A Multi-Breed Genome-Wide Association Analysis for Canine Hypothyroidism Identifies a Shared Major Risk Locus on CFA12

Matteo Bianchi[1☯], Stina Dahlgren[2☯], Jonathan Massey[3], Elisabeth Dietschi[4], Marcin Kierczak[1], Martine Lund-Ziener[2], Katarina Sundberg[5], Stein Istre Thoresen[2], Olle Kämpe[6], Göran Andersson[5], William E. R. Ollier[3], Åke Hedhammar[7], Tosso Leeb[4], Kerstin Lindblad-Toh[1,8], Lorna J. Kennedy[3‡], Frode Lingaas[2‡], Gerli Rosengren Pielberg[1‡]*

1 Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden, 2 Department of Basic Sciences and Aquatic Medicine, Norwegian University of Life Sciences, Oslo, Norway, 3 Centre for Integrated Genomic Medical Research, The University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom, 4 Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland, 5 Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, 6 Department of Medicine (Solna), Karolinska Institutet, Stockholm, Sweden, 7 Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden, 8 Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* gerli.pielberg@imbim.uu.se

Abstract

Hypothyroidism is a complex clinical condition found in both humans and dogs, thought to be caused by a combination of genetic and environmental factors. In this study we present a multi-breed analysis of predisposing genetic risk factors for hypothyroidism in dogs using three high-risk breeds—the Gordon Setter, Hovawart and the Rhodesian Ridgeback. Using a genome-wide association approach and meta-analysis, we identified a major hypothyroidism risk locus shared by these breeds on chromosome 12 ($p = 2.1 \times 10^{-11}$). Further characterisation of the candidate region revealed a shared ~167 kb risk haplotype (4,915,018–5,081,823 bp), tagged by two SNPs in almost complete linkage disequilibrium. This breed-shared risk haplotype includes three genes (*LHFPL5*, *SRPK1* and *SLC26A8*) and does not extend to the dog leukocyte antigen (DLA) class II gene cluster located in the vicinity. These three genes have not been identified as candidate genes for hypothyroid disease previously, but have functions that could potentially contribute to the development of the disease. Our results implicate the potential involvement of novel genes and pathways for the development of canine hypothyroidism, raising new possibilities for screening, breeding programmes and treatments in dogs. This study may also contribute to our understanding of the genetic etiology of human hypothyroid disease, which is one of the most common endocrine disorders in humans.

Introduction

The domestic dog occupies an important position as a companion animal for humans, however for research scientists it also provides a unique comparative resource for genetic studies of phenotypic variation and development of diseases comparable to those in humans [1–4]. The genomic structure of domestic dogs, formed through domestication and intense breed-creation, is highly amenable to the identification of causal genetic loci in the same way as studies performed using other domestic animals [5–7]. Dogs share environmental factors with humans to a greater extent than any other domestic species, making them particularly suitable for comparative studies of complex diseases. The canine genome has been sequenced [5] and subsequently molecular tools for large-scale genotyping and analysis have been developed [8, 9]; large datasets describing gene expression are also publicly available [10, 11].

One of the most frequent endocrine diseases affecting both humans and dogs is hypothyroidism, a disorder in which the thyroid gland fails to produce sufficient amounts of thyroid hormones [12, 13]. Symptoms in human as well as in dogs are generally non-specific, including tiredness, weight gain and poor ability to tolerate cold, reflecting the key function of thyroid hormones in regulating the metabolism of the body [14, 15]. A common cause of human hypothyroidism worldwide is insufficient levels of dietary iodine [16]. However, the autoimmune disease called Hashimoto's thyroiditis (HT) is also a common cause of hypothyroidism, especially in countries with sufficient daily intake of iodine [17]. The equivalent of HT in dogs is called canine lymphocytic thyroiditis (CLT). CLT is characterised by infiltration of B and T lymphocytes into the thyroid gland and the presence of circulating auto-antibodies against thyroglobulin (TgAA), resulting in the progressive destruction of the thyroid [18–21]. Hypothyroidism in dogs may also be caused by thyroid idiopathic atrophy, characterised by a degenerative rather than autoimmune process, thought to potentially represent an end-stage of CLT [22]. The observation that some purebred dog breeds (primarily medium to large size breeds) are more predisposed to develop hypothyroidism suggests an underlying genetic risk for developing the disease. Breeds with high risk to develop hypothyroidism include the English Setter, Rhodesian Ridgeback, Giant Schnauzer, Hovawart, Old English Sheepdog, Boxer, Doberman Pinscher, Gordon Setter and the Beagle [18, 23–27]. The disease also shows notable familial clustering, further supporting the presence of a hereditary component [28]. Identifying the underlying genetic etiology of hypothyroidism in both humans and dogs is of major interest. Associations between major histocompatibility complex class II gene polymorphisms and hypothyroidism have previously been revealed in both humans (HLA) [29, 30] and dogs (DLA) [31, 32]. However, the reported DLA allele associations with CLT do not fully explain the susceptibility seen in all dog breeds. Furthermore, the contribution of DLA risk alleles to the development of hypothyroidism in specific breeds only explains a proportion of the underlying genetic risk for CLT. This indicates the presence of additional genetic risk factors and thus suggests a complex nature for the disease [22]. Congenital hypothyroidism in humans is associated with variants in genes including *DUOX2*, *PAX8*, *SLC5A5*, *TG*, *TPO*, *TSHB* and *TSHR* [33]. Genetic variants in the *TPO* gene are also associated with this rare type of hypothyroidism in Toy Fox and Tenterfield Terriers [34, 35].

Here we present a primary genome-wide association (GWA) study of hypothyroidism in three high-risk dog breeds, designed to identify novel susceptibility genes of the disease. This is the first time a major risk locus contributing to a complex disease has been identified across several dog breeds and confirmed as shared by using a meta-analysis approach.

Materials and Methods

Collection of samples and DNA preparation

EDTA blood and serum samples were collected from three hypothyroidism high-risk breeds. These breeds were Gordon Setter (GS) from Norway, Hovawart (HV) from several European countries, and Rhodesian Ridgeback (RR) from the United States of America (S1 Table). Preliminary characterisation of RR cohort is presented by Massey J.P. [36].

Samples were collected at veterinary clinics after obtaining owner's written consent. Sampling routines appropriate for each of the different countries were followed (Norway: FOR-2010-08-06-1147; Sweden: Swedish Animal Ethical Committee No. C139/9 and C2/12 and Swedish Animal Welfare Agency No. 31-4714/09 and 31-998/12; Switzerland: Canton of Bern No. 23/10; United Kingdom and United States of America: ethical permit not needed for leftover samples originally collected for veterinarian purposes). Genomic DNA was extracted from the EDTA blood samples using QIAamp DNA Blood Mini/Midi extraction kit, QiaSymphony DNA midikit (both from Qiagen, Hilden, Germany) or E.Z.N.A blood DNA kit (Omega Bio-Tek, Norcross, GA, USA) following the manufacturers' recommendations, and subsequently stored at -20°C. Serum was separated from clotted blood by centrifugation and stored at -20°C.

Phenotyping

Dogs were initially classified as cases or controls based on clinical diagnoses from expert veterinarians certified in veterinary internal medicine. The clinical diagnoses of cases were subsequently supported by thyroid typical serological measurements (thyroid-stimulating hormone (TSH) and free thyroxine (fT4)) using Siemens IMMULITE Immunoassay System [31, 37]. In order to be classified as cases, dogs had to present with increased levels of TSH (> 40 mU/L) and reduced levels of fT4 (< 7 pmol/L), whereas controls had to be older than seven years of age. A subsequent careful survey of clinical records and/or questionnaires completed by dog owners was performed to exclude cases with another condition potentially influencing the dog's wellbeing as well as controls with a history of immune-mediated conditions.

SNP genotyping and quality control

The initial genetic datasets related to 165 GS, 74 HV and 92 RR individuals (S1 Table), and were genotyped using the Illumina 170 k CanineHD BeadChip (Illumina, San Diego, CA, USA) at the same technology platform. All SNP-positions are given according to the dog Can-Fam3.1 assembly [11].

