Manuscript Click here to view linked References

The shared selection landscape of dog and human cancers

Authors: Diane P. Genereux[1*C], Kate Megquier[1*], Ross Swofford[1], Jason Turner-Maier[1], Michelle White[1], Vista Sohrab[1,2,3], Christopher Husted[1,2,3], Heather Gardner[4], Corrie Painter[1,5], Cheryl London[4], Elinor K. Karlsson[1,2,6,7C]

* Co-first

[C] Corresponding author

1 Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

2 Genomics and Computational Biology, University of Massachusetts Chan Medical School, Worcester, MA 01655, USA

3 Morningside Graduate School of Biomedical Sciences, University of Massachusetts Chan Medical School, Worcester, MA 01655, USA.

4 Cummings School of Veterinary Medicine at Tufts University, North Grafton, MA

5 Precede Biosciences, Boston, MA

6 Darwin's Ark Foundation, Seattle, WA 98026, USA

7 Program in Molecular Medicine, University of Massachusetts Chan Medical School, Worcester, MA 01655, USA

Supplementary figures

Supplemental tables

Abstract

Cancers in pet dogs are prevalent, progress rapidly, and closely resemble human cancers, positioning them as powerful models for precision oncology. While genetic drivers of human cancer often transcend histologic boundaries, most comparative studies have focused on matched cancer types, leaving the broader scope of genomic similarity unresolved. We performed the first exome-wide, histology-agnostic comparison of canine and human cancers, analyzing 429 dog and 14,966 human tumors across 39 types. Mutational signatures and genes under selection are widely shared between species, and cancer types are as genomically similar between species as within species, with no greater similarity within dog breeds than between breeds. Machine-learning models identify genetic features shared by dog and human tumors of different histologies, mirroring cross-histology patterns in human cancer. These findings establish dog cancer as a powerful system for genomics-informed precision oncology and support pan-cancer approaches to discover translationally relevant models beyond histologic classification.

Introduction

Genomics-informed precision medicine has yielded promising treatments by targeting somatic mutations in tumor cells[1–6]. Some therapies initially approved for one cancer type have later proven effective against other cancers with mutations in the same genes[7–9]. For example, vemurafenib was initially developed to treat melanomas with V600E mutations in the *BRAF* gene. Subsequent pan-cancer studies revealed that it is also effective against tumors of 13 other cancer types that have the *BRAF* mutation, including several for which targeted treatments were not yet available[7]. Identifying genomic alterations shared across cancer types can reveal new treatment options for rare cancers, for which therapeutic development is often limited by small patient populations[10].

Even so, effective genomics-informed treatments remain unavailable for many patients. Existing drugs target mutations in only a small set of genes, and acquired resistance limits long-term efficacy[11,12]. Identifying targetable driver mutations amid the background of passenger mutations often requires large patient cohorts, which are difficult to assemble[13,14]. Once a promising target is found, drug development is slow and costly, and efficacy of small molecule inhibitors is not guaranteed. Clinical trials can span a decade and cost as much as $2.8 billion[15], and only about 10% of candidate drugs are ultimately approved[16]. Barriers are even steeper for the so-called “long tail” of rare driver mutations that occur in <5% of tumors, making it difficult or impossible to assemble patient cohorts adequate for target discovery and clinical testing[17–19].

Pet dogs offer a powerful and underutilized opportunity to accelerate development of targeted therapies in a widely accepted preclinical model with a very large patient population [20–24]. An estimated four million dogs are diagnosed with cancer each year in the United States[25]. Many pet dogs receive sophisticated care at veterinary specialty hospitals that parallels best practices in human oncology, including imaging, surgery, chemotherapy, immunotherapy, and radiation[26]. Furthermore, the less restrictive regulatory frameworks in veterinary medicine mean that new therapies can be tested in treatment-naive tumors, and dogs' faster disease progression expedites evaluation of outcomes. The unique population structure of pet dogs includes hundreds of genetically isolated subpopulations (breeds)[27,28], some of which have markedly high cancer risk[29–32].

Cancer in pet dogs recapitulates many key features of human disease that can affect drug response and disease outcome[33–35]. Like other mammals, dogs acquire somatic mutations at a rate inversely correlated with their lifespan, and an adult dog has an equivalent number of somatic mutations in normal tissues as an adult human[36]. Dogs encounter many of the same environmental and lifestyle risk factors as humans and develop genetically heterogeneous, spontaneous tumors within the context of an intact immune system[37–40]. Cancer in laboratory mice, in contrast, is typically induced through genetic manipulation in animals with suppressed or absent immune function, and even immunocompetent transgenic mouse models do not recapitulate the heterogeneity of human tumors[41,42]. Alternatives such as human cancer xenografts or humanized mice address some of these limitations, but still fail to capture the full complexity of an immunocompetent host[41,43].

Despite its potential to accelerate the development of genomically-informed targeted therapeutics, the dog model remains underexplored. Most studies to date have chosen dog cancers as models for human cancers based on matching by tissue of origin or clinical features[44–54]. However, it is increasingly clear that dog and human tumors from different tissues can share genomic features, including some relevant to treatment response. Clinical trials of tyrosine kinase inhibitors in dogs with mast cell tumors informed therapies for human gastrointestinal stromal tumors, as both cancer types have *KIT* gene mutations[55–58], and dog osteosarcomas often have mutations in the genes *SETD2* and *DMD*[59], which are rarely mutated in human osteosarcoma but

commonly mutated in other human cancers[60,61]. Leveraging shared genomic features between dogs and humans, including across different histologies, may help to inform how interactions among somatic mutations influence responses to targeted therapy. Dogs with *BRAF*-mutant urothelial carcinoma exhibit lower rates of clinical responses to the targeted agent vermurafenib[62], and could potentially help to elucidate the mechanism behind treatment resistance in *BRAF*-mutant human cancers with lower response rates, such as colorectal carcinomas[63].

Here, we comprehensively evaluate the genomic similarity of dog to human cancers, and assess the potential of dog cancers to reveal new models for developing genomics-informed targeted therapeutics.

Results

Constructing a comparative cancer dataset

We constructed the largest cross-species comparative cancer dataset to date by compiling publicly available somatic mutation data for human and dog tumor exomes. Each tumor sample included in our set had a paired sample from normal tissue, enabling accurate detection of somatic (non-germline) mutations. We considered exclusively exonic variants (both single nucleotide variants (SNVs) and short insertions/deletions) because these were the mutation calls available for most tumors (77% in humans and 82% in dogs)(**Table S1, Fig. S1)** . In total, we included 14,966 human tumors of 32 different cancer types from the Cancer Genome Atlas (TCGA)[64], International Cancer Genome Consortium (ICGC)[65] and CBioPortal[66], as well as 429 dog tumors of seven different types of cancer[59,67–73] (**Table S2, Fig. 1A).** The number of human tumors of each cancer type ranges from 45 angiosarcomas to 1784 breast cancers (mean=468±413 tumors per cancer type). The number of dog tumors per cancer type ranges from 23 mammary tumors to 109 B-cell lymphomas (mean=61±31)

For six of the seven dog cancer types, our dataset includes a human cancer type considered to be a histological match: osteosarcoma, melanoma, mammary tumors (breast cancer in humans), glioma (high-grade glioma in humans), B-cell lymphoma (diffuse large B-cell lymphoma in humans) and hemangiosarcoma (angiosarcoma in humans).

We analyzed 15,315 genes for damaging protein-coding mutations in dogs and humans. This included 15,289 genes with a single ortholog in dogs and 26 genes with two orthologs in dogs that are known cancer driver genes in humans (**Table S3**). In total, our list includes 1,098 genes known to drive cancer in humans (hereafter referred to as “driver genes”)[74–76] and 4852 genes identified as potentially druggable targets by OncoKB[74], including 464 driver genes (hereafter referred to as “druggable genes”). Most driver genes have not yet been functionally validated to also drive cancer in dogs. We classified as damaging all nonsynonymous mutations predicted to be damaging using either Ensembl Variant Effect Predictor or SnpEff[77,78].

Dog cancers model the genomic heterogeneity of human cancers

Overall, we found that cancer in dogs is a remarkably close genomic model for human cancers, with mutations in many of the same genes, often at the same genomic position, and a similar degree of genomic heterogeneity. We compared dog and human cancers in four dimensions: (i) the total number of mutations present; (ii) the distribution of mutations across the genome; (iii) the specific sets of genes mutated; and (iv) mutational signatures detected.

Similar number of damaging mutations

The number of damaging mutations in dog versus human tumors is indistinguishable after accounting for cancer type. Overall, dog tumors have significantly fewer damaging mutations than human tumors (Fig. 1B). However, in both species, cancer type has a significant effect on mutation rate (dog ANOVA effect=0.11, F=8.7, $p=6.9x10^{-9}$; human ANOVA effect=0.09, F=47.0, $p=1.8x10^{-273}$). $p=6.9x10^{-9}$; human ANOVA effect=0.09, F=47.0, $p=1.8x10^{-273}$). When analyzed by cancer type, dog and human tumors have similar median mutation rates (p_{wilcox}= 0.27) and have similar variability as measured by interquartile range (p_{wilcox}=0.19) (Fig. 1C). This suggests that the higher overall mutation burden observed in the human dataset is explained by a greater representation of tumor types with higher mutation rates.

We defined a set of human cancer types with mutation rates comparable to those observed in dogs ("low mutation rate human cancer types") by excluding nearly all human cancer types with a median mutation rates higher than the most highly mutated dog cancer type (dog osteosarcoma; median = 0.43 mutations/Mb) (Fig. 1C), with two exceptions. Both high-grade glioma (0.57 mutations/Mb) and angiosarcoma (0.62 mutations/Mb) had mutation rates slightly above our threshold (0.57 and 0.62 mutations/Mb respectively), but were also histiological counterparts for dog cancer types. We did exclude two histological matches that were among the most mutated human cancer types: diffuse large B-cell lymphoma (1.06 mutations/Mb) and melanoma (3.14 mutations/Mb). By retaining high-grade glioma and angiosarcoma, we were able to more robustly compare histologically matched and unmatched cancer types across species. Mutation rates in our "low mutation rate human cancer types" dataset are indistinguishable from dog cancers (Fig. 1B).

The rates at which genes are mutated is remarkably similar in dog and human tumors. Nearly all genes (15,313/15,315) were mutated in at least one of the 15,008 human tumors; a much smaller percentage of genes are mutated in at least one dog tumor (5462 genes; 36%; ks.test D=0.72, p=0). This difference was driven by the larger sample numbers and high mutation rate cancer types in humans, as it vanishes when we randomly downsample the human dataset to match the numbers of samples and cancer types in dogs, and exclude high mutation rate human cancer types (**Fig. 1D**)[79]. Two genes were not mutated in any human tumor: RPP21, which encodes a subunit of nuclear ribonuclease P and was also not mutated in dogs, and CD24, which encodes a cell adhesion protein and was mutated in a single dog tumor (a highly mutated melanoma).

Genes mutated in dog and human tumors are largely the same

With regard to the specific genes mutated, dog and human tumors of different cancer types are as similar as human tumors of different cancer types. We quantified mutational similarity for pairings of tumors using cosine similarity, which ranges from 0 (none of the same genes mutated) to 1 (all the same genes mutated)(**Table S4**). After excluding highly mutated cancer types, we randomly sampled up to 200 tumors per cancer type, yielding 3,031,953 human/human pairs, 91,806 dog/dog pairs and 1,056,627 dog/human pairs. The similarity of dog/human pairs (mean=0.005) is comparable to the similarity of tumors from different cancer types within species (dog/dog mean=0.0061; human/human mean=0.0039)(**Table S5**; **Fig. S2**). Highly similar tumor pairs are only slightly more likely to come from the same species (odds ratio = 1.09; CI 1.07-1.10; p_{fisher}=5x10-54), and this enrichment disappears when considering only driver genes (odds ratio = 1.0; CI 0.99-1.01; p=0.94) (**Fig. 1E**; **Table S6**).

The heterogeneity of mutated genes within and between cancer types is also similar in dogs compared to humans[80]. Tumors are more similar within a given cancer type than between cancer types in both dogs (mean=0.0172 vs. 0.0061; p_{wilcox}=0) and in humans (mean = 0.0228 vs 0.0039; p_{wilcox}=0). There is also a significant enrichment for highly similar tumor pairs within cancer types in both dogs (OR=2.8; CI 2.7-3.0;

p_{fisher}=0) and humans (OR=7.0; CI 6.9-7.1; p_{fisher}=0)(**Fig. 1E**). The highest mean similarity for human tumors was between low-grade gliomas (median=0.056 for 19,900 pairs) and for dogs between pairs of hemangiosarcomas (mean=0.025 for 946 pairs)(**Table S7**, **Fig. S2**).

Minimal effect of dog breed on spectrum of damaging mutations

We found no evidence that somatic mutations in tumors differ substantially between dog breeds. Our dataset includes dogs from over 50 dog breeds. When considering both cancer type and breed, breed had no significant effect on the number of mutations (Anova ges=0.047, p=0.31 for 16 breeds with three or more dogs; ges=0.051, p=0.004 for cancer type) (Fig. S3). There is no difference in the cosine similarity of tumors from different breeds compared to those from the same breed (p=0.06 for all genes and p=0.9 for driver genes)(Fig. S2). Tumors of the same type from the same breed are also no more enriched for highly similar pairs than tumors of the same type from different breeds when considering all genes (OR=1.05; p=0.33) or just driver genes (OR=0.97; p=0.55) (Fig. 1E).

Genes shared between dogs and humans are enriched for driver genes

Genes with higher rates of mutation in both dog and human tumors are more enriched for driver genes than in either species independently, illustrating the potential for information from dogs to reveal genes relevant to cancer progression. We considered a gene mutated if it had one or more predicted-damaging mutations. Overall, genes mutated in both human and dog tumors are enriched for both druggable genes (binomial p= $2x10^{-4}$, enrichment=1.1) and driver genes (binomial p = $1.9x10^{-25}$, enrichment=1.4). In both dogs and humans, enrichment for driver genes is most pronounced among the genes most frequently mutated (**Fig. 2A**). Even after correcting for gene-length, the fraction of tumors with mutations in a given gene is also correlated between the two species for genes overall (R^2= 0.53; p=0), and especially for driver genes (R^2=0.78; p= $4.2x10^{-222}$).

Our finding that the genes driving human cancer are particularly overrepresented among genes under positive selection in both dog and human tumors suggests that mutations in these genes confer a proliferative advantage in both species. Tests for positive selection can distinguish genes with mutations that confer growth advantage from passenger mutations[81,82]. For each gene, in each species, we calculated the ratio of nonsynonymous missense mutations to synonymous mutations (dN/dS) using the software dNdScv, which controls for variation in gene length and mutation rates[79]. Considering whether a gene has a high dN/dS in either species yields a stronger enrichment for driver genes than considering each species on its own (**Fig. 2B**). Across all genes, those with higher dN/dS score in human tumors tend to have higher dN/dS scores in dogs as well (**Fig. 2C**), and the strength of selection on each gene is correlated between the two species, especially for driver genes (**Fig. 2D**).

Some driver genes rarely mutated in human tumors are commonly mutated in dogs

Some driver genes that are rarely mutated in human tumors are more commonly mutated in dog tumors, indicating potential for assembling larger study cohorts to evaluate targeted therapeutics. *POT1*, which drives angiosarcoma, a cancer rare in humans[83], is mutated in 4.4% of the dog tumors in our set, compared to just 1% of all human tumors (p=$4.8x10^{-7}$), *FBXW7*, a driver gene in uterine carcinosarcoma[84], is mutated in approximately 6% of the dog tumors in our set, but only 1.9% of human tumors (p=$5.8x10^{-7}$)(**Table S8**), and

DDX3X, which is often mutated in B- and T-cell lymphoma and melanoma[85], is mutated in ~4% of all dog tumors, but just ~1% of human tumors ($p=7.7x10^{-5}$). The prevalence of these mutations in part reflects the large proportion of dog B-cell lymphomas (25.4% of all dog tumors) in our dataset. All three of these genes are under positive selection (dN/dS>1 and p_{FDR}<=0.05) in dog B-cell lymphomas and frequently mutated (17.4%, 21.1%, and 12.8% for *POT1*, *FBXW7* and *DDX3X* respectively). In humans, mutation rates peak for these three genes at 17.8% for *POT1* in angiosarcomas, 17.8% for *FBXW7* in uterine & endometrial carcinomas and 6.2% for *DDX3X* in pediatric medulloblastomas (**Table S9**).

Many genes are under stronger selection in a dog cancer type than in any human cancer type. About 16.1% of all genes mutated in dog tumors (882 genes), and 11.9% of driver genes (67 genes), have higher dN/dS in at least one dog cancer type than in any of the 32 human cancer types we analyzed (**Table S9**). Fourteen genes are under significant selection (p_{FDR}<0.05) in at least one dog cancer type, despite the relatively small sample numbers for each dog cancer type. These include *SATB1,* which is under selection in both human thyroid cancer (THCA) and dog T-cell lymphoma, but is rarely mutated in humans (frequency=0.006) and commonly mutated in dogs (freq=0.29). It also includes *PIK3CA,* which is mutated in 31.8% of dog hemangiosarcomas, a higher frequency than in any human cancer type except uterine & endometrial carcinomas.

Mutation hotspots are shared.

In addition to sharing many of the same frequently mutated genes, dog cancers often have mutations at the same frequently mutated "hotspot" positions within those genes (**Fig. 3A**; **Table S10**). We here define a site as a hotspot if it was either previously reported as a human hotspot[86,87] and or mutated in five or more samples in our set.

The five genes with the most commonly shared hotspots are driver genes: *TP53* (52/130 sites), *PIK3CA* (6/24 sites), *FBXW7* (6/12 sites), *PTEN* (3/24 sites) and *NRAS* (3/3 sites). This confirms previous work that described this phenomenon in smaller numbers of tumors[69,73] and candidate genes[88]. We found 148 hotspots in humans that are also mutated in at least one of the 429 dog tumors. This is a relatively small percentage of all human hotspots identified from previous studies[87,89,90] (2.7%; 148 out of 5377 hotspots in 3450 genes), but this is expected given the smaller size of the dog dataset. Nearly all human hotspots (99.1%; 5428 hotspots) have a frequency below our detection threshold in dogs (1/429 dogs). This suggests that if we had data for more dog tumors, and thus more power to detect lower frequency hotspots, we would discover more shared hotspots. Consistent with this, hotspots with higher frequencies in human tumors are more likely to also be found in dogs ($p_{wilcox}=1.9x10^{-5}$)(**Fig. 3B**).

Hotspots shared between dogs and humans are strongly enriched for cancer driver genes (hypergeometric $p=5.9x10^{-37}$), and overall the frequency of hotspot mutations is correlated in driver genes but not other genes (**Fig. 3C**). Of the 10 most common hotspots in dogs, including some more common in specific dog cancer types, only four have a frequency greater than 0.02 in any human cancer type, and all four are in driver genes (**Fig. 3D**). Despite the smaller size of the dog dataset, we identified six "dog-specific" hotspots mutated in at least five dog tumors and no human tumors. This includes a hotspot in *NOC3L*, a gene involved in the initiation of DNA replication, found in 11 dog tumors, including 5 out of 41 T-cell lymphomas. Another dog-specific hotspot alters the adhesion G-protein coupled receptor *ADGRE* and is found in tumors from three different dog cancer types (B-cell lymphoma, T-cell lymphoma and melanoma). In humans, aberrant expression of *ADGRE* is associated with poor outcomes in patients with chronic myeloid leukemia[91].

Mutational signatures are shared.

Dog and human tumors exhibit the same mutational signatures and at similar frequencies, suggesting that the biological processes, environmental exposures, and DNA repair deficiencies underlying cancer development are also shared. We screened each sample for all of the 43 mutational signatures detected in at least one tumor in our dataset. Within both dog and human cancer types, we find heterogeneity in the abundance of each signature (**Fig. S4, Fig. S5**). We considered a signature to be present in a sample if its value exceeded 0.05, consistent with published guidelines[92](**Table S11**). Twenty-eight of the 43 signatures detected in human tumors were also detected in at least one dog tumor. The fraction of tumors with a given mutational signature was correlated between dogs and humans (Spearman's rho=0.61; $p=1.2x10^{-5}$)(**Fig. 4A**).

Seven of the top ten most common mutational signatures were shared between the two species. The aging-associated signature SBS1 was the most common in both, appearing in 31% of human and 15% of dog tumors. Three other signatures were present in at least ten dog tumors: SBS6 (17% in humans, 5% in dogs) and SBS15 (6% in humans, 3% in dogs), both implicated in defective DNA mismatch repair, and SBS10b, which is linked to defective proofreading activity of POLE (4% in humans, 3% in dogs)(**Fig. 4A**).

Overall, mutational signature values are broadly similar between dogs and humans. Although the maximum signature value per tumor is often higher in human tumors (**Fig. 4B**), linear regression indicates no overall effect of species ($p = 1$) or mutation count ($p = 1$) on signature values. However, there is a strong interaction between species and specific signatures ($p < 2.5x10^{-214}$), explaining 9.1% of the total variance in signature values (**Fig. 4C**). The species effect varies by signature: values are significantly higher in dogs for 25 signatures (FDR < 0.05), including four with effect sizes >0.005, and significantly higher in humans for six signatures, also including four with effect sizes >0.005. For example, SBS1, associated with spontaneous deamination of 5-methylcytosine and cellular aging, is higher in humans, possibly reflecting their longer lifespan[93]. SBS2 and SBS13, linked to APOBEC activity, are also more abundant in human tumors[94].

The mutational signatures detected in dog tumors suggests many, but not all, environmental risk factors for cancer are shared between dogs and humans **(Fig. 4D)**. UV-associated signatures like SBS7a are common in humans but rare in dogs (detected in fewer than 1% of dog tumors), consistent with the protective effect of fur[95]. In turn, dog tumors tend to have higher values for SBS38, which is linked to indirect damage from UV-light[96](**Fig. 4C**). However, when UV-associated signatures are detected in dogs, it is in cancer types linked to UV exposure in humans. The two human cancer types where SBS7a signatures are most common are melanoma (80%) and angiosarcoma (36%). In dogs, we detect SBS7a in cancers from two tissues not typically exposed to UV,oral melanoma and visceral hemangiosarcoma, the histological analog of angiosarcoma[97,98], a phenomenon previously described in human lung melanomas[99]. We do not detect SBS4, a signature primarily associated with the human activity of tobacco smoking, in any dog tumors, although our dataset does not include dog nasal carcinomas, which have been associated with second-hand smoke exposure[100]. In four dog osteosarcomas, we detect SBS24, a signature proposed to be linked in humans to exposure to the mycotoxin aflatoxin[101], a known possible contaminant in dry pet food[102]. We also find signatures associated with exposure to the carcinogenic herb *Aristolochia* in humans and mice[103,104] (SBS22 in one dog B-cell lymphoma) and with prior treatment with azathioprine in humans (SBS32 in one dog glioma)[105]. Azathioprine is a purine anti-metabolite also used to treat autoimmune diseases in dogs[106]. We note, though, that while mutational signatures are known to be shared across mammals[36], the associations of specific signatures with environmental and endogenous exposures have primarily been discovered in humans[96] and have yet to be validated in dogs.

Dog cancers resemble human cancers in variation of mutational signatures present.

Dog tumors are almost indistinguishable from human tumors in terms of mutational signature values once the high mutation rate human cancer types are excluded. We calculated cosine similarity for all tumor pairs (3,031,953 human/human pairs, 91,806 dog/dog pairs and 1,056,627 dog/human pairs) . The median similarity of tumors of different types was similar in dogs (median=0.955 +/- IQR 0.053; N_{pairs}=76,030), humans (median=0.950 +/- IQR 0.117, N_{pairs}=2,823,437) and when comparing dog and human tumors (median=0.953 +/- IQR 0.075; N_{pairs}=1,056,627) (Table S5).

Tumors pairs from the same species are slightly enriched for close matches (OR=1.09; CI 1.07-1.10; $p_{fisher}=5x10^{-55}$), but much less so than same-species tumor pairs from the same cancer type in both dogs (OR=2.03; CI 1.89-2.17; $p_{fisher}=4x10^{-84}$) and humans (OR= 1.60; CI 1.58-1.63; p_{fisher}=0) (**Fig. 1E**). In dogs, breed has little effect. Tumors of the same type from the same breed have a slightly lower proportion of high-similarity pairs than do tumors of the same type from different breeds (OR=0.870; CI 0.763 - 0.991; p_{fisher}=0.03), suggesting that dog's breed does not strongly influence the mutational signatures in the tumor (Table S7).

