Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science

Genomic Analysis of Hydrocephalus in Friesian Horses

Astri Olivia Herlino

Examensarbete / Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics,
419
Uppsala 2013

Master's Thesis, 30 HEC

Erasmus Mundus Programme – European Master in Animal Breeding and Genetics

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Genomic Analysis of Hydrocephalus in Friesian Horses

Astri Olivia Herlino

Supervisors:
Dr. Ir. Anouk Schurink, Wageningen University
Dr. Ir. Bart Ducro, Wageningen University
Gabriella Lindgren, SLU, Department of Animal Breeding and Genetics
Examiner:
Erling Strandberg, SLU, Department of Animal Breeding and Genetics

Credits: 30 HEC
Course title: Degree project in Animal Science
Course code: EX0556
Programme: Erasmus Mundus Programme - European Master in Animal Breeding and Genetics
Level: Advanced, A2E

Place of publication: Uppsala
Year of publication: 2013
Name of series: Examensarbete / Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, 419
On-line publication: http://epsilon.slu.se

Key words: Friesian, horses, genomic, analysis, hydrocephalus

European Master in Animal Breeding and Genetics

Genomic Analysis of Hydrocephalus in Friesian Horses

Astri Olivia Herlino
890512328030

THESIS ANIMAL BREEDING AND GENETICS (ABG-80436)
July 2013

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Animal Breeding and Genomics Centre (ABGC)

SUPERVISORS:

Dr. Ir. Anouk Schurink, Wageningen University

Dr. Ir. Bart Ducro, Wageningen University

CO-SUPERVISOR:

Dr. Ir. Gabriella Lindgren, Swedish University of Agricultural Sciences

Education and Culture
Erasmus Mundus

PREFACE

This thesis is written on behalf of Animal Breeding and Genetics, Animal Science Department, Wageningen University, The Netherlands.

Hydrocephalus is an abnormal accumulation of cerebrospinal fluid (CSF) in the cavities or ventricles of the brain. However, hydrocephalus is an uncommon disease in horses, and it observed more often in the Friesian horse breed than in other breeds. Several studies of phenotype and genetic background of hydrocephalus mostly have been done in human. Only few phenotype studies of hydrocephalus have been done in horse.

This thesis title is "Genomic analysis of hydrocephalus in Friesian horses". This research will contribute to understand the genetic background of hydrocephalus in Friesian horses. The thesis is divided into five chapters. Chapter one discusses the introduction of Friesian horses and hydrocephalus. Chapter two provides the materials and methods of the research. Chapter three provides the results of the research. Chapter four discusses the results of the research and suggestions for further study. Chapter five presents the conclusion of the study.

First and foremost, I would like to thank my supervisors and co-supervisors, Dr. Ir. Anouk Schurink (Wageningen University), Dr. Bart Ducro (Wageningen University) and Dr. Ir. Gabriella Lindgren (Swedish University of Agricultural Sciences), who have provided tremendous insight and guidance on a variety of topics, methods, technical analysis, and other elements required for the research. In addition, I really appreciated all of those individuals in Animal Breeding and Genetics Group. I would also like to thank Erasmus Mundus and all Universities that contribute in European Master in Animal Breeding and Genetics (EM-ABG) program for giving me a chance to join in this program. Last but not least, I would like to extend my appreciation to my parent, family and friends for all supports during my study.

Astri Olivia Herlino

July, 2013

SUMMARY

Hydrocephalus is an uncommon disorder in horse. However, hydrocephalus is observed more often in the Friesian horse breed than in other breeds. Due to the Friesian horse population has been closed for outside breeding and has a limited genetic pool. It leads to high rate of inbreeding. The general objective of this study is to get better understanding of genetic background of hydrocephalus in Friesian horses, which will be used to develop a DNA-test using genetic markers for hydrocephalus in the Friesian horse breed.

Genomic analysis based on case-control study was performed on 20 cases and 47 controls. Data used in this study was provided by Faculty of Veterinary Medicine at Utrecht University, the Netherlands. Population stratification seems to be present between cases and controls because 55.2% of some cases and controls coming from different sires. This population structure was corrected using genomic control method.

Hydrocephalus mutation in Friesian horses was found on chromosome 1:57,760,860-87,079,561 (29.3 Mb region) using genome-wide association study (GWAS). 68 associated SNPs based on genotype frequency differences included the 29 associated SNPs based on allele frequency differences between cases and controls were identified in this region. 85-90% of affected Friesian horses were homozygous genotypes of unfavorable alleles for these following SNPs: BIEC2-27351, BIEC2-27352, BIEC2-27588, BIEC2-28874, BIEC2-28875, BIEC2-28876, BIEC2-29359, and BIEC2-31514.

Haplotype association study was performed on chromosome 1. The result of this study showed that 54 associated haplotype blocks were also identified in the 29.3 Mb region. These haplotype blocks contained all associated SNPs that were found from GWAS. 228 genes were found in this region, but these genes were not the same as identify genes of hydrocephalus in human and other species.

Further studies are necessary to be able to develop a DNA-test for hydrocephalus in Friesian horses. These further studies are to validate the identified associations, to narrow down the associated region, to identify causal mutation. Furthermore, this DNA-test can be used in the Friesian horse breeding program to reduce or remove hydrocephalus in the population.

Key words: Friesian horses, genomic analysis, hydrocephalus

TABLE OF CONTENTS

PREFACE i

SUMMARY iii

LIST OF FIGURES vii

LIST OF TABLES ix

CHAPTER 1. INTRODUCTION 1

1.1. Background 1

1.1.1. The Friesian horse 1

1.1.2. Hydrocephalus 2

1.1.3. Hardy-Weinberg equilibrium 4

1.2. Research Questions 5

CHAPTER 2. MATERIALS AND METHODS 7

2.1. Data 7

2.2. Quality control 8

2.3. Population stratification 9

2.3.1. Population structure 9

2.3.2. Genomic control 10

2.4. Genome-wide association study 11

2.5. Homozygous segments 12

2.6. Haplotype construction and association 12

2.7. Candidate genes 13

CHAPTER 3. RESULTS 15

3.1. Population stratification 15

3.1.1. Detection of population stratification 15

3.1.2. Genomic control 16

3.2. Genome-wide association study 17

3.2.1. GWAS based on allele frequency 17

3.2.2. GWAS based on genotype frequency 20

3.3. Haplotype construction and association 25

3.4. Linkage disequilibrium 26
3.5. Candidate genes of hydrocephalus 27
3.5.1. Identification of genes from Genome-wide association study 29
3.5.2. Haplotype block 31
CHAPTER 4. DISCUSSION 33
4.1. Population structure 33
4.2. Genome-wide association study 34
4.3. Haplotype association study 34
4.4. Linkage disequilibrium 35
4.5. Candidate genes 35
4.6. Phenotype 36
4.7. DNA-test 37
4.8. Implementation of DNA-test in breeding program 40
CHAPTER 5. CONCLUSIONS 43
REFERENCES 45
APPENDIX 49

LIST OF FIGURES

Figure 1. Friesian foal with hydrocephalus (Sipma *et al.*, 2011). 2

Figure 2. The classical multidimensional scaling plot of genetic distances between cases and controls. Filled circles are the cases and open circles are the controls. The first two principle components are PC1 and PC2. 15

Figure 3. Quantile-quantile plots of chi-square statistics testing allele frequency differences between cases and controls. Line 2 is expected under no inflation and association. Line 1 is fitted slope. A is the Q-Q plot without the correction of *P-values* and B is the Q-Q plot with corrected *P-values*..... 16

Figure 4. The quantile-quantile plot hydrocephalus in Friesian horses from chi-square statistics testing genotype frequency differences between cases and controls. Line 2 is expected under no inflation and association. Line 1 is fitted slope..... 17

Figure 5. Manhattan plot of hydrocephalus in Friesian horses. Association of 29,935 SNPs with hydrocephalus represented by $-\log_{10}$ corrected *P-values* from chi-square statistics testing allele frequency differences between cases and controls plotted by chromosome and sorted by chromosomal position. 18

Figure 6. Manhattan plot of hydrocephalus in Friesian horses. Association of 29,935 SNPs with hydrocephalus represented by $-\log_{10}$ *P-values* from chi-square statistics testing genotype frequency differences between cases and controls plotted by chromosome and sorted by chromosomal position. 20

Figure 7. Genotype data of 30 most significant associated SNPs with hydrocephalus in Friesian horses. Orange colour represent homozygous for the unfavorable allele. Blue colour represent heterozygous. Red colour represent homozygous for the favorable allele. White colour with '--' represent no genotype information..... 23

Figure 8. The region contained 68 associated SNPs with hydrocephalus in Friesian horses on chromosome 1. $-\log_{10}$ *P-values* from chi-square statistics testing allele frequency differences between cases and controls before permutation testing with corrected *P-values*. 24

Figure 9. Regional plots of BIEC2-31300 (73,545,254 bp) with three different sizes of down-stream and up-stream (20 and 30 Mbp) of Friesian horses on chromosome 1. $-\log_{10}$ *P-values* from chi-square statistics testing allele frequency differences between cases and controls before permutation testing with corrected *P-values*..... 25

Figure 10. Associated haplotype blocks with hydrocephalus in Friesian horses on chromosome 1. 27

Figure 11. Genes in region BIEC2-23808–BIEC2-37017 (29.3 Mb) with most significant associated SNPs..... 29

Figure 12. Synteny of horse chromosome 1 in human and dog chromosomes... 30

Figure 13. Steps overview of development of a DNA-test for hydrocephalus in the Friesian horse breed. 38

Figure 14. Breeding strategy of the Friesian horse breed. 41

LIST OF TABLES

Table 1. Genetic crosses between paternal carrier and maternal carrier..... 4

Table 2. Number of case and control offspring per sire..... 7

Table 3. SNPs significantly (*P-value* < 0.05) associated with hydrocephalus in Friesian horses based on allele frequency differences between cases and controls after permutation testing..... 19

Table 4. 30 SNPs (lowest *P-value*) associated with hydrocephalus in Friesian horses based on genotype frequency differences between cases and controls after permutation testing..... 21

Table 5. 12 associated haplotype blocks (lowest *P-values*) with hydrocephalus in Friesian horses on chromosome 1 after 5 million permutation tests..... 26

Table 6. 16 different hits that related to hydrocephalus were identified in animal by Online Mendelian Inheritance in Animal (OMIA) database. 27

Table 7. Candidate genes of hydrocephalus from Ensembl 71..... 28

Table 8. Associated haplotype blocks with associated SNPs, including candidate genes, and synteny with human and dog..... 31

Table 9. 68 SNPs significantly (*P-value* < 0.05) associated with hydrocephalus in Friesian horses based on genotype frequency after permutation testing..... 49

Table 10. 54 associated haplotype blocks (*P-value* < 0.05) with hydrocephalus in Friesian horses based on chromosome 1 after 5 permutation testing..... 52

CHAPTER 1. INTRODUCTION

1.1. Background

1.1.1. The Friesian horse

The Friesian horse breed is one of the oldest original Dutch breeds that originated in Friesland, the Netherlands. It is characterized by its distinctive black coat color, high-stepping trot, powerful general build and good bone structure. Friesian horses have been registered since the original Friesian studbook was established in 1879. The current active breeding population amounts to roughly 100 breeding stallions and 8,000 breeding mares (Ducro *et al.*, 2011). However, inbreeding rate has been beyond 1% per generation for many generations in the past (Sevinga *et al.*, 2004) as the population has been closed for outside breeding and has a limited genetic pool. This inbreeding rate of the Friesian horse breed is higher than the standard of 1% per generation that has been set by FAO (1998). It is important to keep the inbreeding rate below 1% to prevent the loss of genetic variation and maintain genetic diversity. Maintaining genetic diversity within species is important for adaptation to a new environment that is constantly changing. Moreover, loss of genetic diversity can lead to inbreeding depression, which reduces performance, health and fertility. Also, loss of genetic variation can increase the chance of genetic disorders.

The high rate of inbreeding within the Friesian horse population has increased the chance of genetic disorders. One of these genetic disorders is hydrocephalus (i.e., “water on the brain”), which is a lethal disease. The incidence rate of hydrocephalus has been estimated at 0.25% (Sipma *et al.*, 2011). However, possibly not all affected individuals are reported to the studbook by the breeder, which means the incidence rate is likely to be underestimated.

1.1.2. Hydrocephalus

Hydrocephalus is defined as an active distension of the ventricular system of the brain resulting from inadequate passage of cerebrospinal fluid (CSF) from its point of production within the cerebral ventricles to its point of absorption into the systemic circulation (Rekate, 2008). Rekate (2008) has defined hydrocephalus definition because it is a complicated neurologic disorder with many causes and methods of treatment. As hydrocephalus is an active condition, the definition does not require specific pathology processes, concept of CSF dynamics, and how or where CSF is produced and absorbed (Rekate, 2008). Therefore, classification of hydrocephalus can be made based on Retake's definition.

In human, the incidence rate of hydrocephalus has been estimated varying between 0.48 and 0.81 affected children per 1000 births (Fernell *et al*., 1994; Hoppe-Hirsch *et al*., 1998; Schurr *et al*., 1953). The Online Mendelian Inheritance in Man (OMIM) and Animal (OMIA) databases recognized, respectively 322 and 16 different hits that relate to hydrocephalus. About 4.5 to 14.2% of cases of hydrocephalus in human are caused by the mutation of *L1CAM* (L1 cell adhesion molecule) gene on chromosome X with a recessive mode of inheritance (Haverkamp *et al.* 1999; Schrandel-Stumpel and Vos, 2004).

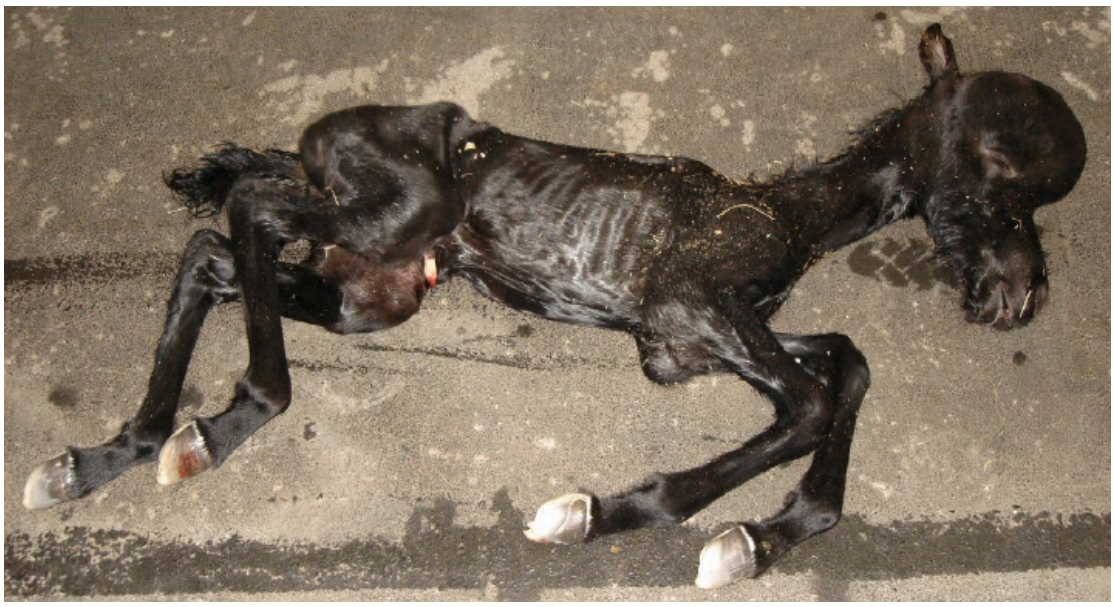

Figure 1. Friesian foal with hydrocephalus (Sipma *et al.*, 2011).