Genotyping data quality control (QC) was carried out for each breed separately, using R v3.0.2 [38] and GenABEL v1.8–0 [39]. Firstly, an individual-based QC step was performed to identify potential duplicated samples and samples with gender discrepancies. Secondly, a marker-based QC was performed, including: pruning of the total set of SNPs according to minor allele frequency thresholds (MAF) (0.05 for all breeds), SNP and individual call rates (95% for all breeds), p-values ($1\text{x}10^{-5}$ in GS, $1\text{x}10^{-8}$ in HV, $1\text{x}10^{-3}$ in RR) and false discovery rates for Hardy-Weinberg equilibrium (0.2 in all breeds). Moreover, each breed dataset was also checked for correlation between disease status and gender distribution. Fisher's exact test and a phi coefficient were used to evaluate statistical significance and magnitude of correlation between such dichotomous variables (*i.e.* disease status and gender) [40, 41].

Genome-wide association (GWA) analysis

A GWA analysis was performed on the quality controlled SNP datasets for GS, HV and RR breeds separately. All analytical steps were carried out using R v3.0.2 [38] and GenABEL v1.8–

0 [39]. Using 2,000 randomly selected autosomal markers a genomic kinship matrix weighted by allele frequencies was computed in every breed. In all the breeds, we applied a standard linear mixed model, which was fitted using the polygenic_hglm function from the hglm package ver 2.0–8 [42], including the genomic kinship matrix as random effect. The mixed model approach is able to deal with both population structure and relatedness [43]. Breed-specific genomic kinship matrices were also used to project genetic distance between individuals into a plane using multidimensional scaling (MDS) and for subsequent plotting. For HV population, where samples had different geographic origins, we wanted to test whether this could have introduced any structure into the population. For this purpose, we followed an approach suggested by Tengvall and colleagues [3]. Shortly, we used K-means clustering to assign individuals to a predefined number of subpopulations. The number of clusters K = 2 (here subpopulations) was determined using a so-called scree test on a within-cluster sum of squares in a function of K (for details see [3]). Next, we used a linear mixed model with population as a fixed effect and genomic kinship as random effect. The statistical significance thresholds were evaluated as follows: (a) empirical genome-wide significance levels (P_{genome}) obtained after 1000 permutations of the mixed model residuals (residualY returned by polygenic function in GenABEL) and (b) 95% empirical SNP distributions confidence intervals (CI_{95}) as proposed by Karlsson and colleagues [1]. By permuting mixed model residuals, we maintained the connection between phenotypes and fixed effects [44], thus being able to evaluate the significance of only the genetic effects. For each single-breed GWA study, a quantile-quantile (QQ) plot was produced in R v3.0.2 and a Manhattan plot was generated using the R package qqman [45]. The independence of the signal was verified by association analysis conditioned on the genotype of the most significantly associated SNP (top SNP) for each breed separately.

Breed-specific associated loci were defined based on pairwise linkage disequilibrium (LD) estimates ($R^2 \geq 0.7$) of the three breed-specific top SNPs to SNPs in CFA12.

Meta-analysis of genome-wide association

GWA meta-analysis of the three independent datasets (breeds) was carried out using MetABEL v0.2.0 [39], a part of the R statistical suite v3.0.2 [38]. Assuming the associated shared allelic effect being the same in each dataset, MetABEL performs a fixed effects meta-analysis, where each study is weighted according to the inverse of its' squared standard error in order to maximise the power of discovery [46]. We created an MDS plot, displaying the samples belonging to the three different breeds as subpopulations, a QQ plot, showing the degree of deviation of the associated SNPs compared to their null distribution, and a Manhattan plot, showing the genome-wide association signals, as described above.

Haplotype analysis

The minimal risk haplotype, shared across breeds, was identified in the associated locus from the meta-analysis. Firstly, genotypes of the shared associated region were imputed (if missing) and phased into haplotypes in each breed separately using fastPHASE [47]. At this stage the phenotype of each individual was used as a covariate, in order to avoid prediction of spurious haplotypes. Thereafter, the risk haplotypes present in cases and non-risk haplotypes present in controls were identified based on the genotype at the meta-analysis top SNP. Starting from the meta-analysis top SNP and walking both up- and downstream, we then identified the SNP-positions where the risk haplotype was broken by a recombination event (*i.e.* two alternative alleles were present on both risk and non-risk haplotypes). This was done separately for each breed, and thereafter the minimal shared risk haplotype across breeds defined.

Two SNPs tagging the associated risk haplotype across breeds were analysed for association with the phenotype as both haplotypes and genotypes using Pearson's Chi-squared and Fisher's exact tests respectively [40, 48].

Results

GWA study sample cohort

We performed genotyping of 165 GS, 74 HV and 92 RR individuals for ~170,000 SNP markers distributed across the entire canine genome. The individual-based QC indicated 1 HV and 18 GS samples due to gender discrepancies between phenotypic and genetic data, retaining 147 GS (63 cases and 84 controls), 73 HV (44 cases and 29 controls) and 92 RR (38 cases and 54 controls) individuals for further analysis (S1 Table).

The marker-based QC removed the majority of SNPs due to MAF < 0.05, whereas SNPs were also removed due to call rate < 95% and deviation from the Hardy-Weinberg equilibrium (S2 Table). The final datasets consisted of 110,221 markers in GS, 100,864 in HV and 103,612 in RR cohorts.

We did not identify a significant correlation between phenotype and gender distribution in any of the breeds (p = 0.9, phi coefficient = 0.02 in GS; p = 0.1, phi coefficient = 0.19 in HV and p = 0.39, phi coefficient = 0.11 in RR).

Single-breed GWA studies show association to CFA12 in all breeds

Association analysis for hypothyroidism was performed in the three breeds separately. Population stratification was not detected in any of the breeds, as confirmed by the even distribution of cases and controls on the MDS plots (Fig 1a–1c). After the application of the mixed model approach, to account for population structure and cryptic relatedness common in dog breeds, no genetic inflation was observed in any of the breed-specific association tests ($\lambda = 1.019$ in GS, $\lambda = 0.995$ in HV and $\lambda = 0.996$ in RR) as presented on the QQ plots (Fig 1d–1f). The QQ plots also indicate the breed-specific statistical significance levels (see materials and methods).

We detected a significant genetic association in the GS and HV and a suggestive association in the RR to a region on CFA12 (Figs 1 and 2a–2c) with the following breed-specific top SNPs: GS 4,456,564 bp ($p_{raw} = 2.4x10^{-8}$), HV 5,158,474 bp ($p_{raw} = 4.2x10^{-6}$) and RR 9,336,752 bp ($p_{raw} = 2.0x10^{-5}$) (Table 1). We found the most significant association in the GS breed cohort, reaching a Bonferroni corrected alpha = 5% significance level ($p < 4.5x10^{-7}$), which is in any case considered excessively stringent in dog GWA studies. In the HV breed the association significance exceeded both empirical genome-wide and empirical CI_{95} levels, whereas in the RR breed the association was suggestive towards the CI_{95} empirical cut-off.

The GS and HV breed-specific top SNPs were present in the GWA analysis of all breeds, whereas the RR breed-specific top SNP was removed from GS and HV populations in marker-based QC process due to MAF < 0.05 (in GS population: MAF = 0.030, $MAF_{Cases} = 0.047$, $MAF_{Controls} = 0.018$; in HV population: MAF = 0.027, $MAF_{Cases} = 0.034$, $MAF_{Controls} = 0.017$) (S3 Table).

