



OPEN

Integrated analysis of canine soft tissue sarcomas identifies recurrent mutations in *TP53*, *KMT* genes and *PDGFB* fusions

Sunetra Das^{1,2}, Rupa Idate^{1,2}, Susan E. Lana^{1,2}, Daniel P. Regan^{2,3} & Dawn L. Duval^{1,2,4}✉

Soft tissue sarcomas (STS) are a heterogeneous group of mesenchymal tumors representing over 50 distinct types with overlapping histological features and non-specific anatomical locations. Currently, localized sarcomas are treated with surgery +/– radiation in both humans and dogs with few molecularly targeted therapeutic options. However, to improve precision-based cancer therapy through trials in pet dogs with naturally occurring STS tumors, knowledge of genomic profiling and molecular drivers in both species is essential. To this purpose, we sought to characterize the transcriptomic and genomic mutation profiles of canine STS subtypes (fibrosarcoma, undifferentiated pleomorphic sarcoma, and peripheral nerve sheath tumors), by leveraging RNAseq, whole exome sequencing, immunohistochemistry, and drug assays. The most common driver mutations were in cell cycle/DNA repair (31%, *TP53*-21%) and chromatin organization/binding (41%, *KMT2D*-21%) genes. Similar to a subset of human sarcomas, we identified fusion transcripts of platelet derived growth factor B and collagen genes that predict sensitivity to PDGFR inhibitors. Transcriptomic profiling grouped these canine STS tumors into 4 clusters, one PNST group (H1), and 3 FSA groups selectively enriched for extracellular matrix interactions and *PDGFB* fusions (H2), homeobox transcription factors (H3), and elevated T-cell infiltration (H4). This multi-omics approach provides insights into canine STS sub-types at a molecular level for comparison to their human counterparts, to improve diagnosis, and may provide additional targets for chemo- and immuno-therapy.

Soft tissue sarcomas (STS) are a heterogeneous group of mesenchymal neoplasms which represent approximately 8–12% of malignant tumors in dogs. Predominant within this category are hemangiosarcoma (HSA), fibrosarcoma (FSA), peripheral nerve sheath tumors (PNST), perivascular wall tumors (PWT) and histiocytic sarcomas, with myxosarcoma, liposarcoma, rhabdomyosarcoma, leiomyosarcoma, synovial cell sarcomas, and lymphangiosarcomas occurring less frequently. Approximately 6000–25,000 cases are diagnosed in dogs yearly compared to approximately 12,500 STS cases diagnosed annually in adults and children in the United States¹. The subset of canine cutaneous and subcutaneous STS profiled here parallels current diagnostic and prognostic conventions in veterinary medicine (FSA, PNST, and PWT) and excludes those sarcomas with more definitive features (hemangiosarcoma, lymphangiosarcoma, histiocytic sarcoma, synovial sarcoma, leiomyosarcoma, and rhabdomyosarcoma). The FSA, PNST, and PWT tumors typically arise from subcutaneous tissues, are locally aggressive with low to moderate metastatic rates and are usually treated with surgical excision and/or radiation therapy². Previous studies have used immunohistochemistry, microarray, and RT-PCR techniques to identify markers to distinguish STS sub-types^{3–5}. Histologically, these STS sub-types can be difficult to discern due to overlapping morphology and their controversial cell of origin⁶. Thus, genomic, and transcriptomic profiling of canine STS may prove valuable to differentiate these sub-types of mesenchymal neoplasms and understand their biological and clinical behavior.

To date a limited number of genome wide studies have been conducted in canine STS. Whole exome sequence analysis of tumor and matched normal samples from 10 dogs with STS (seven PNST, one FSA, one myxosarcoma, and one synovial cell sarcoma) identified single nucleotide substitutions in *NF1*, *MLL3*, and *PTCH1*, and

¹Department of Clinical Sciences, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO 80523, USA. ²Flint Animal Cancer Center, Colorado State University, Fort Collins, CO 80523, USA. ³Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO 80523, USA. ⁴University of Colorado Cancer Center, Anschutz Medical Campus, Aurora, CO 80045, USA. ✉email: dawn.duval@colostate.edu

amplification of *MDM4* in 4 separate tumors⁷. The only recurrently mutated gene was *ATP7B* ($n=2$), while *AIG1* was amplified in 2 tumors. Another study identified recurrent changes in the TP53 pathway among 6 primary malignant PNST⁸. In contrast, human PNST is strongly associated with neurofibromatosis type 1 with a 10–15% lifetime risk of PNST development⁹. Recent next-generation sequencing of human malignant PNST identified recurrent mutations in *NF1* (87.5%), *TP53* (40%), and *CDKN2A* (75%) in addition to a dysregulated Ras pathway¹⁰.