Matching dog and human tumors with high genomic similarity

Unsupervised clustering finds no distinct clusters

No dog cancer type is distinct from human tumors or closely aligned to any single human cancer type when tumor relationships are visualized using t-distributed Stochastic Neighbor Embedding (t-SNE). To cluster based on mutational signatures, we ran t-SNE on the principal components that together explained more than 70% of the variance (**Table S12**, **Table S13**). To cluster based on driver gene mutations, we applied t-SNE directly to the matrix of mutated driver gene counts. We used this approach because the mutation matrix is sparse, and each principal component explains only a small fraction of the variance, even using methods optimized for sparse data. Consistent with our comparisons of driver mutations and signatures in dog and human cancers, dog tumors overlap low mutation rate human cancer types in the t-SNE analysis (**Fig. S7**). The lack of separation of dog tumors from low mutation rate human tumors is even more evident in a t-SNE analysis that excludes the high mutation rate human cancer types (**Fig. 5AB; Fig. S8; Fig. S9**).

Clustering tumors using the mutational signature principal components similarly showed that dog tumors are intermingled with human tumors. We performed k-means clustering and selected the number of clusters using the elbow method based on the within-cluster sum of squares. When clustering all tumors (10 clusters), 94% of dog tumors were assigned to the largest single cluster, together with half of all human tumors (**Fig. S10A,B**). This proportion decreased to 43% when we included only low mutation rate human tumors (13 clusters), and the number of clusters containing ten or more dog tumors increased from two to five (**Fig. 5C**; **Fig. S10B**). Each dog cancer type was distributed across at least four clusters, and clusters with a more heterogeneous representation of human cancer types, as measured by the Shannon entropy, were also more heterogeneous in dogs (**Fig. S10C**)(Spearman rho=0.93, $p=4x10^{-5}$). We observed some enrichment of certain cancer types within specific clusters. For example, both dog osteosarcomas (observed = 50, expected = 16.6, $p_{FDR} = 7x10^{-19}$) and human thymomas (observed = 13, expected = 3.6, p_{FDR} = 0.009) were enriched in cluster 5 (**Fig. S10D**).

Supervised machine learning finds cross-species genomic similarities between cancer types

We used supervised machine learning to identify pairs of dog and human cancer types enriched for shared genetic features, even in the absence of broader genomic similarity. Our approach, which was previously used to classify human cancers[107], could identify dog cancer types that are potential models for trialling targeted therapies. We first learn the genomic features of the histology-labeled human tumors (our training set), then use those features to classify each dog tumor as a human cancer type. To quantify the relative contribution of individual features to classification, we computed Shapley values[108,109]. This approach is well-suited for classifying samples using genomic features common in the human cancer types well-represented in our data set. It may miss pairs previously reported to share genomic features if there are few human samples in the training set or if the shared genomic features are not included in our model (e.g. noncoding, structural, and epigenetic changes[70]). It will also miss histology-matches with few shared genomic features because of differing etiology (e.g. angiosarcoma and melanoma in humans frequently occur in sun exposed skin).

To implement machine learning, we use eXtreme Gradient Boosting (XGBoost[110]) models with up to 40,189 features, including mutated genes and pathways, mutational signature values, and overall mutation count. XGBoost is a decision tree–based algorithm that is well-suited for somatic mutation data, as it efficiently models non-linear relationships and interactions among sparse, high-dimensional features while filtering out uninformative variables. We trained models on two groups of human cancers: (1) 17 low mutation rate cancer types and (2) all cancer types. For each group, we evaluated three different feature sets: the "full model" (655 features retained as informative by machine learning), a "driver gene mutations model" (all 1098 driver genes), and a "mutational signatures model" (all 43 signatures). We calculated the average accuracy of the classifier on held-out human data across 5 folds using the k-fold cross validation procedure[111]. Using the full model, tumor classification accuracy was 63.5% for low mutation rate cancer types and 63.4% for all cancer types (**Fig. S11**, **Fig. S12, Table S14**). This is comparable to previous classifiers trained on somatic mutations in exome sequencing data[112,113].

We also assessed the performance of a model trained on the much smaller dataset of dog cancers. Accuracy under our dog-trained model was slightly lower (57.7%) than under our human-trained model (**Table S14, Fig. S11C**, **Fig. S12**). Classification accuracy for individual dog cancer types was positively correlated with the number of tumors of that type in the dataset ($R=0.87$; $p = 0.018$). Variation in classification accuracy across cancer types was explained in part by the number of samples available (11% of variance explained; $p=0.047$) , and not by species (0% of variance explained; $p=0.99$).

Classifying dog tumors as human cancer types

We applied the human-trained classifier to assign each dog tumor to a human cancer type, based on genomic features. Using models trained on low mutation rate cancer types, tumors of five of the seven dog cancer types were assigned to fewer human cancer types than expected by chance. We quantified this phenomenon of “clumpiness” using Shannon’s entropy, and measured the statistical significance via permutation (**Table S14**). Of the seven dog cancer types, hemangiosarcoma, osteosarcoma and T-cell lymphoma are clumpy in both the full model and at least one other model (**Fig. 6A**). Glioma and B-cell lymphoma are clumpy only in the mutational signatures model. Using models trained on all cancer types, six of the seven cancer types were clumpy in at least one model (**Fig. S13A**).

Dog tumors of all types were classified to multiple different human cancer types, consistent with the genomic heterogeneity of both human and dog cancer types. Even so, some dog cancer types were classified as a specific human cancer type more often than expected by chance (**Fig. 6B**, **Fig. S13B, Table S15)**. For

example, using the full model trained on low mutation-rate cancer types, dog hemangiosarcoma samples were often classified as high-grade glioma ($p=2x10^{-4}$) or breast cancer (p=0.038), and dog T-cell lymphoma was classified as prostate adenocarcinoma ($p=6x10^{-5}$). Under all three models, tumor classifications were informed by multiple features.

Low sample counts likely limited classification of some dog tumors to histology-matched human cancer types. Dog tumors were more often assigned to human cancers that were better represented in the training set (Pearson's r=0.49; $p=1.6x10^{-8}$ for the full model). Histology-matched cancer types were among those with fewest samples, including B-cell lymphoma (N = 172), osteosarcoma (N = 117), and angiosarcoma (N = 45), which may explain why machine learning did not assign dog tumors to these cancer types any more than expected by chance, even when the dog cancer type was previously reported to share genomic features (e.g. diffuse large B-cell lymphoma[67] (N=172), osteosarcoma[59,68] (N=117), and angiosarcoma[45,69](N=45)). In contrast, the histology-matched type for canine mammary tumors, human breast cancer, was the most abundant cancer type in the dataset (N = 1,784), and canine mammary tumors were assigned to this type at a frequency higher than expected by chance in the driver-only model (4 of 23, p = 0.0218), consistent with previous reports[73].

The classification of most dog hemangiosarcomas as both breast cancer (14/44) and high-grade glioma (16/44) was informed by a combination of genomic features, illustrating the capacity of machine learning to integrate diverse signals. Classification of most hemangiosarcomas as either glioma or breast cancer was informed by mutations in the TCGA Glioblastoma Mutated[114] gene set (**Fig. 6C-D, Fig. S14, Table S16**, **Table S17**), consistent with the established role of the PI3K pathway, which includes many of these genes, in these cancer types[45,69,115,116]. Classification as breast cancer versus high-grade glioma depended in part on which genes were mutated. *TP53* was mutated in every hemangiosarcoma classified as high-grade glioma (N=16/16, mean Shapley = 0.784), and multiple genes in this pathway were also mutated in seven tumors (mean subset Shapley=0.599). Mutations in the TCGA_Glioblastoma_Mutated gene set were present in most hemangiosarcomas classified as breast cancer (mean Shapley value in N=13/14 samples where this feature was positive (mean subset Shapley) = 0.384), most commonly *PIK3CA* (N=11/14). However, this classification was also informed by the APOBEC signatures in a subset of samples SBS13 (N=3/14, mean subset Shapley=0.548) and SBS2 (N=2/14, mean subset Shapley=0.557), with SBS13 being the most important feature in a single sample. APOBEC signatures are found in more than 75% of human breast cancers and promote metastasis and treatment resistance[117–119].

In contrast, the classification of 18 out of 41 dog T-cell lymphoma as prostate adenocarcinoma was informed primarily by mutational signatures. For all 18 tumors, signatures were among the top explanatory features. Multiple signatures contributed to each classification. The only non-signature feature informative for most samples was low to moderate mutational burden ($mean_{dog}$=34 mut/Mb) (N=17/18, mean subset Shapley =0.313), while the signatures informative for most samples included higher levels of SBS28, which is linked to decreased DNA proofreading by polymerase epsilon[120] (N=17, mean subset Shapley = 0.386), the mismatch repair deficiency signatures SBS6 (N=16/18, mean subset Shapley =0.278) and SBS15 (N=16/18, mean subset Shapley=0.103), and the aging signature SBS1 (N=17/18, mean subset Shapley =0.198). In addition, a subset of classifications were informed by higher levels of UV light signature SBS7d[121] (N=6, mean subset Shapley = 0.056) and lower levels of signatures related to reactive oxygen species[122,123] (SBS17a, N=10, mean subset Shapley = 0.145) and alkylating agents[124] (SBS11, N=18, mean subset Shapley = 0.057).

DISCUSSION

Cancer is unequivocally the same phenomenon in dogs and humans. Genomic similarities exist across several dimensions: the number and frequency of mutations, the specific genes and hotspots mutated, and the mutational signatures present. Together these traits confer similarities so extensive that pairs of dog and human tumors are, on average, no more genomically distinct than pairs of samples of different cancer types within each species. The strength of positive selection on individual genes is correlated between the two species, suggesting that mutation sharing is not merely a byproduct of similar mutation rates for orthologous genes, but instead a reflection of cancer mechanisms shared between dogs and humans. Though some human cancer types, particularly those induced by high carcinogen exposure, have higher mean mutation counts and concomitantly higher mutational signature values than observed in any dog cancer, mutation counts for dog cancer types fall within the overall range for human cancers.

The similarity of somatic mutational landscapes in dog and human cancers is consistent with the deep conservation of canonical cancer pathways across eukaryotes. Dogs and humans are mammalian species that diverged approximately 94 million years ago[125], with total evolutionary branch length of 1.99 substitutions per site[126]. The distance is only slightly greater than the distance of house mice (*Mus musculus*) to human (1.64 substitutions per site; about 87 million years). In contrast, most canonical cancer pathways, including those involved in DNA damage response and repair, are conserved across vertebrates (588 million years), and often across all eukaryotes (1598 million years)[127]. While some species may have evolved lineage-specific modifications to suppress cancer development[128], a comparison of 16 mammalian species found that somatic mutation rates in normal intestinal tissue across mammals are evolutionarily constrained and inversely correlated with lifespan (thus, an old dog is comparable to an old human), and result from common mutational processes that produce shared mutational signatures[36].

This positions dogs as a powerful natural model for development of new genomics-informed targeted cancer therapies. Unlike induced cancer models in mouse, for example, cancer in dogs is as genomically heterogeneous as in humans, with mutational landscapes that mirror human cancers. Some driver mutations that are rare in human cancers are common in dogs, offering a unique opportunity to develop therapeutics for genomically defined subsets of human patients[129]. Dog and human cancers also share mutational signatures, many of which have yet to be conclusively connected to a specific exposure or genetic risk feature. Dogs could therefore help connect environmental exposures to cancer risk, especially given their short lifespans and accelerated disease course.

Pet dogs could be particularly useful for investigating tumor evolution during patient treatment[130]. In dogs, new therapeutics can be more easily trialed in treatment-naive tumors, and longitudinal studies completed more quickly because of the faster disease course[131–133]. Our comparisons show that the strength of selection acting on cancer driver genes is correlated between dogs and humans, and that dog tumors may have mutational signatures associated with prior treatment or treatment resistance. Trials in dogs would support the development of approaches to improve accuracy in predicting therapeutic resistance, tumor progression, and recurrence[134].

Our machine learning results demonstrate the value of dog cancers as models for precision oncology. For example, we find that dog hemangiosarcomas commonly share genomic features with two different cancer types in humans, breast cancer and high-grade glioma, including mutations in *PIK3CA* and *TP53*, respectively. Given that hemangiosarcoma is estimated to affect up to 300,000 pet dogs annually in the United States, with few effective treatment options available, studies in dogs with hemangiosarcoma could address the recognized need for a model to discern how *PIK3CA* mutations impact vascular cancer[135,136]. Similarly, shared mutational signatures between dog T-cell lymphoma and human prostate adenocarcinoma suggest shared mutagenic

pressures. However, genomic heterogeneity within dog cancer types makes tumor histology alone an unreliable predictor of which dogs are most appropriate for inclusion in clinical trials of targeted therapies.

Our observation of genomic heterogeneity within histologically-defined cancer types of both dog and human aligns with the well-established phenomenon of molecular cancer subtypes within[137,138] and across[139] cancer types. This heterogeneity likely affected the performance of our machine-learning models, for which training data are labeled only by cancer type, and not by subtype. Mislabeled or ambiguous data can lead the model to learn incorrect patterns, thereby reducing predictive power and increasing error rates[140]. Unfortunately, while subtype annotations are available from some of the data sources used here, subtype labels are difficult to assign consistently across data sources[141]. Furthermore, for cancer types with relatively few samples, subtype labeling is often less reliable[137]. Subdivision reduces sample sizes even more, undermining model performance[142]. As datasets grow and standardized subtype annotations are widely adopted, incorporating them into cross-species machine learning approaches could reveal genomic similarities between dog and human subtypes missed when labelling only by cancer type and potentially identify genomic models for human cancer subtypes that currently lack animal models[135].

Somewhat unexpectedly, we find that, while the prevalence of specific cancer types differs across dog breeds, the somatic mutational profile of each cancer type does not. This is in contrast to earlier studies in smaller cohorts[67,143], and has important implications for the design of comparative oncology studies. For studies targeting a specific cancer type, focusing recruitment on breeds with higher prevalence will yield a study population with a more uniform germline background, potentially reducing confounding factors common in human trials, while still capturing relevant somatic mutation heterogeneity. For many studies, though, particularly those focused on specific somatic mutations, restricting by breed is unnecessary. This will enable the recruitment of larger, genetically diverse cohorts that better mirror the diversity of human patient populations.

Our study focused on single base substitutions and short insertions/deletions in coding regions, the variant types for which data were most commonly available for both dog and human. This may have contributed to why, for example, dog osteosarcomas were not more often classified as human osteosarcoma. In both species, osteosarcoma has an unusually high burden of somatic copy-number alterations and structural rearrangements[59,68,144–149] that, in humans, have been used to define molecular subtypes with clinical relevance [150,151]. Our approach also does not capture non-coding variants that can activate oncogenes, as with *TERT* promoter mutations in human melanoma[152,153]. Although driver mutations in non-coding genes and regulatory sequence are likely less frequent than in protein coding genes, more continue to be discovered as human whole-genome cancer datasets grow[154]. Regulatory elements are often conserved across placental mammals[155,156], and mutations in conserved regulatory elements may affect activity of driver genes[157]. Extending machine learning training data to consider structural variants and non-coding mutations could identify additional dog cancer types as useful genomic models among, potentially expanding cohorts for pre-clinical testing of targeted therapeutics[158].

Achieving this will require expanding the scale and scope of dog cancer genomic data. Human pan-cancer discoveries have been enabled by major collaborative efforts to share genomic data for thousands of patients. Here, we leveraged TCGA[64], ICGC[65] and CBioPortal[66] to include data for 14,966 tumors from human patients. Despite dogs having a nearly two-fold higher annual cancer incidence than humans[159], with approximately 4 million diagnosed cases per year in the United States[25], available genomic data for dogs is very limited, even for just coding regions. Assembling exome data for the 429 dog tumors we consider here required aggregating results across eight different studies[59,67–73].

Dogs represent a unique opportunity for translational cancer research, as veterinary oncology routinely uses many of the same drugs, treatment modalities, and clinical metrics as human oncology. Dog owners are highly engaged partners, motivated by past experiences, the hope for better care, and the opportunity to contribute to medical advances. Today, mortality rates for dogs remain high—even with treatment: hemangiosarcoma kills 90% of dogs within a year[160]; osteosarcoma has a <20% two-year survival rate[161]; oral malignant melanoma has a one-year survival of 30%[162]; and median survival times are 6.5 months for high-grade glioma[163], 6–8 months for T-cell lymphoma, and ~12 months for B-cell lymphoma[164]. Establishing genomics as a routine and clinically valuable tool in veterinary oncology could rapidly expand available datasets, especially if data is shared through platforms like the NCI's Integrated Canine Data Commons[165], and provide patient populations for preclinical trials. By integrating dog cancer genomics into translational research pipelines, we can accelerate the development of targeted therapies for both human and veterinary oncology.

Resource Availability

Acknowledgements

The results published or shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

FIGURES

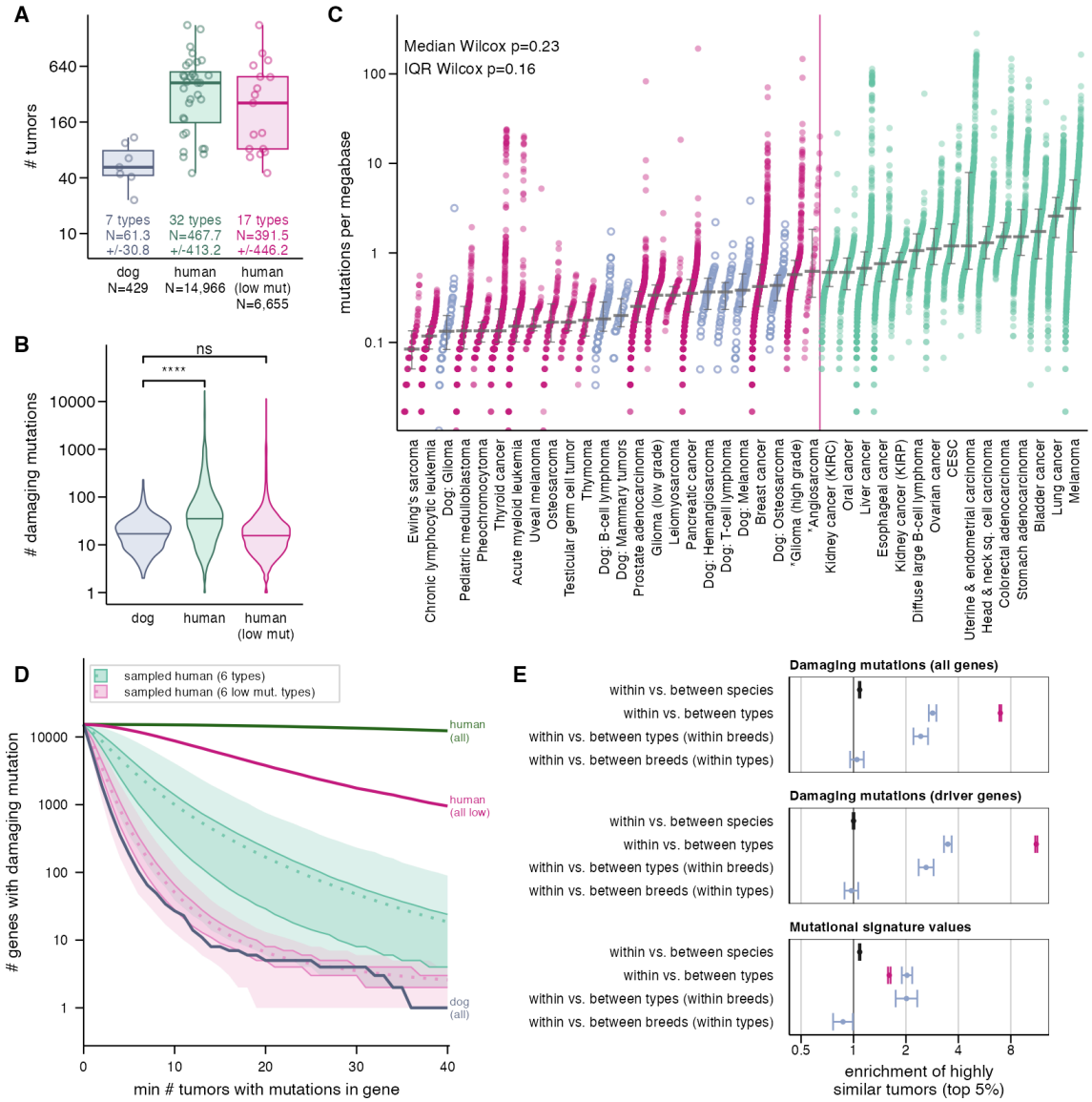


Figure 1. Differences between tumors are attributable to cancer type rather than species differences

A. Dataset includes 7 dog (blue) and 32 human (green) cancer types, including 17 we categorized as “low mutation rate” in our analysis (see panel C). Boxplots show median (line), IQR (box), and 1.5× IQR (whiskers); points represent individual tumors (jittered). **B.** Dog tumors harbor fewer damaging mutations than human tumors overall ($p = 2.8\times10^{-47}$), but are not significantly different from low mutation rate human tumors ($p = 0.19$, two-sided Wilcoxon rank-sum). “Low mut” set excludes human cancer types with unusually high mutation rates. **C.** Mutation frequency per tumor, ordered by cancer-type median, modeled on prior work in human cancer[79]. Thick gray bars indicate medians; thin bars indicate IQRs. Species differences are not significant after accounting for cancer type. **D.** The higher number of damaging mutations in human cancers is explained by larger sample sizes and the high mutation rate cancer types. Downsampling to match dog sample sizes

(1,000 permutations) shows comparable mutation frequencies between species. Dashed lines represent the median; the outer shaded ribbon the 95% range and the inner ribbon the 50% range. **E.** Tumors are no more similar within species than between species when considering either damaging mutations or mutational signatures, but are similar within cancer types. In dogs, tumors are not more similar within breeds. Colors: blue = dog; green = all human cancers; pink = low mutation rate human cancers.

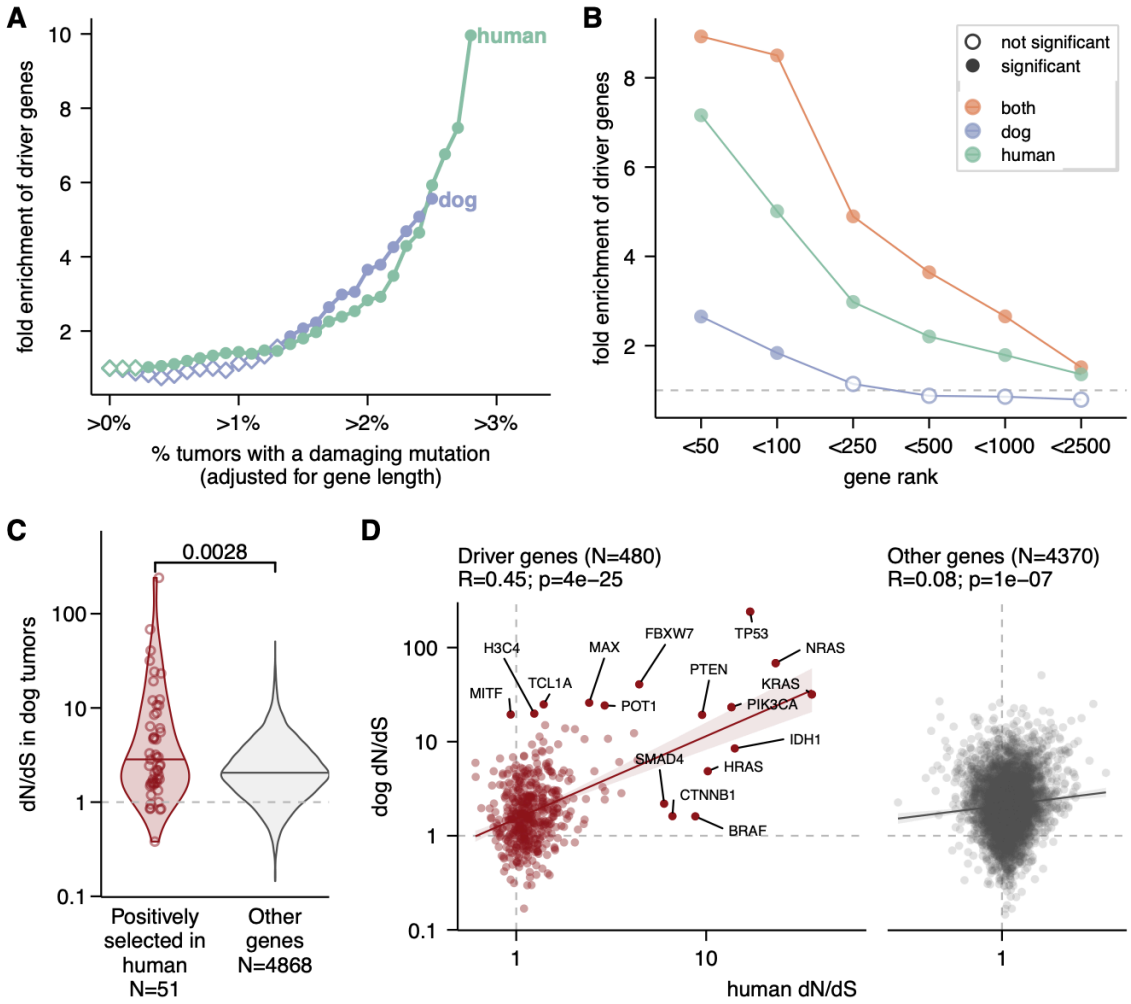


Figure 2. Driver genes are enriched for selection in both species.