Breed-specific associated regions overlap

To define the location of associated regions on CFA12 in each of the three dog breeds, we performed an LD-analysis based on R^2-values. By including the breed-specific SNP sets used for association analysis we determined the R^2-value of each SNP on CFA12 to the respective breed-specific top SNP. Using a cutoff of $R^2 \geq 0.7$ as a signal of LD we defined the associated regions on CFA12 as following: GS 3,022,017–9,637,877 bp, HV 4,110,982–6,635,695 bp and

MDS plots
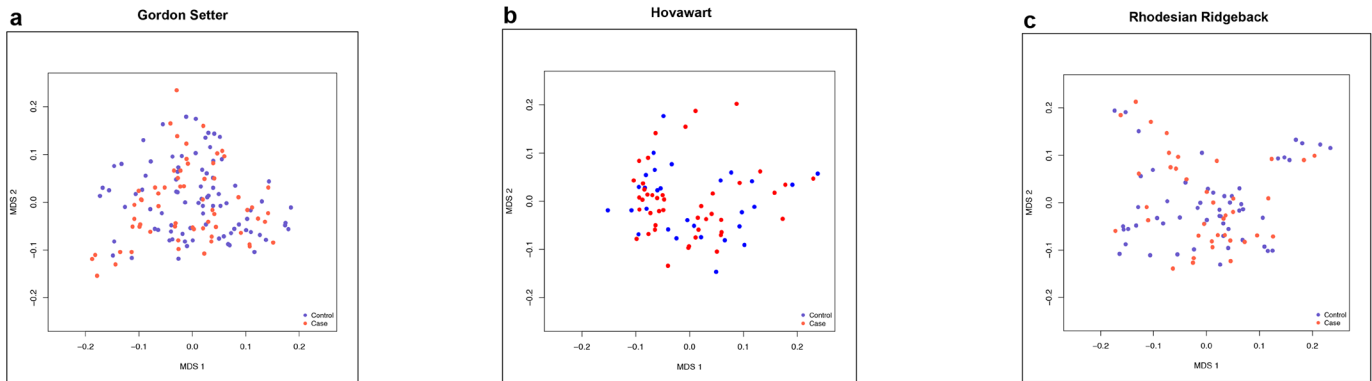
QQ plots
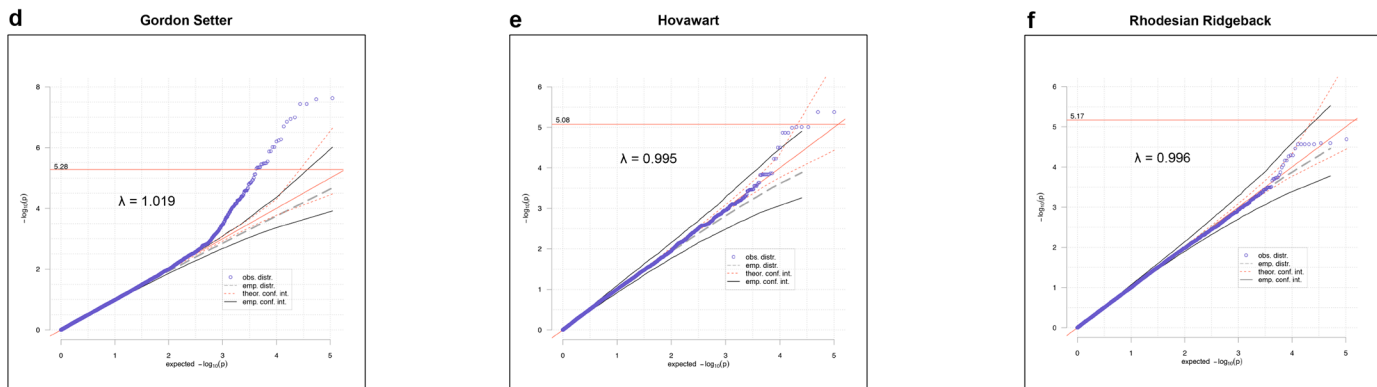
Fig 1. (a-c) MDS plots showing homogenous sample sets without presence of clustered samples in the (a) Gordon Setter (GS), (b) Hovawart (HV) and (c) Rhodesian Ridgeback (RR) breed cohorts. (d-f) QQ plots showing no inflation after the mixed model approach, and the empirical genome-wide significance thresholds (indicated by red lines and their corresponding values) and empirical 95% confidence intervals (CI_{95}) in (d) GS λ = 1.019, (e) HV λ = 0.995 and (f) RR λ = 0.996.

RR 3,561,421–9,336,752 bp (Fig 3). The long associated regions in each breed (~2.5–6.6 Mb) were confirmed as single signals by conditional association analysis (S1 Fig). The shared associated region was defined as 4,110,982–6,635,695 bp (~2.5 Mb), corresponding to the associated region defined for the HV sample cohort (Fig 3).

Meta-analysis confirms one major risk locus shared in all three breeds

Meta-analysis was performed across breeds (MDS plot Fig 4a), in order to identify the shared associated region and the top SNP. No genetic inflation was observed (λ = 0.977, QQ plot Fig 4b) and the most significant association (P_{raw} = 2.1x10^{-11}) was detected at CFA12: 5,039,806 bp, located in the shared associated region identified above from the breed-specific GWA analyses (Fig 2d). Therefore, CFA12: 4,110,982–6,635,695 bp represents a candidate region for increased susceptibility to canine hypothyroidism in the GS, HV and the RR breeds. This

a

Gordon Setter

b

Hovawart

c

Rhodesian Ridgeback

d
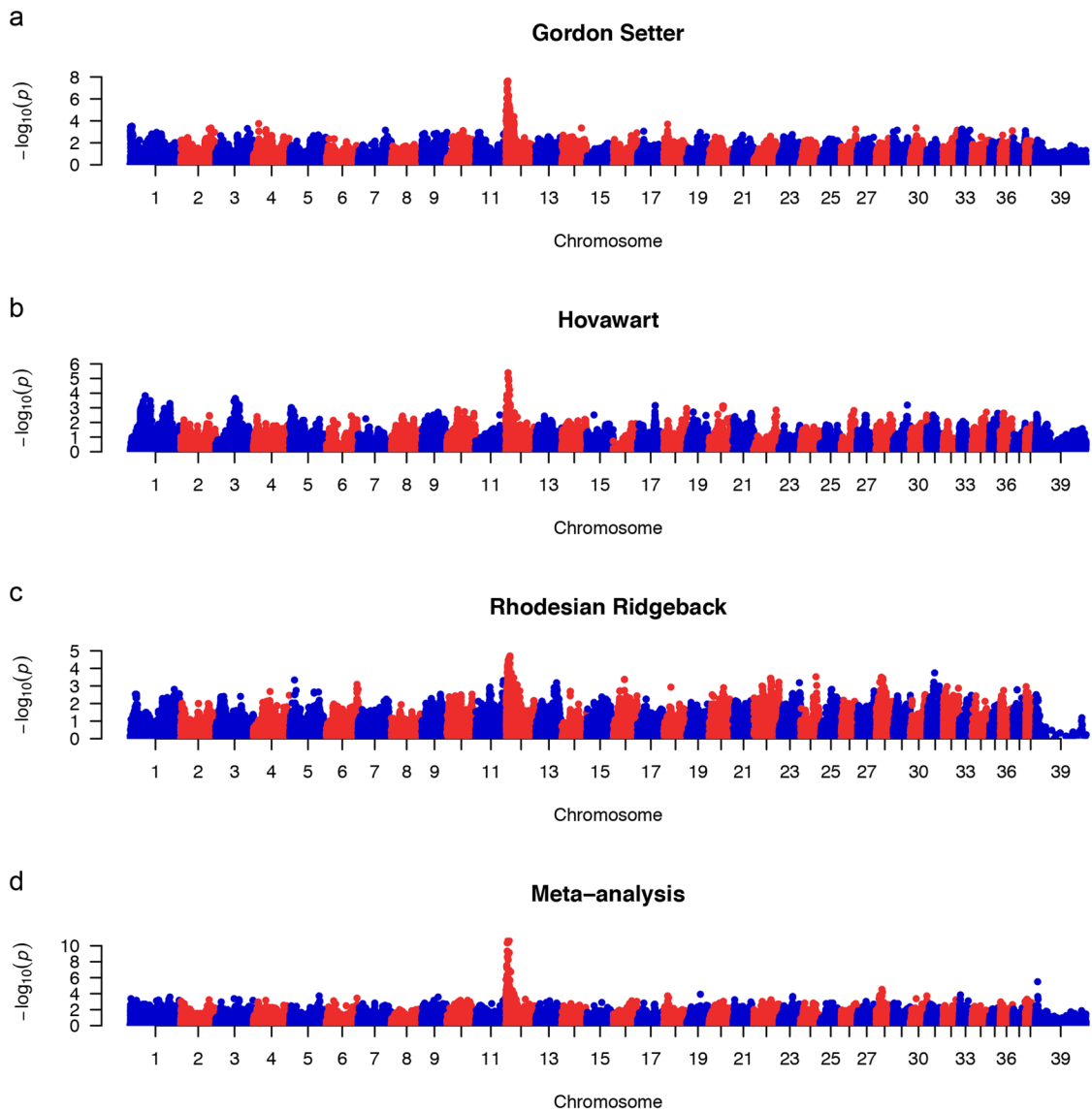
Meta–analysis

Fig 2. Manhattan plots showing a common peak on CFA12 in (a) GS $p_{raw} = 2.4 \times 10^{-8}$, (b) HV $p_{raw} = 4.2 \times 10^{-6}$, (c) RR $p_{raw} = 2.0 \times 10^{-5}$ and (d) multi-breed meta-analysis $p_{raw} = 2.1 \times 10^{-11}$.

doi:10.1371/journal.pone.0134720.g002

Table 1. Summary of the three single-breed GWA analysis results.