In horses, hydrocephalus is an uncommon disorder. In the Friesian horse breed, hydrocephalus has been observed more often than in other breeds. Hydrocephalus has also been observed in Thoroughbred horses (Bowman, 1980), Standardbred trotter (Ojala and Ala-Huikku, 1992), Przewalski horses (Frackowiak *et al*., 2006), and American Miniature horses (Ferris *et al*., 2010) however, with no estimate of incidence rate given. Ojala and Ala-Huikku (1992) investigated the mode of inheritance of hydrocephalus based on family pedigree (only three generations) in Standardbred trotter horses. One Standardbred trotter stallion (first generation) had seven affected and three unaffected offspring of both sexes within 6 years (second generation). The dam of the foal in the third generation was also affected from hydrocephalus, where the sire of this dam was from the stallion of the first generation. Their results showed that the mode of inheritance of hydrocephalus was neither an autosomal recessive disorder nor X-linked disorder. It was found that internal hydrocephalus in Standardbred trotter horses most likely was due to a dominant mutation in the germ-line in one of the parents or in an embryo (Ojala and Ala-Huikku, 1992).

The cases of hydrocephalus in Friesian horses can be caused by an abnormal narrowing of the passage of the jugular foramen, thus leading to compression of the internal jugular vein, perturbed CSF drainage and increased CSF accumulation (Boerma *et al.*, 2012; Sipma *et al.*, 2011). But this cause is not proven for 100% certainty. The genetic basis of hydrocephalus in Friesian horse breed has not been investigated yet. However, the mode of inheritance of hydrocephalus was autosomal recessive in golden hamster (Yoon and Slaney, 1972).

Yoon and Slaney (1972) investigated the mode inheritance of hydrocephalus in golden hamster by mating golden hamsters that are carrier. The parents previously already produced one or more offspring with hydrocephalus. Golden hamster progeny were checked at 12-14 days old using abnormal head size and shape as the criteria. Usually affected golden hamsters have dome-shaped heads. They determined the phenotype by dissecting the brain, if the size of the lateral ventricles is bigger than normal; they were classified with the possibility to suffer from hydrocephalus. They conclude that

hydrocephalus in Golden hamster is autosomal recessive based on the examination of the phenotype at age 4, 6, 10, and 12-14 days old after birth. These different ages were also used to see the development of hydrocephalus.

1.1.3. Hardy-Weinberg equilibrium

Hydrocephalus in Friesian horses is expected to be autosomal recessive disease. However, it is not possible to discriminate between the carrier (heterozygous) and the non-carrier of hydrocephalus (normal homozygous) based on phenotype. Furthermore, once a horse has sired a hydrocephalus offspring, it should be carrier of the unfavorable allele. When this occurs the Hardy-Weinberg equilibrium is assumed to apply. The carrier frequency of hydrocephalus mutation can be estimate from Hardy–Weinberg equation p^2 + 2pq + q^2 (Table 1); it states that the genotype frequencies AA (p^2), 2Aa (2pq), and aa (q^2) will not change if the allele frequencies remain constant from generation to generation. Consider when a Friesian horse population has alleles frequencies for favorable allele is 50% (p = 0.5) and other 50% (q = 0.5) is unfavorable allele. When two carriers are mated (Table 1) there is 25% chance of AA (unaffected), 50% of Aa (carrier), and 25% chance of aa (affected). So, when allele frequency of the unfavorable allele in Friesian horses is high, the chance of an affected foal will be born is higher. For example: If p = 85% and q = 15%, then the genotype distribution in the population will be 72.25% AA (unaffected), 2.25% aa (affected) and 25.5% (12.75 + 12.75) will be Aa (carrier).

Table 1. Genetic crosses between paternal carrier and maternal carrier

		Maternal allele(carrier)	
		A (p) = 50% | 85%	a (q) = 50% | 15%
Paternal allele (carrier)	A (p) = 50% | 85%	AA (p^2) = 25% | 72.25%	Aa (pq) =25% | 12.75%
	a (q) = 50% | 15%	Aa (pq) = 25% | 12.75%	aa (q^2) = 25% | 2.25%

A = the favorable allele
a = the unfavorable allele

Therefore, to have better information of the percentage of carriers in the population and to prevent hydrocephalus appears in the population, it is necessary to have a good estimate of the frequency of unfavorable allele. Hence, this study is important as a first step to get better information of hydrocephalus genetic background in Friesian horses. Furthermore, it can be used to develop a DNA-test that can identify a carrier of hydrocephalus in the population. Afterward, this DNA-test can be implemented in breeding program that can be used gradually to reduce or remove hydrocephalus from the Friesian horse population.

1.2. Research Questions

The general objective of the investigation of the genetic basis of hydrocephalus is to develop a DNA-test for hydrocephalus in the Friesian horse breed. Using genetic markers (e.g., single-nucleotide polymorphisms or SNPs), a DNA-test can be used to identify mutation carriers within the Friesian horse population and remove these carriers from breeding or provide other breeding strategies to prevent the genetic defect from being passed along to subsequent generations.

The specific objective of this thesis is to identify SNP(s) and haplotype(s) associated with hydrocephalus in Friesian horses using a genome-wide association study (GWAS) and haplotype association study, and to identify the location of homozygous chromosomal segments in affected horses by looking at the individual genotypes. Therefore, research questions of this thesis are: (a) Is there any association between disease status and SNP(s) or haplotypes? (b) Are homozygous regions shared by all affected horses and are these regions not homozygous for the disease allele in unaffected horses?

CHAPTER 2. MATERIALS AND METHODS

Genomic analysis based on case-control data was performed as the first step to develop a DNA-test for detecting carriers of hydrocephalus in Friesian horses. Material was from ear tissue samples of fetuses and stillborn foals, as the disorder is lethal. Several methods that were used in this study were genome-wide association study (GWAS) and haplotype association study. Moreover, Genes that related to hydrocephalus were sought among other things using information from other species. These methods were used to identify the location and the region within the chromosome that causes hydrocephalus in Friesian horses. Before these methods were performed, the data needed to be properly cleaned by performing quality control, check for the structure of the samples, and correct for the population structure to prevent false-positive and false-negative associations.

2.1. Data

Data on hydrocephalus in Friesian horses contained 20 cases and 47 controls.

Table 2. Number of case and control offspring per sire

Number of case offspring per sire	Number of controls offspring per sire	Number of sires
0	1	19
0	1	2*
0	2	2
0	3	2
1	1	4
2	1	2
4	1	1
2	2	1
1	6	1
1	0	1
1	0	1*
2	0	2

*Unknown sire.

44.8% of the data contained half-sibs horses descending from 9 sires with both cases and controls among their offspring (Table 2). 55.2% of the data contained horses descending from 29 sires with only cases or controls among their offspring. DNA samples were mostly from ear tissue samples. Data and DNA samples were collected by Utrecht University. During several years, hydrocephalus cases were actively sought through advertisements. Breeders voluntarily sent in tissue from their affected foal, often via their own veterinarian. Samples were genotyped using the Illumina® EquineSNP50 Genotyping BeadChip containing 54,602 SNPs.

2.2. Quality control

The identification of true genetic associations in genome-wide association study depends upon the overall quality of the data. If the genotypic data has not been properly cleaned, it can lead to false-negative and false-positive associations. Therefore, the GenABEL package (Aulchenko *et al.*, 2007) in R 1.7-4 (http://www.genabel.org/packages/GenABEL) was used to perform quality control by applying the *check.marker* function. This procedure identifies samples and SNPs that should be removed from the data according to certain criteria set by the researcher. The criteria that were used in this study were based on default settings of Haploview 4.2. There were two steps of the iterative procedure of the *check.marker* function to exclude SNPs based on a) per SNP statistics and b) per individual and between individual statistics. In the first step (a), SNPs with a call rate ≤75% across the samples, SNPs with minor allele frequency ≤0.05, and SNPs with P-value for Hardy-Weinberg equilibrium ≤0.001 were removed using the *summary.snp.data* function. In the second step (b), individuals with a call rate ≤90% across SNPs were removed using the *perid.summary* function. The quality control procedure was repeated iteratively until no further SNPs were eliminated and all criteria were met.

371 SNPs were excluded due to failed genotyping (no genotype information, which were labeled with '--'). 153 SNPs with a call rate ≤75%, 24,019 SNPs with minor allele frequency ≤0.05, and 232 SNPs with P-value for Hardy-Weinberg equilibrium ≤0.001 were removed. 29,935 SNPs (54.8%) from

54,602 SNPs were used in the analyses. And all individuals had a call rate ≥90% and were therefore included in the analyses.

2.3. Population stratification

Population stratification occurs when the samples consists of multiple subgroups of individuals who systematically differ in allele frequency within a population. This might be due to different ancestry or nonrandom mating between groups or differences in the proportion of causal allele/genotype in the subgroups. It can be a problem for association studies because false-positive associations could be detected. Two methods were used, to check the structure of the sample and to correct for stratification. The first method was used to check the structure of the samples, which is described in section 2.3.1. The second method is genomic control that was used to account for population structure, which is described in section 2.3.2.

2.3.1. Population structure

Population structure can be determined from the kinship coefficient among individuals in the population. A matrix of genomic kinship was generated from the case-control data in the population using genotypes of SNPs that passed quality control. Differences between cases and controls in kinship might result from stratification and might lead to false-positive associations. The kinship was computed by the *ibs* function of the GenABEL package in R 1.7-4 (Aulchenko *et al.*, 2007) as:

$$f_{i,j} = \Sigma_k \frac{(x_{i,k} - p_k) * (x_{j,k} - p_k)}{(p_k * (1 - p_k))}$$

where $f_{i,j}$ = genomic kinship (identity-by-state) between horse *i* and *j*, based on *k* = 28,536 autosomal SNPs, where X-chromosome was excluded. $x_{i,k}$ and $x_{j,k}$ are the genotypes (AA = 0, Aa = 0.5, aa = 1 which were arbitrarily chosen) of the *i*th or *j*th horse for SNP *k* and p_k is the allele frequency.

The matrix of genomic kinship was converted to a distance matrix, which was used to carry out classical multidimensional scaling (MDS) (Gower, 1966). The genomic kinship matrix was converted to a distance matrix by the *as.dist* function as:

$$d_{ij}^2 = \sum_{r=1}^{v} (x_{ir} - x_{jr})^2,$$

where $d^2_{i,j}$ = distance between individual points of *i* and *j* horse in a Euclidean space. x_{ir} and x_{jr} where the individual points of *i* and *j* horse have coordinates that referred to rectangular axes. After a distance matrix was made, classical MDS was constructed by the method of principal components. Principal components analysis was regarded as a rotation of the axes in this space, consider the variance of all individuals when projected at right-angles onto an axis, these projections are called the "scores" for that axis (Weale, 2010). The first principal component (PC) axis was the one that has the maximum possible variance of its PC scores. The second PC axis has the maximum possible variance of its PC scores, conditional on it being at right-angles to the first PC axis (Weale, 2010). The plot visualized distances between the group of cases and controls.

2.3.2. Genomic control

In genomic control, test statistics are corrected for the observed inflation, which was caused by stratification. This stratification was corrected by adjusting the test statistics at each SNP by an equal overall inflation factor. The quantile-quantile (Q-Q) was used to visualize the observed test statistics (based on associations between genotyped SNPs and hydrocephalus in Friesian horses) versus test statistics expected under the null hypothesis with no association with hydrocephalus. Observed test statistics (chi-square) were calculated, then ranked in order from smallest to largest on the y-axis and plotted against the distribution of expected chi-square under the null hypothesis of no association (on the x-axis). The Q-Q plot also showed deviations, which can be seen within the two lines that were produced. The first line is the fitted slope and the second line is expected under no inflation and association. This deviations from the null

hypothesis were assumed to be a result of either an incorrect distribution of expected chi-square under the null hypothesis or true associations with the investigate phenotype (Pearson and Manolio, 2008).

Inflation factor was produced by the *estlambda* function of the R 1.7-4 package GenABEL (Aulchenko *et al*., 2007). Estimated inflation factor is $\lambda = 1$ when no population stratification is present. An estimated inflation factor of $\lambda > 1$, indicates presence of population stratification. The standard error (se) is also produced by the *estlambda* function. However, according to the manual, the se cannot be used to test if inflation is significant and make conclusions about presence of stratification (Aulchenko *et al*., 2007).

2.4. Genome-wide association study

Genome-wide association study (GWAS) was used to identify genetic associations between SNPs and hydrocephalus in Friesian horses using allele and genotype frequency differences between cases and controls. GWAS based on allele frequency difference was used to find the allele that was more frequent in cases, which was associated with hydrocephalus. Furthermore, GWAS based on genotype frequency differences is more important because it was used to identify the homozygous individuals for the unfavorable allele. The mode of inheritance of hydrocephalus is expected to be autosomal recessive in Friesian horses. All cases are therefore expected to be homozygous for the unfavorable allele, whereas controls are expected to be heterozygous or homozygous for the favorable allele.

GWAS was carried out on the case-control data using GenABEL package in R 1.7-4 (Aulchenko *et al*., 2007). The *ccfast* function was used to perform case-control analysis by computing chi-square test from 2x2 (allelic) or 2x3 (genotypic) tables. After association was found between SNPs and the phenotype, permutation testing was required to determine how often the SNPs would appear by chance if the study were repeated and if there were true-positive associations. Moreover, permutation testing was used to set the threshold of significance while accounting for multiple testing. Associated SNPs

were ascertained through multiple testing by doing permutation testing (*n* = 1,000,000) to find significant association (p-value < 0.05).

2.5. Homozygous segments

Individual genotypes (of each horse, cases and controls) of highest association SNPs (with the lowest *P-values*) were investigated to identify the locations of homozygosity because the associated SNPs were expected to be homozygous in all affected Friesian horses.

2.6. Haplotype construction and association

Haploview 4.2 software (http://www.broadinstitute.org/haploview) was used to construct haplotype blocks, calculate haplotype population frequency, and perform haplotype association tests (Barrett *et al*., 2005). Haplotype construction and subsequent association testing provide information on ancestral chromosome segments that may harbor alleles that influence disease phenotypes.

A haplotype block was created from a region of comparisons among informative SNP pairs that show strong linkage disequlibrium (LD) and strong evidence for historical recombination. For strong LD, the pair of informative SNP was defined if the one-sided upper 95% confidence interval bound on D prime was >0.98 (no historical recombination) and the lower bound was >0.7. The fraction of the strong LD in informative comparisons should be at least D prime >0.95. For strong evidence for historical recombination, if the upper 95% confidence interval bound on D prime was <0.9. For both strong LD and strong evidence for historical recombination, markers with MAF <0.05 was excluded (Gabriel *et al*., 2002). In this study, haplotype construction was only performed in the chromosome in which the most significant SNPs associated with hydrocephalus were found.

Associated haplotype blocks were ascertained by doing permutation testing (n = 5,000,000) to find significant association (p-value < 0.05).

2.7. Candidate genes

Ensembl 71 is a genome database that contains genomic information for vertebrates and other eukaryotic species (http://www.ensembl.org/index.html) and is freely available online. It was used to find genes associated with hydrocephalus in other species, and also to identify candidate genes and their functions that were located in the region of SNPs and haplotype blocks associated with hydrocephalus in our study. Ensembl 71 database was also used to identify the synteny of horse chromosomes, finding similar blocks of genes in the same relative positions in the genome within different species. It helps to find the homologous genes between different species with the same region.

Online Mendelian Inheritance in Man (OMIM) is a database of human genes and genetic phenotypes that is freely available online (http://www.ncbi.nlm.nih.gov/omim). Online Mendelian Inheritance in Animal (OMIA) is a database of genes, inherited disorders and traits in more than 135 animal species (other than human and mouse), (http://www.ncbi.nlm.nih.gov/omia). OMIM and OMIA were used to find genes and phenotypes of hydrocephalus in, respectively, human and animal species.

CHAPTER 3. RESULTS

3.1. Population stratification

3.1.1. Detection of population stratification

The classical multidimensional scaling (MDS) plot showed genetic distances between 20 cases and 47 controls based on the genomic kinship matrix (Figure 2). The distribution of Friesian horses over the two-dimensional space did show some overlap between cases and controls in the first principal component (PC1). It shows on right-hand side Figure 2: PC1 between 0.0 and 0.05. This overlaps because some cases and controls offspring descending from the same sires. In the second principal component (PC2), there was no population stratification. From this result, population stratification between cases and controls seems to be present due to some affected and unaffected Friesian horses descending from different sires.

