ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

FULL PAPER

doi: 10.1111/age.13047

Polymorphisms in canine immunoglobulin heavy chain gene cluster: a double-edged sword for diabetes mellitus in the dog

C. R. Sanz*, N. Sevane*, M. D. Pérez-Alenza†, M. Valero-Lorenzo* and S. Dunner*

*Department of Animal Production, Veterinary Faculty, Complutense University of Madrid, Madrid 28040, Spain. †Department of Animal Medicine, Surgery and Pathology, Veterinary Faculty, Complutense University of Madrid, Madrid 28040, Spain.

Summary

Insulin deficiency diabetes (IDD) in dogs is an endocrine disease similar to human type 1 diabetes. There are breeds more commonly affected, such as Yorkshire Terrier and Samoyed, suggesting an underlying genetic component. However, the genetic basis for canine diabetes mellitus (DM) is not fully established. We conducted both whole-genome scans for selection signatures and GWASs to compare the genomes of 136 dogs belonging to 29 breeds previously described at low or high risk for developing DM. Candidate variants were tested in dogs with a diagnosis of IDD and controls attending the Complutense Veterinary Teaching Hospital. The only genomic region under selection (CFA8:72 700 000–74 600 000; CanFam3.1) retrieved by our analyses is included in the immunoglobulin heavy chain gene cluster, which has already been related to human human type 1 diabetes susceptibility. This region contains two non-synonymous variants, rs852072969 and rs851728071, showing significant associations with high or low risk for IDD, respectively. The first variant, rs852072969, alters a protein poorly characterised in the dog. In contrast, rs851728071 was predicted to block the synthesis of an immunoglobulin variable (V) domain in breeds at low risk for DM. Although a large and diverse V gene repertoire is thought to offer a fitness advantage, we suggest that rs851728071 prevents the formation of an auto-reactive immunoglobulin V domain probably involved in the pathophysiology of IDD and, thus, decreases the risk for the disease. These results should be interpreted with caution until the functional roles of the proposed variants have been proved in larger studies.

Keywords auto-reactive, *Canis lupus familiaris*, genetic variant, genome-wide association studies, resistance, susceptibility, whole-genome, XP-EHH

Introduction

Diabetes mellitus (DM) is one of the most common endocrine disorders in dogs, with a prevalence ranging from 0.0005 to 1.5% (Wilkinson 1960; Mattheeuws *et al.* 1984; Catchpole *et al.* 2005). Canine DM can be classified into insulin deficiency diabetes (IDD) or insulin resistance diabetes (IRD). Most diabetic dogs are thought to have IDD, resulting mainly from a progressive pancreatic β-cell destruction that leads to insulin deficiency and persistent hyperglycaemia, with only a small percentage of them suffering from IRD, primarily related to the progesterone-dominated phase of dioestrus, Cushing's disease or the presence of acromegaly (Guptill *et al.* 2003; Catchpole *et al.* 2005). The clinical consequences of both types of DM include polyuria, polydipsia, polyphagia, weight loss, lethargy and diabetic ketoacidosis. To control clinical signs and prevent complications, these patients usually require life-long insulin treatment.

Even though IDD resembles many of the hallmarks of human type 1 diabetes (T1D), the underlying cause of pancreatic β-cell destruction is less clear than in T1D, in which an immune-mediated pathogenesis is well established (American Diabetes Association 2014). However, evidence for autoimmunity in the pathogenesis of canine diabetes is present in multiple studies, which detected serum anti-β-cell antibodies and insulin antibodies, as well as proinsulin autoantibodies, in untreated diabetic dogs (Hoenig & Dawe 1992; Davison *et al.* 2008a, 2008b, 2011; Holder *et al.* 2015). Additionally, autoantibodies against GAD65 or IA-2,

Address for correspondence

C. R. Sanz, Department of Animal Production, Veterinary Faculty, Complutense University of Madrid, Madrid 28040, Spain.
E-mail: caroro07@ucm.es

Accepted for publication 19 January 2021

two major autoantigens in T1D, were found in newly diagnosed, untreated diabetic dogs (Davison *et al*. 2008a, 2008b). Other indirect evidence of autoimmunity was the decrease in stimulated insulin release and the fast lysis of islet cells when mice were exposed to serum from diabetic dogs *in vitro* (Sai *et al*. 1984). Along these lines, several genes involved in immune response processes have been implicated in determining susceptibility to canine diabetes, suggesting not only an immune-mediated pathogenesis, but also the existence of an underlying genetic component (O'Kell *et al*. 2017).

The contribution of genetic variants to immune-mediated DM has already been evidenced in other species. So far, there are more than 60 loci established as risk factors for human T1D (Pociot & McDermott 2002), with variants located within immune genes conferring the greatest genetic risk (Ounissi-Benkalha & Polychronakos 2008), especially those coding for human leukocyte antigen. Some other relevant genes, such as *INS*, *CTLA4*, *IL2RA*, *PTPN22* and *IFIH1*, may also contribute to disease susceptibility (Noble & Valdes 2011; Tandon 2015; Nyaga *et al*. 2018; Bakay *et al*. 2019). In non-obese diabetic mice, additional genes have been associated with increased risk for diabetes, such as genes coding for IL-2, IL-21, T-cell receptor, CD30, TNFR2 and β2-microglobulin (Driver *et al*. 2011; Jayasimhan *et al*. 2014).