	GWA top SNP (bp)	P_{raw}	OR	CI_{OR}	P_{genome}	CI_{95}
Gordon Setter	4,456,564	2.4×10^{-8}	4.0	2.1–7.0	$< 5.2 \times 10^{-6}$	$< 2.0 \times 10^{-3}$
Hovawart	5,158,474	4.2×10^{-6}	5.0	2.1–13.1	$< 8.3 \times 10^{-6}$	$< 3.2 \times 10^{-5}$
Rhodesian Ridgeback	9,336,752	2.0×10^{-5}	3.4	1.8–6.7	$< 6.8 \times 10^{-6}$	NA

P_{raw}—uncorrected p-value for the strongest association, OR—Odds ratio, CI_{OR}—OR 95% Confidence Intervals, P_{genome}—empirical genome-wide significance levels obtained after 1000 permutations, CI_{95}—95% empirical SNP distributions confidence intervals, NA—not applicable.

doi:10.1371/journal.pone.0134720.t001
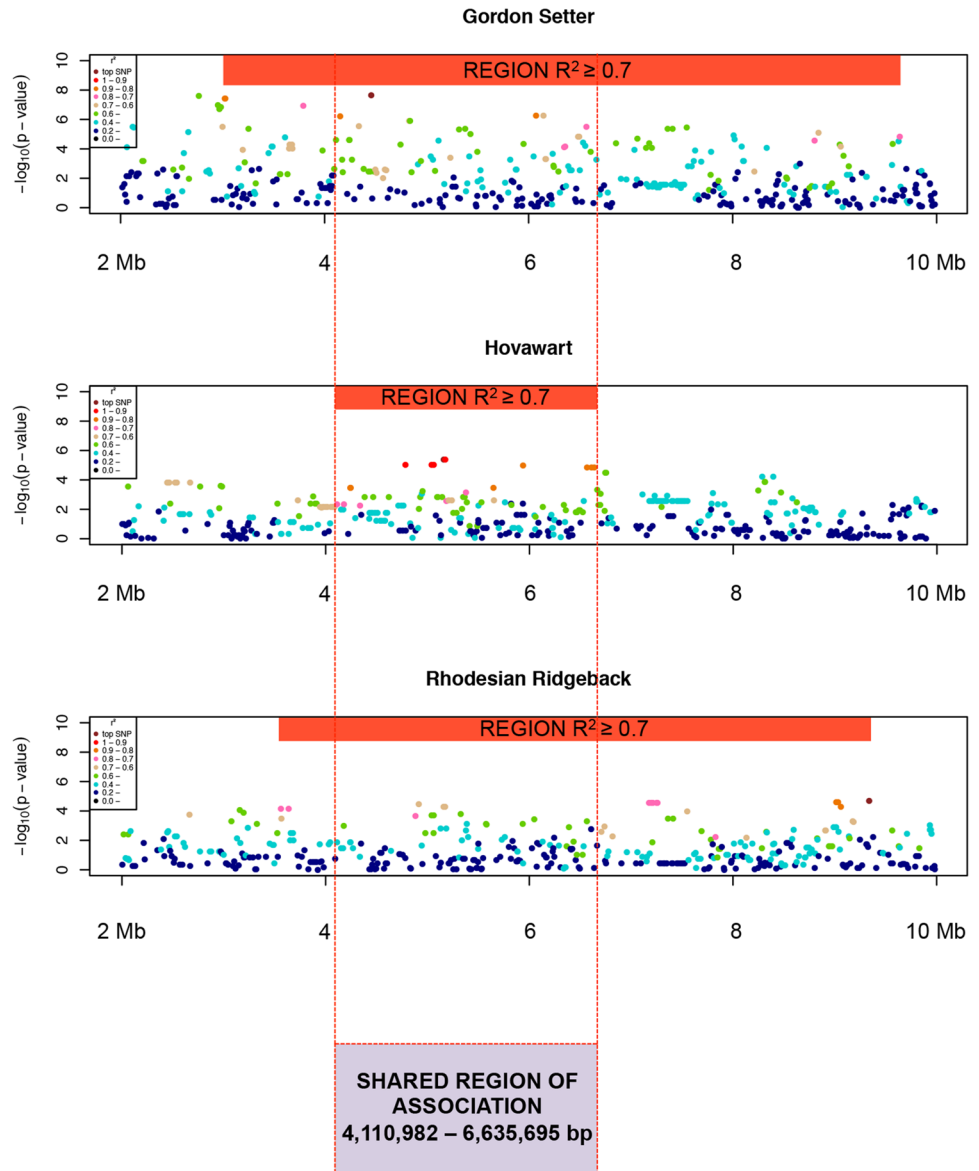
Fig 3. Breed-specific LD Manhattan plots of relevant region on CFA12, indicating R^2-values of each SNP in respect to the highest associated markers. Orange bars are highlighting the associated regions ($R^2 \geq 0.7$) in GS (3,022,017–9,637,877 bp), HV (4,110,982–6,635,695 bp) and RR (3,561,421–9,336,752 bp). Dashed lines indicate the shared region of association (4,110,982–6,635,695 bp).

doi:10.1371/journal.pone.0134720.g003

region contains many genes (n > 40), which include candidate genes involved in cell survival, apoptosis and immunity (*e.g. DEF6*, *MAPK14*, *STK38* and *CDKN1A*).

Haplotype analysis across breeds identifies a shared risk haplotype

Haplotype analysis was performed in each breed separately using the meta-analysis top SNP as a tagging marker for risk and control haplotypes. We identified risk haplotypes for each breed as following: GS 4,915,018–5,081,823 bp, HV 4,906,914–5,081,823 bp and RR 3,496,085–5,158,474 bp; thereby defining the risk haplotype shared across breeds as 4,915,018–5,081,823 bp. The defined risk haplotype spans ~167 kb and harbours 9 SNPs. Fig 5 shows all risk

Meta-analysis
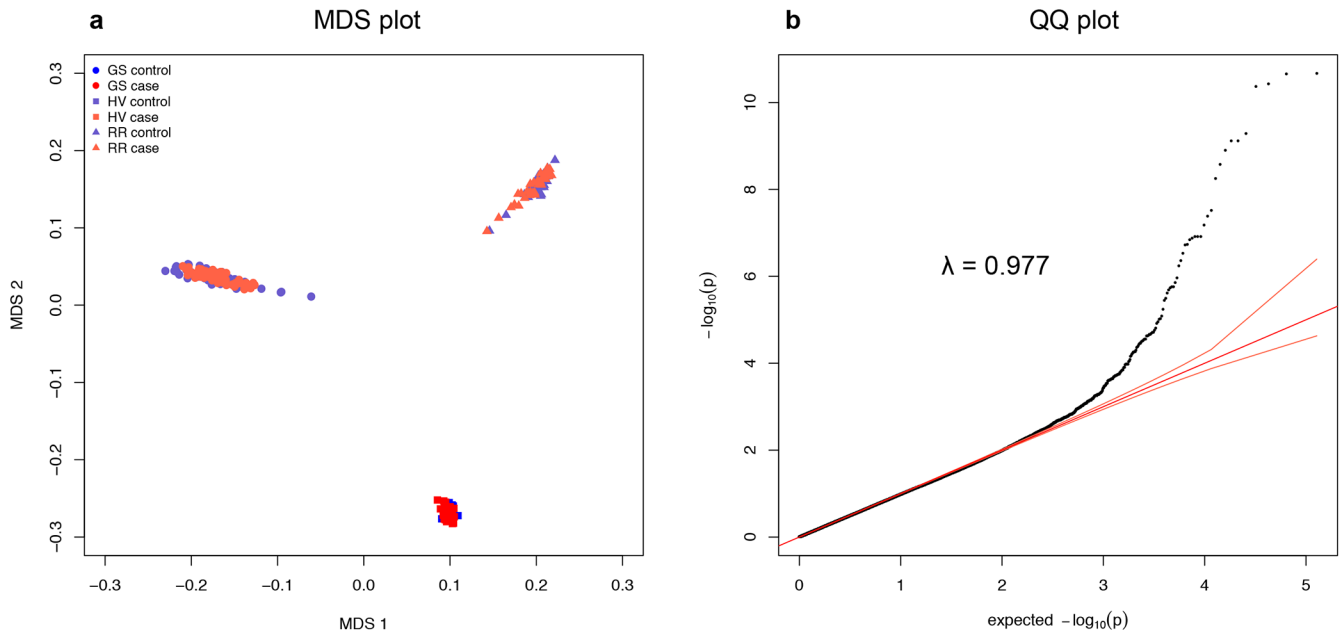
Fig 4. Meta-analysis (a) MDS plot showing genetic distances between GS, HV and RR sample sets and (b) QQ plot showing no inflation (λ = 0.977). The red lines show theoretical distribution in absence of association and theoretical confidence intervals.

