



**HAL**  
open science

## Genome-wide analyses reveals an association between invasive urothelial carcinoma in the Shetland sheepdog and NIPAL1

Heidi G Parker, Alexander C Harris, Jocelyn Plassais, Deepika Dhawan, Erika M. Kim, Deborah Knapp, Elaine Ostrander

### ► To cite this version:

Heidi G Parker, Alexander C Harris, Jocelyn Plassais, Deepika Dhawan, Erika M. Kim, et al.. Genome-wide analyses reveals an association between invasive urothelial carcinoma in the Shetland sheepdog and NIPAL1. *npj Precision Oncology*, 2024, 8 (1), pp.112. 10.1038/s41698-024-00591-0 . hal-04596540

**HAL Id: hal-04596540**

**<https://hal.science/hal-04596540>**

Submitted on 31 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

<https://doi.org/10.1038/s41698-024-00591-0>

# Genome-wide analyses reveals an association between invasive urothelial carcinoma in the Shetland sheepdog and *NIPAL1*

Check for updates

Heidi G. Parker<sup>1</sup>, Alexander C. Harris<sup>1</sup>, Jocelyn Plassais<sup>2</sup>, Deepika Dhawan<sup>3</sup>, Erika M. Kim<sup>4</sup>, Deborah W. Knapp<sup>3,5</sup> & Elaine A. Ostrander<sup>1</sup> ✉

Naturally occurring canine invasive urinary carcinoma (iUC) closely resembles human muscle invasive bladder cancer in terms of histopathology, metastases, response to therapy, and low survival rate. The heterogeneous nature of the disease has led to the association of large numbers of risk loci in humans, however most are of small effect. There exists a need for new and accurate animal models of invasive bladder cancer. In dogs, distinct breeds show markedly different rates of iUC, thus presenting an opportunity to identify additional risk factors and overcome the locus heterogeneity encountered in human mapping studies. In the association study presented here, inclusive of 100 Shetland sheepdogs and 58 dogs of other breeds, we identify a homozygous protein altering point mutation within the *NIPAL1* gene which increases risk by eight-fold (OR = 8.42, CI = 3.12–22.71), accounting for nearly 30% of iUC risk in the Shetland sheepdog. Inclusion of six additional loci accounts for most of the disease risk in the breed and explains nearly 75% of the phenotypes in this study. When combined with sequence data from tumors, we show that variation in the MAPK signaling pathway is an overarching cause of iUC susceptibility in dogs.

Bladder cancer is the sixth most common human cancer in the United States, accounting for over 80,000 new cases per year (<http://globocan.iarc.fr>)<sup>1</sup>. Ninety percent of bladder cancers are urothelial carcinomas (UC), and while the noninvasive form is most common, approximately 25–30% are muscle invasive (MIBC) at the time of diagnosis<sup>2,3</sup>. Distant metastases are present in 5% of all cases at diagnosis, a number that ultimately grows to 50% for muscle-invasive disease<sup>2,4,5</sup>. Despite improvements in treatment, such as immune-based therapy and new methods for early diagnosis, the five-year survival rate for MIBC patients is only 58%, and 5% if the disease has metastasized. The risk of developing bladder cancer in humans can be attributed to several factors including smoking<sup>6,7</sup>, pollutants<sup>8,9</sup>, and genetic risk variants<sup>10,11</sup>.

Domestic dogs develop cancer naturally as they age, just as humans do, are diagnosed as symptoms arise, and are treated with surgery, radiation,

and/or chemotherapy. The increased risk of specific cancers observed in subsets of breeds indicates the presence of inherited mutations of high penetrance that are conducive to genome wide association studies (GWAS). The reasons for this are several-fold. Importantly, breed structure requires that any dog who is a registered member of a breed must be descended from registered members of the same breed, thus forming closed populations segregating limited numbers of mutations. Also, most breeds feature popular sires whose genomes are over-represented in the breed population, further ensuring that small numbers of variants are likely responsible for increased susceptibility to a particular cancer. Together, this means that canine cancer genetics represents a unique approach for dealing with the problem of locus heterogeneity that is so common in genetic studies of human cancer susceptibility. Finally, canine cancers are not induced. This ensures that the loci identified contribute to susceptibility and disease

<sup>1</sup>Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Center, National Institutes of Health, Bethesda, MD, USA. <sup>2</sup>Institut de Génétique et Développement de Rennes, CNRS-UMR6290, University of Rennes, 35000 Rennes, France. <sup>3</sup>Department of Veterinary Clinical Sciences, College of Veterinary Medicine, Purdue University, West Lafayette, IN, USA. <sup>4</sup>Center for Biomedical Informatics & Information Technology, National Cancer Institute, National Institutes of Health, Frederick, MD, USA. <sup>5</sup>Purdue University Center for Cancer Research, West Lafayette, IN, USA. ✉e-mail: [eostrand@mail.nih.gov](mailto:eostrand@mail.nih.gov)

development rather than other aspects of the tumor life cycle. Indeed, while other approaches, such as those based on murine/PDX, knock-out mice, CRISPR generated mutation driven tumors, etc., exist for studies of tumor metastasis, growth rate, clonality, and treatment response, the canine system proves most useful in revealing the genetic underpinnings of disease susceptibility, especially for traits where multiple genes contribute differentially, such as cancer.

Invasive urinary carcinoma (iUC) is the most common cancer of the urinary bladder in dogs, accounting for 90% of bladder cancers, with 40,000 newly diagnosed cases a year<sup>12,13</sup>. Naturally occurring canine iUC closely mimics human high grade MIBC in terms of age of onset, pathology, cellular and molecular features, metastatic profiles, and treatment response<sup>14</sup>. Approximately 50% of affected dogs will experience metastases to distant organs<sup>15</sup> and median survival times are typically < one year, although dogs receiving multiple sequential protocols may live beyond a year<sup>16,17</sup>. As with humans, there are multiple environmental factors that reportedly increase risk of canine iUC<sup>18</sup>. Approximately 80% of canine iUC tumors carry a somatic BRAF<sup>V595E</sup> mutation, which has been shown to increase activity of the MAPK signaling pathway<sup>19,20</sup>. Similarly, more than one third of human MIBC tumors carry activating mutations in the MAPK pathway<sup>21</sup>. Tumor cells treated with therapeutics that target a specific somatic mutations can develop resistance through multiple mechanisms (reviewed in<sup>22</sup>). For example, in response to MAPK pathway inhibitors that target the BRAF<sup>V600E</sup> mutation, tumor cells have been found to reactivate the MAPK pathway through mutations in the tyrosine kinase receptors or RAS genes or by duplicating and further mutating the BRAF gene<sup>23,24</sup>. Treatment options are improved when the pathway is treated as a whole.

Amongst purebred dogs, there is a subset of breeds at increased risk for developing iUC, including the Scottish terrier (OR = 21.1, 95% CI = 16.2–27.5), West Highland white terrier (OR = 5.8 95% CI = 4.2–8.1), and Shetland sheepdog (OR = 6.1, 95% CI = 4.8–7.7)<sup>12</sup>. Shetland sheepdogs (Shelties), the focus of this study, are a small herding breed and popular family pet. Multiple studies demonstrate their extraordinary disease risk<sup>12,17,25</sup>, with one survey reporting that of 3359 Shelties (<https://ofa.org/about/health-surveys/?breed=SS>), iUC comprised 12% of all Sheltie cancer diagnoses, which is six times the expected rate of 2% for all dogs<sup>13</sup>. Of note, dogs are more likely to get invasive disease rather than non-invasive disease and tumors are typically stage T2 or higher at diagnoses<sup>25–27</sup>. These facts, combined with the high level of veterinary scrutiny dogs receive, are additional attributes making domestic dog breeds an ideal system for iUC genetic mapping studies.

In this study we analyze one hundred Shelties and 55 dogs from closely related breeds and identify a single major risk locus associated with iUC, featuring *NIPA-like domain containing 1* (*NIPALI*), a strong candidate gene, and associated mutations. By segregating samples according to genotype, we identify six additional loci that modify individual disease risk when combined with a putative pathogenic protein altering mutation in *NIPALI*. Merging the GWAS loci with tumor sequence identifies the mitogen activated protein kinase (MAPK) signaling pathway as a key component of canine iUC genetic susceptibility.

## Results

### Genome-wide association study

Affected Shelties were submitted to the study through the Purdue University Veterinary Hospital or directly from the owners. Samples from Purdue had complete pathology including tumor grade and stage data, which confirmed that 92.3% of affected dogs had high grade (3–4) tumors at the time of diagnosis, and 44% had distant metastases at the time of death. Prior to the association analysis, we tested the samples for population substructure as well as correct breed assignment, disqualifying four dogs from breed-specific analyses and three dogs from all analyses (Fig. 1a and Supplementary Figs. 1 and 2). PCA showed clustering that appeared to correlate with affection status within the Sheltie cohort. One-way ANOVA confirmed that PC2 correlated with disease status ( $P = 0.0046$ , Fig. 1b). We therefore corrected for population substructure in both the within and across breed analyses.