In dogs, genetic susceptibility to DM has likewise been reported (Table 1). Similar to humans, some dog leukocyte antigen (analogous to human leukocyte antigen) haplotypes have been shown to confer susceptibility or protection in a variety of dog breeds (Kennedy *et al*. 2006), although these results were not corrected for multiple testing and population structure could have contributed to the associations. Recently, 37 SNPs, located in the candidate genes *PTPN22*, *IL10*, *IL12B*, *IL6*, *IL4*, *CCL5*, *IFNG*, *INS*, *IL1A*, *TNFA*, *IGF2* and *CTLA4*, all involved in the immune response, have also been linked to susceptible and protective phenotypes (Short *et al*. 2007, 2009, 2010). However, the implication of other genes in the pathogenesis of IDD is expected, since it shares much of the complexity that is noted in T1D (Short *et al*. 2010; O'Kell *et al*. 2017; Moshref *et al*. 2019). Next-generation sequencing approaches, such as WGS, have provided new insights into the molecular basis of a wide range of diseases (Precone *et al*. 2015) and could help disentangle genetic influences on canine DM.

Table 1 Breeds, their risk for developing diabetes mellitus (DM) as previously described in the literature, and number of animals (*n*) included in each group for the extended haplotype homozygosity (XP-EHH) and GWAS analyses.

Group	Breed	Risk	*n*	Reference
LR1	Golden Retriever	Low	20	Hess *et al*. (2000), Davison *et al*. (2005), Mattin *et al*. (2014), Yoon *et al*. (2020)
	Pekingese	Low	1	Guptill *et al*. (2003)
	Shih Tzu	Low	1	Guptill *et al*. (2003)
LR2	German Shepherd	Low	15	Hess *et al*. (2000), Guptill *et al*. (2003), Davison *et al*. (2005), Mattin *et al*. (2014), Yoon *et al*. (2020)
	Norwegian Elkhound	Low	2	Guptill *et al*. (2003)
	Shetland Sheepdog	Low	3	Guptill *et al*. (2003)
LR3	Beagle	Low	4	Guptill *et al*. (2003)
	Boston Terrier	Low	1	Guptill *et al*. (2003)
	Brittany Spaniel	Low	1	Guptill *et al*. (2003)
	Bulldog	Low	1	Guptill *et al*. (2003)
	Dalmatian	Low	2	Guptill *et al*. (2003)
	English Setter	Low	3	Guptill *et al*. (2003)
	Great Dane	Low	5	Guptill *et al*. (2003)
	Irish Setter	Low	1	Guptill *et al*. (2003)
HR1	Australian Terrier	High	1	Guptill *et al*. (2003), Fall *et al*. (2007), Wiles *et al*. (2017) and Yoon *et al*. (2020)
	Border Terrier	High	4	Mattin *et al*. (2014)
	Cairn Terrier	High	1	Guptill *et al*. (2003) and Davison *et al*. (2005)
	Miniature Schnauzer	High	2	Hess *et al*. (2000), Guptill *et al*. (2003) and Yoon *et al*. (2020)
	Standard Schnauzer	High	4	Guptill *et al*. (2003) and Yoon *et al*. (2020)
	Tibetan Terrier	High	2	Davison *et al*. (2005)
	West Highland White Terrier	High	11	Wiles *et al*. (2017) and Yoon *et al*. (2020)
HR2	Border Collie	High	14	Mattin *et al*. (2014)
	Keeshond	High	1	Guptill *et al*. (2003)
	Miniature Poodle	High	2	Hess *et al*. (2000), Guptill *et al*. (2003), Wiles *et al*. (2017) and Yoon *et al*. (2020)
	Pug	High	1	Hess *et al*. (2000)
	Samoyed	High	2	Hess *et al*. (2000), Guptill *et al*. (2003), Davison *et al*. (2005) and Fall *et al*. (2007)
	Siberian Husky	High	4	Guptill *et al*. (2003) and Yoon *et al*. (2020)
	Toy Poodle	High	2	Hess *et al*. (2000), Guptill *et al*. (2003) and Yoon *et al*. (2020)
HR3	Yorkshire Terrier	High	25	Mattin *et al*. (2014)

Low risk: OR < 1; high risk: OR > 1.

The aim of this work was thus to analyse WGS data (Plassais *et al.* 2019) of various dog breeds at low and high risk for developing DM (Hess *et al.* 2000; Guptill *et al.* 2003; Catchpole *et al.* 2005) to find additional genetic variants influencing the disease. Both genome scans for selection signatures and GWASs were performed. Firstly, we detected regions under selection, significant variants and candidate genes by comparing groups of susceptible and protected breeds. Then, we identified non-synonymous variants located in candidate genes. Finally, these variants were tested in dogs diagnosed with IDD and controls attending the Complutense Veterinary Teaching Hospital.

Materials and methods

Whole-genome sequencing data from 136 dogs belonging to 29 breeds previously found to be at low (OR < 1) or high risk (OR > 1) for developing DM (Table 1) were obtained from Plassais *et al.* (2019; accession no. PRJNA448733) after removing samples with <20× coverage.

Three whole-genome scans for selection signatures based on cross-population extended haplotype homozygosity (XP-EHH) were performed to characterise genetic differences between breeds at low or high risk for developing DM (Sabeti *et al.* 2007). The XP-EHH test detects selective sweeps in which the selected allele has achieved fixation in one canine population but remains polymorphic in the canine population as a whole (Sabeti *et al.* 2007). To conduct these XP-EHH scans, the WGS data was split into six subgroups (Table 1). Breeds at low risk were randomly divided into three subgroups (LR1, LR2 and LR3), while breeds at high risk were classified based on the presence (HR1 and HR3) or absence of furnishings (HR2). All bearded breeds at high risk for developing DM, except the Yorkshire Terrier, were included in group HR1. The Yorkshire Terrier constituted a subgroup by itself (HR3) owing to the high number of samples available. Selective sweeps were identified using the software REHH v3.0.1 (Gautier *et al.* 2017), comparing pairs of groups as follows: HR1–LR1, HR2–LR2 and HR3–LR3. This triple comparison allows the exclusion of some selective sweeps not linked to the predisposition to canine DM, especially those related to the presence of furnishings, since we found in exploratory analyses that mutations in the *RSPO2* gene, associated with moustache and eyebrow growth pattern (Cadieu *et al.* 2009), could generate biased signals of selection (data not available). We used default options for all analyses, except for *p.adjust.method*, which was set to Bonferroni. We used a window size of 1 Mb and the minimum number of extremal markers to define a window was 2. Significant SNPs were defined by selecting the top 0.1% of XP-EHH scores for each cross-population comparison and, if the distance between them was below 100 000 bp, the SNPs were grouped into a single region under selection. The choice of this threshold was based on the work by Sabati *et al.* (2007), in order to avoid false positives outliers owing to the number of samples analysed and their genetic heterogenicity.