Canine FSA is frequently associated with rearrangements of chromosome 11 including loss of heterozygosity in the region containing *CDKN2B-CDKN2A*¹¹. In contrast, FSA in both children and adults are increasingly categorized based on frequently observed gene fusions. For example, congenital FSA often bears *ETV6-NTRK3* gene fusions that activate the Ras-MAP kinase and PI3 kinase signaling cascades¹². The majority of adult human low-grade fibromyxoid sarcomas have *FUS-CREB3L2* or *FUS-CREB3L1* fusions, with the highly aggressive sclerosing epithelioid fibrosarcomas prevalently bearing *EWSR1-CREB3L1* fusions¹³. In addition, greater than 90% of Dermatofibrosarcoma Protuberans (DP) tumors carry *COL1A1-PDGFB* fusions leading to both improved diagnosis and neoadjuvant therapy with PDGFR inhibitors for patients with recurrent, nonresectable, or metastatic disease¹⁴.

In this study, we have utilized a multi-omics strategy to delineate transcriptomic, genomic, and immunogenomic profiles of canine STS tumors. These 29 STS tumors identified histologically as FSA and PNST, were grouped by gene expression profiles into a single PNST cluster and 3 FSA clusters. The genomic profile identified *TP53* and *KMT2D* as the top two recurrently mutated genes with other identified driver mutations in cell cycle, DNA repair, and chromatin remodeling pathways. One of the major findings was the identification of *PDGFB* fusions with collagen genes in 10% of the STS tumors. Both immunohistochemical and genomic immune profiling identified a set of tumors with high infiltration of immune cells. Taken together, exploration of these STS tumors reveals distinct groups of tumors with discernable genomic, transcriptomic, and immune-genomic profiles.

Results

Canine patients and tumor overview. For this multi-omics study, STS tumors excised from 29 canine patients were reviewed by a veterinary pathologist and diagnosed into two major histologic sub-types of STS based on morphology alone: fibrosarcoma (FSA, $n=15$) and peripheral nerve sheath tumor/perivascular wall tumor (PNST $n=14$) (Table 1). While we recognize that the latter subtype represents a single morphological entity whose cellular origin remains controversial, investigating this distinction was not a primary goal of this study. Most dogs in this study were mixed breed ($n=11$) with Labrador retrievers ($n=5$) representing the single largest breed group and 41% of the 29 patients were female. Twenty-four patients had sub-cutaneous tumors of the extremities, with the remaining tumors in the oral cavity. These oral cavity tumors tested negative for melanoma markers. Of the tumors selected for this study, (44%) were grade III, with increased potential for recurrence and metastasis¹⁵. Histologically, FSAs were characterized by proliferations of spindle shaped cells arranged in streaming interwoven bundles or herringbone patterns in a background of dense collagenous stroma, and a lack of whorling. A subset of 7 FSA samples (T1124, T1238, T1292, T1490, T1834, T1923, and T416) were poorly differentiated/anaplastic, characterized by sheets of neoplastic cells with marked cellular and nuclear pleomorphism, polygonal to ovoid cellular morphology and containing large vesicular nuclei. Subsequent immunohistochemical staining and transcript analysis of the 7 anaplastic/undifferentiated sarcomas for the histiocyte markers, *CD204* and *Iba1*, identified T1923 as an undifferentiated pleomorphic sarcoma (UPS). In contrast, tumors diagnosed as PNST/PWT were characterized spindle shaped cells arranged in whorling, and storiform or palisading (Antoni A) growth patterns, or loosely arranged hypocellular regions within a myxoid matrix (Antoni B). No definitive histological features suggesting origination from vessels or nerves (besides whorling around central unidentifiable structures) was observed. However, to determine whether gene expression could further distinguish the tumors histologically identified as PNST/PWTs, the expression of pericyte markers was extracted¹⁶. No significant up-regulation of these 17 genes were observed (Supplementary Fig. S1), while many samples demonstrated enrichment of genes and pathways consistent with peripheral nerve sheath origin, as detailed below.

Transcriptomic profiling of STS. *Gene expression defined tumor clusters.* The RNAseq reads ranging from 45.8 to 64.5 million were trimmed and 94% of these reads were mapped against CanFam3.1 genome to obtain gene expression data (Supplementary Table S1). Unsupervised hierarchical clustering of 1926 genes grouped these samples into 4 optimal hierarchical (H) clusters with 11, 7, 5, and 6 samples in H1, H2, H3, and H4 clusters, respectively (Fig. 1, Supplementary Fig. S2)¹⁷. Comparison of histological categorization and tumor clustering revealed that all tumors in cluster H1 were PNSTs. However, three samples histologically identified as PNSTs were placed in cluster H2 (T738, T479) or H3 (T232) (Fig. 1A). The multi-dimensional scaling plot shows that clusters H1 and H2 are separated while H3 and H4 display some overlap (Fig. 1B).