A. The most frequently mutated genes are strongly enriched for known driver genes, with a similar rate of increase in enrichment as the percentage of mutated tumors rises in both dog (blue) and human (green) tumors. **B.** Considering whether a gene has a high dN/dS in either species yields stronger enrichment for driver genes than analyzing each species independently. The dashed line indicates no enrichment. **C.** Genes under positive selection in human tumors (red; defined as having a dN/dS ratio for missense mutations > 1 and FDR-adjusted p-value < 0.05) show higher dN/dS values in dog tumors than other genes (gray). **D.** The dN/dS ratio for missense mutations is significantly correlated between dog and human tumors for driver genes (red), but not for other genes (gray). Dashed lines indicate dN/dS = 1. Linear regression lines are shown for each category, with shaded bands representing 95% confidence intervals. Pearson correlation coefficients (R) and corresponding p-values from two-sided correlation tests are shown above each plot. Genes ranked in the top ten in either species are labeled.

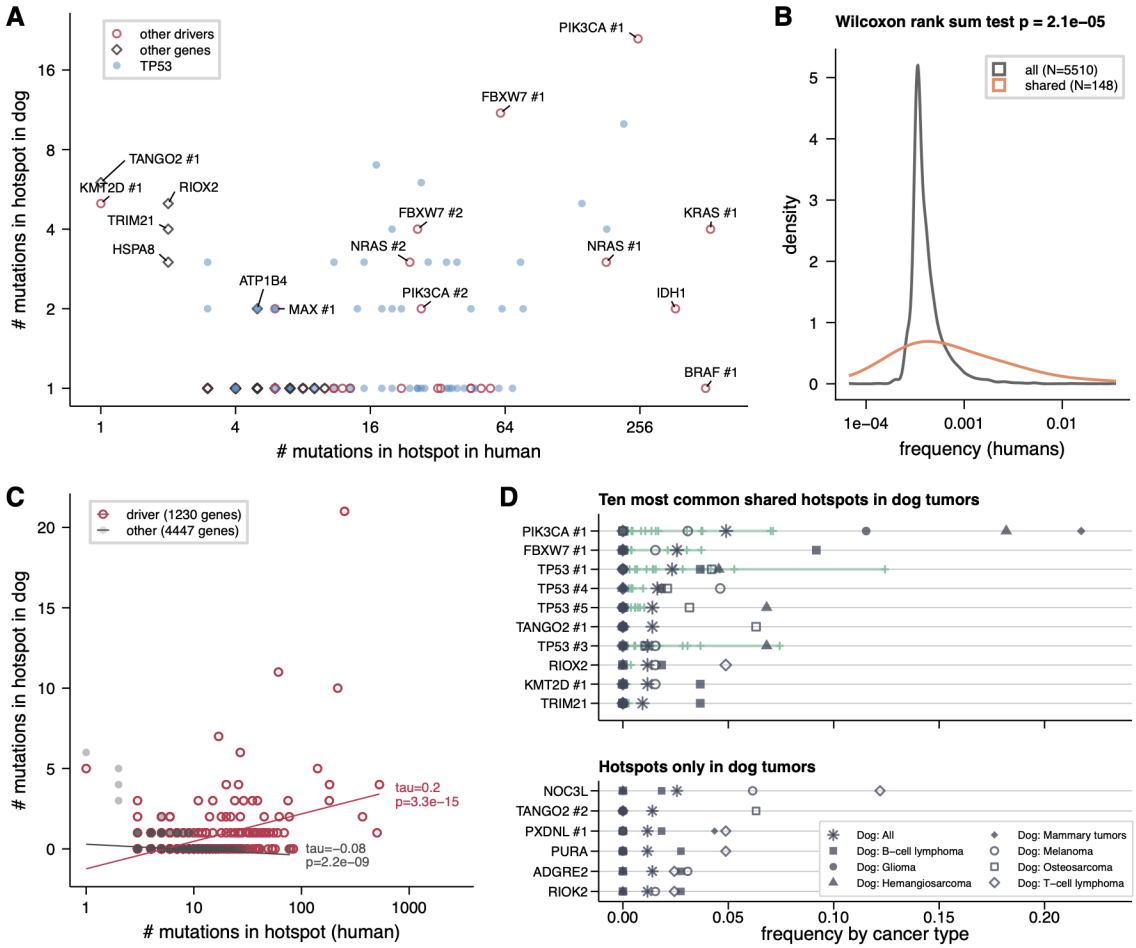


Figure 3. The same mutation hotspots are found in dog and human tumors.

A. Nearly all mutation hotspots that are common in human tumors are also observed in at least one dog tumor, despite the much smaller number of dog tumors analyzed. Some mutations that are rare in humans appear more frequently in dogs. **B.** Shared hotspots tend to occur at significantly higher frequency in human tumors, suggesting that the smaller size of the dog dataset may limit detection of less frequent hotspots. **C.** For hotspots in driver genes (red), the number of mutations is significantly correlated between dogs and humans, while no such correlation is observed for hotspots in other genes (grey). The strength and significance of the association are reported as Kendall's τ and the corresponding p-value. **D.** Of the ten most common hotspots in dogs that are also seen in human tumors, five occur at frequencies above 1% in at least one human cancer type. Six hotspots were observed in at least five dog tumors but in no human tumors. Dog cancer types are shown as grey symbols. The green horizontal line shows the range of frequencies across human cancer types, with each type indicated by a short vertical tick mark.

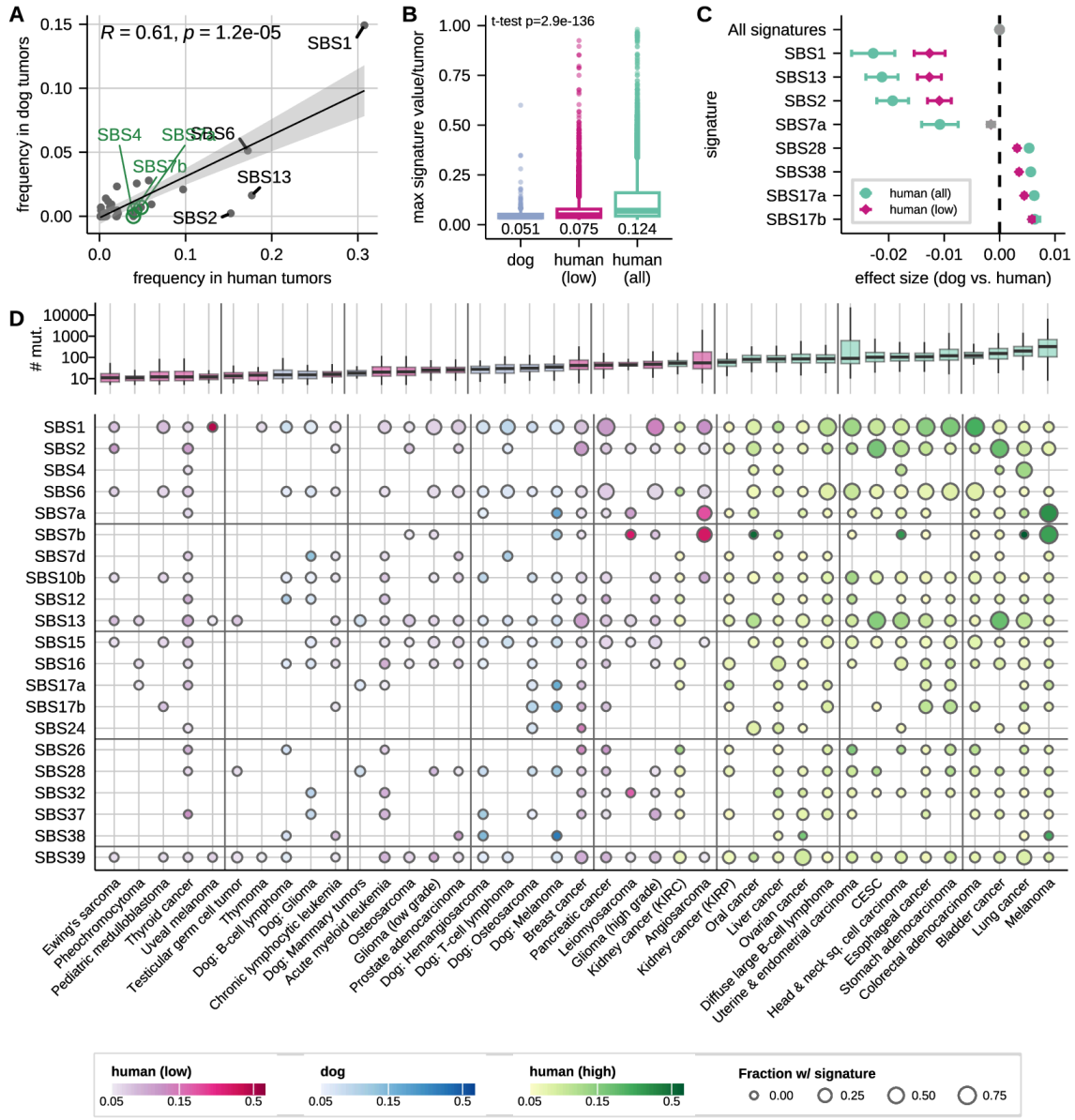


Figure 4. The same mutational signatures are found at similar frequencies in dog and human tumors

A. Correlation of mutational signature frequency between dog tumors and human tumors. The four most common signatures in human tumors are labeled in grey. Signatures linked to environmental exposures are labeled in green and occur at somewhat lower frequencies in dogs. These include SBS4 (tobacco smoke) and SBS7a and SBS7b (ultraviolet light). **B.** Human tumors show higher maximum signature values than dog tumors overall, but are more similar when only low mutation rate human cancer types are considered. Boxplots show the median (center line), interquartile range (box), and whiskers extending to 1.5× the interquartile range. Outliers are shown as individual points. **C.** Overall, species has no effect on signature values, but some signatures tend to be higher in humans (negative effect size) or higher in dogs (positive effect size). Non-significant (p_{FDR}<0.05) effect sizes in gray. **D.** Mutational signatures detected in dog and human tumors.

The most common signatures in dogs, and those in panel C, are shown here. Plot of all signatures shown in Fig. S6.

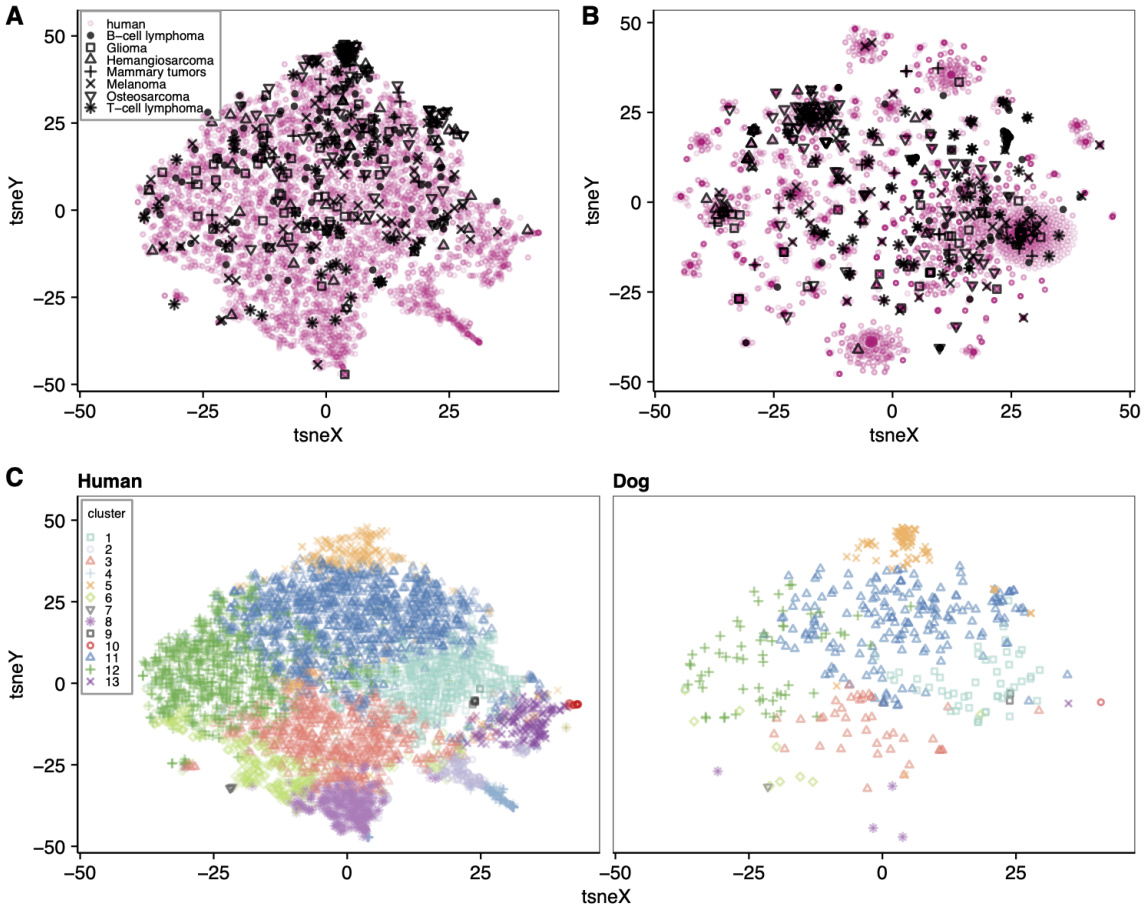


Figure 5. Unsupervised clustering does not separate dog cancers from human cancers.

Dog cancers overlap low mutation rate human cancers and do not separate by cancer type in t-SNE visualizations based on **(A)** principle components explaining 70% of sample variance in mutational signatures and **(B)** which driver genes are mutated. (**C)** K-means clustering using the mutational signature principal components yields 13 clusters, 11 of which include both human and dog tumors.

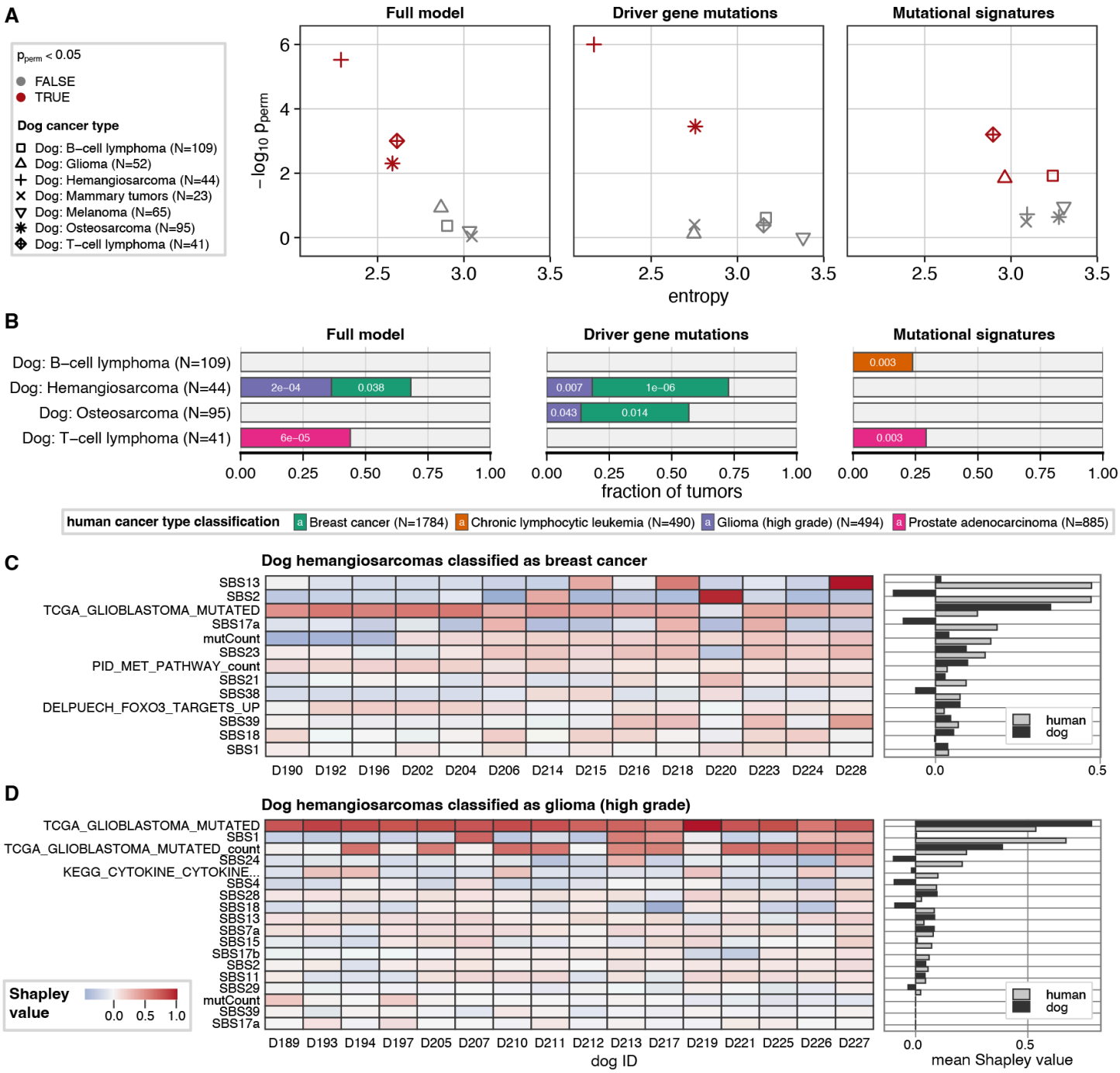

A
Full model
Driver gene mutations
Mutational signatures
p_{perm} < 0.05
FALSE
TRUE
Dog cancer type
Dog: B−cell lymphoma (N=109)
Dog: Glioma (N=52)
Dog: Hemangiosarcoma (N=44)
Dog: Mammary tumors (N=23)
Dog: Melanoma (N=65)
Dog: Osteosarcoma (N=95)
Dog: T−cell lymphoma (N=41)
$-\log_{10} p_{perm}$
entropy
B
Dog: B−cell lymphoma (N=109)
Dog: Hemangiosarcoma (N=44)
Dog: Osteosarcoma (N=95)
Dog: T−cell lymphoma (N=41)
2e−04
0.038
0.007
1e−06
0.043
0.014
6e−05
0.003
0.003
fraction of tumors
human cancer type classification
Breast cancer (N=1784)
Chronic lymphocytic leukemia (N=490)
Glioma (high grade) (N=494)
Prostate adenocarcinoma (N=885)
C
Dog hemangiosarcomas classified as breast cancer
SBS13
SBS2
TCGA_GLIOBLASTOMA_MUTATED
SBS17a
mutCount
SBS23
PID_MET_PATHWAY_count
SBS21
SBS38
DELPUECH_FOXO3_TARGETS_UP
SBS39
SBS18
SBS1
D190 D192 D196 D202 D204 D206 D214 D215 D216 D218 D220 D223 D224 D228
human
dog
D
Dog hemangiosarcomas classified as glioma (high grade)
TCGA_GLIOBLASTOMA_MUTATED
SBS1
TCGA_GLIOBLASTOMA_MUTATED_count
SBS24
KEGG_CYTOKINE_CYTOKINE...
SBS4
SBS28
SBS18
SBS13
SBS7a
SBS15
SBS17b
SBS2
SBS11
SBS29
mutCount
SBS39
SBS17a
D189 D193 D194 D197 D205 D207 D210 D211 D212 D213 D217 D219 D221 D225 D226 D227
dog ID
Shapley value
mean Shapley value

Figure 6. Supervised machine learning finds significant matches between human and dog cancer types using genomic features.

A. Dog cancers in red were classified as fewer human cancer types than expected by chance, suggesting enrichment for shared genomic features. **B.** A subset of dog cancer types were classified to specific human cancer types more than expected by chance, and this varied depending on the type of genomic features included in the model. **C.** The genomic features explaining the classification of dog hemangiosarcomas as either breast cancer or **D.** high grade glioma varies between tumor samples. The label for the KEGG Cytokine–Cytokine Receptor Interaction pathway is truncated.

References

1. Chae, Y.K., Pan, A.P., Davis, A.A., Patel, S.P., Carneiro, B.A., Kurzrock, R., and Giles, F.J. (2017). Path toward precision oncology: Review of targeted therapy studies and tools to aid in defining “actionability” of a molecular lesion and patient management support. Mol. Cancer Ther. *16*, 2645–2655.

2. Carter, P., Presta, L., Gorman, C.M., Ridgway, J.B., Henner, D., Wong, W.L., Rowland, A.M., Kotts, C., Carver, M.E., and Shepard, H.M. (1992). Humanization of an anti-p185HER2 antibody for human cancer therapy. Proc. Natl. Acad. Sci. U. S. A. *89*, 4285–4289.

3. Goldenberg, M.M. (1999). Trastuzumab, a recombinant DNA-derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer. Clin. Ther. *21*, 309–318.

4. Bulbul, A., and Husain, H. (2018). First-Line Treatment in EGFR Mutant Non-Small Cell Lung Cancer: Is There a Best Option? Front. Oncol. *8*, 94.

5. Ciardiello, F., and Tortora, G. (2001). A novel approach in the treatment of cancer: targeting the epidermal growth factor receptor. Clin. Cancer Res. *7*, 2958–2970.

6. Sosman, J.A., Kim, K.B., Schuchter, L., Gonzalez, R., Pavlick, A.C., Weber, J.S., McArthur, G.A., Hutson, T.E., Moschos, S.J., Flaherty, K.T., et al. (2012). Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. N. Engl. J. Med. *366*, 707–714.

7. Subbiah, V., Puzanov, I., Blay, J.-Y., Chau, I., Lockhart, A.C., Raje, N.S., Wolf, J., Baselga, J., Meric-Bernstam, F., Roszik, J., et al. (2020). Pan-Cancer Efficacy of Vemurafenib in BRAFV600-Mutant Non-Melanoma Cancers. Cancer Discov *10*, 657–663.

8. Subbiah, V., Cassier, P.A., Siena, S., Garralda, E., Paz-Ares, L., Garrido, P., Nadal, E., Vuky, J., Lopes, G., Kalemkerian, G.P., et al. (2022). Pan-cancer efficacy of pralsetinib in patients with RET fusion–positive solid tumors from the phase 1/2 ARROW trial. Nature Medicine *28*, 1640–1645.

9. Drilon, A., Laetsch, T.W., Kummar, S., DuBois, S.G., Lassen, U.N., Demetri, G.D., Nathenson, M., Doebele, R.C., Farago, A.F., Pappo, A.S., et al. (2018). Efficacy of Larotrectinib in TRK Fusion–Positive Cancers in Adults and Children. New England Journal of Medicine. https://doi.org/10.1056/NEJMoa1714448.

10. Takamizawa, S., Koyama, T., Sunami, K., Sudo, K., Hirata, M., Kubo, T., Tao, K., Cho, H., Narita, Y., Kato, K., et al. (2024). Identification of barriers to implementation of precision oncology in patients with rare cancers. Cancer Sci *115*, 2023–2035.

11. Goyal, L., DiToro, D., Facchinetti, F., Martin, E.E., Peng, P., Baiev, I., Iyer, R., Maurer, J., Reyes, S., Zhang, K., et al. (2025). A model for decoding resistance in precision oncology: acquired resistance to FGFR inhibitors in cholangiocarcinoma. Annals of oncology : official journal of the European Society for Medical Oncology *36*. https://doi.org/10.1016/j.annonc.2024.12.011.

12. Subbiah, V., Gouda, M.A., Iorgulescu, J.B., Dadu, R., Patel, K., Sherman, S., Cabanillas, M., Hu, M., Castellanos, L.E., Amini, B., et al. (2024). Adaptive Darwinian off-target resistance mechanisms to selective RET inhibition in RET driven cancer. npj Precision Oncology *8*, 1–9.

