Norwegian University of Life Sciences
Faculty of Veterinary Medicine
Department of Preclinical Sciences and Pathology

Philosophiae Doctor (PhD)
Thesis 2024:02

Genetic studies of distichiasis in the Staffordshirebull terriers and cataract in Norwegian buhunds

Genetiske studier av distichiasis hos Staffordshire bull terrier og katarakt hos Norsk buhund

Dina Turid Ulstein Jørgensen

Genetic studies of distichiasis in the Staffordshire bull terriers and cataract in Norwegian buhunds

Genetiske studier av distichiasis hos Staffordshire bull terrier og katarakt hos Norsk buhund

Philosophiae Doctor (PhD) Thesis

Dina Turid Ulstein Jørgensen

Norwegian University of Life Sciences
Faculty of Veterinary Medicine
Department of Preclinical Sciences and Pathology

Ås 2023

Thesis number 2024:02
ISSN 1894-6402
ISBN 978-82-575--2113-4

Supervisors

Main supervisor

Frode Lingaas
Medical Genetics Unit, department of Preclinical Sciences and Pathology
Faculty of Veterinary Medicine
Norwegian University of Life Sciences
1433 Ås, Norway

Co-supervisor

Ernst Otto Ropstad
Evidensia Oslo Dyresykehus
Ensjøveien 14
0655 Oslo

Acknowledgements

The present work was carried out at the Unit of Medical Genetics, Department of Preclinical Sciences and Pathology, at the Norwegian University of Life Science, NMBU, between January 2019 and October 2023. Financial support was granted through a university scholarship.

I want to thank my supervisors, Frode Lingaas and Ernst-Otto Ropstad, for giving me this unique opportunity and sharing their knowledge and expertise. A special thanks to Frode for all the support, invaluable inputs and nice discussions during these last years.

I want to thank all the colleagues at the genetics unit for making this a special time. I want to share my special gratitude towards Ellen, Ole and Liv at the lab for guiding me through the rewarding but sometimes frustrating lab procedures. Thanks to Bjørn for the knowledgeable input. To Karin, thank you for letting me take part in your last paper and for sharing joys and worries. To Kim for always being positive and your everlasting enthusiasm towards canine genetics. A very special thanks to Mariella for always being cheerful and supportive.

Furthermore, I want to thank my friends, former officemates and fellow PhD students for great lunch breaks and inspiring conversations. A special thanks to my codriver Tonje for the early morning chats and shared worries about closed tunnels.

This project would not be possible without all the work laid down by the breed clubs. The contribution from the Norwegian Buhund club and especially Elin Kristiansen and Stein Dahl have been priceless. I'm grateful to all the dog owners who contributed with their dogs' DNA samples.

Thank you to Ida Marie for the excellent accommodation during sample collection at Buhundens Nasjonaldag and for just being a great friend.

Without the support from my family, this journey would not be possible. Thank you, Lars, for keeping up with me during the last periods and for all

the support and believing in me, to Frida and Ole for giving me needed breaks and being the joy of my life. Thanks to all the grandparents for taking care of sick children so I could go to work. Last but not least, a special thanks to my mother for inspiring me to choose an academic path. Without your support, this would not have been possible.

Table of Contents

1 Abbreviations and definitions

AUC	Area under the curve
BLUP	Best linear unbiased prediction
CFA	Notation of canine chromosome - Canis Familiaris
CI	Confidence interval
DoGA	Dog Genome Annotation
e	Residual error
ECVO	European College of Veterinary Ophthalmologists
GCTA	Genome-wide Complex Trait Analysis software
Gb	Gigabases
GBLUP	Genomic best linear unbiased prediction
GV	Genetic variance
GRM	Genetic relationship matrix
GWAS	Genome-wide association study
H^2	Broad-sense heritability
h^2	Narrow-sense heritability
HTS	High- throughput DNA sequencing
Kb	Kilobases
LD	Linkage disequilibrium
LDAK	Linkage Disequilibrium Adjusted Kinship software
MB	Megabases
MCMC	Marcov chain Monte Carlo
MLMA	Mixed linear model based association analysis
NKK	Norwegian Kennel Club
OMIA	Online Mendelian inheritance in animals
PCA	Principal component analysis
PGR	Polygenic risk (score)

PNC	Pulverulent Nuclear Cataract
PNK	Puverulent nukleær katarakt
PRA	Progressive retinal atrophy
QTL	Quantitative trait loci
REML	Restricted maximum likelihood
ROC curve	Receiver operating characteristic curve
SE	Standard Error
SNP	Single nucleotide polymorphism
V_A	Variance in the additive genetic effects
V_E	Variance in the environment
V_G	Genetic variance
V_p	Total phenotypic variance
WGS	Whole genome sequencing

2 List of papers

Paper I
Heritability estimates of distichiasis in Staffordshire bull terriers using pedigrees and genome-wide SNP data
Dina Jørgensen, Per Madsen, Ernst-Otto Ropstad, Frode Lingaas
Acta Vet Scand. 2022 Nov 21;64(1):30.

Paper II
Genomic analysis and prediction of genomic values for distichiasis in Staffordshire bull terriers
Dina Jørgensen, Ernst-Otto Ropstad, Theodorus Meuwissen, Frode Lingaas
Canine Med Genet. 2023 Jul 24;10(1):9

Paper III
A genome-wide study identifies a region on CFA37 associated with cataract in the Norwegian buhund
Dina Jørgensen, Ernst-Otto Ropstad, Elin Kristiansen, Stein Dahl, Frode Lingaas
Manuscript

3 Abstract

This thesis addresses the two most common eye disorders in the dog breeds, Norwegian buhund and Staffordshire bull terrier.

The first two papers focus on the eye disease distichiasis in Staffordshire bull terriers. Distichiasis is a condition where abnormal eyelashes grow inside the row of the regular eyelashes. These misplaced eyelashes can cause irritation and lesions to the cornea. Given the increased occurrence in certain dog breeds, distichiasis is believed to be a hereditary condition. The condition is relatively common in dogs but has also been observed in various other species, such as humans, cats, ferrets, horses, and cattle.

Paper I examines the prevalence and heritability of distichiasis in Staffordshire bull terriers in Norway. We studied eye examination data and pedigrees from 4177 dogs. Over half of the dogs underwent an eye examination around eight weeks of age. We investigated whether this early screening provided a reliable diagnosis of distichiasis. We found a significant difference in the risk of being diagnosed with distichiasis in the eye examination performed as a puppy and the eye examination of the same dogs as adult dogs. The overall prevalence of distichiasis in puppies was around 3%, while the prevalence in adult dogs was approximately 19%.

Based on these findings, we concluded that there is considerable uncertainty around a negative diagnosis in puppies. Therefore, heritability calculations were based on dogs with eye examination data after one year of age. By using both pedigree data and genetic marker data from microarrays, the heritability ranged from 0.37 to 0.48, which is considered moderately to highly heritable. With this knowledge, there should be good prospects for reducing the prevalence of distichiasis in Staffordshire bull terriers through increased use of dogs without displaced eyelashes.

Paper II describes a genetic association study where we used genetic markers distributed throughout the entire genome to find an association between genetic markers and distichiasis. We identified four potential genetic regions on chromosomes 1, 18, 32, and 34. Further studies are needed to verify these regions. Additionally, we explored the possibility of using all genetic markers on the microarray to predict the likelihood of a dog developing displaced eyelashes. By using the genetic markers, we could rank dogs based on genetic risk at the population level. However, the results were not precise enough on an individual level. Still, the method can be used like traditional breeding values with indices calculated from pedigrees.

The last paper, paper III, focuses on a specific type of cataract affecting Norwegian buhund, called pulverulent nuclear cataract (PNC). This disorder affects the nucleus of the lens. Around half of all examined buhund dogs are affected by PNC.

In paper III genetic markers distributed throughout the genome were used to look for an association between PNC and these genetic markers. We identified a region on chromosome 37 associated with the cataract. This region contains more than 50 genes, including five genes that have previously caused similar changes in humans and other species. We sequenced four of these genes in a subset of cases and controls using Sanger sequencing, but we did not find any causative mutation. The fifth gene was not known as a candidate gene at the time when the sequencing was performed. Whole genome sequencing of affected dogs and controls is in progress to identify risk variants that can be used to support breeding and hopefully develop a genetic test.

4 Norsk sammendrag

Denne avhandlingen tar for seg de to vanligste øyesykdommene hos hunderasene Norsk buhund og Staffordshire bull terrier.
De to første artiklene i avhandlingen omhandler øyesykdommen distichiasis hos Staffordshire bull terrier. Distichiasis er en sykdom der det vokser feilstilte øyehår innenfor raden av de normale øyevippene. De feilstilte øyehårene kan føre til irritasjon og skader på hornhinnen. Fordi det er sett en økt forekomst hos enkelte hunderaser, antas distichiasis å ha en arvelig bakgrunn. Lidelsen er relativt vanlig hos hund, men er også sett hos en rekke andre arter som menneske, katt, hest, ilder og storfe.

Den første artikkelen ser på forekomsten og arvegraden av distichiasis hos Staffordshire bull terrier i Norge. Vi har studert øyelysningsdata og stamtavler fra 4177 hunder. Mer enn halvparten av hundene fikk gjennomført en øyeundersøkelse ved åtte ukers alder, før valpen ble levert fra oppdretter til eieren. Vi undersøkte om denne tidlige undersøkelsen ga en sikker distichiasis diagnose. Vi fant en signifikant forskjell i sjansen for en positiv distichiasis diagnose hos de hundene som kun ble undersøkt som valper og de hundene som også ble undersøkt som voksne hunder, etter fylte ett år. Forekomsten av distichiasis hos valpene var rundt 3%, mens forekomsten hos voksne hunder var på om lag 19%. Basert på disse funnene konkluderte vi med at det er stor usikkerhet rundt en negativ diagnose hos valpene. Arvegradsberegningene ble derfor basert på hunder med øyelysningsdata etter fylte ett år. For å beregne arvegraden brukte vi både stamtavledata og data fra genetiske markører fra mikromatriser. Arvegraden lå mellom 0,37 - 0,48, noe som ansees som moderat til høy arvelighet. Med denne kunnskapen bør det være gode muligheter til å redusere forekomsten av distichiasis hos Staffordshire bull terrierne ved økt bruk av hunder uten feilstilte øyehår i avl.

Artikkel II er en genetisk assosiasjonsstudie der vi har brukt genetiske markører jevnt fordelt utover hele genomet for å finne en sammenheng

mellom genetiske markører og distichiasis. Vi identifiserte fire potensielt relevante genetiske regioner på kromosom 1, 18, 32 og 34. Videre studier må til for å verifisere disse regionene. I tillegg så vi på muligheten for å bruke alle genetiske markørene fra micromatrisen til å forutse hvor stor sannsynlighet det er for at en hund vil utvikle feilstilte øyehår eller ikke. Ved bruk av de genetiske markørene kunne vi rangere hundene basert på genetisk risiko på rasenivå. På individ-nivå var resultatene ikke nøyaktige nok, men metoden kan brukes på samme måte som tradisjonelle avlsverdier og avlsindekser beregnet fra stamtavler.

Den siste artikkelen, artikkel III, omhandler en spesiell type katarakt («grå stær») som er vanlig forekommende hos Norsk buhund. Denne sykdommen er kalt pulverulent nukleær katarakt (PNK) eller buhundkatarakt. Forandringen affiserer den indre kjernen til linsen. Sykdommen forekommer hos om lag halvparten av alle undersøkte buhunder.

I artikkel III har vi brukt genetiske markører fordelt utover hele genomet for å finne en assosiasjon mellom katarakt hos buhund og de genetiske markørene. Vi fant en region på kromosom 37 som er assosiert med buhundkatarakt. Området inneholder mer enn 50 gener, inklusive fem aktuelle gener som tidligere har vist å være assosiert med lignende forandringer hos andre arter, blant annet mennesker. Vi har sekvensert fire av disse genene med sangersekvensering uten å identifisere en mutasjon som har sammenheng med sykdommen. Det femte genet hadde ingen kjent sammenheng med PNK da sekvenseringen ble utført. Videre sekvensering av regionen må gjøres for å finne en mutasjon som har sammenheng med sykdommen og som kan brukes som hjelpemiddel i avlsarbeidet, og forhåpentligvis til å utvikle en gentest.

5 Synopsis

5.1 Introduction

5.1.1 A historical perspective

The resemblance among family members and the transfer of information from one generation to the next has intrigued scientists and philosophers since the distant past. This fascination traces back to ancient times, and in Aristotle's work "*On the Generation of Animals*," Aristotle (~350BC) extensively explored the very foundations of this concept. Though the mechanism behind the transfer of information from one generation to the next remained a puzzle until the 20th century, it was utilized by animal and plant breeders to improve traits in different breeding programs (Lynch and Walsh, 2018, pp 3-16).

Mendel's study of garden peas and the publication of the famous law of segregation in 1865 laid the foundation of modern genetics (Griffiths et al., 2002, pp 1-3; Jorde et al., 2016, pp 1-5). From the 20th century until today, the field of genetics has been revolutionized. Some highlights are Fisher's work within statistics and quantitative genetics (Fisher, 1918), Watson and Crick's description of the DNA's structure in 1953 (Watson and Crick, 1953), and the sequencing techniques developed by Sanger (Sanger et al., 1977). These contributions, among others, culminated in the human genome project – a large collaboration lasting for more than twenty years and resulting in the publication of the first draft of the human genome in 2001 (Lander et al., 2001; Venter et al., 2001) and partly completed in 2003 (International Human Genome Sequencing, 2004). The human genome project relied on Sanger sequencing and lasted for more than two decades. Advances in sequencing technology that led to the development of next-generation sequencing techniques have been remarkable, making it now possible to sequence genomes within hours (Giani et al., 2020). This technological development has also benefited the study of other species, and

in 2005, the first high-quality draft of a canine genome was published (Lindblad-Toh et al., 2005).

5.1.2 The dog in the humane society

Dogs are presumed to be the first animals to be domesticated; thus, the exact time and place are not agreed upon (Vilà et al., 1997; Savolainen et al., 2002; Freedman et al., 2014; Skoglund et al., 2015; Perri, 2016; Wang et al., 2016; Bergström et al., 2020). Recent studies indicate that the first dogs were domesticated more than 10,000 years ago, probably descending from one or a few closely related extinct wolf populations (Breen, 2008; Bergström et al., 2020). Dogs have lived and evolved alongside humans for thousands of years. They have become essential partners both as purpose-breeds used for hunting, herding, service dogs, military, and police dogs, but also as pure companions. The dog as a companion plays a vital role in human society, and the number of dogs in Norway was estimated to be around 565,000 in 2021 (DyreID, 2023). During the Covid-19 pandemic, the number of new dogs registered in the Norwegian Kennel Club (NKK) increased by about 30% compared with previous years (NKK, 2021b).

Health problems in dogs can lead to animal welfare issues and a substantial economic cost. The American Pet Products Association estimated that around $34.3 billion in the USA was spent on veterinary care and products in 2021 (APPA, 2022). A large number of canine disorders are of known genetic origin, and the compendium of inherited disorders, Online Mendelian Inheritance in Animals (OMIA), had, by September 2023, noted 903 Mendelian traits in dogs, most of these are disease traits (Nicholas et al., 1995; OMIA, 2023). Expanding the knowledge concerning the canine genome and the genetic basis of hereditary disorders can be used to increase the welfare of dogs.

The dog has been suggested to be a valuable model organism for studying human disorders (Karlsson and Lindblad-Toh, 2008; Davis and Ostrander, 2014), both due to the occurrence of similar naturally occurring diseases in humans and dogs but also owing to some unique features of the canine

genome facilitating the identification of genes associated with disease and behavioural traits.

5.1.3 The canine genome

The canine genome comprises around 2.4 gigabases (Gb) and more than 20,000 genes (Lindblad-Toh et al., 2005; Wang et al., 2021), distributed on 38 pairs of autosomal chromosomes and the X and Y chromosomes (Gustavsson, 1964; Lindblad-Toh et al., 2005; Breen, 2008). The first sequence of the canine genome was published in 2003, with a 1.5x coverage of approximately 78% of the genome of a standard Poodle (Kirkness et al., 2003), soon followed by a high-quality draft of the canine reference genome, canFam-2.0 (Lindblad-Toh et al., 2005). A whole genome shotgun approach was used, with a 7.5x coverage, including 99% of the genome. To further improve the accuracy and close gaps of canFam-2.0, a new version of the canine genome, canFam-3.1, was published in 2014 (Hoeppner et al., 2014). These first assemblies relied on short-read sequencing technology, and some gaps in the sequence persisted in certain regions that were difficult to sequence. Several new assemblies, including new long-read technologies, have recently been published (Edwards et al., 2021; Halo et al., 2021; Jagannathan et al., 2021; Wang et al., 2021). Utilizing the new sequencing technology, researchers have identified critical gaps, resulting in the creation of a more precise reference genome (Wang et al., 2021).

To further increase the knowledge of canine genetics, multicenter collaborations have been initiated, such as the DOG10K consortium and the Dog Genome Annotation project, DoGA. The DoGA aims to create a functional annotation atlas of the dog and wolf in an open database (DoGA consortium, 2023). Inspired by the Human 1000 Genomes project, the Dog10K genomes project has the goal of whole genome sequencing 10,000 dogs, including all known dog breeds, village dogs and wild canids (Dog10K, 2015; Ostrander et al., 2019; Wang et al., 2019). Making this database openly available provides a valuable resource to the scientific community.

Special features of the canine genome

The genetic basis of the dog was set when the species was created from a descent wolf population. Only a subset of variants from the original gene pool was brought on to the descendant dogs, leading to loss of genetic variation and the first of two main genetic "bottlenecks" in the dog's history (Lindblad-Toh et al., 2005). The second "bottleneck" occurred during the formation of the modern dog breeds. Dogs were selected based on appearance or physical skills, laying the foundation for future breeds (Sutter et al., 2004; Lindblad-Toh et al., 2005). Around the 19th and 20th centuries, kennel clubs and closed studbooks were established, creating small, closed breeding populations (NKK, 2023a; The Kennel Club, 2023) (Figure 1).

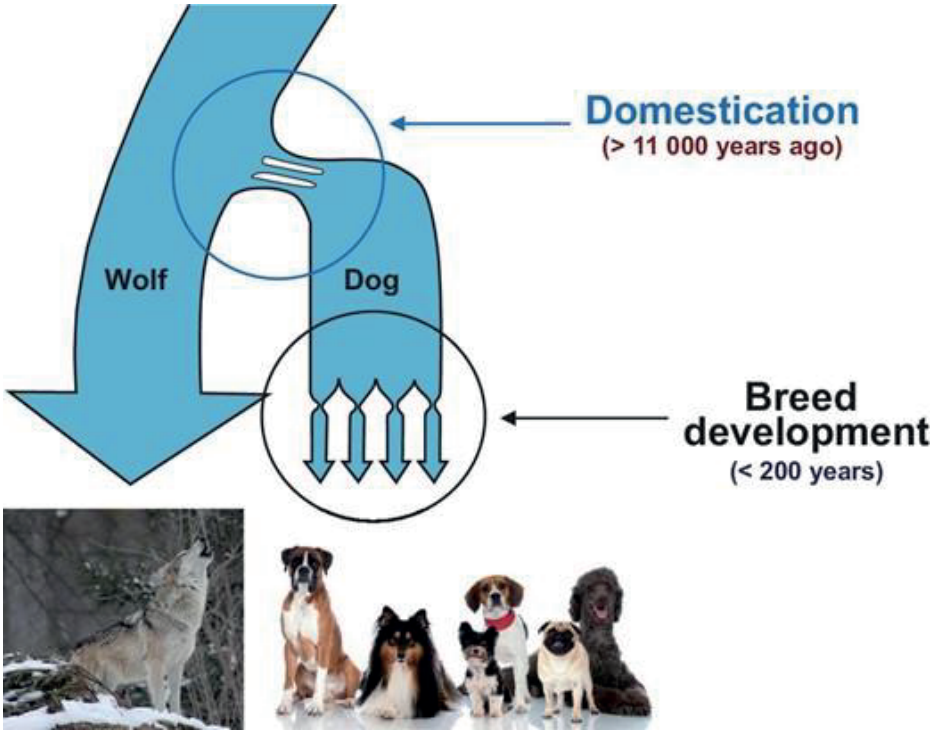


Figure 1. Illustration of the two first bottlenecks in the history of the dog. The illustration is reused from Ostrander et al. (2019) with permission under the Creative Commons licence.

In addition to the two main bottlenecks, several dog breeds, such as the puffin hunter, the Norwegian lundehund (Melis et al., 2013), have suffered reduced genetic diversity due to "private" bottlenecks resulting from disasters such as war, illness and reduced popularity. Also, using a low number of sires within a breed, known as the 'popular sire syndrome', has led to reduced genetic diversity in many breeds (Leroy, 2011).

Due to the recent genetic bottleneck, extensive areas of linkage disequilibrium[1] (LD) have been created. Within a dog breed, the LD has been found to range between 0.3 to 3.2 megabases (Mb) (Sutter et al., 2004) compared to the much shorter range of 11- 22 kb in humans, depending on the population (Gabriel et al., 2002). The LD measured in dogs between different breeds is primarily influenced by the LD resulting from the domestication of the wolf, and the LD blocks are relatively short, typically spanning only tens of kilobases (kb), and are comparable to the LD observed in the human population (Lindblad-Toh et al., 2005).

Small, closed breeding populations, such as dog breeds, can face the challenge of genetic drift and mating among close relatives (Leroy, 2011). This can lead to increased homozygosity in individual dogs, which can expose recessive detrimental mutations, potentially manifesting as diseases and inbreeding depression[2] (Charlesworth and Willis, 2009; Leroy, 2011). Inbreed depression has been documented in dogs, where increased levels of inbreeding have been linked to smaller litter sizes (Leroy et al., 2015; Chu et al., 2019) and decreased life expectancy (Leroy et al., 2015; Yordy et al., 2020). Genetic diseases are highly prevalent in some heavily bred dog

[1] Linkage disequilibrium is the non-random linkage of alleles at different loci in a population, often due to the alleles being physically close on the same chromosome or having limited recombination between them (Jorde et al., 2016, pp 161-164).

[2] Inbreed depression is the diminished fertility and survival of offspring born to closely related individuals (Charlesworth and Willis, 2009)

breeds (Axelsson et al., 2021). Inflation of harmful variants within a breed may be the result if such variants are linked to regions under positive selection. For instance, the wrinkled skin, characteristic of the Chinese shar-pei dog, is associated with a gene causing a condition known as periodic fever syndrome in this breed (Olsson et al., 2011). Such "hitch-hiking" effects may be one of the reasons why dogs seem to carry more deleterious variants compared to wolves (Marsden et al., 2016).

Despite the previously described challenges closed breeding populations can cause, the highly detailed pedigree records and the unique breed structures could provide excellent opportunities to explore the connection between phenotypes and relationships in dogs. This concept will be further elaborated upon in the following section.

5.1.4 Genetic architecture

In order to provide sound advice and make informed decisions regarding animal breeding, it is crucial to understand the extent to which genetic or environmental factors contribute to the variations in a particular trait. The genetic architecture refers to the underlying genetic basis of a trait and includes a description of the number of loci affecting the trait, the effect size of alleles at the different loci and how these interact.

A distinction is often made between qualitative and quantitative traits (Falconer and Mackay, 1996, pp 1-3; Lynch and Walsh, 1998, pp 3-16). In qualitative traits, the phenotype is usually determined by alleles on a single locus, thus having a monogenetic origin, often called Mendelian traits. The phenotype often exhibits distinct, non-overlapping categories or classes. There is a high number of examples of Mendelian disease traits, e.g. hereditary cataracts (Mellersh et al., 2006) and hereditary nasal parakeratosis in Labrador retrievers (Jagannathan et al., 2013), and many more are published on OMIA (Nicholas et al., 1995; OMIA, 2023). Nonetheless, many important traits are of a quantitative nature and exhibit continuous variation, where individuals in a population show a range of values rather than falling into distinct categories. These traits are often influenced by multiple genes and environmental factors, which contribute to

the observed variation. Such traits are also known as continuous or complex traits. Two examples of complex traits are height in humans (Wood et al., 2014) and behaviour in dogs (Ilska et al., 2017; Morrill et al., 2022). The literature often mentions a third group of traits known as threshold traits. Threshold traits are complex in nature and influenced by multiple genetic factors. However, the phenotype usually displays a binary or dichotomous distribution similar to Mendelian traits. The trait becomes evident only when a certain threshold of liability is reached (Falconer and Mackay, 1996, pp 300-311). Typical threshold traits are disorders where animals are classified as affected or not affected, e.g. cranial cruciate rupture (Baker et al., 2017) and cancer (osteosarcoma) (Letko et al., 2021). Due to the underlying continuity, threshold traits can be studied in a similar manner as complex traits by transformation to an underlying liability scale (Dempster and Lerner, 1950; Falconer, 1965).

5.1.5 Quantitative genetics and heritability estimates

Quantitative genetics addresses the study of continuous traits. The foundation of quantitative genetics is based on the research performed by Fisher (Visscher and Goddard, 2019), who introduced the term variation and the infinitesimal model in 1918 (Fisher, 1918). Fisher tied together the Mendelian theory of inherence with quantitative genetics by suggesting that complex traits were determined by an infinite number of unlinked additive loci, each with an infinitesimal small effect on the trait. Although the assumptions made in the infinitesimal model do not hold true in real biological traits (the genome is not infinite), it has shown to be a valuable tool in quantitative genetics. The variance in the phenotype (V_P) is determined by variances in the genotype (V_G) and the environment (V_E): $V_P = V_G + V_E$. The environmental variance can be partitioned into several different factors contributing to the observed phenotype, like the litter or siblings, feeding, housing, breeder, measurement error, and residual error (Visscher et al., 2008). The genetic variance is broadly made up of the additive genetic effects, interactions of alleles within a locus (dominance) and interactions of alleles between different loci (epistasis), as well as

interactions between the genotype and environment (Caballero, 2020, pp 42-66).

Heritability

The heritability of a trait is an estimate of the proportion of the total phenotypic variance that may be explained by genetic variance within a population (Mills et al., 2020, pp 3-30). Heritability is expressed on a scale between zero and one. Values near zero suggest that the environment is the primary cause of variation, whereas values close to one indicate that genetic factors are the primary cause of variation.

A distinction is made between broad-sense heritability, also referred to as; $H^2 = \frac{V_G}{V_P}$, including all genetic factors (additive, dominance, and epistasis) and narrow-sense heritability, $h^2 = \frac{V_A}{V_P}$, including only the additive genetic effects (V_A) (Mills et al., 2020, pp 3-30). Although debated, the theory is that dominance and epistatic effects in most complex traits are small and can be neglected (Hill et al., 2008). One argument is that the complete genotypes that determine a trait are not inherited as a single unit from the parent; instead, each allele is inherited independently (to a greater or lesser extent). The exception is estimates based on twins and full siblings sharing more genetic material (Falconer and Mackay, 1996, pp 146-184). Even though the effect of dominance and epistasis are expected to be minor, it has been shown that the non-additive effects can vary between traits (Sun et al., 2014). Heritability estimates are often connected with a large standard error (Falconer and Mackay, 1996, pp 163-184). The estimates are population-specific and depend on the distribution of alleles and the particular environment (Visscher et al., 2008).

In human medical genetics, heritability estimates can provide important insight into the nature of a disease and may be helpful in estimating the risk (and recurrence rate) of developing complex disorders within families and populations, such as diabetes (Willemsen et al., 2015) and Alzheimer's disease (Gatz et al., 2006).

The foundation of quantitative genetics was developed without knowledge about the genes influencing the traits studied (Lynch and Walsh, 2018, pp 3-16) and, therefore, based on observations of the variation in the phenotype among different groups of relatives. If a particular trait is more commonly observed among relatives or if relatives tend to exhibit greater similarity in the trait compared to more distant relatives or the general population, this may indicate shared genetic factors underlying the trait (Falconer and Mackay, 1996, pp 148-161). In the beginning, heritability estimates were usually based on pedigree information and studied in balanced designs by regressing the phenotypic value of the offspring on parents or by comparing siblings (Falconer and Mackay, 1996 pp 163-184; Lynch and Walsh, 1998, pp 779-803). As the statistical methods improved, complex pedigrees with different levels of relationships were included, using mixed models such as "the animal model". In the animal model, the relationship coefficients (also known as the additive genetic covariance) are fitted between each pair of individual animals within the whole dataset. The variance component can be estimated using methods like restricted maximum likelihood (REML) or Bayesian approaches (Caballero, 2020, pp 122-147).

As genomic technology has advanced and the costs of genotyping and genome sequencing have decreased, new approaches utilizing genomic data have been developed and integrated into the field of quantitative genetics (Meuwissen et al., 2001). The statistical methods applied to genomic data are, to a large extent, the same approaches as when using pedigree data, for example, mixed linear models similar to the animal model. However, the relationship coefficients from pedigrees are replaced by estimates of the genomic relationship between pairs of individuals calculated from genetic markers (VanRaden, 2008; Yang et al., 2010).

There are several benefits of using genomic heritability estimates compared to traditional estimates based on pedigrees. One is that estimates can be made based on genomic data alone when pedigree information is unknown (Ritland, 1996). In human genetics, heritability estimates based on twin studies may overestimate the heritability due to shared environment. Genomic estimates on unrelated individuals can reduce the biases

introduced by shared environments (Yang et al., 2010; Mills et al., 2020, pp 3-30). Combining pedigree and genomic data can increase the precision of the heritability estimates (Haile-Mariam et al., 2013) and be useful when genotype data of some animals are missing (Christensen and Lund, 2010).

Examples where heritability estimates have been based on genomic markers in dogs are cranial cruciate rupture in Labrador retrievers (Cook et al., 2020), osteosarcoma in Leonberger (Letko et al., 2021), episodic exercise-induced collapse in border collies (Norton et al., 2021), and fear of sounds in standard poodles (Handegård et al., 2023).

5.1.6 Identification of genes and variants associated with phenotypes

Numerous approaches have been developed to search for genetic variants associated with traits of interest. The method of choice depends on the genetic architecture of the trait and available material.

Linkage and candidate gene studies

Before high-density gene maps were developed, linkage studies of high-risk families were common. Pedigrees of families with segregating disease phenotypes were studied, and individuals of two or more generations were genotyped using panels, usually of highly polymorphic short tandem repeat markers[3]. Using this strategy, one could identify linkage to sets of markers in specific chromosomal regions (Jorde et al., 2016, pp154-166). Followed with fine mapping, genomic regions containing causal variants could be identified. A limitation was the requirement of large families segregating a

[3] Short tandem repeat markers are repetitive DNA sequences that typically consist of 2-5 base pairs repeated in a consecutive manner. These markers often exhibit high variability and uniqueness among individuals (Jorde et al., 2016, pp 45-46)

disease caused by the same genetic variant. Therefore, high locus heterogeneity could represent a challenge. Long ranges of LD within a family, like in many dog breeds, could also make it difficult to pinpoint the causal variant. Linkage studies were, for instance, used to identify genomic regions associated with hereditary renal cancer syndrome (Jónasdóttir et al., 2000) and neuronal ceroid lipofuscinosis in English setter and border collie (Lingaas et al., 1998; Melville et al., 2005).

A candidate gene study is a direct approach used to identify causal variants linked to a particular trait or disease. This method can be used when prior knowledge of the trait provides a list of candidate genes likely to be involved in the phenotype of interest. Genes are usually chosen based on biological function, association with similar phenotypes in other species, or genes within a genomic region identified by a previous study. Sequence variants are compared between healthy and diseased individuals. Candidate gene studies have been criticized because false positives are often identified for (human) complex disorders and because of problems replicating the findings (Border et al., 2019). However, several disease-associated genes have been identified in dogs, for example, the nonsense mutation in the *PDE6B* gene associated with rod/cone dysplasia in Irish setter (Suber et al., 1993) and a deletion in *RPE65* linked with retinal dystrophy of Swedish briard (Veske et al., 1999).

Genome-wide Association Studies

Genome-wide association studies (GWAS) usually use high-resolution microarrays to compare thousands to millions of single nucleotide polymorphism (SNP)-markers across the whole genome for association with a given trait by regressing each SNP marker on the phenotype one SNP at a time (Mills et al., 2020 pp 77-97). In GWAS, causal variants are usually not directly detected; instead, the study identifies markers in LD with causal variants. This is primarily because most SNPs on the array represent common natural variants often unrelated to disease. GWAS is usually conducted in populations, including unrelated individuals, making gathering

material somewhat simpler than family studies. Some of the first GWASs in human medicine were fulfilled at the beginning of the 21st century (Ozaki et al., 2002; Klein et al., 2005). Due to its simple design and applicability in the study of complex diseases, GWAS soon became an attractive method. By July 2023, the human GWAS catalogue (https://www.ebi.ac.uk/gwas/home) had registered 6499 publications and more than 500,000 SNPs associated with different traits (Sollis et al., 2023) (Figure *2*).

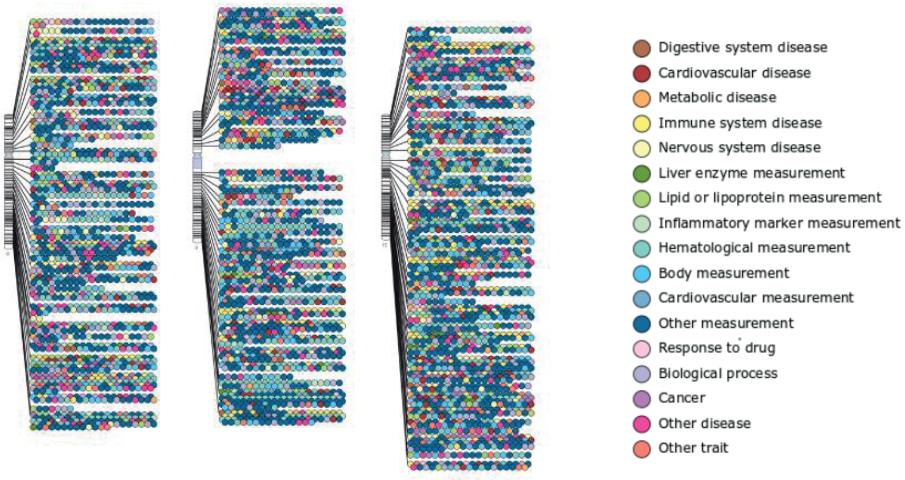


Figure 2. The diagram displays the human chromosomes 8, 9, 10. The coloured dots represent associations with p-values less than 5x10^{-8}. Different trait categories are coloured differently and displayed on the right side of the figure. The diagram with the summary statistics was downloaded from the NHGRI-EBI GWAS Catalog (Sollis et al., 2023) on 28/06/2023, available from https://www.ebi.ac.uk/gwas/diagram. With permission to use under the general EMBL-EBI terms (https://www.ebi.ac.uk/about/terms-of-use).

Despite the success of GWASs, there are several challenges. The most obvious is that the number of tested markers usually exceeds the number of individuals. A stringent significance level is set to correct for multiple testing and avoid false positive findings. Bonferroni correction and permutation are commonly used to set an acceptable significance level (Sham and Purcell,

2014). Permutation can be used to determine the significance level by comparing the observed test statistic to a distribution of test statistics generated by shuffling the phenotype labels multiple times. A traditional Bonferroni correction accounts for multiple testing by defining a significance level, often 0.05, divided by the number of independent statistical tests performed. In human GWAS, there is a general agreement on a significance level of $5.2x10^{-8}$ reflecting a Bonferroni correction estimated on the number of independent haploblocks[4] in the European population (Gabriel et al., 2002; Altshuler et al., 2005; Xu et al., 2014). As previously mentioned, the LD level within a dog breed is more extensive than in humans, and it can span across regions of several Mb, depending on the breed's history (Sutter et al., 2004; Lindblad-Toh et al., 2005). Karlsson et al. (2013) suggested a significance level of $2x10^{-5}$ based on the average size of 1Mb of independent haploblocks in a 2.4 Gb dog genome.

GWAS is usually the preferred design when studying complex traits (Mills et al., 2020, pp 77-97). Because complex traits are influenced mainly by multiple loci with a small effect, a stringent significance level requires large study populations to detect causal variants (Sham and Purcell, 2014). In such designs, an excess of false negatives is expected. The power to detect a causal variant in a GWAS depends on several factors, including the allele frequency in the population, the effect size of the alleles, and how close a marker is linked to the causal variant. Causal variants with a small effect size and a low allele frequency (a low number of dogs carry the causal variant) can be challenging to detect and require large sample sizes (Figure 3).

[4] Haploblock or haplotype blocks are genetic region (haplotypes) that tends to be inherited together due to a low level of recombination. This results in a relatively consistent pattern of genetic variation within the block across generations, and with a strong correlation between different SNPs in the block (Gabriel et al., 2002)

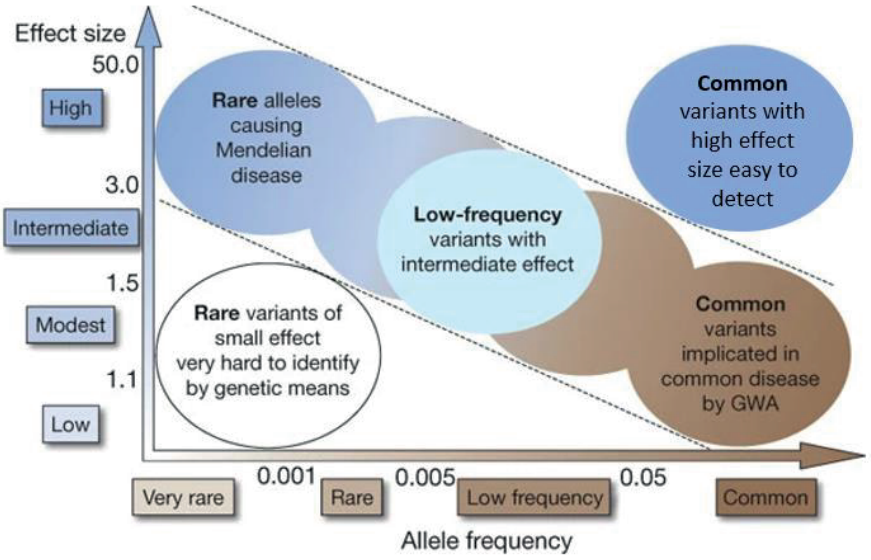


Figure 3. The figure illustrates how the effect size and allele frequency affect the identification of causal variants. The figure is a modified version of Manolio et al. (2009) with permission from Springer Nature. The original figure of Manolio et al. was based on the figure of McCarthy et al. (2008) with permission from Nature Reviews Genetics.