Comparisons were made between samples from one cluster and combined samples from the remaining clusters to identify the genes and pathways distinguishing these 4 clusters. The significantly up-regulated genes (H1 = 575, H2 = 94, H3 = 119, H4 = 430, Supplementary Table S2) in each cluster were analyzed to identify pathways and gene ontologies driving the four clusters (Supplementary Table S3). The top 10 upregulated genes in each of the 4 clusters are plotted in Fig. 2A. The biological themes identified through pathway analysis for H1, H2, H3, and H4 clusters were associated with nervous system-, extracellular matrix-, development-, and immune-related genes, respectively (Fig. 2B). The heatmap of selected genes significantly upregulated in H1 samples involved in neuronal cell body (GO:0043025), neurogenesis (GO:0022008) and neuron differentiation (GO:0030182), includes *GAL*, *CHRNA4*, *NRXN1*, *MAST1*, *ROBO1*, *TNR*, *CBLN1*, *ACTL6B*, *CRMP*,

Sample ID	Breed	Sex	Age at diagnosis	Tumor sub-type	Tumor location	Tumor grade	Gene expression cluster
T1860	Cock-a-poo	Male	13.5	PNST	Neck	II	H1
T1921	Golden Retriever	Female	11.1	PNST	Flank	I	H1
T2063	Mix	Female	10.2	PNST	Thigh	II	H1
T148	Mix	Male	14.8	PNST	Hock	I	H1
T218	Mix	Female	11.5	PNST	Hip	II	H1
T230	Mix	Female	13.3	PNST	Knee region	I	H1
T311	Mix	Male	7.5	PNST	Inguinal region	I	H1
T320	Mix	Male	15.5	PNST	Flank	II	H1
T336	Mix	Female	11.3	PNST	Rib region	II	H1
T344	Basset Hound	Female	10.8	PNST	Humerus	I	H1
T697	Labrador Retriever	Female	8.9	PNST	Thorax	III	H1
T1238	Labrador Retriever	Male	9.1	FSA	Hock	III	H2
T167	Mix	Male	8.6	FSA	Cranial sternum	III	H2
T1891	Labrador Retriever	Male	9.7	FSA	Femur	II	H2
T1989	Labrador Retriever	Female	11.1	FSA	Neck	II	H2
T416	Alaskan Malamute	Male	10.2	FSA	Elbow joint	III	H2
T479	German Shepherd	Female	9.3	PNST	Hip	II	H2
T738	Pointer	Male	5.8	PNST	Tarsus	III	H2
T1124	Greyhound	Male	8.3	FSA	Side, lateral	III	H3
T1407	Mix	Male	3.9	FSA	Maxilla	III	H3
T1490	Golden Retriever	Male	5.4	FSA	Hip	III	H3
T1834	Mix	Male	12.6	FSA	Mandible	III	H3
T232	Poodle	Male	6.9	PNST	Femur	III	H3
T1292	Beagle	Male	12.4	FSA	Maxilla	III	H4
T1462	Labrador Retriever	Male	11.9	FSA	Thigh	III	H4
T1585	Beagle	Female	10.0	FSA	Abdomen	III	H4
T1923	Siberian Husky	Male	10.5	UPS	Rib region	II	H4
T282	Border Collie	Female	12.5	FSA	Lip	II	H4
T485	Mix	Female	6.3	FSA	Maxilla	I	H4

Table 1. Patient and tumor characteristics. *PNST* Peripheral nerve sheath tumor, *FSA* Fibrosarcoma, *UPS* Undifferentiated pleomorphic sarcoma.

NEFH, and *NYAP*. Upregulated extracellular matrix (GO:0031012) genes in H2 samples include: *COL11A1*, *COL24A1*, *COL2A1*, *EMILIN3*, *SFRP1*, and *VIT*. In H3, the upregulated genes *BMPR1B*, *DNAH*, *HOXD11*, *SIX1*, *SIX2*, *NPNT*, *WNT7A*, were associated with pattern specification (GO:0007389) and tissue morphogenesis (GO:0048729). In cluster H4, a significant proportion of enriched pathways were immune system related and included genes like *CCL3*, *CD27*, *CD79B*, *CD8A*, *CTLA4*, *IL21R*, *IL23R*, and *IL7R*, among others (Supplementary Fig. S3 and Supplementary Table S3).

Genomic profile of STS. *Mapping and variant calling statistics.* Whole exome sequencing of 29 primary STS tumors and corresponding matched normals was mapped against the CanFam3.1 genome and somatic short variants and copy number variants (CNVs) were identified. Following trimming, 90.7% ($\pm 0.44\%$) reads were retained and 99.9% ($\pm 0.0003\%$) of those reads were mapped to the CanFam3.1 genome (Supplementary Table S4). The average coverage of mapping was 95.9X ($\pm 28.3X$) and ranged from 22.5X (T282) to 132.7X (T1238) (Supplementary Fig. S4).