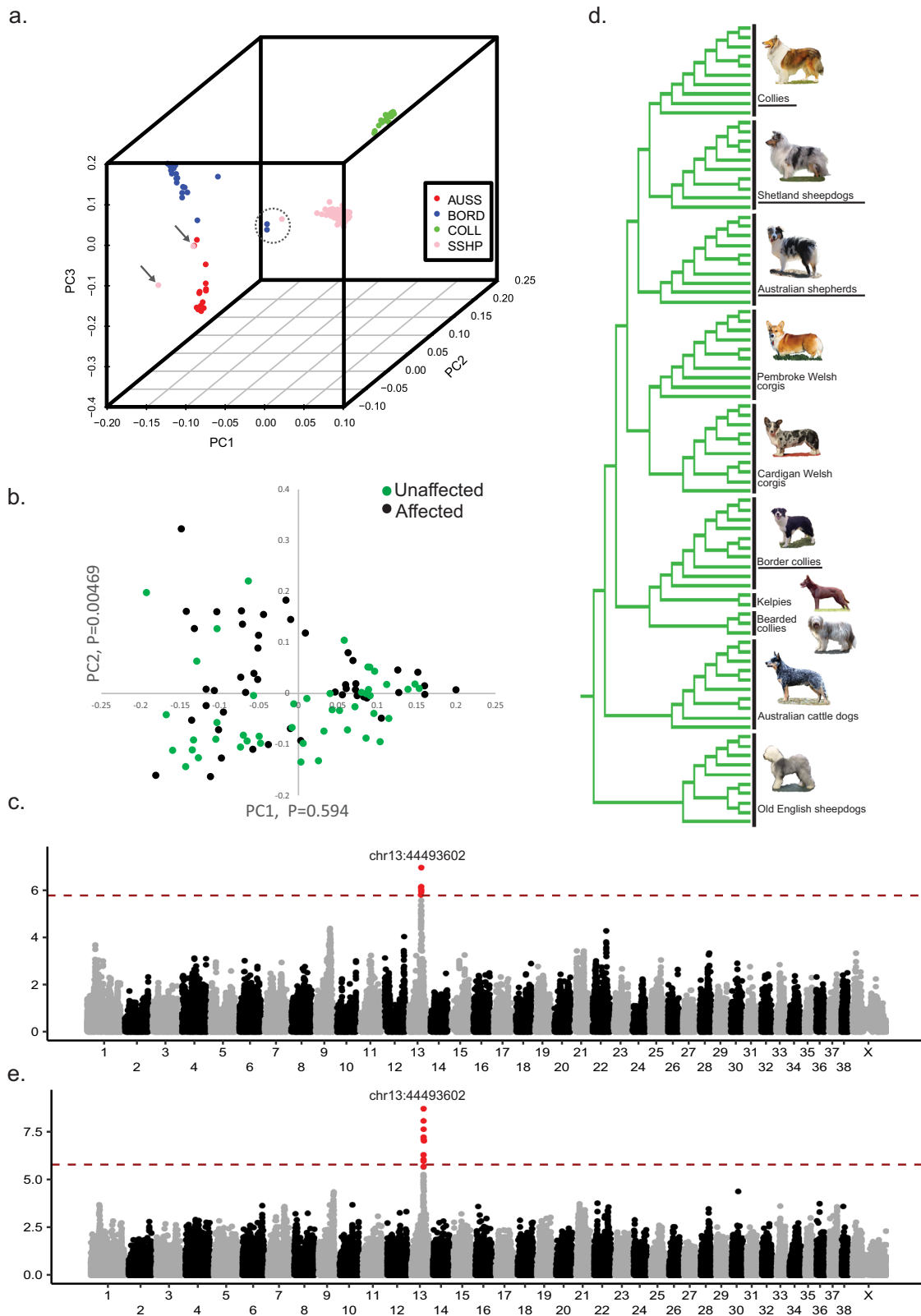
The association analysis of 49 affected Shelties and 51 unaffected Shelties, aged 10 years and older (Table 1) revealed a single locus tagged by three SNPs, chr13: 44493602, 44508476, and 44520164, which was significantly associated with iUC ( $P = 1.05 \times 10^{-7}$ ). The result remained significant after 10,000 permutations ( $P\text{-perm}_{0.05} = 1.37 \times 10^{-6}$ ) (Fig. 1c, Q-Q plot in Supplementary Figure 3). The GWAS was repeated after adding an additional 21 affected and 34 unaffected dogs from three related breeds: Collie, Border collie, and Australian shepherd (Fig. 1d). These breeds were chosen because they are closely related to Shetland sheepdogs<sup>28</sup>, share other deleterious mutations<sup>29,30</sup> and we had a sufficient number of affected dogs on hand. The addition of other breeds that share association within the same region of the genome has been shown to greatly reduce the size of associated haplotypes by providing additional recombination events, thus improving the chance to find a shared disease associated mutation<sup>29,31–33</sup>. The same three SNPs were associated with iUC in the multi-breed cohort,  $P = 5.36 \times 10^{-10}$  (Fig. 1e, Q-Q plot in Supplementary Figure 3).

We calculated linkage disequilibrium (LD) from the peak SNP to all others on the chromosome and identified a region between base-pairs (bp) 42943427 and 45635207 that included markers with a pairwise  $r^2 \geq 0.6$  to the peak SNPs and an association with iUC of  $P \leq 1 \times 10^{-5}$  (Fig. 2). This locus spans a region of significant dysregulation previously identified in an RNAseq analysis of multiple iUC tumors<sup>34</sup> (Supplementary Fig. 4). Genotypes were phased across the locus and haplotypes estimated. More than 25% of the affected Shelties were homozygous for a single haplotype across the entire 2.69 Mb region. A core haplotype was identified between two sites of predicted recombination (43459581 and 44717379) (Fig. 2) that was found in 40 of the 49 affected dogs, ~60% of which were homozygous. This haplotype was also the most common across all Shelties, accounting for 53% of total haplotypes: 68% in cases and 39% in controls, (OR 2.61, CI (1.01–6.66),  $\chi^2 P = 3.20 \times 10^{-15}$ ). The core haplotype was also homozygous in five of six affected Collies.

### Genome sequence analysis

We sequenced the genomes of seven affected and five aged, unaffected Shelties (Supplementary Table 1). These were aligned to CanFam3.1 and UU\_GSD1.0 reference sequences, and genotypes were called in conjunction with dogs from multiple, diverse breeds to test for possible causative mutations. The samples were chosen based on their haplotype across the chr13 associated locus. Using a combination of three canine gene annotation models, a list of curated enhancer regions derived from 131 human biosamples<sup>35</sup>, and ChIPseq data from two canine iUC cell lines, we assessed the single nucleotide variants (SNVs) on the core-associated haplotype for possible functionality (Fig. 3). There were 1968 variants that displayed the same allele distribution as the peak SNPs from the GWAS. Of these, 115 were in regulatory or non-coding elements and 47 were in protein coding transcripts. These could be further categorized into 27 variants in three-prime and five-prime UTRs, 12 synonymous variants, and eight missense variants (Supplementary Table 2, complete list can be found at [https://research.nhgri.nih.gov/dog\\_genome/data\\_release/index.shtml](https://research.nhgri.nih.gov/dog_genome/data_release/index.shtml)). Four of the missense variants were in non-conserved regions of genes and therefore less likely to be damaging. The sites of the four remaining variants were highly conserved between dog and human in both genomic sequence and protein sequence therefore they were converted to human coordinates and assessed for potential pathogenicity using 11 different variant effect predictors. Only one of the coding mutations, a G > A missense mutation at position 43897196/CanFam3.1 (chr13:45186252/UU\_GSD1.0), was predicted to be potentially pathogenic. It encoded a G256D change in the *NIPALI* gene. This highly conserved amino acid is within the sixth transmembrane domain of the protein and is predicted to disrupt the helix structure ( $P = 6.0 \times 10^{-3}$  MutPred2) (Supplementary Fig. 5). Excluding the breeds analyzed in this study, this mutation was found in only one dog from a data set of 1330 for which WGS was available, and in zero out of 1894 dogs published by the Dog10K consortium<sup>36</sup>.

Because our previous study showed that this region harbors numerous upregulated genes, including *NIPALI*, we also evaluated putative regulatory



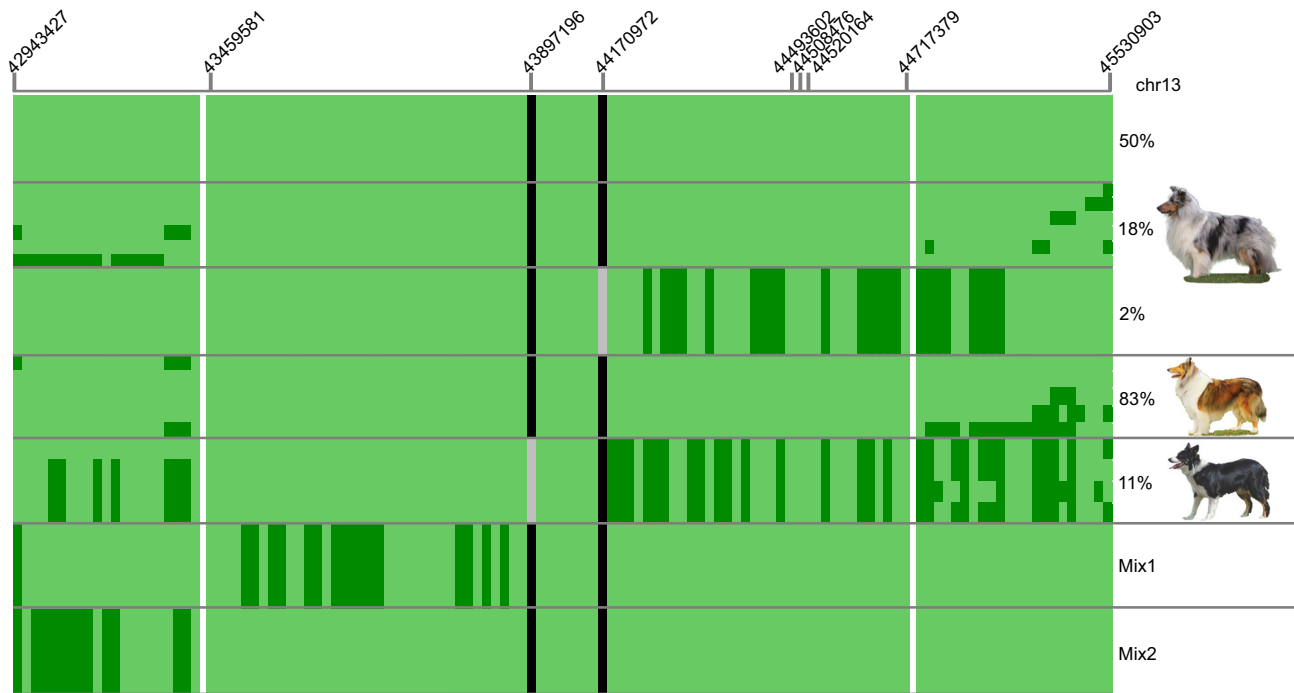
**Fig. 1 | Sample choice and results from GWAS of iUC in Shelties and other herding breeds. a** PC plot of four herding breeds: Sheltie, Collie, Border collie and Australian shepherd. Arrows indicate mis-assigned dogs that were removed from the study. Arrows indicate mis-assigned dogs that were removed from the study. Circle encloses three samples in between breeds determined atypical of their reported breed. For clarity, Supplementary Figure 1 shows two-dimensional plots of the same three PCs in all three combinations. **b** PC plot of only Shetland sheepdogs with iUC cases and controls highlighted.  $P$ -values for correlation with disease status are given along the axes. **c** Manhattan plot showing association values of 90,530 SNPs with iUC in 100 Shetland sheepdogs.

**d** Breed tree showing the breeds that form the UK herding clade from (Parker et al. 2017) with breeds used in the GWAS analysis underlined. **e** GWAS of 156 herding dogs from the four related breeds. The x-axis comprises SNP positions from the beginning of chr1 through the end of the X chromosome. Shading of black and gray shows chromosome boundaries. The y-axis displays the  $-\log$  of the  $p$ -value. The red line indicates the Bonferroni significance estimate,  $1.7 \times 10^{-6}$ . Position of the top SNP at each significant locus is indicated above it. SNPs with suggestive  $p$ -value less than  $1 \times 10^{-5}$  are highlighted in red.

**Table 1 | Samples included in GWAS analyses**

Breed	Affected			Unaffected		
	Number	Age <sup>1</sup> (range)	Sex (%F)	Number	Age <sup>1</sup> (range)	Sex (%F)
Shetland sheepdog	49	11 (5–16)	58%	51	12 (10–16)	57%
Collie	6	10 (6–11)	29%	12	4 (0.1–12)	50%
Border collie	12	10 (7–13)	73%	12	8 (1–14)	45%
Australian shepherd	4	10 (10)	100%	10	8 (3–12)	50%

<sup>1</sup>Age is given as the average across all samples in the group.



**Fig. 2 | iUC associated haplotypes across the chr13 locus.** The haplotypes comprise 120 markers represented by vertical bars. White vertical lines indicate the boundaries of the core haplotype. Alleles present in the Sheltie iUC associated haplotype are light green and the alternate allele is dark green. The putative causative variants are shaded black when present on the haplotype or grey when

not present. The percent in affected dogs of each haplotype, or cluster of haplotypes sharing the same core, is listed on the left followed by the image of the breed top to bottom: Shetland sheepdog, Collie, Border collie. Mix 1 and 2 are atypical Border collies 1 and 2, respectively.

