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

Detection and characterization of genomic regions associated to Dilated Cardiomyopathy in Great Danes

Ana Letícia Kolicheski

Examensarbete / Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics,
391
Uppsala 2012

Master's Thesis, 30 HEC

Erasmus Mundus Programme – European Master in Animal Breeding and Genetics

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

Detection and characterization of genomic regions associated to Dilated Cardiomyopathy in Great Danes

Ana Letícia Kolicheski

Supervisors:
Göran Andersson, SLU, Department of Animal Breeding and Genetics
Jennifer Meadows, SLU, Department of Animal Breeding and Genetics
Susanne Björnerfeldt, SLU, Department of Animal Breeding and Genetics
Examiner:
Stefan Marklund, SLU, Department of Animal Breeding and Genetics

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

Place of publication: Uppsala
Year of publication: 2012
Name of series: Examensarbete / Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, 391
On-line publication: http://epsilon.slu.se

Key words: Great Dane, DCM, cardiomyopathy, dog, GWAS

Acknowledgements

First and foremost, I would like to thank all Erasmus Mundus Animal Breeding and Genetics (EMABG) staff, with a special mention of the program coordinator Prof. Johan van Arendonk and the SLU program coordinator Prof. Birgitta Malmfors. For giving me the opportunity, assistance, and making my participation in this spectacular master's program possible. My Brazilian referees: Dr. Ricardo Lehtonen and Dr. Paulo Brandão. My supervisors, Dr. Jennifer Meadows, Dr. Susanne Björnefeldt and Prof. Göran Andersson; for all the motivation, corrections, advice, guidance and support. It is important to mention Prof. Jens Häggström and all veterinarians involved in the phenotyping and providing the data for this study, and the dog owners that collaborated to the study. Also, I would like to give a special thank you Emma, Ben and Maja, who shared the same office and were very helpful when I interrupted their work with my questions! Off course, my family and friends support was essential and they couldn't be left out. Even far away, my mother, father and brothers were a key piece of motivation. I have met many good people along these two years of study and I would like to thank Fede Fröstl, Anya Bubnova, Aliki Ilona, Alexandre Barrozo, Cindy Canton, Agnese Viluma, Merina Shrestha, Astri Olivia, Batmagnai Enkhbaatar and all from the international student group, for making my stay in Sweden pleasant and fun. Also, thanks to all my EMABG colleagues and last but not least, my friends that supported me online from far away: my boyfriend Nick Grobben, who tried to understand my thesis and even gave advice, though many times he isn't sure if he even understands what I study; Stefan Norenius, Thor Magni, Nathaly Gasparin and Celso Pedri.
I have certainly learned a lot in this short period, and am really thankful to everyone who was part of my life at this important time.

http://www.dogsonline.com/

"*Knowing how to think empowers you far beyond those who know only what to think*." Neil deGrasse Tyson

Index

Abstract

Canine dilated cardiomyopathy (DCM) is a lethal myocardial disease prevalent in large and giant-sized dog breeds, including the Great Dane (GD). The goal of this study was to identify genetic risk factors for DCM and to contribute to a future development of DCM genetic markers in order to facilitate disease diagnosis and risk allele detection. Furthermore, in the long term, the identification of novel genes and pathways that may contribute for development of future therapies. Two different approaches of genome wide association study (GWAS) using data from ~170,000 Single Nucleotide Polymorphism (SNP) markers were conducted in order to identify disease-associated regions of the genome. A previous GWAS to this study used a different cohort of individuals and analytical software. This methodology detected six candidate regions on chromosomes 3, 16, 19, 29 and 30 (two associated regions on chr19). The current GWAS tested nine different models in a dataset containing 127 Great Danes (62 cases and 65 controls). The GenABEL package for R was used for the analysis. The elected model 3 included population stratification and had an inflation factor lambda (λ) equal to 1.04. Two regions on chromosome 19 overlapped with the first GWAS. Region 1: between 23-23.3 Mb containing two rRNAs; Region 2: between 44-46Mb containing three genes: *SPOPL, NXPH2* and a *LRP1B* orthologue. No marker reached Bonferroni statistical significance, and the low power could be due to insufficient number of cases and controls. It is suggested that SNP typing of extra Great Danes and a better use of the available phenotype data would help to improve the statistical power. The correction for complex gene-gene interactions such as epistasis may be an interesting approach to detect novel candidates. The regions detected in the precedent GWAS were re-sequenced using Illumina Next generation sequencing technology in five animals (three controls and two cases), and the data was available for analysis in the current study. Six candidate non-synonymous mutations were detected in the *DMXL2, TMC3, ZMAT4, MESDC2* genes and in a paralogue to *RPL10A*. *DMXL2* encodes for a protein important for the Notch pathway in mammals. In humans, a SNP in this gene has been associated with ischemic stroke. The polymorphism in *DMXL2* was confirmed through the genotyping of 319 individuals and the non-reference allele is slightly more frequent on verified cases as opposed to verified controls, however, the difference is not statistically significant (t-test p-value= 0.366; 2X3 Fisher test p-value= 0.258). Further analysis on both regions on chr19 should be carried out in order to detect possible novel mutations, copy number variants, genes and regulatory elements that could influence the DCM phenotype.

Key words: Great Dane, DCM, cardiomyopathy, dog, GWAS

1. Introduction

Dilated cardiomyopathy (DCM) is a myocardial disease that affects different species including dogs (*Canis familiaris*). Several authors describe DCM pathology, and it generally consists of cardiac dilatation with atrio-ventricular incompetence and congestive heart failure; and consequently due to compensatory mechanisms: pulmonary edema, pleural effusion and ascites [1-4]. The prognosis is poor hence an affected dog has a ~34% chance of surviving one year after the initial diagnosis, with a median of only 126 days according to a retrospective study of 62 dogs [2]. Histopathologically there are two described types of DCM with probable different etiologies: *attenuated wavy fiber type* and *fatty infiltration-degenerative type*. The "wavy fiber type" is the most commonly found in dogs [4].

DCM in dogs has an overall prevalence of approximately 0.5%. It affects mostly pedigree dogs (0.65%) compared with crossbreeds (0.16%). Dogs from large- and giant-sized breeds are the most predisposed to the disease, including the Great Danes, Newfoundlands, Boxers, Irish Wolfhounds and Doberman Pinschers among others [5]. In this study we focus on Great Danes (GD) and it is the second most commonly affected dog breed with a clinical prevalence of 3.9%[5]. The DCM epidemiology facts suggest that even though the disease etiology is highly influenced by environmental factors (e.g. nutritional deficiencies), there is also a strong genetic factor underlying this pathology. There is also an apparent sex predisposition as male dogs are affected more often than female dogs [6]. The pedigree analysis of 17 Great Danes done by Meurs et al. [7] has suggested that DCM in Great Danes has a X-linked recessive mode of inheritance, which could resemble the X-linked DCM in humans. However, an autosomal recessive mode of inheritance hasn't been ruled out, therefore the type of inheritance for this particular breed is still unknown [7].

There are several DCM candidate genes and loci across breeds, but so far no actual causative mutations have been confirmed. The criteria for determination of candidate genes are generally based on (1) molecular information from a similar disease in a unique species such as human or mouse; (2) the prevalence of the candidate gene in a tissue of interest (myocardium, valve, etc); or (3) the specific function of the gene (i.e., structural protein, involved in channel regulation, etc)[8]. In humans, many causative genes have been identified, and these genes appear to encode mutant proteins of the cytoskeleton, sarcolemma and sarcomere, ion channels and transcription factors. [9-10]. Gene mutations in the lamin (*LMNA)* gene are the most prevalent in humans, and more than 20 other genes have been implicated in DCM etiology: myosin heavy chain (*MYH7)*; troponin (*TNNT2* and *TNNC2)*; actin (*ACTC1* and *ACTN2)*; dystrophin (*DMD*) genes and others. The mode of inheritance in familial dilated cardiomyopathy in humans varies, and it can be autosomal dominant (most common), autosomal recessive or X-linked [10]. Mutations in critical cytoskeletal proteins such as

metavinculin *VCL*, α-dystroglycan *DAG1*, α and γ- sarcoglycan *SGCA* and *SGCG*, and muscle LIM protein may also result in DCM. The canine homologues for each of these structural protein genes are potential candidates for canine DCM in general. In a study carried out by Wiersma et al., an evaluation of 15 candidate genes was made in four Newfoundland dog families [3]. Candidates like *TNNT2* and *LMNA* were not associated to DCM, suggesting that the genetic cause of DCM for this dog breed may be different from the currently known mutations in humans and experimental models of disease [3]. This might also be the case for the Great Danes. Several candidate genes were proposed for GD DCM by previous studies, but confirmed causative mutations have yet to be described. For example, Oyama et al. conducted a transcriptional study where they analyzed gene expression patterns of left ventricular tissue samples using three Great Dane dogs with end-stage DCM, and three large breed control dogs [11-12]. The study demonstrated that Great Danes with DCM showed an abnormal castalbin2 and triadin expression, exposing the ryanodine receptor 2 (*RYR2)* gene as a candidate. Another expression study using two Doberman Pinschers with end-stage DCM and five healthy dogs, identified 478 transcripts that were differentially expressed (≥ 2.5 fold change). The identified transcripts were divided in groups that revealed that pathways involving cellular energy production, signaling and communication, and cell structure were downregulated, whereas pathways involving cellular defense and stress responses were upregulated [12]. This represents how complex the DCM trait is, and that compensatory mechanisms and regulatory mutations are no less important than causative structural mutations.