BIOMART v3.8 (Durinck *et al.* 2009) was used to identify positively selected genes located within or in close proximity (distance <100 kb) to these regions. Then, to minimise false discovery owing to genetic drift resulting from the small effective population size of dog breeds and the strong human-driven selection for specific breed traits, only the genes within positively selected regions detected by the three comparisons were considered as high-confidence candidate genes.

Subsequently, a GWAS was carried out to identify variants associated with low or high risk for developing DM within the candidate genes. For this aim, standard methodology for case–control analysis (χ^2 test and logistic regression) was applied using PLINK v1.90 with default parameters (Purcell *et al.* 2007). Obtained *P*-values were corrected for multiple testing using the Benjamini–Hochberg method implemented in the R STATS package version 3.6.2 (R Core Team 2020). The significance threshold for the GWAS was set at a suggestive level of significance ($P < 0.05$; false discovery rate, FDR, <0.1) owing to the relatively small size of our testing population (136 individuals) coupled with stringent thresholds applied to account for multiple testing error (>300 000 variants were evaluated). The main difference between the GWAS analysis and the XP-EHH test is that the GWAS evaluates SNPs and small INDEL frequencies individually, while the XP-EHH approach assesses frequencies of long haplotypes. The combination of these two methods contributes to a more precise identification of candidate variants.

Significant variants located in candidate genes were further investigated for their potential effects on proteins using VEP 98 with default parameters (McLaren *et al.* 2016). Although regulatory or non-deleterious variants could also play a role, their potential effects over the risk for developing DM were not investigated here because this information is still poorly annotated in the dog genome. VEP 98 calculates the SIFT score, which predicts whether an amino acid substitution is likely to affect protein function based on sequence homology and the physico-chemical similarity between the alternate amino acids. The score is the normalised probability that the amino acid change is tolerated so scores nearer zero are more likely to be deleterious. The qualitative prediction is derived from this score such that substitutions with a score <0.05 are called 'deleterious' and all others are called 'tolerated' (Kumar *et al.* 2009). Using IGV v2.6.3 (Robinson *et al.* 2011; Thorvaldsdóttir *et al.* 2013) and manual inspection, the coverage at the identified positions of non-synonymous substitutions was evaluated.

A pilot study to validate the significant SNPs was accomplished by direct sequencing of PCR products from a total of 12 dogs (eight diabetic and four control animals) belonging to seven different breeds: Labrador Retriever,

Miniature Poodle, Miniature Schnauzer, Pug, Samoyed, West Highland White Terrier and Yorkshire Terrier (Table S1). Diabetic dogs were diagnosed with primary IDD based on consistent clinical signs (polyuria, polydipsia, polyphagia and weight loss), documented hyperglycaemia and the exclusion of other causes of secondary DM, including dioestrus, gestational diabetes, Cushing's disease, acromegaly and pancreatitis. Control animals were dogs older than 11 years who were attending the Complutense Veterinary Teaching Hospital for regular health checks and had never been diagnosed with diabetes. Genomic DNA was isolated from residual pathology blood samples preserved in Magic Buffer® (Biogen) by standard phenol–chloroform method (Sambrook & Russell 2006). Primers for variants rs852072969 (5′-CACGGGGTTCCATCTTGCATA-3′ and 5′-GTTCCTTTATGCATCTCCAGCACAT-3′) and rs851728071 (5′-CTGGCAGAACCAGCTCATGTA-3′ and 5′-CAAATCGGG CGCTCCCTAAA-3′) were designed based on the *Canis lupus familiaris* publicly available genome (CanFam3.1) using Primer-blast (Ye *et al.* 2012). The expected amplicon sizes were 364 and 404 bp, respectively. The obtained PCR fragments were subjected to Sanger sequencing (Macrogen). Allele frequencies between our study population ($n = 12$) and a previously described reference population ($n = 238$; accession no. PRJEB24066; Hunt *et al.* 2018) were compared by χ^2 tests.

Results and discussion

The XP-EHH method was applied to detect signals that may reflect the different risk for developing DM observed in some dog breeds. Figure 1 depicts the genome-wide distribution of outliers on each autosome detected by XP-EHH analysis. Distribution plots of raw XP-EHH values are shown in Fig. S1. As expected, the distributions of XP-EHH values were close to Gaussian. The numbers of significant SNPs and selective sweeps identified per scan are listed in Table S2. Using the criteria of contiguous blocks of at least two significant SNPs from the XP-EHH analysis confirmed with the three cross-population comparisons, we retrieved only one genomic region under selection, located on chromosome 8 (CFA8:72 700 000–74 600 000; CanFam3.1) (Fig. 2; Table S2). This region, included within the immunoglobulin heavy chain (IGH) cluster (Lefranc *et al.* 2009; Massari *et al.* 2009; Matiasovic *et al.* 2009; Bao *et al.* 2010; Mineccia *et al.* 2012; Hwang *et al.* 2018; Martin *et al.* 2018), harbours the *immunoglobulin heavy constant mu* (*IGHM*) gene, previously associated with autoantibody reactivity and T1D (Rolim *et al.* 2017), and 36 novel Ensembl genes, predicted based on cDNA sequence information from dog, with 29 of them containing immunoglobulin-like domains (Table S2).