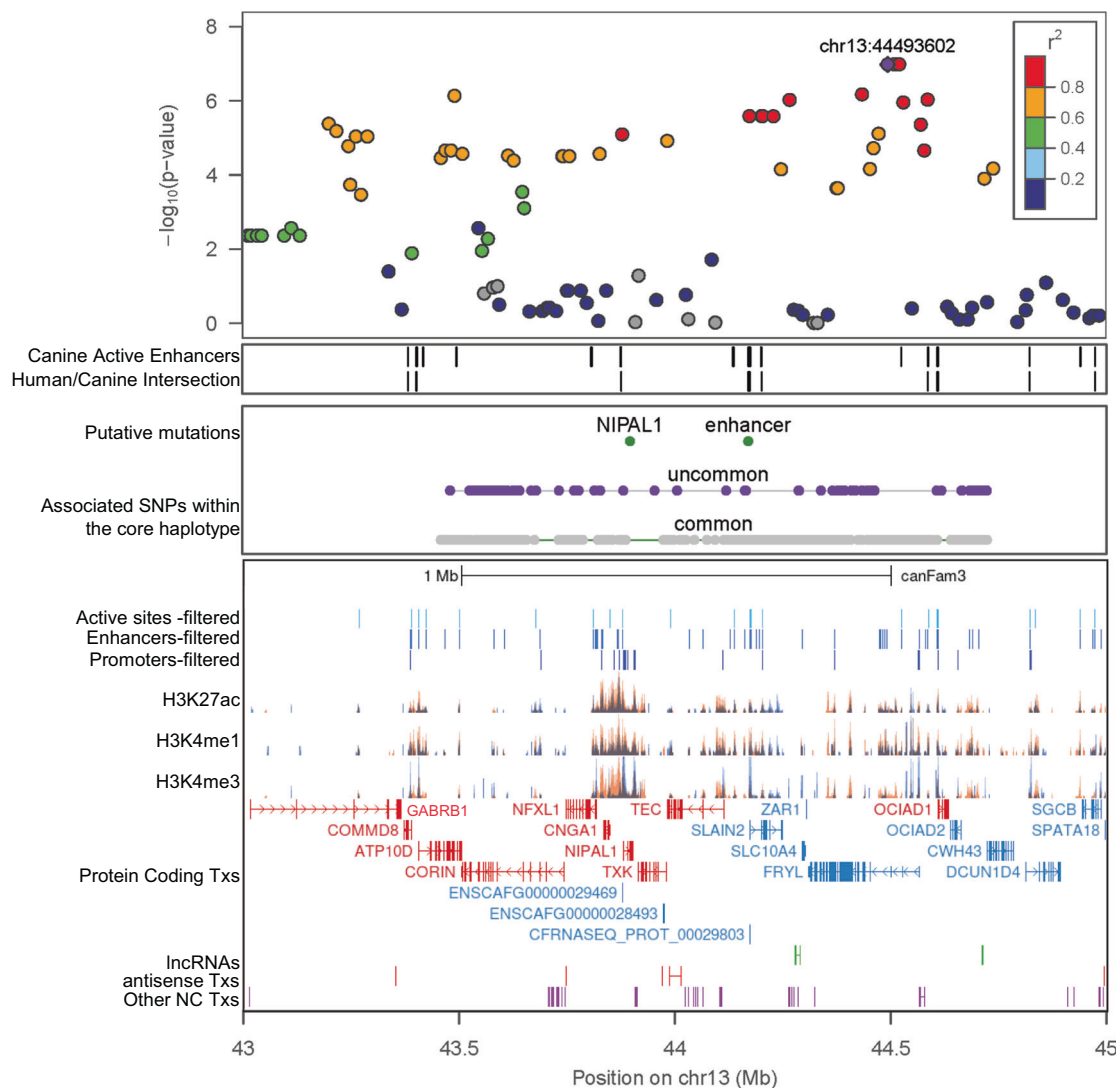
variants. We compared the sites of known human enhancers to the active enhancer regions predicted from ChIP-seq in canine iUC cell lines (Fig. 3). We identified 19 variants that were within sites corresponding to active enhancers present in both species that, in humans, are predicted to affect multiple genes (Supplementary Table 2). Because markers across this region had identical genotypes in the 12 sequenced dogs, we assessed the frequencies of the iUC associated alleles in other dog breeds. Among the 1237 dogs and 93 wolves, twelve of the 19 iUC associated enhancer alleles were major alleles in the general population and another six had allele frequencies > 0.34 in dogs and > 0.24 in wolves. One variant, located at chr13:44170972, 1260 bp upstream of the *SLAIN2* gene, had an allele frequency of only 0.05 among dog breeds not in this study, and was not found in wolves. The change of alleles at this variant, chr13:44170972 G > A, results in the creation of a binding site for 14 different transcription factors or transcription factor complexes, as ascertained using two different prediction methods (Supplementary Table 3). Increased affinity for any of these transcription factors to the region could lead to dysregulation of local genes.

We sequenced the two mutations within the chr13 locus in 47 affected and 47 unaffected Shelties and 150 dogs from related breeds, including 34 with iUC. The missense mutation at chr13:43897916, which we refer to as

NIPAL<sup>G256D</sup>, was found in 85% of the affected Shelties and was homozygous in 60%. The odds ratio for developing iUC with at least one copy of the mutation was 2.95 (CI 1.08–8.05), and if homozygous for the mutation, the OR increased to 8.42 (CI 3.12–22.71). The mutation within the putative multi-gene enhancer was found in 83% of affected Shelties and was homozygous in 57%. The OR for having the disease with at least one copy of the enhancer mutation was 2.76 (CI 1.05–7.26), and in the homozygous state, OR = 7.71 (CI 2.87–20.75). Recombination between these two variants was observed in only two Shelties, one affected and one unaffected; both carried NIPAL1<sup>G256D</sup> but not the enhancer variant. These risk ratios are much higher than any described for variants associated with human bladder cancer.

The NIPAL<sup>G256D</sup> variant was also found in seven affected Collies and two affected Border collies. The enhancer variant was found in the same nine dogs and in two additional affected Border collies. Neither variant was found in the other related breeds (Table 2).

We noted that the two affected Border collies who carried NIPAL1<sup>G256D</sup> were highlighted in the population analyses as atypical Border collies because they did not cluster with the other members of the breed (Fig. 1a and Supplementary Figs. 1 & 2). To determine if these two dogs might be



**Fig. 3 | Detailed graph of the associated locus on chr13.** LD between the top SNP and all others in the region is indicated by the color of the markers with red indicating highest value. Positions of predicted active regulatory regions are in the second panel. The third panel contains positions of the associated SNPs within the shared haplotype with the presumed functional SNP highlighted in green. The

bottom panel contains regulatory regions from ChIPseq as well as gene annotations. Each gene is represented by the longest transcript from the Broad Improved Canine Annotation v.1 (UCSC). The components are plotted against the CanFam3.1 genome using custom tracks in the UCSC browser (<http://genome.ucsc.edu>). Genes highlighted in red were upregulated in iUC tumors<sup>34</sup>.

admixed, we phased their genomes and calculated haplotypes predicted to be identical-by-descent (IBD) with the dogs from 250 breeds in the 1330 dataset, plus the breeds from this study. Both dogs displayed the highest amount of IBD haplotype sharing with Shelties and a unique mixture of other breeds (Supplementary Fig. 6). Neither showed any Border collie ancestry beyond what would be expected for a Sheltie. In addition to the two mutations, both carried the Sheltie iUC-associated core haplotype. This haplotype was not found in any purebred Border collies. This data, therefore, identifies NIPAL1<sup>G256D</sup> in Shelties, Collies and in affected mixed breed dogs with Sheltie ancestry.

Though NIPAL<sup>G256D</sup> was not found in purebred Border collies, the enhancer variant was. We phased the chr13 locus in a total of 56 Border collies and genotyped the two iUC associated variants. We found 17 Border collies that carried the enhancer variant and none carried NIPAL<sup>G256D</sup>. Interestingly, all Border collies with the enhancer mutation carried a haplotype that matched the Sheltie iUC-associated haplotype between chr13:43453329 and 44174029, which is approximately half of the core haplotype including the sites of both putative disease-causing variants. This suggests NIPAL1<sup>G256D</sup> appeared on a haplotype that carried the enhancer mutation, possibly after the division of the breeds in the late 1800s.

### Additional GWAS

The alleles on chr13 contribute to a substantial portion (16–41% in this dataset) of iUC cancer risk in Shelties. However, in a group of 87 healthy Shelties of all ages, some of which will likely develop cancer as they age, the NIPAL1<sup>G256D</sup> allele frequency was 0.41, a larger percentage than we would expect for iUC. The difference between cases and controls lies in the frequency of homozygous mutations. Approximately 60% of the affected dogs are homozygous for the NIPAL1 mutation compared to <15% of the unaffected dogs.

Though homozygosity plays a large role in the risk conferred by NIPAL1<sup>G256D</sup>, there are undeniably additional contributing loci. To locate genomic regions that moderate risk within the Shelties, we performed three additional GWAS, grouping dogs based on their genotype at chr13:43897196: all dogs that are homozygous for the mutation, those that are heterozygous, and those that do not carry the mutation (Table 2). Subsetting the dataset in this manner removes the contribution of the chr13 mutations, thus revealing additional risk loci. These three analyses, for which cases and controls have the same genotype at NIPAL1<sup>G256D</sup>, revealed six additional loci on five different chromosomes associated with iUC ( $P < 2.06 \times 10^{-6}$ ) (Fig. 4).

Calculating LD and from the peak SNP at each locus produced associated regions ranging from 100 kbp to 2.5 Mbp (Supplementary Table 4). The loci on chr1 and 28 were gene poor. The locus on chr4 lies within an F-box gene, *FBXL7*, which plays a role in cell cycle control, and the locus surrounding chr21:34085269 comprises four protein coding genes, three of which have a role in cancer growth. This locus overlaps the associated locus surrounding chr21:33595359, but the two tagging markers, chr21:34085269 and 33595359, are not in LD ( $r_2 = 0.047$ ). The second locus spans more than two megabases and encompasses 30 genes. The locus surrounding chr9:22495330 encompasses 26 genes (Supplementary Table 4).