Detecting the causal variant can also be difficult when the SNP markers are distant from or in poor LD with the causal variant. In addition, locus heterogeneity, where different loci can cause the same phenotype, may pose a challenge**.** Genotypic heterogeneity is well known in Mendelian inherited disorders such as cataracts (Shiels and Hejtmancik, 2017) and has been observed in a number of complex psychiatric disorders such as major depression in humans (Nguyen et al., 2022).

There has been a tremendous increase in sample size in many human GWAS to overcome these challenges. Multicentre consortiums have been created, such as the Psychiatric Genomics Consortium (https://pgc.unc.edu/) and the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (https://www.hgsc.bcm.edu/human/charge-consortium) (Psaty et al.,

2009). With large multicentre collaborations, sample sizes with several hundred thousand and up to a million participants are achieved in the study of human disorders such as Alzheimer's disease (Wightman et al., 2021), schizophrenia (Trubetskoy et al., 2022) and diabetes type 2 (Mahajan et al., 2022). In canine genetics, there has been a lack of big open biobanks sharing phenotypes and genotypes, but there are some attempts, such as the project Darwin's Ark, where volunteering dog owners answer a questionnaire and submit DNA from their dog to be used in science (https://darwinsark.org).

Another challenge in GWAS is false positives due to spurious associations. A spurious association refers to a statistically significant correlation between a genetic marker and a trait that appears to be real but is actually false and not causally related to the trait under investigation. It occurs when there is no true functional biological link between the genetic marker and the trait under investigation (Mills et al., 2020, p 403). Poor quality control of markers can lead to spurious associations. Therefore, implementing strict quality control measures is crucial, particularly in case-control studies. Spurious associations may also occur due to batch effects, variations in array types, or technical artefacts, leading to differences in allele frequency between cases and controls not caused by an actual biological difference (Marees et al., 2018). Population structure and relationship are other common sources of spurious associations and must be addressed to avoid false positives. Standard methods are to include a principal components analysis (PCA) (Price et al., 2006) or a genetic relationship matrix (GRM) in a linear model. PCA is a statistical procedure that reduces the dimensionality of a dataset by identifying distinct axes of variation and reducing the data's variance into a number of dimensions that capture the maximum amount of variability. While PCA can capture population structure resulting from population admixture, it is less effective in correcting for relationships between individuals in the dataset (Price et al., 2010). In a dataset with a high degree of relationship between study objects, which is often the case within a dog breed, a mixed linear model associations study (MLMA) including a GRM is recommended (Yang et al., 2011). The incorporation of a GRM makes it possible to adjust for population stratification and

relationships by quantifying the genetic similarity between pairs of individuals.

As mentioned earlier, the dog was predicted to be an excellent model organism for genetic studies, and there were great expectations owing to the unique features of the canine genome, with extensive LD, low genetic variance, and accumulation of disease alleles within a breed (Sutter and Ostrander, 2004; Karlsson and Lindblad-Toh, 2008). Karlsson et al. (2007) estimated that 20 dogs could be enough to map a trait of recessive inherence, and around 40 dogs could be sufficient for a trait of dominant inheritance. The low number of participants needed to detect a causal variant has been confirmed in several GWAS, such as the associated study of Dandy-Walker-like malformation in Eurasier dogs, which located a mutation in the *VLDLR* gene, including only nine cases and eleven controls (Gerber et al., 2015). The study of hereditary nasal parakeratosis in Labrador Retrievers included 13 cases and 23 controls and identified a mutation in the *SUV39H2* gene (Jagannathan et al., 2013), and a GWAS concerning congenital stationary night blindness in Beagle included only twelve affected and eleven unaffected dogs to identify a causal variant in the *LRIT3* gene (Das et al., 2019).

On the other hand, as previously stated, complex traits require a greater sample size. However, the required sample size was expected to be relatively moderate due to the limited genetic variability within specific dog breeds and the high disease prevalence in certain breeds. Around 100 cases and 100 controls were expected to be sufficient to identify loci with a fivefold increased risk, while approximately 500 cases and 500 controls were needed to detect variants with a twofold increased risk (Karlsson and Lindblad-Toh, 2008). The low number of dogs needed to detect significant markers associated with a complex disease was later confirmed, exemplified in the study by Wilbe et al. (2010) using 81 cases and 57 controls to identify five genomic regions associated with canine systemic lupus erythematosus in Nova Scotia duck tolling retriever.

Even though there have been many successful GWAS within canine genetics, long haploblocks and extensive LD can be a challenge as they give low resolution, and identifying the causal variant can be difficult. Sutter et al. (2004) suggested that by including closely related breeds with similar phenotypes, who share the same causal variants resulting from mutations that occurred before the breeds diverged, it is possible to disrupt the long ranges of LD seen within a breed and increase the resolution of the GWAS. Increasing the number of markers can also boost statistical power by reducing the average distance between markers and potential causal variants. However, studies in genomic prediction have shown that this improvement might be limited in enhancing accuracy within a breed or specific breeding lines due to long LD and because adding more markers can introduce noise since most of these new markers will not have an impact on the trait (van Binsbergen et al., 2015; Zhang et al., 2018).

Bayesian approaches to GWAS

Bayesian approaches in genetics were first applied to genomic predictions but have later also been used in GWAS (Meuwissen et al., 2001; Fernando and Garrick, 2013, pp 237-274; Wray et al., 2019; Tengvall et al., 2022; Lingaas et al., 2023). In the Bayesian GWAS methods, Bayes' theorem is applied to predefine a prior distribution of the SNP effect, and then this information is combined with the observed data (Fernando and Garrick, 2013, pp 237-274; Wolc and Dekkers, 2022).

In a traditional linear model, all SNPs are assumed to come from one normal distribution of the SNP effect, where most SNPs have no or little effect on the phenotype, and only a few SNPs exhibit a large effect on the trait. In a Bayesian approach, more flexible prior information on the SNP effect can be incorporated into the model. These priors can be constructed from an assumed or known distribution of the SNP effects. The development of different sets of prior distributions of SNP effects has led to the term "Bayesian alphabet" (Gianola et al., 2009). Examples are BayesA and BayesB (Meuwissen et al., 2001), BayesC (Habier et al., 2011), and BayesR (Erbe et

al., 2012). The main difference between these models is how they model the distribution of SNP effects. BayesA incorporates a t-distribution for the SNP effect, and BayesB uses a mixture of a t-distribution with a probability of π and an additional t-distribution with zero effect with a probability of 1-π (Meuwissen et al., 2001). BayesC assumes a common variance of all SNPs and a mixture of two normal distributions, one with the probability of π and one distribution with zero effect and a probability of 1-π (Habier et al., 2011). BayesR includes four normal distributions with SNP effect, with zero, small, medium and large effects (Erbe et al., 2012). When available, these models can be further improved by including information from previous studies or biological knowledge concerning certain genomic regions or SNPs in the model; an example of this is BayesRC (MacLeod et al., 2016). By including such information, different SNPs can be differentiated according to their assumed effect on the phenotype. The best methods will, however, depend on the assumed or known distribution of the SNP effect in the given dataset; a distribution of the true SNP effect matching the prior distribution in the model is an advantage (Wolc and Dekkers, 2022).

The Bayesian models fit all marker effects simultaneously, and with this, they address the issues of population structure and relationship (Wolc and Dekkers, 2022). By fitting all SNPs simultaneously instead of testing one SNP at a time, like traditional GWAS, the problem of a strict significance level and a low power to detect an association is claimed to be reduced (Fernando and Garrick, 2013, pp 237-274).

Post GWAS

A successful GWAS aims to identify one or more genomic regions associated with a phenotype. The next step is usually to search for potential candidate genes within the associated region. Usually, the first approach is to look for relevant candidate genes in the identified regions using developed gene maps. Targeted resequencing of a single region is an attractive method that can reveal potential causal variants, and in situations with several candidate loci, whole genome sequencing of a few cases and controls may be cost-

efficient. Detected variants also need to be verified in unrelated individuals or populations. Identifying such variants can give further insight into the biology of the disease and make it possible to establish genetic tests. An example is the research on parakeratosis in Labrador retrievers, where a mutation in the *SUV39H2* gene was identified (Jagannathan et al., 2013).

However, most common disorders are complex traits influenced by several loci, sometimes with interactions within and between loci and environmental factors, as shown in studies of atopic dermatitis (Tengvall et al., 2022) and hip dysplasia in dogs (Mikkola et al., 2019). Single variants can be strongly associated with a phenotype but may only explain a small portion of the variance in the phenotype, with the rest of the variance caused by numerous loci with a small effect on the trait. Gathering all information obtained by the GWAS makes it possible to create a risk profile based on the accumulated effect of all risk loci. This approach will be further elaborated in section 5.1.7 concerning genomic prediction.

When multiple genomic regions are identified, a list of genes can be defined within these regions. The gene set can be analysed and possibly identify enriched biological processes or pathways. Examples of pathways analysis performed in canine genetics are anterior cruciate ligament rupture, where c-type lectin pathways were enriched (Baker et al., 2017), and osteosarcoma (Karlsson et al., 2013).

The identification of genes and pathways can give a deeper understanding of the biology of a trait and makes it possible to move on from statistical risk associations to functional studies by exploring how genes and intergenic regions affect different biological processes.

High- throughput DNA sequencing

The development of new efficient sequencing technology and computer power has made it possible to sequence whole genomes at a reasonable cost and time (Giani et al., 2020). It can give high coverage and high resolution and the possibility to directly identify the causal variant. While whole exome sequencing focuses on only the protein-coding exones, whole genome

sequencing (WGS) sequences the entire genome and may therefore also include causal variants in intergenic regions and regulatory elements (Jorde et al., 2016, pp 173-179).

Sanger sequencing is often referred to as first-generation sequencing and is considered the golden standard of DNA sequencing (Shendure and Ji, 2008). It has high accuracy, but it is time-consuming and costly. With massive parallel sequencing, also known as next-generation sequencing, amplification of the DNA gives a large number of short overlapping sequences that are sequenced simultaneously, followed by aligning to the reference genome (Giani et al., 2020). This gives the opportunity to sequence large genomes in a relatively short time and with high accuracy. There are, however, some challenges in the alignment process when using technologies based on short sequences. Genomic regions with a large number of tandem repeats, like telomeres and centromeres, long repeated elements and duplicates within the genome, could be bioinformatical demanding to place accurately. In addition, it has been difficult to sequence cytosine and guanine-rich regions (Ebbert et al., 2019).

With long-read technology, also known as third-generation sequencing, reads of several Mb pairs are possible to obtain. This has made it possible to sequence whole chromosomes from telomere to telomere (Miga et al., 2020). Two common methods are nanopore sequencing (Oxford Nanopore Technologies, https://nanoporetech.com/) and single-molecule real-time sequencing (Pacific Biosciences, https://www.pacb.com). With the long-read technology, repeated elements and duplicates within the DNA can be sequenced and correctly placed. In addition, Nanopore can detect base-pair modifications such as methylation, which is important in regulating gene expression (epigenetic effects) (Amarasinghe et al., 2020).

With the increasing number of dogs being sequenced, new insight into the canine genome and existing naturally presumed harmless variants are provided, which might make variant calling and filtering variants more efficient.

So far, an important application of high-throughput DNA sequencing (HTS) in medical genetics has been in family studies of monogenetic disorders of recessive inheritance, for example, in trio studies with two healthy parents and one or more affected offspring (Bellamy et al., 2022). Due to the high cost and the challenge with the storage and handling of a large amount of data, the use of HTS in population studies has been limited. However, studies based on whole genome data in populations have started to emerge, for example, in the study of Parkinson's disease in humans (Billingsley et al., 2023).

Low-pass whole genome sequencing and imputation

In low-pass whole genome sequencing (low-pass WGS), samples are sequenced with a low average number of reads (down to one), first of all, to reduce costs. Ambiguous and missing sequences (due to gaps during the sequencing process) can be picked up or corrected from databases of known haplotypes by imputation from high-quality reference genomes. A similar approach is to impute missing genotypes from microarrays. This can increase the number of markers compared with standardized reference panels, such as customized SNP arrays. Low-pass WGS and imputations make it possible to obtain close to whole genome sequence data on a large number of individuals at a low cost to be used in association studies at the population level (van den Berg et al., 2017; Buckley et al., 2022).

Global expression analysis

Global expression analysis is a quantitative study of gene expression in different tissues, organs, or cell types (Lovén et al., 2012). It involves quantifying messenger RNA transcripts in specific cells and can be used to compare the gene expression of multiple genes in affected and unaffected individuals. The RNA level can be measured using various methods, such as microarrays and complementary DNA sequencing with massive parallel sequencing techniques. Expression analysis can provide valuable information on gene activity under specific conditions, including the gene

activity of different developmental stages, and allow for further insights into the underlying biology of a disease. Additionally, it has the potential to uncover potential biomarkers for diagnostic purposes.

5.1.7 Genomic prediction and breeding values

In medical genetics, it is of great importance to be able to predict the disease phenotype of individuals or relatives as well as the recurrence risk in families with affected relatives. Both in human and veterinary medicine, predicting a disease enables the implementation of early interventions and could minimize the impact of a disease. In animals, there are additional important possibilities to use the knowledge of disease-associated variants to support the selection of the best animals for use in breeding. Because most complex traits are caused by multiple genes with a small effect on the trait, simple gene tests are not possible. However, quantifying the total genetic effect of associated markers that contribute to the trait can be used to predict the phenotype. Such quantification or risk estimation is the basis for genomic breeding values in animals and polygenic risk scores (PGR) in humans (Wray et al., 2019). Breeding values in animals are commonly transformed into indexes based on the population mean. The estimated breeding value is used to rank the different breeding animals by their genetic merits, i.e., how likely they will bring on a trait to their offspring.

Estimating the breeding values and PGR has two main goals. In animals, especially livestock animals, the aim is usually to predict breeding values to produce the best offspring and to monitor the selection responses. In human genetics, the goal is to predict a phenotype in an individual person (Wray et al., 2019). In dog breeding, like livestock breeding, there is a need for improved accuracy in predicting and selecting the best breeding animals. Additionally, like in humane medicine, there is a great interest in predicting the phenotype of individual dogs. However, sometimes, the low prevalence of a disease, high cost, and logistical challenges may not necessarily justify implementing complex genomic selection systems compared to traditional selection methods in the breeding of dogs.

Breeding values can be estimated with high accuracy using an extensive dataset of pedigree information. A preferred approach among animal breeders was the best linear unbiased prediction (BLUP), developed by Henderson in the 1950s (Robinson, 1991).

Genomic selection

The main idea of genomic selection is to summarize and weigh the genetic effects of all loci affecting a trait to improve the accuracy of predicted breeding values. The challenge is that the loci and the relative effect size of the alleles are usually unknown. Quantitative trait loci (QTL) were included in marker-assisted selection work more than 30 years ago (Fernando and Grossman, 1989; Lande and Thompson, 1990). However, predictions based only on known QTLs could have relatively low accuracy in complex traits. In 2001, Meuwissen et al. showed, in a simulation study, that it was possible to use genomic SNP markers from the whole genome to predict breeding values. Several approaches have been developed, such as the Genomic BLUP (GBLUP) and Bayesian approaches, such as those previously described in the section about Bayesian approaches to GWAS.

The predictive performance of genomic prediction methods varies significantly depending on the specific characteristics of the study population and the underlying genetic architecture of the trait. For instance, BayesGC demonstrated higher accuracy than GBLUP in most maternal traits in sows, but in traits like the number of stillborn piglets, there was no difference between BayesGC and GBLUP (Kjetså et al., 2022). In addition to the influence of the methods used, the accuracy of genomic selection depends on the heritability of the trait, the reference population size and structure, and the number of independent loci affecting the trait (Daetwyler et al., 2008; Wray et al., 2013; Kjetså et al., 2022). Increasing the number of genotyped and phenotyped individuals may increase the accuracy of the estimates (Wray et al., 2019).

With the implementation of breeding values, systematic selection has become more precise. Genomic selection adds further improvements due to

the possibility of shortening the generation interval and improving accuracy. As a result, livestock populations have seen significant gains in several production traits of economic value (Zuidhof et al., 2014; García-Ruiz et al., 2016; Hickey et al., 2017).

Examples of traits where traditional breeding values often support selection in dog breeding are the breeding programs aiming to reduce the prevalence of hip and elbow dysplasia (Oberbauer et al., 2017; Hedhammar, 2020). Genomic prediction has so far not been used to a large extent in dog breeding. Nevertheless, there are some studies describing the possibilities of the use in complex traits such as cranial cruciate ligament rupture (Baker et al., 2020), hip dysplasia (Guo et al., 2011; Sánchez-Molano et al., 2015; Jiang et al., 2021) and in kidney disease (Lingaas et al., 2023).

5.1.8 The eye

Genetics of eye disorders in dogs

Ocular disorders in certain dog breeds are relatively common, and many of these conditions can cause pain, reduced vision and impaired quality of life for the affected dogs (Gelatt et al., 2013). Combined with the fact that the eye is an organ easily approachable for examination due to its location (Mellersh, 2014), many breed clubs have implemented official eye screening programs to monitor and prevent breeding dogs with hereditary eye disorders (NKK, 2023b). The eye screenings in Norway are performed by veterinarians with a special certification from the European College of Veterinary Ophthalmologists (ECVO), using standardised schemes to record the eye diagnosis (https://www.ecvo.eu). To be able to register the dog within the Kennel Clubs, eye screening is mandatory in some breeds, such as the Norwegian buhund (NKK, 2023b), while it is only a recommendation in other breeds, such as the Staffordshire bull terrier (The Norwegian Terrier Club, 2021).

The relatively high prevalence and detailed records of eye diseases in many breeds have provided unique opportunities for genetic studies of eye disorders. These studies have increased the knowledge about the aetiology of eye diseases, including the heritability. Collection of DNA samples from dogs with high-quality diagnoses (clinically affected or unaffected) has facilitated genetic mapping of associated genes and made it possible to develop genetic tests. Some examples are the mutation in the *HSF4* found to be causative of hereditary cataracts in Staffordshire bull terriers, Boston terriers and French bulldogs (Mellersh et al., 2006; Mellersh et al., 2007), the mutation in the *PRCD* gene causing progressive retinal atrophy (PRA) in multiple breeds (Acland et al., 1998; Zangerl et al., 2006), canine multifocal retinopathy due to a mutation in *BEST1* (Guziewicz et al., 2007), and collie eye anomaly-choroidal hypoplasia caused by a mutation in the *NHEJ1* gene (Lowe et al., 2003; Parker et al., 2007). Genetic tests in dogs can lower the incidence rate of inherited disorders by identifying and using unaffected animals in breeding. They also enable the safe use of carriers of recessive disorders when mated to non-carriers, thus preserving genetic variance while reducing the disorder's impact.

The anatomy of the eye

The following two sections describe the anatomy of the eye based on the anatomy books: "*Lehrbuch der Anatomie der Haustiere"* (Nickel et al., 2004, pp 405-444) and "*The Textbook of Veterinary Anathomy*" (Singh et al., 2018, pp 318-331). The canine eye lies within the *orbital fossa* and is protected by the skull and the eyelids. The eye consists of the eyeball and the *adnexa* with the eyelids, musculature, tear system, blood, nervous system, and a posterior fat pad. The eyeball, *Bulbus oculi,* has three tunic layers: the *tunica fibrosa bulbi, tunica vasculaosa bulbi* and *tunica interna bulbi*. The outer *tunica fibrosa bulbi*, surrounds the eyeball and protects and gives the eyeball its round shape. It consists of the sclera in the posterior and the transparent cornea on the anterior part of the eye. The next layer *Tunica vasculaosa bulbi* or uvea, consists of the choroid, ciliary bodies and the iris. The main function of the vascular tunic is to nourish the eye, but it also holds the lens in place

with the zonular fibres. The ciliary muscle regulates the form of the lens and enables accommodation. The iris rests on the anterior part of the lens, forming the pupil. The iris splits the eye into the anterior chamber between the cornea and the lens and the posterior chamber between the lens and the retina. The posterior eye chamber is filled with a gel-like substance, *corpus vitreum*. The *tunica interna bulbi,* or the retina, is the optical part of the eye. The photoreceptors: cones and rods are located in the posterior two-thirds, known as the *pars optica,* of the retina. The photoreceptors are specialized neuroepithelial cells able to absorb light photons that trigger a membrane potential in the retinal cells, which is transformed into nerve signals. See also Figure 4 for an overview of the anatomy of the eye.

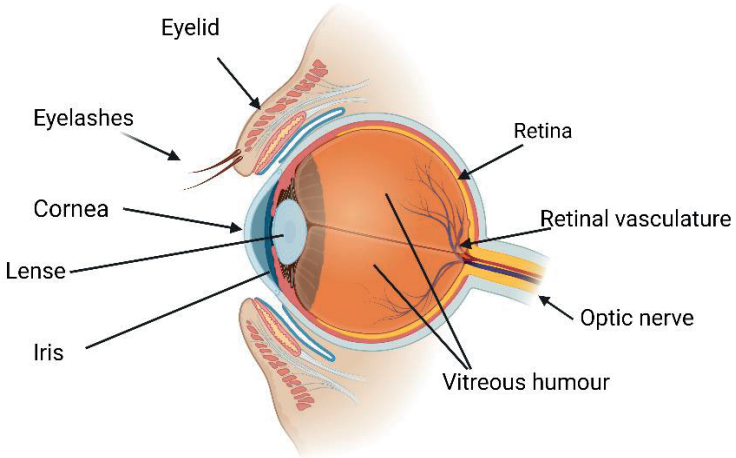


Figure 4. A sagittal view of the eye, the illustration is created in Biorender.com

The crystalline lens is a highly specialized, avascular, transparent, biconvex structure in the eye that focuses light onto the retina. On the outer surface, the lens is surrounded by the lens capsule. The anterior part of the lens is covered by the lens epithelium. At the equator of the lens, the epithelium divides and gives rise to the lens fibres. The lens fibres make up the largest bulk of the lens. They are long and thin, highly organised cells arranged in concentric sheets, from anterior to posterior, like layers of an onion. The

ends of the fibres meet at the pole-axis, forming the y-shaped suture lines. The strict organization of the lens fibres ensures transparency. In the centre of the eye, the fibres are thin and firm, forming the hard nucleus, while the fibres in the outer cortical layers are softer, allowing the lens to change shape during accommodation.

The eyelids are musculofibrous folds consisting of three layers: the inner mucous membrane or palpebral conjunctiva, a middle musculofibrous layer and the skin on the outer eyelid (Singh et al., 2018, pp 318-331). The outer eyelid is covered by epidermal tissue and a thin fur coat. The transition between the skin and conjunctiva forms the free eyelid margins. Along the eyelid margins, the openings of the meibomian, Moll and Zeiss glands can be seen. In front of these gland openings are the eyelashes, in dogs, primarily located from the medial quarter to the lateral cantus of the upper eyelid (Gelatt, 2011, pp 89- 111; Gelatt et al., 2013, pp 832-835). The eyelashes are usually the same colour as the coat on the eyelid. The function of the eyelid and eyelashes is to protect the eye.

Embryology of the eye

This section will give a brief introduction to the embryology of the eye, with an emphasis on the development of the lens and the eyelids. The first signs of the optic primordia emerge as paired depressions of the neural ectoderm proximal on the prosencephalon (McGeady et al., 2016, pp 298-312). As the depression becomes deeper, the optic vesicle becomes visible around day 15 of gestation in the dog (Aguirre et al., 1972). The optic vesicle elongates towards the ectoderm and the lens placode, and a thickening of specialised ectoderm invaginate into the neural ectoderm (Lovicu and McAvoy, 2005), forming the lens vesicle and the optic cup. As the lens vesicle buds off from the surface ectoderm, it consists of a single layer of cuboid cells enclosed by a basal membrane. The basement membrane later develops into the lens capsule (Danysh and Duncan, 2009). The polarity of the lens is determined by the primitive retina, and the cells on the posterior part of the lens start to elongate, forming the primary lens fibres that fill the embryonic lens nucleus

(Lovicu and McAvoy, 2005). In the transformation into the highly specialised lens fibres, proteins specific to the lens, like lens crystalline, are synthesized, and the organelles are degraded. At birth, the lens consists mainly of the lens nucleus, only surrounded by a small cortex. The anterior cuboid epithelial cells at the cortex continue to divide throughout life and differentiate into secondary lens fibres at the lens equator, where they elongate posterior and anterior around the embryonal nucleus, forming the Y sutures.

The eyelids are developed at the end of the embryonic period from the surface ectoderm (McGeady et al., 2016, pp 298-312). The ectoderm grows from each side of the corona and fuses in the middle, covering the developing cornea. Around 8-14 days post-natal, the eyelids separate again in puppies (Aguirre et al., 1972; Hyttel et al., 2009, pp 163-172). Eyelashes are developed from modified sebaceous glands along the eyelid margin (Hyttel et al., 2009, pp 163-172).

Ocular disorders

The eye can be affected by numerous disorders stemming from environmental and genetic factors. The latter are often enriched within specific breeds with confirmed or presumed hereditary origins (ACVO Genetics Committee, 2021). The following text will describe the two ocular disorders studied in this thesis.

Distichiasis

Distichiasis is an ocular condition with extra hairs appearing along the margin of the eyelid resembling a second row of eyelashes (Lawson, 1973). Thereby, the name distichiasis originating from the Greeks *di* (two) and *stichos* (rows). The *distichiae* can occur uni- or bilaterally and in both upper and lower eyelids (Lawson, 1973). Studies indicate that distichiasis is most prevalent in the upper eyelid (Zimmerman and Reinstein, 2019; Palella

Gómez et al., 2020). The hairs arise from ectopic hair follicles close to the sebaceous glands of the eyelid, usually emerging from the excretory duct openings of the meibomian glands on the eyelid margin as single or multiple hairs (Raymond-Letron et al., 2012; Palella Gómez et al., 2020). The displaced hairs can vary in size but are often thinner and softer than the normal eyelashes and can sometimes be difficult to detect (Lawson, 1973) (Figure *5*).

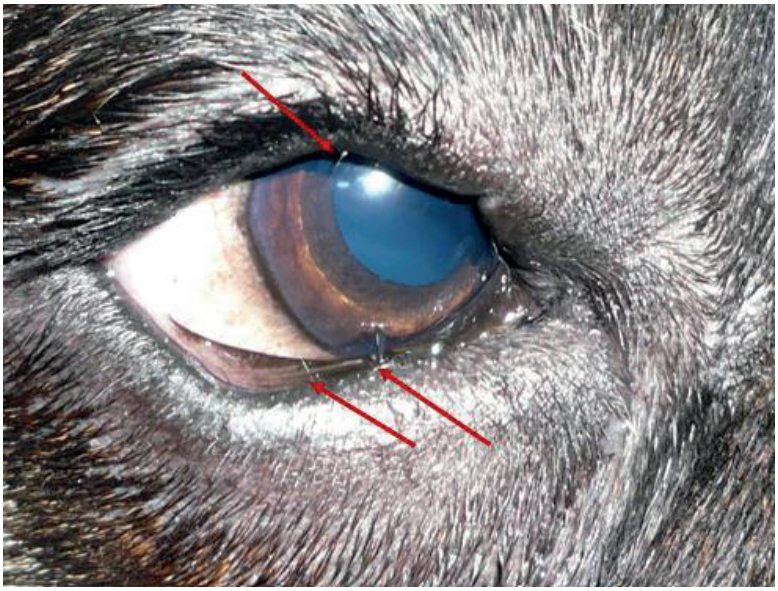


Figure 5. Distichiasis in a dog, the red arrows point at the aberrant eye hairs in the upper and lower eyelid. By Joel Mills - with permission to use under the Creative Commons licence CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=3438089

The clinical symptoms caused by distichiasis are due to physical irritation of the misplaced hairs directed towards the cornea. The symptoms are often mild (Jondeau et al., 2023). Still, they can lead to eye irritation, inflammation, excessive tearing and discharge, blepharospasm and conjunctivitis (Lawson,

1973; Gelatt, 2011, pp 89- 111; Gelatt et al., 2013, pp 840-843). In more severe cases, corneal ulceration and keratitis can be seen.

The management of the condition depends on the severity of the symptoms. Aberrant eye hairs causing eye irritation can be removed. The simplest form is the manual epilation of the hair. The *distichiae* can be permanently removed with surgical procedures, such as electroepilation and cryoepilation. Complications have been reported for all techniques, with the most common being incomplete removal and regrowth of hairs, fibrosis of the eyelid margin, entropion, and local loss of pigmentation (Gelatt, 2011, pp 89- 111; Zimmerman and Reinstein, 2019; Ioannides et al., 2020; Palella Gómez et al., 2020).

Distichiasis is a common condition in dogs, and it has been registered in both mixed and pure-breed dogs (Lawson, 1973; D'Anna et al., 2007; Raymond-Letron et al., 2012; Jondeau et al., 2023). The prevalence varies between breeds and populations (ACVO Genetics Committee, 2021; NKK, 2021a). For example, the prevalence of English cocker spaniels has been reported to be 49.31% in Denmark (Petersen et al., 2015) and 12.2% in a French study population (Jondeau et al., 2023). Due to the increased occurrence within some breeds, distichiasis is assumed to be an inherited condition with a genetic involvement. The heritability has been established to be moderate to high in the dog breeds Havanese (Bellamy et al., 2021), Elo (Kaufhold et al., 2006), and English cocker spaniels (Engelhardt et al., 2007; Petersen et al., 2015), but low in the Tibetan terrier (Ketteritzsch et al., 2004).

The condition is less common in other species than dogs. Distichiasis has been reported in ferrets (Verboven et al., 2014), cats (Reinstein et al., 2011), horses (Hisey et al., 2020) and cattle (Allais-Bonnet et al., 2013; Arteaga and Crasta, 2021). In cattle, distichiasis has been linked to the autosomal dominant *polled* locus on the bovine chromosome 1 (Allais-Bonnet et al., 2013). In Friesian horses, an association between distichiasis and an intergenic region on the equine chromosome 13 was identified (Hisey et al.,

2020). In Friesian horses, the inheritance is believed to be autosomal dominant with incomplete penetrance.

In humans, congenital distichiasis is a rare condition. It is often associated with a syndrome combined with lymphedema (Hoover and Kelley, 1971; Erickson et al., 2001; Brice et al., 2002; Brooks et al., 2003; Tavian et al., 2016). This syndrome has an autosomal dominant inheritance associated with mutations in the *FOXC2* gene. Distichiasis has also been seen together with the rare syndromes Blepharocheilodontic syndrome, associated with mutations in *CTNND1* and *CDH1* gene (Kievit et al., 2018), and facial dermal dysplasia linked with a mutation in the *TWIST2* gene (Cervantes-Barragán et al., 2011).

Despite being a common condition in dogs, to the author's best knowledge, no study concerning the molecular genetic background of distichiasis in dogs has been published.

Cataract

Cataract is an ocular condition where the lens becomes opaque, leading to reduced transparency and potentially impaired vision (Shiels and Hejtmancik, 2019). The previously described strict organization of the lens fibres is essential for the lens's transparency. Any disruption of this structure can cause cataracts. Because the mature lens fibres lack most organelles and a nucleus, the response to damage is limited. Cataracts can be caused by various factors, including genetic mutations, trauma, inflammation, metabolic disorders, nutritional deficiencies, exposure to toxins, and old age. Cataracts are often progressive and commonly seen with increasing age.

Cataracts are found to be the most common cause of blindness in humans (Burton et al., 2021; Steinmetz et al., 2021). According to the World Health Organization, by October 2022, 94 million persons were affected by cataracts (WHO, 2022).

Cataract in humans is often a complex disorder that arises due to the intricate interplay between genetic and environmental risk factors.

However, a wide range of mutations inherited as monogenetic disorders have been associated with cataracts. Autosomal dominant inheritance is the most common form in congenital and juvenile hereditary cataracts, but X-linked and autosomal recessive inheritance are also described (Shiels et al., 2010; Cat-Map, 2023). More than a hundred genes have been associated with cataracts in humans, including genes encoding membrane proteins, cytoskeletal proteins, crystallines and transcription factors. Generally, gene mutations affecting the critical structure of lens proteins, such as crystalline, often cause congenital and juvenile cataracts. In contrast, age-related cataracts often are due to mutations causing increased vulnerability to damage caused by environmental factors such as light, hyperglycaemia and oxidative stress (Shiels and Hejtmancik, 2019).

In dogs, cataract is a common ocular condition and a leading cause of blindness. The prevalence varies greatly between breeds and populations (Gelatt and MacKay, 2005; Engelhardt et al., 2007; Park et al., 2009; Donzel et al., 2017; Kristiansen et al., 2017). Numerous different forms of cataracts can occur in dogs, and various classification schemes exist to categorize them (Gelatt et al., 2013, pp 1206-1208; Maggs et al., 2018, pp 312-322). These schemes may be based on the location of the cataract within the lens, the age of onset, the underlying causes, as well as the severity and progression stage of the condition. The clinical signs of cataracts vary from subtle changes with little impact on the dog to severe cataracts affecting the entire lens and resulting in impaired vision. Lens-induced uveitis can be seen in the more severe forms due to leakage of lens proteins. Treatment for cataracts depends on the symptoms. In severe cases, surgery is recommended (Gelatt, 2011, pp 305-330).

Certain dog breeds are predisposed to cataracts, and two genes are found to be associated with cataracts in dogs. *HSF4* is associated with early-onset cataracts in Boston terriers, Australian shepherds, and French bulldogs and primary cataracts in Staffordshire bull terriers (Mellersh et al., 2006; Mellersh et al., 2007; Mellersh et al., 2009). Juvenile cataracts in Wirehaired Pointing Griffon dogs have been associated with the *FYCO1* gene (Rudd Garces et al., 2022). A GWAS found an association between cataracts and a

locus on chromosome 13 in Australian shepherds, but no further information has been published (Ricketts et al., 2015). Recently, a preliminary GWAS identified two loci on CFA20 and CFA21 associated with posterior polar cataracts in Havanese (Bellamy and Lingaas, 2023).

Pulverulent nuclear cataracts

Nuclear cataracts are those cataracts that only affect the central part of the lens, the nucleus. Pulverulent nuclear cataracts (PNC) are characterized by the presence of a fine powdery appearing opacification in the nucleus.

At least ten different genes have been discovered to be associated with PNC in humans (Shiels et al., 2010; Cat-Map, 2023): *CRYBB1* (Meyer et al., 2009), *CRYBA1* (Bateman et al., 2000; Lu et al., 2007), *CRYGC* and CRYGD (Kumar et al., 2011), *GJA3* (Ding et al., 2011), *GJA8* (Arora et al., 2008; Yan et al., 2008), *MIP* (Wang et al., 2010), *MAF* (Jamieson et al., 2002), *LIM2* (Berry et al., 2020), and *PIKFYVE* (Mei et al., 2022).

Pulverulent nuclear cataracts in the Norwegian Buhund

Bjerkås and Haaland (1995) described the initial changes of PNC in Buhund as small opacities along the suture line behind the nucleus. With time, these opacities gradually advanced to affect the entire foetal nucleus (Figure 6). An autosomal dominant inheritance with a high degree of penetrance was assumed. The prevalence among the examined dogs was around 50%. In 2009, “the buhund project” was initiated as a collaboration between the Nordic buhund associations, Hønefoss Dyrehospital and the Medical Genetics Unit at the Veterinary Faculty, NMBU (Norsk Buhundklubb, 2020). The project's main goal is to develop a genetic test for cataracts in the buhund (Kristiansen et al., 2017). The high prevalence of PNC persisted despite the breed club's increased focus on the disease. Additionally, Kristiansen et al. (2017). found a correlation between increasing age and a positive PNC diagnosis, indicating that young dogs initially classified as unaffected might develop the disease at an advanced age.

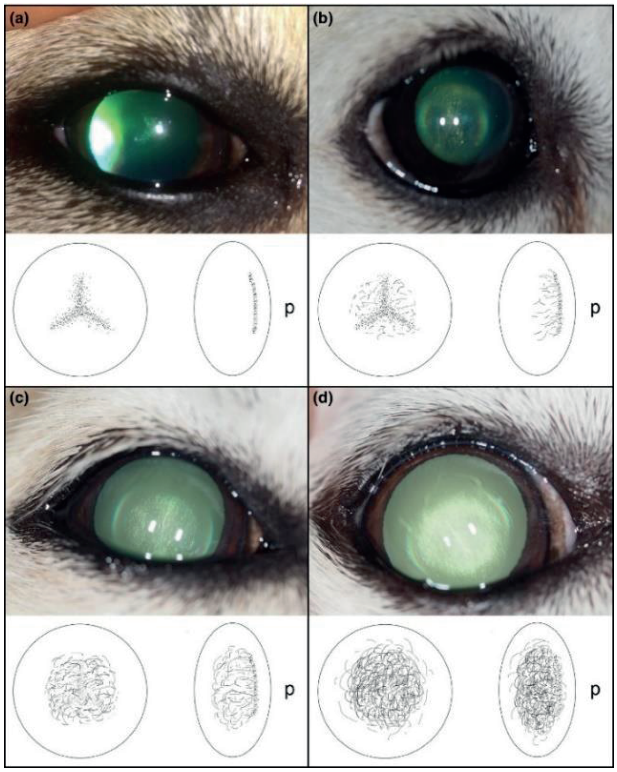


Figure 6. Each facet shows a different stage of PNC. A frontal view photograph is on top and a blueprint of the lens changes (frontal and transversal views) at the bottom. a) minimal changes in a two-year-old dog, b) mild changes in an eleven-year-old dog, c) moderate lenticular changes in a ten-year-old dog, d) pronounced changes in a ten-year-old dog. The picture is from the article by Kristiansen et al., with permission to reuse from John Wiley and Sons.