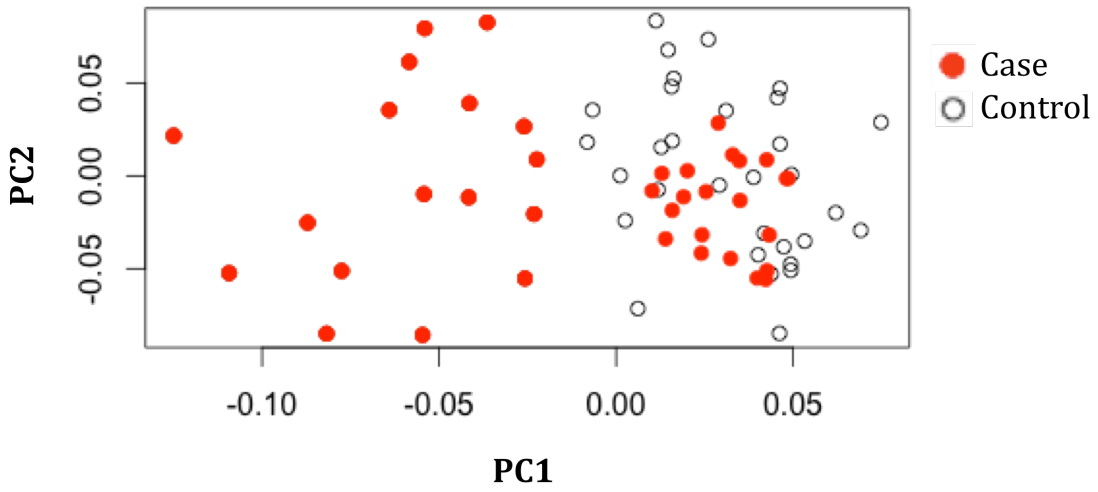

Figure 2. The classical multidimensional scaling plot of genetic distances between cases and controls. Filled circles are the cases and open circles are the controls. The first two principle components are PC1 and PC2.

3.1.2. Genomic control

Quantile-quantile (Q-Q) plot showed the distribution of the observed vs. expected chi-square statistics of 29,935 SNPs in Friesian horses (Figure 3). Figure 3A, the Q-Q plot shows deviations from the expected distribution, with a genomic inflation factor of λ = 1.53 (standard error = 0.003).

It can be seen within the two lines that were produced. This deviation from the expected distribution is observed, which was assumed to originate from an incorrect distribution or true associations with hydrocephalus. The inflation of observed test statistics indicates population stratification. Correction for this inflation was by simple division of the unadjusted observed *P-value* with the inflation factor, which reduces observed test statistics from line 1 to line 2. Genomic control successfully corrected population stratification, as λ = 1.00 (Figure 3B). After the inflation was corrected, a sharp deviation above expected chi-square value of approximately 7.5 was observed, which could be due to strong associations with hydrocephalus in Friesian horses. Moreover, false-positive associations are possibly not be detected in this study.

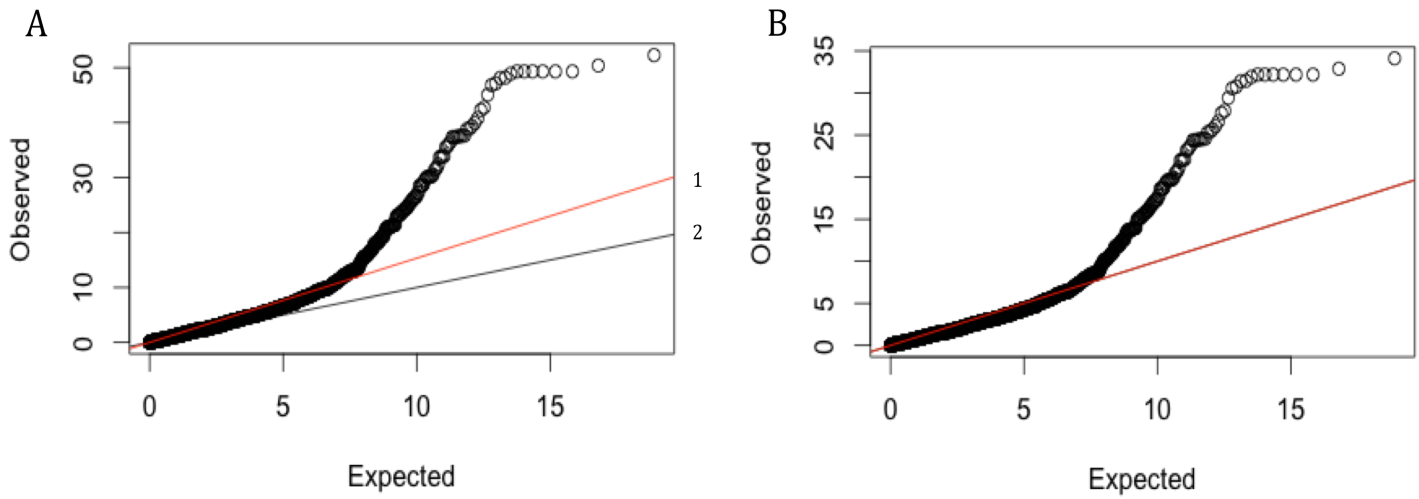

Figure 3. Quantile-quantile plots of chi-square statistics testing allele frequency differences between cases and controls. Line 2 is expected under no inflation and association. Line 1 is fitted slope. A is the Q-Q plot without the correction of *P-values* and B is the Q-Q plot with corrected *P-values*.

Figure 4 shows the Q-Q plot based on *P-values* for an association of genotype frequency. Deviation from expected distribution is observed with a genomic inflation factor λ = 1.38 (standard error = 0.002). This inflation indicated population stratification. However, this inflation could not be corrected, which means false-positive associations are possibly detected in this study.

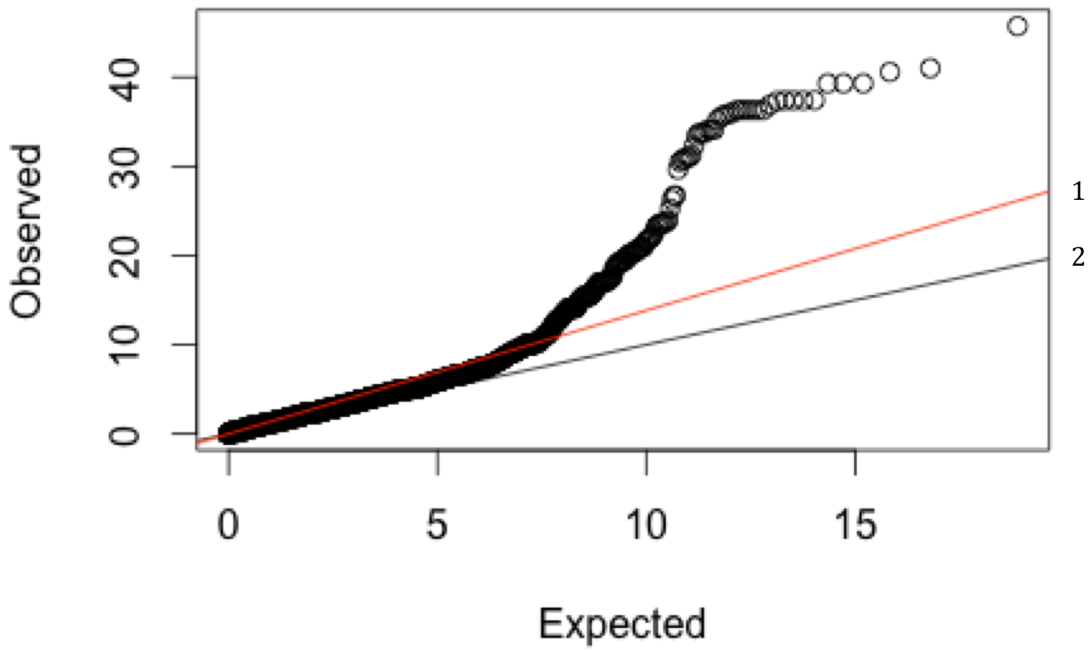

Figure 4. The quantile-quantile plot hydrocephalus in Friesian horses from chi-square statistics testing genotype frequency differences between cases and controls. Line 2 is expected under no inflation and association. Line 1 is fitted slope.

3.2. Genome-wide association study

3.2.1. GWAS based on allele frequency

The genome-wide association study plot based on allele frequency differences between cases and controls, sorted by chromosome, is shown in Figure 5. This figure shows the association before permutation testing, and corrected for population stratification. SNPs with *P-value* < 10^{-5} associated with hydrocephalus in Friesian horses were only identified on chromosome 1.

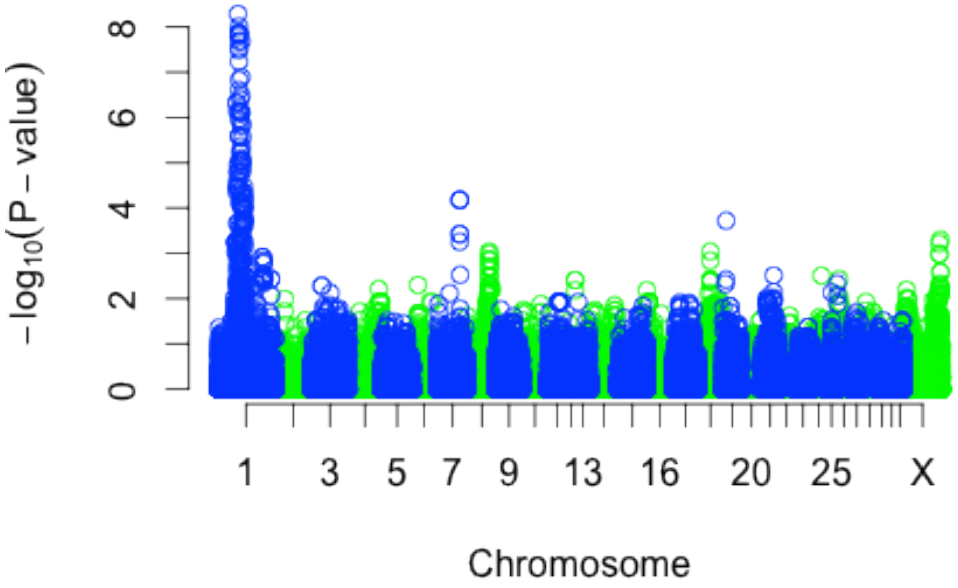

Figure 5. Manhattan plot of hydrocephalus in Friesian horses. Association of 29,935 SNPs with hydrocephalus represented by -$\log_{10}$ corrected *P-values* from chi-square statistics testing allele frequency differences between cases and controls plotted by chromosome and sorted by chromosomal position.

One million permutation testing was performed to correct for multiple testing to reduce false-positive associations. 29 SNPs on chromosome 1 were found to be significantly associated (*P-value* <0.05) with hydrocephalus in Friesian horses (Table 3). The most significant SNP after permutation testing is BIEC2-27588 at position 66,296,523 bp with *P-value* = 4 x 10^{-6} and *corrected P-value* = 11 x 10^{-4}, with the frequency of unfavorable allele 0.95 for cases and 0.26 for controls. The odds ratio of unfavorable allele effect for BIEC2-27588 was 5.59, which means the odds of Friesian horses heterozygous for unfavorable allele is increased by 5.59 times (= $e^{1.72}$, 1.72 was the log-effB, which effB corresponds to the odds ratio estimate for the SNP) compared to Friesian horses homozygous for favorable allele. Moreover, the odd of Friesian horses homozygous for unfavorable allele is increased by 31.19 times (= $e^{2x1.72}$) compared to Friesian horses homozygous for favorable allele.

Frequencies of the unfavorable alleles were ≥90% in cases and ≤33% in controls for BIEC2-27588 (66,296,523 bp), BIEC2-28464 (68,708,147 bp), BIEC2-28874 (69,326,275 bp), BIEC2-28875 (69,339,109 bp), BIEC2-28876

(69,339,208 bp), and BIEC2-29359 (70,378,596 bp) (Table 3). So, these unfavorable alleles were expected to be mostly present in cases than in controls. Odds ratios for the significant SNPs were between 0.15 and 5.59 (Table 3). BIEC2-28464 has the lowest odds ratio (0.15) of unfavorable allele effect, but it has allele frequency of 0.98 for cases.

Table 3. SNPs significantly (*P-value* < 0.05) associated with hydrocephalus in Friesian horses based on allele frequency differences between cases and controls after permutation testing

						Allele frequency*	
SNP name	Position (bp)	N	Corrected P-value	Allele	OR (95% CI)°	Cases	Controls
BIEC2-25335	61,311,028	67	0.0288	A:G	3.30 (2.52-4.08)	0.85	0.27
BIEC2-25357	61,479,005	67	0.0288	A:C	3.30 (2.52-4.08)	0.85	0.27
BIEC2-27351	65,841,804	67	0.0372	A:G	3.38 (2.54-4.22)	0.88	0.30
BIEC2-27352	65,842,030	67	0.0372	A:G	3.38 (2.54-4.22)	0.88	0.30
BIEC2-27588	66,296,523	66	0.0011	A:G	5.59 (3.90-7.27)	0.95	0.26
BIEC2-28302	68,597,487	67	0.0018	G:A	3.65 (2.77-4.52)	0.78	0.15
BIEC2-28459	68,701,400	67	0.0018	A:C	3.65 (2.77-4.52)	0.78	0.15
BIEC2-28464	68,708,147	67	0.0046	C:A	0.15 (-5.21-5.51)	0.98	0.34
BIEC2-28506	68,778,089	67	0.0018	A:G	3.65 (2.77-4.52)	0.78	0.15
BIEC2-28507	68,778,169	67	0.0061	A:G	3.40 (2.62-4.18)	0.78	0.17
BIEC2-28749	69,062,965	66	0.0025	G:A	3.61 (2.74-4.47)	0.78	0.16
BIEC2-28874	69,326,275	67	0.0119	C:A	3.85 (2.80-4.90)	0.90	0.29
BIEC2-28875	69,339,109	67	0.0021	A:C	4.24 (3.06-5.41)	0.90	0.25
BIEC2-28876	69,339,208	67	0.0162	A:G	3.77 (2.75-4.79)	0.90	0.30
BIEC2-28907	69,505,589	67	0.0042	G:A	3.52 (2.69-4.35)	0.78	0.16
BIEC2-28908	69,543,750	66	0.0015	G:A	3.75 (2.83-4.67)	0.78	0.14
BIEC2-28912	69,561,176	67	0.0018	G:A	3.65 (2.77-4.52)	0.78	0.15
BIEC2-28931	69,697,043	67	0.0018	A:G	3.65 (2.77-4.52)	0.78	0.15
BIEC2-28955	69,736,571	67	0.0018	A:G	3.65 (2.77-4.52)	0.78	0.15
BIEC2-29359	70,378,596	67	0.0453	A:G	3.53 (2.59-4.48)	0.90	0.33
BIEC2-31300	73,545,254	67	0.0365	G:A	3.02 (2.40-3.64)	0.78	0.22
BIEC2-31328	73,575,357	67	0.0365	A:C	3.02 (2.40-3.64)	0.78	0.22
BIEC2-31580	73,992,006	67	0.0218	G:A	3.11 (2.45-3.76)	0.78	0.20
BIEC2-32202	75,458,447	67	0.0382	G:A	3.06 (2.41-3.70)	0.80	0.23
BIEC2-32205	75,461,775	67	0.0382	G:A	3.06 (2.41-3.70)	0.80	0.23
BIEC2-32520	76,154,827	67	0.0382	A:C	3.06 (2.41-3.70)	0.80	0.23
BIEC2-32706	76,631,203	67	0.0189	A:C	3.11 (2.45-3.76)	0.75	0.17
BIEC2-32912	76,963,647	67	0.0025	A:G	3.57 (2.72-4.41)	0.75	0.14
BIEC2-32970	77,220,026	67	0.0106	A:G	3.25 (2.54-3.96)	0.73	0.15

*Allele frequency of the unfavorable (hydrocephalus) allele.