Putative mutational signatures of canine STS. The distribution of 6 types of single nucleotide changes shows that C>T was most prevalent, followed by C>A (Supplementary Fig. S5A). Non-negative factorization (NMF) generated three de novo signatures that were assigned to the tumors and named based on their similarity to human COSMIC signatures. The SBS5-like signature, (etiology: correlated with age of patients) was assigned to 52% of the tumors, 38% of the tumors were SBS6-like (etiology: defective DNA mismatch repair), and 10% were considered SBS29-like (etiology: tobacco chewing habit; Fig. 3, Supplementary Fig. S5B)¹⁸. There was no difference in signature distribution across the 4 clusters of tumors, however, the SBS29-like signature was restricted to clusters H3 and H4. The estimation of cosine similarities between individual tumor mutational matrices and 60 COSMIC signatures showed that in addition to SBS5 (n = 14), SBS6 (n = 3) and SBS29 (n = 1), some samples had greater similarity to SBS3 (defective homologous recombination-based DNA damage repair, n = 3), SBS30 (deficiency in base excision repair due to inactivating mutations in NTHL1, n = 2), SBS40 (etiology: unknown, n = 5) and SBS94 (etiology: unknown, n = 1) (Fig. 3C).

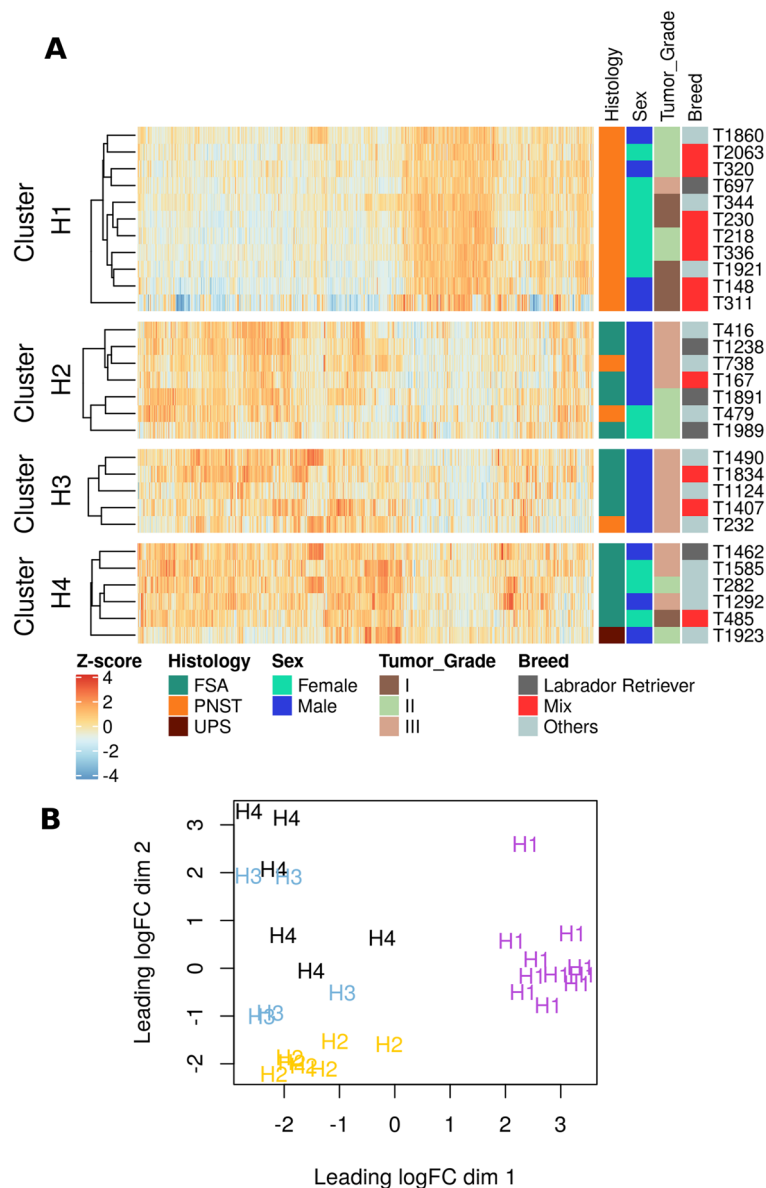


Figure 1. Clustering of canine soft tissue sarcoma samples. **(A)** Unsupervised hierarchical clustering of 29 STS tumor samples using genes with variance $> \log_2 5$ and mean $> \log_2 2$. The samples were clustered using the spearman correlation distance method and grouped into 4 clusters. The clinical features associated with samples were plotted on the right side of the heatmap. **(B)** Multidimensional scaling (MDS) plot visualizing principal component analysis in two dimensions. Tumor samples with similar gene expression profiles were clustered together.

Mutations in STS tumors. *SNVs and INDELS.* The number of SNVs ranged from 193 (T2063) to 3744 (T282) and INDELS range was 43 (T1238) to 1281 (T311) (Supplementary Fig. S6). A total of 6439 non-synonymous coding sequence (CDS) variants in 4254 genes were extracted from 29 samples ranging from 13 (T1238) to 1329 (T282) (Supplementary Table S5). The top recurrently mutated genes, present in 21% of samples, included *KMT2D*, *TP53*, *SAFB*, *ZFN398*, *TTN*, *OR13L2*, and *COR52A16* (Supplementary Fig. S7). The median purity of tumors was estimated to be 0.54 (range: 0.31–1), although, there was no correlation between purity and number of variants identified (Pearson's correlation coefficient: 0.16, p -val = 0.39).