5.2 Aims

The overall aim of this thesis was to improve the knowledge about the genetic background of common ocular disorders in the dog breeds, the Staffordshire bull terrier and the Norwegian buhund. Additionally, we aimed to provide basic knowledge that might have the potential to be helpful in future breeding programs.

Sub goals:

- Estimate the heritability of distichiasis in Staffordshire bull terrier.

- Identify genomic regions associated with distichiasis in Staffordshire bull terriers and investigate the possibility of using genomic data to predict a complex disorder such as distichiasis in a dog breed.

- Identify genomic regions associated with PNC in Norwegian buhund and evaluate if genetic variants in the region can potentially be used for risk estimation.

5.3 Materials and Methods

5.3.1 Material

Paper I and II

The genetic material in papers I and II consists of DNA samples from 893 dogs with eye examination data. Illumina CanineHD Whole-Genome Genotyping BeadChip (Illumina, 2010) was used for genotyping the dogs. We genotyped 775 dogs with Illuminas CanineHD BeadChip containing approximately 220,000 markers. Additionally, 118 dogs from a previous study genotyped using the Illumina CanineHD BeadChip with around 170,000 SNP markers were added. The genotyping was carried out by Neogene, USA. The samples were collected at the biobank of the Medical Genetics Unit at the Veterinary faculty-NMBU. All samples were collected with the owner's written consent and in agreement with all applicable ethical guidelines. DNA was extracted from EDTA-blood using E.Z.N.A.® Blood DNA Mini Kit from Omega Bio-Tek, Norcross, USA. The quality of the DNA was measured using Epoch from BioTek. After isolation, the DNA was stored at - 20 ˚C until further analysis.

Pedigree information and eye examination records, including all diagnostic data, were provided by NKK and are openly available at the NKK's database "dogweb" (www.dogweb.no) (NKK, 2021a). Only dogs with a secure distichiasis diagnosis after an examination performed by an ECVO-certified veterinarian were included in the studies, and genotyped dogs without a secure distichiasis diagnosis were not included.

Paper III

DNA samples for the Norwegian buhund PNC project were collected from the biobank at the Veterinary Faculty, NMBU. To increase the number of DNA samples from unaffected dogs, we collected additional samples at dog shows and through direct recruitment of dog owners via telephone. All new DNA samples were collected using DNA-cheek swabs (Performagene™; DNA

Genotek Inc). All samples were collected with the owner's consent. DNA from EDTA blood was isolated with the previously described approach for papers I and II. DNA from the buccal swabs were isolated as recommended by the manufacturer.

Two hundred and three samples were initially genotyped on Illuminas CanineHD BeadChip with 170.000 markers, while Illuminas CanineHD BeadChip with 220.000 SNP markers was used for the last 83 samples. After genotyping, strict quality control was performed, and low-quality markers and samples were removed, as well as dogs without a secure PNC diagnosis. The final dataset consisted of 100662 SNP markers and 160 dogs, where 121 dogs were affected with PNC and 39 dogs fulfilled the requirements as unaffected.

PNC diagnosis was confirmed after eye examinations performed by an ECVO-certified eye examiner. The records were registered in the NKK's open database, "Dogweb", or the breed club's register. Dogs were either classified as affected or unaffected. To be classified as unaffected in our study, the dog must have undertaken at least one eye examination by an ECVO-certified examiner after the age of seven with no signs of PNC.

5.3.2 Methods

Statistics

The descriptive statistics for all three papers were performed using R (R Core Team, 2021). In paper I, R version 4.2.0, and in papers II and III, R version 4.2.3 were used. Post-processing of genomic analysis was performed in R. All plots were conducted in R.

The heritability estimates

The pedigree-based heritability estimates in paper I were based on 1391 dogs with an eye examination record after the age of one year. The youngest age group (puppies under eight weeks) was excluded due to the low

prevalence, while dogs between eight and 52 weeks were not included in the analysis because of a limited number of observations in that age range. We used a mixed model:

$$Y = \mu + a + e.$$

Where Y = distichiasis status, μ = the mean term, a = the additive genetic effect, and e = the residual error. The data was analysed using an REML approach in the DMU package version 6, release 5.3 (Madsen and Jensen, 2013). The heritability was converted to a theoretical underlying continuous scale using the methods of Dempster and Lerner (Dempster and Lerner, 1950).

The genomic heritability estimates were based on a subgroup of dogs (N=498) from the pedigree-based heritability estimates. These dogs were selected because they were included in the pedigree-based estimate and had DNA available from the biobank.

Strict quality control was performed prior to the analysis. For further details, see paper I. A mixed linear model was used to estimate the genomic heritability using a genomic restricted maximum likelihood approach in the Genome-wide Complex Trait Analysis software (GCTA) version 1.93.2. (Yang, 2023). The mixed model was similar to the model used for the pedigree-based estimates:

$$Y = \mu + g + e$$

Y = distichiasis status, μ = the mean term, g = the animal effect based on GRM, and e = the residual error. The heritability estimates on the observed scale were transformed to a modified underlying scale (Yang et al., 2011). Two different GRMs were used in the REML, one calculated in GCTA and one in Linkage Disequilibrium Adjusted Kinship software (LDAK) version ldak.5.1.Linux (http://dougspeed.com/downloads). The main difference is how the programs deal with LD between SNPs. In LDAK, SNPs are weighted based on the amount of LD in the region (Speed et al., 2012), while in GCTA, all SNPs are assumed to contribute equally (Yang et al., 2011). Ten different models were initially evaluated, adding alternate fixed effects like sex, year

of birth, year of examination, age at examination, and the interaction between sex and year of birth and age at examination. None of the fixed effects affected the heritability estimates and were left out in the final model in paper I.

Genomic prediction

Two approaches were used in the genomic prediction: GBLUP in GCTA version 1.93.2 (Yang et al., 2011) and a Bayesian mixed model in BayesR software version 01/04/2021 (Moser et al., 2015). BayesR fits the model:

$$y = 1_n\mu + XA + e$$

Where *y* is a vector of the phenotypes, 1_n is a n-dimensional vector of ones, μ is the general mean term, *X* is a matrix of the genotypes of the markers, *A* is a vector of the effect of the SNPs, and *e* is the residual errors (Moser et al., 2015). BayesR includes a mixture of four classes of SNP effect (nulls, small, medium and large effect). The following distribution was used in paper II: null $N(0,0 * \sigma_g^2)$, low $N(0,0.0001 * \sigma_g^2)$, medium $N(0,0.001 * \sigma_g^2)$ and high effect $N(0,0.01 * \sigma_g^2)$. A Markov chain Monte Carlo (MCMC) method was used to draw samples from the posterior density. to find the SNP effect

In contrast to BayesR, GBLUP uses one normal distribution of SNP effects $N(0, A\sigma_g^2)$. GBLUP used the following mixed model:

$$y = \alpha + g + e$$

Where *y* is a vector of the phenotype distichiasis, α the mean term, *g* equals the genetic value, and *e* is the residual error. The values of *g* and *e* were estimated from the two equations: $\hat{g} = VgAV^{-1}y$ and $\hat{e} = VeV^{-1}y$, where A is the GRM, *Vg* is the genetic variance, and *Ve* is the residual variance, V is the variance matrix of the records y, and calculated as $V=AV_g+IV_e$ (Yang et al., 2011; Yang, 2023).

Before the genomic prediction, strict quality control was performed as described in the paper and full siblings were removed. The dataset comprised 607 dogs (248 cases and 359 controls) and 94697 SNPs.

The two methods, BayesR and GBLUP, were compared using a six-fold cross-validation and calculating the area under the curve (AUC) from a receiver operating characteristic curve (ROC curve) in the R package pROC (Robin et al., 2011). To compare the GV, the odds ratio between the 25 % of dogs with a GV indicating that they were at highest risk and the 25% of dogs with a GV indicating that they were least likely to develop distichiasis were calculated.

GWAS

The genome-wide association studies (GWAS) for papers II and III were conducted in GCTA (Yang et al., 2011). In addition, a GWAS in paper II was conducted using the Bayesian model, BayesR (Moser et al., 2015).

GCTA uses an MLMA, testing one SNP at a time by using the following model:

$$y = \alpha + \beta X + g + e$$

Where y is a vector of the different phenotypes (distichiasis and PNC), α is the general mean term, β the additive genetic effect of the SNP (fixed effect), X is the genotype of the SNP, g equals the effect from the GRM (random effect), and e is the residual error. The GRM was included to adjust for a potential population structure and relationship.

The BayesR GWAS approach in paper two employed the same mixture model used in the genomic prediction, estimating the absolute SNP effect for distichiasis.

Sanger sequencing

Sanger sequencing of all coding sequences of the four candidate genes in paper III was performed on four cases and four controls with the 3500xL Genetic Analyzer from Applied Biosystems, Life Technologies, Thermo Fisher Scientific. Primers were designed in Primer3Plus (Untergasser et al., 2007) and were based on the Can.Fam4 reference genome (Wang et al., 2021) using in the University of California, Santa Cruz genome browser, http://genome.ucsc.edu (Kent et al., 2002). Sequences were analysed in Sequencher™ 5.1 software from Gene Codes Corporations. The odds ratio

between alternative alleles was calculated in R version 4.2.3. (R Core Team, 2021).

5.4 Summary of papers

Paper I: Heritability estimates of distichiasis in Staffordshire bull terriers using pedigrees and genome-wide SNP data

Distichiasis is the most prevalent eye disorder in the Norwegian Staffordshire bull terrier population, with an overall prevalence of 8.38% (Confidence interval, CI 7.56-9.26). The purpose of the study was to evaluate the reliability of eye examinations in young dogs and to estimate the heritability of the disorder.

We observed a significant difference in the prevalence of distichiasis in puppies examined at around eight weeks of age and dogs examined after one year of age. The puppies examined at around eight weeks had a prevalence of 2.86% (CI 2.18 - 3.37), while the dogs examined after 52 weeks (> one year of age) had a prevalence of 18.22% (CI 16.71- 20.87).

We, therefore, compared a group of dogs examined both around eight weeks and after 52 weeks and found a significantly increased risk of being diagnosed with distichiasis after 52 weeks.

The heritability was estimated with two methods: a pedigree-based approach and a genomic-based approach. The heritability estimates from the pedigree data (1391 dogs) was $h^2_{ped0} \sim 0.22$ (SE 0.05) on the observed scale and, after transforming to the underlying scale, $h^2_{pedT} \sim 0.48$ (SE0.11). The genomic-based heritability estimates were calculated with the two different GRMs, each respectively calculated in either GCTA or LDAK. Two models were used, the first excluding full siblings and the second including close relatives. A summary of these genomic-based heritability estimates is represented in Table 1. The two GRMs calculated in GCTA and LDAK gave similar results for each model. The model including full siblings gave higher heritability estimates than the model excluding full siblings. The dataset excluding full siblings consisted of 498 dogs (207 affected and 291 unaffected), while the dataset including all dogs comprised 548 dogs (228 were affected and 320 unaffected),

Table 1. Genomic-based heritability estimates of distichiasis in Staffordshire bull terriers.

Model	GRM	Number of dogs	Heritability estimates observed scale (h^2_{gO})	Heritability estimate transformed scale (h^2_{gT})
Full siblings with equal affection status excluded	GCTA	498	0.275(SE 0.082)	0.368 (SE 0.109)
Full siblings with equal affection status excluded	LDAK	498	0.281 (SE 0.083)	0.375(SE 0.112)
All dogs included	GCTA	548	0.345 (SE 0.078)	0.461 (SE 0.104)
All dogs included	LDAK	548	0.357 (SE 0.079)	0.476 (SE 0.106)

The heritability estimates are estimated for the different models, including and excluding full siblings. Two different genomic relationship matrices (GRM) were used, calculated in GCTA and LDAK. SE is Standard Error. The table is a modified version of table 2 in paper I, Heritability estimates of distichiasis in Staffordshire bull terriers using pedigrees and genome-wide SNP data.

Our findings revealed a significant increase in disease prevalence from eight weeks to one year of age. This suggests that screening young puppies around eight weeks may not yield an accurate diagnosis and could underestimate the prevalence of disease in the breed. The heritability estimates suggest that distichiasis has a moderate to high heritability, indicating good possibilities of reducing its prevalence through selective breeding. It is noteworthy that genome-based heritability estimates are comparable with pedigree-based estimates despite a relatively small sample size and higher standard errors.

Paper II: Genomic analysis and prediction of genomic values for distichiasis in Staffordshire bull terriers

The primary objective of this study is to explore the genetic basis of distichiasis in Staffordshire bull terriers using a GWAS. We also assessed two approaches for genomic prediction, namely GBLUP and BayesR. We used the results to explore the potential usefulness of utilizing genomic data in predicting genomic values for distichiasis in Staffordshire bull terriers.

The GWAS was conducted on a material consisting of 731 genotyped Staffordshire bull terriers, where 324 dogs were affected with distichiasis and 407 dogs were unaffected. Two approaches were used in the association study: an MLMA in GCTA and a Bayesian model in BayesR software. Four genomic regions were associated with distichiasis on chromosomes: CFA1, CFA18, CFA32 and CFA34 in MLMA, (and three regions; CFA1, CFA18 and CFA34 in BayesR). The cumulative risk allele load was significantly different between affected and unaffected dogs. On average, affected dogs carried one more risk allele than the unaffected dogs.

Genomic values were predicted for a subset of the material, counting 607 dogs (248 affected and 359 unaffected). The genomic prediction estimated using GBLUP and BayesR showed a significant difference in the estimated breeding values between affected and unaffected. By comparing the 25% of dogs with genetic values most likely to develop distichiasis with the 25% of dogs with genetic values least likely to develop distichiasis, the first groups had a ~4.0 times increased risk of developing distichiasis. There was no significant difference between the two methods used, and the AUC was 0.651 (CI 0.607-0.695) in BayesR and 0.655 (CI 0.612-0.699) for GBLUP.

The four associated genomic regions imply that distichiasis in Staffordshire bull terriers is a complex trait influenced by multiple genetic loci. Nevertheless, further research is crucial to validate and strengthen these results.

The genomic prediction shows promise in aiding selective breeding to reduce distichiasis prevalence in Staffordshire bull terriers. However, its accuracy in predicting individual phenotypes is relatively low.

Paper III: A genome-wide study identifies a region on CFA37 associated with cataract in the Norwegian buhund

In the third paper, we aimed to find a causal variant associated with PNC in the Norwegian buhund. We conducted a GWAS involving 184 dogs affected by PNC and 39 dogs that tested negative for PNC upon eye screening after reaching seven years of age. Using an MLMA in GCTA we identified a significant region on CFA37 associated with PNC. The significance level was set to $5.09x10^{-07}$, in accordance with the Bonferroni correction, which involves dividing 0.05 by the total number of SNP used in the model. The significant region spanned a distance of about 6 Mb. The region enclosed ~50 genes, including five candidate genes that have been shown to cause PNC or other forms of cataract in humans and other species: *PIKFYVE* and the four gamma crystalline genes *CRYGA, CRYGB, CRYGC, CRYGD.* We also performed a risk assessment of five out of the six most significant SNPs. We found a positive correlation between the number of risk alleles and developing PNC. Carrying five or more risk alleles gave an odds ratio of 8.25 (95% CI 3.73 - 19.47).

The exons of the four gamma crystalline genes were sequenced using Sanger sequencing technology. No causal variants were disclosed within the exons of the four genes. A single SNP in the second exon of *CRYGD* exhibited a distinct distribution pattern between affected and unaffected individuals, with an odds ratio of around six for carrying the risk variant. However, the variant was found to be benign.

Currently, no causal variant has been identified. Additional sequencing of the region may uncover a putative variant in the region. SNP markers in the associated region may be used for marker-assisted selection but must be verified in an independent sample.

5.5 Discussion

This thesis focuses on two distinct eye disorders that differ both in their physical manifestations and underlying genetic foundations. While distichiasis is a polygenic trait, PNC may be influenced by a major locus with a large effect on the phenotype. These main findings are discussed in detail in the three papers. This section contains an overall discussion, including some methodological and ethical considerations.

Evaluation of the diagnostic data

Eye examination data gathered by the breed clubs, as part of the official eye screening programs, is available from the Norwegian kennel club and presents valuable data to study hereditary eye disorders in dogs. However, it is essential to assess the reliability of the diagnostic data and phenotypes since misclassification of dogs can cause serious bias in the results. An incorrect diagnosis can affect the heritability estimates, reduce the accuracy of the genomic prediction and lead to a lack of power in a genetic study such as GWAS.

According to dogweb, many Staffordshire bull terrier puppies underwent an eye examination before leaving the breeder at about eight weeks. We, therefore, first evaluated whether this early examination provided an accurate "lifelong diagnosis" to find an optimal examination time. As the results show in paper I, there was a large discrepancy between the prevalence among the youngest dogs and those examined after one year of age (> 52 weeks). Our findings revealed a high level of uncertainty associated with a negative distichiasis diagnosis in young puppies, and diagnostics conducted after the dogs reach one year of age are more reliable. Including the eye examination records of young dogs (<1 year; <52 weeks) can grossly underestimate the prevalence if the uncertainty of the diagnosis is not considered.

In the third paper, the study of PNC in the Norwegian buhund, the cut-off age for controls was set to seven years of age. The cut-off age in the controls was based on the findings from a previous study by Kristiansen et al. (2017),

who found a positive correlation between increasing age and the development of PNC. The high cut-off age made it challenging to obtain a large number of controls. This difficulty is partly due to the high prevalence of the disease in the breed but also due to a lack of diagnostic data on dogs after seven years of age. As a routine, most dogs are examined at a young age and before breeding. Because the number of dogs bred after age seven is low, there is a lack of eye examination records for older dogs.

Both studied traits, PNC and distichiasis, exhibit a varying degree of severity. While some dogs are mildly affected, others are more severely affected, with symptoms appearing at an early age. We have defined both of the studied traits as binary, affected or unaffected. This was done because there was an inconsistency in the grading of the eye examination records. An accurate grading could have been beneficial because the transformation from a continuous scale to a binary scale can have led to a loss of some nuances. For example, it would be interesting to know whether homozygotes for the identified risk variants are more severely affected by the disorders than heterozygotes.

Heritability estimates

We used two approaches to estimate the heritability of distichiasis: a traditional pedigree-based and a genomic-based approach. Both methods were based on a mixed linear model using a REML to determine the variance component. REML is a well-established method that can handle complex relationships and allows for the estimation of heritability by considering the covariance between relatives and accounts for the different types of relationships within the pedigree (Falconer and Mackay, 1996, pp 163-184; Lynch and Walsh, 1998, pp 779-803).

Adjustment for relevant environmental factors may increase the precision of the estimates. A Danish study found that female dogs were more prone to develop distichiasis than males (Petersen et al., 2015). However, we found no effect of sex after correcting for age in our material, and sex did not affect the heritability estimates. Sex was, therefore, left out of the final model. We

also initially tested other covariates, such as birth year, to look at potential changes in the population or environmental factors over a timespan. There were no significant differences in the heritability estimates, including these covariates, and they were not included in the final model.

Pedigree-based heritability estimates in human studies are often higher than genomic heritability estimates (Yang et al., 2010; Mills et al., 2020, p 3-30). The discrepancy between the pedigree-based and genomic-based heritability has been suggested to be caused by the lack of linkage between a marker and the causal variants or due to low allele frequency of the causal variants where rare variants are only carried by a few individuals (Yang et al., 2010; Wainschtein et al., 2022). Because of the high levels of LD and the low heterogeneity within a dog breed, it can be comparatively easier to obtain more accurate genomic-based heritability estimates in dogs than in humans. In our estimates, we observed that the genomic-based estimates were close to the pedigree-based estimates. This agrees with other heritability estimates in canine complex disorders (Baker et al., 2017).

A challenge with SNP-based estimates is that they can be affected by differences in allele frequencies between different subpopulations. Appropriate methods, such as PCA, can be incorporated to address this issue. Our estimates came in large part from one Staffordshire bull terrier population, and no obvious population structure between dogs from different countries was observed in our analysis; also, the inclusion of PCA in the model did not affect the results, indicating that the effect of the population structure captured in the PCA was minimal.

One advantage of the use of genomic data compared with pedigree-based is that there is no need for extensive, detailed pedigree data. This can be advantageous in human population studies and wildlife populations, where pedigree information can be hard to obtain. A lack of detailed pedigrees is less of a problem in dogs as most dog breeds keep detailed stud books. However, the lack of detailed ancestry records can pose a challenge, for instance, in populations with numerous imported dogs, such as the Norwegian Staffordshire bull terrier population. Another challenge in

pedigree-based heritability estimates can arise due to registration inaccuracies, undiscovered mating occurrences or dishonest registration of matings (Leroy, 2011). The estimated pedigree error rate ranges between 1% and 9%, depending on the dog breed (Leroy et al., 2012). Such errors can be surpassed using genomic-based heritability estimates.

Another benefit of genomic-based heritability estimates is the possibility of estimating the heritability among unrelated individuals. This can reduce the bias of shared environmental factors and, to a greater extent, reduce the impact on epistasis and dominance. Due to limited population sizes within a dog breed, finding a large number of unrelated individuals with a phenotype can be challenging as most dogs are somewhat related within a dog breed. Our materials included some sibling pairs, which could potentially bias the heritability estimate upwards due to shared early-life environments, epistasis and dominance. Our genomic heritability estimates, including full siblings (0.47), are higher than those without (0.37). In the pedigree-based heritability estimates, full siblings were included. In retrospect, looking at the potential effects of including litter and breeder as covariates in this model would have been interesting. Including these covariates could have aided in identifying the reasons behind the variations in the estimates, both when including or excluding full siblings, whether the differences in estimates are attributed to shared environments, epistasis, dominance, or mere chance. An alternative to correct for close relationships in heritability estimates is to include pedigree information or a relationship matrix from genomic data in the model, as proposed by Zatilen (Zaitlen et al., 2013).

We employed two distinct GRMs for the heritability estimates: one generated using LDAK and another using GCTA. We found it intriguing to incorporate the GRM from LDAK because it takes into account the LD between SNPs and is claimed to enhance the accuracy compared to the GRM derived from GCTA (Speed et al., 2012). Given the notable levels of LD prevalent within a dog breed, we hypothesized that LD might exert an influence on the heritability estimates. However, in our material, the two GRMs preformed very similarly. It is worth noting that other studies have

reported slightly higher heritability estimates in dogs when employing an LDAK-generated GRM (Cook et al., 2020).

The high estimate of heritability shows that it is possible to improve the trait through breeding and is important for developing breeding schemes and conducting further genetic studies and risk assessments. Additionally, the heritability level establishes an upper limit on potential findings in genetic studies, such as in the GWAS in paper II. Moreover, it sets a limit on the precision of predicted breeding values.

GWAS

In the second and third papers, we employed a genome-wide association approach to identify genomic regions associated with the two disorders, distichiasis and PNC. As discussed previously, due to the special features of the canine genome, associated regions can often be discovered in relatively small materials.

In the second paper, we investigated the complex disorder distichiasis. Four genomic regions were identified through the association analysis using MLMA in GCTA, and three of the same genomic regions were identified in BayesR. Although significant, none of the four regions received a very high significance level in the MLMA in GCTA. The lack of highly significant findings indicates that the study is under powered. Other plausible reasons for the lack of highly significant regions include causal variants with small effect sizes, genomic heterogeneity, and challenges with diagnostic misclassification.

Even though a low genetic heterogeneity and possibly accumulation of disease alleles within a dog breed may be expected, the collection of a large number of dogs is sometimes difficult due to factors such as limitations in population size, a lack of dogs with well-documented phenotypes, and limited research funding. Expanding the material to include dogs from other countries could be beneficial if a similar eye examination is performed as the standardized ECVO Eye scheme.

Combining data from multiple breeds can also increase the sample size. However, some of the benefits of studying dogs diminish. Two different dog breeds may exhibit different linkages to a causal variant. Phenotypic heterogeneity presents another challenge. A similar phenotype does not necessarily have the same genetic background. Therefore, utilizing two closely related breeds might offer more advantages than employing two breeds that are phylogenetically distant from one another. By including different dog breeds, the distinct genetic backgrounds and population structures must be considered, as population stratification can introduce confounding effects and spurious associations if not adequately addressed.

In our study, most dogs affected by distichiasis were only mildly affected. Comparing more severely affected dogs with unaffected dogs may possibly exhibit larger genomic differences. However, severely affected dogs were scarce in our material, and the grading was inconsistent; this added uncertainty to the use of the grades, and thus, we decided to use a binary model.

The third paper demonstrates some of the benefits of using pedigree dogs for genetic studies. Only a small sample size was sufficient to detect a strong association with variants in a region with a large effect on the phenotype. The identification of the genomic region associated with PNC was obtained using only 39 controls and 121 cases. This is consistent with the sample size suggested by Karlsson et al. to detect a monogenetic dominant disease in dogs (Karlsson et al., 2007). Nevertheless, paper III also sheds light on another challenge in canine genetic studies: The high level of LD can give rise to associated regions spanning several million base pairs, making it challenging to pinpoint the specific causal variant.

The Sanger sequencing of the exons of the four crystalline genes did not disclose any causal variants. Although the four genes are highly relevant, it remains to see if a causal variant resides outside the exons, for example, in regulatory elements. There might also be undetected structural variations in the region. We have, therefore, initiated high-quality WGS to be able to study the candidate region on CFA37 in detail. The fifth candidate gene, PIKFYVE,

was not sequenced via Sanger sequencing, first of all since the association between PNC and *PIKFYVE* reported by Mei et al. (2022) was published after the candidate gene study included in paper III was completed. Furthermore, *PIKFYVE* is a lengthy gene with 42 exons and multiple splice variants, making it more efficient to incorporate the study of this gene into the WGS analysis.

With the identification of causal variants, it can be possible to use this information to develop methods for genetic testing. It can also give a deeper insight into the biology of a pathological process and make it possible to design optimal strategies for precision medicine, where individuals are treated not only for a clinical disease but according to the presence of specific disease-causing mutations (Collins et al., 2016; Jorde et al., 2016 pp 292-300). Another application is to use identified genomic regions in genomic prediction.

Genomic prediction

In the second paper, we looked at the genomic prediction of distichiasis in Staffordshire bull terriers. Genomic prediction is in its infancy as a support for dog breeding, especially with the recognition that many of the most common disorders in canines are complex traits. The reduced costs and advancements in DNA sequencing technology make large-scale genomic data easier available, and genomic prediction may become an attractive tool. Given the low genetic diversity and long ranges of LD within a dog breed, the dog should be well suited for genomic prediction.

We compared two approaches to predict genomic values from SNP data, GBLUP and the Bayesian method BayesR. Bayesian approaches usually achieve higher accuracy than GBLUP (Meuwissen et al., 2001; Kjetså et al., 2022), particularly in traits with high heritability and traits with alleles with large effect size (Moser et al., 2015; Meher et al., 2022). GBLUP generally performs well for traits influenced by numerous variants with minor effects following a normal distribution, with most variants having small effect sizes on the trait (Meher et al., 2022). Considering the moderate to high

heritability and the identification of three to four associated variants with a moderate to large effect on the distichiasis phenotype, we expected BayesR to perform better than GBLUP in our data. However, in our study, both methods gave a very similar result. As discussed in paper II, the limited sample size could be a contributing factor, as the Bayesian model may not effectively distinguish between variants of small effect size and variants with no effect on the trait. Another contributing factor, as discussed by Kjetså et al. (2022), is that in datasets where numerous markers exhibit strong LD with a major effect allele, multiple markers can collectively account for the effect and resemble the distribution of SNP effects assumed in GBLUP. Given the high LD observed within the studied dog breeds, this could be one of the reasons why GBLUP performed well in our study.

We utilized a six-fold cross-validation technique to estimate the AUC of the ROC curve for comparison of the GBLUP and BayesR methods. Cross-validation is a well-recognized method to evaluate the ability of a test to discriminate between affected and unaffected individuals (Wray et al., 2010). Although the AUC gives a good measurement for comparing the two methods, it does not provide a complete picture of the accuracy of the estimates. The accuracy also depends on the trait's genetic architecture, heritability and prevalence. Also, the effective population size and structure will affect the accuracy (Kemper and Goddard, 2012; Wray et al., 2019).

The AUC estimated in our data was ~ 0.66; this is lower than the AUC predicted for distichiasis by Thorsrud et al. (2022)., ~0.96. However, their study population of guide dogs was very different from our Staffordshire bull terrier population. For example, the disease prevalence was much lower in their population (4.9%) compared to the Staffordshire bull terrier population (18.7%). A lower prevalence has been shown to increase the accuracy of genomic prediction (Wray et al., 2010).

Because complex disorders are affected by multiple factors, estimates of total risk can be improved by adjusting for environmental factors that affect the diseases (Baker et al., 2020). As previously discussed, in our study, like in many other studies, there were no known environmental factors to

include. Further studies of possible confounding environmental factors could probably provide useful knowledge and more accurate results.

The extensive use of imported dogs among Staffordshire bull terrier breeders has led to many incomplete pedigrees. This frequent use of foreign dogs in breeding can lead to a higher genetic variability within the population and less accurate predictions. However, it can also make the results valid across additional Staffordshire bull terrier populations in other countries, particularly when there is a substantial degree of interchange of breeding dogs. It has previously been shown that including closely related dog populations from different countries can be beneficial for the estimation of traditional breeding values (Arvelius and Klemetsdal, 2013).

As a diagnostic tool for the individual dog, the estimated GV was not very accurate, which is a disadvantage since dog breeders, first and foremost, are concerned about the accuracy of the estimates in the individual dog. Nevertheless, on a population level, it can rank animals according to their genetic merits and thereby help reduce the disease prevalence in the same manner as traditional breeding values estimated from pedigrees. Comparing the performance of pedigree-based predictions with the genomic-based predictions of distichiasis would have been intriguing.

It would also be interesting to assess the predictive ability of the four identified alleles on CAF1, CFA18, CFA32, and CFA34 from the GWAS with the GV predicted from all 95K markers in GBLUP and BayesR. However, the four markers must first be validated, and, in our case, the predictive value based on the four markers could be overestimated since the GWAS is conducted in the same dataset as the prediction. The GV from all 95K markers is, however, based on six-fold cross-validation on independent subsets of the material. For a robust comparison of the performance of the two approaches, an independent test sample would be required. Once the genomic markers are validated, it may also be relevant to consider combining the information from the four markers with the whole-genome SNP data in a Bayesian model.

Ethical considerations

Using private pet dogs with naturally occurring disorders for this project has offered valuable insight into the genetic background of disorders affecting both dogs and humans. Utilizing such spontaneous animal models can potentially decrease the reliance on conventional research animals, such as laboratory mice or rats. Furthermore, the general public might accept this approach more readily.

Despite the benefits of using pet dogs in research, it can raise ethical considerations that must be carefully addressed. The well-being of the dogs and the principles of responsible animal research must be upheld. The use of animals in research in Norway is governed by the Animal Welfare Act ("Lov om dyrevelferd," 2010) and the regulations regarding the use of animals in experiments ("Forskrift om bruk av dyr i forsøk," 2015). In addition, NMBU has specific guidelines when researchers involve animals in research (NMBU, 2015). These guidelines emphasise that the principles of the “three Rs” for animal experimentation (“reduce, refine, replace”) should guide all instances of the use of animals in research.

The impact on the dogs attending the project has been limited; eye examination data were already available in connection with the eye screening recommended by the respective breed clubs, and most of the DNA material was obtained from the biobank at NMBU. To reduce harm and discomfort to the dogs, we have used buccal swabs to collect DNA samples.

All collection of samples has included written consent from the owner. Personal information has been anonymized to ensure the integrity of the guidelines (NMBU, 2015) and the act relating to the processing of personal data ("Act relating to the processing of personal data," 2022).

5.6 Future prospectives

Future applications of the results in dog breeding

The studies and findings have shown that there is a potential to obtain genetic progress for both studied diseases. However, due to the different inheritance of the diseases, distinct approaches will be necessary to address these two disorders in their respective breeds. The association study of PNC in buhund revealed a significantly associated region on CFA37. Although a causal variant is not yet found, the high significance and the low number of dogs needed to identify the region indicate a causal variant with a strong effect on the disease. By using WGS data and variant calling, a causal variant may be identified, and a genetic test may be developed and used as support for the selection of dogs for breeding.

While waiting for a genetic test, it may be possible to utilize SNP markers or haplotypes in LD with the putative causal variant in the associated region to assess a dog's risk of carrying PNC-associated variants. A genetic risk test could give the opportunity to identify dogs at a high risk of producing affected offspring, reduce the prevalence of risk alleles and simultaneously preserve genetic variation. However, it is important to state the need to verify the risk estimates in an independent material since the risk estimates are based on the same material where we found the associations.

Due to the high prevalence in the buhund population, excluding all dogs with PNC from breeding can drastically reduce the genetic pool within the breed, and a careful consideration of how to implement a genetic test in the future must be done.

According to the buhund club's breed advice, it is recommended to mate two unaffected dogs or, in special cases, one unaffected with one affected dog (Norsk Buhundklubb, 2023). Assuming PNC have a dominant inheritance, using combinations with one affected parent and one unaffected parent will, on average, produce half of the offspring affected with PNC, and the overall prevalence using such a mating would remain constant. By studying the pedigrees, we have found indications that PNC have an incomplete

penetrance, and there are several examples where two unaffected parents (examined after the age of seven) have produced affected offspring. In addition, the late onset can increase the chance of using affected dogs in breeding before they develop symptoms. Relying solely on eye examinations as a diagnostic tool for screening PNC may not be sufficient to reduce the prevalence. With a genetic test, an early diagnosis of breeding dogs could be possible, and the combination of two affected dogs for breeding could be avoided. A genetic test is particularly beneficial by identifying dogs homozygous for the causal variant, which for PNC may be common due to the high prevalence of the disease. Additionally, it might be possible to reduce the prevalence of the more severely affected PNC-dogs, if homozygotes for the causal variants have more advanced symptoms or an earlier onset of the disease than heterozygotes.

The study of cataracts in the buhund in this thesis primarily focuses on PNC. However, it is essential to acknowledge that buhunds are also highly susceptible to other forms of cataracts, such as cortical cataracts and posterior polar cataracts. Given the increasing occurrence of these other forms of cataracts in buhunds, future genetic studies in the breeds should include all forms of cataracts within the breed. Preliminary GWAS studies have been performed but have not revealed any potential genetic regions, and the gathering of a larger number of affected dogs is required.

For complex disorders like distichiasis, we have shown that genomic prediction may be an alternative method to identify the genetically best animals for breeding and reduce the disease prevalence. The current approach to reducing the occurrence of distichiasis involves eye screening and restricting the use of affected animals in breeding.

With moderate to high heritability and a relatively accurate diagnostic method available, there are good opportunities to reduce the occurrence of the disorder. It is, however, necessary that the dogs are diagnosed before they are used for breeding and that the screening is done when the dog is at least one year of age. Despite the existence of screening programs, the prevalence of distichiasis in the Norwegian Staffordshire bull terriers has

persisted over the last 20 years. To achieve faster progress in reducing the occurrence of distichiasis, breeders need to put more weight on this trait or consider alternative methods such as pedigree-based breeding values or genomic prediction.

Breeding values and genomic prediction in livestock have, as previously discussed, been very successful. However, systematic use of genomic prediction requires extensive databases containing phenotypes, pedigrees, and genotypes. Achieving the same level of accuracy in dog breeding may not be easily achievable as of today. There is a lack of big, accessible databases with phenotype and genotype information, and it is not known whether the cost of extensive genotyping and implementing complex analysis will justify faster breeding progress. In addition, dog breeders usually have a personal attachment to their dogs, regardless of the dog's genomic value.

Several studies on genomic prediction in dogs demonstrate that the accuracy on the individual level is currently low (Guo et al., 2011; Sánchez-Molano et al., 2015; Baker et al., 2020; Jiang et al., 2021). This may present a challenge if such methods are to be recommended to dog breeders. While genomic selection shows potential benefits for certain diseases and breeds, the cost, logistical challenges, and practical implications for breeders must be carefully evaluated and weighed against the potential increase in breeding progress for each breed and disorder. Only after such assessments can the implementation of these methods be confidently recommended.

Future applications of the results in research

The GWAS conducted on distichiasis in Staffordshire bull terriers in paper II is a valuable first step toward revealing the biological background of distichiasis in canines. As noted earlier, the genomic regions must be verified, and a larger sample size is probably necessary. It can also be informative to study dog breeds in which distichiasis presents a notable issue, such as the miniature poodle, where the disorder is so widespread

that identifying two unaffected individuals for breeding may be challenging (NKK, 2021a).

The next step to reveal a casual variant for PNC in the buhund is to use WGS to sequence a number of cases and controls and to use variant calling focusing on the associated region on CFA37. This will also facilitate the analysis of the candidate gene *PIKFYVE*. The WGS is already in progress, but the work is not completed, and the results are therefore not included in this thesis.

5.7 Conclusions

Certain hereditary disorders such as PNC and distichiasis exhibit a high prevalence in breeds like the buhund and Staffordshire bull terrier. This thesis provides new information concerning heritability estimates, the identification of associated genetic regions, and the evaluation of genomic prediction. These findings can help monitor and manage PNC and distichiasis and ultimately ensure the best possible welfare of the dogs.

Genetic testing for monogenetic traits has been extensively utilized in dog breeding to reduce the prevalence of ocular disorders. As of today, other approaches, such as genomic prediction, have received less attention. With that in mind, it is important for researchers in the field of dog genetics to explore tools used in other animal species and consider the cost and benefit of implementing new approaches in future breeding practices for dogs.