More recently with the rapid advances in sequencing technology, the detection of candidate genes and associated regions got a whole new approach. High-resolution single nucleotide polymorphism (SNP) arrays allow the mapping of the whole genome for many individuals simultaneously. When whole genome and good phenotypic data information is available, it is now possible to detect differences in the genome between cases and controls. This gives the opportunity to find novel genes and mutations that may contribute to a better understanding of the pathology not only in the dog but also in all species affected by the disease.

Genome-wide association studies (GWAS) have been shown to be an efficient method to identify candidate genes for mono- and polygenetic diseases [13]. The dog is an optimal species when it comes to GWAS, and that is because it possess a large extent of linkage disequilibrium (LD) within dog breeds (distances of several megabases) and between breeds there is a shorter LD (over tens of kilobases) [14]. Because of this, as few as 10,000-30,000 SNPs are sufficient for a GWAS in the dog [15].

Based on this background, a GWAS was performed using the high density ~170,000 SNP array in 127 Great Danes (62 cases and 65 controls). A precedent GWAS detected six candidate regions in Great Danes (in chromosomes 3, 16, 19, 29 and 30). Two candidate regions were present on chromosome 19. All these associated regions were then re-sequenced by Illumina Next generation sequencing technology.

In the present study, a GWAS using different individuals was run. Two different datasets were used and in total nine different models were tested; the variables of the models included affection status (case or control), gender (male or female), population stratification and kinship matrix. A dataset including all 127 animals and a separated male dataset to examine the X chromosome was used, with the expectation of increasing the reliability of the associated regions detected in the previous analysis. Furthermore, the Illumina sequences of the candidate regions from the previous GWAS were analyzed in order to detect genetic variants between the affected and control Great Dane dogs that could be functional causative mutations.
The goal of this study was to contribute to the future development of GD DCM markers in order to facilitate its diagnosis, and risk assessment. In addition, the long-term goal is to identify novel genes and pathways that may contribute for a better understanding of the disease etiology, and for the development of future therapies.

2. Materials and methods

2.1. GWAS analysis:

The genotype data was obtained from an Illumina 170k SNP assay of 127 individuals (62 cases and 65 controls) from 7 different countries (24 from Sweden, 49 Germany, 12 Holland, 7 USA, 19 UK, 1 Finland, 1 France and 4 unknown). The criteria to determine which animals are cases and controls was based on proposed guidelines for DCM diagnosis [5]:

(1) DCM, verified PAD: The dog was diagnosed after autopsy and histological examination of myocardium;
(2) DCM, CHF: The dog's medical history was consistent with DCM and had developed clear signs of heart failure and pulmonary edema, and ascites leverstas;
(3) DCM, verified echo: The diagnosis was confirmed by ultrasound, according to a standardized method developed specifically for the disease [5];

Two different phenotypical datasets (Appendix 1) were analyzed using the GenABEL package (http://www.genabel.org/) [16-17] for the RStudio™ open source integrated development environment for R. The datasets and had different analytical purposes:

Dataset 1: 127 animals (62 cases and 63 controls) and (74 males and 53 females);

Dataset 2: 67 individuals, all males (44 cases and 23 controls).

The purpose of dataset 1 was to include as many animals as possible to test for association significance on autosomal markers. The dataset 2 allowed testing

markers on the X chromosome. The phenotype data input contained information such as animal ID, gender (male or female) and affected status (case or control). The data went through two quality controls:

Quality Control 1: Preliminary control in order to remove markers with call rate (< 0.95); markers with minor allele frequency (MAF<10^{-8}); markers which are very strongly out of Hardy-Weinberg equilibrium (p-level <10^{-8}) and individuals with more than 5% missing genotypes.

Quality Control 2: Eliminates markers that are out of Hardy-Weinberg equilibrium in a more strict approach. In this quality control, the rate of unacceptably high individual heterozygosity false discovery rate (FDR <0.2) was considered. This was applied only within the control group, because in cases the deviation can be the actual signal.

Statistical models: An exploratory analysis of the data was performed, including a t-test to check whether the number of affected males is significantly different than the number of affected females, followed by a Fisher's exact test to calculate the proportion of affected males compared to affected females. For the genetic test of markers associated to the trait, the functions 'qtscore' and 'egscore' from the GenABEL package were used (Table 1). The parameters included in the model were corrected until the inflation factor lambda (λ) was equal to/the closest to 1. Values below 1 were considered overcorrected. In all models, the fixed factor is disease affection status (case or control) and this parameter is called "AffStat" in the formula. Gender was used as a covariate in the models 2,4 and 6 (dataset 1). The subpopulation structure was estimated using a within sum of squares (wss) versus cluster graph. The population substructure (K=3) was then included in the models 3, 4 (dataset 1) and 2 (dataset 2).

The IBS for each pair of individuals *i* and *j* was computed as:

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

Where *k* changes from 1 to *N* number of SNPs genome wide, x_{i,k} is a genotype of *i*th individual at the *k*th SNP, coded as 0, 1/2, 1 and p_k is the frequency of the "+" allele. This apparently provides an unbiased estimate of the kinship coefficient (how related the individuals in the population are to each other). The kinship matrix, which contained these relatedness values, was then determined (as returned by IBS function) and considered in the models 5,6 (dataset 1) and 3 (dataset 2).

Table 1. Statistical models implemented with the different datasets.

	Dataset 1
1	qtscore(AffStat, data, trait= "binomial")
2	qtscore(AffStat~sex, data, trait= "binomial")
3	qtscore(AffStat, data, strata=pop, trait="binomial")
4	qtscore(AffStat~sex, data, strata=pop, trait="binomial")
5	egscore(AffStat, data, kinship=data.gkin)
6	egscore(AffStat~sex, data, kinship=data.gkin)
	Dataset 2
1	qtscore(Affstat, data2, trait="binomial")
2	qtscore(AffStat, data2, strata=pop, trait="binomial")
3	egscore(AffStat, data2, kinship=data.gkin)

The lambda was computed by regression in a Quantile-Quantile (Q-Q) plot for each model. The raw and the genomic control corrected p-values were plotted as a Manhattan plot. A Bonferroni correction limit line (-log10(0.05/nsnps) = 6.39) was set and added to the plot. An additional threshold (-log10(1e-05) = 5) for model 3 was determined based on inflated markers visualized in the Q-Q plot, and included in the Manhattan plot. Lastly, the top 20 SNPs according to their genomic control p-value (Pc1df) were summarized.

2.2. Resequence Analysis and polymorphism detection:

Re-sequence data from five Great Danes (two cases and three controls) was available for this methodology. The data visualization was performed using the software IGV [18]. The software allowed an overview of the sequence quality (coverage and number of reads). Furthermore, this visualization allowed the confirmation of haplotype blocks and key SNPs that were detected in the previous GWAS and also the discovery of polymorphisms that segregate in opposite directions between cases and controls.

The SNPs and indels were then scored by conservation across species using the software SeqScoring (http://www.seqscoring.net/) [19]. The genome browser UCSC (http://genome.ucsc.edu/) was used in order to visualize the location of these polymorphisms, and in this way it was determined whether they are located in conserved regions (exons, 5' and 3'-UTRs, transcription factor binding sites and other regulatory elements).

Finally, the "Variant Effect Predictor" tool from Ensembl (http://www.ensembl.org/index.html) was applied to the polymorphisms with a pattern score ≥3. However, pattern scores higher than 3 don't necessarily mean allele segregation between the case/control groups (Table 2). Therefore, variant alleles exclusively found in cases (either fixed or heterozygote) for an opposite allele as the one found in controls, were preferred as candidates.

Table 2. Example of SeqScoring output showing how the relevant polymorphisms were detected.

Example	Reference	Control 1	Control 2	Case 1	Case 2	*Pattern Score
1	A	A	A	G	G	5
2	G	R	R	A	A	4
3	G	R	R	A	G	4

*A pattern score equal to 5 and a clear segregation on case vs control was considered the best case scenario. The two last examples above have a pattern score equal to 4, but the last one is not segregating, meaning that pattern score value alone is not sufficient to determine how different cases are from controls.

2.3. PCR and Sanger Sequencing

2.3.1. Primer design:

Polymorphisms matching the criteria of section 2.2 were detected in the genes *TMC3, MESDC2, ZMAT4* and *DMXL-2*. In order to confirm these polymorphisms, PCR primers for genotyping the candidate mutations were designed using the Primer3 software (**http://frodo.wi.mit.edu/primer3/**).

In total, 4 primer pairs were designed (Table 3). The melting temperatures of the primers were between 59.8°C and 61.0°C.