°Odds ratio estimate for unfavorable allele effect calculated as $e^{log\text{-}effB}$; 95% confidence interval between parentheses.

3.2.2. GWAS based on genotype frequency

GWAS based on genotype frequency differences between cases and controls was performed to determine the mode of inheritance of hydrocephalus, which is expected to be autosomal recessive in Friesian horses. It was also used to look if the significant associated SNPs based on allele frequency were also significantly associated based on genotype frequency.

Figure 6 shows the genome-wide association study plot based on genotype frequency differences between cases and controls, which were sorted by chromosome. SNPs significantly associated (*P-value* < 10^{-5}) with hydrocephalus in Friesian horses were again identified on chromosome 1.

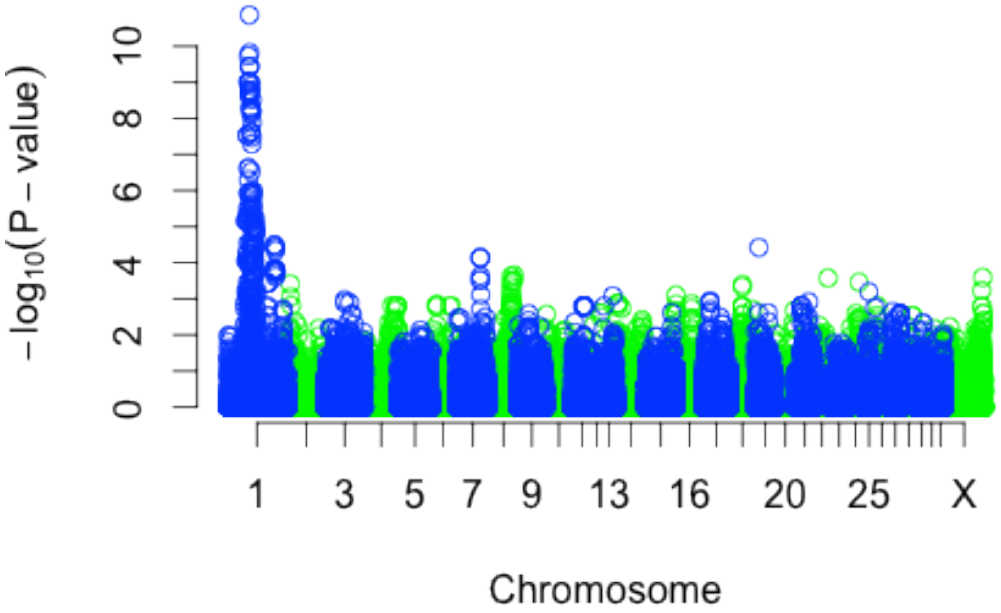

Figure 6. Manhattan plot of hydrocephalus in Friesian horses. Association of 29,935 SNPs with hydrocephalus represented by -$\log_{10}$ *P-values* from chi-square statistics testing genotype frequency differences between cases and controls plotted by chromosome and sorted by chromosomal position.

68 associated SNPs (*P-value* < 0.05) on chromosome 1 were found to be the most significant SNPs after 1,000,000 permutation testing (Appendix, Table 9). Table 4 only shows 30 SNPs with the lowest p-value. All 68 associated SNPs included the 29 associated SNPs that showed a significant difference on genotype frequency between cases and controls. However, these 68 SNPs were

not corrected for population stratification. It indicated that false associations with hydrocephalus could be present within these 68 SNPs, except the 29 SNPs associated based on allele frequency differences because their *P-values* had already been corrected. 18 associated SNPs at position between 65,841,804 to 70,378,596 bp were found to have the lowest *P-value* = 10^{-6} (Table 4).

Based on genotype frequency differences, Table 4 shows that BIEC2-27588 (66,296,523 bp) and BIEC2-28874 (69,326,275 bp) were found ≥85% homozygous genotype for the unfavorable allele in affected Friesian horses and ≤45% heterozygous genotype in unaffected Friesian horses. It indicates that these two SNPs (homozygous genotype of unfavorable alleles) were more present in cases rather than in controls. Odds ratios of heterozygous for the significant SNPs were between 0.11 and 54.9 (Table 3). BIEC2-28464 has the lowest odds ratio of heterozygous genotype of 0.11, and it has heterozygous frequency of 0.05 for cases.

Table 4. 30 SNPs (lowest *P-value*) associated with hydrocephalus in Friesian horses based on genotype frequency differences between cases and controls after permutation testing

					Genotype count		
SNP name	Position (bp)	N	P-value	OR*	All	Controls	Cases
						AA AG GG	
BIEC2-25335	61,311,028	67	4.90E-05	2.06	18 23 26	3 19 25	15 4 1
						AA AC CC	
BIEC2-25357	61,479,005	67	4.90E-05	2.06	18 23 26	3 19 25	15 4 1
						AA AG GG	
BIEC2-27351	65,841,804	67	1.00E-06	0.74	20 23 24	3 22 22	17 1 2
						AA AG GG	
BIEC2-27352	65,842,030	67	1.00E-06	0.74	20 23 24	3 22 22	17 1 2
						AA AG GG	
BIEC2-27588	66,296,523	66	1.00E-06	54.6	20 20 26	3 18 26	17 2 0
						GG AG AA	
BIEC2-28302	68,597,487	67	1.00E-06	2.16	13 19 35	0 14 33	13 5 2
						AA AC CC	
BIEC2-28459	68,701,400	67	1.00E-06	2.16	13 19 35	0 14 33	13 5 2
						CC AC AA	
BIEC2-28464	68,708,147	67	1.00E-06	0.11	22 26 19	3 25 19	19 1 0
						AA AG GG	
BIEC2-28506	68,778,089	67	1.00E-06	2.16	13 19 35	0 14 33	13 5 2
						AA AG GG	
BIEC2-28507	68,778,169	67	1.00E-06	1.98	13 21 33	0 16 31	13 5 2
						GG AG AA	
BIEC2-28749	69,062,965	66	1.00E-06	2.13	13 19 34	0 14 32	13 5 2
BIEC2-28874	69,326,275	67	1.00E-06	1.41		CC AC AA	

					20 23 24	3 21 23 AA AC CC	17 2 1
BIEC2-28875	69,339,109	67	1.00E-06	1.55	19 21 27	2 19 26 AA AG GG	17 2 1
BIEC2-28876	69,339,208	67	1.00E-06	1.35	20 24 23	3 22 22 GG AG AA	17 2 1
BIEC2-28907	69,505,589	67	1.00E-06	2.07	13 20 34	0 15 32 GG AG AA	13 5 2
BIEC2-28908	69,543,750	66	1.00E-06	2.23	13 18 35	0 13 33 GG AG AA	13 5 2
BIEC2-28912	69,561,176	67	1.00E-06	2.16	13 19 35	0 14 33 AA AG GG	13 5 2
BIEC2-28931	69,697,043	67	1.00E-06	2.16	13 19 35	0 14 33 AA AG GG	13 5 2
BIEC2-28955	69,736,571	67	1.00E-06	2.16	13 19 35	0 14 33 AA AG GG	13 5 2
BIEC2-29359	70,378,596	67	1.00E-06	1.20	20 27 20	3 25 19 GG AG AA	17 2 1
BIEC2-31300	73,545,254	67	4.40E-05	1.80	14 23 30	1 18 28 AA AC CC	13 5 2
BIEC2-31328	73,575,357	67	4.40E-05	1.80	14 23 30	1 18 28 GG AG AA	13 5 2
BIEC2-31514	73,881,378	66	1.50E-05	0.18	21 28 17	4 26 16 GG AG AA	17 2 1
BIEC2-31580	73,992,006	67	4.10E-05	1.88	14 22 31	1 17 29 GG AG AA	13 5 2
BIEC2-32202	75,458,447	67	1.00E-05	1.51	15 24 28	1 20 26 GG AG AA	14 4 2
BIEC2-32205	75,461,775	67	1.00E-05	1.51	15 24 28	1 20 26 CC AC AA	14 4 2
BIEC2-32520	76,154,827	67	1.00E-05	1.51	15 24 28	1 20 26 AA AC CC	14 4 2
BIEC2-32706	76,631,203	67	6.10E-05	1.55	14 19 34	1 15 31 AA AG GG	13 4 3
BIEC2-32912	76,963,647	67	2.00E-06	1.72	13 17 37	0 13 34 AA AG GG	13 4 3
BIEC2-32970	77,220,026	67	2.30E-05	1.81	12 19 36	0 14 33	12 5 3

*Odds ratio (without standard error) estimate for heterozygous genotype effect calculated as $e^{log\text{-}effAB}$; in all Friesian horses of this study.

Individual genotypes of 30 most significantly associated SNPs with hydrocephalus in 67 Friesian horses are shown in Figure 7. In unaffected Friesian horses, 39 horses did not have SNP with homozygous genotype for the unfavorable allele. However, 8 unaffected Friesian horses were found to have some SNPs with homozygous genotype for the unfavorable allele. It can be that the linkage disequilibrium between SNPs and gene is not 100%, so unaffected Friesian horses with homozygous genotype for the unfavorable allele might be observed. Based on this result, 10 affected Friesian horses had homozygous genotype for the unfavorable allele for all 30 SNPs.

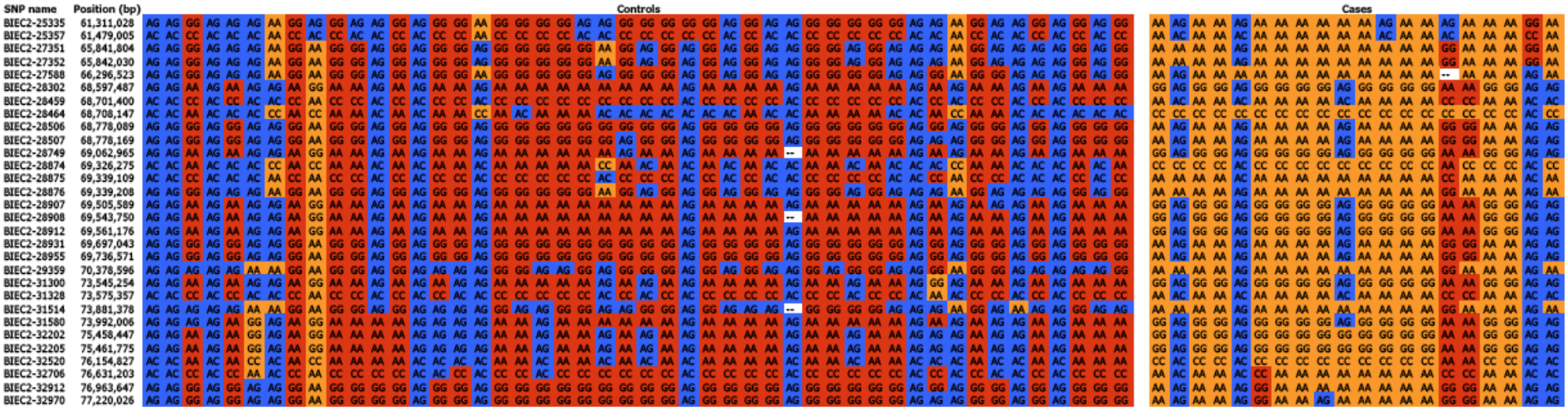

Figure 7. Genotype data of 30 most significant associated SNPs with hydrocephalus in Friesian horses. Orange colour represent homozygous for the rare-allele. Blue colour represent heterozygous. Red colour represent homozygous for the common-allele. White colour with '--' represent no genotype information.

A 29.3 Mb region starting from BIEC2-23808 (57,760,860 bp) to BIEC2-37017 (87,079,561 bp) was investigated in more detail (Figure 8). This region contained 693 SNPs including the 68 significantly associated SNPs based on genotype frequency differences.

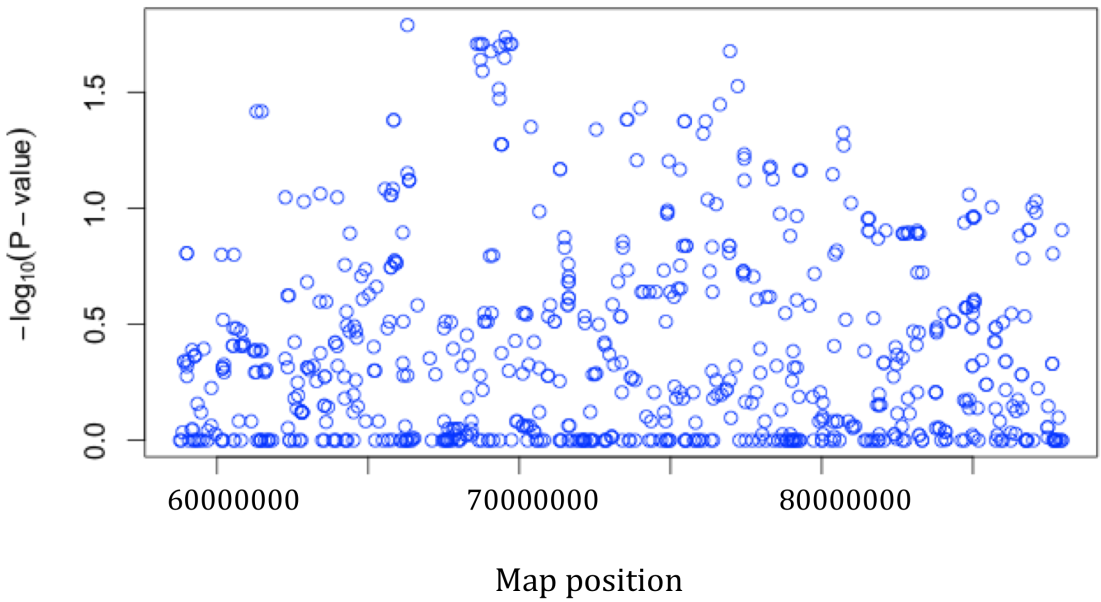

Figure 8. The region contained 68 associated SNPs with hydrocephalus in Friesian horses on chromosome 1. $-\log_{10}$ *P-values* from chi-square statistics testing allele frequency differences between cases and controls after permutation testing with corrected *P-values*.

Figure 9 shows regions of BIEC2-31300 (73,545,254 bp) with two different sizes of down-stream and up-stream, 68 associated SNPs were included. 952 SNPs, and 1,428 SNPs were found in the region of 20 Mb and 30 Mb, respectively.

Figure 8 and 9 showed the zoomed area of the peak that was found on chromosome 1 (Figure 5). It showed that this peak contained the significant associated SNPs (*P-values* < 0.05).

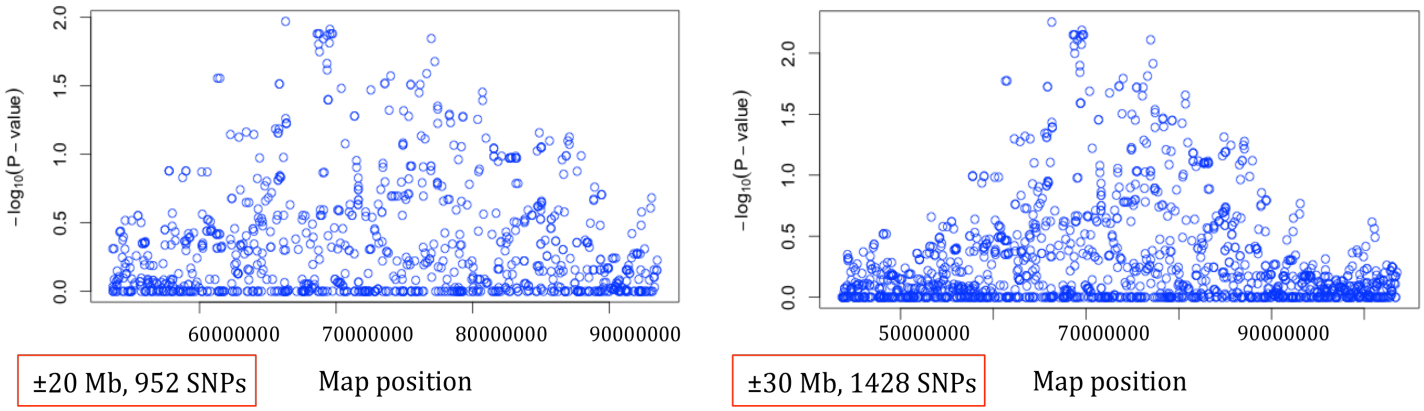

Figure 9. Regional plots of BIEC2-31300 (73,545,254 bp) with three different sizes of down-stream and up-stream (20 and 30 Mbp) of Friesian horses on chromosome 1. -$\log_{10}$ *P-values* from chi-square statistics testing allele frequency differences between cases and controls after permutation testing with corrected *P-values*.