This thesis aimed to improve the understanding of distichiasis in Staffordshire bull terriers and PNC in the buhunds. The subsequent goal was to provide breed clubs with new insights that may be helpful in future breeding programs.

In paper I, we established that the distichiasis has a moderate to high heritability in the Staffordshire bull terriers. Additionally, our findings indicated that the optimal time for an eye examination regarding distichiasis is after one year of age. With this knowledge, there are excellent opportunities to reduce the occurrence of the disorder through selective breeding.

In paper II we identified four novel genomic regions associated with distichiasis, however, the exact genes involved remain unidentified, and the genomic regions require verification. Nevertheless, the findings give further support to the hypothesis that distichiasis is a polygenetic trait in dogs. Despite the limited statistical power, this is the first published GWAS on canine distichiasis, a highly prevalent condition in dogs. Moreover, in paper II, utilizing genomic markers on microarrays, we have explored the potential of genomic prediction to predict a complex disorder like distichiasis in a dog breed such as Staffordshire bull terrier. Our findings suggest that genomic

prediction may be a valuable alternative to support the selection of breeding dogs. Still, at the individual level, the accuracy is low. Implementing genomic prediction into dog breeding may be challenging because dog breeders have a personal interest in the specific individual dogs. A system of genomic prediction will also require extra efforts and costs for dog owners and breeders, necessitating extensive and expensive genomic prediction systems for breeding clubs.

Paper III deals with cataracts in the Norwegian buhund. The study identified a highly significant region on CFA37. The region contains more than 50 genes, of which at least five stand out as candidate genes. The Sanger sequencing of the coding regions of four of these candidates (gamma crystalline genes) has so far not revealed a causal variant. Further sequencing of the region is warranted to pinpoint a causal variant, and this work is currently in progress. Until a causal variant is identified, using SNP markers in the associated region has the potential to be used in marker-assisted selection to identify dogs at risk of producing offspring with PNC in the Norwegian buhund. Such a DNA-test will be helpful as some dogs may display symptoms of PNC only after being used for breeding.

In conclusion, this thesis has increased the insight concerning the two common ocular disorders, distichiasis and PNC. The exploration and discussion in this work shed light on how this new knowledge could be implemented in future breeding programs to improve the health of the two respective dog breeds, Staffordshire bull terrier and Norwegian buhund. By embracing advancements in genetics and applying new approaches, we can work towards healthier and happier dogs while preserving the integrity of different breeds. The journey toward improved canine welfare is an ongoing process, and continued research in the field of dog genetics is essential for its success.

6 References

Acland, G.M., Ray, K., Mellersh, C.S., Gu, W., Langston, A.A., et al., (1998). Linkage analysis and comparative mapping of canine progressive rod-cone degeneration (prcd) establishes potential locus homology with retinitis pigmentosa (RP17) in humans. Proc Natl Acad Sci USA 95, 3048-53.

Act relating to the processing of personal data. (2022), https://lovdata.no/dokument/NLE/lov/2018-06-15-38.

ACVO Genetics Committee, (2021). The Blue Book, Oculare diorders presumed to be inherited in purebred dogs 2021, In: Genetics Committee of the American college of veterinary ophthalmologists (Ed.).

Aguirre, G.D., Rubin, L.F., Bistner, S.I., (1972). Development of the canine eye. Am J Vet Res 33, 2399-414.

Allais-Bonnet, A., Grohs, C., Medugorac, I., Krebs, S., Djari, A., et al., (2013). Novel insights into the bovine polled phenotype and horn ontogenesis in Bovidae. PLoS One 8, e63512.

Altshuler, D., Donnelly, P., The International HapMap, C., (2005). A haplotype map of the human genome. Nature 437, 1299-320.

Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., Gouil, Q., (2020). Opportunities and challenges in long-read sequencing data analysis. Genome Biol 21, 30.

APPA, American Pet Products Association, (2022) Pet Industry Market Size, Trends & Ownership Statistics. https://www.americanpetproducts.org/press_industrytrends.asp. Accessed 18.10.2022.

Aristotle, (~350BC). On the Generation of Animals. Transl by Arthur Platt, Wikisource. https://en.wikisource.org/wiki/On_the_Generation_of_Animals, Accessed 19.09.2023

Arora, A., Minogue, P.J., Liu, X., Addison, P.K., Russel-Eggitt, I., et al., (2008). A novel connexin50 mutation associated with congenital nuclear pulverulent cataracts. J Med Genet 45, 155-60.

Arteaga, K., Crasta, M., (2021). Successful treatment of distichiasis in a cow using a direct-contact 810-nm diode laser. Vet Surg 50, 1164-8.

Arvelius, P., Klemetsdal, G., (2013). How Swedish breeders can substantially increase the genetic gain for the English Setter's hunting traits. J Anim Breed Genet 130, 142-53.

Axelsson, E., Ljungvall, I., Bhoumik, P., Conn, L.B., Muren, E., et al., (2021). The genetic consequences of dog breed formation-Accumulation of deleterious genetic variation and fixation of mutations associated with myxomatous mitral valve disease in cavalier King Charles spaniels. PLoS Genet 17, e1009726.

Baker, L.A., Kirkpatrick, B., Rosa, G.J., Gianola, D., Valente, B., et al., (2017). Genome-wide association analysis in dogs implicates 99 loci as risk variants for anterior cruciate ligament rupture. PLoS One 12, e0173810.

Baker, L.A., Momen, M., Chan, K., Bollig, N., Lopes, F.B., et al., (2020). Bayesian and Machine Learning Models for Genomic Prediction of Anterior Cruciate Ligament Rupture in the Canine Model. G3 (Bethesda) 10, 2619-28.

Bateman, J.B., Geyer, D.D., Flodman, P., Johannes, M., Sikela, J., et al., (2000). A new betaA1-crystallin splice junction mutation in autosomal dominant cataract. Invest Ophthalmol Vis Sci 41, 3278-85.
Bellamy, K.K.L., Lingaas, F., (2023). Cataracts in Havanese: genome wide association study reveals two loci associated with posterior polar cataract. Canine Med Genet 10, 5.
Bellamy, K.K.L., Lingaas, F., Madsen, P., (2021). Heritability of distichiasis in Havanese dogs in Norway. Canine Med Genet 8, 11.
Bellamy, K.K.L., Skedsmo, F.S., Hultman, J., Arnet, E.F., Guttersrud, O.A., et al., (2022). A 1 bp deletion in HACE1 causes ataxia in Norwegian elkhound, black. PLoS One 17, e0261845.
Bergström, A., Frantz, L., Schmidt, R., Ersmark, E., Lebrasseur, O., et al., (2020). Origins and genetic legacy of prehistoric dogs. Science 370, 557-64.
Berry, V., Pontikos, N., Dudakova, L., Moore, A.T., Quinlan, R., et al., (2020). A novel missense mutation in LIM2 causing isolated autosomal dominant congenital cataract. Ophthalmic Genet 41, 131-4.
Billingsley, K.J., Ding, J., Jerez, P.A., Illarionova, A., Levine, K., et al., (2023). Genome-Wide Analysis of Structural Variants in Parkinson Disease. Ann Neurol.
Bjerkås, E., Haaland, M.B., (1995). Pulverulent nuclear cataract in the Norwegian buhund. J Small Anim Pract 36, 471-4.
Border, R., Johnson, E.C., Evans, L.M., Smolen, A., Berley, N., et al., (2019). No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. Am J Psychiatry 176, 376-87.
Breen, M., (2008). Canine cytogenetics--from band to basepair. Cytogenet Genome Res 120, 50-60.
Brice, G., Mansour, S., Bell, R., Collin, J.R., Child, A.H., et al., (2002). Analysis of the phenotypic abnormalities in lymphoedema-distichiasis syndrome in 74 patients with FOXC2 mutations or linkage to 16q24. J Med Genet 39, 478-83.
Brooks, B.P., Dagenais, S.L., Nelson, C.C., Glynn, M.W., Caulder, M.S., et al., (2003). Mutation of the FOXC2 gene in familial distichiasis. J AAPOS 7, 354-7.
Buckley, R.M., Harris, A.C., Wang, G.D., Whitaker, D.T., Zhang, Y.P., Ostrander, E.A., (2022). Best practices for analyzing imputed genotypes from low-pass sequencing in dogs. Mamm Genome 33, 213-29.
Burton, M.J., Ramke, J., Marques, A.P., Bourne, R.R.A., Congdon, N., et al., (2021). The Lancet Global Health Commission on Global Eye Health: vision beyond 2020. Lancet Glob Health 9, e489-e551.
Caballero, A., (2020). Quantitative Genetics. Cambridge university press.
Cat-Map, (2023) Cat-Map. https://cat-map.wustl.edu/. Accessed 06.07.2023.
Cervantes-Barragán, D.E., Villarroel, C.E., Medrano-Hernández, A., Durán-McKinster, C., Bosch-Canto, V., et al., (2011). Setleis syndrome in Mexican-Nahua sibs due to a homozygous TWIST2 frameshift mutation and partial expression in heterozygotes: review of the focal facial dermal dysplasias and subtype reclassification. J Med Genet 48, 716-20.
Charlesworth, D., Willis, J.H., (2009). The genetics of inbreeding depression. Nature Reviews Genetics 10, 783-96.
Christensen, O.F., Lund, M.S., (2010). Genomic prediction when some animals are not genotyped. Genet Sel Evol 42, 2.
Chu, E.T., Simpson, M.J., Diehl, K., Page, R.L., Sams, A.J., Boyko, A.R., (2019). Inbreeding depression causes reduced fecundity in Golden Retrievers. Mamm Genome 30, 166-72.

Collins, H., Calvo, S., Greenberg, K., Forman Neall, L., Morrison, S., (2016). Information Needs in the Precision Medicine Era: How Genetics Home Reference Can Help. Interact J Med Res 5, e13.
Cook, S.R., Conzemius, M.G., McCue, M.E., Ekenstedt, K.J., (2020). SNP-based heritability and genetic architecture of cranial cruciate ligament rupture in Labrador Retrievers. Anim Genet 51, 824-8.
D'Anna, N., Sapienza, J.S., Guandalini, A., Guerriero, A., (2007). Use of a dermal biopsy punch for removal of ectopic cilia in dogs: 19 cases. Vet Ophthalmol 10, 65-7.
Daetwyler, H.D., Villanueva, B., Woolliams, J.A., (2008). Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach. PLOS One 3, e3395.
Danysh, B.P., Duncan, M.K., (2009). The lens capsule. Experimental Eye Research 88, 151-64.
Das, R.G., Becker, D., Jagannathan, V., Goldstein, O., Santana, E., et al., (2019). Genome-wide association study and whole-genome sequencing identify a deletion in LRIT3 associated with canine congenital stationary night blindness. Sci Rep 9, 14166.
Davis, B.W., Ostrander, E.A., (2014). Domestic dogs and cancer research: a breed-based genomics approach. Ilar J 55, 59-68.
Dempster, E.R., Lerner, I.M., (1950). Heritability of Threshold Characters. Genetics 35, 212-36.
Ding, X., Wang, B., Luo, Y., Hu, S., Zhou, G., et al., (2011). A novel mutation in the connexin 46 (GJA3) gene associated with congenital cataract in a Chinese pedigree. Mol Vis 17, 1343-9.
Dog10K, (2015) Dog 10K genomes project. http://www.dog10kgenomes.org/. Accessed 06.08.2023.
DoGA consortium, (2023) The dog genome annotation project. https://www.doggenomeannotation.org/. Accessed 06.08.2023.
Donzel, E., Arti, L., Chahory, S., (2017). Epidemiology and clinical presentation of canine cataracts in France: a retrospective study of 404 cases. Vet Ophthalmol 20, 131-9.
DyreID, (2023). Kjæledyrrapporten 2022 - Nasjonal rapport for landets kjæledyr, p. 17.
Ebbert, M.T.W., Jensen, T.D., Jansen-West, K., Sens, J.P., Reddy, J.S., et al., (2019). Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. Genome Biol 20, 97.
Edwards, R.J., Field, M.A., Ferguson, J.M., Dudchenko, O., Keilwagen, J., et al., (2021). Chromosome-length genome assembly and structural variations of the primal Basenji dog (Canis lupus familiaris) genome. BMC Genomics 22, 188.
Engelhardt, A., Stock, K.F., Hamann, H., Brahm, R., Grussendorf, H., et al., (2007). Analysis of systematic and genetic effects on the prevalence of primary cataract, persistent pupillary membrane and distichiasis in the two color variants of English Cocker Spaniels in Germany. Berl Munch Tierarztl Wochenschr 120, 490-8.
Erbe, M., Hayes, B.J., Matukumalli, L.K., Goswami, S., Bowman, P.J., et al., (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci 95, 4114-29.
Erickson, R.P., Dagenais, S.L., Caulder, M.S., Downs, C.A., Herman, G., et al., (2001). Clinical heterogeneity in lymphoedema-distichiasis with FOXC2 truncating mutations. J Med Genet 38, 761-6.
Falconer, D.S., (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann Hum Genet 29, 51-76.

Falconer, D.S., Mackay, T.F.C., (1996). Introduction to quantitative genetics. 4th ed. Longman Group.
Fernando, R., Grossman, M., (1989). Marker assisted selection using Best Linear Unbiased Prediction. Genet Sel Evol 21, 467.
Fernando, R.L., Garrick, D., (2013). Bayesian Methods Applied to GWAS. Humana Press.
Fisher, R.A., (1918). The correlation between relatives on the supposition of Mendelian inheritance. Earth Environ Sci Trans R Soc Edinb 52, 399-433.
Forskrift om bruk av dyr i forsøk. (2015), FOR-2015-06-18-761, Lovdata, https://lovdata.no/dokument/SF/forskrift/2015-06-18-761.
Freedman, A.H., Gronau, I., Schweizer, R.M., Ortega-Del Vecchyo, D., Han, E., et al., (2014). Genome sequencing highlights the dynamic early history of dogs. PLoS Genet 10, e1004016.
Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., et al., (2002). The Structure of Haplotype Blocks in the Human Genome. Science 296, 2225-9.
García-Ruiz, A., Cole, J.B., VanRaden, P.M., Wiggans, G.R., Ruiz-López, F.J., Van Tassell, C.P., (2016). Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc Natl Acad Sci USA 113, E3995-4004.
Gatz, M., Reynolds, C.A., Fratiglioni, L., Johansson, B., Mortimer, J.A., et al., (2006). Role of Genes and Environments for Explaining Alzheimer Disease. Archives of General Psychiatry 63, 168-74.
Gelatt, K.N., Gelatt, J. P., Plummer, C (2011). Veterinary ophthalmic surgery. Elsevier Saunders. 1 ed.
Gelatt, K.N., Gilger, B.C., Kern, T.J., (2013). Veterinary ophthalmology : Volume 2. 5th ed. Blackwell.
Gelatt, K.N., MacKay, E.O., (2005). Prevalence of primary breed-related cataracts in the dog in North America. Vet Ophthalmol 8, 101-11.
Gerber, M., Fischer, A., Jagannathan, V., Drögemüller, M., Drögemüller, C., et al., (2015). A Deletion in the VLDLR Gene in Eurasier Dogs with Cerebellar Hypoplasia Resembling a Dandy-Walker-Like Malformation (DWLM). PLoS One 10, e0108917.
Giani, A.M., Gallo, G.R., Gianfranceschi, L., Formenti, G., (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. Comput Struct Biotechnol J 18, 9-19.
Gianola, D., de los Campos, G., Hill, W.G., Manfredi, E., Fernando, R., (2009). Additive Genetic Variability and the Bayesian Alphabet. Genetics 183, 347-63.
Griffiths, A.J.F., (2002). Modern genetic analysis : integrating genes and genomes. 2nd ed. W.H. Freeman and Co. p 1-22.
Guo, G., Zhou, Z., Wang, Y., Zhao, K., Zhu, L., et al., (2011). Canine hip dysplasia is predictable by genotyping. Osteoarthritis Cartilage 19, 420-9.
Gustavsson, I., (1964). The chromosomes of the dog. Hereditas 51, 187-9.
Guziewicz, K.E., Zangerl, B., Lindauer, S.J., Mullins, R.F., Sandmeyer, L.S., et al., (2007). Bestrophin gene mutations cause canine multifocal retinopathy: a novel animal model for best disease. Invest Ophthalmol Vis Sci 48, 1959-67.
Habier, D., Fernando, R.L., Kizilkaya, K., Garrick, D.J., (2011). Extension of the bayesian alphabet for genomic selection. BMC Bioinform 12, 186.

Haile-Mariam, M., Nieuwhof, G.J., Beard, K.T., Konstatinov, K.V., Hayes, B.J., (2013). Comparison of heritabilities of dairy traits in Australian Holstein-Friesian cattle from genomic and pedigree data and implications for genomic evaluations. J Anim Breed Genet 130, 20-31.

Halo, J.V., Pendleton, A.L., Shen, F., Doucet, A.J., Derrien, T., et al., (2021). Long-read assembly of a Great Dane genome highlights the contribution of GC-rich sequence and mobile elements to canine genomes. Proc Natl Acad Sci USA 118.

Handegård, K.W., Storengen, L.M., Joergensen, D., Lingaas, F., (2023). Genomic analysis of firework fear and noise reactivity in standard poodles. Canine Med Genet 10, 2.

Hedhammar, Å., (2020). Swedish Experiences From 60 Years of Screening and Breeding Programs for Hip Dysplasia-Research, Success, and Challenges. Front Vet Sci 7, 228.

Hickey, J.M., Chiurugwi, T., Mackay, I., Powell, W., Hickey, J.M., et al., (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. Nat Geneti 49, 1297-303.

Hill, W.G., Goddard, M.E., Visscher, P.M., (2008). Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. PLoS Genet 4, e1000008.

Hisey, E.A., Hermans, H., Lounsberry, Z.T., Avila, F., Grahn, R.A., et al., (2020). Whole genome sequencing identified a 16 kilobase deletion on ECA13 associated with distichiasis in Friesian horses. BMC Genomics 21, 848.

Hoeppner, M.P., Lundquist, A., Pirun, M., Meadows, J.R., Zamani, N., et al., (2014). An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. PLoS One 9, e91172.

Hoover, R.E., Kelley, J.S., (1971). Distichiasis and lymphedema: a hereditary syndrome with possible multiple defects. A report of a family. Trans Am Ophthalmol Soc 69, 293-306.

Hyttel, P., Sinowatz, F., Vejlsted, M., (2009). Essentials of domestic animal embryology. Saunders.

Illumina, (2010) CanineHD BeadChip. Illumina Inc,. https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_caninehd.pdf. Accessed 24.07.2023.

Ilska, J., Haskell, M.J., Blott, S.C., Sánchez-Molano, E., Polgar, Z., et al., (2017). Genetic Characterization of Dog Personality Traits. Genetics 206, 1101-11.

International Human Genome Sequencing, C., (2004). Finishing the euchromatic sequence of the human genome. Nature 431, 931-45.

Ioannides, J., Everson, R., Matas Riera, M., Dawson, C., (2020). A description of blend electrolysis for treatment of canine distichiasis: 78 cases (2012-2017). Vet Rec 187, e102.

Jagannathan, V., Bannoehr, J., Plattet, P., Hauswirth, R., Drögemüller, C., et al., (2013). A mutation in the SUV39H2 gene in Labrador Retrievers with hereditary nasal parakeratosis (HNPK) provides insights into the epigenetics of keratinocyte differentiation. PLoS Genet 9, e1003848.

Jagannathan, V., Hitte, C., Kidd, J.M., Masterson, P., Murphy, T.D., et al., (2021). Dog10K_Boxer_Tasha_1.0: A Long-Read Assembly of the Dog Reference Genome. Genes (Basel) 12.

Jamieson, R.V., Perveen, R., Kerr, B., Carette, M., Yardley, J., et al., (2002). Domain disruption and mutation of the bZIP transcription factor, MAF,associated with cataract, ocular anterior segment dysgenesis and coloboma. Hum Mol Genet 11, 33-42.
Jiang, L., Li, Z., Hayward, J.J., Hayashi, K., Krotscheck, U., et al., (2021). Genomic Prediction of Two Complex Orthopedic Traits Across Multiple Pure and Mixed Breed Dogs. Front Genet 12, 666740.
Jónasdóttir, T.J., Mellersh, C.S., Moe, L., Heggebø, R., Gamlem, H., et al., (2000). Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs. Proc Natl Acad Sci USA 97, 4132-7.
Jondeau, C., Gounon, M., Bourguet, A., Chahory, S., (2023). Epidemiology and clinical significance of canine distichiasis: A retrospective study of 291 cases. Vet Ophthalmol 26, 339-346.
Jorde, L.B., Carey, J.C., Bamshad, M.J., (2016). Medical genetics. 5th ed. Elsevier.
Karlsson, E.K., Baranowska, I., Wade, C.M., Salmon Hillbertz, N.H., Zody, M.C., et al., (2007). Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet 39, 1321-8.
Karlsson, E.K., Lindblad-Toh, K., (2008). Leader of the pack: gene mapping in dogs and other model organisms. Nat Rev Genet 9, 713-25.
Karlsson, E.K., Sigurdsson, S., Ivansson, E., Thomas, R., Elvers, I., et al., (2013). Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. Genome Biol 14, R132.
Kaufhold, J., Hamann, H., Steinbach, G., Gordon, S., Brahm, R., et al., (2006). Analysis of the prevalence of distichiasis in the dog breed Elo. Berl Munch Tierarztl Wochenschr 119, 233-7.
Kemper, K.E., Goddard, M.E., (2012). Understanding and predicting complex traits: knowledge from cattle. Hum Mol Genet 21, R45-51.
Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., et al., (2002). The human genome browser at UCSC. Genome Res 12, 996-1006.
Ketteritzsch, K., Hamann, H., Brahm, R., Grussendorf, H., Rosenhagen, C.U., Distl, O., (2004). Genetic analysis of presumed inherited eye diseases in Tibetan Terriers. Vet J 168, 151-9.
Kievit, A., Tessadori, F., Douben, H., Jordens, I., Maurice, M., et al., (2018). Variants in members of the cadherin-catenin complex, CDH1 and CTNND1, cause blepharocheilodontic syndrome. Eur J Hum Genet 26, 210-9.
Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K., et al., (2003). The Dog Genome: Survey Sequencing and Comparative Analysis. Science 301, 1898-903.
Kjetså, M.V., Gjuvsland, A.B., Nordbø, Ø., Grindflek, E., Meuwissen, T., (2022). Accuracy of genomic prediction of maternal traits in pigs using Bayesian variable selection methods. J Anim Breed Genet 139, 654-65.
Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., et al., (2005). Complement factor H polymorphism in age-related macular degeneration. Science 308, 385-9.
Kristiansen, E., Revold, T., Lingaas, F., Narfstrom, K., Pedersen, P.B., et al., (2017). Cataracts in the Norwegian Buhund-current prevalence and characteristics. Vet Ophthalmol 20, 460-7.

Kumar, M., Agarwal, T., Khokhar, S., Kumar, M., Kaur, P., et al., (2011). Mutation screening and genotype phenotype correlation of α-crystallin, γ-crystallin and GJA8 gene in congenital cataract. Mol Vis 17, 693-707.
Lande, R., Thompson, R., (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124, 743-56.
Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., et al., (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.
Lawson, D.D., (1973). Canine distichiasis. J Small Anim Pract 14, 469-78.
Leroy, G., (2011). Genetic diversity, inbreeding and breeding practices in dogs: Results from pedigree analyses. Vet J 189, 177-82.
Leroy, G., Danchin-Burge, C., Palhiere, I., Baumung, R., Fritz, S., et al., (2012). An ABC estimate of pedigree error rate: application in dog, sheep and cattle breeds. Anim Genet 43, 309-14.
Leroy, G., Phocas, F., Hedan, B., Verrier, E., Rognon, X., (2015). Inbreeding impact on litter size and survival in selected canine breeds. Vet J 203, 74-8.
Letko, A., Minor, K.M., Norton, E.M., Marinescu, V.D., Drögemüller, M., et al., (2021). Genome-Wide Analyses for Osteosarcoma in Leonberger Dogs Reveal the CDKN2A/B Gene Locus as a Major Risk Locus. Genes (Basel) 12.
Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., et al., (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438, 803-19.
Lingaas, F., Tengvall, K., Jansen, J.H., Pelander, L., Hurst, M.H., et al., (2023). Bayesian mixed model analysis uncovered 21 risk loci for chronic kidney disease in boxer dogs. PLoS Genet 19, e1010599.
Lingaas, F., Aarskaug, T., Sletten, M., Bjerkås, I., Grimholt, U., et al., (1998). Genetic markers linked to neuronal ceroid lipofuscinosis in English setter dogs. Anim Genet 29, 371-6.
Lov om dyrevelferd, (2010), https://lovdata.no/dokument/NL/lov/2009-06-19-97.
Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., et al., (2012). Revisiting global gene expression analysis. Cell 151, 476-82.
Lovicu, F.J., McAvoy, J.W., (2005). Growth factor regulation of lens development. Dev Biol 280, 1-14.
Lowe, J.K., Kukekova, A.V., Kirkness, E.F., Langlois, M.C., Aguirre, G.D., et al., (2003). Linkage mapping of the primary disease locus for collie eye anomaly. Genomics 82, 86-95.
Lu, S., Zhao, C., Jiao, H., Kere, J., Tang, X., et al., (2007). Two Chinese families with pulverulent congenital cataracts and deltaG91 CRYBA1 mutations. Mol Vis 13, 1154-60.
Lynch, M., Walsh, B., (1998). Genetics and analysis of quantitative traits. Sinauer Associates.
Lynch, M., Walsh, B., (2018). Evolution and Selection of Quantitative Traits. Oxford: Oxford University Press.
MacLeod, I.M., Bowman, P.J., Vander Jagt, C.J., Haile-Mariam, M., Kemper, K.E., et al., (2016). Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genomics 17, 144.
Madsen, P., Jensen, J., (2013). A user's guide to DMU. A package for analysing multivariate mixed models version 6, 1-33.
Maggs, D.J., Miller, P.E., Ofri, R., Slatter, D.H., (2018). Slatter's fundamentals of veterinary ophthalmology. 6th ed. Elsevier.

Mahajan, A., Spracklen, C.N., Zhang, W., Ng, M.C.Y., Petty, L.E., et al., (2022). Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet 54, 560-72.
Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., et al., (2009). Finding the missing heritability of complex diseases. Nature 461, 747-53.
Marees, A.T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., et al., (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res 27, e1608.
Marsden, C.D., Ortega-Del Vecchyo, D., O'Brien, D.P., Taylor, J.F., Ramirez, O., et al., (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc Natl Acad Sci USA 113, 152-7.
McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., et al., (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9, 356-69.
McGeady, T.A., Quinn, P.J., Fitzpatrick, E.S., (2016). Veterinary embryology. 2nd ed. Wiley-Blackwell.
Meher, P.K., Rustgi, S., Kumar, A., (2022). Performance of Bayesian and BLUP alphabets for genomic prediction: analysis, comparison and results. Heredity 128, 519-30.
Mei, S., Wu, Y., Wang, Y., Cui, Y., Zhang, M., et al., (2022). Disruption of PIKFYVE causes congenital cataract in human and zebrafish. eLife 11, e71256.
Melis, C., Borg Å, A., Espelien, I.S., Jensen, H., (2013). Low neutral genetic variability in a specialist puffin hunter: the Norwegian Lundehund. Anim Genet 44, 348-51.
Mellersh, C.S., (2014). The genetics of eye disorders in the dog. Canine Genet Epidemiol 1, 3.
Mellersh, C.S., Graves, K.T., McLaughlin, B., Ennis, R.B., Pettitt, L., et al., (2007). Mutation in HSF4 associated with early but not late-onset hereditary cataract in the Boston Terrier. J Hered 98, 531-3.
Mellersh, C.S., McLaughlin, B., Ahonen, S., Pettitt, L., Lohi, H., Barnett, K.C., (2009). Mutation in HSF4 is associated with hereditary cataract in the Australian Shepherd. Vet Ophthalmol 12, 372-8.
Mellersh, C.S., Pettitt, L., Forman, O.P., Vaudin, M., Barnett, K.C., (2006). Identification of mutations in HSF4 in dogs of three different breeds with hereditary cataracts. Vet Ophthalmol 9, 369-78.
Melville, S.A., Wilson, C.L., Chiang, C.S., Studdert, V.P., Lingaas, F., Wilton, A.N., (2005). A mutation in canine CLN5 causes neuronal ceroid lipofuscinosis in Border collie dogs. Genomics 86, 287-94.
Meuwissen, T.H., Hayes, B.J., Goddard, M.E., (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-29.
Meyer, E., Rahman, F., Owens, J., Pasha, S., Morgan, N.V., et al., (2009). Initiation codon mutation in betaB1-crystallin (CRYBB1) associated with autosomal recessive nuclear pulverulent cataract. Mol Vis 15, 1014-9.
Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., et al., (2020). Telomere-to-telomere assembly of a complete human X chromosome. Nature 585, 79–84.
Mikkola, L., Holopainen, S., Pessa-Morikawa, T., Lappalainen, A.K., Hytönen, M.K., et al., (2019). Genetic dissection of canine hip dysplasia phenotypes and osteoarthritis reveals three novel loci. BMC genomics 20, 1027.

Mills, M.C., Barban, N., Tropf, F.C., (2020). An introduction to statistical genetic data analysis. The MIT Press.
Morrill, K., Hekman, J., Li, X., McClure, J., Logan, B., et al., (2022). Ancestry-inclusive dog genomics challenges popular breed stereotypes. Science 376, eabk0639.
Moser, G., Lee, S.H., Hayes, B.J., Goddard, M.E., Wray, N.R., Visscher, P.M., (2015). Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. PLoS Genet 11, e1004969.
Nguyen, T.-D., Harder, A., Xiong, Y., Kowalec, K., Hägg, S., et al., (2022). Genetic heterogeneity and subtypes of major depression. Mol Psychiatry 27, 1667-75.
Nicholas, F., Tammen, I., Sydney Informatics, H., (1995). Online Mendelian Inheritance in Animals (OMIA), The University of Sydney.
Nickel, R., Schummer, A., Seiferle, E., (2004). Lehrbuch der Anatomie der Haustiere Band 4, Parey, Stuttgart.
NKK, The Norwegian Kennel Club (2021a) Dog web, www.dogweb.no. Accessed 12.05.2021.
NKK, The Norwegian Kennel Club (2021b) NKK statestikk. https://www.nkk.no/statistikk/category1098.html. Accessed 04.03.2021.
NKK, The Norwegian Kennel Club (2023a) The Norwegian Kennel Club. https://www.nkk.no/english/category1045.html. Accessed 12.07.2023.
NKK, The Norwegian Kennel Club (2023b) Raser med registreringsrestriksjoner. https://www.nkk.no/raser-med-krav-for-registrering/category964.html. Accessed 04.07.2023.
NMBU, (2015) Etiske retningslinjer for NMBU. https://main-bvxea6i-kdsvgmpf4iwws.eu-5.platformsh.site/sites/default/files/pdfattachments/160111_nye-etiske_retningslinjer_vedtatt_26_11_12015_0_4_ny_mal.pdf#page=13&zoom=100,112,76. Accessed 20.08.2023.
Norsk Buhundklubb, (2020) Buhundprosjektet. https://buhund.no/buhundprosjektet/. Accessed 06.07.2023.
Norsk Buhundklubb, (2023) Regler og retningslinjer for avl av norsk buhund https://buhund.no/wp-content/uploads/2023/04/Regler-og-retningslinjer-for-avl-av-norsk-buhund.pdf. Accessed 13.07.2023.
Norton, E.M., Minor, K.M., Taylor, S.M., McCue, M.E., Mickelson, J.R., (2021). Heritability and Genomic Architecture of Episodic Exercise-Induced Collapse in Border Collies. Genes (Basel) 12.
Oberbauer, A.M., Keller, G.G., Famula, T.R., (2017). Long-term genetic selection reduced prevalence of hip and elbow dysplasia in 60 dog breeds. PLoS One 12, e0172918.
Olsson, M., Meadows, J.R., Truvé, K., Rosengren Pielberg, G., Puppo, F., et al., (2011). A novel unstable duplication upstream of HAS2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. PLoS Genet 7, e1001332.
OMIA, (2023) Online Mendelian Inheritance in Animals, OMIA. Sydney School of Veterinary Science. Available at: https://omia.org/. Accessed 15.09.2023.
Ostrander, E.A., Wang, G.D., Larson, G., vonHoldt, B.M., Davis, B.W., et al., (2019). Dog10K: an international sequencing effort to advance studies of canine domestication, phenotypes and health. Natl Sci Rev 6, 810-24.
Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., et al., (2002). Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. Nat Genet 32, 650-4.

Palella Gómez, A., Mazzucchelli, S., Scurrell, E., Smith, K., Pinheiro de Lacerda, R., (2020). Evaluation of partial tarsal plate excision using a transconjunctival approach for the treatment of distichiasis in dogs. Vet Ophthalmol 23, 506-14.
Park, S.A., Yi, N.Y., Jeong, M.B., Kim, W.T., Kim, S.E., et al., (2009). Clinical manifestations of cataracts in small breed dogs. Vet Ophthalmol 12, 205-10.
Parker, H.G., Kukekova, A.V., Akey, D.T., Goldstein, O., Kirkness, E.F., et al., (2007). Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. Genome Res 17, 1562-71.
Perri, A., (2016). A wolf in dog's clothing: Initial dog domestication and Pleistocene wolf variation. J Archaeol Sci 68, 1-4.
Petersen, T., Proschowsky, H.F., Hardon, T., Rasch, S.N., Fredholm, M., (2015). Prevalence and heritability of distichiasis in the English Cocker spaniel. Canine Genet Epidemiol 2, 11.
Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904-9.
Price, A.L., Zaitlen, N.A., Reich, D., Patterson, N., (2010). New approaches to population stratification in genome-wide association studies. Nat Rev Genet 11, 459-63.
Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., et al., (2009). Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet 2, 73-80.
R Core Team, (2021) R: A Language and Environment for Statistical Computing. https://www.R-project.org/.
Raymond-Letron, I., Bourges-Abella, N., Rousseau, T., Douet, J.Y., de Geyer, G., Regnier, A., (2012). Histopathologic features of canine distichiasis. Vet Ophthalmol 15, 92-7.
Reinstein, S.L., Gross, S.L., Komáromy, A.M., (2011). Successful treatment of distichiasis in a cat using transconjunctival electrocautery. Vet Ophthalmol 14 Suppl 1, 130-4.
Ricketts, S.L., Pettitt, L., McLaughlin, B., Jenkins, C.A., Mellersh, C.S., (2015). A novel locus on canine chromosome 13 is associated with cataract in the Australian Shepherd breed of domestic dog. Mamm Genome 26, 257-63.
Ritland, K., (1996). A Marker-Based Method for Inferences About Quantitative Inheritance in Natural Populations. Evolution 50, 1062-73.
Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., et al., (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinform 12, 77.
Robinson, G.K., (1991). That BLUP is a Good Thing: The Estimation of Random Effects. Stat Sci 6, 15-32, 18.
Rudd Garces, G., Christen, M., Loechel, R., Jagannathan, V., Leeb, T., (2022). FYCO1 Frameshift Deletion in Wirehaired Pointing Griffon Dogs with Juvenile Cataract. Genes (Basel) 13.
Sánchez-Molano, E., Pong-Wong, R., Clements, D.N., Blott, S.C., Wiener, P., Woolliams, J.A., (2015). Genomic prediction of traits related to canine hip dysplasia. Front Genet 6, 97.
Sanger, F., Nicklen, S., Coulson, A.R., (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74, 5463-7.
Savolainen, P., Zhang, Y.-p., Luo, J., Lundeberg, J., Leitner, T., (2002). Genetic Evidence for an East Asian Origin of Domestic Dogs. Science 298, 1610-3.
Sham, P.C., Purcell, S.M., (2014). Statistical power and significance testing in large-scale genetic studies. Nat Rev Genet 15, 335-46.

