repeats. The mean distance between these microsatellites in the horse genome is 112 kb. This collection of sequences was used to establish a highly polymorphic minimal microsatellite set for whole genome scans in warm- and coldblood horse breeds (Mittmann et al. 2010a). Markers from this panel have been used in whole genome scan for chronic pastern dermatitis in German draft horses (Mittmann et al. 2010b) and guttural pouch tympany in Arabian and German warmblood horses (Zeitz et al. 2009; Metzger et al. 2012).

### *Tiling arrays*

In the past few years, specialized tiling arrays have been constructed for the study of CNVs in the horse genome. The first such array comprised 418,000 60 bp oligonucleotides representing exons and UTRs of ~21,000 annotated horse protein coding genes RNA genes, pseudogenes, transposons, and retrotransposons. The array was produced by Illumina and has so far been used in one study (Doan et al. 2012a). Very recently, Wang and colleagues reported about the production of a NimbleGen WG

tiling array with 1.4 million probes. The array was used to study CNVs in 6 indigenous horse breeds from China and Mongolia (Wang et al. 2014b). Overall, tiling arrays are 'newcomers' in equine genomics and have, as yet, found limited application. This topic will be discussed in more detail in the following section "Present status of CNV research in the horse".

**Genomics of equine diseases and traits**

The goals of equine genomics worldwide are to identify mutations responsible for genetic disorders, and discover genes associated with simple (coat color) and complex (athletic performance, reproduction) traits. This will allow the improvement of breeding practices and the development of better tests for diagnostics (Chowdhary et al. 2008; Chowdhary and Raudsepp 2008; Finno et al. 2009; Swinburne et al. 2009). Before 'genome sequencing era' causative genes and mutations for a handful of monogenic equine traits (Finno et al. 2009) were discovered using gene maps, comparative approaches and pedigree analyses. Genetic tests are currently available for 14 coat colors (see Table 2 below), 9 monogenic disorders (see Table 3 below) and for the recently discovered gait-mutation in *DMRT3* (Andersson et al. 2012a). However, given that there over 200 equine traits and disorders listed in the database of Online Mendelian Inheritance in Animals (OMIA), much more work needs to be done. Furthermore, a large proportion of the traits important to the equine industry are complex being controlled by multiple genes and environmental factors. Examples of such conditions are

recurrent airway obstruction (RAO) and other respiratory and allergic disorders, osteochondrosis, laminitis, reproductive and behavioral disorders. The availability of horse genome reference sequence and new cutting edge pan-genomic analysis tools have made a breakthrough in the study of these conditions facilitating hypothesis free genome-wide discovery of associated genes or genomic regions. Genome wide SNP association studies have been conducted for about 20 equine conditions and disorders. Although, causative mutations have been found for just a few - Lavender Foal Syndrome (Brooks et al. 2010), Fell Pony Syndrome (Fox-Clipsham et al. 2011a; Fox-Clipsham et al. 2011b) and gait (Andersson et al. 2012a). Additionally, likely candidate/contributing genes have been proposed for some complex disorders – *MSTN* for racing distance (Hill et al. 2012; McGivney et al. 2012), *FKBP6* for impaired acrosome reaction (Raudsepp et al. 2012), and MHC Class II for insect bite hypersensitivity (Andersson et al. 2012b) (Table 3). For others, GWAS has identified one or more potentially associated genomic regions where the search for candidate genes will require additional efforts.

Taken together, despite the rapid and qualitative progress in equine disease and trait genomics, contributing molecular factors for many economically important complex traits and conditions remain unknown. This justifies the launch of alternative directions in research, taking advantage of other forms of inter-individual genome variation, such as copy number variants (CNVs).

**Table 2: Summary of genetic test available for 14 coat colors in horses.**

| Coat Color | Mode of inheritance | Mutation | References |
|---|---|---|---|
| Red Factor/Black Factor (Extension) | Autosomal recessive epistatic over black, | Missense mutation and loss-of-function mutation on *MCIR* gene in ECA 3 | (Rieder et al. 2001) |
| Agouti (Bay) | Autosomal dominant | Allele combinations in *MC1R* in ECA 3, and in *ASIP* in ECA 22 | (Rieder et al. 2001) |
| Agouti (Black) | Autosomal recessive | 11bp deletion at position 2174–2184; frameshift loss-of-function mutation in *ASIP* in ECA 22 | (Rieder et al. 2001) |
| Champagne Dilution | Autosomal dominant | Mutation in Exon 2 of *SLC36A1* in ECA 14 | (Cook et al. 2008) |
| Cream Dilution | Autosomal codominant | Mutation in *MATP* gene in ECA 21 | (Mariat et al. 2003) |
| Pearl Dilution, Barlink Factor | Autosomal recessive | Mutation in *MATP* gene in ECA 21 | |
| Silver Dilution | Autosomal dominant | Missense mutation *PMEL17* gene in ECA 6 | (Brunberg et al. 2006) |
| Silver Dilution | Incomplete dominance, homozygous for disease allele | SNP A>T and C>T at *PMEL* gene in ECA7 | (Andersson et al. 2013) |
| Gray | Autosomal dominant | 4.6-kb duplication in intron 6 of *STX17* gene in ECA 25 | (Rosengren Pielberg et al. 2008) |
| Appaloosa Coat Pattern / Leopard Print | Autosomal – incompletely – dominant | Autosomal dominant 1378 bp insertion in intron 1 of *TRPM1* in ECA 1. | (Bellone et al. 2010; Bellone et al. 2013) |
| Overo Lethal White Foal Syndrome (OWLS) | Autosomal recessive. | Point mutation that results in an isoleucine/lysine substitution at codon 118 of the *EDNRB* gene located in ECA 17. | (Hultgren 1982; McCabe et al. 1990; Metallinos et al. 1998) |
| Sabino 1 | Autosomal dominant | Single nucleotide polymorphism caused by a base substitution for T with A in intron 16 of *KIT* gene in ECA 3. | (Brooks and Bailey 2005) |
| Splashed White Overo (SW-1, SW-2, SW-3) | Autosomal dominant | Breed-specific mutations in *KIT* gene in ECA 3 | (Mau et al. 2004) |
| Tobiano | Autosomal dominant | A large chromosomal inversion in ECA 3 near the *KIT* gene of Tobiano horses. | (Brooks et al. 2007; Haase et al. 2008) |
| Dominant White (W1-W11) | Autosomal dominant | Carries the specific *KIT* mutation in ECA 3 | (Haase et al. 2007; Haase et al. 2009) |

**Table 3: List of conditions in horses where diagnostic tests are available to detect.**

| Condition name | Breeds | Mode of inheritance | Mutation | Reference |
|---|---|---|---|---|
| Hyperkalemic periodic paralysis (HYPP) | Quarter horse | Autosomal dominant | A point mutation in *SCN44* gene in ECA17 | (Rudolph et al. 1992) |
| Glycogen Branching Enzyme Deficiency (GBED) | Quarter horse | Autosomal recessive | A point mutation in exon 1 changes a tyrosine to a premature stop codon in the *GBE1* gene in ECA26. | (Ward et al. 2004) |
| Polysaccharide Storage Myopathy (PSSM) | Quarter horse, Belgians, Percherons, Morgans, Mustangs and some Warmblood breeds | Autosomal dominant | Point mutation that results in an arginine to histidine substitution in the *GYS1* gene in ECA10. | (McCue et al. 2008) |
| Malignant Hyperthermia (MH) | Quarter horse | Autosomal dominant | Point mutation that results in an arginine to glycine substitution in the *RYR1* gene in ECA10. | (McCue et al. 2009) |
| Hereditary Equine Regional Dermal Asthenia (HERDA or HC) | Quarter horse | Autosomal recessive | Point mutation that results in a glycine to arginine substitution in the equine cyclophilin B gene *PPIB* in ECA1. | (Tryon et al. 2007) |
| Overo Lethal White Foal Syndrome (OWLS) | American Paint horses | Autosomal recessive | Point mutation that results in an isoleucine/lysine substitution at codon 118 of the *EDNRB* gene located in ECA 17. | (Hultgren 1982; McCabe et al. 1990; Metallinos et al. 1998) |
| Junctional Epidermolysis bullosa (JEB) | Belgian Draft horses, Breton, Comtois, Vlaams Paard, Belgische, Koudbloed, Flander draft, American Saddlebreds. | Autosomal recessive | Drafts have a cytosine insertion (1368insC) creating a premature stop codon in the *LAMC2* gene in ECA 5, Saddlebreds has a 6589-bp deletion spanning exons 24-27 in the *LAMA3* gene in ECA 8. | (Spirito et al. 2002) |
| Cerebellar Abiotrophy (CA) | Arabian horses Miniature horses, Gotland Pony, Oldenburg. | | | (Cavalleri et al. 2013) |
| Appaloosa Coat Pattern / Leopard Print/ Congenital Stationary Night Blindness (CSNB) | Appaloosa | Autosomal dominant | 1378 bp insertion in intron 1 of *TRPM1* in ECA 1. | (Bellone et al. 2010; Bellone et al. 2013) |
| Myotonia | New Forest ponies | Autosomal recessive | A total of three polymorphisms were identified, two single nucleotide substitutions in *CLCN1* exon 15 (c.1593A>G and c.1775A>C) and a three base pair insertion (c.2652_2653insCTT) in exon 23 in ECA 4. | (Wijnberg et al. 2012) |

**Table 3 continued.**

| Condition name | Breeds | Mode of inheritance | Mutation | Reference |
|---|---|---|---|---|
| Lavender Foal Syndrome (LFS) | Arabian | Autosomal recessive | A single base deletion in exon 30 of *MYO5A* in ECA 1 that changes the reading frame and introduces a premature stop codon. | (Brooks et al. 2010) |
| Severe Combined Immunodeficiency (SCID) | Arabian | Autosomal recessive | A 5-basepair deletion in the gene encoding DNA-protein kinase catalytic subunit *DNA-PK* in ECA 9. | (Shin et al. 1997) |
| Foal immunodeficiency syndrome (FIS) | Fell and Dales ponies | Autosomal recessive | Mutation in the *SLC5A3* gene, in ECA 26, this causes a P446L substitution in the protein. | (Fox-Clipsham et al. 2011a; Fox-Clipsham et al. 2011b) |
| Equine insect bite hypersensitivity (IBH) | Icelandic horse Exmoor pony | | SNPs associated with *IGHE, FCER1A, IL4, IL4R, IL10, IL1RA*, and *JAK2* genes | (Andersson et al. 2012b) |

## Genomic Copy Number Variation (CNV)

### Background

Genomes of individuals of the same species differ from each other at many different levels – from single nucleotide and microsatellite polymorphism as discussed in previous paragraphs to large scale chromosomal variations which are studied by cytogenetics. There is also a type of structural variation which involves DNA segments of intermediate length, typically from 1 kilobase-pair (kb) to several megabase-pairs (Mb), although variants as small as 50 bp have been detected (Li et al. 2004; Feuk et al. 2006; Redon et al. 2006; Conrad et al. 2010b). These DNA segments can be inserted or deleted giving rise to copy number gains or losses or complex rearrangements involving both. Collectively, these types of variants are known as *copy number variants* (CNVs).

Copy number variants in vertebrate genomes were first acknowledged by Susumu Ohno over 40 years ago in the book "Evolution by Gene Duplication" (Ohno 1970). It was, however, only from 2004 onwards that the presence of CNVs across the genomes of mice (Li et al. 2004; She et al. 2008) and humans (Iafrate et al. 2004; Sebat et al. 2004; Tuzun et al. 2005; Feuk et al. 2006; Fiegler et al. 2006; Redon et al. 2006) was shown using contemporary molecular methods. Thereafter, CNVs have been studied in the genomes of primates (Cheng et al. 2005; Dumas et al. 2007; Perry et al. 2008), rat (Guryev et al. 2008) and more recently, in domestic species (Clop et al. 2012; Liu and Bickhart 2012).

Comparison of CNVs with SNPs showed that CNVs constitute 5 to 12% of the mammalian genome (Redon et al. 2006; Pielberg and Andersson 2007; Kim et al. 2008; Hastings et al. 2009) and contribute approximately twice the amount of total genetic differences between individuals when compared to SNPs (Sharp 2009).

## Genomic distribution and mechanisms of formation of CNVs

Studies in humans and animals show that CNVs are not randomly distributed in the genome (Hastings et al. 2009; Hall and Quinlan 2012). They are significantly enriched in regions with segmental duplications (SD), also known as low copy number repeats (Hastings et al. 2009; Stankiewicz and Lupski 2010; Bickhart et al. 2012; Du et al. 2012) transposable elements (Cordaux and Batzer 2009; Xing et al. 2009), and other regions with complex architectural features, such as subtelomeric regions (DeScipio et

al. 2008; Riethman 2008) and the Y chromosome (Skaletsky et al. 2003). Association of CNVs with structurally complex regions suggests that the formation of CNVs is caused by the same general mechanisms that can cause other structural changes in chromosomes – occurrence of chromosomal breaks followed by repair mechanisms using homologous and non-homologous recombination (Hastings et al. 2009). As follows, is a summary of possible mechanisms leading to CNV formation.

***Non-Allelic Homologous Recombination (NAHR)***

Typically occurs in regions enriched with low copy number repeats or segmental duplications. These are sequences larger than 1 kb that share over 95% sequence identity (Hurles and Lupski 2006; Hastings et al. 2009). Double stranded breaks in these regions can be repaired by different mechanisms involving both gene conversion and crossing over (Hurles and Lupski 2006) resulting in deletions, duplications, inversions or translocations. Examples of NAHR resulting in the formation of CNV are shown in Figure 1.

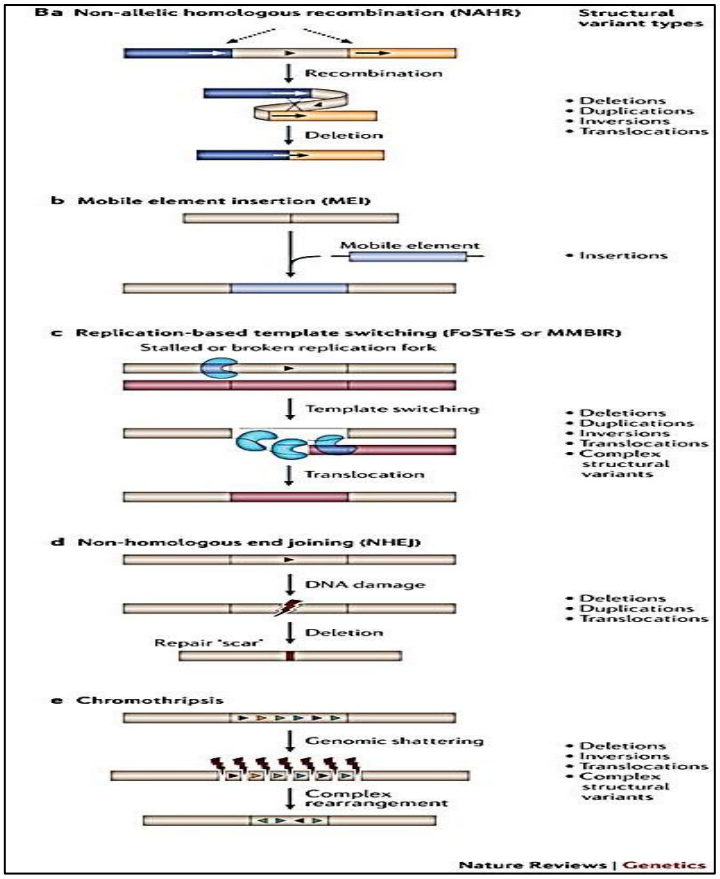


**Figure 1. Different mechanisms of CNV formation.**
(a) Recurrent structural variants often result from non-allelic homologous recombination (NAHR) which involves recombination between long highly similar low-copy-number repeats (blue and orange segments). (b) Novel genomic insertions can involve mobile element insertion of transposable elements by retrotransposition. (c) DNA-replication-associated template-switching events, involving the fork-stalling and template switching (FoSTeS) and microhomology-mediated break-induced replication (MMBIR) mechanisms. (d) Non-homologous end joining (NHEJ), which is a process that repairs DNA double-strand breaks in the absence of extensive sequence homology and is often accompanied by the addition or deletion of several nucleotides in the form of a 'repair-scar' (small red bar). (e) Chromothripsis — which is a phenomenon that seems to involve chromosome shattering leading to numerous breakpoints, followed by error-prone DNA repair (Weischenfeldt et al. 2013).

***Non-homologous recombination (NHR)***

It is a DNA break repair mechanism which uses very limited or no sequence homology and can lead to the formation of CNVs. NHRs can be divided into non-replicative and replicative mechanisms (Hastings et al. 2009). The former are not related to DNA replication and include phenomena such as non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), and breakage–fusion–bridge cycle (Hastings et al. 2009). For example, in NHEJ, double stand breaks are repaired without sequence homology and can result in deletion, insertion or duplication of a small number of nucleotides (1-4 bp), termed as replication scar. Sources of free DNA that can be inserted are mitochondrial DNA and retrotransposons (LINEs). Suggested mechanisms of CNV formation that are due to events during DNA replication are replication slippage, fork stalling and template switching (FoSTeS), and microhomology-mediated break-induced replication (MMBIR) (Hurles and Lupski 2006; Hastings et al. 2009). A recently discovered form of NHR is chromothripsis which is althought to involve both MMBIR and NHEJ, leading to the formation of multiple and massive structural rearrangements in a very short time which is typically observed in cancers (Stephens et al. 2011; Liu and Bickhart 2012). It has been observed that small CNVs usually originate from mobile element insertion and NHEJ, whereas large CNVs are caused by NAHR (Conrad et al. 2010a).

Taken together, even although some mechanisms of CNV formation are understood, the overall mechanistic aspects leading to CNVs are not completely clear and require further research. The study of CNVs is further complicated by possible

somatic mosaicism where CNVs are formed in some but not in other somatic cells or tissues (Hall and Quinlan 2012).

Methods for CNV discovery and analysis

Several tools and approaches have been used to examine CNVs. These include arrays of large genomic clones (Fiegler et al. 2006), SNP genotyping microarrays (Shaikh et al. 2009; Hou et al. 2012a; Yang et al. 2014), whole genome oligonucleotide tiling arrays (also known as comparative genomic hybridization or CGH arrays) (Carter 2007; Gresham et al. 2008; Mitra et al. 2009; Conrad et al. 2010b; Ueno et al. 2012; Mizuno et al. 2014), and whole genome sequencing (Kidd et al. 2008; Alkan et al. 2009; Ni et al. 2013; Poultney et al. 2013; Iben and Maraia 2014; Tan et al. 2014). Although sequencing certainly provides the most comprehensive information about CNVs, array based CGH is currently the most widely used, efficient and relatively cost-effective way of detecting and measuring structural variations in the genome of humans (Carter 2007; Gresham et al. 2008; Mitra et al. 2009; Conrad et al. 2010b; McCarroll 2010) and animals (Liu et al. 2009; Fadista et al. 2010; Wang et al. 2010). The only prerequisite for designing a CGH array is the availability of whole genome sequence information for the species of interest. Once the arrays have been designed and produced, they are analyzed by array CGH. The method is based on co-hybridization of two differently fluorescently (usually with Cy3 and Cy5) labeled genomic DNA samples, of which one is always a reference and another, the sample of interest (case). The case/reference signal ratios are

measured and copy number variations are detected according to the ratio. Thereafter, the data can be analyzed using a variety of software packages, often provided by the companies producing the arrays: e.g., Agilent Technologies – Agilent Genomics Workbench, NimbleGen – NimbleGen SignalMap software and Affymetrix-Chromosome Analysis Suite (ChAS) Software.

In recent years, next generation sequencing (NGS) technology has been used widely to detect CNVs (Medvedev et al. 2009; Mills et al. 2011). CNVs identified by NGS are about 50 bp in size (Alkan et al. 2009). A new approach whole exome sequencing (WES), is also considered because in compare to NGS it is cost effective to study large population (Ng et al. 2009). WES focuses on protein-coding regions or exomes or customer defined target regions, which only encompass about 1% of the entire genome.

CNV research in humans and primates

Studies of CNVs in humans started a decade ago (Iafrate et al. 2004), first a small number of unrelated individuals (Iafrate et al. 2004; Feuk et al. 2006; Fiegler et al. 2006; Freeman et al. 2006; Redon et al. 2006), but soon expanding to larger cohorts representing specific human populations: European (Vogler et al. 2010), African (Matsuzaki et al. 2009), Japanese (Takahashi et al. 2008), Chinese (Lin et al. 2009), Korean (Moon et al. 2011). Additional information about human CNVs was provided by

the Human HapMap consortium and the 1000 Genome Sequence Project. The data are available from Ensembl and Database of Genomic Variants (DGV) (Table 4).

The DGV shows that approximately half of the 109,863 human CNVs (July 2013) fall into the size range from 1 to 10 kb; about 21% overlap with genomic regions listed in OMIM (Online Mendelian Inheritance in Human), and according to DECIPHER database (Firth et al. 2009), CNVs relate to 70 known human syndromes.

**Table 4: An overall summary of the DGV database.**
(July 2013 update, mapped to GRCh37 assembly) (MacDonald et al. 2014)

| Database content | Number of entries |
|---|---|
| Studies | 55 |
| Unique samples | 14316 |
| Variant regions | 202431 |
| Deletion | 77268 |
| Duplication | 668 |
| Loss | 64185 |
| Gain | 24891 |
| Gain+loss | 3850 |
| Insertion | 24140 |
| Inversion | 1149 |
| Complex | 4090 |
| Unknown | 2189 |
| Variant calls | 2393718 |
| CNV | 2391408 |
| Inversion | 2310 |
| Filtered variants | 3900253 |

CNVs formed in both genic and intergenic region. It was expected to find more CNVs in intergenic region as they contain large repeats, but most CNVs were present in gene containing region or inside known coding region (Iafrate et al. 2004; Perry et al. 2006). CNVs present in intergenic region mostly containing non-coding RNAs which are responsible for regulation of normal gene expression as well as cause various diseases and disorders (Khalil et al. 2009; Huarte et al. 2010; Walker and Scherer 2013).

Human CNV studies were performed to study various aspects of genomic variations and their effect on normal human population. Gene containing CNVs cause changes in gene dosage (Jaradat et al. 2013; Polan et al. 2014; Vittori et al. 2014), gene expression regulation (Haraksingh and Snyder 2013), and changes in recessive allele exposure (Boone et al. 2013) . Duplication and deletion of cis-regulatory elements have been shown to greatly influence phenotype, particularly developmental genes (Spielmann and Klopocki 2013). CNVs have great impact on genomic selection (Conrad et al. 2006; Redon et al. 2006; Itsara et al. 2010; Gokcumen et al. 2011; Stenberg and Larsson 2011), gene duplication and evolution (Popesco et al. 2006; Dumas et al. 2007; Nozawa et al. 2007; Hasin et al. 2008; Young et al. 2008; Waszak et al. 2010; Jarick et al. 2011; Shadravan 2013; Veerappa et al. 2013), and changes in innate and adaptive immunity (Kulski et al. 2002; She et al. 2008; Orange et al. 2011; Bergen et al. 2012; Lee et al. 2012; Llaurens et al. 2012). From these outcomes a common CNV map was generated where most of the CNVs are associated with cell adhesion, sensory perception, neurophysiological processes, kinase- and phosphorylation-related categories, and cell signaling.

CNV studies were performed with primates to look for genome-wide structural microheterogeneity, breakpoints for chromosomal rearrangements between their genomes and humans. CNVs in primate genome might be indicative of the gaining of new functions and therefore of adaptive evolution. Comparative studies between human, chimpanzee (Kehrer-Sawatzki and Cooper 2007), gorilla (Wilson et al. 2006), orangutan genomes (Yohn et al. 2005) might help to determine the ancestral state of primate genome, thereby indicating the divergences between them.

CNV research in mice

The close association of CNVs with complex diseases and disorders (Weischenfeldt et al. 2013) identified several new structural mutations contributed to develop human disorders, and evolutionary drive of genes and species. It is important to generate accurate CNV maps not only in humans but in many different organisms and cellular contexts, so that the biological significances and molecular mechanism of CNV formation can be evaluated (Hall and Quinlan 2012). In this context CNV studies were performed in model mammalian animals; mouse, rat (Guryev et al. 2008; Charchar et al. 2010) and dog (described later). CNV work in mouse was started in 2004 (Li et al. 2004). CNV study in mouse was initiated by array CGH method (Li et al. 2004; Graubert et al. 2007; She et al. 2008) and later moved to NGS-based study (Quinlan et al. 2010; Yalcin et al. 2011). In mouse studies the genomic variation is more than human (7196 CNVs in single study) (Quinlan et al. 2010) because of the presence of high level

of transposable elements (Akagi et al. 2008) in mouse genome. Mouse and other mammalian studies showed that overall levels of structural variations are roughly similar. This information will help to analyze all species on a common CNV discovery platform in the future (Hall and Quinlan 2012).

Phenotypes, genetic diseases, and disorders associated with CNVs

The association of DNA copy number variation with phenotypic traits, adaptations and evolution was proposed over 40 years ago by Susumu Ohno in '*Evolution by Gene Duplication*' (Ohno 1970; Conant and Wolfe 2008) and has found compelling evidence from recent CNV studies. For example, comparison of humans and other primates has revealed several human lineage specific gene amplifications: copy number increase of the salivary amylase gene (*AMY1*) in some human populations gives an adaptive advantage in digesting starch-rich food (Perry et al. 2007); copy number expansion of aquaporin7 (*AQP7*) is thought to increase glycerol transport and endurance, or, CNVs in genes expressed specifically in neurons (*DUF1220/NBPF*) are thought to be related to cognition (Popesco et al. 2006; Lupski 2007b). The best known CNV related phenotypes in animals are coat colors, such as grey in horses (Rosengren Pielberg et al. 2008), dominant white in pigs (Giuffra et al. 2002), agouti locus in goats (Fontanesi et al. 2009) or color sidedness in cattle (Durkin et al. 2012).

### *Microscopic CNVs*

Association of CNVs with diseases and disorders has been known for years since early cytogenetic studies. It is well known that chromosomal aneuploidies, as well as deletions and duplications which are large enough to be observed under light microscope are associated with diseases and disorders. The best known examples for aneuploidy in humans are the Down syndrome with trisomy of chromosome 21 and the Turner syndrome with X chromosome monosomy (Beckmann et al. 2008). Examples for large microscopic rearrangements are Smith-Magenis syndrome/ Potocki-Lupski syndrome associated with deletion/ duplication in Chr17p11.2 (Ricard et al. 2010); deletions in a complex imprinted region in Chr15q11-13 which causes Angelman syndrome if the deletion is maternal or Prader-Willi syndrome if deletion occurs on the paternal chromosome (Sanders et al. 2011; Rangasamy et al. 2013; Weischenfeldt et al. 2013); Thromcytopenia-absent radius (TAR) syndrome is caused by a deletion in Chr1q21.1, or *Cri-du-Chat* and Williams syndromes with large deletions in Chr5p and 7q, respectively (Sanders et al. 2011). Also, it is well established that multiple deletions in the AZF regions in the human Y chromosome are responsible for a spectrum of male fertility disorders (Lahn and Page 1997; Skaletsky et al. 2003; Rozen et al. 2012).

### *Submicroscopic CNVs*

During the past decade tools have become available for the discovery and study of submicroscopic CNVs. This is directly attributed to the availability of whole genome sequence assemblies which have allowed the construction of tiling and SNP arrays for

CNV discovery, whereas the resolution and sensitivity of these tools is gradually increasing, so that CNVs as small as 50 bp can be discovered (Conrad et al. 2010b). Even more developments in the field are accompanying the application of NGS.

One of the central goals of CNV research is to determine their association with genome instability, genetic diseases, and congenital disorders. CNVs are considered as major sources of inter-individual genetic variation that could explain variable penetrance of Mendelian and polygenic diseases and variation in the phenotypic expression of complex traits (Beckmann et al. 2007; Beckmann et al. 2008). Human CNVs have been associated with a few rare conditions, such as CHARGE syndrome (Monfort et al. 2008) and many common diseases (Conrad and Antonarakis 2007; Hollox et al. 2008; Zhang et al. 2009; Choy et al. 2010; Fanciulli et al. 2010; Lee and Scherer 2010). Among the latter, most are complex and polygenic disorders affecting a broad range of processes of biomedical importance (Ermakova et al. 2011). Gene ontology analyses show that CNV regions are significantly enriched for genes involved in immune response and inflammation, cognition and perception, but also genes involved in reproduction, drug metabolism and other complex traits (Redon et al. 2006; Cooper et al. 2007). As follows is a summary of human biological systems and corresponding clinical conditions that have been shown to be influenced by or associated with CNVs.

### *Immune and autoimmune disorders*

Because of the outstanding bias of CNVs in immunity-related genes, contribution of CNVs has been shown for several human infectious diseases (Choy et al. 2010),

autoimmune disorders, and allergies. For example, higher genomic copy number for beta-defensin genes has been associated with risk of psoriasis (Hollox et al. 2008); copy number of alpha-defensin genes is a strong risk factor for Crohn's disease, HIV progression and multiple sclerosis (McCarroll et al. 2008a; Choy et al. 2010; Khan et al. 2013) and deletion variants of *RABGAP1L*, 10q21.3, and cytochrome 4 (*C4)* are associated with the risk of systemic lupus erythematosus (Kim et al. 2013). Furthermore, 21% of 270 candidate genes for asthma are residing near or within a CNV (Rogers et al. 2013)

***Pharmacogenomics***

A relatively new field of CNV research is pharmacogenomics with a focus on genes that are responsible for the metabolism of drugs and pharmaceuticals or are involved in the genetics of pain and analgesia (He et al. 2011; Madian et al. 2012; Mogil 2012). The aim to discover human genes and their variants that affect response to drugs with an ultimate goal to devise novel personalized pharmacological treatment strategies (Madian et al. 2012). Genes of interest include glutathione S-transferases (*GST*s) which encode enzymes that catalyze the detoxification of carcinogens, therapeutic chemicals and environmental toxins (Hayes et al. 2005); cytochrome P450 2D6 (*CYP2D6*) which is predominantly expressed in human liver and metabolizes over 25% of drugs currently used in the clinic (Zhou 2009), and opioid receptors (Hayashida et al. 2008), to mention some.

### *Neuropsychiatric and behavioral disorders*

Among the most multifactorial traits affected by CNVs is human behavior. A number of CNVs have been associated with psychiatric and neurological disorders, mental retardation, autism and obsessive-compulsive disorders (Zhang et al. 2009), and are althought to be part of the development of cognition (Lupski 2007a). About 9.5% of patients with schizophrenia, bipolar or cognitive disorders have chromosomal abnormalities (van Karnebeek et al. 2005) including X chromosome aneuploidy and rearrangement in chromosome 9 (Ponnudurai et al. 2012), micro deletion in chromosome 22 (Vassos et al. 2010) or rearrangements in chromosomes 1, 15 and 16 (Stefansson et al. 2008). There are multiple studies about the role of CNVs in various autism spectrum disorders (Sanders et al. 2011; Krumm et al. 2013; Poultney et al. 2013). For example, of 300 kb exonic deletion in neurexin 1 gene (*NRXN1)* in chromosome 2 is associated with autism in some families (Grayton et al. 2012; Walker and Scherer 2013; Tansey et al. 2014), while duplications in chromosomes 2 (Barbosa-Goncalves et al. 2008; Lo-Castro et al. 2009; Shim et al. 2014), 17 (Curry et al. 2013) and 22 (Marchani et al. 2012), in others. Further, CNVs have been associated with hyperactivity and attention deficit disorders (Elia et al. 2012), and vaiants in chromosomes 1, 15 and 16 with epilepsy (Mefford et al. 2010; Vassos et al. 2010).

### *Metabolic and cardiovascular disorders*

CNV research is rapidly expanding in the study of complex metabolic and cardiovascular disorders (Fanciulli et al. 2010; Lee and Scherer 2010). Copy number

changes in *GATA4* and *NODAL* have been identified as candidate loci for congenital heart disease (Warburton et al. 2014). Like in neurodevelopmental disorders, multiple regions in the human genome are associated with obesity: a 45 kb deletion near neuronal growth regulator gene (*NEGR1*) (Speliotes et al. 2010); a 25 kb deletion upstream of *GPRC5B* gene; a CNV in chromosomes 10, 11 and 16 (Sha et al. 2009; Willer et al. 2009; Walters et al. 2010; Jarick et al. 2011; D'Angelo and Koiffmann 2012). CNVs involved in developmental delay or intellectual disabilities are also associated with obesity. The examples are Prader-Willi syndrome and Smith-Magenis syndrome (D'Angelo and Koiffmann 2012). CNVs in alpha-defensin genes are strong risk factors both for obesity and type I diabetes (Choy et al. 2010; Khan et al. 2013). Finally, transgenic mouse model with an extra copy of obesity candidate gene, G protein β3 (*GNB3*), implicates *GNB3* duplication in a childhood obesity syndrome and obesity-related syndromes, such as type 2 diabetes, cardiovascular diseases, hypertension, and cancer (Goldlust et al. 2013).

### *CNVs involved in cancer*

Cancer genomes are characterized by multiple forms of genome instability, including genetic and epigenetic alterations which cumulatively trigger oncogenic processes (Lupski 2013). The diverse genetic changes include also CNVs and have been characterized in many different forms of cancer: breast cancer (Stephens et al. 2009), chronic lymphocytic leukemia (Puente et al. 2011), colorectal cancer (Bass et al. 2011; Stephens et al. 2011), lung cancer (Campbell et al. 2008; Stephens et al. 2011),

melanoma (Pleasance et al. 2010), pancreatic cancer (Campbell et al. 2010), prostate cancer (Berger et al. 2011), renal cancer (Stephens et al. 2011) and thyroid cancer (Stephens et al. 2011), to list a few.

***Disorders of sexual development and reproduction***

A number of studies associate CNVs with human disorders of sexual development (DSD) and reproduction. This is a diverse group of complex disorders which include urogenital abnormalities, sex reversal, hypospadias, cryptorchidism (Smyk et al. 2007; Ledig et al. 2010b; Tannour-Louet et al. 2010; Li et al. 2011; Tuttelmann et al. 2011; White et al. 2011), premature ovarian failure (Aboura et al. 2009; Quilter et al. 2010), and male subfertility (Carrell and Aston 2011). Several of these conditions are regulated by dosage-sensitive genes, thus being particularly responsive to the effect of CNVs (Conrad and Antonarakis 2007; Sharp 2009; Tian et al. 2009; Quilter et al. 2010). Gonadal dysgenesis has been associated with gains or losses in regions with known sex determining genes, such as *DMRT1, DAX1* and *SOX9* (Tannour-Louet et al. 2010), with a 35 kb deletion that completely removes *NEIL2* gene in human chromosome 8p (Ledig et al. 2006; Ledig et al. 2010b), and with deletions in the short arm of human chromosome 9 affecting *KANK1, DOCK8* and *DMRT1* genes (Tuttelmann et al. 2011).

Besides, the studies have identified novel CNV genes, such as *FGFR2* and *ZEB2 that* are potentially associated with some DSD phenotypes (Ledig et al. 2010b). DSD-associated CNVs have been identified also in non-coding regions of the genome. It has

been postulated that these regions harbor unidentified genes and non-coding RNAs that might be involved in sexual development and gonadal functions (White et al. 2011). While Y chromosome deletions are long known causes of spermatogenic failure in men (Lahn and Page 1997; Rozen et al. 2012), copy number variants in autosomal regions have also been associated with sperm abnormalities (Tuttelmann et al. 2011). Also, X chromosome aberrations are more frequent in azoospermic males compared to males with normal sperm count. Deletions and duplications in chromosomes 3, 6 and 12 and have been found in patients with severe oligozoospermia and Sertoli-cell-only-syndrome. These regions contain *EPHA3*, *ANKS1A* and *ASNKS1B* –genes that are expressed in Sertoli cells or germ cells (Tannour-Louet et al. 2010).

In summary, the list of human clinical conditions that are associated or althought to be associated with CNVs is long and the corresponding literature overwhelming. Although the exact functions of CNVs in most complex diseases and disorders remain elusive (Weischenfeldt et al. 2013), studies in humans have encouraged and paved way for initiating similar research in domestic species.

## CNV research in domestic animals

Microscopically visible large scale genome variations in domestic animals have been known for years through extensive cytogenetic analysis (Chowdhary and Raudsepp 2000; Ducos et al. 2008; Raudsepp and Chowdhary 2010). Many of these variants caught attention because of relationship with infertility, abnormal sexual development or

congenital defects (Fechheimer 1971; Ducos et al. 2008). However, with the development of approaches to identify submicroscopic variations between genomes, it became clear that only a fraction of CNVs might be associated with diseases and congenital abnormalities, while the majority of CNVs are neutral or contribute to normal phenotypic variation between individuals, populations, and breeds (Liu et al. 2010).

Studies of sub-microscopic structural variations in animal genomes started when genome draft sequences (Ensembl, http://uswest.ensembl.org/index.html) became available. This allowed the design of suitable probes for whole genome tiling arrays and SNP beadchips, both of which can be used for the detection of CNVs. It was equally important to have a reference genome as a common platform to compare CNV findings among individuals and populations. To date, CNV research is ongoing or initiated in all main domestic species: cattle, pig, horse, sheep, goat, dog, and chicken; although, no CNV studies have yet been conducted in alpacas, camels, and cats. The current status of CNV research in domestic species is summarized in Table 5 and described in more detail in the following sections.

## Table 5: A summary of CNV studies in domestic species.

| Species; # studies | Genome coverage | Platform | Method | Reference |
|---|---|---|---|---|
| **Horse 5** | Exons and UTRs | 2X400K Tiling array, Agilent | aCGH | (Doan et al. 2012a) |
| | WG | Illumina | NGS | (Doan et al. 2012b) |
| | WG | SNP50 Beadchip, Illumina | Genotyping | (Dupuis et al. 2013) |
| | WG | SNP50 Beadchip, Illumina | Genotyping | (Metzger et al. 2013a) |
| | WG | 1.3M Tiling array, NimbleGen | aCGH | (Wang et al. 2014b) |
| **Cattle 14** | WG | 385K tiling Array, Nimblegen | aCGH | (Liu et al. 2008) |
| | WG | SNP50 BeadChip, Illumina | Genotyping | (Matukumalli et al. 2009) |
| | WG | SNP50 BeadChip, Illumina | Genotyping | (Bae et al. 2010) |
| | WG | 6.3M tiling array, NimbleGen | aCGH | (Fadista et al. 2010) |
| | WG | 385K tiling Array, Nimblegen | aCGH | (Liu et al. 2010) |
| | WG | SNP50 BeadChip, Illumina | Genotyping | (Hou et al. 2011) |
| | WG | Paired end sequencing, Illumina; High-Density SNP BeadChip and SNP50 BeadChip, Illumina; 6.3M tiling array, NimbleGen | NGS, Genotyping, aCGH | (Zhan et al. 2011) |
| | WG | SOLiD 3 (ABI); Bovine HD SNP Beadchip, Illumina | NGS, Genotyping | (Stothard et al. 2011) |
| | WG | GAIIx, Illumina | NGS | (Bickhart et al. 2012) |
| | WG | SNP50 BeadChip, Illumina | Genotyping | (Jiang et al. 2012) |
| | WG | 385K tiling Array, Nimblegen | aCGH | (Hou et al. 2012a) |
| | WG | Applied Biosystems SOLiD System | NGS | (Choi et al. 2013) |
| | WG | High-Density SNP BeadChip, Illumina | Genotyping | (Jiang et al. 2013) |
| | WG | SNP50 BeadChip, Illumina | Genotyping | (Cicconardi et al. 2013) |
| | WG | Illumina, Hiseq 2000 | NGS | (Shin et al. 2014) |
| **Pig 11** | WG | 385K tiling Array, Nimblegen | aCGH | (Fadista et al. 2008) |
| | WG | SNP60 BeadChip, Illumina | Genotyping | (Ramayo-Caldas et al. 2010) |
| | WG | SNP60 BeadChip, Illumina | Genotyping | (Chen et al. 2012) |
| | WG | 3 × 720K tiling array, Nimblegen | aCGH | (Li et al. 2012) |
| | WG | SNP60 BeadChip, Illumina | Genotyping | (Wang et al. 2012) |
| | WG | Infinium II Multisample assay, Illumina | Genotyping | (Wang et al. 2013a) |
| | WG | SNP60 BeadChip, Illumina | Genotyping | (Wang et al. 2013b) |
| | WG | SNP60 BeadChip, Illumina | Genotyping | (Fowler et al. 2013) |
| | WG | HiSeq, Illumina | NGS | (Paudel et al. 2013) |
| | WG | 2.1M tiling array, NimbleGen | aCGH | (Wang et al. 2014a) |
| | WG | SNP60 BeadChip, Illumina | Genotyping | (Fernandez et al. 2014) |
| **Dog 8** | WG | 385K tiling array, Nimblegen | aCGH | (Chen et al. 2009) |
| | WG | 2.1M tiling array, NimbleGen | aCGH | (Nicholas et al. 2009) |
| | WG | HD BeadChip assay, Illumina | Genotyping | (Alvarez and Akey 2012) |
| | WG | 2.1M array, NimbleGen | aCGH | (Berglund et al. 2012) |
| | WG | v2 SNP Chip, Affymetrix | Genotyping | (Karyadi et al. 2013) |
| | WG | 720K tiling array, NimbleGen | Genotyping | (Jung et al. 2013) |
| | WG | HD BeadChip, Illumina | Genotyping | (Gurgul et al. 2014) |
| | WG | HD Genotyping array | Genotyping | (Molin et al. 2014) |
| **Sheep 2** | WG | *****Bovine 385K tiling Array, Nimblegen | aCGH | (Fontanesi et al. 2010a) |
| | WG | SNP50 BeadChip array, Illumina | Genotyping | (Liu et al. 2013) |
| **Goat 1** | WG | *****Bovine 385K tiling Array, Nimblegen | aCGH | (Fontanesi et al. 2010b) |
| **Chicken 8** | WG | 385K tiling Array, Nimblegen | aCGH | (Griffin et al. 2008) |
| | WG | 385K tiling Array, Nimblegen | aCGH | (Skinner et al. 2009) |
| | WG | 385K tiling Array, Nimblegen | aCGH | (Volker et al. 2010) |

**Table 5 continued.**

| Species; # studies | Genome coverage | Platform | Method | Reference |
|---|---|---|---|---|
| **Chicken** | WG | 385K tiling Array, Nimblegen | aCGH | (Wang et al. 2010) |
| | WG | 60K SNP BeadChip, Illumina | Genotyping | (Jia et al. 2012) |
| | WG | 4×180K tiling array, Agilent | aCGH | (Abe et al. 2013) |
| | WG | 244K tiling array, Agilent | aCGH | (Crooijmans et al. 2013) |
| | WG | 385K tiling Array, Nimblegen | aCGH | (Skinner et al. 2014) |

***** The array platforms used for CNV discovery in different species were species-specific, with the exception of sheep and goat where bovine arrays were used.

### *CNV research in the canine genome*

Dog is an important model species for biomedical research. This is probably why the first WG CNV studies in animals were carried out in dogs using a large insert clone (BAC) array to detect CNVs associated with canine cancers (Thomas et al. 2008). Thereafter, two studies used NimbleGen high-resolution ~380K custom tiling arrays to demonstrate that segmental duplications encompass about 4.21% of the canine genome and that many are breed or breed class specific (Chen et al. 2009; Nicholas et al. 2009). Interestingly, genomic imbalances in canine transmissible venereal tumors involve CNVs in centromeric and telomeric sequences (Chen et al. 2009), thus underscoring that genomic imbalances leading to disease are not necessarily located in genes. Prevalence of breed-specific CNVs in dogs has also been shown by two recent studies, one using a 2.1 M NimbleGen canFam2 WG CGH array (Berglund et al. 2012), and another, a CanineHD 170 K SNP array (Molin et al. 2014). The latter analyzed 359 dogs from 30 different breeds and with very stringent criteria identified 72 CNVRs of which 60% overlapped with previously reported CNVs in dogs. Notably, 15 CNVRs were defined as strictly breed specific among 12 different breeds. Altogether, over 400 CNV regions

have been found in the dog genome (Berglund et al. 2012; Molin et al. 2014) and like in humans, many canine CNVs are related to phenotypic traits and diseases (Chen et al. 2009; Molin et al. 2014).

### *CNV research in the cattle genome*

Among domestic mammals, the cattle genome is probably the most extensively studied for CNVs. Since 2009, at least 14 studies report about WG CNV analysis in the cattle genome using array CGH, SNP beadchip genotyping, or NGS platforms (Table 5).

The first CNV study in cattle was published in 2008, involved 3 animals and identified 25 CNVs (Liu et al. 2008). A year later, using bovine SNP50 Beadchip, 79 deletion variants were detected in 556 animals of various breeds (Matukumalli et al. 2009). However, the first systematic genome-wide CNV analysis using 385K Nimblegen array was reported by Liu and colleagues in 2010. The study included 90 animals representing taurine, indicine, and mixed breeds. Over 200 CNVs, covering 1.07% of cattle genome, were identified, whereas there were more CNVs in indicine breeds than in taurine breeds. These CNVs involved 400 annotated cattle genes and were significantly enriched with genes with functions in immunity, lactation, reproduction, and rumination. A strong association was observed between CNVs and evolutionary breakpoint regions (EBRs), whereas ~20% of cattle specific and ~52% of artiodactyl-specific EBRs were significantly enriched with CNVs (Liu et al. 2010). This was a pioneering CNV study in cattle and domestic animals and established a foundation for further research. In the same year, Fadista and colleagues (2010) identified 304 CNVs in

20 animals of 4 breeds. Majority of CNV genes were involved in environmental response or were associated with human orthologs for known diseases and disorders. Importantly, over 50% of cattle CNVs have been associated with segmental duplications which are well known hotspots for CNV formation (Fadista et al. 2010; Liu et al. 2010).

The most comprehensive information about CNVs in the cattle genome has been obtained by NGS. Bickhart and colleagues (2012) used WG sequence data for 5 taurine and one indicine cattle and identified 1265 CNVs of which 476 were novel. Interestingly, Nellore cattle showed copy number gain in genes responsible for pathogen and parasite resistance, while genes involved in lipid transport and metabolism were duplicated in taurine breeds of beef cattle (Bickhart et al. 2012). Thereafter, many more CNV studies have been conducted in different cattle breeds and populations from different parts of the World (Hou et al. 2012a; Jiang et al. 2012; Choi et al. 2013; Cicconardi et al. 2013; Jiang et al. 2013). Despite differences in methodology and animal cohorts, all studies agree that the bovine genome has hundreds of CNVs, many of which span known protein coding genes with functions in immune response and defense, sensory perception, reproduction, rumination and lactation. Interestingly, the absence of CNVs in housekeeping genes is thought to be an indication of strong refining selection (Liu and Bickhart 2012).

***CNV research in sheep and goat genomes***

CNV research in sheep and goat were initiated by cross-species hybridization to bovine WG 385K array (Fontanesi et al. 2010a; Fontanesi et al. 2010b). Analysis of 9

goats of 4 breeds identified 127 CNVs, while in 11 sheep of 6 breeds 135 CNVs were detected. Importantly, goat and sheep CNVs significantly overlapped with those previously known for cattle, suggesting that several chromosome regions might contain recurrent interspecies CNVs. Like in other mammals, goat and sheep CNVs affected genes with functions in immunity and disease resistance, environmental response, but also genes involved in lipid metabolism and behavior (Fontanesi et al. 2010a; Fontanesi et al. 2010b). More recently, an ovine SNP50 BeadChip was constructed and used for CNV analysis in 329 animals of 3 different sheep breeds (Liu et al. 2013). A total of 238 CNVs were identified covering 2.27% of the sheep genome. Gene Ontology analysis showed significant relationship of CNVs with olfactory receptors, G-protein coupled receptor protein signaling pathway, signal transduction, and plasma membrane components (Liu et al. 2013).

***CNV research in the porcine genome***

The first snapshot of copy number variation in the pig genome was obtained by CGH analysis of 12 Duroc boars and one unrelated Hampshire boar on a custom made oligoarray for pig chromosomes 4, 7, 14 and 17 (Fadista et al. 2008). The authors identified 37 CNVRs of which some overlapped with known segmental duplications. This initial study has by now been followed by 10 WG CNV analyses in a variety of pig breeds and populations using CGH platforms, SNP Beadchip and NGS (Table 5). Studies relying on array based platforms, such as SNP60 BeadChip (Chen et al. 2012) or 720K NimbleGen WG tiling array (Li et al. 2012), have identified hundreds of CNVs in

the porcine genome. In contrast, the recent NGS analysis of 16 animals from Europe and Asia, including wild boars, revealed over 3000 CNVs, 545 of which involved genes (Paudel et al. 2013). However, the large number of intergenic CNVs in the pig genome implies that the majority of variants are likely neutral. Observation that there are more CNVs in large porcine populations suggests that like SNPs and microsatellites, the CNVs reflect demographic history rather than phenotypic diversity (Paudel et al. 2013). Nevertheless, GO analysis of CNV genes indicates that, like in other species studied so far, copy number variable genes are related to sensory perception, various adaptations, and behavioral changes during domestication and, specific for pigs, to omnivorous lifestyle. A small number of CNVs are uniquely present in domestic pigs proposing their selection during domestication (Paudel et al. 2013).

### *CNV research in avian genomes*

CNVs have been studied in several economically important poultry species, such as chicken (Table 5), turkey (Griffin et al. 2008), and Pekin duck (Skinner et al. 2009), and model species, like zebra finch (Volker et al. 2010). Very recently, CNV analysis was conducted in key avian clades including Galliformes, Anseriformes, Passeriformes, Gruiformes and Falconiformes (Skinner et al. 2014). The methodology involved a chicken WG 385 K tiling array (NimbleGen) and cross-species array CGH with the genomic DNA from 16 different species. The results, for the first time provide a global overview of the patterns of CNVs in birds. While avian genomes are approximately one third the size of a typical mammalian genome, the size and abundance of CNVs is

similar to that of mammals. Although, it is possible that in avian genomes there is slightly higher association between genes and CNVs than it is in mammals. Also, as the CNVs have an equal range of sizes across chromosomes, the proportion of a microchromosome affected by CNVs is greater than the comparable proportion of a macrochromosome. This explains why bird genomes are enriched with genic CNVs because about two thirds of avian genes are located in microchromosomes. Gene ontology analysis shows that, like in mammals, avian CNVs are predominantly associated with immune response and antigen presentation genes. Probably the most interesting finding was that 62% of all CNVRs were unique to individual species. For example, CNVs were associated with muscle activity and speed in falcons, migratory behavior in quails, brain development, and neuronal functions in turkey, and immune functions in the common quail and silver pheasant. Overall, this is the first broad survey of CNVs in avian species and establishes an important foundation for future research.

***Summary of CNV research in domestic species***

Taken together, CNV research in domestic species is rapidly gaining momentum and the attention of researchers worldwide. Recent methodological advances, such as high resolution array platforms and NGS, have essentially refined the existing data and facilitated the discovery of new CNVs. Despite this, the current research is restricted to the description of the genomic landscape and functional categories of CNVs in different species. At the same time, the discovery of variants that are associated with complex traits and disorders of biological and economic interest is, as yet, at a very initial stage.

**Present status of CNV research in the horse**

Initial knowledge about the involvement of CNVs in equine biology dates back to 1970s when cytogenetic studies revealed that several developmental and reproductive disorders in horses are caused by chromosomal aneuploidies – gains or losses of whole chromosomes (Chowdhary and Raudsepp 2000). Here the best known are various sex chromosome aneuploidies (63,XO; 65,XXX; 65,XXY) and autosomal trisomies. Thereafter, gene mapping revealed traits and conditions that are caused by smaller CNVs in genes or genomic regions. For example, a 4.6-kb duplication in intron 6 of *STX17* (syntaxin-17) in ECA25 causes gray coat color (Swinburne et al. 2002; Rosengren Pielberg et al. 2008; Sundstrom et al. 2012); deletions in the male specific region of the Y (MSY) chromosome that involve the *SRY* gene, cause some forms of Y-linked male-to-female sex reversal syndrome (Raudsepp et al. 2010); or duplication of the Y chromosome via the formation of a Y isochromosome causes abnormal sexual development (Das et al. 2012). However, genome-wide CNV research in horses started only after the horse whole genome draft assembly became available (Wade et al. 2009). The genome sequence showed that the horse genome is relatively poor for segmental duplications – genomic regions that have been frequently associated with CNVs in other species (Hastings et al. 2009; Stankiewicz and Lupski 2010; Bickhart et al. 2012; Du et al. 2012). Segmental duplications make up only about 0.5% of the equine genome and are most abundant in ECA25 (Wade et al. 2009).

The first systematic CNV research in horses was conducted by array CGH using a custom-made tiling array which specifically targeted gene exons (Doan et al. 2012a). The study involved 16 horses of 15 different breeds and a donkey as an out group. Altogether, 775 CNVRs involving 1707 protein- and RNA-coding genes were identified, whereas 96.4% of genes in CNVs were protein coding. Functional categories of CNV genes included sensory perception, signal transduction, metabolism, regulating blood group antigens, fecundity, coat color, keratin formation, neuronal homeostasis, lactation, and height.

This work was followed by the discovery of 282 CNVRs in the genome of a Quarter Horse mare by NGS (Doan et al. 2012b). Among the 282 CNVRs, 192 involved genes and 90 were located in intergenic regions. Similarly to CNV studies in other mammals (see above) and the previous study in horses (Doan et al. 2012a), most of the CNVs were enriched with genes involved in sensory perception, signal transduction, and immunity and defense pathways.

Besides the contribution to the CNV research in horses, this was the second published horse WG sequence and the first genomic sequence of a Quarter Horse. The identified SNPs, INDELs and CNVs are a resource for future studies of genetic variation in horses (Doan et al. 2012b).

The primary goal of the third CNV study in horses was to identify variants associated with recurrent laryngeal neuropathy (RLN), an important equine upper airway disease compromising performance (Dupuis et al. 2013). This genome-wide CNV study used genotyping on the Illumina Equine SNP50 BeadChip and involved 477 horses of

diverse breeds. The authors identified 2797 CNVs corresponding to 478 CNVRs. Of these, 67 common CNVRs were tested for association with RLN but no significant associations were found. Nevertheless, duplication in ECA10 was detected in 10 cases of 3 breeds, but in none of the controls. Functional significance of this finding regarding RLN, however, remained unclear and requires further studies.

The primary focus of the study by Metzger and colleagues (2013) was to compare the efficiency of different bioinformatics approaches and algorithms (CNVPartition, PennCNV and QuantiSNP) for CNV detection. The authors used the Illumina Equine SNP50 BeadChip and analyzed 717 horses of 17 breeds, the latter representing mainly various Thoroughbred-based European warmblood horses. Collectively, the three algorithms identified 50 common CNVs associated with 153 genes. Like in earlier studies (Doan et al. 2012a; Doan et al. 2012b; Dupuis et al. 2013), the prevailing GO terms for CNV genes were sensory perception, signal transduction, and cellular components. Additionally, the authors carried out GWAS and identified significant losses in ECA1, 8, and 9 that could be related to body size in horses. Notably, homeologous regions to these CNVs in HSA1 and HSA9 have been associated with height in humans (Dauber et al. 2011).

The most recent study analyzed CNVs in the genomes of indigenous Chinese horses using a custom made NimbleGen WG tiling array (Wang et al. 2014b). The array comprised 1,402,459 probes with 1.6 kb average distance between the probes. The horses were from five Chinese breeds: Mongolian, Abaga, Hequ, and Kazakh horses from the plateau, and the Debao horse from plains. The Thoroughbred was the reference.

A total of 353 CNVRs were identified across all autosomes. The size of CNVRs ranged from 6.1 kb to 1.45 Mb. The most interesting finding was the discovery of CNVs in 7 heme binding genes: *CYP4A11*, *CYP4X1*, *EIF2AK1*, *CYP2C18*, *CYP4F22*, *NOS2*, *CYP4B1*. These CNVs were present mainly in the Hequ horse and in some other plateau breeds but not in the plain horses. The authors theorize that CNVs in heme binding genes might be related to the adaption to the severe environment of the plateau.

In summary, while the 5 above described studies set an important foundation for CNV research in horses, the current information is not adequate for efficient discovery of variants affecting equine health and disorders. This is because the studies have used different CNV discovery platforms, the number of breeds and individuals is very limited, and the majority of CNVs have not been validated by multiple independent studies. Also, the available information has not been integrated to facilitate the analysis of known and the discovery of new CNVs in the horse.

**Prospective equine complex traits for CNV research**

Growing knowledge about the association of CNVs with human complex traits (see section: "Phenotypes, genetic disease and disorders associated with CNVs") justifies and guides the launch of similar research in the horse. As follows is an overview about biomedically and economically important equine traits that have a complex and likely polygenic genetic component, and share similarity with human conditions associated with CNVs.

Recurrent Airway Obstruction (RAO)

Recurrent airway obstruction (RAO) or heaves is a common, lifelong, inflammatory disease of respiratory airways which predominantly affects middle-aged and older horses worldwide (Buechner-Maxwell et al. 1994; Robinson 2001).

### *Characterization of RAO phenotype*

RAO is characterized by coughing and increased breathing effort due to cholinergic bronchospasm, and neutrophil and mucus accumulation in the airways as a result of hypersensitivity reactions to various inhaled allergens, in particular fungal spores (Art et al. 1999; Robinson 2001). Among the clinical symptoms are nasal discharge, flaring of the nostrils, labored expiratory effort, and abnormal respiratory sounds such as crackles and wheezes. Severely affected horses typically have higher neutrophil number in tracheobronchial secretion (TBS) and bronchoalveolar fluid (BALF), as well as airway hyper-responsiveness. Additionally, they may exhibit anorexia, weight loss, and dyspnea or shortness of breath (Armstrong et al. 1986; Robinson et al. 1996). As inflammation and obstruction lead to respiratory distress and exercise intolerance, the disease has a devastating impact on equine health and welfare, as well as to the economy of the equine industry. This is why American Association of Equine Practitioners has ranked RAO and other respiratory ailments as the second highest priority disorders (after laminitis) requiring research.