Table 3.Primers

Primer Name	primer sequence	Chr	Sense	Start	End
MESDC2-U1	GTCCTGCGAGAACTTTGGTC	3	+	59569235	59569255
MESDC2-U2	CAGTTGAGCACCCACCTTTT	3	-	59569867	59569887
TMC3-U1	ACCACCTGCTTTGTCTCTGG	3	+	59313913	59313932
TMC3-D1	CAGCTCACCTGTGAGTGTCC	3	-	59314394	59314413
ZMAT4-U1	TCAGTGAGAAAGTCTGCCGTAA	16	+	27691973	27691994
ZMAT4-D1	AAGAGCCTTGCAGAAGATGC	16	-	27692474	27692493
DMXL2-U2	TGGGAATAAGAGTCGGAAGC	30	+	20222840	20222859
DMXL2-D2	CTCGGATTCCTGTTGCATTT	30	-	20223492	20223511

2.3.2. PCR

Available DNA from seven Great Danes (four cases and three controls) was used for the PCR and subsequent Sanger sequencing reaction. The total reaction volume of the PCR was 20 µl, composed of:
Applied Biosystems 1X PCR Buffer (50 mM potassium chloride and 10 mM Tris-HCl, pH 7,5) + $MgCl_2$ [1.5 mM]+ dNTP[0.20µM] + primer pair [0.25 µM per primer] + 0.14 µl Taq polymerase [0.035 U/µl]+ 1 ng/µl DNA and addition of milliQ water until the completion of the 20 µl reaction volume. PCR reactions were performed in Applied Biosystems 2720 Thermal Cycler machines using the program

represented in Figure 1. The products were then quantified on 1.5% agarose gel electrophoresis using standard 100bp precision marker (Fermentas).

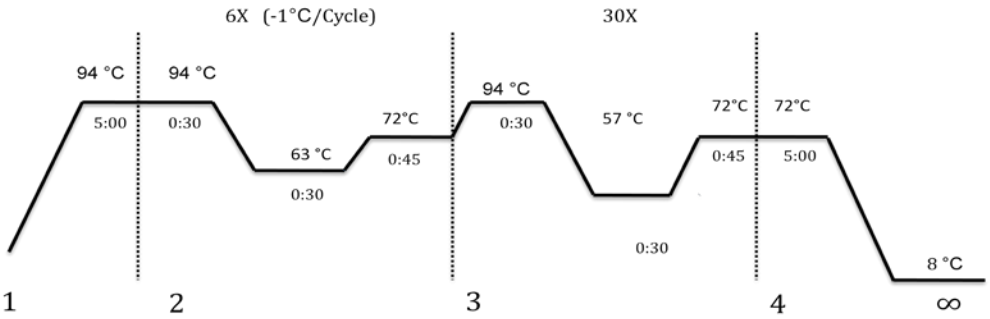

Figure 1. PCR cycle. 1: denaturing at 94°C for 5 minutes; 2: Touchdown cycle (-1 °C per cycle for 6 cycles) of denaturation, annealing and extension; 3. 30 cycles of denaturation, annealing and extension; 4. End of cycle.

2.3.3. Sanger sequencing reaction

PCR products were treated with approximately 1.6 U of calf intestinal alkaline phosphatase (CIAP) + 2U of Exonuclease 1 (Exo1) in a 1X Exo1 Buffer (Fermentas). The reactions were incubated at 37°C for one hour and inactivated at 85°C for 15 minutes in the PCR machine.

The PCR products with the CIAP mix were then diluted and distributed into a sequencing plate. Per well, each reaction contained 18 µl of total volume. It included 12 µl of H_2O, 3 µl PCR product and 3 µl of primer (5µM). The samples were then submitted to the Uppsala Genome Center for sequencing.

2.3.4. Sequence Analysis

Obtained nucleotide sequences were analyzed using the software CodonCode Aligner version 3.5.6 (CodonCode Corporation). The sequence ends were trimmed and the sequences aligned. The polymorphisms were detected and identified according to position (Figure 2). Also, the Dog (*Canis lupus familiaris*) genome browser UCSC BLAST tool (Broad/canFam2 assembly) was used to determine the exact genomic position.

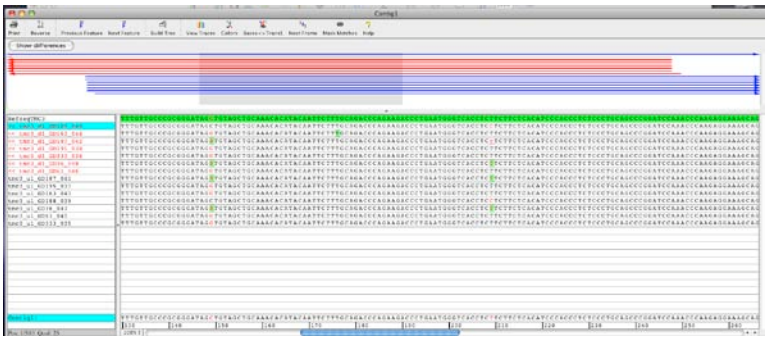

Figure 2. Sequence Analysis using CodonCode Aligner software. An illustrative representation of the visualization of two polymorphisms on *TMC3* gene

2.4. TaqMan genotyping:

In total, 332 individuals: 178 females, 142 males, 12 unknown; 55 cases (43 verified), 77 controls (36 verified) and 11 unknown were genotyped for five different polymorphisms in four genes (*ZMAT4, TMC3, MESDC2 and DMXL2*). The 332 dogs originated from nine different countries: (205 from Sweden, 6 Finland, 48 Norway, 3 Denmark, 47 Germany, 2 The Netherlands, 1 France, 10 USA and 10 unknown).
The genotyping assays were designed according to custom manufacturer specifications (Applied Biosystems, Life technologies).
The reaction was carried out in a 384 well plate. The total volume reaction per well was 5 µl, consisting of 1X of Applied Biosystems ®TaqMan Genotyping Master Mix ; 0.125 µl 1X Applied Biosystems ®TaqMan Genotyping Assay (primers + probes); 1.375 µl MQ H_2O and 10 ng of DNA. A provided allelic discrimination guide (Appendix 2) was used. The software SDS 2.3 connected the machine Applied Biosystems, 7900HT Fast Real-Time PCR system generated the results. The cycle conditions are represented in Figure 3.

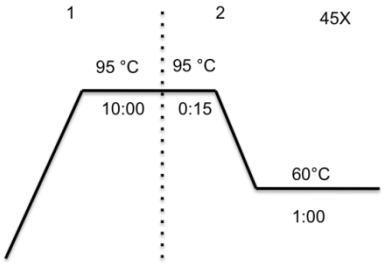

Figure 3. TaqMan cycling conditions. 1. Denaturing at 95°C for 10 minutes 2. 45X 95°C for 15 seconds denaturing followed by annealing temperature of 60 °C for 1 minute.

The output produced by the software SDS 2.3 was exported into a .txt and graphs as .png files. The allele frequencies calculations and a simple statistics: Student t-test and Fisher exact (2X2, 2X3 and recessive inheritance test) on the alternative allele frequency difference between cases and controls was performed using the website http://vassarstats.net/.

3. Results

3.1. GWAS Analysis:

In total, 127 dogs (62 cases and 65 controls) composed the dataset. Where 74 are males (23 controls and 51 cases), and 53 are females (42 controls and 11 cases). In our data set the number of affected males is significantly different compared to the number of affected females, and that is proven with a simple t-test (Table 4). A Fisher exact test was used to determine the odds ratio, which is approximately eight affected males for each affected female (Table 5).

Table 4. t.test (H_0 = true difference in means is equal to zero)

t	df	p-value	95% CI	
-6.06	121.33	1.51E-08	-0.62	-0.31

Table 5. Fisher exact test (H_0 = true odds ratio is equal to 1)

p-value	95% CI		Odds ratio
1.00E-07	3.46	21.31	8.30

After the first quality control, six individuals were excluded due to high identity by state IBS (>=0.95): "cHD8_1004" "cHD8_1016" "cHD8_1055" "cHD8_942" "cHD8_951" and "cHD8_965". The excluded dogs are actually repeated measures of three individuals (one male case and one female control from Germany and one female control from Finland).

The subpopulations were determined. According to the wss vs cluster graph (Appendix 1), K is between 2 and 4. When applying K=2 or K=4 to the statistical models, the λ is either inflated (λ> 1.2) or overcorrected (λ<1) in comparison to K=3. Therefore, the number of clusters and subpopulations is more likely equal to 3 (λ = 1.04) (Figure 4).

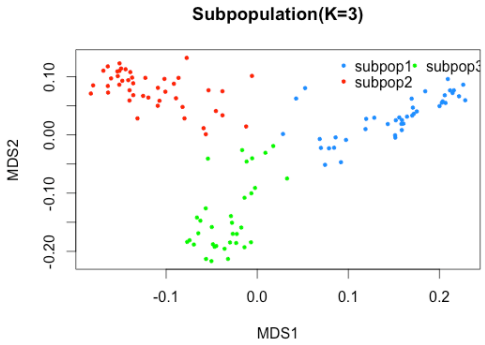

Figure 4: Subpopulation clusters (K=3)

The proportion of cases and controls (females and males) per subpopulation cluster is represented in Table 6.

Table 6. Number of cases and controls per subpopulations 1,2 and 3.

