

Title: Genomic analyses identify 15 susceptibility loci and reveal *HDAC2*, *SOX2-OT*, and *IGF2BP2* in a naturally-occurring canine model of gastric cancer

Shawna R. Cook^{1,2}, Sanne Hugen³, Jessica J. Hayward², Thomas R. Famula⁴, Janelle M. Belanger⁴, Elizabeth McNiel⁵, Hille Fieten³, Anita M. Oberbauer⁴, Peter A.J. Leegwater³, Elaine A. Ostrander⁶, Paul J.J. Mandigers³, Jacquelyn M. Evans^{1,2*}

¹ Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA

² Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA

³ Expertisecentre of Genetics, Department of Clinical Sciences, Utrecht University, Utrecht, The Netherlands

⁴ Department of Animal Science, University of California, Davis, CA, USA

⁵ Cummings School of Veterinary Medicine, Tufts University, Grafton, Massachusetts, USA

⁶ Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Center, National Institutes of Health, Bethesda, MD, USA

*Correspondence: jme255@cornell.edu

Keywords: adenocarcinoma, Bayesian GWAS, selection, *PDZRN3*, Tervuren, Malinois, Belgian Sheepdog, dog, canine, Groenendael

SUMMARY

Gastric cancer (GC) is the fifth most common human cancer worldwide, but the genetic etiology is largely unknown. We performed a Bayesian genome-wide association study and selection analyses in a naturally-occurring canine model of GC, the Belgian Tervuren and Sheepdog breeds, to elucidate underlying genetic risk factors. We identified 15 loci with over 90% predictive accuracy for the GC phenotype. Variant filtering revealed germline putative regulatory variants for the *EPAS1* (*HIF2A*) and *PTEN* genes and a coding variant in *CD101*. Although closely related to Tervuren and Sheepdogs, Belgian Malinois rarely develop GC. Across-breed analyses uncovered protective haplotypes under selection in Malinois at *SOX2-OT* and *IGF2BP2*. Among Tervuren and Sheepdogs, *HDAC2* putative regulatory variants were present at comparatively high frequency and were associated with GC. Here, we describe a complex genetic architecture governing GC in a dog model, including genes such as *PDZRN3*, that have not been associated with human GC.

INTRODUCTION

Gastric cancer (GC) is the fifth most common malignancy and fifth leading cause of cancer deaths in humans globally.¹ Treatment options for GC are limited and typically include gastrectomy and chemotherapy, with a 5-year age-adjusted survival rate of 20-40%.²⁻⁵ Gastric adenocarcinoma comprises approximately 90% of all GCs⁵ with non-cardia involvement in >80%.⁶ The two major histological subtypes are intestinal and diffuse, based on Lauren classification.⁷ A stepwise progression from chronic inflammation through metaplasia and dysplasia characterizes the intestinal subtype.^{8,9} *Helicobacter pylori* infection is a strong environmental risk factor for intestinal GC in humans.¹⁰ The genetic basis of familial intestinal GC cases is poorly understood and likely polygenic.¹¹ The diffuse subtype is highly infiltrative and has a worse prognosis.¹² Diffuse GC may be caused by germline hereditary mutations in *CDH1*, encoding E-cadherin, or *CTNNA1*, which encodes part of the adherens junction protein complex.^{13,14} Approximately 10-20% of GCs show familial aggregation,¹⁵ but a genetic cause has only been identified for 1-3% of GCs.¹⁶ Human GC genome-wide association studies (GWASs) have identified common polymorphisms associated with sporadic cases of GC in some populations.¹⁷⁻¹⁹ However, the genetic risk factors underlying most cases are unknown.¹⁵

Gastric adenocarcinoma represents less than 1% of all neoplasias in dogs but disproportionately affects the Belgian Tervuren and Belgian Sheepdog breeds, indicating a genetic predisposition. In the Dutch

Tervuren and Sheepdog populations, GC incidence is very high at 5-8% (unpublished data), and predisposition in these two breeds has been described in multiple countries spanning many years.²⁰⁻²⁴ GC in the Tervuren and Sheepdog is clinically and histopathologically similar to human gastric adenocarcinoma, impacting the lesser curvature of the stomach.^{25,26} Both intestinal and diffuse GC occur at approximately equal rates in Tervuren and Sheepdogs.²⁶ Associations with *Helicobacter* species have not been identified in dogs.^{25,27} Consistent with the middle aged to older onset observed in humans, GC age at diagnosis is commonly 8-10 years old in these breeds.^{25,28} Humans are typically asymptomatic in early tumor stages, hindering early diagnosis and curative treatment.²⁹ In dogs, initial presentation of clinical signs also occurs at advanced disease and, like humans, includes inappetence, weight loss, and vomiting.²⁵ Definitive diagnosis is made by histological evaluation of endoscopic tissue biopsies.^{26,30} Surgery, with or without adjuvant chemotherapy, is the only potential curative treatment. However, even with surgery, the prognosis is very poor, with a median survival time of 178 days in dogs.³¹ Given the late stage of diagnosis and poor treatment outcomes, humane euthanasia is often elected for dogs.

The genetic structure of purebred dogs, shaped by population bottlenecks, founder events, and popular sire effects, has resulted in high within-breed homogeneity and, in some cases, enrichment for deleterious alleles.³²⁻³⁴ These factors facilitate detection of disease alleles compared to the confounding effects of heterogeneous populations in human genetic studies,^{35,36} with far fewer dogs needed than in human studies to identify high impact variants underlying complex traits.³⁷⁻³⁹ Causal genes and variants are often shared between dogs and humans,⁴⁰⁻⁴⁴ and dogs are a proven model for discovery of novel genes in human disorders.⁴⁵⁻⁴⁷ The Sheepdogs (also known as Groenendael) and Tervuren are members of the closely related Belgian shepherd dog group, which consists of four breeds, or varieties, based on coat type and color who share founder individuals. The most popular variety is the Belgian Malinois,⁴⁸ which is notably not at increased risk for GC.²⁰⁻²⁴ The contrasting frequencies of spontaneous GC in Belgian shepherd breeds and their shared genetic background make these dogs an ideal model to elucidate the genetic basis of GC.

The aim of this study was to identify loci and candidate variants underlying GC susceptibility in Tervuren and Sheepdogs. We hypothesized that multiple loci at variable frequencies would contribute to GC risk, and that Belgian Malinois may possess a protective haplotype(s) that is at, or approaching, fixation within the breed. Applying Bayesian mixture model GWAS and cross-population number of segregating sites by

length (XP-nSL) analyses to a Tervuren and Sheepdog cohort of 200 cases vs. 270 controls, we identified 15 loci that have approximately 92% predictive accuracy for GC phenotype. Across-breed selection analyses in an independent cohort of Tervuren and Sheepdogs compared to Malinois revealed additional risk and protective haplotypes.

RESULTS

SNP array imputation and study populations

SNP array data were imputed for 426 Belgian Tervuren, 259 Belgian Sheepdogs, and 185 Belgian Malinois, using a multi-breed reference panel of high coverage whole genome sequence (WGS) from 1,143 canines, including 20 Tervuren, 13 Sheepdogs, and 8 Malinois (Table S1). Genotype concordance was assessed by comparing imputed datasets between individuals included on multiple SNP array platforms (See Methods). Average concordance per chromosome was >85% for all allele frequency bins. The average concordance per chromosome increased to >98% for variants with INFO score ≥ 0.9 ($n=42,223,441$); these variants were used for subsequent analyses (Figure S1).

GC phenotypes were known for 200 cases (151 Tervuren, 49 Sheepdog) and 270 controls (183 Tervuren, 87 Sheepdog; Table S2). Median age at GC diagnosis or most recent health update was 9.2 and 13.2 years for cases and controls, respectively (Figure 1A). No significant association of sex with phenotype was observed ($p=0.22$). Phenotypes were unavailable for the remaining 400 samples, and this cohort was used in the selection scan analyses.

Principal component analysis (PCA) of cases and controls revealed population separation by breed/variety, as well as partial overlap between samples of European (119 cases and 105 controls) and North American (81 cases and 165 controls) origin on the first two principal components (Figure 1B). Eleven dogs reported as Tervuren or Sheepdog clustered with the opposite breed. Segregation of coat color alleles in these breeds allows fawn or mahogany Tervuren to produce solid black puppies, who may be reported as Sheepdogs, and vice versa.⁴⁹ Breed classifications were based on PCA clusters for subsequent analyses.

Bayesian GWAS and selection scan analysis identify multiple associated loci

Bayesian mixture model GWAS was performed for 200 cases and 270 controls, using 576,754 variants after

minor allele frequency and LD pruning. Bayesian mixture modeling via BayesR⁵⁰ simultaneously estimates the effect size of all variants rather than individual markers as in traditional GWASs, thus this approach has greater power for identifying segregating genetic risk factors for complex traits.^{51–53} Eight independent associated loci classified as having a moderate (effect size ≥ 0.001) effect on GC phenotype were identified across six autosomes (Figure 2A, Table 1). The locus of greatest effect size was on CFA20 and spanned only the *PDZRN3* gene. On CFA12, a 2Mb gene-dense region in high LD ($r^2 > 0.8$) included major histocompatibility complex (MHC) class II and some class I genes. Three independent signals surpassed moderate effect on CFA26 and were proximal to the *PCDH15* and *PTEN* genes (Figure 2B). Additional signals approaching moderate effect were identified on CFA10, CFA11, and CFA24 (Table 1, Table S3).

XP-nSL analysis⁵⁴ was used to compare cases and controls to identify regions of increased homozygosity and haplotype length in either group (Figure 2C, Table 1). The strongest signal replicated the CFA26:38Mb association from the GWAS at *PTEN*. An additional selection signal was identified proximal to the CFA26:32Mb GWAS signal. These regions demonstrated extended haplotype homozygosity among cases and are associated with GC risk (Table 1, Table S4). Regions within the 99.99th percentile (Table S4) also included loci on CFA13:9.9Mb, CFA17:54Mb, and CFA17:55.3Mb, demonstrating increased homozygosity among controls (CFA13) or cases (CFA17; Table 1). Gene ontology of the CFA17 regions indicated enrichment for immunoglobulin genes (FDR $p=0.024$), including *IGSF3*, *CD2*, *CD58*, and *CD101*.

Two loci are associated with intestinal gastric cancer subtype

For a subset of the cases, the diffuse ($n=25$) vs. intestinal ($n=26$) histopathological tumor subtype was known.²⁶ Fisher's exact tests were performed at the lead variant/haplotype for each of the aforementioned loci to detect significant associations with tumor subtype. The CFA17:54Mb (Fisher's 2-tailed $p=0.016$, OR=3.61, 95% CI=1.28-10.18) and CFA28:24.8Mb (Fisher's 2-tailed $p=0.028$, OR=2.62, 95% CI=1.17-5.86) loci were each associated with intestinal tumor subtype (Table S5). Significant associations with the diffuse subtype were not detected.

Fifteen loci predict gastric cancer phenotype with greater than 90% discriminatory accuracy

Bayesian GWAS and XP-nSL case-control analyses identified 15 independent loci (See Methods). At these

loci, the total number of risk alleles/haplotypes ($2n=30$) observed among cases averaged 20.3 compared to 15.9 among controls (two-tailed t -test $p=1.31 \times 10^{-61}$). Malinois ($n=149$), who are at low risk for GC, possessed an average of 15.4 risk alleles, clustering with the controls (Figure 3A). In the imputation reference panel, representing 227 other breeds, the average number of risk alleles per individual ($n=1,102$) was 13.6, which is within the range observed in controls (Table S6). Age at GC diagnosis was negatively correlated with total number of risk alleles ($p=0.00184$), *i.e.*, dogs possessing high numbers of risk alleles were diagnosed with GC at younger ages (Figure S2).

To identify the best model for predicting GC phenotype, additive models including top GWAS variants and three-SNP haplotypes of case vs. control XP-nSL loci were compared using ANOVA Chi-squared tests (Table S7). Significant evidence of epistatic interactions between the 15 loci was not observed, thus additive models weighting the effects of each locus were applied, rather than interactive models. A 13-locus model including the eight loci of moderate effect and five loci approaching moderate effect threshold from the GWAS was a significant improvement over the eight moderate loci alone ($p=4.32 \times 10^{-11}$). Adding two independent loci identified through XP-nSL significantly improved GC phenotype prediction ($p=0.025$), thus including the top 15 loci from the genomic analyses produces the simplest and most accurate predictive model. Five-fold cross-validation of this 15-locus additive model within the GWAS case-control cohort yielded an ROC-AUC of 0.918 (95% CI=0.893-0.943), indicating that this model predicts GC phenotype with greater than 90% discriminatory accuracy (Figure 3B). CFA12 and CFA20 genotypes were the most influential predictors of phenotype in the model (Figure S3).

Belgian shepherd across-breed comparisons identify GC risk and protective loci under selection

We hypothesized that some loci under selection in Malinois may confer protection from developing GC, based on the breed's low risk despite their close genetic relationship to the high-risk Tervuren and Sheepdogs.^{23,24,26,55} Further, some loci at high frequency in both Tervuren and Sheepdogs may contribute to their GC predisposition. An independent population of dogs (non-GWAS cohort) was used to identify regions at, or approaching, fixation in the breeds. After relationship pruning, 69 Tervuren, 82 Sheepdogs, and 149 Malinois were included in XP-nSL (Figure 4A, Table S8), F_{st} (Figure S4, Table S9), and runs of homozygosity (ROH; Figure S5, Table S10) analyses.

The strongest evidence for shared selection among the Tervuren and Sheepdog was at CFA4:24.2-24.4Mb (haplotype frequency = 95.4%), with the locus observed in all three selection analyses (Table 2). The gene-rich region includes genes that are dysregulated in human GC tumors, such as *PLAU*, *FUT11*, and *VCL*. Strong selection among Tervuren and Sheepdogs was also identified by XP-nSL and Fst on CFA12:70Mb, including *HS3ST5* and the upstream region and first exon of *HDAC2* (Table 2). These regions show significant enrichment for genes involved in the biosynthesis of heparan sulfate and heparin (FDR $p=0.023$) based on *NDST2* on CFA4 and *HS3ST5* on CFA12.