The median protein coding mutations per callable megabase (coverage $> 10X$) was 2.7 and ranged from 0.03 (T1238) to 37.3 (T282) (Fig. 3). Missense mutations were most prevalent (median 62% of all CDS mutations), followed by frameshift and UTR regions variants (13% each) (Supplementary Fig. S8). Using the SIFT scoring method, the identified deleterious missense mutations ranged from 2 in T1238 to 553 in T282 with a median of 29 across all tumor samples. Functional annotation of CDS variant genes identified 116 pathways (FDR < 0.05 , Supplementary Table S6). Relevant pathways included: PI3K-Akt signaling (72% tumors), focal adhesion, RHO

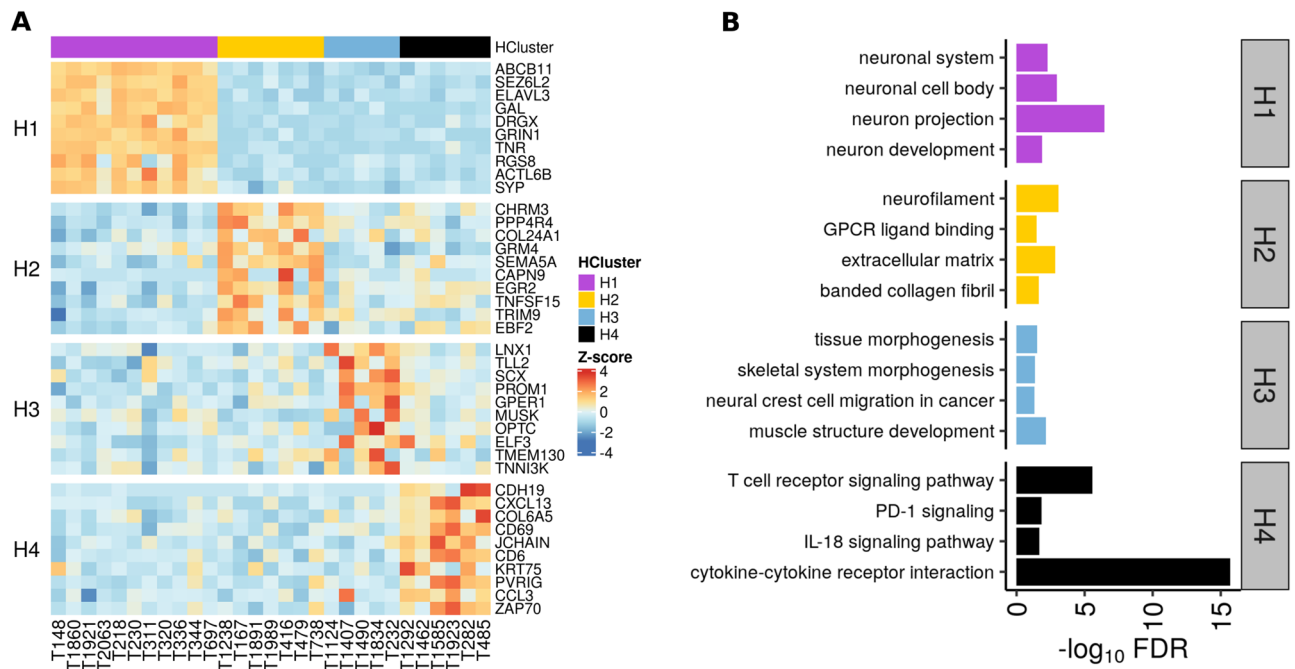


Figure 2. Enriched genes and pathways in 4 clusters. **(A)** Heatmap of top 10 significant differentially expressed genes (positive \log_2 fold change > 2) in each cluster. **(B)** Graphical representation of false discovery rate of representative pathways and gene ontologies identified by functional annotation of DEGs in each cluster. The main biological theme in cluster H1, H2, H3, and H4 were nervous system, extracellular matrix, development, and immune system-related, respectively. The enriched pathways were identified using R package gprofiler2.

GTPase cycle (58% tumors), signaling by PDGF (52% tumors) and chromatin binding (83% tumors). Since most somatic variants likely represent passenger mutations, we identified putative driver genes by cross-referencing canine somatic variants to likely human oncogenic variants. About 6.6% ($n = 427$) of CDS variants were within OncoKB designated cancer genes with 55 variants in 37 genes identified as putative drivers¹⁹. The median number of cancer genes in each tumor was 9 (range 0–75). These genes were grouped into 7 biological categories: cell cycle and DNA repair, chromatin organization/binding, PI3K-Akt-mTOR-signaling pathway, Wnt signaling pathway, transcription regulator activity, nervous system development, and kinase binding (Fig. 4). The bulk of the putative driver variants (37%) were in chromatin organization/binding genes, followed by 18% categorized in the cell cycle and DNA repair and nervous system development gene groups. About 28% of samples carried mutations in *KMT* genes, most prevalent was *KMT2D* mutated in 6 samples: T311, T167, T416, T282, T1462, T1891 (Supplementary Fig. S9). There were 10 STS tumors with alterations in cell cycle and DNA repair genes that included recurrently mutated *TP53* (5 samples: T1124, T1490, T336, T416, T1834, T1462). Among the 8 putative driver genes grouped in nervous system development, five genes were mutated in tumors that were histologically identified as PNSTs.