	Control	Cases	Total
1	19(12,7)	23(6,17)	42
2	12(7,5)	22(2,20)	34
3	30(19,11)	15(3,12)	45
Total	61	60	**121**

*The values in parentheses are the number of females and males respectively.

The number of markers that were eliminated in the quality controls for the dataset 1 is expressed in Table 7. The Table also contains the lambda(λ) values for each model, and which chromosomes contain the top 20 SNPs according to their genomic control p-value (Pc1df).

Table 7. Quality control output for dataset 1 and Inflation factor λ and Chromosomes containing the top 20 associated SNPs.

Model	Quality Control	Total SNPs	Total SNPs Out HWE (X=1e-04)	*SNPs in Controls Out of HWE (X=1e-04)	*SNPs in Cases Out of HWE (X=1e-04)	**SNP with Low call rate (<95%)	**Low MAF(<1e-06%)	*Markers out of HWE (P <1e-08)	Excluded Animals	Lambda (λ)	Chr
Raw data		174375	4475	1375	3572	-	-	-	-		
1	QC1	134300	1484	152	867	2351	34644	2537	6		15,16,171 9,20
	QC2	124157	1113	0	824	0	9265	877	0	1.28	
2					Same as for model 1						1,2,3,6,8, 10,11,16, 21,22,29, 30,32
	QC2									1.12	
3					Same as for Model 1						15,16,17, 19,20
	QC2									1.04	
4					Same as for model 1						1,2,3,8, 10,21,22
	QC2									0.97	
5					Same as for model 1						15,16,17, 19
	QC2									1.11	
6					Same as for model 1						1,3,10,14, 19,22
	QC2									1.05	

*Number of SNPs out of HWE that were still kept in the data after the quality control.

** Number of SNPs eliminated in the quality control due to low call rate, low minor allele frequency and out of Hardy-Weinberg Equilibrium.

Model 3, which included population stratification, was considered the best λ value (1.04) according to the analysis criteria. Model 3 is the elected model for determining the associated markers. No marker had a significant Bonferroni p-value. The Q-Q plot and Manhattan plot of this model can be visualized in Figure 3, and its corresponding top 20 SNPs are shown in Table 8.

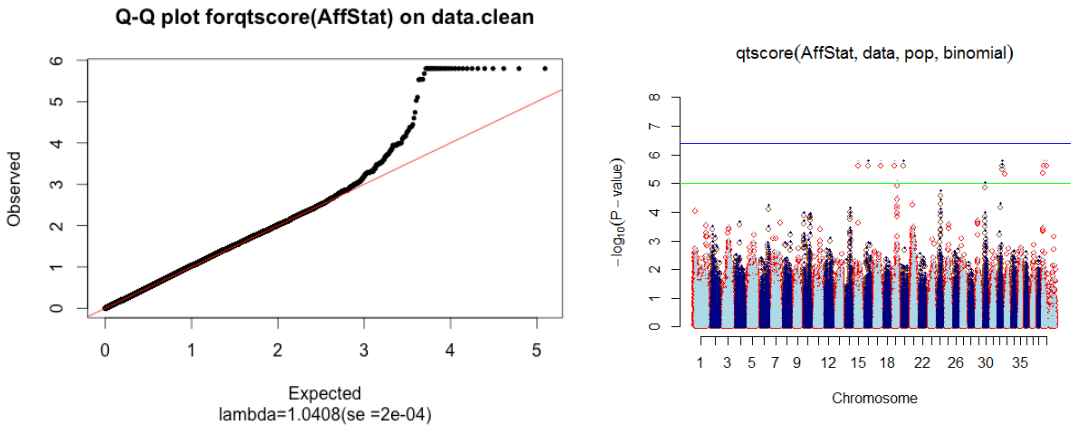

Figure 5. Q-Q plot and Manhattan plot for the statistical model number 3.
Blue line: Bonferroni corrected -log10(0.05/nsnps). Green line: -log10(1e-05).
Blue and light blue dots: raw p-value (P1df) Red dots: genomic control p-value (Pc1df)

Table 8. Top 20 SNPs according to the statistical model number 3.

Marker	Chr	Position	A1	A2	Pc1df*
BICF2P785060	15	32998454	T	C	2.54E-06
BICF2P1365875	16	36531600	G	A	2.54E-06
BICF2S23755960	17	67326820	T	C	2.54E-06
TIGRP2P261690	19	23061287	C	T	2.54E-06
BICF2S236384	19	23071429	G	A	2.54E-06
BICF2G63042752	19	23074950	G	A	2.54E-06
BICF2G63042757	19	23094737	T	C	2.54E-06
TIGRP2P261969	19	23123709	C	A	2.54E-06
TIGRP2P262022	19	23134751	T	G	2.54E-06
BICF2P15989	19	23153883	G	A	2.54E-06
BICF2P9773	19	23181081	A	C	2.54E-06
BICF2G63042843	19	23193716	C	T	2.54E-06
TIGRP2P262674	19	23265091	C	T	2.54E-06
BICF2G63042904	19	23281246	A	C	2.54E-06
TIGRP2P262866	19	23301862	G	A	2.54E-06
TIGRP2P262902	19	23312578	A	G	2.54E-06
BICF2P137634	19	23318694	A	G	2.54E-06
TIGRP2P262951	19	23323193	G	A	2.54E-06
BICF2P12405	19	23348192	A	G	2.54E-06
BICF2G630233713	20	24868338	A	G	2.54E-06

*P-values from the 1-d.f. test for association between SNP and trait; the statistics is corrected for possible inflation (genomic control)

The Q-Q plots and Manhattan plots for the alternative models can be visualized in the following Figures 6,7,8,9 and 10. The tables with the top markers can be found in the Appendix section.

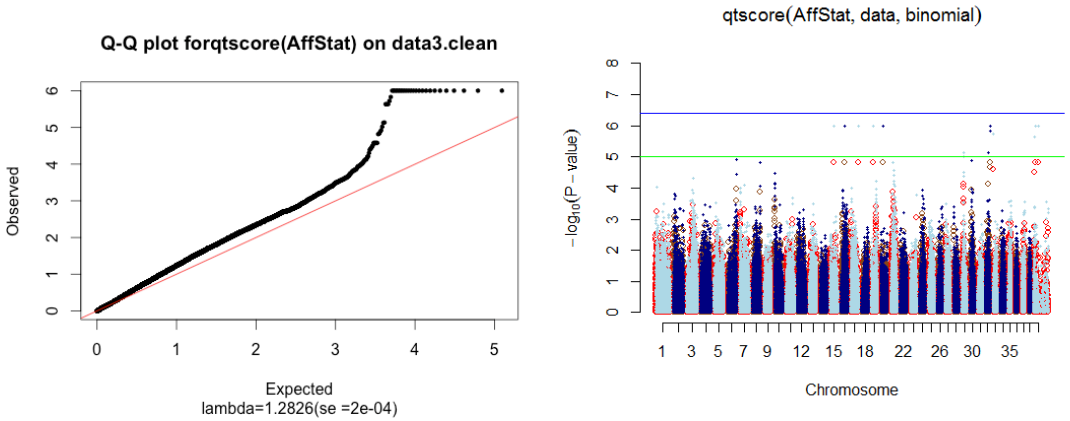

Figure 6. Q-Q plot and Manhattan plot for the statistical model number 1.
Blue line: Bonferroni corrected -log10(0.05/nsnps). Green line: -log10(1e-05).
Blue and light blue dots: raw p-value (P1df) Red dots: genomic control p-value (Pc1df)

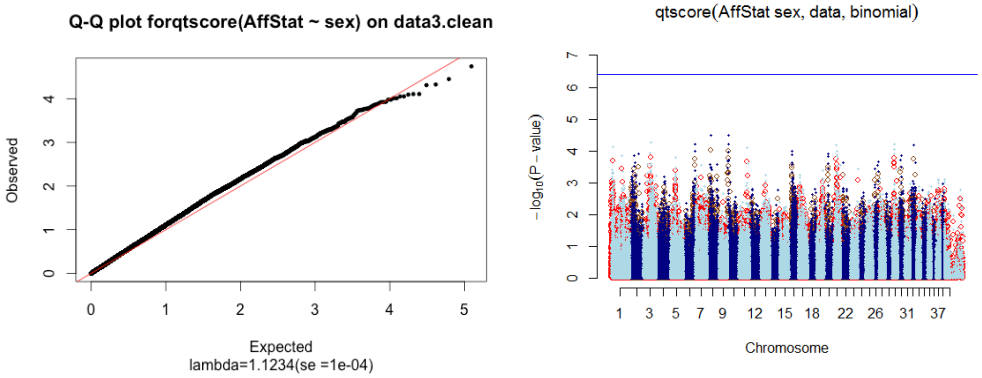

Figure 7. Q-Q plot and Manhattan plot for the statistical model number 2.
Blue line: Bonferroni corrected -log10(0.05/nsnps). Green line: -log10(1e-05).
Blue and light blue dots: raw p-value (P1df) Red dots: genomic control p-value (Pc1df)

* λ values <1 are rounded to 1 by default by the program.