3.3. Haplotype construction and association

Haplotype construction was used to identify combination of alleles (SNPs) at different loci on chromosome 1. As the Friesian horse breed is highly inbred, it is expected to have large haplotype blocks. 2,717 SNPs (62.1%) on chromosome 1 were used in the analyses after quality control. 54 haplotype blocks (*P-value* < 0.05) on chromosome 1 were found to be significantly associated with hydrocephalus in Friesian horses after five million permutation tests (Appendix, Table 10).

Table 5 only shows 12 associated haplotype blocks with the lowest p-value. These associated haplotype blocks include the top 30 SNPs that showed significant differences in genotype frequency between cases and controls. Haplotype block 124, 129 (GAGCAA), 130, 131, and 148 were found to have the lowest *P-value* after permutation testing (Table 5).

Based on haplotype block frequency results (Table 5), it showed that haplotype blocks 112, 122 (AA), and 124 were found to have haplotype frequency ≥85% in affected Friesian horses. However, not all alleles in these haplotype blocks are associated with hydrocephalus, instead it contained some associated SNPs in the block. The most significant SNP (BIEC2-27588) was found on haplotype block 124 with *P-value* after permutation testing of 0.00. Two

significant associated SNPs BIEC2-29359 and BIEC2-31514 (Table 8) were not found in any associated haplotype blocks because neighboring of these SNPs had low D prime, which indicated weak LD.

Table 5. 12 associated haplotype blocks (lowest *P-values*) with hydrocephalus in Friesian horses on chromosome 1 after 5 million permutation tests

Name	Position (bp)	Size (kb)	Multiallelic D prime	P-value[a]	P-value[b]	Haplotype frequency	
						Cases	Controls
Block 112: GAAAGG	61,436,332-61,636,387	200.055	0.709	8.19E-10	8.00E-07	0.850	0.272
Block 122: AA	65,841,804-65,842,030	0.226	0.371	1.62E-09	1.20E-06	0.875	0.304
Block 122: GG	65,841,804-65,842,030	0.226	0.371	1.62E-09	1.20E-06	0.125	0.696
Block 124: ACAGA	66,296,523-66,358,865	62.324	0.860	3.03E-11	0.00E+00	0.874	0.250
Block 129: ACGAGG	68,597,487-68,778,947	180.682	0.915	5.38E-09	3.40E-06	0.025	0.567
Block 129: GAGCAA	68,597,487-68,778,947	180.682	0.915	3.99E-12	0.00E+00	0.775	0.152
Block 130: GGAGGAGC	68,835,676-69,326,275	490.599	0.986	2.93E-12	0.00E+00	0.774	0.149
Block 131: AAAGAAGGGCAGA	69,339,109-69,736,571	397.462	0.868	4.08E-12	0.00E+00	0.775	0.152
Block 142: CCGA	73,417,368-73,575,357	157.989	1.000	5.43E-10	6.00E-07	0.774	0.204
Block 147: AGGAG	75,409,929-75,523,947	114.018	0.991	7.36E-10	6.00E-07	0.800	0.228
Block 148: AGGACGGAGAGA	76,508,159-76,961,647	455.488	0.943	7.00E-12	0.00E+00	0.750	0.141
Block 149: AAAGGAAG	77,220,026-77,435,722	215.696	0.942	5.27E-10	6.00E-07	0.700	0.152

[a]*P-values* before permutation testing.
[b]*P-values* after 5 million permutation testing..

3.4. Linkage disequilibrium

Figure 10 shows the linkage disequilibrium (LD) of the 54 associated haplotype blocks and it included the 68 associated SNPs, which showed significant difference in genotype frequency between cases and controls. Higher D prime-values were indicated by darker colour (red). All associated haplotype blocks with the lowest *P-values* (Table 5) had strong LD presented by a multiallelic D prime (from multi-SNPs) between 0.7 and 1.0. However, haplotype block 122 had strong evidence of historical recombination. Haplotype blocks 130, 142, and 147 (Table 5) have the strongest LD (multiallelic D prime >0.98). Associated haplotype block 130 contained SNP BIEC2-28874, for which the homozygous genotype for the unfavorable allele was present more in cases than in controls.

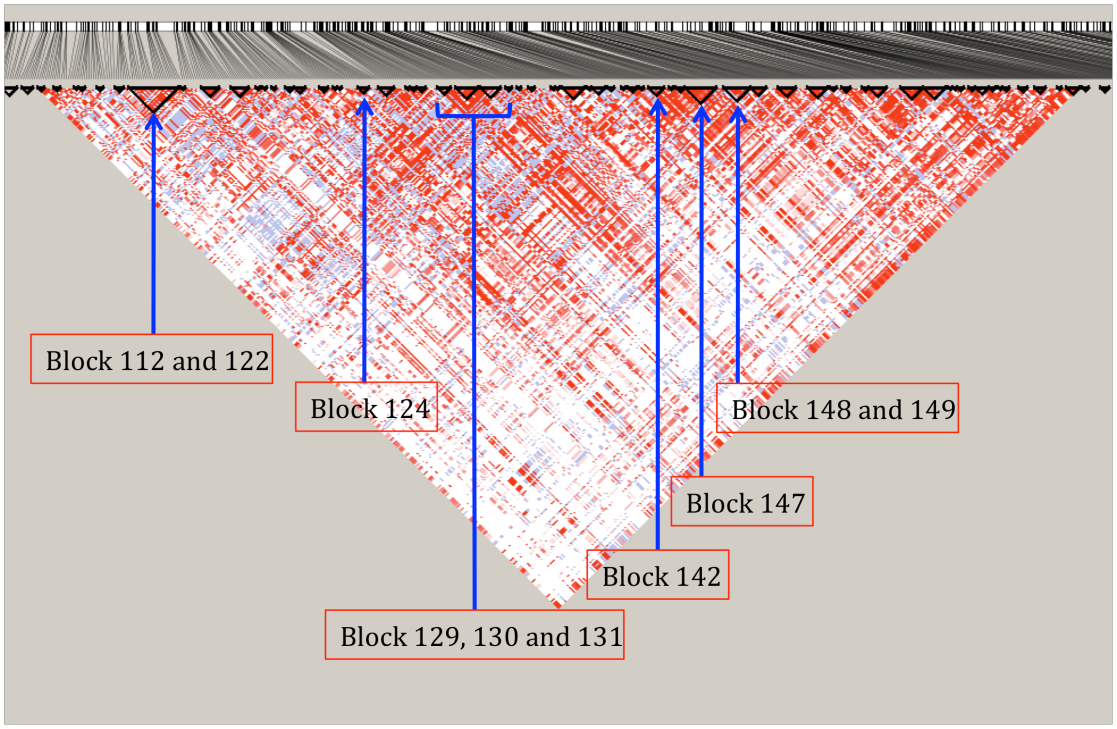

Figure 10. Associated haplotype blocks with hydrocephalus in Friesian horses on chromosome 1.

3.5. Candidate genes of hydrocephalus

322 hits were related to hydrocephalus in human and 16 hits (Table 6) in animals using Online Mendelian Inheritance in Man (OMIM) and Animal (OMIA) databases. These different hits have different phenotypes of hydrocephalus, and it included combination of hydrocephalus with other diseases or syndromes.

Table 6. 16 different hits that related to hydrocephalus were identified in animal by Online Mendelian Inheritance in Animal (OMIA) database

OMIA ID	Animals	Diseases
3334	Yellow-crowned parrot	Hydrocephalus*
3277	Cattle	Retinal dysplasia and internal hydrocephalus
2468	Llama	Hydrocephalus*
2369	Dog	Hydrocephalus, internal
1199	Domestic cat	Mannosidosis, alpha*
1974	Golden hamster	Hydrocephalus*
845	Cattle	Hydrocephalus, internal
841	Sheep	Hydrocephalus*
840	Rabbit	Hydrocephalus*
839	Domestic pig	Hydrocephalus*
838	Horse	Hydrocephalus*
837	Dog	Hydrocephalus*

836	Cattle	Hydrocephalus*
835	Domestic cat	Hydrocephalus*
500	Sheep	Dandy-Walker syndrome°
499	Domestic cat	Dandy-Walker syndrome°

*Enlargement of the cranium caused by accumulation of fluid.
ǂA lysosomal storage disease in which there is a buildup (storage) of mannose-rich compounds, due to the lack of the enzyme alpha-mannosidase, whose task is to cleave mannose from such compounds. Clinical signs include ataxia, head tremor, aggression, and finally paralysis and death.
°Congenital hydrocephalus (enlargement of the cranium due to accumulation of fluid; "water on the brain") due to obstruction of the foramina of Magendie and Luschka (openings within the brain).

Table 7 shows candidate genes of hydrocephalus in human and other species from Ensembl 71 database. Candidate genes were found in OMIM/A and Ensembl 71 databases and will be used to check the genes that are found based on GWAS and haplotype analysis.

Table 7. Candidate genes of hydrocephalus from Ensembl 71

Genes	Species		Horse chromosome
HYDIN	Chimpanzee, Ferret, Human, Macaque, Mouse, Panda, Rat		3:22599664-22895359 (1)°
VAC14	Platypus		3:22909412-23013323 (1)°
si:ch211-245p7.3	Zebrafish		-
Novel gene	Anole lizard, C. savignyi (5), Cat, Dog, Fugu, Marmoset, Medaka, Platypus, Rat, Turkey, Xenopus, Zebra finch (2)		-
CCDC88C		14:91737667-91884188 (-1)*	24:34496196-34625346 (-1)*
FANCB		X:14861529-14891191 (-1)*	X:10380764-10398001 (-1)*
HYDIN		16:70841281-71264625 (-1)*	3:22599664-22895359 (1)°
L1CAM		X:153126969-153174677 (-1)*	16:15165246-15247062 (-1)* X:122310356-122321631 (-1)*
MPDZ		9:13105703-13279589 (-1)*	3:78695744-78785664 (1)° 23:33398081-33526976 (-1)*
PIK3R2		19:18263626-18288927 (1)°	21:2863966-2872709 (1)° 1:40515961-40601946 (-1)* 2:33626216-33640340 (-1)*
PTEN		10:89622870-89731687 (1)°	17
ZIC3	Human chromosome	X:136648301-136659850 (1)°	22 X:109456759-109462626 (1)°

*-1 is reverse strand.
°1 is forward strand.

3.5.1. Identification of genes from Genome-wide association study

Based on the result of GWAS, Figure 11 shows 228 genes that were found in region from BIEC2-23808 (57,760,860 bp) to BIEC2-37017 (87,079,561 bp) in horse chromosome 1. 105 genes with forward strand and 123 genes with reverse strand. The 228 genes were compared with identified genes of hydrocepahlus in human and other species. Unfortunately, the genes in the most significantly associated region are not the same to any already identified hydrocephalus genes in other species.

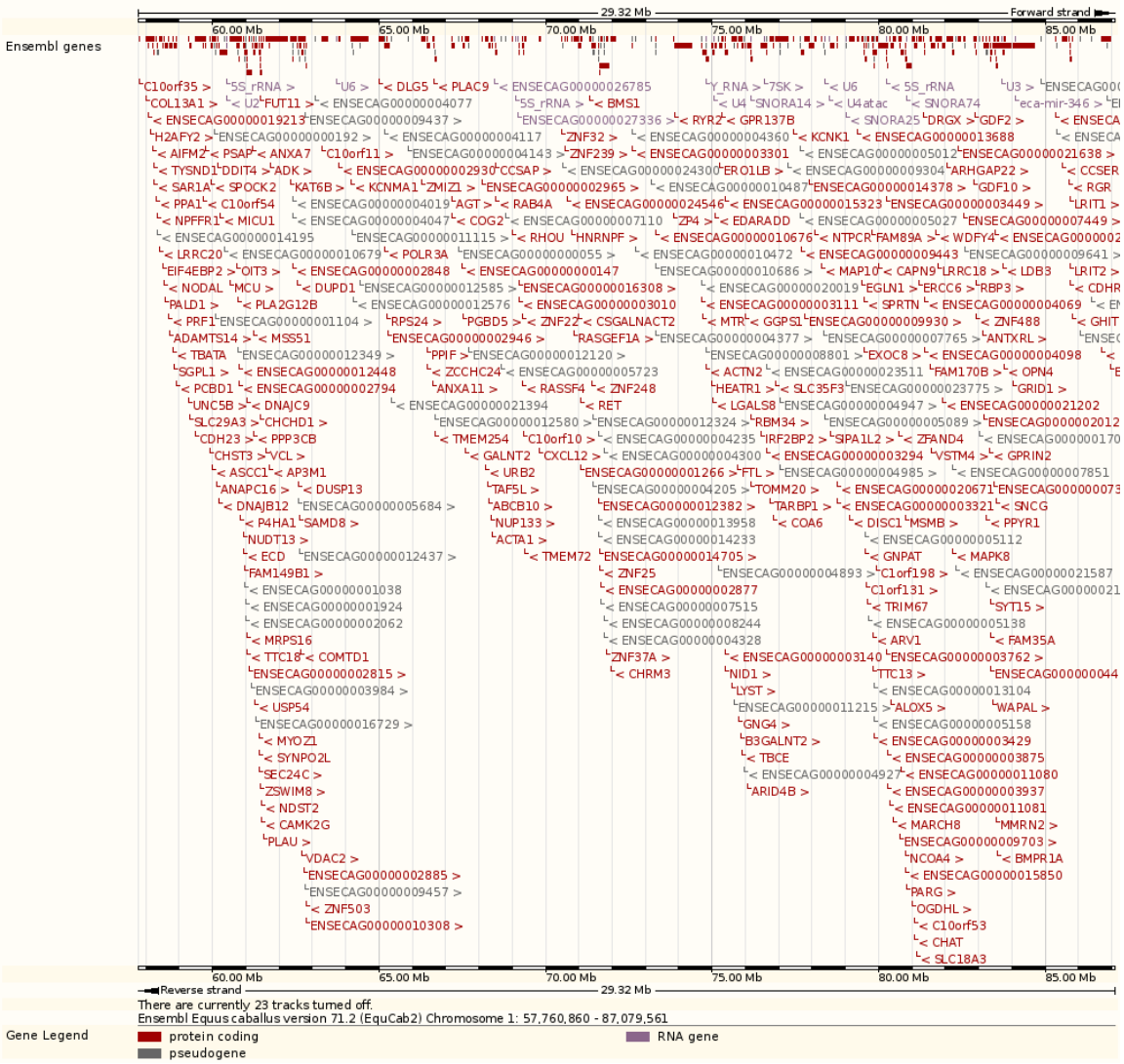

Figure 11. Genes in region BIEC2-23808–BIEC2-37017 (29.3 Mb) with most significant associated SNPs.

Figure 12 shows syntenies of horse chromosme 1 with human chromosomes 1, 10, 14, 15, and 17; and dog chromosomes 3, 4, 8, 13, 15, 26, 28, and 30. The red bar on horse chromosome 1 is the region that was identified to be associated with hydrocepahlus in Friesian horses. The synteny of this region was found on human chromosome 10 and dog chromosome 4.