Six loci under strong selection in the Malinois were identified through XP-nSL (Figure 4A). A haplotype including *SOX2-OT* (CFA34:14.3-14.4Mb) was present among 100% of Malinois, with 91.3% haplotype frequency (Table 2). The Malinois haplotype was observed at significantly higher frequency among the GWAS control cohort (46% in controls vs. 34% in cases, Fisher's 2-tailed $p=0.00047$, OR=0.61, 95% CI=0.47-0.80), suggesting a protective effect. A second locus under selection in Malinois at CFA34:18.3-18.5Mb, harboring *IGF2BP2*, is also overrepresented among controls (24.1% haplotype frequency in controls vs. 11.5% in cases, Fisher's 2-tailed $p=1.22 \times 10^{-6}$, OR=0.41, 95% CI=0.28-0.59). The remaining Malinois XP-nSL haplotypes were not significantly associated with GC phenotype in the GWAS cohort. Adding the CFA4, 12, and 34 loci individually did not significantly improve upon the 15-locus GC prediction model (Table S7). The model including the CFA4 locus was slightly preferred by leave one out cross validation information criterion (loocv) but was not significantly better than the base model.

Rare variants confer risk or protection for gastric cancer

To identify candidate variants driving GC susceptibility in the Tervuren and Sheepdog, critical intervals defined through the above genomic analyses were interrogated in high-coverage WGS data for coding or regulatory variants unique to these breeds or rarely observed in other breeds (See Methods). Variants present on a risk or protective haplotype were pruned according to their effect on protein or PhyloP conservation scores and their presence in human (hg38) syntenic regulatory elements.

Two coding and nine non-coding variants were associated with GC in the case-control cohort, having *P*-values less than or equal to those of the lead variants from GWAS or selection scans (Table 3). Within the GWAS identified regions, coding variants predicted to impact protein function were present in *CD101*

(XP_533019.3:p.Leu934Pro) and *DCLRE1A* (XP_535018.3:p.Leu326Arg). The former is on the risk haplotype and replicates the lead variant association with intestinal tumor subtype. The *DCLRE1A* variant was identified on a control haplotype and did not replicate the intestinal subtype association of the CFA28 lead variant. The CFA28 locus may harbor both risk and protective haplotypes relevant to GC. Variants within GWAS identified regions were present in multiple syntenic human enhancer elements for *EPAS1* (*HIF2A*) and were associated with GC (Table 3). Putative regulatory variants of *PTEN* were also identified, including within an enhancer demonstrating *HDAC2* binding,⁵⁶ and within 3'UTR of human *PTEN* transcripts (Figure 2B, Table 3). The only lead GWAS variant to meet the allele frequency filtering criterion was on CFA12 (alternate allele AF=8.3%); however, it was filtered out at the conservation score step (PhyloP = -0.658). The variant does lie within a human syntenic enhancer (GH06J031896).

Within the CFA4 region of strong selection in Tervuren and Sheepdogs, one non-coding variant approached significantly increased frequency among GC cases ($p=0.064$). This variant is syntenic to a human super enhancer (SEdb2.0 ID: SE_02_152500349) expressed in stomach tissue and interacting with *PLAU* and *FUT11*. The CFA12:70Mb Tervuren and Sheepdog haplotype harbors two putative regulatory variants of *HDAC2* that were significantly associated with GC in the GWAS cohort ($p=0.0020-0.013$). These variants lie within a syntenic human super enhancer (SEdb2.0 ID: SE_02_019600073) and promoter element expressed in the stomach. The CFA12:70,107,159 variant lifts over to an untranslated region of human antisense-RNA (*HDAC2-AS2*) regulating *HDAC2* and *HS3ST5* (Figure 4B). Incorporating the candidate variants individually into the 15-locus model did not significantly improve upon utilizing the lead variants (Table S11).

DISCUSSION

GC is a leading cause of cancer-related deaths in humans worldwide with poor treatment options and prognosis. The genetic etiology is largely unknown. In dogs, Belgian Tervuren and Belgian Sheepdog breeds are at a uniquely high risk of developing GC, while their close breed relative, the Belgian Malinois, is at low risk. These breeds represent a naturally occurring model to elucidate the genetic basis of GC. Here, we have conducted the first GC GWAS in dogs, uncovering 15 loci that predict the phenotype with 92% discriminatory accuracy. These include the tumor suppressor gene *PTEN*, as well as genes that have not yet been associated with human GC, such as *PDZRN3* and *KLHL29*. Leveraging the contrasting GC incidences

between these breeds, we performed across-breed selection scans and uncovered an additional risk locus at *HDAC2* approaching fixation in Tervuren and Sheepdogs and two protective loci at *SOX2-OT* and *IGF2BP2* at high frequency in Malinois. These genes have critical roles in human GC progression and represent potential therapeutic targets.^{57–60}

Bayesian GWAS for GC in Tervuren and Sheepdogs revealed 13 loci associated with GC susceptibility. The top variant is intronic to *PDZRN3* (*LN3*), an E3 ubiquitin ligase with roles in angiogenesis and cell migration through non-canonical Wnt signaling pathways.^{61,62} In humans, *PDZRN3* ubiquitinates *DVL3*,⁶¹ which is overexpressed in diffuse GC,⁶³ and *PDZRN3* is reported to regulate connexin 43 (Cx43) protein stability in gap junctions.⁶⁴ Connexins are linked to multiple cancers,^{65,66} and Cx43 is dysregulated in GC tumors.^{67–69} *PDZRN3* is a promising therapeutic target in colorectal cancer.⁷⁰ A second locus of moderate effect harbors *KLHL29*, a member of the Kelch-like (KLHL) protein family that bind to specific substrate protein targets for ubiquitination by the cullin3 E3 ubiquitin ligase complex. KLHLs have been implicated in human gastrointestinal tumors,⁷¹ including *KLHL5*, *KLHL9*, and *KLHL19/KEAP1* in GC.^{72–74} A target of *KLHL29*, *DDX3*, regulates E-cadherin,^{75–78} which is often inactivated in diffuse GC,⁷⁹ and germline variants in *CDH1*, which encodes E-cadherin, are associated with hereditary diffuse GC in humans.^{13,14}

Three independent GWAS signals were detected on CFA26, including a locus at *PTEN*, which was also identified in the case-control XP-nSL analysis. *PTEN* is a well-known tumor suppressor gene whose loss has been identified in many tumors, including human GC.^{80,81} Decreased expression of *PTEN* has also been reported in the gastric mucosa of first-degree relatives to GC patients.⁸² We identified variants syntenic to two human enhancers that interact with *PTEN*, one of which is known to be expressed in the stomach, and a 3'UTR variant. Additional putative regulatory variants were identified within the CFA10 GWAS critical interval syntenic to multiple human enhancers of *EPAS1*, also known as *HIF2A*. *EPAS1* expression is induced under hypoxia, and it is upregulated in many malignancies, including GC where it contributes to metastasis.^{83–90}

The CFA12 locus has a strong effect on GC phenotype prediction. This gene-dense region harbors major histocompatibility complex class II, as well as some of the class I genes. Alleles of MHC class II HLA genes have been associated with risk or protection from human GC,^{91,92} as well as autoimmune gastritis,^{93,94} which can progress to GC.⁹⁵ The relationship between gastritis and GC in dogs is unclear.²⁵ Further work is necessary to identify the specific variants at this locus contributing to GC predisposition in Tervuren and

Sheepdogs. Additional associations with immune-related genes were identified through our case vs. control selection analyses. A CFA13 region of extended homozygosity among controls contains *EBAG9*, also known as tumor associated antigen *RCAS1*, which is involved in tumor immune evasion and demonstrates aberrant localized expression in human GC tumors.^{96–99} There may be a recessive protective variant in this region or a dominant risk factor present among cases, which were more heterozygous in this region. Enrichment of immunoglobulin genes was observed within an extended case haplotype on CFA17:54Mb. Variant filtering revealed a rare protein coding variant within *CD101* that disrupts a conserved amino acid, is enriched among cases, and is associated with intestinal tumor subtype in our cohort. *CD101* has roles in inhibiting T cell proliferation, promoting development of regulatory T cells, and controlling inflammation, and germline *CD101* coding variants have been shown to have a proinflammatory effect in humans.^{100–102}

Within dog breeds, many loci are under strong artificial selection for breed-defining physical and behavioral traits, such as body size, coat color and type, intelligence, and trainability.¹⁰³ As a result, deleterious alleles can increase in frequency through hitchhiking or pleiotropic effects.^{43,104,105} We performed across-breed analyses to identify regions under divergent selection in the high-risk Tervuren and Sheepdog compared to the low-risk Malinois. The close genetic relatedness of these breeds⁵⁵ minimized the number of loci resulting from this approach, highlighting a subset of genomic regions contributing to GC susceptibility. In Malinois, a haplotype including the lncRNA *SOX2-OT* is under strong selection and is associated with protection from GC. A SNP intronic to *SOX2-OT* was associated with decreased risk of developing GC in humans.¹⁰⁶ *SOX2-OT* dysregulation occurs in various malignancies,¹⁰⁷ including GC where it is overexpressed, leading to epithelial-mesenchymal transition.⁵⁷ A key role of *SOX2-OT* is modulating expression of the *SOX2* gene.¹⁰⁸ *SOX2* possesses antitumor effects in GC and demonstrates positive regulation of *PTEN*.^{109,110} A gradual reduction in *SOX2* expression has been observed during GC tumorigenesis.¹⁰⁹ In addition to critical roles in the stomach,^{111,112} *SOX2-OT* and *SOX2* are important for neural differentiation,^{108,113,114} and this locus may be under selection in Malinois for cognitive traits.

A second protective locus was identified in Malinois at *IGF2BP2*. In addition to roles in metabolism and type 2 diabetes, *IGF2BP2* contributes to cancer proliferation and cancer stem cell maintenance.¹¹⁵ Overexpression of *IGF2BP2* promotes GC progression and epithelial to mesenchymal transition in human GC tumors and cell lines,^{58,116,117} and SNPs have been associated with chemotherapeutic response in GC.¹¹⁸

IGF2BP2 is a recognition binding protein for N6-methyladenosine (m⁶A) RNA methylation, involved in mRNA stability, and its interaction with multiple targets promotes human GC tumor progression and represents a potential therapeutic target.^{58,60,119,120} The use of small molecule inhibitors of IGF2BP2 in other cancers have shown promising anti-cancer effects.^{121,122}

Among Tervuren and Sheepdogs, a shared haplotype on CFA4 is under strong selection and harbors variants within human syntenic enhancer elements of *PLAU* and *FUT11* that are skewed toward GC cases ($p=0.064$). *PLAU* is a driver of tumor progression in multiple cancers and is overexpressed in GC.^{123,124} *FUT11* is a fucosyltransferase that is overexpressed in GC and a predictor of poor prognosis.^{125,126} Gene amplification and missense mutations of *FUT11* are commonly observed in GC tumors.¹²⁶ This region also contains *NDST2*, which plays a role in modification of heparan sulfate chains via deacetylation and sulfation of glucosamine.¹²⁷

A second shared haplotype under strong selection in Tervuren and Sheepdogs is on CFA12 and includes *HS3ST5* and the upstream region and first exon of *HDAC2*. Like *NDST2*, *HS3ST5* is in the heparan sulfate metabolism pathway. In GC, heparan sulfate proteoglycan levels decrease with tumor progression,¹²⁸ and dysregulation of other heparan sulfate modifying genes is associated with GC progression and prognosis.¹²⁹ However, the specific role of *HS3ST5* in GC is unclear.¹³⁰ *HDAC2* is a class I histone deacetylase that has roles in transcriptional regulation, cell proliferation, and tumorigenesis.⁵⁹ We identified putative regulatory variants, syntenic to a human super enhancer, that are associated with GC and are rarely found outside of the Tervuren and Sheepdog breeds. *HDAC2* is upregulated in human GC tumors, and knockdown reduces cell proliferation and restores *CDKN2A* expression in GC cell lines.^{131–135} *HDAC2* interacts with CDH1 (E-cadherin), complete loss of which causes diffuse GC.¹³⁶ HDAC inhibitors have been shown to have anti-proliferative effects in GC cell lines and are a potential therapy for GC.¹³⁵

Multiple genes identified herein have reported gene/protein interactions (Figure S6), including downregulation of *PTEN* by *HDAC2* and formation of HDAC2-SOX2 protein complexes in cancer stemness.⁵⁹ While dysregulation of many of these genes has been described in human GC, this work supports the need for functional investigation of their interactions in GC tumors, which may facilitate development of targeted therapies.

Limitations of the study

The five-fold cross validation approach applied herein indicates 92% predictive accuracy of the model; however, additional samples will be necessary to assess model accuracy in an independent validation cohort to ensure the model is not overfitting. This study does not account for environmental risk factors for GC. Population substructure was present based on geographic origin and breed; however, we applied a relationship matrix in the GWAS to correct for this. Statistical evidence of genotype interactions was not observed. This is likely due to limitations in sample size to robustly represent all possible combinations of genotypes across the 15 loci. Because GC is rare across dog breeds, we applied a variant filtering approach that identified candidate variants at low frequency outside of Tervuren and Sheepdogs. However, there may be common polymorphisms at one or more risk loci that rarely exist in combination in other breeds. Further studies will be necessary to assess relevant common polymorphisms. Next steps also include functional validation of putative regulatory variants identified herein. While this work highlights shared genetic risk factors for GC in two high-risk breeds, future studies should include definitive Lauren tumor subtypes to determine whether additional loci distinguish predisposition to intestinal vs. diffuse GC.

Conclusions

This is the first study in dogs to identify genetic variants associated with GC, revealing 15 loci predictive of GC phenotype. The shared genetic background and contrasting incidences of GC among high-risk Tervuren and Sheepdogs compared to low-risk Malinois enabled identification of additional genetic risk or protective factors at high frequency in either population. This work establishes a genetic risk model for these breeds that will enable selective breeding to reduce disease frequency, as well as routine monitoring for high-risk individuals. Genes identified herein overlap with those dysregulated in human GC tumors and include previously unassociated genes, such as *PDZRN3* and *KLHL29*. Here we demonstrate that GC in Tervuren and Sheepdogs reflects the genetic complexity observed in humans, in addition to clinical heterogeneity, and thus has tremendous translational potential to impact human health.

MATERIALS & METHODS

Samples

Whole blood or buccal samples were obtained from purebred Belgian Shepherd breeds with informed owner consent in accordance with regulations of each institution's animal care and use committee. DNA was isolated according to the Puregene DNA Isolation protocol (Gentra) or standard phenol-chloroform extraction. Cases were diagnosed via clinical signs and ultrasound, histopathology, and/or cytology. Approximately 84% of cases had diagnoses confirmed by histopathology. For inclusion as controls, dogs were at least 10 years of age at sample collection or last health update with no owner-reported history of GC or clinical signs. A total of 224 samples were collected from Europe (119 cases and 105 controls) and 246 samples from North America (81 cases and 165 controls). Additional samples and/or SNP array genotype data were collected irrespective of health status and age at collection for breed-based population comparisons (Belgian Tervuren, n=95; Belgian Sheepdog, n=120; Belgian Malinois, n=185).