Copy number variants. We identified 1203 genes that were located within significantly amplified or deleted regions of canine genome (Supplementary Table S7). The median number of altered copy number variation (CNV) genes across 29 samples was 576 (range 8 (T230) to 1093 (T1834) Supplementary Fig. S10). There were significantly more amplified genes (77.6%) (Student T-test, p -value $1.3e-07$). The only significantly enriched functional pathway was the Hippo signaling pathway, and enriched gene ontologies included blood vessel morphogenesis possibly associated with angiogenesis (Supplementary Table S8). Cancer genes, as designated in the OncoKB database, were 5% of all CNV genes with 51 and 13 significantly amplified and deleted genes, respectively. To evaluate the functional effect of these CNVs, we used Pearson correlations between CNV gene amplitude values and Z-scored gene expression values. This identified 242 amplified and 68 deleted genes positively correlated with corresponding expression levels (Supplementary Table S9). Of these, 16 were known cancer genes restricted to chromosomes 6, 22, 13, 14 and 36. This included 8 amplified genes from CFA13: *AGO2*, *PABPC1*, *UBR5* (each altered in 35% of tumors), *RAD21*, *NDRG1*, *EXT1* (38%), *FIP1L1*, *CHIC2* (41%); 3 CFA14 genes: *POT1* (35%), *HOXA9*, *HOXA11* (62%), and *ITGAV* (CFA36, 79%). Of the 4 deleted genes: *DIS3*, *RB1*, *SETDB2* (35%), were on CFA22, while *SBDS* (45%) was on CFA6 (Fig. 4). The CFA13 amplified genes were largely restricted to clusters H2-H4 except for T697.

Fusion variants. Using the RNAseq data, 234 unique fusion genes were identified, where 121 were inter-chromosomal and 113 were intrachromosomal fusions. To identify putative drivers, fusions with gene partners from the same gene family and genes without annotations were eliminated and the remaining fusions were cross