In human synteny, in total there were 245 genes found. 33 novel genes were found to have no homology with human chromosomes and 14 genes were found only in human chromosomes. It might be that these genes have not been identified yet in horse chromosome 1. In dog synteny, 233 genes were found. 43 novel genes were found to have no homology with dog chromosomes and 2 genes were found only in dog chromosomes. Unfortunately, these genes are also not the same to any identify genes of hydrocephalus in other species.

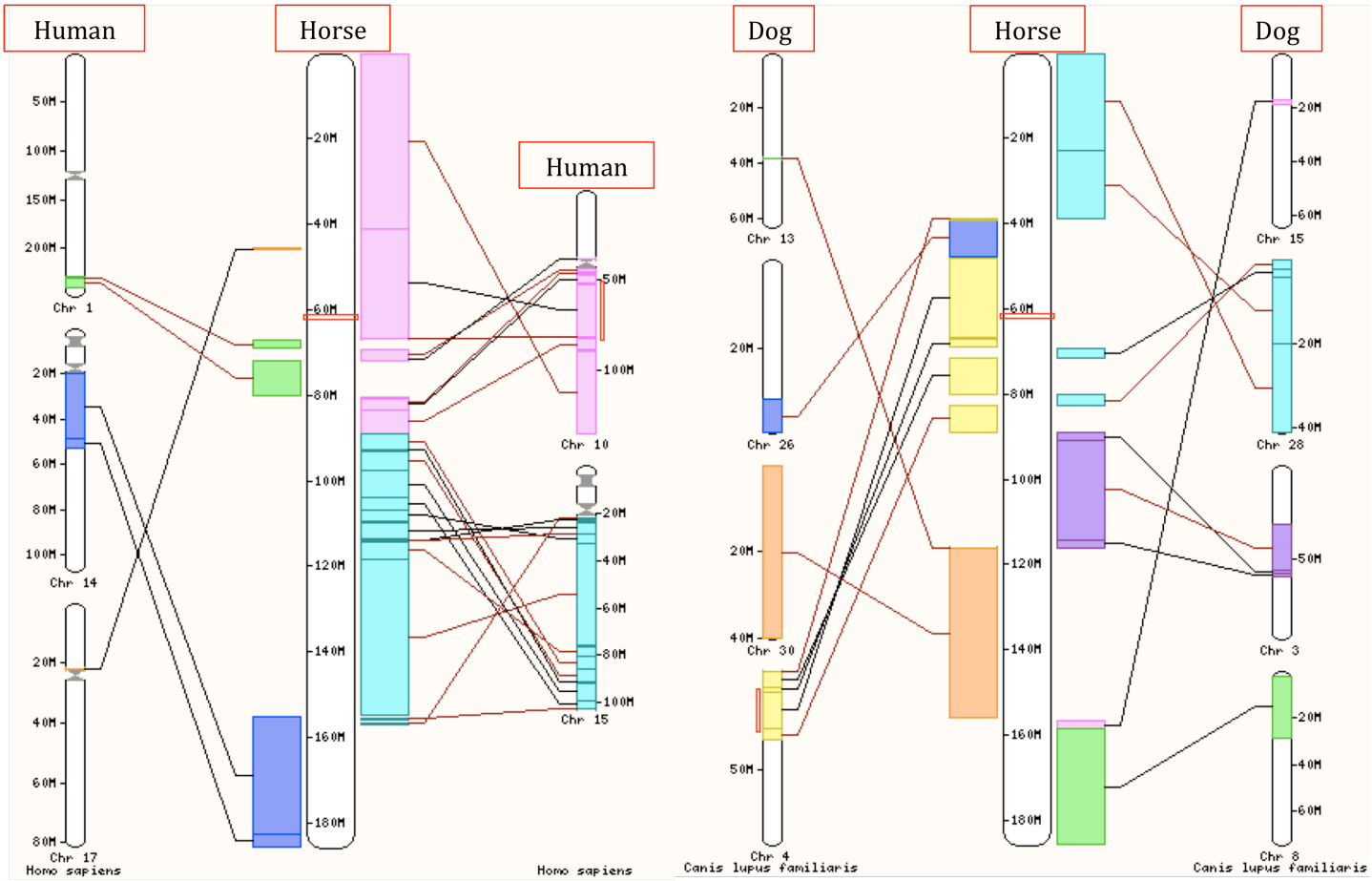

Figure 12. Synteny of horse chromosome 1 in human and dog chromosomes.

3.5.2. Haplotype block

Based on haplotype block analysis, candidate genes (Table 8) were found in 12 associated haplotype blocks (lowest *P-value*) with hydrocephalus in Friesian horses, except haplotype block 122, 129, and 142. Candidate genes that have 'ENSECAG' symbol are novel genes (uncharacterized protein). Unfortunately, these candidate genes are not related to any genes of hydrocephalus that already identified in other species.

Syntenies of all candidate genes (Table 8) that were found within associated haplotype blocks are human chromosomes 1 and 10; and dog chromosomes 4 and 28, except some novel genes on haplotype blocks 130, 147, and 148 were found to have no homology in human and dog chromosomes (only found in horse chromsome 1). Unfortunately, these novel genes are not genes previously identified to be causing hydrocephalus in other species.

Table 8. Associated haplotype blocks with associated SNPs, including candidate genes, and synteny with human and dog

Block	Size (kb)	Distance (kb)	Associated SNPs	Genes	Synteny
112	200.055		BIEC2-25357	FUT11, CHCHD1, ZSWIM8, NDST2, CAMK2G, PLAU, VCL	Human (ch.10), Dog (ch.4)
122	0.226	4205.417	BIEC2-27351, BIEC2-27352	-	-
124	62.324	454.493	BIEC2-27588	ZMIZ1	Human (ch.10), Dog (ch.4)
129	180.682	2238.622	BIEC2-28507, BIEC2-28464, BIEC2-28506, BIEC2-28459, BIEC2-28302	-	-
130	490.599	57.507	BIEC2-28874, BIEC2-28749	RHOU, ENSECAG00000016308, ENSECAG00000003010, ZNF22, RASSF4	Human (ch.1 & 10), Dog (ch.4 & 28), no homologous for ENSECAG00000003010
131	397.462	12.834	BIEC2-28955, BIEC2-28908, BIEC2-28912, BIEC2-28931, BIEC2-28875, BIEC2-28907, BIEC2-28876	RASSF4, TMEM72	Human (ch.10), Dog (ch.28)
142	157.989	3680.797	BIEC2-31328, BIEC2-31300	-	-
147	114.018	1834.572	BIEC2-32205, BIEC2-32202	GPR137B, ENSECAG00000015323, ENSECAG00000003140, NID1	Human (ch.1), Dog (ch.4), no homologous
148	455.488	984.212	BIEC2-32912, BIEC2-32706	IRF2BP2, ENSECAG00000003294, TARBP1, COA6, SLC35F3	Human (ch.1), Dog (ch.4), no homologous
149	215.696	256.379	BIEC2-32970	SLC35F3	Human (ch.1), Dog (ch.4)

CHAPTER 4. DISCUSSION

Genomic associations and genes associated with hydrocephalus in Friesian horses have not been identified before. Therefore, the aims of this study were to identify SNPs and haplotype blocks associated with hydrocephalus in Friesian horses using genome-wide association study (based on allele and genotype frequency differences between cases and controls) and haplotype association study, and also to identify homozygous chromosomal segments in affected Friesian horses.

4.1. Population structure

In this study, data on hydrocephalus in Friesian horses was checked for population structure. The classical multidimensional scaling plot (Figure 2) showed some overlap because 44.8% of the samples were half-sibs horses descending from 9 sires with both cases and controls among their offspring (Table 2). However, the cause of population stratification seems to be present between cases and controls because some cases and controls coming from different sires. The case-control of this study is bad when the samples used has subgroups of individuals that cases are on average more related to each other than controls in a population. On the other hand, it is better when cases and controls are on average more related to each other because the association could be due to a disease associated locus instead of the population structure.

Genomic control method was used in this study to correct for the population structure by correction *P-values* for the estimated inflation factor. The inflation based on allele frequency differences was corrected in this study. However, correction for the inflation was not possible because there is no genomic control possible for *P-values* based on genotype frequency differences, it only possible for allele frequency differences. Therefore, false associations between SNPs and hydrocephalus could be detected in genotype frequency analysis because its *P-values* were inflated.

4.2. Genome-wide association study

Genome-wide association study (GWAS) of hydroceohalus in Friesian horses identified a region on chromosome 1 containing 29 significantly (*P-value* < 0.05 with the correction for stratification) associated SNPs based on allele freqency differences on region between 61,311,028 and 77,220,026 bp (15.91 Mb) and 68 significantly (*P-value* < 0.05 without the correction for stratification) associated SNPs based on genotype frequency differences on region between 57,760,860-87,079,561 bp (29.3 Mb). Moreover, these 68 associated SNPs based on genotype frequency differences included the 29 associated SNPs based on allele frequency differences between cases and controls.

Homozygous chromosomal segments in affected Friesian horses were identified using individual genotypesof cases and controls. Based on the most 30 significantly associated SNPs, 15-35% of unaffected Friesian horses have genotypes of unfavorable alleles and 50% of affected Friesian horses have homozygous genotypes for unfavorable alleles. Moreover, 85-90% of affected Friesian horses were homozygous for these following SNPs: BIEC2-27351, BIEC2-27352, BIEC2-27588, BIEC2-28874, BIEC2-28875, BIEC2-28876, BIEC2-29359, and BIEC2-31514. These unvaforable alleles are expected to be alleles associated with hydrocephalus. To confirm this finding and also to narrow down the region, autozygosity mapping can be performed in the further study.

4.3. Haplotype association study

Haplotype association study of hydroceohalus in Friesian horses identified 54 significantly (*P-value* < 0.05) associated haplotype blocks on chromosome 1. Furthermore, these 54 associated haplotype blocks included the 68 associated SNPs based on genotype frequency differences between cases and controls. To confirm this finding and also to narrow down the region, resequencing of 54 associated haplotype blocks can be performed and the identification of the region associated with hydrocephalus in other horse breeds could help.

4.4. Linkage disequilibrium

All associated haplotype blocks were constructed based on D prime in Haploview software, which indicated high LD. However, the associated haplotype blocks also contained many SNPs marker that were not individually associated to hydrocepahlus. Moreover, Haploview software also included a pair SNPs that have strong hostorical recombination. So, it might gives false-positive associations with hydrocepahlus. Therefore, other software packages can be considered to perform haplotype reconstruction in the further study. These softwares are SNPHAP, PHASE, and FASTPHASE (Balding, 2006).

4.5. Candidate genes

The size of the region of all associated SNPs on chromosome 1 is 29.3 Mb (57,760,860-87,079,561 bp). 228 genes were found in this region that included 33 and 43 novel genes that had no homology with human chromosomes and dog chromosomes. There is a possibility that one of these novel genes is responsible for hydrocephalus. However, these 228 genes are not the same as identify genes of hydrocephalus in human and other species. It might be that there is no one gene that causes hydrocephalus in all species, but several genes in many species or different genes in different species.

Hydrocephalus in human is caused by several genes mutation, and these mutated genes showed different phenotypes. Identified genes of hydrocephalus in human are *CCDC88C* (coiled-coil domain containing 88C), *FANCB* (fanconi anemia, complementation group B), *HYDIN* (axonemal central pair apparatus protein), *L1CAM* (L1 cell adhesion molecule), *MPDZ* (multiple PDZ domain protein), *PIK3R2* (phosphoinositide-3-kinase, regulatory subunit 2 (beta)), *PTEN* (phosphatase and tensin homolog), *ZIC3* (zic family member 3), etc.

The most common gene that causes hydrocephalus in human (4.5-14.2%) is *L1CAM* gene on chromosome Xq28 (153,126,969-153,174,677 bp) with a recessive inheritance mode (Haverkamp *et al.* 1999; Schrandel-Stumpel and Vos, 2004). L1 cell adhesin molecule has an important role in nervous system

development, including neuronal migration and differentiation. The cause of *L1CAM* gene mutation can be due to congenital stenosis aqueduct of sylvius (HSAS) or linkage of HSAS and MASA syndrome (mental retardation, aphasia, shuffling gait, and adducted thumbs) with phenotypes of enlarged cerebral ventricles and mental retardation, and often includes spastic paraparesis and adducted thumbs (Schrander-Stumpel *et al.*, 1990; Rosenthal *et al.*, 1992). In the horse, *L1CAM* gene is located on chromosome 16 (15,165,246-15,247,062 bp) and X (122,310,356-122,321,631 bp).

On the other hand, hydrocephalus that caused by mutation of *HYDIN* gene is found not only in human but also in other species i.e. chimpanzee, ferret, macaque, mouse, panda, and rat. However, this gene was not found in the region of this study. *HYDIN* gene is located on chromosome 16q22.2 with recessive mode of inheritance in human (Davy and Robinson, 2003; Olbrich *et al*., 2012). It has gene phenotype relationship with primary ciliary dyskinesia-5 without situs inversus (CILD5), which identified a homozygous splice site mutation in the *HYDIN* gene (Olbrich *et al*., 2012). *HYDIN* gene has phenotype of chronic wet cough, recurrent bronchitis, pneumonia, otitis media, chronic rhinosinusitis, and bronchiectasis in human (Olbrich *et al*., 2012). In horse, *HYDIN* gene is located on chromosome 3 in the region 22,599,664-22,895,359 bp.

4.6. Phenotype

In this study, the phenotype of cases and controls were examined by the Faculty of Veterinary Medicine at Utrecht University, the Netherlands. So, the cases were expected to have similar phenotype within cases, and also the controls have similar phenotype within controls. Vetrenarians probably defined phenotype characteristics based on the literature study of Sipma *et al*. (2011) as they werethe only and the first study that characterized phenotype of Friesian horses with hydrocephalus.

Sipma *et al*. (2011) analyzed phenotypic characteristics of hydrocephalus in four stillborn Friesian foals with hydrocephalus and compared them with two

control stillborn Friesian foals without hydrocephalus. They were examined macroscopically and microscopically. Macroscopical examinations that were used in their study were post-mortem examination, latex perfusion of cranial blood vessels and computed tomography. For microscopical examination, histological examination was performed on brain tissue. They found malformation of the petrosal bone and the jugular foramen based on the result of computed tomography, and also dilated blood vessels on the brain cortex based on microscopic examination. A combination of these results indicated hydrocephalus in Friesian horses, which causes disruption of the absorption of cerebrospinal fluid (CSF) in the systemic circulation at the venous sinuses. It might be due to the damaged blood drainage from the cerebral venous to the jugular vein, which is caused by an abnormal narrowing of the passage of the jugular foramen. However, Sipma *et al*. (2011) only performed their analysis based on four cases. So, it cannot be proven that this is the cause of hydrocephalus in Friesian horses.

Based on other studies in human, there were 322 hits that have different phenotypes of hydrocepahlus, which included the combination of hydrocephalus with other disorders. Therefore for further study, it is better to characterize more hydrocephalus cases to get more certainty abouth the phenotypes of hydrocephalus in Friesian horses, to prevent false-positive associations in finding genetic background of hydrocephalus.

4.7. DNA-test

This study is the first step to increase our understanding of genetic background of hydrocepahlus in Friesian horses before a DNA-test can be developed. The steps that need to be taken to develop a DNA-test for hydrocephalus are shown in Figure 13.

a. Develop the first DNA-test based on haplotype reconstrction and autozygosity mapping

Haplotype reconstruction and autozygosity mapping steps were chosen because based on the result of this study, it was found that the genomic region of hydrocephalus in the Friesian horse breed is relatively huge in length (29.3 Mb)

due to high inbreeding rate. Therefore, narrowing down the region is necessary to be done. These two steps will be used to narrow down the length of genomic regions.