To identify specific polymorphisms associated with susceptible and protected phenotypes, a GWAS was also conducted to evaluate a total of 332 593 variants. This resulted in the detection of 123 768 significant ($P < 0.05$; FDR < 0.1) variants (Table S3). Among them, 1323 were located within 16 of the 36 candidate genes found with XP-EHH (Table 2). A Q–Q plot is provided in Figure S2, which shows a distinct excess of large *P*-values compared with the theoretical expectation, consistent with the discovery of real associations. These polymorphisms were investigated for their potential effects on the encoded protein structure using vep, and the results suggested that only two of the 1323 variants (Table 3) were deleterious (SIFT score < 0.05; Kumar *et al.* 2009). Both of them were located at novel Ensembl genes: rs852072969 in *ENSCAFG00000049412*, and rs851728071 in *ENSCAFG00000043645*. Haplotype bifurcation diagrams are presented in Fig. S3, showing the breakdown of LD at increasing distances in both directions from the selected core regions.

The first one, rs852072969, is associated with the susceptible phenotype ($P = 0.002$; FDR = 0.012), with 31.6% of the 76 dogs from the WGS data belonging to breeds at high risk for developing DM carrying the variant. Breeds that contributed the most to its identification were Yorkshire Terrier (46.2% of the carriers at high risk, including one homozygous individual) and West Highland White Terrier (23.1%, also including one homozygous), followed by Standard Schnauzer and Border Collie (each of them representing a 7.7% and all heterozygous). The variant consisted of a non-synonymous substitution T to A at CFA8:74 291 885 (CanFam3.1), changing an aspartic acid for a valine in the amino acid position 365 of the protein encoded by *ENSCAFG00000049412*. This candidate gene is an orthologue of the *ADAM* gene family. In human and mouse, *ADAM* genes are found in the IGH cluster, although this potential orthologue has not been added to the IGH annotation in dogs given the limited knowledge of this gene family in dogs (Martin *et al.* 2018). To check the association between the polymorphism rs852072969 and the predisposition towards developing IDD, the region containing this variant was sequenced in both diabetic ($n = 8$) and control ($n = 4$) dogs included in this pilot study and the allele frequencies were compared with the reference population ($n = 238$). The diabetic group was composed of dogs diagnosed with primary IDD, excluding IRD, given that the candidate SNPs were located within an immune cluster and probably influence the pathogenesis of IDD. In the diabetic group, the frequency of the A allele (25%) triplicated the highest MAF observed in the reference population (8%; Hunt *et al.* 2018), while in the control group, it remained near the highest MAF value of the reference population (12%), suggesting a significant association ($P = 0.011$) between the A allele and a higher susceptibility to IDD (Table S1).

The second polymorphism, rs851728071, was an A to G transversion at position CFA8:74 080 057 (CanFam3.1) associated with the protected phenotype ($P = 0.029$; FDR = 0.088). It was detected in 48.3% of the dogs from