**Table 2 | Genotypes of putative causative mutations within dogs of each breed**

		43897196			44170972		
		AA	AG	GG	AA	AG	GG
Shetland sheepdog	Affected	28	12	7	27	12	8
	Un-Affected	7	24	16	7	23	17
	Gen Pop <sup>1</sup>	20	27	35	22	27	33
Collie	Affected	7	0	1	7	0	1
	Gen Pop <sup>1</sup>	5	13	14	4	14	14
	Gen Pop >10	0	1	1	0	1	1
Border collie	Affected <sup>2</sup>	0	0	11	0	2	9
	Gen Pop <sup>1</sup>	0	0	61	2	16	43
	Gen Pop >10	0	0	17	0	4	13
Sheltie Mix	Affected	0	2	0	0	2	0
Other related <sup>3</sup>	both	0	0	36	0	0	36

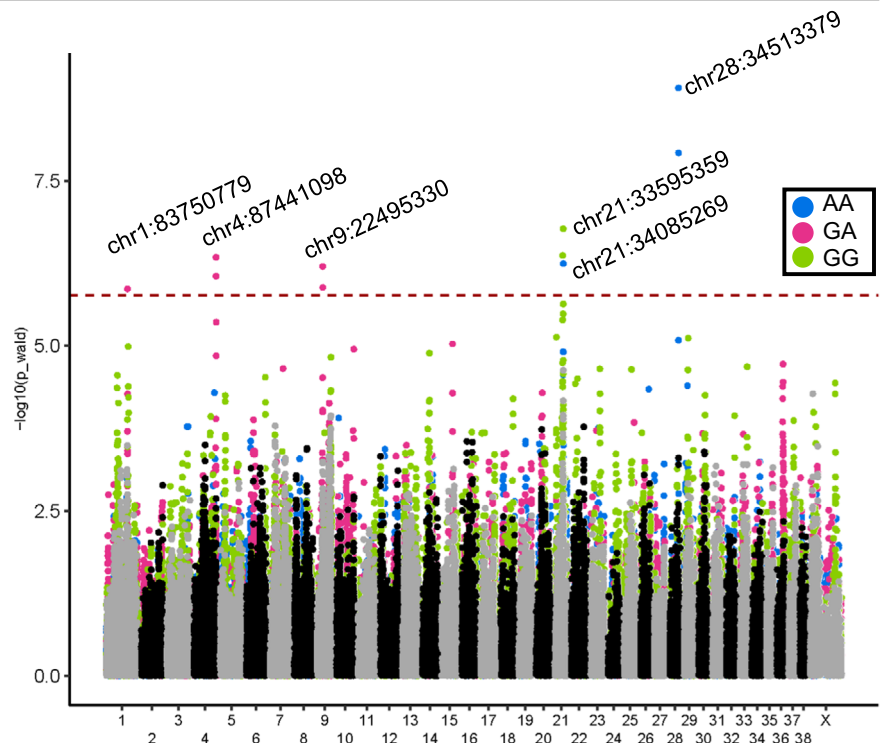
<sup>1</sup>Gen Pop or general population dogs are all ages and were not diagnosed with iUC at the time of collection. Gen Pop > 10 are general population dogs greater than 10 years old. <sup>2</sup>Two of the affected Border collies were determined to be Shetland sheepdogs mixes reducing the Border collie affected counts. <sup>3</sup>Other related breeds (13 affected and 23 unaffected) include Australian shepherd, Bearded collie, Pembroke Welsh corgi, Australian cattle dog.

By calculating the OR for the multiple genotype combinations, we determined the effect each locus is likely to have on disease risk in Shelties (Fig. 5, Table 3). We find that homozygosity of the major allele for at least two of the chr1: 83750779, chr4: 87441098 or chr9: 22495330 loci offers a protective effect to individuals that are heterozygous for the *NIPAL1*<sup>G256D</sup> variant, while the opposite, carrying the minor allele at two of the loci, increases the odds of developing iUC (Table 3). The other loci provide only risk or protection, not both. These combined data provide a platform for identifying interacting genes/mutations and common pathways that lead to cancer susceptibility in Shelties and are foundational in continued searches for iUC disease-associated mutations in other breeds.

**Somatic mutation**

Two of the supporting loci, one at chr9:22381987-23201516 and the other on chr21: 31405230-33951574 include the oncogene *ERBB2* and a tumor suppressor *WEE1*, respectively. These genes are frequently mutated and/or dysregulated in human bladder cancers. We did not find germline mutations in the coding regions of either *WEE1* or *ERBB2* in the WGS of seven Sheltie cases. However, when we analyzed the sequence of these two genes in three canine iUC whole tumor sequences and 15 previously published whole transcript sequences, we found somatic mutations in each gene. Specifically, after applying multiple variant effect predicting programs, the *ERBB2* gene had two putatively damaging missense variants, chr9:22775561 C > A and chr9:22775767 C > G, each found in two tumor alignments, which create the D277Y and D251H protein changes in *ERBB2*. There was also a single putatively damaging C > G variant in the *WEE1* gene at chr21:32843843, leading to a predicted P175R protein change (Supplementary Table 5). A second predicted damaging variant at chr21:32854760, V468A within the kinase domain of *WEE1*, was identified in another tumor however evidence from RNAseq was not sufficient to rule out error and further sequencing is required to confirm. An additional mutation affecting *ERBB2*, G292R, was found in one canine transcript. This germline mutation creates a G292S change in humans which has slightly reduced effect predictions than the dog variant. Each *WEE1* variant was found in a single tumor while the *ERBB2* variants were each found in two different tumors for a minimum putatively damaging

**Fig. 4 | GWAS results from the Shetland sheepdog cohort divided by *NIPAL1* variant genotype.** The results from three independent GWAS are combined for comparison. Each is colored by genotype at *NIPAL1*<sup>G256D</sup> according to the key provided. The significance line is set at  $1.7 \times 10^{-6}$ .



gene mutation frequency of 6% and 24% respectively. Additional coding variants with benign predictions were found in each of these genes.

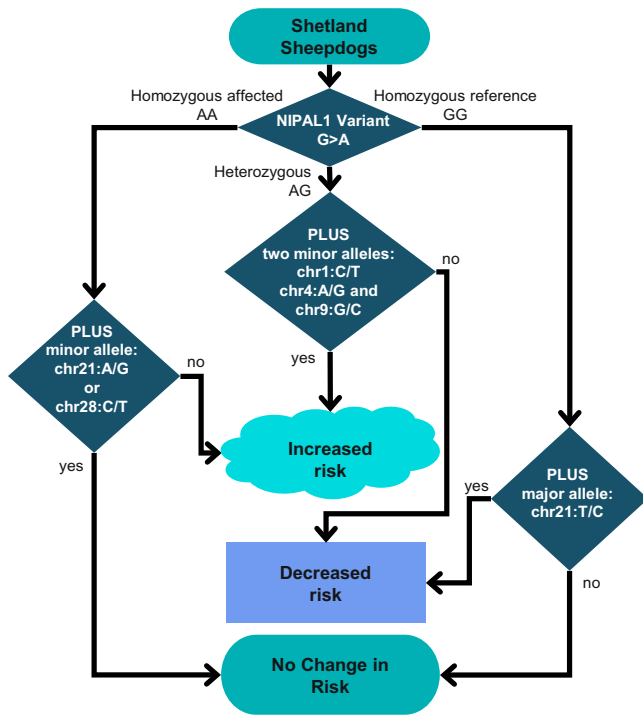
**Discussion**

There are over 350 recognized dog breeds in the world, each a unique population whose members share common ancestors, traits, and disease susceptibilities. Often multiple breeds will develop from a common working population due to geographic separation or regional preferences in size, coat color or working style. The Collie, Border collie, and Australian shepherd are closely related to the Sheltie and are all descended from the original working collies of the British Isles, prized for their ability to herd sheep. Overall these farm dogs have given rise to least ten independent yet related dog breeds<sup>38</sup>

that share disease risk associated genetic variants<sup>29,30,37</sup>. In this study (Fig. 6) we examined susceptibility to iUC in the above four breeds, one of which, the Sheltie, demonstrates a particularly high disease risk<sup>12</sup>. We observe that the Sheltie and Collie carry the same disease-associated coding variant NIPAL1<sup>G256D</sup> and, along with the Border collie, share a putative regulatory variant associated with iUC located 273 kb downstream. These variants may also exist at low frequency in additional related or descendant breeds.

Dog breeds are developed by reproductively isolating a small group of individuals from the general dog population, therefore we can treat each breed as a large extended family, rather than a geographic or ethnic population, and assume a common risk variant will be responsible for the majority of disease in any given breed<sup>38,39</sup>. In essence, the number of breeds or lineages is the major determinant of power, and not the number of individual dogs. In a high-risk breed, the disease associated variant is often found at high frequency and can be used to further subset the population into genotype-specific clusters to find additional disease loci. Secondary loci may modify the effect of the primary locus or may themselves be primary risk loci that only exist in a small subset of the breed, perhaps in the presence of a particular aspect of the genetic background. We have taken advantage of breed structure to identify a total of seven iUC-associated loci by segregating samples based on their genotype at an initial, primary locus, NIPAL1. Although the additional loci likely play only secondary roles in Sheltie iUC risk, they may be part of the primary path to disease development in other breeds. Within the Sheltie iUC-associated loci we identified germline or somatic variants in three genes that, combined with the previously identified BRAF<sup>V595E</sup> somatic mutation, point towards the disruption of the MAPK signaling pathway and its components as a likely mechanism for iUC development in dogs.

The most compelling mutation found in the strongly associated chr13 locus is a variant within the NIPAL1 gene that causes the non-polar glycine at the 256<sup>th</sup> position to be replaced with a negatively charged aspartic acid. The G256D mutation lies within one of the transmembrane domains of the NIPAL1 protein and is predicted to disrupt the helical structure, possibly by changing the steric configuration through size or charge. The NIPAL1 protein is a transporter of multiple divalent cations, most notably magnesium (Mg<sup>2+</sup>). Magnesium levels and the disruption of magnesium transporters have been associated with cell proliferation, metastasis, and angiogenesis as well as effectiveness and toxicity of chemotherapeutic treatments in bladder and other cancers<sup>40,41</sup>. In human bladder cancer cell lines, 24 h treatment with MgCl<sub>2</sub> is shown to alter the expression and phosphorylation of multiple components of the MAPK signaling pathway compared to untreated cells<sup>42</sup>. The pathway is crucial for tumor development and is frequently dysregulated in both human (reviewed in refs. 21,43),