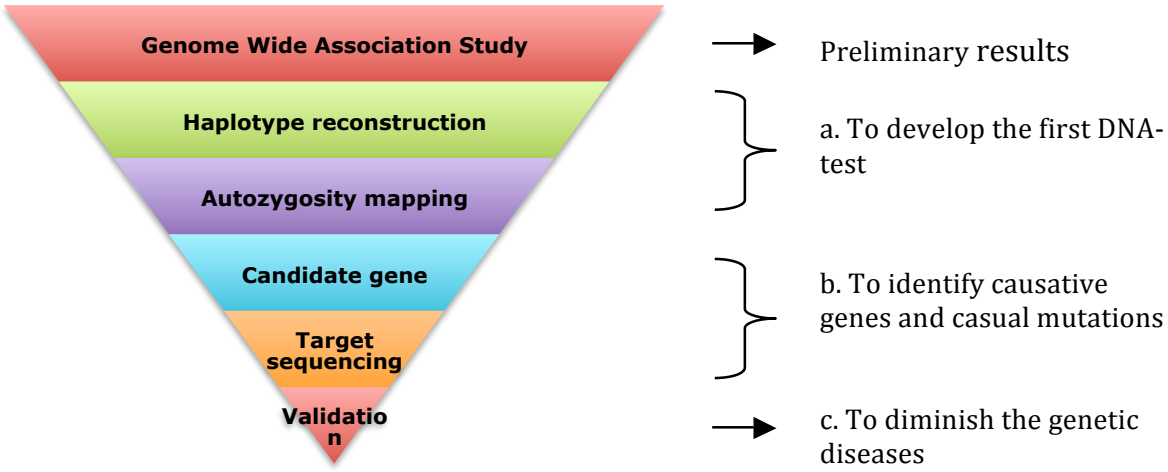

Figure 13. Steps overview of development of a DNA-test for hydrocephalus in the Friesian horse breed.

- Haplotype reconstruction

Based on the result of this study, genomic region of hydrocephalus was found on chromosome 1:57,760,860-87,079,561. However, many SNPs that were not associated with hydrocephalus were found within the associated haplotype blocks during analysis and it is difficult to detect genetic effects. Therefore, to make the detection in genetic associations more powerful, haplotype block needs to be reconstruct and these finding need to be compared with results from genetic analysis of hydrocephalus in other horse breeds that also suffer from hydrocepahlus. Several software packages that can be considered to perform haplotype reconstruction are SNPHAP, PHASE, and FASTPHASE (Balding, 2006).

First techniques to narrow down the length of genomic region in haplotype reconstruction is long-range PCR, followed by sequencing using Next Generation Sequencing (NGS) (additional option is EquineSNP777 array). Afterwards, the data will be analysed with haplotype reconstruction software. Long-range polymerase chain reaction (PCR) can amplify DNA fragments that excess of 5 kb, and it relies on overlapping PCR to generate a constant band (QIAGEN, 2010; Ozcelik *et al.*, 2012). NGS has high accuracy to detect SNPs at low coverage and a low false positive rate (Wang *et al.*, 2011). In the mean while, if

the Equine SNP777 array will be released during the further study, it is better to use this assay because it is more specific for horse genome.

- Autozygosity mapping

After haplotypes have been reconstructed, continue with autozygosity mapping by finding homologous regions shared by all affected horses that are homozygous for the unfavorable allele and heterozygous in unaffected horses. In this study, homozygous chromosomal segments were already identified by looking at individual genotypes. However, only 85-90% of affected Friesian horses shared homozygous regions for the unfavorable allele. Therefore, autozygosity mapping needs to be performed in the further study because the difference between what is performed in this study and autozygosity mapping is the identification of the strength of homozygosity of each chromosomal segment between cases and controls. Autozygosity mapping relies on the recognition of regions IBD that are homozygous in affected individuals by plotting a cumulative LOD score (statistical test for linkage analysis) together with overlaid homozygosity blocks by considering the allele frequencies for all SNPs. The software that can be used is AutoSNPa (Carr *et al*., 2006).

Haplotype reconstruction and autozygosity mapping followed by long-range PCR and sequencing is already good enough as the first development of DNA-test because it can be used to estimate the frequency of hydrocephalus on a small sample of the Friesian horse population. Furthermore, the frequency of hydrocephalus can be monitored and a better selection program can be designed. However, in scientific perspective, to identify candidate genes and casual mutations is more preferable because it will give good molecular basis information and more reliable DNA-tests.

b. Identify candidate gene and casual mutation

Furthermore after the first DNA-test is developed, wherein genomic regions become more specific, candidate gene approach and targeted sequencing can be performed to find a causative gene and a mutation within blocks.

- Candidate gene approach

Identification of candidate genes of hydrocephalus in human and other species already performed in this study using online databases, Online Mendelian Inheritance in Man or Animals (OMIM/OMIA) and Ensembl. However, all identify genes of hydrocephalus in human and other species were not identified in the region that was found in this study.

After the region is narrow down, candidate gene approach can be used to select candidate genes in genomic regions that are highly associated with hydrocephalus. The gene can be select based on its function and the relation with the phenotype of affected Friesian horses.

- Targeted sequencing

After a closer look in more specific regions of interest within the genome, targeted sequencing can be performed. The use of targeted sequencing is to accurately identify SNPs and distinguish them from false positives. PCR and sequencing techniques will be used in this step.

c. Validation

The next development of DNA-tests will also be performed through more reliable genetic markers (SNPs). The validation can be performed by carry out a DNA-test to the parents of affected individuals, affected individuals, and controls not previously included in our study.

4.8. Implementation of DNA-test in breeding program

A DNA-test can be implemented in selection to diminish hydrocephalus from the Friesian horse population, and a better breeding program can be designed. The DNA-test can determine whether an individual horse is at risk of becoming affected while carrying two copies of the unfavorable allele or at no risk of becoming affected while carrying the favorable alleles.

Breeding strategy that can be considered to reduce or remove hydrocephalus in the Friesian horse population is shown in Figure 14. The goals of this breeding strategy are to gradually reduce hydrocephalus in the population and to gradually increase genetic diversity.

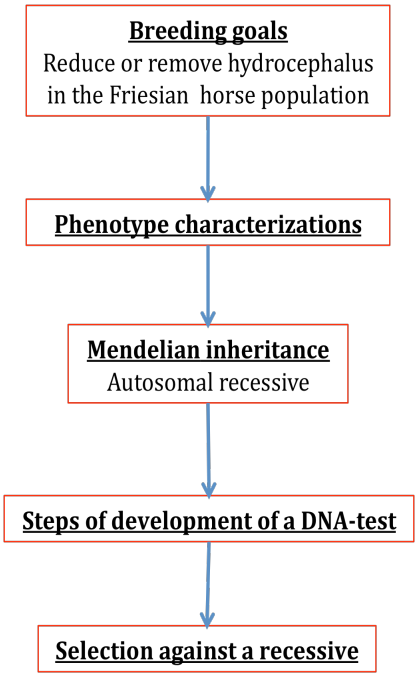

Figure 14. Breeding strategy of the Friesian horse breed.

The selection type for this breeding program is selection against a recessive. This DNA-test can be used to identify the carrier. Three suggestions for breeding decisions depending on the ratio of homozygous healthy horses to heterozygous horses (that is, carriers) that likely to be produced: a) first suggestion is mating a homozygous healthy horse to a homozygous healthy horse, b) second suggestion is mating a homozygous healthy horse to a heterozygous horse (carrier), c) the last suggestion is mating a heterozygous horse (carrier) to a heterozygous horse (carrier). However, the breeder needs to prevent mating heterozygous with heterozygous Friesian horses because they will have affected offspring with the segregation frequency of 0.25. To eliminate completely recessive allele in the population is by mating healthy homozygous (favorable alleles) with healthy homozygous (favorable alleles). But it is not necessary to remove recessive (unfavorable) allele completely, as long as both

heterozygous are never mated together, so the heterozygous still can be used for mating. In this way, the recessive allele will remain in the population but the disease will not occurs because homozygous recessive (unfavorable allele) never occur.

Moreover, there are many other aspects that also need to consider i.e. genetic drift, inbreeding, and loss in genetic variation, etc. The Friesian horse breed faced a genetic bottleneck with small effective population size in the beginning of the establishment of Friesian horse studbook. There are some undesirable effects in the current population with low number of individual, which are carrying undesirable genes/alleles. Loss in genetic variation also needs to be considered because by eliminating recessive gene completely in the population that gene will be lost forever and the selected gene will become fixed in the population.

CHAPTER 5. CONCLUSIONS

This study is used as the first step for our understanding of genetic background of hydrocephalus in Friesian horses.

Hydrocephalus mutation in Friesian horses was found on chromosome 1:57,760,860-87,079,561 (29.3 Mb region). 68 associated SNPs and 54 associated haplotype blocks were identified in this region. Homozygosity of 8 SNPs were shared by more than 85% in affected Friesian horses. These 8 SNPs were homozygous favorable alleles and heterozygous in 85-90% of unaffected Friesian horses.

Two hundred and twenty eight genes were found in this region. 33 and 43 novel genes had no homology with human chromosomes and dog chromosomes. However, none of the genes were similiar to candidate genes that have been found already in other species.

This study gives suggestions for further study, which can be used to develop a DNA-test for hydrocephalus in the Friesian horse breed. As the first step is to narrow down the region that has been found in this study. And continue with finding the causal mutation of hydrocepahlus in Friesian horses. Furthermore, when a causal gene(s) has been identified, a DNA-test can be used to identify carriers of hydrocephalus within the Friesian horse population. It will be used to either remove these carriers from breeding and/or to give recommendation to breeding strategies that can prevent this genetic defect to appear in the next generation.

REFERENCES

Aulchenko, Y.S., Ripke, S., Isaacs, A., and van Duijen C.M. 2007. GenABEL: an R library for genome-wide association analysis. Bioinformatics, 23(10): 1294-1296.

Balding, D.J. 2006. A tutorial on statistical methods for population association studies. Nature Genet. 7:781-791.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics, 21(2): 263-265.

Boerma, S., Back, W., and van Oldruitenborgh-Oosterbaan, M.M.S. 2012. Clinical Commentary, The Friesian horse breed: A clinical challange to the equine veterinarian? Equine Veterinary Education, 24(2):66-71.

Bowman, R. 1980. Congenital hydrocephalus in 2 foals. Mod. Vet. Pract., 61:862-864.

Carr, I.M., Flintoff, K., Taylor, G.R., Markham, A,F., and Bonthron, D.T. 2006. Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families. Human Mutation, 27:1041-6.

Davy, B. E., and Robinson, M. L. 2003. Congenital hydrocephalus in hy3 mice is caused by a frameshift mutation in Hydin, a large novel gene. Hum. Molec. Genet. 12:1163-1170.

Ducro, B.J., Bovenhuis, H., Oliehoek, P.A., and van Arendonk, J.A.M. 2011. Chapter 6: Genetic diversity in the Friesian Horse population using pedigree analysis. PhD thesis, Wageningen University, Wageningen, the Netherlands.

FAO. 1998. Secondary Guidelines for Development of National Farm Animal genetic Resources Management Plans: Management of Small Populations at Risk. FAO: Rome.

Fernell, E., Hagberg, G., and Hagberg, B. 1994. Infantile hydrocephalus epidemiology: an indicator of enhanced survival. Arch Dis Child Fetal Neonatal, 70:F123-128.

Ferris, R.A., Sonnis, J., Webb, B., Lindholm, A., and Hassel, D. 2010. Hydrocephalus in an American Miniature Horse Foal: A case report and review. Journal of Equine Veterinary Science, 31(11):611-614.

Frackowiak, H., Botko, P., and Stanislwaski, D. 2006. A case of hydrocephalus in the Przewalski's wild horse.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J., and Altshuler, D. The structure of haplotype blocks in the human genome. Science, 296 (5576): 2225-2229.

Gower, J.C. 1996. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika, 53:325-338.

Haverkamp, F., Wolfle, J., Aretz, M., Kramer, A., Hohmann, B., Fahnenstich, H., and Zerres, K. 1999. Congenital hydrocephalus internus and aqueduct stenosis: aetiology and implications for genetic counselling. Eur J Pediat, 158:474-478.

Hoppe-Hirsch, E., Laroussinie, F., Brunet, L., Sainte-Rose, C., Renier, D., Cinalli, G., Zerah, M., and Pierre-Kahn, A. 1998. Late outcome of the surgical treatment of hydrocephalus. Childs Nerv Syst ,14:97-99.

Johnston, I. and Teo, C. 2000. Disorders of CSF hydrodynamics. Childs Nerv Syst 16, 776-799.

Ojala, M. and Ala-Huikku, J. 1992. Inheritance of hydrocephalus in horses. Equine vet. J. 24(2):140-143.

Olbrich, H., Schmidts, M., Werner, C., Onoufriadis, A., Loges, N. T., Raidt, J., Banki, N. F., Shoemark, A., Burgoyne, T., Al Turki, S., Hurles, M. E., UK10K Consortium, and 10 others. 2012. Recessive HYDIN mutations cause primary ciliary dyskinesia without randomization of left-right body asymmetry. Am. J. Hum. Genet. 91:672-684.

Online Mendelian Inheritance in Animal, OMIA. Frank Nicholas Faculty of Veterinary Science, The University of Sydney and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 19-05-2013. World Wide Web URL: http://www.ncbi.nlm.nih.gov/omia/

Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 19-05-2013. World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/

Ozcelik, H., Shi, X., Chang, M.C., Tram, E., Vlasschaert, M., Di Nicola, N., Kiselova, A., Yee, D., Goldman, A., Dowar, M., Sukhu, B., Kandel, R., and Siminovitch, K. 2012. Long-Range PCR and Next-Generation Sequencing of BRCA1 and BRCA2 in Breast Cancer. The Journal of Molecular Diagnostics, Vol. 14(5):467-475.

Pearson, T.A. and Manolio, T.A. 2008. How to interpret a genome-wide association study. JAMA, 299(11):1335-1344.

QIAGEN. 2010. Long range PCR handbook, for reliable and accurate long-range PCR up t0 40 kb. http://www.qiagen.com/products/qiagenlongrangepcrkit.aspx#Tabs=t1

Rekate, H.L. 2008. The definition and classification of hydrocephalus: a personal recommendation to stimulate debate. Cerebrospinal Fluid Res 5, 2.

Rosenthal, A., Joulet, M., and Kenwrick, S. 1992. Aberrant splicing of neural cell adhesion molecule L1 mRNA in a family with X-linked hydrocephalus. Nature Genet. 2:107-112.

Schrander-Stumpel, C., Legius, E., Fryns, J. P., and Cassiman, J. J. 1990. MASA syndrome: new clinical features and linkage analysis using DNA probes. J. Med. Genet. 27:688-692.

Schrander-Stumpel, C.T. and Vos, Y.J. 2004. From gene to disease; X-linked hydrocephalus and L1CAM. Ned Tijdschr Geneeskd, 148:1441-1443

Schurr, P.H., McLaurin, R.L., and Ingraham, F.D. 1953. Experimental studies on the circulation of the cerebrospinal fluid and methods of producing communicating hydrocephalus in the dog. J Neurosurg, 10:515-525.

Sevinga, M., Vrijenhoek, T., Hesselink, J.W., Barkema, H.W. , and Groen, A.F. 2004. Effect of inbreeding on the incidence of retained placenta in Friesian horses. Journal of Animal Science, 82:851-859.

Sipma, K.D., Cornillie, P., Saulez, M.N., Stout, T., Voorhout, G., and Back., W. 2011. Phenotypic characteristics of hydrocephalus in Friesian stillborn foals. 12th congress of the world equine veterinary association, 2-5 November 2011, Hyderabad, India.

Strachan, T. and Read, A.P. 2005. Human molecular genetics, 2nd edition. New York: Wiley-Liss.

Wang, W., Wei, Z., Lam, T.W., and Wang, J. 2011. Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. Scientific Report, 1:55.

Yoon, C.H. and Slaney, J. 1972. Hydrocephalus : A new mutation in the Syrian golden hamster. J Hered, 63(6):344-346.