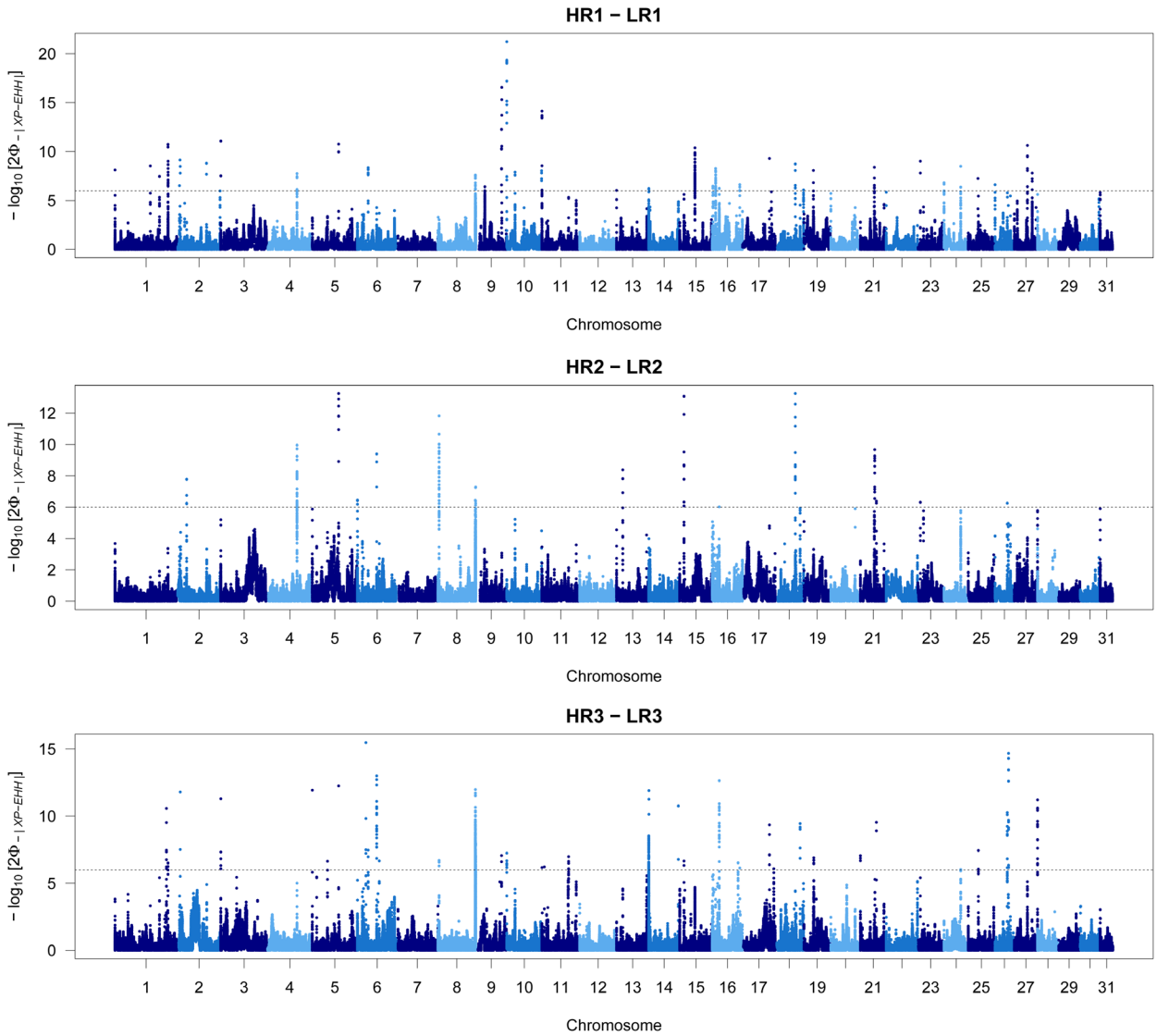

Figure 1 Manhattan plots of genome-wide distribution of selection signatures detected with extended haplotype homozygosity (XP-EHH) for groups of breeds at high risk for developing diabetes mellitus (DM) (HR1–HR3) when compared with groups of breeds at low risk (LR1–LR3). The threshold is set at $-\log_{10}(P\text{-XP-EHH}) = 6$.

the WGS data belonging to breeds at low risk for developing DM, and it is worth noting that all of the carriers were heterozygous. German Shepherd (41.4% of the low-risk carriers) and Golden Retriever (41.4%) were the breeds that contributed the most to the identification of this variant, followed by Beagle, Brittany, Bulldog, Pekingese and Shih Tzu (each of them representing a 3.5% of the low-risk carriers). The variant was predicted to cause a start codon loss of an *ENSCAFG00000043645* transcript (Ensembl transcript ID: ENSCAFT00000049715), which codifies a protein with immunoglobulin variable (V) domains. The immunoglobulin V domains are responsible for providing the specificity to react with the range of foreign antigens, and a large and diverse V gene repertoire is thought to offer the organism a fitness advantage. Nevertheless, not all V genes confer an advantage as some could be auto-reactive (Martin *et al.* 2018). This could be the case, since the reference allele frequency is significantly higher in the breeds at high risk for developing DM, while the start codon loss could be conferring protection in some breeds. This relationship between the polymorphism rs851728071 and the protected phenotype was verified ($P = 0.034$) in the same way as described with the previous variant. In this case, the frequency of the G allele in the control group from this pilot study (50%) duplicated the highest MAF value of the reference population (20%), whereas it was in accordance with this MAF in the diabetic group (19%). The increased frequency of the minor allele (G) in the control group strengthens the idea that this allele confers some degree of protection against IDD, since control animals were old dogs from breeds at high risk for developing DM that did not develop the disease (Table S1).