Shendure, J., Ji, H., (2008). Next-generation DNA sequencing. Nat Biotechnol 26, 1135-45.
Shiels, A., Bennett, T.M., Hejtmancik, J.F., (2010). Cat-Map: putting cataract on the map. Mol Vis 16, 2007-15.
Shiels, A., Hejtmancik, J.F., (2017). Mutations and mechanisms in congenital and age-related cataracts. Exp Eye Res 156, 95-102.
Shiels, A., Hejtmancik, J.F., (2019). Biology of Inherited Cataracts and Opportunities for Treatment. Annu Rev Vis Sci 5, 123-49.
Singh, B., Dyce, K.M., Wensing, C.J.G., Sack, W.O., (2018). Dyce, Sack, and Wensing's textbook of veterinary anatomy. 5th ed. Elsevier.
Skoglund, P., Ersmark, E., Palkopoulou, E., Dalén, L., (2015). Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds. Curr Biol 25, 1515-9.
Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., et al., (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res 51, D977-d85.
Speed, D., Hemani, G., Johnson, M.R., Balding, D.J., (2012). Improved heritability estimation from genome-wide SNPs. Am J Hum Genet 91, 1011-21.
Steinmetz, J.D., Bourne, R.R.A., Briant, P.S., Flaxman, S.R., Taylor, H.R.B., et al., (2021). Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. Lancet Glob Health 9, e144-e60.
Suber, M.L., Pittler, S.J., Qin, N., Wright, G.C., Holcombe, V., et al., (1993). Irish setter dogs affected with rod/cone dysplasia contain a nonsense mutation in the rod cGMP phosphodiesterase beta-subunit gene. Proc Natl Acad Sci USA 90, 3968-72.
Sun, C., VanRaden, P.M., Cole, J.B., O'Connell, J.R., (2014). Improvement of prediction ability for genomic selection of dairy cattle by including dominance effects. PLoS One 9, e103934.
Sutter, N.B., Eberle, M.A., Parker, H.G., Pullar, B.J., Kirkness, E.F., et al., (2004). Extensive and breed-specific linkage disequilibrium in Canis familiaris. Genome Res 14, 2388-96.
Sutter, N.B., Ostrander, E.A., (2004). Dog star rising: the canine genetic system. Nat Rev Genet 5, 900-10.
Tavian, D., Missaglia, S., Maltese, P.E., Michelini, S., Fiorentino, A., et al., (2016). FOXC2 disease-mutations identified in lymphedema-distichiasis patients cause both loss and gain of protein function. Oncotarget 7, 54228-39.
Tengvall, K., Sundström, E., Wang, C., Bergvall, K., Wallerman, O., et al., (2022). Bayesian model and selection signature analyses reveal risk factors for canine atopic dermatitis. Commun Biol 5, 1348.
The Kennel Club, (2023) History of The Kennel Club. https://www.thekennelclub.org.uk/about-us/about-the-kennel-club/history-of-the-kennel-club/. Accessed 11.07.2023.
The Norwegian Terrier Club, (2021) Avlsanbefalinger. https://norskterrierklub.no/Userfiles/Sites/files/Rasegruppens%20anbefalinger%20for%20avl(4).pdf. Accessed 05.04.2022.
Thorsrud, J.A., Evans, K.M., Holle, D.M., Huson, H.J., (2022). 394. Comparison of GBLUP and machine learning breeding values for binary disease traits in working dogs, Proceedings of 12th WCGALP, Wageningen Academic Publishers, pp. 1643-6.
Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., et al., (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 604, 502-8.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A., (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35, W71-4.
van Binsbergen, R., Calus, M.P.L., Bink, M.C.A.M., van Eeuwijk, F.A., Schrooten, C., Veerkamp, R.F., (2015). Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. Genet Sel Evol 47, 71.
van den Berg, I., Bowman, P.J., MacLeod, I.M., Hayes, B.J., Wang, T., et al., (2017). Multi-breed genomic prediction using Bayes R with sequence data and dropping variants with a small effect. Genet Sel Evol 49, 70.
VanRaden, P.M., (2008). Efficient Methods to Compute Genomic Predictions. J Dairy Sci 91, 4414-23.
Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., et al., (2001). The sequence of the human genome. Science 291, 1304-51.
Verboven, C.A., Djajadiningrat-Laanen, S.C., Kitslaar, W.J., Grinwis, G.C., Schoemaker, N.J., Boevé, M.H., (2014). Distichiasis in a ferret (Mustela putorius furo). Vet Ophthalmol 17, 290-3.
Veske, A., Nilsson, S.E., Narfström, K., Gal, A., (1999). Retinal dystrophy of Swedish briard/briard-beagle dogs is due to a 4-bp deletion in RPE65. Genomics 57, 57-61.
Vilà, C., Savolainen, P., Maldonado, J.E., Amorim, I.R., Rice, J.E., et al., (1997). Multiple and Ancient Origins of the Domestic Dog. Science 276, 1687-9.
Visscher, P.M., Goddard, M.E., (2019). From R.A. Fisher's 1918 Paper to GWAS a Century Later. Genetics 211, 1125-30.
Visscher, P.M., Hill, W.G., Wray, N.R., (2008). Heritability in the genomics era — concepts and misconceptions. Nat Rev Genet 9, 255-66.
Wainschtein, P., Jain, D., Zheng, Z., Aslibekyan, S., Becker, D., et al., (2022). Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nat Genet 54, 263-73.
Wang, C., Wallerman, O., Arendt, M.L., Sundström, E., Karlsson, Å., et al., (2021). A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. Commun Biol 4, 185.
Wang, G.D., Larson, G., Kidd, J.M., vonHoldt, B.M., Ostrander, E.A., Zhang, Y.P., (2019). Dog10K: the International Consortium of Canine Genome Sequencing. Natl Sci Rev 6, 611-3.
Wang, G.D., Zhai, W., Yang, H.C., Wang, L., Zhong, L., et al., (2016). Out of southern East Asia: the natural history of domestic dogs across the world. Cell Res 26, 21-33.
Wang, W., Jiang, J., Zhu, Y., Li, J., Jin, C., et al., (2010). A novel mutation in the major intrinsic protein (MIP) associated with autosomal dominant congenital cataracts in a Chinese family. Mol Vis 16, 534-9.
Watson, J.D., Crick, F.H.C., (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 171, 737-8.
WHO, (2022) Blindness and vision impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. Accessed 06.07.2023.
Wightman, D.P., Jansen, I.E., Savage, J.E., Shadrin, A.A., Bahrami, S., et al., (2021). A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. Nat Genet 53, 1276-82.
Wilbe, M., Jokinen, P., Truvé, K., Seppala, E.H., Karlsson, E.K., et al., (2010). Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. Nat Genet 42, 250-4.

Willemsen, G., Ward, K.J., Bell, C.G., Christensen, K., Bowden, J., et al., (2015). The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. Twin Res Hum Genet 18, 762-71.

Wolc, A., Dekkers, J.C.M., (2022). Application of Bayesian genomic prediction methods to genome-wide association analyses Genet Sel Evol 54, 31.

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., et al., (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 46, 1173-86.

Wray, N.R., Kemper, K.E., Hayes, B.J., Goddard, M.E., Visscher, P.M., (2019). Complex Trait Prediction from Genome Data: Contrasting EBV in Livestock to PRS in Humans: Genomic Prediction. Genetics 211, 1131-41.

Wray, N.R., Yang, J., Goddard, M.E., Visscher, P.M., (2010). The Genetic Interpretation of Area under the ROC Curve in Genomic Profiling. PLOS Genet 6, e1000864.

Wray, N.R., Yang, J., Hayes, B.J., Price, A.L., Goddard, M.E., Visscher, P.M., (2013). Pitfalls of predicting complex traits from SNPs. Nat Rev Genet 14, 507-15.

Xu, C., Tachmazidou, I., Walter, K., Ciampi, A., Zeggini, E., et al., (2014). Estimating genome-wide significance for whole-genome sequencing studies. Genet Epidemiol 38, 281-90.

Yan, M., Xiong, C., Ye, S.Q., Chen, Y., Ke, M., et al., (2008). A novel connexin 50 (GJA8) mutation in a Chinese family with a dominant congenital pulverulent nuclear cataract. Mol Vis 14, 418-24.

Yang, J., (2023) GCTA a tool for Genome-wide Complex Trait Analysis. https://yanglab.westlake.edu.cn/software/gcta. Accessed 19.04.2023.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., et al., (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42, 565-9.

Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88, 76-82.

Yordy, J., Kraus, C., Hayward, J.J., White, M.E., Shannon, L.M., et al., (2020). Body size, inbreeding, and lifespan in domestic dogs. Conserv Genet 21, 137-48.

Zaitlen, N., Kraft, P., Patterson, N., Pasaniuc, B., Bhatia, G., et al., (2013). Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet 9, e1003520.

Zangerl, B., Goldstein, O., Philp, A.R., Lindauer, S.J., Pearce-Kelling, S.E., et al., (2006). Identical mutation in a novel retinal gene causes progressive rod-cone degeneration in dogs and retinitis pigmentosa in humans. Genomics 88, 551-63.

Zhang, C., Kemp, R.A., Stothard, P., Wang, Z., Boddicker, N., et al., (2018). Genomic evaluation of feed efficiency component traits in Duroc pigs using 80K, 650K and whole-genome sequence variants. Genet Sel Evol 50, 14.

Zimmerman, K.L., Reinstein, S.L., (2019). Evaluation of transconjunctival thermal electrocautery for treatment of canine distichiasis: 88 eyelids (2013-2016). Vet Ophthalmol 22, 50-60.

Zuidhof, M.J., Schneider, B.L., Carney, V.L., Korver, D.R., Robinson, F.E., (2014). Growth, efficiency, and yield of commercial broilers from 1957, 1978, and 2005. Poult Sci 93, 2970-82.

7 Papers I - III

Paper I

Joergensen *et al.*
Acta Veterinaria Scandinavica (2022) 64:30
https://doi.org/10.1186/s13028-022-00650-1

Acta Veterinaria Scandinavica

RESEARCH

Open Access

Heritability estimates of distichiasis in Staffordshire bull terriers using pedigrees and genome-wide SNP data

Dina Joergensen[1*], Per Madsen[2], Ernst-Otto Ropstad[3] and Frode Lingaas[1]

Abstract

Background: Distichiasis is the most frequently recorded eye disorder in the Norwegian Staffordshire bull terrier (SBT). The condition is often mild but can, in severe cases, lead to pain and blindness. The current study's main purpose was to estimate the heritability based on pedigree information as well as single nucleotide polymorphisms (SNPs) to evaluate whether it is realistic to reduce the frequency by systematic breeding. The majority of the dogs had only one examination as a young puppy. To evaluate whether this early screening gave a reliable representation of the disease burden in the population, we compared the diagnosis in puppies and adult dogs.

Results: Our material consisted of data from 4177 dogs with an overall prevalence of distichiasis of 8.38% (CI 7.56–9.26). The prevalence in puppies examined around eight weeks of age was significantly lower than in dogs examined after 52 weeks (2.87%, CI 2.29–3.54 versus 18.72%, CI 16.71–20.87). The heritability was estimated in dogs examined after 52 weeks. We used both pedigree (1391 dogs) and genotype (498 dogs) information for the estimates. The pedigree-based heritability was ~ 0.22 (on the underlying scale ~ 0.48), while the genomic-based heritability (on the underlying scale) was ~ 0.47, and ~ 0.37 when excluding close relatives with equal affection status.

Conclusions: Screening for distichiasis in puppies before eight weeks of age is not sufficient to give an accurate estimate of the prevalence, and an additional examination after one year is recommended. The heritability of distichiasis is medium to high, showing that it should be possible to reduce the prevalence by selective breeding.

Keywords: Aberrant eyelash, Canine, Genomic heritability, Ocular disorder, SNP-based heritability

Background

Distichiasis is a condition of the eyelid with displaced eyelashes [1]. The aberrant hairs arise from ectopic hair follicles close to the meibomian glands in the eyelid and mostly emerge from the duct openings of the meibomian glands on the margin of the eyelid as single or multiple hairs [2, 3]. The disorder is common in dogs but has also been reported in cats [4], ferrets [5], and horses [6]. It is seen in both purebred and mixed-breed dogs [1, 2, 7]. The prevalence varies strongly between breeds, with 49.3% in English cocker spaniels [8], 11.4% in Tibetan terriers [9], and 27.9% in Elos [10]. Distichiasis is the most frequently diagnosed ocular disorder in the Norwegian Staffordshire bull terrier (SBT) population [11].

Distichiasis occurs in puppies as young as six weeks of age [12]. The majority of dogs diagnosed with distichiasis are only mildly affected, with subtle or no apparent clinical signs except for aberrant eye hair. Common clinical signs are conjunctivitis, irritation and rubbing of the eye, increased blinking, and lacrimation. In severe cases, the abnormal hair growth can lead to corneal lesions such as ulceration and keratitis [1, 12, 13]. There are several

*Correspondence: dijo@nmbu.no

[1] Department of Preclinical Sciences and Pathology, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, NMBU Veterinærhøgskolen, Oluf Thesens Vei 22, 1433 Ås, Norway
Full list of author information is available at the end of the article

different treatment methods, but recurrence and complications are associated with all procedures [3, 13–15]. Therefore, a reduction of the incidence by selective breeding would be advantageous.

Heritabilities (h^2) for distichiasis have been reported for: the Tibetan terrier ($h^2=0.043$, 849 dogs) [9], Elo ($h^2=0.238\pm0.122$, 234 dogs) [10], English cocker spaniels ($h^2=0.22$ and 0.51, 799 dogs) [8], and Havanaise ($h^2=0.276$ linear model and $h^2=0.720$ Bayesian threshold model, 1156 dogs) [16].

The main purpose of the present study was to estimate the additive heritability (h^2) and the prevalence of distichiasis in the SBTs in Norway and to explore whether it is realistic to reduce the incidence of distichiasis by selective breeding. In addition, we were interested in investigating the possibility of using genomic data to estimate the heritability of distichiasis in dogs. So far, the heritability of distichiasis has been estimated by using pedigree information only, and to the best of our knowledge, this is the first study to include genomic data in heritability estimates of an ocular disorder in dogs.

Methods

The study was based on official data from eye examination records registered by the Norwegian Kennel Club (NKK). ECVO-certified veterinarians performed the examinations using a bio-microscopical examination of the adnexal structures of the eye. The results were stored and are publicly available in "Dogweb",—a pedigree database maintained by the NKK (www.dogweb.no).

The primary dataset contained records from 2005 until May 2021 and comprised a total of 4752 eye examinations recorded in 4177 SBTs. 499 (10.5%) dogs had more than one eye examination (the cumulative numbers were: 1 dog = 5 examinations, 11 dogs = 4 examinations, 64 dogs = 3 examinations, and 499 dogs = 2 examinations). All dogs, both uni- and bilaterally affected, were counted once. The proportion of female dogs examined was 2196 (53%) and male dogs 1981 (47%). The age of the examined dogs ranged from 4.5 weeks to twelve years. 2894 (69%) of the dogs were examined before 58 days (~ 8 weeks).

At present, a positive distichiasis diagnosis recorded in the NKK will persist, regardless of the findings on later examinations. The affected dogs are graded as mildly or severely affected; earlier, the grade moderately affected was also included. We have treated the phenotype as a binary trait (affected/unaffected) due to missing categorisation in about 16% of the affected dogs, low numbers of the more severely affected dogs and previous practice in heritability estimates of distichiasis [8, 9, 16].

Descriptive statistics

The descriptive statistics calculating the prevalence and the effect of age and sex were performed using base R, the R package tidyverse, and epiR [17–19]. P values below 0.05 were considered statistically significant with a 95% confidence interval (CI).

The data was unbalanced, with a large group of dogs examined around eight weeks of age and several dogs with multiple examinations. Therefore, the data were stratified into three age groups: D1: dogs with a single standing examination between 0 and 58 days, D2: dogs examined between 59 and 364 days, and D3: dogs examined after 365 days (Table 1). There was missing information on the age at the examination in twelve dogs (one affected), and these were excluded from further analysis. Other age groups were assessed by splitting the dogs into six different age classes, 0–1, 1–2, 2–3, 3–4, 4–5 and > 5 years. The oldest age groups (dogs > 5 years) were merged due to small numbers of observations in this age

Table 1 Age distribution and the prevalence of distichiasis in the different age groups

Age group		Age (weeks)	Number of dogs	Number of dogs affected with distichiasis	Prevalence of distichiasis in %	Standard error	Confidence interval (95%)	Odds
	All Young examined ≤ 0.16 year	≤ 8.3	2894	83	2.87	0.003	2.29–3.54	0.03
D1	Young only examined once ≤ 0.16 year	≤ 8.3	2508	76	3.03	0.003	2.39–3.78	0.03
D2	Last examination between > 0.16 < 1 year	8.3–52	263	12	4.56	0.013	2.38 -7.83	0.05
D3	All dogs examined after one year	≥ 52	1394	261	18.72	0.01	16.71—20.87	0.23

The distribution of the examined dogs in the different age groups: All dogs < 0.16 years, D1, D2 and D3. The prevalence of distichiasis is given in the four age groups, including the standard error and the 95% confidence interval of the prevalence. The last column shows the odds of being affected with distichiasis in the different age groups

span. There was no significant difference in distichiasis status between any age groups above one year, and they were joined into one class, D3.

The effect of age and sex between the two groups D1 and D3 were estimated using logistic regression, including distichiasis as a response variable and the age groups and sex as the explanatory variables. We also compared the diagnosis in dogs with multiple examinations, once before 58 days and at least one examination after 364 days using the McNemar's test.

The effect of age at examination within the age groups D2 and D3 was estimated using logistic regression, including distichiasis as a response variable and age as the explanatory variable. Only one observation per dog was used. In dogs with multiple examinations, we used the age at the last presentation in unaffected and the age at first positive diagnosis in affected dogs. In case of inconsistency in the distichiasis status between two examinations, the more severe diagnosis was kept, affected individuals were considered once affected, always affected.

Estimates of pedigree-based heritability

The pedigree-based heritability (h^2_{ped}) estimating additive effects was conducted on 1391 dogs in age group D3. The youngest (D1) and the middle age group (D2) were excluded from further analysis due to the low prevalence in D1 and the low number of observations in D2.

The h^2_{ped} was estimated in the age group D3 using an average information restricted maximum likelihood approach (REML), analysed in the DMU package [20]. Afterwards, the heritability was converted to a theoretical underlying continuous scale (h^2_{pedT}) [21]. The model used was $Y = \mu + a + e$. Where Y = distichiasis status, μ = the mean term (fixed effect), a = the additive genetic effect (random effect), and e = the residual error (random effect).

Genomic heritability

The biological material was based on samples from a biobank established in collaboration between The Norwegian University of Life Sciences and the NKK. DNA was extracted using E.Z.N.A Blood DNA Mini Kit from Omega. The quality of the DNA was measured using Epoch from BioTek. A total of 681 dogs were genotyped (118 dogs with Illumina 170k CanineHD Bead chip and 629 dogs with Illumina 220k CanineHD Bead chip). All material was gathered in agreement with all relevant ethical guidelines and with the owners' written consent.

Quality control

Plink 1.9 and R were used for data management and quality control (QC) [17, 22, 23]. We performed a QC on each dataset before merging. Dogs with more than 5% missingness, a heterozygosity rate above three standard deviations from the mean, sex mismatches and duplicates were removed. Markers with a call rate below 95% and a minor allele frequency of ≤ 0.04 were removed. After QC, 611 dogs and 129,217 markers from the 220k and 92 dogs and 101,806 markers remained in the 170k dataset. After merging the two datasets, only markers in common between the two datasets were kept (93,973 markers), and a post-merge QC was performed with the same parameters as pre-merge QC, removing 37 markers with a minor allele frequency below 0.04 and 20 duplicated individuals. Multidimensional scaling plots were conducted to inspect potential differences (batch effects) between the two datasets (170k and 220k). Only dogs in D3 were kept for further analyses to make the genomic heritability estimates comparable to the pedigree-based estimates, excluding 133 individuals. The final dataset consisted of 93,936 markers and 548 dogs, where 228 were affected, and 320 were unaffected.

Genomic heritability estimates

The genomic heritability (h^2_g) was estimated using a genomic restricted maximum likelihood (GREML) model in Genome-wide Complex Trait Analysis software (GCTA). Using the model: $Y = \mu + g + e$, where Y = distichiasis status, μ = the mean term, g = the genetic effect based on the genomic relationship matrixes (GRM), and e = the residual. The variance estimates explained on the observed scale are transformed by GCTA to a modified version of the underlying scale adjusting for sample ascertainment caused by an increased number of cases in the sample compared to the actual population [21, 24].

As there is little knowledge about the genetic architecture of distichiasis, two GRMs were used. The first was calculated in GCTA, where the GRM is calculated from autosomal single nucleotide polymorphisms (SNP), and the SNPs are assumed to contribute an equal amount to the trait [24]. Due to the high level of linkage disequilibrium (LD) in dogs [25], we calculated an alternative GRM in Linkage Disequilibrium Adjusted Kinship software (LDAK). In LDAK, SNPs are weighted depending on the degree of LD in the region. SNPs in regions with high levels of LD receive a lower weight than SNPs in regions with a lower LD level, thus avoiding underestimating causal variants in areas with low levels of LD and overestimating variants in areas with high levels of LD [26].

To look at potential biases introduced by closely related individuals, we ran one analysis including all individuals and one analysis where siblings with equal affection status were removed (50 dogs). In sibling pairs with both affected and unaffected siblings, one of each was kept (36 sibling pairs), leaving 207 affected and 291 unaffected

individuals. The disease prevalence was set to 0.187 according to the prevalence among dogs examined after one year of age.

Results

Prevalence and grading

The proportion of dogs diagnosed with distichiasis each year ranged from 2.89% (2009) to 13.5% (2016) (Fig. 1). The total number of dogs diagnosed with distichiasis was 350, giving an overall prevalence of 8.38% (CI 7.56–9.26). The majority, 256 (73.14%) of the affected dogs, were marked as mildly affected, 31 (8.86%) dogs as moderate and, seven (2%) dogs severely affected, 56 (16%) dogs had no grading marked on the examination scheme.

Effect of age and distribution

The age distribution is displayed in Table 1 and Fig. 2. The prevalence of distichiasis was significantly higher in D3 (18.72%) compared with D1 (3.03%). Thus, there is a significantly increased risk of being diagnosed with distichiasis in D3 compared to D1 (P = < 2e−16, CI 1.71–2.24). Comparing 370 dogs examined both before 53 days and after one year of age gave a significantly increased risk of being diagnosed with distichiasis at the second examination, after one year of age (62 affected dogs) compared with the first examination (6 affected dogs) (P = 9.41e−15). There was a significantly increased risk of distichiasis with increasing age within D2 (P = 3.89e−05, CI 2.17–5.98), but not in D3 (P = 0.9, CI −0.10 to 0.08).

Re-examination

The diagnostic consistency among the 499 dogs with multiple examinations was investigated. The diagnosis changed in 82 (16.43%) of the re-examined dogs. Three dogs were first diagnosed with distichiasis, then diagnosed as free, while 79 dogs were first classified as unaffected and then as affected on a later examination. Most of these dogs (n = 63; 12.62%) were younger than eight weeks at the first examination, while only 11 dogs (2.2%) were older than one year at the initial assessment. In most cases, the grading is consistent, only two out of 350 affected dogs changed from mild to moderate, and one dog changed from moderately to mildly affected.

Sex

The number of female cases was 220, and male cases were 130 (Fig. 3). There was a greater number of female dogs (877) examined after one year of age than males (517). Female dogs were also more frequently re-examined than male dogs; 301 females had a second examination as opposed to only 198 males. The proportion of affected dogs in the two sexes was approximately the same in the

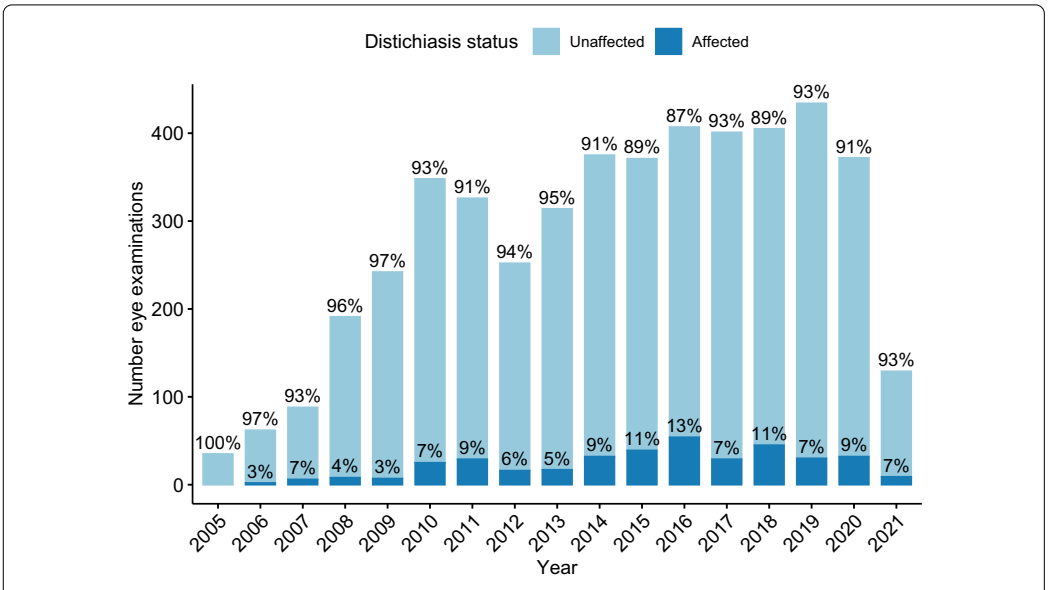


Fig. 1 Number of eye examinations pr year

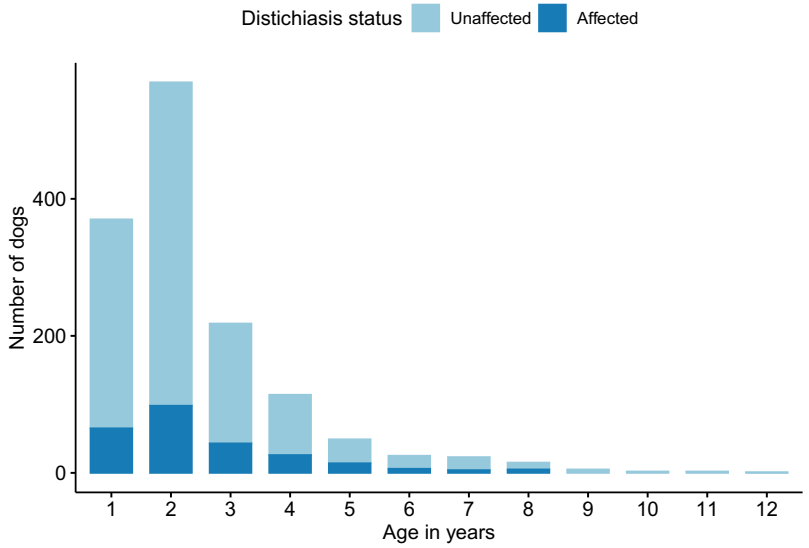


Fig. 2 Age distribution of dogs examined after 1 year of age

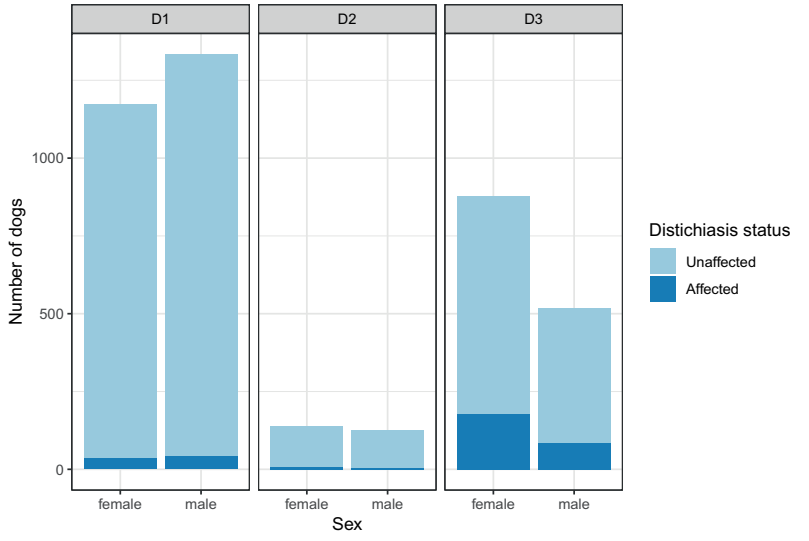


Fig. 3 The distribution of sex in the age groups. D1 (0 < days ≤ 58), D2 (58 > days < 365) and D3 (days ≥ 365)

three age groups. The logistic regression model showed no significant effect of sex when correcting for age (P = 0.14, CI −0.43 to 0.06).

Estimates of heritability

The additive heritability estimate using pedigree data (age group D3, linear model) was on the observed scale

$h^2_{pedO} \sim 0.22$, SE 0.05, and after transforming to the underlying scale, $h^2_{pedT} \sim 0.48$, SE 0.11. The genomic heritability estimates using the different GRMs and the two different models, including all dogs and excluding siblings of equal affection status, are represented in Table 2. The average estimated genomic heritability in age group D3 was on the underlying scale $h^2_{gT} \sim 0.47$ SE 0.10, and after removing siblings, $h^2_{gT} \sim 0.37$ SE 0.11.

Discussion

The current study showed a high prevalence of distichiasis in the Norwegian SBT population. Most cases were only mildly affected. More than 50% of the dogs in this study were examined at around eight weeks of age as part of a screening for inherited eye disorders. This screening captures inherited eye diseases such as distichiasis and persistent hyperplastic *tunica vasculosa lentis*/persistent hyperplastic primary vitreous (PHTVL/PHPV). Our study indicates that this early screening has a limited predictive value for a distichiasis diagnosis in grown-up dogs, and the probability of being diagnosed with distichiasis is significantly higher after one year compared with young puppies. After one year of age, there is no clear relationship between increased age and a positive diagnosis. However, the data about dogs older than five years is sparse.

Due to the low prevalence and limited predictive value in the youngest age groups, the pedigree-based heritability estimates were based on dogs examined after 52 weeks. We based our estimates on a linear model, as Bellamy et al. showed a good agreement between heritability estimates from the Bayesian threshold model and heritability from linear models converted to the underlying scale [16]. The estimated heritability of $h^2_{pedO} = \sim 0.22$ and $h^2_{pedT} \sim 0.48$ on the underlying scale agrees with other studies of dogs [8, 10], even if there is some variation between breeds. While Ketteritzsch estimated a lower heritability in Tibetan terriers [9], our estimated pedigree-based heritability is slightly lower than in the study of Bellamy et al. [16].

The genomic-based heritability was estimated to be $h^2_{gT} \sim 0.47$, and after removing siblings with equal affection status $h^2_{gT} \sim 0.37$ on the underlying scale. There were only minor variations between the methods. Using the GRM calculated in LDAK gave a slightly higher h^2_{gLDAK} value than the h^2_{gGCTA} using the GRM calculated in GCTA. The genomic heritability estimates have a relatively large standard error, and a larger sample size could reduce the standard error. The genomic-based heritability is estimated on a subset of the dogs used in the pedigree bases estimates, and the estimates are on the same level as the pedigree-based heritability when siblings with equal affection status are included. This is interesting as genomic heritability estimates tend to underestimate the heritability compared to traditional methods using pedigree data [27, 28]. A possible bias in our genomic heritability estimates is the degree of relationship in the data. In humans, genomic heritability estimates usually only include unrelated individuals [24, 27]. Within a dog breed, the average relationship is usually much higher than in humans. Removing unrelated individuals would lead to a low sample size and decreased power. Including close relatives can bias the results upwards due to a shared environment, but epistasis and dominance might also have an effect [29–31]. We removed sibling pairs with the same affected state to account for some biases introduced by close relationships. Including all siblings in the calculation increased the estimates by around 10%. In livestock, it is not uncommon to combine pedigree and genomic data in a single-step analysis to give a more accurate heritability estimate without having the cost of genotyping the whole breeding stocks [32]. The single-step method is an attractive method as long as there is no genomic selection of the trait [33]. Including pedigree information in addition to the genomic data has been shown to improve the precision of heritability estimates in dairy cattle [34].

Table 2 The genomic heritability estimates including all models in age group D3

GRM	Model	Heritability estimates observed scale (h^2_{gO})	Standard error	Heritability estimate transformed scale (h^2_{gT})	Standard error
GCTA	All dogs included	0.345	0.078	0.461	0.104
LDAK	All dogs included	0.357	0.079	0.476	0.106
GCTA	Full siblings with equal affection status excluded	0.275	0.082	0.368	0.109
LDAK	Full siblings with equal affection status excluded	0.281	0.083	0.375	0.112

The heritability estimates for the different models with and without full siblings, using two different genomic relationship matrices (GRM) calculated in GCTA and LDAK. h^2_{gO} is the genetic heritability estimate on the observed linear scale, and h^2_{gT} is the genetic heritability estimate on the transformed underlying scale

The material for the pedigree-based estimates consisted of 1391 dogs, a relatively low proportion of the total population. There is no data about the population size, but in the last years, there have been approximately 1000 newly registered SBTs per year in Norway [35]. With an overall low number of dogs examined after one year of age, we cannot be certain that our data are representative of the whole population. We assume that most dogs intended for breeding undergo at least one eye examination as an adult. Thus, we believe our results are valid for the breeding population, the genetic basis for the next generation. We have no reason to think that breeding dogs are more prone to develop distichiasis than other family dogs. As most of the affected dogs are only mildly affected with no clinical signs, distichiasis would not commonly be a reason for an eye examination.

There are several challenges related to the diagnostics of distichiasis, which might lead to a false-negative diagnosis, including patient cooperation, the experience of the veterinarians [9], and the lifecycle of the eyelash, with shedding and regrowth. Also, single hairs not detectable at the time of examination may appear later. Gómez found that previously undetected hair could emerge by manipulating the eyelid during surgery [3]. Also, Lawson noted that it is not uncommon that more delicate hair and those just starting to appear are not observed at the initial examination [1]. At last, there is the possibility that the owner intentionally removes the *distichiae.*

The overall prevalence of distichiasis seems persistent over the years. A reduction of the frequency could be expected by following the breeding advice stated by The Norwegian Terrier Club (NTC) and ECVO: excluding severely affected dogs from breeding and only breed mildly affected dogs with unaffected dogs [36, 37]. The number of newly registered SBTs has increased 19 folds in Norway over the last 20 years [35]. The rapid growth in the popularity of the SBT might have led to high pressure on the breeding stock and less stringent selection against unfavourable traits like distichiasis, thus maintaining the high prevalence of the condition. According to the NTC's breeding strategy, distichiasis is not considered a main concern, but breeders are encouraged to exercise prudence [36]. Distichiasis usually has few clinical implications and might not be the main focus of the breeders. Among the 350 affected dogs, NKK has registered 68 affected dogs used in breeding, where most of these dogs were only mildly affected (52), and no breeding dogs were severely affected [11].

The present information about the medium to high heritability of distichiasis shows that it is possible to reduce the prevalence of the disease by selective breeding. We recommend excluding all severely affected dogs, and careful consideration must be given when breeding mildly affected dogs. In a future breeding strategy, further restrictions than those implemented by NKK and NTC at present can be evaluated after carefully considering the clinical impact of distichiasis and the presence of other important traits under selection.

Conclusions

We found a significant increase in the disease prevalence from eight weeks to one year of age, showing that screening young puppies around eight weeks is insufficient to give an accurate "lifetime" diagnosis and might underestimate the prevalence of distichiasis in the breed. Estimates of heritability based on pedigree data and genomic SNP data indicate that the heritability of distichiasis is moderate to high. This shows that it can be possible to reduce the prevalence by weighting the trait when selecting parents for breeding. Interestingly, the genome-based heritability estimates are on the same level as the pedigree-based heritability estimates, even though the sample size is small, and the standard errors are relatively high.

Acknowledgements
Not applicable.

Prior publication
The data have not been published previously.

Author contributions
DJ performed the descriptive statistics, prepared the samples, analysed the genomic data, and drafted the manuscript, FL had the conceptual idea, prepared the pedigree data and was a major contributor to writing the manuscript, PM performed the pedigree-based heritability estimates and contributed to writing the manuscript, EOR contributed to writing the manuscript. All authors have taken part in discussing and revising the manuscript. All authors read and approved the final manuscript.

Funding
The project received financial support from Svenska Kennelklubben (SKK) and Agria Pet Insurance (project no N2021-0020). The funders had no role in the project nor in the decision to publish.

Availability of data and materials
The eye examination records and pedigree data are available through the NKKs online database, "dogweb", at: www.dogweb.no. The genomic datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
All data has been gathered in agreement with all applicable ethical guidelines. The phenotype and pedigree data were based on publicly available data, while the genomic material was based on samples from the biobank at The Norwegian University of Life Sciences. By submission of the blood samples, the owners sign a written consent allowing the DNA to be used in research.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Preclinical Sciences and Pathology, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, NMBU Veterinærhøgskolen, Oluf Thesens Vei 22, 1433 Ås, Norway. [2]Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark. [3]Evidensia, Oslo Dyresykehus, Ensjøveien 14, 0655 Oslo, Norway.