Whole genome sequence reference panel

Whole genome sequence data were generated for 20 Belgian Tervuren (9 cases, 8 controls, and 3 of unknown phenotype) and 13 Belgian Sheepdogs (7 cases, 4 controls, and 2 of unknown phenotype), ranging from 9.6 to 48.2X coverage, with a median coverage of 16X. Raw data were aligned to CanFam3.1, and these genomes were included in a multibreed joint-called reference panel comprising 1,143 canines representing 230 breeds, including 8 Belgian Malinois and 1 Belgian Laekenois (Table S1). Alignment to CanFam3.1 and jointcalling were performed as described by Plassais et al. (2019).¹³⁷

Phasing was performed for each chromosome, using SHAPEIT5¹³⁸ phase_common command in 30Mb chunks with 10Mb overlapping, followed by ligation with SHAPEIT5 ligate command. The output was converted to VCF and indexed using BCFtools¹³⁹ prior to input into imp5Converter to create an IMPUTE5¹⁴⁰ file for use in the imputation pipeline.

Pre-imputation SNP array concordance and quality control

SNP array genotypes were generated on either the Illumina CanineHD BeadChip array (173k markers), Embark's custom high-density Illumina 220k SNP arrays,¹⁴¹ or Affymetrix Axiom Canine HD Array (670k

markers; Table S2). Each array dataset (target panel) was imputed independently prior to merging all three datasets. First, the conform-gt tool was used to unify allele strand between the target and reference panels and to remove variants that were only present in the target panel.¹⁴² Fifteen dogs were genotyped on both the 173k and 220k arrays, 23 dogs on the 173k and 670k arrays, and 7 dogs on the 220k and 670k arrays. Three of these dogs were genotyped on all three arrays. To identify discordant SNPs between duplicate dogs on the arrays, we generated transposed (--export A-transpose command using PLINK v2.00a3.7LM (24 Oct 2022)) files for each chromosome of each array dataset, where the genotypes were coded 0, 1, 2 for homozygous alternate, heterozygous, and homozygous reference, respectively.¹⁴³ The transposed files were used as input in the python script "imputation_accuracy.py," which compares all sites in common across every dataset type for all duplicate dogs. The script removes all sites with allele mismatches. The output is a matrix of values calculated by subtracting the genotype codes (0,1,2) in one dataset from the other. A value of 0 for a given site means that the genotypes for the two dogs in the two datasets are the same. A site that was discordant in any of the comparisons was removed from all three genotype datasets. These matrix outputs were used in the python script "analysisOfTSV.py" to calculate concordance across sites in common for each chromosome per dog, and then concordance across dogs in common per site. A site that was discordant in any of the genotype comparisons in at least one of the duplicate dogs was removed from all genotype datasets, resulting in 5,300 markers removed from 173k, 5,518 markers removed from 220k, and 5,602 markers removed from the 670k.

SNP array imputation

Genotype datasets were filtered in PLINK 1.9¹⁴³ to remove SNPs with MAF <0.01 and/or call rates <90%, and then split into chromosomes. PLINK files were then converted into VCF files in VcfCooker v1.1.1 (<https://genome.sph.umich.edu/wiki/VcfCooker>) then indexed using BCFtools index. Phasing was performed using SHAPEIT5 v5.1.0 phase_common command incorporating available parent-offspring information (four parent-offspring pairs and one trio for 173k; one parent and five offspring for 220k) and genetic map files for the 173k and 220k datasets. Default SHAPEIT5 options were used with the exception of pbwt-depth, which was set to eight, to improve accuracy while not compromising speed. For the imputation step, imp5Chunker, part of the IMPUTE5 package, was first used to split the chromosomes into ~5-10Mb chunks with ~500kb buffer region, which were used to impute across using the reference panel (--h option in IMPUTE5 v1.1.5). The

imputed chunks were ligated back together using BCFtools concat. Finally, PLINK2 was used to generate pgen format PLINK files.

Imputation concordance

To calculate imputation concordance between dogs included in multiple datasets, we used the same scripts as above. The imputed pgen output files were converted into .raw files and fed into “imputation_accuracy.py” and the output matrix files were fed into “analysisOfTSV.py.” Concordance values for all variants shared across the imputed output files were calculated in non-reference allele frequency bins of 0.05 increments. After imputation and merging, variants with INFO score ≥ 0.9 (Table S12), variants missing in at least one array dataset ($n=17,222,525$) were removed using PLINK1.9,¹⁴⁴ resulting in 42,223,441 remaining variants.

Genome-wide association analyses

Pruning for minor allele frequency <0.05 in the GC GWAS cohort yielded 5,998,047 variants. For PCA, variants were pruned for linkage disequilibrium ($r^2 > 0.4$) using the --indep-pairwise option with 50 kb windows in PLINK1.9, resulting in 130,872 variants. PCA was conducted using eigensoft’s smartpca.¹⁴⁵ For GWAS, a genomic relatedness matrix was created in GEMMA¹⁴⁶ using the LD-pruned dataset of 130,872 variants. The eigenvectors were subsequently extracted and used as covariates for GWAS analyses to correct for population substructure. BayesR⁵⁰ was first used to determine effect sizes of the 5,998,047 variants, using 300,000 iterations, 100,000 burn-in, permutations, and GEMMA eigenvector covariates. The overall estimated effect each SNV has on the phenotype is updated iteratively. In each iteration, BayesR estimates the probability of an individual variant belonging to one of four normal distributions: zero, low, moderate, and high effect, each explaining 0%, 0.01%, 0.1%, and 1% of the genetic variance, respectively. Critical intervals were defined by the most centromeric and telomeric variants in pairwise LD ($r^2 \geq 0.8$) with the lead variant in each independent GWAS peak. To obtain an accurate estimation of effect sizes at each locus, which are diluted by variants in strong LD,^{50,147} LD-based clumping in PLINK1.9¹⁴⁴ was used to filter variants with $r^2 \geq 0.9$ relative to the lead variant within 50 kb windows, while maintaining the lead variant of greatest effect size. In total, 576,754 variants were retained, and BayesR analysis was repeated using the aforementioned parameters.

Genotype models

For regions identified in GWAS and XP-nSL that overlapped or were in high LD, a single representative variant or haplotype was used for subsequent analyses, specifically the CFA17:54,006,088 three-SNP haplotype, CFA26:33,035,867 variant, and CFA26:38,180,502 variant (Table 1). Preference was given to GWAS lead variants. The case predominant allele or haplotype was assigned as “risk,” and the total number of risk alleles/haplotypes present in each individual in the case-control population was determined. A two-tailed *t*-test was used to calculate the difference between the risk allele load of cases and controls. The total risk allele/haplotype count in the Belgian Malinois (n=149) was calculated using the same methodology. Lastly, the total risk allele count was calculated across canines in the imputation reference panel, excluding Tervuren, Sheepdogs, and Malinois (n=1102).

Pairwise interactions between BayesR and XP-nSL lead variants were analyzed using the “--epistasis” command in PLINK1.9,¹⁴⁴ and multiple testing correction was applied to determine significance. Generalized linear models were created in R¹⁴⁸ to determine the model that best explains GC risk in the GWAS cohort. These models were compared via ANOVA chi-square testing. Additionally, models were generated using a Bayesian generalized linear model via the `stan_glm` command in the `rstanarm` package¹⁴⁹ and evaluated using leave-one-out cross-validation information criterion (LOOIC) performed using 25,000 iterations, 5,000 burn-in, and a thinning parameter of 25.

K-fold cross-validation was used to evaluate the additive model which incorporated the 15 loci identified from the GWAS and case vs. control XP-nSL analyses. The GWAS cohort was randomly partitioned into five subsets. An iterative approach was applied wherein four subsets were used to train the model and the remaining subset was used to test the model. The receiver operating characteristic curve and area under the curve (ROC-AUC) were calculated using the `pROC` package in R for each of the five iterations and the average. Variable importance of each locus was calculated using the `caret` package in R to determine the loci having the greatest influence within the model.¹⁵⁰

Selection scans

Cross-population number of segregating sites by length (XP-nSL)

Two populations of Belgian Shepherd dogs, the GWAS cohort (n=470) and an independent cohort with

unknown GC phenotypes (n=401), were examined for evidence of selective sweeps that may contribute to overall GC risk. For the independent cohort with unknown GC phenotypes, dogs were split by breed (n=185 Malinois, 120 Sheepdogs, and 95 Tervuren), and a minor allele frequency of 0.1 was applied prior to genetic relatedness pruning with GCTA grm-cutoff¹⁵¹ of 0.3, resulting in 149 Malinois, 82 Sheepdogs, and 69 Tervuren for the across-breed analysis.

The initial genotyping dataset comprising 42,223,441 variants was LD-pruned using a pairwise r^2 of 0.99 and a sliding window of 50 kb, with 2,039,330 variants remaining for analysis. SelScan default settings were used to perform XP-nSL followed by “norm”^{54,152} to identify regions under selection in cases or controls in the GWAS cohort, as well as to compare Tervuren and Sheepdogs to the Malinois in the cohort of unknown GC phenotypes. Variants within the top 99.99% threshold were further examined and used to identify critical intervals. Haplotype frequencies were calculated using the lead variant and two flanking variants reported by XP-nSL. Fisher’s exact association tests were performed in the case-control GWAS cohort comparing the frequency of the predominant haplotype in the population under selection to the frequency of all other haplotypes.

Runs of Homozygosity & Fst

Runs of homozygosity (ROH) were assessed separately in the Tervuren, Sheepdogs, and Malinois in the cohort of dogs with unknown GC phenotype. The LD-pruned dataset of 2,039,330 variants was evaluated using the sliding window method in detectRuns.¹⁵³ The sliding window used 40 variants, of which one was allowed to be heterozygous, and a minimum ROH length was 500kb. As the data were imputed and sufficiently dense, minimum density and maximum gap size parameters were set to be ignored. The default window threshold value of 0.05 was used. Regions of interest were defined per breed as the genomic areas that were most commonly shared within that breed, specifically, those in the 99.9th percentile.

Fst was calculated in PLINK1.9¹⁴⁴ using the same population of dogs, comparing the Tervuren and Sheepdogs to the Malinois. For this analysis, the initial dataset of 42,223,441 variants was used and monomorphic variants were discarded (n=7,836,277 variants remaining). The top 0.01% of Fst values were extracted and variants within 1Mb were collapsed into one region.

Variant filtration

All variants present in the WGS reference dataset of 1,143 canids (n=33 Tervuren and Sheepdog) within the critical intervals identified through GWAS were extracted. Additionally, all variants within the XP-nSL critical intervals +/-100kb buffer on either side were extracted. Variants were filtered per individual for genotype quality score ≥ 20 via BCFtools. SNPs and INDELS with alternate allele frequency (AF) ≥ 0.25 in Tervuren and Sheepdogs and alternate AF ≤ 0.1 in all other canids were retained. To allow for the reference allele being a candidate GC variant, filtering was also performed to retain variants with alternate AF ≤ 0.75 in Tervuren and Sheepdogs and alternate AF ≥ 0.9 in other canids. Variant effect predictor (VEP) was used to annotate variants.¹⁵⁴ PhyloP scores from 241 vertebrates¹⁵⁵ were extracted, and variants with scores >1 were retained. Putative regulatory variants were identified from syntenic human regulatory elements using the hg38 GeneHancer v5.20 track and evidence of super enhancers was examined in SEdb 2.0.¹⁵⁶ SNPs and INDELS meeting the above criteria were manually examined in relation to all imputed variants within the critical region to determine if the identified variants segregated with a risk or non-risk haplotype as identified in the GWAS cohort. Fisher's exact two-tailed p -values were calculated for candidate variants in the imputed GWAS dataset (2n=940) to determine if allele frequencies were significantly different between cases and controls. Allele frequencies (Table 3) are reported for the entire reference population (2n=2286).

Enrichment analysis

All gene set enrichment analyses were conducted in STRING version 12.0.¹⁵⁷ Regions of interest were converted to their equivalent human positions using UCSC's LiftOver¹⁵⁸ and input into STRING using *Homo sapiens* as the organism.

DATA AVAILABILITY

The python scripts “imputation_accuracy.py” and “analysisOfTSV.py” used in the concordance part of the imputation pipeline are publicly available on github at:

https://github.com/JessicaHayward/imputation_accuracy_GC. SNP array and imputed data will be available at datadryad.org upon publication. SRA accession numbers for high coverage WGS are available in Table S1.

ACKNOWLEDGEMENTS

We thank the many owners, breeders, and veterinarians who contributed samples and clinical data for this study. We also thank Embark Veterinary, Inc. for providing anonymized array data. This material is based upon work supported by the Cornell Richard P. Riney Canine Health Center Research Grants Program, a grant made available to the College of Veterinary Medicine, Cornell University. This work was supported by grants from the American Kennel Club Canine Health Foundation (AMO), the Intramural Program of the National Human Genome Research Institute at the National Institutes of Health (EAO), The Hartwell Foundation (SRC), Albert C. Bostwick Foundation, Belgian Sheepdog Club of America, American Belgian Tervuren Club, Nederlandse Vereniging van Belgische Herdershonden, Belgische Herder Club Nederland, Deutscher Klub für Belgische Schäferhunde, the Tjech KCHBO, and the Nederlands Kankerfonds voor Dieren.

AUTHOR CONTRIBUTIONS

Writing-original draft - JME, SRC, SH, JJH

Writing-review & editing - JME, SRC, EAO, AMO, JMB, JJH, HF, SH, PJJM, PAJL

Investigation & Formal analysis - SRC, JME, SH, JJH, TRF

Data curation - JJH, SRC, SH, JMB, JME

Validation - SRC, TRF, JME

Visualization - SRC, JJH, JME

Resources - EAO, AMO, JMB, EM, HF, PAJL, PJJM, JME

Conceptualization - JME, SRC, EAO, AMO, PAJL, PJJM

Funding acquisition - JME, EAO, PJJM, SRC, AMO

DECLARATION OF INTERESTS

The authors declare no competing interests.

Table 1. Lead variants from Bayesian mixture model GWAS and XP-nSL of 200 case and 270 control Tervuren and Sheepdogs.