13. Kan, Z., Jaiswal, B.S., Stinson, J., Janakiraman, V., Bhatt, D., Stern, H.M., Yue, P., Haverty, P.M., Bourgon, R., Zheng, J., et al. (2010). Diverse somatic mutation patterns and pathway alterations in human cancers. Nature *466*, 869–873.

14. Russnes, H.G., Navin, N., Hicks, J., and Borresen-Dale, A.-L. (2011). Insight into the heterogeneity of

breast cancer through next-generation sequencing. J. Clin. Invest. *121*, 3810–3818.

15. Wouters, O.J., McKee, M., and Luyten, J. (2020). Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. JAMA *323*, 844–853.

16. Hay, M., Thomas, D.W., Craighead, J.L., Economides, C., and Rosenthal, J. (2014). Clinical development success rates for investigational drugs. Nat. Biotechnol. *32*, 40–51.

17. Armenia, J., Wankowicz, S.A.M., Liu, D., Gao, J., Kundra, R., Reznik, E., Chatila, W.K., Chakravarty, D., Han, G.C., Coleman, I., et al. (2018). The long tail of oncogenic drivers in prostate cancer. Nat. Genet. *50*, 645–651.

18. Mohsen, H., Gunasekharan, V., Qing, T., Seay, M., Surovtseva, Y., Negahban, S., Szallasi, Z., Pusztai, L., and Gerstein, M.B. (2021). Network propagation-based prioritization of long tail genes in 17 cancer types. Genome Biol. *22*, 287.

19. Baldridge, D., Wangler, M.F., Bowman, A.N., Yamamoto, S., Undiagnosed Diseases Network, Schedl, T., Pak, S.C., Postlethwait, J.H., Shin, J., Solnica-Krezel, L., et al. (2021). Model organisms contribute to diagnosis and discovery in the undiagnosed diseases network: current state and a future vision. Orphanet J. Rare Dis. *16*, 206.

20. LeBlanc, A.K., Mazcko, C.N., and Khanna, C. (2016). Defining the value of a comparative approach to cancer drug development. Clin. Cancer Res. *22*, 2133–2138.

21. Khanna, C., London, C., Vail, D., Mazcko, C., and Hirschfeld, S. (2009). Guiding the optimal translation of new cancer treatments from canine to human cancer patients. Clin. Cancer Res. *15*, 5671–5677.

22. LeBlanc, A.K., and Mazcko, C.N. (2020). Improving human cancer therapy through the evaluation of pet dogs. Nat. Rev. Cancer *20*, 727–742.

23. Gutmann, D.H., Boehm, J.S., Karlsson, E.K., Padron, E., Seshadri, M., Wallis, D., and Snyder, J.C. (2024). Precision preclinical modeling to advance cancer treatment. J. Natl. Cancer Inst., djae249.

24. Somarelli, J.A., Boddy, A.M., Gardner, H.L., DeWitt, S.B., Tuohy, J., Megquier, K., Sheth, M.U., Hsu, S.D., Thorne, J.L., London, C.A., et al. (2020). Improving cancer drug discovery by studying cancer across the tree of life. Mol. Biol. Evol. *37*, 11–17.

25. Schiffman, J.D., and Breen, M. (2015). Comparative oncology: what dogs and other species can teach us about humans with cancer. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20140231.

26. Gardner, H.L., Fenger, J.M., and London, C.A. (2016). Dogs as a Model for Cancer. Annu Rev Anim Biosci *4*, 199–222.

27. Morrill, K., Hekman, J., Li, X., McClure, J., Logan, B., Goodman, L., Gao, M., Dong, Y., Alonso, M., Carmichael, E., et al. (2022). Ancestry-inclusive dog genomics challenges popular breed stereotypes. Science *376*, eabk0639.

28. Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature *438*, 803–819.

29. 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.

30. Karlsson, E.K., Sigurdsson, S., Ivansson, E., Thomas, R., Elvers, I., Wright, J., Howald, C., Tonomura, N.,

Perloski, M., Swofford, R., et al. (2013). Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. Genome Biol. *14*, R132.

31. Evans, J.M., Parker, H.G., Rutteman, G.R., Plassais, J., Grinwis, G.C.M., Harris, A.C., Lana, S.E., and Ostrander, E.A. (2021). Multi-omics approach identifies germline regulatory variants associated with hematopoietic malignancies in retriever dog breeds. PLoS Genet. *17*, e1009543.

32. Benson, J., Neff, S., Tincher, E.M., and Spadafori, G. (2022). Oodles of Doodles: Popularity and health (Nationwide Insurance).

33. Proietto, M., Crippa, M., Damiani, C., Pasquale, V., Sacco, E., Vanoni, M., and Gilardi, M. (2023). Tumor heterogeneity: preclinical models, emerging technologies, and future applications. Front. Oncol. *13*, 1164535.

34. Kashyap, A., Rapsomaniki, M.A., Barros, V., Fomitcheva-Khartchenko, A., Martinelli, A.L., Rodriguez, A.F., Gabrani, M., Rosen-Zvi, M., and Kaigala, G. (2022). Quantification of tumor heterogeneity: from data acquisition to metric generation. Trends Biotechnol. *40*, 647–676.

35. Cahill, J.A., Smith, L.A., Gottipati, S., Torabi, T.S., and Graim, K. (2024). Bringing the Genomic Revolution to Comparative Oncology: Human and Dog Cancers. Annu Rev Biomed Data Sci *7*, 107–129.

36. Cagan, A., Baez-Ortega, A., Brzozowska, N., Abascal, F., Coorens, T.H.H., Sanders, M.A., Lawson, A.R.J., Harvey, L.M.R., Bhosle, S., Jones, D., et al. (2022). Somatic mutation rates scale with lifespan across mammals. Nature *604*, 517–524.

37. Backer, L.C., Grindem, C.B., Corbett, W.T., Cullins, L., and Hunter, J.L. (2001). Pet dogs as sentinels for environmental contamination. Sci. Total Environ. *274*, 161–169.

38. Wise, C.F., Hammel, S.C., Herkert, N., Ma, J., Motsinger-Reif, A., Stapleton, H.M., and Breen, M. (2020). Comparative exposure assessment using silicone passive samplers indicates that domestic dogs are sentinels to support human health research. Environ. Sci. Technol. *54*, 7409–7419.

39. Hegedus, C., Andronie, L., Uiuiu, P., Jurco, E., Lazar, E.A., and Popescu, S. (2023). Pets, genuine tools of environmental pollutant detection. Animals (Basel) *13*, 2923.

40. Wise, C.F., Herkert, N.J., Hoffman, K., Vaden, S., Breen, M., and Stapleton, H.M. (2025). Environmental exposures and canine bladder cancer: A case control study using silicone passive samplers. Environ. Sci. Technol. *59*, 1121–1132.

41. Day, C.-P., Merlino, G., and Van Dyke, T. (2015). Preclinical mouse cancer models: a maze of opportunities and challenges. Cell *163*, 39–53.

42. Onaciu, A., Munteanu, R., Munteanu, V.C., Gulei, D., Raduly, L., Feder, R.-I., Pirlog, R., Atanasov, A.G., Korban, S.S., Irimie, A., et al. (2020). Spontaneous and induced animal models for cancer research. Diagnostics (Basel) *10*, 660.

43. Yin, L., Wang, X.-J., Chen, D.-X., Liu, X.-N., and Wang, X.-J. (2020). Humanized mouse model: a review on preclinical applications for cancer immunotherapy. Am. J. Cancer Res. *10*, 4568–4584.

44. Paoloni, M., Davis, S., Lana, S., Withrow, S., Sangiorgi, L., Picci, P., Hewitt, S., Triche, T., Meltzer, P., and Khanna, C. (2009). Canine tumor cross-species genomics uncovers targets linked to osteosarcoma progression. BMC Genomics *10*, 625.

45. Wang, G., Wu, M., Maloneyhuss, M.A., Wojcik, J., Durham, A.C., Mason, N.J., and Roth, D.B. (2017).

Actionable mutations in canine hemangiosarcoma. PLoS One *12*, e0188667.

46. Hendricks, W.P.D., Zismann, V., Sivaprakasam, K., Legendre, C., Poorman, K., Tembe, W., Perdigones, N., Kiefer, J., Liang, W., DeLuca, V., et al. (2018). Somatic inactivating PTPRJ mutations and dysregulated pathways identified in canine malignant melanoma by integrated comparative genomic analysis. Preprint, https://doi.org/10.1371/journal.pgen.1007589 https://doi.org/10.1371/journal.pgen.1007589.

47. Kim, K.-K., Seung, B.-J., Kim, D., Park, H.-M., Lee, S., Song, D.-W., Lee, G., Cheong, J.-H., Nam, H., Sur, J.-H., et al. (2019). Whole-exome and whole-transcriptome sequencing of canine mammary gland tumors. Sci Data *6*, 147.

48. Wong, K., van der Weyden, L., Schott, C.R., Foote, A., Constantino-Casas, F., Smith, S., Dobson, J.M., Murchison, E.P., Wu, H., Yeh, I., et al. (2019). Cross-species genomic landscape comparison of human mucosal melanoma with canine oral and equine melanoma. Nat. Commun. *10*, 353.

49. Wang, G., Wu, M., Durham, A.C., Radaelli, E., Mason, N.J., Xu, X., and Roth, D.B. (2020). Molecular subtypes in canine hemangiosarcoma reveal similarities with human angiosarcoma. PLoS One *15*, e0229728.

50. Kim, T.-M., Yang, I.S., Seung, B.-J., Lee, S., Kim, D., Ha, Y.-J., Seo, M.-K., Kim, K.-K., Kim, H.S., Cheong, J.-H., et al. (2020). Cross-species oncogenic signatures of breast cancer in canine mammary tumors. Nat. Commun. *11*, 3616.

51. Wong, K., Ludwig, L., Krijgsman, O., Adams, D.J., Wood, G.A., and van der Weyden, L. (2021). Comparison of the oncogenomic landscape of canine and feline hemangiosarcoma shows novel parallels with human angiosarcoma. Dis. Model. Mech. *14*. https://doi.org/10.1242/dmm.049044.

52. Das, S., Idate, R., Regan, D.P., Fowles, J.S., Lana, S.E., Thamm, D.H., Gustafson, D.L., and Duval, D.L. (2021). Immune pathways and TP53 missense mutations are associated with longer survival in canine osteosarcoma. Commun Biol *4*, 1178.

53. Chu, S., Skidmore, Z.L., Kunisaki, J., Walker, J.R., Griffith, M., Griffith, O.L., and Bryan, J.N. (2021). Unraveling the chaotic genomic landscape of primary and metastatic canine appendicular osteosarcoma with current sequencing technologies and bioinformatic approaches. PLoS One *16*, e0246443.

54. Alsaihati, B.A., Ho, K.-L., Watson, J., Feng, Y., Wang, T., Dobbin, K.K., and Zhao, S. (2021). Canine tumor mutational burden is correlated with TP53 mutation across tumor types and breeds. Nat Commun *12*, 4670.

55. London, C.A., Hannah, A.L., Zadovoskaya, R., Chien, M.B., Kollias-Baker, C., Rosenberg, M., Downing, S., Post, G., Boucher, J., Shenoy, N., et al. (2003). Phase I dose-escalating study of SU11654, a small molecule receptor tyrosine kinase inhibitor, in dogs with spontaneous malignancies. Clin. Cancer Res. *9*, 2755–2768.

56. Demetri, G.D., van Oosterom, A.T., Garrett, C.R., Blackstein, M.E., Shah, M.H., Verweij, J., McArthur, G., Judson, I.R., Heinrich, M.C., Morgan, J.A., et al. (2006). Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial. Lancet *368*, 1329–1338.

57. London, C.A., Malpas, P.B., Wood-Follis, S.L., Boucher, J.F., Rusk, A.W., Rosenberg, M.P., Henry, C.J., Mitchener, K.L., Klein, M.K., Hintermeister, J.G., et al. (2009). Multi-center, placebo-controlled, double-blind, randomized study of oral toceranib phosphate (SU11654), a receptor tyrosine kinase inhibitor, for the treatment of dogs with recurrent (either local or distant) mast cell tumor following surgical excision. Clin. Cancer Res. *15*, 3856–3865.

58. Pryer, N.K., Lee, L.B., Zadovaskaya, R., Yu, X., Sukbuntherng, J., Cherrington, J.M., and London, C.A. (2003). Proof of target for SU11654: inhibition of KIT phosphorylation in canine mast cell tumors. Clin. Cancer Res. *9*, 5729–5734.

59. Gardner, H.L., Sivaprakasam, K., Briones, N., Zismann, V., Perdigones, N., Drenner, K., Facista, S., Richholt, R., Liang, W., Aldrich, J., et al. (2019). Canine osteosarcoma genome sequencing identifies recurrent mutations in DMD and the histone methyltransferase gene SETD2. Commun Biol *2*, 266.

60. Li, J., Duns, G., Westers, H., Sijmons, R., van den Berg, A., and Kok, K. (2016). SETD2: an epigenetic modifier with tumor suppressor functionality. Oncotarget *7*, 50719–50734.

61. Jones, L., Naidoo, M., Machado, L.R., and Anthony, K. (2021). The Duchenne muscular dystrophy gene and cancer. Cell Oncol (Dordr) *44*, 19–32.

62. Rossman, P., Zabka, T.S., Ruple, A., Tuerck, D., Ramos-Vara, J.A., Liu, L., Mohallem, R., Merchant, M., Franco, J., Fulkerson, C.M., et al. (2021). Phase I/II trial of vemurafenib in dogs with naturally occurring, BRAF-mutated urothelial carcinoma. Mol. Cancer Ther. *20*, 2177–2188.

63. Kopetz, S., Desai, J., Chan, E., Hecht, J.R., O'Dwyer, P.J., Maru, D., Morris, V., Janku, F., Dasari, A., Chung, W., et al. (2015). Phase II pilot study of vemurafenib in patients with metastatic BRAF-mutated colorectal cancer. J. Clin. Oncol. *33*, 4032–4038.

64. Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. *45*, 1113–1120.

65. Zhang, J., Baran, J., Cros, A., Guberman, J.M., Haider, S., Hsu, J., Liang, Y., Rivkin, E., Wang, J., Whitty, B., et al. (2011). International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. Database *2011*. https://doi.org/10.1093/database/bar026.

66. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, l1.

67. Elvers, I., Turner-Maier, J., Swofford, R., Koltookian, M., Johnson, J., Stewart, C., Zhang, C.-Z., Schumacher, S.E., Beroukhim, R., Rosenberg, M., et al. (2015). Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. Genome Res. *25*, 1634–1645.

68. Sakthikumar, S., Elvers, I., Kim, J., Arendt, M.L., Thomas, R., Turner-Maier, J., Swofford, R., Johnson, J., Schumacher, S.E., Alföldi, J., et al. (2018). SETD2 Is Recurrently Mutated in Whole-Exome Sequenced Canine Osteosarcoma. Cancer Res. *78*, 3421–3431.

69. Megquier, K., Turner-Maier, J., Swofford, R., Kim, J.-H., Sarver, A.L., Wang, C., Sakthikumar, S., Johnson, J., Koltookian, M., Lewellen, M., et al. (2019). Comparative genomics reveals shared mutational landscape in canine hemangiosarcoma and human angiosarcoma. Mol. Cancer Res. https://doi.org/10.1158/1541-7786.MCR-19-0221.

70. Amin, S.B., Anderson, K.J., Boudreau, C.E., Martinez-Ledesma, E., Kocakavuk, E., Johnson, K.C., Barthel, F.P., Varn, F.S., Kassab, C., Ling, X., et al. (2020). Comparative Molecular Life History of Spontaneous Canine and Human Gliomas. Cancer Cell *37*, 243–257.e7.

71. White, M.E., Hayward, J.J., Hertafeld, S.R., Castelhano, M.G., Leung, W., Dave, S.S., Bhinder, B.H., Elemento, O.L., Boyko, A.R., Richards, K.L., et al. (2020). Consensus-based somatic variant-calling

method correlates FBXW7 mutations with poor prognosis in canine B-cell lymphoma. bioRxiv, 2020.08.16.250100. https://doi.org/10.1101/2020.08.16.250100.

72. Prouteau, A., Mottier, S., Primot, A., Cadieu, E., Bachelot, L., Botherel, N., Cabillic, F., Houel, A., Cornevin, L., Kergal, C., et al. (2022). Canine Oral Melanoma Genomic and Transcriptomic Study Defines Two Molecular Subgroups with Different Therapeutical Targets. Cancers *14*. https://doi.org/10.3390/cancers14020276.

73. Arendt, M.L., Sakthikumar, S., Melin, M., Elvers, I., Rivera, P., Larsen, M., Sällström, S., Lingaas, F., Rönnberg, H., and Lindblad-Toh, K. (2022). PIK3CA is recurrently mutated in canine mammary tumors, similarly to in human mammary neoplasia. https://doi.org/10.21203/rs.3.rs-1801127/v1.

74. Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol *2017*. https://doi.org/10.1200/PO.17.00011.

75. Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat. Rev. Cancer *18*, 696–705.

76. Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371–385.e18.

77. 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.

78. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly *6*, 80–92.

79. Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer genes. Nature *499*, 214.

80. Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell *173*, 291.

81. Martincorena, I., Raine, K.M., Gerstung, M., Dawson, K.J., Haase, K., Van Loo, P., Davies, H., Stratton, M.R., and Campbell, P.J. (2018). Universal Patterns of Selection in Cancer and Somatic Tissues. Cell *173*, 1823.

82. Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B., Tubio, J.M., et al. (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science *348*, 880–886.

83. Wu, Y., Poulos, R.C., and Reddel, R.R. (2020). Role of POT1 in Human Cancer. Cancers *12*, 2739.

84. Cuevas, I.C., Sahoo, S.S., Kumar, A., Zhang, H., Westcott, J., Aguilar, M., Cortez, J.D., Sullivan, S.A., Xing, C., Hayes, D.N., et al. (2019). Fbxw7 is a driver of uterine carcinosarcoma by promoting epithelial-mesenchymal transition. Proceedings of the National Academy of Sciences *116*, 25880–25890.

85. Gadek, M., Sherr, E.H., and Floor, S.N. (2023). The variant landscape and function of DDX3X in cancer

and neurodevelopmental disorders. Trends Mol Med *29*, 726–739.

86. GitHub - griffithlab/docm: Retired Project: Rails frontend to The Genome Institute's database of curated mutations (DoCM) GitHub. http://www.docm.info/.

87. Hess, J.M., Bernards, A., Kim, J., Miller, M., Taylor-Weiner, A., Haradhvala, N.J., Lawrence, M.S., and Getz, G. (2019). Passenger hotspot mutations in cancer. Cancer Cell *36*, 288–301.e14.

88. Rodrigues, L., Watson, J., Feng, Y., Lewis, B., Harvey, G., Post, G., Megquier, K., White, M.E., Lambert, L., Miller, A., et al. (2023). Shared hotspot mutations in oncogenes position dogs as an unparalleled comparative model for precision therapeutics. Sci. Rep. *13*, 10935.

89. Ellrott, K., Bailey, M.H., Saksena, G., Covington, K.R., Kandoth, C., Stewart, C., Hess, J., Ma, S., Chiotti, K.E., McLellan, M., et al. (2018). Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. Cell Syst. *6*, 271–281.e7.

90. Priestley, P., Baber, J., Lolkema, M.P., Steeghs, N., de Bruijn, E., Shale, C., Duyvesteyn, K., Haidari, S., van Hoeck, A., Onstenk, W., et al. (2019). Pan-cancer whole-genome analyses of metastatic solid tumours. Nature *575*, 210–216.

91. Huang, D., Yu, Z., Lu, H., Jiang, P., Qian, X., Han, Y., and Qian, P. (2024). Adhesion GPCR ADGRE2 maintains proteostasis to promote progression in acute myeloid leukemia. Cancer Res. *84*, 2090–2108.

92. Degasperi, A., Amarante, T.D., Czarnecki, J., Shooter, S., Zou, X., Glodzik, D., Morganella, S., Nanda, A.S., Badja, C., Koh, G., et al. (2020). A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. Nat. Cancer *1*, 249–263.

93. Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al. (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993.

94. Chan, K., Roberts, S.A., Klimczak, L.J., Sterling, J.F., Saini, N., Malc, E.P., Kim, J., Kwiatkowski, D.J., Fargo, D.C., Mieczkowski, P.A., et al. (2015). An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat. Genet. *47*, 1067–1072.

95. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans (1992). Studies of Cancer in Animals (International Agency for Research on Cancer).

96. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94–101.

97. Painter, C.A., Jain, E., Tomson, B.N., Dunphy, M., Stoddard, R.E., Thomas, B.S., Damon, A.L., Shah, S., Kim, D., Gómez Tejeda Zañudo, J., et al. (2020). The Angiosarcoma Project: enabling genomic and clinical discoveries in a rare cancer through patient-partnered research. Nat. Med. *26*, 181–187.

98. Ike, E., Mai, J.Z., Sargen, M.R., Schonfeld, S.J., and Cahoon, E.K. (2024). Ambient UV radiation is associated with cutaneous angiosarcoma incidence in the United States, 1992 to 2020. J. Am. Acad. Dermatol. *91*, 102–104.

99. Gniadecki, R., O'Keefe, S., Hennessey, D., and Iyer, A. (2023). Is Cutaneous T-Cell Lymphoma Caused by Ultraviolet Radiation? A Comparison of UV Mutational Signatures in Malignant Melanoma and Mycosis Fungoides. Cells *12*, 1616.

100. Reif, J.S., Bruns, C., and Lower, K.S. (1998). Cancer of the nasal cavity and paranasal sinuses and exposure to environmental tobacco smoke in pet dogs. Am. J. Epidemiol. *147*, 488–492.

101. Chawanthayatham, S., Valentine, C.C., 3rd, Fedeles, B.I., Fox, E.J., Loeb, L.A., Levine, S.S., Slocum, S.L., Wogan, G.N., Croy, R.G., and Essigmann, J.M. (2017). Mutational spectra of aflatoxin B1 in vivo establish biomarkers of exposure for human hepatocellular carcinoma. Proc. Natl. Acad. Sci. U. S. A. *114*, E3101–E3109.

102. Boermans, H.J., and Leung, M.C.K. (2007). Mycotoxins and the pet food industry: toxicological evidence and risk assessment. Int. J. Food Microbiol. *119*, 95–102.

103. Poon, S.L., Pang, S.-T., McPherson, J.R., Yu, W., Huang, K.K., Guan, P., Weng, W.-H., Siew, E.Y., Liu, Y., Heng, H.L., et al. (2013). Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. Sci. Transl. Med. *5*, 197ra101.

104. Lu, Z.-N., Luo, Q., Zhao, L.-N., Shi, Y., Wang, N., Wang, L., and Han, Z.-G. (2020). The mutational features of aristolochic acid-induced mouse and human liver cancers. Hepatology *71*, 929–942.

105. Inman, G.J., Wang, J., Nagano, A., Alexandrov, L.B., Purdie, K.J., Taylor, R.G., Sherwood, V., Thomson, J., Hogan, S., Spender, L.C., et al. (2018). The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature. Nat. Commun. *9*, 3667.

106. Viviano, K.R. (2022). Glucocorticoids, cyclosporine, azathioprine, chlorambucil, and mycophenolate in dogs and cats: Clinical uses, pharmacology, and side effects. Vet. Clin. North Am. Small Anim. Pract. *52*, 797–817.

107. Moon, I., LoPiccolo, J., Baca, S.C., Sholl, L.M., Kehl, K.L., Hassett, M.J., Liu, D., Schrag, D., and Gusev, A. (2023). Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary. Nat. Med. *29*, 2057–2067.

108. Lundberg, S., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. arXiv [cs.AI].

109. Shapley, L.S. (1953). A Value for n-Person Games.” Contribution to the Theory of Games 2.

110. Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD ’16. (Association for Computing Machinery), pp. 785–794.

111. Vanwinckelen, G., and Blockeel, H. (2012). On estimating model accuracy with repeated cross-validation. 39–44.

112. Chen, Y., Sun, J., Huang, L.-C., Xu, H., and Zhao, Z. (2015). Classification of Cancer Primary Sites Using Machine Learning and Somatic Mutations. Biomed Res. Int. *2015*, 491502.

113. Marquard, A.M., Birkbak, N.J., Thomas, C.E., Favero, F., Krzystanek, M., Lefebvre, C., Ferté, C., Jamal-Hanjani, M., Wilson, G.A., Shafi, S., et al. (2015). TumorTracer: a method to identify the tissue of origin from the somatic mutations of a tumor specimen. BMC Med. Genomics *8*, 58.

114. Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature *455*, 1061–1068.

115. Vasan, N., Toska, E., and Scaltriti, M. (2019). Overview of the relevance of PI3K pathway in HR-positive breast cancer. Ann. Oncol. *30*, x3–x11.

116. Langhans, J., Schneele, L., Trenkler, N., von Bandemer, H., Nonnenmacher, L., Karpel-Massler, G.,

Siegelin, M.D., Zhou, S., Halatsch, M.-E., Debatin, K.-M., et al. (2017). The effects of PI3K-mediated signalling on glioblastoma cell behaviour. Oncogenesis *6*. https://doi.org/10.1038/s41389-017-0004-8.

117. Roelofs, P.A., Martens, J.W.M., Harris, R.S., and Span, P.N. (2023). Clinical implications of APOBEC3-mediated Mutagenesis in breast cancer. Clin. Cancer Res. *29*, 1658–1669.

118. Olson, M.E., Harris, R.S., and Harki, D.A. (2018). APOBEC enzymes as targets for virus and cancer therapy. Cell Chem. Biol. *25*, 36–49.

119. Nik-Zainal, S., and Morganella, S. (2017). Mutational signatures in breast cancer: The problem at the DNA level. Clin. Cancer Res. *23*, 2617–2629.

120. Hodel, K.P., Sun, M.J.S., Ungerleider, N., Park, V.S., Williams, L.G., Bauer, D.L., Immethun, V.E., Wang, J., Suo, Z., Lu, H., et al. (2020). POLE mutation spectra are shaped by the mutant allele identity, its abundance, and mismatch repair status. Mol. Cell *78*, 1166–1177.e6.

121. Boot, A., Huang, M.N., Ng, A.W.T., Ho, S.-C., Lim, J.Q., Kawakami, Y., Chayama, K., Teh, B.T., Nakagawa, H., and Rozen, S.G. (2018). In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. Genome Res. *28*, 654–665.

122. Secrier, M., Li, X., de Silva, N., Eldridge, M.D., Contino, G., Bornschein, J., MacRae, S., Grehan, N., O'Donovan, M., Miremadi, A., et al. (2016). Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. Nat. Genet. *48*, 1131–1141.

123. Manaka, Y., Kusumoto-Matsuo, R., Matsuno, Y., Asai, H., and Yoshioka, K.-I. (2024). Single base substitution signatures 17a, 17b, and 40 are induced by γ-ray irradiation in association with increased reactive oxidative species. Heliyon *10*, e28044.

124. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

125. Kumar, S., Suleski, M., Craig, J.M., Kasprowicz, A.E., Sanderford, M., Li, M., Stecher, G., and Hedges, S.B. (2022). TimeTree 5: An expanded resource for species divergence times. Mol. Biol. Evol. *39*. https://doi.org/10.1093/molbev/msac174.

126. Foley, N.M., Mason, V.C., Harris, A.J., Bredemeyer, K.R., Damas, J., Lewin, H.A., Eizirik, E., Gatesy, J., Karlsson, E.K., Lindblad-Toh, K., et al. (2023). A genomic timescale for placental mammal evolution. Science *380*, eabl8189.

127. Nischwitz, E., Schoonenberg, V.A.C., Fradera-Sola, A., Dejung, M., Vydzhak, O., Levin, M., Luke, B., Butter, F., and Scheibe, M. (2023). DNA damage repair proteins across the Tree of Life. iScience *26*, 106778.

128. Tollis, M., Schiffman, J.D., and Boddy, A.M. (2017). Evolution of cancer suppression as revealed by mammalian comparative genomics. Curr. Opin. Genet. Dev. *42*, 40–47.

129. Harada, G., Yang, S.-R., Cocco, E., and Drilon, A. (2023). Rare molecular subtypes of lung cancer. Nat. Rev. Clin. Oncol. *20*, 229–249.

130. McGranahan, N., and Swanton, C. (2017). Clonal heterogeneity and tumor evolution: Past, present, and the future. Cell *168*, 613–628.

131. van der Heiden, A.D., Pensch, R., Agger, S., Gardner, H.L., Hendricks, W., Zismann, V., Wong, S.,

Briones, N., Turner, B., Forsberg-Nilsson, K., et al. (2025). Characterization of the genomic landscape of canine diffuse large B-cell lymphoma reveals recurrent H3K27M mutations linked to progression-free survival. Sci. Rep. *15*, 4724.

132. Mee, M.W., Faulkner, S., Wood, G.A., Woods, J.P., Bienzle, D., and Coomber, B.L. (2024). Longitudinal study of transcriptomic changes occurring over six weeks of CHOP treatment in canine lymphoma identifies prognostic subtypes. Vet. Sci. *11*, 540.

133. Megquier, K., Husted, C., Rhoades, J., White, M.E., Genereux, D.P., Chen, F.L., Xiong, K., Kwon, E., Swofford, R., Painter, C., et al. (2024). Impact of preanalytical factors on liquid biopsy in the canine cancer model. Genomics.

134. Lipinski, K.A., Barber, L.J., Davies, M.N., Ashenden, M., Sottoriva, A., and Gerlinger, M. (2016). Cancer evolution and the limits of predictability in precision cancer medicine. Trends Cancer *2*, 49–63.

135. Benton, A., Liu, B., Gartenhaus, L.E., and Hanna, J.A. (2025). Genomic landscape and preclinical models of angiosarcoma. Mol Oncol *19*, 965–983.

136. Lee, D., Kozurek, E.C., Abdullah, M., Wong, E.J., Li, R., Liu, Z.S., Nguyen, H.D., Dickerson, E.B., and Kim, J.H. (2025). PIK3CA mutation fortifies molecular determinants for immune signaling in vascular cancers. Cancer Gene Ther *32*, 254–267.

137. Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

138. Cancer Genome Atlas Network (2015). Genomic classification of cutaneous melanoma. Cell *161*, 1681–1696.

139. Hoadley, K.A., Yau, C., Wolf, D.M., Cherniack, A.D., Tamborero, D., Ng, S., Leiserson, M.D.M., Niu, B., McLellan, M.D., Uzunangelov, V., et al. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell *158*, 929–944.

140. Wei, Y., Deng, Y., Sun, C., Lin, M., Jiang, H., and Peng, Y. (2024). Deep learning with noisy labels in medical prediction problems: a scoping review. J. Am. Med. Inform. Assoc. *31*, 1596–1607.

141. Ellrott, K., Wong, C.K., Yau, C., Castro, M.A.A., Lee, J.A., Karlberg, B.J., Grewal, J.K., Lagani, V., Tercan, B., Friedl, V., et al. (2025). Classification of non-TCGA cancer samples to TCGA molecular subtypes using compact feature sets. Cancer Cell *43*, 195–212.e11.

142. Tarca, A.L., Carey, V.J., Chen, X.-W., Romero, R., and Drăghici, S. (2007). Machine learning and its applications to biology. PLoS Comput. Biol. *3*, e116.

143. Thomas, R., Wang, H.J., Tsai, P.-C., Langford, C.F., Fosmire, S.P., Jubala, C.M., Getzy, D.M., Cutter, G.R., Modiano, J.F., and Breen, M. (2009). Influence of genetic background on tumor karyotypes: evidence for breed-associated cytogenetic aberrations in canine appendicular osteosarcoma. Chromosome Res. *17*, 365–377.

144. Angstadt, A.Y., Thayanithy, V., Subramanian, S., Modiano, J.F., and Breen, M. (2012). A genome-wide approach to comparative oncology: high-resolution oligonucleotide aCGH of canine and human osteosarcoma pinpoints shared microaberrations. Cancer Genet. *205*, 572–587.

145. Shao, Y.W., Wood, G.A., Lu, J., Tang, Q.-L., Liu, J., Molyneux, S., Chen, Y., Fang, H., Adissu, H., McKee, T., et al. (2019). Cross-species genomics identifies DLG2 as a tumor suppressor in osteosarcoma. Oncogene *38*, 291–298.

146. Angstadt, A.Y., Motsinger-Reif, A., Thomas, R., Kisseberth, W.C., Guillermo Couto, C., Duval, D.L., Nielsen, D.M., Modiano, J.F., and Breen, M. (2011). Characterization of canine osteosarcoma by array comparative genomic hybridization and RT-qPCR: signatures of genomic imbalance in canine osteosarcoma parallel the human counterpart. Genes Chromosomes Cancer *50*, 859–874.

147. Chen, X., Bahrami, A., Pappo, A., Easton, J., Dalton, J., Hedlund, E., Ellison, D., Shurtleff, S., Wu, G., Wei, L., et al. (2014). Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. Cell Rep. *7*, 104–112.

148. Squire, J.A., Pei, J., Marrano, P., Beheshti, B., Bayani, J., Lim, G., Moldovan, L., and Zielenska, M. (2003). High-resolution mapping of amplifications and deletions in pediatric osteosarcoma by use of CGH analysis of cDNA microarrays. Genes Chromosomes Cancer *38*, 215–225.

149. Perry, J.A., Kiezun, A., Tonzi, P., Van Allen, E.M., Carter, S.L., Baca, S.C., Cowley, G.S., Bhatt, A.S., Rheinbay, E., Pedamallu, C.S., et al. (2014). Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. Proc. Natl. Acad. Sci. U. S. A. *111*, E5564–E5573.

150. Jiang, Y., Wang, J., Sun, M., Zuo, D., Wang, H., Shen, J., Jiang, W., Mu, H., Ma, X., Yin, F., et al. (2022). Multi-omics analysis identifies osteosarcoma subtypes with distinct prognosis indicating stratified treatment. Nat. Commun. *13*, 7207.

151. Southekal, S., Shakyawar, S.K., Bajpai, P., Elkholy, A., Manne, U., Mishra, N.K., and Guda, C. (2023). Molecular subtyping and survival analysis of osteosarcoma reveals prognostic biomarkers and key canonical pathways. Cancers (Basel) *15*, 2134.

152. Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K., et al. (2013). TERT promoter mutations in familial and sporadic melanoma. Science *339*, 959–961.

153. Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G.V., Chin, L., and Garraway, L.A. (2013). Highly recurrent TERT promoter mutations in human melanoma. Science *339*, 957–959.

154. Rheinbay, E., Nielsen, M.M., Abascal, F., Wala, J.A., Shapira, O., Tiao, G., Hornshøj, H., Hess, J.M., Juul, R.I., Lin, Z., et al. (2020). Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature *578*, 102–111.

155. Sullivan, P.F., Meadows, J.R.S., Gazal, S., Phan, B.N., Li, X., Genereux, D.P., Dong, M.X., Bianchi, M., Andrews, G., Sakthikumar, S., et al. (2023). Leveraging base-pair mammalian constraint to understand genetic variation and human disease. Science *380*, eabn2937.

156. Christmas, M.J., Kaplow, I.M., Genereux, D.P., Dong, M.X., Hughes, G.M., Li, X., Sullivan, P.F., Hindle, A.G., Andrews, G., Armstrong, J.C., et al. (2023). Evolutionary constraint and innovation across hundreds of placental mammals. Science *380*, eabn3943.

157. Roy, A., Sakthikumar, S., Kozyrev, S.V., Nordin, J., Pensch, R., Mäkeläinen, S., Pettersson, M., Zoonomia Consortium, Karlsson, E.K., Lindblad-Toh, K., et al. (2023). Using evolutionary constraint to define novel candidate driver genes in medulloblastoma. Proc. Natl. Acad. Sci. U. S. A. *120*, e2300984120.

158. Roberts, R.D., Lizardo, M.M., Reed, D.R., Hingorani, P., Glover, J., Allen-Rhoades, W., Fan, T., Khanna, C., Sweet-Cordero, E.A., Cash, T., et al. (2019). Provocative questions in osteosarcoma basic and translational biology: A report from the Children's Oncology Group. Cancer *125*, 3514–3525.

159. Baioni, E., Scanziani, E., Vincenti, M.C., Leschiera, M., Bozzetta, E., Pezzolato, M., Desiato, R., Bertolini,

S., Maurella, C., and Ru, G. (2017). Estimating canine cancer incidence: findings from a population-based tumour registry in northwestern Italy. BMC Vet. Res. *13*, 203.

160. Borgatti, A., Fahrenkrug, A.M., and Modiano, J.F. (2017). Studying Cancer in Dogs as a Path Towards a World Where We No Longer Fear Cancer – A Contemporary Update on Canine Hemangiosarcoma (Animal Care and Cancer Research of the University of Minnesota).

161. Frimberger, A.E., Chan, C.M., and Moore, A.S. (2016). Canine osteosarcoma treated by post-amputation sequential accelerated doxorubicin and carboplatin chemotherapy: 38 cases. J. Am. Anim. Hosp. Assoc. *52*, 149–156.

162. Piras, L.A., Riccardo, F., Iussich, S., Maniscalco, L., Gattino, F., Martano, M., Morello, E., Lorda Mayayo, S., Rolih, V., Garavaglia, F., et al. (2017). Prolongation of survival of dogs with oral malignant melanoma treated by en bloc surgical resection and adjuvant CSPG4-antigen electrovaccination. Vet. Comp. Oncol. *15*, 996–1013.

163. Merickel, J.L., Pluhar, G.E., Rendahl, A., and O'Sullivan, M.G. (2021). Prognostic histopathologic features of canine glial tumors. Vet. Pathol. *58*, 945–951.

164. Fan, T.M. (2023). Lymphoma in Dogs. Merck Veterinary Manual. https://www.merckvetmanual.com/circulatory-system/lymphoma-in-dogs/lymphoma-in-dogs.

165. London, C.A., Gardner, H., Zhao, S., Knapp, D.W., Utturkar, S.M., Duval, D.L., Chambers, M.R., Ostrander, E., Trent, J.M., and Kuffel, G. (2023). Leading the pack: Best practices in comparative canine cancer genomics to inform human oncology. Vet. Comp. Oncol. *21*, 565–577.

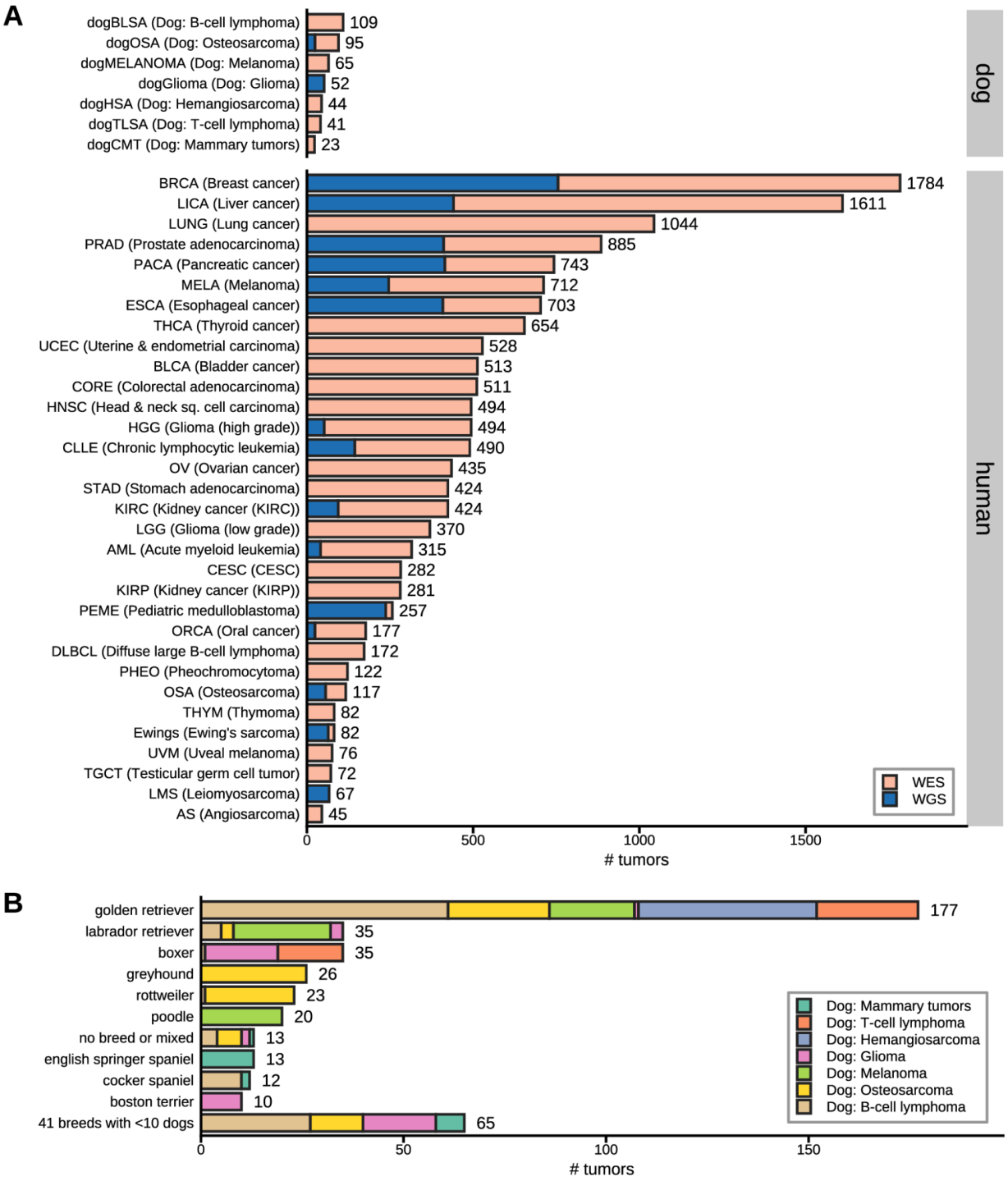


Figure S1. Cancer types and breeds represented in the comparative pan-cancer data set. **(A)** To construct our comparative data set, we compiled existing, publicly available somatic mutation data from whole genome (blue) and whole exome (orange) sequencing from 429 dog cancers and 14,966 human cancers of multiple histologies. **(B)** Dog tumors are from multiple dog breeds, most of which contributed tumors of multiple cancer types.

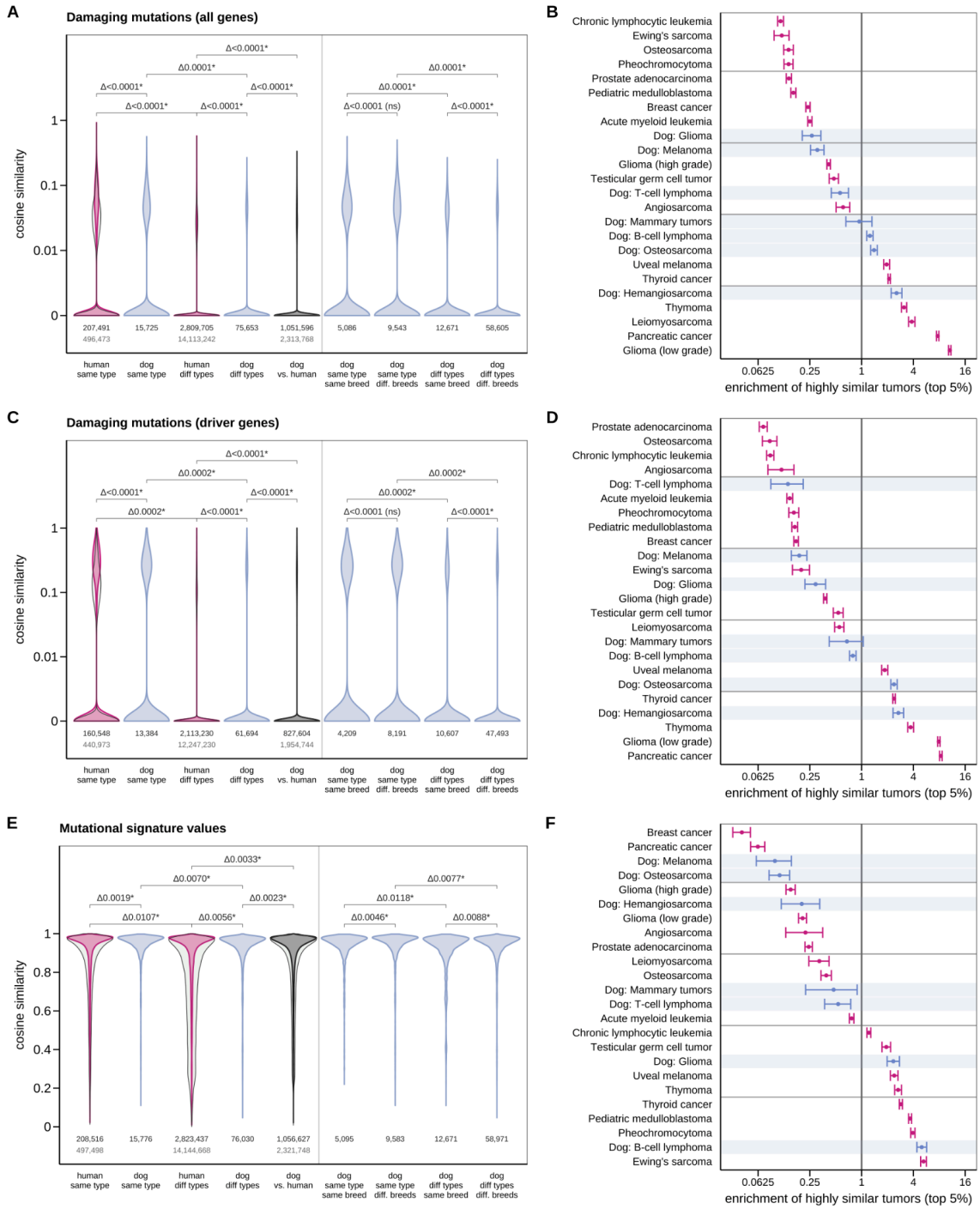


Figure S2. Dog and human tumors are, on average, no more different than dog/dog or dog/human tumors pairs of different cancer types.

Cosine similarity scores pairs for pairs of tumors from human (pink) dog (blue), and for dog/human comparisons (black) based on (A) mutations in all genes, with (B) dog cancer types interspersed with human cancer types. The same is seen for (C,D) mutations in driver genes; and (E,F) mutational signatures. Black lines in violin plots show the distributions when the high mutation rate cancer types are not removed.

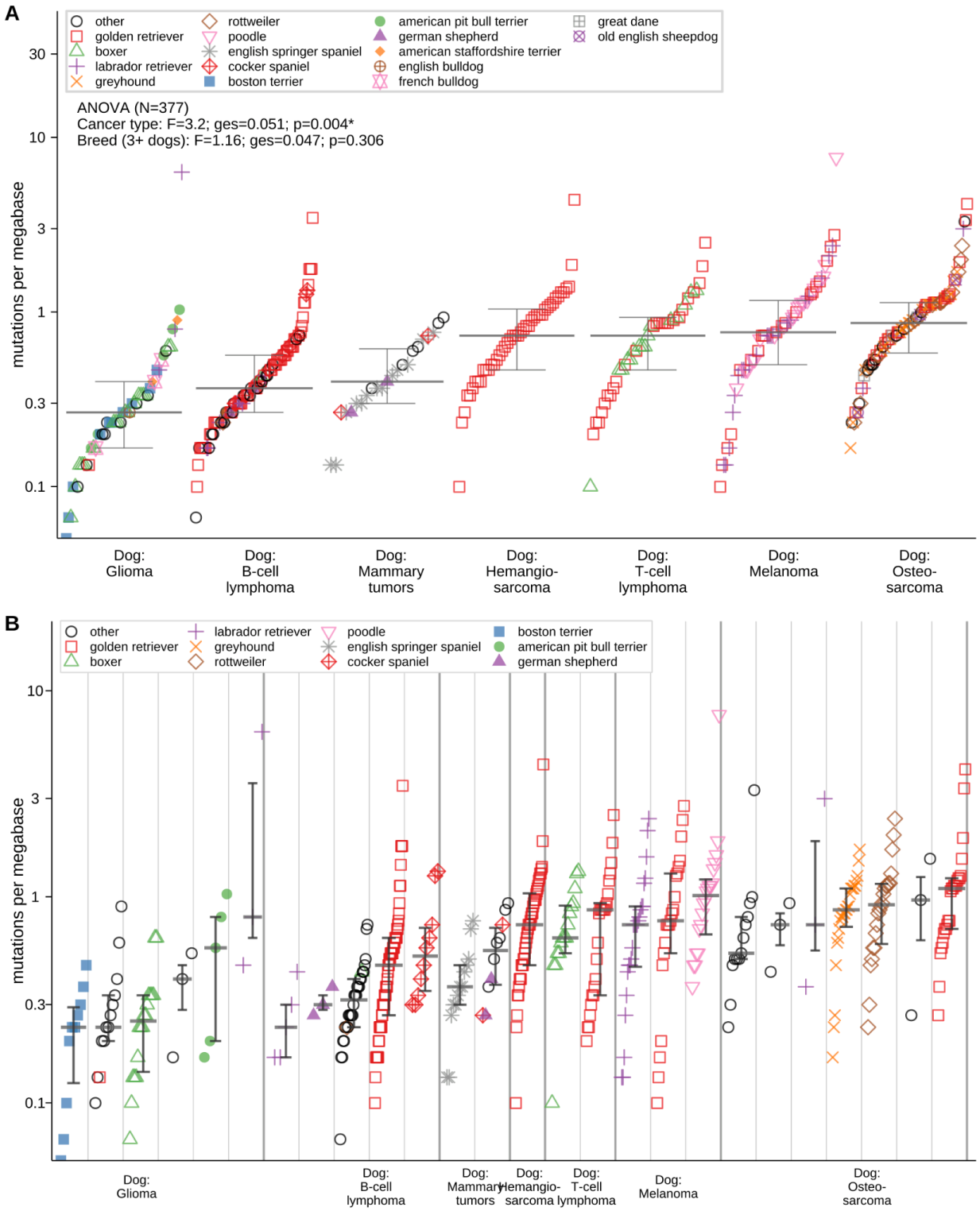


Figure S3. Mutational burden in canine tumors varies between cancer types but not between breeds,

Distribution of mutational burden in mutations per megabase **(A)** across all canine cancer types, and **(B)** separated by breed and cancer type.