doi:10.1371/journal.pone.0134720.g004

haplotypes identified in cases and all non-risk haplotypes identified in controls from all three breeds. The ~167 kb haplotype harbours three genes, *LHFPL5*, *SRPK1* and *SLC26A8*, which have not previously been implicated, and are thereby novel, with respect to the development of hypothyroidism.

The meta-analysis top SNP (5,039,806 bp, G/A, located between *SRPK1* and *SLC26A8* genes) and another SNP in the vicinity (5,060,994 bp, C/T, located in the last intron of *SLC26A8*) showed a complete LD in risk haplotypes and nearly complete LD in control haplotypes across all three breeds (recombination indicated on only one chromosome, see Fig 5). We used these two SNPs and the phenotypes for the following association tests: haplotype-based (non-risk = GC, risk = AT) and genotype-based (homozygous non-risk = G/G and C/C, heterozygous = G/A and C/T, and homozygous risk = A/A and T/T). We identified a significant enrichment of the risk (AT) haplotype in cases versus controls both in each breed separately and across breeds (see Table 2 for p-values). Similarly, a significant association was also observed when associating genotypes of these SNPs to the phenotypic classes in each breed separately and across breeds (see Table 2 for p-values).

Discussion

In this study we present the first GWA analysis identifying a major shared risk locus for the development of canine hypothyroidism in three high-risk dog breeds. By adapting a multi-breed approach, we have identified a shared risk haplotype for a complex disease. The possibility of using such an approach was proposed in 2007, when Karlsson and colleagues [8] used a monogenic white coat colour as an example trait. The authors used a one-breed GWA study, followed by an approach to narrow down the candidate region by including another breed with the same phenotype. Since then, studies have used an inter-breed approach to map canine
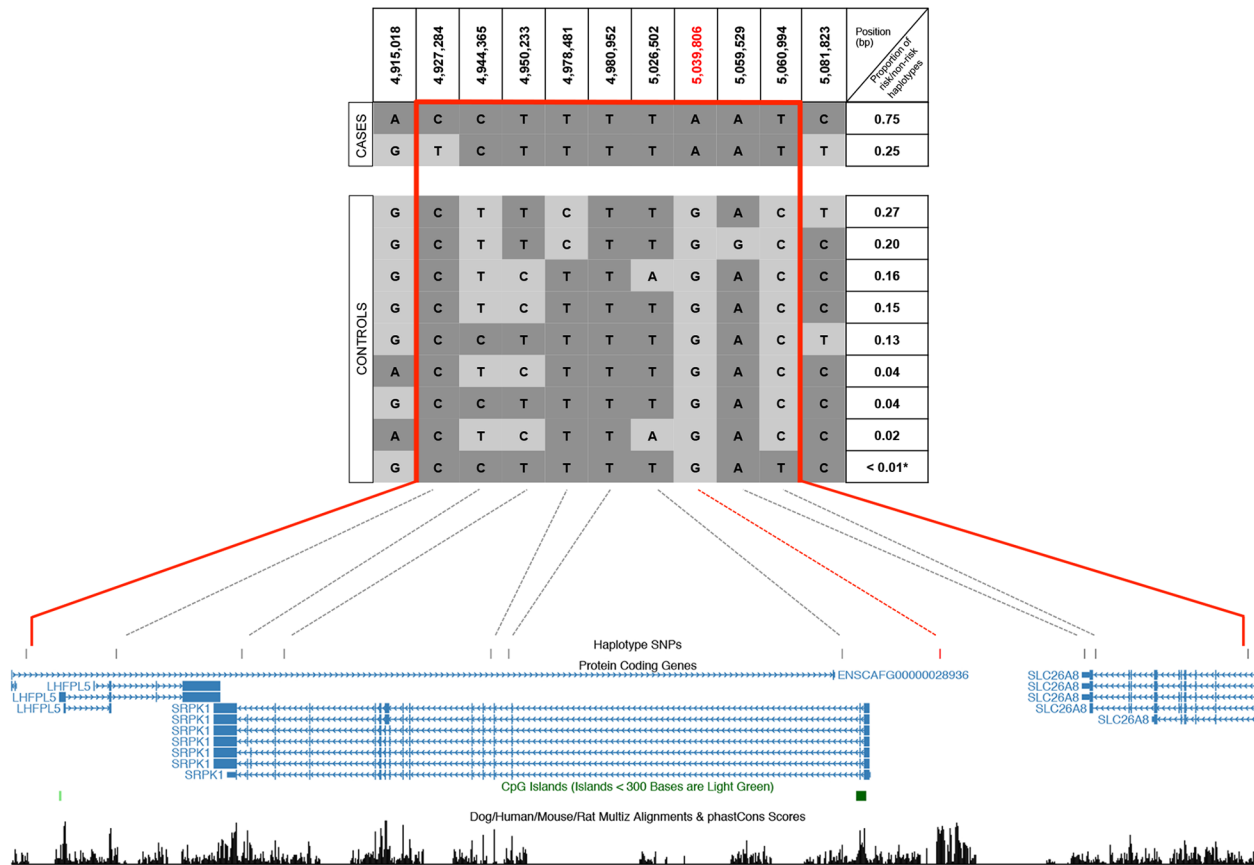
Fig 5. Haplotype analysis from all breeds identified two risk haplotypes in cases and nine non-risk haplotypes in controls. The upper panel indicates the alternative alleles in different grey shades, and the defined ~167 kb risk haplotype (4,915,018–5,081,823 bp) with red boundaries. Proportion of risk/non-risk haplotypes indicates the relative proportions of corresponding haplotypes among risk haplotypes determined in cases or non-risk haplotypes determined in controls, respectively. Star indicates a rare haplotype represented by only one chromosome. The lower panel shows the genomic location of the SNPs defining the risk haplotype in relation to protein coding genes (for each gene all the transcripts isoforms are included), CpG islands and conservation scores (figure adapted from: http://genome.ucsc.edu)

doi:10.1371/journal.pone.0134720.g005

brachycephaly [49], and a multi-breed common pathway analysis to identify genes behind canine osteosarcoma [1]. The integrated GWA and meta-analysis approach presented here is to our knowledge the first successful study identifying a risk locus involved in the development of a complex disease across dog breeds. Such meta-analysis has been widely used in human genetics, in which the detected loci explain only a small proportion of the genetic contribution to the respective complex traits [46, 50].

In our study, the meta-analysis validated and corroborated the single-breed GWA analysis results, as well as identified a ~167 kb risk haplotype for canine hypothyroidism shared among three high-risk breeds. Although the identified risk haplotype explains a large proportion of the cases in our study cohorts, there are likely additional risk factors contributing to the development of hypothyroidism in dogs and even in the breeds included in our study. The risk haplotype is located on CFA12, starting approximately 2 Mb downstream of the DLA gene cluster. DLA haplotypes on CFA12 in dogs [31, 32] and HLA haplotypes on HSA6 in humans [51–53] have already been associated with an increased risk of developing hypothyroidism in both species. Wilbe and colleagues [32] investigated the role of DLA haplotypes in hypothyroid Hovawart dogs without identifying any association and observing low genetic variation in the region. Such association analyses are considered difficult due to complex LD structure of the

Table 2. Haplotype and genotype association to phenotypic classes.