The phenotype of RAO can be determined by a standardized questionnaire or by detailed clinical tests. The questionnaire gathers information about the horse's history of chronic coughing and includes data about gender, age, and signs of respiratory disease, i.e., coughing, nasal discharge, type of breathing and performance. Also, there are detailed questions on management and feeding, time spent outdoors, de-worming strategies and specific compounds, as well as other diseases and the use of the horse. This information is combined into a Horse Owner Assessed Respiratory Signs Index (HOARSI) (Ramseyer et al. 2007; Laumen et al. 2010). The HOARSI index scores from 1 to 4 where HOARSI 1 denotes unaffected normal horses; HOARSI 2 denotes individuals with ambiguous phenotype, while HOARSI 3 and HOARSI 4 denote clearly RAO-affected horses (Laumen et al. 2010). However, the most accurate way for determining RAO phenotype is to document the pathophysiological changes by conducting a full clinical exam (Gerber et al. 2003; Gerber et al. 2004). This includes records on the respiratory and heart rate, coughing or nasal discharge, rectal temperature, the submandibular lymph nodes, the sensitivity of the pharynx/larynx area and the coughing reflex. The clinical test also involves estimation of lung functions by FOM (forced oscillation mechanics) (Ryhner et al. 2008), scoring of mucus accumulation by endoscopy (Gerber et al. 2004), and cytological analysis of tracheo-bronchial secretion (TBS) and bronchio-alveolar lavage fluid (BALF) by haemocytometer and light microscopy (Gerber et al. 2004; Ryhner et al. 2008).

### *Etiology of RAO*

RAO is a complex disease as both environmental and genetic factors are involved. While the environmental causes of RAO are established, little is known about the intrinsic factors and mechanisms in equine body that lead to RAO. The presence of a polygenic genetic component in equine RAO was postulated over 70 years ago (Schaeper 1939). More recent studies agree with this and suggest a complex inheritance mode (Schaeper 1939; Gerber et al. 2009). RAO has been studied using a combination of genomics approaches, such as microsatellite- and SNP-based GWAS (Swinburne et al. 2009; Shakhsi-Niaei et al. 2010; Klukowska-Rotzler et al. 2012a; Shakhsi-Niaei et al. 2012) and candidate gene and WG gene expression profiling (Kachroo et al. 2010; Venugopal et al. 2010; Cote et al. 2012; Klukowska-Rotzler et al. 2012b; Lanz et al. 2013). In one study pathways analysis of candidate genes was combined with proteomic data from bronchoalveolar lavage fluid of affected and control horses (Racine et al. 2011). While the findings provide some insight into the multi-genic and complex nature of the condition and show the likely involvement of interleukins and other innate immunity related genes (Swinburne et al. 2009; Shakhsi-Niaei et al. 2010; Klukowska-Rotzler et al. 2012b; Shakhsi-Niaei et al. 2012), little is known about the genetic predisposition of horses to RAO, and molecular genetic mechanisms modulating pathogenesis of the disease.

Several studies have been conducted for the search of candidate genes and/or genomic regions responsible for RAO. Mucin glycoprotein alteration and mucus accumulation during RAO suggests elevated expression of mucin genes in affected

horses. Expression analysis of *MUC5AC* and *MUC2* showed upregulation of *MUC5AC*, while no relation between *MUC2* and RAO was detected (Gerber et al. 2003). Later it was shown that the observed upregulation of *MUC5AC* and several other genes is not related to mucus formation in RAO affected horses (Ryhner et al. 2008). The complex genetic nature of RAO is further illustrated by a microsatellite analysis in two Swiss Warmblood horse families, where association between interleukin 4 alpha receptor (*IL4RA*) gene in ECA13 was found in one family but not in the other (Jost et al. 2007). Additional studies of the genetic mode of RAO inheritance in the two Swiss Warmblood families confirmed that in one family the condition is associated with a 20 kb quantitative trait locus (QTL) in ECA13 and inherited in an autosomal recessive mode (Swinburne et al. 2009), whereas in the other family, RAO is associated with ECA15 and inherited in an autosomal dominant mode (Gerber et al. 2009; Swinburne et al. 2009). At the same time, RAO phenotypes in the two families in the HOARSI scale (Laumen et al. 2010) were very similar, indicating genetic heterogeneity of the phenotype. Several follow-up studies confirm that the *IL4R* gene which regulates IgE production and stimulates Th2-cells (Lebman and Coffman 1988) is associated with RAO in some horse families (Klukowska-Rotzler et al. 2012b). However, no genetic variation has been found in the coding region of the gene and it is althought that probably the association is due to a variant in the non-coding regulatory region of *IL4R* (Shakhsi-Niaei et al. 2012). The QTL in ECA13 has also been analyzed for other candidate genes among which *ITGAX* caught attention because of its role in immune-complex mediated hypersensitivity. However, no association between RAO and *ITGAX*

was found (Shakhsi-Niaei et al. 2010). Overall, while these studies have identified some genes and genomic regions associated with RAO, the complex genetics of predisposition and pathogenesis of RAO remain elusive.

### *RAO and human asthma*

Many of the clinical symptoms observed in RAO are strikingly similar with those of human asthma (Zhang et al. 2008; Swinburne et al. 2009). Similarities are also in the genetics of the two conditions: both conditions show variable expression and genetic heterogeneity across populations, and the genetic factors underlying RAO and asthma remain elusive. However, while just a few candidate genes have been proposed for RAO (Jost et al. 2007; Shakhsi-Niaei et al. 2012), there are hundreds of candidate genes for asthma (Rogers et al. 2013). Both disorders involve immune system genes, such as interleukins *IL-4* and *IL-5*, and interferon-gamma (*IFN-γ*) elevated levels of which have been observed in RAO and asthma (Lavoie et al. 2001; Lavoie-Lamoureux et al. 2010). However, compared to humans, the population structure of many horse breeds is better suited for identifying genotypes associated with RAO. Therefore, equine RAO may prove to be a good animal model for human asthma. Furthermore, RAO represents the only natural model of asthma (Swinburne et al. 2009; Scharrenberg et al. 2010) besides feline asthma, and has been investigated much more intensively than the latter (Snapper 1986; Anton et al. 2005).

***RAO and CNVs***

In humans, a large proportion of the established disease-related CNVs are associated with immune system disorders including asthma (Walsh et al. 2010; Lee et al. 2011; Rogers et al. 2013). Among these associations, one of the best characterized is an association of a common deletion in *GSTM1* gene with asthma and airflow obstruction (Rogers et al. 2009; Rogers et al. 2013). Likely involvement of CNVs in asthma is further supported by a recent study where 58 genes (21%) out of 270 candidate genes for asthma were located within or near a CNV (Rogers et al. 2013). The similarity of equine RAO with human asthma (Swinburne et al. 2009) and evidence of the association of human asthma with CNVs, strongly justifies the initiation of research aimed at identifying CNVs potentially associated with RAO.

Disorders of Sexual Development (DSDs) and reproduction

Disorders of sexual development and reproduction include a broad variety of complex conditions that affect sex determination, sexual differentiation, the development and function of gonads, and fertility (Wilhelm et al. 2007). The latter has a direct bearing on reproduction and production and is therefore an economically important trait. In contrast to the importance of the problem, very little is known about the genetic component of DSDs in horses or other species. This is largely because sexual development and reproduction are regulated by almost 20% of the genes (~5000) in the mammalian genome (Hargreave 2000; Matzuk and Lamb 2002; Carrell 2007; Krausz

and Giachini 2007) which function in complex spatio-temporally regulated gene networks. Difficulties are also encountered in precise clinical characterization of DSDs phenotypes and the collection of research samples (Blanchard et al. 2000; Turner and Casas-Dolz 2002).

### *Mammalian sex determination and disorders of sexual development*

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 and Strong 1959). The search for the testis-determining factor (*TDF*) on the Y chromosome lasted for several decades with numerous candidate genes being studied and discounted (Silvers and Wachtel 1975; 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). 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 (Marshall Graves 2008; Sekido and Lovell-Badge 2008).

Furthermore, recent studies have identified genes (*RSPO1*, *WNT4*, β-catenin) (Figure 2) necessary for the molecular pathways in female sex determination and ovarian differentiation (Chassot et al. 2008; Sekido and Lovell-Badge 2009) . Until then it was thought that while *SRY* initiates the male differentiation pathway, female development

happens as a default in the absence of the Y chromosome and *SRY*, and does not require specific genetic triggers. 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.

Disruption of the many genes and genetic pathways that regulate sexual development can cause DSDs at many different levels. For example, disruption of the very early steps of sex determination at the stage of a bi-potential gonad can result in abnormal gonadal development in both sexes (Marshall Graves 2008; Sekido and Lovell-Badge 2009) (Figure 2).

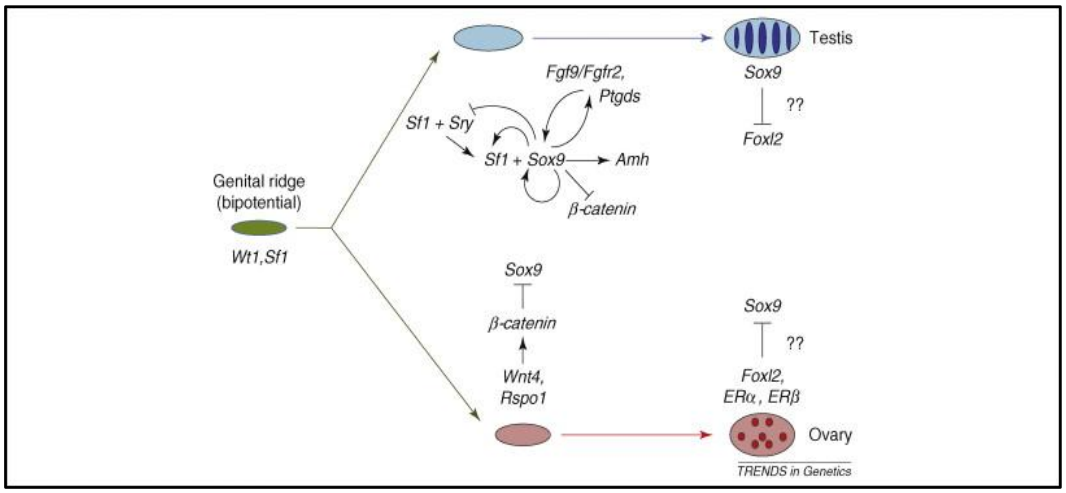


**Figure 2: The molecular and genetic events in mammalian sex determination.** The bipotential genital ridge is established by genes including Sf1 and Wt1, the early expression of which might also initiate that of Sox9 in both sexes. β-catenin can begin to accumulate as a response to Rspo1–Wnt4 signaling at this stage. In XX supporting cell precursors, β-catenin levels could accumulate sufficiently to repress SOX9 activity, either through direct protein interactions leading to mutual destruction, as seen during cartilage development (Akiyama et al. 2004), or by a direct effect on Sox9 transcription. However, in XY supporting cell precursors, increasing levels of SF1 activate Sry expression and then SRY, together with SF1, boosts Sox9 expression. Once SOX9 levels reach a critical threshold, several positive regulatory loops are initiated, including autoregulation of its own expression and formation of feed-forward loops via FGF9 or PGD2 signaling. If SRY activity is weak, low or late, it fails to boost Sox9 expression before β-catenin levels accumulate sufficiently to shut it down. At later stages, FOXL2 increases, which might help, perhaps in concert with ERs, to maintain granulosa (follicle) cell differentiation by repressing Sox9 expression. In the testis, SOX9 promotes the testis pathway, including Amh activation, and it also probably represses ovarian genes, including Wnt4 and Foxl2. However, any mechanism that increases Sox9 expression sufficiently will trigger Sertoli cell development, even in the absence of SRY (Sekido and Lovell-Badge 2009).

Later on, disruption of genes involved in steroid hormone metabolism by the developing gonads, can affect sexual differentiation, gonadal maturation and the development of secondary sexual characteristics, such as sex specific behavior and external sexual phenotypes - the presence of penis and descended testes in males, and vulva and vagina in females (Auchus 2004; Hughes 2008; Fluck et al. 2011; Biason-Lauber et al. 2013) (Figure 3). Many forms of DSDs manifest as discrepancies between the genetic sex (sex chromosomes), gonadal sex (testes or ovaries), and phenotypic sex. In literature, the resulting phenotypes appear under terms like intersexuality, gonadal dysgenesis, pseudohermaphroditism, testicular feminization, true hermaphroditism, and sex reversal syndromes (Howden 2004; Villagomez et al. 2011; Lear and McGee 2012), depending on at which levels (genetic, gonadal, phenotypic) sexual characteristics are altered or observed.

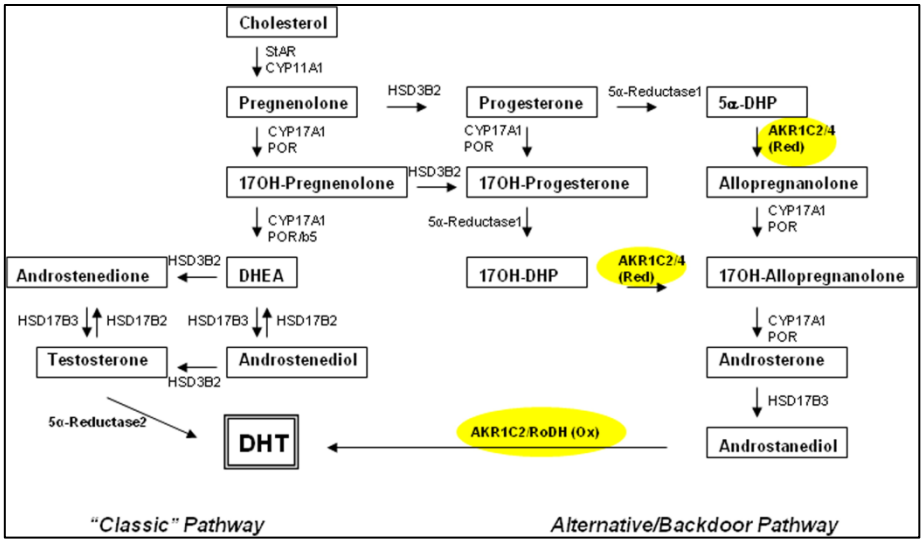


**Figure 3: Synthesis of dihydrotestosterone via the classic and alternative pathways.** The classic pathway of steroidogenesis leading to dihydrotestosterone is shown on the left, and the alternative pathway is shown on the right. The factors in the classic pathway are CYP11A1 (cholesterol side-chain cleavage enzyme, P450scc), StAR (steroidogenic acute regulatory protein), CYP17A1 (17α-hydroxylase/17,20-lyase, P450c17), HSD3B2 (3β-hydroxysteroid dehydrogenase, type 2), HSD17B3 (17β-HSD3 [17β-hydroxysteroid dehydrogenase, type 3] and 5α-reductase, type 2 [5a-reductase 2, encoded by SRD5A2]). The alternative pathway is characterized by the presence of additional enzymes: 5α-reductase, type 1 (5a-reductase 1, encoded by SRD5A1), AKR1C2 3 (3α-reductase, type 3) and possibly AKR1C4 (3α-reductase, type 1) and RoDH (3-hydroxyepimerase, encoded by HSD17B6). Most steroids are identified by their trivial names; 17-hydroxy-dihydroprogesterone (17OH-DHP) is 5α-pregnane-17α-ol-3,20-dione; 17-hydroxy-allopregnanolone (17OH-allo) is 5α-pregnan-3α,17α-diol-20-one; 5α-dihydroprogesterone (5α-DHP) is 5α-pregnane-3,20-dione, and allopregnanolone is 3α-hydroxy-dihydroprogesterone (3α-OH-DHP) or 5α-pregnane-3α-ol-20-one (Fluck et al. 2011).

### *Sex reversal syndromes*

Sex reversal is a situation where the genetic sex (the karyotype, sex chromosomes) disagrees with the gonadal and/or the phenotypic sex. The affected individuals are sterile with various degrees of abnormalities in sexual development. Cases of both male-to-female and female-to-male sex reversal have been reported and studied in mice, human and most of the domestic species (Vaiman and Pailhoux 2000; Villagomez et al. 2009; Jimenez et al. 2013) including the horse (Kent et al. 1986; Raudsepp et al. 2010; Pujar and Meyers-Wallen 2012; Jimenez et al. 2013). In horses, sex reversal syndrome is of particular concern because it not only negatively affects fertility but might raise issues regarding high-performance horses. For example, recently two elite Standardbred trotters with questionable sexual identity were disqualified from races in the female category and made headlines in New York Times (Finely 2009) and Horse.com (Lear and McGee 2012).

In horses, the male-to-female sex reversal where phenotypic mares or mare-like individuals have 64,XY karyotype is probably the only true sex reversal condition (Raudsepp et al. 2010). A reversed situation where phenotypic males have normal female karyotype, as it has been described in humans (Zenteno-Ruiz et al. 2001), is probably not existing in horses. So far, all cases reported as equine 64,XX sex reversal, actually categorize as cases of intersex, hermaphroditism or male pseudohermaphroditism (Milliken et al. 1995; Meyers-Wallen et al. 1997; Buoen et al. 2000; Chowdhary and Raudsepp 2000; Vaughan et al. 2001; Bannasch et al. 2007; Villagomez et al. 2009; Raudsepp et al. 2013). Chromosomes of these animals are

normal, and very little is known regarding molecular causes of 64,XX conditions in horses (Pujar and Meyers-Wallen 2012). Although, mutations in several genes, such as *SOX3, SOX9, RSPO1, PISRT1, FOXL2, WNT4* have been reported for XX female-to-male sex reversal syndromes in other species (Pujar and Meyers-Wallen 2012).

Slightly more is known about the equine 64,XY sex reversal syndrome. This is probably because it is the second most frequent sex chromosome abnormality after X chromosome monosomy (Das et al. 2012) and accounts for approximately 12-30% of all cytogenetic abnormalities in horses (Power 1986; Bowling et al. 1987; Lear and Bailey 2008; Raudsepp et al. 2010). The affected individuals 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 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; Raudsepp et al. 2010; Villagomez et al. 2011). All described cases have been infertile except one that produced a normal 64,XX filly (Sharp et al. 1980). With regards the Y chromosome, the equine 64,XY sex reversal has two forms: the *SRY*-negative and the *SRY*-positive forms. Recent studies showed that the *SRY*-negative form is typically caused by various deletions on the Y chromosome, all of which include the *SRY* gene (Raudsepp et al. 2010). In contrast, the *SRY*-positive condition is likely genetically heterogeneous and involves multiple genes (Switonski et al. 2005; Raudsepp et al. 2010;

Lear and McGee 2012; Pujar and Meyers-Wallen 2012). In one Friesian 64,XY *SRY*-positive horse with hypospadias a missense mutation was found in *MAMLD1* gene, although without clear causative relationship to the sex reversal condition (De Lorenzi et al. 2010); in one Quarter Horse family, a c.1A>G mutation was found in the start codon of the *AR* gene (Revay et al. 2012). The latter is to date the only known causative mutation for *SRY*-positive sex reversal in horses, whereas the genetic causes of the majority of similar conditions remain, as yet, undefined.

### *Cryptorchidism*

Cryptorchidism (CO) is the most common non-lethal developmental defect in mammals with prevalence varying in a wide range between species, breeds, and populations (Amann and Veeramachaneni 2006; Amann and Veeramachaneni 2007). A generally accepted frequency of CO in full term male birth is 2-8% for humans (Agoulnik et al. 2012) and horses (Hayes 1986; Amann and Veeramachaneni 2006; Amann and Veeramachaneni 2007; Foster and Ladds 2007; Arighi 2011; Russell and Pollock 2011), while a broader range of variation has been reported for CO in dogs and pigs (Dolf et al. 2008). The condition manifests in the failure of one (unilateral) or both (bilateral) testes to descend into the scrotum at the time typical for a species (Amann and Veeramachaneni 2006; Amann and Veeramachaneni 2007). Undescended testis can be retained at different locations along the path of testis descent from the abdomen through inguinal canal to a proper position in the scrotum. In most mammals, a retained testis most commonly is located in an abdominal position. In horses, retention of testes within

the inguinal canal is equally common (Amann and Veeramachaneni 2006; Amann and Veeramachaneni 2007). Overall, CO usually is unilateral and retention of the testes occurs with similar frequency on the left and right sides. In stallions, however, left-sided CO testes are more frequently located in the abdomen (75.2%), and right-sided CO testes in the inguinal canal (58.2%) (Stickle and Fessler 1978). Cryptorchidism might occur as an isolated condition or as a part of other developmental disorders, known as syndromic CO (Amann and Veeramachaneni 2006; Amann and Veeramachaneni 2007).

The retained testis undergoes dysgenesis, rendering bilateral CO sterile (Amann and Veeramachaneni 2006; Agoulnik and Feng 2007; Amann and Veeramachaneni 2007; Massart and Saggese 2010; Arighi 2011), while unilateral CO are generally fertile. In stallions, where unilateral CO accounts for 81-93% of diagnosed cases (Hayes 1986; Amann and Veeramachaneni 2007; Arighi 2011), the undescended testis is of reduced testicular mass with no exocrine capability, resulting in reduced numbers of ejaculated sperm (Foster and Ladds 2007; Russell and Pollock 2011). However, the affected stallions generally produce sufficient normal sperm to yield acceptable pregnancy rates in commercial breeding programs. Surgical removal of a retained testis might be indicated because studies in humans (Wood and Elder 2009), dogs (Liao et al. 2009) and to a lesser extent in horses (Pratt et al. 2003; Foster and Ladds 2007; Arighi 2011) reveal that a maldescended testis poses an increased risk for the development of primary testicular neoplasms, such as seminomas, Leydig cell tumors and Sertoli cell tumors. Surgical removal of an undescended testis, depending on the position of the testis, might be challenging and further increase the health risk (Russell and Pollock 2011). Also, the

undescended testes continue testosterone production, due to which males with bilateral cryptorchidism show the same behavioral characteristics as males with two scrotal testes. Therefore, castration of CO stallions is necessary to eliminate associated behavioral characteristics (Russell and Pollock 2011). However, the surgical procedures required for castratration of a uni-or bilateral cryptorchid stallion are more complicated and expensive, as compared to routine castration (Arighi 2011; Russell and Pollock 2011).

Because of possible infertility, increased risk for gonadal tumors, clinical complications, vertical transmissibility, and increased medical/veterinary costs, CO has been a focus of clinical and genetic research for a long time. The majority of studies involve humans (Agoulnik and Feng 2007; Harris et al. 2010; Agoulnik et al. 2012; Cannistraci et al. 2013) and model species, such as mice and rats (Barthold 2008; Agoulnik et al. 2012; Barthold et al. 2013). Among domestic species, CO has been studied mainly in horses (Hayes 1986; Arighi 2011; Russell and Pollock 2011), followed by pigs (Rothschild et al. 1988; Dolf et al. 2008) and dogs (Amann and Veeramachaneni 2007; Dolf et al. 2008), and to a lesser extent in cattle, sheep and goats (Amann and Veeramachaneni 2007). Despite this, the etiology of cryptorchidism remains, for the most part, unknown. It is generally accepted that the causes of cryptorchidism are complex and multifactorial. These include environmental factors, such as endocrine-disrupting chemicals with anti-androgenic and/or estrogenic effects, but also epigenetic and genetic factors (Barthold 2008).

Abundant evidence supports the presence of a genetic component in CO. For example, about 25% of CO cases in humans are inherited (Barthold 2008); there are multiple reports about familial CO in humans (Agoulnik and Feng 2007; Massart and Saggese 2010) and animals (Diribarne et al. 2009; Zhao et al. 2010); reports about pedigrees where multiple generations have been affected (Perrett and O'Rourke 1969), and case reports of unilateral CO where the contralateral, normally descended testis may also be altered (Foresta et al. 1996). Some breeds of horses, like Percherons, American Saddlebreds, American Quarter Horses, and ponies, are more prone to have CO than, for example Thoroughbreds (Hayes 1986), thereby pointing to a likely genetic predisposition. Furthermore, it is well known that syndromic CO is frequently a part of complex congenital disorders that are caused by chromosomal aneuploidies and aberrations (Massart and Saggese 2010), although no chromosomal abnormalities have been consistently associated with isolated cryptorchidism.

Despite these observations, the mechanisms of genetic contributions to cryptorchidism are not well understood. There is no consensus about the mode of CO inheritance – an indication that the genetic component of CO is heterogeneous and involves multiple genes. Indeed, the majority of studies in humans (Klonisch et al. 2004; Massart and Saggese 2010), pigs (Rothschild et al. 1988) and dogs (Nielen et al. 2001) propose a polygenic recessive model involving abnormalities in over 20 genes, while a small number of studies support autosomal dominant or Y-linked inheritance for some forms of CO in humans (Pardo-Mindan et al. 1975; Barthold 2008).

Difficulties to identify the genetic component of CO are largely due to the complexity of the gene network regulating sexual differentiation and testes development (Sekido and Lovell-Badge 2009). Testes descent requires an interaction of multiple genes with testicular hormones testosterone, and its more potent derivative dihydrotestosterone (DHT) (Fluck et al. 2011). Testosterone, DHT and their receptor *AR* (Androgen receptor) are critically involved in multiple stages of testes development, including the transinguinal phase (Agoulnik and Feng 2007; Barthold 2008; Hughes and Acerini 2008). The classic biosynthetic pathway from cholesterol to testosterone and the subsequent conversion of testosterone to DHT is well established (Fluck et al. 2011). However more recently, an alternative pathway for DHT synthesis that does not involve testosterone was discovered in marsupials (Wilson et al. 2003), and thereafter shown to be present in humans and other eutherian mammals (Auchus 2004; Fluck et al. 2011; Biason-Lauber et al. 2013; Fukami et al. 2013). The pathway is called 'the backdoor pathway' (Auchus 2004) and involves genes and enzymes, such as 3α-reductases (*AKR1C* gene family) that are not present in the 'classic' DHT synthesis pathway. The importance of the 'backdoor pathway' and *AKR1C* genes in male sexual development was recently demonstrated by a study in humans showing that mutations in *AKR1C2* and *AKR1C4* genes are associated with various disorders of sexual development (DSDs) including CO (Figure 3) (Fluck et al. 2011; Biason-Lauber et al. 2013).

Another important hormone/receptor system in testes development involves the insulin-like factor 3 (*INSL3*) and its G-protein-coupled receptor *RXFP2* which control gubernaculum differentiation, and are essentially involved in trans-abdominal phase of

testis descent. Testes migration depends also on calcitonin-related polypeptide alpha (*CALCA)* which is a chemoattractant and induces the growth of the developing tip of the gubernaculum (Ng et al. 2005). The latter enlarges under stimulation of *INSL3* to anchor the testes in place during gradual abdominal translocation. Further, the production of testosterone during testicular descent is regulated by genes involved in the steroidogenic pathway such as luteinizing hormone (*LH*) and its receptor (*LHCGR*) that stimulate Leydig cells to produce testosterone. Testosterone synthesis, in turn, depends on 3β-steroid delta-isomerase 1 (*HSD3B1)* and aromatase (C*YP19A1*) – the key enzymes which convert testosterone to estradiol and control its availability (Agoulnik and Feng 2007). Additionally, several other loci, such as homeobox A10 (*HOXA10*) and A11 (*HOXA11*), zinc finger proteins *ZNF214* and *ZNF215* implicate testicular descent (Massart and Saggese 2009; Massart and Saggese 2010).

The complex genetic regulation of sexual differentiation and testes development is probably the reason why only limited success has accompanied studies using candidate gene approach. Sequencing and mutation analysis of over 20 candidate genes in humans (Massart and Saggese 2009; Massart and Saggese 2010), mice (Agoulnik and Feng 2007; Agoulnik et al. 2012) and dogs (Pathirana et al. 2010; Zhao et al. 2010) have not revealed any mutations or single nucleotide polymorphisms (SNPs) that are consistently and significantly associated with CO. Likewise, no significant associations were detected by microsatellite genotyping in seven CO candidate genes (*AR, CALCA, ESR1, HOXA10, INSL3, NR5A1, RXFP2*) in a Thoroughbred pedigree composed of 23 CO and 24 unaffected horses (Diribarne et al. 2009).

Some progress, although no true breakthrough, has accompanied genome-wide association studies (GWAS) by genotyping case-control cohorts on SNP beadchips. An association to *TGFBR3* and *BMP7* genes was shown by GWAS in men with testicular dysgenesis syndrome (TDS) – a condition which links testicular germ cell cancer, cryptorchidism and some cases of hypospadias (Dalgaard et al. 2012). GWAS in Siberian Huskies associated CO with chromosomes 6, 9, 24, 27 and X (Zhao et al. 2013). However, as these studies have not been followed up by in depth analysis of positional candidate genes, no causative mutations or risk genotypes for CO have been as yet determined. Nevertheless, identification of multiple associated genes and genomic regions is in agreement with the proposed polygenic nature of CO.

***Disorders of sexual development and CNVs***

Recently, another form of genome variation – copy number variation - has come to the spotlight in the genomics of DSDs and CO. Indeed, studies in humans associate CNVs with disorders of sexual development and reproduction, including urogenital abnormalities, sex reversal, hypospadias and CO (Smyk et al. 2007; Tannour-Louet et al. 2010; Li et al. 2011). For example, in CO individuals, imbalances have been found in 9 genomic regions of which 5, *viz.*, 5p12, 10p14, 12q24, 15q11, and Xq28, were clinically significant (Tannour-Louet et al. 2010). The findings encourage continuing in depth studies of the candidate regions in humans and the initiation of similar research in other species, including the horse.

**Goals of this study**

The focus of this dissertation is the study of copy number variation in the horse genome. The overall goal is to improve the knowledge of CNVs in equine health and disease by characterizing CNVs in normal horse breeds and populations, and in equine congenital disorders and diseases. The findings are expected to facilitate the development or improvement of molecular diagnostic tools for horses, and might also serve as useful models for related biomedical conditions in other domestic species and humans.

We hypothesize that CNVs are associated with normal phenotypic variation, as well as with complex diseases and genetic disorders in the horse. The immediate goals of this dissertation research are:

<u>Objective #1: Investigate naturally occurring CNVs in diverse horse breeds and generate a baseline for future studies.</u>

We argue that in order to find CNVs that that are associated with genetic abnormalities, the first step is to obtain an in depth knowledge about the CNVs in the genomes of normal horses of diverse horse breeds and populations. We will carry out genome-wide CNV analysis using array CGH in diverse horse breeds, each represented by at least 2 individuals. The findings will be compared and merged with the currently available CNV data for the horse. We intend to generate a composite CNV database and,

thus, lay a critical foundation for the discovery of CNVs that are associated with equine diseases, disorders, and traits of importance.

Objective #2: Comparative analysis of genomic copy number variation among horses susceptible and resistant to Recurrent Airway Obstruction (RAO).

Array CGH will be conducted in severely RAO affected (HOARSI 4) and healthy control (HOARSI 1). The goal is to determine the role of CNVs in the genetics of complex equine diseases, such as RAO.

Objective #3: CNV research in equine disorders of sexual development (DSDs) – XY *SRY*-positive sex reversal and cryptorchidism.

Array CGH will be used to study CNVs in two disorders *SRY*-positive sex reversal 64XY females and in equine bilateral abdominal isolated cryptorchidism. The findings will contribute to expand the currently limited knowledge about the genetic regulation of sexual development in horses and other mammals.

# CHAPTER II

# GENOMIC COPY NUMBER VARIATION IN EQUINE POPULATIONS

## Introduction

The outstanding role of DNA copy number variants (CNVs) in phenotypic diversity, adaptations, and evolution was first recognized over 40 years ago by Susumu Ohno (Ohno 1970). Yet, systematic genome-wide discovery and functional interpretation of CNVs started in the past decade with foundational studies in humans (Iafrate et al. 2004; Redon et al. 2006) and mice (Li et al. 2004), followed by genome-wide (GW) CNV discovery in chicken (Griffin et al. 2008), cattle (Liu et al. 2010) and other domestic species (Clop et al. 2012). It is now well established that CNVs are a common feature of vertebrate genomes. They provide a genetic basis for normal phenotypic variations and adaptations, but can also be molecular signatures for congenital disorders and diseases (Girirajan and Eichler 2010; Hall and Quinlan 2012).

According to the current definition, CNVs are DNA sequence variants from at least 50 base-pairs (bp) to over several megabase-pairs (Mb) in size that are involved in deletions, insertions, duplications and translocations, causing structural differences between genomes (Conrad et al. 2010b; Weischenfeldt et al. 2013). Because CNVs are larger than single nucleotide polymorphisms (SNPs), they are responsible for more heritable sequence differences (0.5-1%) between individuals than SNPs (0.1%) (Conrad et al. 2010b; Pang et al. 2010; Weischenfeldt et al. 2013).

One of the central goals of CNV research has been to determine their association with genome instability, genetic diseases, and congenital disorders. It is assumed that CNVs, as a major source of inter-individual genetic variation, could explain variable penetrance of Mendelian and polygenic diseases, and variation in the phenotypic expression of complex traits (Beckmann et al. 2007; Beckmann et al. 2008). Indeed, CNVs have been associated with common complex and polygenic disorders in humans affecting a broad range of biological processes, such as immune response, autoimmunity and inflammation (Redon et al. 2006; Cooper et al. 2007; Rogers et al. 2013); musculoskeletal (Robinson et al. 2012; Alvarado et al. 2013) and cardiovascular systems (Hitz et al. 2012; Peng et al. 2013); neurodevelopment, cognition and behavior (Lupski 2007b; Grayton et al. 2012), and sexual development and reproduction (Ledig et al. 2010a; Tannour-Louet et al. 2010; Carrell and Aston 2011; Tuttelmann et al. 2011; White et al. 2011).

The availability of whole genome (WG) sequence draft assemblies, combined with the advances in array-based technologies and next generation sequencing (NGS), has prompted CNV research in all main domestic species with the most advanced information currently available for the cattle (Matukumalli et al. 2009; Fadista et al. 2010; Liu et al. 2010; Bickhart et al. 2012; Hou et al. 2012a; Hou et al. 2012b; Hou et al. 2012c; Liu and Bickhart 2012), followed by pigs (Chen et al. 2012; Li et al. 2012; Wang et al. 2012; Wang et al. 2013a; Wang et al. 2013b; Wang et al. 2014a), dogs (Chen et al. 2009; Nicholas et al. 2009; Alvarez and Akey 2012; Berglund et al. 2012; Gurgul et al. 2014; Molin et al. 2014), sheep (Fontanesi et al. 2010a; Liu et al. 2013) and goats

(Fontanesi et al. 2010b). Five CNV studies have been conducted in horses reporting the discovery of copy number variants in the whole genome (Doan et al. 2012b; Dupuis et al. 2013; Metzger et al. 2013a; Wang et al. 2014b) or in gene exons (Doan et al. 2012a), and attempting to associate CNVs with diseases (Dupuis et al. 2013), adaptations (Wang et al. 2014b) or phenotypic traits (Doan et al. 2012a; Metzger et al. 2013a). While these studies set a foundation for understanding the role of CNVs in horse biology, the current information is inadequate for efficient discovery of variants affecting equine health and disorders. This is because the studies have used different CNV discovery platforms, the number of breeds and individuals in some studies is very limited, and the majority of reported CNVs are unique and not validated by two or more independent studies. Also, the available information has not been integrated into a composite dataset to facilitate the analysis of known, and the discovery of new CNVs.

The aim of this study is to improve the current rather limited knowledge of CNVs in horses by their genome-wide discovery in multiple individuals of additional diverse horse breeds. Using a custom-made WG tiling array, we generate a CNV map for the horse genome and integrate this with the previous CNV studies into a composite dataset.

## Material and methods

### Array design

Recently, as collaboration between Texas A&M University (Drs. Raudsepp and Chowdhary) and the University of Adelaide (Dr. Adelson), a whole-genome tiling array for the horse was constructed (Qu et al. 2011 PAG abstract). The array comprises 417,377 60-mer oligonucleotides distributed in gene containing and intergenic sequences of the horse genome. Median distance between the tiles on the array is 7.5 kb, reducing to 4 kb in subtelomeric regions, and increasing to ~ 20 kb in the Y chromosome. The array was designed using horse genome draft sequence (EcuCab2); (Wade et al. 2009), Oligowiz2.0, ArrayOligoSelector, and ArrayDesign (Rouillard et al. 2002) software packages. Array sequences for the autosomes and the X chromosome originated from the horse reference genome (Thoroughbred mare *Twilight*; Wade et al. 2009). Sequences for the Y chromosome originated from the genome of the DNA donor of the CHORI-241 horse genomic BAC library (Thoroughbred male, *Bravo*) (Paria et al. 2011; our unpublished data). Before inclusion to the array, the specificity of all sequences were analysed with BLAT and BLAST against EcuCab2 reference genome sequence. Probes with more than one hit in the genome were discarded. Possible cross-hybridization of the probes was further evaluated using Kane's parameters (Kane et al. 2000) and all probes that had a total percent identity >75-80% with a non-target sequence, or probes with

contiguous stretches of identity >15 nucleotides with a non-target sequence were discarded.The array was designed to target the following regions of interest:

***Genes***

The array consists of 85,852 probes corresponding to one or more exons of 18,763 annotated equine genes. Genes with one to four exons have a probe corresponding to each exon. The number of probes (N) for genes with more than four exons was determined by the formula N = [(M-4)/5-4}] (where M is the number of exons). Probes could not be designed for the exons of 967 genes. Instead, introns and 300 bp upstream or downstream sequences were used to select suitable oligonucleotides. Despite these efforts, no probes could be designed for 336 equine genes. The majority of these are members of large gene families, such as olfactory receptors, keratins and ribosomal proteins.

***Intergenic chromosomal regions (excluding subtelomeres)***

These regions are represented by 305,416 probes and correspond to all horse autosomes, the X chromosome and chromosome Un. The latter represents sequence scaffolds that have not yet been assigned to chromosomes.

### *Subtelomeric regions*

A total of 5,716 probes were designed from subtelomeric sequences defined as the terminal 1 Mb of each chromosome arm. These regions were specifically targeted because subtelomeres contain transcribed gene families and are hotspots of DNA breakage and repair, and undergo structural rearrangements more frequently than the rest of the genome (DeScipio et al. 2008; Riethman 2008). Enrichment of the array with subtelomeric sequences is an important and unique feature because one of the pitfalls of most animal and many human arrays is the under-representation of probes in the subtelomeric regions (Stankiewicz and Beaudet 2007). At the same time, subtelomeric regions which are extremely difficult to study by cytogenetic methods, are frequently involved in structural imbalances associated with congenital disorders and diseases both in humans (Tannour-Louet et al. 2010) and animals (Thomas et al. 2009).

### *The Y chromosome*

Another unique feature of the array is the incorporation of 519 sequences from horse Y chromosome (unpublished data). Typically, the Y chromosome is missing from tiling arrays because Y sequences are not available for most species.

A Cytoband file was generated to align the horse draft sequence assembly with the cytogenetic map (ISCNH 1997). This allows the generation of files showing the distribution of SVs along individual chromosomes. For bioinformatics purposes, fictional cytoband files were generated also for chromosome Un (probes with sequence information but no chromosomal assignment) and for the Y chromosome (probes with

chromosomal location but not included in WG sequence assembly). Thus, the CNV analysis will encompass the entire horse genome including the Un and the Y chromosomes.

Array production

The array, designated as ***Texas-Adelaide horse WG tiling array***, was fabricated by Agilent Technologies using Agilent SurePrint G3 technology and 2 x 400K chip format (two arrays on a single slide). The array is available at Agilent Technologies; Design ID #030025, Cat. No G4124A.

Selection of breeds and animals

The CNV discovery cohort comprised 38 horses representing 16 diverse breeds and the Przewalski's horse (Table 6). An additional cohort of 52 horses representing the same 16 breeds was used for quantitative PCR validation of CNVs.

Horse breeds were selected according to the recent population studies (Wade et al. 2009; McCue et al. 2012; Petersen et al. 2013a; Petersen et al. 2013b) with an aim to maximize the genetic diversity among samples and to encompass the common warm blood, cold blood (draft) and native pony breeds (Gorrie et al. 2008). The warm blood horses were represented by Akhal-Teke, American Quarter Horse, Arabian, Standardbred, Swiss Warmblood, and the Thoroughbred – well-known breeds used for

racing, riding, show jumping, dressage, and other show competitions. The draft horses were represented by Belgian, Clydesdale, Friesian, and Percheron breeds. Among the indigenous ponies, the Mongolian native horse is used for riding/racing, but also for meat and milk. Caspian, Fell and Exmoor ponies are distinguished by small size and hardiness, and used for agriculture, dressage, and as pets. The American Miniature Horse is characterized by extremely small body size (Petersen et al. 2013b), and the Sorraia horse represents Iberian horse breeds and is used for riding and as a pet (Gorrie et al. 2008).

Geographically, the selection included breeds from Asia/Orient, such as Akhal-Teke, Arabian, Caspian, and Mongolian; from Europe, such as Belgian, Clydesdale, Exmoor Pony, Fell Pony, Friesian, Percheron, Sorraia, Swiss Warmblood, and Thoroughbred, and breeds of American origin, such as the American Miniature Horse, American Standardbred, and American Quarter Horse (Gorrie et al. 2008).

**Table 6: Horse breeds (n=16) and individuals (n=38) used in this study.**

| **Domestic horse, *Equus caballus*** | | | |
|---|---|---|---|
| **Breed** | **ID** | **Sex** | **Source of DNA** |
| Akhal-Teke1 | BP131 | M | Blood |
| Akhal-Teke2 | BP132 | M | Blood |
| American Miniature Horse1 | BP7 | M | Blood |
| American Miniature Horse2 | BP353 | F | Hair |
| American Quarter Horse1 | BP3 | M | Blood |
| American Quarter Horse2 | H261 | F | Blood |
| American Quarter Horse3 | H528 | M | Blood/hair |
| Arabian1 | BP35 | M | Blood |
| Arabian2 | BP290 | F | Hair |
| Belgian Draft1 | BP117 | M | Blood |
| Belgian Draft2 | BP291 | F | Hair |
| Caspian Pony1 | BP60 | M | Blood |
| Caspian Pony2 | BP288 | F | Hair |
| Clydesdale1 | BP65 | M | Blood |
| Clydesdale2 | BP322 | F | Hair |
| Exmoor Pony1 | BP197 | M | Blood |
| Exmoor Pony2 | BP297 | F | Hair |
| Fell Pony1 | BP85 | M | Blood |
| Fell Pony2 | BP285 | F | Hair |
| Friesian1 | H519 | M | Blood |
| Friesian2 | H481 | F | Blood |
| Friesian3 | H525 | M | Blood |
| Friesian4 | H526 | M | Blood |
| Mongolian Native Horse1 | BP304/GC52020 | M | Hair |
| Mongolian Native Horse2 | BP287/GC52022 | F | Hair |
| Percheron1 | H520 | M | Blood |
| Percheron2 | BP351 | F | Hair |
| Sorraia1 | BP303/ECAS32 | M | Blood/hair |
| Sorraia2 | BP295 | F | Hair |
| Standardbred1 | BP356/M1054 | M | Blood |
| Standardbred2 | BP357/M1039 | F | Blood |
| Swiss Warmblood1 | HOARSI1-154 | M | Blood |
| Swiss Warmblood2 | HOARSI1-146 | F | Blood |
| Swiss Warmblood3 | HOARSI1-140 | F | Blood |
| Thoroughbred1 | Bravo* | M | Blood |
| Thoroughbred2 | Twilight** | F | Blood |
| **Przewalski horse, *Equus przewalskii*** | | | |
| Przewalski's Horse1 | KB4064 | M | Fibroblasts |
| Przewalski's Horse2 | KB4070 | F | Fibroblasts |

* Male Thoroughbred *Bravo* was the DNA donor for the CHORI-241 horse whole genome BAC library (http://bacpac.chori.org/equine241.htm);

** Female Thoroughbred *Twilight* was the DNA donor for the horse genome sequence draft assembly EcuCab2 (Wade et al. 2009).

DNA isolation & quality control

DNA was isolated from peripheral blood and/or hair using standard PCI (Phenol/ Chloroform/ Isoamyl alcohol) method (Sambrook et al. 1989) or Gentra Puregene Blood Kit (QIAGEN, USA) according to manufacturer's protocol. The latter was modified to adjust it for DNA isolation from hair follicles. Briefly, hair follicles were incubated in cell lysis solution for at least 72 hrs. After isopropanol treatment to precipitate DNA pellet, the sample was centrifuged at 16,000 x g for 20-30 minutes, instead of 3 minutes at 2000 x g, according to the original protocol. The same applied after treatment with 70% ethanol. The DNA was cleaned with DNeasy Blood and Tissue kit (Qiagen) with the following modifications to the manufacturer's protocol: during DNA purification, 80% ethanol was used instead of solution AW2, and the final DNA pellet was eluted in water instead of solution AE. The final quality and quantity of the DNA samples were checked by gel electrophoresis on a 1% agarose gel and by Nanodrop spectrophotometry (Thermo Scientific).

Array Comparative Genomic Hybridization

Probe labeling and array CGH experiments were performed according to Agilent Technologies Protocol Version 6.2.1. All hybridizations comprised of a pair of differently labeled probes, one of which was always the reference DNA – a Thoroughbred mare *Twilight* for females and a Thoroughbred stallion *Bravo* for males.

Briefly, 1µg of purified genomic DNA was digested with restriction enzymes *RsaI* and *AluI* for 2 hours at 37°C and 20 minutes at 65°C to produce 200-500 bp fragments. Cleaved samples were labeled with fluorescence dyes Cy3 (for the reference DNA) or Cy5 (for sample DNA) by random priming using Genomic DNA Enzymatic Labeling Kit (Agilent Technologies). The final labeling reaction contained random primer (5 µl), 1X dNTP mix (5µl), 1X reaction buffer (10 µl), Exo-Klenow enzyme (1 µl), and Cy3- or Cy5-dUTP (3 µl) and was conducted for 2 hours at 37°C, 10 minutes at 65°C. The products were cleaned with 30 kDa filters (Amicon) and the yield and specific activity of labeled DNA was determined with a Nanodrop spectrophotometer. Typical yield for 1 µg of starting DNA was 6-8 µg; specific activity for Cy3 was 25-40 pmol/µg and for Cy5 20-35 pmol/µg. The hybridization mixture was prepared using Agilent Oligo aCGH Hybridization Kit and contained equal quantity of Cy3 and Cy5 labeled probes, 1 µg/µL horse Cot1 DNA, 10X blocking agent, and 2X Hi-RPM buffer. Denatured and pre-annealed probe mixture was applied onto gasket slide, placed in Agilent SureHyb hybridization chamber, 'sandwiched' with an array slide and incubated in Agilent hybridization oven at 65°C for 40 hours. The array slides were washed with Agilent aCGH Wash Buffers 1 and 2, and dried with Acetonitrile and Stabilization and Drying Solutions. Washing in Wash Buffer 1 was done for 10 minutes instead of the 5 minutes recommended by the Agilent protocol.

Array CGH data analysis

The array slides were scanned with Agilent SureScan DNA Microarray Scanner and Agilent Scanner Control software v8.3 with the following settings: slide id - auto detect; channels - R+G; resolution – 2 µm; tiff – 16bit; R PMT and G PMT are 100%, and XDR set to "0.05". The settings can be changed according to labeling and hybridization performance.

The data were extracted and normalized with Agilent Feature Extraction software v10.10.1.1 and saved in .fep format. The Feature Extraction software also checks the quality of aCGH by measuring Derivative $Log_2$ Ratio Standard Deviation (DLRSD), Signal-To-Noise Ratio (SNR) and Background Noise (BGNoise).

The data were further analyzed with Agilent Genomic Workbench 5.0 software. In each array spot $log_2$ ratios of Cy3 versus Cy5 were computed with the default *P*-value threshold 0.05 and overlap threshold value 0.9. The CNVs were represented by gains and losses of normalized fluorescence intensities relative to the reference and called by conservative criteria which required alternations of >0.5 $log_2$ ratios over 5 neighboring probes. Homozygous losses were called when signal $log_2$ ratio was <-2.0. Copy number variable regions (CNVRs) were determined by ADM-2 algorithm (Lipson et al. 2006) by combining overlapping CNVs in all samples across the CGH experiments.

Parameters used in this analysis were as follows: Threshold of ADM-2: 6.0; Centralization: ON (Threshold: 6.0, Bin Size: 10); Fuzzy Zero: ON; Aberration Filters: ON (minProbes = 5 & minAvgAbsLogRatio = 0.5 & maxAberrations = 10000 AND

percentPenetrance = 0); Feature Level Filters: ON, IntraArray: ON). Output files were generated with genomic coordinates and cytoband locations for all CNVs. The raw data were submitted to NCBI Gene Expression Omnibus (GEO) accession GSE55266.

Array performance evaluation

To evaluate baseline variations and determine false discovery rate (FDR) (Benjamini and Hochberg 1995; Wang et al. 2005) female and male self-to-self, and female-to-male control hybridizations were conducted using blood DNA from one female and one male Thoroughbred horses. The female Thoroughbred, *Twilight*, was the DNA donor for the horse reference sequence EcuCab2 (Wade et al. 2009) and the origin of the probes on the tiling array. The male Thoroughbred, *Bravo*, a half-sibling to *Twilight*, was the DNA donor for the CHORI-241 BAC library and the origin of all Y chromosome probes on the array. The FDR was calculated as a percentage of the ratio of CNVs in self-to-self hybridization to the total number of CNVs in all experiments. Additionally, array performance was evaluated by self-to-self hybridizations with blood and hair DNA from one Quarter Horse (H528, Table S1). Hybridization quality was assessed by DLRSD which calculates probe-to probe log ratio noise of an array: DLRSD $< 0.2$ was considered excellent; $0.2 \geq \text{DLRSD} \leq 0.3$ was good, and values $> 0.3$ indicated poor quality hybridization.

Array CGH data validation by qualitative and quantitative PCR (qPCR)

Genomic copy number changes as detected by aCGH were validated by quantitative PCR (qPCR) for 18 selected CNVRs using 22 probe-specific primers. Additionally, 8 putative homozygous deletions were validated by regular (qualitative) PCR. Primers (Table S2) were designed inside CNVRs using array probe sequences and the horse whole genome sequence information (EquCab2 at UCSC and Ensembl) (Karolchik et al. 2003; Stalker et al. 2004) and Primer3 software (Koressaar and Remm 2007; Untergasser et al. 2012). Specific criteria for the design of primers for qPCR were: a) GC content over 50%, b) complementarity = 4, c) qPCR product size 100-200 bp, and d) annealing temperature between 55°C and 60°C.

The best DNA concentration for qPCR was 25 ng/µL which was determined by serial dilutions of 200ng/µL, 100 ng/µL, 50 ng/µL, 25 ng/µL, 12.5 ng/µL and 6.25 ng/µL. The qPCR experiments were performed with LightCycler® 480 (Roche Diagnostics) in triplicate assays. Each assay was done in triplicate 20 µL reactions containing 50 ng of template DNA, 10 µM primers, and the SYBR Green PCR kit (Roche).

Relative copy numbers of the selected regions were determined in comparison to the reference sample (Thoroughbred and Quarter Horse) and normalized to an autosomal reference gene *GAPDH*. The cycling conditions were 1 cycle 5 min at 95°C; 45 cycles 10 sec at 95°C, 5 sec at 58°C, and 10 sec at 72°C; 1 cycle for melting curve 30 sec 95°C, 30 sec 65°C and final cooling 20 sec at 50°C.

Quantification of the copy number was carried out using the comparative $C_T$ method ($2^{\Delta\Delta}$Ct) (Livak and Schmittgen 2001; Bodin et al. 2005) with $p<0.05$ as a cut-off threshold for statistical significance. Qualitative PCR results were analyzed by agarose gel electrophoresis.

Array CGH data validation by fluorescence in situ hybridization (FISH)

CNV specific primers were used to screen CHORI-241 BAC library (CHORI-241) by PCR (Table S2). The BAC DNA was isolated by Plasmid Midiprep kit (Qiagen), labeled with biotin-16-dUTP or digoxigenin-11-dUTP using Biotin- or DIG-Nick Translation Mix (Roche), and hybridized to metaphase chromosomes of CNV carriers and control horses following standard protocols (Raudsepp and Chowdhary 2008a). A BAC clone representing a non-CNV region was used as a control in each FISH experiment. Images for a minimum of 20 metaphase and/or interphase cells were captured for each experiment and analyzed with a Zeiss Axioplan2 fluorescent microscope equipped with Isis v5.2 (MetaSystems GmbH) software.

Gene ontology enrichment analysis

Ensembl gene list (Ensembl Genebuild 73.2) along with their position in the horse genome was added to Agilent Genomic Workbench as a custom track to determine the genic and intergenic CNVs. Gene Ontology analysis (GO) and Kyoto Encyclopedia

of Genes and Genomes (KEGG) pathway analysis of the genes present in CNVs were performed using DAVID (The Database for Annotation, Visualization and Integrated Discovery) bioinformatics tool with default settings (Huang da et al. 2009b; Huang da et al. 2009a). TheGO annotations used were: BP (Biological Process), MF (Molecular Function) and CC (Cellular Component).

Because only a limited number of genes in the horse genome have been annotated, horse gene IDs were converted to orthologous human Ensembl gene IDs by BioMart, followed by GO and pathway analyses, as described above. Biological functions of the genes in CNVRs were further analyzed manually by data mining in Ensembl (Stalker et al. 2004), UCSC (Karolchik et al. 2003) and NCBI (Sayers et al. 2010) Genome Browsers searching for data for equine orthologs in other mammalian species. CNVs present in intergenic regions were analyzed in UCSC genome browser and NCBI and GeneCards (MacDonald et al. 2014) for similarities to known mammalian genes.

Chromosome CNVR enrichment

Horse chromosome enrichment percentage was determined by the total length of CNVRs present each chromosome, divided by length of the chromosome. The genome Enrichment was calculated by the division of the total length of all CNVRs with the length of the assembled horse genome. Enriched chromosomes were acknowledged if

their enrichment percentage was more than the enrichment percentages of the whole horse genome.

Generation of a composite CNV dataset for the horse

Genomic positions of CNVs/CNVRs from this and all previously published studies (Doan et al. 2012a; Doan et al. 2012b; Dupuis et al. 2013; Metzger et al. 2013a; Wang et al. 2014b) were aligned, and partially or completely overlapping and adjacent CNVs (the end position of a previous CNV and the start position of the next CNV are the same) were consolidated into one CNVR.

Phylogenetic analysis

Genotypes for 15 microsatellite loci (Khanshour et al. 2013); E.G. Cothran, unpublished) were available for 32 out of 41 horse breeds involved in CNV studies. Majority-rule consensus of Restricted Maximum Likelihood (RML) trees were constructed and visualized as described elsewhere (Khanshour et al. 2013). The Przewalski horse population was used as an out-group.

## Results

Performance of the Texas-Adelaide horse WG tiling array

Self-to-self control hybridizations (Figure 4) showed 1.55% of FDR (4/258*100) - an indication that array design, fabrication, and aCGH procedures were optimal. As a proof-of principle, female-to-male hybridizations between two half-sib Thoroughbreds, *Twilight* (female) and *Bravo* (male), showed massive loss in the X chromosome and a gain in the Y chromosome in the male (Figure 5), whereas only one CNV was detected in an autosome, chr3. The DLRSD values for all hybridizations with blood DNA from *Twilight* and *Bravo* were < 0.2. Therefore, and because the oligonucleotides on the array were derived from the sequences of these two horses, DNA of *Twilight* and *Bravo* was used as a reference for all aCGH experiments: *Twilight* for females and *Bravo* for males.

Further, because the DNA collection from horse breeds contained samples isolated from blood and hair, an additional self-to-self hybridization was conducted using DNA from blood and hair of one male American Quarter Horse QH3-H528 (Table 6). Hybridization quality was evaluated by measuring Derivative LogRatio Standard Deviation (DLRSD) that assesses the log ratio variances between successive probes divided by the square of 2, and calculates probe to probe noise across the array. Blood DNA gave good quality results with DLRSD = 0.14, whereas consistent and high level hybridization noise was observed for hair DNA (DLRSD = 0.41) (Figure 6). Due to this, CNVs in all experiments were called with stringent criteria: $\log_2$ ratio alternations higher

than 0.5 over 5 neighboring probes. With median probe spacing of 7.5 kb on the array, this allowed detection CNVs of about 30 kb, and in probe-dense regions even smaller. We concluded that the performance of the equine 400K Texas-Adelaide whole-genome CGH array was optimal for the discovery of CNVs in the horse genome.