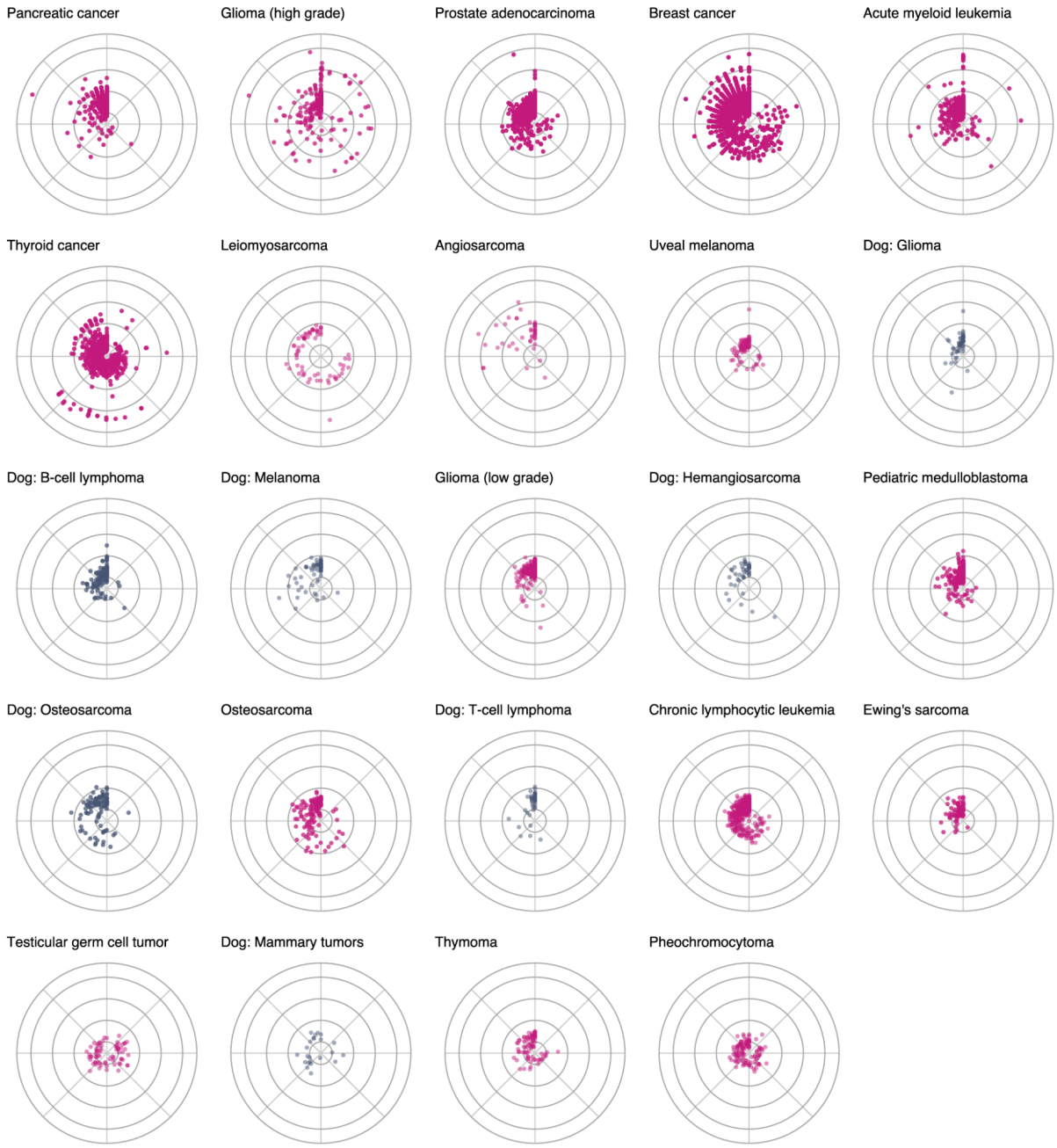


Figure S4. Relative abundance of mutation signatures is heterogeneous within individual cancer types of dog and low-mutation human cancers.
Radial plots of each tumor from dog (pink) and human (gray) positioned an angle determined by the relative abundance of 43 mutational signatures detected in at least one human and one dog. Distance from center is log10 of mutation count. Cancer types are sorted by their maximum mutation count.

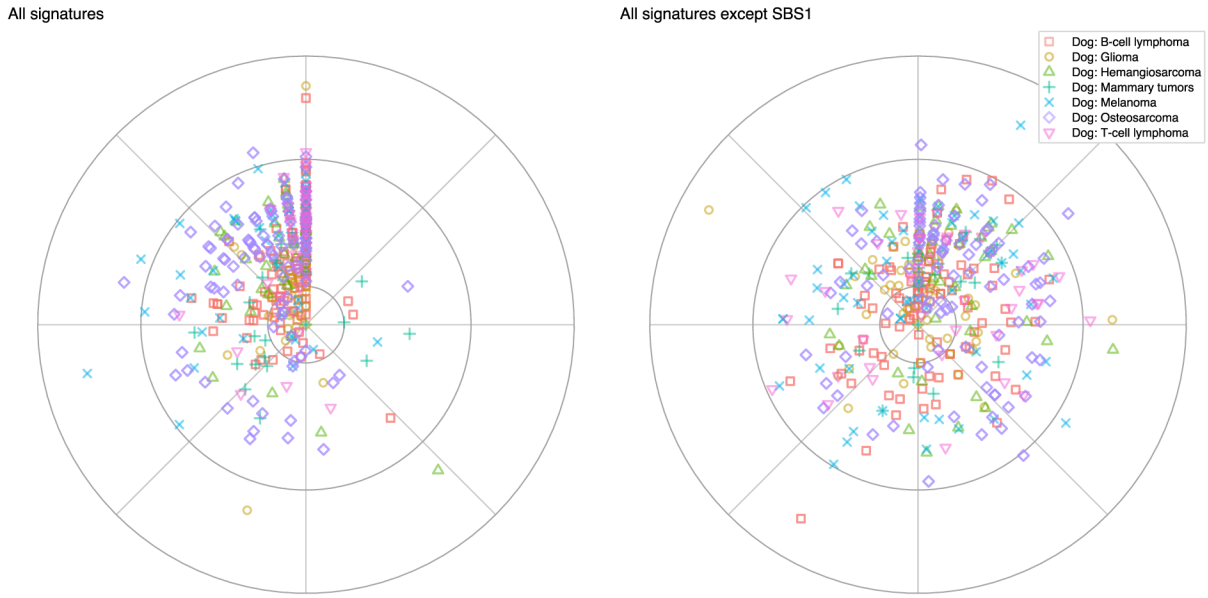


Figure S5. Relative abundance of mutation signatures is heterogeneous within individual dog cancer types.
Radial spectrum plots place each tumor into one of 43 (left) or 42 (right, excluding SBS1) sectors, positioned based on its most abundant mutational signature. The angle within each sector reflects the relative contributions of the remaining signatures, and the distance from the center represents the log10 of the mutation count.

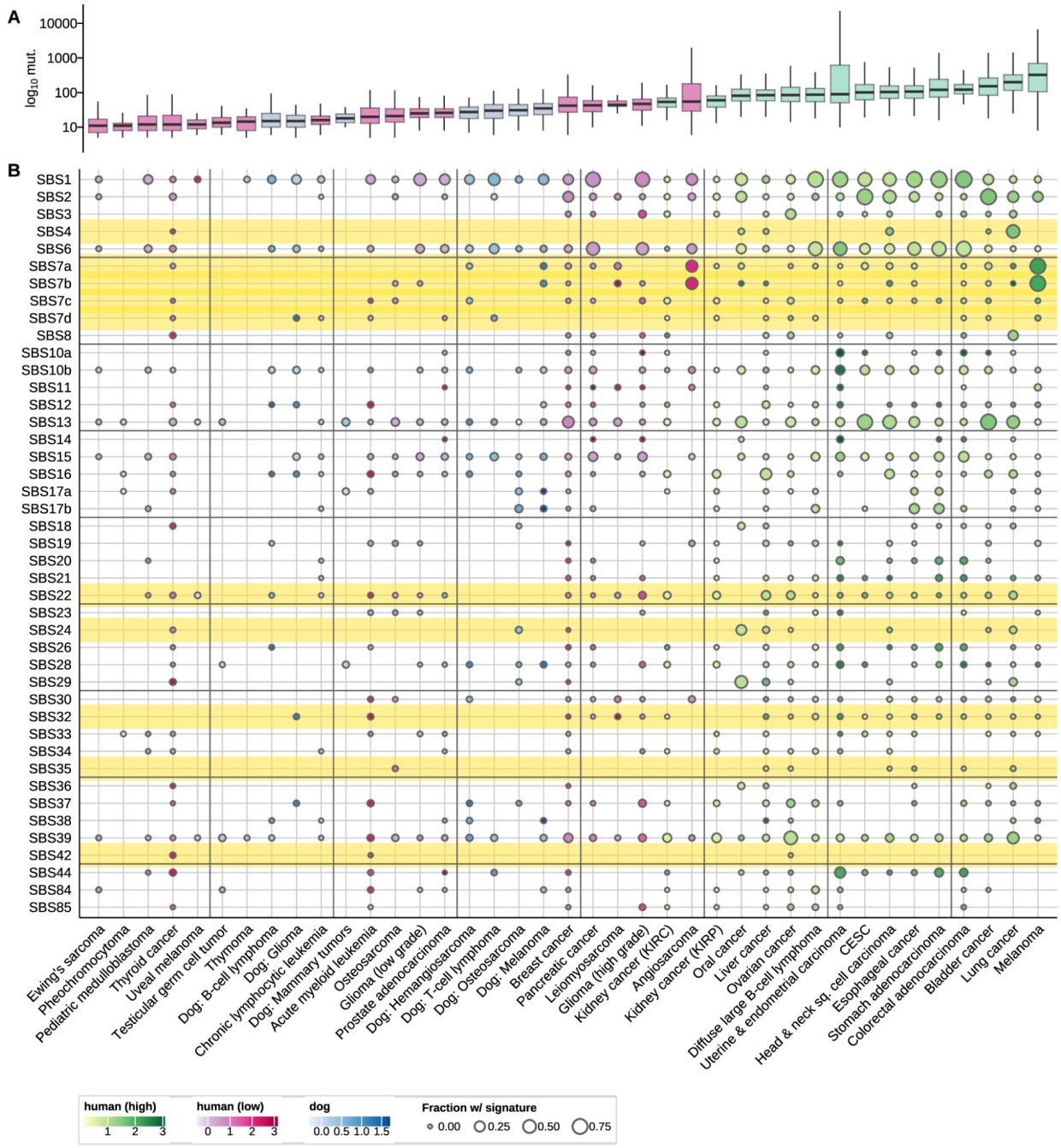


Figure S6. Full version of figure 4C.

Signatures detect in dog and human (both low and high mutation rate) cancer types. Boxplot shows the mutation count per tumor. Circle plot shows how frequent each signature was detected at a value >0.05 in each tumor type. Yellow shading indicates signatures putatitively connected to environmental exposures.

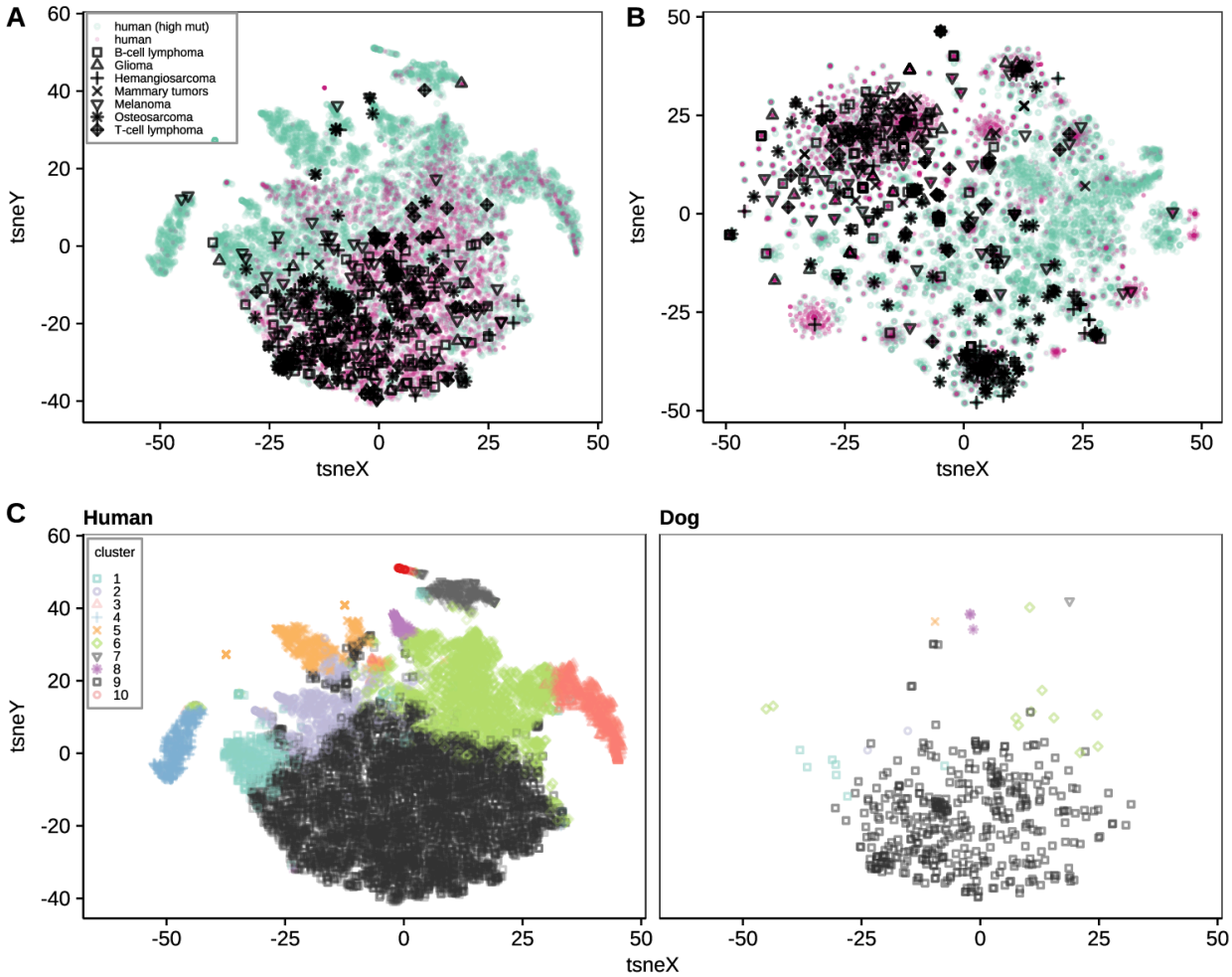


Figure S7. Unsupervised clustering of all dog and human cancer types groups dog tumors with human tumors from low-mutation cancer types.

A. t-SNE visualization of all dog and human tumors clustered using k-means based on mutational signatures, after dimensionality reduction to principal components explaining 70% of the variance. **B.** Clustering of all dog tumors and human tumors using binary mutation profiles of driver genes. **C.** Joint clustering of all human (left) and dog (right) tumors based on mutational signatures. The position of each tumor is identical to panel A. Cluster 9 contains mostly dog tumors and tumors from low-mutation human cancer types. Similar plots for just the low mutation rate human cancer types and the dog cancer types are in Figure 5.

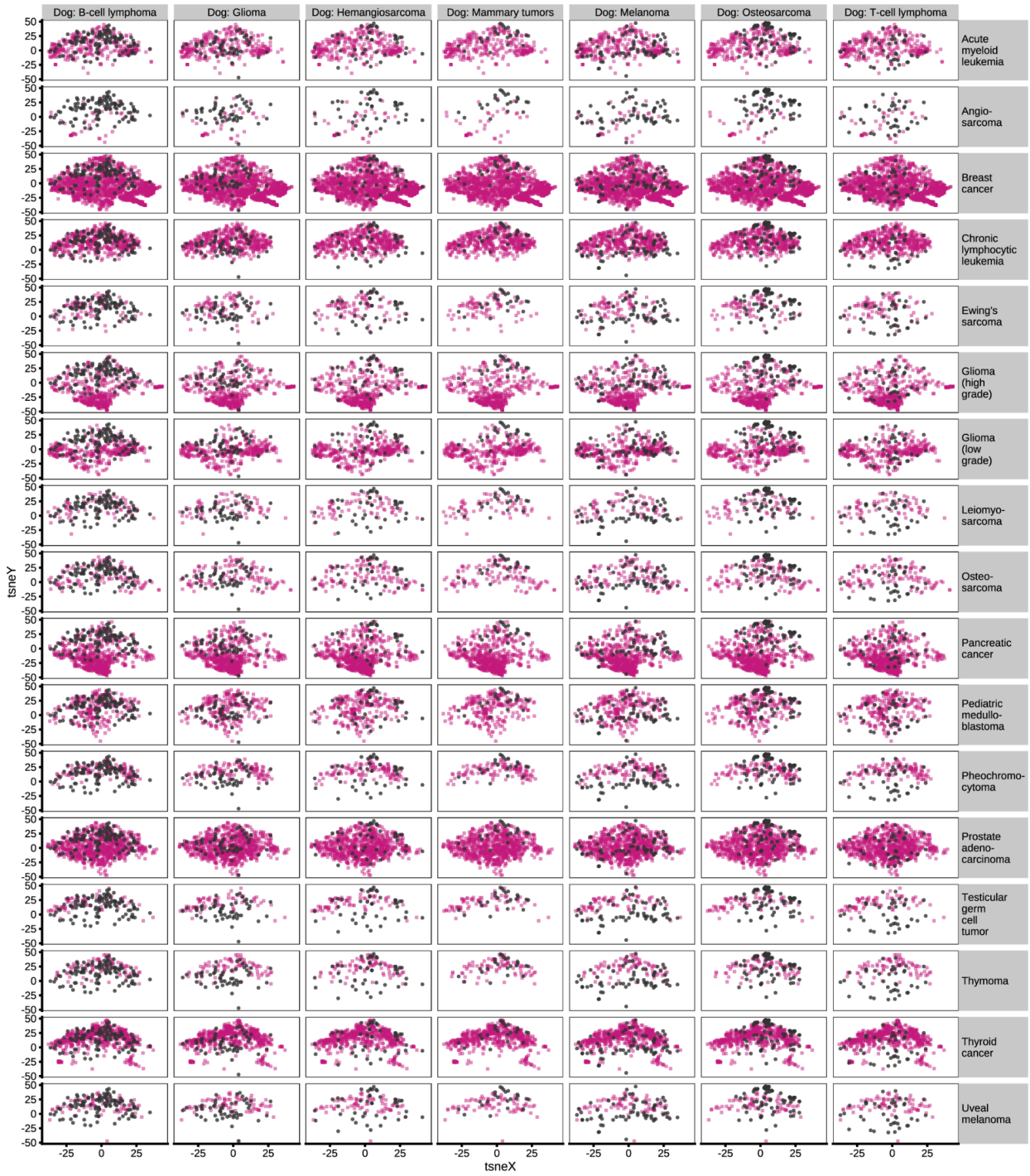


Figure S8. Dog and human cancer types do not form one-to-one matches using mutational signature principal components.

Dog tumors cluster with multiple human cancer types and vice versa. Pairwise overlaps are shown for all dog cancer types (black) and low-mutation human cancer types (pink) using k-means clustering on mutational signatures, based on principal components explaining 70% of the variance.

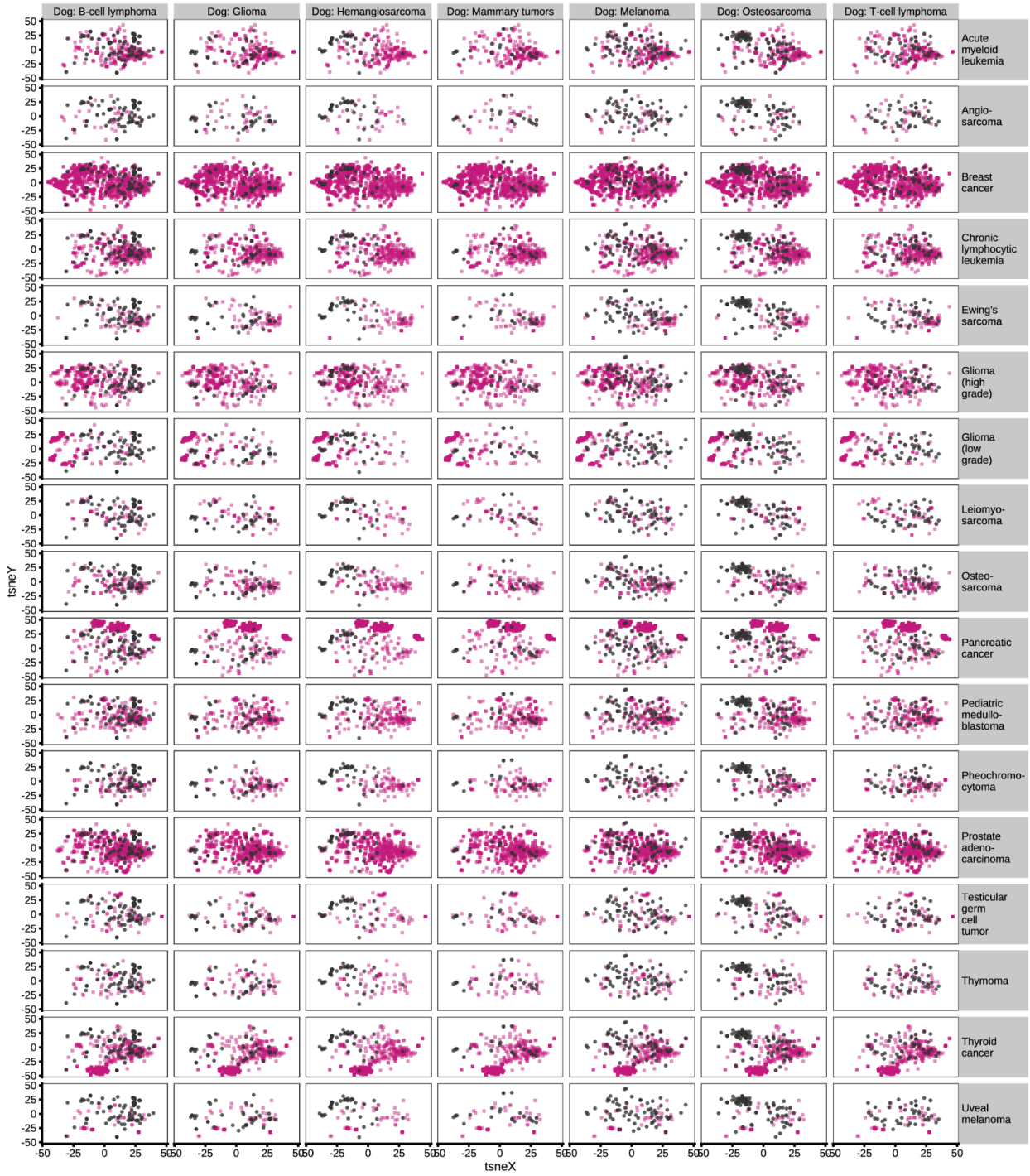


Figure S9. Dog and human cancer types do not form one-to-one matches using driver gene mutations.