		Haplotype				Genotype (5,039,806)*				Genotype (5,060,994)*			
		GC	AT	p-value	OR	G/G	G/A	A/A	p-value	C/C	C/T	T/T	p-value
Gordon Setter	cases 63	98 (0.77)	28 (0.23)	7.3x10^{-4}	3.4 (CI = 1.6–7.5)	32 (0.52)	29 (0.48)	0 (0)	2.3x10^{-5}	34 (0.54)	29 (0.46)	0 (0)	3.3x10^{-5}
	controls84	154$^{\#}$ (0.92)	13 (0.08)			68 (0.85)	11 (0.14)	1 (0.01)		71 (0.85)	12 (0.14)	1 (0.01)	
Hovawart	cases 44	48 (0.55)	40 0.45)	5.3x10^{-5}	6.0 (CI = 2.4–17.4)	10 (0.22)	28 (0.64)	6 (0.14)	1.0x10^{-5}	10 (0.22)	28 (0.64)	6 (0.14)	1.0x10^{-5}
	controls29	51 (0.88)	7 (0.12)			22 (0.77)	7 (0.23)	0 (0)		22 (0.77)	7 (0.23)	0 (0)	
Rhodesian Ridgeback	cases 38	43 (0.57)	33 (0.43)	8.0x10^{-4}	3.2 (CI = 1.6–6.5)	13 (0.34)	17 (0.45)	8 (0.21)	3.0x10^{-3}	13 (0.34)	17 (0.45)	8 (0.21)	3.0x10^{-3}
	controls54	87 (0.81)	21 (0.19)			35 (0.65)	17 (0.31)	2 (0.04)		35 (0.65)	17 (0.31)	2 (0.04)	
Combined breeds	cases 145	189 (0.65)	101 (0.35)	4.5x10^{-11}	3.8 (CI = 2.5–5.9)	55 (0.38)	74 (0.52)	14 (0.1)	1.5x10^{-11}	57 (0.39)	74 (0.51)	14 (0.1)	3.0x10^{-11}
	controls167	292 (0.88)	41 (0.12)			125 (0.77)	35 (0.21)	3 (0.02)		128 (0.77)	36 (0.21)	3 (0.02)	

Association analysis of two tagging SNPs as haplotypes and genotypes to phenotypic classes in breeds separately and combined. Numbers in columns of haplotype indicate the number of chromosomes, whereas numbers in columns of genotypes indicate the number of individuals. Numbers in brackets show proportion of chromosomes/genotypes in cases and controls. OR—odds ratios, CI—confidence intervals.

* Differences in the number of genotypes for SNP 5,039,806 and 5,060,994 are due to missing genotypes

$^{\#}$ One missing chromosome due to a recombination event (*i.e.* AC-haplotype)

region [54], and also because of potentially spurious association signals obtained when the number of haplotypes in the region is limited [55]. Our study provides strong evidence of a canine hypothyroidism risk locus on a region of CFA12 not harbouring the DLA region.

We identified a ~167 kb risk haplotype associated with hypothyroidism in all the three breeds included in our study. Thus, we have determined a haplotype carrying the putative causative mutation(s), which could be located either in coding or regulatory region(s). The haplotype contains three genes (*LHFPL5*, *SRPK1* and *SLC26A8*) with functions that are not explicitly obvious in the development of hypothyroidism. However, they could pave the way to the characterisation of entirely novel pathways and mechanisms having a role in the etiology of this disease. *LHFPL5* is a gene encoding for lipoma HMGIC fusion partner-like 5. Mutations in this gene result in deafness in humans [56, 57], and in mice [58]. *SRPK1* (SRSF protein kinase 1) encodes a serine/arginine protein kinase specific for the SR (serine/arginine-rich domain) family of splicing factors. It has been shown as being an important factor in tumorigenesis [59], viral infection [60, 61] and apoptosis [62]. *SLC26A8* (solute carrier family 26 member 8) is a member of the *SLC26* gene family of anion transporters, which are well-conserved in both gene structure and protein length. Variants in one member of this gene family, specifically in *SLC26A4*, have been shown to cause a genetic disorder called Pendred syndrome, characterised by goitre and occasionally also hypothyroidism [63]. Variants in *SLC26A8* have been implicated to cause asthenozoospermia (reduced sperm motility) via altered interaction with CFTR (cystic fibrosis transmembrane conductance regulator) [64]. Variants in the *CFTR* gene are

known to cause cystic fibrosis, often comorbid with iodine deficiency and subclinical hypothyroidism [65, 66], thereby indicating a potential functional link to our phenotype of interest. Even though the expression of *SLC26A8* gene is reported to be restricted to spermatocytes [67], several publicly available databases (the Human Protein Atlas (www.proteinatlas.org) [68], BioGPS database (www.biogps.org) [69] suggest expression in a wide range of tissues and cell-lines. Consequently, *SLC26A8* emerges as the strongest candidate gene within the associated risk haplotype.

The associated haplotype contains many conserved elements based on analysis of 29 different mammalian species [70] and some GC-rich regions potentially functioning as CpG islands. The corresponding human region indicates abundant chromatin functional structures and transcription factor binding sites. It is possible that, despite being located in the ~167 kb haplotype identified in our study, the putative causative mutation(s) for canine hypothyroidism may be regulating the expression of a gene which may lie outside the borders of the haplotype.

On average we would expect haplotypes within dog breeds being around 1 Mb, and across several dog breeds about 10–100 kb [5]. However the breed-specific associated regions and the associated haplotype identified in our study are longer. One of the reasons for these long haplotypes may be selection *i.e.* favorable genetic information in the region keeping the haplotypes intact. The three genes located in the ~167 kb risk haplotype are all involved in basic physiological processes, suggesting that any of them might have represented a favourable target for selection. Therefore, it is possible that the detrimental hypothyroidism risk factor may have been "hitchhiking" together with the selected locus during dog domestication or breed-creation. In a study designed to identify domestication selection signals, the authors reported no sweep signals between wolves and dogs in the ~167 kb hypothyroidism risk haplotype implicated in our study [71]. Indeed, the hypothyroidism risk alleles are absent from the wolf population used in the Axelsson study, thereby not supporting the possible prior domestication origin of these alleles (http://genome.ucsc.edu, public track hub: Broad Improved Canine Annotation v1, track: Axelsson SNPs) [11, 71]. Additionally, hypothyroidism risk alleles have been reported in dog breeds with low risk for developing hypothyroidism [9]. Therefore we propose that the putative canine hypothyroidism risk factor appeared after domestication and before breed-creation, since gene flow between breeds included in our study is very unlikely. Further research on the evolutionary and demographic history of the hypothyroidism risk factor in different dog breeds is of utmost interest and should help us to strengthen this hypothesis.

In summary, we have demonstrated the notable potential of the integrated GWA and meta-analysis approach for detecting genetic loci underlying complex diseases in dogs. Further characterisation of the risk haplotype for canine hypothyroidism present in the Gordon Setter, Hovawart and the Rhodesian Ridgeback populations used in this study is necessary in order to extensively and deeply characterise the locus from a genomic and a functional point of view. We expect that future work focusing on this genomic region may identify a shared functional variant(s) increasing the risk of developing hypothyroidism in dogs. The identification of the functional variant(s) may contribute to the wellbeing not only of dogs, via breeding strategies, but also benefit human research, via identification of new potential genetic risk factors, pathways and treatment strategies for hypothyroidism.

Supporting Information

S1 Fig. QQ and Manhattan plots after conditioning breed-specific association analyses for respective top SNP.
(TIF)

S1 Table. Overview of samples used in the study.
(DOCX)

S2 Table. SNP-based quality control summary.
(DOCX)

S3 Table. Associations of breed-specific top SNP genotypes to phenotypic classes in the different breeds.
(DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: MB SD JM ED MK MLZ KS SIT OK GA WERO ÅH TL KLT LJK FL GRP. Performed the experiments: MB SD JM ED MLZ KS. Analyzed the data: MB SD JM MK MLZ KS KLT LJK FL GRP. Contributed reagents/materials/analysis tools: MB SD JM ED MK MLZ KS SIT OK GA WERO ÅH TL KLT LJK FL GRP. Wrote the paper: MB SD JM ED MK MLZ KS SIT OK GA WERO ÅH TL KLT LJK FL GRP.