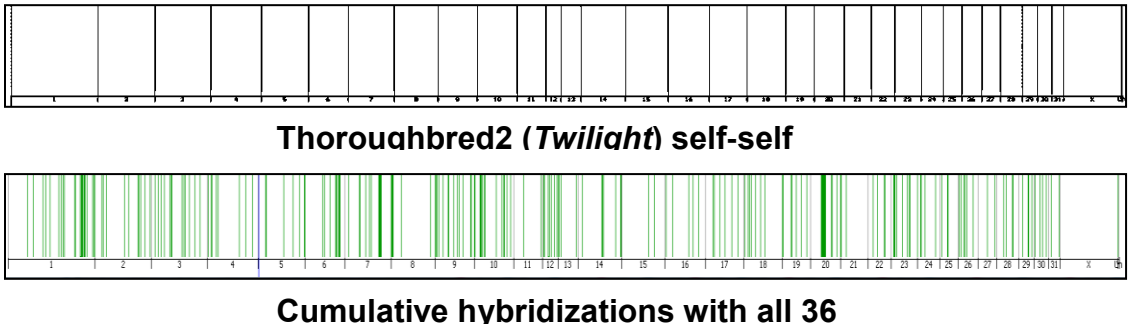


**Figure 4: Genome-wide distribution of CNVs in self-to-self hybridization.** (upper) compared to cumulative hybridizations with all animals (lower) to determine FDR; green vertical lines denote CNVs

CNV discovery and construction of a whole-genome CNV map for the horse

Altogether, 950 CNV calls were made across 36 animals (Appendix 2.1) with an average of 26.4 calls (19.1 losses and 7.3 gains) per individual (Table 2.2). The number of CNV calls was the highest in the two American Miniature Horses (59 and 46) and the lowest in American Quarter Horses (12 and 14) and a Caspian Pony (12) (Table 7; Appendix 2.1). Because the Thoroughbred served as a reference, no calls were made in

Thoroughbreds. The number and distribution of CNVRs in the two Przewalski horses were similar to those in domestic horses (Table 7; Appendix 2.1).

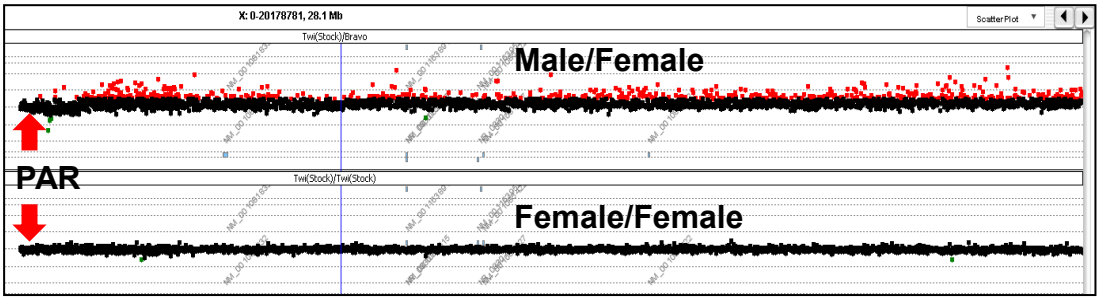


**Figure 5: Male-to-female aCGH for the X chromosome.**
Massive loss (red dots) was seen in the X chromosome of the male (upper), whereas no gains and losses were detected in the pseudoautosomal region (PAR) between the male and female genomes.

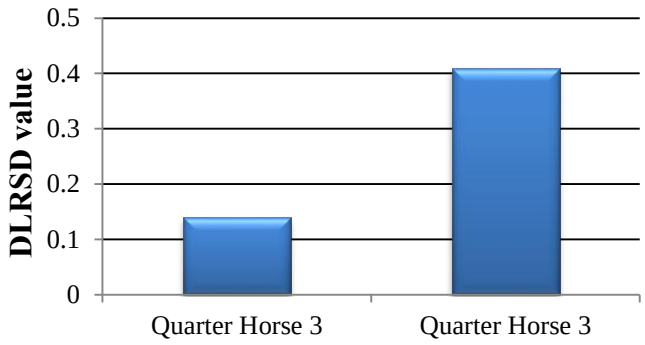


**Figure 6: Comparative DLRSD values.**
aCGH using DNA from blood (left) and from hair (right) of the same individual.

**Table 7: Breed- and individual-wise summary of CNVRs in horses.**

| Horse breed | CNVRs per individual | Gains | Losses |
|---|---|---|---|
| Akhal-teke 1 | 37 | 12 | 25 |
| Akhal-teke 2 | 26 | 13 | 13 |
| American Miniature Horse 1 | 59 | 16 | 43 |
| American Miniature Horse 2 | 46 | 4 | 42 |
| American Quarter Horse 1 | 12 | 0 | 12 |
| American Quarter Horse 2 | 21 | 2 | 19 |
| American Quarter Horse 3 | 14 | 10 | 4 |
| Arabian 1 | 21 | 17 | 4 |
| Arabian 2 | 17 | 0 | 17 |
| Belgian 1 | 31 | 14 | 17 |
| Belgian 2 | 14 | 1 | 13 |
| Caspian Pony 1 | 40 | 16 | 24 |
| Caspian Pony 2 | 12 | 1 | 11 |
| Clydesdale 1 | 25 | 6 | 19 |
| Clydesdale 2 | 16 | 1 | 15 |
| Exmoor Pony 1 | 29 | 15 | 14 |
| Exmoor Pony 2 | 18 | 8 | 10 |
| Fell Pony 1 | 25 | 11 | 14 |
| Fell Pony 2 | 47 | 11 | 36 |
| Friesian 1 | 29 | 6 | 23 |
| Friesian 2 | 39 | 10 | 29 |
| Friesian 3 | 41 | 9 | 32 |
| Friesian 4 | 22 | 12 | 10 |
| Mongolian Native Horse1 | 22 | 1 | 21 |
| Mongolian Native Horse2 | 18 | 2 | 16 |
| Percheron 1 | 17 | 11 | 6 |
| Percheron 2 | 12 | 1 | 11 |
| Przewalski's Horse 1 | 21 | 5 | 16 |
| Przewalski's Horse 2 | 21 | 3 | 18 |
| Sorraia 1 | 36 | 8 | 28 |
| Sorraia 2 | 18 | 1 | 17 |
| Standardbred 1 | 17 | 7 | 10 |
| Standardbred 2 | 44 | 13 | 31 |
| Swiss Warmblood 1 | 23 | 1 | 22 |
| Swiss Warmblood 2 | 30 | 6 | 24 |
| Swiss Warmblood 3 | 29 | 9 | 20 |
| **Average** | **26.4** | **7.3** | **19.1** |

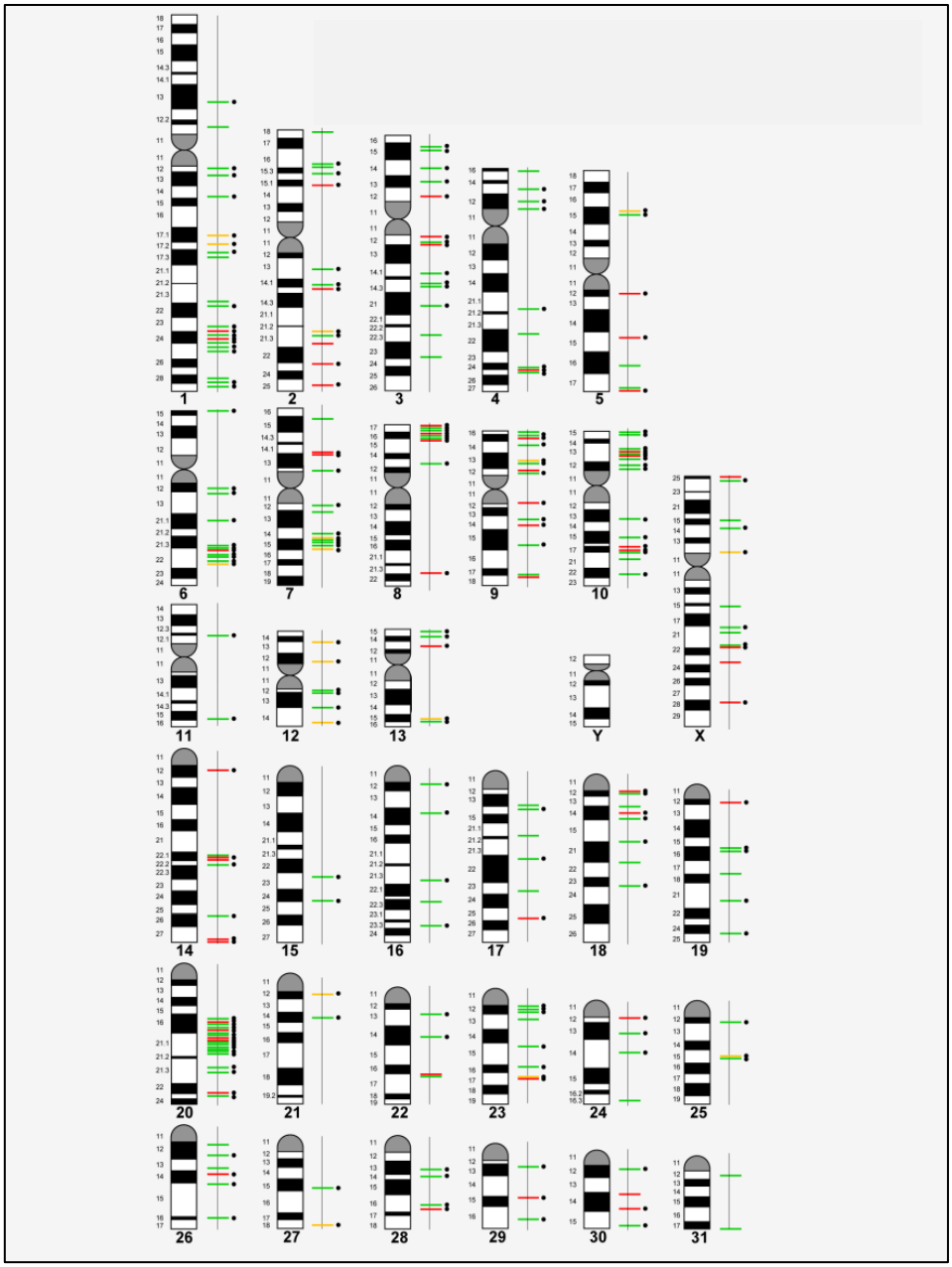

**Figure 7: A CNVR map of the horse genome.**
Green line – loss; red line – gain; yellow line – complex; black dots – genes involved.

**Table 8: Chromosome-wise CNVR statistics for the horse genome.**
Shared – found in 2 or more individuals; private – in one horse only; novel – not reported before; horse genome statistics retrieved from Ensembl.

| Chr | #CNVR | Shared | Private | Novel | Gains | Losses | Complex | Genic | Intergenic | Sub-telomeric | Mean CNVR size (bp) | CNVR length (bp) | Chr. size (bp) | Enrich-ment, % | Gene/Mb |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 21 | 8 | 13 | 9 | 2 | 17 | 2 | 14 | 7 | 0 | 241,828 | 5,078,379 | 185,838,109 | 2.73 | 6.63 |
| 2 | 13 | 3 | 10 | 5 | 5 | 7 | 1 | 9 | 4 | 1 | 69,711 | 906,246 | 120,857,687 | 0.75 | 6.64 |
| 3 | 14 | 3 | 11 | 3 | 3 | 11 | 0 | 11 | 3 | 0 | 80,328 | 1,124,591 | 119,479,920 | 0.94 | 5.19 |
| 4 | 9 | 4 | 5 | 3 | 1 | 8 | 0 | 7 | 2 | 0 | 119,780 | 1,078,019 | 108,569,075 | 0.99 | 4.94 |
| 5 | 7 | 5 | 2 | 2 | 3 | 3 | 1 | 4 | 3 | 2 | 60,379 | 422,653 | 99,680,356 | 0.42 | 8.14 |
| 6 | 11 | 5 | 6 | 3 | 1 | 9 | 1 | 9 | 2 | 0 | 206,650 | 2,273,151 | 84,719,076 | 2.68 | 9.08 |
| 7 | 12 | 4 | 8 | 3 | 2 | 8 | 2 | 8 | 4 | 0 | 119,155 | 1,429,861 | 98,542,428 | 1.45 | 9.36 |
| 8 | 9 | 5 | 4 | 1 | 4 | 5 | 0 | 8 | 1 | 1 | 170,025 | 1,530,227 | 94,057,673 | 1.63 | 5.66 |
| 9 | 14 | 6 | 8 | 11 | 5 | 8 | 1 | 4 | 10 | 1 | 45,785 | 640,986 | 83,561,422 | 0.77 | 5.66 |
| 10 | 16 | 7 | 9 | 7 | 4 | 12 | 0 | 13 | 3 | 1 | 72,954 | 1,167,268 | 83,980,604 | 1.39 | 8.16 |
| 11 | 2 | 0 | 2 | 1 | 0 | 2 | 0 | 2 | 0 | 0 | 5,235 | 10,470 | 61,308,211 | 0.02 | 14.71 |
| 12 | 6 | 5 | 1 | 1 | 0 | 3 | 3 | 5 | 1 | 0 | 538,645 | 3,231,871 | 33,091,231 | 9.77 | 14.97 |
| 13 | 5 | 2 | 3 | 1 | 1 | 3 | 1 | 5 | 0 | 0 | 23,837 | 119,184 | 42,578,167 | 0.28 | 11.69 |
| 14 | 8 | 3 | 5 | 3 | 5 | 3 | 0 | 5 | 3 | 1 | 55,523 | 444,184 | 93,904,894 | 0.47 | 5.41 |
| 15 | 2 | 0 | 2 | 2 | 0 | 2 | 0 | 2 | 0 | 0 | 80,429 | 160,857 | 91,571,448 | 0.18 | 5.32 |
| 16 | 5 | 3 | 2 | 4 | 0 | 5 | 0 | 4 | 1 | 0 | 46,120 | 230,599 | 87,365,405 | 0.26 | 5.95 |
| 17 | 6 | 2 | 4 | 3 | 1 | 5 | 0 | 3 | 3 | 0 | 100,527 | 603,159 | 80,757,907 | 0.75 | 3.1 |
| 18 | 8 | 4 | 4 | 4 | 2 | 6 | 0 | 5 | 3 | 0 | 59,831 | 478,649 | 82,527,541 | 0.58 | 3.66 |
| 19 | 6 | 1 | 5 | 5 | 1 | 5 | 0 | 2 | 4 | 0 | 69,818 | 418,908 | 59,975,221 | 0.70 | 5.23 |
| 20 | 19 | 11 | 8 | 4 | 5 | 14 | 0 | 13 | 6 | 0 | 102,575 | 1,948,920 | 64,166,202 | 3.04 | 8.85 |
| 21 | 2 | 1 | 1 | 1 | 0 | 1 | 1 | 2 | 0 | 1 | 230,842 | 461,684 | 57,723,302 | 0.80 | 4.87 |
| 22 | 4 | 2 | 2 | 1 | 1 | 3 | 0 | 2 | 2 | 0 | 48,296 | 193,182 | 49,946,797 | 0.39 | 8.08 |
| 23 | 8 | 2 | 6 | 4 | 1 | 6 | 1 | 5 | 3 | 0 | 122,812 | 982,492 | 55,726,280 | 1.76 | 3.91 |
| 24 | 4 | 1 | 3 | 2 | 1 | 3 | 0 | 3 | 1 | 0 | 104,966 | 419,862 | 46,749,900 | 0.90 | 6.19 |
| 25 | 3 | 2 | 1 | 1 | 0 | 2 | 1 | 2 | 1 | 0 | 46,573 | 139,720 | 39,536,964 | 0.35 | 10.56 |
| 26 | 6 | 4 | 2 | 3 | 1 | 5 | 0 | 2 | 4 | 0 | 105,913 | 635,479 | 41,866,177 | 1.52 | 3.68 |
| 27 | 2 | 1 | 1 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 15,084 | 30,168 | 39,960,074 | 0.08 | 3.91 |
| 28 | 4 | 2 | 2 | 2 | 1 | 3 | 0 | 2 | 2 | 0 | 37,557 | 150,227 | 46,177,339 | 0.33 | 6.42 |
| 29 | 3 | 2 | 1 | 2 | 1 | 2 | 0 | 3 | 0 | 0 | 250,105 | 750,316 | 33,672,925 | 2.23 | 4.17 |

**Table 8 continued.**

| Chr | #CNVR | Shared | Private | Novel | Gains | Losses | Complex | Genic | Intergenic | Sub-telomeric | Mean CNVR size (bp) | CNVR length (bp) | Chr. size (bp) | Enrich-ment, % | Gene/Mb |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 30 | 4 | 1 | 3 | 2 | 2 | 2 | 0 | 2 | 2 | 0 | 67,956 | 271,825 | 30,062,385 | 0.90 | 4.23 |
| 31 | 2 | 1 | 1 | 2 | 0 | 2 | 0 | 0 | 2 | 1 | 34,286 | 68,572 | 24,984,650 | 0.27 | 4.58 |
| X | 12 | 5 | 7 | 10 | 4 | 7 | 1 | 0 | 12 | 1 | 87,146 | 1,045,753 | 124,114,077 | 0.84 | 4.46 |
| Un | 11 | 9 | 2 | 11 | 6 | 0 | 5 | 0 | 11 | 0 | 4,118 | 45,298 | 117,461,955 | 0.04 | - |
| Y | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00 | 0.57 |
| **Total** | **258** | **114** | **144** | **116** | **63** | **173** | **22** | **158** | **100** | **10** | **110,437** | **28,492,760** | **2,484,515,402** | **1.15** | - |

The ADAM-2 algorithm arranged individual CNV calls into 258 CNV regions (CNVRs; Appendix 2.2) of which 114 were shared between at least 2 individuals of the same or different breeds, while 144 were private and found only in one individual Table 8). Two CNVRs were found in two or more individuals of the same breed but not in other breeds and were tentatively considered as breed-specific: a 14 kb loss in chr9 in Exmoor ponies and a 39 kb loss in chr20 in Swiss Warmblood horses (Table 9).

Based on the 258 CNVRs, a whole genome CNV map for the horse was constructed (Figure 7) details of which are summarized in Table 8. The mean size of CNVRs was 110 kb ranging from 1 kb to 2.5 Mb. The CNVRs occupied 1.15 % of equine genome and were distributed over all horse chromosomes, except the Y, with the highest enrichment in chromosomes 12 (9.7%) and 20 (3.0 %). Even although chr12 is the gene richest chromosome in the horse genome (15 genes/ Mb), there was no overall correlation between CNV enrichment and gene density. For example, the enrichment values for the second and third gene densest chromosomes, chr11 and chr13, were 0.02% and 0.28%, respectively (Table 8). Likewise, we did not observe CNV enrichment in sub-telomeres, as previously reported for humans (Riethman 2008): the array contained 5,716 sub-telomeric probes, although only 10 CNVRs were detected in these regions in horses.

**Table 9: Tentative breed-specific CNVRs**

| Breed and individual | Chr | Start | Stop | Size | Gain log2 | Loss log2 | Previous report | Breed | Genes |
|---|---|---|---|---|---|---|---|---|---|
| Exmoor Pony 1 | 9 | 833,013 | 847,731 | 14,718 | 0 | -2.346321 | none | n/a | intergenic |
| Exmoor Pony 2 | 9 | 833,013 | 847,731 | 14,718 | 0 | -0.732666 | none | n/a | intergenic |
| Swiss Warmblood 2 | 20 | 47,378,055 | 47,417,700 | 39,645 | 0 | -0.609726 | none | n/a | mRNA JL626884 |
| Swiss Warmblood 3 | 20 | 47,378,055 | 47,417,700 | 39,645 | 0 | -0.64524 | none | n/a | mRNA JL626884 |

In general, losses (173; 67%) prevailed over gains (63; 24%), although 6 horses had more gains than losses (Table 7). Twenty-two CNVRs (8.5%) were complex involving both losses and gains in different individuals (Table 8). Even although aCGH on diploid samples cannot discriminate between copies of alleles and thus, distinguish between heterozygous and homozygous CNVs, two gains and 14 losses were tentatively considered homozygous because of log2 alterations over 2.0 (Appendix 2.3). Homozygosity of 8 losses was confirmed by qualitative PCR (Figure 8).

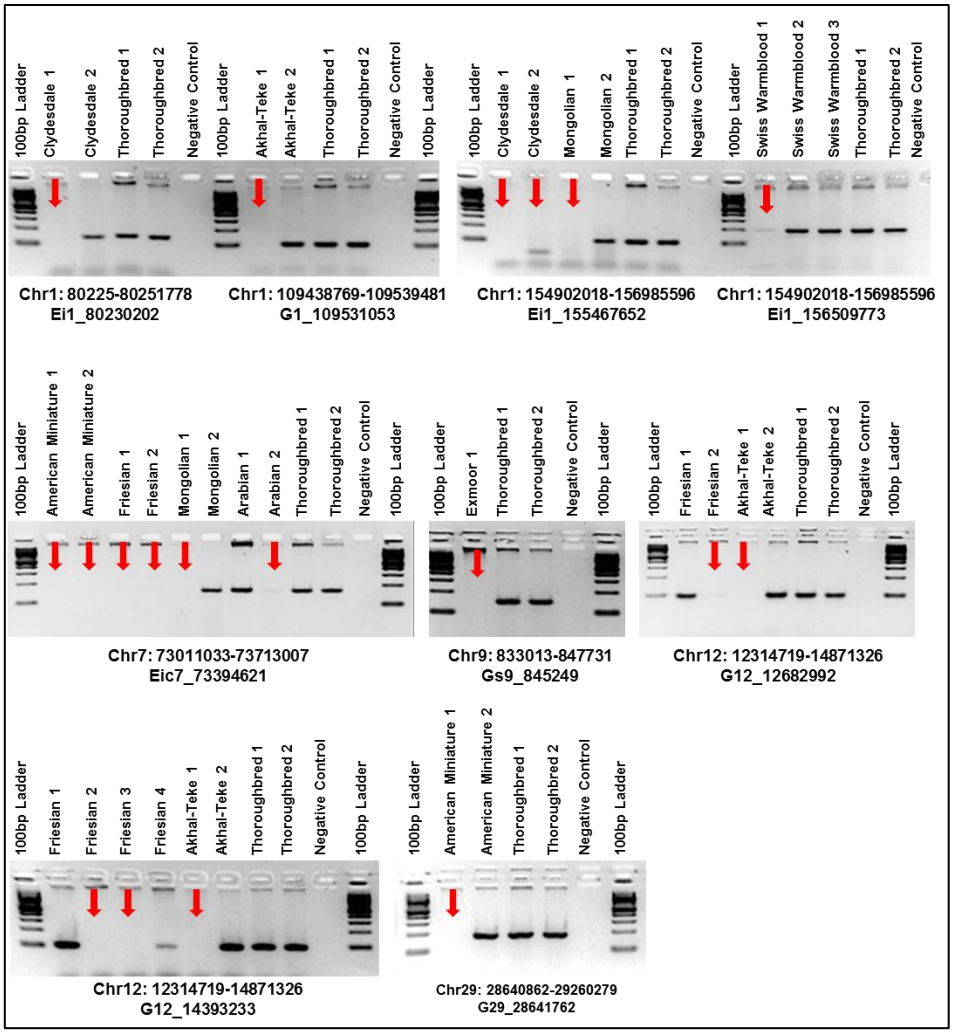


**Figure 8: Confirmation of homozygous deletion CNVs by qualitative PCR.**
Red arrows indicates deleted CNVs.

Gene content of CNVRs and functional categories of copy number variable genes

The majority (82%) of horse CNVRs contained one or more known Ensembl horse genes (158 CNVRs) or non-horse mammalian reference genes (54 CNVRs) (Appendix 2.4), while 46 CNVRs (18%) were located in intergenic regions (Appendix

2.5). Gene containing CNVRs were also predominant in individual chromosomes with the exception of chr31 which was enriched with intergenic variants (Figure 9). However, the calls for intergenic CNVRs are tentative and subject to change as the annotation of the horse genome is still in progress.

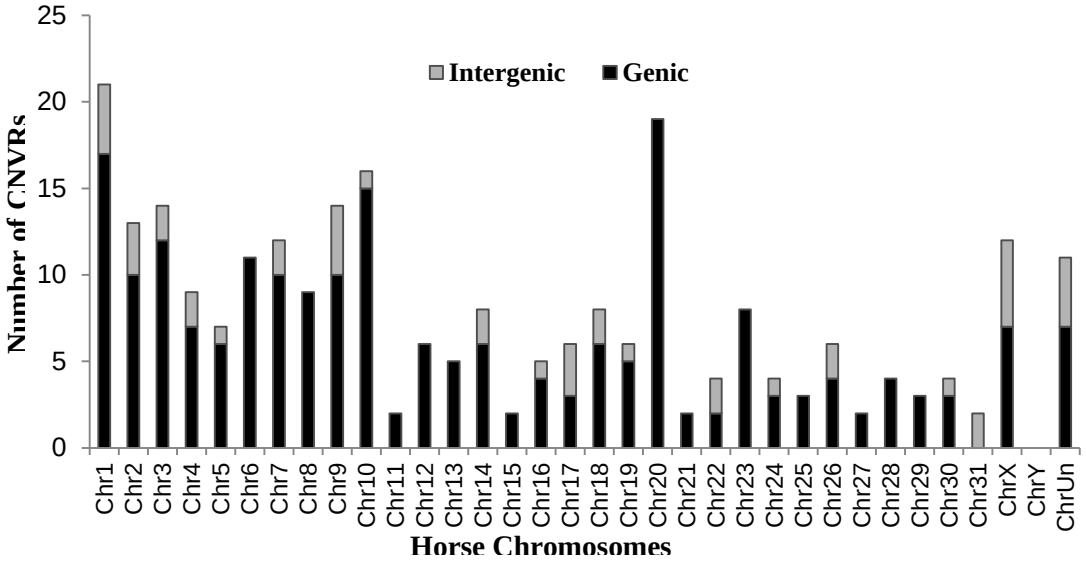


**Figure 9: Chromosome-wise distribution of CNVRs in the horse genome.**

Altogether, the CNVRs involved 805 protein-coding genes (750 Ensembl genes, 33 non-Ensembl genes and 22 horse mRNAs; (Appendix 2.4) but also non-coding small and long RNA genes, and pseudogenes. The largest CNVRs with the highest number of genes corresponded to clusters of olfactory and non-olfactory G-protein coupled

receptors (GPCRs) or to immunity related genes, such as immunoglobulins, T-cell receptors, and MHC protein complex genes - a typical feature of CNVRs in all mammalian genomes studied so far (Redon et al. 2006; Berglund et al. 2012; Bickhart et al. 2012; Doan et al. 2012a; Hou et al. 2012a; Li et al. 2012). Likewise, Gene Ontology (GO) analysis indicated that equine copy number variable genes are predominantly involved in biological processes and molecular functions related to transmembrane signal transduction, chemo-attractant sensory perception, immune response and steroid metabolism (Figure 10; Appendix 2.6). In addition, only one CNVR was involved in the following cellular components and molecular processes: the structural constituent of cytoskeleton; keratin filament; testosterone 17-beta-dehydrogenase activity, and aldo-keto reductase activity of secondary sexual development (Appendix 2.6).

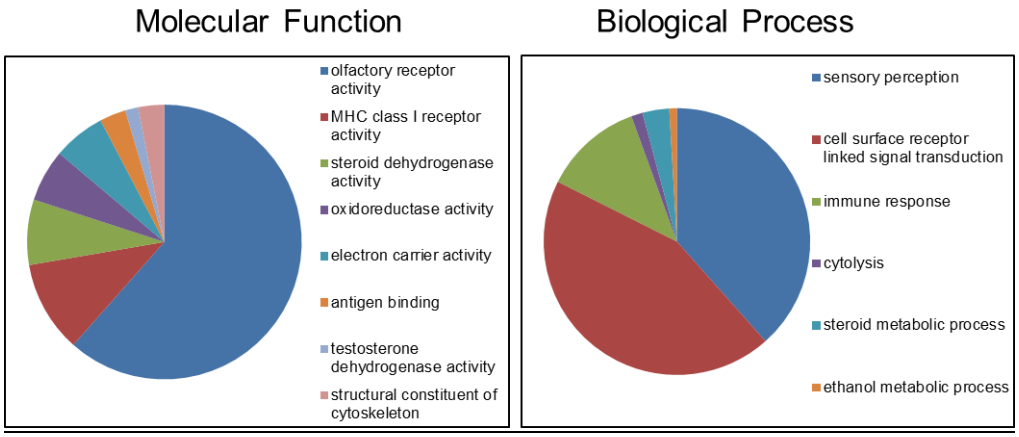


**Figure10: GO classifications of copy number variable genes in horses.**
GO= Gene Ontology

A composite CNV dataset for the horse genome

Comprehensive knowledge of CNVs in normal horse populations, within and across breeds, is a prerequisite for the discovery of variants that contribute to equine genetic diseases and disorders. Therefore, we aligned the 258 CNVRs identified in this study with previously published CNV data for the horse, *viz.*, (Doan et al. 2012a; Doan et al. 2012b; Dupuis et al. 2013; Metzger et al. 2013a; Wang et al. 2014b). Altogether, we found records of about 2041 CNVs and CNVRs (calling criteria vary between studies). These were further consolidated, based on adjacent locations or partial overlaps, into 1476 CNVRs of which 301 CNVRs (20%) were shared between two or more studies (Appendix 2.7, Figure 11). The majority of common CNVRs involved genes associated with olfactory reception (50 CNVRs) and membrane transport (49 CNVRs) but also genes involved in transcription (30 CNVRs), cell cycle regulation (12 CNVRs) and RNA genes (34 CNVRs). Expectedly, CNVRs that were found in more than 100 horses and reported by all 6 studies exclusively involved olfactory receptors. Most of the shared CNVs between studies are autosomal. In contrast, X chromosome CNVs have been reported just in two studies: 21 CNVs by Doan et al. 2012a and 12 in this study. Of these, 2 are shared – one involving a long non-coding RNA gene and another, an ATPase gene (Appendix 2.4 and 2.7).

Comparative analysis also revealed that novel CNVRs predominated over shared ones in all 6 studies (Figure 11). Novel CNVRs of functional interest from this study involved genes related to sperm-egg interaction and fertilization in chr4:19.8-19.9 Mb; a

long non-coding RNA *SOX2-OT* in chr19:20.1 Mb which might regulate the *SOX2* gene and is possibly involved in developmental processes (Amaral et al. 2009); an X-linked region harboring genes of circadian pacemaker function in chrX:83.8-84.0 Mb, and a complex CNVR in chrUn:225-226 kb with cancer related genes. Notably, the latter two CNVRs were found in more than 10 horses each. Details of all novel and shared CNVRs are presented in Appendix 2.7.

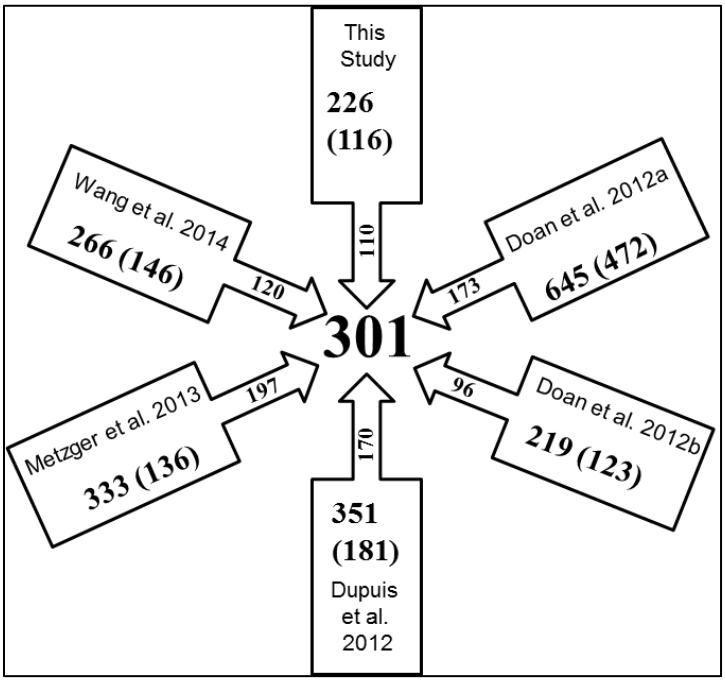


**Figure 11: A summary diagram for the CNVs identified in the horse genome.** Total CNVs=1476. Numbers in arrow-heads denote the contribution of each study to the common pool of 301 shared CNVs/CNVRs; numbers in arrow-tails denote the total and novel (in parentheses) CNVs/CNVRs per study**.**

Comparative analysis of horse CNVRs

Attempts were made to align horse CNVR sequences with the human genome in UCSC table browser to identify evolutionarily conserved CNVRs shared by the two species. In many cases a CNVR in the horse retrieved multiple sequence coordinates in the human genome and confounded the analysis. However, some success accompanied the alignment of horse gene-containing CNVRs. The corresponding human sequences were further analyzed for CNVs in the human Database of Genomic Variants (MacDonald et al. 2014). Altogether, after excluding CNVRs containing clusters of olfactory receptor and immunity related genes, we identified 80 horse genic CNVRs with a homeologous CNVR in the human genome (Appendix 2.8).

Further, genes involved in CNVRs were compared with Online Mendelian Inheritance in Man (OMIM) and Online Mendelian Inheritance in Animals **(**OMIA) databases to see whether any of the equine copy number variable genes are involved in known human or animal phenotypes. Notably, 108 CNV-genes from this study were present in OMIM database, and 5 genes in OMIA database. The latter were associated with known OMIA phenotypes for immune, reproductive, or neuromuscular diseases (Table 10).

**Table 10: Equine copy number variable genes with known mammalian phenotypes.**

| Gene symbol | Biological system and phenotype | OMIA ID | Ensembl ID | Location Chr:Mb | CNVR size, kb | CNVR Type | Discovery breeds | Reference |
|---|---|---|---|---|---|---|---|---|
| *BMPR1B* | *REPRODUCTIVE*: Fecundity in Booroola and Bonpala sheep | 676; 3210 | ENSECAG 00000012140 | **3:**43.57-43.60 | 28 | Loss | Friesian, Quarter Horse, Standardbred, Swiss Warmblood | This study, (Doan et al. 2012a) |
| *BTN1A1* | *IMMUNE*: Resistance to avian sarcoma and leucosis viruses in chicken | 3167 | ENSECAG 00000017948 | **20:**24.22-24.62 | 405 | Gain | American Miniature, Arabian, Belgian, Caspian, Clydesdale, Fell Pony, Friesian, Standardbred, Sorraia, Swiss Warmblood | This study, (Doan et al. 2012a; Dupuis et al. 2013; Wang et al. 2014b) |
| *CFH* | *IMMUNE*: Thrombocytopenia in cattle and dogs | 2691; 1729 | ENSECAG 00000011534 | **30:**24.74-24.87 | 132 | Loss | Sorraia, Clydesdale, Fell Pony, Friesian, Standardbred, Swiss Warmblood | Novel |
| *GLB1* | *NEUROMUSCULAR*: Gangliosidosis in sheep, cattle, dogs and cats with progressive neuromuscular dysfunctions | 700-703 | ENSECAG 00000011942 | **16:**51.36-51.37 | 12 | Loss | Fell Pony | Novel |
| *KRT1* | *IMMUNE*: Epidermolytic hyperkeratosis in dogs | 2618; 2787 | ENSECAG 00000022233 | **6:**69.75-69.77 | 21 | Loss | Akhal-Teke, Belgian, Fell Pony, Friesian, Mongolian, Standardbred, Swiss Warmblood | This study, (Doan et al. 2012a; Metzger et al. 2013a) |

Validation of CNVRs by qPCR

Nineteen CNVRs were validated by quantitative PCR (qPCR) using array probe-specific primers (Appendix 2.9). The regions were selected upon three criteria – size, gene content, and novelty. The tested CNVRs covered a size range from 4 kb to 2 Mb; 13 involved clusters of horse genes, and 6 were novel. Summary of qPCR results are presented in Figures 12 and Appendix 2.10. Overall, qPCR agreed well ($P$-value < 0.05) with the array CGH data. For example, it confirmed a complex CNVR in chr27 involving *CSMD1* gene (CUB and Sushi multiple domains 1) which encodes a transmembrane and a candidate tumor suppressor protein (Shull et al. 2013). Copy numbers in this region were tested on 11 breeds with at least 2 individuals each and showed a gain in native ponies, draft breeds, and the Przewalski horse, and a loss in American Miniature horses in relation to the Thoroughbred (Figure 12 F). Likewise, qPCR confirmed a CNVR in chr20 (Figure 12 C) which has been found only in this study and in indigenous plateau horses (Wang et al. 2014b).

We also found some differences: instead of a loss, qPCR showed a significant gain in chr20:32.0-32.4 Mb and chr17:18.8-19.0 Mb in Swiss Warmblood and Mongolian horses, respectively. Also, a loss in chr7:74.8-74.9 Mb in Swiss Warmblood horses was not significant by qPCR. These minor discrepancies can be attributed to intra-breed variation: array CGH was based on 2 to 4 individuals, while qPCR involved 4 or more horses per breed (Figure 12; Appendix 2.10).

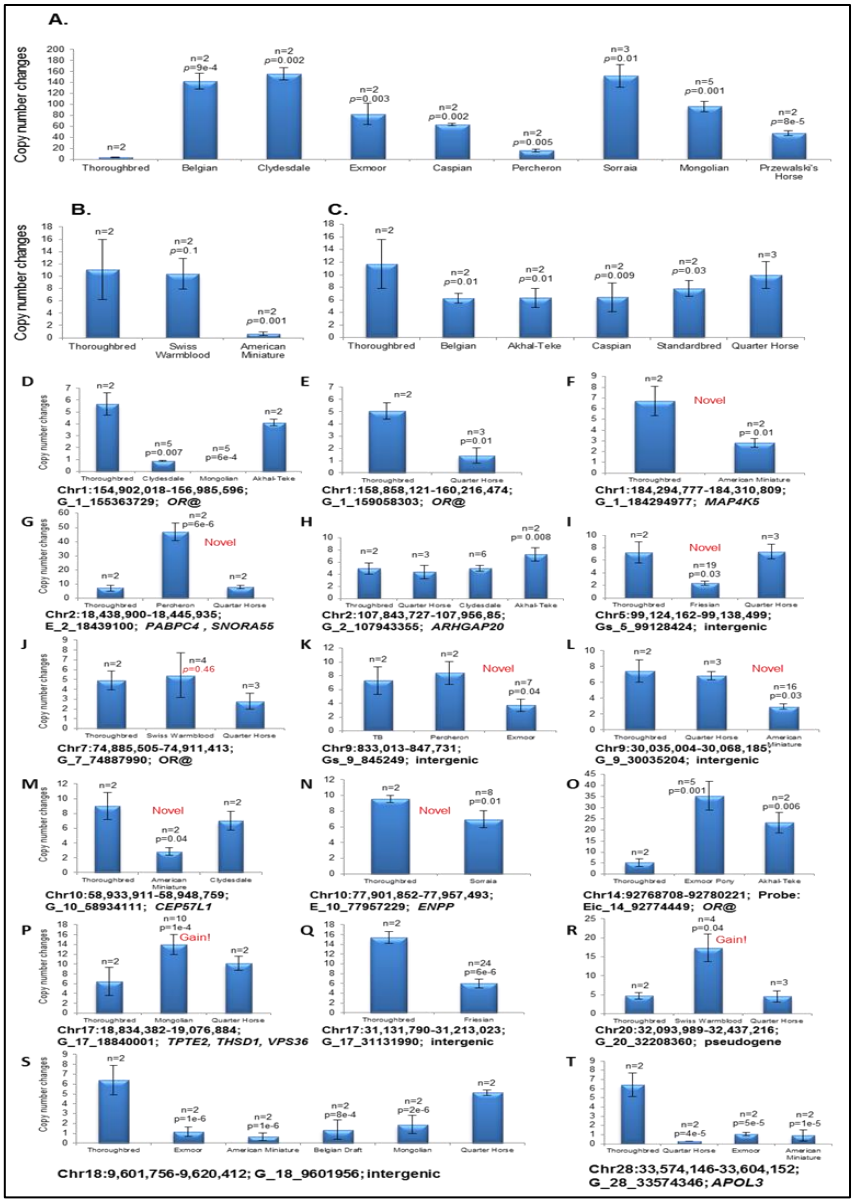


**Figure 12: Validation of CNVRs by qPCR.**
A. gains and B. losses in Chr27 (37.3 Mb; probe Gs_27_37371896) involving CUB and Sushi multiple domains 1 (CSMD1) gene; C. Loss in Chr20 (24.8 Mb; probe Eic_20_24841849) involving olfactory receptors; n – number of individuals analyzed. D-T. Validation of other selected CNVRs.

Validation of CNVRs by FISH

As cell suspensions were not available for most of the horses studied by array CGH, FISH validation was carried out for just a few CNVRs using CNV-containing CHORI-241 BAC clones. These included BAC 132B13 for a complex 200 kb gain-loss region in chr1:114.0-114.2 Mb (Figure 13), and BAC 91B23 for a 2.2 kb gain in chrUn: 529-531 kb (Figure 14). Clear differences in copy numbers between individual horses, as well as between homologous chromosomes of the same horse were observed. Additionally, the CNVR in chrUn was mapped to horse chr19q12-q13 (Figure 14).

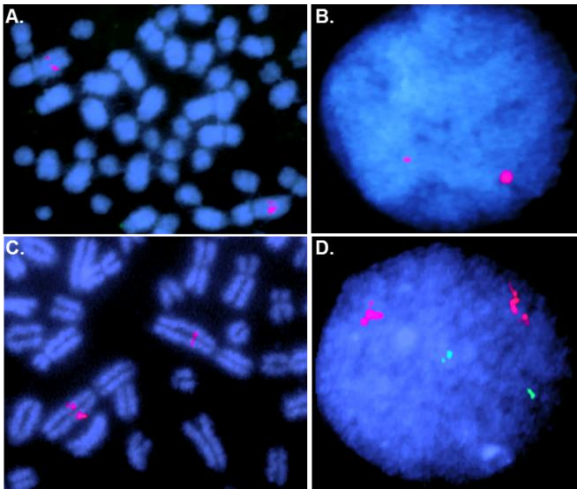


**Figure 13: Validation of a copy number gain in chr1 (114.0 Mb) by FISH.**
A. and B. – metaphase and interphase of the Thoroughbred control; C. and D. metaphase and interphase of a Quarter Horse; red signals - BAC 132B13; green signals in D. – a single-copy control BAC. Note the difference in copy numbers between homologous chromosomes in both horses**.**

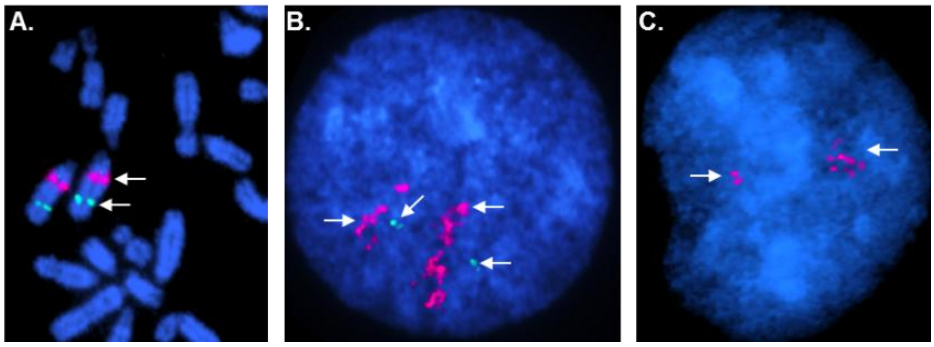


**Figure 14: Chromosomal assignment and validation of a CNVR.**
CNV present in ChrUn(529-531 kb). **A.** Mapping the CNVR to chr19q12-q13 by FISH with BAC 132B13 (red); green - a control BAC with UMPS gene in chr19q21 (Raudsepp and Chowdhary 2008a); **B.** The CNVR (red) in interphase chromosomes of a Percheron; green – a single-copy control probe; C. The CNVR (red) in interphase chromosomes of a Thoroughbred (*Twilight*). Note the difference in copy numbers between the Percheron and the Thoroughbred, as well as between homologous chromosomes.

## Discussion

Copy number variation is an important and characteristic feature of mammalian genomes. It is part of normal genetic variation contributing to phenotypic diversity and adaptations (Redon et al. 2006; McCarroll et al. 2008b; Zhang et al. 2009), but can also be pathogenic and associated with diseases and disorders. For example in humans, CNVs have been associated with autoimmune disorders (Fanciulli et al. 2010; Olsson and Holmdahl 2012; Robinson et al. 2012), psychoneurological conditions such as schizophrenia (Vassos et al. 2010; Crespi and Crofts 2012; Kirov et al. 2014; Tansey et al. 2014) and autism (Krumm et al. 2013; Poultney et al. 2013; Rangasamy et al. 2013;

Walker and Scherer 2013), as well as infectious (Fanciulli et al. 2010) and cardiovascular diseases (Wineinger et al. 2011; Costelloe et al. 2012; Duschek et al. 2013). In addition, CNVs pose interest for pharmacogenomics – a field of genomics studying the molecular differences in drug metabolism, drug sensitivity, toxicity and efficacy between individuals (He et al. 2011). Furthermore, some CNVs can have both adaptive value and pathogenic effect, as recently shown for the CNV in the human salivary amylase gene (*AMY1*) which is associated with diet processing, as well as with obesity phenotypes (Falchi et al. 2014). In order to understand the functions of CNVs and distinguish between normal and pathogenic variants, detailed knowledge about CNVs in the species of interest is needed.

Here we report about the construction of a 400K high-density WG tiling oligoarray for the horse and its application for the discovery of CNVs in 38 normal horses of 16 diverse breeds. Probes on the array were designed to detect CNVs in 18,763 equine autosomal and X-linked genes but also in intergenic, sub-telomeric and Y chromosome sequences. Compared to previous CNV studies in the horse, our CNV discovery platform most closely resembled the WG 1.3 M NimbleGen CGH array by Wang and colleagues (Wang et al. 2014b), but essentially complemented the exon CGH array by Doan and colleagues (Doan et al. 2012a) and the studies based on WG SNP50 BeadChip (McCue et al. 2012; Metzger et al. 2013a). The latter is of a magnitude lower density and not specifically designed for CNV capture.

Also, as shown in humans and cattle, the efficiency of CNV discovery is lower in SNP platforms compared to CNV-focused arrays (Haraksingh et al. 2011; Liu and Bickhart 2012). While the future direction for CNV research in any species is probably next generation sequencing (NGS), the approach has as yet found only limited application in horses: for the discovery of CNVs in the genome of a Quarter Horse mare (Doan et al. 2012b) and for the discovery of segmental duplications in 6 horse breeds and the donkey (Orlando et al. 2013).

A unique feature of our CGH array was the inclusion of probes from the Y chromosome and sub-telomeric regions. This was because CNVs and segmental duplications are known to be an integral part of the architecture of the mammalian Y chromosome (Skaletsky et al. 2003; Li et al. 2013), while sub-telomeres are hotspots of DNA breakage and repair, and undergo structural rearrangements more frequently than the rest of the genome (DeScipio et al. 2008; Riethman 2008). Despite this, only 10 CNVs were detected in sub-telomeres and none in the Y chromosome (Table 8). It is likely that the complex sub-telomeric sequences are missing or underrepresented in the current horse sequence draft assembly (Wade et al. 2009), so the designed probes did not originate from actual sub-telomeres. Also, because the Y chromosome contains ampliconic genes and has acquired sequences from other parts of the genome (Paria et al. 2011), many potential copy number variable Y probes did not pass the 'uniqueness' test by BLAST and were dropped during the array design.

Copy number variable genes and intergenic regions

Our study showed that CNVs tend to be located more frequently in genes than in intergenic regions. We also showed that among CNV genes predominately those involved in transmembrane signal transduction and chemo-attractant sensory perception (olfactory and non-olfactory G-protein coupled receptors, GPCRs), immune response (immunoglobulins, T-cell receptors, MHC protein complexes), and steroid metabolism (Appendix 2.4, 2.6). Notably, predominance of genic CNVs over intergenic variants, and the involvement of the same above mentioned functional gene categories are reported in all other CNV studies in horses (Doan et al. 2012a; Doan et al. 2012b; Dupuis et al. 2013; Metzger et al. 2013a; Wang et al. 2014b), humans (Redon et al. 2006; Hasin et al. 2008), cattle/ruminants (Fontanesi et al. 2010a; Fontanesi et al. 2010b; Bickhart et al. 2012), pigs (Wang et al. 2014a), dogs (Berglund et al. 2012) and chickens (Wang et al. 2010). It is likely that variation in these groups of genes is important for the adaptive plasticity of a species as already proposed by Susumu Ohno (Ohno 1970).

Although the majority of CNVRs involved genes, it is noteworthy that 20% of the CNVRs detected in this study were located in intergenic regions. These CNVRs were relatively small (average 50 kb, median 35 kb) and represented predominantly losses (Appendix 2.5). Prevalence of losses among intergenic CNVRs has also been found in humans (Conrad et al. 2006; Freeman et al. 2006) and dogs (Berglund et al. 2012).

Although there is no information about possible phenotypic effects of these regions in animal genomes, studies in humans show that intergenic deletions are

significantly enriched among so called regulatory CNVs (Schlattl et al. 2011) and probably modulate the expression of nearby genes (Kurth et al. 2009; Kantaputra et al. 2010). This is in line with the outstanding findings of the ENCODE project (Dunham et al. 2012) showing that only 2-3% of the mammalian genome is protein coding. The remaining 97-98% of the genome, earlier although to be junk DNA, is actually transcriptionally active, contains over 400,000 enhancer like elements and over 70,000 promoter-like sequences, and is probably involved in a variety of regulatory functions. It is therefore plausible that functions of some of these regulatory regions are further modulated by CNVs. Thus, with the improvement of genome sequence assembly and annotation in horses, intergenic CNVRs would be of interest for future studies.

Copy number variants and segmental duplications

Studies in human (Redon et al. 2006; Du et al. 2012) and cattle (Bickhart et al. 2012) have noted strong correlation between CNVs and segmental duplications (SDs). This is because SDs share 90% sequence similarity with another genomic location and can promote CNV formation by non-allelic homologous recombination (Stankiewicz and Lupski 2010).

Similar tendency has been observed in horses (Doan et al. 2012a), although horse SDs are relatively small (largest ~ 60 kb) and involve only about 0.5 – 0.6 % of the genome (Wade et al. 2009), thus less than the portion involved in CNVs. An additional 0.4% of the genome shows SDs in unplaced contigs (chrUn), although in this study only

0.04 % of chrUn sequences had CNVs (Table 8). Likewise, chr25 which is the most SD-rich chromosome (1.7%) according to EcuCab2 genome assembly (Wade et al. 2009), was only moderately enriched with CNVs (0.35%) in this study. Yet, findings by us and others support the correlation between CNVs and SDs in other regions of the horse genome. For example, a known large (750 kb) segmental duplication at the boundary of ELA class I and class III (Brinkmeyer-Langford et al. 2010) falls into a large common CNVR in chr20:30,127,886-31,231,182 (Appendix 2.4). Interrelation between the two types of variation is also supported by GO analysis showing that functional categories, such as olfactory reception and immune response, prevail among the genes involved both in CNVs and SDs (Orlando et al. 2013).

Integration and comparison of CNV data for the horse

In order to obtain a comprehensive overview about the status of CNV discovery in horses, we integrated the CNVRs identified in this study with the CNVs/CNVRs from all previous horse studies (Doan et al. 2012a; Doan et al. 2012b; Dupuis et al. 2013; Metzger et al. 2013a; Wang et al. 2014b). A composite dataset of 1476 CNVRs was generated (Figure 11; Appendix 2.7).

It is noteworthy that despite six CNV studies in the horse; only 301 CNVRs are shared between the studies and can be considered as validated. The remaining 1174 CNVs/CNVRs are study-specific and just a few have been confirmed by qRT-PCR and/or FISH. A similar situation has been encountered in human and other mammalian

CNV studies (Perry et al. 2006; Perry et al. 2008; Hou et al. 2011) indicating that methodological differences in CNV detection may essentially affect the results. With this regards, it is important to mention that over 32% (37 out of 116) of novel (study-specific) CNVRs identified by us aligned with a known CNVRs in the human genome (Appendix 2.8), providing an indirect evidence of these being true variants. Nevertheless, further work is necessary to confirm that the 1174 regions in the horse genome are indeed CNVRs and not just a noise caused by methodology. This is of particular importance because the present and all previous CNV studies in horses (Doan et al. 2012a; Doan et al. 2012b; Dupuis et al. 2013; Metzger et al. 2013a; Wang et al. 2014b) differ by discovery platforms, genome coverage, resolution, study cohorts, and methodological tools (Tables 11). Therefore, the overall numbers, size ranges, and chromosomal distribution of CNVs vary between the studies. For example, due to analytical reasons (Wang et al. 2012), all CGH-based studies ((Doan et al. 2012a; Wang et al. 2014b), this study) have detected more losses than gains. In contrast, gains predominate (97%) among the CNVs found by NGS in a Quarter Horse mare (Doan et al. 2012b). Regarding chromosomal representation, CNVs in the X chromosome have been found only in this study and by Doan & colleagues (Doan et al. 2012a), and CNVs in chrUn only in this study. Further, Wang & colleagues (Wang et al. 2014b) did not detect any CNVs in chrs30 and 31, while these two small autosomes show the highest number of CNVs in the Quarter Horse mare (Doan et al. 2012b). Major differences are also in the size, diversity and origin of study cohorts, ranging from just a few breeds and individuals (Doan et al. 2012b; Wang et al. 2014b) to over 15 breeds (this study and

(Metzger et al. 2013a)) and hundreds of individuals (Dupuis et al. 2013; Metzger et al. 2013a) (Table 11).

**Table 11: Summary statistics of all CNV studies in horses**

| | **This study** | **Doan *et al.* 2012a** | **Doan *et al.* 2012b** | **Dupuis *et al.* 2012** | **Metzger *et al.* 2013 **** | **Wang et al. 2014** |
|---|---|---|---|---|---|---|
| **Platform** | Tiling array | Tiling array | - | SNP Beadchip | SNP Beadchip | Tiling array |
| **Genome coverage** | WG | Exons and UTRs | WG | WG | WG | WG |
| **No of probes** | 400K | 400K | - | 50K | 50K | 1.3M |
| **Method** | CGH | CGH | NGS | genotyping | genotyping | CGH |
| **Breeds** | 16 | 15 | 1 | 4 | 17 | 6 |
| **Horses** | 38 | 16 | 1 | 477 | 717 | 6 |
| **CNVs/CNVRs *** | 258 | 775 | 282 | 478 | 166-1090 | 353 |
| **CNV size range** | 1 kb - 2.5 Mb | 197 bp - 3.5 Mb | 3.7 kb - 4.8 Mb | 97 bp - 2.7 Mb | 516 bp – 0.9 Mb | 6.1 kb - 0.5 Mb |
| **CNVs discovered** | WG, X, Un | WG, X | Autosomes | Autosomes | Autosomes | Autosomes, except chr 30, 31 |
| **Most enriched chr.** | 12 | 12 | 12 | 12 | 12 | 12 |
| **Chrs. with highest no. of CNVs** | 1; 20 | 1; 7 | 30; 31 | 1 | 12 | 20 |
| **Genome enrichment %** | 1.15 | 3.65 | 3.53 | 2.32 | 1.7 - 22.0 | 0.61 |

* As reported by original studies and before consolidating overlapping and tandemly located CNVRs into composite dataset

** The data by Metzger and colleagues (2013) vary in a broad range because the authors carried out multiple analyses to test different software packages.

Besides the differences, there are also outstanding similarities between horse CNV studies (Table 11): all agree that CNVs account for about 1 to 3 % of the horse genome; that there are more CNVs in genes than in intergenic regions, and all studies

unanimously find chr12 being the most CNV-enriched - not because of many CNVs, but because of a few very large clusters of olfactory receptors and immunity-related genes (Appendix 2.4, Appendix 2.6).

Breed-specific CNVs

One of the goals of CNV research in horses is to find breed specific variants that could be related to particular phenotypic traits or adaptations. To achieve this goal, CNV research in a large number of diverse horse breeds, each represented by multiple individuals, is needed. In order to evaluate the status of CNV research in horses from this viewpoint, we made a list of all horse breeds studied for CNVs and performed a phylogenetic analysis using population data of 15 microsatellite loci (Khanshour et al. 2013).

Currently, 41 horse breeds, thus less than 10% of the over 400 horse breeds known worldwide (Hendricks 1995), have been studied for CNVs (Table 12). Importantly, 25% (11/41) of the studied breeds were included in this study and represented mainly native ponies and draft horses. Genetic relationships of 32 breeds (the ones for which microsatellite genotyping data was available; E.G. Cothran, unpublished) are presented in Figure 15. The dendrogram shows that while the major clades of domestic horses are represented, there is a clear preponderance of the breeds with Thoroughbred ancestry.

Furthermore, only 7 breeds have been involved in 2 or more studies (Figure 15, Table 12) and several breeds are represented by just one individual (Doan et al. 2012a; Wang et al. 2014b). Therefore at this stage of research, any horse CNV reported to be breed-specific should be taken with caution. For example, our composite CNV dataset (Appendix 2.7) shows that the 18 CNVs reported to be specific for Hanoverians (Metzger et al. 2013a) are present in other breeds. Likewise, only one of the 7 CNVs considered to be specific to plateau breeds (Wang et al. 2014b), is not found in other breeds. The same happened with our data where initially we identified over 10 putative breed-specific CNVs which, after comparison, reduced to 2: one in Exmoor pony, another in Swiss Warmblood horse (Table 9). Interestingly, no specific CNVs were found in the Przewalski horse which shared similarity mainly with ponies and draft breeds (Appendix 2.1). Besides, only 9 of the 25 CNVs in Przewalski horses were shared between the two individuals studied. Similar tendency for intra-breed individual variation was observed for domestic horses where private CNVs predominated over the shared ones. Nevertheless, as suggested by other studies in horses (Doan et al. 2012a), cattle (Liu and Bickhart 2012), pigs (Wang et al. 2014a) and dogs (Molin et al. 2014),we anticipate that a small percentage of CNVs might remain unique to their respective breeds, although this requires analysis of much larger and more diverse equine populations.