Lead Variant Position (CanFam3.1)	Critical Interval ^a	Analysis	BayesR Effect Size	Normalized XP-nSL	Genes within critical intervals ^b
CFA10:28,537,583	CFA10:28,152,671-28,821,673	GWAS	0.000741	-	<i>RBFOX2, APOL5, APOL6, MB, RASD2, MCM5, HMOX1, TOM1</i>
CFA10:48,471,044	CFA10:48,141,927-48,512,574	GWAS	0.000442	-	<i>PRKCE</i>
CFA11:15,714,902	CFA11:15,377,666-15,728,753	GWAS	0.000723	-	<i>GRAMD2B</i>
CFA11:38,106,700	CFA11:38,104,905-38,206,955	GWAS	0.000521	-	<i>ADAMTSL1</i>
CFA12:1,359,325	CFA12:1,359,325-3,451,568	GWAS	0.00225	-	115 genes and non-coding transcripts
CFA13:9,923,659	CFA13:9,831,529-10,417,265	XP-nSL	-	-4.53	<i>NUDCD1, ENY2, PKHD1L1, EBAG9, SYBU, SYBU-AS1</i>
CFA17:18,030,682	CFA17:17,946,446-18,030,682	GWAS	0.00154	-	<i>KLHL29</i>
CFA17:54,006,088	CFA17:53,974,863-54,445,506	XP-nSL	-	4.35	<i>CD58, IGSF3, CD2, PTGFRN, CD101, CD101-AS1</i>
CFA17:55,324,675	CFA17:55,013,848-55,324,755	XP-nSL	-	4.25	<i>GDAP2, WDR3, SPAG17</i>
CFA20:18,681,361	CFA20:18,646,458-18,718,242	GWAS	0.00661	-	<i>PDZRN3</i>
CFA24:41,167,902	CFA24:41,030,826-41,191,757	GWAS	0.000484	-	-
CFA26:31,661,685	CFA26:31,661,685-32,050,496	XP-nSL	-	4.30	-
CFA26:33,035,867	CFA26:32,881,257-33,106,991	GWAS	0.00161	-	<i>PCDH15</i>
CFA26:34,625,869	CFA26:34,612,341-34,848,102	GWAS	0.00201	-	-
CFA26:38,012,570	CFA26:37,794,344-38,275,926	XP-nSL	-	4.83	<i>ATAD1, CFL1P1, KLLN, PTEN, RNLS</i>
CFA26:38,180,502	CFA26:38,180,502	GWAS	0.00200	-	<i>RNLS</i>
CFA28:24,778,275	CFA28:24,295,247-24,847,272	GWAS	0.00120	-	<i>HABP2, NRAP, CASP7, PLEKHS1, DCLRE1A, NHLRC2</i>
CFA37:26,075,098	CFA37:26,075,098	GWAS	0.00105	-	<i>OBSL1</i>

^aGWAS critical intervals were defined by variants in high LD with the lead SNV ($r^2 \geq 0.8$). XP-nSL critical intervals represent regions in the top 0.01%.

^bAll genes within the critical interval, in CanFam3.1 or CanFam4, are reported with the exception of CFA12 because it contains 115 genes.

Table 2. Regions under selection in Tervuren and Sheepdogs or Malinois identified through XP-nSL.

Lead Variant Position (CanFam3.1)	Critical Interval ^a	Under selection ^b	Normalized XP-nSL	P-value ^c	Odds Ratio	95% CI	Genes within critical interval ^d
CFA2:80,184,506	CFA2:80,183,015-80,184,650	Malinois	-3.71	0.23	0.85	0.62-1.10	<i>IGSF21</i>
CFA4:24,246,918	CFA4:24,209,750-24,411,978 ^e	Tervuren & Sheepdog	4.03	0.064	1.77	0.97-3.22	<i>FUT11, CHCHD1, ZSWIM8, NDST2, CAMK2G, PLAU, VCL</i>
CFA4:25,307,748	CFA4:25,277,331-25,328,920	Tervuren & Sheepdog	3.84	0.25	0.83	0.61-1.12	<i>KAT6B, DUSP29</i>
CFA11:46,442,066	CFA11:46,438,442-46,442,066	Malinois	-3.79	0.95	1.01	0.78-1.31	<i>LINGO2</i>
CFA12:70,342,037	CFA12:70,079,430-70,351,863 ^f	Tervuren & Sheepdog	4.03	0.125	1.44	0.93-2.25	<i>HDAC2, HDAC2-AS2, HS3ST5</i>
CFA24:19,684,061	CFA24:19,622,112-19,684,730	Malinois	-4.01	0.37	0.88	0.67-1.15	<i>FKBP1A, SDCBP2</i>
CFA34:14,379,142	CFA34:14,368,043-14,413,380	Malinois	-3.78	0.00047	0.61	0.46-0.80	<i>SOX2-OT</i>
CFA34:15,178,285	CFA34:15,139,760-15,180,010	Malinois	-3.71	2.70x10 ⁻⁶	0.44	0.31-0.62	-
CFA34:18,529,581	CFA34:18,322,441-18,543,138	Malinois	-3.98	1.22x10 ⁻⁶	0.41	0.28-0.59	<i>SENP2, IGF2BP2</i>
CFA35:11,305,275	CFA35:11,305,256-11,305,334	Tervuren & Sheepdog	3.71	0.61	1.10	0.78-1.54	<i>ADTRP</i>

^aXP-nSL critical intervals represent regions in the top 0.01%.

^bBreed(s) possessing the selection signature (extended haplotype homozygosity).

^cFisher's exact association tests for XP-nSL three-variant haplotype with GC in the GWAS cohort. Two-tailed P-values are reported.

^dAll genes within the critical interval are reported.

^eOverlaps with a run of homozygosity in both Tervuren and Sheepdogs, as well as a region of high Fst.

^fOverlaps with a region of high Fst.

Table 3. Coding and putative regulatory candidate variants for gastric cancer susceptibility.

Candidate Variant Position (CanFam3.1)	REF/ALT Allele ^a	ALT AF Tervuren & Sheepdog (2n=66) ^b	ALT AF other canines (2n=2220)	GeneHancer Double Elite Element ID ^c	Element Type	Gene/Gene Target	PhyloP ^d	P-value Lead Variant ^e	P-value of candidate variant ^f	OR	95% CI
CFA4:24,304,634	G/A	0.88	0.034	GH10J073878	Enhancer	<i>PLAU, FUT11</i>	4.84	0.064	0.064	1.82	0.998-3.31
CFA10:48,313,427	A/G	0.29	0.032	GH02J046009	Enhancer	<i>EPAS1/HIF2A</i>	1.86	0.00015	2.92x10 ⁻⁶	1.91	1.46-2.50
CFA10:48,330,465	T/C	0.29	0.069	GH02J046029	Enhancer	<i>EPAS1/HIF2A</i>	3.12	0.00015	2.92x10 ⁻⁶	1.91	1.46-2.50
CFA10:48,378,432	G/T	0.26	0.011	GH02J046085	Enhancer	<i>EPAS1/HIF2A</i>	2.87	0.00015	2.92x10 ⁻⁶	1.91	1.46-2.50
CFA10:48,399,845	G/T	0.29	0.015	GH02J046116	Enhancer	<i>EPAS1/HIF2A</i>	1.05	0.00015	2.92x10 ⁻⁶	1.91	1.46-2.50
CFA12:70,106,510	C/T	0.83	0.073	GH06J113994	Enhancer, Promoter	<i>HDAC2, HDAC2-AS2</i>	1.29	0.125	0.0020	1.88	1.26-2.82
CFA12:70,107,159	C/CT	0.79	0.0059	GH06J113994	Enhancer, Promoter	<i>HDAC2, HDAC2-AS2</i>	1.13	0.125	0.013	1.60	1.11-2.29
CFA17:54,417,724	T/C	0.29	0.012	-	-	<i>CD101</i>	2.85	7.05x10 ⁻¹⁰	8.81x10 ⁻¹²	2.51	1.92-3.28
CFA26:37,880,356	G/GGT	0.53	0.066	GH10J087915	Enhancer	<i>PTEN, LIPJ, Inc-PAPSS2-7</i>	1.36	9.08x10 ⁻⁹	2.08x10 ⁻¹⁰	2.38	1.82-3.13
CFA26:37,917,914	A/G	0.55	0.072	-	3'UTR ^g	<i>PTEN</i>	1.58	9.08x10 ⁻⁹	2.08x10 ⁻¹⁰	2.38	1.82-3.13
CFA26:37,984,898	G/A	0.53	0.086	GH10J088058	Enhancer	<i>PTEN</i>	6.90	9.08x10 ⁻⁹	5.29x10 ⁻¹³	2.69	2.05-3.54
CFA28:24,720,256	A/C	0.44	0.038	-	-	<i>DCLRE1A</i>	2.89	0.00074	0.00027	0.60	0.46-0.79

^aReference allele (REF) is based on WGS reference panel, and alternate allele (ALT) is alternate to the reference.

^bAllele frequencies (AF) of the ALT allele are calculated using the WGS reference panel (n=1143).

^cGeneHancer double elite v5.20 IDs for regulatory elements encompassing the human (hg38) syntenic position of the dog candidate variant.

^dPhyloP scores from 241 vertebrates are based on the human (hg38) variant positions.

^eFisher's exact two-tailed *P*-value of BayesR lead variant or XP-nSL three-variant haplotype were calculated comparing GWAS cases (2n=400) to controls (2n=540).

^fFisher's exact two-tailed *P*-value of the imputed candidate variant genotypes for GWAS cases (2n=400) vs. controls (2n=540).

^gThis variant is syntenic to the 3'UTR of human *PTEN* transcripts per NCBI RefSeq (NM_001304718.2, NM_000314.8, NM_001304717.5)

FIGURE LEGENDS

Figure 1. Population structure of the case-control Tervuren and Sheepdog cohort. A) Box-and-whisker plots of age at diagnosis for cases and age at collection or most recent health update for controls. B) Principal component analysis (PCA) plot for the 470 dogs included in the case vs. control analyses with each point representing an individual.

Figure 2. Bayesian GWAS and selection scan analysis identify multiple loci associated with gastric cancer. A) Manhattan plot showing the results of Bayesian GWAS with 200 cases vs. 270 controls. The absolute value of individual variant effect sizes (y-axis) reported by BayesR for 576,754 variants after minor allele frequency and LD pruning is plotted by position (x-axis). The red line is at the moderate effect size (0.001) threshold. B) Top: A regional Manhattan plot of CFA26 shows three independent regions surpassing moderate effect. Pairwise linkage disequilibrium (r^2) between the top BayesR variant at CFA26:38Mb (purple diamond) and all other variants is color-coded. Bottom: UCSC image of the human (hg38) syntenic region for the CFA26:38Mb XP-nSL interval that overlaps the GWAS signal. Candidate variants meeting filtering criteria are indicated as blue vertical bars in the top track. GeneHancer double elite v5.21 enhancer (gray bars) and promoter (red bars) regulatory elements are below, with darkest colors indicating highest confidence. Clustered interactions of regulatory elements with protein coding genes and a lncRNA are shown; thick bars correspond to the regulatory element that interacts with the gene. The NCBI RefSeq Select and MANE gene track is at the bottom. C) Normalized XP-nSL values are plotted for 2,039,330 variants with dashed black lines demarcating the 99.99th percentiles. Positive values represent regions with extended haplotype homozygosity among cases, whereas negative values indicate extended haplotype homozygosity among controls.

Figure 3. Fifteen loci predict gastric cancer phenotype with greater than 90% discriminatory accuracy. A) Bar graph of the distribution of risk alleles among cases (n=200, red), controls (n=270, gray), and Malinois (n=149, blue). Cases have an average of 20.3 total risk alleles while controls have 15.9 ($p=1.31 \times 10^{-61}$), and Malinois cluster with controls. B) Receiver Operating Characteristics (ROC) curves for the 5-fold cross-

validated 15-locus additive model of GC risk in Tervuren and Sheepdog. Area Under the Curve (AUC) is reported for each fold, with the 95% confidence interval in brackets.

Figure 4. Belgian shepherd across-breed comparisons identify GC risk and protective loci under selection. A) Normalized XP-nSL values are plotted for 2,039,330 variants with dashed black lines demarcating the 99.99th percentiles. Positive values indicate regions under selection in an independent Tervuren (n=69) and Sheepdog (n=82) cohort with unknown GC phenotypes, whereas negative values represent regions under selection in Malinois (n=149). B) Left: UCSC image of the human (hg38) syntenic CFA12 region under selection in Tervuren and Sheepdogs. Two variants significantly associated with GC are denoted, as well as all GeneHancer double elite regulatory elements. Right: zooming in to the human syntenic location of the two identified variants shows overlap with a GeneHancer double elite element and ENCODE candidate cis-regulatory elements. The second variant falls within an untranslated region of *HDAC2-AS2*. Below, the ReMap ChIP-seq density track displays evidence of transcriptional regulators, and the ENCODE H3K27ac track shows enrichment of the H3K27ac histone modification, indicating areas of active enhancers.

SUPPLEMENTARY FIGURES & TABLES

Table S1. WGS imputation reference panel SRA numbers.

Table S2. Sample information for all imputed Belgian Tervuren, Sheepdog, and Malinois data.

Table S3. BayesR GWAS results for all variants with effect size ≥ 0.0001 .

Table S4. XP-nSL 99.99th percentile results of case vs. control selection scan.

Table S5. Fisher's exact tests for association of each case-control lead variant or haplotype with tumor subtype.

Table S6. Total number of risk alleles across the 15 loci included in GC risk model for each individual in the imputation reference panel.

Table S7. Results of model comparisons using ANOVA Chi-squared tests and leave one out cross validation information criterion (LOOIC).

Table S8. XP-nSL 99.99th percentile results of Belgian Tervuren and Sheepdog vs. Belgian Malinois selection scan.

Table S9. Top 99.99th percentile F_{st} values calculated in PLINK1.9 with imputed data for Belgian Tervuren and Belgian Sheepdogs vs. Belgian Malinois.

Table S10. Regions in the 99.9th percentile of per breed runs of homozygosity for Belgian Tervuren, Belgian Sheepdog, and Belgian Malinois breeds.

Table S11. Results of model comparisons using identified variants using ANOVA Chi-squared tests.

Table S12. Number of variants in each SNP array dataset after imputation, INFO score pruning, and merging.

Figure S1. Imputation concordance between dogs genotyped on multiple arrays. A) Average concordance per chromosome of all imputed variants prior to INFO score filtering. B) Average concordance after removing variants with INFO score <0.9 . Each datapoint is the average genotype concordance per alternate allele frequency bin per chromosome. Concordance for dogs genotyped on the 173k and 220k arrays ($n=15$) is shown in red, 173k and 670k ($n=23$) in green, and 220k and 670k ($n=7$) in blue.

Figure S2. Linear model demonstrating the relationship between number of risk alleles and age at gastric cancer diagnosis. The total number of risk alleles, among cases, in the 15-locus model ($2n=30$) is plotted on the x-axis, with age at diagnosis in years on the y-axis.

Figure S3. Importance of number of risk alleles at each locus in the 15-locus predictive model. The importance score (x-axis) ranks the relative contribution of one (1) or two (2) risk alleles at each locus (y-axis) to the model's predictive accuracy, with the highest score indicating the greatest effect within the model. The importance score is based on the absolute value of the t-statistic of each variable when applied to the model and was calculated using the varImp function in the caret R package. Importance is set to zero for zero copies of the risk allele.