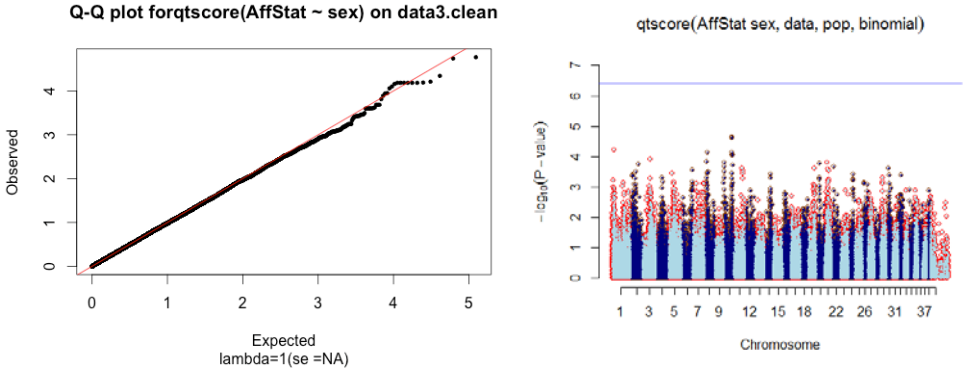

Figure 8. Q-Q plot and Manhattan plot for the statistical model number 4.
Blue line: Bonferroni corrected -log10(0.05/nsnps). Green line: -log10(1e-05).
Blue and light blue dots: raw p-value (P1df) Red dots: genomic control p-value (Pc1df)

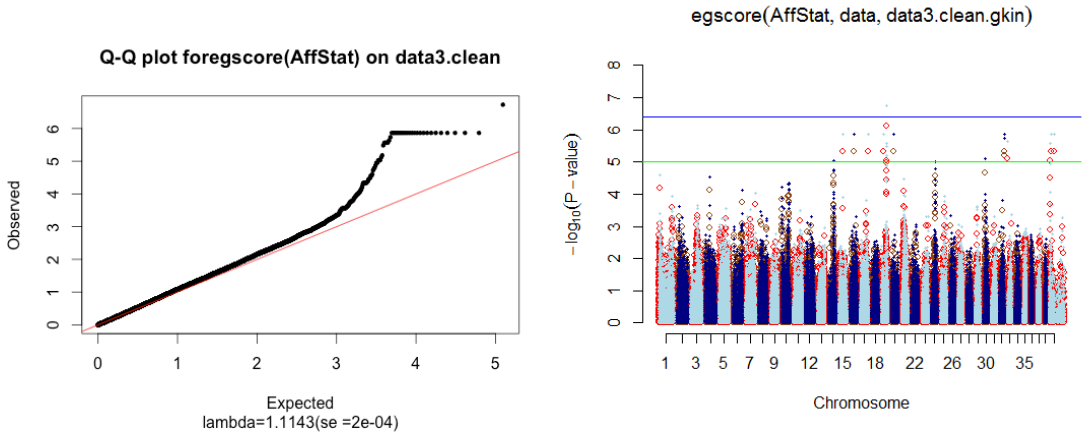

Figure 9. Q-Q plot and Manhattan plot for the statistical model number 5.
Blue line: Bonferroni corrected -log10(0.05/nsnps). Green line: -log10(1e-05).
Blue and light blue dots: raw p-value (P1df) Red dots: genomic control p-value (Pc1df)

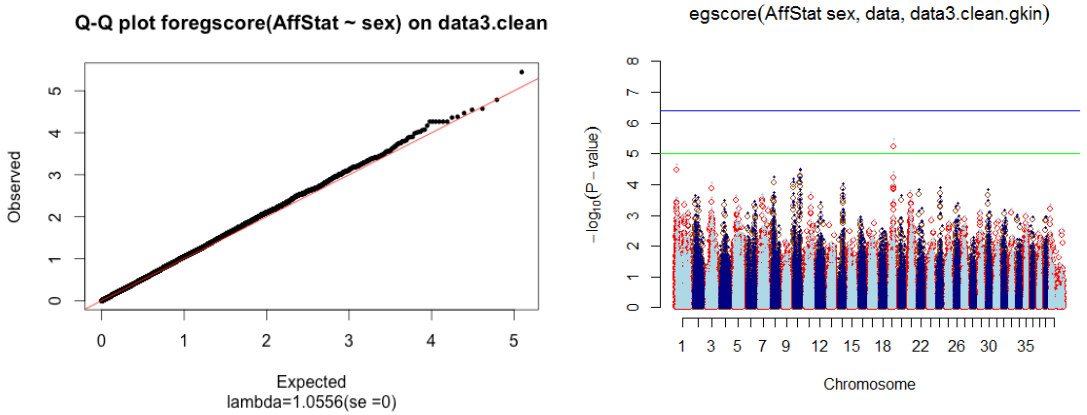

Figure 10. Q-Q plot and Manhattan plot for the statistical model number 6.
Blue line: Bonferroni corrected -log10(0.05/nsnps). Green line: -log10(1e-05).
Blue and light blue dots: raw p-value (P1df) Red dots: genomic control p-value (Pc1df)

The main objective of dataset 2 was to test if there were any associated markers on the X chromosome (chr 39). No Bonferroni significant markers were detected in any of the three tested models for dataset 2. This could be due to the insufficient number of individuals, which may have reduced the statistical power. In Table 9, information about the quality control of the dataset 2, λ values and chromosomes containing the top 20 SNPs for each value is included. The Q-Q and Manhattan plots and top markers are included in the appendix section.

Table 9. Quality control output for dataset 2, and λ values and Chromosomes containing the top 20 SNPs.

Model	Quality Control	Total SNPs	*Total SNPs Out of HWE (X=1e-04)	*SNPs in Controls Out of HWE (X=1e-04)	*SNPs in Cases out of HWE (X=1e-04)	**Low call rate (<95%)	**Low MAF (<1e-06%)	*Markersout of HWE (P <1e-08)	Excluded Animals	λ	Chr
1	Raw	174375	3596	2427	3211	-	-	-	-		1,3,8,10,11, 15,19,21,25 26,27,30,33
	QC1	130340	579	20	284	2333	38373	2837	0	1.07	
2	QC1									0.97	1,11,14,15, 19,25,29,30
3	QC1									1.06	1,3,11,14, 15,19,24,25 36

* Number of SNPs kept in the data after the quality control;

** Number of SNPs eliminated in the quality control due to low call rate, low minor allele frequency and out of Hardy-Weinberg Equilibrium.

The genetic association peak on chromosome 19 is significant (-log10(1e-05)) with most tested models. Two regions in chromosome 19 were also highlighted in the previous GWAS. The top SNPs from model 3 comprised a region between 23.0 Mb and 23.3 Mb which was also highlighted in the previous GWAS and contains two 5S ribosomal RNA and is the main associated region. In Figure 11 the distribution of the markers along chromosome 19 and their significance level can be observed. The peak between 23.0 Mb and 23.3 Mb is confirmed, and the region between 40 Mb and 60 Mb seem also to have an association peak.

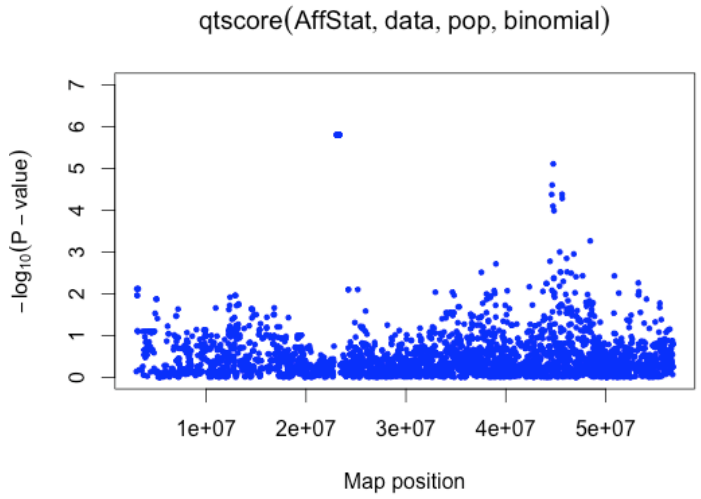

Figure 11. Marker distribution along chromosome 19

When taking a closer look as in Figure 12 and 13, it is observed that the associated markers are located around 23.0Mb and 23.3Mb and 44 Mb and 46 Mb.

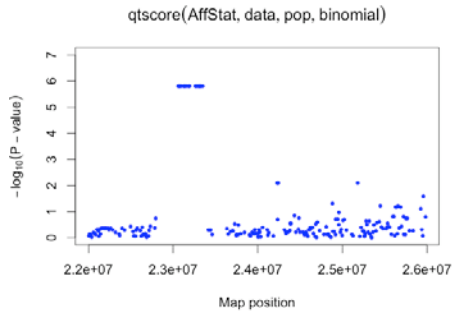

Figure 12. Marker distribution along chromosome 19 from 22.2Mb until 26.2Mb

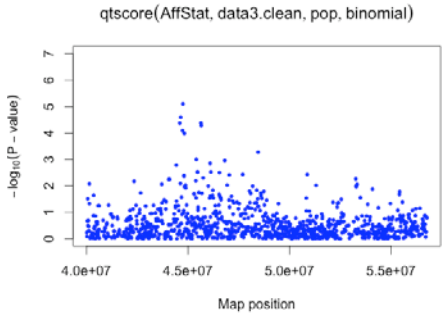

Figure 13. Marker distribution along chromosome 19 from 44Mb until 46Mb.