APPENDIX

Table 9. 68 SNPs significantly (*P-value* < 0.05) associated with hydrocephalus in Friesian horses based on genotype frequency after permutation testing

					Genotype count		
SNP name	Position (bp)	N	P-value	OR*	All	Controls	Cases
BIEC2-23808	57,760,860	67	3.84E-02	0.20	30 28 9	GG AG AA 12 27 8	18 1 1
BIEC2-23811	57,764,046	67	3.84E-02	0.20	30 28 9	GG AG AA 12 27 8	18 1 1
BIEC2-23813	57,764,065	67	3.84E-02	0.20	30 28 9	AA AG GG 12 27 8	18 1 1
BIEC2-25335	61,311,028	67	4.90E-05	2.06	18 23 26	AA AG GG 3 19 25	15 4 1
BIEC2-25357	61,479,005	67	4.90E-05	2.06	18 23 26	AA AC CC 3 19 25	15 4 1
BIEC2-25551	62,260,532	67	2.42E-02	1.61	13 24 30	AA AG GG 2 18 27	11 6 3
BIEC2-25825	62,876,363	67	3.63E-02	0.30	30 24 13	AA AG GG 12 22 13	18 2 0
BIEC2-26016	63,415,184	66	3.05E-02	1.65	13 23 30	CC AC AA 2 17 27	11 6 3
BIEC2-26386	63,975,987	67	2.42E-02	1.61	13 24 30	GG AG AA 2 18 27	11 6 3
BIEC2-27143	65,553,951	67	9.01E-04	1.35	13 26 28	GG AG AA 1 21 25	12 5 3
BIEC2-27280	65,732,134	67	1.59E-02	1.93	19 27 21	AA AG GG 5 22 20	14 5 1
BIEC2-27301	65,769,727	67	1.59E-02	1.93	19 27 21	AA AC CC 5 22 20	14 5 1
BIEC2-27336	65,814,777	67	9.01E-04	1.35	13 26 28	AA AG GG 1 21 25	12 5 3
BIEC2-27351	65,841,804	67	1.00E-06	0.74	20 23 24	AA AG GG 3 22 22	17 1 2
BIEC2-27352	65,842,030	67	1.00E-06	0.74	20 23 24	AA AG GG 3 22 22	17 1 2
BIEC2-27399	65,922,084	67	4.18E-02	0.94	17 30 20	CC AC AA 4 26 17	13 4 3
BIEC2-27588	66,296,523	66	1.00E-06	54.60	20 20 26	AA AG GG 3 18 26	17 2 0
BIEC2-27590	66,302,733	67	2.18E-03	0.27	22 25 20	CC AC AA 6 22 19	16 3 1
BIEC2-27609	66,346,442	67	5.63E-03	0.21	31 23 13	AA AC CC 12 22 13	19 1 0
BIEC2-27617	66,353,633	67	5.63E-03	0.21	31 23 13	GG AG AA 12 22 13	19 1 0
BIEC2-27626	66,358,865	67	5.63E-03	0.21	31 23 13	AA AG GG 12 22 13	19 1 0
BIEC2-28302	68,597,487	67	1.00E-06	2.16		GG AG AA	

					13 19 35	0 14 33 AA AC CC	13 5 2
BIEC2-28459	68,701,400	67	1.00E-06	2.16	13 19 35	0 14 33 CC AC AA	13 5 2
BIEC2-28464	68,708,147	67	1.00E-06	0.11	22 26 19	3 25 19 AA AG GG	19 1 0
BIEC2-28506	68,778,089	67	1.00E-06	2.16	13 19 35	0 14 33 AA AG GG	13 5 2
BIEC2-28507	68,778,169	67	1.00E-06	1.98	13 21 33	0 16 31 AA AG GG	13 5 2
BIEC2-28749	69,062,965	66	1.00E-06	2.13	13 19 34	0 14 32 CC AC AA	13 5 2
BIEC2-28874	69,326,275	67	1.00E-06	1.41	20 23 24	3 21 23 AA AC CC	17 2 1
BIEC2-28875	69,339,109	67	1.00E-06	1.55	19 21 27	2 19 26 AA AG GG	17 2 1
BIEC2-28876	69,339,208	67	1.00E-06	1.35	20 24 23	3 22 22 AA AG GG	17 2 1
BIEC2-28888	69,407,277	67	1.00E-06	0.13	24 32 11	5 31 11 AA AG GG	19 1 0
BIEC2-28892	69,407,512	67	1.00E-06	0.13	24 32 11	5 31 11 AA AG GG	19 1 0
BIEC2-28898	69,410,844	67	1.00E-06	0.13	24 32 11	5 31 11 GG AG AA	19 1 0
BIEC2-28907	69,505,589	67	1.00E-06	2.07	34 20 13	0 15 32 GG AG AA	13 5 2
BIEC2-28908	69,543,750	66	1.00E-06	2.23	13 18 35	0 13 33 GG AG AA	13 5 2
BIEC2-28912	69,561,176	67	1.00E-06	2.16	13 19 35	0 14 33 AA AG GG	13 5 2
BIEC2-28931	69,697,043	67	1.00E-06	2.16	13 19 35	0 14 33 AA AG GG	13 5 2
BIEC2-28955	69,736,571	67	1.00E-06	2.16	13 19 35	0 14 33 AA AG GG	13 5 2
BIEC2-29359	70,378,596	67	1.00E-06	1.20	20 27 20	3 25 19 AA AC CC	17 2 1
BIEC2-29471	70,661,243	67	1.00E-02	0.28	23 28 16	7 25 15 GG AG AA	16 3 1
BIEC2-29857	71,344,522	67	4.00E-06	0.17	21 30 16	4 28 15 AA AC CC	17 2 1
BIEC2-29865	71,348,106	67	4.00E-06	0.17	21 30 16	4 28 15 AA AG GG	17 2 1
BIEC2-31300	73,545,254	67	4.40E-05	1.80	14 23 30	1 18 28 AA AC CC	13 5 2
BIEC2-31328	73,575,357	67	4.40E-05	1.80	14 23 30	1 18 28 AA AG GG	13 5 2
BIEC2-31514	73,881,378	66	1.50E-05	0.18	21 28 17	4 26 16 GG AG AA	17 2 1
BIEC2-31580	73,992,006	67	4.10E-05	1.88	14 22 31	1 17 29 GG AG AA	13 5 2
BIEC2-32009	74,880,142	67	1.29E-03	1.15	17 30 20	3 18 26 GG AG AA	14 4 2
BIEC2-32016	74,897,083	67	6.37E-03	1.22	18 28 21	4 24 19	14 4 2

BIEC2-32043	74,938,827	67	4.00E-06	0.17	21 29 17	AA AG GG 4 27 16	17 2 1
BIEC2-32168	75,308,260	67	1.00E-06	1.02	16 39 21	CC AC AA 1 27 19	15 3 2
BIEC2-32202	75,458,447	67	1.00E-05	1.51	15 24 28	GG AG AA 1 20 26	14 4 2
BIEC2-32205	75,461,775	67	1.00E-05	1.51	15 24 28	GG AG AA 1 20 26	14 4 2
BIEC2-32520	76,154,827	67	1.00E-05	1.51	15 24 28	CC AC AA 1 20 26	14 4 2
BIEC2-32546	76,229,305	67	1.41E-02	0.27	26 24 17	AA AG GG 9 22 16	17 2 1
BIEC2-32688	76,508,159	67	2.72E-02	0.31	24 25 18	AA AC CC 8 22 17	16 3 1
BIEC2-32706	76,631,203	67	6.10E-05	1.55	14 19 34	AA AC CC 1 15 31	13 4 3
BIEC2-32912	76,963,647	67	2.00E-06	1.72	13 17 37	AA AG GG 0 13 34	13 4 3
BIEC2-32970	77,220,026	67	2.30E-05	1.81	12 19 36	AA AG GG 0 14 33	12 5 3
BIEC2-33182	77,433,355	67	4.61E-03	1.62	14 21 32	AA AG GG 2 16 29	12 5 3
BIEC2-33183	77,433,370	66	5.85E-03	1.66	14 20 32	AA AG GG 2 15 29	12 5 3
BIEC2-33188	77,435,722	67	3.33E-02	2.69	16 25 26	GG AG AA 4 18 25	12 7 1
BIEC2-33837	78,303,728	67	1.78E-02	2.47	12 23 32	AA AC CC 2 15 30	10 8 2
BIEC2-33900	78,377,914	67	4.43E-02	2.20	14 23 30	AA AG GG 3 16 28	11 7 2
BIEC2-35047	80,361,702	67	4.96E-03	1.45	12 21 34	GG AG AA 1 16 30	11 5 4
BIEC2-35298	80,711,571	67	6.94E-03	1.97	11 17 39	CC AC AA 1 11 35	10 6 4
BIEC2-35714	81,858,337	67	2.81E-02	0.92	19 26 22	CC AC AA 5 23 19	14 3 3
BIEC2-36578	84,873,250	67	4.81E-02	1.59	8 18 41	AA AG GG 0 12 35	8 6 6
BIEC2-37017	87,079,561	67	3.39E-02	1.68	8 21 38	GG AG AA 0 14 33	8 7 5

*Odds ratio (without standard error) estimate for heterozygous genotype effect calculated as $e^{log\text{-}effAB}$; in all Friesian horses of this study.

Table 10. 54 associated haplotype blocks (*P-value* < 0.05) with hydrocephalus in Friesian horses based on chromosome 1 after 5 permutation testing

Name	Position (bp)	Size (kb)	Multiallelic D prime	P-value[a]	P-value[b]	Haplotype frequency	
						Cases	Controls
Block 102: AAG	57,760,860-57,764,065	3.205	0.285	2.14E-05	0.0099	0.075	0.457
Block 102: GGA	57,760,860-57,764,065	3.205	0.285	2.14E-05	0.0099	0.925	0.543
Block 105: GAA	58,999,235-59,003,649	4.414	0.770	2.11E-05	0.0098	0.100	0.489
Block 107: AGAA	60,145,840-60,189,924	44.084	0.313	3.76E-05	0.0203	0.975	0.630
Block 110: GG	60,556,150-60,562,293	6.143	1.000	3.76E-05	0.0203	0.975	0.630
Block 111: GGGAAAAGG	60,792,512-61,281,408	488.896	0.856	3.32E-07	0.0002	0.925	0.450
Block 112: GAAAGG	61,436,332-61,636,387	200.055	0.709	8.00E-07	8.00E-07	0.850	0.272
Block 113: AG	62,260,532-62,260,637	0.105	0.817	5.26E-07	0.0003	0.700	0.239
Block 115: ACCCCGAG	63,288,726-63,568,838	280.112	0.879	4.27E-07	0.0003	0.700	0.236
Block 117: GG	63,975,987-63,991,102	15.115	0.741	5.26E-07	0.0003	0.700	0.239
Block 120: AGCGA	64,771,094-65,184,880	413.786	0.595	5.84E-05	0.0268	0.750	0.370
Block 121: AAAA	65,732,134-65,769,727	37.593	0.444	4.67E-07	0.0003	0.825	0.348
Block 121: GCGC	65,732,134-65,769,727	37.593	0.444	5.44E-05	0.0244	0.175	0.554
Block 122: AA	65,841,804-65,842,030	0.226	0.371	1.62E-09	1.20E-06	0.875	0.304
Block 122: GG	65,841,804-65,842,030	0.226	0.371	1.62E-09	1.20E-06	0.125	0.696
Block 123: AG	65,869,896-65,869,949	0.053	0.554	3.41E-05	0.0173	0.825	0.435
Block 123: GA	65,869,896-65,869,949	0.053	0.554	3.41E-05	0.0173	0.175	0.565
Block 124: ACAGA	66,296,523-66,358,865	62.324	0.860	3.03E-11	0.00E+00	0.874	0.250
Block 124: GACAG	66,296,523-66,358,865	62.324	0.860	1.63E-07	0.0001	0.025	0.500
Block 125: AGGGGAA	67,222,990-67,647,989	424.999	0.909	1.00E-04	0.0471	0.925	0.587
Block 129: ACGAGG	68,597,487-68,778,947	180.682	0.915	5.38E-09	3.40E-06	0.025	0.567
Block 129: GAGCAA	68,597,487-68,778,947	180.682	0.915	3.99E-12	0.00E+00	0.775	0.152
Block 130: GGAGGAGC	68,835,676-69,326,275	490.599	0.986	2.93E-12	0.00E+00	0.774	0.149
Block 131: AAAGAAGGGCAGA	69,339,109-69,736,571	397.462	0.868	4.08E-12	0.00E+00	0.775	0.152
Block 131: CGGGGGAAACGGG	69,339,109-69,736,571	397.462	0.868	2.60E-05	0.012	0.000	0.338
Block 136: AC	71,344,522-71,348,106	3.584	0.546	3.75E-08	3.24E-05	0.100	0.620
Block 136: GA	71,344,522-71,348,106	3.584	0.546	3.75E-08	3.24E-05	0.900	0.380
Block 137: AAAGGAGGCG	71,495,554-71,635,419	139.865	0.572	7.38E-06	0.0038	0.846	0.424
Block 139: AAGACA	72,424,991-72,823,343	398.352	0.830	5.44E-08	4.16E-05	0.900	0.387
Block 142: AAAC	73,417,368-73,575,357	157.989	1.000	2.61E-05	0.0122	0.099	0.483
Block 142: CCGA	73,417,368-73,575,357	157.989	1.000	5.43E-10	6.00E-07	0.774	0.204
Block 145: AG	74,897,083-74,897,451	0.368	0.875	9.84E-07	0.0006	0.200	0.663
Block 145: GA	74,897,083-74,897,451	0.368	0.875	9.84E-07	0.0006	0.800	0.337
Block 146: ACGCAG	74,985,150-75,333,244	348.094	0.889	6.92E-08	4.38E-05	0.825	0.315
Block 147: AGGAG	75,409,929-75,523,947	114.018	0.991	7.36E-10	6.00E-07	0.800	0.228
Block 147: GAAGA	75,409,929-75,523,947	114.018	0.991	1.20E-05	0.0062	0.125	0.533
Block 148: AGGACGGAGAGA	76,508,159-76,961,647	455.488	0.943	7.00E-12	0.00E+00	0.750	0.141
Block 149: AAAGGAAG	77,220,026-77,435,722	215.696	0.942	5.27E-10	6.00E-07	0.700	0.152

Block 150: ACGGCA	77,854,425-78,280,130	425.705	0.803	4.95E-08	3.90E-05	0.700	0.206
Block 151: AAGAAG	79,173,540-79,592,692	419.152	1.000	1.07E-08	1.00E-05	0.650	0.152
Block 153: AACA	80,429,662-80,724,601	294.939	0.852	3.93E-09	3.20E-06	0.650	0.141
Block 153: GGAG	80,429,662-80,724,601	294.939	0.852	3.41E-05	0.0173	0.175	0.565
Block 154: GGGAGGAC	81,413,523-81,858,337	444.814	0.989	6.92E-06	0.0034	0.750	0.326
Block 155: CGAG	82,052,341-82,371,288	318.947	0.938	8.91E-06	0.0048	0.750	0.331
Block 156: AAAAGCAGAG	82,672,037-83,151,847	479.81	0.940	7.38E-06	0.0039	0.700	0.283
Block 157: GGGCG	83,324,830-83,803,319	478.489	1.000	1.02E-06	0.0006	0.550	0.141
Block 163: GAGAGG	84,808,413-84,977,960	169.547	0.840	4.06E-07	0.0003	0.550	0.130
Block 164: GAA	84,981,382-84,985,403	4.021	0.549	2.16E-06	0.0012	0.600	0.185
Block 166: AAA	85,025,168-85,030,689	5.521	0.829	2.16E-06	0.0012	0.400	0.815
Block 166: GGG	85,025,168-85,030,689	5.521	0.829	2.16E-06	0.0012	0.600	0.185
Block 167: CACA	85,302,089-85,624,691	322.602	0.726	3.60E-07	0.0002	0.549	0.129
Block 170: AAAAAG	86,708,696-87,079,561	370.865	0.828	1.02E-06	0.0006	0.550	0.141
Block 173: AAAGC	89,422,205-89,623,371	201.166	0.853	3.28E-07	0.0002	0.502	0.100

[a]*P-values* before permutation testing.
[b]*P-values* after 5 million permutation testing.