Figure S4. Across-breed fixation index (F_{st}) values. A) F_{st} values of 7,836,277 non-monomorphic variants comparing the Belgian Tervuren ($n=69$) and Belgian Sheepdogs ($n=82$) of unknown phenotype to Belgian Malinois ($n=149$) as reported by PLINK1.9. The dashed black line represents the 99.99% threshold. B) Distribution of F_{st} values; approximately half of all values lie between 0 and 0.05.

Figure S5. Incidence plots of runs of homozygosity (ROH) by breed. ROH were assessed in the Tervuren (n=69), Sheepdog (n=82), and Malinois (n=149) for 2,039,330 variants, with a minimum 500kb interval. The proportion of times each SNV is represented in an ROH within the population is shown. The dashed line represents the 99.9th percentile of runs calculated per breed.

Figure S6. STRING protein network interactions. Proteins identified through GWAS and XP-nSL genomic analyses with reported interactions were included, as well as CDH1, whose expression is often lost in human diffuse GC. Protein interactions as defined by default STRING options are represented by gray lines. Blue shapes indicate downregulation and red shapes indicate upregulation of the protein or gene in human GC, while gray represents unknown. The figure was generated with Cytoscape.

REFERENCES

1. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., and Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* *74*, 229–263. <https://doi.org/10.3322/caac.21834>.
2. Allemani, C., Matsuda, T., Di Carlo, V., Harewood, R., Matz, M., Nikšić, M., Bonaventure, A., Valkov, M., Johnson, C.J., Estève, J., et al. (2018). Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *The Lancet* *391*, 1023–1075. [https://doi.org/10.1016/S0140-6736\(17\)33326-3](https://doi.org/10.1016/S0140-6736(17)33326-3).
3. Arnold, M., Rutherford, M.J., Bardot, A., Ferlay, J., Andersson, T.M.-L., Myklebust, T.Å., Tervonen, H., Thursfield, V., Ransom, D., Shack, L., et al. (2019). Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study. *Lancet Oncol.* *20*, 1493–1505. [https://doi.org/10.1016/S1470-2045\(19\)30456-5](https://doi.org/10.1016/S1470-2045(19)30456-5).
4. Li, Y., Feng, A., Zheng, S., Chen, C., and Lyu, J. (2022). recent estimates and predictions of 5-year survival in patients with gastric cancer: a model-based period analysis. *Cancer Control J. Moffitt Cancer Cent.* *29*, 1–9. <https://doi.org/10.1177/10732748221099227>.
5. Lordick, F., Carneiro, F., Cascinu, S., Fleitas, T., Haustermans, K., Piessen, G., Vogel, A., and Smyth, E.C. (2022). Gastric cancer: ESMO clinical practice guideline for diagnosis, treatment and follow-up. *Ann. Oncol.* *33*, 1005–1020. <https://doi.org/10.1016/j.annonc.2022.07.004>.
6. Arnold, M., Ferlay, J., Henegouwen, M.I. van B., and Soerjomataram, I. (2020). Global burden of oesophageal and gastric cancer by histology and subsite in 2018. *Gut* *69*, 1564–1571. <https://doi.org/10.1136/gutjnl-2020-321600>.
7. Lauren, P. (1965). The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. an attempt at a histo-clinical classification. *Acta Pathol. Microbiol. Scand.* *64*, 31–49. <https://doi.org/10.1111/apm.1965.64.1.31>.
8. Correa, P. (1988). A human model of gastric carcinogenesis. *Cancer Res.* *48*, 3554–3560.
9. Sugano, K., Moss, S.F., and Kuipers, E.J. (2023). gastric intestinal metaplasia: real culprit or innocent bystander as a precancerous condition for gastric cancer? *Gastroenterology* *165*, 1352-1366.e1. <https://doi.org/10.1053/j.gastro.2023.08.028>.
10. Hanada, K., and Graham, D.Y. (2014). *Helicobacter pylori* and the molecular pathogenesis of intestinal-type gastric carcinoma. *Expert Rev. Anticancer Ther.* *14*, 947–954. <https://doi.org/10.1586/14737140.2014.911092>.
11. Garcia-Pelaez, J., Barbosa-Matos, R., São José, C., Sousa, S., Gullo, I., Hoogerbrugge, N., Carneiro, F., and Oliveira, C. (2022). Gastric cancer genetic predisposition and clinical presentations: Established heritable causes and potential candidate genes. *Eur. J. Med. Genet.* *65*, 104401. <https://doi.org/10.1016/j.ejmg.2021.104401>.
12. Díaz Del Arco, C., Ortega Medina, L., Estrada Muñoz, L., García Gómez de Las Heras, S., and Fernández Aceñero, M.J. (2021). Is there still a place for conventional histopathology in the age of molecular medicine? Laurén classification, inflammatory infiltration and other current topics in gastric cancer diagnosis and prognosis. *Histol. Histopathol.* *36*, 587–613. <https://doi.org/10.14670/HH-18-309>.
13. Decourtye-Espiard, L., and Guilford, P. (2023). Hereditary diffuse gastric cancer. *Gastroenterology* *164*, 719–735. <https://doi.org/10.1053/j.gastro.2023.01.038>.
14. Guerra, J., Pinto, C., Pinto, P., Pinheiro, M., Santos, C., Peixoto, A., Escudeiro, C., Barbosa, A., Porto, M., Francisco, I., et al. (2023). Frequency of CDH1, CTNNA1 and CTNND1 germline variants in families with diffuse and mixed gastric cancer. *Cancers* *15*, 4313. <https://doi.org/10.3390/cancers15174313>.
15. Lott, P.C., and Carvajal-Carmona, L.G. (2018). Resolving of gastric cancer aetiology: an update in genetic predisposition. *Lancet Gastroenterol. Hepatol.* *3*, 874–883. [https://doi.org/10.1016/S2468-1253\(18\)30237-1](https://doi.org/10.1016/S2468-1253(18)30237-1).
16. Oliveira, C., Pinheiro, H., Figueiredo, J., Seruca, R., and Carneiro, F. (2015). Familial gastric cancer: genetic susceptibility, pathology, and implications for management. *Lancet Oncol.* *16*, e60–e70. [https://doi.org/10.1016/S1470-2045\(14\)71016-2](https://doi.org/10.1016/S1470-2045(14)71016-2).
17. Saeki, N., Ono, H., Sakamoto, H., and Yoshida, T. (2013). Genetic factors related to gastric cancer susceptibility identified using a genome-wide association study. *Cancer Sci.* *104*, 1–8. <https://doi.org/10.1111/cas.12042>.

18. Tanikawa, C., Kamatani, Y., Toyoshima, O., Sakamoto, H., Ito, H., Takahashi, A., Momozawa, Y., Hirata, M., Fuse, N., Takai-Igarashi, T., et al. (2018). Genome-wide association study identifies gastric cancer susceptibility loci at 12q24.11-12 and 20q11.21. *Cancer Sci.* *109*, 4015–4024. <https://doi.org/10.1111/cas.13815>.
19. Hess, T., Maj, C., Gehlen, J., Borisov, O., Haas, S.L., Gockel, I., Vieth, M., Piessen, G., Alakus, H., Vashist, Y., et al. (2023). Dissecting the genetic heterogeneity of gastric cancer. *eBioMedicine* *92*, 104616. <https://doi.org/10.1016/j.ebiom.2023.104616>.
20. Fonda, D., Gualtieri, M., and Scanziani, E. (1989). Gastric carcinoma in the dog: A clinicopathological study of 11 cases. *J. Small Anim. Pract.* *30*, 353–360. <https://doi.org/10.1111/j.1748-5827.1989.tb01579.x>.
21. Scanziani, E., Giusti, A.M., Gualtieri, M., and Fonda, D. (1991). Gastric carcinoma in the Belgian shepherd dog. *J. Small Anim. Pract.* *32*, 465–469. <https://doi.org/10.1111/j.1748-5827.1991.tb00991.x>.
22. Lubbes, D., Mandigers, P.J.J., Heuven, H.C.M., and Teske, E. (2009). Incidence of gastric carcinoma in Dutch Tervueren shepherd dogs born between 1991 and 2002. *Tijdschr. Diergeneesk.* *134*, 606–610.
23. Seim-Wikse, T., Jörundsson, E., Nødtvedt, A., Grotmol, T., Bjornvad, C.R., Kristensen, A.T., and Skancke, E. (2013). Breed predisposition to canine gastric carcinoma--a study based on the Norwegian canine cancer register. *Acta Vet. Scand.* *55*, 25. <https://doi.org/10.1186/1751-0147-55-25>.
24. Candido, M.V., Syrjä, P., Kilpinen, S., and Spillmann, T. (2018). Canine breeds associated with gastric carcinoma, metaplasia and dysplasia diagnosed by histopathology of endoscopic biopsy samples. *Acta Vet. Scand.* *60*, 37. <https://doi.org/10.1186/s13028-018-0392-6>.
25. Hugen, S., Thomas, R.E., German, A.J., Burgener, I.A., and Mandigers, P.J.J. (2017). Gastric carcinoma in canines and humans, a review. *Vet. Comp. Oncol.* *15*, 692–705. <https://doi.org/10.1111/vco.12249>.
26. Kijan, C., Hugen, S., Thomas, R.E., Oberbauer, A.M., Leegwater, P.A.J., Fieten, H., German, A.J., and Mandigers, P.J.J. (2023). The histopathological characteristic of gastric carcinoma in the Belgian Tervueren and Groenendael dog: a comparison of two classification methods. *Anim. Open Access J. MDPI* *13*, 1532. <https://doi.org/10.3390/ani13091532>.
27. Husnik, R., Klimes, J., Kovarikova, S., and Kolorz, M. (2022). Helicobacter Species and Their Association with Gastric Pathology in a Cohort of Dogs with Chronic Gastrointestinal Signs. *Animals* *12*, 1254. <https://doi.org/10.3390/ani12101254>.
28. Pocurull, A., Herrera-Pariente, C., Carballal, S., Llach, J., Sánchez, A., Carot, L., Botargues, J.M., Cuatrecasas, M., Ocaña, T., Balaguer, F., et al. (2021). Clinical, molecular and genetic characteristics of early onset gastric cancer: analysis of a large multicenter study. *Cancers* *13*, 3132. <https://doi.org/10.3390/cancers13133132>.
29. Smyth, E.C., Nilsson, M., Grabsch, H.I., van Grieken, N.C., and Lordick, F. (2020). Gastric cancer. *The Lancet* *396*, 635–648. [https://doi.org/10.1016/S0140-6736\(20\)31288-5](https://doi.org/10.1016/S0140-6736(20)31288-5).
30. von Babo, V., Eberle, N., Mischke, R., Meyer-Lindenberg, A., Hewicker-Trautwein, M., Nolte, I., and Betz, D. (2012). Canine non-hematopoietic gastric neoplasia. Epidemiologic and diagnostic characteristics in 38 dogs with post-surgical outcome of five cases. *Tierarztl. Prax. Ausg. K Klientiere Heimtiere* *40*, 243–249.
31. Abrams, B., Wavreille, V.A., Husbands, B.D., Matz, B.M., Massari, F., Liptak, J.M., Cray, M.T., de Mello Souza, C.H., Wustefeld-Janssens, B.G., Oblak, M.L., et al. (2019). Perioperative complications and outcome after surgery for treatment of gastric carcinoma in dogs: A Veterinary Society of Surgical Oncology retrospective study of 40 cases (2004-2018). *Vet. Surg. VS* *48*, 923–932. <https://doi.org/10.1111/vsu.13239>.
32. Marsden, C.D., Ortega-Del Vecchyo, D., O'Brien, D.P., Taylor, J.F., Ramirez, O., Vilà, C., Marques-Bonet, T., Schnabel, R.D., Wayne, R.K., and Lohmueller, K.E. (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci.* *113*, 152–157. <https://doi.org/10.1073/pnas.1512501113>.
33. Ostrander, E.A., Dreger, D.L., and Evans, J.M. (2019). Canine cancer genomics: lessons for canine and human health. *Annu. Rev. Anim. Biosci.* *7*, 449–472. <https://doi.org/10.1146/annurev-animal-030117-014523>.
34. Meadows, J.R.S., Kidd, J.M., Wang, G.-D., Parker, H.G., Schall, P.Z., Bianchi, M., Christmas, M.J., Bougiouri, K., Buckley, R.M., Hitte, C., et al. (2023). Genome sequencing of 2000 canids by the Dog10K consortium advances the understanding of demography, genome function and architecture. *Genome Biol.* *24*, 187. <https://doi.org/10.1186/s13059-023-03023-7>.

35. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* *20*, 467–484. <https://doi.org/10.1038/s41576-019-0127-1>.
36. Visscher, P.M., Yengo, L., Cox, N.J., and Wray, N.R. (2021). Discovery and implications of polygenicity of common diseases. *Science* *373*, 1468–1473. <https://doi.org/10.1126/science.abi8206>.
37. Tang, R., Noh, H.J., Wang, D., Sigurdsson, S., Swofford, R., Perloski, M., Duxbury, M., Patterson, E.E., Albright, J., Castelhana, M., et al. (2014). Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. *Genome Biol.* *15*, R25. <https://doi.org/10.1186/gb-2014-15-3-r25>.
38. Evans, J.M., Noorai, R.E., Tsai, K.L., Starr-Moss, A.N., Hill, C.M., Anderson, K.J., Famula, T.R., and Clark, L.A. (2017). Beyond the MHC: A canine model of dermatomyositis shows a complex pattern of genetic risk involving novel loci. *PLOS Genet.* *13*, e1006604. <https://doi.org/10.1371/journal.pgen.1006604>.
39. Hadji Rasouliha, S., Barrientos, L., Anderegg, L., Klesty, C., Lorenz, J., Chevallier, L., Jagannathan, V., Rösch, S., and Leeb, T. (2019). A RAPGEF6 variant constitutes a major risk factor for laryngeal paralysis in dogs. *PLoS Genet.* *15*, e1008416. <https://doi.org/10.1371/journal.pgen.1008416>.
40. Zangerl, B., Goldstein, O., Philp, A.R., Lindauer, S.J.P., Pearce-Kelling, S.E., Mullins, R.F., Graphodatsky, A.S., Ripoll, D., Felix, J.S., Stone, E.M., et al. (2006). Identical mutation in a novel retinal gene causes progressive rod–cone degeneration in dogs and retinitis pigmentosa in humans. *Genomics* *88*, 551–563. <https://doi.org/10.1016/j.ygeno.2006.07.007>.
41. Seppälä, E.H., Reuser, A.J.J., and Lohi, H. (2013). A Nonsense Mutation in the Acid α -Glucosidase Gene Causes Pompe Disease in Finnish and Swedish Lapphunds. *PLoS ONE* *8*, e56825. <https://doi.org/10.1371/journal.pone.0056825>.
42. Rinz, C.J., Levine, J., Minor, K.M., Humphries, H.D., Lara, R., Starr-Moss, A.N., Guo, L.T., Williams, D.C., Shelton, G.D., and Clark, L.A. (2014). A COLQ Missense Mutation in Labrador Retrievers Having Congenital Myasthenic Syndrome. *PLoS ONE* *9*, e106425. <https://doi.org/10.1371/journal.pone.0106425>.
43. Abrams, S.R., Hawks, A.L., Evans, J.M., Famula, T.R., Mahaffey, M., Johnson, G.S., Mason, J.M., and Clark, L.A. (2020). Variants in FtsJ RNA 2'-O-Methyltransferase 3 and Growth Hormone 1 are associated with small body size and a dental anomaly in dogs. *Proc. Natl. Acad. Sci. U. S. A.* *117*, 24929–24935. <https://doi.org/10.1073/pnas.2009500117>.
44. Cook, S., Hooser, B.N., Williams, D.C., Kortz, G., Aleman, M., Minor, K., Koziol, J., Friedenber, S.G., Cullen, J.N., Shelton, G.D., et al. (2023). Canine models of Charcot-Marie-Tooth: *MTMR2*, *MPZ*, and *SH3TC2* variants in golden retrievers with congenital hypomyelinating polyneuropathy. *Neuromuscul. Disord.* *33*, 677–691. <https://doi.org/10.1016/j.nmd.2023.06.007>.
45. Grall, A., Guaguère, E., Planchais, S., Grond, S., Bourrat, E., Hausser, I., Hitte, C., Le Gallo, M., Derbois, C., Kim, G.-J., et al. (2012). PNPLA1 mutations cause autosomal recessive congenital ichthyosis in golden retriever dogs and humans. *Nat. Genet.* *44*, 140–147. <https://doi.org/10.1038/ng.1056>.
46. Wolf, Z.T., Leslie, E.J., Arzi, B., Jayashankar, K., Karmi, N., Jia, Z., Rowland, D.J., Young, A., Safra, N., Sliskovic, S., et al. (2014). A LINE-1 Insertion in DLX6 Is Responsible for Cleft Palate and Mandibular Abnormalities in a Canine Model of Pierre Robin Sequence. *PLOS Genet.* *10*, e1004257. <https://doi.org/10.1371/journal.pgen.1004257>.
47. Wolf, Z.T., Brand, H.A., Shaffer, J.R., Leslie, E.J., Arzi, B., Willet, C.E., Cox, T.C., McHenry, T., Narayan, N., Feingold, E., et al. (2015). Genome-Wide Association Studies in Dogs and Humans Identify ADAMTS20 as a Risk Variant for Cleft Lip and Palate. *PLOS Genet.* *11*, e1005059. <https://doi.org/10.1371/journal.pgen.1005059>.
48. Haid, M. The Most Popular Dog Breeds of 2023. Am. Kennel Club. <https://www.akc.org/expert-advice/news/most-popular-dog-breeds-2023/>.
49. Dreger, D.L., Hooser, B.N., Hughes, A.M., Ganesan, B., Donner, J., Anderson, H., Holtvoigt, L., and Ekenstedt, K.J. (2019). True Colors: Commercially-acquired morphological genotypes reveal hidden allele variation among dog breeds, informing both trait ancestry and breed potential. *PLOS ONE* *14*, e0223995. <https://doi.org/10.1371/journal.pone.0223995>.
50. Moser, G., Lee, S.H., Hayes, B.J., Goddard, M.E., Wray, N.R., and Visscher, P.M. (2015). Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model. *PLOS Genet.* *11*, e1004969. <https://doi.org/10.1371/journal.pgen.1004969>.
51. Tengvall, K., Sundström, E., Wang, C., Bergvall, K., Wallerman, O., Pederson, E., Karlsson, Å., Harvey,

- N.D., Blott, S.C., Olby, N., et al. (2022). Bayesian model and selection signature analyses reveal risk factors for canine atopic dermatitis. *Commun. Biol.* 5, 1348. <https://doi.org/10.1038/s42003-022-04279-8>.
52. Brander, G., Rohdin, C., Bianchi, M., Bergvall, K., Andersson, G., Ljungvall, I., Hultin Jäderlund, K., Häggström, J., Hedhammar, Å., Lindblad-Toh, K., et al. (2023). Multiple Genetic Loci Associated with Pug Dog Thoracolumbar Myelopathy. *Genes* 14, 385. <https://doi.org/10.3390/genes14020385>.
53. Lingaas, F., Tengvall, K., Jansen, J.H., Pelander, L., Hurst, M.H., Meuwissen, T., Karlsson, Å., Meadows, J.R.S., Sundström, E., Thoresen, S.I., et al. (2023). Bayesian mixed model analysis uncovered 21 risk loci for chronic kidney disease in boxer dogs. *PLOS Genet.* 19, e1010599. <https://doi.org/10.1371/journal.pgen.1010599>.
54. Szpiech, Z.A., Novak, T.E., Bailey, N.P., and Stevison, L.S. (2021). Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. *Evol. Lett.* 5, 408–421. <https://doi.org/10.1002/evl3.232>.
55. Parker, H.G., Dreger, D.L., Rimbault, M., Davis, B.W., Mullen, A.B., Carpintero-Ramirez, G., and Ostrander, E.A. (2017). Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Rep.* 19, 697–708. <https://doi.org/10.1016/j.celrep.2017.03.079>.
56. Hammal, F., de Langen, P., Bergon, A., Lopez, F., and Ballester, B. (2022). ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res.* 50, D316–D325. <https://doi.org/10.1093/nar/gkab996>.
57. Wei, R., Ding, C., Rodríguez, R.A., and Del Mar Requena Mullor, M. (2018). The SOX2OT/miR-194-5p axis regulates cell proliferation and mobility of gastric cancer through suppressing epithelial-mesenchymal transition. *Oncol. Lett.* 16, 6361–6368. <https://doi.org/10.3892/ol.2018.9433>.
58. Ji, R., Wu, C., Yao, J., Xu, J., Lin, J., Gu, H., Fu, M., Zhang, X., Li, Y., and Zhang, X. (2023). IGF2BP2-mediated m6A modification of CSF2 reprograms MSC to promote gastric cancer progression. *Cell Death Dis.* 14, 1–14. <https://doi.org/10.1038/s41419-023-06163-7>.
59. Jo, H., Shim, K., Kim, H.-U., Jung, H.S., and Jeoung, D. (2023). HDAC2 as a target for developing anti-cancer drugs. *Comput. Struct. Biotechnol. J.* 21, 2048–2057. <https://doi.org/10.1016/j.csbj.2023.03.016>.
60. Li, Y., Wang, Z., Gao, P., Cao, D., Dong, R., Zhu, M., Fei, Y., Zuo, X., and Cai, J. (2024). CircRHBD1 promotes immune escape via IGF2BP2/PD-L1 signaling and acts as a nanotherapeutic target in gastric cancer. *J. Transl. Med.* 22, 704. <https://doi.org/10.1186/s12967-024-05498-9>.
61. Sewduth, R.N., Jaspard-Vinassa, B., Peghaire, C., Guillabert, A., Franzl, N., Larrieu-Lahargue, F., Moreau, C., Fruttiger, M., Dufourcq, P., Couffignal, T., et al. (2014). The ubiquitin ligase PDZRN3 is required for vascular morphogenesis through Wnt/planar cell polarity signalling. *Nat. Commun.* 5, 4832. <https://doi.org/10.1038/ncomms5832>.
62. Konopelski Snavelly, S.E., Susman, M.W., Kunz, R.C., Tan, J., Srinivasan, S., Cohen, M.D., Okada, K., Lamb, H., Choi, S.S., Karuna, E.P., et al. (2021). Proteomic analysis identifies the E3 ubiquitin ligase Pdzrn3 as a regulatory target of Wnt5a-Ror signaling. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2104944118. <https://doi.org/10.1073/pnas.2104944118>.
63. Sremac, M., Paic, F., Grubelic Ravic, K., Serman, L., Pavicic Dujmovic, A., Brcic, I., Krznaric, Z., and Nikuseva Martic, T. (2021). Aberrant expression of SFRP1, SFRP3, DVL2 and DVL3 Wnt signaling pathway components in diffuse gastric carcinoma. *Oncol. Lett.* 22, 822. <https://doi.org/10.3892/ol.2021.13083>.
64. Jiao, B., Zhang, H., Jiang, H., Liu, S., Wang, Y., Chen, Y., Duan, H., Niu, Y., Shen, M., Wang, H., et al. (2024). Granulysin-mediated reduction of PDZRN3 induces Cx43 gap junctions activity exacerbating skin damage in trichloroethylene hypersensitivity syndrome. *Ecotoxicol. Environ. Saf.* 274, 116174. <https://doi.org/10.1016/j.ecoenv.2024.116174>.
65. Bonacquisti, E.E., and Nguyen, J. (2019). Connexin 43 (Cx43) in cancer: Implications for therapeutic approaches via gap junctions. *Cancer Lett.* 442, 439–444. <https://doi.org/10.1016/j.canlet.2018.10.043>.
66. Zhou, M., Zheng, M., Zhou, X., Tian, S., Yang, X., Ning, Y., Li, Y., and Zhang, S. (2023). The roles of connexins and gap junctions in the progression of cancer. *Cell Commun. Signal.* CCS 21, 8. <https://doi.org/10.1186/s12964-022-01009-9>.
67. Tang, B., Peng, Z., Yu, P.-W., Yu, G., Qian, F., Zeng, D., Zhao, Y., Shi, Y., Hao, Y., and Luo, H. (2013). Aberrant Expression of Cx43 Is Associated with the Peritoneal Metastasis of Gastric Cancer and Cx43-Mediated Gap Junction Enhances Gastric Cancer Cell Diapedesis from Peritoneal Mesothelium. *PLOS*

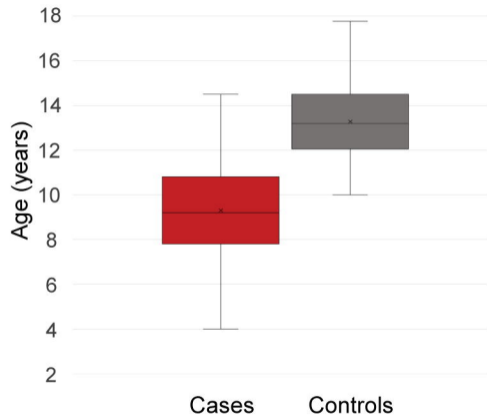
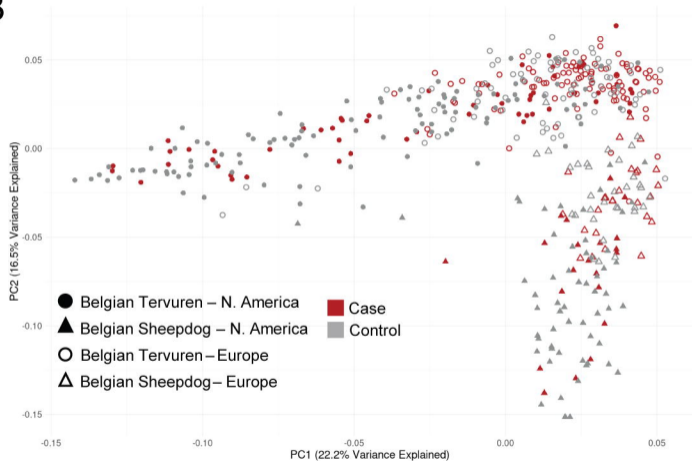
- ONE 8, e74527. <https://doi.org/10.1371/journal.pone.0074527>.
68. Kim, E.Y., Jun, K.H., and Yim, K. (2020). The Roles of Connexin 26, 32, and 43 as Prognostic Factors for Gastric Cancer. *Anticancer Res.* 40, 4537–4545. <https://doi.org/10.21873/anticancer.14459>.
 69. Li, C.-H., Hao, M.-L., Sun, Y., Wang, Z.-J., and Li, J.-L. (2020). Ultrastructure of gap junction and Cx43 expression in gastric cancer tissues of the patients. *Arch. Med. Sci. AMS* 16, 352–358. <https://doi.org/10.5114/aoms.2020.92859>.
 70. Luo, J., Chen, J.-W., Zhou, J., Han, K., Li, S., Duan, J.-L., Cao, C.-H., Lin, J.-L., Xie, D., and Wang, F.-W. (2022). TBX20 inhibits colorectal cancer tumorigenesis by impairing NHEJ-mediated DNA repair. *Cancer Sci.* 113, 2008–2021. <https://doi.org/10.1111/cas.15348>.
 71. Fu, A., Xiang, S., He, Q., and Ying, M. (2023). Kelch-like proteins in the gastrointestinal tumors. *Acta Pharmacol. Sin.* 44, 931–939. <https://doi.org/10.1038/s41401-022-01007-0>.
 72. Yoo, N.J., Kim, H.R., Kim, Y.R., An, C.H., and Lee, S.H. (2012). Somatic mutations of the KEAP1 gene in common solid cancers. *Histopathology* 60, 943–952. <https://doi.org/10.1111/j.1365-2559.2012.04178.x>.
 73. Lee, B., Yoon, K., Lee, S., Kang, J.M., Kim, J., Shim, S.H., Kim, H.-M., Song, S., Naka, K., Kim, A.K., et al. (2015). Homozygous deletions at 3p22, 5p14, 6q15, and 9p21 result in aberrant expression of tumor suppressor genes in gastric cancer. *Genes. Chromosomes Cancer* 54, 142–155. <https://doi.org/10.1002/gcc.22226>.
 74. Wu, Q., Yin, G., Lei, J., Tian, J., Lan, A., and Liu, S. (2020). KLHL5 Is a Prognostic-Related Biomarker and Correlated With Immune Infiltrates in Gastric Cancer. *Front. Mol. Biosci.* 7. <https://doi.org/10.3389/fmolb.2020.599110>.
 75. Sun, M., Song, L., Zhou, T., Gillespie, G.Y., and Jope, R.S. (2011). The Role of DDX3 in Regulating Snail. *Biochim. Biophys. Acta* 1813, 438–447. <https://doi.org/10.1016/j.bbamcr.2011.01.003>.
 76. Chen, H.-H., Yu, H.-I., Cho, W.-C., and Tarn, W.-Y. (2015). DDX3 modulates cell adhesion and motility and cancer cell metastasis via Rac1-mediated signaling pathway. *Oncogene* 34, 2790–2800. <https://doi.org/10.1038/onc.2014.190>.
 77. Mo, J., Liang, H., Su, C., Li, P., Chen, J., and Zhang, B. (2021). DDX3X: structure, physiologic functions and cancer. *Mol. Cancer* 20, 38. <https://doi.org/10.1186/s12943-021-01325-7>.
 78. Yao, L., Hao, Q., Wang, M., Chen, Y., Cao, H., Zhang, Q., Yu, K., Jiang, Y., Shao, Z., Zhou, X., et al. (2023). KLHL29-mediated DDX3X degradation promotes chemosensitivity by abrogating cell cycle checkpoint in triple-negative breast cancer. *Oncogene* 42, 3514–3528. <https://doi.org/10.1038/s41388-023-02858-5>.
 79. Zou, G., Huang, Y., Zhang, S., Ko, K.-P., Kim, B., Zhang, J., Venkatesan, V., Pizzi, M.P., Fan, Y., Jun, S., et al. (2024). E-cadherin loss drives diffuse-type gastric tumorigenesis via EZH2-mediated reprogramming. *J. Exp. Med.* 221, e20230561. <https://doi.org/10.1084/jem.20230561>.
 80. Kang, Y.-H., Lee, H.S., and Kim, W.H. (2002). Promoter Methylation and Silencing of PTEN in Gastric Carcinoma. *Lab. Invest.* 82, 285–291. <https://doi.org/10.1038/labinvest.3780422>.
 81. Kim, B., Kang, S.Y., Kim, D., Heo, Y.J., and Kim, K.-M. (2020). PTEN Protein Loss and Loss-of-Function Mutations in Gastric Cancers: The Relationship with Microsatellite Instability, EBV, HER2, and PD-L1 Expression. *Cancers* 12, 1724. <https://doi.org/10.3390/cancers12071724>.
 82. Fei, G., Ebert, M.P.A., Mawrin, C., Leodolter, A., Schmidt, N., Dietzmann, K., and Malfertheiner, P. (2002). Reduced PTEN expression in gastric cancer and in the gastric mucosa of gastric cancer relatives. *Eur. J. Gastroenterol. Hepatol.* 14, 297–303. <https://doi.org/10.1097/00042737-200203000-00015>.
 83. Bangoura, G., Liu, Z.-S., Qian, Q., Jiang, C.-Q., Yang, G.-F., and Jing, S. (2007). Prognostic significance of HIF-2 α /EPAS1 expression in hepatocellular carcinoma. *World J. Gastroenterol.* 13, 3176–3182. <https://doi.org/10.3748/wjg.v13.i23.3176>.
 84. Wang, Y., Li, Z., Zhang, H., Jin, H., Sun, L., Dong, H., Xu, M., Zhao, P., Zhang, B., Wang, J., et al. (2010). HIF-1 α and HIF-2 α correlate with migration and invasion in gastric cancer. *Cancer Biol. Ther.* 10, 376–382. <https://doi.org/10.4161/cbt.10.4.12441>.
 85. Wu, X.-H., Qian, C., and Yuan, K. (2011). Correlations of hypoxia-inducible factor-1 α /hypoxia-inducible factor-2 α expression with angiogenesis factors expression and prognosis in non-small cell lung cancer. *Chin. Med. J. (Engl.)* 124, 11–18.
 86. Tong, W.-W., Tong, G.-H., Chen, X.-X., Zheng, H.-C., and Wang, Y.-Z. (2015). HIF2 α is associated with poor prognosis and affects the expression levels of survivin and cyclin D1 in gastric carcinoma. *Int. J. Oncol.* 46, 233–242. <https://doi.org/10.3892/ijo.2014.2719>.
 87. Tong, W.-W., Tong, G.-H., Kong, H., and Liu, Y. (2016). The tumor promoting roles of HSP60 and HIF2 α