Dog tumors cluster with multiple human cancer types and vice versa. Pairwise overlaps are shown for all dog cancer types (black) and low-mutation human cancer types (pink) using k-means clustering on driver gene mutations.

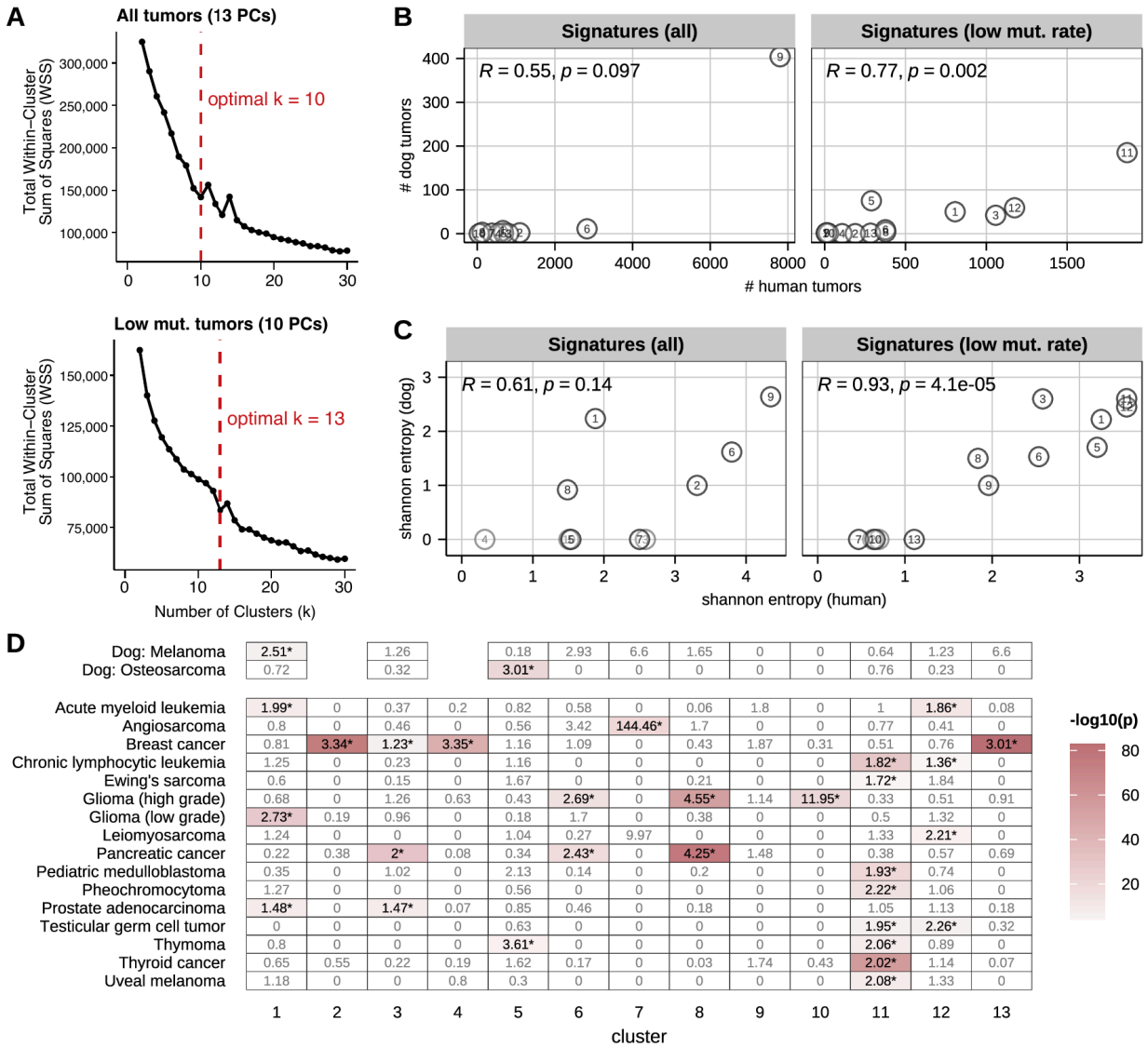


Figure S10. k-Means Unsupervised clustering of dog and human tumors based on principal components explaining 70% of the variance in mutational signatures.

A. Elbow plot showing the total within-cluster sum of squares (WCSS) as a function of the number of clusters (k) used in k-means clustering. The red dashed line marks the "elbow" point at which adding more clusters yields diminishing returns, suggesting an optimal number of clusters. **B.** When clustering all tumors, most tumors fall in cluster 9. Removing high mutation rate human cancer types yields more clusters with more tumors in both species and **C.** more heterogeneous clusters, as measured with Shannon's entropy. **D.** Tumors of some cancer types are enriched in some clusters compared to the null expectation of random distribution.

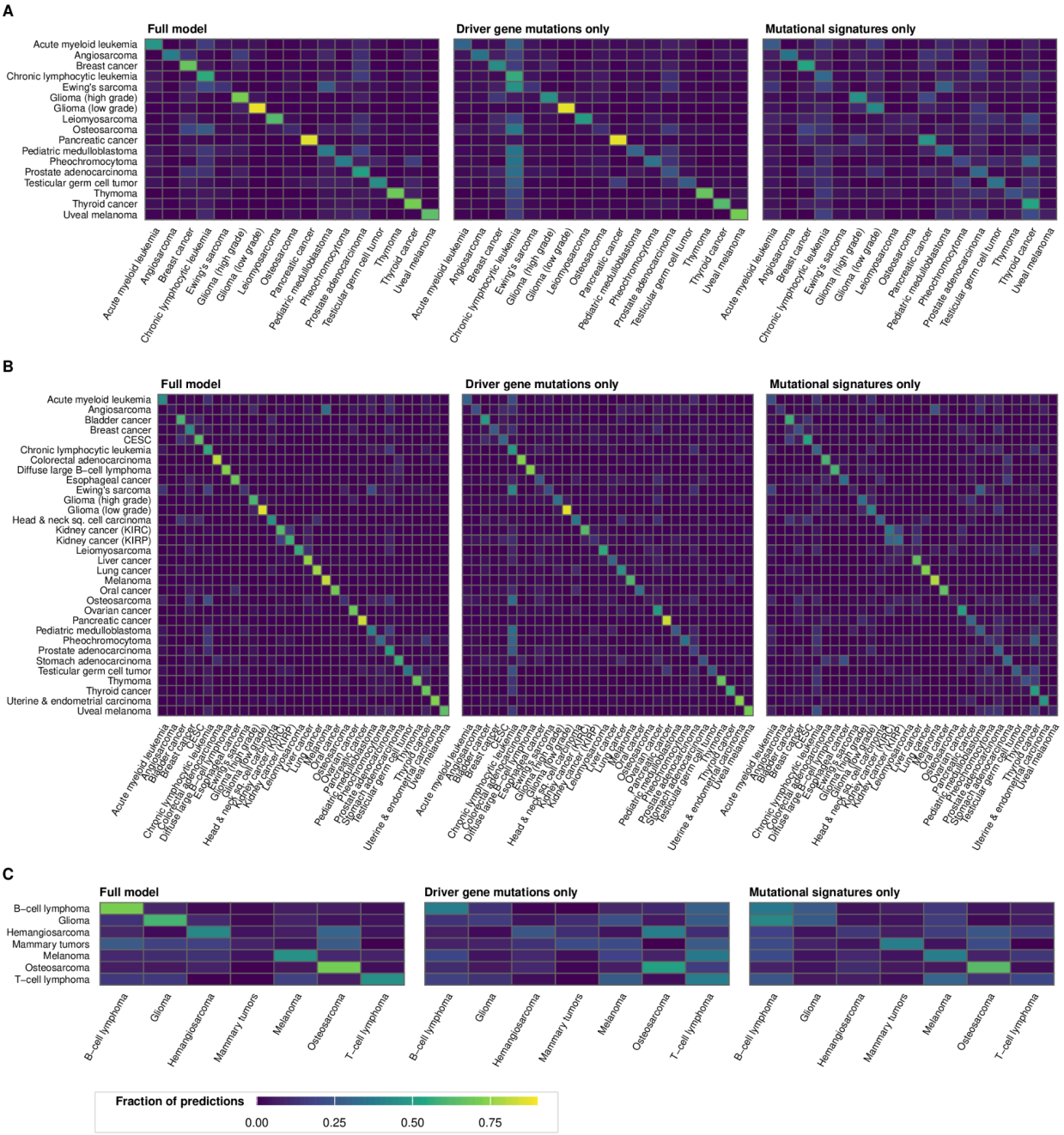


Figure S11. Confusion matrices.

Confusion matrices depicting the proportion of tumors of each cancer type (y-axis) classified as each cancer type (x-axis) for each machine learning model. **(A)** Human trained classifiers on low mutation rate human cancer types only. **(B)** Human-trained classifiers on all human cancer types. **(C)** Dog-trained models on all dog cancer types.

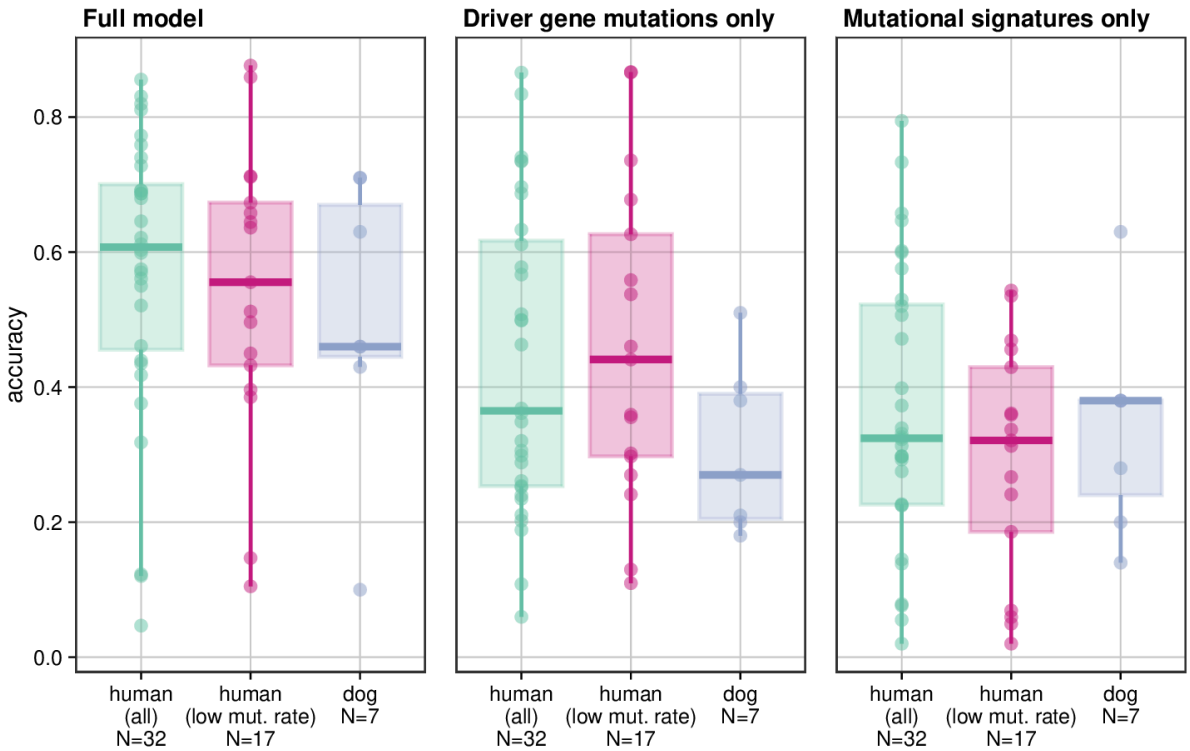


Figure S12. Average accuracy per cancer type.
Boxplots depict the average accuracy across five folds for each cancer type in each machine learning model. Green - human-trained, all cancer types; Pink - human-trained, low mutation rate cancer types; Blue - dog-trained.

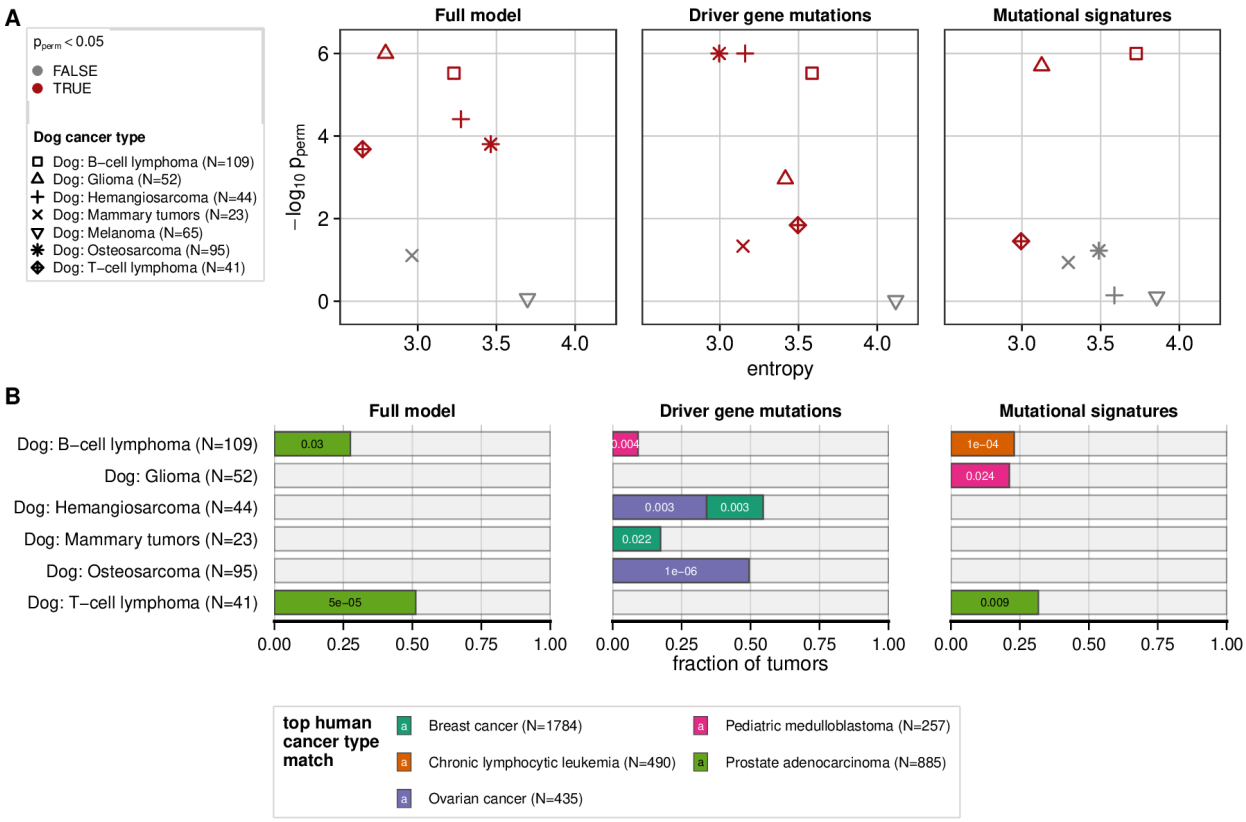
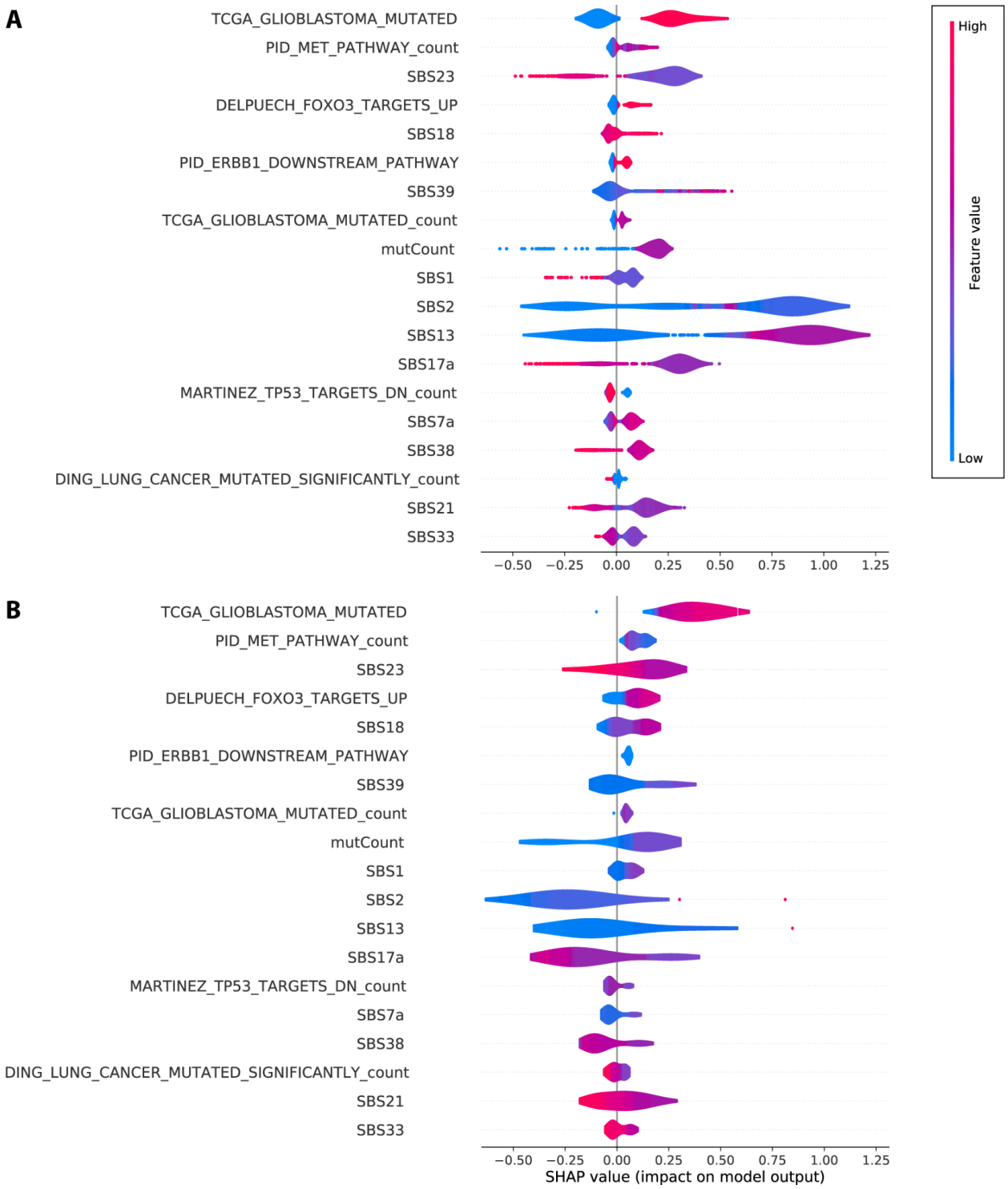


Figure S13. Clumpiness and one-to-one cancer type matching in the human-trained, all cancer types models.

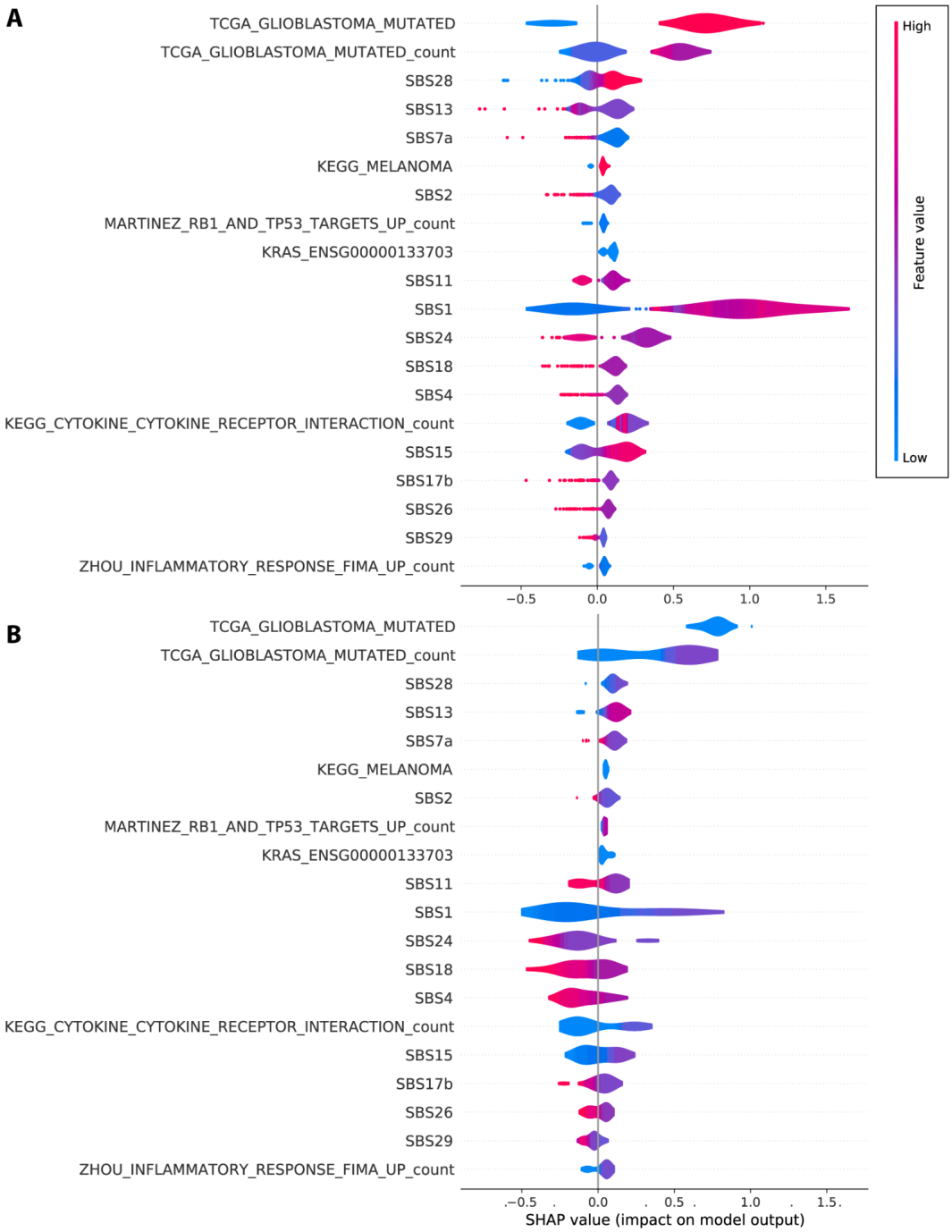
(A) Dog cancers in red were classified as fewer human cancer types than expected by chance. (B) A subset of dog tumors were classified as specific human cancer types more than expected by chance in each model.

Figure S14



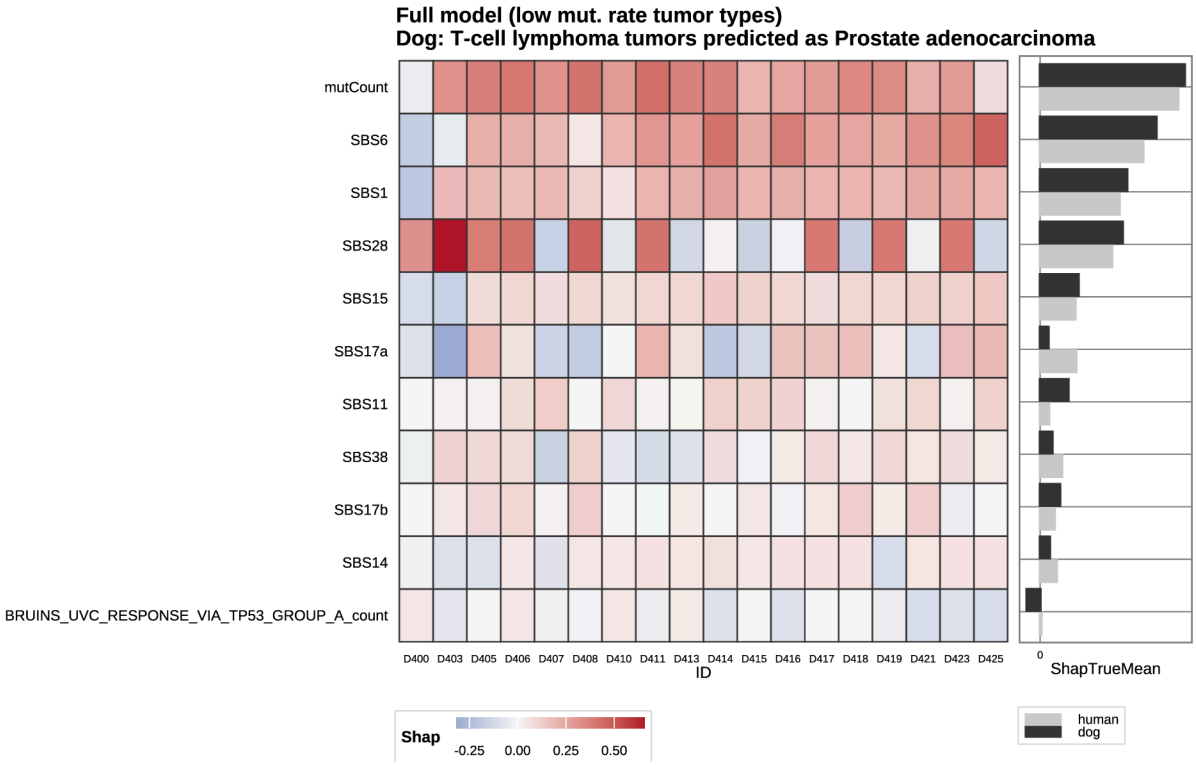
Shapley and feature value violin plots (breast cancer). The top and bottom ten most important features for classification of dog hemangiosarcoma tumors as breast cancer, visualized in (A) human breast cancer tumors classified as breast cancer, and (B) dog hemangiosarcoma tumors classified as breast cancer.