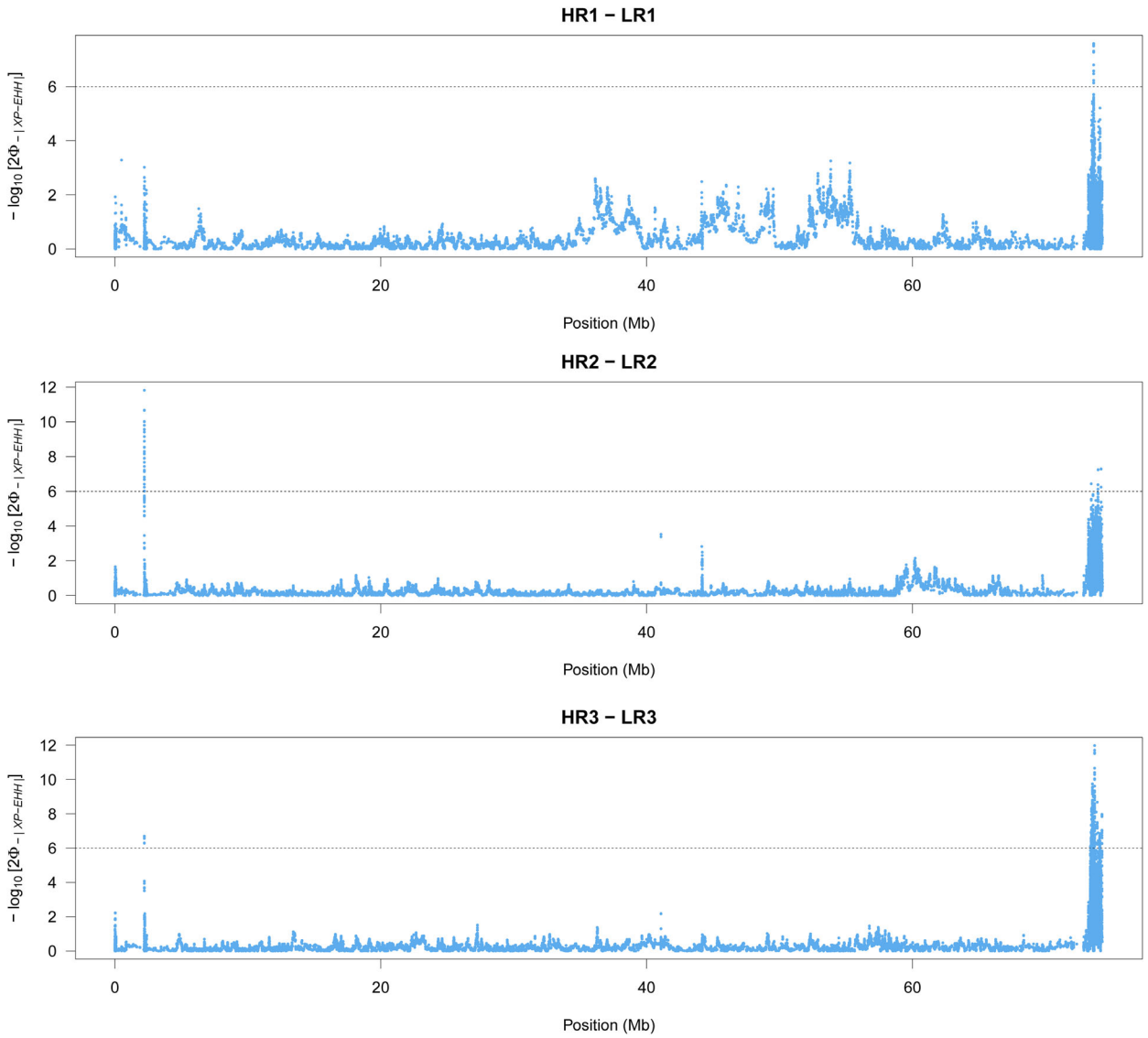

Figure 2 Manhattan plots of the distribution of selection signatures within chromosome 8 detected with XP-EHH for groups of breeds at high risk for developing DM (HR1–HR3) when compared with groups of breeds at low risk (LR1–LR3). The threshold is set at $-\log_{10}(P\text{-XP-EHH}) = 6$.

Recently, Rolim *et al.* (2017) performed a human case–control study and a family-based cohort comprising a total of 240 T1D patients, 172 first-degree relatives and 130 unrelated healthy controls, and they proved that variants at the IGH cluster are genetic determinants of autoantibodies against pancreatic antigens and contribute to T1D susceptibility. Our results are in line with these findings and further support the idea that control of autoantibody generation by polymorphisms in the IGH cluster may be a component of the genetic susceptibility to DM not only in humans, but also in dogs. In addition, the dual action of the IGH cluster is reflected in the results: while some variants may help to promote resistance against diseases by eliminating pathogenic microorganisms or damaged cells, others could lead to increased risk for developing autoimmune disorders as a consequence of the autoantibody production.

The main criterion used here to consider a significant SNP as a candidate causal variant is that it had a potential deleterious effect on the protein. However, it should be noted that non-deleterious variants might also be influencing the pathogenesis of canine DM. In fact, epigenetic modifications have been shown to contribute to human T1D and constitute a promising research area to find alternative ways to control T1D in human medicine (Akil *et al.* 2020). Future research on epigenetic modifications associated with canine DM could uncover novel therapeutic strategies in veterinary medicine as well.

In summary, the identification of a significant association between the IGH cluster and breed predisposition to DM for the first time in dogs provides further support for the existence of a genetic component on the pathogenesis of this disease. In particular, variants at this cluster may

 doi: 10.1111/age.13047

Table 2 Candidate genes included in genomic regions under selection identified with XP-EHH and containing significant variants detected in the GWAS.

Candidate gene	CFA position	Significant variants	Description
ENSCAFG00000030258	8:72 881 377–73 388 517	196	Immunoglobulin heavy constant gamma
ENSCAFG00000044395	8:72 927 338–72 928 918	2	T complex 1 subunit beta
ENSCAFG00000018499	8:73 088 147–73 091 095	1	Disintegrin
ENSCAFG00000018468	8:73 093 030–73 096 291	2	
ENSCAFG00000047346	8:73 152 989–73 159 817	1	Immunoglobulin heavy chain V region
ENSCAFG00000045674	8:73 215 492–73 222 954	4	
ENSCAFG00000030900	8:73 235 300–73 266 913	20	
ENSCAFG00000041353	8:73 386 962–73 466 787	156	
ENSCAFG00000028509	8:73 477 691–73 886 755	709	
ENSCAFG00000045680	8:73 586 711–73 595 785	13	
ENSCAFG00000044861	8:73 837 451–73 956 704	157	
ENSCAFG00000049081	8:73 845 247–73 851 259	23	
ENSCAFG00000045461	8:73 905 454–73 940 065	1	
ENSCAFG00000045080	8:73 980 495–74 052 834	68	
ENSCAFG00000043645	8:74 067 754–74 243 130	168	
ENSCAFG00000049412	8:74 290 728–74 293 206	3	Disintegrin

Table 3 Summary of the polymorphisms with potential effects on the encoded protein structure identified with VEP (Kumar *et al*. 2009).

SNP	Location	Gene (transcript)	Base change	Amino acid change	Protein position	Consequence	SIFT
rs852072969	8:74 291 885–74 291 885	*ENSCAFG00000049412* (ENSCAFT00000070080)	T>A	D>V	365	Missense variant	Deleterious (0)
rs851728071	8:74 080 057–74 080 057	*ENSCAFG00000043645* (ENSCAFT00000049715)	A>G	M>T	1	Start lost	Deleterious (0)

contribute to the generation of autoantibodies in the course of IDD, probably jeopardising the pancreatic function of affected individuals and explaining the development of clinical signs. These results should be interpreted with caution until the functional roles of the proposed variants have been proved in larger studies.