These same regions were detected in the previous GWAS and are strong candidate loci. There are three genes within 44Mb and 46 Mb: *SPOPL, NXPH2* and a novel transcript orthologue to *LRP1B* in humans. There is also a large intergenic region between *NXPH2* and *LRP1B,* which can potentially harbor regulatory elements.

3.2. Re-Sequence Analysis: Polymorphism detection:

In total, 73 polymorphisms with pattern score ≥3 were detected in the re-sequence dataset of five individuals (three controls and two cases). The Variant Effect Predictor tool of Ensembl was used to determine the consequence of each of these 73 SNPs (Table 10).

Table 10. Polymorphisms and consequences.

Chr	Synonymous	Non-synonymous	Intronic	Intergenic/ Downstream	3' UTR	Total
3	20	5	1	4	-	30
16	2	1	2	10	-	15
29	0	1	0	6	-	7
30	5	2	1	11	2	21
Total	27	9	4	31	2	73

From the 73 polymorphisms, six non-synonymous mutations were highlighted. The affected genes, mutation consequence, codon change, aminoacid change and exons where these mutations were found are listed in Table 11.

Table 11. Main polymorphisms

Chr	Gene	Mutation Consequence	Codon change	*Aminoacid change	Exon	Total nº of exons
3	*TMC3:* transmembrane channel-like 3	Non Synonymous Coding	cCt /cTt	Pro /Leu	Exon 21	22
3	*MESDC2:* mesoderm development candidate *2*	Non Synonymous Coding	Aac /Gac	Asn /Asp	Exon 1	2
16	*ZMAT4*: zinc finger, matrin-type 4	Non Synonymous Coding	aCc /aAc	Thr /Asn	Exon 3	6
29	Paralogue to *RPL10A* (self specie): ribosomal protein L10a	Non Synonymous Coding	tCt /tAt	Ser /Tyr	Exon 2	6
30	***DMXL2*: Dmx-like 2	Stop loss	tCg /tAg	Ser /Stop	Exon 12	43/14
30	***DMXL2*: Dmx-like 2	Non Synonymous Coding	aTa/aCa	Ile/Thr	Exon 10	43/14

*Aminoacid change codes. Leu: Leucine, Pro: Proline, Asp: Aspartic acid, Asn: Asparagine, Thr: Threonine, Tyr: Tyrosine, Ser: Serine, Ile: Isoleucine.
** *DMXL2* has two transcripts.

3.3. Sanger sequence

Sanger sequencing confirmed the polymorphisms in *TMC3* and *MESDC2* genes. The SNPs in *ZMAT4* and *DMXL2* stop loss mutation were not detected, and the primer pair did not cover the non-synonymous mutation in *DMXL2*. However, it is possible that the allele frequency for the non-reference allele is very low, and the re-sequence information supports the presence of these mutations with >3 reads (Appendix). *TMC3, MESDC2, DMXL2* and *ZMAT4* polymorphisms were genotyped with the TaqMan assay. However, only the non-synonymous coding SNP on *DMXL2* was successfully genotyped due to technical problems.

3.4. TaqMan Genotyping

332 individuals were genotyped for five different probe sets. The non-synonymous mutation in *DMXL2* gene was successfully genotyped in 319 animals. The allelic discrimination plot is represented on Figure 14. The allele frequencies for T and C are represented on Table 12.

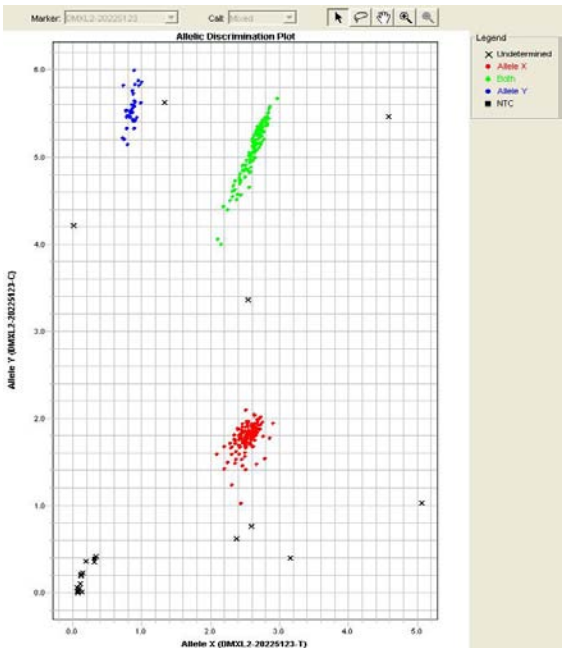

Figure 14. Allelic discrimination plot

Table 12. Allele frequencies for DMXL2.

	T/T	T/C	C/C	T	C	p	q	Total nº individuals
Verified Cases	24	15	4	63	23	0.73	0.26	**43**
Verified Controls	26	9	1	61	11	0.84	0.15	**36**
All Cases	40	30	7	110	44	0.73	0.27	**77**
All Controls	27	19	4	73	27	0.71	0.28	**50**
Population	142	128	49	412	226	0.64	0.35	**319**

Where T/T is the number of animals homozygous for the allele T, T/C heterozygous, C/C homozygous for the allele C. T the total number of the allele T for each group, C the total number of the allele C for each group, p represents the frequency (%) of the common allele T and q is the frequency (%) of the alternative allele C.

The alternative allele frequency C is slightly higher in cases as opposed to controls. However, a simple t-test when comparing allele frequencies between cases and controls shows no significance with a p-value=0.3866, and a Fisher's exact test (2x3) p-value= 0.258.

4. Discussion

Cases of dilated cardiomyopathy (DCM) in the present data are significantly more prevalent in males than females, with an odds ratio approximately eight male cases to each female case. This is consistent with the literature [6-7]. However, data bias cannot be completely ruled out. No associated markers peak in the X chromosome were detected by the GWAS. This could be either because the disease in Great Danes is in fact not X-linked, or due to low statistical power. It can also be that there are different types of inheritance within this breed, as is observed in human patients [7]. The fact that the study made by Meurs et al. used dogs from the USA [7], while in this study the great majority of the dogs are European is relevant. A retrospective study on the pedigree of the animals from this data could help to determine the mode of inheritance. In case the disease is not X-linked, the male prevalence could maybe be explained by physiological differences and epistatic interactions with genes or copy number variants in the Y chromosome. In humans, males are also prevalent in heart diseases. The human Y chromosome was associated with risk of coronary artery disease in men of European ancestry, possibly through interactions of immunity and inflammation [20].

In total, nine statistical models were approached in the GWAS. Model number 3 included population stratification and had an inflation factor lambda equal to 1.04 and showed association (genomic control p-value = 2.54E-06) on chromosomes 15,16,17,19 and 20. It was noticed that when including gender to the models the lambda is generally deflated. It could be because this is controlling for all the markers that differ between males and females, and since there are more affected males in our data, this could take away markers that were actually associated to the disease, however a closer look into the data must still be carried out to confirm this suspicion.
Also in model 4, when including population stratification together with gender, there is a slight overcorrection of the inflation factor λ=0.98. It is possible that the population clusters are already correcting for the gender differences, as can be observed in Table 6, as cluster 1 is dominated by female cases, cluster 2 by male cases and cluster two is gender neutral. In that way, applying structure to the model accounts for both the gender and genetic stratification of the data and by again including gender, corrects for this value twice. Different chromosomal peaks are observed on different models, and no bonferroni significant markers were present. In model 5 (Figure 9) there is one marker that is above the Bonferroni line, however that is the raw p-value (P1df). These factors could happen for different reasons, (1) there are not enough animals in the dataset for a good statistical power; (2) the phenotype information is not well explored; (3) there is data bias (not independent); (4) there are too many complex interactions such as epistasis; (5) confounding environmental factors, i.e; nutrition, common environment, physical activity, exposure to infectious diseases, stress, etc. (6) strict quality control leading to elimination of truly associated markers. Most of

these problems are difficult to solve without performing studies on environmental risk factors for DCM.

The SNP sequence technology is still costly and the amount of genotype data is dependent on available funds for the research. Furthermore, a well-described phenotype data is critical for a good GWAS. Unfortunately it is a big challenge, mostly because confirmatory diagnosis relies on post-mortem histopathological information, and this relevant data depends highly on the dog owner/breeding societies cooperation. The diagnosis can also vary according to different veterinarians, and clinical diagnosis mistakes can happen. For this study such mistakes are very unlikely, since the available phenotypic data was very well characterized, with postmortem histopathology information, echocardiogram and in depth veterinary records. However, most of this extra was not included in the models due to the complexity of the analysis. It would be interesting to include information such as histopathology status, intensity of the clinical signs, age of onset, originating country and breeding company/owner into the GWAS analysis. For an accurate association and good statistical power it is ideal to use unrelated (independent) and balanced data, for example, an equal number of unrelated female and male cases. Though the data will most likely be biased when a disease is predominant in males and affects dogs from a breed with considerably low genetic variation. The analysis of the GWAS of complex polygenic diseases needs to be adapted and corrected for the available data, and that is the reason why population stratification and a kinship coefficient matrix were used in the statistical models for this study. A closer look into kinship matrix and a retrospective pedigree study can be done to determine the rate of inbreeding of the animals used in the database.