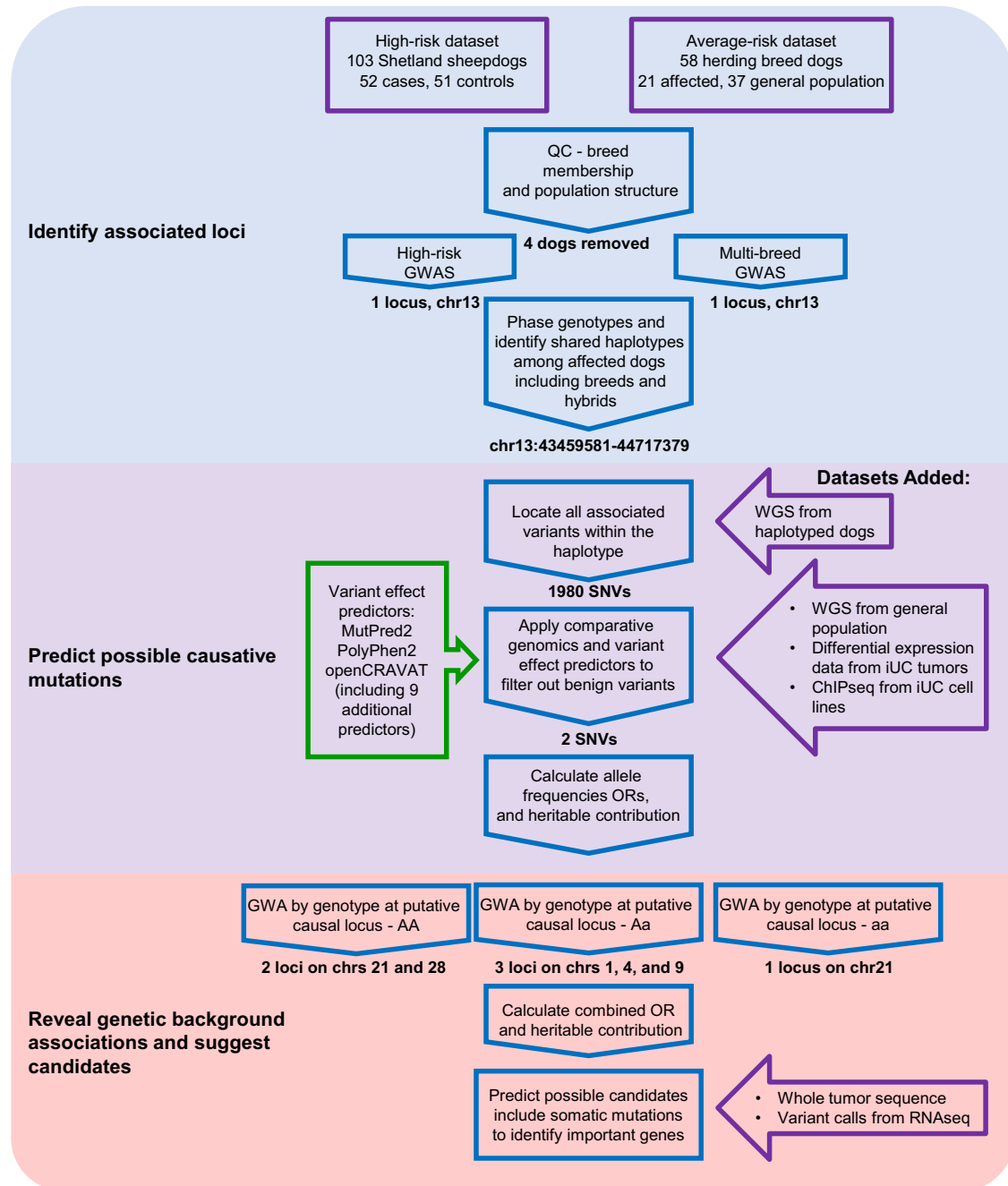
**Fig. 5 | A flow chart of predicted risk for variant combinations leading to the development of iUC.** Combinations of alleles at six loci increase the risk of or the protection from iUC in dogs carrying one of the three possible genotypes for NIPAL1G256D.

**Table 3 | Odds ratios for disease risk given combinations of genotypes at the NIPAL1G256D and associated SNPs at six additional loci**

NIPAL1G256D Genotype	AA		AG		GG				
Chr <sup>1</sup>	21 (1)	28	1	4	9	21 (2)	OR	CI	risk
Associated SNP genotypes	AA	**					18.5	3.1–22.7	increase
	**	CC					58.2	7.4–457.7	increase
			C*	A*	**				
			C*	**	G*		19.1	1.1–342.8	increase
			**	A*	G*				
			TT	GG	**				
			TT	**	CC		0.06	0.02–0.22	decrease
		**	GG	CC					
					T*	0.13	0.04–0.49	decrease	
Disease associated allele frequency <sup>2</sup>	0.92	0.82	0.05	0.2	0.08	0.37			

<sup>1</sup>Chromosomes numbers indicate loci associated with iUC on each chromosome as described. 21(1) is chr21:34085269, 21(2) is chr21:33595359. <sup>2</sup>Disease associated allele frequency calculated in 201 Shetland sheepdogs of all ages with no known cancer diagnosis. OR=Odds Ratio, CI= 95% confidence interval, \*=any allele, \*\*=any allele on both chromosomes.





**Fig. 6 | The order of analyses employed in this study.** These steps were taken to identify possible genetic variants predisposing dogs to iUC and candidate genes that may play an important role in promoting or preventing disease development.

Analysis stages are indicated by changing background color. Datasets used are listed inside purple shapes and predictive software in green.

and dog bladder cancers<sup>20,44</sup>. The mutation is expected to alter the ability of the gene to transport magnesium, and possibly other ions, across cellular membranes, thus affecting magnesium concentration within cells.

Interestingly, *NIPALI* is also associated with gout in a GWAS of more than 4000 Japanese men<sup>45</sup>. Gout is a form of hyperuricemia that can occur in the presence of reduced renal urate excretion, and affected individuals have a 1.18 fold increased risk of developing bladder cancer (95% CI = 1.05–1.32)<sup>46</sup>. Uric acid levels are significantly increased in the urine of dogs with bladder cancer compared to matched controls<sup>47</sup>, and while this relationship has been reported in multiple studies, the underlying cause remains unknown<sup>48</sup>. *NIPALI* may also play a role in regulating the transport kinetics of urea in bladder cells<sup>49</sup>.

In addition to the above, we identified a putative enhancer variant at chr13:44170972 that is in near perfect LD with the *NIPALI* mutation. Based on synteny with human chr4:47547066-48980084 (hg38), this enhancer may affect the expression of 10 genes in the region, including *NIPALI*<sup>35</sup>. Multiple studies show that *NIPALI* is overexpressed in canine iUC tumors<sup>34,50,51</sup>, as are 14 other genes within the primary risk-associated chr13 locus<sup>34</sup>. The enhancer variant alters putative transcription factor binding sites at the locus, creating new sites that bind AP-1 complexes, which are heterodimers of the JUN (Jun proto-oncogene AP-1 transcription factor subunit) and FOS (Fos proto-oncogene AP-1 transcription factor subunit) protein families. These complexes are frequently linked to tumorigenesis, potentially accounting for increased expression across the locus<sup>52</sup>. The

MAPK signaling pathway regulates AP-1 through activation and by controlling expression<sup>53</sup>.

Secondary loci found on chr9:22381987-23201516 and chr21:31405230-33951574 comprise many potential cancer genes, but two, *Erb-B2 receptor tyrosine kinase 2 (ERBB2)* and *WEE1 G2 checkpoint kinase (WEE1)*, stand out as likely candidates for disease risk, as both are frequently mutated, amplified or dysregulated in human bladder tumors. *ERBB2* is a member of the epidermal growth factor receptor family and functions upstream of signaling pathways frequently implicated in cancer, such as the MAPK or the mechanistic target of rapamycin (mTOR) signaling pathways. Multiple studies of canine iUC tumors show that *ERBB2* is over expressed in canine bladder tumors compared to normal bladder tissues<sup>34,51,54</sup>. We did not find any *ERBB2* disease-associated germline coding mutations in the WGS from seven Sheltie cases; however, we did find potentially pathogenic somatic alterations in whole genome sequence and RNAseq from iUC tumors. Mutations at chr9:22775767 and chr9:22775561 correspond to the protein alterations D251H and D277Y in *ERBB2*, respectively. Both variants are predicted to be damaging by multiple metrics (Supplementary Table 5). Twelve percent of human bladder tumors have mutations in *ERBB2*, and the D277Y protein variant is among them<sup>55</sup>. Though the D251H variant has not yet been found in human cancers, multiple mutations in the adjacent amino acid at position 250 have been reported. An exome study of shed DNA from dogs with suspected bladder cancer also found the nearby *ERBB2* mutation, G292R<sup>20</sup> which was observed in one of our transcript sequences. These variants are found in the furin-like domain of the protein which includes the most common human somatic variant in the gene, S310F. In aggregate this argues that *ERBB2* plays an important role in canine iUC disease risk.

An additional gene implicated in human bladder cancer, the tumor suppressor *WEE1*, is located in the disease associated chr21:31405230-33951574 region. *WEE1* regulates the G2/M checkpoint through an activation loop with cell division kinases, CDK1 and CDK2, within the MAPK signaling pathway<sup>56,57</sup>. Activating BRAF mutations, which are found in the majority of canine iUC tumors, increase cellular levels of *WEE1* protein<sup>58</sup>. Through phosphorylation of the CDKs, *WEE1* plays a role in homologous recombination DNA repair, and because cell cycle and DNA damage repair pathways are frequently interrupted during tumor development, *WEE1* inhibitors are effective cancer therapeutics<sup>59,60</sup>. We again identified at least one somatic mutation, P175R, in an iUC tumor. This mutation is in the N-terminal regulatory domain of *WEE1*, adjacent to the nuclear export signal and is predicted to affect the function of the gene. In humans bladder urothelial carcinomas *WEE1* is frequently dysregulated through copy number loss (13.7%)<sup>57</sup>.

Combined, these findings point to repeated disruption of the cellular proliferation and DNA damage repair pathways as mechanisms for development of iUC in dogs. Our data suggests that increased risk is caused through germline variants that provide the initial step in the cascade. We have identified one common germline mutation (*NIPAL1*<sup>D256G</sup>), disease associated sites in the regions of *ERBB2* and *WEE1*, and the presence of somatic mutation in the same genes suggesting they are important in canine iUC progression. Combined with the previously identified BRAF<sup>V595E</sup> mutation, our findings point toward the MAPK signaling pathway and its components as a key driver of iUC in dogs.

By combining genotypes from seven associated loci we can account for the development of iUC in 75% of Shelties. These predictions could improve after analysis of a larger cohort. For example, there were no unaffected dogs heterozygous for the *NIPAL1* variant that carry the minor allele at the chr1 locus. Similarly, only affected dogs that do not carry the *NIPAL1* variant are homozygous for the alternate allele at chr21:33595359. To calculate the odds ratios, we assume a frequency of >0 for these combinations. If these combinations are not seen at increased numbers, they would account for disease risk in an additional 10% of cases, bringing the ability to predict disease as high as 85%. While this would make an excellent start in genetic testing for iUC risk in Shelties, these findings need to be verified in an unrelated set of dogs.

There are factors that limit the generalizability of these results: limited sample sizes and a lack of direct functional assessments for the putative mutations. The modest sample size is appropriate for finding variants of moderate to large effect in dogs given their population structure, but may lead to false negatives. We are unlikely to recognize variants with very small effect sizes given the small number of individuals included in the GWAS and the WGS was targeted specifically for finding high-effect variants on chr13. It is also possible that there are low penetrance variants contributing to the disease that are fixed within this population. We expect these low-effect and low-penetrance variants to reveal themselves as additional high-risk breeds are analyzed and datasets are combined. The dog genome project has advanced substantially within the last decade with the addition of four new reference sequences and 19 new assemblies (NCBI). There is also a finely filtered and vetted variant dataset of over 2000 whole genome sequences<sup>36</sup>. However, there remains a dearth of experimentally confirmed protein function data, thus comparative genomics provides the majority of data regarding gene function<sup>61-63</sup>. This is not necessarily a problem, however, as comparative genomics between dog and human features numerous successful studies that identify altered genes that share function in the same disorders in both species (reviewed in<sup>64-66</sup>). The next step in our studies will be to perform functional analyses on the newly identified and predicted pathogenic mutations using protein assays to determine specific activity, allele-specific functional studies in cell cultures, or by creating mouse models. There are also non-coding RNAs in the dog genome that have yet to be annotated due to a lack of sequence conservation. For example, there is a lncRNA, *OCA1D1-AS1* that has been linked to bladder cancer prognoses in humans<sup>67</sup> within our primary associated region on chr13. This lncRNA is weakly conserved and not annotated in dogs and though there are no associated mutations in the corresponding locus, we may be missing changes in expression that correlate with tumor behavior.