**Table 12: List of horse breeds studied for CNVs**

| Breed | Reference |
|---|---|
| Abaga horse | Wang et al. 2014 |
| Akhal-Teke | This study |
| American Miniature Horse | Doan et al. 2012a, this study |
| American Quarter Horse | Doan et al. 2012a, Doan et al. 2012b, this study |
| Andalusian | Doan et al. 2012a |
| Anglo-Arabian | Metzger et al. 2013 |
| Arabian | Doan et al. 2012a, Metzger et al. 2013, this study |
| Arabian (Shagya) | Doan et al. 2012a |
| Belgian Draft | This study |
| Brandenburger | Metzger et al. 2013 |
| Caspian Pony | This study |
| Clydesdale | This study |
| American Baskhir Curly | Doan et al. 2012a |
| Debao | Wang et al. 2014 |
| Exmoor | This study |
| Fell Pony | This study |
| Friesian | This study |
| German Riding Pony | Metzger et al. 2013 |
| Gypsy Vanner | Doan et al. 2012a |
| Hanoverian | Doan et al. 2012a, Metzger et al. 2013 |
| Hegu | Wang et al. 2014 |
| Holsteiner | Metzger et al. 2013 |
| Hungarian | Doan et al. 2012a |
| Kasakh | Wang et al. 2014 |
| Lucitano | Doan et al. 2012a, Metzger et al. 2013 |
| Maremanno | Metzger et al. 2013 |
| Mongolian Native | Wang et al. 2014, this study |
| Oldenburg | Metzger et al. 2013 |
| Paso Fino | Doan et al. 2012a |
| Percheron | This study |
| Peruvian Paso | Doan et al. 2012a |
| Rhinelander | Metzger et al. 2013 |
| Rhenish-German Cold Blood | Metzger et al. 2013 |
| Selle Francais | Metzger et al. 2013 |
| Shire | Doan et al. 2012a |
| Sorraia | This study |
| Standardbred | This study |
| Swiss Warmblood | This study |
| Thoroughbred | Doan et al. 2012a, Dupuis et al. 2012, Metzger et al. 2013, Wang et al. 2014, this study |
| Trakehner | Metzger et al. 2013 |
| Welsh-Arabian pony | Doan et al. 2012a |
| Westphalian | Metzger et al. 2013 |
| Zweibrücker | Metzger et al. 2013 |
| Przewalski horse | Metzger et al. 2013, this study |

*No individual breeds were specified by Dupuis and colleagues (2012) where horses were classified as ponies, trotters, warmblood or draft breeds.

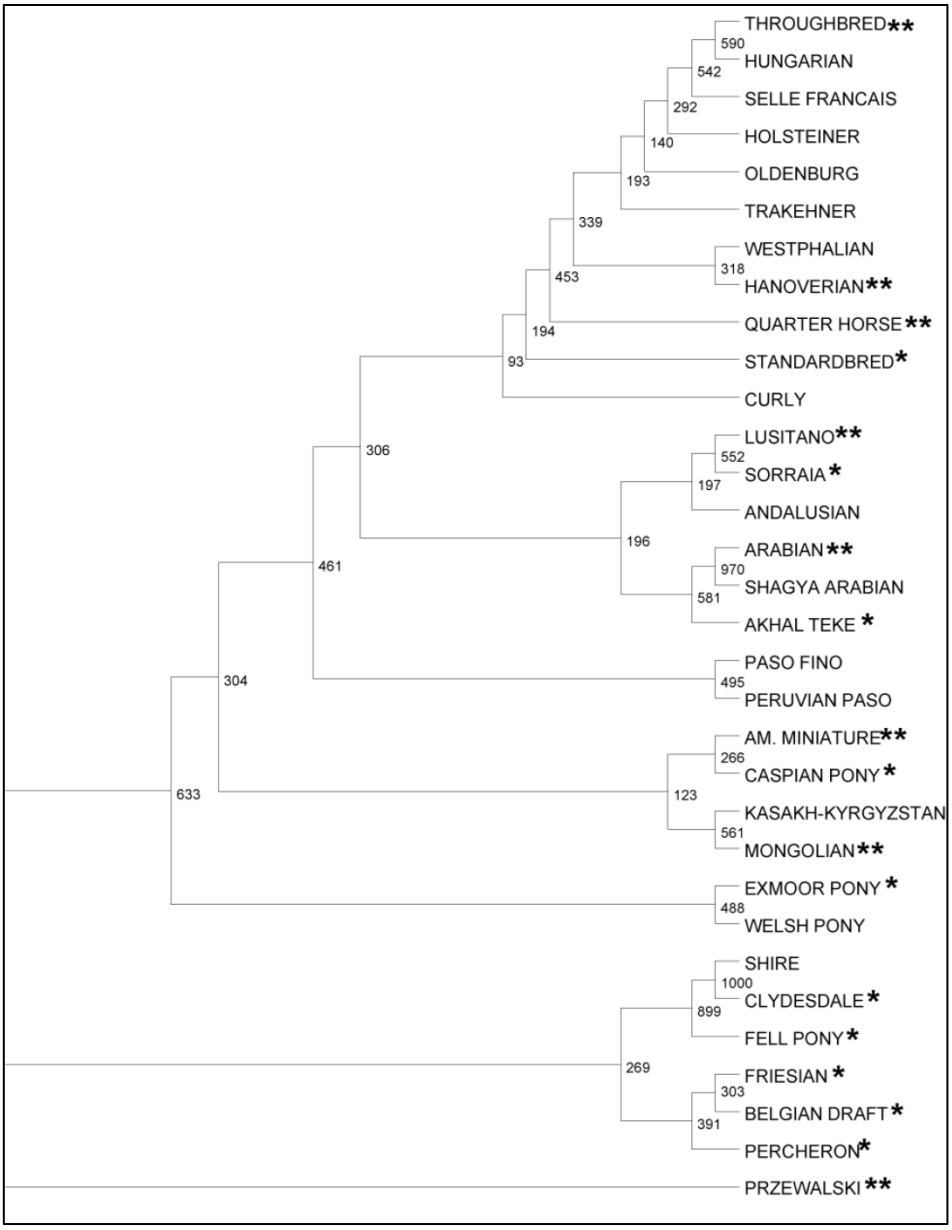


**Figure 15. Genetic relationships of the horse breeds studied for CNVs.**
Phylogenetic analysis done by Maximum Likelihood.

* New breeds added in this study (except Swiss Warmblood);

** Breeds involved in 2 or more studies. Numbers denote bootstrap values.

In summary, this study contributes to CNV research in horses by analyzing genomes of 16 horse breeds of which 11 were studied for CNVs for the first time. We took an important step towards the discovery of variants of biomedical importantce by carrying out a comparative analysis and integration of all available CNV data for the horse. We generated an integarted datset of 1476 equine CNVRs which is a necessary resource for the study of sequence properties and functional profiles of equine CNVs, as well as for the discovery of new variants, particularly those with possible pathogenic effects. We conclude that despite progress, the majority of the CNVs reported for the horse require further validation by invloving more diverse breeds and individual animals, and using comparable methodological platforms.

# CHAPTER III

# ANALYSIS OF GENOMIC COPY NUMBER VARIATION IN EQUINE RECURRENT AIRWAY OBSTRUCTION

## Introduction

Respiratory disorders directly affect the health, well-being, and performance of horses and have a major economic burden on the owners. Affected horses cannot compete for key events and require constant treatment or prevention that is expensive and has limited efficacy. That is why the equine industry and practitioners have listed respiratory ailments as disorders of the highest research priority in the horse (Oke 2013).

Recurrent airway obstruction (RAO) is a common chronic pulmonary disease that affects the respiratory system of adult horses of diverse breeds' worldwide (Robinson et al. 1996). The condition has two forms. One form, usually referred to as just RAO or chronic obstructive pulmonary disease (COPD-RAO) (Venugopal et al. 2010), affects horses living in poorly ventilated stalls in cold climate. Continuous inhalation of organic dusts and endotoxins causes hypersensitivity reaction of the lower respiratory tract and results in inflammation, airway-blockage, bronchial muscle contraction, and difficulty in breathing during resting (Derksen et al. 1988; McGorum et al. 1993; Pirie et al. 2003). The other form is summer pasture associated obstructive pulmonary disease (SPAOPD) which is caused by air born allergens in pasture-living horses in warm and humid climates (Seahorn et al. 1994). Clinical signs of SPAOPD

include recurrent nasal discharge, cough, labored expiratory effort, and crackles and wheezes on auscultation (Seahorn and Beadle 1993). Overall, the pathophysiologic changes for COPD-RAO and SPAOPD-RAO are similar, characterized by mucus production, neutrophil accumulation in the airway, bronchial hyperactivity, and bronchospasm (Venugopal et al. 2010). In literature, the cold climate form is typically referred to as RAO, and the Southern form as SPAOPD-RAO (Venugopal et al. 2010). Notably, many of the clinical symptoms observed in RAO are indicative of human asthma due to which this equine condition has been proposed to serve as a natural model for asthma in humans (Swinburne et al. 2009; Scharrenberg et al. 2010).

Being a complex disease, etiology of RAO involves both an environmental and a genetic component. The latter is thought to be polygenic (Schaeper 1939) with a complex inheritance mode (Schaeper 1939; Gerber et al. 2009), variable expressivity and genetic heterogeneity across equine populations (Swinburne et al. 2009). Due to this complexity, the current knowledge of RAO genetics is limited. The condition has been studied using a combination of genomics approaches, such as microsatellite- and SNP-based whole-genome association studies (Swinburne et al. 2009; Shakhsi-Niaei et al. 2010; Shakhsi-Niaei et al. 2012), candidate gene and whole genome gene expression profiling (Kachroo et al. 2010; Venugopal et al. 2010; Cote et al. 2012; Klukowska-Rotzler et al. 2012b; Lanz et al. 2013) and pathway analysis of candidate genes combined with proteomic data (Racine et al. 2011). While the findings provide some insight into the multi-genic and complex nature of the condition and show the likely involvement of interleukins and other innate immunity related genes (Shakhsi-Niaei et

al. 2010; Klukowska-Rotzler et al. 2012a; Shakhsi-Niaei et al. 2012; Lanz et al. 2013), still very little is known about the genetic predisposition of horses to RAO, and molecular genetic mechanisms modulating pathogenesis of the disease.

So far, candidate gene and association studies of RAO in horses have primarily focused on SNPs, and not considered possible role of another form of common genetic polymorphism – copy number variation (CNVs). This involves deletions, duplications, and complex rearrangements typically larger than 50 base-pairs that contribute to phenotypic plasticity and adaptations, but can also be pathogenic and associated with diseases and disorders (Conrad et al. 2010b; Weischenfeldt et al. 2013). For example, recent studies show that over 20% of asthma candidate genes are located near or within CNVs, suggesting possible role of CNVs in the susceptibility and/or pathogenesis of human asthma (Rogers et al. 2013; Ceroni et al. 2014).

Building upon these findings in humans, the aim of this study is to investigate potential involvement of CNVs in equine RAO. Using a custom-made whole-genome tiling array, we identify CNVs in the genomes of severely RAO affected horses and matching controls. The findings add to our insights about the genetic component of this respiratory disorder and the role of CNVs in equine complex traits.

## Material and methods

Horses and phenotypes

The study population (n=63) comprised of severely RAO affected (n=39) and age- and/or breed-matched healthy control horses (n=24) of various breeds (Table 13). The horses originated from three different study populations and were phenotyped as follows: i) RAO affected (n=20) and control (n=19) Swiss Warmblood horses were phenotyped at the University of Bern, Switzerland (Dr. Gerber) by measured exposure to agents that trigger RAO, followed by detailed clinical examination using the Horse Owner Assessed Respiratory Signs Index (HOARSI) (Laumen et al. 2010)(Lanz et al. 2013). All RAO affected horses in this group categorized as HOARSI-4 and the controls as HOARSI-1; ii) RAO affected horses (n=13) of different breeds from The University of Montreal, Canada (Dr. Lavoie) were phenotyped by exposure to moldy hay to induce airway obstruction, followed by full clinical exam which included tests for airway reactivity, mucus accumulation, and transpulmonary pressure, and cytological analysis of tracheo-bronchial and bronchio-alveolar fluids (Lavoie-Lamoureux et al. 2012; de LAGARDE et al. 2014), and iii) SPAOPD-RAO (n=6) and matched control horses (n=5) of different breeds were phenotyped according to HOARSI (Laumen et al. 2010) at the Mississippi State University (Dr. Swiderski) as HOARSI-4 (RAO affected) and HOARSI-1 (controls). Twenty-two horses (16 RAO affected and 6 controls, Table 13)

were studied by array CGH, while all 63 animals were analyzed by quantitative PCR (qPCR).

**Table 13: Details about the RAO study cohort.**
H4 – HOARSI4; H1 – HOARSI1; SPA – SPAOPD; * - horses used for array CGH

| Horse ID | Breed | Gender | Sample origin |
|---|---|---|---|
| **RAO affected horses** | | | |
| *H4-1912 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| H4-1883 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| *H4-1892 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| *H4-1943 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H4-1970 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| *H4-1987 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H4-1996 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H4-2004 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| H4-2015 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H4-2021 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| *H4-1998 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| *H4-1999 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H4-2007 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H4-2017 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| *H4-2029 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H4-2035 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| *H4-2109 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H4-2134 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H4-2139 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H4-2147 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| SPA-H4-1 | Appaloosa | Male | Mississippi State University, USA |
| *SPA-H4-2 | Quarter Horse | Male | Mississippi State University, USA |
| *SPA-H4-3 | Tennessee Walking | Male | Mississippi State University, USA |
| *SPA-H4-4 | Missouri Foxtrotter | Female | Mississippi State University, USA |
| *SPA-H4-5 | Tennessee Walking | Male | Mississippi State University, USA |
| SPA-H4-6 | Mustang | Male | Mississippi State University, USA |
| *RAO-40 | Standardbred | Female | University of Montreal, Canada |
| *RAO-354 | Appaloosa | Male | University of Montreal, Canada |
| *RAO-361 | Thoroughbred | Female | University of Montreal, Canada |
| RAO-363 | Quarter Horse | Female | University of Montreal, Canada |
| *RAO-372 | Belgian draft | Female | University of Montreal, Canada |

**Table 13 continued**

| Horse ID | Breed | Gender | Sample origin |
|---|---|---|---|
| **RAO affected horses** | | | |
| RAO-378 | Flea-bitten Grey Horse | Female | University of Montreal, Canada |
| RAO-379 | Canadian | Female | University of Montreal, Canada |
| RAO-380 | Standardbred | Male | University of Montreal, Canada |
| RAO-382 | Paint | Male | University of Montreal, Canada |
| RAO-383 | Arabian | Female | University of Montreal, Canada |
| RAO-384 | Quarter Horse | Female | University of Montreal, Canada |
| RAO-391 | Appaloosa | Female | University of Montreal, Canada |
| RAO-392 | Quarter Horse | Female | University of Montreal, Canada |
| **Control horses** | | | |
| *H1-126* | Swiss-Warmblood | Female | University of Bern, Switzerland |
| *H1-140 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| *H1-146 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| *H1-154 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H1-160 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H1-191 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H1-225 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| H1-1502 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H1-1864 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| H1-192 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H1-1145 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| H1-1460 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| H1-1509 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H1-1511 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H1-1519 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H1-1605 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| H1-1610 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| H1-1941 | Swiss-Warmblood | Female | University of Bern, Switzerland |
| H1-2008 | Swiss-Warmblood | Male | University of Bern, Switzerland |
| SPA-Hl-1 | Quarter Horse | Male | Mississippi State University, USA |
| SPA-Hl-2 | Quarter Horse | Male | Mississippi State University, USA |
| *SPA-Hl-3 | Quarter Horse | Male | Mississippi State University, USA |
| SPA-Hl-4 | Quarter Horse | Female | Mississippi State University, USA |
| *SPA-Hl-5 | Thoroughbred | Female | Mississippi State University, USA |

DNA isolation & quality control

DNA was isolated from peripheral blood and/or hair using standard PCI (Phenol/ Chloroform/ Isoamyl alcohol) method (Sambrook et al. 1989) or Gentra Puregene Blood Kit (QIAGEN, USA) according to manufacturer's protocol. The latter was modified to adjust it for DNA isolation from hair follicles. Briefly, hair follicles were incubated in cell lysis solution for at least 72 hrs. After isopropanol treatment to precipitate DNA pellet, the sample was centrifuged at 16,000 x g for 20-30 minutes, instead of 3 minutes at 2000 x g, according to the original protocol. The same applied after treatment with 70% ethanol. The DNA was cleaned with DNeasy Blood and Tissue kit (Qiagen) with the following modifications to the manufacturer's protocol: during DNA purification, 80% ethanol was used instead of solution AW2, and the final DNA pellet was eluted in water instead of solution AE. The final quality and quantity of the DNA samples were checked by gel electrophoresis on a 1% agarose gel and by Nanodrop spectrophotometry (Thermo Scientific).

Array comparative genomic hybridization

***Array CGH experimental procedures***

Probe labeling and array CGH experiments were performed according to Agilent Technologies Protocol Version 6.2.1. All hybridizations comprised of a pair of differently labeled probes, one of which was always the reference DNA (Thoroughbred).

Briefly, 1µg of purified genomic DNA was digested with restriction enzymes *RsaI* and *AluI* for 2 hours at 37°C and 20 minutes at 65°C to produce 200-500 bp fragments. Cleaved samples were labeled with fluorescence dyes Cy3 (for the reference DNA) or Cy5 (for sample DNA) by random priming using Genomic DNA Enzymatic Labeling Kit (Agilent Technologies). The final labeling reaction contained random primer (5µl), 1X dNTP (5µl) mix, 1X reaction buffer (10µl), Exo-Klenow enzyme (1µl) and Cy3- or Cy5-dUTP (3µl), and was conducted for 2 hours at 37°C, 10 minutes at 65°C. The products were cleaned with 30 kDa filters (Amicon) and the yield and specific activity of labeled DNA was determined with a Nanodrop spectrophotometer. Typical yield for 1 µg of starting DNA was 6-8 µg; specific activity for Cy3 was 25-40 pmol/µg and for Cy5 20-35 pmol/µg. The hybridization mixture was prepared using Agilent Oligo aCGH Hybridization Kit and contained equal quantity of Cy3 and Cy5 labeled probes, 1 µg/µL horse Cot1 DNA, 10X blocking agent, and 2X Hi-RPM buffer. Denatured and pre-annealed probe mixture was applied onto gasket slide, placed in Agilent SureHyb hybridization chamber, 'sandwiched' with an array slide and incubated in Agilent hybridization oven at 65°C for 40 hours. The array slides were washed with Agilent aCGH Wash Buffers 1 and 2, 10 min in each, and dried with Acetonitrile and Agilent Stabilization and Drying Solutions.

***Array data analysis***

The array slides were scanned with Agilent SureScan DNA Microarray Scanner and Agilent Scanner Control software v8.3. The data were extracted and normalized with Agilent Feature Extraction software v11.0.1.1 and saved in .fep format. The Feature Extraction software checks the quality of aCGH by measuring Derivative $Log_2$ Ratio Standard Deviation (DLRSD), Signal-To-Noise Ratio (SNR) and Background Noise (BGNoise). The data were further analyzed with Agilent Genomic Workbench 5.0 software. In each array spot $log_2$ ratios of Cy3 versus Cy5 were computed with the default *P*-value threshold 0.05 and overlap threshold value 0.9. The CNVs were represented by gains and losses of normalized fluorescence intensities relative to the reference and called by conservative criteria which required alternations of >0.5 $log_2$ ratios over 5 neighboring probes. Homozygous losses were called when signal $log_2$ ratio was <-2.0. Copy number variable regions (CNVRs) were determined by ADM-2 algorithm (Lipson et al. 2006) by combining overlapping CNVs in all samples across the CGH experiments. Parameters used in this analysis were as follows: Threshold of ADM-2: 6.0; Centralization: ON (Threshold: 6.0, Bin Size: 10); Fuzzy Zero: ON; Aberration Filters: ON (minProbes = 5 & minAvgAbsLogRatio = 0.5 & maxAberrations = 10000 AND percentPenetrance = 0); Feature Level Filters: ON, IntraArray: ON). Output files were generated with genomic coordinates and cytoband locations for all CNVs. The raw data will be submitted to NCBI Gene Expression Omnibus (GEO).

Quantitative PCR (qPCR)

Genomic copy number changes in horse genes *SPI2* and *SERPINA1* in chr24, as detected by aCGH, were validated and further analyzed by qualitative and quantitative PCR. Primers for array probes, genomic segments and gene exons and introns (Table 14) were designed using horse whole genome sequence information (EquCab2), UCSC Genome Browser, and Primer3 software. Optimal DNA concentration for qPCR, 25 ng/µL, was determined by serial dilutions. The qPCR experiments were performed with LightCycler® 480 (Roche Diagnostics) in triplicate assays. Each assay was done in triplicate 20 µL reactions containing 50 ng of template DNA, 10 µM primers, and the SYBR Green PCR kit (Roche). Relative copy numbers of the selected regions were determined in comparison to the reference sample (Thoroughbred) and normalized to an autosomal reference gene *GAPDH*. The cycling conditions were 1 cycle 5 min at 95°C; 45 cycles 10 sec at 95°C, 5 sec at 58°C, and 10 sec at 72°C; 1 cycle for melting curve 30 sec 95°C, 30 sec 65°C and final cooling 20 sec at 50°C. Quantification of the copy number was carried out using the comparative $C_T$ method ($2^{\Delta\Delta}$Ct) (Livak and Schmittgen 2001; Bodin et al. 2005) with $p<0.05$ as a cut-off threshold for statistical significance. Qualitative PCR results were analyzed by agarose gel electrophoresis.

**Table 14: Primers for qualitative and qPCR to analyze a CNVR in chr24.**
(see also Figure 3.1 for explanations)

| No | Primer Location | Forward 5'-3' | Reverse 5'-3' | Product size, bp |
|---|---|---|---|---|
| 1 | *SPI2* 3' | TCGGAGCCTCAGTTTCTTCT | ACATGAGGGATGACGGTGA | 133 |
| 2 | *SPI2* 3' | GTTATGCGGAGTTGCACACA | ATCACCGCCTAAGCCACTC | 100 |
| 3 | *SPI2* 3' | TTCAGGAAGCTGTGGGTTTC | CCTTGCCTGTGTCCACTGTA | 157 |
| 4 | Exon | GTGCCAGGTCAACCATTCTT | ACCACGATGTGGGAAATCAT | 216 |
| 5 | Exon | TGCCCAAACTGTCCATTTCT | GTCAAGGGCACTTCCTCAGT | 121 |
| 6 | Intron | GTCAGGCCAGTAGGTTGGAA | AGCCTGAGTGGTGCTACCTG | 135 |
| 7 | Intron | AGAGGCTTTGCAGGACAAGA | TCTCCCACTTAGCCCCCTAT | 113 |
| 8 | Exon | TCGAGCCTGAGTCTACGACA | GTCGGGCAGGATGAAGAAG | 105 |
| 9 | Intron | TGCACAACCAACCTGAGCTA | CCTTCTTTGCTGGTGGAAAC | 103 |
| 10 | Intron | TACAAAGAACACGGGACTCG | ACAGGTTCCCGACAGTTTCA | 115 |
| 11 | Intron | TTAGCCCCGTTTACAGATGG | GGAGAGAAGGGAAGGTGGAC | 159 |
| 12 | Array probe | GCAAAGCCCTCATCTTCTTG | GTGGGTTTCAGCCAACATTT | 167 |
| 13 | Exon | TTCCTTGGGTTCCCTTTTCT | GCTTCCAGCATCTCCTGAAC | 219 |
| 14 | Array probe | AGAATATCCCAAGCCACCAG | CCTCACCTTCCCCATCTGTA | 199 |
| 15 | Array probe | GTGCTGACCATTGACGAGAA | TTAAGACGAAGGGCCTGTTG | 109 |
| 16 | Array probe | CAAATATCCCAAGCCACCAG | GACGGATTTTGCGAGGATAA | 148 |
| 17 | Array probe | GCTCCTGAACCACAGAGTCC | GGTAGCTCTTGCTCCTCGTG | 134 |
| 18 | Array probe | TAACGACAGGCCATTCTTCC | AAGAAGAGCTTCCCCTGAGC | 194 |
| 19 | Exon | TGGGGAGAAGAAGATGTTGG | CTGAGGATCTGCAGGGAGAC | 173 |
| 20 | Exon | TCGAGCCTGAGTCTACGACA | GTCGGGCAGGATGAAGAAG | 176 |
| 21 | Array probe | GTGCTGACCATTGACGAGAA | TTAAGACGAAGGGCCTGTTG | 109 |
| 22 | Exon | GTTTGCCCAAACTGTCCATT | ACACTGTCAGGGACGTTTCC | 129 |
| 23 | Array probe | AGAACCGTCACCTGACAAGG | CCATCTGGCTGTGCAGAGTA | 153 |
| 24 | Exon | GCTGAACCAGAGAGGCAGTT | GACGAGAAAGGCACTGAAGC | 183 |
| 25 | Array probe | GTGTCCCAGGTCTTCTCGTC | CTGAACCACACACAGGAAGC | 155 |

Gene content analysis of CNVs

Ensembl gene list (Ensembl Genebuild 73.2) along with their position in the horse genome was added to Agilent Genomic Workbench as a custom track to determine the genic and intergenic CNVs. Gene Ontology analysis (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the genes present in CNVs were performed using DAVID (The Database for Annotation, Visualization and Integrated Discovery) bioinformatics tool with default settings (Huang da et al. 2009b; Huang da et al. 2009a). TheGO annotations used were: BP (Biological Process), MF (Molecular Function) and CC (Cellular Component). Because only a limited number of genes in the horse genome have been annotated, horse gene IDs were converted to orthologous human Ensembl gene IDs by BioMart, followed by GO and pathway analyses, as described above. Biological functions of the genes in CNVRs were further analyzed manually by data mining in Ensembl (Stalker et al. 2004), UCSC (Karolchik et al. 2003) and NCBI (Sayers et al. 2010) Genome Browsers searching for data for equine orthologs in other mammalian species. CNVs present in intergenic regions were analyzed in UCSC genome browser and NCBI and GeneCards for similarities to known mammalian genes. Genes present within equine CNVRs were also aligned with the human genome variant database (HGVD) and Online Mendelian Inheritance databases for Man (OMIM) and Animals (OMIA).

## Results

### CNV discovery and analysis

Array CGH experiments in 22 horses, 16 RAO cases and 6 controls, resulted in 760 CNV calls (Table 15; Appendix 3.1) with an average of 35.9 calls per individual. This number is slightly higher than the 26.4 calls per individual as reported by us for normal horse populations in Chapter II. Among individual horses, the number of CNV calls was the highest (69) in a SPAOPD affected Quarter Horse SPA-H4-2 and the lowest (10) in an RAO-affected Standardbred RAO-40 (Table 15). While there were no significant differences between the numbers of CNV calls between RAO cases and controls, a clear tendency for fewer calls was observed for experiments using DNA isolated from hair follicles compared to blood DNA (Table 15).

Losses clearly dominated over gains which is a typical feature for aCGH experiments and reported by us (Chapter II) and others (Doan et al. 2012a; Wang et al. 2012; Wang et al. 2014b) before. Completely and partially overlapping CNV calls were merged into 245 CNV regions (CNVRs) (Appendix 3.2).

Next, in order to find CNVs specific to RAO cases, we compared the CNVRs detected in this study with the composite dataset of 1476 CNVs/CNVRs for the horse (Chapter II) and with the CNVRs detected in horses with disorders of sexual development (Chapter III). Comparison revealed that the majority (197) of CNVRs

found in this study were shared with other equine populations, thus validating those as common CNVRs in the horse genome.

Importantly, 48 CNVRs (Appendix 3.3) were found only in the RAO study cohort (16 cases, 6 controls) and, thus, represented novel CNVRs for the horse. Among these, 18 were present in both RAO cases and controls, or controls only, and contributed mainly to the horse CNV dataset. However, 30 variants were exclusive to RAO cases and subjected for further analysis.

Novel CNVRs in RAO affected horses

After the initial analysis of all CNVRs, we focused on the 30 novel CNVRs that were found only in RAO affected horses (Table 16; Appendix 3.4) and were tentatively referred to as '*RAO-specific'*. Such CNVRs regions were detected in 9 out of 16 RAO affected horses and the majority were private, thus present in only one individual (Table 16; Appendix 3.4). One novel CNVR, a 36 kb loss in chr10 involving an immunity-related *FUT9* gene (Comelli et al. 2006), was shared by a Swiss Warmblood H4-1998 and a Quarter Horse SPA-H4-2. No RAO specific CNVRs were found in six affected horses, *viz.*, RAO-40, RAO-354, RAO-361, RAO-372, H4-1892, H4-2109, and SPA-H4-3. Chromosome-wise, the RAO specific CNVRs were distributed on 15 autosomes (chr1, 3, 5, 6, 8, 10, 11, 14, 15, 16, 20, 21, 23, 24, 29, and chrUn. The average size of these CNVRs was 37 kb which is smaller than the 110 kb average for the horse genome (Chapter II).

**Table 15: Summary of CNV calls in RAO cases and controls.**
H1 – controls; H4 –cases; @ - horses used in breed CNV study in Chapter II.

| Horse ID & phenotype | #CNVs | Gains | Losses | Novel | DNA source |
|---|---|---|---|---|---|
| **RAO affected horses** | | | | | |
| *H4-1912 | 43 | 11 | 32 | 1 | blood |
| *H4-1892 | 39 | 12 | 27 | 0 | blood |
| *H4-1943 | 42 | 4 | 38 | 1 | blood |
| *H4-1987 | 60 | 10 | 50 | 5 | blood |
| *H4-1998 | 59 | 6 | 53 | 3 | blood |
| *H4-1999 | 61 | 3 | 58 | 4 | blood |
| *H4-2029 | 23 | 0 | 23 | 1 | blood |
| *H4-2109 | 30 | 3 | 27 | 0 | blood |
| *SPA-H4-2 | 69 | 15 | 54 | 12 | blood |
| *SPA-H4-3 | 39 | 14 | 25 | 0 | blood |
| *SPA-H4-4 | 21 | 11 | 10 | 2 | blood |
| *SPA-H4-5 | 24 | 9 | 15 | 1 | blood |
| *RAO-40 | 10 | 4 | 6 | 0 | hair |
| *RAO-354 | 18 | 1 | 17 | 0 | hair |
| *RAO-361 | 11 | 4 | 7 | 0 | hair |
| *RAO-372 | 13 | 3 | 10 | 0 | hair |
| **Control horses** | | | | | |
| *H1-126 | 30 | 8 | 22 | 0 | blood |
| *H1-140@ | 29 | 9 | 20 | 1 | blood |
| *H1-146@ | 30 | 6 | 24 | 1 | blood |
| *H1-154@ | 23 | 1 | 22 | 0 | blood |
| *SPA-H1-3 | 24 | 7 | 17 | 0 | blood |
| *SPA-H1-5 | 62 | 18 | 44 | 0 | blood |
| **Total** | **760** | **159** | **601** | 32 | |
| **Average** | **35.9** | | | | |

Altogether, 24 of the 30 RAO specific CNVRs (80%) involved transcribed genes: 19 protein coding genes and 9 known human or horse mRNA/cDNA sequences (Table 16, Appendix 3.4). Six CNVRs were located in intergenic regions or in chrUn. Functional relevance of the genes to RAO was further analyzed by mining human and animal genome databases (NCBI, Ensembl, GeneCards, GO, HGVD, OMIM, OMIA) and published literature.

Among the CNV genes, several posed interest due their involvement in acquired and innate immunity (*BTNL2* MHC class III; *FUT9*; *RNASE9*), infections (*FUT9; KRT23),* disease susceptibility (*BTNL2*), respiratory disorders (*BTNL2, LINGO2, SPI2, SERPINA1*) and tuberculosis (*BTNL2*, Morais et al. 2012; *RDH16, SIRT4*), autoimmune disorders (*BTNL2* rheumatoid arthritis; Mitsunaga et al. 2013) and drug metabolism (*ELTD1, RDH16*). However, none of the RAO specific CNVRs involved genes which have been previously associated with RAO, such as interleukins or interleukin receptors, (Klukowska-Rotzler et al. 2012b; Shakhsi-Niaei et al. 2012). Likewise, no CNVRs were detected in horse chromosome 11 which is homeologous to HSA17q21 (Raudsepp et al. 1996, Raudsepp et al. 2008) – the strongest candidate region for human asthma by GWAS (Akhabir & Sandford 2011). Even although a CNVR in chr15 at 79 Mb in a Swiss Warmblood H4-1999 fell into one of the RAO candidate regions according to GWAS (Swinburne et al. 2009), the functions of the *NBAS* gene in the CNVR were not related to autoimmunity or inflammation (Table 16).

Eighteen equine RAO specific genic CNVRs had a homeologous CNVR in The Human Genome Variant Database (HGVD) (Table 16) - an indication that these

genomic regions tend to be copy number variable in diverse mammalian genomes. The search of OMIM and OMIA databases revealed that 6 RAO-specific CNVRs involved genes associated with 6 known Mendelian disorders in humans, and one in dogs (Table 16).

Of particular interest was a CNVR in chr24 at 37 Mb in a Swiss Warmblood horse H4-1987 (Table 16). This was the only RAO specific CNVR which corresponded to entries in all three databases analysed: the HDGV, OMIM and OMIA. Most importantly, the region involved two genes, *SPI2* and *SERPINA1*, which encode for alpha-antitrypsin. Deficiency of alpha-antitrypsin causes lung and liver malfunctioning, and human *SERPINA1* is a known candidate gene for Chronic Pulmonary Obstructive Disease (COPD) (Zorzetto et al. 2008; Denden et al. 2010; Papatheodorou et al. 2010; Carroll et al. 2011; Quint et al. 2011; Serapinas et al. 2012; Linja-aho et al. 2013). Furthermore, this CNVR is located less than 40 kb upstream of a cluster of at least 5 other serpin-family genes (Figure 16), including *SERPINA3* which has been associated with human asthma (Rogers et al. 2012). Due to these reasons, this CNVR was analyzed in more detail (see below).

Besides the analysis of putative RAO specific CNVRs, we sorted the CNV data for highly negative aberration values ($\log2 < -2$ ) and identified 15 putative homozygous deletions (appendix 3.5). However, as the majority involved clusters of olfactory receptor genes, no further analysis followed.

**Table 16: Summary information for the 30 "RAO specific" CNVRs.**

| Position, Mb | CNV Type | Horse ID and phenotype | Gene symbol | Associated phenotype/processes | Human DGV ID | OMIM ID, description | OMIA ID, description |
|---|---|---|---|---|---|---|---|
| chr1:22.5 | Loss | SPA-H4-2 | JL635247 | n/a | n/a | n/a | n/a |
| chr1:43.8 | Loss | SPA-H4-2 | intergenic | n/a | n/a | n/a | n/a |
| chr1:99.1 | Loss | SPA-H4-2 | human cDNA | n/a | n/a | n/a | n/a |
| chr1:149.1 | Loss | H4-1999 | AK127847 | n/a | esv2745248 | n/a | n/a |
| chr1:157.4 | Loss | H4-2029 | *RNASE9* | Epididymitis, pancreatitis | nsv901468 | n/a | n/a |
| chr1:185.8 | Gain | H4-1912 | JL637766, JL629243 | n/a | n/a | n/a | n/a |
| chr3:105.5 | Loss | SPA-H4-2 | JL640154 | n/a | n/a | n/a | n/a |
| chr5:5.6 | Loss | H4-1998 | *NME7* | Primary ciliary dyskinesia and *situs inversus* | nsv872531 | n/a | n/a |
| chr5:83.9 | Loss | SPA-H4-2 | *ELTD1* | Drug metabolism. Cannabis dependence and nicotine addiction | nsv871835 | n/a | n/a |
| chr6:16.3 | Loss | SPA-H4-2 | intergenic | n/a | n/a | n/a | n/a |
| chr6:74.6 | Gain | SPA-H4-4 | *RDH16* | Cytochrome p450, tuberculosis. Metabolism and Drug metabolism | esv24053 | 612712, leber congenital amaurosis 13, *RDH12* | n/a |

**Table 16 continued.**

| Position, Mb | CNV Type | Horse ID and phenotype | Gene symbol | Associated phenotype/processes | Human DGV ID | OMIM ID, description | OMIA ID, description |
|---|---|---|---|---|---|---|---|
| chr6:78.7 | Loss | H4-1999 | *USP15* | Spinocerebellar ataxia | n/a | n/a | n/a |
| chr8:13.8 | Loss | SPA-H4-2 | *SIRT4, U4* | Down-regulates insulin secretion. Hyperinsulinism and tuberculosis | nsv899555 | n/a | n/a |
| chr8:89.2 | Gain | H4-1987 | ZNF407 | Ectodermal dysplasia | nsv909919 | n/a | n/a |
| chr10:47.7 | Loss | SPA-H4-2 , H4-1998 | *FUT9*, pseudogene | Malaria, and cytomegalovirus infection | esv2422431 | n/a | n/a |
| chr10:53.7 | Loss | SPA-H4-2 | AK125948 | n/a | nsv912634 | n/a | n/a |
| chr11:21.7 | Loss | H4-1998 | *KRT23* | Aggressive periodontitis and periodontitis | nsv9550 | n/a | n/a |
| chr14:75.4 | Loss | H4-1999 | intergenic | n/a | n/a | n/a | n/a |
| chr15:79.0 | Loss | H4-1999 | *NBAS* | Short stature, optic nerve atrophy, Pelger–Huët anomaly and neuroblastoma | nsv515567 | 614800, short stature, optic nerve atrophy, and Pelger–Huët anomaly | n/a |
| chr16:13.1 | Loss | SPA-H4-2 | *CNTN4* | Autism spectrum disorders and atrioventricular septal defect | esv2422274 | 612540, myopathy, congenital, compton-north, *CNTN1* | n/a |

**Table 16 continued.**

| Position, Mb | CNV Type | Horse ID and phenotype | Gene symbol | Associated phenotype/processes | Human DGV ID | OMIM ID, description | OMIA ID, description |
|---|---|---|---|---|---|---|---|
| chr16:67.8 | Loss | SPA-H4-2 | *CPNE4* | Infertility | Variation_4354 | n/a | n/a |
| chr20:32.6 | Loss | SPA-H4-2 | *BTNL2* | Pulmonary sarcoidosis | nsv830630 | 181000, sarcoidosis, susceptibility | n/a |
| chr21:39.7 | Loss | SPA-H4-5 | *CDH12* | Meconium aspiration syndrome and congenital diaphragmatic hernia | nsv428115 | 137215, gastric cancer, hereditary diffuse; *CDH1* | n/a |
| chr23:45.6 | Loss | SPA-H4-4 | *LINGO2* | Neuronitis and essential tremor | esv270849 | n/a | n/a |
| chr24:37.0 | Loss | H4-1987 | *SPI2*, *SERPINA1* | Alpha 1-antitrypsin deficiency, and liver disease, chronic pulmonary obstructive disease (COPD) | nsv470662 | 611489, corticosteroid-binding globulin deficiency *SERPINA6* | 000032-9615 UID:102, Alpha-1-antitrypsin deficiency in dogs |
| chr29:23.3 | Gain | H4-1987 | *CELF2*, Non-Horse Ref Gene | Neuroblastoma, and childhood epilepsy | nsv5798 | n/a | n/a |

**Table 16 continued.**

| Position, Mb | CNV Type | Horse ID and phenotype | Gene symbol | Associated phenotype/processes | Human DGV ID | OMIM ID, description | OMIA ID, description |
|---|---|---|---|---|---|---|---|
| chrUn:0.11 | Gain | SPA-H4-5 | intergenic | n/a | n/a | n/a | n/a |
| chrUn:0.47 | Loss | H4-1987 | intergenic | n/a | n/a | n/a | n/a |
| chrUn:0.04 | Loss | H4-1987 | intergenic | n/a | n/a | n/a | n/a |
| chrUn:0.08 | Loss | H4-1943 | JL631917 | n/a | n/a | n/a | n/a |

Analysis of a CNVR in *SERPIN* genes in chr 24

The CNVR in chr24: 37,051,746-37,077,730 was a 26 kb size deletion involving the entire *SERPINA1* gene and *SPI2* exon 1 (Figure 16). The CNVR was demarcated by 9 array probes which were located as follows: probe #12 (Ec_24_37051895) in *SPI2* exon 1; probes #21 and #23 (Eic_24_37070988, Eic_24_37071086,) in *SERPINA1* introns 2 and 3; probe #25 (G_24_37077530) in the 5' region of *SERPINA1*, and probes #14-18 (Eic_24_37054678, G_24_37061721, G_24_37066083, Eic_24_37068005, Ec_24_37068288) between *SPI2* and *SERPINA1* (Figure 16B, C; note that both genes are transcribed from the reverse strand (Ensembl). The deletion aberration values were moderate ($-\log_{2average}$ -0.5; $\log_{2max}$ -0.8) and the CNVR was detected in one RAO affected Swiss Warmblood male horse (H4-1987).

As a first step, we confirmed the deletion by qPCR using primers for four CNVR specific probes (#15, #18, #21, #23; Figure 17) and showed significant difference ($p < 0.05$) in copy numbers between H4-1987 and the Thoroughbred reference.

Next, the copy numbers were compared between 38 RAO cases and 23 controls (Table 13; note that DNA for one RAO case and one control was finished) using primers for 8 array probes. While clear differences were observed between individual horses in both groups, there was no significant difference in copy numbers between cases and controls (Figure 18).

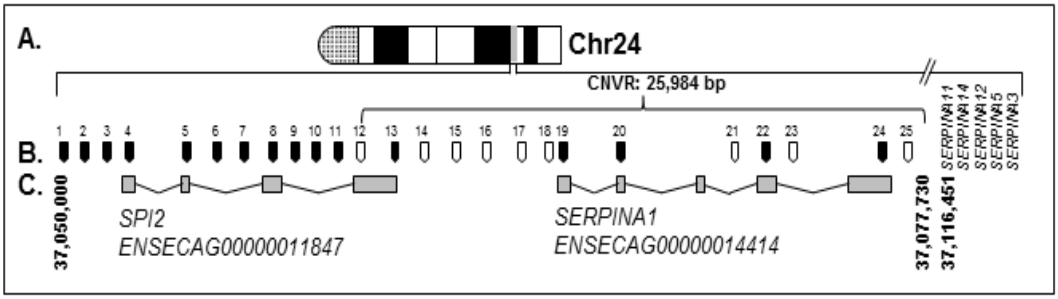


**Figure 16: Schematic diagram of a region in chr24.**
The region flanking the CNVR and serpin gene cluster A. Cytogenetic location of the region (grey line) in Chr24; B. Arrows showing the location of PCR primers #1-#25 (Table 3.2; open arrows correspond to tiling array probes); brackets denote the span of a 25,984 bp RAO specific CNVR; C. Schematic of the genomic structure of SPI2 and SERPINA1 (Ensembl).

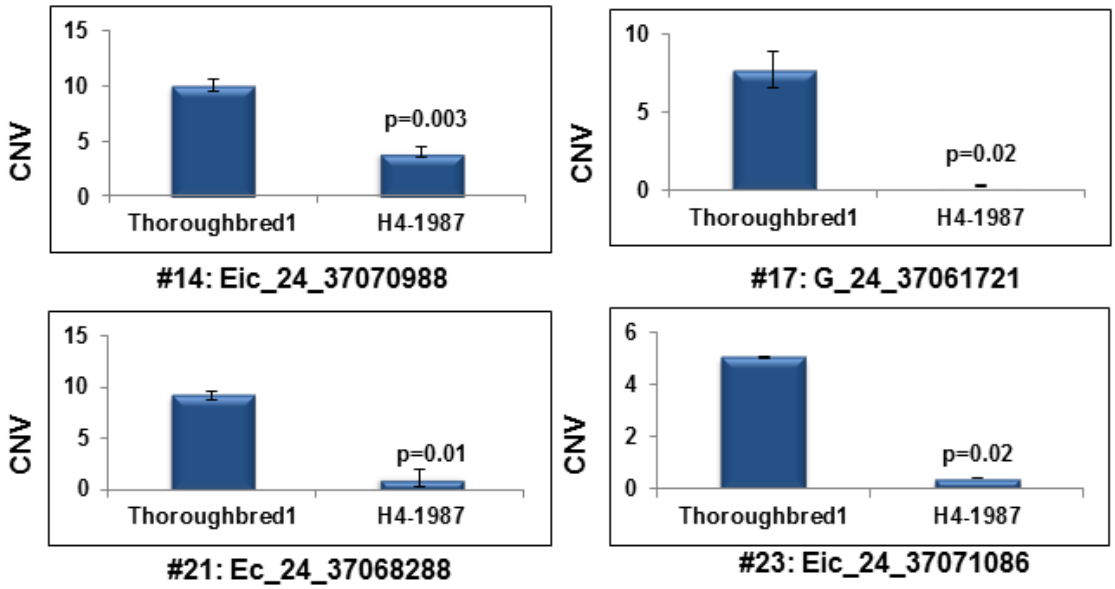


**Figure 17: Validation of the deletion in chr24 by qPCR.**
qPCR was performed in RAO affected Swiss Warmblood H4-1987 and the Thoroughbred reference.

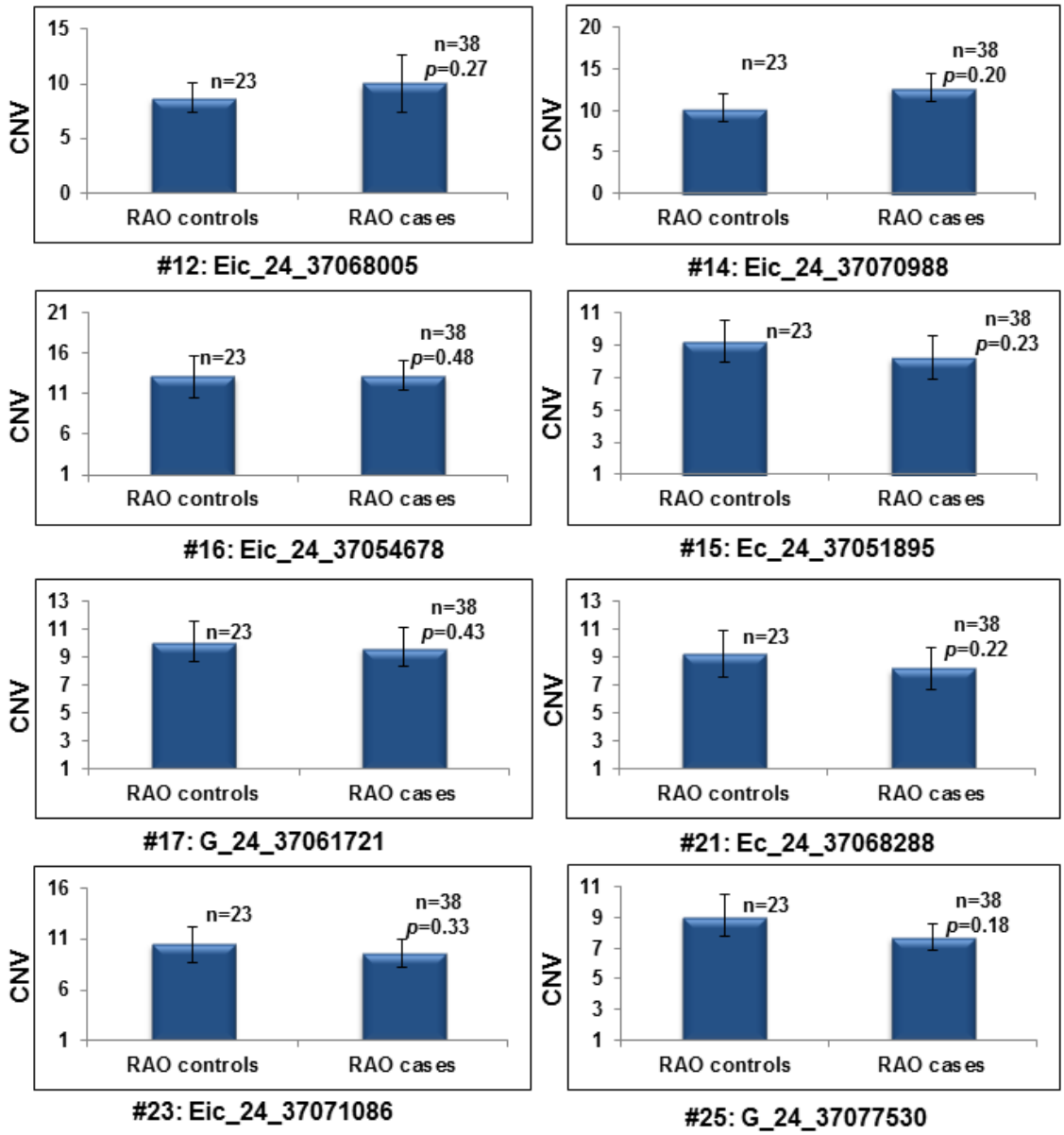


**Figure 18: CNV analysis in chr24 by qPCR using array probe specific primers.**

The array probes in chr24 region of interest were mainly intergenic or intronic (Figure 16), thus we designed primers for all *SERPINA1* and *SPI2* exons, all *SPI2* introns and flanking regions (Figure 16). While optimizing the primers by regular qualitative PCR, we noticed that primer #13 from *SPI2* exon 1 (ENSECAE00000074130; Figure 16) amplified differently from individual horses. This difference was expressed as a very strong, medium, or weak band in gel electrophoresis (Figure 19). We conducted the analysis twice on the entire study cohort and confirmed that these differences were consistent and repeatable, although the CNV was present in cases and controls. No similar variability by qualitative PCR was observed for any other *SPI2* or *SERPINA1* exons.

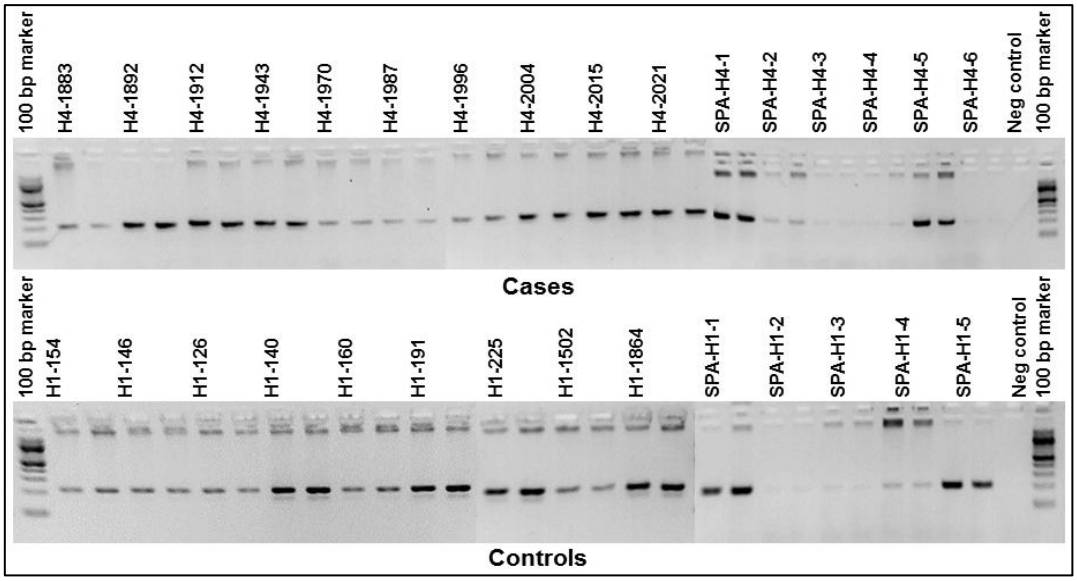


**Figure 19: Regular (qualitative) PCR with *SPI2* exon 1 primers.**
(primer #13; (ENSECAE00000073450) showing CNV in both cases and controls. Each horse was tested in duplicate (2 adjacent lanes per horse).

Copy number variation in *SPI2* exon 1 was further refined by qPCR (Figure 20) showing outstanding differences between individual horses but, again, the variation was present in both RAO cases and controls.

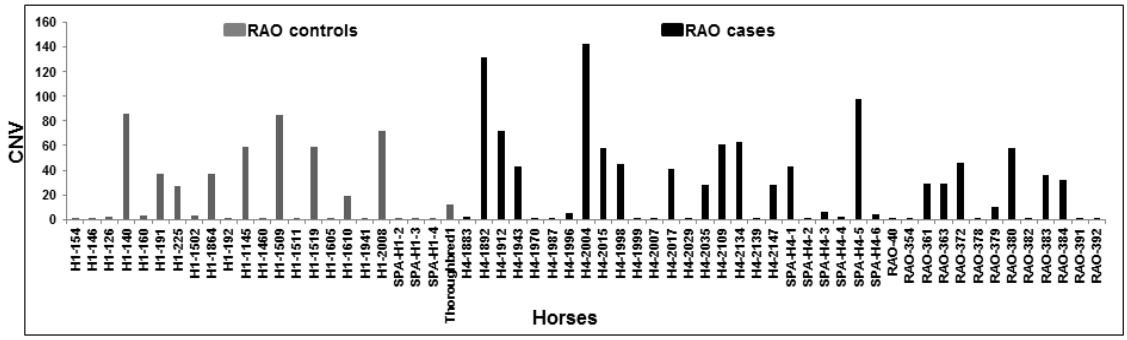


**Figure 20: Analysis of CNV in *SPI2* exon 1.**
(primer #13; ENSECAE00000073450) by qPCR in all RAO cases and controls

Finally, all *SPI2* and *SERPINA1* exon specific primers were tested by qPCR in the study cohort and no significant differences between the two groups were found (Figure 21). We conclude that the region in chr24 containing serpin genes is copy number variable in horses but this variation is not associated with the RAO phenotype.

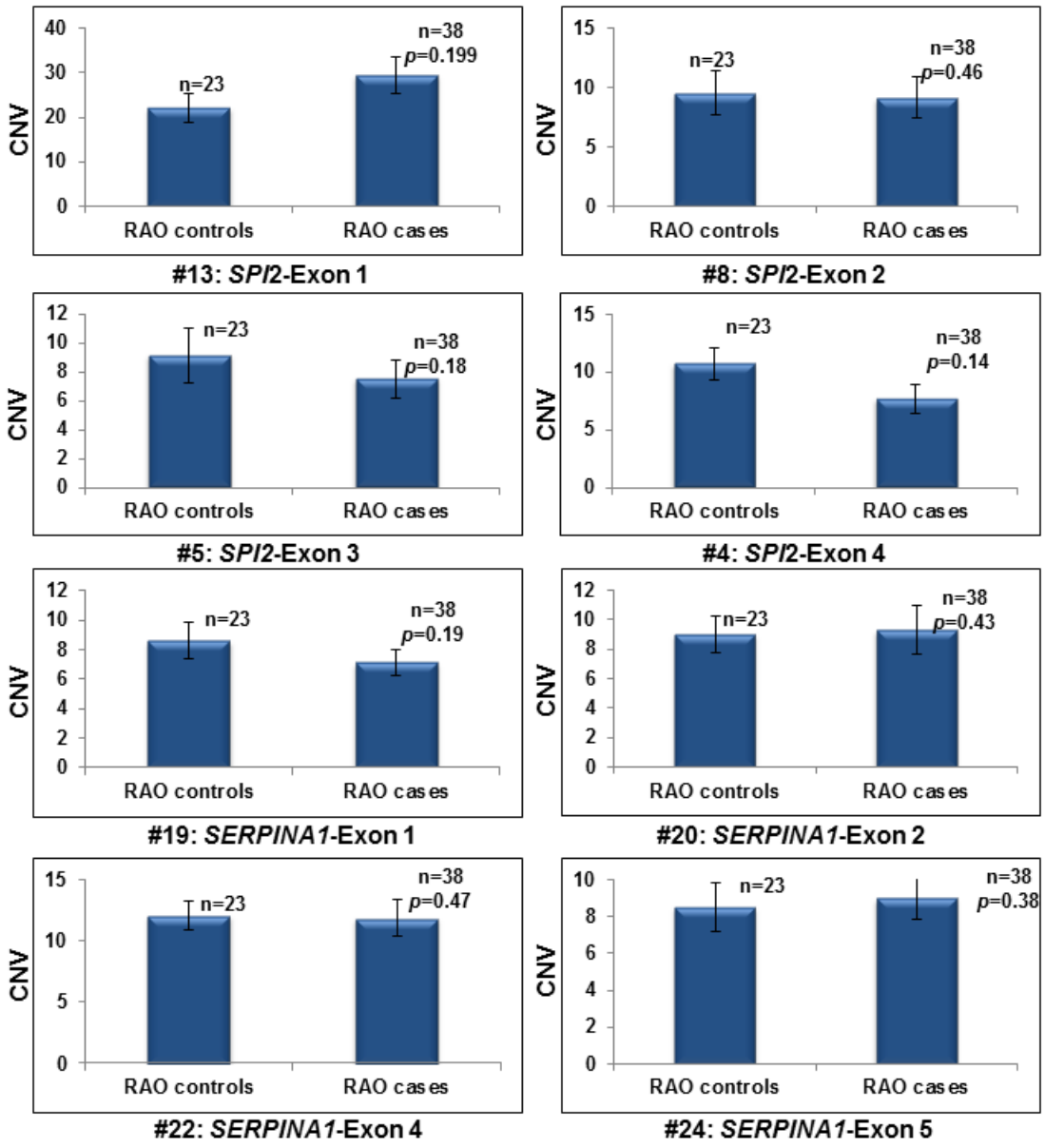


**Figure 21: Quantitative PCR with *SPI2* and *SERPINA1* exon primers.**

## Discussion

Substantial advances have been made in recent years in the discovery and analysis of CNVs in the horse genome by studying populations of normal horses ((Doan et al. 2012a; Doan et al. 2012b; Dupuis et al. 2013; Metzger et al. 2013a; Wang et al. 2014b) and Chapter II). Altogether, over 1400 CNVRs have been identified in 41 diverse breeds (Chapter II). This dataset was effectively used in this study for the analysis of CNVs in RAO affected and control horses, and the detection of CNVs that were present only in RAO affected horses. Furthermore, as the CNVs in RAO control group largely overlapped with those already known for the horse, array CGH for similar studies in the future can be conducted in cases only and compared with the composite dataset.