Two quality controls were performed to remove markers with low call rate, low minor allele frequency and strongly deviating from HWE. The first control used the same criteria for both, cases and controls. The second quality control was stricter and eliminated all markers out of HWE in controls. Perhaps for this dataset the false discovery rate (FDR <0.2) in the control group should have not been applied. It could be that the risk allele is segregating lowly in controls, so the association signal would show up in controls instead of in cases. By eliminating all the markers out of HWE in controls, that signal is completely lost.

As mentioned before, when possible, the data should be improved by adding more individuals, and detailed and well-characterized trait information. However, the more information available, more challenging the analysis is. Large datasets with several variables require powerful computers and complex modeling. Statistical and computational knowledge is required to solve the genetic pathways involved in complex diseases. An interesting approach for a more complete GWAS for a polygenic trait would include algorithms that account for gene-gene interactions (epistasis). There are many efforts to include epistasis detection in GWAS using different methodologies as described by different authors, using from regression, machine learning and Bayesian approaches [21-27]. Most of these algorithms require substantial computing time in one trait, and an understanding of the differences between the models is important when

choosing which will be the best for a given data [22]. To date, very few publications have incorporated interaction testing in GWAS data, since it is a very new approach. The first exhaustive genome-wide epistasis search on a real dataset was performed by by Gayan et al., via their developed tool called "Hypothesis Free Clinical Cloning" (HFCC) on a cohort of Parkinson's disease patients and controls [28]. Another even more ambitious suggestion is the use of expert knowledge extracted from protein interaction databases. And even further, the use protein and gene expression data from these important structural/ion channel proteins as a covariate in the GWAS model, for example castalbin2 and triadin. This approach may facilitate the biological interpretation of the data, however there is still a need to develop a logical method to evaluate the information in these protein databases and also the metrics developed from this information in order to incorporate this type of expert knowledge into the analysis, which is another great challenge [29].

Even though there are all these complicating factors, the GWAS performed in this study was enough to show two possibly associated regions on chromosome 19. The regions between 23.0 and 23.3 Mb and between 44 and 46 Mb overlapped with the results from the previous GD GWAS. The region around 23 Mb contains two rRNAs, and the region between 44 and 46 Mb contains a large intergenic region and three genes: speckle-type POZ protein-like (*SPOPL*), neurexophilin 2 (*NXPH2)* and a novel transcript orthologue to the low density lipoprotein receptor-related protein 1B (*LRP1B*) in humans. None of these genes were previously associated to DCM or other cardiac diseases. The non-coding RNA may have interesting regulatory proprieties and must be further investigated (ref). *SPOPL* may be involved in ubiquitination and proteosomal degradation, and *NXPH2* may encode signaling molecules that resemble neuropeptides. The re-sequence data analysis on chromosome 19 included *SPOPL* and *NXPH2* genes, but no relevant SNPs were found within them. All the detected polymorphisms that opposed between cases and controls in chr 19 fell into intergenic regions. Intergenic regions can potentially harbor non-coding RNA (ncRNA), long non-coding RNA (lncRNA), cromatin markers, enhancers and copy number variants. ncRNAs controls the expression of target genes at the posttranscriptional level. It has been described as part of the "dark matter of the genome", due to the fact that although we are now able to detect its presence, its function and activity remains poorly understood. Cromatin markers regulate epigenetic factors; epigenetic modifications are heritable alterations of the genome, which can govern gene expression without altering the DNA sequence and can be relevant in complex diseases [30]. The *LRP1B* gene may encode potential cell surface proteins and was not included in the re-sequence analysis. Because of the very large intergenic region in this chromosome, it would also be relevant to run a coverage analysis in order to detect possible copy number variants. Further analysis in both candidate regions should be carried out, with the possibility to discover novel genes or regulatory elements that could participate in the pathogenesis. No other regions detected on the previous GWAS were re-associated. It may be because the previous analysis used smaller individual

database, equal quality control for cases and controls, and a different method for controlling population stratification.

So far many negative aspects of the GWAS were highlighted. But the great advantage of GWAS in dogs is that they possess a large extent of linkage disequilibrium (LD) within breeds (distances of several megabases) and between breeds there is a shorter LD (over tens of kilobases)[14]. These characteristics represent the two bottlenecks of the modern purebred dogs (domestication and breed creation), which led to a high genetic variation between breeds, but low variation within breeds. This means that a common disease in different breeds can be genetically traced in two stages: genome-wide mapping of a single breed and subsequently narrowing the detected regions by association with the different breeds [14]. Newfoundland dog phenotype and SNP typing data is available. No overlapping regions where found between Newfoundlands and Great Danes. However, the Newfoundland data could be used for a second GWAS following similar methodology as this study, perhaps in this way some regions may overlap. Other dog breeds could also be used.

Regarding the resequence data, non-synonymous polymorphisms within exons that presented allele segreation between cases and controls were detected in six genes. The candidate genes are: (chr3) transmembrane channel-like 3 (*TMC3*) and mesoderm development candidate 2 (*MESDC2*); (chr16) zinc finger, matrin-type 4 (*ZMAT4*); (chr29) a paralogue to RPL10A: ribosomal protein L10a; and (chr 30) Dmx-like 2 (*DMXL2*). The IGV visualization figures of the polymorphisms from three control animals (011,012 and 013) and two cases (022 and 023) are attached to the Appendix section.
The *TMC3* gene belongs to the TMC novel family, which encodes transmembrane proteins. In humans,*TMC3* mRNA is detected mostly in neuronal organs[31]. *MESDC2* encodes for an assisting protein in the beta folding of low-density lipoprotein receptors (LDLR) [32]. *ZMAT4* is a zinc ion binding protein; *RPL10A* encodes for a ribosomal protein that is component of the 60S subunit (L10a). However the detected polymorphism is in a *RPL10A* paralogue, and is important to consider that it is typical that multiple processed pseudogenes are dispersed through the genome for genes encoding ribosomal proteins [33]. *DMXL2*, or its alternate name Rabonnectin-3, is a functional regulator of the Notch signaling pathway in mammals. Notch signaling plays an important role in the development and differentiation of many cell types in diverse organisms. Abnormal activity of the Notch pathway has been linked to several developmental disorders and malignant diseases [34]. Furthermore, a SNP in *DMXL2* has been associated with ischemic stroke in afro-descendent humans [35]. Because of its function and two detrimental mutations within *DMXL2,* it is considered one of the main candidate genes of this study. The non-synonymous mutation was confirmed, however the mutation results in a stop loss in *DMXL2* could not be confirmed in the 332 animals genotype due to technical problems.

The non-synonymous polymorphism on *DMXL2* is located in exon 10, and changes an ATA codon to an ACA (isoleucine to threonine). The alternative allele C was slightly more frequent in cases than controls, however, this difference is not statistically significant with a p-value= 0.386. Even though it is not statistically significant, this could be because of technical mistakes during the genotyping. It can also be that some controls are false negatives (due to age of onset or other factors), or the phenotype is only expressed when other mutations are also present. Genetic risk penetrance should also be taken in consideration before completely discarding this gene as a candidate. Further evaluations should be made in order to check if this mutation has any participation in DCM pathology pathway whatsoever. A second attempt to genotype the candidate gene *TMC3*, which had been confirmed in the Sanger sequence of the 7 individuals should be made. In case a statistically significant mutation is finally detected, a gene expression study using the available tissues, as well as the functional study, pathways and gene interactions/regulation analysis should be carried on.

The detection of the genetic basis of multifactor and complex diseases is still a big challenge, even with the advent of high throughput sequencing technologies. There are many different ways of analyzing the large amount of data generated, and many variables need to be considered. The validation of the candidate genes is no less of a challenge. It is therefore important to intertwine knowledge of different specialized areas (clinical diagnosis, bioinformatics, statistics, biological pathways, among others) to achieve the goal of understanding the genetics of DCM.

5. Conclusion

The current GWAS methodology showed a genomic re-association of two regions on chromosome 19 (23.0-23.3Mb and 44-46Mb) to Great Dane DCM. No other regions were re-associated. Both regions may contain potential novel genes and regulatory factors that can possibly be involved in the GD DCM etiology pathway. From the Illumina re-sequencing data, six candidate mutations in five different genes were detected. A non-synonymous mutation in *DMXL2* was confirmed through the successful genotyping of 319 individuals. The alternative allele was slightly more frequent in cases than controls, however not statistically significant (p-value=0.38). Further investigation of the detected candidate mutations and regions, and the addition of extra phenotype information and consideration of gene-gene interactions in a GWAS should be carried out in future studies in order to determine the relevance of the present findings.