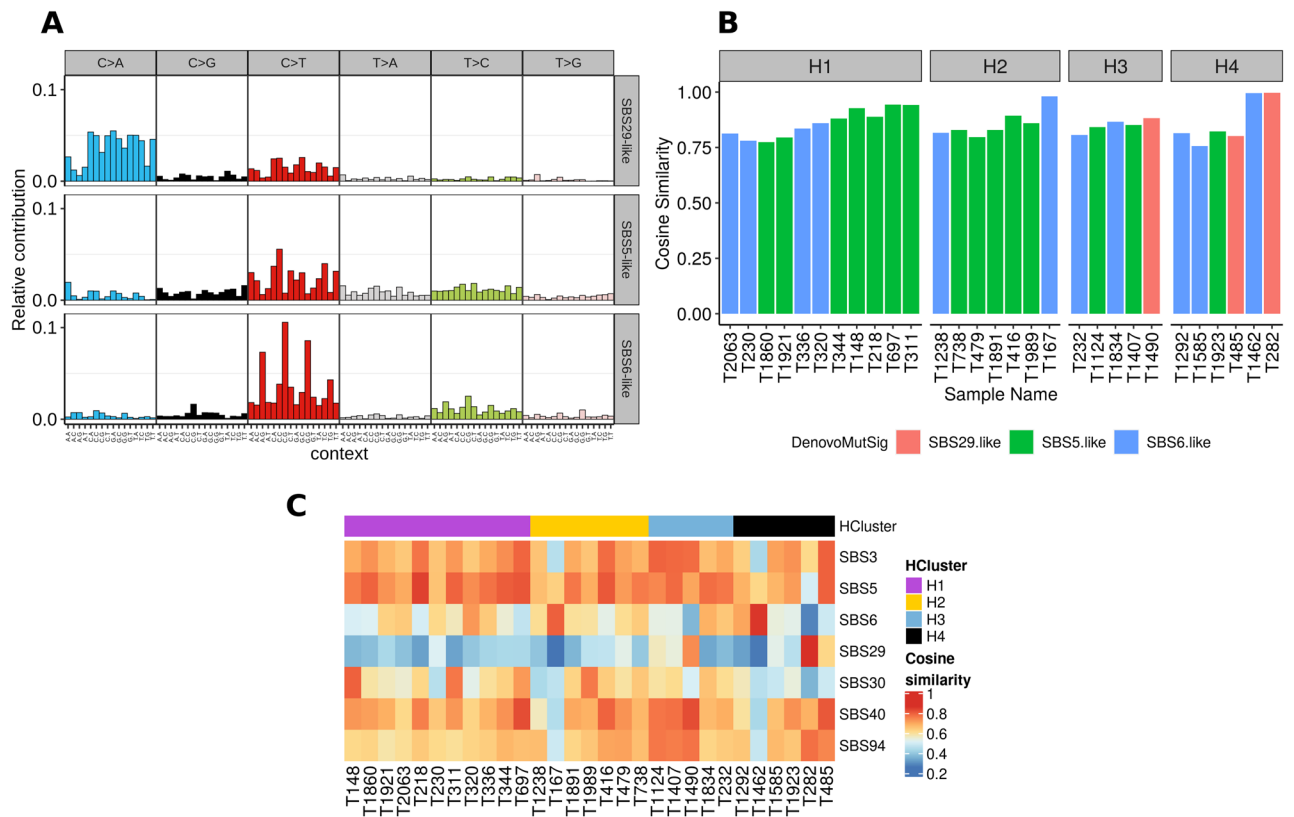


Figure 3. COSMIC mutational signature analysis in canine STS tumors. **(A)** Relative contribution of 96 trinucleotide contexts across three predominant de novo signatures as identified by non-negative matrix factorization method of SNV frequencies from samples. The three de novo signatures were similar to COSMIC signatures SBS5, SBS6, and SBS29. **(B)** Distribution of de novo signatures across 29 STS samples. **(C)** Heatmap of cosine similarities between COSMIC mutational signature profiles and the single nucleotide mutation frequency profiles of 29 STS tumors. For each sample, the signature with the highest cosine similarity was chosen to plot. This resulted in the selection of 7 COSMIC mutational signatures.

referenced against the Cbioportal database to identify likely oncogenic changes due to loss of TSG or activation of oncogenes. This identified 15 putative oncogenic fusions within 9 samples (Supplementary Table S10) and a subset of these are depicted in Fig. 4. Most of these putative driver fusions ($n=12$) were identified in clusters H2 and H3, with a single fusion (CAND2-PPARG) in H1 and two (CNOT2-MSH2, SPRED1-FAM98B) in the H4 clusters.

Three tumors had in-frame fusions of collagen genes (exon 6 *COL1A2*, exon 20 *COL1A2* and exon 35 *COL3A1*) with exon 2 of platelet derived growth factor subunit B (*PDGFB*) which were confirmed by Sanger sequencing of amplified cDNA (T1238/T416 and T738, respectively, Supplementary Fig. S11). These fusions were associated with significantly increased expression of *PDGFB* (Z-score: 4.22 in T416, 2.16 in T738, 1.11 in T1238) in these tumors compared to the cohort (Supplementary Fig. S12). Of these 3 tumors, T1238 and T738 also have reduced *CDKN2A* expression (Z-score: -0.69 , -0.65), while tumor T416 bears a TP53 missense mutation and *RB1* CNV loss (Supplementary Fig. S13). These 3 tumors were in the H2 FSA cluster associated with extracellular matrix and G-protein coupled receptor signaling pathways.

Similarly, increased *EGR2* expression in T1238, T167, and TT416 was associated with *EGR2-ATXN7* fusions. Perhaps more relevantly, we also observed loss of TSG transcripts associated with NF2 fusions in T1834, a CD81-MEN1 fusion in T1238, and reduced *RB1* from the *RB1-VWC2* fusion in T1124 (Fig. 4). While the CNOT-MSH2 fusion in T282 was associated with elevated *MSH2* transcripts, the resulting frameshift in the *MSH2* DNA mismatch repair gene may produce nonfunctional protein. In fact, T282 carried the highest mutational burden of any tumor in this study with 37.63 mutations/Mb.

A newly developed canine soft tissue sarcoma cell line, FACC-19CSTS36, with a Sanger sequence confirmed *COL3A1-PDGFB* fusion was tested for sensitivity to imatinib, a tyrosine kinase inhibitor capable of inhibiting v-Abl, c-Kit, and PDGFR. Also included in this analysis as positive and negative controls were the canine C2 mast cell line, with an internal tandem duplication that activates c-Kit, and the canine STSA-1 soft tissue sarcoma cell line with an inactivating frameshift mutation in NF-1²⁰. The C2 cell line, as expected, was sensitive with an interpolated LD50 of 24 nM, while LD50 values for the STSA-1 and FACC-19CSTS36 cells were approximately 1000-fold higher at 17 and 45 mM, respectively (Supplementary Fig. S14).