Despite this progress, very little is known about CNVs that might influence complex equine traits or susceptibility to complex diseases. Recently, CNVs were studied in recurrent laryngeal neuropathy (RLN), common upper airway pathology in the horse (Dupuis et al. 2012). The authors used SNP genotyping data for CNV discovery but found no association between CNVs and RLN, although GWAS on the same data had previously identified chr 21 and 31 as candidate regions for RLN (Dupuis et al. 2011).

The present study is, to our best knowledge, the second attempt to detect CNVs that might influence the risk to complex diseases in horses, and the first systematic evaluation of CNVs in equine RAO. Although this study and the one for RLN (Dupuis et al. 2012) differ regarding the phenotype, CNV detection methodology and the size of

study cohorts, a few similarities are worth mentioning. Both phenotypes were equine complex respiratory disorders and both studies detected a few functionally relevant candidate CNVs, but no association to the disease. In RLN the candidate regions by GWAS did not overlap with candidate CNVs (Dupuis et al. 2011, 2012). Likewise, the 30 putative RAO specific CNVRs in this study (Table 16) did not overlap with the known RAO associated regions in chr13 and 15 as revealed by GWAS (Swinburne et al. 2009; Shakhsi-Niaei et al. 2012). Horse chr13 showed no study specific (Appendix 3.3) or RAO specific (Table 16, Appendix 3.4) CNVRs, although a quantitative trait locus (QTL) including the interleukin 4 receptor (*IL4R*) gene, one of the candidate regions for RAO, maps to this chromosome (Shakhsi-Niaei et al. 2012). The only CNVR in chr15 was a small deletion in one RAO affected horse and located at 79 Mb, thus outside the candidate region at 40-60 Mb as detected by Swinburne and colleagues (2009).

The absence of overlap or close proximity between SNPs and CNVs in RLN (Dupuis et al. 2011, 2012) likely signifies the known limitations of using SNP arrays to tag multiallelic and complex CNVs (Redon et al. 2006, Beckman et al. 2007). Due to this, it has been recommended to combine GWAS with array CGH (Beckman et al. 2007). However, in the case of RAO, different platforms were applied for GWAS (Swinburne et al. 2009; Shakhsi-Niaei et al. 2012) and for CNV discovery (this study). No intersection between SNP and CNV data might indicate that different mechanisms regulate the many genomic loci contributing to RAO.

Here we identified 30 CNVRs that were exclusive to RAO affected horses, although these CNVRs were predominantly private (present in one animal) and could

not be associated with the disease phenotype (Table 16). Prevalence of private CNVRs can have several explanations. First, these might be just novel CNVRs for the horse genome and not related to RAO. Indeed, 18 out of 30 RAO specific CNVRs were found in breeds that are poorly (Swiss Warmblood) or not at all (Missouri Foxtrotters, Tennessee Walking horses) represented in the composite CNVR dataset of 41 horse breeds (Chapter II, Table 12). However, 11 (37%) RAO specific CNVRs were found in an affected Quarter Horse SPA-H4-2 (Table 13, Table 16) – a breed well represented in the CNVR dataset. This suggests that at least some of the RAO specific CNVRs might be related to the condition and were found in single individuals due to the heterogeneous nature of RAO (Swinburne et al. 2009), or because of a small study cohort, or both. Finally, it is also possible that a few of these CNVRs are false positives. Therefore, all RAO specific CNVRs will be validated by qPCR and/or FISH in the nearest future.

Although we did not identify CNV association with equine RAO, recent findings in humans (Rogers et al. 2013) do not exclude CNVs as factors modulating susceptibility to complex diseases, such as asthma. It is just not yet clear how. Similarly to our study, the vast majority of CNVs in human asthma candidate genes were rare (< 5%) and not statistically significant - this despite the involvement of 1212 subjects. Also, there was no intersection between CNVs and previously detected asthma-associated SNPs (reviewed by Akhabir et al. 2011) – a situation similar to RLN (Dupuis et al. 2011, 2012) and RAO (this study, Swinburne et al. 2009, Shakhsi-Niaei et al. 2012) in horses.

On the other hand, analysis of over 270 asthma candidate genes showed that 69 CNVs mapped to within 50 kb of 58 (21%) asthma genes, whereas in six cases the

candidate genes resided within the CNVR boundaries, and modest association was found with two genes – *NOS1* and *SERPINA3* (Rogers et al. 2013). The latter was of interest regarding RAO because a CNVR in one affected Swiss Warmblood horse involved two *SERPIN* genes - *SPI2* and S*ERPINA1* (Table 16), whereas several other serpin family members including *SERPINA3* were in close vicinity (Figure 16). These genes are appealing candidates for RAO because of the known association of *SERPINA1* with pulmonary health in asthma (Thun et al. 2013) and COPD (Denden et al. 2010; Saunders et al. 2012; Thun et al. 2013). The gene encodes for alpha1-antitrypsin which is a highly polymorphic protein, and its deficiency predisposes human adults to lung emphysema and chronic liver disease (Denden et al. 2010). Although we were not able to associate *SERPINA1* with RAO, we showed that the equine genes for alpha1-antitrypsin are highly copy number variable. It is therefore surprising that the CNVR in horse chr24 involving *SPI2* and *SERPINA1* was not detected by any of the previous CNV studies in the horse (Doan et al. 2012a; Doan et al. 2012b; Dupuis et al. 2013; Metzger et al. 2013a; Wang et al. 2014b), including ours (Chapter II). Partially, this can be explained by difficulties to design array probes in a region containing a cluster of genes that share sequence similarity. Indeed, only one array probe was designed in *SPI2* (Figure 16). However, this does not explain why NGS (Doan et al. 2012b) or SNP beachips (Metzger et al. 2013, Dupuis et al. 2012) did not detect the CNVRs, or why we found it now and not before. It is plausible that the region is variable in particular breeds, like Swiss Warmblood, which has been used in only one previous study (Chapter II).

Besides the serpins, the other genes of most functional interest among the RAO specific CNVRs were probably *BTNL2* and *FUT9* (Table 16). Butyrophilin-like 2 (*BTNL2*) belongs to MHC class III, is functionally associated with class II genes in regulating T-cell proliferation, and is involved in immunity disorders such as sarcoidosis related lung inflammation and rheumatoid arthritis (Morais et al. 2012; Aigner et al. 2013; Mitsunaga et al. 2013; Wennerstrom et al. 2013). Fucosyltransferase 9 (*FUT9*) is needed for terminal fucosylation of important molecules (KID, BRN) in innate and acquired immunity (Comelli et al. 2006) and was the only CNV gene that showed up in two RAO affected horses (Table 16). Detailed analysis of the two genes in the entire RAO study cohort of 63 horses is underway. It would also be worth to explore the possibility of somatic mosaicism (Pham et al. 2014) and check whether the CNVRs of interest in this study are present in tissues other than blood. Maybe this answers also why so few CNVRs were found in DNA isolated from hair follicles (Table 15). Future studies should also investigate whether copy number differences in functionally relevant genes to RAO are correlated with different gene expression. Last but not least: the search for CNVs in RAO identified 48 new CNVRs in the horse, thus contributing to the composite horse CNV dataset and to better understanding the role of these variants in equine biology.

# CHAPTER IV

# IDENTIFICATION OF CNVS IN EQUINE DISORDERS OF SEXUAL DEVELOPMENT (DSDS) – CRYPTORCHIDISM AND XY *SRY*-POSITIVE SEX REVERSAL

## Introduction

Disorders of sexual development (DSDs) and reproduction are not uncommon in horses and involve a broad variety of conditions that affect sex determination, sexual differentiation, the development and function of gonads, and fertility (Lear and Bailey 2008; Villagomez and Pinton 2008; Villagomez et al. 2009). Among these, probably the most common are cryptorchidism and sex reversal syndromes (Villagomez and Pinton 2008; Villagomez et al. 2009; Raudsepp et al. 2010; Raudsepp et al. 2013). Both are complex disorders involving non-genetic and genetic factors and both negatively affect the fertility and performance of horses. While etiology of DSDs largely remains unclear, contribution of a genetic component is acknowledged for all (Villagomez and Pinton 2008; Villagomez et al. 2009; Raudsepp et al. 2013). Therefore, molecular genetic research is needed to better understand the etiology of the conditions and devise tests for diagnostics and prevention.

Cryptorchidism affects about 8% of full term male foals (Hayes 1986; Amann and Veeramachaneni 2006; Amann and Veeramachaneni 2007; Arighi 2011; Russell and Pollock 2011) and manifests in the failure of one (unilateral) or both (bilateral) testes to

descend into the scrotum. Because undescended testis is typically dysfunctional, bilateral cryptorchids are sterile (Foster and Ladds 2007; Russell and Pollock 2011), while unilateral cryptorchids are usually fertile because of normal spermatogenesis in the descended testis. Nevertheless, surgical removal of a retained testis is indicated, because of increased risk for the development of primary testicular neoplasms (Pratt et al. 2003; Foster and Ladds 2007; Arighi 2011) and because it continues the production of testosterone and induction of stallion behavior (Arighi 2011; Russell and Pollock 2011). Thus, while cryptorchidism may or may not affect stallion fertility, it poses risk on health and performance. At the same time, very little is known about the genetic component of cryptorchidism. It is thought to be heterogeneous and polygenic, although candidate gene studies in humans (Massart and Saggese 2009; Massart and Saggese 2010), mice (Agoulnik and Feng 2007; Agoulnik et al. 2012), pigs (Rothschild et al. 1988), dogs (Pathirana et al. 2010; Zhao et al. 2010; Zhao et al. 2013) and horses (Diribarne et al. 2009) have not revealed any consistent and significant associations with cryptorchidism.

Sex reversal is a situation where the genetic sex (the karyotype, sex chromosomes) disagrees with the gonadal and/or the phenotypic sex. In horses, male-to-female 64,XY sex reversal syndrome is the second most frequent sex chromosome abnormality after X chromosome monosomy (Raudsepp et al. 2010; Das et al. 2012). The affected individuals are genetically male with female-like external genitalia. Gonadal phenotypes vary from underdeveloped and retained testes (like in bilateral cryptorchidism) to underdeveloped ovaries and ambiguous forms in between. One form

of XY sex reversal is associated with various deletions in the Y chromosome, all involving the *SRY* gene (Raudsepp et al. 2010). Another form is *SRY*-positive and is genetically heterogeneous (Switonski et al. 2005; Raudsepp et al. 2010; Lear and McGee 2012; Pujar and Meyers-Wallen 2012). A point mutation in the androgen receptor (*AR*) gene is to date the only known causative mutation for this form, although found only in one affected Quarter Horse family (Revay et al. 2012). Genetic causes of the majority of *SRY*-positive cases of XY sex reversal remain, as yet, undefined.

Recent studies in human DSDs, including sex reversal, hypospadias and cryptorchidism, suggest critical involvement of dosage sensitive genes and genomic copy number variants (CNVs) (Smyk et al. 2007; Ledig et al. 2010b; Li et al. 2010; Tannour-Louet et al. 2010; Tuttelmann et al. 2011; White et al. 2011) encouraging the initiation of similar research in horses.

The aim of this study is to investigate the possible involvement of CNVs in equine cryptorchidism and *SRY*-positive XY sex reversal. Whole genomes of affected horses are studied for CNVs by array comparative genomic hybridization (CGH). The findings are compared with a composite CNV database for normal horse populations (Chapter II), so that variants specific to DSDs can be identified and analyzed.

## Material and methods

Horses and phenotypes

The study involved 12 bilateral cryptorchid horses of various breeds (Table 17). Eight of these had testes retained in abdomen, one in inguinal canal and no information about the position of testes was available for 3 horses. The horses were phenotyped and samples procured with owner's consent at Texas A&M Large Animal Clinic by Dr. Carolyn Arnold.

**Table 17: List of bilateral cryptorchid horses used for the study.**

| **Horse ID** | **Breed** | **Karyotype** | ***SRY*-PCR** | **Cryptorchid phenotype** |
|---|---|---|---|---|
| H291 | Mixed breed | 64,XY | pos | Bilateral, inguinal |
| H304 | N/A | 64,XY | pos | Bilateral, abdominal |
| H324 | Appaloosa | 64,XY | pos | Bilateral, abdominal |
| H349 | Lipizzaner | 64,XY | pos | Bilateral, abdominal |
| H354 | Pony | 64,XY | pos | Bilateral, abdominal |
| H361 | Mustang | 64,XY | pos | Bilateral, abdominal |
| H365 | N/A | 64,XY | pos | Bilateral, abdominal |
| H390 | American Quarter Horse | 64,XY | pos | Bilateral, abdominal |
| H397 | Mixed breed | 64,XY | pos | Bilateral, position n/a |
| H441 | American Quarter Horse | 64,XY | pos | Bilateral, position n/a |
| H451 | American Quarter Horse | 64,XY | pos | Bilateral, abdominal |
| H614 | American Quarter Horse | 64,XY | pos | Bilateral, position n/a |

The study also involved 8 *SRY*-positive sex reversal horses (Table 18) whose phenotypes and karyotypes were available at the depository of The Molecular Cytogenetics and Genomics Laboratory at Texas A&M University, or supplied by collaborators Dr. Teri Lear (The University of Kentucky) and Dr. Sue McDonnell (The University of Pennsylvania). As a rule, the identity of horses remained confidential. The exceptions were Standardbreds H369 (Martha) and H348 (Helen) (Table 18), two elite racers whose problematic sexual identity has been released in press making headlines in New York Times (Finely 2009) and Horse.com (Lear and McGee 2012).

DNA isolation & quality control

DNA was isolated from peripheral blood using standard PCI (Phenol/ Chloroform/ Isoamyl alcohol) method (Sambrook et al. 1989) or Gentra Puregene Blood Kit (QIAGEN, USA) according to manufacturer's protocol. The DNA was cleaned with DNeasy Blood and Tissue kit (Qiagen) with the following modifications to the manufacturer's protocol: during DNA purification, 80% ethanol was used instead of solution AW2, and the final DNA pellet was eluted in water instead of solution AE. The final quality and quantity of the DNA samples were checked by gel electrophoresis on a 1% agarose gel and by Nanodrop spectrophotometry (Thermo Scientific).

**Table 18: List of *SRY*-positive sex- reversal horses used for array CGH.**

| Horse ID | Breed | Karyotype | SRY-PCR | Phenotype of gonads and external genitalia |
|---|---|---|---|---|
| H369 | American Standardbred (Martha) | 64,XY | pos | Male pseudohermaphrodite: rudimentary abdominal testicles and abnormal female external genitalia |
| H348 | American Standardbred (Helen) | 64,XY | pos | Male pseudohermaphrodite: rudimentary abdominal testicles and abnormal female external genitalia |
| H169 | Appaloosa | 64,XY | pos | Normal female external genitalia |
| H252 | Mixed breed | 64,XY | pos | Abnormal female external genitalia |
| H543 | American Standardbred | 64,XY | pos | Female-like sexual behavior; gelding-like external genitalia with penis and sheath |
| H544 | Tennessee Walking Horse | 64,XY | pos | Abnormal female external genitalia |
| H545 | American Standardbred | 64,XY | pos | Female-like sexual behavior; gelding-like external genitalia with penis and sheath |
| H546 | Thoroughbred | 64,XY | pos | Hypoplastic testes with abnormal female external genitalia |

Array comparative genomic hybridization and data analysis

Probe labeling and array CGH experiments were performed according to Agilent Technologies Protocol Version 6.2.1. All hybridizations comprised of a pair of differently labeled probes, one of which was always the reference DNA – a Thoroughbred mare *Twilight* for females and a Thoroughbred stallion *Bravo* for males.

Briefly, 1µg of purified genomic DNA was digested with restriction enzymes *RsaI* and *AluI* for 2 hours at 37°C and 20 minutes at 65°C to produce 200-500 bp fragments. Cleaved samples were labeled with fluorescence dyes Cy3 (for the reference DNA) or Cy5 (for sample DNA) by random priming using Genomic DNA Enzymatic Labeling Kit (Agilent Technologies). The final labeling reaction contained random primer (5 µl), 1X dNTP mix (5µl), 1X reaction buffer (10 µl), Exo-Klenow enzyme (1 µl), and Cy3- or Cy5-dUTP (3 µl) and was conducted for 2 hours at 37°C, 10 minutes at 65°C. The products were cleaned with 30 kDa filters (Amicon) and the yield and specific activity of labeled DNA was determined with a Nanodrop spectrophotometer. Typical yield for 1 µg of starting DNA was 6-8 µg; specific activity for Cy3 was 25-40 pmol/µg and for Cy5 20-35 pmol/µg. The hybridization mixture was prepared using Agilent Oligo aCGH Hybridization Kit and contained equal quantity of Cy3 and Cy5 labeled probes, 1 µg/µL horse Cot1 DNA, 10X blocking agent, and 2X Hi-RPM buffer. Denatured and pre-annealed probe mixture was applied onto gasket slide, placed in Agilent SureHyb hybridization chamber, 'sandwiched' with an array slide and incubated in Agilent hybridization oven at 65°C for 40 hours. The array slides were washed with Agilent aCGH Wash Buffers 1 and 2, and dried with Acetonitrile and Stabilization and Drying Solutions. Washing in Wash Buffer 1 was done for 10 minutes instead of the 5 minutes recommended by the Agilent protocol.

Data analysis was done using the Agilent Genomic Workbench software (v5.0). The CNVs were represented by gains and losses of normalized fluorescence intensities relative to the reference and identified using the ADM-2 aberration algorithm (Lipson et

al. 2006). Parameters used in this analysis were as follows: Threshold of ADM-2: 6.0; Centralization: ON (Threshold: 6.0, Bin Size: 10); Fuzzy Zero: ON; Aberration Filters: ON (minProbes = 5 & minAvgAbsLogRatio = 0.5 & maxAberrations = 10000 AND percentPenetrance = 0); Feature Level Filters: ON, IntraArray: ON). The CNVs were called with an average $\log_2$ ratio 0.5 over 5 neighboring probes (Chapter II). Output files of CNVs were generated with genomic coordinates, cytoband chromosomal locations, and sequence description showing whether the CNV was present in a gene, intergenic segment, and subtelomeric region or in the Y chromosome. Signal log2 ratio less than -2.0 was considered an indication of a homozygous deletion.

Qualitative and Quantitative PCR

Quantitative PCR on the genomic DNA of cases and controls was used for demarcating breakpoints for deletions discovered by array CGH. Primers for regular PCR were designed using Primer3 software (Table 19) (Koressaar and Remm 2007; Untergasser et al. 2012).

Quantitative PCR (qPCR) was performed to validate selected CNVs. Primers were designed using horse whole genome sequence information (EquCab2), UCSC Genome Browser (Karolchik et al. 2003) and Primer3 software (Koressaar and Remm 2007; Untergasser et al. 2012). The criteria for qPCR primer design were as follows: GC content > 50%; complementarity = 4; product size 100-200 bp, and annealing temperature 55°C - 60°C (Table 19).

The DNA concentration was optimized by serial dilutions and qPCR reactions were carried out with 25 ng/µL of DNA using the SYBR Green PCR kit (Roche) and a Roche LightCycler® 480 (Roche Applied Sciences) according to the manufacturer's instructions. The CNVs were tested in respective cases with the controls in triplicate reactions in duplicate experiments, and normalized to reference gene *GAPDH* (representing autosome with no CNVs in a ~ 1 kb window). Quantification of the copy number was carried out using the comparative Ct method ($2^{\Delta\Delta}$Ct) (Livak and Schmittgen 2001). Cut-off threshold for statistical significance was set at $p<0.05$.

Fluorescence *in situ* Hybridization (FISH)

Selected CNVs were validated by fluorescence *in situ* hybridization (FISH). The CHORI-241 equine BAC library (CHORI-241) was screened by PCR using CNV-specific primers. The corresponding BAC clones were identified and their DNA isolated as described elsewhere (Gustafson et al. 2003). The BAC DNA was labeled with biotin-16-dUTP or digoxigenin-11-dUTP using Biotin- or DIG-Nick Translation Mix (Roche) and hybridized to metaphase and interphase chromosomes. Images for a minimum of 10 metaphase spreads and 20 interphase cells were captured for each experiment and analyzed with a Zeiss Axioplan2 fluorescent microscope equipped with Isis V5.2 (MetaSystems GmbH) software. Cell cultures, chromosome preparations, BAC library screening, BAC DNA isolation, FISH, microscopy and image analysis were carried out according to our standard protocols (Raudsepp and Chowdhary 2008a).

**Table 19: Primers for qualitative and quantitative PCR.**

| Primer/ Probe Name | Forward 5'-3' | Reverse 5'-3' | Product size, bp |
|---|---|---|---|
| **Primers for Chr29 deletion** | | | |
| **Start of deletion** | | | |
| S1 | GTAAGCCTGGGACCGATGTA | TGCTCCCCAAAGTCATTTTC | 789 |
| S2 | TGGTCTGTCTTCCCTCAACC | GGAAGATGGGGTAGGAGAGG | 808 |
| S3 | CAGGTGCTTCTGTGTCTCCA | TAGTGGCAGCCCTCAGAGAC | 718 |
| S4 | GTTTCAGGTAGCTGCCCAAG | TGAGTCTCTGTGGCCCTCTT | 760 |
| S5 | GCAAGGTCAGGAGCAGTAGG | TGAAGAAAGCCTGCCTGTTT | 831 |
| S6 | GCTGTGATGCAGGAAGAACA | CATCGCTCATTCAGGAGACA | 732 |
| S7 | GTAACCACAGGTGGGTGGAC | TGTGGTGAGTTAGGGGGAAA | 711 |
| S8 | CAGCATCGAGCACTGAAGAA | GGGGTCTGGCATACATGAAC | 761 |
| S9 | GAAGACTGCAGTGCCAGTGA | TGCACGTCCCTGTAGAAGTC | 705 |
| S10 | CACCCCCAGTTAGAGAGCTG | GCAGGGTGGATGGGTTAGT | 863 |
| S11 | TATGTGAAGCCCTCCCTGAC | CAAGCCCCCAGTACACAGTT | 705 |
| **End of deletion** | | | |
| E1 | TGAGACAGGGATTGGAAAGG | TGAGCCCTCTCTTTCCTTCA | 727 |
| E2 | TGTCACCCCCATATGGAATC | GCCACATAAAAATCGAGTGC | 709 |
| E3 | TGGAGACAAAAGCACCATTG | AGTCCCTGCCAAGAACTCAA | 779 |
| E4 | CTGGTTGCTTTCTTGACAGAAG | CTGGATATGGAATCTGCCTTT | 182 |
| E5 | ACATCAGCATGGGTTCCTTC | TGTCCAGGGTGATTTGTTCA | 773 |
| E6 | AGCAGCTCAGGGGATCACTA | TGGGCCATTATTGTGTCAAA | 736 |
| E7 | GTTGTAACCACTGGGCCTTC | AATTAGCGAACGCCAAGAAT | 701 |
| E8 | CAAATCACCCTGGACATTGA | TCCTGTGTCAGAGGCAGCTA | 923 |
| E9 | GGCAACCTAAGACAGGCATC | AGCGTCCCATTAACAAACCA | 814 |
| E10 | TTTGGGGAGCTCTCAATCAG | GGAGCAGTAGGCAAGACAGG | 806 |
| E11 | TTAACGACTTGGGGAACAGC | AACATTCCATTTGCCTCAGC | 750 |
| E12 | AAAGCCGCCTTCTAAACCAT | CGGGAGAGTCTTTCCACAAA | 871 |
| **qPCR primers for Chr29 deletion** | | | |
| G_29_28703419 | TAGGTTCGTTTGGGCTTCAC | ACAGGCTGGTCTGGATGACT | 159 |
| G_29_28711378 | CATCTCCTTCATTAAGCTGTGAAC | ATTCTATGGGGCGGGACTT | 84 |
| G_29_28730858 | CTCTCCCTTCGTTTTCATCG | TTCCTGTGGACAGCCTTTCT | 77 |
| Ei_29_28791319 | CTTGTGGTCAATCCCAGACA | TTCCTGTGGACAGCCTTTCT | 166 |
| **qPCR Primers for CNVR in PAR region** | | | |
| AKAP17A- Exon1 | GCTTCTCCGACATCCTCAAG | TCCCCAAACTTCTCGAACAC | 235 |
| AKAP17A- Exon4 | GAAACCTGAACGGGAGTGTG | GGTTGCACTTGTCTTGCTCA | 213 |
| ASMT-Ex 2 | GAGCTGGGGGTGTTCGAC | TTCCTCTCCTCGTGTCCACT | 154 |
| ASMT-Ex 8 | CGTCCTGGTGGTCGAAGG | CAGGACGGCGTGGTACAG | 193 |

**Table 19 continued.**

| Primer/ Probe Name | Forward 5'-3' | Reverse 5'-3' | Product size, bp |
|---|---|---|---|
| **qPCR Primers for CNVR in PAR region** | | | |
| ZBED1-Ex A | GACAACCAGAACCGCTCCTA | CGAACACCTTGTTGCTGATG | 190 |
| ZBED1-Ex B | CGAGATCGACATGTTCCTCA | TAGCTGCCGTCCTTGATCTT | 150 |
| XG- Ex 9 | CGATCTTGCTTTTCGACTCC | ACAATTCCTCCTCCTGTTGC | 129 |
| GYG2-Ex 1 | CAGTGTCTGACCAGGCCTTT | GTGATCAGCACCACCAGCTT | 121 |
| Gs_X_55441 | CTTTTGATGAGGTGGGCAAC | ATTCAGAGCCCGAGATGAGA | 174 |
| Gs_X_66198 | GAACGGAATGTCGCTCTTTC | TGAGTTTCCAGATGGCTCCT | 168 |
| Gs_X_72077 | GGAATCAGCCAGGTGAGTGT | TAGAGGCACGGCTGTTATGA | 167 |
| G_X_133651 | ACGGGGACCGAAAAATCTAC | GGAGGGAATTCACACCTTGA | 188 |
| Gs_X_147232 | GTCAGGTTGTGGGGAGAAAA | AATCCAAGAGGAACCCAACC | 176 |
| **qPCR Primers for sex-reversal and cryptorchid specific common CNVRs** | | | |
| Eic_8_13132401 | ACAGCTTGGCTCTTGGTCAT | CACCCAGAAAAGGGAACAAA | 159 |
| Eic_28_18836887 | TCATCTTACGGCCTGCTACC | GTTCAAGATCTCCCAATTACAGG | 75 |

Gene content analysis of CNVs

The Ensembl annotated horse gene list with their genomic coordinates in Ensembl-73 was added to Agilent Genomic Workbench as a custom track to identify CNVs involving genes. Regions that did not contain Ensembl horse genes were manually analyzed in UCSC genome browser and NCBI to find similarity to known mammalian genes, mRNAs or cDNAs. Gene Ontology analysis (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the genes present in CNVs were performed using DAVID bioinformatics tool with default settings (Huang da

et al. 2009b; Huang da et al. 2009a) to determine enrichment for particular biological processes. Additional search for gene functions was conducted with NCBI Gene and GeneCards.

## Results

### CNVs in bilateral-cryptorchid horses

Array CGH analysis in 12 bilateral cryptorchid horses resulted in 330 CNV calls (Appendix 4.1.1) with an average of 27.5 calls per individual (Table 20). Overlapping and adjacent calls were arranged into 130 CNVRs located across all autosomes, the X chromosome, and chrUn (Appendix 4.1.2). In order to identify CNVRs present specifically in cryptorchid horses, the 130 CNVRs were aligned with the composite dataset of 1476 CNVRs (Chapter II) and the CNVRs identified in RAO affected horses (Chapter III). Ultimately, 16 CNVRs were present in bilateral cryptorchid horses only, and were considered as novel and putatively cryptorchid specific (Table 21., Appendix 4.1.3). However, only one small gain in chrUn was shared between two cryptorchid Quarter Horses (H441 and H614), while the remaining cryptorchid specific CNVRs were private, *i.e.*, present in one horse only (Tables 20, 21; Appendix 4.1.3).

**Table 20: Summary of CNVs in the 12 bilateral cryptorchid horses.**

| Horse ID | Breed | CNVs | Gains | Losses | Novel CNVRs |
|---|---|---|---|---|---|
| H291 | Mixed breed | 40 | 20 | 20 | none |
| H304 | N/A | 33 | 19 | 14 | 2 |
| H324 | Appaloosa | 24 | 11 | 13 | none |
| H349 | Lipizzaner | 45 | 23 | 22 | 3 |
| H354 | Pony | 43 | 16 | 27 | 4 |
| H361 | Mustang | 31 | 9 | 22 | none |
| H365 | N/A | 31 | 11 | 20 | 2 |
| H390 | American Quarter Horse | 8 | 1 | 7 | none |
| H397 | Mixed breed | 10 | 4 | 6 | 1 |
| H441 | American Quarter Horse | 21 | 3 | 18 | 1; shared with H614 |
| H451 | American Quarter Horse | 17 | 2 | 15 | none |
| H614 | American Quarter Horse | 27 | 9 | 18 | 2; 1 shared with H441 |
| **Total** | | **330** | **128** | **202** | **16** |
| **Average** | | **27.5** | | | |
| **Median** | | **29** | | | |

Analysis of the 16 cryptorchid specific CNVRs for gene content and functions showed that all CNVRs, except those located in chrUn, involved known genes, cDNAs or mRNAs (Table 21). Functions of the genes were variable and related to cell-cell interactions, cell defense, and various cellular functions, such as transcription, DNA repair, and apoptosis. Two CNVRs harbored genes that could be very conditionally

related to reproduction: a loss in chr31 containing *STXBP5* which is involved in membrane fusions and calcium-dependent exocytosis, thus possibly in sperm acrosome reaction, and a loss in chr11 involving *ZNF599* – a transcription factor which mutations have been associated with hypospadias (Table 21).

Seven CNVRs that were found only in cryptochid horses had orthologous variable sequences in the human genome (human genome variant database), suggesting that these variants are a normal part of mammalian genomes and might be more prevalent in horses as well. Four CNV genes corresponded to Mendelian disorders in man and one in dogs, although the disorders were related to nervous sytem and anemia (Table 21).

Taken together, there was no associaotion between the bilateral cryptochid phenotype and the CNVRs, and none of the CNVRs specifically found in cryptorchid horses involved genes with known functions in development or reproduction.

## Table 21: Novel CNVRs in bilateral cryptorchid horses.

| Position, chr:Mb | Type | Horse ID | Gene symbol | Description | DGV | OMIM | OMIA |
|---|---|---|---|---|---|---|---|
| 1:24.0 | Loss | H354_bi crypt | Pseudogene | N/A | N/A | N/A | N/A |
| 4:0.9 | Loss | H614_bi crypt | GNAO1 | Increase or decrease in cAMP levels | nsv833242 | 615473. Epileptic encephalopathy, early infantile, 17; EIEE17 | N/A |
| 4:56.6 | Gain | H304_bi crypt | JL632581 | Horse mRNA | N/A | N/A | N/A |
| 8:74.4 | Loss | H354_bi crypt | WDR7 | Amelogenesis imperfecta, and alcohol dependence. | nsv527474 | N/A | N/A |
| 8:88.0 | Gain | H304_bi crypt | JT478681 | non-Horse mRNA | N/A | N/A | N/A |
| 11:172 bp | Loss | H365_bi crypt | ZNF599 | Transcriptional regulation. Diseases: Hypospadias, and mental retardation | dgv1074e1 | 606937. Spinocerebellar ataxia, autosomal recessive 5; SCAR5, ***ZNF592*** | N/A |
| 20:9.7 | Gain | H349_bi crypt | JL631755 | Horse mRNA | N/A | N/A | N/A |
| 20:19.7 | Loss | H397_bi crypt | RPL5 | Diseases: Diamond-Blackfan anemia, and Diamond-Blackfan anemia. | esv32853 | 615550. Diamond-Blackfan anemia 12; ***RPL15*** | OMIA 000041-9615 UID:119 Diamond-Blackfan anaemia. Phenotype in dog (Canis lupus familiaris). |
| 29:15.4 | Loss | H614_bi crypt | PLXDC2 | May play a role in tumor angiogenesis. Diseases: Sotos syndrome, and endothelitis. | dgv131n67 | N/A | N/A |
| 31:19.1 | Loss | H614_bi crypt | STXBP5 | Regulatory role in calcium-dependent exocytosis and neurotransmitter release. Diseases: type 1 von Willebrand disease, and von Willebrand's disease. | nsv886752 | 612164, epileptic encephalopathy, early infantile, 4; ***Stxbp1*** | N/A |
| Un:0.01 | Gain | H349_bi crypt | N/A | N/A | N/A | N/A | N/A |
| Un:0.05 | Gain | H441_bi crypt, H614_bi crypt | N/A | N/A | N/A | N/A | N/A |
| Un:0.2 | Loss | H365_bi crypt | N/A | N/A | N/A | N/A | N/A |
| Un:0.4 | Loss | H349_bi crypt | N/A | N/A | N/A | N/A | N/A |
| X:85.1 | Gain | H354_bi crypt | TSC22D3 | Protects T-cells from IL2 deprivation-induced apoptosis. Diseases: vascular dementia, and brain edema | N/A | N/A | N/A |
| X:118.4 | Loss | H354_bi crypt | EIF2S1 | Initiation of protein synthesis. Diseases: brain ischemia, and Newcastle disease. | esv2751285 | N/A | N/A |

CNVs in XY *SRY*-positive sex reversal horses

A total of 345 CNV calls were made in the 8 sex reversal horses (Table 22; Appendix 4.2.1). The lowest number of CNVs were detected in H252 (17) and the highest in H543 (88) with a median of 32.5 CNVs per horse. Based on the overlaps, the CNVs were arranged into 172 CNVRs that were distributed on all chromosomes except the Y (Appendix 4.2.2). After comparison of this data with the composite dataset of 1476 horse CNVRs (Chapter II) and with the CNVRs identified in the RAO study cohort (Chapter III), 41 CNVRs were confined to sex reversal horses only and considered novel (Table 23, Appendix 4.2.3). All DSD horses, except H348, had novel CNVRs (Table 22, 23) of which 8 were shared between two or more horses and the remaining 33 were private (Table 23, Appendix 4.2.3).

Notably, all 41 novel CNVRs involved genes, pseudogenes, or expressed sequences such as mRNAs, cDNAs or ESTs (Table 23), and 15 CNVRs shared genes with the human variation database (Table 23 and Appendix 4.2.3). Analysis of OMIM and OMIA databases showed that 9 novel CNVRs contained genes corresponding to Mendelian conditions in man and one in animals. The latter was a CNVR in chr1 involving the *PTEN* gene which is associated with hydrocephalus in cattle and dogs.

**Table 22: Summary of CNVs in the 8 sex reversal horses.**

| Horse ID | Breed | CNVs | Gains | Losses | Novel CNVRs |
|---|---|---|---|---|---|
| H169 | Appaloosa | 26 | 13 | 13 | 1 |
| H252 | Mixed breed | 17 | 5 | 12 | 1 |
| H348 | American Standardbred (*Arizona Helen*) | 31 | 9 | 22 | 0 |
| H369 | American Standardbred (*Martha Maxine*) | 34 | 9 | 25 | 1 |
| H543 | American Standardbred | 88 | 4 | 84 | 20 |
| H544 | Tennessee Walking Horse | 29 | 1 | 28 | 6 |
| H545 | American Standardbred | 78 | 5 | 73 | 17 |
| H546 | Thoroughbred | 42 | 3 | 39 | 6 |
| **Total** | | **345** | **49** | **296** | **31** |
| **Average** | | **43.1** | | | |
| **Median** | | **32.5** | | | |

The known protein coding genes in novel CNVRs were analyzed for possible functional significance in DSDs with the primary focus on the 8 CNVRs that were shared between affected horses. The majority was involved in functions typical to CNV genes in general population of horses and other mammals (Redon et al. 2006). These include G-protein coupled receptors or olfactory receptors (4), signal transduction (4), cell-cycle regulation (4), and transcription factors (5). Three CNV genes (*BCL2L11*, *DMTN* and *PTEN*) were involved in tumor suppression and apoptosis. Interestingly, a CNVR involving an X-linked stromal antigen *STAG2,* was found in three sex reversal horses. Although the gene has important functions in sister chromatid separation, cell cycle and development, no DSD phenotypes have been as yet associated with *STAG2* mutations.

**Table 23: Summary of genes involved in the 41 novel CNVRs in sex reversal horses.**

| Position | Type | Horse ID | Gene symbol | Function, description |
|---|---|---|---|---|
| 1:40.51 | Loss | H543_SEX REV | *PTEN* | Tumor suppressor by negatively regulating AKT/PKB signaling pathway. |
| 3:57.38 | Loss | H169_SEX REV | *ETF1* | Termination of mRNA translation. |
| 3:72.25 | Loss | H543_SEX REV, H545_SEX REV | *LPHN3* | G-protein coupled receptors (GPCR), cell adhesion and signal transduction. |
| 4:44.10 | Loss | H545_SEX REV | *PHF14* | N/A |
| 4:68.08 | Loss | H545_SEX REV | JL626060 | horse mRNA |
| 4:76.49 | Loss | H543_SEX REV | AK154362 | mouse mRNA |
| 4:81.74 | Loss | H545_SEX REV | AK190472 | mouse mRNA |
| 4:95.02 | Loss | H546_SEX REV | *PRSS2* | T cell receptor |
| 5:22.27 | Loss | H545_SEX REV | pseudogene | N/A |
| 5:77.99 | Loss | H543_SEX REV | *COL24A1* | Fibrillogenesis at specific anatomical locations during fetal development |
| 7:27.96 | Loss | H545_SEX REV | *ARHGEF12* | G protein-coupled receptors. |
| 7:72.35 | Loss | H543_SEX REV | *OR52I1*, *pseudogene*, *OR52I1* | Olfactory receptors |
| 7:72.76 | Loss | H543_SEX REV | *OR51F2*, *OR52R1* | Olfactory receptors |
| 8:54.75 | Loss | H543_SEX REV | *NOL4* | This gene clusters with an RNA gene, lncRNA |
| 9:36.38 | Loss | H543_SEX REV, H545_SEX REV | JV181555 | mRNA, *Junco hyemalis* (Aves) |
| 9:53.30 | Loss | H543_SEX REV | BX647886 | Human mRNA |
| 9:77.98 | Loss | H543_SEX REV | FAM135B | N/A |
| 10:35.77 | Loss | H543_SEX REV | JL628991 | horse mRNA |
| 12:0.62 | Loss | H545_SEX REV | *CSTF3* | Alternative splicing results in multiple transcript variants encoding different isoforms. |
| 14:29.41 | Loss | H544_SEX REV | *TET2* | Defects in this gene have been associated with several myeloproliferative disorders. Associated with anemia, myeloid leukemia |
| 14:66.39 | Loss | H543_SEX REV | JL627555, *DMTN* | Tumor suppressor and inhibits malignant cell transformation |
| 15:14.24 | Loss | H369_SEX REV | *BCL2L11* | Programmed cell death or apoptosis. |
| 16:19.96 | Loss | H543_SEX REV | KF274565 | human mRNA |
| 17:34.06 | Loss | H545_SEX REV | BC034915 | human mRNA |

**Table 23 continued.**

| Position | Type | Horse ID | Gene symbol | Function, description |
|---|---|---|---|---|
| 18:14.06 | Loss | H543_SEX REV, H544_SEX REV | JU909423 | bison mRNA |
| 18:60.93 | Loss | H546_SEX REV | DUSP19 | Congenital cataracts, facial dysmorphism, and neuropathy, and Lafora disease. |
| 19:41.02 | Loss | H543_SEX REV | JU918700 | bison mRNA |
| 20:18.92 | Gain | H252_SEX REV | *E2F3* | Transcription factors. Diseases associated with E2F3 include retinoblastoma, and Wilms tumor |
| 20:36.50 | Loss | H546_SEX REV | *SRSF3* | Affiliated with the lncRNA class. Diseases include open-angle glaucoma, and primary open angle glaucoma |
| 20:52.04 | Loss | H543_SEX REV | JU726520 | bison mRNA |
| 23:27.81 | Gain | H546_SEX REV, H545_SEX REV | *GLDC* | Electron carrier activity and lyase activity. |
| 23:41.23 | Loss | H543_SEX REV, H545_SEX REV | CX600759 | N/A |
| 23:46.90 | Loss | H544_SEX REV | JO239254 | non-Horse mRNA |
| 25:7.30 | Loss | H545_SEX REV | Processed pseudogenes | N/A |
| 25:15.79 | Loss | H543_SEX REV | *MUSK* | Transmembrane receptor protein. Diseases include musk-related congenital myasthenic syndrome, and progressive bulbar palsy |
| 25:22.61 | Loss | H543_SEX REV, H544_SEX REV, H545_SEX REV | JO239254 | non-Horse mRNA |
| 26:5.80 | Loss | H544_SEX REV, H543_SEX REV | *TIGD1* | DNA binding. |
| 26:17.64 | Loss | H545_SEX REV | JL641040 | horse mRNA |
| X:57.61 | Loss | H545_SEX REV | *ATRX* | Diseases include thalassemia, Juberg Marsidi syndrome, and sex reversal. Involved in mammalian sexual development, development of gonads. |
| X:72.17 | Loss | H545_SEX REV | *PCDH11X* | Diseases include schizoaffective disorder, and Cornelia de Lange syndrome. |
| X:98.51 | Loss | H546_SEX REV, H544_SEX REV, H545_SEX REV | *STAG2* | Regulates the separation of sister chromatids during cell division. Diseases include tracheitis and laryngotracheitis |

Functionally the most interesting and relevant among novel CNV genes was the X-linked *ATRX*. Although the gene has mainly been associated with several human X-linked mental retardation syndromes (Picketts et al. 1996), recent studies show that *ATRX* has a critical and evolutionarily conserved role in normal development of testes and ovary (Huyhn et al. 2011). In humans, deletions or mutations in *ATRX* display varying degrees of sex reversal, implicating *ATRX* in the development of the human testis (Biason-Lauber 2010). Therefore, and despite being present in only one sex reversal horse (American Standardbred H545), this CNVR will be subject for further studies.

Besides novel CNVRs, of particular interest was a large, over 360 kb deletion in the pseudoautosomal region (PAR; chrX: 203-366,729) of a mixed breed sex reversal horse H252 (Appendix 4.2.2). A small 35 kb intergenic gain within this region was previously detected in a Caspian pony (chrX: 34,725-70,176; Chapter II) and that is why this CNVR was not included as novel. However, deletion of this magnitude in the PAR was found in horses for the first time. The deletions was probably heterozygous (log = -0.5; Appendix 4.2.1) and included 4 PAR genes tel-*AKAP17A-ASMT-ZBED1-XG* as ordered starting from the telomere, with *GYG2* being immediately proximal to the CNVR (Fig 22).

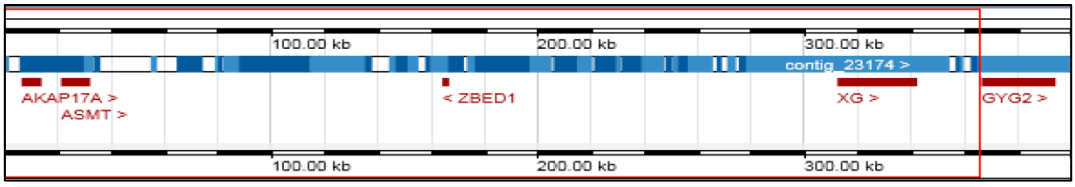


**Figure 22: Ensembl output for chrX:2 03-366,729.**
Showing the gene content of the deletion (red frame) in H252 sex reversal mixed breed horse.

The deletion was validated by qPCR using primers for 5 array probes and the involved genes (Table 19 for primers). Comparison of copy numbers between H252 and *Bravo*, the male reference for aCGH, showed statistical significance for two array probes (GsX_147232, GsX_55441), *ASMT* exon8 and *ZBED1* exon A (Figure 23). Further, copy numbers were compared between H252 and *Twilight*, the female reference, and significant or highly significant copy number differences were observed for three array probes (GsX_72077, GX_133651, and GsX_66198) and *AKAP17A*, *ASMT*, and *ZBED1*.

Significant differences were also present between the two reference horses (Figure 23). We concluded that the CNVR in the PAR of H252 is true and heterozygous. We also concluded that the region is probably more variable among horses than revealed by aCGH and requires further analysis by qPCR, FISH, and/or NGS. The significance of the deletion with regards the sex reversal phenotype, however, remained as yet unclear.

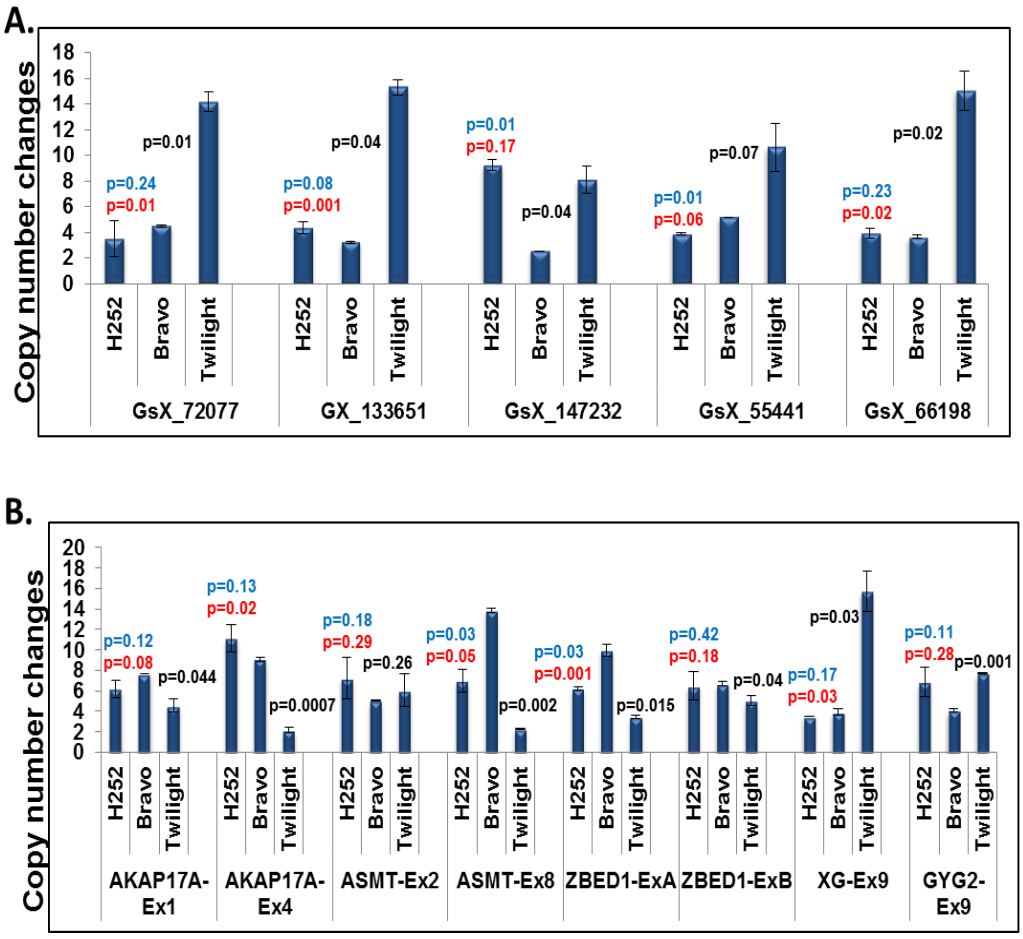


**Figure 23: Validation of the CNVR in the PAR of H252 by qPCR.**
A. ArrayCGH probe specific qPCR and B. Gene exon specific qPCR. *p*-values in blue relate H252 to Thoroughbred male reference *Bravo* and those in red, to Thoroughbred female reference *Twilight*. The *p*-values in black relate the two reference genome

Novel CNVRs shared between bilateral cryptorchid and XY sex reversal horses

Two novel CNVRs, Chr8:13,128,936-13,134,708 and Chr28:18,833,995-18,846,757, were shared between two or more XY sex-reversal and bilateral cryptorchid horses. The CNVRs included 3 genes, all involved at different stages of cell cycle regulation and transcription (Table 24). Validation by qPCR confirmed a statistically significant loss in chr28 in all 5 affected horses (Figure 24B). A significant copy number change was confirmed also for chr8, although qPCR showed a gain and not a loss as aCGH (Figure 24A).

**Table 24: Novel CNVRs present in sex-reversal and bilateral cryptorchid horses.**

| **Position, chr:Mb** | **Type** | **Horse ID** | **Gene symbol** | **Gene name** | **General description** |
|---|---|---|---|---|---|
| 8:13.12 | Loss; 5.7 kb | H390_bi crypt, H451_bi crypt, H369_SEX REV | *MZT2B, TUBA3D* | mitotic spindle organizing protein 2B, tubulin, alpha 3d | Maintain cellular structure, function in intracellular transport; spindle formation during mitosis. |
| 28:18.83 | Loss; 12.7 kb | H390_bi crypt, H369_SEX REV H544_SEX REV, H545_SEX REV, H546_SEX REV, | *UBE2N* | Ubiquitin-Conjugating Enzyme E2N | Mediates transcriptional activation of target genes; control of progress through the cell cycle and differentiation; error-free DNA repair pathway; contributes to the survival of cells after DNA damage. Diseases include Riddle syndrome, and Machado-Joseph disease. |

Thus, further research is needed to elucidate the CNV profile of this genomic region. It is possible that primers for qPCR recognize sequences in other parts of the genome and confound the results. Overall, because the involved genes regulate cell cycle and transcription, their contribution to DSDs is possible but remains elusive at the current status of knowledge about gene functions.

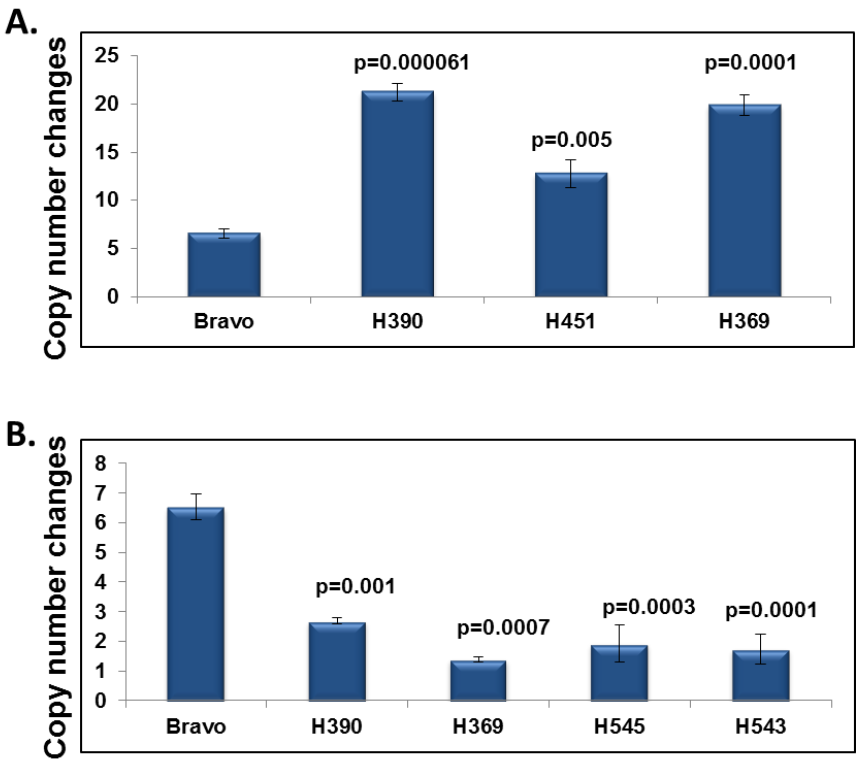


**Figure 24: Quantitative PCR to validate shared CNVRs.**
A. CNVR in chr8; B. CNVR in chr 28 shared between cryptorchid and sex reversal horses.. *Note:* as DNA for H544 and H546 was finished, DNA from H543 was used instead; this horse has a very similar phenotype to H544 and H546 (Table 4. 2).

A homozygous deletion in chromosome 29 involving *AKR1C* genes

In addition to detailed analysis of novel CNVRs, all 130 CNVRs in bilateral cryptorchid horses and the 172 CNVRs in XY sex reversal horses were studied for log2 aberration values to identify putative homozygous deletions of functional significance (Appendices 4.1.4 & 4.2.4). Of particular interest was an approximately 200 kb loss in Chr29:28.6-28.8 Mb showing deeply negative log2 values in one bilateral cryptorchid Pony H354 ($log_{2average}$ = -3.8) and in two XY sex reversal (male pseudohermaphrodites) Standardbreds H348($log_{2average}$ = -3.5) and H369 ($log_{2average}$ = -3.1). These aberration values were significantly lower than those for normal horses ($log_{2average}$ =-0.7), suggesting that the deletion is homozygous. Most importantly, the deletion involved members of the *AKR1C* aldo-keto reductase gene family, known to be critical in the backdoor pathway of dihydrotestosterone (DHT) synthesis and sexual development (Biason-Lauber et al. 2013; Fukami et al. 2013). A schematic overview of the CNVR, including the aberration profiles of all array probes in this region, is presented in Figure 25.

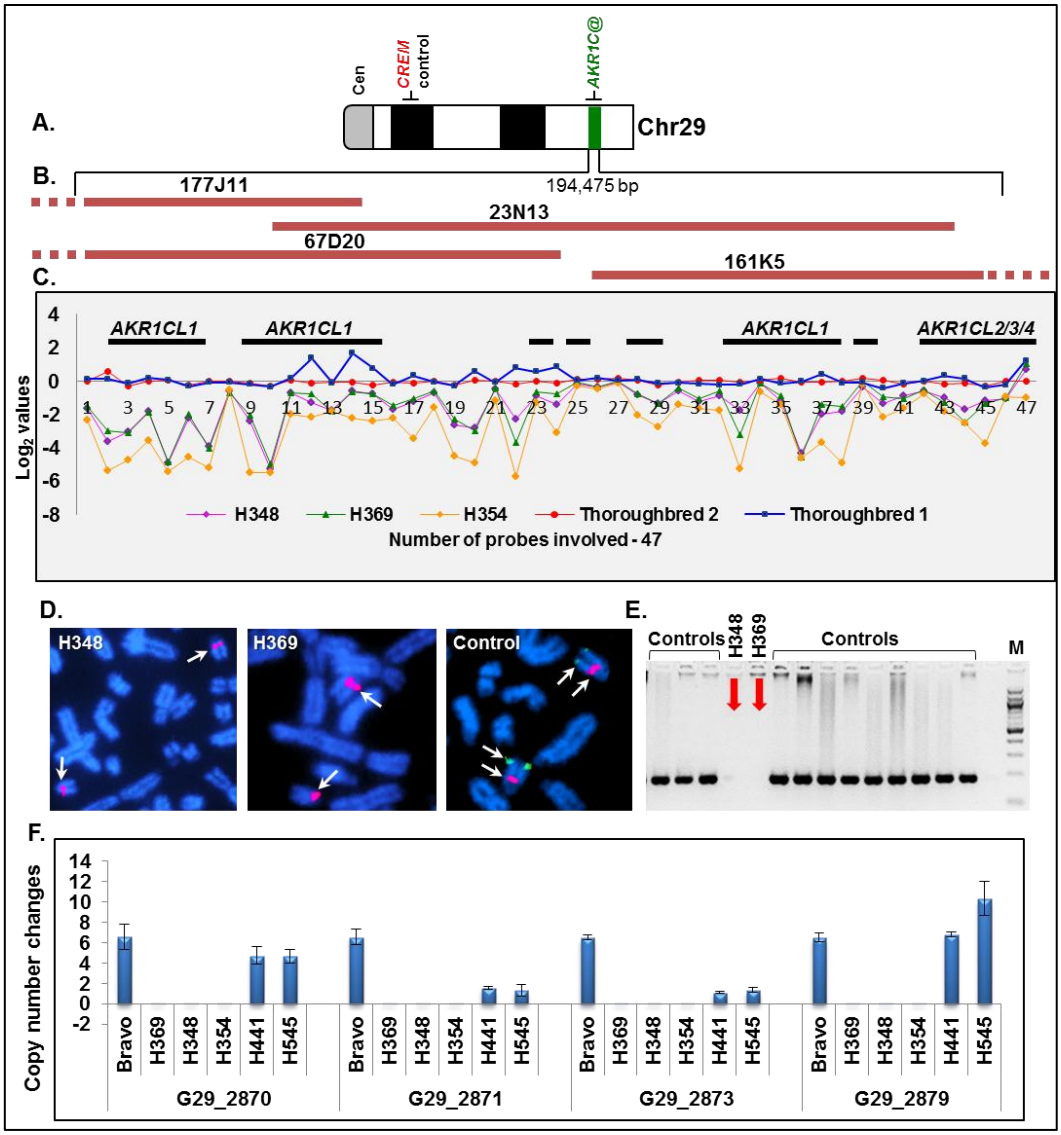


**Figure 25: Schematic of the homozygous deletion in chr29 in DSD horses.**
A. chr29 ideogram showing the location of *AKR1C* gene cluster and a control gene *CREM*; B. Four BAC clones spanning the deletion CNVR (28,640,862-28,835,337); C. Detailed map of the CNVR showing the location of genes (black horizontal bars) and signal $\log_2$ values for 47 array probes in DSD and reference horses; D. FISH results with a CHORI-241 BAC 23N13 spanning the deletion (green signal) and a control BAC 76H13 for *CREM* gene from a non-CNVR (red signal); E. PCR with CNVR-specific primers in the DSD and control horses; F. qPCR with CNVR-specific primers in the DSD and control horses.