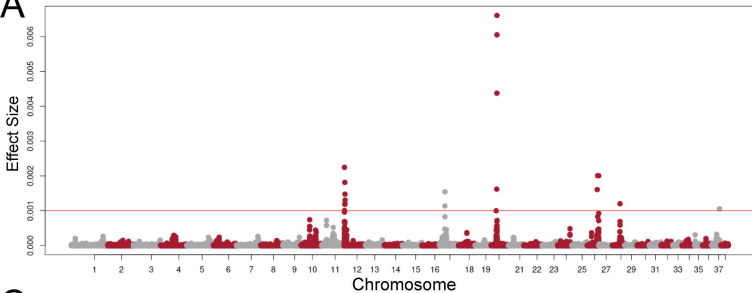
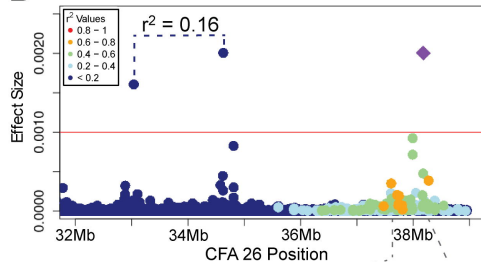
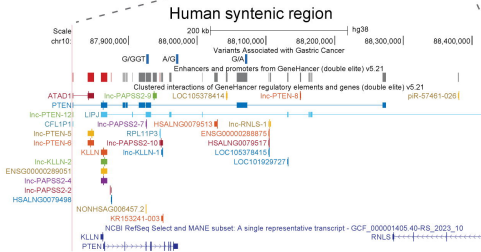
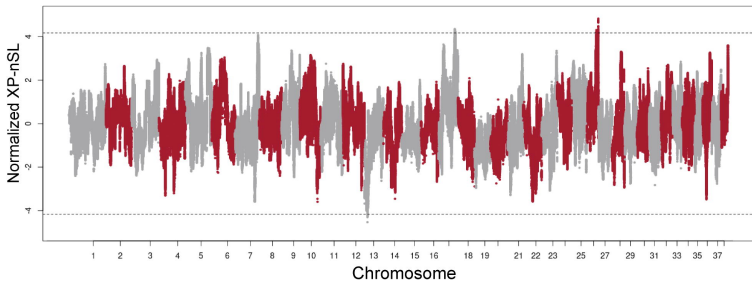
- in gastric cancer cells. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* 37, 9849–9854. <https://doi.org/10.1007/s13277-015-4783-2>.
88. Zhang, Q., Lou, Y., Zhang, J., Fu, Q., Wei, T., Sun, X., Chen, Q., Yang, J., Bai, X., and Liang, T. (2017). Hypoxia-inducible factor-2 α promotes tumor progression and has crosstalk with Wnt/ β -catenin signaling in pancreatic cancer. *Mol. Cancer* 16, 119. <https://doi.org/10.1186/s12943-017-0689-5>.
89. Islam, F., Gopalan, V., Law, S., Lam, A.K., and Pillai, S. (2020). Molecular Deregulation of EPAS1 in the Pathogenesis of Esophageal Squamous Cell Carcinoma. *Front. Oncol.* 10. <https://doi.org/10.3389/fonc.2020.01534>.
90. Islam, F., Gopalan, V., Lu, C.T., Pillai, S., and Lam, A.K. (2021). Identification of novel mutations and functional impacts of EPAS1 in colorectal cancer. *Cancer Med.* 10, 5557–5573. <https://doi.org/10.1002/cam4.4116>.
91. Wu, M., Hsieh, R., Huang, S., Chang, Y., Lin, M., Chang, M., Shun, C., Sheu, J., and Lin, J. (2002). Association of HLA-DQB1*0301 and HLA-DQB1*0602 with Different Subtypes of Gastric Cancer in Taiwan. *Jpn. J. Cancer Res. Gann* 93, 404–410. <https://doi.org/10.1111/j.1349-7006.2002.tb01271.x>.
92. Mahmud, M.T., Ahmed, F., Rana, M.J., Rahman, Md.A., Atta, A., and Saif-Ur-Rahman, K. (2024). Association of HLA gene polymorphisms with Helicobacter pylori related gastric cancer—a systematic review. *HLA* 103, e15394. <https://doi.org/10.1111/tan.15394>.
93. Lahner, E., Spoletni, M., Buzzetti, R., Corleto, V.D., Vannella, L., Petrone, A., and Annibale, B. (2010). HLA-DRB1*03 and DRB1*04 are associated with atrophic gastritis in an Italian population. *Dig. Liver Dis. Off. J. Ital. Soc. Gastroenterol. Ital. Assoc. Study Liver* 42, 854–859. <https://doi.org/10.1016/j.dld.2010.04.011>.
94. Oksanen, A.M., Haimila, K.E., Rautelin, H.I.K., and Partanen, J.A. (2010). Immunogenetic characteristics of patients with autoimmune gastritis. *World J. Gastroenterol.* 16, 354–358. <https://doi.org/10.3748/wjg.v16.i3.354>.
95. Lenti, M.V., Rugge, M., Lahner, E., Miceli, E., Toh, B.-H., Genta, R.M., De Block, C., Hershko, C., and Di Sabatino, A. (2020). Autoimmune gastritis. *Nat. Rev. Dis. Primer* 6, 1–19. <https://doi.org/10.1038/s41572-020-0187-8>.
96. Kubokawa, M., Nakashima, M., Yao, T., Ito, K.I., Harada, N., Nawata, H., and Watanabe, T. (2001). Aberrant intracellular localization of RCAS1 is associated with tumor progression of gastric cancer. *Int. J. Oncol.* 19, 695–700. <https://doi.org/10.3892/ijo.19.4.695>.
97. Fukuda, K., Tsujitani, S., Maeta, Y., Yamaguchi, K., Ikeguchi, M., and Kaibara, N. (2002). The expression of RCAS1 and tumor infiltrating lymphocytes in patients with T3 gastric carcinoma. *Gastric Cancer Off. J. Int. Gastric Cancer Assoc. Jpn. Gastric Cancer Assoc.* 5, 220–227. <https://doi.org/10.1007/s101200200038>.
98. Nakamura, Y., Yamazaki, K., Oizumi, S., Nakashima, M., Watanabe, T., Dosaka-Akita, H., and Nishimura, M. (2004). Expression of RCAS1 in human gastric carcinoma: a potential mechanism of immune escape. *Cancer Sci.* 95, 260–265. <https://doi.org/10.1111/j.1349-7006.2004.tb02213.x>.
99. Giaginis, C., Efkarpidis, T., Alexandrou, P., Patsouris, E., Kouraklis, G., and Theocharis, S. (2013). Increased RCAS1 expression is associated with advanced histopathological stage and poor prognosis in patients with gastric adenocarcinoma. *Dis. Markers* 35, 213–219. <https://doi.org/10.1155/2013/527548>.
100. Bouloc, A., Bagot, M., Delaire, S., Bensussan, A., and Bousmell, L. (2000). Triggering CD101 molecule on human cutaneous dendritic cells inhibits T cell proliferation via IL-10 production. *Eur. J. Immunol.* 30, 3132–3139. [https://doi.org/10.1002/1521-4141\(200011\)30:11<3132::AID-IMMU3132>3.0.CO;2-E](https://doi.org/10.1002/1521-4141(200011)30:11<3132::AID-IMMU3132>3.0.CO;2-E).
101. Schey, R., Dornhoff, H., Baier, J.L.C., Purtak, M., Opoka, R., Koller, A.K., Atreya, R., Rau, T.T., Daniel, C., Amann, K., et al. (2016). CD101 inhibits the expansion of colitogenic T cells. *Mucosal Immunol.* 9, 1205–1217. <https://doi.org/10.1038/mi.2015.139>.
102. Richert-Spuhler, L.E., Mar, C.M., Shinde, P., Wu, F., Hong, T., Greene, E., Hou, S., Thomas, K., Gottardo, R., Mugo, N., et al. (2021). CD101 genetic variants modify regulatory and conventional T cell phenotypes and functions. *Cell Rep. Med.* 2, 100322. <https://doi.org/10.1016/j.xcrm.2021.100322>.
103. Dutrow, E.V., Serpell, J.A., and Ostrander, E.A. (2022). Domestic dog lineages reveal genetic drivers of behavioral diversification. *Cell* 185, 4737–4755.e18. <https://doi.org/10.1016/j.cell.2022.11.003>.
104. Karyadi, D.M., Karlins, E., Decker, B., vonHoldt, B.M., Carpintero-Ramirez, G., Parker, H.G., Wayne, R.K., and Ostrander, E.A. (2013). A Copy Number Variant at the KITLG Locus Likely Confers Risk for Canine Squamous Cell Carcinoma of the Digit. *PLOS Genet.* 9, e1003409. <https://doi.org/10.1371/journal.pgen.1003409>.