Received: 24 July 2022 Accepted: 14 November 2022
Published online: 21 November 2022

References

1. Lawson DD. Canine distichiasis. J Small Anim Pract. 1973;14:469–78.
2. Raymond-Letron I, Bourges-Abella N, Rousseau T, Douet JY, de Geyer G, Regnier A. Histopathologic features of canine distichiasis. Vet Ophthalmol. 2012;15:92–7.
3. Palella Gómez A, Mazzucchelli S, Scurrell E, Smith K, Pinheiro de Lacerda R. Evaluation of partial tarsal plate excision using a transconjunctival approach for the treatment of distichiasis in dogs. Vet Ophthalmol. 2020;23:506–14.
4. Reinstein SL, Gross SL, Komáromy AM. Successful treatment of distichiasis in a cat using transconjunctival electrocautery. Vet Ophthalmol. 2011;14(Suppl 1):130–4.
5. Verboven CA, Djajadiningrat-Laanen SC, Kitslaar WJ, Grinwis GC, Schoemaker NJ, Boevé MH. Distichiasis in a ferret (*Mustela putorius furo*). Vet Ophthalmol. 2014;17:290–3.
6. Hisey EA, Hermans H, Lounsberry ZT, Avila F, Grahn RA, Knickelbein KE, et al. Whole genome sequencing identified a 16 kilobase deletion on eca13 associated with distichiasis in Friesian horses. BMC Genomics. 2020;21:848.
7. D'Anna N, Sapienza JS, Guandalini A, Guerriero A. Use of a dermal biopsy punch for removal of ectopic cilia in dogs: 19 cases. Vet Ophthalmol. 2007;10:65–7.
8. Petersen T, Proschowsky HF, Hardon T, Rasch SN, Fredholm M. Prevalence and heritability of distichiasis in the English cocker spaniel. Canine Genet Epidemiol. 2015;2:11.
9. Ketteritzsch K, Hamann H, Brahm R, Grussendorf H, Rosenhagen CU, Distl O. Genetic analysis of presumed inherited eye diseases in Tibetan terriers. Vet J. 2004;168:151–9.
10. Kaufhold J, Hamann H, Steinbach G, Gordon S, Brahm R, Grussendorf H, et al. Analyse der Prävalenz von Distichiasis bei der Hunderasse Elo. Prakt Tierarzt. 2006;119.
11. Norwegian Kennel Club. Dog web 2021; https://www.dogweb.no. Accessed 12 May 2021.
12. Gelatt KN, Gilger BC, Kern TJ. Veterinary ophthalmology: Volume 2. 5th ed. Ames: Blackwell; 2013. p. 832–93.
13. Gelatt KN. Veterinary ophthalmic surgery. Gelatt JP, editor. S.l.: Elsevier Saunders; 2011. p. 100–3.
14. Ioannides J, Everson R, Matas Riera M, Dawson C. A description of blend electrolysis for treatment of canine distichiasis: 78 cases (2012–2017). Vet Rec. 2020;187: e102.
15. Zimmerman KL, Reinstein SL. Evaluation of transconjunctival thermal electrocautery for treatment of canine distichiasis: 88 eyelids (2013–2016). Vet Ophthalmol. 2019;22:50–60.
16. Bellamy KKL, Lingaas F, Madsen P. Heritability of distichiasis in Havanese dogs in Norway. Canine Med Genet. 2021;8:11.
17. R Core Team. R: A language and environment for statistical computing. 2021; https://www.R-project.org/
18. Wickham, et al. Welcome to the tidyverse. JOSS. 2019;4:1686.
19. Stevenson M. ESw, contributions from Telmo Nunes C, Heuer JM, Javier, Sanchez RT, Jeno Reiczigel, Jim Robison-Cox PS, Peter Solymos KY, Geoff, et al. Epir: Tools for the analysis of epidemiological data. 2021; https://CRAN.R-project.org/package=epiR.
20. Madsen P, Jensen J. A user's guide to DUM. A package for analysing multivariate mixed models version. 2013;6:1–33
21. Dempster ER, Lerner IM. Heritability of threshold characters. Genetics. 1950;35:212–36.
22. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation plink: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:s13742.
23. Shaun Purcell CC. Plink 1.9. https://www.cog-genomics.org/plink/1.9/. Accessed 04 Oct 2021.
24. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.
25. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, et al. Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. Genome Res. 2004;14:2388–96.
26. Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. Am J Hum Genet. 2012;91:1011–21.
27. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565–9.
28. Mills M, Mills MC, Barban N, Tropf FC. An introduction to statistical genetic data analysis. Cambridge: The MIT Press; 2020. p. 4–31.
29. Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th ed. Harlow: Longman Group; 1996.
30. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, et al. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet. 2013;9: e1003520.
31. Oea Z. The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci USA. 2012;109:1193–8.
32. Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. Gen Sel Evol. 2010;42:2.
33. Wang L, Janss LL, Madsen P, Henshall J, Huang CH, Marois D, et al. Effect of genomic selection and genotyping strategy on estimation of variance components in animal models using different relationship matrices. Gen Sel Evol. 2020;52:31.
34. Haile-Mariam M, Nieuwhof GJ, Beard KT, Konstatinov KV, Hayes BJ. Comparison of heritabilities of dairy traits in Australian Holstein-Friesian cattle from genomic and pedigree data and implications for genomic evaluations. J Anim Breed Genet. 2013;130:20–31.
35. Norwegian Kennel Club. 2021; https://www.nkk.no/statistikk/category1098.html. Accessed 4 Mar 2021.
36. The Norwegian Terrier Club. Rasegruppens anbefalinger for avl. 2021; https://norskterrierklub.no/Userfiles/Sites/files/Rasegruppens%20anbefalinger%20for%20avl(4).pdf. Accessed 5 Apr 2022.
37. ECVO, 2021. ECVO Manual: Chapter 8 - The Veterinary ophthalmologists' advice. 2021; https://www.ecvo.org/media/8-vet_advice.pdf. Accessed 7 Aug 2021.

Publisher's Note

Paper II

Jørgensen *et al.*
Canine Medicine and Genetics (2023) 10:9
https://doi.org/10.1186/s40575-023-00132-1

Canine Medicine and Genetics

RESEARCH

Open Access

Genomic analysis and prediction of genomic values for distichiasis in Staffordshire bull terriers

Dina Jørgensen[1*], Ernst-Otto Ropstad[2], Theodorus Meuwissen[3] and Frode Lingaas[1]

Abstract

Background Distichiasis is a condition characterized by aberrant hairs along the eyelid margins. The symptoms are usually mild but can lead to ulcerations and lesions of the cornea in severe cases. It is the most frequently noted ocular disorder in Norwegian Staffordshire bull terriers (SBT), with a prevalence above 18% in the adult population. A complex inheritance is assumed, but there is sparse knowledge about the genetic background of distichiasis in dogs. We have performed a genome-wide association study of distichiasis in SBT and used genomic data in an attempt to predict genomic values for the disorder.

Results We identified four genetic regions on CFA1, CFA18, CFA32 and CFA34 using a mixed linear model association analysis and a Bayesian mixed model analysis. Genomic values were predicted using GBLUP and a Bayesian approach, BayesR. The genomic prediction showed that the 1/4 of dogs with predicted values most likely to acquire distichiasis had a 3.9 -4.0 times higher risk of developing distichiasis compared to the quarter (1/4) of dogs least likely to acquire the disease. There was no significant difference between the two methods used.

Conclusion Four genomic regions associated with distichiasis were discovered in the association analysis, suggesting that distichiasis in SBT is a complex trait involving numerous loci. The four associated regions need to be confirmed in an independent sample. We also used all 95 K SNPs for genomic prediction and showed that genomic prediction can be a helpful tool in selective breeding schemes at breed level aiming at reducing the prevalence of distichiasis in SBTs in the future, even if the predictive value of single dogs may be low.

Keywords Canine, Genomic prediction, Distichiasis, GWAS, Staffordshire bull terrier

Plain English Summary

Distichiasis is a condition where abnormal hairs grow along the margin of the eyelids. It's common in Staffordshire bull terriers and can cause eye problems of variable severity. The abnormal eye hairs can be found during an eye inspection performed by a veterinarian.

We performed a genome-wide association analysis and identified four genomic areas associated with the condition. But more genes may be involved in causing the disease.

We have used genomic data to predict genomic values. Genomic values can be used to predict the total load of disease-associated alleles. Genomic prediction would therefore be helpful at the breed level, similar to pedigree-based

*Correspondence:
Dina Jørgensen
dijo@nmbu.no
Full list of author information is available at the end of the article

breeding values, to reduce the prevalence of dogs with distichiasis, even if the low accuracy to predict phenotypes in individual dogs may be a challenge. More research is needed to confirm these findings and see if genomic prediction could be a helpful tool within dog breeding in the future.

Background

Distichiasis is a condition with abnormal growth of eye hairs along the margins of the eyelid. The aberrant hairs arise from ectopic hair follicles near the meibomian glands and emerge through the excretory duct opening of the sebaceous glands [1, 2]. In most cases, the symptoms caused by distichiasis are mild. Eye irritation with increased lacrimation and conjunctivitis can be seen. In severe cases, the aberrant eye hair can lead to lesions of the cornea with ulcerations and keratitis [3].

Distichiasis is common in dogs [3, 4] and the most frequently noted ocular disorder in Staffordshire bull terriers (SBT) in Norway [5]. In a previous study, we found a prevalence of 18.72% among Norwegian SBTs examined after one year of age, and the heritability was estimated to be moderate to high [6]. The same level of heritability has been seen in the dog breeds; havanais [7], elo [8] and cocker spaniels [9]. A simple Mendelian inheritance was excluded in a segregation analysis in elos, however, the exact mode of inheritance was not defined [10]. A complex mode of inheritance involving multiple genes is assumed. Thus far, little is known about the genetic background of distichiasis in dogs.

Distichiasis is less common in other species than the dog but has been described in cats [11], ferrets [12], cattle [13, 14], and horses [15]. In Friesian horses, Hisey et al. found a 16 kb deletion in an intergenic region on equine chromosome 13 associated with distichiasis, and a dominant inheritance with incomplete penetrance is assumed [15]. In cattle, distichiasis has been associated with the autosomal dominant *Polled* locus on the bovine chromosome 1 [14]. In humans, distichiasis has been associated with an autosomal dominant mutation in the region of the *FOXC2* gene, both alone and as a part of a syndrome with lymphedema [16–18]. Other rare conditions in humans seen in combination with distichiasis are facial dermal dysplasia caused by a frameshift mutation in *TWIST2* [19] and Blepharocheilodontic syndrome linked to mutations in *CTNND1* and *CDH1* encoding proteins in the cadherin–catenin complex [20]. So far, no genes or genetic regions have been found to be associated with distichiasis in dogs.

We were interested in using distichiasis as a model for canine genomic prediction, by estimating the joint effect of all genomic markers to predict a phenotype. Genomic predictions have had great success in livestock breeding [21]. Until now, genomic predictions have received little attention in dog breeding. There have been a few studies using genomic SNP data to predict disorders such as canine hip dysplasia [22, 23], cranial cruciate ligament rupture [24], and kidney disease [25]. Thorsrud et all. compared genomic best linear unbiased prediction (GBLUP) with four different machine learning techniques to predict distichiasis, mandibular distocclusion and acral lick dermatitis in a guide dog population consisting of German shepards, golden retrievers, Labrador retrievers, and Labrador and golden retriever mixes [26].

In the present study, we aim to investigate the genetic background of distichiasis in SBTs through a genome-wide association analysis (GWAS). We have compared two approaches for genomic prediction: GBLUP and BayesR and have used the results to investigate the potential value of using genomic data to predict genomic values for the disorder in SBTs.

Results

Genome-wide association study

The association study was based on 731 SBTs (407 controls and 324 cases), and 94,697 autosomal markers. Four genomic regions were identified using a mixed linear model-based association analysis (MLMA) in Genome-wide Complex Trait Analysis (GCTA) [27], located on CFA1, CFA18, CFA32 and CFA34 (Fig. 1, Table 1). The same genomic regions on CFA1, CFA18 and CFA34 obtained an absolute effect size above 0.005 in BayesR. However, the SNP on CFA32 received a lower signal in the Bayesian model (Fig. 2, the posterior probability is displayed in Supplementary Fig. 1). The total load of the top four risk alleles in cases and controls is presented in Fig. 3. There is a significant difference in the total risk allele load between cases and controls ($p < 2.0 \times 10^{-16}$). On average, affected dogs carry 4.3 risk alleles and unaffected 3.3. risk alleles.

The SNP on CFA1, BICF2P714726 attains the strongest effect size in BayesR. It is situated in an intergenic region, flanked by the genes *GAS1* ~ 260 kb downstream and *TUT7* ~ 210 kb upstream for the SNP (haploblock structure within the region is displayed in Supplementary Fig. 2).

The top SNP on CFA18, BICF2P1386405, is also situated in an intergenic region, the closest gene is *LRRC4C* 123 kb upstream, and the next gene, *API5* is situated ~ 1.2 MP downstream. The top SNP lies in a haploblock with five other SNPs spanning a distance of 134 kb. All six

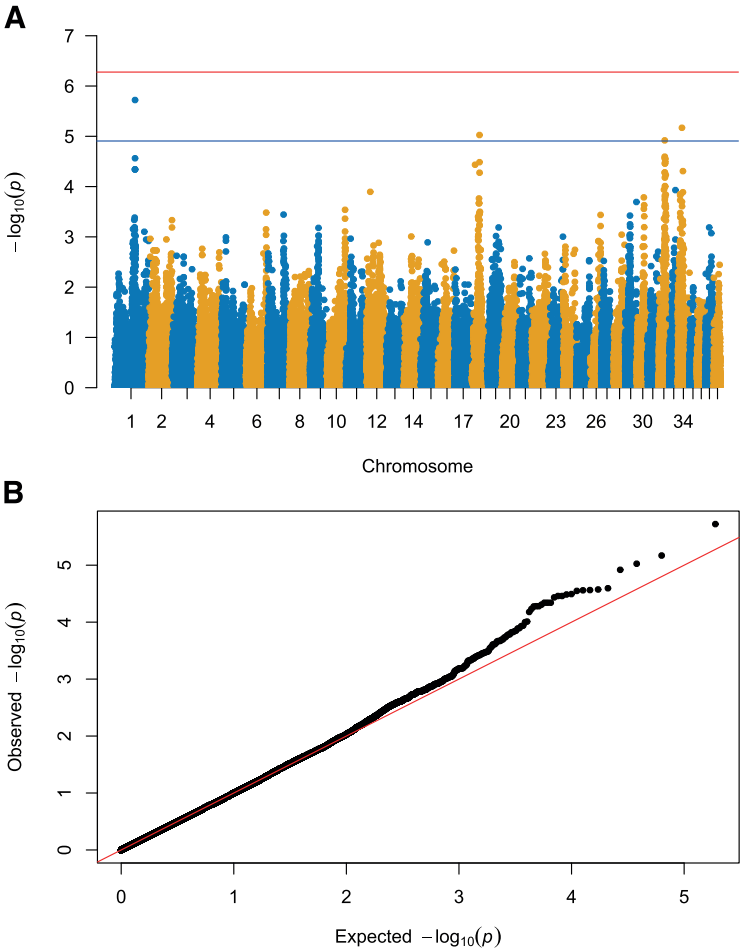


Fig. 1 Mixed linear model association analysis. **A:** A Manhattan plot displaying the MLMA performed in GCTA. The association analysis was based on 94,697 SNP markers and 731 dogs (324 cases and 407 controls). The significance level (blue line) was set to 1.24×10^{-05} using Bonferroni correction to adjust for multiple testing considering the LD, haploblock structure and number of independent SNPs after pruning the data. A second significance level (red line) was set to 5.28×10^{-07}, using Bonferroni correction to adjust for all markers in the data. **B:** A quantile–quantile (q-q) plot showing the expected *p* value against the observed *p*-values of the MLMA

Table 1 The top four associated SNPs

Chr	SNP	BP position	Risk allele / Protective allele	Risk allele frequency affected	Risk allele frequency unaffected	OR	95% CI	*P*-value (MLMA)	Absolute effect size (BayesR)
1	BICF2P714726	73,777,342	G /A	0.94	0.85	2.66	1.84—3.85	1.90×10^{-06}	0.017
18	BICF2P1386405	27,949,474	T/C	0.34	0.20	2.03	1.60—2.57	9.42×10^{-06}	0.007
32	BICF2G630590287	17,605,832	T/C	0.33	0.21	1.84	1.46—2.33	1.21×10^{-05}	0.003
34	BICF2S232639	14,960,862	G/A	0.54	0.38	1.92	1.55—2.36	6.77×10^{-06}	0.010

The four top SNPs identified in the association analysis in the MLMA in GCTA, and with the effect size from BayesR. The base pair position is given in can.fam4 reference genome. The association study was based on 731 phenotyped SBTs (407 controls and 324 cases), and 94,697 autosomal SNP markers

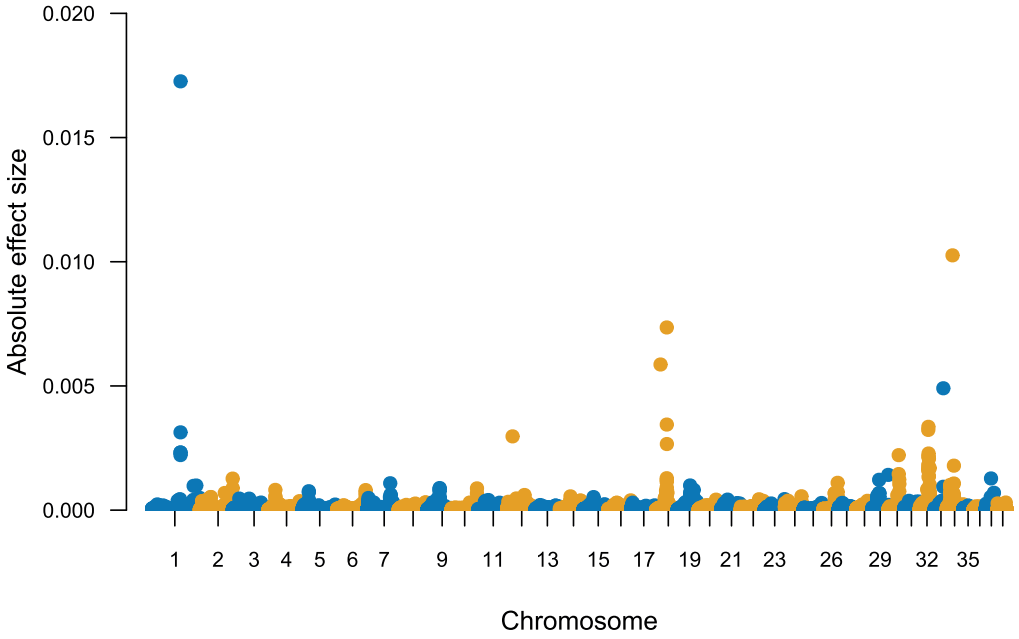


Fig. 2 A Manhattan plot of the absolute SNP effect estimated in BayesR over the 38 autosomal chromosomes. The analysis was based on 97,185 SNP markers and 731 dogs (324 cases and 407 controls)

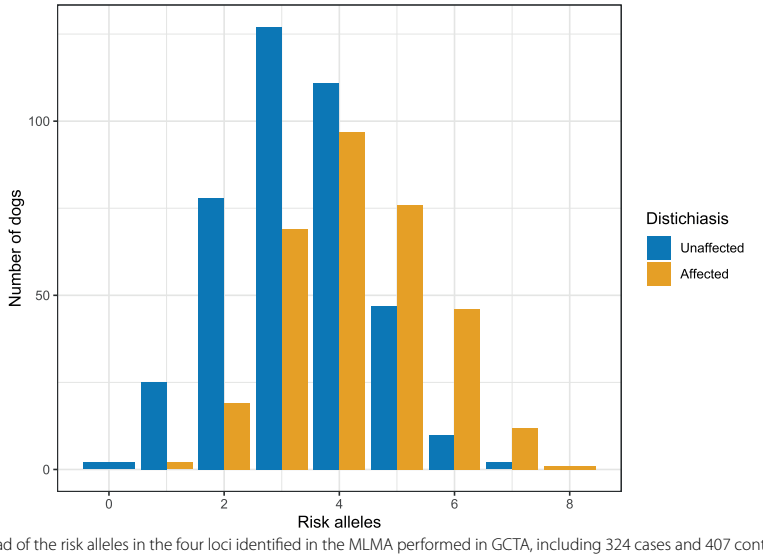


Fig. 3 The total load of the risk alleles in the four loci identified in the MLMA performed in GCTA, including 324 cases and 407 controls

SNPs within the haploblock lie in the same intergenic region (Supplementary Fig. 3).

BICF2G630590287 on CFA32 obtained the lowest signals in the MLMA and was less distinct in BayesR. The top SNP is situated in an intron of the *EMCN* gene (Supplementary Fig. 4).

The SNP on CFA34, BICF2S232639, reached the second highest signal and is situated in an intron in the *TPK1* gene, and lies within a haploblock with nine adjacent SNPs spanning a distance of 850 kb (Supplementary Fig. 5). Other genes within the haploblock are *SOX2* 65 kb (upstream) and *DNJAC* (583 kb), and *FXR1* (582 kb) (both downstream).

Population structure and relationships

Visual inspection of the principal components analysis (PCA) plots showed no general stratification. PCA plots including country of origin, genotyping arrays and cases and controls, are included in Supplementary Figs. 6 and 7. According to the registration number, the genotyped dogs are mainly Norwegian, followed by dogs from Sweden, reflecting the true population. The mean heterozygosity rate among the dogs in the dataset was 0.35. To assess some of the family structures in the dataset, 79 dogs that had an equal affection status as another sibling were excluded from an additional analysis (Supplementary Fig. 8). The SNP on CFA1 and CFA18 had a reduced significance compared with the analysis, including all dogs, while the SNPs on CFA32 and CFA34 had slightly higher significance. The top SNPs in the regions remained constant. The minor allele frequency of the four top SNPs between the two arrays was found to be at a similar level (Supplementary Table 1).

Prediction of genomic values

Genomic values of the dogs were predicted using GBLUP in GCTA and BayesR when their phenotypes were masked in the six-fold cross-validation design. The two methods were compared by calculating the area under the curve (AUC) from a receiver operating characteristic curve (ROC curve). There was no significant difference ($P=0.984$) in the AUC between the two methods. The AUC was 0.655 (CI 0.612–0.699) in GBLUP (Supplementary Fig. 9) and 0.651 (CI 0.607–0.695) in BayesR (Supplementary Fig. 10). Both methods gave a significant difference between the predicted genetic value (GV) in cases and control $P=4.12\times10^{-10}$ in GBLUP and $P=3.56\times10^{-08}$ using BayesR. Comparing the odds of the 25% of dogs with the highest GV estimate (most likely to develop distichiasis), with the 25% of dogs with a GV least likely to develop the disease, the odds ratio was 4.02 (95% CI 2.48–6.63) in GBLUP and 3.86 (95% CI 2.31- 6.55) in BayesR. No covariates had any effect on the performance of the models (Supplementary Table 2).

Discussion

We have identified four potential candidate regions on CFA1, 18, 32 and 34 associated with distichiasis. The four risk alleles represent four novel genomic regions associated with distichiasis. *CTNND1,* seen in connection with "Blepharocheilodontic syndrome" with distichiasis in humans [20], is situated on CFA18 but more than 10 MB upstream from the top SNP on CFA18. The genetic mechanisms for developing distichiasis in humans, bovines and Friesian horses appear different from those in SBTs since none of the identified loci in these species overlaps with the loci identified in this study. The diversity of genes and genetic regions associated with distichiasis implies a significant genetic heterogeneity, where multiple loci lead to similar phenotypes.

There are several genes of interest within the associated genomic regions. *GAS1,* located around 260 kb downstream of the top SNP on *CFA1*, is involved in growth suppression, apoptosis and embryonal development [28]. *SOX2* is located 65 kb away from the top SNP on CFA34. *SOX2* is a transcription factor involved in regulating embryonic development [29]. On CFA18, *API5* is the closest gene to the top SNP. *API5* is an apoptosis inhibitor [30].

A GWAS intends to detect markers in LD with the causal variant. The LD within a single dog breed can be extensive and span regions of several megabases [31, 32]. This makes it challenging to pinpoint the actual causal variant. Adding related dog breeds could break up stretches of LD and might help identify causal variants.

We have set the significance level at 1.24×10^{-05}, according to the number of independent SNPs after pruning the data, and after considering the LD structure and haploblock sizes. This is the same levels as suggested by Karlsson et al. based on the average size of 1 MB of independent haploblocks in a 2.4 GB dog genome [33], and the significase level sugested by Hayward et al. within breeds [34]. This significance level is, however, less stringent than a Bonferroni correction based on the number of all 94,697 markers in the dataset, which assumes all these markers are independent. When using a Bonferroni correction, none of the four genomic regions reaches significance.

The dataset contained genotyped dogs from two Illumina arrays, only the SNPs shared between the two arrays were used. Potential batch effects were assessed during the quality control. Including batch effect as a covariate did not have any effect on the genomic prediction. There is a high level of relationship between the dogs in the dataset. The GRM included in the mixed

linear model account for part of this relationship. However, keeping siblings from the same litter may introduce some bias due to shared environment in early life and maternal effects. To account for such effects, we ran one analysis after removing siblings of equal affection status; however, the same four top SNPs remained.

In our study, the four genomic regions disclosed contribute only with a moderate effect on the phenotype. The top SNP on CAF32 obtained the lowest significance among the top SNPs in the MLMA and received low signals in BayesR. At the top SNP on CFA1, the minor allele (A) is protective, and carrying the major risk allele, G, gave 2.66 increased odds of developing distichiasis. In the other three SNPs, the minor allele is the risk allele and carrying one of the risk alleles gives a twofold increased risk of developing distichiasis. The chance of developing distichiasis increased with the number of risk alleles.

In a previous study conducted on the same SBT population, we found that most SBTs was only mildly affected by distichiasis [6]. Additionally, it has been observed that single *distichia* may be difficult to observe [2, 35]. As a result, some false negative controls are expected. However, the previously estimated heritability from a subset of the same SBT population using both SNP data and pedigree data was between ~ 0.37 and ~ 0.48 [6]. These estimates are consistent with other findings in the literature, which indicates that any recording errors are not so substantial that they reduce the genetic versus error variance ratio.

According to the registration in The Norwegian Kennel Club (NKK), the prevalence of distichiasis in the SBTs has persisted over the last twenty years [5]. Marker-assisted DNA- testing could help identify dogs with an increased risk of carrying disease alleles and identify the best dogs for breeding. The use of DNA-based risk tests for complex traits is challenging due to multiple casual loci with varying effect sizes. Additionally, the predictive value may differ between distinct populations due to differences in LD between the markers and the causal loci.

In complex traits where the effect size of most risk alleles is small and therefore not captured by the association analysis, genomic prediction, including the combined effect of all SNP markers across the whole genome, can be used to predict phenotypes [36]. Since breeding populations in dogs are often small compared to humans and livestock, and there are few examples of genomic prediction in dogs, our intention was to compare the two methods, BayesR and GBLUP, to predict genomic values for distichiasis in SBTs. BayesR has been shown to give more accurate predictions in human disease traits with loci of large effects compared with traditional mixed models [37]. The two approaches, GBLUP and BayesR, performed similarly within our dataset, which may be because our dataset was too small to accurately distinguish non-causal loci from causal loci with moderate to small effects.

There was a significant difference between the GV predicted from all 95 K SNPs between the cases and controls. Comparing the 25% of dogs most likely to acquire distichiasis with the 25% of dogs with the "best" GV (least likely to develop the disease) gave four times increased risk of developing distichiasis in the first group. The predictive accuracy for the individual dog was low and can, therefore, not be used to predict the phenotype in individual dogs. However, using the GVs at the breed level in the same manner as traditional pedigree-based breeding values, it should be possible to reduce the prevalence of distichiasis in the SBT population. Even if, on average, there would be an improvement in the population, the number of dogs with a "false" prediction may represent a challenge for the communication with the breeders.

Prediction of complex traits using genomic data typically requires large training datasets and testing in independent data sets [38–40], which may be challenging in small dog breeds. Edwards et al. [41] demonstrated that combining genomic predictions from two dog populations from different countries, even within the same breed, can reduce the prediction accuracy. This may be due to differences in LD between the two populations, different genes being important for the disease in the populations and recording differences. There is, therefore, a great need to evaluate the benefit of genomic selection in dog populations, and how to combine data across populations.

The material includes only dogs with a phenotype and represents only a subset of the overall SBT population. However, because most dogs used for breeding undergoes an eye examination, we believe the material is representative of the breeding population.

The use of imported SBTs in breeding is extensive in Norway, and more than half of the litters are from combinations where at least one parent is registered in another country. This can increase the genetic variation within the population and reduce the accuracy of genomic prediction.

Thorsrud et al. reported a higher AUC (0.94 with GBLUP) in their genomic prediction of distichiasis compared to our AUC of 0.66 with GPLUP. The divergent results between the SBT and guide dog populations emphasize that the results from genomic predictions of one disease trait may not be easily transferable between different breeds or populations and depend on the heritability and genetic complexity of the disease, number of disease cases and controls, and population structure and effective population size.

Conclusion

Our study indicates that distichiasis in SBT is a complex trait with multiple genetic loci involved. We have identified four potential genomic regions on CFA1, 18, 32 and 34. Further studies must be conducted to validate the findings.

The genomic prediction, estimated from the joint effect of all 94,697 SNP, has the potential to aid in selective breeding, to reduce the prevalence of distichiasis in the SBTs but has a low predictive value for phenotypes in individual dogs. The genomic prediction of distichiasis must be validated in each other target population.

Material and methods

A subset of SBTs with an official eye examination record registered by NKK between 2005 and April 2022 were included. Eye examinations were performed by veterinarians certified by the European College of Veterinary Ophthalmologists (ECVO). The eye examination records are available in "dogweb", an open database established by NKK (www.dogweb.no). The dogs were classified as affected or unaffected according to the diagnosis on the eye examination records. In our study, dogs with a positive distichiasis diagnosis were regarded as affected (case) regardless of examination age and a later negative examination. Dogs were considered unaffected (controls) when diagnosed as negative for distichiasis after one year of age. This is consistent with the findings from a previous study where we found that a negative distichiasis status in puppies did not give a reliable picture of the distichiasis status in the adult dog [6].

Samples were collected from a biobank established in collaboration between the Norwegian University of Life Sciences (NMBU) and the NKK. DNA from EDTA blood was extracted using E.Z.N.A. Blood DNA Mini Kit from Omega, following the manufacturer's description. The DNA quality was measured with Epoch from BioTek. Seven hundred and thirty-four samples were genotyped on the Illumina 220 K CanineHD Bead chip (Neogen Genomics, USA), and 118 samples from a previous study were genotyped on the Illumina 170 K CanineHD bead chip. Only the 170 K markers shared between the two datasets were kept in the joint analysis.

Quality control

Quality control was performed in Plink 1.9 [42, 43] and in R, using base R and the R package Tidyverse [44, 45]. At the individual level, we removed samples with a genotyping rate below 95% and a heterozygosity rate above three standard deviations from the mean. We controlled for sex mismatch to identify potential sample mix-ups and removed duplicates. At the marker level, we eliminated markers with a call rate below 98%, a minor allele frequency below 0.05, and deviation from Hardy–Weinberg equilibrium at a level of -1.0×10^{-6} in controls and -1.0×10^{-10} in cases, using the Fisher exact test incorporated in Plink. Dogs with a missing phenotype were removed. To assess potential batch effects of the two arrays 170 K and 220 K, a PCA plot was constructed. In addition, the SNP markers were regressed on the two batch (170 K and 220 K) to assess differences in the allele frequency in the two batches. After quality control, the material consisted of 97,185 markers and 731 dogs, where 407 were controls and 324 cases, 442 female (206 cases) and 289 males (118 cases); Seventy-six dogs (16 cases) were genotyped on the 170 k array, and 655 dogs (308 cases) were genotyped on the 220 k array. Mean age of last eye examination in the controls was 2.7 years and 1.9 years in the cases.

Population structure and LD

There is no data on the current population size of SBT in Norway. According to NKK there is around 1000 new registrations of SBTs every year. Between 2005 and April 2022, NKK had registered 1481 imported dogs from 30 different countries, the majority imported from Sweden (52,65%), followed by the United Kingdom (10.05%). Among 2339 litters registered during the same time period, 1493 (61.52%) litters were of mattings with at least one parent from another country. Population structure attributable to country of origin (according to pedigree number), was assessed using principal components analysis conducted in Plink. Plots were constructed in R using the R package ggplot2 [46]. In addition, population structure due to stratification between cases and controls was assessed.

The dataset contained 72 families with offspring, and both parents genotyped with an equal number of affected and unaffected offspring. The total number of genotyped full siblings was 220, distributed on 96 different litters. A GWAS excluding 79 dogs with an equal affection status as another litter mate was conducted.

Linkage disequilibrium (LD) and haploblock size were estimated using Plink 1.9 [42, 43] and R [44]. Haploblock sizes were estimated using the *–blocks* function. Plink uses the haploblock definition suggested by Gabriel et al. [47]. To identify markers in pairwise LD, and to estimate the number of independent SNP markers, we used the LD pruning function in Plink: *indep-pairwise* with the options; window size: 50, step size: 5 and r^2: 0.2. The mean heterozygosity rate was estimated in Plink using the *–het* function.

Association analysis

The association analysis was performed in BayesR (v01/04/2021) [37]. BayesR fit all markers simultaneously, and there are indications that the Bayesian model has a higher power to detect true associations and SNP effect than traditional linear models. In addition, BayesR gives information about the genetic architecture of the trait. BayesR uses the model:

$$y = 1_n\mu + Xa + e$$

where y=a vector of the phenotypes, μ is the general mean term, X is a matrix of the genotypes, and a a vector of SNP effects, and e is a vector of residual errors [37]. BayesR uses a prior of four predefined classes of SNP effects, with the normal distributions $N(0, 0 * \sigma_g^2)$, $N(0, 0.0001 * \sigma_g^2)$, $N(0, 0.001 * \sigma_g^2)$ and $N(0, 0.01 * \sigma_g^2)$. The variance of the SNP effect (σ_g^2) is defined by the data, using a Gibbs sampler to draw samples. Our analysis was run with 100.000 iterations and 50.000 burn-in steps.

In addition, a traditional mixed linear model-based association analysis (MLMA) was run in Genome-wide Complex Trait Analysis (GCTA) version 1.93.2 [27, 48]. A relationship matrix (GRM) is used to control for population structure and relationships.

$$y = \alpha + \beta X + g + e$$

where y=a vector of the phenotype, α is the general mean term, β the fixed additive genetic effect of the SNP considered in the analysis, X=genotype of the SNP coded as 0, 1 and 2 for homozygous, heterozygous and opposite homozygous, respectively, g *is* the random effect of background genes assumed distributed as $g \sim$ N(0,GRM), and e is the residual error. We have used two significance levels, one using Bonferroni correction according to number of all SNPs in the dataset ($0.05/94697 = 5.28 \times 10^{-07}$). Bonferroni correction based on all SNP markers is often considered over-conservative because SNPs are in LD and not independent. Therefore, we calculated a second significance level in accordance with the number of independent SNPs after pruning the data ($0.05/4032 = 1.24 \times 10^{-05}$). Manhattan plots were made in R with the qqman package [49].

The genomic position on the arrays was given in CanFam3.1 [50]. To convert the genomic positions from CanFam3.1 to GSD_1.0 /canFam4 reference genome [51], we used the Liftover tool developed by the University of California Santa Cruz Genomes [52]. All genomic positions refer to GSD_1.0 /canFam4.

Candidate regions

A t-test in R was used to assess if there was a significant difference in the mean risk allele load between cases and controls [44]. Haploblocks within the four candidate regions were analyzed and visualized using Haploview [53].

Prediction

Two approaches were used to predict the dogs' genetic value (GV) based on whole genome SNP markers. Genomic best linear unbiased prediction (GBLUP) calculated in GCTA [27] and a Bayesian hierarchical model, BayesR [37, 54]. GBLUP in GCTA assume a normal distribution of the SNP effect and is based on a mixed linear model:

$$y = \alpha + g + e$$

where y is a vector of the phenotypes, α the mean term, g is the genetic value and ε the residual error. The values of g and e were estimated from the formulas: $\hat{g} = V_g A V^{-1} y$ and $\hat{e} = V_e V^{-1} y$, *wh*ere A is the GRM, *Vg* the genetic variance and *Ve* residual variance and $V = A^*V_g + I^*V_e$ is the variance matrix of the records y [27, 48]. The prediction in BayesR is based on the same mixed models as described in the association analysis. The same mixture of four normal distributions of SNP effects was applied for the prediction. For the prediction, 20.000 burn-in steps and 50.000 iterations were used. Full siblings were removed from the dataset prior to the prediction. The dataset consisted of 94,697 markers across the 38 autosomal chromosomes and 607 dogs, including 248 cases and 359 controls. A sixfold cross-validation was used. To assess the two methods' ability to discriminate between cases and controls, we used the R package pROC [55] to compute the AUC from the ROC curves, with sensitivity on the y-axis and specificity on the x-axis. Different models were tested in GBLUP, GCTA; the base model with no covariables, and models including the covariables: sex, examination age, batch effect, ten first PCA and country of origin. None of the tested covariables improved the model. Therefore, in the final estimates, the base model with no covariates was used. Delong's test in pROC was used to detect a significant difference between the two approaches, BayesR and GBLUP in GCTA.

We compared the odds of developing distichiasis in the 25% of dogs with GV predicted to be most likely to develop the disease, with the odds of developing distichiasis of the 25% of dogs with GV least likely to acquire distichiasis.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40575-023-00132-1.

Additional file 1.

Acknowledgements
We would like to express our gratitude to Ole Albert Guttersrud and Ellen Frøysadal Arnet for their excellent technical support.

Authors' contributions
DJ carried out the association analysis and the genomic predictions and drafted the first manuscript. FL was a major contributor to writing the manuscript. All authors contributed to the writing process, discussion of the methods, results, and revision. The final manuscript was read and approved by all writers.

Funding
The project was funded by the Norwegian University of Life Sciences (Veterinary faculty) and received additional funding from Agria Pet Insurance and Svenska Kennelklubben (SKK) (project no N2021-0020). The funders were not involved in the project or in the decision to publish.

Availability of data and materials
The genomic data are available from the corresponding author upon reasonable request. All phenotype data are available from the open database "Dogweb", www.dogweb.no.

Declarations

Ethics approval and consent to participate
The data has been gathered in agreement with all applicable ethical guidelines. The biological material was based on samples from the biobank at The Norwegian University of Life Sciences. By submission of the blood samples, the owners sign a written consent allowing the DNA to be used in research. The phenotype and pedigree data were based on data from "Dogweb", a public database provided by NKK.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Medical Genetics Unit, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, P.O. box 5003, 1432 Ås, Norway. [2]Evidensia Oslo Dyresykehus, Ensjøveien 14, 0655 Oslo, Norway. [3]Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, P.O. box 5003, 1432 Ås, Norway.