Homozygosity of the deletion was confirmed by qualitative and quantitative PCR and by FISH. The genomic DNA of the three DSD horses did not amplify with primers designed inside the deletion (Figure 25D, E), although normal amplification was obtained with primers outside the CNVR. Homozygosity of the deletion was further confirmed by FISH with a BAC clone 23N13 spanning the deletion (Figure 25C). The BAC hybridized to chr29 only in control animals but not in the three DSD horses, whereas a control BAC with *CREM* gene from a non-CNVR in chr29 hybridized equally to chr29 in the DSD horses and controls. The known involvement of *AKR1C* genes in sexual development suggested a causative relationship between the homozygous deletion of this region and the abnormal sexual phenotype.

Because of the likelihood of a causative deletion, it is important to devise tools for the detection of homozygous carriers. Therefore, we designed 23 sets of primers for PCR-walking into the deletion from the left (primers S1-S11) and right (primers E1-E12) to precisely determine the deletion start and end sites (Table 19, Figure 26). According to PCR, putative deletion breakpoints were identified between primers S7 and S8 at the left (Start) and between primers E8 and E9 at the right (End) of the deletion.

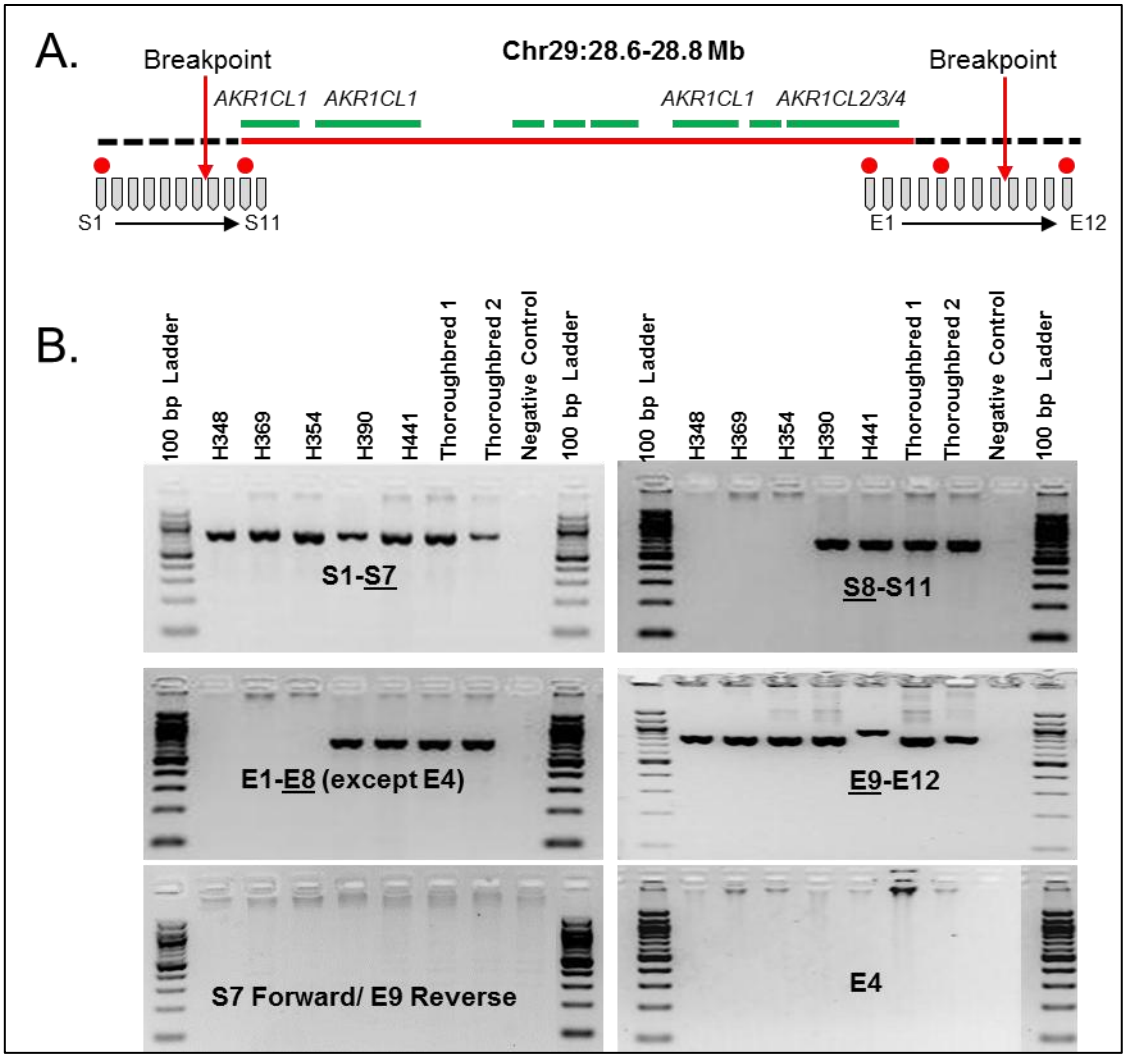


**Figure 26: Breakpoint analysis for the homozygous deletion in chr29.**
A. Schematic of the region showing the deletion (red solid line) as defined by aCGH; green solid bars show the genes involved; grey filled arrowheads S1-S11 and E1-E12 denote the location of PCR primers; arrowhead with red dots denote overlaps of PCR primers with array probes; red arrows indicate the location of deletion breakpoints 1 and 2 as determined by PCR; B. Gel images showing the results of PCR-walking that determined breakpoint-1 between segments S7 and S8, and breakpoint-2 between segments E8 and E9; products of the underlined primers are shown.

This suggested that primers immediately outside the deletion from left and right, *viz.*, S7-forward and E9-reverse, will amplify a small PCR product in the three deletion carriers H348, H369, H354. However, we were not able to amplify any products with these or other primers flanking the deletion (Figure 26B), suggesting that the deletion landscape is more complex, and/or the genome sequence assembly in this region is not accurate. These might also be the reasons why primers E4 did not amplify from any of the affected or control horses (Figure 26B), and why primers E9 amplified a different product in H441 (Figure 26B).

**Discussion**

Here we conducted a genome-wide CNV analysis in horses with two defined disorders of sexual development – bilateral abdominal cryptorchidism and male-to-female XY *SRY*-positive sex reversal syndrome. To our best knowledge, this is the first global analysis of CNVs in XY DSDs in horses and domestic animals, even although cryptorcidism and XY sex reversal have been amply described in several species (Amann and Veeramachaneni 2006; Amann and Veeramachaneni 2007; Villagomez et al. 2009; Raudsepp et al. 2010; Pujar and Meyers-Wallen 2012). Hitherto, contribution of CNVs to various forms of XY DSDs has been to some extent explored only in humans, and clinically relevant gains and losses have been found in known (*DAX1, SOX9, GATA4*) and novel (*FGFR2, KANK1, ADCY2, ZEB2*) candidate genes for gonadal development (Ledig et al. 2010a; Tannour-Louet et al. 2010; White et al. 2011).

An important foundation for the present study was provided by the previous genome-wide CNV discoveries in diverse horse breeds and populations (Chapters II and III, and (Doan et al. 2012a; Doan et al. 2012b; Dupuis et al. 2013; Metzger et al. 2013a; Wang et al. 2014b)). The resulting composite dataset of over 1400 CNVRs for the horse (Chapter II) was a critical resource for evaluating the known CNVRs and identifying those that are novel. In the present study we analyzed the genomes of 20 horses - 12 bilateral cryptorchids and 8 XY *SRY*-positive sex reversals, and identified altogether 302 CNVRs (130 in cryptorchids and 172 in sex reversals (Appendices 4.1.2, 4.2.2 ). Of these, over 80% were shared with previous studies, thus essentially confirming that these genomic regions are CNV in horses. This is important because many of the reported CNVs have been discovered in just a single individual ((Doan et al. 2012a; Wang et al. 2014b), see Chapter II) and require additional supporting evidence.

With regards equine XY DSDs, our primary interest in this study were the remaining 20% of CNVRs that were unique to cryptorchid and/or sex reversal horses and not described in horses before. Such CNVRs were found in both groups of DSD horses, although it was noteworthy that only 16 novel CNVRs were detected in the 12 bilateral cryptorchids (Table 17) compared to 41 in 8 sex reversal horses (Table 18). Furthermore, in cryptorchids only one CNVR was shared between two affected horses while the degree of sharing was 20% in the sex reversal group (Tables 21, 23). Since the cryptorchid cohort comprised 7 breeds and the sex reversals 5 breeds (Tables 17, 18), these differences cannot be explained by breed diversity. It is also noteworthy that the

cryptorchids shared two novel CNVRs with the sex reversal group but only one among themselves.

The two shared CNVRs between the groups involved genes required for protein turnaround and DNA repair (*UBE2N*) and for cell division (*MZT2B, TUBA3D*) (Table 24). These are universal biological processes and likely signify normal variants for the horse, particularly because no DSD phenotypes have been associated with these genes in mice or humans (Safran et al. 2010; Blake et al. 2014). Likewise, the analysis of other novel CNV genes in cryptorchids did not reveal any with functional coherence to male sexual development and testis descent, and none included genes which mutations or CNVs have been associated with cryptorchidism in humans (Tannour-Louet et al. 2010). It is therefore conceivable that CNVs do not contribute to cryptorcidism in this particular study cohort. It is also possible that our limited knowledge about the functions of these CNV genes combined with the genetic heterogeneity of cryptorchidism, counfounded the analysis.

The situation was different in *SRY*-positive sex reversal horses. Despite being phenotypically much more diverse (Table 18) than bilateral abdominal cryptorchidism, more novel CNVRs (8 out of 41, 19.5%) were shared between individuals. Functional relevance of the majority of CNV genes, however, remained unclear, partly because many were transcribed sequences with hitherto uncharacterized functions (Table 23).

Probably the most interesting novel CNVR gene was the transcriptional regulator *ATRX*. Mutations of this gene in humans cause the *ATRX*-syndrome which is characterized by complex developmental disorders including genital and gonadal

abnormalities (Picketts et al. 1996; Huyhn et al. 2011). Due to *ATRX* mutations or deletions, genetically males (XY) display varying degrees of sex reversal, implicating *ATRX* in the development of the human testis (Biason-Lauber 2010). These observations are further supported by studies in mice and marsupials showing that *ATRX* is an evolutionarily conserved regulator of gonadal development in therian mammals and acts downstream of *SRY* and *SOX9* in the differentiation of testes (Huyhn et al. 2011). It is therefore plausible that *ATRX* mutations or deletions, as found in this study, affect normal sexual development also in horses. Detailed analysis of this gene in XY DSD horses should be subject for the future studies.

Of interest was also the large deletion found in the PAR of one XY horse with almost normal female external phenotype. The functions of the 4 genes involved (*AKAP17A, ASMT, ZBED1* and *XG*) are not well understood, although it is known that the PAR genes are dosage sensitive and possibly regulating very early stages of development (Raudsepp et al. 2012; Das et al. 2013b; Bellott et al. 2014). Thus, deletions in the PAR will cause haploinsufficiency for the genes involved like it happens on a larger scale in X-monosomy. However, as the regions showed significant variation by qPCR also in reference horses, the findings must be taken with caution and require further investigation.

However, the most functionally relevant CNVR was not a novel but a common complex variant present in chr29 at 28.6-28.8 Mb (Figs. 25, 26), previously found in normal individuals of at least 10 different breeds as a gain (Doan et al. 2012a) or a loss (Chapter II). In the current study, this region was homozygously deleted in one bilateral

cryptorchid and two sex reversal horses. The deletion involved at least 4 members of the aldo-keto reductase gene family (*AKR1C*), which function in the biochemical pathway that leads to dihydrotestosterone (DHT) synthesis without testosterone intermediate. As opposed to the 'classical' DHT synthesis from cholesterol and testosterone, this pathway is known as 'the backdoor pathway' and was originally discovered in marsupials (Wilson et al. 2003) and thereafter in eutherian mammals (Auchus 2004; Fluck et al. 2011; Biason-Lauber et al. 2013; Fukami et al. 2013). The importance of the 'backdoor pathway' and *AKR1C* genes in male sexual development was recently demonstrated by a study in humans showing that mutations in *AKR1C2* and *AKR1C4* genes are associated with various disorders of sexual development (DSDs) including cryptorchidism and sex reversal (Fluck et al. 2011; Biason-Lauber et al. 2013). Furthermore, CNV studies in human cryptorchidism have found genomic imbalances and one clinically significant *de novo* CNV in HSA10p14 - a region next to the *AKR1C* gene cluster.

Therefore, it is tempting to speculate that the CNVR in chr29 in the horse is a causative or a risk factor for some forms of equine XY DSDs. It is noteworthy that the three horses with the deletion had remarkably similar gonadal phenotypes: the two sex reversal horses (H348, H369, Table 18) were male pseudohermaphrodites with underdeveloped abdominal testes like it is in bilateral abdominal cryptorchidism. Thus, it is possible that some forms of cryptorchidism and sex reversal are due to the same genetic alterations in testis development pathways. However, apparent differences in the external phenotypes of the three horses suggest the involvement of additional genetic and non-genetic factors. Nevertheless, the region in chr29 requires detailed research by

re-sequencing, so that diagnostic tests could be developed for the detection of heterozygous carriers.

Taken together, this is the first attempt to investigate the contribution of CNVs in equine XY DSDs. The findings confirm genetic heterogeneity of the conditions, suggest likely involvement of CNVs in some, and indicate possible overlap in pathways leading to cryptorchidism and male pseudo hermaphroditism.

# CHAPTER V

# CONCLUSIONS AND FUTURE WORK

During just the past two years, several studies have addressed the phenomenon of copy number variation in the horse genome contributing to our knowledge about the genomic landscape of CNVs and their role in inter-individual variation in horses. Despite the progress, efficient biomedical application of this information requires addition of data from many more populations and individuals, as well as validation and analysis of the existing CNVs.

The overall goal of this dissertation was to advance CNV studies in horses by i) constructing a new, efficient and high-density whole genome CNV analysis platform; ii) discovering CNVs in the genomes of diverse horse breeds including those that have not been analyzed before; iii) generating a composite CNV dataset for the horse genome by integrating all hitherto available CNV information, and iv) discovering CNVs potentially contributing to complex equine diseases and congenital disorders.

In Chapter II, a tiling array was designed for the discovery of CNVs in the horse genome. The 400,000 array probes represented the majority of equine autosomal and X-linked genes, had a good representation of intergenic regions, and, as a unique feature, included sequences from the Y chromosome and sub-telomeric regions. The efficiency of the array was illustrated by its use for the discovery of over 100 novel CNVRs, not reported in horses before, and by detecting CNVRs in chrX and Un – the chromosomes largely ignored by previous studies. Importantly, this platform has a prospective use in

many future studies because the array design can easily accommodate the improvements in the horse genome sequence assembly.

The array was successfully applied for CNV discovery in 38 horses representing 16 diverse breeds of which 11 breeds, mainly ponies and draft horses, were analyzed for CNVs for the first time. Of the 258 CNVRs that were identified, 45% were novel and 55% overlapped with those reported previously. Thus, the findings added new information but also validated the existing CNV data for horses.

Probably one of the most important outcomes of this study was the integration of all CNVs and CNVRs from this and previous studies into a composite dataset of 1476 CNVRs for the horse genome. The integrated dataset is a critical resource for in depth analysis of known CNVRs and for new discoveries.

The utility of the tiling array and the integrated dataset was demonstrated in Chapters III and IV by studying CNVs in complex equine conditions: recurrent airway obstruction and disorders of sexual development. Cumulatively, these studies identified 500 CNVRs of which 87 were novel additions to the composite CNVRs dataset. Most importantly, these studies identified copy number variable genes and regions that are likely contributing or risk factors to complex equine disorders and diseases, such as RAO, cryptorchidism and XY sex reversal syndrome. As these studies are the first of their kind in horses, the findings are subject for further research.

One of the immediate future tasks is to properly validate all the novel CNVRs reported in this study. Also, novel variants of functional relevance should be analyzed in

larger cohorts of affected and control horses to determine possible associations with the phenotypes.

Regarding the overall CNV research in horses, many more individuals of diverse breeds should be studied, so that a comprehensive CNV catalogue can be generated for the horse genome. Ideally, this information should be incorporated and available in all main genome browsers, like it is for humans.

With the advances of NGS and the availability of whole genome sequence data for hundreds of individual horses, it is anticipated that array based CNV analysis will soon be replaced by whole genome sequence analysis, although, at the moment it is still cheaper to analyze multiple individuals on array platforms than carrying out high-coverage (40X) NGS and assembly in the same number of individuals. Furthermore, due to the very nature of CNVs, these regions are likely to have short-read NGS assemblies not as accurate as non-variable regions. Thus, the known CNVRs also identify potential targets for genome re-sequencing and –assembly using long-read NGS technologies, such as Pacific Biosciences and other single molecule platforms.

Taken together, exploration of CNVs in the horse genome has started and is promising, although, much more research is needed to properly understand the role of these variants in equine biology – in phenotypic variation, adaptations, and complex traits, in genetic diseases and congenital disorders.

# REFERENCES


Abe, H., K. Nagao and M. Inoue-Murayama (2013). "Short Copy Number Variations Potentially Associated with Tonic Immobility Responses in Newly Hatched Chicks." *PLoS One* **8**(11): e80205, 1-6.

Aboura, A., C. Dupas, G. Tachdjian, M. F. Portnoi, N. Bourcigaux, D. Dewailly, R. Frydman, B. Fauser, N. Ronci-Chaix, B. Donadille, et al. (2009). "Array Comparative Genomic Hybridization Profiling Analysis Reveals Deoxyribonucleic Acid Copy Number Variations Associated with Premature Ovarian Failure." *J Clin Endocrinol Metab* **94**(11): 4540-4546.

Adelson, D. L., J. M. Raison, M. Garber and R. C. Edgar (2010). "Interspersed Repeats in the Horse (*Equus caballus*); Spatial Correlations Highlight Conserved Chromosomal Domains." *Anim Genet* **41 Suppl 2**: 91-99.

Agoulnik, A. I. and S. Feng (2007). The Genetics of Cryptorchidism. *The Genetics of Male Infertility*. D. T. Carrell (ed). Totowa, New Jersey, Humana Press: 185-198.

Agoulnik, A. I., Z. Huang and L. Ferguson (2012). "Spermatogenesis in Cryptorchidism." *Methods Mol Biol* **825**: 127-147.

Aigner, J., S. Villatoro, R. Rabionet, J. Roquer, J. Jimenez-Conde, E. Marti and X. Estivill (2013). "A Common 56-Kilobase Deletion in a Primate-Specific

Segmental Duplication Creates a Novel Butyrophilin-Like Protein." *BMC Genet* **14**: 61, 1-12.

Akagi, K., J. Li, R. M. Stephens, N. Volfovsky and D. E. Symer (2008). "Extensive Variation between Inbred Mouse Strains Due to Endogenous L1 Retrotransposition." *Genome Res* **18**(6): 869-880.

Akiyama, H., J. P. Lyons, Y. Mori-Akiyama, X. Yang, R. Zhang, Z. Zhang, J. M. Deng, M. M. Taketo, T. Nakamura, R. R. Behringer, et al. (2004). "Interactions between *SOX9* and Beta-Catenin Control Chondrocyte Differentiation." *Genes Dev* **18**(9): 1072-1087.

Alkan, C., J. M. Kidd, T. Marques-Bonet, G. Aksay, F. Antonacci, F. Hormozdiari, J. O. Kitzman, C. Baker, M. Malig, O. Mutlu, et al. (2009). "Personalized Copy Number and Segmental Duplication Maps Using Next-Generation Sequencing." *Nat Genet* **41**(10): 1061-1067.

Alvarado, D. M., J. G. Buchan, S. L. Frick, J. E. Herzenberg, M. B. Dobbs and C. A. Gurnett (2013). "Copy Number Analysis of 413 Isolated Talipes Equinovarus Patients Suggests Role for Transcriptional Regulators of Early Limb Development." *Eur J Hum Genet* **21**(4): 373-380.

Alvarez, C. E. and J. M. Akey (2012). "Copy Number Variation in the Domestic Dog." *Mamm Genome* **23**(1-2): 144-163.

Amann, R. P. and D. N. Veeramachaneni (2007). "Cryptorchidism in Common Eutherian Mammals." *Reproduction* **133**(3): 541-561.

Amann, R. P. and D. N. R. Veeramachaneni (2006). "Cryptorchidism and Associated Problems in Animals." *Anim. Reprod.* **3**(2): 108-120.

Amaral, P. P., C. Neyt, S. J. Wilkins, M. E. Askarian-Amiri, S. M. Sunkin, A. C. Perkins and J. S. Mattick (2009). "Complex Architecture and Regulated Expression of the *SOX2OT* Locus During Vertebrate Development." *RNA* **15**(11): 2013-2027.

Andersson, L. and K. Sandberg (1984). "Genetic Linkage in the Horse. II. Distribution of Male Recombination Estimates and the Influence of Age, Breed and Sex on Recombination Frequency." *Genetics* **106**(1): 109-122.

Andersson, L. S., M. Larhammar, F. Memic, H. Wootz, D. Schwochow, C. J. Rubin, K. Patra, T. Arnason, L. Wellbring, G. Hjalm, et al. (2012a). "Mutations in *Dmrt3* Affect Locomotion in Horses and Spinal Circuit Function in Mice." *Nature* **488**(7413): 642-646.

Andersson, L. S., J. E. Swinburne, J. R. Meadows, H. Brostrom, S. Eriksson, W. F. Fikse, R. Frey, M. Sundquist, C. T. Tseng, S. Mikko, et al. (2012b). "The Same ELA Class II Risk Factors Confer Equine Insect Bite Hypersensitivity in Two Distinct Populations." *Immunogenetics* **64**(3): 201-208.

Andersson, L. S., M. Wilbe, A. Viluma, G. Cothran, B. Ekesten, S. Ewart and G. Lindgren (2013). "Equine Multiple Congenital Ocular Anomalies and Silver Coat

Colour Result from the Pleiotropic Effects of Mutant *PMEL*." *PLoS One* **8**(9): e75639, 1-8.

Ansari, H. A., R. Hediger, R. Fries and G. Stranzinger (1988). "Chromosomal Localization of the Major Histocompatibility Complex of the Horse (ELA) by in Situ Hybridization." *Immunogenetics* **28**(5): 362-364.

Anton, F., I. Leverkoehne, L. Mundhenk, W. B. Thoreson and A. D. Gruber (2005). "Overexpression of *ECLCA1* in Small Airways of Horses with Recurrent Airway Obstruction." *J Histochem Cytochem* **53**(8): 1011-1021.

Arighi, M. (2011). Testicular Descent. *Equine Reproduction*. A.O. McKinnon, W.E. Vaala, D. D. Varner (ed). New Delhi, India, Wiley-Blackwell: 1099-1112.

Armstrong, P. J., F. J. Derksen, R. F. Slocombe and N. E. Robinson (1986). "Airway Responses to Aerosolized Methacholine and Citric Acid in Ponies with Recurrent Airway Obstruction (Heaves)." *Am Rev Respir Dis* **133**(3): 357-361.

Art, T., N. Kirschvink, N. Smith, D. Votion and P. Lekeux (1999). "Cardiorespiratory Measurements and Indices of Oxidative Stress in Exercising COPD Horses." *Equine Vet J Suppl* **30**: 83-87.

Auchus, R. J. (2004). "The Backdoor Pathway to Dihydrotestosterone." *Trends Endocrinol Metab* **15**(9): 432-438.

Austen, C., S. Gorrie, P. Roome and N. J. Swinney (2008). *The Complete Illustrated Encyclopedia of Horses*. New York, Metro Books: 1-352.

Bae, J. S., H. S. Cheong, L. H. Kim, S. NamGung, T. J. Park, J. Y. Chun, J. Y. Kim, C. F. Pasaje, J. S. Lee and H. D. Shin (2010). "Identification of Copy Number Variations and Common Deletion Polymorphisms in Cattle." *BMC Genomics* **11**: 232, 1-10.

Bailey, E., K. T. Graves, E. G. Cothran, R. Reid, T. L. Lear and R. B. Ennis (1995). "Synteny-Mapping Horse Microsatellite Markers Using a Heterohybridoma Panel." *Anim Genet* **26**(3): 177-180.

Bannasch, D., C. Rinaldo, L. Millon, K. Latson, T. Spangler, S. Hubberty, L. Galuppo and L. Lowenstine (2007). "*SRY* Negative 64,XX Intersex Phenotype in an American Saddlebred Horse." *Vet J* **173**(2): 437-439.

Barbosa-Goncalves, A., C. B. Vendrame-Goloni, A. L. Martins and A. C. Fett-Conte (2008). "Subtelomeric Region of Chromosome 2 in Patients with Autism Spectrum Disorders." *Genet Mol Res* **7**(2): 527-533.

Barthold, J. S. (2008). "Undescended Testis: Current Theories of Etiology." *Curr Opin Urol* **18**(4): 395-400.

Barthold, J. S., Y. Wang, A. Robbins, J. Pike, E. McDowell, K. J. Johnson and S. M. McCahan (2013). "Transcriptome Analysis of the Dihydrotestosterone-Exposed

Fetal Rat Gubernaculum Identifies Common Androgen and Insulin-Like 3 Targets." *Biol Reprod* **89**(6): 143, 1-12.

Bass, A. J., M. S. Lawrence, L. E. Brace, A. H. Ramos, Y. Drier, K. Cibulskis, C. Sougnez, D. Voet, G. Saksena, A. Sivachenko, et al. (2011). "Genomic Sequencing of Colorectal Adenocarcinomas Identifies a Recurrent *VTILA-Tcf7l2* Fusion." *Nat Genet* **43**(10): 964-968.

Beckmann, J. S., X. Estivill and S. E. Antonarakis (2007). "Copy Number Variants and Genetic Traits: Closer to the Resolution of Phenotypic to Genotypic Variability." *Nat Rev Genet* **8**(8): 639-646.

Beckmann, J. S., A. J. Sharp and S. E. Antonarakis (2008). "CNVs and Genetic Medicine (Excitement and Consequences of a Rediscovery)." *Cytogenet Genome Res* **123**(1-4): 7-16.

Bellone, R. R., G. Forsyth, T. Leeb, S. Archer, S. Sigurdsson, F. Imsland, E. Mauceli, M. Engensteiner, E. Bailey, L. Sandmeyer, et al. (2010). "Fine-Mapping and Mutation Analysis of *TRPM1*: A Candidate Gene for Leopard Complex (Lp) Spotting and Congenital Stationary Night Blindness in Horses." *Brief Funct Genomics* **9**(3): 193-207.

Bellone, R. R., H. Holl, V. Setaluri, S. Devi, N. Maddodi, S. Archer, L. Sandmeyer, A. Ludwig, D. Foerster, M. Pruvost, et al. (2013). "Evidence for a Retroviral

Insertion in *TRPM1* as the Cause of Congenital Stationary Night Blindness and Leopard Complex Spotting in the Horse." *PLoS One* **8**(10): e78280, 1-14.

Bellott, D. W., J. F. Hughes, H. Skaletsky, L. G. Brown, T. Pyntikova, T. J. Cho, N. Koutseva, S. Zaghlul, T. Graves, S. Rock, et al. (2014). "Mammalian Y Chromosomes Retain Widely Expressed Dosage-Sensitive Regulators." *Nature* **508**(7497): 494-499.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *J Roy Statist Soc Ser B (Methodological)* **57**(1): 289-300.

Bergen, S. E., C. T. O'Dushlaine, S. Ripke, P. H. Lee, D. M. Ruderfer, S. Akterin, J. L. Moran, K. D. Chambert, R. E. Handsaker, L. Backlund, et al. (2012). "Genome-Wide Association Study in a Swedish Population Yields Support for Greater CNV and MHC Involvement in Schizophrenia Compared with Bipolar Disorder." *Mol Psychiatry* **17**(9): 880-886.

Berger, M. F., M. S. Lawrence, F. Demichelis, Y. Drier, K. Cibulskis, A. Y. Sivachenko, A. Sboner, R. Esgueva, D. Pflueger, C. Sougnez, et al. (2011). "The Genomic Complexity of Primary Human Prostate Cancer." *Nature* **470**(7333): 214-220.

Berglund, J., E. M. Nevalainen, A. M. Molin, M. Perloski, C. Andre, M. C. Zody, T. Sharpe, C. Hitte, K. Lindblad-Toh, H. Lohi, et al. (2012). "Novel Origins of Copy Number Variation in the Dog Genome." *Genome Biol* **13**(8): R73, 1-18.

Biason-Lauber, A. (2010). "Control of Sex Development." *Best Pract Res Clin Endocrinol Metab* **24**(2): 163-186.

Biason-Lauber, A., W. L. Miller, A. V. Pandey and C. E. Fluck (2013). "Of Marsupials and Men: "Backdoor" Dihydrotestosterone Synthesis in Male Sexual Differentiation." *Mol Cell Endocrinol* **371**(1-2): 124-132.

Bickhart, D. M., Y. Hou, S. G. Schroeder, C. Alkan, M. F. Cardone, L. K. Matukumalli, J. Song, R. D. Schnabel, M. Ventura, J. F. Taylor, et al. (2012). "Copy Number Variation of Individual Cattle Genomes Using Next-Generation Sequencing." *Genome Res* **22**(4):778-790.

Binns, M. M., D. A. Boehler, E. Bailey, T. L. Lear, J. M. Cardwell and D. H. Lambert (2012). "Inbreeding in the Thoroughbred Horse." *Anim Genet* **43**(3): 340-342.

Binns, M. M., D. A. Boehler and D. H. Lambert (2010). "Identification of the Myostatin Locus (*MSTN*) as Having a Major Effect on Optimum Racing Distance in the Thoroughbred Horse in the USA." *Anim Genet* **41 Suppl 2**: 154-158.

Blake, J. A., C. J. Bult, J. T. Eppig, J. A. Kadin and J. E. Richardson (2014). "The Mouse Genome Database: Integration of and Access to Knowledge About the Laboratory Mouse." *Nucleic Acids Res* **42**(Database issue): D810-D817.

Blanchard, T. L., I. Johnson and A. J. Roser (2000). "Increased Germ Cell Loss Rates and Poor Semen Quality in Stallions with Idiopathic Testicular Degeneration." *J Equine Vet Sci* **20**: 263-265.

Blott, S. C., J. E. Swinburne, C. Sibbons, L. Y. Fox-Clipsham, M. Helwegen, L. Hillyer, T. D. Parkin, J. R. Newton and M. Vaudin (2014). "A Genome-Wide Association Study Demonstrates Significant Genetic Variation for Fracture Risk in Thoroughbred Racehorses." *BMC Genomics* **15**(1): 147, 1-8.

Bodin, L., P. H. Beaune and M. A. Loriot (2005). "Determination of Cytochrome P450 2d6 (*Cyp2d6*) Gene Copy Number by Real-Time Quantitative PCR." *J Biomed Biotechnol* **2005**(3): 248-253.

Boone, P. M., I. M. Campbell, B. C. Baggett, Z. T. Soens, M. M. Rao, P. M. Hixson, A. Patel, W. Bi, S. W. Cheung, S. R. Lalani, et al. (2013). "Deletions of Recessive Disease Genes: CNV Contribution to Carrier States and Disease-Causing Alleles." *Genome Res* **23**(9): 1383-1394.

Bower, M. A., M. Whitten, R. E. Nisbet, M. Spencer, K. M. Dominy, A. M. Murphy, R. Cassidy, E. Barrett, E. W. Hill and M. Binns (2013). "Thoroughbred Racehorse Mitochondrial DNA Demonstrates Closer Than Expected Links between Maternal Genetic History and Pedigree Records." *J Anim Breed Genet* **130**(3): 227-235.

Bowling, A. T., L. Millon and J. P. Hughes (1987). "An Update of Chromosomal Abnormalities in Mares." *J Reprod Fertil Suppl* **35**: 149-155.

Breen, M., G. Lindgren, M. M. Binns, J. Norman, Z. Irvin, K. Bell, K. Sandberg and H. Ellegren (1997). "Genetical and Physical Assignments of Equine Microsatellites-

-First Integration of Anchored Markers in Horse Genome Mapping." *Mamm Genome* **8**(4): 267-273.

Bright, L. A., S. C. Burgess, B. Chowdhary, C. E. Swiderski and F. M. McCarthy (2009). "Structural and Functional-Annotation of an Equine Whole Genome Oligoarray." *BMC Bioinformatics* **10 Suppl 11**: S8, 1-8.

Brinkmeyer-Langford, C., T. Raudsepp, A. Gustafson-Seabury and B. P. Chowdhary (2008). "A BAC Contig Map over the Proximal Approximately 3.3 Mb Region of Horse Chromosome 21." *Cytogenet Genome Res* **120**(1-2): 164-172.

Brinkmeyer-Langford, C. L., W. J. Murphy, C. P. Childers and L. C. Skow (2010). "A Conserved Segmental Duplication within ELA." *Anim Genet* **41 Suppl 2**: 186-195.

Brooks, S. A. and E. Bailey (2005). "Exon Skipping in the *KIT* Gene Causes a Sabino Spotting Pattern in Horses." *Mamm Genome* **16**(11): 893-902.

Brooks, S. A. and R. R. Bellone (2013). Coat Color Genomics. *Equine Genomics*. B. P. Chowdhary (ed), Wiley-Blackwell**:** 143-153.

Brooks, S. A., N. Gabreski, D. Miller, A. Brisbin, H. E. Brown, C. Streeter, J. Mezey, D. Cook and D. F. Antczak (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, 1-7.

Brooks, S. A., T. L. Lear, D. L. Adelson and E. Bailey (2007). "A Chromosome Inversion near the *KIT* Gene and the Tobiano Spotting Pattern in Horses." *Cytogenet Genome Res* **119**(3-4): 225-230.

Brunberg, E., L. Andersson, G. Cothran, K. Sandberg, S. Mikko and G. Lindgren (2006). "A Missense Mutation in *PMEL17* is Associated with the Silver Coat Color in the Horse." *BMC Genet* **7**: 46, 1-10.

Buechner-Maxwell, V., C. Zhang, J. Robertson, N. C. Jain, D. F. Antczak, B. F. Feldman and M. J. Murray (1994). "Intravascular Leukostasis and Systemic Aspergillosis in a Horse with Subleukemic Acute Myelomonocytic Leukemia." *J Vet Intern Med* **8**(4): 258-263.

Buoen, L. C., T. Q. Zhang, A. F. Weber and G. R. Ruth (2000). "SRY-Negative, XX Intersex Horses: The Need for Pedigree Studies to Examine the Mode of Inheritance of the Condition." *Equine Vet J* **32**(1): 78-81.

Campbell, P. J., P. J. Stephens, E. D. Pleasance, S. O'Meara, H. Li, T. Santarius, L. A. Stebbings, C. Leroy, S. Edkins, C. Hardy, et al. (2008). "Identification of Somatically Acquired Rearrangements in Cancer Using Genome-Wide Massively Parallel Paired-End Sequencing." *Nat Genet* **40**(6): 722-729.

Campbell, P. J., S. Yachida, L. J. Mudie, P. J. Stephens, E. D. Pleasance, L. A. Stebbings, L. A. Morsberger, C. Latimer, S. McLaren, M. L. Lin, et al. (2010).

"The Patterns and Dynamics of Genomic Instability in Metastatic Pancreatic Cancer." *Nature* **467**(7319): 1109-1113.

Cannistraci, C. V., J. Ogorevc, M. Zorc, T. Ravasi, P. Dovc and T. Kunej (2013). "Pivotal Role of the Muscle-Contraction Pathway in Cryptorchidism and Evidence for Genomic Connections with Cardiomyopathy Pathways in Rasopathies." *BMC Med Genomics* **6**: 5, 1-16.

Carrell, D. T. (2007). The Genetics of Male Infertility in the Era of Genomics. *The Genetics of Male Infertility*. D. T. Carrell (ed). Totowa, New Jersey, Humana Press: 3-27.

Carrell, D. T. and K. I. Aston (2011). "The Search for SNPs, CNVs, and Epigenetic Variants Associated with the Complex Disease of Male Infertility." *Syst Biol Reprod Med* **57**(1-2): 17-26.

Carroll, T., C. O'Connor, O. Floyd, J. McPartlin, D. Kelleher, G. O'Brien, B. Dimitrov, V. Morris, C. Taggart and N. McElvaney (2011). "The Prevalence of Alpha-1 Antitrypsin Deficiency in Ireland." *Respir Res* **12**(1): 91, 1-7.

Carter, N. P. (2007). "Methods and Strategies for Analyzing Copy Number Variation Using DNA Microarrays." *Nat Genet* **39**(7 Suppl): S16-S21.

Cavalleri, J. M., J. Metzger, M. Hellige, V. Lampe, K. Stuckenschneider, A. Tipold, A. Beineke, K. Becker, O. Distl and K. Feige (2013). "Morphometric Magnetic

Resonance Imaging and Genetic Testing in Cerebellar Abiotrophy in Arabian Horses." *BMC Vet Res* **9**: 105, 1-9.

Ceroni, F., A. Sagar, N. H. Simpson, A. J. Gawthrope, D. F. Newbury, D. Pinto, S. M. Francis, D. C. Tessman, E. H. Cook, A. P. Monaco, et al. (2014). "A Deletion Involving *CD38* and *BST1* Results in a Fusion Transcript in a Patient with Autism and Asthma." *Autism Res* ;**7**(2): 254-63.

Charchar, F. J., M. Kaiser, A. J. Bingham, N. Fotinatos, F. Ahmady, M. Tomaszewski and N. J. Samani (2010). "Whole Genome Survey of Copy Number Variation in the Spontaneously Hypertensive Rat: Relationship to Quantitative Trait Loci, Gene Expression, and Blood Pressure." *Hypertension* **55**(5): 1231-1238.

Chassot, A. A., E. P. Gregoire, M. Magliano, R. Lavery and M. C. Chaboissier (2008). "Genetics of Ovarian Differentiation: *RSPO1*, a Major Player." *Sex Dev* **2**(4-5): 219-227.

Chen, C., R. Qiao, R. Wei, Y. Guo, H. Ai, J. Ma, J. Ren and L. Huang (2012). "A Comprehensive Survey of Copy Number Variation in 18 Diverse Pig Populations and Identification of Candidate Copy Number Variable Genes Associated with Complex Traits." *BMC Genomics* **13**(1): 733, 1-10.

Chen, W. K., J. D. Swartz, L. J. Rush and C. E. Alvarez (2009). "Mapping DNA Structural Variation in Dogs." *Genome Res* **19**(3): 500-509.

Cheng, Z., M. Ventura, X. She, P. Khaitovich, T. Graves, K. Osoegawa, D. Church, P. DeJong, R. K. Wilson, S. Paabo, et al. (2005). "A Genome-Wide Comparison of Recent Chimpanzee and Human Segmental Duplications." *Nature* **437**(7055): 88-93.

Choi, J. W., K. T. Lee, X. Liao, P. Stothard, H. S. An, S. Ahn, S. Lee, S. Y. Lee, S. S. Moore and T. H. Kim (2013). "Genome-Wide Copy Number Variation in Hanwoo, Black Angus, and Holstein Cattle." *Mamm Genome* **24**(3-4): 151-163.

Chowdhary, B. P., N. Paria and T. Raudsepp (2008). "Potential Applications of Equine Genomics in Dissecting Diseases and Fertility." *Anim Reprod Sci* **107**(3-4): 208-218.

Chowdhary, B. P. and T. Raudsepp (2008). "The Horse Genome Derby: Racing from Map to Whole Genome Sequence." *Chromosome Res* **16**(1): 109-127.

Chowdhary, B. P., T. Raudsepp, S. R. Kata, G. Goh, L. V. Millon, V. Allan, F. Piumi, G. Guerin, J. Swinburne, M. Binns, et al. (2003). "The First-Generation Whole-Genome Radiation Hybrid Map in the Horse Identifies Conserved Segments in Human and Mouse Genomes." *Genome Res* **13**(4): 742-751.

Chowdhary, B. P., T. Raudsepp, D. Honeycutt, E. K. Owens, F. Piumi, G. Guerin, T. C. Matise, S. R. Kata, J. E. Womack and L. C. Skow (2002). "Construction of a 5000(RAD) Whole-Genome Radiation Hybrid Panel in the Horse and Generation

of a Comprehensive and Comparative Map for ECA11." *Mamm Genome* **13**(2): 89-94.

Chowdhary, B. P. and T. Raudsepp (2000). Cytogenetics and Physical Gene Maps. *The Genetics of the Horse*. A. T. Bowling and A. Ruvinsky (ed), CABI Publishing**:** 171-241.

Choy, K. W., S. R. Setlur, C. Lee and T. K. Lau (2010). "The Impact of Human Copy Number Variation on a New Era of Genetic Testing." *An International Journal of Obstetrics and Gynaecology* **117**(4): 391-398.

Cicconardi, F., G. Chillemi, A. Tramontano, C. Marchitelli, A. Valentini, P. Ajmone-Marsan and A. Nardone (2013). "Massive Screening of Copy Number Population-Scale Variation in *Bos taurus* Genome." *BMC Genomics* **14**: 124, 1-14.

Cieslak, M., M. Pruvost, N. Benecke, M. Hofreiter, A. Morales, M. Reissmann and A. Ludwig (2010). "Origin and History of Mitochondrial DNA Lineages in Domestic Horses." *PLoS One* **5**(12): e15311, 1-13.

Clop, A., O. Vidal and M. Amills (2012). "Copy Number Variation in the Genomes of Domestic Animals." *Anim Genet* **43**(5): 503-517.

Coleman, S. J., Z. Zeng, M. S. Hestand, J. Liu and J. N. Macleod (2013). "Analysis of Unannotated Equine Transcripts Identified by mRNA Sequencing." *PLoS One* **8**(7): e70125, 1-11.

Conant, G. C. and K. H. Wolfe (2008). "Turning a Hobby into a Job: How Duplicated Genes Find New Functions." *Nat Rev Genet* **9**(12): 938-950.

Conrad, B. and S. E. Antonarakis (2007). "Gene Duplication: A Drive for Phenotypic Diversity and Cause of Human Disease." *Annu Rev Genomics Hum Genet* **8**: 17-35.

Conrad, D. F., T. D. Andrews, N. P. Carter, M. E. Hurles and J. K. Pritchard (2006). "A High-Resolution Survey of Deletion Polymorphism in the Human Genome." *Nat Genet* **38**(1): 75-81.

Conrad, D. F., C. Bird, B. Blackburne, S. Lindsay, L. Mamanova, C. Lee, D. J. Turner and M. E. Hurles (2010a). "Mutation Spectrum Revealed by Breakpoint Sequencing of Human Germline CNVs." *Nat Genet* **42**(5): 385-391.

Conrad, D. F., D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T. D. Andrews, C. Barnes, P. Campbell, et al. (2010b). "Origins and Functional Impact of Copy Number Variation in the Human Genome." *Nature* **464**(7289): 704-712.

Cook, D., S. Brooks, R. Bellone and E. Bailey (2008). "Missense Mutation in Exon 2 of *SLC36A1* Responsible for Champagne Dilution in Horses." *PLoS Genet* **4**(9): e1000195, 1-9.

Cooper, G. M., D. A. Nickerson and E. E. Eichler (2007). "Mutational and Selective Effects on Copy-Number Variants in the Human Genome." *Nat Genet* **39**(7 Suppl): S22-S29.

Corbin, L. J., S. C. Blott, J. E. Swinburne, C. Sibbons, L. Y. Fox-Clipsham, M. Helwegen, T. D. Parkin, J. R. Newton, L. R. Bramlage, C. W. McIlwraith, et al. (2012). "A Genome-Wide Association Study of Osteochondritis Dissecans in the Thoroughbred." *Mamm Genome* **23**(3-4): 294-303.

Cordaux, R. and M. A. Batzer (2009). "The Impact of Retrotransposons on Human Genome Evolution." *Nat Rev Genet* **10**(10): 691-703.

Costelloe, S. J., J. S. El-Sayed Moustafa, F. Drenos, J. Palmen, Q. Li, S. Whiting, M. Thomas, M. Kivimaki, M. Kumari, A. D. Hingorani, et al. (2012). "Gene-Targeted Analysis of Copy Number Variants Identifies 3 Novel Associations with Coronary Heart Disease Traits." *Circ Cardiovasc Genet* **5**(5): 555-560.

Cote, O., B. N. Lillie, M. A. Hayes, M. E. Clark, L. van den Bosch, P. Katavolos, L. Viel and D. Bienzle (2012). "Multiple Secretoglobin 1A1 Genes are Differentially Expressed in Horses." *BMC Genomics* **13**: 712, 1-14.

Crespi, B. J. and H. J. Crofts (2012). "Association Testing of Copy Number Variants in Schizophrenia and Autism Spectrum Disorders." *J Neurodev Disord* **4**(1): 15, 1-9.

Crooijmans, R. P., M. S. Fife, T. W. Fitzgerald, S. Strickland, H. H. Cheng, P. Kaiser, R. Redon and M. A. Groenen (2013). "Large Scale Variation in DNA Copy Number in Chicken Breeds." *BMC Genomics* **14**: 398, 1-10.

Curry, C. J., J. A. Rosenfeld, E. Grant, K. W. Gripp, C. Anderson, A. S. Aylsworth, T. B. Saad, V. V. Chizhikov, G. Dybose, C. Fagerberg, et al. (2013). "The Duplication 17p13.3 Phenotype: Analysis of 21 Families Delineates Developmental, Behavioral and Brain Abnormalities, and Rare Variant Phenotypes." *Am J Med Genet A* **161A**(8): 1833-1852.

D'Angelo, C. S. and C. P. Koiffmann (2012). "Copy Number Variants in Obesity-Related Syndromes: Review and Perspectives on Novel Molecular Approaches." *J Obes* **2012**: 845480, 1-15.

Dalgaard, M. D., N. Weinhold, D. Edsgard, J. D. Silver, T. H. Pers, J. E. Nielsen, N. Jorgensen, A. Juul, T. A. Gerds, A. Giwercman, et al. (2012). "A Genome-Wide Association Study of Men with Symptoms of Testicular Dysgenesis Syndrome and its Network Biology Interpretation." *J Med Genet* **49**(1): 58-65.

Das, P. J., S. K. Lyle, D. Beehan, B. P. Chowdhary and T. Raudsepp (2012). "Cytogenetic and Molecular Characterization of Y Isochromosome in a 63XO/64Xi(Yq) Mosaic Karyotype of an Intersex Horse." *Sex Dev* **6**(1-3): 117-127.

Das, P. J., F. McCarthy, M. Vishnoi, N. Paria, C. Gresham, G. Li, P. Kachroo, A. K. Sudderth, S. Teague, C. C. Love, et al. (2013a). "Stallion Sperm Transcriptome Comprises Functionally Coherent Coding and Regulatory RNAs as Revealed by Microarray Analysis and RNA-Seq." *PLoS One* **8**(2): e56535, 1-15.

Das, P. J., D. K. Mishra, S. Ghosh, F. Avila, G. A. Johnson, B. P. Chowdhary and T. Raudsepp (2013b). "Comparative Organization and Gene Expression Profiles of the Porcine Pseudoautosomal Region." *Cytogenet Genome Res* **41**(1): 26-36

Dauber, A., Y. Yu, M. C. Turchin, C. W. Chiang, Y. A. Meng, E. W. Demerath, S. R. Patel, S. S. Rich, J. I. Rotter, P. J. Schreiner, et al. (2011). "Genome-Wide Association of Copy-Number Variation Reveals an Association between Short Stature and the Presence of Low-Frequency Genomic Deletions." *Am J Hum Genet* **89**(6): 751-759.

De Lorenzi, L., V. Genualdo, A. Iannuzzi, G. P. Di Meo, A. Perucatti, R. Mancuso, M. Russo, D. Di Berardino, P. Parma and L. Iannuzzi (2010). "Cytogenetic and Genetic Studies in a Hypospadic Horse (*Equus caballus*, 2n = 64)." *Sex Dev* **4**(6): 352-357.

Denden, S., A. H. Khelil, J. Knani, R. Lakhdar, P. Perrin, G. Lefranc and J. B. Chibani (2010). "Alpha-1 Antitrypsin Gene Polymorphism in Chronic Obstructive Pulmonary Disease (Copd)." *Genet Mol Biol* **33**(1): 23-26.

Derksen, F. J., N. E. Robinson, J. S. Scott and J. A. Stick (1988). "Aerosolized Micropolyspora Faeni Antigen as a Cause of Pulmonary Dysfunction in Ponies with Recurrent Airway Obstruction (Heaves)." *Am J Vet Res* **49**(6): 933-938.

DeScipio, C., N. B. Spinner, M. Kaur, D. Yaeger, L. K. Conlin, A. Ambrosini, S. Hu, S. Shan, I. D. Krantz and H. Riethman (2008). "Fine-Mapping Subtelomeric

Deletions and Duplications by Comparative Genomic Hybridization in 42 Individuals." *Am J Med Genet A* **146A**(6): 730-739.

Diribarne, M., A. Vaiman, M. Péchayre, E. Pailhoux, X. Mata, G. Guérin and S. Chaffaux (2009). "Polymorphism Analysis of Microsatellites Associated with Seven Candidate Genes for Equine Cryptorchidism." *J Equine Vet Sci* **29**: 37-41.

Doan, R., N. Cohen, J. Harrington, K. Veazy, R. Juras, G. Cothran, M. E. McCue, L. Skow and S. V. Dindot (2012a). "Identification of Copy Number Variants in Horses." *Genome Res* **22**(5): 899-907.

Doan, R., N. D. Cohen, J. Sawyer, N. Ghaffari, C. D. Johnson and S. V. Dindot (2012b). "Whole-Genome Sequencing and Genetic Variant Analysis of a Quarter Horse Mare." *BMC Genomics* **13**(1): 78, 1-11.

Dolf, G., C. Gaillard, C. Schelling, A. Hofer and E. Leighton (2008). "Cryptorchidism and Sex Ratio are Associated in Dogs and Pigs." *J Anim Sci* **86**(10): 2480-2485.

Du, R., C. Lu, Z. Jiang, S. Li, R. Ma, H. An, M. Xu, Y. An, Y. Xia, L. Jin, et al. (2012). "Efficient Typing of Copy Number Variations in a Segmental Duplication-Mediated Rearrangement Hotspot Using Multiplex Competitive Amplification." *J Hum Genet* **57**(8): 545-551.

Ducos, A., T. Revay, A. Kovacs, A. Hidas, A. Pinton, A. Bonnet-Garnier, L. Molteni, E. Slota, M. Switonski, M. V. Arruga, et al. (2008). "Cytogenetic Screening of

Livestock Populations in Europe: An Overview." *Cytogenet Genome Res* **120**(1-2): 26-41.

Dumas, L., Y. H. Kim, A. Karimpour-Fard, M. Cox, J. Hopkins, J. R. Pollack and J. M. Sikela (2007). "Gene Copy Number Variation Spanning 60 Million Years of Human and Primate Evolution." *Genome Res* **17**(9): 1266-1277.

Dunham, I., A. Kundaje, S. F. Aldred, P. J. Collins, C. A. Davis, F. Doyle, C. B. Epstein, S. Frietze, J. Harrow, R. Kaul, et al. (2012). "An Integrated Encyclopedia of DNA Elements in the Human Genome." *Nature* **489**(7414): 57-74.

Dupuis, M. C., Z. Zhang, T. Druet, J. M. Denoix, C. Charlier, P. Lekeux and M. Georges (2011). "Results of a Haplotype-Based Gwas for Recurrent Laryngeal Neuropathy in the Horse." *Mamm Genome* **22**(9-10): 613-620.

Dupuis, M. C., Z. Zhang, K. Durkin, C. Charlier, P. Lekeux and M. Georges (2013). "Detection of Copy Number Variants in the Horse Genome and Examination of Their Association with Recurrent Laryngeal Neuropathy." *Anim Genet* **44**(2): 206-208.

Durkin, K., W. Coppieters, C. Drogemuller, N. Ahariz, N. Cambisano, T. Druet, C. Fasquelle, A. Haile, P. Horin, L. Huang, et al. (2012). "Serial Translocation by Means of Circular Intermediates Underlies Colour Sidedness in Cattle." *Nature* **482**(7383): 81-84.

Duschek, S., J. Worsching and G. A. Reyes del Paso (2013). "Interactions between Autonomic Cardiovascular Regulation and Cortical Activity: A CNV Study." *Psychophysiology* **50**(4): 388-397.

Elia, J., J. T. Glessner, K. Wang, N. Takahashi, C. J. Shtir, D. Hadley, P. M. Sleiman, H. Zhang, C. E. Kim, R. Robison, et al. (2012). "Genome-Wide Copy Number Variation Study Associates Metabotropic Glutamate Receptor Gene Networks with Attention Deficit Hyperactivity Disorder." *Nat Genet* **44**(1): 78-84.

Ermakova, O., L. Piszczek, L. Luciani, F. M. Cavalli, T. Ferreira, D. Farley, S. Rizzo, R. C. Paolicelli, M. Al-Banchaabouchi, C. Nerlov, et al. (2011). "Sensitized Phenotypic Screening Identifies Gene Dosage Sensitive Region on Chromosome 11 that Predisposes to Disease in Mice." *EMBO Mol Med* **3**(1): 50-66.

Fadista, J., M. Nygaard, L. E. Holm, B. Thomsen and C. Bendixen (2008). "A Snapshot of CNVs in the Pig Genome." *PLoS One* **3**(12): e3916, 1-9.

Fadista, J., B. Thomsen, L. E. Holm and C. Bendixen (2010). "Copy Number Variation in the Bovine Genome." *BMC Genomics* **11**: 284, 1-12.

Falchi, M., J. S. El-Sayed Moustafa, P. Takousis, F. Pesce, A. Bonnefond, J. C. Andersson-Assarsson, P. H. Sudmant, R. Dorajoo, M. N. Al-Shafai, L. Bottolo, et al. (2014). "Low Copy Number of the Salivary Amylase Gene Predisposes to Obesity." *Nat Genet* **46**(5): 492-497.

Fanciulli, M., E. Petretto and T. J. Aitman (2010). "Gene Copy Number Variation and Common Human Disease." *Clin Genet* **77**(3): 201-213.

Fechheimer, N. S. (1971). "Cytogenetic Considerations in Animal Breeding." *Annales de Génétique et de Sélection Animale* **3**: 43–58.

Fernandez, A. I., C. Barragan, A. Fernandez, M. C. Rodriguez and B. Villanueva (2014). "Copy Number Variants in a Highly Inbred Iberian Porcine Strain." *Anim Genet* **45**(3): 357-66.

Feuk, L., A. R. Carson and S. W. Scherer (2006). "Structural Variation in the Human Genome." *Nat Rev Genet* **7**(2): 85-97.

Fiegler, H., R. Redon, D. Andrews, C. Scott, R. Andrews, C. Carder, R. Clark, O. Dovey, P. Ellis, L. Feuk, et al. (2006). "Accurate and Reliable High-Throughput Detection of Copy Number Variation in the Human Genome." *Genome Res* **16**(12): 1566-1574.

Finely, B. (2009). Another Sex Dispute, but This Athlete Has Four Legs. *New York Times*. New York.

Finno, C. J., S. J. Spier and S. J. Valberg (2009). "Equine Diseases Caused by Known Genetic Mutations." *Vet J* **179**(3): 336-347.

Firth, H. V., S. M. Richards, A. P. Bevan, S. Clayton, M. Corpas, D. Rajan, S. V. Vooren, Y. Moreau, R. M. Pettett and N. P. Carter (2009). "Decipher: Database

of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources." *Am J Hum Genet* **84**(4): 524-533.

Fluck, C. E., M. Meyer-Boni, A. V. Pandey, P. Kempna, W. L. Miller, E. J. Schoenle and A. Biason-Lauber (2011). "Why Boys Will Be Boys: Two Pathways of Fetal Testicular Androgen Biosynthesis Are Needed for Male Sexual Differentiation." *Am J Hum Genet* **89**(2): 201-218.

Fontanesi, L., F. Beretti, P. L. Martelli, M. Colombo, S. Dall'olio, M. Occidente, B. Portolano, R. Casadio, D. Matassino and V. Russo (2010a). "A First Comparative Map of Copy Number Variations in the Sheep Genome." *Genomics* **97**(3): 158-165.

Fontanesi, L., F. Beretti, V. Riggio, E. Gomez Gonzalez, S. Dall'Olio, R. Davoli, V. Russo and B. Portolano (2009). "Copy Number Variation and Missense Mutations of the Agouti Signaling Protein (*ASIP*) Gene in Goat Breeds with Different Coat Colors." *Cytogenet Genome Res* **126**(4): 333-347.

Fontanesi, L., P. L. Martelli, F. Beretti, V. Riggio, S. Dall'Olio, M. Colombo, R. Casadio, V. Russo and B. Portolano (2010b). "An Initial Comparative Map of Copy Number Variations in the Goat (*Capra hircus*) Genome." *BMC Genomics* **11**: 639, 1-15.

Foresta, C., A. Ferlin, A. Garolla, C. Milani, G. Oliva and M. Rossato (1996). "Functional and Cytologic Features of the Contralateral Testis in Cryptorchidism." *Fertility and Sterility* **66**(4): 624-629.

Foster, R. A. and P. W. Ladds (2007). Male Genital System. *Pathology of Domestic Animals*. M. G. Maxie (ed). Philadelphia, Elesevier Saunders: 564-619.