References:

1. Tidholm, A., and Jhonsson, L., *Histologic Characterization of Canine Dilated Cardiomyopathy.* Vet Pathol 2005(42): p. 1-8.
2. Tidholm, A., *Survival in dogs with dilated cardiomyopathy and congestive heart failure treated with digoxin, furosemide and propranolol: A retrospective study of 62 dogs.* Journal of veterinary cardiology: the official journal of the European Society of Veterinary Cardiology 8(1):41-7, 2006. **8**(1): p. 41-7.
3. Wiersma,A.C., Stabej,P., Leegwater,P.A.J, Van Oost, B.A., Ollier, W.E., and Dukes-McEwan, J., *Evaluation of 15 candidate Genes for Dilated cardiomyopathy in the Newfoundland dog.* Journal of Heredity 2008. **99**(1): p. 73-80.
4. Tidholm, A., *Canine Idiopathic Dilated Cardiomyopathy: Epidemiology, Histopathology and Pathophysiology.* Phd Thesis, 2000.
5. Dukes-McEwan, J., Borgarelli,M., Tidholm,A., Vollmar,A.C., Häggström,J. , *Proposed Guidelines for the Diagnosis of Canine Idiopathic Dilated Cardiomyopathy* Journal of Veterinary Cardiology, 2003. **Vol. 5**.
6. Broschk, C., and Distl, O., *Dilated cardiomyopathy (DCM) in dogs--pathological, clinical, diagnosis and genetic aspects.* Dtsch Tierarztl Wochenschr, 2005. **Oct;112(10):380-5**.
7. Meurs, K.M., Miller,M.W., and Wright,N.A., *Clinical features of dilated cardiomyopathy in Great Danes and results of a pedigree analysis: 17 cases (1990-2000).* JAVMA, 2001. **218**(5).
8. Parker, H.G., Meurs, K.M., Ostrander, E.A. *Finding cardiovascular disease genes in the dog.* J Vet Cardiol. , 2006. **8**(2): p. 115-27.
9. Towbin,J.A., and Solaro,J. *Genetics of Dilated Cardiomyopathy: More Genes that kill.* 2004. **44**(10).
10. Ghosh, N., and Haddad,H. *Recent progress in the genetics of cardiomyopathy and its role in the clinical evaluation of patients with cardiomyopathy.* Current Opinion in Cardiology, 2011(26): p. 155-164.
11. Oyama,M.A., Chittur,S.V., and Reynolds,C.A. *Decreased Triadin and Increased Calstabin2 Expression in Great Danes with Dilated Cardiomyopathy.* J Vet Intern Med 2009(23): p. 1014-1019.
12. Oyama, M.A., and Chittur, S. , *Genomic expression patterns of cardiac tissues from dogs with dilated cardiomyopathy.* Am J Vet Res, 2005. **66**(7): p. 1140–55.
13. Roberts, R., Wells,G.A., Stewart,A.F.R., Dandona,S., Chen,L., *The Genome-Wide Association Study—A New Era for Common Polygenic Disorders.* J Cardiovasc Transl Res, 2010. **3**(3): p. 173-82.
14. Lindblad-Toh,K.,Wade,C.M., Mikkelsen,T.S., Karlsson,E.K., Jaffe,D.B., Kamal,M., Clamp,M.,Chang,J.L.,KulbokasIII,E.J.,Zody,M.C.,Mauceli,E.,Xie,X.,Breen, M.,Wayne,R.K.,Ostrander,E.A.,Ponting,C.P.,Galibert,F.,Smith,D.R.,deJong,P.J., Kirkness,E.,Alvarez,P., Biagi,T., Brockman,W., Butler,J., Chin,C., Cook,A., Cuff,J., Daly,M.J.,DeCaprio,D.,Gnerre,S.,Grabherr,M.,Kellis,M.,Kleber,M., Bardeleben,C.,

Goodstadt,L., Heger,A., Hitte,C., Kim,L., Koepfli,K.P., Parker,H.G., Pollinger,J.P., Searle,S.M.J.,Sutter,N.B.,Thomas,R.,Webber,C.,Broad Institute Genome Sequencing Platform & EricS. Lander *Genome sequence, comparative analysis and haplotype structure of the domestic dog. .* Nature, 2005. **438**: p. 803-819.

15. Wayne, R.K., and Ostrander, E.A., *Lessons Learned from the dog Genome* Trends in Genetics, 2007. **23**(11): p. 557–567.
16. Aulchenko, Y., *GenABEL tutorial.* 2011.
17. Kierczak,M., Shen,X.,Tengvall,K.,Aulchenko,Y., and Carlborg,Ö. *GWAS Analysis in Stratified Populations- LUPA Workshop.* 2011.
18. Thorvaldsdottir, H., Robinson,J.T., and Mesirov,J.P. *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. .* Briefings in Bioinformatics 2012.
19. Truvé,K., Eriksson,O., Norling,M., Wilbe,M., Mauceli,E., Lindblad-Toh,K., and Bongcam-Rudloff,E., *SEQscoring: a tool to facilitate the interpretation of data generated with next generation sequencing technologies.* **EMBnet.journal 17.1**.
20. Charchar,F.J., Bloomer,L.D.S., Barnes ,T.A.,Cowley,M.J., Nelson,C.P., Wang, Y., Denniff,M., Debiec,R., Christofidou,P., Nankervis,S., Dominiczak,A.F., Bani-Mustafa,A., Balmforth,A.J., Hall,A.S., Erdmann,J., Cambien,F., Deloukas,P., Hengstenberg,C., Packard,C., Schunkert,H., Ouwehand,W.H., Ford,I., Goodall,A.H., Jobling,M.A., Samani,N.J., and Tomaszewskib, M., *Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome.* Lancet., 2012. **379**(9819): p. 915-922.
21. Gyenesei,A., Moody,J., Semple, C.A.M., Haley, C.S., and Wei,W. *High throughput analysis of epistasis in genome-wide association studies with BiForce.* Bioinformatics Advance Access, 2012.
22. Cordell, H., *Detecting gene-gene interactions that underlie human diseases.* Nat Rev Genet., 2009. **10**(6): p. 392-404.
23. Cattaert,T., Urrea,V., Naj,A.C., De Lobel,L. ,De Wit,V., Fu,M., Jestinah, M., Mahachie, J.,Shen,H., Calle,M.L., Ritchie, M.D.,Edwards,T.L., Van Steen,K., *FAM-MDR: A Flexible Family-Based Multifactor Dimensionality Reduction Technique to Detect Epistasis Using Related Individuals.* Plos One, 2010. **5**(4).
24. Liu, Y., et al, *Genome-Wide Interaction-Based Association Analysis Identified Multiple New Susceptibility Loci for Common Diseases.* Plos Genet, 2011. **7**.
25. Schupbach, T., et al., *FastEpistasis: a high performance computing solution for quantitative trait epistasis.* Bioinformatics Advance Access, 2010. **26**: p. 1468-1469.
26. Wan, X., et al. , *BOOST: A Fast Approach to Detecting Gene-Gene Interactions in Genome-wide Case-Control Studies, American journal of human genetics.* American journal of human genetics, 2010. **87**(325-340).
27. Tang. W., et al., *Epistatic module detection for case-control studies: a Bayesian model with a Gibbs sampling strategy,.* PLos Genet. **5**.
28. Gayán,J.,González-Pérez,A., Bermudo,F., Sáez,M.E., Royo,J.L., Quintas,A., Galan,J.J., Morón,F.J., Ramirez-Lorca,R., Real,L.M., and Ruiz,A., *A method for detecting epistasis in genome-wide studies using case-control multi-locus association analysis.* BMC Genomics, 2008. **9** (360).

29. Pattin, K.A., and Moore,J.H., *Exploiting the proteome to improve the genome-wide genetic analysis of epistasis in common human diseases.* Hum Genet 2008(124): p. 19–29.
30. Baccarelli, A., and Ghosh, S., *Environmental exposures, epigenetics and cardiovascular disease.* Curr Opin Clin Nutr Metab Care, 2012. **15**(4): p. 323-9.
31. Keresztes,G., Mutai,H., and Heller,S., *TMC and EVER genes belong to a larger novel family, the TMC gene family encoding transmembrane proteins.* BMC Genomics 2003: p. 4:24.
32. Hsieh, J.C., Lee, L., Zhang, L., Wefer, S., Brown, K., DeRossi, C., Wines, M. E., Rosenquist, T., Holdener, B. C., *Mesd encodes the LRP5/6 chaperone essential for specification of mouse embryonic polarity.* Cell 2003(112): p. 355-367.
33. Olvera,J., and Wool,I.G., *The Primary Structure of Rat Ribosomal Protein L10a.* Biochemical and Biophysical Research Communications, 1996. **220** (3): p. 954–957.
34. Sethi,N., Yan,Y., Quek,D., Schupbach,T., and Kang,Y., *Rabconnectin-3 Is a Functional Regulator of Mammalian Notch Signaling.* J Biol Chem, 2010. **285**(45): p. 34757–34764.
35. Luke, M.M., O'Meara, E.S., Rowland, C.M., Shiffman, D., Bare, L.A., Arellano, A.R., Longstreth, W.T. Jr, Lumley, T., Rice, K., Tracy, R.P., Devlin, J.J., and Psaty, B.M. *Gene variants associated with ischemic stroke: the cardiovascular health study.* Stroke, 2009. **40**(2): p. 363-8.