Acknowledgements

CRS is a recipient of a Predoctoral Fellowship for Research Personnel in Training UCM funded by Universidad Complutense de Madrid (Spain) and Banco Santander (CT42/18-CT43/18).

Conflict of interest

The authors of the present study declare that there was no conflict of interest in carrying out this work.

Data availability statement

The complete WGS data obtained from Plassais *et al*. (2019) and used in this study are available via the Short Read Archive (ncbi.nlm.nih.gov/sra; BioProject number: PRJNA448733). GWAS phenotypes and genotypes are available at Figshare (https://doi.org/10.6084/m9.figshare.12988004.v1).

References

Akil A.-S.A.-S., Jerman L.F., Yassin E. *et al*. (2020) Reading between the (genetic) lines: how epigenetics is unlocking novel therapies for type 1 diabetes. *Cells* **9**, 2403.

American Diabetes Association (2014) Diagnosis and classification of diabetes mellitus. *Diabetes Care* **37**, S81–90.

Bakay M., Pandey R., Grant S.F. *et al*. (2019) The genetic contribution to type 1 diabetes. *Current Diabetes Reports* **19**, 116.

Bao Y., Guo Y., Xia S. & Zhao Z. (2010) Molecular characterization of the VH repertoire in *Canis familiaris*. *Veterinary Immunology and Immunopathology* **137**, 64–75.

Cadieu E., Neff M.W., Quignon P. *et al*. (2009) Coat variation in the domestic dog is governed by variants in three genes. *Science* **326**, 150–3.

Catchpole B., Ristic J.M., Fleeman L.M. & Davison L.J. (2005) Canine diabetes mellitus: can old dogs teach us new tricks? *Diabetologia* **48**, 1948–56.

Davison L.J., Herrtage M.E. & Catchpole B. (2005) Study of 253 dogs in the United Kingdom with diabetes mellitus. *Veterinary Record* **156**, 467–71.

Davison L.J., Herrtage M.E. & Catchpole B. (2011) Autoantibodies to recombinant canine proinsulin in canine diabetic patients. *Research in Veterinary Science* **91**, 58–63.

 doi: 10.1111/age.13047

Davison L.J., Walding B., Herrtage M.E. & Catchpole B. (2008a) Anti-insulin antibodies in diabetic dogs before and after treatment with different insulin preparations. *Journal of Veterinary Internal Medicine* **22**, 1317–25.

Davison L.J., Weenink S.M., Christie M.R. *et al.* (2008b) Autoantibodies to GAD65 and IA-2 in canine diabetes mellitus. *Veterinary Immunology and Immunopathology* **126**, 83–90.

Driver J.P., Serreze D.V. & Chen Y.-G. (2011) Mouse models for the study of autoimmune type 1 diabetes: a NOD to similarities and differences to human disease. *Seminars in Immunopathology* **33**, 67–87.

Durinck S., Spellman P.T., Birney E. & Huber W. (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols* **4**, 1184–91.

Fall T., Hamlin H.H., Hedhammar A. *et al.* (2007) Diabetes mellitus in a population of 180,000 insured dogs: incidence, survival, and breed distribution. *Journal of Veterinary Internal Medicine* **21**, 1209–16.

Gautier M., Klassmann A. & Vitalis R. (2017) rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources* **17**, 78–90.

Guptill L., Glickman L. & Glickman N. (2003) Time trends and risk factors for diabetes mellitus in dogs: analysis of veterinary medical data base records (1970–1999). *The Veterinary Journal* **165**, 240–7.

Hess R.S., Kass P.H. & Ward C.R. (2000) Breed distribution of dogs with diabetes mellitus admitted to a tertiary care facility. *Journal of the American Veterinary Medical Association* **216**, 1414–7.

Hoenig M. & Dawe D.L. (1992) A qualitative assay for beta cell antibodies. Preliminary results in dogs with diabetes mellitus. *Veterinary Immunology and Immunopathology* **32**, 195–203.

Holder A.L., Kennedy L.J., Ollier W.E.R. & Catchpole B. (2015) Breed differences in development of anti-insulin antibodies in diabetic dogs and investigation of the role of dog leukocyte antigen (DLA) genes. *Veterinary Immunology and Immunopathology* **167**, 130–8.

Hunt S.E., McLaren W., Gil L. *et al.* (2018) Ensembl variation resources. *Database* **2018**, bay119.

Hwang M.-H., Darzentas N., Bienzle D. *et al.* (2018) Characterization of the canine immunoglobulin heavy chain repertoire by next generation sequencing. *Veterinary Immunology and Immunopathology* **202**, 181–90.

Jayasimhan A., Mansour K.P. & Slattery R.M. (2014) Advances in our understanding of the pathophysiology of Type 1 diabetes: lessons from the NOD mouse. *Clinical Science* **126**, 1–18.

Kennedy L.J., Davison L.J., Barnes A. *et al.* (2006) Identification of susceptibility and protective major histocompatibility complex haplotypes in canine diabetes mellitus. *Tissue Antigens* **68**, 467–76.

Kumar P., Henikoff S. & Ng P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4**, 1073–81.

Lefranc M.-P., Giudicelli V., Ginestoux C. *et al.* (2009) IMGT®, the international ImMunoGeneTics information system®. *Nucleic Acids Research* **37**, D1006–12.

Martin J., Ponstingl H., Lefranc M.-P. *et al.* (2018) Comprehensive annotation and evolutionary insights into the canine (*Canis lupus familiaris*) antigen receptor loci. *Immunogenetics* **70**, 223–36.

Massari S., Bellahcene F., Vaccarelli G. *et al.* (2009) The deduced structure of the T cell receptor gamma locus in *Canis lupus familiaris*. *Molecular Immunology* **46**, 2728–36.