Fowler, K. E., R. Pong-Wong, J. Bauer, E. J. Clemente, C. P. Reitter, N. A. Affara, S. Waite, G. A. Walling and D. K. Griffin (2013). "Genome Wide Analysis Reveals Single Nucleotide Polymorphisms Associated with Fatness and Putative Novel Copy Number Variants in Three Pig Breeds." *BMC Genomics* **14**: 784, 1-15.

Fox-Clipsham, L. Y., E. E. Brown, S. D. Carter and J. E. Swinburne (2011a). "Population Screening of Endangered Horse Breeds for the Foal Immunodeficiency Syndrome Mutation." *Vet Rec* **169**(25): 655-659.

Fox-Clipsham, L. Y., S. D. Carter, I. Goodhead, N. Hall, D. C. Knottenbelt, P. D. May, W. E. Ollier and J. E. Swinburne (2011b). "Identification of a Mutation Associated with Fatal Foal Immunodeficiency Syndrome in the Fell and Dales Pony." *PLoS Genet* **7**(7): e1002133, 1-8.

Freeman, J. L., G. H. Perry, L. Feuk, R. Redon, S. A. McCarroll, D. M. Altshuler, H. Aburatani, K. W. Jones, C. Tyler-Smith, M. E. Hurles, et al. (2006). "Copy Number Variation: New Insights in Genome Diversity." *Genome Res* **16**(8): 949-961.

Fritz, K. L., H. J. Kaese, S. J. Valberg, J. A. Hendrickson, A. K. Rendahl, R. R. Bellone, K. M. Dynes, M. L. Wagner, M. A. Lucio, F. M. Cuomo, et al. (2014). "Genetic Risk Factors for Insidious Equine Recurrent Uveitis in Appaloosa Horses." *Anim Genet* **45**(3): 392-399.

Fritz, K. L., M. E. McCue, S. J. Valberg, A. K. Rendahl and J. R. Mickelson (2012). "Genetic Mapping of Recurrent Exertional Rhabdomyolysis in a Population of North American Thoroughbreds." *Anim Genet* **43**(6): 730-738.

Fukami, M., K. Homma, T. Hasegawa and T. Ogata (2013). "Backdoor Pathway for Dihydrotestosterone Biosynthesis: Implications for Normal and Abnormal Human Sex Development." *Dev Dyn* **242**(4): 320-329.

Gebhardt, S., M. Merkl, N. Herbach, R. Wanke, J. Handler and S. Bauersachs (2012). "Exploration of Global Gene Expression Changes During the Estrous Cycle in Equine Endometrium." *Biol Reprod* **87**(6): 136, 1-13.

Gerber, V., D. Baleri, J. Klukowska-Rotzler, J. E. Swinburne and G. Dolf (2009). "Mixed Inheritance of Equine Recurrent Airway Obstruction." *J Vet Intern Med* **23**(3): 626-630.

Gerber, V., N. E. Robinson, R. J. Venta, J. Rawson, A. M. Jefcoat and J. A. Hotchkiss (2003). "Mucin Genes in Horse Airways: *MUC5AC*, but Not *MUC2*, May Play a Role in Recurrent Airway Obstruction." *Equine Vet J* **35**(3): 252-257.

Gerber, V., R. Straub, E. Marti, J. Hauptman, C. Herholz, M. King, A. Imhof, L. Tahon and N. E. Robinson (2004). "Endoscopic Scoring of Mucus Quantity and Quality: Observer and Horse Variance and Relationship to Inflammation, Mucus Viscoelasticity and Volume." *Equine Vet J* **36**(7): 576-582.

Girirajan, S. and E. E. Eichler (2010). "Phenotypic Variability and Genetic Susceptibility to Genomic Disorders." *Hum Mol Genet* **19**(R2): R176-R187.

Giuffra, E., A. Tornsten, S. Marklund, E. Bongcam-Rudloff, P. Chardon, J. M. Kijas, S. I. Anderson, A. L. Archibald and L. Andersson (2002). "A Large Duplication Associated with Dominant White Color in Pigs Originated by Homologous Recombination between LINE Elements Flanking Kit." *Mamm Genome* **13**(10): 569-577.

Glaser, K. E., Q. Sun, M. T. Wells and A. J. Nixon (2009). "Development of a Novel Equine Whole Transcript Oligonucleotide Genechip Microarray and Its Use in Gene Expression Profiling of Normal Articular-Epiphyseal Cartilage." *Equine Vet J* **41**(7): 663-670.

Go, Y. Y., E. Bailey, D. G. Cook, S. J. Coleman, J. N. Macleod, K. C. Chen, P. J. Timoney and U. B. Balasuriya (2011). "Genome-Wide Association Study among Four Horse Breeds Identifies a Common Haplotype Associated with in Vitro Cd3+ T Cell Susceptibility/Resistance to Equine Arteritis Virus Infection." *J Virol* **85**(24): 13174-13184.

Goh, G., T. Raudsepp, K. Durkin, M. L. Wagner, A. A. Schaffer, R. Agarwala, T. Tozaki, J. R. Mickelson and B. P. Chowdhary (2007). "High-Resolution Gene Maps of Horse Chromosomes 14 and 21: Additional Insights into Evolution and Rearrangements of HSA5 Homologs in Mammals." *Genomics* **89**(1): 89-112.

Gokcumen, O., P. L. Babb, R. C. Iskow, Q. Zhu, X. Shi, R. E. Mills, I. Ionita-Laza, E. J. Vallender, A. G. Clark, W. E. Johnson, et al. (2011). "Refinement of Primate Copy Number Variation Hotspots Identifies Candidate Genomic Regions Evolving under Positive Selection." *Genome Biol* **12**(5): R52, 1-11.

Goldlust, I. S., K. E. Hermetz, L. M. Catalano, R. T. Barfield, R. Cozad, G. Wynn, A. C. Ozdemir, K. N. Conneely, J. G. Mulle, S. Dharamrup, et al. (2013). "Mouse Model Implicates *GNB3* Duplication in a Childhood Obesity Syndrome." *Proc Natl Acad Sci USA* **110**(37): 14990-14994.

Gorrie, S., P. Roome, C. Austen and N. J. Swinney (2008). *The Complete Illustrated Encyclopedia of Horses*, Metro Books, New York: 1-352.

Graubert, T. A., P. Cahan, D. Edwin, R. R. Selzer, T. A. Richmond, P. S. Eis, W. D. Shannon, X. Li, H. L. McLeod, J. M. Cheverud, et al. (2007). "A High-Resolution Map of Segmental DNA Copy Number Variation in the Mouse Genome." *PLoS Genet* **3**(1): e3, 1-9.

Grayton, H. M., C. Fernandes, D. Rujescu and D. A. Collier (2012). "Copy Number Variations in Neurodevelopmental Disorders." *Prog Neurobiol* **99**(1): 81-91.

Gresham, D., M. J. Dunham and D. Botstein (2008). "Comparing Whole Genomes Using DNA Microarrays." *Nat Rev Genet* **9**(4): 291-302.

Griffin, D. K., L. B. Robertson, H. G. Tempest, A. Vignal, V. Fillon, R. P. Crooijmans, M. A. Groenen, S. Deryusheva, E. Gaginskaya, W. Carre, et al. (2008). "Whole Genome Comparative Studies between Chicken and Turkey and Their Implications for Avian Genome Evolution." *BMC Genomics* **9**: 168, 1-16.

Gu, W. and A. L. Bertone (2004). "Generation and Performance of an Equine-Specific Large-Scale Gene Expression Microarray." *Am J Vet Res* **65**(12): 1664-1673.

Gubbay, J., J. Collignon, P. Koopman, B. Capel, A. Economou, A. Munsterberg, N. Vivian, P. Goodfellow and R. Lovell-Badge (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**(6281): 245-250.

Guerin, G., E. Bailey, D. Bernoco, I. Anderson, D. F. Antczak, K. Bell, M. M. Binns, A. T. Bowling, R. Brandon, G. Cholewinski, et al. (1999). "Report of the International Equine Gene Mapping Workshop: Male Linkage Map." *Anim Genet* **30**(5): 341-354.

Gurgul, A., K. Zukowski, B. Slaska, E. Semik, K. Pawlina, T. Zabek, I. Jasielczuk and M. Bugno-Poniewierska (2014). "General Assessment of Copy Number Variation in Normal and Tumor Tissues of the Domestic Dog (*Canis lupus familiaris*)." *J Appl Genet* **55**(3): 353-363.

Guryev, V., K. Saar, T. Adamovic, M. Verheul, S. A. van Heesch, S. Cook, M. Pravenec, T. Aitman, H. Jacob, J. D. Shull, et al. (2008). "Distribution and Functional Impact of DNA Copy Number Variation in the Rat." *Nat Genet* **40**(5): 538-545.

Gustafson-Seabury, A., T. Raudsepp, G. Goh, S. R. Kata, M. L. Wagner, T. Tozaki, J. R. Mickelson, J. E. Womack, L. C. Skow and B. P. Chowdhary (2005). "High-Resolution RH Map of Horse Chromosome 22 Reveals a Putative Ancestral Vertebrate Chromosome." *Genomics* **85**(2): 188-200.

Gustafson, A. L., R. L. Tallmadge, N. Ramlachan, D. Miller, H. Bird, D. F. Antczak, T. Raudsepp, B. P. Chowdhary and L. C. Skow (2003). "An Ordered BAC Contig Map of the Equine Major Histocompatibility Complex." *Cytogenet Genome Res* **102**(1-4): 189-195.

Haase, B., S. A. Brooks, A. Schlumbaum, P. J. Azor, E. Bailey, F. Alaeddine, M. Mevissen, D. Burger, P. A. Poncet, S. Rieder, et al. (2007). "Allelic Heterogeneity at the Equine *KIT* Locus in Dominant White (W) Horses." *PLoS Genet* **3**(11): e195, 1-8.

Haase, B., S. A. Brooks, T. Tozaki, D. Burger, P. A. Poncet, S. Rieder, T. Hasegawa, C. Penedo and T. Leeb (2009). "Seven Novel *KIT* Mutations in Horses with White Coat Colour Phenotypes." *Anim Genet* **40**(5): 623-629.

Haase, B., R. Jude, S. A. Brooks and T. Leeb (2008). "An Equine Chromosome 3 Inversion is Associated with the Tobiano Spotting Pattern in German Horse Breeds." *Anim Genet* **39**(3): 306-309.

Hall, I. M. and A. R. Quinlan (2012). "Detection and Interpretation of Genomic Structural Variation in Mammals." *Methods Mol Biol* **838**: 225-248.

Hamann, H., R. Jude, H. Sieme, U. Mertens, E. Topfer-Petersen, O. Distl and T. Leeb (2007). "A Polymorphism within the Equine *CRISP3* Gene is Associated with Stallion Fertility in Hanoverian Warmblood Horses." *Anim Genet* **38**(3): 259-264.

Haraksingh, R. R., A. Abyzov, M. Gerstein, A. E. Urban and M. Snyder (2011). "Genome-Wide Mapping of Copy Number Variation in Humans: Comparative Analysis of High Resolution Array Platforms." *PLoS One* **6**(11): e27859, 1-12.

Haraksingh, R. R. and M. P. Snyder (2013). "Impacts of Variation in the Human Genome on Gene Regulation." *J Mol Biol* **425**(21): 3970-3977.

Harbitz, I., B. P. Chowdhary, H. Saether, J. G. Hauge and I. Gustavsson (1990). "A Porcine Genomic Glucosephosphate Isomerase Probe Detects a Multiallelic Restriction Fragment Length Polymorphism Assigned to Chromosome 10pter in Horse." *Hereditas* **112**(2): 151-156.

Hargreave, T. B. (2000). "Genetics and Male Infertility." *Curr Opin Obstet Gynecol* **12**(3): 207-219.

Harris, R. M., C. Finlayson, J. Weiss, L. Fisher, L. Hurley, T. Barrett, D. Emge, R. A. Bathgate, A. I. Agoulnik and J. L. Jameson (2010). "A Missense Mutation in *LRR8* of *RXFP2* is Associated with Cryptorchidism." *Mamm Genome* **21**(9-10): 442-449.

Hasin, Y., T. Olender, M. Khen, C. Gonzaga-Jauregui, P. M. Kim, A. E. Urban, M. Snyder, M. B. Gerstein, D. Lancet and J. O. Korbel (2008). "High-Resolution Copy-Number Variation Map Reflects Human Olfactory Receptor Diversity and Evolution." *PLoS Genet* **4**(11): e1000249, 1-14.

Hastings, P. J., J. R. Lupski, S. M. Rosenberg and G. Ira (2009). "Mechanisms of Change in Gene Copy Number." *Nat Rev Genet* **10**(8): 551-564.

Hayashida, M., M. Nagashima, Y. Satoh, R. Katoh, M. Tagami, S. Ide, S. Kasai, D. Nishizawa, Y. Ogai, J. Hasegawa, et al. (2008). "Analgesic Requirements after Major Abdominal Surgery are Associated with *OPRM1* Gene Polymorphism Genotype and Haplotype." *Pharmacogenomics* **9**(11): 1605-1616.

Hayes, H. M. (1986). "Epidemiological Features of 5009 Cases of Equine Cryptorchism." *Equine Vet J* **18**(6): 467-471.

Hayes, J. D., J. U. Flanagan and I. R. Jowsey (2005). "Glutathione Transferases." *Annu Rev Pharmacol Toxicol* **45**: 51-88.

He, Y., J. M. Hoskins and H. L. McLeod (2011). "Copy Number Variants in Pharmacogenetic Genes." *Trends Mol Med* **17**(5): 244-251.

Hendricks, B. L. (1995). *International Encyclopedia of Horse Breeds*, University of Oklahoma Press, Oklahoma: 1-512.

Hill, E. W., D. G. Bradley, M. Al-Barody, O. Ertugrul, R. K. Splan, I. Zakharov and E. P. Cunningham (2002). "History and Integrity of Thoroughbred Dam Lines Revealed in Equine mtDNA Variation." *Anim Genet* **33**(4): 287-294.

Hill, E. W., B. A. McGivney, J. Gu, R. Whiston and D. E. Machugh (2010). "A Genome-Wide SNP-Association Study Confirms a Sequence Variant (G.66493737C>T) in the Equine Myostatin (*MSTN*) Gene as the Most Powerful Predictor of Optimum Racing Distance for Thoroughbred Racehorses." *BMC Genomics* **11**: 552, 1-10.

Hill, E. W., D. P. Ryan and D. E. MacHugh (2012). "Horses for Courses: A DNA-Based Test for Race Distance Aptitude in Thoroughbred Racehorses." *Recent Pat DNA Gene Seq* **6**(3): 203-208.

Hitz, M. P., L. P. Lemieux-Perreault, C. Marshall, Y. Feroz-Zada, R. Davies, S. W. Yang, A. C. Lionel, G. D'Amours, E. Lemyre, R. Cullum, et al. (2012). "Rare Copy Number Variants Contribute to Congenital Left-Sided Heart Disease." *PLoS Genet* **8**(9): e1002903, 1-13.

Holl, H. M., T. L. Lear, R. D. Nolen-Walston, J. Slack and S. A. Brooks (2013). "Detection of Two Equine Trisomies Using SNP-CGH." *Mamm Genome* **24**(5-6): 252-256.

Hollox, E. J., U. Huffmeier, P. L. Zeeuwen, R. Palla, J. Lascorz, D. Rodijk-Olthuis, P. C. van de Kerkhof, H. Traupe, G. de Jongh, M. den Heijer, et al. (2008). "Psoriasis is Associated with Increased Beta-Defensin Genomic Copy Number." *Nat Genet* **40**(1): 23-25.

Hou, Y., D. M. Bickhart, H. Chung, J. L. Hutchison, H. D. Norman, E. E. Connor and G. E. Liu (2012a). "Analysis of Copy Number Variations in Holstein Cows Identify Potential Mechanisms Contributing to Differences in Residual Feed Intake." *Funct Integr Genomics* **12**(4): 717-723.

Hou, Y., D. M. Bickhart, M. L. Hvinden, C. Li, J. Song, D. A. Boichard, S. Fritz, A. Eggen, S. DeNise, G. R. Wiggans, et al. (2012b). "Fine Mapping of Copy Number Variations on Two Cattle Genome Assemblies Using High Density SNP Array." *BMC Genomics* **13**: 376, 1-10.

Hou, Y., G. E. Liu, D. M. Bickhart, M. F. Cardone, K. Wang, E. S. Kim, L. K. Matukumalli, M. Ventura, J. Song, P. M. VanRaden, et al. (2011). "Genomic Characteristics of Cattle Copy Number Variations." *BMC Genomics* **12**: 127, 1-11.

Hou, Y., G. E. Liu, D. M. Bickhart, L. K. Matukumalli, C. Li, J. Song, L. C. Gasbarre, C. P. Van Tassell and T. S. Sonstegard (2012c). "Genomic Regions Showing Copy Number Variations Associate with Resistance or Susceptibility to

Gastrointestinal Nematodes in Angus Cattle." *Funct Integr Genomics* **12**(1): 81-92.

Howden, K. J. (2004). "Androgen Insensitivity Syndrome in a Thoroughbred Mare (64, XY-Testicular Feminization)." *Can Vet J* **45**(6): 501-503.

Huang da, W., B. T. Sherman and R. A. Lempicki (2009a). "Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists." *Nucleic Acids Res* **37**(1): 1-13.

Huang da, W., B. T. Sherman and R. A. Lempicki (2009b). "Systematic and Integrative Analysis of Large Gene Lists Using David Bioinformatics Resources." *Nat Protoc* **4**(1): 44-57.

Huang, L., W. Zhu, C. P. Saunders, J. N. Macleod, M. Zhou, A. J. Stromberg and A. C. Bathke (2008). "A Novel Application of Quantile Regression for Identification of Biomarkers Exemplified by Equine Cartilage Microarray Data." *BMC Bioinformatics* **9**: 300, 1-8.

Huarte, M., M. Guttman, D. Feldser, M. Garber, M. J. Koziol, D. Kenzelmann-Broz, A. M. Khalil, O. Zuk, I. Amit, M. Rabani, et al. (2010). "A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response." *Cell* **142**(3): 409-419.

Huddleston, J., S. Ranade, M. Malig, F. Antonacci, M. Chaisson, L. Hon, P. H. Sudmant, T. A. Graves, C. Alkan, M. Y. Dennis, et al. (2014). "Reconstructing

Complex Regions of Genomes Using Long-Read Sequencing Technology." *Genome Res* **24**(4):688-696.

Hughes, I. A. (2008). "Disorders of Sex Development: A New Definition and Classification." *Best Pract Res Clin Endocrinol Metab* **22**(1): 119-134.

Hughes, I. A. and C. L. Acerini (2008). "Factors Controlling Testis Descent." *European Journal of Endocrinology* **159 Suppl 1**: S75-S82.

Hultgren, B. D. (1982). "Ileocolonic Aganglionosis in White Progeny of Overo Spotted Horses." *J Am Vet Med Assoc* **180**(3): 289-292.

Hurles, M. and J. Lupski (2006). Recombination Hotspots in Nonallelic Homologous Recombination. *Genomic Disorders*. J. Lupski and P. Stankiewicz (ed), Humana Press**:** 341-355.

Huyhn, K., M. B. Renfree, J. A. Graves and A. J. Pask (2011). "*ATRX* Has a Critical and Conserved Role in Mammalian Sexual Differentiation." *BMC Dev Biol* **11**: 39, 1-10.

Iafrate, A. J., L. Feuk, M. N. Rivera, M. L. Listewnik, P. K. Donahoe, Y. Qi, S. W. Scherer and C. Lee (2004). "Detection of Large-Scale Variation in the Human Genome." *Nat Genet* **36**(9): 949-951.

Iben, J. R. and R. J. Maraia (2014). "tRNA Gene Copy Number Variation in Humans." *Gene* **536**(2): 376-384.

ISCNH (1997). "International System for Cytogenetic Nomenclature of the Domestic Horse. Bowling, A. T., Breen, M., Chowdhary, B. P., Hirota, K., Lear, T., Millon, L. V., Ponce De Leon, F. A., Raudsepp, T., Stranzinger, G. Report of the Third International Committee for the Standardization of the Domestic Horse Karyotype, Davis, Ca, USA, 1996." *Chromosome Res* **5**(7): 433-443.

Itsara, A., H. Wu, J. D. Smith, D. A. Nickerson, I. Romieu, S. J. London and E. E. Eichler (2010). "De Novo Rates and Selection of Large Copy Number Variation." *Genome Res* **20**(11): 1469-1481.

Jacobs, P. A. and J. A. Strong (1959). "A Case of Human Intersexuality Having a Possible XXY Sex-Determining Mechanism." *Nature* **183**(4657): 302-303.

Jaradat, S. W., C. Hoder-Przyrembel, S. Cubillos, N. Krieg, K. Lehmann, S. Piehler, B. W. Sigusch and J. Norgauer (2013). "Beta-Defensin-2 Genomic Copy Number Variation and Chronic Periodontitis." *J Dent Res* **92**(11): 1035-1040.

Jarick, I., C. I. Vogel, S. Scherag, H. Schafer, J. Hebebrand, A. Hinney and A. Scherag (2011). "Novel Common Copy Number Variation for Early Onset Extreme Obesity on Chromosome 11q11 Identified by a Genome-Wide Analysis." *Hum Mol Genet* **20**(4): 840-852.

Jia, X., S. Chen, H. Zhou, D. Li, W. Liu and N. Yang (2012). "Copy Number Variations Identified in the Chicken Using a 60K SNP Beadchip." *Anim Genet* **44**(3):276-284.

Jiang, L., J. Jiang, J. Wang, X. Ding, J. Liu and Q. Zhang (2012). "Genome-Wide Identification of Copy Number Variations in Chinese Holstein." *PLoS One* **7**(11): e48732, 1-8.

Jiang, L., J. Jiang, J. Yang, X. Liu, J. Wang, H. Wang, X. Ding, J. Liu and Q. Zhang (2013). "Genome-Wide Detection of Copy Number Variations Using High-Density SNP Genotyping Platforms in Holsteins." *BMC Genomics* **14**: 131, 1-10.

Jimenez, R., F. J. Barrionuevo and M. Burgos (2013). "Natural Exceptions to Normal Gonad Development in Mammals." *Sex Dev* **7**(1-3): 147-162.

Jost, U., J. Klukowska-Rotzler, G. Dolf, J. E. Swinburne, A. Ramseyer, M. Bugno, D. Burger, S. Blott and V. Gerber (2007). "A Region on Equine Chromosome 13 is Linked to Recurrent Airway Obstruction in Horses." *Equine Vet J* **39**(3): 236-241.

Jung, S. H., S. H. Yim, H. J. Oh, J. E. Park, M. J. Kim, G. A. Kim, T. M. Kim, J. S. Kim, B. C. Lee and Y. J. Chung (2013). "De Novo Copy Number Variations in Cloned Dogs from the Same Nuclear Donor." *BMC Genomics* **14**: 863, 1-11.

Kachroo, P., A. Gustafson-Seabury, J. Lefebre-Lavoie, J.-P. Lavoie, I. Ivanov and B. P. Chowdhary (2010). Gene Expression Profiling in Equine Recurrent Airway Obstruction (RAO): Detection of Differentially Expressed Genes and Their Role in Pathogenesis of the Disease. *Plant and Animal Genome XVIII*, January 9-13, San Diego, CA, USA.

Kachroo, P., I. Ivanov, A. G. Seabury, M. Liu, B. P. Chowdhary and N. D. Cohen (2013). "Age-Related Changes Following in Vitro Stimulation with *Rhodococcus equi* of Peripheral Blood Leukocytes from Neonatal Foals." *PLoS One* **8**(5): e62879, 1-9.

Kalbfleisch, T., J. Rebolledo-Mendez, A. Ginolhac, L. Orlando and J. N. MacLeod (2013). Strategy, Resources, and Progress Towards Equcab3. *10th Dorothy Russell Havemeyer Foundation, International Equine Genome Mapping* Workshop, Azores, Portugal.

Kamm, J. L., D. D. Frisbie, C. W. McIlwraith and K. E. Orr (2013). "Gene Biomarkers in Peripheral White Blood Cells of Horses with Experimentally Induced Osteoarthritis." *Am J Vet Res* **74**(1): 115-121.

Kane, M. D., T. A. Jatkoe, C. R. Stumpf, J. Lu, J. D. Thomas and S. J. Madore (2000). "Assessment of the Sensitivity and Specificity of Oligonucleotide (50mer) Microarrays." *Nucleic Acids Res* **28**(22): 4552-4557.

Kantaputra, P. N., E. Klopocki, B. P. Hennig, V. Praphanphoj, C. Le Caignec, B. Isidor, M. L. Kwee, D. J. Shears and S. Mundlos (2010). "Mesomelic Dysplasia Kantaputra Type is Associated with Duplications of the *HOXd* Locus on Chromosome 2q." *Eur J Hum Genet* **18**(12): 1310-1314.

Karolchik, D., R. Baertsch, M. Diekhans, T. S. Furey, A. Hinrichs, Y. T. Lu, K. M. Roskin, M. Schwartz, C. W. Sugnet, D. J. Thomas, et al. (2003). "The UCSC Genome Browser Database." *Nucleic Acids Res* **31**(1): 51-54.

Karyadi, D. M., E. Karlins, B. Decker, B. M. vonHoldt, G. Carpintero-Ramirez, H. G. Parker, R. K. Wayne and E. A. Ostrander (2013). "A Copy Number Variant at the *KITlg* Locus Likely Confers Risk for Canine Squamous Cell Carcinoma of the Digit." *PLoS Genet* **9**(3): e1003409, 1-14.

Kehrer-Sawatzki, H. and D. N. Cooper (2007). "Structural Divergence between the Human and Chimpanzee Genomes." *Hum Genet* **120**(6): 759-778.

Kent, M. G., H. E. Schneller, R. L. Hegsted, S. D. Johnston and S. S. Wachtel (1988). "Concentration of Serum Testosterone in XY Sex Reversed Horses." *J Endocrinol Invest* **11**(8): 609-613.

Kent, M. G., R. N. Shoffner, L. Buoen and A. F. Weber (1986). "XY Sex-Reversal Syndrome in the Domestic Horse." *Cytogenet Cell Genet* **42**(1-2): 8-18.

Khalil, A. M., M. Guttman, M. Huarte, M. Garber, A. Raj, D. Rivea Morales, K. Thomas, A. Presser, B. E. Bernstein, A. van Oudenaarden, et al. (2009). "Many Human Large Intergenic Noncoding RNAs Associate with Chromatin-Modifying Complexes and Affect Gene Expression." *Proc Natl Acad Sci USA* **106**(28): 11667-11672.

Khan, F. F., D. Carpenter, L. Mitchell, O. Mansouri, H. A. Black, J. Tyson and J. A. Armour (2013). "Accurate Measurement of Gene Copy Number for Human Alpha-Defensin *DEFA1A3*." *BMC Genomics* **14**: 719, 1-12.

Khanshour, A., E. Conant, R. Juras and E. G. Cothran (2013). "Microsatellite Analysis of Genetic Diversity and Population Structure of Arabian Horse Populations." *J Hered* **104**(3): 386-398.

Khanshour, A. M. and E. G. Cothran (2013). "Maternal Phylogenetic Relationships and Genetic Variation among Arabian Horse Populations Using Whole Mitochondrial DNA D-Loop Sequencing." *BMC Genet* **14**: 83, 1-12.

Kidd, J. M., G. M. Cooper, W. F. Donahue, H. S. Hayden, N. Sampas, T. Graves, N. Hansen, B. Teague, C. Alkan, F. Antonacci, et al. (2008). "Mapping and Sequencing of Structural Variation from Eight Human Genomes." *Nature* **453**(7191): 56-64.

Kim, J. H., S. H. Jung, J. S. Bae, H. S. Lee, S. H. Yim, S. Y. Park, S. Y. Bang, H. J. Hu, H. D. Shin, S. C. Bae, et al. (2013). "Deletion Variants of *RABGAP1l*, 10q21.3, and C4 Are Associated with the Risk of Systemic Lupus Erythematosus in Korean Women." *Arthritis Rheum* **65**(4): 1055-1063.

Kim, P. M., H. Y. Lam, A. E. Urban, J. O. Korbel, J. Affourtit, F. Grubert, X. Chen, S. Weissman, M. Snyder and M. B. Gerstein (2008). "Analysis of Copy Number Variants and Segmental Duplications in the Human Genome: Evidence for a

Change in the Process of Formation in Recent Evolutionary History." *Genome Res* **18**(12): 1865-1874.

Kirov, G., E. Rees, J. T. Walters, V. Escott-Price, L. Georgieva, A. L. Richards, K. D. Chambert, G. Davies, S. E. Legge, J. L. Moran, et al. (2014). "The Penetrance of Copy Number Variations for Schizophrenia and Developmental Delay." *Biol Psychiatry* **75**(5): 378-385.

Klonisch, T., P. A. Fowler and S. Hombach-Klonisch (2004). "Molecular and Genetic Regulation of Testis Descent and External Genitalia Development." *Developmental Biology* **270**(1): 1-18.

Klukowska-Rotzler, J., V. Gerber and T. Leeb (2012a). "Association Analysis of SNPs in the *IL21R* Gene with Recurrent Airway Obstruction (RAO) in Swiss Warmblood Horses." *Anim Genet* **43**(4): 475-476.

Klukowska-Rotzler, J., J. E. Swinburne, C. Drogemuller, G. Dolf, J. Janda, T. Leeb and V. Gerber (2012b). "The Interleukin 4 Receptor Gene and Its Role in Recurrent Airway Obstruction in Swiss Warmblood Horses." *Anim Genet* **43**(4): 450-453.

Koopman, P., J. Gubbay, N. Vivian, P. Goodfellow and R. Lovell-Badge (1991). "Male Development of Chromosomally Female Mice Transgenic for *SRY*." *Nature* **351**(6322): 117-121.

Koressaar, T. and M. Remm (2007). "Enhancements and Modifications of Primer Design Program Primer3." *Bioinformatics* **23**(10): 1289-1291.

Krausz, C. and C. Giachini (2007). "Genetic Risk Factors in Male Infertility." *Arch Androl* **53**(3): 125-133.

Krumm, N., B. J. O'Roak, E. Karakoc, K. Mohajeri, B. Nelson, L. Vives, S. Jacquemont, J. Munson, R. Bernier and E. E. Eichler (2013). "Transmission Disequilibrium of Small Cnvs in Simplex Autism." *Am J Hum Genet* **93**(4): 595-606.

Kulbrock, M., S. Lehner, J. Metzger, B. Ohnesorge and O. Distl (2013). "A Genome-Wide Association Study Identifies Risk Loci to Equine Recurrent Uveitis in German Warmblood Horses." *PLoS One* **8**(8): e71619, 1-6.

Kulski, J. K., T. Shiina, T. Anzai, S. Kohara and H. Inoko (2002). "Comparative Genomic Analysis of the MHC: The Evolution of Class I Duplication Blocks, Diversity and Complexity from Shark to Man." *Immunol Rev* **190**: 95-122.

Kurth, I., E. Klopocki, S. Stricker, J. van Oosterwijk, S. Vanek, J. Altmann, H. G. Santos, J. J. van Harssel, T. de Ravel, A. O. Wilkie, et al. (2009). "Duplications of Noncoding Elements 5' of *SOX9* Are Associated with Brachydactyly-Anonychia." *Nat Genet* **41**(8): 862-863.

Lahn, B. T. and D. C. Page (1997). "Functional Coherence of the Human Y Chromosome." *Science* **278**(5338): 675-680.

Lanz, S., V. Gerber, E. Marti, H. Rettmer, J. Klukowska-Rotzler, B. Gottstein, J. B. Matthews, S. Pirie and E. Hamza (2013). "Effect of Hay Dust Extract and Cyathostomin Antigen Stimulation on Cytokine Expression by *PBMC* in Horses

with Recurrent Airway Obstruction." *Vet Immunol Immunopathol* **155**(4): 229-237.

Laumen, E., M. G. Doherr and V. Gerber (2010). "Relationship of Horse Owner Assessed Respiratory Signs Index to Characteristics of Recurrent Airway Obstruction in Two Warmblood Families." *Equine Vet J* **42**(2): 142-148.

Lavoie-Lamoureux, A., K. Moran, G. Beauchamp, S. Mauel, F. Steinbach, J. Lefebvre-Lavoie, J. G. Martin and J. P. Lavoie (2010). "*IL-4* Activates Equine Neutrophils and Induces a Mixed Inflammatory Cytokine Expression Profile with Enhanced Neutrophil Chemotactic Mediator Release ex-vivo." *Am J Physiol Lung Cell Mol Physiol* **299**(4): L472-L482.

Lavoie, J. P., K. Maghni, M. Desnoyers, R. Taha, J. G. Martin and Q. A. Hamid (2001). "Neutrophilic Airway Inflammation in Horses with Heaves is Characterized by a Th2-Type Cytokine Profile." *Am J Respir Crit Care Med* **164**(8 Pt 1): 1410-1413.

Lear, T. L. and E. Bailey (2008). "Equine Clinical Cytogenetics: The Past and Future." *Cytogenet Genome Res* **120**(1-2): 42-49.

Lear, T. L. and R. B. McGee (2012). "Disorders of Sexual Development in the Domestic Horse, *Equus caballus*." *Sex Dev* **6**(1-3): 61-71.

Lear, T. L., K. A. Trembicki and R. B. Ennis (1992). "Identification of Equine Chromosomes in Horse X Mouse Somatic Cell Hybrids." *Cytogenet Cell Genet* **61**(1): 58-60.

Lebman, D. A. and R. L. Coffman (1988). "Interleukin 4 Causes Isotype Switching to IGE in T Cell-Stimulated Clonal B Cell Cultures." *J Exp Med* **168**(3): 853-862.

Ledig, S., O. Hiort, G. Scherer, M. Hoffmann, G. Wolff, S. Morlot, A. Kuechler and P. Wieacker (2010a). "Array-CGH Analysis in Patients with Syndromic and Non-Syndromic XY Gonadal Dysgenesis: Evaluation of Array CGH as Diagnostic Tool and Search for New Candidate Loci." *Hum Reprod* **25**(10): 2637-2646.

Ledig, S., A. Ropke and P. Wieacker (2010b). "Copy Number Variants in Premature Ovarian Failure and Ovarian Dysgenesis." *Sex Dev* **4**(4-5): 225-232.

Lee, C. and S. W. Scherer (2010). "The Clinical Context of Copy Number Variation in the Human Genome." *Expert Rev Mol Med* **12**: e8, 1-29.

Lee, E. J., T. Raudsepp, S. R. Kata, D. Adelson, J. E. Womack, L. C. Skow and B. P. Chowdhary (2004). "A 1.4-Mb Interval RH Map of Horse Chromosome 17 Provides Detailed Comparison with Human and Mouse Homologues." *Genomics* **83**(2): 203-215.

Lee, K. W., P. S. Woon, Y. Y. Teo and K. Sim (2012). "Genome Wide Association Studies (GWAS) and Copy Number Variation (CNV) Studies of the Major Psychoses: What Have We Learnt?" *Neurosci Biobehav Rev* **36**(1): 556-571.

Lee, S. H., J. S. Park and C. S. Park (2011). "The Search for Genetic Variants and Epigenetics Related to Asthma." *Allergy Asthma Immunol Res* **3**(4): 236-244.

Leeb, T., C. Vogl, B. Zhu, P. J. de Jong, M. M. Binns, B. P. Chowdhary, M. Scharfe, M. Jarek, G. Nordsiek, F. Schrader, et al. (2006). "A Human-Horse Comparative Map Based on Equine BAC End Sequences." *Genomics* **87**(6): 772-776.

Levine, M. A. (2005). Domestication and Early History of Horse. *The Domestic Horse: The Origins, Development and Management of Its Behaviour*. D. S. Mills and S. M. McDonnell(ed). Cabridge, UK, Cambridge University Press: 1-264.

Li, G., B. W. Davis, T. Raudsepp, A. J. Pearks Wilkerson, V. C. Mason, M. Ferguson-Smith, P. C. O'Brien, P. D. Waters and W. J. Murphy (2013). "Comparative Analysis of Mammalian Y Chromosomes Illuminates Ancestral Structure and Lineage-Specific Evolution." *Genome Res* **23**(9): 1486-1495.

Li, H., K. W. Choy, Y. P. Lei, W. Wang, H. Y. Wang and Y. Chen (2011). "A Novel 15bp Micro-Duplication in *SF-1* Gene Showing Diverse Phenotypic Spectrum in a Chinese Family." *J Matern Fetal Neonatal Med* **24**(1): 132-136.

Li, J., T. Jiang, J. H. Mao, A. Balmain, L. Peterson, C. Harris, P. H. Rao, P. Havlak, R. Gibbs and W. W. Cai (2004). "Genomic Segmental Polymorphisms in Inbred Mouse Strains." *Nat Genet* **36**(9): 952-954.

Li, Y., S. Mei, X. Zhang, X. Peng, G. Liu, H. Tao, H. Wu, S. Jiang, Y. Xiong and F. Li (2012). "Identification of Genome-Wide Copy Number Variations among Diverse Pig Breeds by Array CGH." *BMC Genomics* **13**(1): 725, 1-9.

Liao, A. T., P. Y. Chu, L. S. Yeh, C. T. Lin and C. H. Liu (2009). "A 12-Year Retrospective Study of Canine Testicular Tumors." *J Vet Med Sci* **71**(7): 919-923.

Lin, C. H., Y. C. Lin, J. Y. Wu, W. H. Pan, Y. T. Chen and C. S. Fann (2009). "A Genome-Wide Survey of Copy Number Variations in Han Chinese Residing in Taiwan." *Genomics* **94**(4): 241-246.

Lindgren, G., K. Sandberg, H. Persson, S. Marklund, M. Breen, B. Sandgren, J. Carlsten and H. Ellegren (1998). "A Primary Male Autosomal Linkage Map of the Horse Genome." *Genome Res* **8**(9): 951-966.

Linja-aho, A., W. Mazur, T. Toljamo, P. Nieminen, S. Ohlmeier, M. Ronty and V. L. Kinnula (2013). "Distribution and Levels of Alpha-1-Antitrypsin in the Lung and Plasma in Smokers and Chronic Obstructive Pulmonary Disease." *Acta Pathologica, Microbiologica, et Immunologica Scandinavica* **121**(1): 11-21.

Lipson, D., Y. Aumann, A. Ben-Dor, N. Linial and Z. Yakhini (2006). "Efficient Calculation of Interval Scores for DNA Copy Number Data Analysis." *J Comput Biol* **13**(2): 215-228.

Liu, G. E. and D. M. Bickhart (2012). "Copy Number Variation in the Cattle Genome." *Funct Integr Genomics* **12**(4): 609-624.

Liu, G. E., Y. Hou, B. Zhu, M. F. Cardone, L. Jiang, A. Cellamare, A. Mitra, L. J. Alexander, L. L. Coutinho, M. E. Dell'Aquila, et al. (2010). "Analysis of Copy Number Variations among Diverse Cattle Breeds." *Genome Res* **20**(5): 693-703.

Liu, G. E., C. P. Van Tassel, T. S. Sonstegard, R. W. Li, L. J. Alexander, J. W. Keele, L. K. Matukumalli, T. P. Smith and L. C. Gasbarre (2008). "Detection of Germline and Somatic Copy Number Variations in Cattle." *Dev Biol (Basel)* **132**: 231-237.

Liu, J., L. Zhang, L. Xu, H. Ren, J. Lu, X. Zhang, S. Zhang, X. Zhou, C. Wei, F. Zhao, et al. (2013). "Analysis of Copy Number Variations in the Sheep Genome Using 50K SNP Beadchip Array." *BMC Genomics* **14**: 229, 1-11.

Liu, W., S. Laitinen, S. Khan, M. Vihinen, J. Kowalski, G. Yu, L. Chen, C. M. Ewing, M. A. Eisenberger, M. A. Carducci, et al. (2009). "Copy Number Analysis Indicates Monoclonal Origin of Lethal Metastatic Prostate Cancer." *Nat Med* **15**(5): 559-565.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-DDCT}$ Method." *Methods* **25**(4): 402-408.

Llaurens, V., M. McMullan and C. van Oosterhout (2012). "Cryptic MHC Polymorphism Revealed but Not Explained by Selection on the Class IIb Peptide-Binding Region." *Mol Biol Evol* **29**(6): 1631-1644.

Lo-Castro, A., G. Giana, M. Fichera, L. Castiglia, L. Grillo, S. A. Musumeci, C. Galasso and P. Curatolo (2009). "Deletion 2p25.2: A Cryptic Chromosome Abnormality in a Patient with Autism and Mental Retardation Detected Using aCGH." *Eur J Med Genet* **52**(1): 67-70.

Lupski, J. R. (2007a). "An Evolution Revolution Provides Further Revelation." *Bioessays* **29**(12): 1182-1184.

Lupski, J. R. (2007b). "Genomic Rearrangements and Sporadic Disease." *Nat Genet* **39**(7 Suppl): S43-S47.

Lupski, J. R. (2013). "Genetics. Genome Mosaicism-One Human, Multiple Genomes." *Science* **341**(6144): 358-359.

Lykkjen, S., N. I. Dolvik, M. E. McCue, A. K. Rendahl, J. R. Mickelson and K. H. Roed (2010). "Genome-Wide Association Analysis of Osteochondrosis of the Tibiotarsal Joint in Norwegian Standardbred Trotters." *Anim Genet* **41 Suppl 2**: 111-120.

Lykkjen, S., N. I. Dolvik, M. E. McCue, A. K. Rendahl, J. R. Mickelson and K. H. Roed (2013). "Equine Developmental Orthopaedic Diseases-a Genome-Wide Association Study of First Phalanx Plantar Osteochondral Fragments in Standardbred Trotters." *Anim Genet* **44**(6): 766-769.

MacDonald, J. R., R. Ziman, R. K. Yuen, L. Feuk and S. W. Scherer (2014). "The Database of Genomic Variants: A Curated Collection of Structural Variation in the Human Genome." *Nucleic Acids Res* **42**(Database issue): D986-D992.

Madian, A. G., H. E. Wheeler, R. B. Jones and M. E. Dolan (2012). "Relating Human Genetic Variation to Variation in Drug Responses." *Trends Genet* **28**(10): 487-495.

Makinen, A., B. Chowdhary, E. Mahdy, L. Andersson and I. Gustavsson (1989). "Localization of the Equine Major Histocompatibility Complex (ELA) to Chromosome 20 by in Situ Hybridization." *Hereditas* **110**(1): 93-96.

Marchani, E. E., N. H. Chapman, C. Y. Cheung, K. Ankenman, I. B. Stanaway, H. H. Coon, D. Nickerson, R. Bernier, Z. Brkanac and E. M. Wijsman (2012). "Identification of Rare Variants from Exome Sequence in a Large Pedigree with Autism." *Hum Hered* **74**(3-4): 153-164.

Mardis, E. R. (2008). "Next-Generation DNA Sequencing Methods." *Annu Rev Genomics Hum Genet* **9**: 387-402.

Mariat, D., S. Taourit and G. Guerin (2003). "A Mutation in the *MATP* Gene Causes the Cream Coat Colour in the Horse." *Genet Sel Evol* **35**(1): 119-133.

Marshall Graves, J. A. (2008). "Weird Animal Genomes and the Evolution of Vertebrate Sex and Sex Chromosomes." *Annu Rev Genet* **42**: 565-586.

Massart, F. and G. Saggese (2009). "Sex Steroidal Targets & Genetic Susceptibility to Idiopathic Cryptorchidism." *Pediatric Endocrinology Reviews* **6**(4): 481-490.

Massart, F. and G. Saggese (2010). "Morphogenetic Targets and Genetics of Undescended Testis." *Sexual development : Genetics, Molecular biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation* **4**(6): 326-335.

Mathai, C. K., S. Ohno and E. Beutler (1966). "Sex-Linkage of the Glucose-6-Phosphate Dehydrogenase Gene in Equidae." *Nature* **210**(31): 115-116.

Matsuzaki, H., P. H. Wang, J. Hu, R. Rava and G. K. Fu (2009). "High Resolution Discovery and Confirmation of Copy Number Variants in 90 Yoruba Nigerians." *Genome Biol* **10**(11): R125, 1-18.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. Smith, T. S. Sonstegard, et al. (2009). "Development and Characterization of a High Density SNP Genotyping Assay for Cattle." *PLoS One* **4**(4): e5350, 1-13.

Matzuk, M. M. and D. J. Lamb (2002). "Genetic Dissection of Mammalian Fertility Pathways." *Nat Cell Biol* **4 Suppl**: S41-S49.

Mau, C., P. A. Poncet, B. Bucher, G. Stranzinger and S. Rieder (2004). "Genetic Mapping of Dominant White (W), a Homozygous Lethal Condition in the Horse (*Equus caballus*)". *Journal of Animal Breeding and Genetics* **121**(6): 374-383.

McCabe, L., L. D. Griffin, A. Kinzer, M. Chandler, J. B. Beckwith and E. R. McCabe (1990). "Overo Lethal White Foal Syndrome: Equine Model of Aganglionic Megacolon (Hirschsprung Disease)." *Am J Med Genet* **36**(3): 336-340.

McCarroll, S. A. (2010). "Copy Number Variation and Human Genome Maps." *Nat Genet* **42**(5): 365-366.

McCarroll, S. A., A. Huett, P. Kuballa, S. D. Chilewski, A. Landry, P. Goyette, M. C. Zody, J. L. Hall, S. R. Brant, J. H. Cho, et al. (2008a). "Deletion Polymorphism Upstream of *IRGM* Associated with Altered *IRGM* Expression and Crohn's Disease." *Nat Genet* **40**(9): 1107-1112.

McCarroll, S. A., F. G. Kuruvilla, J. M. Korn, S. Cawley, J. Nemesh, A. Wysoker, M. H. Shapero, P. I. de Bakker, J. B. Maller, A. Kirby, et al. (2008b). "Integrated Detection and Population-Genetic Analysis of SNPs and Copy Number Variation." *Nat Genet* **40**(10): 1166-1174.

McCue, M., et al (2009). The Horse Gentrain Project: Initial Evaluation of the Equine SNP50 Beadcips. *The 8th Dorothy Russell Havemeyer Foundation International Equine Genome Mapping Workshop*, The West Wing at Ickworth, Suffolk, UK.

McCue, M. and J. Mickelson (2013). Genomic Tools and Resources: Developement and Applications of an Equine SNP Genotyping Array. *Equine Genomics*. B. P. Chowdhary (ed), Wiley-Blackwell: 113-124.

McCue, M. E. (2013). New SNP Array *Plant and Animal Genome XXI*, San Diego, CA, USA.

McCue, M. E., D. L. Bannasch, J. L. Petersen, J. Gurr, E. Bailey, M. M. Binns, O. Distl, G. Guerin, T. Hasegawa, E. W. Hill, et al. (2012). "A High Density SNP Array for the Domestic Horse and Extant Perissodactyla: Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies." *PLoS Genet* **8**(1): e1002451, 1-14.

McCue, M. E., S. J. Valberg, M. Jackson, L. Borgia, M. Lucio and J. R. Mickelson (2009). "Polysaccharide Storage Myopathy Phenotype in Quarter Horse-Related Breeds is Modified by the Presence of an *RYR1* Mutation." *Neuromuscul Disord* **19**(1): 37-43.

McCue, M. E., S. J. Valberg, M. B. Miller, C. Wade, S. DiMauro, H. O. Akman and J. R. Mickelson (2008). "Glycogen Synthase (*GYS1*) Mutation Causes a Novel Skeletal Muscle Glycogenosis." *Genomics* **91**(5): 458-466.

McGivney, B. A., J. A. Browne, R. G. Fonseca, L. M. Katz, D. E. Machugh, R. Whiston and E. W. Hill (2012). "*MSTN* Genotypes in Thoroughbred Horses Influence Skeletal Muscle Gene Expression and Racetrack Performance." *Anim Genet* **43**(6): 810-812.

McGorum, B. C., P. M. Dixon and R. E. Halliwell (1993). "Evaluation of Intradermal Mould Antigen Testing in the Diagnosis of Equine Chronic Obstructive Pulmonary Disease." *Equine Vet J* **25**(4): 273-275.

Medvedev, P., M. Stanciu and M. Brudno (2009). "Computational Methods for Discovering Structural Variation with Next-Generation Sequencing." *Nat Methods* **6**(11 Suppl): S13-S20.

Mefford, H. C., H. Muhle, P. Ostertag, S. von Spiczak, K. Buysse, C. Baker, A. Franke, A. Malafosse, P. Genton, P. Thomas, et al. (2010). "Genome-Wide Copy Number Variation in Epilepsy: Novel Susceptibility Loci in Idiopathic Generalized and Focal Epilepsies." *PLoS Genet* **6**(5): e1000962, 1-9.

Metallinos, D. L., A. T. Bowling and J. Rine (1998). "A Missense Mutation in the Endothelin-B Receptor Gene is Associated with Lethal White Foal Syndrome: An Equine Version of Hirschsprung Disease." *Mamm Genome* **9**(6): 426-431.

Metzger, J., B. Ohnesorge and O. Distl (2012). "Genome-Wide Linkage and Association Analysis Identifies Major Gene Loci for Guttural Pouch Tympany in Arabian and German Warmblood Horses." *PLoS One* **7**(7): e41640, 1-8.

Metzger, J., U. Philipp, M. S. Lopes, A. da Camara Machado, M. Felicetti, M. Silvestrelli and O. Distl (2013a). "Analysis of Copy Number Variants by Three Detection Algorithms and Their Association with Body Size in Horses." *BMC Genomics* **14**: 487, 1-15.

Metzger, J., R. Schrimpf, U. Philipp and O. Distl (2013b). "Expression Levels of *LCORl* are Associated with Body Size in Horses." *PLoS One* **8**(2): e56497, 1-9.

Meyers-Wallen, V. N., J. Hurtgen, D. Schlafer, E. Tulleners, W. R. Cleland, G. R. Ruth and G. M. Acland (1997). "*SRY*-Negative XX True Hermaphroditism in a Pasa Fino Horse." *Equine Vet J* **29**(5): 404-408.

Mienaltowski, M. J., L. Huang, D. D. Frisbie, C. W. McIlwraith, A. J. Stromberg, A. C. Bathke and J. N. Macleod (2009). "Transcriptional Profiling Differences for Articular Cartilage and Repair Tissue in Equine Joint Surface Lesions." *BMC Med Genomics* **2**: 60, 1-14.

Milenkovic, D., A. Oustry-Vaiman, T. L. Lear, A. Billault, D. Mariat, F. Piumi, L. Schibler, E. Cribiu and G. Guerin (2002). "Cytogenetic Localization of 136 Genes in the Horse: Comparative Mapping with the Human Genome." *Mamm Genome* **13**(9): 524-534.

Milliken, J. E., D. L. Paccamonti, S. Shoemaker and W. H. Green (1995). "XX Male Pseudohermaphroditism in a Horse." *J Am Vet Med Assoc* **207**(1): 77-79.

Mills, R. E., K. Walter, C. Stewart, R. E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S. C. Yoon, K. Ye, R. K. Cheetham, et al. (2011). "Mapping Copy Number Variation by Population-Scale Genome Sequencing." *Nature* **470**(7332): 59-65.

Mitra, A., G. Liu and J. Song (2009). "A Genome-Wide Analysis of Array-Based Comparative Genomic Hybridization (CGH) Data to Detect Intra-Species Variations and Evolutionary Relationships." *PLoS One* **4**(11): e7978, 1-9.

Mitsunaga, S., K. Hosomichi, Y. Okudaira, H. Nakaoka, N. Kunii, Y. Suzuki, M. Kuwana, S. Sato, Y. Kaneko, Y. Homma, et al. (2013). "Exome Sequencing Identifies Novel Rheumatoid Arthritis-Susceptible Variants in the *BTNL2*." *J Hum Genet* **58**(4): 210-215.

Mittmann, E. H., V. Lampe, S. Momke, A. Zeitz and O. Distl (2010a). "Characterization of a Minimal Microsatellite Set for Whole Genome Scans Informative in Warmblood and Coldblood Horse Breeds." *J Hered* **101**(2): 246-250.

Mittmann, E. H., S. Momke and O. Distl (2010b). "Whole-Genome Scan Identifies Quantitative Trait Loci for Chronic Pastern Dermatitis in German Draft Horses." *Mamm Genome* **21**(1-2): 95-103.

Mittmann, E. H., J. Wrede, J. Pook and O. Distl (2010c). "Identification of 21781 Equine Microsatellites on the Horse Genome Assembly 2.0." *Anim Genet* **41**(2): 222.

Mizuno, K., Y. Kojima, H. Kamisawa, Y. Moritoki, H. Nishio, A. Nakane, S. Kurokawa, K. Kohri and Y. Hayashi (2014). "Elucidation of Distinctive Genomic DNA Structures in Patients with 46,XX Testicular Disorders of Sex Development Using Genome-Wide Analyses." *J Urol* **192**(2):535-541.

Mogil, J. S. (2012). "Pain Genetics: Past, Present and Future." *Trends Genet* **28**(6): 258-266.

Molin, A. M., J. Berglund, M. T. Webster and K. Lindblad-Toh (2014). "Genome-Wide Copy Number Variant Discovery in Dogs Using the CanineHD Genotyping Array." *BMC Genomics* **15**(1): 210, 1-10.

Monfort, S., M. Rosello, C. Orellana, S. Oltra, D. Blesa, K. Kok, I. Ferrer, J. C. Cigudosa and F. Martinez (2008). "Detection of Known and Novel Genomic Rearrangements by Array Based Comparative Genomic Hybridisation: Deletion of *ZNF533* and Duplication of Charge Syndrome Genes." *J Med Genet* **45**(7): 432-437.

Moon, S., Y. J. Kim, C. B. Hong, D. J. Kim, J. Y. Lee and B. J. Kim (2011). "Data-Driven Approach to Detect Common Copy-Number Variations and Frequency Profiles in a Population-Based Korean Cohort." *Eur J Hum Genet* **19**(11): 1167-1172.

Morais, A., B. Lima, M. J. Peixoto, H. Alves, A. Marques and L. Delgado (2012). "*BTNL2* Gene Polymorphism Associations with Susceptibility and Phenotype Expression in Sarcoidosis." *Respir Med* **106**(12): 1771-1777.

Morais da Silva, S., A. Hacker, V. Harley, P. Goodfellow, A. Swain and R. Lovell-Badge (1996). "*SOX9* Expression During Gonadal Development Implies a

Conserved Role for the Gene in Testis Differentiation in Mammals and Birds." *Nat Genet* **14**(1): 62-68.

Moyo, N. A., E. Marchi and F. Steinbach (2013). "Differentiation and Activation of Equine Monocyte-Derived Dendritic Cells Is Not Correlated with Cd206 or Cd83 Expression." *Immunology* **139**(4):472-483.

Ng, S. B., E. H. Turner, P. D. Robertson, S. D. Flygare, A. W. Bigham, C. Lee, T. Shaffer, M. Wong, A. Bhattacharjee, E. E. Eichler, et al. (2009). "Targeted Capture and Massively Parallel Sequencing of 12 Human Exomes." *Nature* **461**(7261): 272-276.

Ng, S. L., S. S. Bidarkar, M. Sourial, P. J. Farmer, S. Donath and J. M. Hutson (2005). "Gubernacular Cell Division in Different Rodent Models of Cryptorchidism Supports Indirect Androgenic Action Via the Genitofemoral Nerve." *Journal of Pediatric Surgery* **40**(2): 434-441.

Ni, X., M. Zhuo, Z. Su, J. Duan, Y. Gao, Z. Wang, C. Zong, H. Bai, A. R. Chapman, J. Zhao, et al. (2013). "Reproducible Copy Number Variation Patterns among Single Circulating Tumor Cells of Lung Cancer Patients." *Proc Natl Acad Sci U S A* **110**(52): 21083-21088.

Nicholas, T. J., Z. Cheng, M. Ventura, K. Mealey, E. E. Eichler and J. M. Akey (2009). "The Genomic Architecture of Segmental Duplications and Associated Copy Number Variants in Dogs." *Genome Res* **19**(3): 491-499.

Nielen, A. L., S. van der Beek, G. J. Ubbink and B. W. Knol (2001). "Population Parameters to Compare Dog Breeds: Differences between Five Dutch Purebred Populations." *The Veterinary Quarterly* **23**(1): 43-49.

Noschka, E., M. L. Vandenplas, D. J. Hurley and J. N. Moore (2009). "Temporal Aspects of Laminar Gene Expression During the Developmental Stages of Equine Laminitis." *Vet Immunol Immunopathol* **129**(3-4): 242-253.

Nozawa, M., Y. Kawahara and M. Nei (2007). "Genomic Drift and Copy Number Variation of Sensory Receptor Genes in Humans." *Proc Natl Acad Sci U S A* **104**(51): 20421-20426.

Oakenfull, E. A., V. J. Buckle and J. B. Clegg (1993). "Localization of the Horse (*Equus caballus*) Alpha-Globin Gene Complex to Chromosome 13 by Fluorescence in Situ Hybridization." *Cytogenet Cell Genet* **62**(2-3): 136-138.

Ohno, S. (1970). *Evolution by Gene Duplication*, Springer-Verlag, New York: 1-160.

Oke, S. (2013). "Heaves in Horses." from http://www.thehorse.com/free-reports/30024/heaves-in-horses.

Olsson, L. M. and R. Holmdahl (2012). "Copy Number Variation in Autoimmunity--Importance Hidden in Complexity?" *Eur J Immunol* **42**(8): 1969-1976.

Orange, J. S., J. T. Glessner, E. Resnick, K. E. Sullivan, M. Lucas, B. Ferry, C. E. Kim, C. Hou, F. Wang, R. Chiavacci, et al. (2011). "Genome-Wide Association

Identifies Diverse Causes of Common Variable Immunodeficiency." *J Allergy Clin Immunol* **127**(6): 1360-1367.