References

1. 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. doi: 10.1186/gb-2013-14-12-r132 PMID: 24330828

2. Olsson M, Meadows JR, Truve K, Rosengren Pielberg G, Puppo F, Mauceli E, et al. 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. 2011; 7(3):e1001332. doi: 10.1371/journal.pgen. 1001332 PMID: 21437276

3. Tengvall K, Kierczak M, Bergvall K, Olsson M, Frankowiack M, Farias FH, et al. Genome-wide analysis in German shepherd dogs reveals association of a locus on CFA 27 with atopic dermatitis. PLoS Genet. 2013; 9(5):e1003475. doi: 10.1371/journal.pgen.1003475 PMID: 23671420

4. Wilbe M, Jokinen P, Truve K, Seppala EH, Karlsson EK, Biagi T, et al. Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. Nat Genet. 2010; 42(3):250–4. doi: 10.1038/ng.525 PMID: 20101241

5. 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. doi: 10.1038/nature04338 PMID: 16341006

6. Sutter NB, Ostrander EA. Dog star rising: the canine genetic system. Nat Rev Genet. 2004; 5(12):900–10. doi: 10.1038/nrg1492 PMID: 15573122

7. Andersson L, Georges M. Domestic-animal genomics: deciphering the genetics of complex traits. Nat Rev Genet. 2004; 5(3):202–12. doi: 10.1038/nrg1294 PMID: 14970822

8. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, et al. Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet. 2007; 39(11):1321–8. doi: 10.1038/ng.2007.10 PMID: 17906626

9. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, et al. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. PLoS Genet. 2011; 7(10):e1002316. doi: 10.1371/journal.pgen.1002316 PMID: 22022279

10. Briggs J, Paoloni M, Chen QR, Wen X, Khan J, Khanna C. A compendium of canine normal tissue gene expression. PLoS One. 2011; 6(5):e17107. doi: 10.1371/journal.pone.0017107 PMID: 21655323

11. 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. doi: 10.1371/journal.pone.0091172 PMID: 24625832

12. Ferguson DC. Testing for hypothyroidism in dogs. Vet Clin North Am Small Anim Pract. 2007; 37 (4):647–69, v. doi: 10.1016/j.cvsm.2007.05.015 PMID: 17619004

13. Gaitonde DY, Rowley KD, Sweeney LB. Hypothyroidism: an update. Am Fam Physician. 2012; 86 (3):244–51. PMID: 22962987

14. Dixon RM, Reid SW, Mooney CT. Epidemiological, clinical, haematological and biochemical characteristics of canine hypothyroidism. Vet Rec. 1999; 145(17):481–7. PMID: 10596870

15. Khandelwal D, Tandon N. Overt and subclinical hypothyroidism: who to treat and how. Drugs. 2012; 72 (1):17–33. doi: 10.2165/11598070-000000000-00000 PMID: 22191793

16. Laurberg P, Cerqueira C, Ovesen L, Rasmussen LB, Perrild H, Andersen S, et al. Iodine intake as a determinant of thyroid disorders in populations. Best Pract Res Clin Endocrinol Metab. 2010; 24(1):13–27. doi: 10.1016/j.beem.2009.08.013 PMID: 20172467

17. Jacobson DL, Gange SJ, Rose NR, Graham NM. Epidemiology and estimated population burden of selected autoimmune diseases in the United States. Clin Immunol Immunopathol. 1997; 84(3):223–43. PMID: 9281381

18. Beierwaltes WH, Nishiyama RH. Dog thyroiditis: occurrence and similarity to Hashimoto's struma. Endocrinology. 1968; 83(3):501–8. doi: 10.1210/endo-83-3-501 PMID: 5695421

19. Graham PA, Refsal KR, Nachreiner RF. Etiopathologic findings of canine hypothyroidism. Vet Clin North Am Small Anim Pract. 2007; 37(4):617–31, v. doi: 10.1016/j.cvsm.2007.05.002 PMID: 17619002

20. Happ GM. Thyroiditis—a model canine autoimmune disease. Adv Vet Sci Comp Med. 1995; 39:97–139. PMID: 8578979

21. Lucke VM, Gaskell CJ, Wotton PR. Thyroid pathology in canine hypothyroidism. J Comp Pathol. 1983; 93(3):415–21. PMID: 6688431

22. Mooney CT. Canine hypothyroidism: a review of aetiology and diagnosis. N Z Vet J. 2011; 59(3):105–14. doi: 10.1080/00480169.2011.563729 PMID: 21541883

23. Egenvall A, Bonnett BN, Olson P, Hedhammar A. Gender, age, breed and distribution of morbidity and mortality in insured dogs in Sweden during 1995 and 1996. Vet Rec. 2000; 146(18):519–25. PMID: 11321213

24. Nachreiner RF, Refsal KR, Graham PA, Bowman MM. Prevalence of serum thyroid hormone autoantibodies in dogs with clinical signs of hypothyroidism. J Am Vet Med Assoc. 2002; 220(4):466–71. PMID: 11860240

25. Scott DW, Paradis M. A survey of canine and feline skin disorders seen in a university practice: Small Animal Clinic, University of Montreal, Saint-Hyacinthe, Quebec (1987–1988). Can Vet J. 1990; 31 (12):830–5. PMID: 17423707

26. Benjamin SA, Stephens LC, Hamilton BF, Saunders WJ, Lee AC, Angleton GM, et al. Associations between lymphocytic thyroiditis, hypothyroidism, and thyroid neoplasia in beagles. Vet Pathol. 1996; 33(5):486–94. PMID: 8885174

27. Kennedy LJ, Huson HJ, Leonard J, Angles JM, Fox LE, Wojciechowski JW, et al. Association of hypothyroid disease in Doberman Pinscher dogs with a rare major histocompatibility complex DLA class II haplotype. Tissue Antigens. 2006; 67(1):53–6. doi: 10.1111/j.1399-0039.2005.00518.x PMID: 16451201

28. Graham PA, Nachreiner RF, Refsal KR, Provencher-Bolliger AL. Lymphocytic thyroiditis. Vet Clin North Am Small Anim Pract. 2001; 31(5):915–33, vi–vii. PMID: 11570132

29. Ban Y, Davies TF, Greenberg DA, Concepcion ES, Tomer Y. The influence of human leucocyte antigen (HLA) genes on autoimmune thyroid disease (AITD): results of studies in HLA-DR3 positive AITD families. Clin Endocrinol (Oxf). 2002; 57(1):81–8.

30. Roman SH, Greenberg D, Rubinstein P, Wallenstein S, Davies TF. Genetics of autoimmune thyroid disease: lack of evidence for linkage to HLA within families. J Clin Endocrinol Metab. 1992; 74(3):496–503. doi: 10.1210/jcem.74.3.1740483 PMID: 1740483

31. Kennedy LJ, Quarmby S, Happ GM, Barnes A, Ramsey IK, Dixon RM, et al. Association of canine hypothyroidism with a common major histocompatibility complex DLA class II allele. Tissue Antigens. 2006; 68(1):82–6. doi: 10.1111/j.1399-0039.2006.00614.x PMID: 16774545

32. Wilbe M, Sundberg K, Hansen IR, Strandberg E, Nachreiner RF, Hedhammar A, et al. Increased genetic risk or protection for canine autoimmune lymphocytic thyroiditis in Giant Schnauzers depends on DLA class II genotype. Tissue Antigens. 2010; 75(6):712–9. doi: 10.1111/j.1399-0039.2010.01449. x PMID: 20210920

33. Szinnai G. Genetics of normal and abnormal thyroid development in humans. Best Pract Res Clin Endocrinol Metab. 2014; 28(2):133–50. doi: 10.1016/j.beem.2013.08.005 PMID: 24629857

34. Fyfe JC, Kampschmidt K, Dang V, Poteet BA, He Q, Lowrie C, et al. Congenital hypothyroidism with goiter in toy fox terriers. J Vet Intern Med. 2003; 17(1):50–7. PMID: 12564727

35. Dodgson SE, Day R, Fyfe JC. Congenital hypothyroidism with goiter in Tenterfield terriers. J Vet Intern Med. 2012; 26(6):1350–7. doi: 10.1111/j.1939-1676.2012.01015.x PMID: 23113744

36. Massey JP. Mapping the genes for complex canine autoimmune diseases [Dphil thesis]: the University of Manchester; 2012.