Finally, we note that the study of iUC in canine populations may fill in gaps in studies of bladder cancer in humans. The numerous cohorts amassed for studies of human bladder cancer risk feature, on average, > 80% males and 80–85% non-muscle invasive disease<sup>68-71</sup>, even though muscle invasive bladder cancer carries a much higher risk of early mortality<sup>72</sup>. Women are underrepresented in these studies because they are less likely to get bladder cancer than men, however they more often get invasive disease and their relative survival expectancy is reduced by five percentage points or more across all stages of the disease (<https://seer.cancer.gov/>). In dogs, females are either equally susceptible or more likely to get iUC than males, depending on the breed. Our dataset included 59% female and 41% male affected dogs and >90% invasive disease. Because of these factors, the dog is likely to provide new information about genes and pathways important to a highly destructive subset of bladder cancers in humans.

## Methods

### Sample collection and DNA extraction

Affected and unaffected dogs from the high-risk breed were solicited through the Purdue University Veterinary hospital as well as through dog clubs and at dog events. Whole blood samples were collected from dogs with iUC through either the Purdue University Veterinary Hospital, where they had biopsy-confirmed iUC, or via direct mail from owners of affected dogs. Owners of dogs not diagnosed at Purdue submitted pathology reports with diagnosis from the treating veterinarian. Affected dogs had no reports of other cancers. Unaffected dogs were  $\geq 10$  years old at the time of sampling with no reports of any cancer. Both groups were filtered for relatedness so that no two dogs shared grandparents and they were matched for sex. Affected dogs from average risk breeds were collected through the Purdue University Veterinary hospital and unaffected dogs were chosen from population studies (Table 1). Similar human studies often recruit from high-risk clinics, standardized screening programs, or make use of data from HMOs or insurance companies. However, few such programs exist in the veterinary world, e.g., there are no pet HMOs and pet insurance is still in its infancy. However, the very nature of being a Shetland sheepdog puts dogs of this breed in a high-risk category. Further, samples were collected from

Purdue Veterinary hospital department of oncology, which specializes in the treatment of urinary carcinoma, thus insuring a high-quality dataset. Collection was performed according to the NHGRI Animal Care and Use committee approved protocol GFS-05-1 or the Purdue Institutional Animal Care and Use committee approved protocol 1111000169. Dog owners signed informed consent documents prior to collection. DNA was extracted from whole blood samples using cell lysis followed by phenol-chloroform, as described previously<sup>73</sup>, with phase separation performed in 15 mL Phase Lock tubes (5-Prime, Inc. Gaithersburg, MD, USA). DNA was resuspended in a 10 mMolar (mM) tris, 0.1 mM EDTA solution and stored at  $-80^{\circ}\text{C}$ .

### Sample filtering

In order to include non-pedigreed dogs, all samples were assessed for breed membership by combining the cancer genotyping data with previously published population dataset including 157 breeds in addition to the iUC study breeds<sup>28</sup>, and using principal components analysis (PCA) and distance measurements calculated in PLINK v.1.90<sup>74</sup>. The all-breed data can be obtained at [https://research.nhgri.nih.gov/dog\\_genome/data\\_release/](https://research.nhgri.nih.gov/dog_genome/data_release/) or through the NCBI Gene Expression Omnibus (GEO) data repository under accession numbers: GSE90441, GSE83160, GSE70454, and GSE96736. Significant PCs were determined using the Tracy-Widom distribution<sup>75</sup>. The first three PCs significantly separated the four UK herding breeds: Shetland sheepdog, Collie, Border collie, and Australian shepherd (Fig. 1 and Supplementary Fig. 1). Samples that did not cluster with their assigned breed on PCA were marked for follow-up. Samples were then added to a neighbor joining tree which included an additional 157 previously published breeds<sup>28</sup> using a 1-ibs distance matrix calculated in PLINK. Of the samples marked for follow-up after PCA, three putative Shetland sheepdogs did not cluster with the pedigreed Shetland sheepdogs on the multi-breed tree and were therefore excluded from the Shetland sheepdog analysis. Samples that did not cluster with any of the UK herding breeds (two of the above Shetland sheepdogs and one Australian shepherd) were excluded from all analyses. Samples that clustered with the UK herding breeds, but not with their assigned breed, were retained for the multi-breed herding dog analyses (one Australian shepherd, two Border collies and one Shetland sheepdog). All other samples clustered with the submitted breed in both analyses (Supplementary Fig. 2).

The ancestry of the four herding breed dogs, designated atypical, was determined by calculating shared haplotypes, estimated to be identical by descent (IBD) using the BEAGLE v4.1 `ibd=true` option. Total shared haplotypes were summed between all pairs of dogs as described previously<sup>28</sup> this time comparing breed dogs to the unknown dog rather than breed to breed. Breeds with average total haplotype sharing with the unknown dog above the 95% percentile of all comparisons were considered possible contributors. Final relationships were visualized using Circos in R<sup>76</sup> (Supplementary Fig. 6).

### Single Nucleotide Polymorphism (SNP) genotyping and GWAS

DNA samples from 103 Shelties and 58 dogs from related herding breeds were genotyped at 170k loci using the Illumina CanineHD bead chip with standard protocols (Illumina, San Diego, CA). Genotypes were called in Genome Studio 2.0.4 and exported using the CanFam3.1 forward strand orientation. Data was imported into pLINK v1.9 for filtering and reformatting. Single nucleotide polymorphisms (SNPs) missing more than 10% of genotypes, or with minor allele frequencies (MAF) < 5% in Shetland sheepdogs, were removed from the analyses. After filtering the final dataset included 90,530 SNPs and 155 dogs. Power analysis was carried out using `genpwr`<sup>77</sup> in R (Supplementary Fig. 7).

Genomic inflation was calculated from the  $p$ -values from uncorrected association analysis in R with  $\lambda_{GC} > 1.1$ , indicating genome-wide inflation likely due to confounding population structure. PCA was run on the dataset and one-way ANOVA test was calculated in R for each significant PC to identify stratification by disease. To correct for the population structure, GWAS was performed in GEMMA<sup>78</sup> using DNA from 49 Sheltie cases and 51 aged ( $\geq 10$  yrs) Sheltie controls. A second GWAS included additional

DNA samples from 21 affected and 34 unaffected dogs from three related herding breeds (Collie, Border collie and Australian shepherd) (Table 1). Correction for multiple testing was applied in two ways; Bonferroni correction based on independent loci (pair-wise  $r^2 < 0.80$ )  $P$ -Bonferroni<sub>05</sub> =  $1.7 \times 10^{-6}$ , and by permuting the phenotypes 10,000 times and identifying the limit at which 95% of random associations could be excluded ( $P$ -perm 0.05 =  $1.32 \times 10^{-6}$ ).

Associated regions were defined as the area encompassing SNPs in LD ( $r^2 > 0.6$ ) with the most associated SNP and associated with the disease with a  $P \leq 1 \times 10^{-5}$ . All SNPs within this region were phased using PHASE v2.1 with recombination<sup>79</sup>. The common Sheltie case haplotype was included as a known haplotype when phasing non-risk breed data.

Additional association analyses were performed on subsets of samples based on their genotype at chr13:43897196. The three sets include 29 cases and seven controls that were homozygous for the *NIPALI*<sup>G256D</sup> mutation, 11 cases and 31 controls all heterozygous for the mutation, and eight cases and 20 controls, none of which carry the mutation. The association analysis was also repeated in the full set of 101 Shetland sheepdog samples using the number of alternate alleles at chr13:43897196, the variant within the *NIPALI* gene, as a covariate. All analyses were corrected for underlying population structure as described above.

### Whole genome sequence (WGS)

Twelve Shelties were chosen for further genotyping because of the haplotype they carried across the chr13 associated locus. We chose six samples that were homozygous for the disease associated haplotype, five with one copy of the disease associated haplotype and one that was homozygous for the most common non-disease associated haplotype. Germline whole genome sequencing of these 12 samples was carried out at the NIH Intramural Sequencing Center (NISC) using the Illumina TruSeq DNA PCR-Free Protocol (Illumina, San Diego, CA, USA) on an Illumina Novaseq6000. These data are available in the NCBI Short Read Archive (SRA) under BioProject numbers PRJNA448733 and PRJNA685036 (Supplementary Table 1). Paired-end data was aligned to the CanFam3.1 reference genome and variants called using GATK as described previously<sup>80,81</sup>. Annotation was applied to the variant sets using `snpeff` v5.0<sup>82</sup> with the Ensembl gene set CanFam3.1.99 and also using VEP<sup>83</sup> with the CanFam3.1plus gene annotation<sup>84</sup>. Polymorphic variants from the associated region were extracted using `bcftools view` command<sup>85</sup>. The alignment was repeated using the reference build UU\_GSD1.0<sup>86</sup> and the corresponding annotation to look for missing exons.

In addition to the 12 germline sequences described above, three canine iUC tumors were sequenced at 60x coverage and aligned to CanFam3.1. Somatic variants were called using Mutect2<sup>87</sup> with both matched normal sequence and a population-based panel of common variants to rule out germline polymorphism. In addition, RNAseq data published previously<sup>34</sup> was analyzed for somatic mutation. Bam files were prepped following GATK best practices workflow for RNAseq short variant discovery. Variants were called using Mutect2 with matched normal RNAseq or a panel of normal RNAseq if no matched normal was available. A population-based panel of variants was used to filter out common polymorphism. All variants within the genes of interest were confirmed manually in IGV<sup>88</sup>.

### Mutation identification

Variants identified in the 12 WGSs that followed the same pattern as the most significantly associated SNPs identified in the GWAS and appearing on the disease associated haplotype were retained for further investigation. The position of each variant in relation to annotated genes was assessed using `SNPeff` and the CanFam3.1.99 gene model, VEP with the CanFam3.1plus gene model<sup>84</sup>, and by overlap using a bed file of exon positions from the Broad canine improved annotation v.1<sup>89</sup>. Variants within coding transcripts predicted to alter amino acids were converted to human genome positions using lift-over (UCSC, hg19)<sup>90,91</sup>. Within coding sequences, only sites for which both the DNA sequence and predicted amino acid were conserved between dogs and humans were considered for further

evaluation. To confirm that no coding variants were missing due to gaps in the reference sequence we realigned the sequence reads to the UU\_GSD1.0 reference sequence<sup>92</sup> and searched for variation within the updated gene annotation. Pathogenicity of variants was predicted using openCRAVAT<sup>93,94</sup>, which encompasses multiple predictive tools, MutPred2<sup>95</sup>, and PolyPhen2<sup>96</sup>.

Disease-associated non-coding mutations were filtered for overlap with sites of human enhancers that had confirmed local gene interactions<sup>35</sup>, and with predicted canine enhancers and promoters determined from ChIP-seq analysis of canine iUC tumor cell lines. The human enhancer loci were lifted to the dog assembly using UCSC lift-over<sup>90</sup>. The allele frequency of each variant was determined using a set of 1330 dogs representing >200 breeds and 139 wild canids, none of which had a known diagnosis of iUC. Variants were considered potentially disease contributing if they were within a canine active enhancer, in the corresponding site of a human cis-active enhancer, and had an allele frequency of <10% in unaffected, average risk breeds.

Transcription factor binding sites (TFBS) were identified in regions of interest using TFBind and AIModules. Significant matches were identified using an empirically determined cut-off value for sequence similarity based on position weighted matrices from TRANSFAC R.3.4 database implemented by TFBind<sup>97</sup>, or an odds ratio greater than seven and an odds ratio deficit less than eight, compared to the canonical sequence, as determined by similarity calculations from position weighted matrices from the JASPAR 2022 database for the AIModule predictions<sup>98</sup>.