105. Weich, K., Affolter, V., York, D., Rebhun, R., Grahn, R., Kallenberg, A., and Bannasch, D. (2020). Pigment Intensity in Dogs is Associated with a Copy Number Variant Upstream of KITLG. *Genes* 11, 75. <https://doi.org/10.3390/genes11010075>.
106. Hong, J.H., Jin, E.-H., Sung, J.K., Chang, I.A., Kang, H., and Lee, S.-I. (2023). Association of lncRNA SOX2OT rs9839776 polymorphism with gastric cancer risk in Korean: Case-control study. *Medicine (Baltimore)* 102, e35103. <https://doi.org/10.1097/MD.00000000000035103>.
107. Wang, Y., Wu, N., Luo, X., Zhang, X., Liao, Q., and Wang, J. (2020). SOX2OT, a novel tumor-related long non-coding RNA. *Biomed. Pharmacother.* 123, 109725. <https://doi.org/10.1016/j.biopha.2019.109725>.
108. Li, P.-Y., Wang, P., Gao, S.-G., and Dong, D.-Y. (2020). Long Noncoding RNA SOX2-OT: Regulations, Functions, and Roles on Mental Illnesses, Cancers, and Diabetic Complications. *BioMed Res. Int.* 2020, 2901589. <https://doi.org/10.1155/2020/2901589>.
109. Wang, S., Tie, J., Wang, R., Hu, F., Gao, L., Wang, W., Wang, L., Li, Z., Hu, S., Tang, S., et al. (2015). SOX2, a predictor of survival in gastric cancer, inhibits cell proliferation and metastasis by regulating PTEN. *Cancer Lett.* 358, 210–219. <https://doi.org/10.1016/j.canlet.2014.12.045>.
110. Sarkar, A., Huebner, A.J., Sulahian, R., Anselmo, A., Xu, X., Flattery, K., Desai, N., Sebastian, C., Yram, M.A., Arnold, K., et al. (2016). Sox2 Suppresses Gastric Tumorigenesis in Mice. *Cell Rep.* 16, 1929–1941. <https://doi.org/10.1016/j.celrep.2016.07.034>.
111. Otsubo, T., Akiyama, Y., Yanagihara, K., and Yuasa, Y. (2008). SOX2 is frequently downregulated in gastric cancers and inhibits cell growth through cell-cycle arrest and apoptosis. *Br. J. Cancer* 98, 824–831. <https://doi.org/10.1038/sj.bjc.6604193>.
112. Carrasco-Garcia, E., Santos, J.C., Garcia, I., Brianti, M., García-Puga, M., Pedrazzoli, J.J., Matheu, A., and Ribeiro, M.L. (2016). Paradoxical role of SOX2 in gastric cancer. *Am. J. Cancer Res.* 6, 701–713.
113. Zhang, S., and Cui, W. (2014). Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J. Stem Cells* 6, 305–311. <https://doi.org/10.4252/wjsc.v6.i3.305>.
114. Amador-Arjona, A., Cimadamore, F., Huang, C.-T., Wright, R., Lewis, S., Gage, F.H., and Terskikh, A.V. (2015). SOX2 primes the epigenetic landscape in neural precursors enabling proper gene activation during hippocampal neurogenesis. *Proc. Natl. Acad. Sci.* 112, E1936–E1945. <https://doi.org/10.1073/pnas.1421480112>.
115. Cao, J., Mu, Q., and Huang, H. (2018). The Roles of Insulin-Like Growth Factor 2 mRNA-Binding Protein 2 in Cancer and Cancer Stem Cells. *Stem Cells Int.* 2018, 4217259. <https://doi.org/10.1155/2018/4217259>.
116. Liu, D., Xia, A.-D., Wu, L.-P., Li, S., Zhang, K., and Chen, D. (2022). IGF2BP2 promotes gastric cancer progression by regulating the IGF1R-RhoA-ROCK signaling pathway. *Cell. Signal.* 94, 110313. <https://doi.org/10.1016/j.cellsig.2022.110313>.
117. Ouyang, J., Li, J., Li, D., Jiang, J., Hao, T., Xia, Y., Lu, X., Zhang, C., and He, Y. (2022). IGF2BP2 Promotes Epithelial to Mesenchymal Transition and Metastasis through Stabilizing HMGA1 mRNA in Gastric Cancer. *Cancers* 14, 5381. <https://doi.org/10.3390/cancers14215381>.
118. Liu, X., Chen, Z., Zhao, X., Huang, M., Wang, C., Peng, W., Yin, J., Li, J., He, G., Li, X., et al. (2015). Effects of IGF2BP2, KCNQ1 and GCKR polymorphisms on clinical outcome in metastatic gastric cancer treated with EOF regimen. *Pharmacogenomics* 16, 959–970. <https://doi.org/10.2217/pgs.15.49>.
119. Zhang, Z., Xing, Y., Gao, W., Yang, L., Shi, J., Song, W., and Li, T. (2022). N6-methyladenosine (m6A) reader IGF2BP2 promotes gastric cancer progression via targeting SIRT1. *Bioengineered* 13, 11541–11550. <https://doi.org/10.1080/21655979.2022.2068920>.
120. Zang, Y., Tian, Z., Wang, D., Li, Y., Zhang, W., Ma, C., Liao, Z., Gao, W., Qian, L., Xu, X., et al. (2024). METTL3-mediated N6-methyladenosine modification of STAT5A promotes gastric cancer progression by regulating KLF4. *Oncogene* 43, 2338–2354. <https://doi.org/10.1038/s41388-024-03085-2>.
121. Dahlem, C., Abuhaliema, A., Kessler, S.M., Kröhler, T., Zoller, B.G.E., Chanda, S., Wu, Y., Both, S., Müller, F., Lepikhov, K., et al. (2022). First Small-Molecule Inhibitors Targeting the RNA-Binding Protein IGF2BP2/IMP2 for Cancer Therapy. *ACS Chem. Biol.* 17, 361–375. <https://doi.org/10.1021/acscchembio.1c00833>.
122. Feng, P., Chen, D., Wang, X., Li, Y., Li, Z., Li, B., Zhang, Y., Li, W., Zhang, J., Ye, J., et al. (2022). Inhibition of the m6A reader IGF2BP2 as a strategy against T-cell acute lymphoblastic leukemia. *Leukemia* 36, 2180–2188. <https://doi.org/10.1038/s41375-022-01651-9>.
123. Ai, C., Zhang, J., Lian, S., Ma, J., Györfy, B., Qian, Z., Han, Y., and Feng, Q. (2020). FOXM1 functions collaboratively with PLAU to promote gastric cancer progression. *J. Cancer* 11, 788–794.

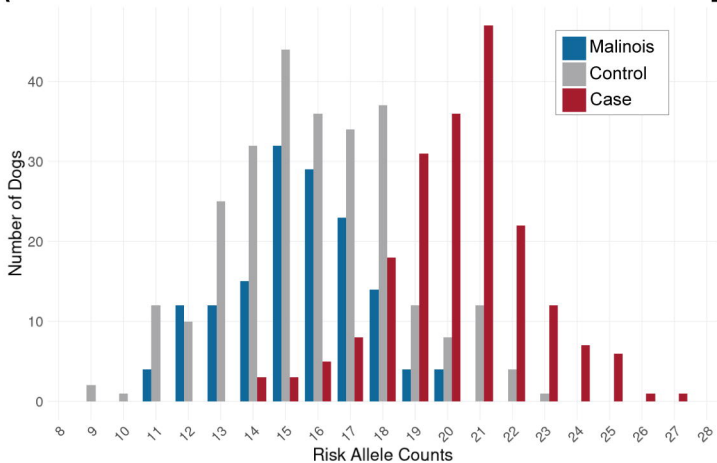
- <https://doi.org/10.7150/jca.37323>.
124. Sun, J., Zhou, F., Xue, J., Ji, C., Qu, Y., and Pan, Y. (2021). Long non-coding RNA TRPM2-AS regulates microRNA miR-138-5p and PLAU (Plasminogen Activator, Urokinase) to promote the progression of gastric adenocarcinoma. *Bioengineered* 12, 9753–9765. <https://doi.org/10.1080/21655979.2021.1995101>.
 125. Cao, W., Zeng, Z., Lan, J., Yang, Y., Lu, M., and Lei, S. (2023). Knockdown of FUT11 inhibits the progression of gastric cancer via the PI3K/AKT pathway. *Heliyon* 9, e17600. <https://doi.org/10.1016/j.heliyon.2023.e17600>.
 126. Huang, Y., Yang, X., Wei, M., Yang, X., Yuan, Z., Huang, J., Wei, J., and Tian, L. (2024). FUT11 expression in gastric cancer: its prognostic significance and role in immune regulation. *Discov. Oncol.* 15, 250. <https://doi.org/10.1007/s12672-024-01120-y>.
 127. Duncan, M.B., Liu, M., Fox, C., and Liu, J. (2006). Characterization of the *N*-deacetylase domain from the heparan sulfate *N*-deacetylase/*N*-sulfotransferase 2. *Biochem. Biophys. Res. Commun.* 339, 1232–1237. <https://doi.org/10.1016/j.bbrc.2005.11.142>.
 128. Lo, S.-L., Thike, A.A., Tan, S.-Y., Lim, T.K.-H., Tan, I.B.-H., Choo, S.-P., Tan, P.-H., Bay, B.-H., and Yip, G.W.-C. (2011). Expression of heparan sulfate in gastric carcinoma and its correlation with clinicopathological features and patient survival. *J. Clin. Pathol.* 64, 153–158. <https://doi.org/10.1136/jcp.2010.087478>.
 129. Yang, H., and Wang, L. (2023). Chapter Nine - Heparan sulfate proteoglycans in cancer: Pathogenesis and therapeutic potential. In *Advances in Cancer Research Novel Methods and Pathways in Cancer Glycobiology Research.*, K. L. Abbott and C. J. Dimitroff, eds. (Academic Press), pp. 251–291. <https://doi.org/10.1016/bs.acr.2022.08.001>.
 130. Denys, A., and Allain, F. (2019). The Emerging Roles of Heparan Sulfate 3-O-Sulfotransferases in Cancer. *Front. Oncol.* 9, 507. <https://doi.org/10.3389/fonc.2019.00507>.
 131. Song, J., Noh, J.H., Lee, J.H., Eun, J.W., Ahn, Y.M., Kim, S.Y., Lee, S.H., Park, W.S., Yoo, N.J., Lee, J.Y., et al. (2005). Increased expression of histone deacetylase 2 is found in human gastric cancer. *APMIS Acta Pathol. Microbiol. Immunol. Scand.* 113, 264–268. https://doi.org/10.1111/j.1600-0463.2005.apm_04.x.
 132. Kim, J.K., Noh, J.H., Eun, J.W., Jung, K.H., Bae, H.J., Shen, Q., Kim, M.G., Chang, Y.G., Kim, S.-J., Park, W.S., et al. (2013). Targeted Inactivation of HDAC2 Restores p16INK4a Activity and Exerts Antitumor Effects on Human Gastric Cancer. *Mol. Cancer Res.* 11, 62–73. <https://doi.org/10.1158/1541-7786.MCR-12-0332>.
 133. Wisnieski, F., Calcagno, D.Q., Leal, M.F., Chen, E.S., Gigeck, C.O., Santos, L.C., Pontes, T.B., Rasmussen, L.T., Payão, S.L.M., Assumpção, P.P., et al. (2014). Differential expression of histone deacetylase and acetyltransferase genes in gastric cancer and their modulation by trichostatin A. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* 35, 6373–6381. <https://doi.org/10.1007/s13277-014-1841-0>.
 134. Orenay-Boyacioglu, S., Kasap, E., Gerceker, E., Yuceyar, H., Demirci, U., Bilgic, F., and Korkmaz, M. (2018). Expression profiles of histone modification genes in gastric cancer progression. *Mol. Biol. Rep.* 45, 2275–2282. <https://doi.org/10.1007/s11033-018-4389-z>.
 135. Sun, J., Piao, J., Li, N., Yang, Y., Kim, K.-Y., and Lin, Z. (2020). Valproic acid targets HDAC1/2 and HDAC1/PTEN/Akt signalling to inhibit cell proliferation via the induction of autophagy in gastric cancer. *FEBS J.* 287, 2118–2133. <https://doi.org/10.1111/febs.15122>.
 136. Aghdassi, A., Sandler, M., Guenther, A., Mayerle, J., Behn, C.-O., Heidecke, C.-D., Friess, H., Büchler, M., Evert, M., Lerch, M.M., et al. (2012). Recruitment of histone deacetylases HDAC1 and HDAC2 by the transcriptional repressor ZEB1 downregulates E-cadherin expression in pancreatic cancer. *Gut* 61, 439–448. <https://doi.org/10.1136/gutjnl-2011-300060>.
 137. Plassais, J., Kim, J., Davis, B.W., Karyadi, D.M., Hogan, A.N., Harris, A.C., Decker, B., Parker, H.G., and Ostrander, E.A. (2019). Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* 10, 1489. <https://doi.org/10.1038/s41467-019-09373-w>.
 138. Hofmeister, R.J., Ribeiro, D.M., Rubinacci, S., and Delaneau, O. (2023). Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. *Nat. Genet.* 55, 1243–1249. <https://doi.org/10.1038/s41588-023-01415-w>.
 139. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10,

- giab008. <https://doi.org/10.1093/gigascience/giab008>.
140. Rubinacci, S., Delaneau, O., and Marchini, J. (2020). Genotype imputation using the Positional Burrows Wheeler Transform. *PLOS Genet.* *16*, e1009049. <https://doi.org/10.1371/journal.pgen.1009049>.
 141. Kawakami, T., Raghavan, V., Ruhe, A.L., Jensen, M.K., Milano, A., Nelson, T.C., and Boyko, A.R. (2022). Early onset adult deafness in the Rhodesian Ridgeback dog is associated with an in-frame deletion in the EPS8L2 gene. *PLOS ONE* *17*, e0264365. <https://doi.org/10.1371/journal.pone.0264365>.
 142. Browning, B.L., Tian, X., Zhou, Y., and Browning, S.R. (2021). Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* *108*, 1880–1890. <https://doi.org/10.1016/j.ajhg.2021.08.005>.
 143. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* *4*, s13742-015-0047–0048. <https://doi.org/10.1186/s13742-015-0047-8>.
 144. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* *81*, 559–575. <https://doi.org/10.1086/519795>.
 145. Patterson, N., Price, A.L., and Reich, D. (2006). Population Structure and Eigenanalysis. *PLOS Genet.* *2*, e190. <https://doi.org/10.1371/journal.pgen.0020190>.
 146. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* *44*, 821–824. <https://doi.org/10.1038/ng.2310>.
 147. Mollandin, F., Rau, A., and Croiseau, P. (2021). An evaluation of the predictive performance and mapping power of the BayesR model for genomic prediction. *G3 GenesGenomesGenetics* *11*, jkab225. <https://doi.org/10.1093/g3journal/jkab225>.
 148. R Core Team (2023). R: A Language and Environment for Statistical Computing.
 149. Goodrich, B., Gabry, J., and Brilleman, S. (2024). rstanarm: Bayesian applied regression modeling via Stan. Version R package version 2.32.1.
 150. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* *28*, 1–26. <https://doi.org/10.18637/jss.v028.i05>.
 151. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* *88*, 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>.
 152. Szpiech, Z.A., and Hernandez, R.D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* *31*, 2824–2827. <https://doi.org/10.1093/molbev/msu211>.
 153. Biscarini, F., Cozzi, P., Gaspa, G., and Marras, G. (2019). detectRUNS: An R Package to Detect Runs of Homozygosity and Heterozygosity in Diploid Genomes. Version 0.9.6.
 154. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* *17*, 122. <https://doi.org/10.1186/s13059-016-0974-4>.
 155. Genereux, D.P., Serres, A., Armstrong, J., Johnson, J., Marinescu, V.D., Murén, E., Juan, D., Bejerano, G., Casewell, N.R., Chemnick, L.G., et al. (2020). A comparative genomics multitool for scientific discovery and conservation. *Nature* *587*, 240–245. <https://doi.org/10.1038/s41586-020-2876-6>.
 156. Wang, Y., Song, C., Zhao, J., Zhang, Y., Zhao, X., Feng, C., Zhang, G., Zhu, J., Wang, F., Qian, F., et al. (2023). SEdb 2.0: a comprehensive super-enhancer database of human and mouse. *Nucleic Acids Res.* *51*, D280–D290. <https://doi.org/10.1093/nar/gkac968>.
 157. Szklarczyk, D., Kirscher, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* *51*, D638–D646. <https://doi.org/10.1093/nar/gkac1000>.
 158. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* *34*, D590–D598. <https://doi.org/10.1093/nar/gkj144>.

A**B**

A**B****C**

A



B