37. Ferm K, Bjornerfeldt S, Karlsson A, Andersson G, Nachreiner R, Hedhammar A. Prevalence of diagnostic characteristics indicating canine autoimmune lymphocytic thyroiditis in giant schnauzer and hovawart dogs. J Small Anim Pract. 2009; 50(4):176–9. doi: 10.1111/j.1748-5827.2008.00696.x PMID: 19320811

38. Ihaka R, Gentleman R. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics. 1996; 5(3):299–314.

39. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007; 23(10):1294–6. doi: 10.1093/bioinformatics/btm108 PMID: 17384015

40. Fisher RA. On the Interpretation of χ2 from Contingency Tables, and the Calculation of P. Journal of the Royal Statistical Society. 1922; 85(1):87–94.

41. Fleiss JL, editor. Statistical methods for rates and proportions. 2nd ed. New York: Wiley; 1981.

42. Rönnegård L, Shen X, Alam M. hglm: A package for fitting hierarchical generalized linear models. The R Journal. 2010; 2:20–8.

43. Hoffman GE. Correcting for population structure and kinship using the linear mixed model: theory and extensions. PLoS One. 2013; 8(10):e75707. doi: 10.1371/journal.pone.0075707 PMID: 24204578

44. Belonogova NM, Svishcheva GR, van Duijn CM, Aulchenko YS, Axenovich TI. Region-based association analysis of human quantitative traits in related individuals. PLoS One. 2013; 8(6):e65395. doi: 10.1371/journal.pone.0065395 PMID: 23799013

45. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots2014. Available from: biorxiv.org/content/early/2014/05/14/005165.

46. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet. 2013; 14(6):379–89. doi: 10.1038/nrg3472 PMID: 23657481

47. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet. 2006; 78(4):629–44. doi: 10.1086/502802 PMID: 16532393

48. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series 5. 1900; 50(302):157–75.

49. Bannasch D, Young A, Myers J, Truve K, Dickinson P, Gregg J, et al. Localization of canine brachycephaly using an across breed mapping approach. PLoS One. 2010; 5(3):e9632. doi: 10.1371/journal.pone.0009632 PMID: 20224736

50. Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. Pharmacogenomics. 2009; 10(2):191–201. doi: 10.2217/14622416.10.2.191 PMID: 19207020

51. Levin L, Ban Y, Concepcion E, Davies TF, Greenberg DA, Tomer Y. Analysis of HLA genes in families with autoimmune diabetes and thyroiditis. Hum Immunol. 2004; 65(6):640–7. doi: 10.1016/j.humimm.2004.02.026 PMID: 15219384

52. Menconi F, Monti MC, Greenberg DA, Oashi T, Osman R, Davies TF, et al. Molecular amino acid signatures in the MHC class II peptide-binding pocket predispose to autoimmune thyroiditis in humans and in mice. Proc Natl Acad Sci U S A. 2008; 105(37):14034–9. doi: 10.1073/pnas.0806584105 PMID: 18779568

53. Moens H, Farid NR, Sampson L, Noel EP, Barnard JM. Hashimoto's thyroiditis is associated with HLA-DRw3. N Engl J Med. 1978; 299(3):133–4. doi: 10.1056/NEJM197807202990306 PMID: 78445

54. Seddon JM, Berggren KT, Fleeman LM. Evolutionary history of DLA class II haplotypes in canine diabetes mellitus through single nucleotide polymorphism genotyping. Tissue Antigens. 2010; 75(3):218–26. doi: 10.1111/j.1399-0039.2009.01426.x PMID: 20047645

55. Safra N, Pedersen NC, Wolf Z, Johnson EG, Liu HW, Hughes AM, et al. Expanded dog leukocyte antigen (DLA) single nucleotide polymorphism (SNP) genotyping reveals spurious class II associations. Vet J. 2011; 189(2):220–6. doi: 10.1016/j.tvjl.2011.06.023 PMID: 21741283

56. Kalay E, Li Y, Uzumcu A, Uyguner O, Collin RW, Caylan R, et al. Mutations in the lipoma HMGIC fusion partner-like 5 (LHFPL5) gene cause autosomal recessive nonsyndromic hearing loss. Hum Mutat. 2006; 27(7):633–9. doi: 10.1002/humu.20368 PMID: 16752389

57. Shabbir MI, Ahmed ZM, Khan SY, Riazuddin S, Waryah AM, Khan SN, et al. Mutations of human TMHS cause recessively inherited non-syndromic hearing loss. J Med Genet. 2006; 43(8):634–40. doi: 10.1136/jmg.2005.039834 PMID: 16459341

58. Longo-Guess CM, Gagnon LH, Cook SA, Wu J, Zheng QY, Johnson KR. A missense mutation in the previously undescribed gene Tmhs underlies deafness in hurry-scurry (hscy) mice. Proc Natl Acad Sci U S A. 2005; 102(22):7894–9. doi: 10.1073/pnas.0500760102 PMID: 15905332

59. Toker A, Chin YR. Akt-ing up on SRPK1: oncogene or tumor suppressor? Mol Cell. 2014; 54(3):329–30. doi: 10.1016/j.molcel.2014.04.020 PMID: 24813709

60. Nousiainen L, Sillanpaa M, Jiang M, Thompson J, Taipale J, Julkunen I. Human kinome analysis reveals novel kinases contributing to virus infection and retinoic-acid inducible gene I-induced type I and type III IFN gene expression. Innate Immun. 2013; 19(5):516–30. doi: 10.1177/1753425912473345 PMID: 23405030

61. Prescott EL, Brimacombe CL, Hartley M, Bell I, Graham S, Roberts S. Human papillomavirus type 1 E1^E4 protein is a potent inhibitor of the serine-arginine (SR) protein kinase SRPK1 and inhibits phosphorylation of host SR proteins and of the viral transcription and replication regulator E2. J Virol. 2014; 88(21):12599–611. doi: 10.1128/JVI.02029-14 PMID: 25142587

62. Kamachi M, Le TM, Kim SJ, Geiger ME, Anderson P, Utz PJ. Human autoimmune sera as molecular probes for the identification of an autoantigen kinase signaling pathway. J Exp Med. 2002; 196 (9):1213–25. PMID: 12417631

63. Dror AA, Lenz DR, Shivatzki S, Cohen K, Ashur-Fabian O, Avraham KB. Atrophic thyroid follicles and inner ear defects reminiscent of cochlear hypothyroidism in Slc26a4-related deafness. Mamm Genome. 2014; 25(7–8):304–16. doi: 10.1007/s00335-014-9515-1 PMID: 24760582

64. Dirami T, Rode B, Jollivet M, Da Silva N, Escalier D, Gaitch N, et al. Missense mutations in SLC26A8, encoding a sperm-specific activator of CFTR, are associated with human asthenozoospermia. Am J Hum Genet. 2013; 92(5):760–6. doi: 10.1016/j.ajhg.2013.03.016 PMID: 23582645

65. Azizi F, Bentley D, Vagenakis A, Portnay G, Bush JE, Shwachman H, et al. Abnormal thyroid function and response to iodides in patients with cystic fibrosis. Trans Assoc Am Physicians. 1974; 87:111–9. PMID: 4456727

66. De Luca F, Trimarchi F, Sferlazzas C, Benvenga S, Costante G, Mami C, et al. Thyroid function in children with cystic fibrosis. Eur J Pediatr. 1982; 138(4):327–30. PMID: 6813123

67. Toure A, Morin L, Pineau C, Becq F, Dorseuil O, Gacon G. Tat1, a novel sulfate transporter specifically expressed in human male germ cells and potentially linked to rhogtpase signaling. J Biol Chem. 2001; 276(23):20309–15. doi: 10.1074/jbc.M011740200 PMID: 11278976

68. Berglund L, Bjorling E, Oksvold P, Fagerberg L, Asplund A, Szigyarto CA, et al. A genecentric Human Protein Atlas for expression profiles based on antibodies. Mol Cell Proteomics. 2008; 7(10):2019–27. doi: 10.1074/mcp.R800013-MCP200 PMID: 18669619

69. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. 2009; 10(11):R130. doi: 10.1186/gb-2009-10-11-r130 PMID: 19919682

70. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. Nature. 2011; 478(7370):476–82. doi: 10.1038/nature10530 PMID: 21993624

71. Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature. 2013; 495(7441):360–4. doi: 10.1038/nature11837 PMID: 23354050