### Chromatin immunoprecipitation followed by sequencing (ChIP-seq)

ChIP-seq was performed as described in<sup>99</sup> from canine iUC cell lines K9TCC\_Mx and K9TCC\_Ab, which were provided by the Knapp lab at the Werling Comparative Oncology Research Center at Purdue University, College of Veterinary Medicine<sup>100,101</sup>. The cell lines were grown to 70–80% confluency in DMEM/F12 media as previously described<sup>100,101</sup>. Approximately  $5 \times 10^7$ – $1 \times 10^8$  cells were used in each assay. Cells were cross-linked using 1/10 volume of 11% formaldehyde quenched with 1/20 volume of 1 M glycine after 10 min. Cells were washed twice with 1x PBS and harvested using a Cell Lifter (Costar #3008) in PBS. Cells were pelleted and lysed in three stages with a protease inhibitor cocktail (Roche #11873580001) added at each step. DNA in pelleted nuclei was sheared on ice using a GEX130 Ultrasonic processor (Sonics, Newtown, CT) in 3 ml volumes with the following settings: power set at 30% output, 25 s ON, then 60 s OFF, repeated for a total of eight minutes. Magnetic beads (Invitrogen Dynabeads Protein G, Thermo Fisher Scientific, Waltham, MA), previously bound with an antibody to one of the histone modifications, were washed with phosphate-buffered saline solution (PBS), resuspended, then added to the fragmented DNA on ice and incubated at 4 °C overnight. Beads with bound DNA strands were washed five times, then rinsed with PBS. The final rinse was aspirated and 115  $\mu$ l of elution buffer added. DNA was eluted at 65 °C for 15 min, vortexing every two minutes, followed by overnight incubation on a Thermomixer R (Eppendorf, Enfield, CT). DNA was purified with phenol-chloroform after treatment with RNaseA and proteinase K.

Immunoprecipitation (IP) was repeated using 10  $\mu$ g of Abcam (Cambridge, United Kingdom) antibodies to H3K4me1 (rabbit polyclonal-AbCam, #ab8895), H3K4me3 (rabbit polyclonal-AbCam, #ab8580), H3K27ac (rabbit polyclonal AbCam #ab4729) and one negative control (IgG rabbit, AbCam#46540). A 100  $\mu$ l input control sample was taken from shorn DNA prior to each IP and sequenced to control for background signal in the IP. DNA was quantified using a Qubit fluorometer (ThermoFisher, Waltham, MA, USA).

The purified, IP DNA was submitted for sequencing at NISC, where paired-end libraries were made with inserts averaging 320 bps and run on an HiSeq2000 Sequencing System (Illumina) with 100 bp reads. The reads were aligned to the CanFam3.1 reference sequence using BWA-MEM<sup>102</sup> and converted to bam files using samtools 1.10<sup>85</sup>. Alignments were filtered as described previously in ref. 103 and histone binding sites called by counting

the number of mapped reads overlapping the peak using MACS2 2.2.1<sup>104</sup>, and the spatial clustering approach from SICER2<sup>105</sup>. Only those that were present in two different canine iUC cell lines, overlapping each other by at least 50%, and called by both SICER and Macs2, were retained for comparisons.

### Targeted genotyping

Primers were designed to amplify the region around the two variants of interest on chr13, chr13:43897196 (F- GCTCCCAAGAAGGGACAGAC; R- TTGGAATAAGATGACAGAGCAAG) and chr13:44170972 (F- AAACGATCCAGAGAGCAGATTAC; R- GAGCCCAGGCCAAGAGTG) using Primer3<sup>106</sup>. PCR was carried out on a SimplyAmp thermocycler or GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA) using AmpliTaq Gold (Thermo Fisher, Waltham, MA) using standard reaction protocols with a touch-down thermocycler program starting at 65 °C annealing temperature and decreasing 0.5 °C each cycle for 20 cycles then an additional 20 cycles at 55 °C annealing temperature. Denaturing and extending temperatures were constant as per standard protocols.

Amplified DNA was sequenced using BigDye Terminator 3.1 (Applied Biosystems, Foster City, CA) and run on a 3730xl sequence analyzer (Applied Biosystems). The variants were genotyped using Sequencher 5.4.6 (GeneCodes Corporation, Ann Arbor, MI). Odds ratios were calculated using genotypes from 47 affected and 47 unaffected Shelties.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The datasets presented in this study can be found in public online repositories. Specifically: Illumina SNP chip data and ChIPseq data from cell lines has been submitted to the NCBI GEO database under accession GSE241367 and GSE254079 respectively. Whole tumor sequence is in the NCBI short read archive (SRA) under BioProject PRJNA1007700. Whole genome sequence can be found in the SRA under BioProjects PRJNA288568 and PRJNA685036. RNAseq was previously submitted to the SRA under BioProjects PRJNA559406 and PRJNA308949.

### Code availability

Code generated to support primary data analysis and build figures is available upon request. Software, versions, and associated parameters used can be found in the methods section.

Received: 29 August 2023; Accepted: 14 April 2024;

Published online: 22 May 2024

### References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **72**, 7–33 (2022).
2. Saginala, K. et al. Epidemiology of bladder cancer. *Med. Sci. (Basel)* **8**, 15 (2020).
3. Witjes, J. A. et al. EAU guidelines on muscle-invasive and metastatic bladder cancer: summary of the 2013 guidelines. *Eur. Urol.* **65**, 778–792 (2014).
4. Milowsky, M. I. et al. Guideline on muscle-invasive and metastatic bladder cancer (European Association of Urology guideline): American Society of Clinical Oncology clinical practice guideline endorsement. *J. Clin. Oncol.* **34**, 1945–1952 (2016).
5. Yi, L., Ai, K., Li, X., Li, Z. & Li, Y. Bone metastasis in bladder cancer. *J. Pers. Med.* **13**, 54 (2023).
6. Pang, L. et al. Causal relationship between smoking status, smoking frequency and bladder cancer: a Mendelian randomization study. *Genes Genomics* **45**, 203–213 (2023).
7. Burger, M. et al. Epidemiology and risk factors of urothelial bladder cancer. *Eur. Urol.* **63**, 234–241 (2013).

8. Mushtaq, J., Thuraija, R. & Nair, R. Bladder cancer. *Surg. (Oxf.)* **37**, 529–537 (2019).
9. Beane Freeman, L. E. et al. Disinfection by-products in drinking water and bladder cancer: Evaluation of risk modification by common genetic polymorphisms in two case-control studies. *Environ. Health Perspect.* **130**, 57006 (2022).
10. Al-Husseini, M. J. et al. Trends in the incidence and mortality of transitional cell carcinoma of the bladder for the last four decades in the USA: a SEER-based analysis. *BMC cancer* **19**, 46 (2019).
11. Murta-Nascimento, C. et al. Risk of bladder cancer associated with family history of cancer: do low-penetrance polymorphisms account for the increase in risk? *Cancer Epidemiol. Biomark. Prev.* **16**, 1595–1600 (2007).
12. Knapp, D. W. et al. Urinary bladder cancer in dogs, a naturally occurring model for cancer biology and drug development. *ILAR J.* **55**, 100–118 (2014).
13. Fulkerson, C. M. & Knapp, D. W. Management of transitional cell carcinoma of the urinary bladder in dogs: a review. *Vet. J.* **205**, 217–225 (2015).
14. Dhawan, D. et al. Identification of a naturally-occurring canine model for early detection and intervention research in high grade urothelial carcinoma. *Front Oncol.* **12**, 1011969 (2022).
15. Knapp, D. W., Dhawan, D. & Ostrander, E. “Lassie,” “Toto,” and fellow pet dogs: poised to lead the way for advances in cancer prevention. *Am. Soc. Clin. Oncol. Educ. Book* **35**, e667–e672 (2015).
16. Mutsaers, A. J., Widmer, W. R. & Knapp, D. W. Canine transitional cell carcinoma. *J. Vet. Intern Med.* **17**, 136–144 (2003).
17. Knapp, D. W. et al. Naturally-occurring invasive urothelial carcinoma in dogs, a unique model to drive advances in managing muscle invasive bladder cancer in humans. *Front Oncol.* **9**, 1493 (2019).
18. Glickman, L. T., Raghavan, M., Knapp, D. W., Bonney, P. L. & Dawson, M. H. Herbicide exposure and the risk of transitional cell carcinoma of the urinary bladder in Scottish Terriers. *J. Am. Vet. Med. Assoc.* **224**, 1290–1297 (2004).
19. Decker, B. et al. Homologous mutation to human BRAF V600E is common in naturally occurring canine bladder cancer—Evidence for a relevant model system and urine-based diagnostic test. *Mol. Cancer Res.* **13**, 993–1002 (2015).
20. Thomas, R. et al. Whole exome sequencing analysis of canine urothelial carcinomas without BRAF V595E mutation: Short in-frame deletions in BRAF and MAP2K1 suggest alternative mechanisms for MAPK pathway disruption. *PLoS Genet.* **19**, e1010575 (2023).
21. Clark-Garvey, S. & Kim, W. Y. RAF1 amplification: an exemplar of MAPK pathway activation in urothelial carcinoma. *J. Clin. Invest.* **131**, e154095 (2021).
22. Song, Y. et al. Targeting RAS-RAF-MEK-ERK signaling pathway in human cancer: Current status in clinical trials. *Genes Dis.* **10**, 76–88 (2023).
23. Awad, M. M. et al. Acquired resistance to KRAS(G12C) inhibition in cancer. *N. Engl. J. Med.* **384**, 2382–2393 (2021).
24. Scardaci, R. et al. Novel RAF-directed approaches to overcome current clinical limits and block the RAS/RAF node. *Mol. Oncol.* (2024).
25. de Brot, S. et al. The dog as an animal model for bladder and urethral urothelial carcinoma: Comparative epidemiology and histology. *Oncol. Lett.* **16**, 1641–1649 (2018).
26. Valli, V. E. et al. Pathology of canine bladder and urethral cancer and correlation with tumour progression and survival. *J. Comp. Pathol.* **113**, 113–130 (1995).
27. Knapp, D. et al. Naturally-occurring canine transitional cell carcinoma of the urinary bladder: A relevant model of human invasive bladder cancer. *Urol. Oncol.* **5**, 47–59 (2000).
28. Parker, H. G. et al. Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. *Cell Rep.* **19**, 697–708 (2017).
29. Parker, H. G. et al. Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res.* **17**, 1562–1571 (2007).
30. Neff, M. W. et al. Breed distribution and history of canine mdr1-1Delta, a pharmacogenetic mutation that marks the emergence of breeds from the collie lineage. *Proc. Natl. Acad. Sci. USA* **101**, 11725–11730 (2004).
31. Sutter, N. B. et al. Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Res.* **14**, 2388–2396 (2004).
32. Lindblad-Toh, K. et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
33. Salmon Hillbertz, N. H. C. et al. Duplication of FGF3, FGF4, FGF19 and ORO1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nat. Genet.* **39**, 1318–1320 (2007).
34. Parker, H. G. et al. RNAseq expression patterns of canine invasive urothelial carcinoma reveal two distinct tumor clusters and shared regions of dysregulation with human bladder tumors. *BMC cancer* **20**, 251 (2020).
35. Nasser, J. et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
36. Meadows, J. R. S. et al. Dog10K: Genome sequencing of 2,000 canids advances the understanding of demography, genome function and architecture *Genome Biology in press*, (2023).
37. Evans, J. M. et al. Beyond the MHC: A canine model of dermatomyositis shows a complex pattern of genetic risk involving novel loci. *PLoS Genet.* **13**, e1006604 (2017).
38. American Kennel Club. *The Complete Dog Book*. 20th edn, (Ballantine Books, 2006).
39. Stockard, C. R. *The Genetic and Endocrinic Basis for Differences in Form and Behavior*. 775 (The Wistar Institute of Anatomy and Biology, 1941).
40. Castiglioni, S. & Maier, J. A. Magnesium and cancer: a dangerous liaison. *Magnes. Res.* **24**, S92–S100 (2011).
41. Trapani, V. & Wolf, F. I. Dysregulation of Mg(2+) homeostasis contributes to acquisition of cancer hallmarks. *Cell Calcium* **83**, 102078 (2019).
42. Li, T. et al. Magnesium in combination with valproic acid suppressed the proliferation and migration of human bladder cancer cells. *Front Oncol.* **10**, 589112 (2020).
43. Guo, Y. J. et al. ERK/MAPK signalling pathway and tumorigenesis. *Exp. Ther. Med.* **19**, 1997–2007 (2020).
44. Mochizuki, H., Shapiro, S. G. & Breen, M. Detection of BRAF mutation in urine DNA as a molecular diagnostic for canine urothelial and prostatic carcinoma. *PLoS One* **10**, e0144170 (2015).
45. Sasahira, T. et al. NIPA-like domain containing 1 is a novel tumor-promoting factor in oral squamous cell carcinoma. *J. Cancer Res. Clin. Oncol.* **144**, 875–882 (2018).
46. Feng, D., Hu, X., Yang, Y., Yang, L. & Wei, W. The association between gout and the risk of urological cancers: A pooled analysis of population-based studies. *Int J. Rheum. Dis.* **24**, 281–283 (2021).
47. Zhang, J. et al. NMR-based metabolomics study of canine bladder cancer. *Biochim Biophys. Acta* **1822**, 1807–1814 (2012).
48. Allegrini, S., Garcia-Gil, M., Pesi, R., Camici, M. & Tozzi, M. G. The good, the bad and the new about uric acid in cancer. *Cancers (Basel)* **14**, 4959 (2022).
49. Nakayama, A. et al. GWAS of clinically defined gout and subtypes identifies multiple susceptibility loci that include urate transporter genes. *Ann. Rheum. Dis.* **76**, 869–877 (2017).
50. Ramsey, S. A. et al. Cross-species analysis of the canine and human bladder cancer transcriptome and exome. *Genes Chromosomes Cancer* **56**, 328–343 (2017).
51. Maeda, S. et al. Comprehensive gene expression analysis of canine invasive urothelial bladder carcinoma by RNA-Seq. *BMC cancer* **18**, 472 (2018).