Received: 4 May 2023 Accepted: 16 July 2023
Published online: 24 July 2023

References

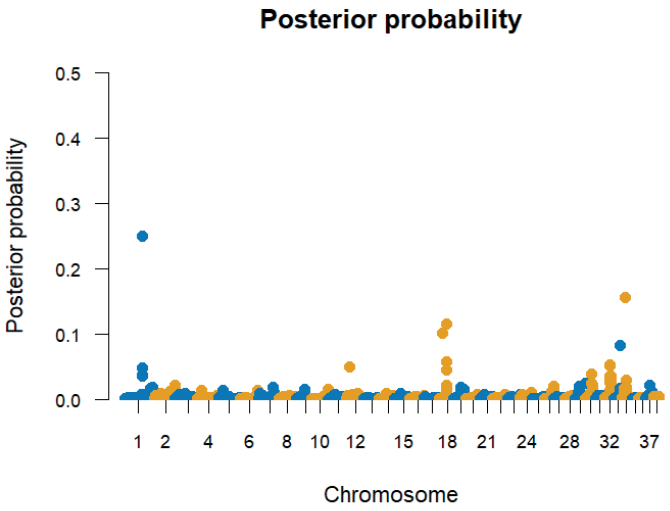
1. Raymond-Letron I, Bourges-Abella N, Rousseau T, Douet JY, de Geyer G, Regnier A. Histopathologic features of canine distichiasis. Vet Ophthalmol. 2012;15(2):92–7.
2. Palella Gómez A, Mazzucchelli S, Scurrell E, Smith K, Pinheiro de Lacerda R. Evaluation of partial tarsal plate excision using a transconjunctival approach for the treatment of distichiasis in dogs. Vet Ophthalmol. 2020;23(3):506–14.
3. Gelatt KN, Gilger BC, Kern TJ. Veterinary ophthalmology : Volume 2. 5th ed. ed. Ames, Iowa: Blackwell; 2013.832–893
4. ACVO Genetics Committee. The blue book, oculare diorders presumed to be inherited in purebred dogs. 2021. https://ofa.org/diseases/eye-disease/blue-book. Accsessed 19 Feb 2023.
5. Norwegian Kennel Club. 2022. www.dogweb.no. Accessed 12 Sept 2022.
6. Joergensen D, Madsen P, Ropstad EO, Lingaas F. Heritability estimates of distichiasis in staffordshire bull terriers using pedigrees and genome-wide snp data. Acta Vet Scand. 2022;64(1):30.
7. Bellamy KKL, Lingaas F, Madsen P. Heritability of distichiasis in havanese dogs in norway. Canine Med Genet. 2021;8(1):11.
8. Kaufhold J, Hamann H, Steinbach G, Gordon S, Brahm R, Grussendorf H, et al. Analyse der Prävalenz von Distichiasis bei der Hunderasse Elo. Prakt Tierarzt. 2006;119.
9. Petersen T, Proschowsky HF, Hardon T, Rasch SN, Fredholm M. Prevalence and heritability of distichiasis in the english cocker spaniel. Canine Genet Epidemiol. 2015;2:11.
10. Kaufhold J. Populationsgenetische untersuchungen zur entwicklung der hunderasse elo sowie zum auftreten und zur vererbung von distichiasis. Hannover: Tierärztliche Hochschule; 2004.
11. Reinstein SL, Gross SL, Komáromy AM. Successful treatment of distichiasis in a cat using transconjunctival electrocautery. Vet Ophthalmol. 2011;14(Suppl 1):130–4.
12. Verboven CA, Djajadiningrat-Laanen SC, Kitslaar WJ, Grinwis GC, Schoemaker NJ, Boevé MH. Distichiasis in a ferret (mustela putorius furo). Vet Ophthalmol. 2014;17(4):290–3.
13. Arteaga K, Crasta M. Successful treatment of distichiasis in a cow using a direct-contact 810-nm diode laser. Vet Surg. 2021;50(5):1164–8.
14. Allais-Bonnet A, Grohs C, Medugorac I, Krebs S, Djari A, Graf A, et al. Novel insights into the bovine polled phenotype and horn ontogenesis in bovidae. PLoS ONE. 2013;8(5): e63512.
15. Hisey EA, Hermans H, Lounsberry ZT, Avila F, Grahn RA, Knickelbein KE, et al. Whole genome sequencing identified a 16 kilobase deletion on eca13 associated with distichiasis in friesian horses. BMC Genom. 2020;21(1):848.
16. Brooks BP, Dagenais SL, Nelson CC, Glynn MW, Caulder MS, Downs CA, et al. Mutation of the foxc2 gene in familial distichiasis. J Pediatr Ophthalmol Strabismus. 2003;7(5):354–7.
17. Patil BB, Bell R, Brice G, Jeffery S, Desai SP. Distichiasis without lymphoedema? EYE (Lond). 2004;18(12):1270–2.
18. Tavian D, Missaglia S, Maltese PE, Michelini S, Fiorentino A, Ricci M, et al. Foxc2 disease-mutations identified in lymphedema-distichiasis patients cause both loss and gain of protein function. Oncotarget. 2016;7(34):54228–39.
19. Cervantes-Barragán DE, Villarroel CE, Medrano-Hernández A, Durán-McKinster C, Bosch-Canto V, del-Castillo V, et al. Setleis syndrome in mexican-nahua sibs due to a homozygous twist2 frameshift mutation and partial expression in heterozygotes: Review of the focal facial dermal dysplasias and subtype reclassification. J. Med. Genet. 2011;48(10):716–20.
20. Kievit A, Tessadori F, Douben H, Jordens I, Maurice M, Hoogeboom J, et al. Variants in members of the cadherin-catenin complex, cdh1 and ctnnd1, cause blepharocheilodontic syndrome. Eur J Hum Genet. 2018;26(2):210–9.
21. García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP. Changes in genetic selection differentials and generation intervals in us holstein dairy cattle as a result of genomic selection. Proc Natl Acad Sci USA. 2016;113(28):E3995-4004.
22. Guo G, Zhou Z, Wang Y, Zhao K, Zhu L, Lust G, et al. Canine hip dysplasia is predictable by genotyping. Osteoarthr Cartil. 2011;19(4):420–9.
23. Sánchez-Molano E, Pong-Wong R, Clements DN, Blott SC, Wiener P, Woolliams JA. Genomic prediction of traits related to canine hip dysplasia. Front Genet. 2015;6:97.
24. Baker LA, Momen M, Chan K, Bollig N, Lopes FB, Rosa GJM, et al. Bayesian and machine learning models for genomic prediction of anterior cruciate ligament rupture in the canine model. G3 (Bethesda). 2020;10(8):2619–28.
25. Lingaas F, Tengvall K, Jansen JH, Pelander L, Hurst MH, Meuwissen T, et al. Bayesian mixed model analysis uncovered 21 risk loci for chronic kidney disease in boxer dogs. PLoS Genet. 2023;19(1): e1010599.
26. Thorsrud JA, Evans KM, Holle DM, Huson HJ. Comparison of gblup and machine learning breeding values for binary disease traits in working dogs. Proceedings of 12th world congress on genetics applied to livestock production (wcgalp). Wageningen Academic Publishers; 2022. p. 1643–6.
27. Yang J, Lee SH, Goddard ME, Visscher PM. Gcta: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76–82.
28. Martinelli DC, Fan C-M. The role of gas1 in embryonic development and its implications for human disease. Cell Cycle. 2007;6(21):2650–5.
29. Novak D, Hüser L, Elton JJ, Umansky V, Altevogt P, Utikal J. Sox2 in development and cancer biology. Semin Cancer Biol. 2020;67:74–82.

30. Morris EJ, Michaud WA, Ji J-Y, Moon N-S, Rocco JW, Dyson NJ. Functional identification of api5 as a suppressor of e2f-dependent apoptosis in vivo. PLoS Genet. 2006;2(11):e196.
31. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005;438(7069):803–19.
32. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, et al. Extensive and breed-specific linkage disequilibrium in canis familiaris. Genome Res. 2004;14(12):2388–96.
33. Karlsson EK, Sigurdsson S, Ivansson E, Thomas R, Elvers I, Wright J, et al. Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near cdkn2a/b. Genome Biol. 2013;14(12):R132.
34. Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, Baxter TL, et al. Complex disease and phenotype mapping in the domestic dog. Nat Commun. 2016;7:10460.
35. Lawson DD. Canine distichiasis. J Small Anim Pract. 1973;14(8):469–78.
36. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157(4):1819–29.
37. Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM. Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. PLoS Genet. 2015;11(4): e1005069.
38. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from snps. Nat Rev Genet. 2013;14(7):507–15.
39. Wray NR, Kemper KE, Hayes BJ, Goddard ME, Visscher PM. Complex trait prediction from genome data: Contrasting ebv in livestock to prs in humans: Genomic prediction. Genetics. 2019;211(4):1131–41.
40. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013;9(3): e1003348.
41. Edwards SM, Woolliams JA, Hickey JM, Blott SC, Clements DN, Sánchez-Molano E, et al. Joint genomic prediction of canine hip dysplasia in uk and us labrador retrievers. Front Genet. 2018;9:101.
42. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation plink: Rising to the challenge of larger and richer datasets. GigaScience. 2015;4(1):7.
43. Purcell S. CC. Plink web page. 2023; www.cog-genomics.org/plink/1.9/. Accessed 30.01.2023.
44. R Core Team. R: A language and environment for statistical computing. 2022; https://www.R-project.org/.
45. Wickham, et al. Welcome to the tidyverse. J Open Source Softw. 2019;4:1686.
46. Wickham H. Ggplot2: Elegant graphics for data analysis: Springer-Verlag New York; 2016
47. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science. 2002;296(5576):2225–9.
48. Yang J. Gcta a tool for genome-wide complex trait analysis. 2023; https://yanglab.westlake.edu.cn/software/gcta. Accessed 19.04.2023
49. Turner SD. Qqman: An r package for visualizing gwas results using q-q and manhattan plots. bioRxiv. 2014:005165.
50. Hoeppner MP, Lundquist A, Pirun M, Meadows JR, Zamani N, Johnson J, et al. An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. PLoS ONE. 2014;9(3): e91172.
51. Wang C, Wallerman O, Arendt ML, Sundström E, Karlsson Å, Nordin J, et al. A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. Commun Biol. 2021;4(1):185.
52. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The ucsc genome browser database: Update 2006. Nucleic Acids Res. 2006;34(Database issue):D590-8.
53. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of ld and haplotype maps. J Bioinform. 2005;21(2):263–5.
54. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci. 2012;95(7):4114–29.
55. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. Proc: An open-source package for r and s+ to analyze and compare roc curves. BMC Bioinformatics. 2011;12(1):77.

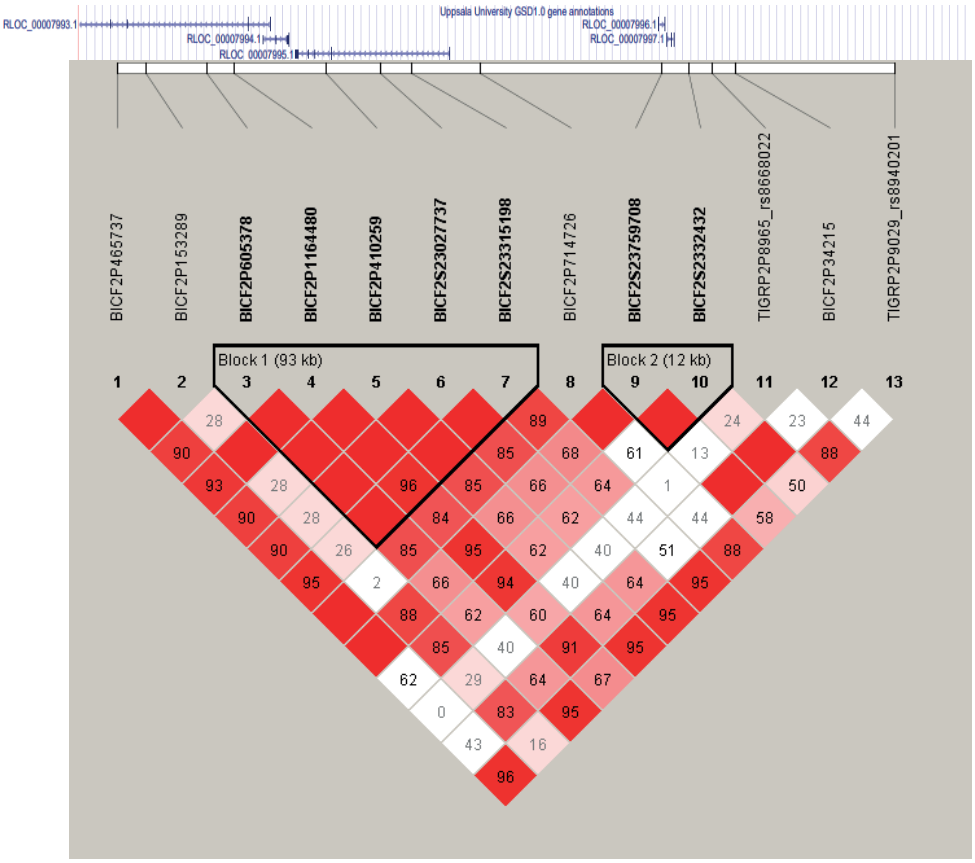
Publisher's Note

Supplementary material Paper II

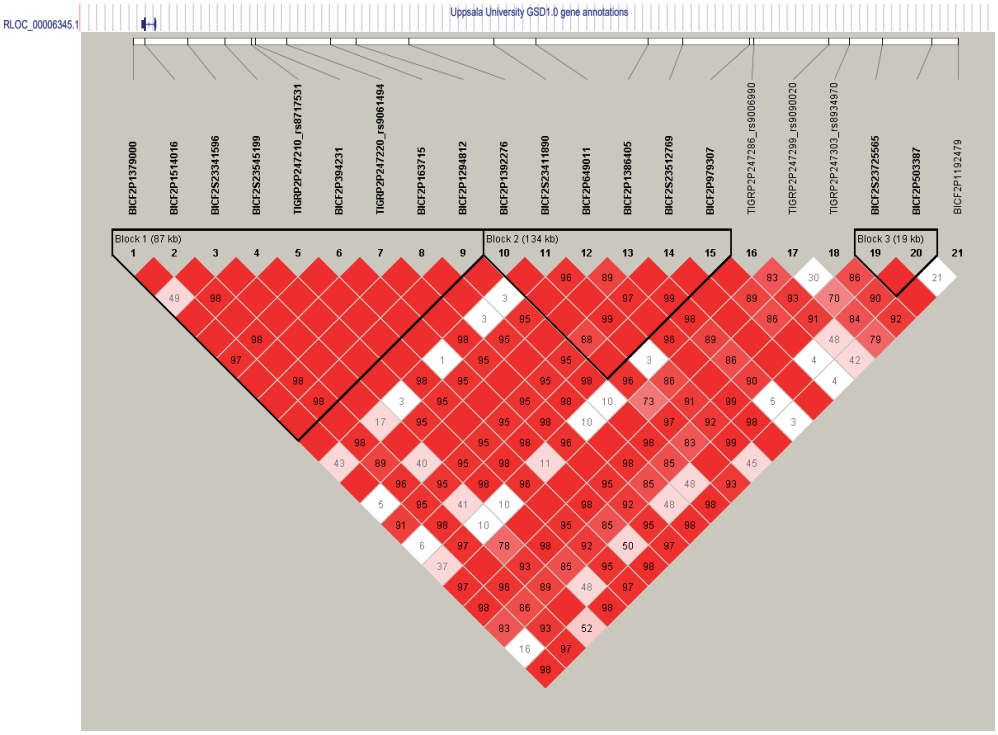
Additional material



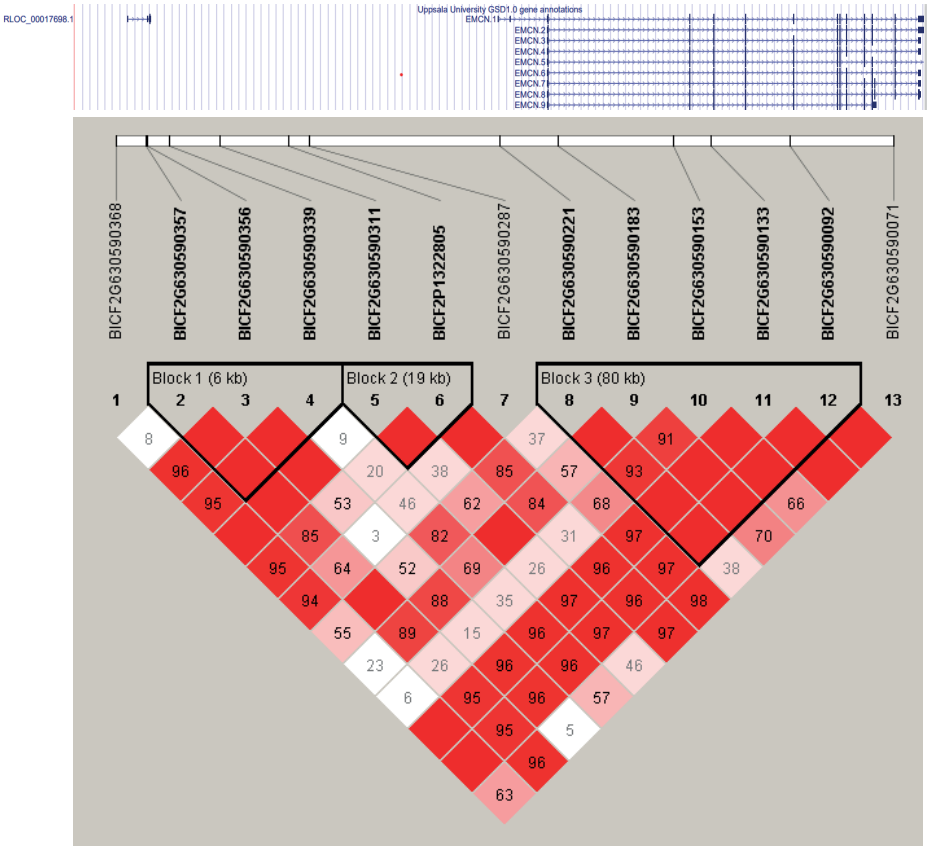
Supplementary Figure 1. *The posterior inclusion probabilities of the SNP in the fourth mixture class (N(0,0.01*σ_(g))^2) from the analysis conducted in BayesR.*



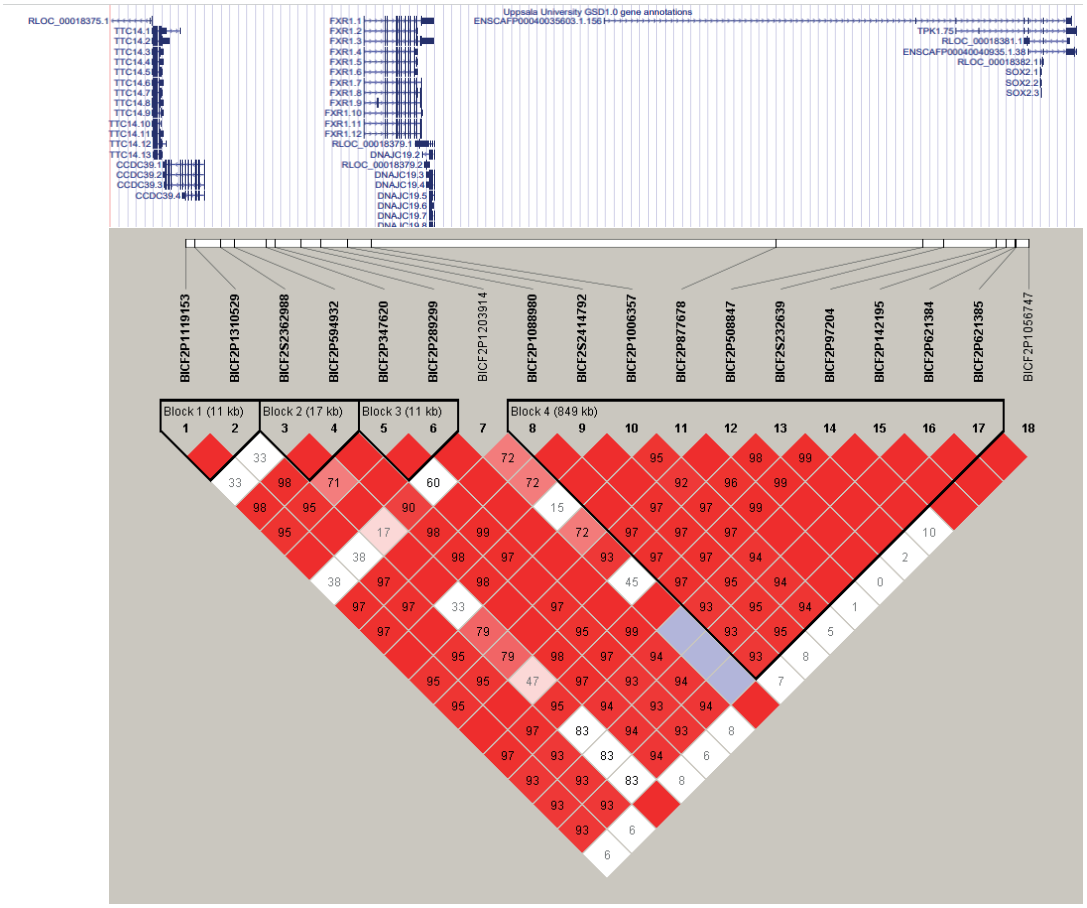
Supplementary Figure 2. *Haploblock from the candidate region on chr 1. shows the top SNP, BICF2P714726 situated between two haploblocks. The numbers in the squares show the level of LD, based on the D prime (D'). The colour shadings are based on the Lod score and D' estimated between two SNPs. A D' < 1 and LOD < 2 gives a white colour, D' < 1 and LOD ≥ 2 gives shades of pink. D' = 1 and LOD < 2 give blue colour. D' = 1 and LOD ≥ 2 give dark bright red colour. (https://www.broadinstitute.org/haploview/ld-display). On the top of the figure is a picture of the corresponding genomic region from UCSCs genome browser (https://genome.ucsc.edu), with the position from canFam4 reference genome.*



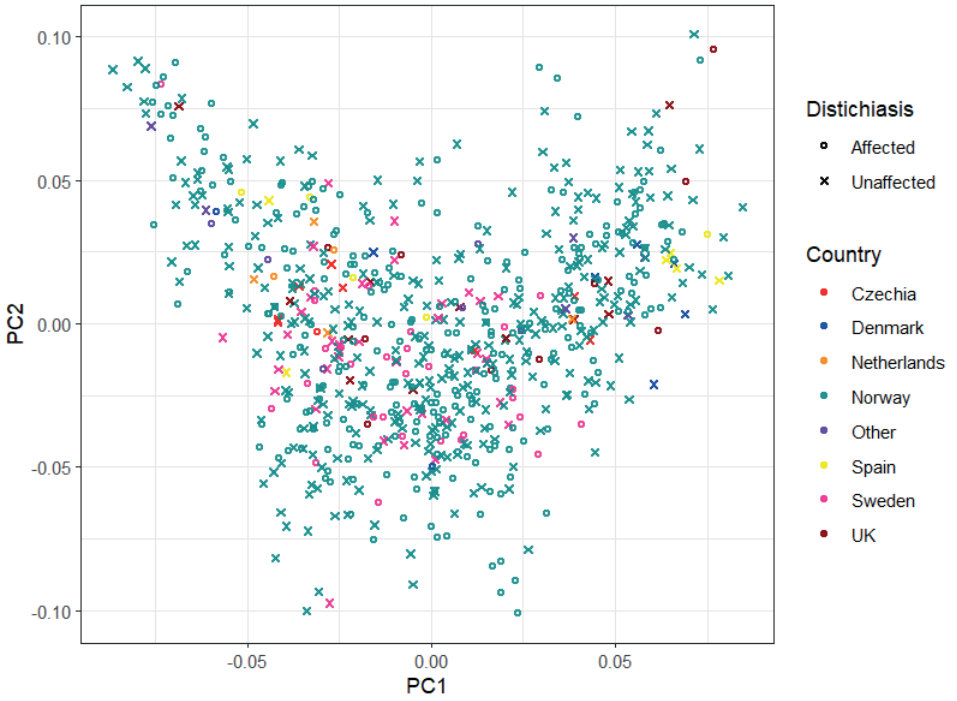
Supplementary Figure 3. *Haploblocks in the candidate region on chr 18, the top SNP BICF2P1386405 is situated in the second haploblock with five other SNPs. The numbers in the squares show the level of LD, based on the D prime (D'). The colour shadings are based on the Lod score and D' estimated between two SNPs. A D' < 1 and LOD < 2 gives a white colour, D' < 1 and LOD ≥ 2 gives shades of pink. D' = 1 and LOD < 2 give blue colour. D' = 1 and LOD ≥ 2 give dark bright red colour. (https://www.broadinstitute.org/haploview/ld-display). On the top of the figure is a picture of the corresponding genomic region from UCSCs genome browser (https://genome.ucsc.edu), with the position from canFam4 reference genome.*



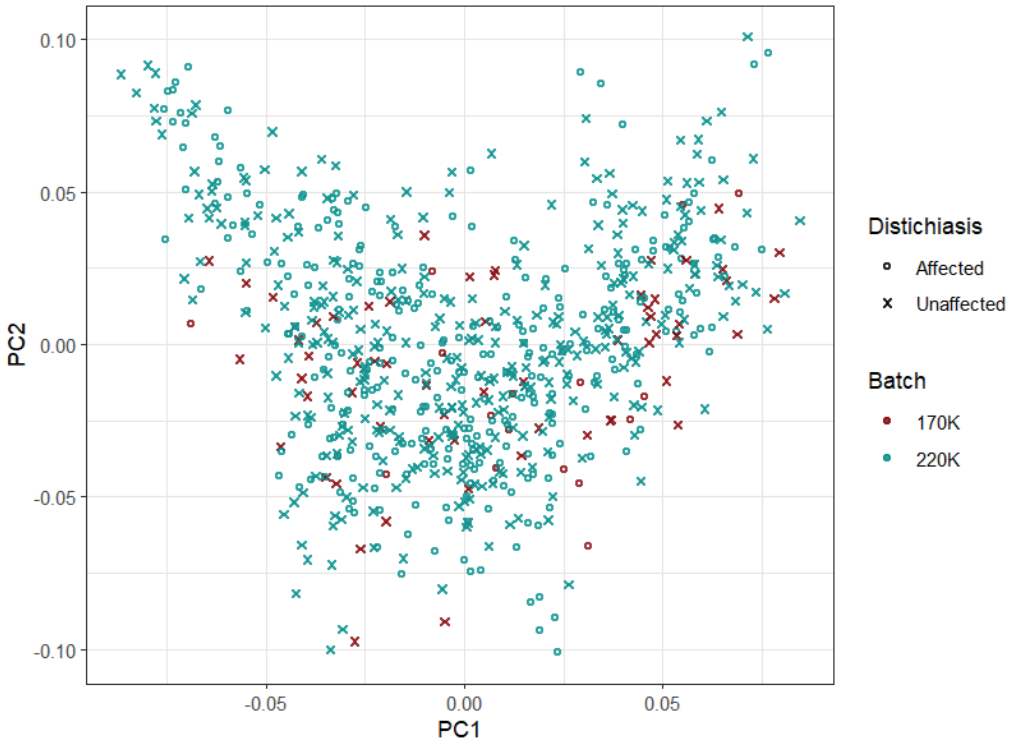
Supplementary Figure 4. *Haploblocks chr 32, the top SNP BICF2G630590287 lays between haploblock 2 and 3. The numbers in the squares show the level of LD, based on the D prime (D'). The color shadings are based on the Lod score and D' estimated between two SNPs. A D' < 1 and LOD < 2 gives a white colour, D' < 1 and LOD ≥ 2 gives shades of pink. D' = 1 and LOD < 2 give blue colour. D' = 1 and LOD ≥ 2 give dark bright red colour. (https://www.broadinstitute.org/haploview/ld-display). On the top of the figure is a picture of the corresponding genomic region from UCSCs genome browser (https://genome.ucsc.edu), with the position from canFam4 reference genome.*



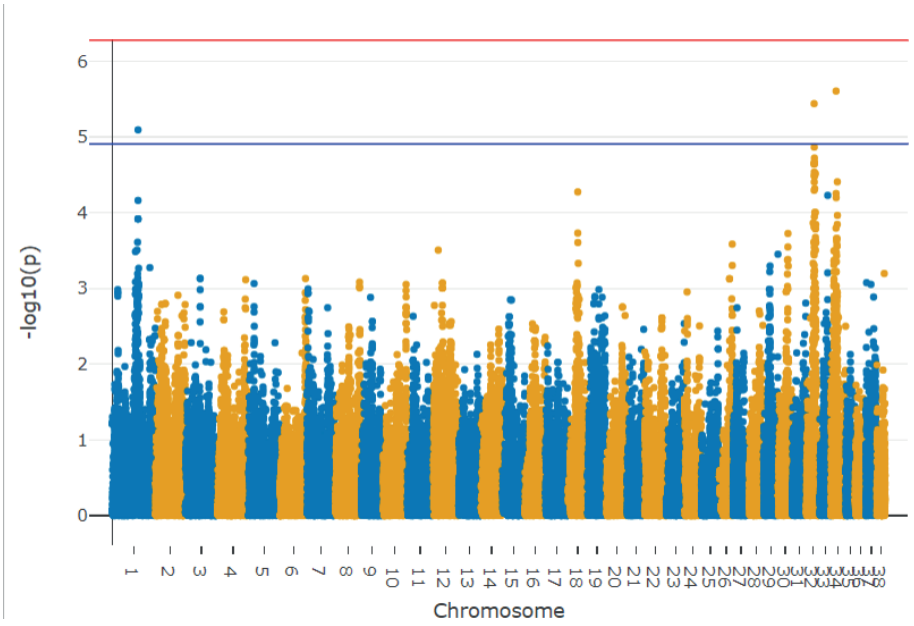
Supplementary Figure 5. *Haploblocks chr 34, the top SNP BICF2S232639 is situated in the middle of haploblock 4 with nine other SNPs. The numbers in the squares show the level of LD, based on the D prime (D'). The colour shadings are based on the Lod score and D' estimated between two SNPs. A D' < 1 and LOD < 2 gives a white colour, D' < 1 and LOD ≥ 2 gives shades of pink. D' = 1 and LOD < 2 give blue colour. D' = 1 and LOD ≥ 2 give dark bright red colour.*
(https://www.broadinstitute.org/haploview/ld-display). On the top of the figure is a picture of the corresponding genomic region from UCSCs genome browser (https://genome.ucsc.edu), with the positions from canFam4 reference genome.



Supplementary Figure 6. *PCA plot, including the country of origin. Other countries are countries with less than five genotyped dogs. This includes dogs from Finland, Germany, Greece, Poland, Malta, Austria, and Hungary.*



***Supplementary Figure 7**. A PCA plot showing the distribution of affected and unaffected dogs in the two arrays, 170K and 220K.*

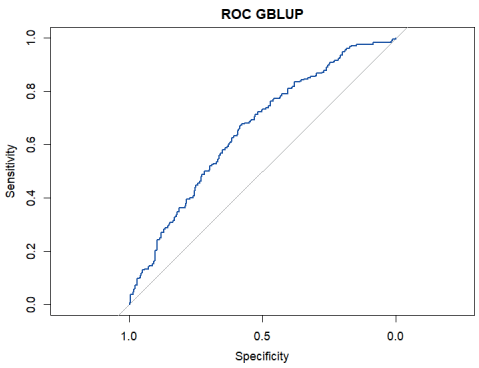


Supplementary Figure 8. *A Manhattan plot where 79 siblings with an equal affection status as full siblings are removed. Includes 652 dogs, 377 are controls, and 275 are cases. The significance level (represented by the blue line) is set to 1.24 x 10^{-05} using a Bonferroni correction to account for multiple testing using the number of independent markers (4030), a second significance level of 5.28 x 10^{-07} (represented by the red line) is the Bonferroni correction based on all markers 94697.*

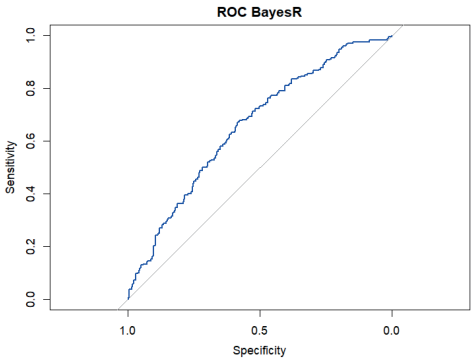
Supplementary Table 1. Minor allele frequency (MAF) of the four top SNPs on the two arrays.

CHR	SNP	MAF 170K Array	MAF 220K Array
1	BICF2P714726	0.12	0.11
18	BICF2P1386405	0.20	0.27
32	BICF2G630590287	0.24	0.26
34	BICF2S232639	0.43	0.45

The 170K array includes 76 dogs (16 cases and 60 controls), and the 220K array includes 655 dogs (308 cases and 347 controls).



***Supplementary Figure 9**. A ROC curve calculated from the estimated GV using GBLUP.*



Supplementary Figure 10. A ROC curve calculated from the estimated GV using BayesR.

Supplementary Table 2. 6-fold cross validation in GBLUP including covariables

Covariable	AUC	Confidence interval
Sex	0.66	0.61-0.70
Batch	0.65	0.61-0.69
Country of origin	0.66	0.61-0.70
Age	0.65	0.60-0.69
PCA	0.64	0.60-0.69

The area under the curve (AUC) calculated from the receiver operating characteristic curve (ROC curve) from all estimated genetic values in the GBLUP. Batch indicates the 170K and 220K array, country of origin according to the registration number, and age is the examination age. PCA includes the first teen first principal components analysis (PCA).

Paper III

A genome-wide study identifies a region on CFA37 associated with cataract in the Norwegian buhund

Dina Jørgensen[1], Elin Kristiansen[2], Stein Dahl[2], Ernst-Otto Ropstad[3] and Frode Lingaas[1]

1) *Medical Genetics Unit,* Faculty of Veterinary Medicine
Norwegian University of Life Sciences
P.O. box 5003
1432 Ås
Norway

2) *AniCura Hønefoss Dyrehospital*
Osloveien 67
3511 Hønefoss
Norway

3) Evidensia, Oslo Dyresykehus
Ensjøveien 14
0655 Oslo
Norway

Corresponding author: Dina Jørgensen, e-mail: dijo@nmbu.no

Abstract

Pulverulent nuclear cataract (PNC) is a common ocular disorder observed in about 50% of the Norwegian buhund population. We have used a genome-wide association study to investigate the association between PNC and genomic markers in Norwegian buhunds. The material consists of 160 genotyped buhunds with eye examination data, including 121 affected dogs and 39 unaffected dogs (negative examination after the age of seven). We identified an associated genomic region on chromosome 37. The genomic region spans approximately six megabases (CAF37:11422235-17384184) with more than 50 annotated genes, including five candidate genes (CRYGA, CRYGB, CRYGC, CRYGD, and PIKFYVE) that have previously been associated with PNC in humans. We conducted Sanger sequencing on four out of the five candidate genes. However, no causal variant was identified. By selecting five of the most significant SNPs in the region and comparing the risk allele load in cases and control, dogs with more than 5 risk alleles was significantly more likekly to get a PNC diagnosis (OR=8,25. 955 CL 3.73-19.47)

Our findings provide a valuable foundation for further research on the genetic background of PNC in Norwegian buhunds. The SNP markers within the identified genomic region hold promise for future marker-assisted selection strategies to reduce the incidence of PNC within this breed. Additional sequencing of the associated region may point out a causative variant.

Keywords

Buhund cataract, canine, GWAS, Norwegian Buhund, PNC, pulverulent nuclear cataract.

Background

The crystal lens is a transparent, highly organized avascular tissue that refracts light onto the retina. Cataract is defined as any opacity that disrupts the transparency of the lens (Liu et al., 2017). Cataracts are one of the main reasons for blindness in humans in middle and low-income countries (WHO; Flaxman et al., 2017). The pathogenesis of cataract is often complex and involve both environmental and genetic factors (Shiels and Hejtmancik, 2019). In humans hereditary congenital and juvenile cateracts are often Mendelian, and the most common mode of inheritance is autosomal dominant, but autosomal recessive and X-linked inheritance and complex is also seen (Hejtmancik, 2008). At least 116 genes in humans have been shown to be associated with cataracts, including genes encoding membrane proteins, cytoskeletal proteins, crystalline and transcription factors (https://cat-map.wustl.edu/) (Shiels et al., 2010; Berry et al., 2020a). For pulverulent nuclear cataract (PNC), at least ten different genes have been shown to be associated with the disorder in humans (Shiels et al., 2010): *GJA3* (Ding et al., 2011)*, GJA8 (Arora et al., 2008; Yan et al., 2008), MIP (Wang et al., 2010), MAF (Jamieson et al., 2002), LIM2 (Berry et al., 2020b), CRYBB1* (Meyer et al., 2009)*, CRYBA1* (Bateman et al., 2000; Lu et al., 2007)*, CRYGC* and CRYGD (Kumar et al., 2011)*,* and *PIKFYVE* (Mei et al., 2022)*.*

In dogs, cataracts are one of the most common intraocular disorders and the most common cause of blindness (Gelatt et al., 2013). Cataracts in dogs are presumed to be hereditary unless associated with known trauma, intoxication, ocular inflammation, specific metabolic disorders, nutritional deficiencies or age (ACVO, 2020; ECVO, 2021). The mode of inheritance has only been documented in a small number of dog breeds, and autosomal recessive inherence is believed to be the most common (Gelatt et al., 2013).

Two genes have so far been shown to be associated with cataracts in dogs. Mellers et al. found an association between heat-shock transcription factor 4 (*HSF4*) and primary cataracts in Staffordshire bull terriers, early-onset cataracts in Boston terriers, Australian shepherds, and French bulldogs. (Mellersh et al., 2006; Mellersh et al., 2007; Mellersh et al., 2009). Later, a frameshift deletion in the *FYCO1* gene was found to be associated with juvenile cataracts in Wirehaired Pointing Griffon dogs (Rudd Garces et al., 2022). In Australian shepherds, a locus on chromosome 13 was found to be associated with cataracts (Ricketts et al., 2015). Recently, Bellamy and Lingaas (2023) found two loci on CFA20 and CFA21 associated with posterior polar cataracts in the dog breed Havanese.

PNC was first described in Norwegian buhunds by Bjerkås and Haaland (1995). Among 102 examined buhunds, they found 52 dogs affected by PNC. The first signs were observed in puppies at around six weeks of age as small opacities adjacent to the posterior suture line. With increasing age, the opacities progressed to involve the complete foetal nucleus. The changes were, in most cases, bilateral with equally affected lenses. Through studies of the pedigrees, an autosomal dominant inherence was suggested (Bjerkås and Haaland, 1995). In 2009, the "Buhund project" was initiated to look at the current prevalence of different cataracts in the breed and find associated mutations (Kristiansen et al., 2017). The project aimed to develop genetic tests to both control the occurrence of cataracts and maintain good general health status in the dogs. Kristiansen et al. found that the prevalence of PNC had not changed since 1995. They found a high age of onset of PNC in many dogs, where several dogs found unaffected at a young age later developed PNC. Another key finding was a high prevalence of other kinds of cataracts, where cortical cataracts and posterior polar cataracts were the most common types. The prevalence of these other forms of cataracts, some of which have greater impact on vision than PNC, appeared to have increased since the initial investigation in 1995. With the persistently high prevalence of PNC and the fact that not all potential cases are detected before dogs are used in breeding, a genetic test or marker-assisted selection would be helpful to reduce the prevalence of PNC in the Norwegian buhund population.