Matiasovic J., Andrysikova R., Karasova D. *et al.* (2009) The structure and functional analysis of canine T-cell receptor beta region. *Veterinary Immunology and Immunopathology* **132**, 282–7.

Mattheeuws D., Rottiers R., Kaneko J.J. & Vermeulen A. (1984) Diabetes mellitus in dogs: relationship of obesity to glucose tolerance and insulin response. *American Journal of Veterinary Research* **45**, 98–103.

Mattin M., O'Neill D., Church D. *et al.* (2014) An epidemiological study of diabetes mellitus in dogs attending first opinion practice in the UK. *Veterinary Record* **174**, 349.

McLaren W., Gil L., Hunt S.E. *et al.* (2016) The ensembl variant effect predictor. *Genome Biology* **17**, 122.

Mineccia M., Massari S., Linguiti G. *et al.* (2012) New insight into the genomic structure of dog T cell receptor beta (TRB) locus inferred from expression analysis. *Developmental & Comparative Immunology* **37**, 279–93.

Moshref M., Tangey B., Gilor C. *et al.* (2019) Concise review: canine diabetes mellitus as a translational model for innovative regenerative medicine approaches. *Stem Cells Translational Medicine* **8**, 450–5.

Noble J.A. & Valdes A.M. (2011) Genetics of the HLA region in the prediction of type 1 diabetes. *Current Diabetes Reports* **11**, 533–42.

Nyaga D.M., Vickers M.H., Jefferies C. *et al.* (2018) The genetic architecture of type 1 diabetes mellitus. *Molecular and Cellular Endocrinology* **477**, 70–80.

O'Kell A.L., Wasserfall C., Catchpole B. *et al.* (2017) Comparative pathogenesis of autoimmune diabetes in humans, NOD mice, and canines: has a valuable animal model of type 1 diabetes been overlooked? *Diabetes* **66**, 1443–52.

Ounissi-Benkalha H. & Polychronakos C. (2008) The molecular genetics of type 1 diabetes: new genes and emerging mechanisms. *Trends in Molecular Medicine* **14**, 268–75.

Plassais J., Kim J., Davis B.W. *et al.* (2019) Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nature Communications* **10**, 1–14.

Pociot F. & McDermott M.F. (2002) Genetics of type 1 diabetes mellitus. *Genes & Immunity* **3**, 235–49.

Precone V., Del Monaco V., Esposito M.V. *et al.* (2015) Cracking the code of human diseases using next-generation sequencing: applications, challenges, and perspectives. *BioMed Research International* **2015**, 161648.

Purcell S., Neale B., Todd-Brown K. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–75.

R Core Team (2020) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Robinson J.T., Thorvaldsdóttir H., Winckler W. *et al.* (2011) Integrative genomics viewer. *Nature Biotechnology* **29**, 24–6.

Rolim I., Duarte N., Barata G. *et al.* (2017) Immunoglobulin M gene association with autoantibody reactivity and type 1 diabetes. *Immunogenetics* **69**, 429–37.

Sabeti P.C., Varilly P., Fry B. *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–8.

Sai P., Debray-Sachs M., Jondet A. *et al.* (1984) Anti-beta-cell immunity in insulinopenic diabetic dogs. *Diabetes* **33**, 135–40.

Sambrook J. & Russell D.W. (2006) Purification of nucleic acids by extraction with phenol:chloroform. *Cold Spring Harbor Protocols* **2006**, pdb.prot4455.

Short A.D., Catchpole B., Kennedy L.J. *et al.* (2007) Analysis of candidate susceptibility genes in canine diabetes. *Journal of Heredity* **98**, 518–25.

Short A.D., Catchpole B., Kennedy L.J. *et al.* (2009) T cell cytokine gene polymorphisms in canine diabetes mellitus. *Veterinary Immunology and Immunopathology* **128**, 137–46.

Short A.D., Saleh N.M., Catchpole B. *et al.* (2010) CTLA4 promoter polymorphisms are associated with canine diabetes mellitus. *Tissue Antigens* **75**, 242–52.

Tandon N. (2015) Understanding type 1 diabetes through genetics: Advances and prospects. *Indian Journal of Endocrinology and Metabolism* **19**, S39–43.

Thorvaldsdóttir H., Robinson J.T. & Mesirov J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178–92.

Wiles B.M., Llewellyn-Zaidi A.M., Evans K.M. *et al.* (2017) Large-Scale survey to estimate the prevalence of disorders for 192 kennel Club registered breeds. *Canine Genetics and Epidemiology* **4**, 8.

Wilkinson J.S. (1960) Spontaneous diabetes mellitus. *Veterinary Record* **72**, 548–53.

Ye J., Coulouris G., Zaretskaya I. *et al.* (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134.

Yoon S., Fleeman L.M., Wilson B.J. *et al.* (2020) Epidemiological study of dogs with diabetes mellitus attending primary care veterinary clinics in Australia. *Veterinary Record* **187**, e22.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Distribution plots of raw XP-EHH values from comparisons of groups HR1–LR1 (C1), HR2–LR2 (C2) and HR3–LR3 (C3).

Figure S2. Q–Q plot of the GWAS data.

Figure S3. Haplotype bifurcation plots for rs852072969 and rs851728071 variants.

Table S1. Characteristics of breed, sex, neuter status, age and genotypes of dogs included in the pilot study to validate the results.

Table S2. Detail on the genomic regions under positive selection detected with XP-EHH analysis for groups of breeds at high risk for developing DM (HR1–HR3) when compared with groups of breeds at low risk (LR1–LR3).

Table S3. Results of the GWAS analysis comparing breeds at low and high risk for developing DM for the identification of variants associated with breed predisposition to DM.