52. Eferl, R. & Wagner, E. F. AP-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer* **3**, 859–868 (2003).
53. Whitmarsh, A. J. Regulation of gene transcription by mitogen-activated protein kinase signaling pathways. *Biochim Biophys. Acta* **1773**, 1285–1298 (2007).
54. Millanta, F. et al. Overexpression of HER-2 via immunohistochemistry in canine urinary bladder transitional cell carcinoma - A marker of malignancy and possible therapeutic target. *Vet. Comp. Oncol.* **16**, 297–300 (2018).
55. Forbes, S. A. et al. The catalogue of somatic mutations in cancer (COSMIC). *Curr. Protoc. Hum. Genet.* **Chapter 10**, Unit 10 11, (2008).
56. Smith, H. L., Southgate, H., Tweddle, D. A. & Curtin, N. J. DNA damage checkpoint kinases in cancer. *Expert Rev. Mol. Med* **22**, e2 (2020).
57. Ghelli Luserna di Rora, A., Cerchione, C., Martinelli, G. & Simonetti, G. A WEE1 family business: regulation of mitosis, cancer progression, and therapeutic target. *J. Hematol. Oncol.* **13**, 126 (2020).
58. Sharma, A. et al. Identification of aurora kinase B and Wee1-like protein kinase as downstream targets of (V600E)B-RAF in melanoma. *Am. J. Pathol.* **182**, 1151–1162 (2013).
59. Kong, A. & Mehanna, H. WEE1 inhibitor: Clinical development. *Curr. Oncol. Rep.* **23**, 107 (2021).
60. Murakami, K. et al. Antitumor effect of WEE1 blockade as monotherapy or in combination with cisplatin in urothelial cancer. *Cancer Sci.* **112**, 3669–3681 (2021).
61. Stephan, T. et al. Darwinian genomics and diversity in the tree of life. *Proc. Natl. Acad. Sci. USA* **119**, e2115644119 (2022).
62. Sullivan, P. F. et al. Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science* **380**, eabn2937 (2023).
63. Zoonomia, C. A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020).
64. London, C. A. et al. Leading the pack: Best practices in comparative canine cancer genomics to inform human oncology. *Vet. Comp. Oncol.* **21**, 565–577 (2023).
65. Pallotti, S., Piras, I. S., Marchegiani, A., Cerquetella, M. & Napolioni, V. Dog-human translational genomics: state of the art and genomic resources. *J. Appl Genet* **63**, 703–716 (2022).
66. Bunel, M. et al. Natural models for retinitis pigmentosa: progressive retinal atrophy in dog breeds. *Hum. Genet.* **138**, 441–453 (2019).
67. Li, W. X. & Zhang, Y. L. Novel long non-coding RNA markers for prognostic prediction of patients with bladder cancer. *Chin. Med Sci. J.* **35**, 239–247 (2020).
68. Pietzak, E. J. et al. Inherited germline cancer susceptibility gene variants in individuals with non-muscle-invasive bladder cancer. *Clin. Cancer Res.* **28**, 4267–4277 (2022).
69. Figueroa, J. D. et al. Genome-wide association study identifies multiple loci associated with bladder cancer risk. *Hum. Mol. Genet.* **23**, 1387–1398 (2014).
70. Figueroa, J. D. et al. Identification of a novel susceptibility locus at 13q34 and refinement of the 20p12.2 region as a multi-signal locus associated with bladder cancer risk in individuals of European ancestry. *Hum. Mol. Genet.* **25**, 1203–1214 (2016).
71. Rafnar, T. et al. Genome-wide association study yields variants at 20p12.2 that associate with urinary bladder cancer. *Hum. Mol. Genet* **23**, 5545–5557 (2014).
72. Bilim, V. et al. Advanced bladder cancer: Changing the treatment landscape. *J. Pers. Med.* **12**, 1745 (2022).
73. Bell, G. I., Selby, M. J. & Rutter, W. I. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeated sequences. *Nature* **295**, 31–35 (1982).
74. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
75. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
76. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
77. Moore, C. M., Jacobson, S. A. & Fingerlin, T. E. Power and sample size calculations for genetic association studies in the presence of genetic model misspecification. *Hum. Hered.* **84**, 256–271 (2019).
78. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
79. Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet* **68**, 978–989 (2001).
80. Parker, H. G., Whitaker, D. T., Harris, A. C. & Ostrander, E. A. Whole genome analysis of a single Scottish deerhound dog family provides independent corroboration that a SGK3 coding variant leads to hairlessness. *G3 (Bethesda)* **10**, 293–297 (2020).
81. Plassais, J. et al. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* **10**, 1489 (2019).
82. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly. (Austin)* **6**, 80–92 (2012).
83. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
84. Wucher, V. et al. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res* **45**, e57 (2017).
85. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
86. Wang, C. et al. A new long-read dog assembly uncovers thousands of exons and functional elements missing in the previous reference. *bioRxiv*, (2020). 2020.2007.2002.185108.
87. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
88. Robinson, J. T., Thorvaldsdottir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant review with the Integrative Genomics Viewer. *Cancer Res.* **77**, e31–e34 (2017).
89. Hoepfner, M. P. et al. An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. *PLoS One* **9**, e91172 (2014).
90. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res* **12**, 996–1006 (2002).
91. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
92. Wang, C. et al. A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. *Commun. Biol.* **4**, 185 (2021).
93. Douville, C. et al. CRAVAT: cancer-related analysis of variants toolkit. *Bioinformatics* **29**, 647–648 (2013).
94. Masica, D. L. et al. CRAVAT 4: Cancer-Related Analysis of Variants Toolkit. *Cancer Res.* **77**, e35–e38 (2017).
95. Pejaver, V. et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* **11**, 5918 (2020).
96. Carter, H., Douville, C., Stenson, P. D., Cooper, D. N. & Karchin, R. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* **14**, S3 (2013).
97. Tsunoda, T. & Takagi, T. Estimating transcription factor bindability on DNA. *Bioinformatics* **15**, 622–630 (1999).

98. Aydinli, M., Liang, C. & Dandekar, T. Motif and conserved module analysis in DNA (promoters, enhancers) and RNA (lncRNA, mRNA) using AIModules. *Sci. Rep.* **12**, 17588 (2022).
99. Stitzel, M. L. et al. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab.* **12**, 443–455 (2010).
100. Dhawan, D., Ramos-Vara, J. A., Stewart, J. C., Zheng, R. & Knapp, D. W. Canine invasive transitional cell carcinoma cell lines: In vitro tools to complement a relevant animal model of invasive urinary bladder cancer. *Urol. Oncol.* **27**, 284–292 (2009).
101. Shapiro, S. G., Knapp, D. W. & Breen, M. A cultured approach to canine urothelial carcinoma: molecular characterization of five cell lines. *Canine Genet. Epidemiol.* **2**, 15 (2015).
102. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595, (2010).
103. Evans, J. M. et al. Multi-omics approach identifies germline regulatory variants associated with hematopoietic malignancies in retriever dog breeds. *PLoS Genet* **17**, e1009543 (2021).
104. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
105. Zang, C. et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952–1958 (2009).
106. Untergasser, A. et al. Primer3–new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).

## Acknowledgements

We thank the many Shetland sheepdog owners who provided samples for this study, the owners who provided samples from additional breed dogs and the many veterinarians who provided blood samples and clinical data for this study. We especially thank Dr. Elizabeth McNeil for providing samples and the staff of the Werling Comparative Oncology Research Center at Purdue University for case evaluation and sample collection. We thank Dr. Dayna Dreger for careful reading of the manuscript and thoughtful comments. We gratefully acknowledge the NIH Intramural Sequencing Center for DNA sequencing. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). We gratefully acknowledge funding from the National Institutes of Health through the Intramural Program of the National Human Genome Research Institute at NIH (E.A.O and H.G.P.) and the National Cancer Institute grant U01CA272280 (D.W.K. and D.D.). All individuals who qualify for authorship as per the International Committee of Medical Journal Editors as outlined by Oxford Academic have been so designated.

## Author contributions

Experiments were conceived by H.G.P., D.W.K. and E.A.O. Case samples, cell lines and associated data were provided by D.W.K. and D.D.; Laboratory experiments were executed by H.G.P., A.C.H., and E.M.K.; Analysis was conducted by H.G.P. and J.P. Project oversight and funding were provided by E.A.O.; H.G.P. wrote the initial draft of the manuscript. All authors reviewed and edited the manuscript and approved the final submission.

## Funding

Open access funding provided by the National Institutes of Health.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41698-024-00591-0>.

**Correspondence** and requests for materials should be addressed to Elaine A. Ostrander.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024