Figure S15 (HGG shapley)



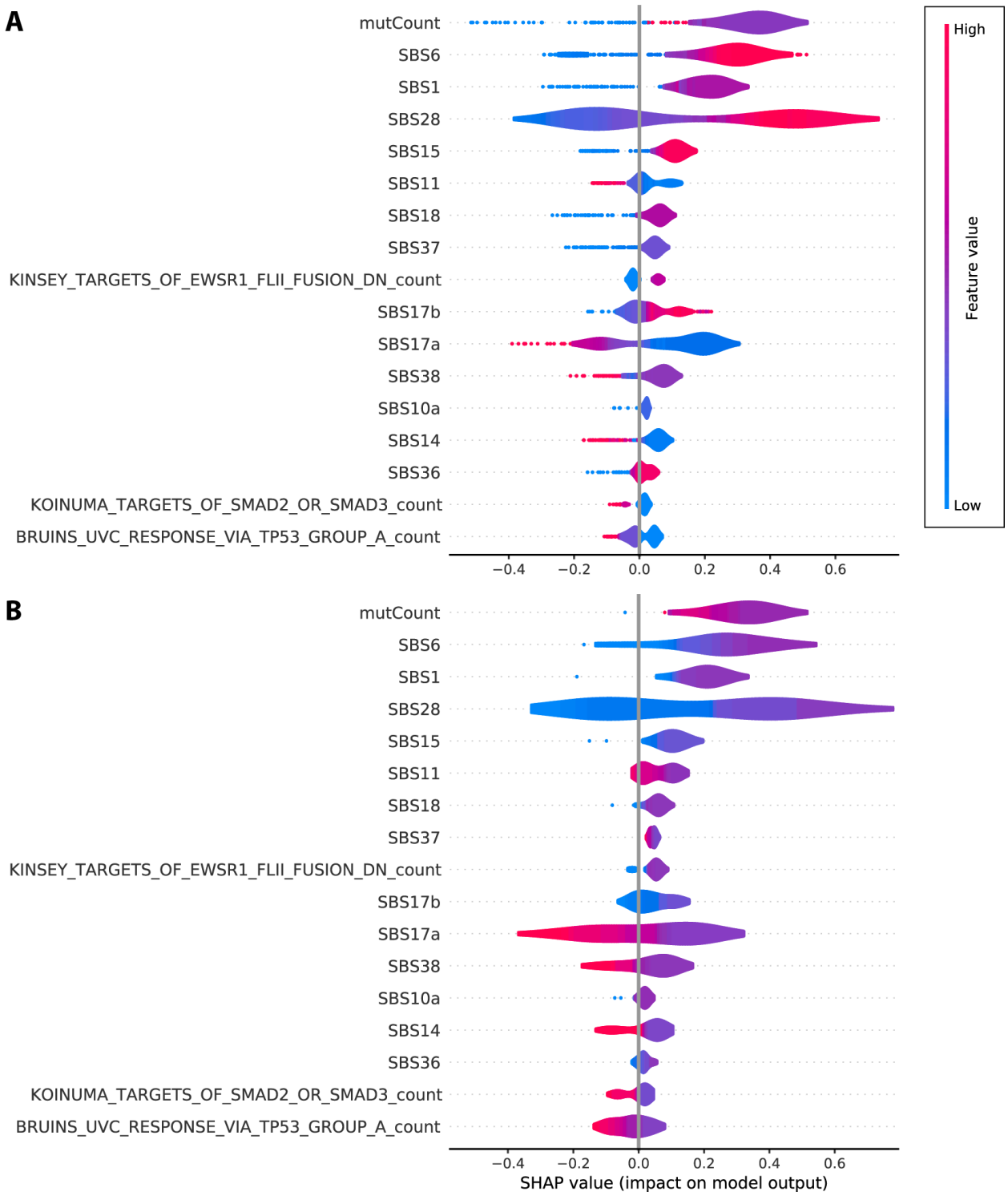
Shapley and feature value violin plots (high-grade glioma). The top and bottom ten most important features for classification of dog hemangiosarcoma tumors as high-grade glioma, visualized in (A) human high-grade glioma tumors classified as high-grade glioma, and (B) dog hemangiosarcoma tumors classified as high-grade glioma.

Fig S16. TLSA



Heatmap of Shapley values (dog T-cell lymphoma classified as prostate adenocarcinoma). The top and bottom most important features for classification of dog T-cell lymphoma tumors as prostate adenocarcinoma, by tumor.

Fig S17.



Shapley and feature value violin plots (prostate adenocarcinoma). The top and bottom ten most important features for classification of dog T-cell lymphoma tumors as prostate adenocarcinoma, visualized in (A) human prostate adenocarcinoma tumors classified as prostate adenocarcinoma, and (B) dog T-cell lymphoma tumors classified as prostate adenocarcinoma.

Supplementary Methods for The shared selection landscape of dog and human cancers

Genereux and Megquier, et al.

Acquisition of human data

We obtained publicly available human somatic mutation data from The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium (ICGC)[1], and cBioPortal[2]. We downloaded data from the 33 projects in the TCGA program on the Genomic Data Commons (GDC). Individual study data from the ICGC Data Coordination Center (DCC) Data Release 28 (excluding TCGA studies) were downloaded from the DCC Data Releases page. In addition, we downloaded data from individual studies from cBioPortal. Data sources are listed in **Table S1**.

Acquisition of dog data

We compiled somatic mutation data from WES and WGS sequencing studies of seven canine cancer types and subtypes. We leveraged mutational calls from previously published studies or preprints. For B-cell lymphoma[3,4], T-cell lymphoma[3], osteosarcoma[5], and hemangiosarcoma[6], we downloaded published mutational data for osteosarcoma[7], and glioma[8]. Mutational calls were provided pre-publication by the authors for oral melanoma[9] and mammary tumors[10]. NCBI, SRA, or ENA sample identifiers are provided in **Table S19**.

Cancer data sets included

We included all studies from the TCGA and ICGC data sets that considered a single cancer type. For studies that contained data from multiple cancer types, we separated samples by cancer type when feasible, otherwise, the entire study was excluded. To include samples for the human cancer types for which we had canine data from the histology-matched human cancer (angiosarcoma, osteosarcoma, and diffuse large B-cell lymphoma), or to increase the number of samples included, we also included data from studies hosted on cBioPortal.

Data QC

We applied consistent criteria in compiling published data from multiple sources. Overall, we favored data from primary tumors rather than metastatic sites, and from earlier- rather than later-stage cancers. We also sought to include no more than one sample from any individual patient. The metadata types available differed somewhat among sample sources, so we filtered as follows:

Human cancer QC:

cBioPortal studies: We removed samples marked as being from relapse time points, as well as those that, based on ID, were duplicates of samples in the included ICGC studies. Single angiosarcoma samples per patient were selected as a subset of the data_clinical_sample.txt file provided with the study data. The first available time point (based on the BX_TIME_DAYS column, annotated as "Sample collection time in days from primary diagnosis") was selected for each patient, regardless of site or procedure. In one case, two samples labeled T1A and T1B were both annotated from the same time point, so the first listed (T1A) was selected. Samples from the Columbia University pediatric study (study identifier: mixed_pipseq_2017) were selected if "osteosarcoma" appeared in the "Cancer Type Detailed" column, “WES” appeared in the "NGS Test" column, and “Diagnostic” appeared in the “Sample_Timepoint” column. Ten samples were found to be shared between the DFCI (study identifier: dlbcl_dfci_2018) and Broad Institute (study identifier: dlbc_broad_2012) DLBCL studies. The data for these samples

from the DFCI study was retained, and the corresponding data from the Broad Institute was excluded.

***ICGC studies**:* We excluded samples that did not have either whole genome or whole exome sequencing data and/or were not from a primary or metastatic tumor sample. When more than one sample was present from a given patient, we prioritized (in order of importance): solid tissue primary tumors to other types of primary tumor; primary tumors to metastatic tumors; if listed, higher percentage cellularity to lower; whole genome sequencing to whole exome; if listed, earlier BINET grade to later BINET grade; enrichment by flow sorting to enrichment by laser capture microdissection; and samples included in the Pan-Cancer Analysis of Whole Genomes over not. For most ICGC sample sets, we excluded all samples if multiple samples remained following application of these criteria. Two studies, LMS-FR and UTCA-FR, contained samples from cancer types for which few samples were available from other studies, so we retained one randomly selected sample per patient.

Dog cancer QC:

We excluded eight samples from the canine melanoma dataset because they were outliers for the number of somatic mutations called. The average mutational count for the remaining 69 samples was 465 (range: 136-1435). One excluded sample had a single somatic mutation. The other seven excluded samples had mean 36125 mutations (range: 11652-56641). Glioma samples were excluded if whole genome or whole exome sequencing was not available or if there was no matched normal. An additional sample was excluded due to being annotated as not having tumor content. Twenty-three of the canine mammary tumors were selected for our analysis based on histology (including carcinomas and mixed tumors, while excluding hyperplasia, adenoma, fibrolipoma, other cancer types (*e.g.* chondrosarcoma), and samples without histology data.

Data harmonization and processing

We defined a set of features for comparison across cancer types and species.

Identifying genes orthologous between human and dog.

We used Ensembl Biomart and NCBI Gene[11] databases to identify a set of 15,315 orthologous genes to compare between human and dog. In cases where the two sources disagreed in defining a given ortholog, all genes identified by either source were considered orthologs of one another. To define the start and end positions of the coding region for each gene, we used annotations from Ensembl (version 104 for canFam3 and hg38; version 87 for hg19). Genes that were predicted to have a length of zero in either species were removed.

Defining known cancer-driver genes

From among the 15,315 orthologous genes, we defined a subset of 1,098 known or suspected driver genes from the Catalog of Somatic Mutations in Cancer (COSMIC) Gene Census[12], as well as genes known to be affected by clonal hematopoiesis in humans or commonly affected in canine cancers. We also designated 4,852 genes that are targeted by existing drugs based on the OncoKB database[13] (**Table S3**).

Identifying predicted-damaging variants.

Variants were parsed from Variant Call Format, Mutation Annotation Format (MAF), or custom formats using a custom python script leveraging the pysam[14] wrapper of the Samtools[15,16] and htslib[17] tools. We included in our analysis variants that fell within the coding regions of our 15,315 orthologous genes and that were predicted to be damaging. In the human datasets, “masked” variants where both reference and alternate alleles were set to reference were

removed. We defined as damaging variants that were classified as moderate or high impact by SnpEff, or those that had ontologies in Variant Effect Predictor (**Table S18**).

Defining mutated genes and pathways.

To define mutated genes, we opted for a binary classification of damaged or not damaged. We used the output of VEP or SNPEff, as available, to determine whether or not a gene was damaged in a sample. We used this same concept to define mutated pathways. Using the C2 database of curated pathways from MSigDB, we defined a pathway as mutated if any of its constituent genes was damaged. We removed from the analysis any samples that had fewer than five mutations remaining after data harmonization.

Defining and assessing hot spot mutations

We defined mutational hotspots at the codon level, considering a mutation in any of the three bases of a given codon to be a hit to that codon. To construct a set of known hotspots for which to tally mutations, we used sites listed in the Database of Curated Mutations (DoCM)[18] and from Hess, *et al*.[19], using a cutoff of Q_LNP < 0.1. We also considered as a hotspot any site mutated in five or more of our samples. To enable consideration of the same codons across all samples, we used UCSC's liftOver tool to liftover variants from samples aligned to canFam3 and hg19 samples to hg38.

Calling mutational signatures.

We used SigFit v.2.2[20] to quantify mutational signature exposures for each sample We tallied trinucleotide context mutational opportunities in the orthologous regions included in our data across hg19, hg38, and CanFam3.1 using the Biostrings package v.2.62[21]. We used mutations observed in each sample, along with the relevant reference genome, to generate trinucleotide context mutation counts for each sample. We fit all of the samples' trinucleotide counts to the normalized COSMIC v3 single base substitution (SBS) signature set[22] using 100,000 iterations and 50,000 warm-up iterations, normalizing each sample by its genome's mutational opportunities. We retained all signatures that were significantly different from zero (lower bound estimate was at least 0.01, the default value for SigFit v.2.2) in at least one sample and refit the samples using only those signatures. To calculate signature values for each species as a whole, we performed the same process as above using the sum of the trinucleotide counts across all samples in that species as the input.

Determining granularity of cancer types Several of the human cancer types represented in our set (*e.g.,* breast cancer, uterine, and endometrial cancers) have multiple known subtypes. To determine at what level of granularity to group the samples, we assessed classification accuracy under machine learning models trained on human data, using extreme gradient boosting (XGBoost). If classification accuracy for a given subtype was < 0.25, it was removed. If a subtype within a given cancer type had classification accuracy ≥ 0.25 but lower than for other subtypes within that same cancer type, it was either made its own category or removed. Specific samples that could not be classified into the resulting categories were excluded from further analysis.

Defining low-mutation cancer types. To define a set of human cancer types with mutation rates comparable to those observed in dogs ("low mutation rate human cancer types"), we applied a conservative threshold of excluding any human cancer type whose median mutation rate exceeded that of the most highly mutated dog cancer type (dog osteosarcoma; median = 0.43 mutations/Mb). However, to enable comparisons between histologically matched and unmatched cancer types, we included two human cancer types which are histologic counterparts to dog cancers and have slightly higher mutation rates: high-grade glioma (0.57 mutations/Mb) and angiosarcoma (0.62 mutations/Mb). We did not include the two histological

matches that were amongst the most mutated human cancer types: diffuse large B-cell lymphoma (1.06 mutations/Mb) and melanoma (3.14 mutations/Mb).

Computing pairwise similarities of samples. To quantify pairwise genomic similarity of tumors, we used the cosine function as implemented in the R package *coop*. Cosine similarity scores can range from 0 (indicating that the tumors are maximally distinct) to 1 (indicating that the two tumors are identical). We computed cosine similarity scores using both damaging mutations and signature values as input.

Computing angles for radial spectrum plots of signature values for each tumor. To display signature values and mutation counts for each tumor, we used radial spectrum plots, using an approach adapted from the methods that Lawrence et al.[23] introduced for plotting tumors according to the relative abundance of classes of mutations. For each tumor, we computed $\alpha ngle = 2\pi\Sigma^{K}_{r=1} i_r(1/K)^r$. Here, K refers to one of the 43 circle sectors to which a given tumor is assigned, Each sector represents one of the 43 signatures we quantified for each sample. In turn, r indicates the rank for that signature in a given sample. This approach places each tumor within the sector that represents its most abundant signature, and within a series of nested subsectors corresponding to the remaining signatures, by declining rank. Each tumor is placed at distance from the circle equal to $\log_{10}$ of its mutations count.

Unsupervised clustering

We implemented unsupervised k-means clustering of all dog and human tumors using the kmeans function in stats version 4.4.3 library in R0. To cluster tumors based on signature values, we first inferred principal components, then used as input for k-means clustering the minimal set of components required to explain 70% in sample-wide variance. To cluster tumors based on driver-gene mutation states, we applied k-means directly to binary mutation data, as, for these sparse data, principal components analysis did not reduce the number of features required to capture 70% of total variance. For each data type, we used the elbow method to determine the number of clusters that minimized within-cluster sums of squares. To enable visualization of these high-dimensionality data, we use t-distributed Stochastic Neighbor Embedding as implemented using the Rtsne library in R.

Machine-learning classification.

We trained extreme gradient boosting (XGBoost) machine-learning models to classify tumors from humans and dogs. We trained models using human data and using dog data for both all cancer types for exclusively low-mutation cancer types. Feature selection. We used a random forest approach for recursive feature elimination with cross validation to select the most informative from among the full set of mutated genes and pathways, mutational signature values, and overall mutation count. This process uses a Bayesian optimization process to repeatedly train models and narrows down optimum values. In each round, this process eliminates 100 features, until a minimum of 400 features are chosen. Model training. We randomly subsetted our data into five folds. For each of these folds, we used 80% of the samples for training the XGBoost model, holding out 20% for testing.

Calculating fraction of samples correctly classified.

We computed the overall accuracy of each model as the fraction of samples assigned to the correct cancer type in the classifier trained on data from that species.

Shapley values

We computed per-sample/per-classification Shapley values to identify the impact of each feature on the probability mass assigned to a given cancer type for a given sample using the (TreeExplainer) function from the (fasttreeshap) package[24] in python. As input, we used the

pruned set of feature types used in making predictions, and the output probabilities per sample per classification.

Assessing classification clumpiness for dog cancers.

To measure clumpiness—the extent to which the best matches for each dog tumor were concentrated in particular human tumor types—we used Shannon's entropy, and assessed the significance of clumping by label permutation. We used a permutation-based approach to assess, for samples of each histological label, whether the human- and dog-trained classifier assigned distribution probability mass to a smaller number of cancer predictions than anticipated by chance. To quantify distribution of probability mass for the true samples of a given histological label, we summed probability-mass distributions across samples, then computed the Shannon entropy ((-1*sum(p_i*ln(p_i)). Larger entropy values indicate greater evenness of density across categories, with maximum entropy equal to the total number of categories. Therefore, the lower the entropy for a given histological label, the more "clumped" the probability mass assigned by the classifier. To assess the significance of clumping for each histological label, we drew one million sets of samples at random and with replacement from the full sample set, with sample size matched to the true sample size for each histological label. We computed the entropy for each of these random draws. To compute a one-sided p-val for the probability of observing entropy as low as observed for the true samples by chance alone, we computed the fraction of randomized samples that yielded an entropy at least as low as that for the true sample set.

Assessing "significantly associated" histological label / classifier label pairs

To assess which histological labels are associated with a particular classifier label more than expected by chance, we again used a permutation test. For each histological label, we first summed the probability weights outputted by the classifier grouped by classification label to get a cumulative probability sum for each histological label / classifier label pair. For each histological label we then generated one million sets of samples at random and with replacement from the full sample set for that species, where the size of each sample was equal to the number of samples with that histological label. For each of these sets, we computed the cumulative probability sums as we did for the original set and retained the maximum sum for that set. To compute a one-sided p-value for the probability of a given histological label being associated with a given classification label, we calculated the fraction of sets drawn for that histological label whose maximum sum is at least as high as the sum for the histological label / classification label pair.

References

1. Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., Stein, L.D., and Ferretti, V. (2019). The International Cancer Genome Consortium Data Portal. Nat. Biotechnol. *37*, 367–369.

2. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, l1.

3. Elvers, I., Turner-Maier, J., Swofford, R., Koltookian, M., Johnson, J., Stewart, C., Zhang, C.-Z., Schumacher, S.E., Beroukhim, R., Rosenberg, M., et al. (2015). Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. Genome Res. *25*, 1634–1645.

4. White, M.E., Hayward, J.J., Hertafeld, S.R., Castelhano, M.G., Leung, W., Dave, S.S., Bhinder, B.H., Elemento, O.L., Boyko, A.R., Richards, K.L., et al. (2020). Consensus-based somatic variant-calling method correlates FBXW7 mutations with poor prognosis in canine B-cell lymphoma. bioRxiv, 2020.08.16.250100. https://doi.org/10.1101/2020.08.16.250100.

5. Sakthikumar, S., Elvers, I., Kim, J., Arendt, M.L., Thomas, R., Turner-Maier, J., Swofford, R., Johnson, J., Schumacher, S.E., Alföldi, J., et al. (2018). SETD2 Is Recurrently Mutated in Whole-Exome Sequenced Canine Osteosarcoma. Cancer Res. *78*, 3421–3431.

6. Megquier, K., Turner-Maier, J., Swofford, R., Kim, J.-H., Sarver, A.L., Wang, C., Sakthikumar, S., Johnson, J., Koltookian, M., Lewellen, M., et al. (2019). Comparative genomics reveals shared mutational landscape in canine hemangiosarcoma and human angiosarcoma. Mol. Cancer Res. https://doi.org/10.1158/1541-7786.MCR-19-0221.

7. Gardner, H.L., Sivaprakasam, K., Briones, N., Zismann, V., Perdigones, N., Drenner, K., Facista, S., Richholt, R., Liang, W., Aldrich, J., et al. (2019). Canine osteosarcoma genome sequencing identifies recurrent mutations in DMD and the histone methyltransferase gene SETD2. Commun Biol *2*, 266.

8. Amin, S.B., Anderson, K.J., Boudreau, C.E., Martinez-Ledesma, E., Kocakavuk, E., Johnson, K.C., Barthel, F.P., Varn, F.S., Kassab, C., Ling, X., et al. (2020). Comparative Molecular Life History of Spontaneous Canine and Human Gliomas. Cancer Cell *37*, 243–257.e7.

9. Prouteau, A., Mottier, S., Primot, A., Cadieu, E., Bachelot, L., Botherel, N., Cabillic, F., Houel, A., Cornevin, L., Kergal, C., et al. (2022). Canine Oral Melanoma Genomic and Transcriptomic Study Defines Two Molecular Subgroups with Different Therapeutical Targets. Cancers *14*. https://doi.org/10.3390/cancers14020276.

10. Arendt, M.L., Sakthikumar, S., Melin, M., Elvers, I., Rivera, P., Larsen, M., Sällström, S., Lingaas, F., Rönnberg, H., and Lindblad-Toh, K. (2022). PIK3CA is recurrently mutated in canine mammary tumors, similarly to in human mammary neoplasia. https://doi.org/10.21203/rs.3.rs-1801127/v1.

11. Brown, G.R., Hem, V., Katz, K.S., Ovetsky, M., Wallin, C., Ermolaeva, O., Tolstoy, I., Tatusova, T., Pruitt, K.D., Maglott, D.R., et al. (2015). Gene: a gene-centered information resource at NCBI. Nucleic Acids Res. *43*, D36–D42.

12. Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat. Rev. Cancer *18*, 696–705.

13. Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol *2017*. https://doi.org/10.1200/PO.17.00011.

14. pysam: Pysam is a Python package for reading, manipulating, and writing genomics data such as SAM/BAM/CRAM and VCF/BCF files. It's a lightweight wrapper of the HTSlib API, the same one that powers samtools, bcftools, and tabix (Github).

15. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

16. 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*. https://doi.org/10.1093/gigascience/giab008.

17. Bonfield, J.K., Marshall, J., Danecek, P., Li, H., Ohan, V., Whitwham, A., Keane, T., and Davies, R.M. (2021). HTSlib: C library for reading/writing high-throughput sequencing data. Gigascience *10*. https://doi.org/10.1093/gigascience/giab007.

18. GitHub - griffithlab/docm: Retired Project: Rails frontend to The Genome Institute's database of curated mutations (DoCM) GitHub. http://www.docm.info/.

19. Hess, J.M., Bernards, A., Kim, J., Miller, M., Taylor-Weiner, A., Haradhvala, N.J., Lawrence, M.S., and Getz, G. (2019). Passenger hotspot mutations in cancer. Cancer Cell *36*, 288–301.e14.

20. Gori, K., and Baez-Ortega, A. (2018). sigfit: flexible Bayesian inference of mutational signatures. bioRxiv.

21. Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2023). Biostrings: efficient manipulation of biological strings. Biostrings, R package version 2.62. Preprint.

22. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. *47*, D941–D947.

23. Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

24. Yang, J. (2021). Fast TreeSHAP: Accelerating SHAP value computation for trees. arXiv [cs.LG].