Orlando, L., A. Ginolhac, M. Raghavan, J. Vilstrup, M. Rasmussen, K. Magnussen, K. E. Steinmann, P. Kapranov, J. F. Thompson, G. Zazula, et al. (2011). "True Single-Molecule DNA Sequencing of a Pleistocene Horse Bone." *Genome Res* **21**(10): 1705-1719.

Orlando, L., A. Ginolhac, G. Zhang, D. Froese, A. Albrechtsen, M. Stiller, M. Schubert, E. Cappellini, B. Petersen, I. Moltke, et al. (2013). "Recalibrating *Equus* Evolution Using the Genome Sequence of an Early Middle Pleistocene Horse." *Nature* **499**(7456): 74-78.

Orr, N., W. Back, J. Gu, P. Leegwater, P. Govindarajan, J. Conroy, B. Ducro, J. A. Van Arendonk, D. E. MacHugh, S. Ennis, et al. (2010). "Genome-Wide SNP Association-Based Localization of a Dwarfism Gene in Friesian Dwarf Horses." *Anim Genet* **41 Suppl 2**: 2-7.

Outram, A. K., N. A. Stear, R. Bendrey, S. Olsen, A. Kasparov, V. Zaibert, N. Thorpe and R. P. Evershed (2009). "The Earliest Horse Harnessing and Milking." *Science* **323**(5919): 1332-1335.

Page, D. C., R. Mosher, E. M. Simpson, E. M. Fisher, G. Mardon, J. Pollack, B. McGillivray, A. de la Chapelle and L. G. Brown (1987). "The Sex-Determining

Region of the Human Y Chromosome Encodes a Finger Protein." *Cell* **51**(6): 1091-1104.

Pang, A. W., J. R. MacDonald, D. Pinto, J. Wei, M. A. Rafiq, D. F. Conrad, H. Park, M. E. Hurles, C. Lee, J. C. Venter, et al. (2010). "Towards a Comprehensive Structural Variation Map of an Individual Human Genome." *Genome Biol* **11**(5): R52, 1-14.

Papatheodorou, A., P. Makrythanasis, M. Kaliakatsos, A. Dimakou, D. Orfanidou, C. Roussos, E. Kanavakis and M. Tzetis (2010). "Development of Novel Microarray Methodology for the Study of Mutations in the *SERPINA1* and *ADRB2* Genes--Their Association with Obstructive Pulmonary Disease and Disseminated Bronchiectasis in Greek Patients." *Clin Biochem* **43**(1-2): 43-50.

Pardo-Mindan, F. J., F. Vargas Torcal, G. Garcia Julian and M. T. Virto Ruiz (1975). "Letter: Familial Cryptorchidism." *Pediatrics* **56**(4): 616.

Paria, N., T. Raudsepp, A. J. Pearks Wilkerson, P. C. O'Brien, M. A. Ferguson-Smith, C. C. Love, C. Arnold, P. Rakestraw, W. J. Murphy and B. P. Chowdhary (2011). "A Gene Catalogue of the Euchromatic Male-Specific Region of the Horse Y Chromosome: Comparison with Human and Other Mammals." *PLoS One* **6**(7): e21374, 1-13.

Parker, K. L., D. A. Rice, D. S. Lala, Y. Ikeda, X. Luo, M. Wong, M. Bakke, L. Zhao, C. Frigeri, N. A. Hanley, et al. (2002). "Steroidogenic Factor 1: An Essential Mediator of Endocrine Development." *Recent Prog Horm Res* **57**: 19-36.

Pathirana, I. N., K. Tanaka, N. Kawate, M. Tsuji, K. Kida, S. Hatoya, T. Inaba and H. Tamada (2010). "Analysis of Single Nucleotide Polymorphisms in the 3' Region of the Estrogen Receptor 1 Gene in Normal and Cryptorchid Miniature Dachshunds and Chihuahuas." *The Journal of Reproduction and Development* **56**(4): 405-410.

Paudel, Y., O. Madsen, H. J. Megens, L. A. Frantz, M. Bosse, J. W. Bastiaansen, R. P. Crooijmans and M. A. Groenen (2013). "Evolutionary Dynamics of Copy Number Variation in Pig Genomes in the Context of Adaptation and Domestication." *BMC Genomics* **14**(1): 449, 1-13.

Penedo, M. C., L. V. Millon, D. Bernoco, E. Bailey, M. Binns, G. Cholewinski, N. Ellis, J. Flynn, B. Gralak, A. Guthrie, et al. (2005). "International Equine Gene Mapping Workshop Report: A Comprehensive Linkage Map Constructed with Data from New Markers and by Merging Four Mapping Resources." *Cytogenet Genome Res* **111**(1): 5-15.

Peng, Y., M. Chen, X. J. Liu, W. Liu, Q. Li, H. Chai, X. Ren, X. Q. Wang, Z. G. Zhao, C. Zhang, et al. (2013). "The *CYP2C19* Genotype Does Not Impact the Long-

Term Prognosis of Patients with Coronary Artery Disease." *Atherosclerosis* **227**(1): 106-111.

Perrett, L. J. and D. A. O'Rourke (1969). "Hereditary Cryptorchidism." *The Medical Journal of Australia* **1**(25): 1289-1290.

Perry, G. H., N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, et al. (2007). "Diet and the Evolution of Human Amylase Gene Copy Number Variation." *Nat Genet* **39**(10): 1256-1260.

Perry, G. H., J. Tchinda, S. D. McGrath, J. Zhang, S. R. Picker, A. M. Caceres, A. J. Iafrate, C. Tyler-Smith, S. W. Scherer, E. E. Eichler, et al. (2006). "Hotspots for Copy Number Variation in Chimpanzees and Humans." *Proc Natl Acad Sci USA* **103**(21): 8006-8011.

Perry, G. H., F. Yang, T. Marques-Bonet, C. Murphy, T. Fitzgerald, A. S. Lee, C. Hyland, A. C. Stone, M. E. Hurles, C. Tyler-Smith, et al. (2008). "Copy Number Variation and Evolution in Humans and Chimpanzees." *Genome Res* **18**(11): 1698-1710.

Petersen, J. L., J. R. Mickelson, E. G. Cothran, L. S. Andersson, J. Axelsson, E. Bailey, D. Bannasch, M. M. Binns, A. S. Borges, P. Brama, et al. (2013a). "Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data." *PLoS One* **8**(1): e54997, 1-15.

Petersen, J. L., J. R. Mickelson, A. K. Rendahl, S. J. Valberg, L. S. Andersson, J. Axelsson, E. Bailey, D. Bannasch, M. M. Binns, A. S. Borges, et al. (2013b). "Genome-Wide Analysis Reveals Selection for Important Traits in Domestic Horse Breeds." *PLoS Genet* **9**(1): e1003211, 1-17.

Pham, J., C. Shaw, A. Pursley, P. Hixson, S. Sampath, E. Roney, T. Gambin, S. H. Kang, W. Bi, S. Lalani, et al. (2014). "Somatic Mosaicism Detected by Exon-Targeted, High-Resolution aCGH in 10 362 Consecutive Cases." *Eur J Hum Genet* **22**(8):969-978.

Picketts, D. J., D. R. Higgs, S. Bachoo, D. J. Blake, O. W. Quarrell and R. J. Gibbons (1996). "*ATRX* Encodes a Novel Member of the *SNF2* Family of Proteins: Mutations Point to a Common Mechanism Underlying the *ATRX* Syndrome." *Hum Mol Genet* **5**(12): 1899-1907.

Pielberg, G. and L. Andersson (2007). "Gene Copy Number Detection in Animal Studies." *Methods Mol Biol* **373**: 147-156.

Pirie, R. S., P. M. Dixon and B. C. McGorum (2003). "Endotoxin Contamination Contributes to the Pulmonary Inflammatory and Functional Response to *Aspergillus fumigatus* Extract Inhalation in Heaves Horses." *Clin Exp Allergy* **33**(9): 1289-1296.

Pleasance, E. D., R. K. Cheetham, P. J. Stephens, D. J. McBride, S. J. Humphray, C. D. Greenman, I. Varela, M. L. Lin, G. R. Ordonez, G. R. Bignell, et al. (2010). "A

Comprehensive Catalogue of Somatic Mutations from a Human Cancer Genome." *Nature* **463**(7278): 191-196.

Polan, M. B., M. T. Pastore, K. Steingass, S. Hashimoto, D. L. Thrush, R. Pyatt, S. Reshmi, J. M. Gastier-Foster, C. Astbury and K. L. McBride (2014). "Neurodevelopmental Disorders among Individuals with Duplication of 4p13 to 4p12 Containing a Gabaa Receptor Subunit Gene Cluster." *Eur J Hum Genet* **22**(1): 105-109.

Ponnudurai, R., B. Srinivasan, R. Sumitha, T. Koshy, S. S. Paul, Thiruvikraman and A. Rani (2012). "Klinefelter's Syndrome (Mosaic) with Chromosome 9 Inv and Schizophrenia." *Indian J Psychiatry* **54**(1): 88-90.

Popesco, M. C., E. J. Maclaren, J. Hopkins, L. Dumas, M. Cox, L. Meltesen, L. McGavran, G. J. Wyckoff and J. M. Sikela (2006). "Human Lineage-Specific Amplification, Selection, and Neuronal Expression of DUF1220 Domains." *Science* **313**(5791): 1304-1307.

Poultney, C. S., A. P. Goldberg, E. Drapeau, Y. Kou, H. Harony-Nicolas, Y. Kajiwara, S. De Rubeis, S. Durand, C. Stevens, K. Rehnstrom, et al. (2013). "Identification of Small Exonic CNV from Whole-Exome Sequence Data and Application to Autism Spectrum Disorder." *Am J Hum Genet* **93**(4): 607-619.

Power, M. M. (1986). "XY Sex Reversal in a Mare." *Equine Vet J* **18**(3): 233-236.

Pratt, S. M., B. A. Stacy, M. B. Whitcomb, J. D. Vidal, H. E. De Cock and W. D. Wilson (2003). "Malignant Sertoli Cell Tumor in the Retained Abdominal Testis of a Unilaterally Cryptorchid Horse." *J Am Vet Med Assoc* **222**(4): 486-490.

Puente, X. S., M. Pinyol, V. Quesada, L. Conde, G. R. Ordonez, N. Villamor, G. Escaramis, P. Jares, S. Bea, M. Gonzalez-Diaz, et al. (2011). "Whole-Genome Sequencing Identifies Recurrent Mutations in Chronic Lymphocytic Leukaemia." *Nature* **475**(7354): 101-105.

Pujar, S. and V. N. Meyers-Wallen (2012). "Sequence Variations in Equine Candidate Genes for XX and XY Inherited Disorders of Sexual Development." *Reprod Domest Anim* **47**(5): 827-834.

Quilter, C. R., A. C. Karcanias, M. R. Bagga, S. Duncan, A. Murray, G. S. Conway, C. A. Sargent and N. A. Affara (2010). "Analysis of X Chromosome Genomic DNA Sequence Copy Number Variation Associated with Premature Ovarian Failure (POF)." *Hum Reprod* **25**(8): 2139-2150.

Quinlan, A. R., R. A. Clark, S. Sokolova, M. L. Leibowitz, Y. Zhang, M. E. Hurles, J. C. Mell and I. M. Hall (2010). "Genome-Wide Mapping and Assembly of Structural Variant Breakpoints in the Mouse Genome." *Genome Res* **20**(5): 623-635.

Quint, J. K., G. C. Donaldson, M. Kumari, P. J. Talmud and J. R. Hurst (2011). "*SERPINA1* 11478g->a Variant, Serum Alpha1-Antitrypsin, Exacerbation Frequency and Fev1 Decline in COPD." *Thorax* **66**(5): 418-424.

Racine, J., V. Gerber, M. M. Feutz, C. P. Riley, J. Adamec, J. E. Swinburne and L. L. Couetil (2011). "Comparison of Genomic and Proteomic Data in Recurrent Airway Obstruction Affected Horses Using Ingenuity Pathway Analysis(R)." *BMC Vet Res* **7**: 48, 1-10.

Ramayo-Caldas, Y., A. Castello, R. N. Pena, E. Alves, A. Mercade, C. A. Souza, A. I. Fernandez, M. Perez-Enciso and J. M. Folch (2010). "Copy Number Variation in the Porcine Genome Inferred from a 60 K SNP Beadchip." *BMC Genomics* **11**: 593, 1-10.

Ramseyer, A., C. Gaillard, D. Burger, R. Straub, U. Jost, C. Boog, E. Marti and V. Gerber (2007). "Effects of Genetic and Environmental Factors on Chronic Lower Airway Disease in Horses." *J Vet Intern Med* **21**(1): 149-156.

Raney, N. E., K. T. Graves, E. G. Cothran, E. Bailey and L. Coogle (1998). Synteny Mapping of the Horse Using a Heterohybridoma Panel. *Plant and Animal Genome VI*, San Diego, CA, USA.

Rangasamy, S., S. R. D'Mello and V. Narayanan (2013). "Epigenetics, Autism Spectrum, and Neurodevelopmental Disorders." *Neurotherapeutics* **10**(4): 742-756.

Raudsepp, T. and B. P. Chowdhary (2008a). "FISH for Mapping Single Copy Genes." *Methods Mol Biol* **422**: 31-49.

Raudsepp, T. and B. P. Chowdhary (2008b). "The Horse Pseudoautosomal Region (PAR): Characterization and Comparison with the Human, Chimp and Mouse PARs." *Cytogenet Genome Res* **121**(2): 102-109.

Raudsepp, T. and B. P. Chowdhary (2010). Cytogenetics and Chromosome Maps. *The Genetics of the Pig*. M. F. Rothschild and A. Ruvinsky (ed), CABI**:** 134-178.

Raudsepp, T. and B. P. Chowdhary (2013). Physical and Comparative Maps. *Equine Genomics*. B. P. Chowdhary (ed), Wiley-Blackwell: 49-72.

Raudsepp, T., P. J. Das and B. P. Chowdhary (2013). Genomics of Reproduction and Fertility. *Equine Genomics*. B. P. Chowdhary (ed), Wiley- Blackwell**:** 199-216.

Raudsepp, T., K. Durkin, T. L. Lear, P. J. Das, F. Avila, P. Kachroo and B. P. Chowdhary (2010). "Molecular Heterogeneity of XY Sex Reversal in Horses." *Anim Genet* **41 Suppl 2**: 41-52.

Raudsepp, T., A. Gustafson-Seabury, K. Durkin, M. L. Wagner, G. Goh, C. M. Seabury, C. Brinkmeyer-Langford, E. J. Lee, R. Agarwala, E. Stallknecht-Rice, et al. (2008). "A 4,103 Marker Integrated Physical and Comparative Map of the Horse Genome." *Cytogenet Genome Res* **122**(1): 28-36.

Raudsepp, T., E. J. Lee, S. R. Kata, C. Brinkmeyer, J. R. Mickelson, L. C. Skow, J. E. Womack and B. P. Chowdhary (2004a). "Exceptional Conservation of Horse-Human Gene Order on X Chromosome Revealed by High-Resolution Radiation Hybrid Mapping." *Proc Natl Acad Sci USA* **101**(8): 2386-2391.

Raudsepp, T., M. E. McCue, P. J. Das, L. Dobson, M. Vishnoi, K. L. Fritz, R. Schaefer, A. K. Rendahl, J. N. Derr, C. C. Love, et al. (2012). "Genome-Wide Association Study Implicates Testis-Sperm Specific *FKBP6* as a Susceptibility Locus for Impaired Acrosome Reaction in Stallions." *PLoS Genet* **8**(12): e1003139, 1-14.

Raudsepp, T., A. Santani, B. Wallner, S. R. Kata, C. Ren, H. B. Zhang, J. E. Womack, L. C. Skow and B. P. Chowdhary (2004b). "A Detailed Physical Map of the Horse Y Chromosome." *Proc Natl Acad Sci USA* **101**(25): 9321-9326.

Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, et al. (2006). "Global Variation in Copy Number in the Human Genome." *Nature* **444**(7118): 444-454.

Revay, T., D. A. Villagomez, D. Brewer, T. Chenier and W. A. King (2012). "GTG Mutation in the Start Codon of the Androgen Receptor Gene in a Family of Horses with 64,XY Disorder of Sex Development." *Sex Dev* **6**(1-3): 108-116.

Ricard, G., J. Molina, J. Chrast, W. Gu, N. Gheldof, S. Pradervand, F. Schutz, J. I. Young, J. R. Lupski, A. Reymond, et al. (2010). "Phenotypic Consequences of Copy Number Variation: Insights from Smith-Magenis and Potocki-Lupski Syndrome Mouse Models." *PLoS Biol* **8**(11): e1000543, 1-12.

Rieder, S., S. Taourit, D. Mariat, B. Langlois and G. Guerin (2001). "Mutations in the Agouti (*ASIP*), the Extension (*MC1R*), and the Brown (*TYRP1*) Loci and Their

Association to Coat Color Phenotypes in Horses (*Equus caballus*)." *Mamm Genome* **12**(6): 450-455.

Riethman, H. (2008). "Human Subtelomeric Copy Number Variations." *Cytogenet Genome Res* **123**(1-4): 244-252.

Robinson, J. I., I. M. Carr, D. L. Cooper, L. H. Rashid, S. G. Martin, P. Emery, J. D. Isaacs, A. Barton, A. G. Wilson, J. H. Barrett, et al. (2012). "Confirmation of Association of *FCGR3B* but Not *FCGR3A* Copy Number with Susceptibility to Autoantibody Positive Rheumatoid Arthritis." *Hum Mutat* **33**(4): 741-749.

Robinson, N. E. (2001). "International Workshop on Equine Chronic Airway Disease. Michigan State University 16-18 June 2000." *Equine Vet J* **33**(1): 5-19.

Robinson, N. E., F. J. Derksen, M. A. Olszewski and V. A. Buechner-Maxwell (1996). "The Pathogenesis of Chronic Obstructive Pulmonary Disease of Horses." *Br Vet J* **152**(3): 283-306.

Rogers, A. J., C. Brasch-Andersen, I. Ionita-Laza, A. Murphy, S. Sharma, B. J. Klanderman and B. A. Raby (2009). "The Interaction of Glutathione S-Transferase M1-Null Variants with Tobacco Smoke Exposure and the Development of Childhood Asthma." *Clin Exp Allergy* **39**(11): 1721-1729.

Rogers, A. J., J. H. Chu, K. Darvishi, I. Ionita-Laza, H. Lehmann, R. Mills, C. Lee and B. A. Raby (2013). "Copy Number Variation Prevalence in Known Asthma

Genes and Their Impact on Asthma Susceptibility." *Clin Exp Allergy* **43**(4): 455-462.

Rosengren Pielberg, G., A. Golovko, E. Sundstrom, I. Curik, J. Lennartsson, M. H. Seltenhammer, T. Druml, M. Binns, C. Fitzsimmons, G. Lindgren, et al. (2008). "A Cis-Acting Regulatory Mutation Causes Premature Hair Graying and Susceptibility to Melanoma in the Horse." *Nat Genet* **40**(8): 1004-1009.

Rothschild, M. F., L. L. Christian and W. Blanchard (1988). "Evidence for Multigene Control of Cryptorchidism in Swine." *J Hered* **79**(4): 313-314.

Rouillard, J. M., C. J. Herbert and M. Zuker (2002). "Oligoarray: Genome-Scale Oligonucleotide Design for Microarrays." *Bioinformatics* **18**(3): 486-487.

Rozen, S. G., J. D. Marszalek, K. Irenze, H. Skaletsky, L. G. Brown, R. D. Oates, S. J. Silber, K. Ardlie and D. C. Page (2012). "*AZFC* Deletions and Spermatogenic Failure: A Population-Based Survey of 20,000 Y Chromosomes." *Am J Hum Genet* **91**(5): 890-896.

Rubes, J., A. Pinton, A. Bonnet-Garnier, V. Fillon, P. Musilova, K. Michalova, S. Kubickova, A. Ducos and M. Yerle (2009). "Fluorescence in Situ Hybridization Applied to Domestic Animal Cytogenetics." *Cytogenet Genome Res* **126**(1-2): 34-48.

Rudolph, J. A., S. J. Spier, G. Byrns, C. V. Rojas, D. Bernoco and E. P. Hoffman (1992). "Periodic Paralysis in Quarter Horses: A Sodium Channel Mutation Disseminated by Selective Breeding." *Nat Genet* **2**(2): 144-147.

Russell, T. M. and P. J. Pollock (2011). Cryptochid Castration. *Equine Reproduction*. A. O. McKinnon, W.E. Vaala, D.D. Varner (ed). New Delhi, India, Wiley-Blackwell. 1: 1531-1539.

Ryhner, T., N. Muller, V. Balmer and V. Gerber (2008). "Increased Mucus Accumulation in Horses Chronically Affected with Recurrent Airway Obstruction Is Not Associated with up-Regulation of *CLCA1, EGFR, MUC5AC, BCL-2, IL-13* and *INF-Gamma* Expression." *Vet Immunol Immunopathol* **125**(1-2): 8-17.

Safran, M., I. Dalah, J. Alexander, N. Rosen, T. Iny Stein, M. Shmoish, N. Nativ, I. Bahir, T. Doniger, H. Krug, et al. (2010). "Genecards Version 3: The Human Gene Integrator." *Database (Oxford)* **2010**: baq020 1-16.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: 1054-1649.

Sandberg, K. (1974). "Linkage between the K Blood Group Locus and the 6-PGD Locus in Horses." *Anim Blood Groups Biochem Genet* **5**(3): 137-141.

Sandberg, K. and L. Andersson (1984). "Genetic Linkage in the Horse. I. Linkage Relationships among 15 Blood Marker Loci." *Hereditas* **100**(2): 199-208.

Sanders, S. J., A. G. Ercan-Sencicek, V. Hus, R. Luo, M. T. Murtha, D. Moreno-De-Luca, S. H. Chu, M. P. Moreau, A. R. Gupta, S. A. Thomson, et al. (2011). "Multiple Recurrent De Novo Cnvs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism." *Neuron* **70**(5): 863-885.

Saunders, D. N., E. A. Tindall, R. F. Shearer, J. Roberson, A. Decker, J. A. Wilson and V. M. Hayes (2012). "A Novel *SERPINA1* Mutation Causing Serum Alpha(1)-Antitrypsin Deficiency." *PLoS One* **7**(12): e51762, 1-5.

Sayers, E. W., T. Barrett, D. A. Benson, E. Bolton, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. Dicuccio, S. Federhen, et al. (2010). "Database Resources of the National Center for Biotechnology Information." *Nucleic Acids Res* **38**(Database issue): D5-D16.

Schaeper, W. (1939). "Untersuchungen U¨Ber Die Erblichkeit Und Das Wesen Des Lungendampfes Beim Pferd [Investigation into the Nature and Heritability of Heaves in the Horse] " *Tieraerztl Rundschau* **31**: 595-599.

Scharrenberg, A., V. Gerber, J. E. Swinburne, A. D. Wilson, J. Klukowska-Rotzler, E. Laumen and E. Marti (2010). "*IGE, IGGA, IGGB* and *IGG*(T) Serum Antibody

Levels in Offspring of Two Sires Affected with Equine Recurrent Airway Obstruction." *Anim Genet* **41 Suppl 2**: 131-137.

Schlattl, A., S. Anders, S. M. Waszak, W. Huber and J. O. Korbel (2011). "Relating Cnvs to Transcriptome Data at Fine Resolution: Assessment of the Effect of Variant Size, Type, and Overlap with Functional Regions." *Genome Res* **21**(12): 2004-2013.

Schmahl, J., Y. Kim, J. S. Colvin, D. M. Ornitz and B. Capel (2004). "Fgf9 Induces Proliferation and Nuclear Localization of *FGFR2* in Sertoli Precursors During Male Sex Determination." *Development* **131**(15): 3627-3636.

Schroder, W., A. Klostermann and O. Distl (2011). "Candidate Genes for Physical Performance in the Horse." *Vet J* **190**(1): 39-48.

Schroder, W., A. Klostermann, K. F. Stock and O. Distl (2012). "A Genome-Wide Association Study for Quantitative Trait Loci of Show-Jumping in Hanoverian Warmblood Horses." *Anim Genet* **43**(4): 392-400.

Seahorn, T. L. and R. E. Beadle (1993). "Summer Pasture-Associated Obstructive Pulmonary Disease in Horses: 21 Cases (1983-1991)." *J Am Vet Med Assoc* **202**(5): 779-782.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi, et al. (2004). "Large-Scale Copy Number Polymorphism in the Human Genome." *Science* **305**(5683): 525-528.

Sekido, R. and R. Lovell-Badge (2008). "Sex Determination Involves Synergistic Action of *SRY* and *SF1* on a Specific *SOX9* Enhancer." *Nature* **453**(7197): 930-934.

Sekido, R. and R. Lovell-Badge (2009). "Sex Determination and Sry: Down to a Wink and a Nudge?" *Trends Genet* **25**(1): 19-29.

Serapinas, D., B. Sitkauskiene and R. Sakalauskas (2012). "Inflammatory Markers in Chronic Obstructive Pulmonary Disease Patients with Different Alpha1 Antitrypsin Genotypes." *Arch Med Sci* **8**(6): 1053-1058.

Sha, B. Y., T. L. Yang, L. J. Zhao, X. D. Chen, Y. Guo, Y. Chen, F. Pan, Z. X. Zhang, S. S. Dong, X. H. Xu, et al. (2009). "Genome-Wide Association Study Suggested Copy Number Variation May Be Associated with Body Mass Index in the Chinese Population." *J Hum Genet* **54**(4): 199-202.

Shadravan, F. (2013). "Sex Bias in Copy Number Variation of Olfactory Receptor Gene Family Depends on Ethnicity." *Front Genet* **4**: 32, 1-13.

Shaikh, T. H., X. Gai, J. C. Perin, J. T. Glessner, H. Xie, K. Murphy, R. O'Hara, T. Casalunovo, L. K. Conlin, M. D'Arcy, et al. (2009). "High-Resolution Mapping and Analysis of Copy Number Variations in the Human Genome: A Data Resource for Clinical and Research Applications." *Genome Res* **19**(9): 1682-1690.

Shakhsi-Niaei, M., J. Klukowska-Rotzler, C. Drogemuller, J. Swinburne, C. Ehrmann, D. Saftic, A. Ramseyer, V. Gerber, G. Dolf and T. Leeb (2012). "Replication and

Fine-Mapping of a QTLfor Recurrent Airway Obstruction in European Warmblood Horses." *Anim Genet* **43**(5): 627-631.

Shakhsi-Niaei, M., J. Klukowska-Rotzler, C. Drogemuller, J. E. Swinburne, V. Gerber and T. Leeb (2010). "Characterization of the Equine *ITGAX* Gene and Its Association with Recurrent Airway Obstruction in European Warmblood Horses." *Anim Genet* **41**(5): 559-560.

Sharp, A. J. (2009). "Emerging Themes and New Challenges in Defining the Role of Structural Variation in Human Disease." *Hum Mutat* **30**(2): 135-144.

Sharp, A. J., S. S. Wachtel and K. Benirschke (1980). "H-Y Antigen in a Fertile XY Female Horse." *J Reprod Fertil* **58**(1): 157-160.

She, X., Z. Cheng, S. Zollner, D. M. Church and E. E. Eichler (2008). "Mouse Segmental Duplication and Copy Number Variation." *Nat Genet* **40**(7): 909-914.

Shim, S. H., J. S. Shim, K. Min, H. S. Lee, J. E. Park, S. H. Park, E. Hwang and M. Kim (2014). "Siblings with Opposite Chromosome Constitutions, Dup(2q)/Del(7q) and Del(2q)/Dup(7q)." *Gene* **534**(1): 100-106.

Shin, D. H., H. J. Lee, S. Cho, H. J. Kim, J. Y. Hwang, C. K. Lee, J. Jeong, D. Yoon and H. Kim (2014). "Deleted Copy Number Variation of Hanwoo and Holstein Using Next Generation Sequencing at the Population Level." *BMC Genomics* **15**(1): 240, 1-16.

Shin, E. K., L. E. Perryman and K. Meek (1997). "Evaluation of a Test for Identification of Arabian Horses Heterozygous for the Severe Combined Immunodeficiency Trait." *J Am Vet Med Assoc* **211**(10): 1268-1270.

Shiue, Y. L., L. A. Bickel, A. R. Caetano, L. V. Millon, R. S. Clark, M. L. Eggleston, R. Michelmore, E. Bailey, G. Guerin, S. Godard, et al. (1999). "A Synteny Map of the Horse Genome Comprised of 240 Microsatellite and RAPD Markers." *Anim Genet* **30**(1): 1-9.

Shull, A. Y., M. L. Clendenning, S. Ghoshal-Gupta, C. L. Farrell, H. V. Vangapandu, L. Dudas, B. J. Wilkerson and P. J. Buckhaults (2013). "Somatic Mutations, Allele Loss, and DNA Methylation of the Cub and Sushi Multiple Domains 1 (*CSMD1*) Gene Reveals Association with Early Age of Diagnosis in Colorectal Cancer Patients." *PLoS One* **8**(3): e58731, 1-11.

Signer-Hasler, H., M. Neuditschko, C. Koch, S. Froidevaux, C. Flury, D. Burger, T. Leeb and S. Rieder (2014). "A Chromosomal Region on ECA13 is Associated with Maxillary Prognathism in Horses." *PLoS One* **9**(1): e86607, 1-5.

Silvers, W. K. and S. S. Wachtel (1975). "H-Y Antigen: Behavior and Function." *Science*(195): 956-960.

Sinclair, A. H., P. Berta, M. S. Palmer, J. R. Hawkins, B. L. Griffiths, M. J. Smith, J. W. Foster, A. M. Frischauf, R. Lovell-Badge and P. N. Goodfellow (1990). "A Gene

from the Human Sex-Determining Region Encodes a Protein with Homology to a Conserved DNA-Binding Motif." *Nature* **346**(6281): 240-244.

Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier, L. G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri, et al. (2003). "The Male-Specific Region of the Human Y Chromosome Is a Mosaic of Discrete Sequence Classes." *Nature* **423**(6942): 825-837.

Skinner, B. M., A. Al Mutery, D. Smith, M. Volker, N. Hojjat, S. Raja, S. Trim, P. Houde, W. J. Boecklen and D. K. Griffin (2014). "Global Patterns of Apparent Copy Number Variation in Birds Revealed by Cross-Species Comparative Genomic Hybridization." *Chromosome Res* **22**(1): 59-70.

Skinner, B. M., L. B. Robertson, H. G. Tempest, E. J. Langley, D. Ioannou, K. E. Fowler, R. P. Crooijmans, A. D. Hall, D. K. Griffin and M. Volker (2009). "Comparative Genomics in Chicken and Pekin Duck Using FISH Mapping and Microarray Analysis." *BMC Genomics* **10**: 357, 1-11.

Smith, K. J., A. L. Bertone, S. E. Weisbrode and M. Radmacher (2006). "Gross, Histologic, and Gene Expression Characteristics of Osteoarthritic Articular Cartilage of the Metacarpal Condyle of Horses." *Am J Vet Res* **67**(8): 1299-1306.

Smyk, M., J. S. Berg, A. Pursley, F. K. Curtis, B. A. Fernandez, G. A. Bien-Willner, J. R. Lupski, S. W. Cheung and P. Stankiewicz (2007). "Male-to-Female Sex

Reversal Associated with an Approximately 250 Kb Deletion Upstream of *NR0B1* (*DAX1*)." *Hum Genet* **122**(1): 63-70.

Snapper, J. R. (1986). "Large Animal Models of Asthma." *Am Rev Respir Dis* **133**(3): 351-352.

Speliotes, E. K., C. J. Willer, S. I. Berndt, K. L. Monda, G. Thorleifsson, A. U. Jackson, H. Lango Allen, C. M. Lindgren, J. Luan, R. Magi, et al. (2010). "Association Analyses of 249,796 Individuals Reveal 18 New Loci Associated with Body Mass Index." *Nat Genet* **42**(11): 937-948.

Spielmann, M. and E. Klopocki (2013). "CNVs of Noncoding Cis-Regulatory Elements in Human Disease." *Curr Opin Genet Dev* **23**(3): 249-256.

Spirito, F., A. Charlesworth, K. Linder, J. P. Ortonne, J. Baird and G. Meneguzzi (2002). "Animal Models for Skin Blistering Conditions: Absence of Laminin 5 Causes Hereditary Junctional Mechanobullous Disease in the Belgian Horse." *J Invest Dermatol* **119**(3): 684-691.

Stalker, J., B. Gibbins, P. Meidl, J. Smith, W. Spooner, H. R. Hotz and A. V. Cox (2004). "The Ensembl Web Site: Mechanics of a Genome Browser." *Genome Res* **14**(5): 951-955.

Stankiewicz, P. and A. L. Beaudet (2007). "Use of Array Cgh in the Evaluation of Dysmorphology, Malformations, Developmental Delay, and Idiopathic Mental Retardation." *Curr Opin Genet Dev* **17**(3): 182-192.

Stankiewicz, P. and J. R. Lupski (2010). "Structural Variation in the Human Genome and Its Role in Disease." *Annu Rev Med* **61**: 437-455.

Stefansson, H., D. Rujescu, S. Cichon, O. P. Pietilainen, A. Ingason, S. Steinberg, R. Fossdal, E. Sigurdsson, T. Sigmundsson, J. E. Buizer-Voskamp, et al. (2008). "Large Recurrent Microdeletions Associated with Schizophrenia." *Nature* **455**(7210): 232-236.

Stenberg, P. and J. Larsson (2011). "Buffering and the Evolution of Chromosome-Wide Gene Regulation." *Chromosoma* **120**(3): 213-225.

Stephens, P. J., C. D. Greenman, B. Fu, F. Yang, G. R. Bignell, L. J. Mudie, E. D. Pleasance, K. W. Lau, D. Beare, L. A. Stebbings, et al. (2011). "Massive Genomic Rearrangement Acquired in a Single Catastrophic Event During Cancer Development." *Cell* **144**(1): 27-40.

Stephens, P. J., D. J. McBride, M. L. Lin, I. Varela, E. D. Pleasance, J. T. Simpson, L. A. Stebbings, C. Leroy, S. Edkins, L. J. Mudie, et al. (2009). "Complex Landscapes of Somatic Rearrangement in Human Breast Cancer Genomes." *Nature* **462**(7276): 1005-1010.

Stickle, R. L. and J. F. Fessler (1978). "Retrospective Study of 350 Cases of Equine Cryptorchidism." *J Am Vet Med Assoc* **172**(3): 343-346.

Stothard, P., J. W. Choi, U. Basu, J. M. Sumner-Thomson, Y. Meng, X. Liao and S. S. Moore (2011). "Whole Genome Resequencing of Black Angus and Holstein Cattle for SNP and CNV Discovery." *BMC Genomics* **12**: 559, 1-14.

Sundstrom, E., F. Imsland, S. Mikko, C. Wade, S. Sigurdsson, G. R. Pielberg, A. Golovko, I. Curik, M. H. Seltenhammer, J. Solkner, et al. (2012). "Copy Number Expansion of the *STX17* Duplication in Melanoma Tissue from Grey Horses." *BMC Genomics* **13**: 365, 1-14.

Swinburne, J., C. Gerstenberg, M. Breen, V. Aldridge, L. Lockhart, E. Marti, D. Antczak, M. Eggleston-Stott, E. Bailey, J. Mickelson, et al. (2000). "First Comprehensive Low-Density Horse Linkage Map Based on Two 3-Generation, Full-Sibling, Cross-Bred Horse Reference Families." *Genomics* **66**(2): 123-134.

Swinburne, J. E., H. Bogle, J. Klukowska-Rotzler, M. Drogemuller, T. Leeb, E. Temperton, G. Dolf and V. Gerber (2009). "A Whole-Genome Scan for Recurrent Airway Obstruction in Warmblood Sport Horses Indicates Two Positional Candidate Regions." *Mamm Genome* **20**(8): 504-515.

Swinburne, J. E., M. Boursnell, G. Hill, L. Pettitt, T. Allen, B. Chowdhary, T. Hasegawa, M. Kurosawa, T. Leeb, S. Mashima, et al. (2006). "Single Linkage Group Per Chromosome Genetic Linkage Map for the Horse, Based on Two Three-Generation, Full-Sibling, Crossbred Horse Reference Families." *Genomics* **87**(1): 1-29.

Swinburne, J. E., A. Hopkins and M. M. Binns (2002). "Assignment of the Horse Grey Coat Colour Gene to ECA25 Using Whole Genome Scanning." *Anim Genet* **33**(5): 338-342.

Switonski, M., A. Chmurzynska, I. Szczerbal, A. Lipczynski, F. Yang and A. Nowicka-Posluszna (2005). "Sex Reversal Syndrome (64,XY; *SRY*-Positive) in a Mare Demonstrating Masculine Behaviour." *J Anim Breed Genet* **122 Suppl 1**: 60-63.

Takahashi, N., N. Tsuyama, K. Sasaki, M. Kodaira, Y. Satoh, Y. Kodama, K. Sugita and H. Katayama (2008). "Segmental Copy-Number Variation Observed in Japanese by Array-CGH." *Ann Hum Genet* **72**(Pt 2): 193-204.

Tan, R., Y. Wang, S. E. Kleinstein, Y. Liu, X. Zhu, H. Guo, Q. Jiang, A. S. Allen and M. Zhu (2014). "An Evaluation of Copy Number Variation Detection Tools from Whole-Exome Sequencing Data." *Hum Mutat* **35**(7):899-907.

Tannour-Louet, M., S. Han, S. T. Corbett, J. F. Louet, S. Yatsenko, L. Meyers, C. A. Shaw, S. H. Kang, S. W. Cheung and D. J. Lamb (2010). "Identification of De Novo Copy Number Variants Associated with Human Disorders of Sexual Development." *PLoS One* **5**(10): e15392, 1-13.

Tansey, K. E., J. J. Rucker, D. H. Kavanagh, M. Guipponi, N. Perroud, G. Bondolfi, E. Domenici, D. M. Evans, J. Hauser, N. Henigsberg, et al. (2014). "Copy Number Variants and Therapeutic Response to Antidepressant Medication in Major Depressive Disorder." *Pharmacogenomics J* **14**(4):395-399.

Teyssedre, S., M. C. Dupuis, G. Guerin, L. Schibler, J. M. Denoix, J. M. Elsen and A. Ricard (2012). "Genome-Wide Association Studies for Osteochondrosis in French Trotter Horses." *J Anim Sci* **90**(1): 45-53.

Thomas, R., S. E. Duke, E. K. Karlsson, A. Evans, P. Ellis, K. Lindblad-Toh, C. F. Langford and M. Breen (2008). "A Genome Assembly-Integrated Dog 1 Mb BAC Microarray: A Cytogenetic Resource for Canine Cancer Studies and Comparative Genomic Analysis." *Cytogenet Genome Res* **122**(2): 110-121.

Thomas, R., C. Rebbeck, A. M. Leroi, A. Burt and M. Breen (2009). "Extensive Conservation of Genomic Imbalances in Canine Transmissible Venereal Tumors (CTVT) Detected by Microarray-Based CGH Analysis." *Chromosome Res* **17**(7): 927-934.

Thun, G. A., M. Imboden, I. Ferrarotti, A. Kumar, M. Obeidat, M. Zorzetto, M. Haun, I. Curjuric, A. Couto Alves, V. E. Jackson, et al. (2013). "Causal and Synthetic Associations of Variants in the Serpina Gene Cluster with Alpha1-Antitrypsin Serum Levels." *PLoS Genet* **9**(8): e1003585, 1-16.

Tian, X., G. Pascal, S. Fouchecourt, P. Pontarotti and P. Monget (2009). "Gene Birth, Death, and Divergence: The Different Scenarios of Reproduction-Related Gene Evolution." *Biol Reprod* **80**(4): 616-621.

Trujillo, J. M., B. Walden, P. O'Neil and H. B. Anstall (1965). "Sex-Linkage of Glucose-6-Phosphate Dehydrogenase in the Horse and Donkey." *Science* **148**: 1603-1604.

Tryon, R. C., S. D. White and D. L. Bannasch (2007). "Homozygosity Mapping Approach Identifies a Missense Mutation in Equine Cyclophilin B (*PPIB*) Associated with Herda in the American Quarter Horse." *Genomics* **90**(1): 93-102.

Turner, R. M. and R. Casas-Dolz (2002). "Differential Gene Expression in Stallions with Idiopathic Testicular Degeneration." *Theriogenology* **58**: 421-434.

Tuttelmann, F., M. Simoni, S. Kliesch, S. Ledig, B. Dworniczak, P. Wieacker and A. Ropke (2011). "Copy Number Variants in Patients with Severe Oligozoospermia and Sertoli-Cell-Only Syndrome." *PLoS One* **6**(4): e19426, 1-11.

Tuzun, E., A. J. Sharp, J. A. Bailey, R. Kaul, V. A. Morrison, L. M. Pertz, E. Haugen, H. Hayden, D. Albertson, D. Pinkel, et al. (2005). "Fine-Scale Structural Variation of the Human Genome." *Nat Genet* **37**(7): 727-732.

Ueno, T., M. Emi, H. Sato, N. Ito, M. Muta, K. Kuroi and M. Toi (2012). "Genome-Wide Copy Number Analysis in Primary Breast Cancer." *Expert Opin Ther Targets* **16 Suppl 1**: S31-S35.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm and S. G. Rozen (2012). "Primer3--New Capabilities and Interfaces." *Nucleic Acids Res* **40**(15): e115, 1-12.

Vaiman, D. and E. Pailhoux (2000). "Mammalian Sex Reversal and Intersexuality: Deciphering the Sex-Determination Cascade." *Trends Genet* **16**(11): 488-494.

van Karnebeek, C. D., M. C. Jansweijer, A. G. Leenders, M. Offringa and R. C. Hennekam (2005). "Diagnostic Investigations in Individuals with Mental Retardation: A Systematic Literature Review of Their Usefulness." *Eur J Hum Genet* **13**(1): 6-25.

Vassos, E., D. A. Collier, S. Holden, C. Patch, D. Rujescu, D. St Clair and C. M. Lewis (2010). "Penetrance for Copy Number Variants Associated with Schizophrenia." *Hum Mol Genet* **19**(17): 3477-3481.

Vaughan, L., W. Schofield and S. Ennis (2001). "*SRY*-Negative XX Sex Reversal in a Pony: A Case Report." *Theriogenology* **55**(5): 1051-1057.

Veerappa, A. M., S. Vishweswaraiah, K. Lingaiah, M. Murthy, D. S. Manjegowda, R. Nayaka and N. B. Ramachandra (2013). "Unravelling the Complexity of Human Olfactory Receptor Repertoire by Copy Number Analysis across Population Using High Resolution Arrays." *PLoS One* **8**(7): e66843, 1-14.

Venugopal, C. S., L. C. Mendes, J. R. Peiro, S. S. Laborde, A. M. Stokes and R. M. Moore (2010). "Transcriptional Changes Associated with Recurrent Airway Obstruction in Affected and Unaffected Horses." *Am J Vet Res* **71**(4): 476-482.

Vila, C., J. A. Leonard, A. Gotherstrom, S. Marklund, K. Sandberg, K. Liden, R. K. Wayne and H. Ellegren (2001). "Widespread Origins of Domestic Horse Lineages." *Science* **291**(5503): 474-477.

Villagomez, D. A., T. L. Lear, T. Chenier, S. Lee, R. B. McGee, J. Cahill, R. A. Foster, E. Reyes, E. St John and W. A. King (2011). "Equine Disorders of Sexual Development in 17 Mares Including XX, SRY-Negative, XY, SRY-Negative and XY, SRY-Positive Genotypes." *Sex Dev* **5**(1): 16-25.

Villagomez, D. A., P. Parma, O. Radi, G. Di Meo, A. Pinton, L. Iannuzzi and W. A. King (2009). "Classical and Molecular Cytogenetics of Disorders of Sex Development in Domestic Animals." *Cytogenet Genome Res* **126**(1-2): 110-131.

Villagomez, D. A. and A. Pinton (2008). "Chromosomal Abnormalities, Meiotic Behavior and Fertility in Domestic Animals." *Cytogenet Genome Res* **120**(1-2): 69-80.

Vittori, A., C. Breda, M. Repici, M. Orth, R. A. Roos, T. F. Outeiro, F. Giorgini and E. J. Hollox (2014). "Copy-Number Variation of the Neuronal Glucose Transporter Gene *SLC2A3* and Age of Onset in Huntington's Disease." *Hum Mol Genet* **23**(12):3129-3137.

Vogler, C., L. Gschwind, B. Rothlisberger, A. Huber, I. Filges, P. Miny, B. Auschra, A. Stetak, P. Demougin, V. Vukojevic, et al. (2010). "Microarray-Based Maps of Copy-Number Variant Regions in European and Sub-Saharan Populations." *PLoS One* **5**(12): e15246, 1-8.

Volker, M., N. Backstrom, B. M. Skinner, E. J. Langley, S. K. Bunzey, H. Ellegren and D. K. Griffin (2010). "Copy Number Variation, Chromosome Rearrangement,

and Their Association with Recombination During Avian Evolution." *Genome Res* **20**(4): 503-511.

Wade, C. M., E. Giulotto, S. Sigurdsson, M. Zoli, S. Gnerre, F. Imsland, T. L. Lear, D. L. Adelson, E. Bailey, R. R. Bellone, et al. (2009). "Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse." *Science* **326**(5954): 865-867.

Walker, S. and S. W. Scherer (2013). "Identification of Candidate Intergenic Risk Loci in Autism Spectrum Disorder." *BMC Genomics* **14**: 499, 1-6.

Walsh, K. M., M. B. Bracken, W. K. Murk, J. Hoh and A. T. Dewan (2010). "Association between Reduced Copy-Number at T-Cell Receptor Gamma (TCRgamma) and Childhood Allergic Asthma: A Possible Role for Somatic Mosaicism." *Mutat Res* **690**(1-2): 89-94.

Walters, R. G., S. Jacquemont, A. Valsesia, A. J. de Smith, D. Martinet, J. Andersson, M. Falchi, F. Chen, J. Andrieux, S. Lobbens, et al. (2010). "A New Highly Penetrant Form of Obesity Due to Deletions on Chromosome 16p11.2." *Nature* **463**(7281): 671-675.

Wang, J., J. Jiang, W. Fu, L. Jiang, X. Ding, J. F. Liu and Q. Zhang (2012). "A Genome-Wide Detection of Copy Number Variations Using SNP Genotyping Arrays in Swine." *BMC Genomics* **13**: 273, 1-10.

Wang, J., J. Jiang, H. Wang, H. Kang, Q. Zhang and J. F. Liu (2014a). "Enhancing Genome-Wide Copy Number Variation Identification by High Density Array CGH Using Diverse Resources of Pig Breeds." *PLoS One* **9**(1): e87571, 1-8.

Wang, J., H. Wang, J. Jiang, H. Kang, X. Feng, Q. Zhang and J. F. Liu (2013a). "Identification of Genome-Wide Copy Number Variations among Diverse Pig Breeds Using SNP Genotyping Arrays." *PLoS One* **8**(7): e68683, 1-8.

Wang, L., X. Liu, L. Zhang, H. Yan, W. Luo, J. Liang, D. Cheng, S. Chen, X. Ma, X. Song, et al. (2013b). "Genome-Wide Copy Number Variations Inferred from SNP Genotyping Arrays Using a Large White and Minzhu Intercross Population." *PLoS One* **8**(10): e74879, 1-8.

Wang, P., Y. Kim, J. Pollack, B. Narasimhan and R. Tibshirani (2005). "A Method for Calling Gains and Losses in Array CGH Data." *Biostatistics* **6**(1): 45-58.

Wang, W., S. Wang, C. Hou, Y. Xing, J. Cao, K. Wu, C. Liu, D. Zhang, L. Zhang, Y. Zhang, et al. (2014b). "Genome-Wide Detection of Copy Number Variations among Diverse Horse Breeds by Array CGH." *PLoS One* **9**(1): e86860, 1-9.

Wang, X., S. Nahashon, T. K. Feaster, A. Bohannon-Stewart and N. Adefope (2010). "An Initial Map of Chromosomal Segmental Copy Number Variations in the Chicken." *BMC Genomics* **11**: 351, 1-10.

Warburton, D., M. Ronemus, J. Kline, V. Jobanputra, I. Williams, K. Anyane-Yeboa, W. Chung, L. Yu, N. Wong, D. Awad, et al. (2014). "The Contribution of De Novo

and Rare Inherited Copy Number Changes to Congenital Heart Disease in an Unselected Sample of Children with Conotruncal Defects or Hypoplastic Left Heart Disease." *Hum Genet* **133**(1): 11-27.

Ward, T. L., S. J. Valberg, D. L. Adelson, C. A. Abbey, M. M. Binns and J. R. Mickelson (2004). "Glycogen Branching Enzyme (*GBE1*) Mutation Causing Equine Glycogen Storage Disease IV." *Mamm Genome* **15**(7): 570-577.

Waszak, S. M., Y. Hasin, T. Zichner, T. Olender, I. Keydar, M. Khen, A. M. Stutz, A. Schlattl, D. Lancet and J. O. Korbel (2010). "Systematic Inference of Copy-Number Genotypes from Personal Genome Sequencing Data Reveals Extensive Olfactory Receptor Gene Content Diversity." *PLoS Comput Biol* **6**(11): e1000988, 1-20.

Weischenfeldt, J., O. Symmons, F. Spitz and J. O. Korbel (2013). "Phenotypic Impact of Genomic Structural Variation: Insights from and for Human Disease." *Nat Rev Genet* **14**(2): 125-138.

Wennerstrom, A., A. Pietinalho, J. Lasota, K. Salli, I. Surakka, M. Seppanen, O. Selroos and M. L. Lokki (2013). "Major Histocompatibility Complex Class II and *BTNL2* Associations in Sarcoidosis." *Eur Respir J* **42**(2): 550-553.

White, S., T. Ohnesorg, A. Notini, K. Roeszler, J. Hewitt, H. Daggag, C. Smith, E. Turbitt, S. Gustin, J. van den Bergen, et al. (2011). "Copy Number Variation in

Patients with Disorders of Sex Development Due to 46,XY Gonadal Dysgenesis." *PLoS One* **6**(3): e17793, 1-10.

Wijnberg, I. D., M. Owczarek-Lipska, R. Sacchetto, F. Mascarello, F. Pascoli, W. Grunberg, J. H. van der Kolk and C. Drogemuller (2012). "A Missense Mutation in the Skeletal Muscle Chloride Channel 1 (*CLCN1*) as Candidate Causal Mutation for Congenital Myotonia in a New Forest Pony." *Neuromuscul Disord* **22**(4): 361-367.

Wilhelm, D., S. Palmer and P. Koopman (2007). "Sex Determination and Gonadal Development in Mammals." *Physiol Rev* **87**(1): 1-28.

Willer, C. J., E. K. Speliotes, R. J. Loos, S. Li, C. M. Lindgren, I. M. Heid, S. I. Berndt, A. L. Elliott, A. U. Jackson, C. Lamina, et al. (2009). "Six New Loci Associated with Body Mass Index Highlight a Neuronal Influence on Body Weight Regulation." *Nat Genet* **41**(1): 25-34.

Williams, H., C. M. Richards, B. A. Konfortov, J. R. Miller and E. M. Tucker (1993). "Synteny Mapping in the Horse Using Horse-Mouse Heterohybridomas." *Anim Genet* **24**(4): 257-260.

Wilson, G. M., S. Flibotte, P. I. Missirlis, M. A. Marra, S. Jones, K. Thornton, A. G. Clark and R. A. Holt (2006). "Identification by Full-Coverage Array CGH of Human DNA Copy Number Increases Relative to Chimpanzee and Gorilla." *Genome Res* **16**(2): 173-181.

Wilson, J. D., R. J. Auchus, M. W. Leihy, O. L. Guryev, R. W. Estabrook, S. M. Osborn, G. Shaw and M. B. Renfree (2003). "5Alpha-Androstane-3Alpha,17Beta-Diol is Formed in Tammar Wallaby Pouch Young Testes by a Pathway Involving 5Alpha-Pregnane-3Alpha,17Alpha-Diol-20-One as a Key Intermediate." *Endocrinology* **144**(2): 575-580.

Wineinger, N. E., A. Patki, K. J. Meyers, U. Broeckel, C. C. Gu, D. C. Rao, R. B. Devereux, D. K. Arnett and H. K. Tiwari (2011). "Genome-Wide Joint SNP and CNV Analysis of Aortic Root Diameter in African Americans: The Hypergen Study." *BMC Med Genomics* **4**: 4, 1-8.

Wood, H. M. and J. S. Elder (2009). "Cryptorchidism and Testicular Cancer: Separating Fact from Fiction." *J Urol* **181**(2): 452-461.

Xing, J., Y. Zhang, K. Han, A. H. Salem, S. K. Sen, C. D. Huff, Q. Zhou, E. F. Kirkness, S. Levy, M. A. Batzer, et al. (2009). "Mobile Elements Create Structural Variation: Analysis of a Complete Human Genome." *Genome Res* **19**(9): 1516-1526.

Xu, X. and U. Arnason (1994). "The Complete Mitochondrial DNA Sequence of the Horse, *Equus caballus*: Extensive Heteroplasmy of the Control Region." *Gene* **148**(2): 357-362.

Xu, Y., B. Peng, Y. Fu and C. I. Amos (2011). "Genome-Wide Algorithm for Detecting CNV Associations with Diseases." *BMC Bioinformatics* **12**: 331, 1-10.

Yalcin, B., K. Wong, A. Agam, M. Goodson, T. M. Keane, X. Gan, C. Nellaker, L. Goodstadt, J. Nicod, A. Bhomra, et al. (2011). "Sequence-Based Characterization of Structural Variation in the Mouse Genome." *Nature* **477**(7364): 326-329.

Yang, B., Z. Mo, C. Wu, H. Yang, X. Yang, Y. He, L. Gui, L. Zhou, H. Guo, X. Zhang, et al. (2014). "A Genome-Wide Association Study Identifies Common Variants Influencing Serum Uric Acid Concentrations in a Chinese Population." *BMC Med Genomics* **7**: 10, 1-10.

Yohn, C. T., Z. Jiang, S. D. McGrath, K. E. Hayden, P. Khaitovich, M. E. Johnson, M. Y. Eichler, J. D. McPherson, S. Zhao, S. Paabo, et al. (2005). "Lineage-Specific Expansions of Retroviral Insertions within the Genomes of African Great Apes but Not Humans and Orangutans." *PLoS Biol* **3**(4): e110, 1-11.

Young, A. E., L. P. Bower, V. K. Affolter, H. E. De Cock, G. L. Ferraro and D. L. Bannasch (2007). "Evaluation of FOXC2 as a Candidate Gene for Chronic Progressive Lymphedema in Draft Horses." *Vet J* **174**(2): 397-399.

Young, J. M., R. M. Endicott, S. S. Parghi, M. Walker, J. M. Kidd and B. J. Trask (2008). "Extensive Copy-Number Variation of the Human Olfactory Receptor Gene Family." *Am J Hum Genet* **83**(2): 228-242.

Zeitz, A., A. Spotter, I. Blazyczek, U. Diesterbeck, B. Ohnesorge, E. Deegen and O. Distl (2009). "Whole-Genome Scan for Guttural Pouch Tympany in Arabian and German Warmblood Horses." *Anim Genet* **40**(6): 917-924.

Zenteno-Ruiz, J. C., S. Kofman-Alfaro and J. P. Mendez (2001). "46,XX Sex Reversal." *Arch Med Res* **32**(6): 559-566.

Zhan, B., J. Fadista, B. Thomsen, J. Hedegaard, F. Panitz and C. Bendixen (2011). "Global Assessment of Genomic Variation in Cattle by Genome Resequencing and High-Throughput Genotyping." *BMC Genomics* **12**: 557, 1-20.

Zhang, F., W. Gu, M. E. Hurles and J. R. Lupski (2009). "Copy Number Variation in Human Health, Disease, and Evolution." *Annu Rev Genomics Hum Genet* **10**: 451-481.

Zhang, J., P. D. Pare and A. J. Sandford (2008). "Recent Advances in Asthma Genetics." *Respir Res* **9**: 4, 1-8.

Zhao, X., Z. Q. Du and M. F. Rothschild (2010). "An Association Study of 20 Candidate Genes with Cryptorchidism in Siberian Husky Dogs." *J Anim Breed Genet* **127**(4): 327-331.

Zhao, X., S. Onteru, M. Saatchi, D. Garrick and M. Rothschild (2013). "A Genome-Wide Association Study for Canine Cryptorchidism in Siberian Huskies." *J Anim Breed Genet* **131**(3): 202-209.

Zhou, S. F. (2009). "Polymorphism of Human Cytochrome P450 2D6 and Its Clinical Significance: Part I." *Clin Pharmacokinet* **48**(11): 689-723.

Zorzetto, M., E. Russi, O. Senn, M. Imboden, I. Ferrarotti, C. Tinelli, I. Campo, S. Ottaviani, R. Scabini, A. von Eckardstein, et al. (2008). "*SERPINA1* Gene Variants in Individuals from the General Population with Reduced Alpha1-Antitrypsin Concentrations." *Clin Chem* **54**(8): 1331-1338.

# APPENDIX

Appendix 2 contains supplemental material (CNV details, gene function, gene ontology analysis, database of genome variant analysis, comparative study with other horse CNV datasets, CNVs with high/low log values, and qPCR primer list) for Chapter II

- Appendix 2_Chapter II.xlsx

Appendix 3 contains supplemental material (CNV details, gene function details, and CNVs with high/low log value of RAO study) for Chapter III

- Appendix 3_Chapter III.xlsx,

Appendix 4 contains supplemental material (CNV details, gene function details, and CNVs with high/low log value of DSD study) for Chapter IV

- Appendix 4.1 _Chapter IV.xlsx
- Appendix 4.2_Chapter IV.xlsx