PNC is not common in other breeds of dogs but has also been observed in a few other dog breeds, such as the flat-coated retrievers and the German shepherds (NKK, 2023). In Leonberger, a non-progressive posterior nuclear cataract similar to PNC in buhund is registered (Heinrich et al., 2006).

This study aimed to identify genomic regions and potential genetic variants associated with PNC in buhund through a genome-wide association study (GWAS).

Materials and methods

Samples were selected from the Biobank of The Veterinary faculty at the Norwegian University of Life Sciences (NMBU), collected as part of clinical examinations and from dog shows in Norway, Sweden, and Denmark and through direct recruitment. The material has been collected with the owner's consent and in agreement with ethical guidelines.

All dogs included in the study were examined by veterinarians certified by the European College of Veterinary Ophthalmologists (ECVO). The records of diagnostic data have been collected and maintained through the Buhund project since 2009 as a joint initiative between a private clinic (Hønefoss Dyrehospital), the Norwegian Buhund club, and the NMBU, and partly through an openly available database, "dogweb", maintained by the Norwegian kennel club (NKK). Included dogs were classified as PNC positive (cases) or negative (controls). Since PNC may have a high age of onset (Kristiansen et al., 2017), only PNC-negative dogs with an eye examination after the age of seven were included as controls. Dogs with posterior polar cataracts or cortical cataracts were accepted in the control group, as Kristiansen et al. found no correlation between PNC and the two other forms of cataracts (Kristiansen et al., 2017).

The material consists of EDTA blood and buccal swabs (Performagene™, DNA Genotek Inc). The DNA from EDTA blood was isolated using the "E.Z.N.A.® Blood DNA Mini Kit" (Omega), while DNA from buccal swabs was isolated as recommended by the manufacturer. The DNA was stored at -20 until prepared for further analysis. The DNA quality was assessed with Epoch from BioTek. Two hundred and three samples were genotyped using the Illumina 170k CanineHD Bead chip, and 83 samples using the Illumina 220k CanineHD Bead chip.

Quality control

Plink 1.9 and R were used for data management and quality control (Chang et al., 2015; R Core Team, 2021; Purcell S., 2023). Only common markers between the two datasets (170K and 220K) were kept for analysis. Samples with a genotyping rate below 97% were eliminated. Individuals with a heterozygosity rate above three standard deviations from the mean were removed. Potential sex mismatch was controlled to identify potential sample mix-ups. Dogs with an uncertain PNC status and duplicates were also taken out. At the marker level, we eliminated markers with a call rate below 96% (n=7579 markers), markers with a minor allele frequency of 0.05 (n=64288), and markers that deviated from Hardy-Weinberg equilibrium at a level of -1e-6 were removed. The final dataset consisted of 100662 markers and 160 dogs, where 39 were controls and 121 dogs with confirmed PNC. There were 77 males (61 cases) and 83 females (60 cases). Plots including the two first principal components analysis (PCA) were generated to visualize the distribution among cases and controls in each dataset, to identify potential clusters of dogs from Norway, Sweden, and Denmark, and one PCA plot to look at any potential batch effect between the 170k and 220k datasets. PCA, batch and country were also included as covariates in the initial association analysis.

Genome-wide association study

The association analysis was performed using Genome-wide Complex Trait Analysis software (GCTA) with a mixed linear model-based association (MLMA) analysis, including a

genetic relationship matrix (GRM) as a random effect to control for close relationships and population structure (Yang et al., 2011). We applied the leaving-one-chromosome-out option (LOCO) to avoid overcorrection of the candidate SNPs. Bonferroni correction was used to correct for multiple testing, dividing 0.05 by the number of tested SNPs (98179), resulting in a significance level of 5.09×10^{-07}. Plots were generated in R using the manhattanly packages (Bhatnagar, 2021). The LocusZoom-like plots were based on the script by Major et al. (Major, 2021) performed in R. The odds ratio and MAF of top associated SNP markers were calculated in Plink control (Chang et al., 2015; Purcell S., 2023). The total risk allele load was calculated and plotted in R using base R (R Core Team, 2021) and tidyverse (Wickham et al., 2019; R Core Team, 2021).

Sequencing of candidate genes in the associated region

Exons of the four gamma crystalline candidate genes, *CRYGA*, *CRYGB*, *CRYGC*, and *CRYGD*, were sequenced using Sanger sequencing. Primers (Table 1 supplementary material) were constructed in Primer3Plus (Untergasser et al., 2007) based on the gene annotation of NCBI RefSeq (O'Leary et al., 2016) and Can.Fam4 (Wang et al., 2021) using the genome browser of the University of California Santa Cruz (UCSC), http://genome.ucsc.edu (Kent et al., 2002; USCS, 2023). The PCR products were sequenced with Sanger sequencing and read with the 3500xL Genetic Analyzer from Applied Biosystems, Life Technologies, Thermo Fisher Scientific. The sequencing results were analyzed using Sequencher™ 5.1., Gene Codes Corporations. For the initial assessment of the candidate genes, four unaffected dogs and four affected dogs were sequenced. In sequences with observed variations, the number of sequenced dogs was increased to 29 (14 affected cases and 15 unaffected dogs). The odds ratio was calculated in R using base R (R Core Team, 2021) and Tidyverse (Wickham et al., 2019).

Results

Association study

The association analysis revealed a significant region on chromosome 37 associated with PNC (Figures 1 and 2). The associated region contains eighteen SNPs that reached a significance level of 5.09×10^{-07}, and the significant region spans approximately 6 Megabases (Mb), Table 1. The region includes more than 50 characterized genes, including five genes; *CRYGA*, *CRYGB*, *CRYGC*, *CRYGD*, and *PIKFYVE,* that have previously been associated with PNC and other forms of cataracts in humans.

Five SNPs (BICF2S230689, BICF2P165614, BICF2P194369, BICF2P375607, TIGRP2P419549) were selected to assess the distribution and overall risk load among cases and controls. The total risk load of the selected SNPs is displayed in Figure 3. On average, counting the total number of risk alleles in these five selected loci showed that unaffected dogs carry 2.08 risk alleles and affected dogs 5.68 risk alleles. Carrying five or more risk alleles gave an increased risk of developing PNC, with an odds ratio of 8.25 (95% CI 3.73 - 19.47).

195 Table 1. The 18 significant SNPs identifieid in the MLMA

SNP	Position canFam4	Minor allele/ Major allele	MAF affected	MAF unaffected	Effect size (SE)	P	OR	Gene
BICF2S230689	Chr37:11422235	A/G	0.380	0.731	-0.280 (0.050)	2.62897e-08	0.226	Non-coding
BICF2P165614	Chr37:13780001	T/C	0.554	0.180	0.260 (0.050)	3.88697e-08	5.672	*PARD3B* intron
TIGRP2P419024	Chr37:13954500	A/C	0.554	0.180	0.260 (0.047)	3.88697e-08	5.672	*PARD3B* intron
BICF2P194369	Chr37:12443448	C/T	0.554	0.192	0.266 (0.049)	5.27288e-08	5.211	Non-coding
BICF2P375607	Chr37:14467850	A/G	0.566	0.205	0.258 (0.048)	6.6893e-08	5.056	*INO80D.1* intron
TIGRP2P419549	Chr37:15685620	T/C	0.546	0.192	0.251 (0.047)	1.05566e-07	5.04	Non-coding
BICF2S23522154	Chr37:11675873	A/G	0.380	0.705	-0.268 (0.051)	1.31716e-07	0.257	*CARF* intron
BICF2P1176294	Chr37:11714638	A/C	0.380	0.705	-0.268 (0.051)	1.31716e-07	0.257	*CARF* intron
TIGRP2P418550	Chr37:11717376	C/T	0.380	0.705	-0.268 (0.051)	1.31716e-07	0.257	*CARF* intron
BICF2P728550	Chr37:11668061	A/G	0.386	0.705	-0.268 (0.051)	1.63881e-07	0.262	*CARF* intron
TIGRP2P419482	Chr37:15510457	A/G	0.554	0.218	0.251 (0.048)	1.75735e-07	4.452	Uncharacterized: LOC102156884 intron
BICF2P1198058	Chr37:15515730	T/C	0.554	0.218	0.251 (0.048)	1.75735e-07	4.452	Uncharacterized: LOC102156884 intron
BICF2S23158338	Chr37:15881973	A/G	0.554	0.218	0.251 (0.048)	1.75735e-07	4.452	Uncharacterized: LOC102157126 intron
BICF2P811332	Chr37:17260018	G/A	0.517	0.167	0.240 (0.046)	1.93071e-07	5.342	*MAP2* intron intron
BICF2S23211477	Chr37:17370039	T/C	0.512	0.167	0.241 (0.046)	2.17082e-07	5.254	*MAP2* intron
BICF2S2294485	Chr37:17384184	C/T	0.512	0.167	0.241 (0.046)	2.17082e-07	5.254	*MAP2* intron
BICF2S2337543	Chr37:15530660	A/G	0.417	0.756	-0.246 (0.048)	3.18347e-07	0.231	Uncharacterized: LOC102156884 - intron / LOC111094350 - exon
BICF2P1254082	Chr37:17024702	C/T	0.554	0.205	0.236 (0.047)	4.25577e-07	4.808	Non-coding

196 *Table 1. The 18 significant SNPs. The Effect size* with *standard error (SE)* and P- values *is estimated using MLMA in GCTA.*
197 Minor allele frequency (MAF) for affected and unaffected, and odds ratio (OR) was calculated in Plink. *Genes are identified*
198 *using the UCSC Genome Browser with the UU_Cfam_GSD_1.0/canFam4 reference sequence. For the gene annotation, we*
199 *used NCBI Canis lupus familiaris Annotation Release 106 (2021-01-11).*

200

Sequencing of candidate genes

Sequencing of exons in the four crystalline genes did not reveal any causal variants. However, we identified one SNP in exon two of CRYGD, located at position CAF37:16289129 (CanFam4.0). The SNP is a previously known missense variant (**A**TG->**C**TG / M->L). An analysis of the protein sequence in PolyPhen-2 v2.2.3r406 (Adzhubei et al., 2010) predicted that the effect of this mutation was benign.

Our analysis of Sanger sequences from 29 dogs (15 controls and 14 cases) showed, however, a higher frequency of the A-allele in the controls compared to cases where most dogs had the C-allele. It appears that the minor allele A is associated with a lower risk for PNC when compared to the C allele, and even if the distance to the top SNP is several Megabases, there is an increased risk of developing PNC associated with the C allele (Odds ratio of 6.17 (95% CI 1.40- 38.88)).

Population structure

We did not identify any population structure among the dogs in the different batches (170K and 220K). The cases and controls were evenly distributed on the PCA plots (see Supplementary Figures 1 and 2). A small group of dogs from Denmark clustered in one corner (5-6 cases and one control). However, by including countries and the teen first PCA as a covariate in the GWAS, the associated region on CAF37 remained stable, indicating a low effect of any population structure.

Discussion

We have identified a significant region on CFA37 with 18 SNPs associated with PNC in the Norwegian buhund. The region includes five relevant candidate genes, *CRYGA*, *CRYGB*, *CRYGC*, *CRYGD*, and *PIKFYVE* previously associated with PNC. The five candidate genes are 4-5 Mb apart from the top SNP (Figure 1). However, most of the significant SNPs in the associated region have p-values at a similar level as the top SNP.

Crystallins are important lens proteins and belong to the soluble proteins in the eye. Crystallins constitute about 85-90% of the lens protein content, depending on the species (Maggs et al., 2018). Crystallins can be classified into two superfamilies: α-crystallins and βϒ -crystallins (Slingsby et al., 2013). βϒ-crystallins, known primarily as structural proteins, exhibit high expression levels in the lens but are also found in other parts of the body, such as the brain (Slingsby et al., 2013). Mutations in crystallin genes tend to result in nuclear or lamellar lens opacities (Shiels et al., 2010).

Sanger sequencing was initially performed on the exons of the four crystalline genes but did not reveal any causal variants. However, one SNP was identified in *CRYGD*'s second exon, showing a difference in allele frequency between cases and controls. This variant is already known and predicted to be benign. It appears likely that the difference is due to a distinct linkage between this variant and a potential causal variant, as observed in the GWAS. Although no causal variants were found among the sequenced candidate genes, we cannot exclude the possibility of a causal variant existing in non-coding or regulatory regions of the candidate genes.

Another candidate gene is *PIKFYVE* (not yet sequenced). This gene encodes the enzyme FYVE finger phosphoinositide kinase. The FYVE finger phosphoinositide kinase phosphorylates phosphatidylinositol (PtdIns). PtdIns serve multiple functions and participate in signal transduction, regulating membrane traffic, modulating membrane permeability, and influencing cytoskeletal functions (Di Paolo and De Camilli, 2006; Burke et al., 2023). A mutation in *PIKFYVE* has been found to be associated with congenital nuclear pulverulent cataracts in humans and early-onset cataracts in zebrafish (Mei et al., 2022).

The two genes closest to the top SNP are *BMPR2* and *FAM117B.1.* As far as we know, neither of these two genes has been previously associated with cataracts. *BMPR2* belongs to the bone morphogenetic protein (BMP) receptor family. The BMP receptor family are transmembrane serine/threonine kinases and serves as a ligand of the TGF-beta superfamily, binding TGF-beta receptors to activate SMAD family transcription factors (Chen et al., 2004). They play a crucial role in regulating gene expression and are involved in the embryological development of the heart, neurons and cartilage (Chen et al., 2004). In addition, it has been shown that TGFβ receptors play a part in lens fibre differentiation in mice (de Iongh et al., 2001). Notably, other members of the BMP family, such as *BMP4*, have been associated with cataracts in humans (Vidya et al., 2018).

The associated region also contains several uncharacterized protein-coding genes as well as different forms of non-coding RNA-loci.

PIKFYVE and BMPR2 were not included in the candidate gene study. However, we have initiated whole-genome sequencing of selected cases and controls, which will cover non-coding areas and relevant genes within the associated region, including *PIKFYVE* and *BMPR2*, and enable variant calling (work in progress).

Even without an identified casual variant, the significant SNPs and haplotypes in the associated region have the potential to be used in marker-assisted selection. In dogs, considering the presumed high effect size of variants in the identified region, this may be a more realistic alternative to introduce and follow up than using genomic prediction based on a high number of SNPs covering the whole genome. Both GWAS and Sanger sequencing have identified SNP markers associated with an increased risk of PNC, and risk estimates of single SNPs have shown an odds ratio of 4-6 (done in the same material). The odds ratio of the SNP identified in *CRYGD* exon two has a slightly higher odds ratio than the top SNP; this could be due to the random selection of dogs, as the candidate gene sequencing is based on a subsample of the dogs in the GWAS.

We have examined the cumulative effect of five selected risk alleles in cases and controls. An increase in the number of risk alleles was associated with an elevated risk of PNC. The five SNPs are among the top six most significant SNPs and have a distance of about 1 Mb between each SNPs. This covers an area of about 4 Mb in the significant region. The distance between the SNPs corresponds well with the estimated mean haploblock size in the canine genome of 1 Mb (Karlsson et al., 2013). The potential utility of these markers in marker-assisted selection is intriguing. However, it is essential to validate these markers using an independent dataset.

One unaffected dog stands out for carrying a significant number of risk alleles. Interestingly, this dog produced an affected offspring when paired with an unaffected male. In the study of pedigrees available in the NKKs database, we have found a couple of other examples where the combinations of two unaffected parents (examined after the age of seven) have produced affected offspring. This may indicate an incomplete dominance of the trait.

Kristiansen et al. pointed out that there are differences in the lenticular changes observed among affected dogs (Kristiansen et al., 2017). Some dogs show subtle alterations visible only at an advanced age, while others exhibit noticeable changes from puppyhood, progressing to more advanced alterations involving the entire lens nucleus. Several factors may contribute to the varying degrees of the disease in dogs. The more severe manifestations could result from homozygosity for the variant, the influence of environmental factors, or the presence of other mutations in a different locus causing similar disorders. Additionally, it is possible that PNC is influenced by several loci. However, the association analysis revealed no additional significant genomic regions, indicating that if there are any other additional variants, they likely have a smaller effect size and may not be detectable with our current sample size. At the present time, there is no consistent grading system for PNC. Introducing a grading system could also be helpful in gaining more insight into the disease.

A limitation of the study is an overrepresentation of affected dogs (121 affected relative to 39 unaffected) in the data material. This disparity arises because most dogs are examined as part of the breed's breeding scheme at a young age and prior to breeding, and generally, a lower number of dogs are examined after the age of seven. Additionally, the high prevalence of the disease in the study population has made it challenging to obtain an equal number of cases and controls.

Conclusion

We have identified a significant associated region on CFA37 spanning from 11422235-17384184. The region contains several significant SNPs and five relevant candidate genes: *CRYGA*, *CRYGB*, *CRYGC*, *CRYGD*, and PIKFYVE. A causal variant has not been identified in the region, which may partly be due to the challenge associated with the classification of (genetic) healthy controls in a situation with dominant inheritance with reduced penetrance and late time of onset. The identified SNPs and haplotypes in the region have however the potential to be valuable in marker-assisted selection to contribute to a lower prevalence of PNC. Further sequencing of the region should be performed to reveal a potential causal variant.

References

ACVO, (2020). "The blue book" - Ocular disorders, proven or presumed to be inherited in purebred dogs, https://www.ofa.org/diseases/eye-certification/blue-book.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., et al., (2010). A method and server for predicting damaging missense mutations. Nat Methods 7, 248-9.

Arora, A., Minogue, P.J., Liu, X., Addison, P.K., Russel-Eggitt, I., et al., (2008). A novel connexin50 mutation associated with congenital nuclear pulverulent cataracts. J Med Genet 45, 155-60.

Bateman, J.B., Geyer, D.D., Flodman, P., Johannes, M., Sikela, J., et al., (2000). A new betaA1-crystallin splice junction mutation in autosomal dominant cataract. Invest Ophthalmol Vis Sci 41, 3278-85.

Bellamy, K.K.L., Lingaas, F., (2023). Cataracts in Havanese: genome wide association study reveals two loci associated with posterior polar cataract. Canine Med Genet 10, 5.

Berry, V., Georgiou, M., Fujinami, K., Quinlan, R., Moore, A., Michaelides, M., (2020a). Inherited cataracts: molecular genetics, clinical features, disease mechanisms and novel therapeutic approaches. 104, 1331-7.

Berry, V., Pontikos, N., Dudakova, L., Moore, A.T., Quinlan, R., et al., (2020b). A novel missense mutation in LIM2 causing isolated autosomal dominant congenital cataract. Ophthalmic Genetics 41, 131-4.

Bhatnagar, S. manhattanly: Interactive Q-Q and Manhattan Plots Using 'plotly.js'. (2021); R package version 0.3.0. Available at: https://CRAN.R-project.org/package=manhattanly. Accessed 11. july 2022.

Bjerkås, E., Haaland, M.B., (1995). Pulverulent nuclear cataract in the Norwegian buhund. J Small Anim Pract 36, 471-4.

Burke, J.E., Triscott, J., Emerling, B.M., Hammond, G.R.V., (2023). Beyond PI3Ks: targeting phosphoinositide kinases in disease. Nature Reviews Drug Discovery 22, 357-86.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4.

Chen, D., Zhao, M., Mundy, G.R., (2004). Bone Morphogenetic Proteins. Growth Factors 22, 233-41.

de Iongh, R.U., Lovicu, F.J., Overbeek, P.A., Schneider, M.D., Joya, J., et al., (2001). Requirement for TGFbeta receptor signaling during terminal lens fiber differentiation. Development 128, 3995-4010.

Di Paolo, G., De Camilli, P., (2006). Phosphoinositides in cell regulation and membrane dynamics. Nature 443, 651-7.

Ding, X., Wang, B., Luo, Y., Hu, S., Zhou, G., et al., (2011). A novel mutation in the connexin 46 (GJA3) gene associated with congenital cataract in a Chinese pedigree. Mol Vis 17, 1343-9.

ECVO. ECVO Manual: Chapter 5 - Definitions (2021). (2021); https://www.ecvo.org/media/5-definitions.pdf. Accessed May 09 2022.

Flaxman, S.R., Bourne, R.R.A., Resnikoff, S., Ackland, P., Braithwaite, T., et al., (2017). Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. Lancet Glob Health 5, e1221-e34.

Gelatt, K.N., Gilger, B.C., Kern, T.J., (2013). Veterinary ophthalmology : Volume 2. 5th ed. ed. Blackwell, Ames, Iowa.

366 Heinrich, C.L., Lakhani, K.H., Featherstone, H.J., Barnett, K.C., (2006). Cataract in the UK
367 Leonberger population. Vet Ophthalmol 9, 350-6.
368 Hejtmancik, J.F., (2008). Congenital cataracts and their molecular genetics. Seminars in Cell
369 & Developmental Biology 19, 134-49.
370 Jamieson, R.V., Perveen, R., Kerr, B., Carette, M., Yardley, J., et al., (2002). Domain
371 disruption and mutation of the bZIP transcription factor, MAF,associated with
372 cataract, ocular anterior segment dysgenesis and coloboma. Hum Mol Genet 11, 33-
373 42.
374 Karlsson, E.K., Sigurdsson, S., Ivansson, E., Thomas, R., Elvers, I., et al., (2013). Genome-wide
375 analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory
376 variants near CDKN2A/B. Genome Biol 14, R132.
377 Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., et al., (2002). The human
378 genome browser at UCSC. Genome Res 12, 996-1006.
379 Kristiansen, E., Revold, T., Lingaas, F., Narfstrom, K., Pedersen, P.B., et al., (2017). Cataracts
380 in the Norwegian Buhund-current prevalence and characteristics. Vet Ophthalmol
381 20, 460-7.
382 Kumar, M., Agarwal, T., Khokhar, S., Kumar, M., Kaur, P., et al., (2011). Mutation screening
383 and genotype phenotype correlation of α-crystallin, γ-crystallin and GJA8 gene in
384 congenital cataract. Mol Vis 17, 693-707.
385 Liu, Y.-C., Wilkins, M., Kim, T., Malyugin, B., Mehta, J.S., (2017). Cataracts. The Lancet 390,
386 600-12.
387 Lu, S., Zhao, C., Jiao, H., Kere, J., Tang, X., et al., (2007). Two Chinese families with
388 pulverulent congenital cataracts and deltaG91 CRYBA1 mutations. Mol Vis 13, 1154-
389 60.
390 Maggs, D.J., Miller, P.E., Ofri, R., (2018). Slatter's fundamentals of veterinary ophthalmology.
391 Sixth edition. ed. Elsevier, St. Louis, Missouri.
392 Major, T.J.T., Riku. LocusZoom-like Plots for GWAS Results. (2021);
393 https://github.com/Geeketics/LocusZooms. Accessed 05.12.2022.
394 Mei, S., Wu, Y., Wang, Y., Cui, Y., Zhang, M., et al., (2022). Disruption of PIKFYVE causes
395 congenital cataract in human and zebrafish. Elife 11.
396 Mellersh, C.S., Graves, K.T., McLaughlin, B., Ennis, R.B., Pettitt, L., et al., (2007). Mutation in
397 HSF4 associated with early but not late-onset hereditary cataract in the Boston
398 Terrier. J Hered 98, 531-3.
399 Mellersh, C.S., McLaughlin, B., Ahonen, S., Pettitt, L., Lohi, H., Barnett, K.C., (2009). Mutation
400 in HSF4 is associated with hereditary cataract in the Australian Shepherd. Vet
401 Ophthalmol 12, 372-8.
402 Mellersh, C.S., Pettitt, L., Forman, O.P., Vaudin, M., Barnett, K.C., (2006). Identification of
403 mutations in HSF4 in dogs of three different breeds with hereditary cataracts. Vet
404 Ophthalmol 9, 369-78.
405 Meyer, E., Rahman, F., Owens, J., Pasha, S., Morgan, N.V., et al., (2009). Initiation codon
406 mutation in betaB1-crystallin (CRYBB1) associated with autosomal recessive nuclear
407 pulverulent cataract. Mol Vis 15, 1014-9.
408 NKK. NKK Hundedatabasen / Dogweb. (2023); https://www.dogweb.no/hundedatabase.
409 Accessed 25.08.2023.
410 O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., et al., (2016). Reference
411 sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and
412 functional annotation. Nucleic Acids Res 44, D733-45.

Purcell S., C.C. Plink Web Page. (2023); www.cog-genomics.org/plink/1.9/. Accessed 30.01.2023.
R Core Team. R: A Language and Environment for Statistical Computing. (2021); https://www.R-project.org/.
Ricketts, S.L., Pettitt, L., McLaughlin, B., Jenkins, C.A., Mellersh, C.S., (2015). A novel locus on canine chromosome 13 is associated with cataract in the Australian Shepherd breed of domestic dog. Mamm Genome 26, 257-63.
Rudd Garces, G., Christen, M., Loechel, R., Jagannathan, V., Leeb, T., (2022). FYCO1 Frameshift Deletion in Wirehaired Pointing Griffon Dogs with Juvenile Cataract. Genes (Basel) 13.
Shiels, A., Bennett, T.M., Hejtmancik, J.F., (2010). Cat-Map: putting cataract on the map. Mol Vis 16, 2007-15.
Shiels, A., Hejtmancik, J.F., (2019). Biology of Inherited Cataracts and Opportunities for Treatment. Annu Rev Vis Sci 5, 123-49.
Slingsby, C., Wistow, G.J., Clark, A.R., (2013). Evolution of crystallins for a role in the vertebrate eye lens. Protein Sci 22, 367-80.
Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A., (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35, W71-4.
USCS. UCSC Genome Browser. (2023); http://genome.ucsc.edu Accessed 18.08.2023.
Vidya, N.G., Vasavada, A.R., Rajkumar, S., (2018). Evaluating the association of bone morphogenetic protein 4-V152A and SIX homeobox 6-H141N polymorphisms with congenital cataract and microphthalmia in Western Indian population. J Postgrad Med 64, 86-91.
Wang, C., Wallerman, O., Arendt, M.L., Sundström, E., Karlsson, Å., et al., (2021). A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. Commun Biol 4, 185.
Wang, W., Jiang, J., Zhu, Y., Li, J., Jin, C., et al., (2010). A novel mutation in the major intrinsic protein (MIP) associated with autosomal dominant congenital cataracts in a Chinese family. Mol Vis 16, 534-9.
WHO. Fact sheets Blindness and vision impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. Accessed 28.01.2019.
Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4, 1686.
Yan, M., Xiong, C., Ye, S.Q., Chen, Y., Ke, M., et al., (2008). A novel connexin 50 (GJA8) mutation in a Chinese family with a dominant congenital pulverulent nuclear cataract. Mol Vis 14, 418-24.
Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88, 76-82.

Figures

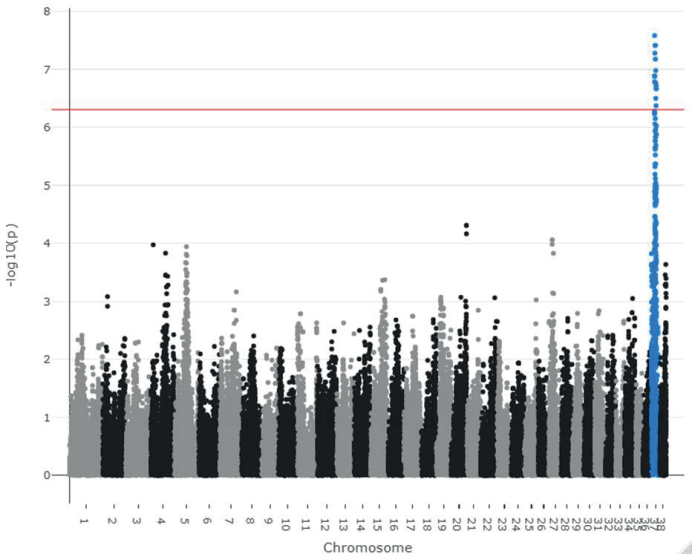


Figure 1A. A Manhattan plot showing the results of the MLMA in GCTA with the autosomal chromosomes on the X-axis and the -log p-value on the Y-axis. Chromosome 37 is marked blue. The red line indicates the significance level using a Bonferroni correction at 5.09 x10^{-07} according to the number of SNP markers in the dataset.

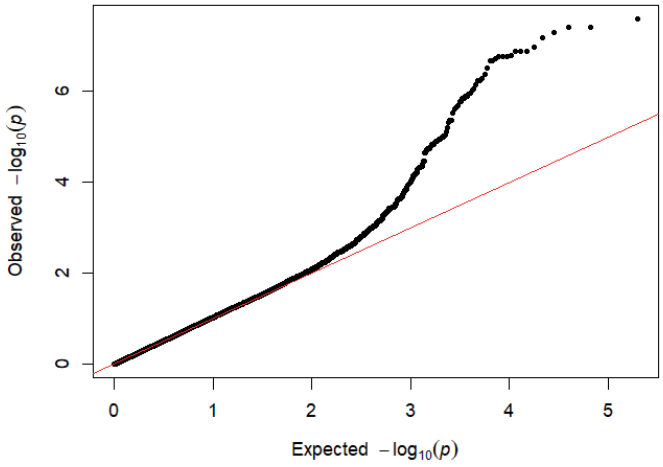


Figure 1B. A quantile-quantile plot (QQ-plot) showing the expected p-value plotted against the observed p-value. The p-values were estimated in GCTA.

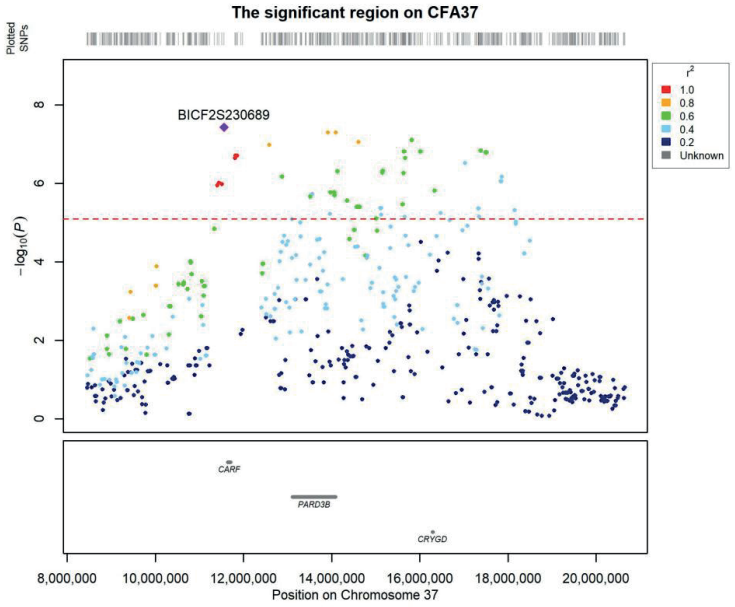


Figure 2. A LocusZoom similar plot is displaying the significant region on CFA37. The top SNP BICF2S230689 is labelled purple, each datapoint represents one SNP, and the colour of the points indicates the level of linkage disequilibrium (r^2) between the top SNP and the respective variant, according to the legend on the top right corner. r^2 is calculated in Plink. In the lower window, three of the genes in the associated region are included.

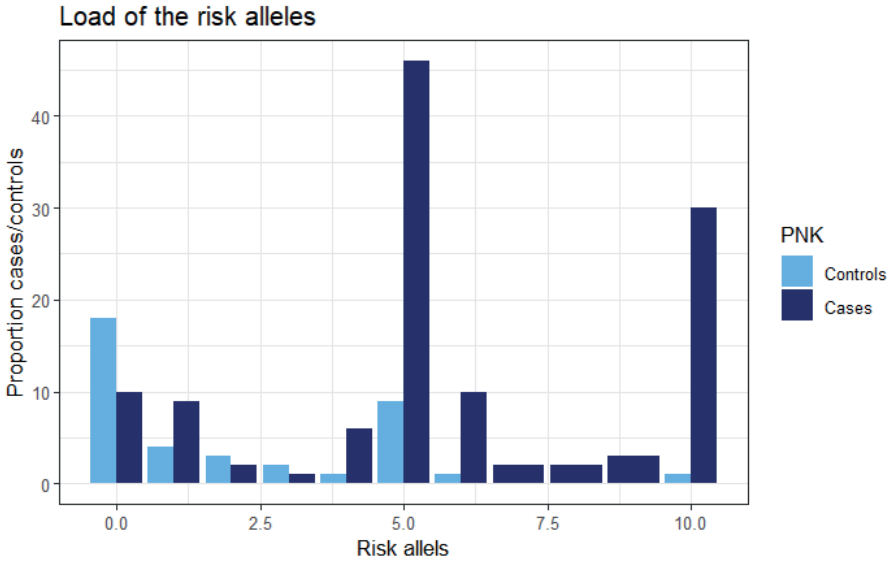
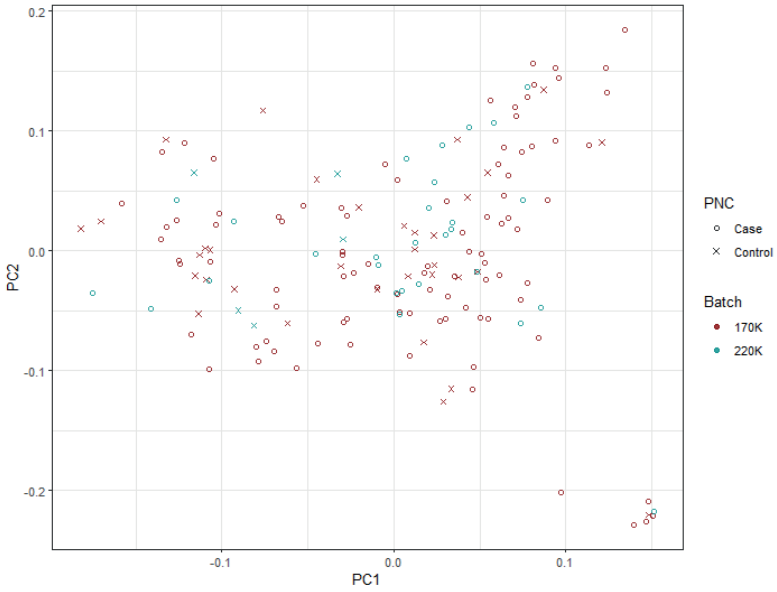


Figure 3. The total load of risk alleles in cases and controls distributed over five significant loci (BICF2S230689, BICF2P165614, BICF2P194369, BICF2P375607, TIGRP2P419549).

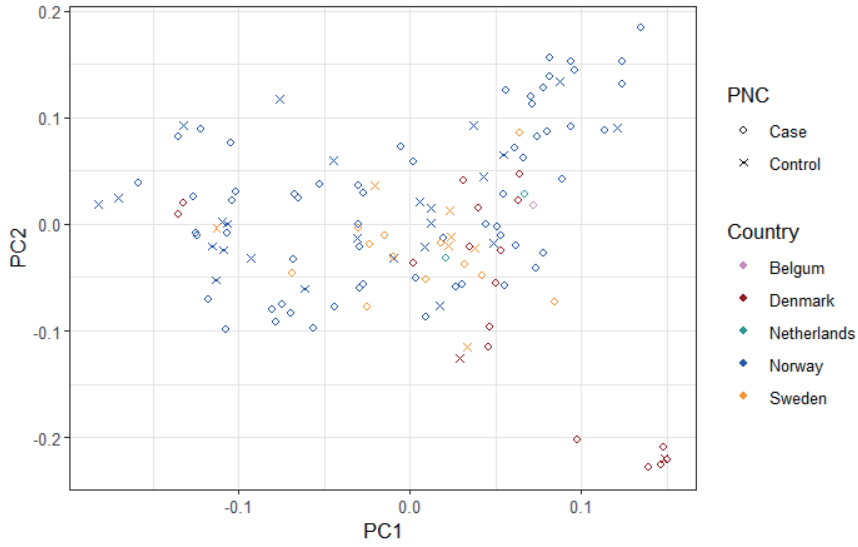
Supplementary material Paper III

Supplementary Table 1. Primers used in the candidate gene study of the four crystalline genes on CFAF37.

Gene	Exon	Forward	Reverse
CRYGA	1	ccaaccccatctcagcagtt	tgaagtagggctgcaggttg
CRYGA	2	catataccggctgccacact	tcccgagtggaaacaaaggg
CRYGA	3	gaggacatgtatccggcagg	cagggtcaggtatgttgcca
CRYGB	1	ttactcaccgaaatgggccc	gctttcgtgtccaaagagcc
CRYGB	2	atgggaaaggtaagtgcggg	agcaagcaccacagagttca
CRYGB	3	ttctctctgcctctccagca	aggctttaaaagaatgctggca
CRYGB	3	ggggcaggctctttaacact	gcaggactccatagtgccag
CRYGC	1	gcagacagcctctggaatca	tgaagtagggctgcaggttg
CRYGC	2	gggccccttttgtgttgttc	ccaaagcccactctaccctg
CRYGC	3	tccatacacggcacacacag	caacctttaccggggagcat
CRYGC	3	cgtgcctcaacctaccaagt	gacaggccacctcctctttc
CRYGD	1	agtgcagcacgtaaaaggga	tactggtggccctggtagtt
CRYGD	2	ccttttgtgtggttcctgcc	cctttgcttctgtttggggc
CRYGD	3	aggtactgaatctcgagctttga	accaatgccaggaacacact
CRYGD	3	tcatgctttgctttcccccct	atggctcggtggttgaacat



Supplementary Figure 1. Distribution of cases and controls in the two baches with 170K and 220K SNP markers.



Supplementary Figure 2. Distribution of cases and controls and the dogs from different countries.

ISBN: 978-82-575--2113-4
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

1 19212 / MEDIEHUSET ANDVORD.NO