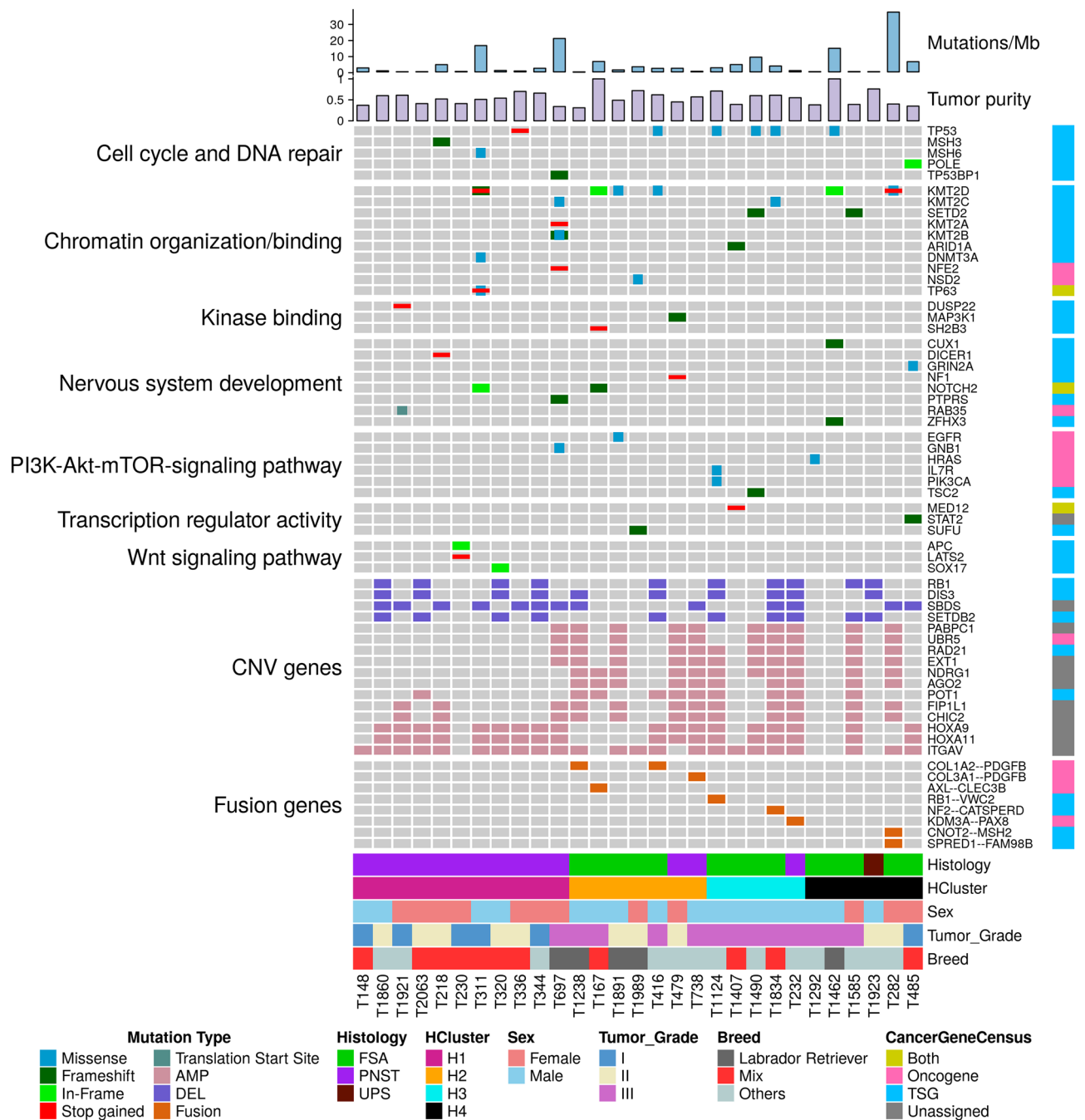


Figure 4. Mutational landscape of canine STS. Oncoplot of selected putative cancer driving variants. INDELS or SNVs are arranged in biological pathways, cancer genes with copy number variations significantly correlated to gene expression (CNV; AMP: amplifications, DEL: deletions) and fusion genes are illustrated. The clinical parameters for each of the 29 samples were represented in the bottom of the Oncoplot, along in silico defined cluster ID. The cancer genes were categorized into oncogene, tumor suppressor gene (TSG), both or unassigned as defined by OncoKB database. Additionally, the top two bar plots included mutational burden scores and tumor purity values per sample.

Markers for FSA and PNST. Previous studies have identified 15 and 7 marker genes for PNST and FSA, respectively^{4,5}. We confirmed significantly elevated expression of 7 of the 15 genes in PNST samples (n = 14): *CLEC3B*, *GLI1*, *FMN2*, *NMUR2*, *DOK4*, *HMG20B* and *GAP43* (Fig. 5A). The remainder of the marker genes were either significantly down-regulated in PNST samples (*NES*, *ROBO1*, *NGFR*) or unchanged (*EGR2*, *MBP*, *MPZ*, *KIF1B*, *SOX10*, Supplementary Fig. S15).

The seven genes reported as markers from FSA samples were *FHL2*, *PLAGL1*, *FBNP1L*, *BAG2*, *HK1*, *CSK* and *COX5A*⁴. Parallel to the previous publication, mean expression of these genes was significantly higher in FSA

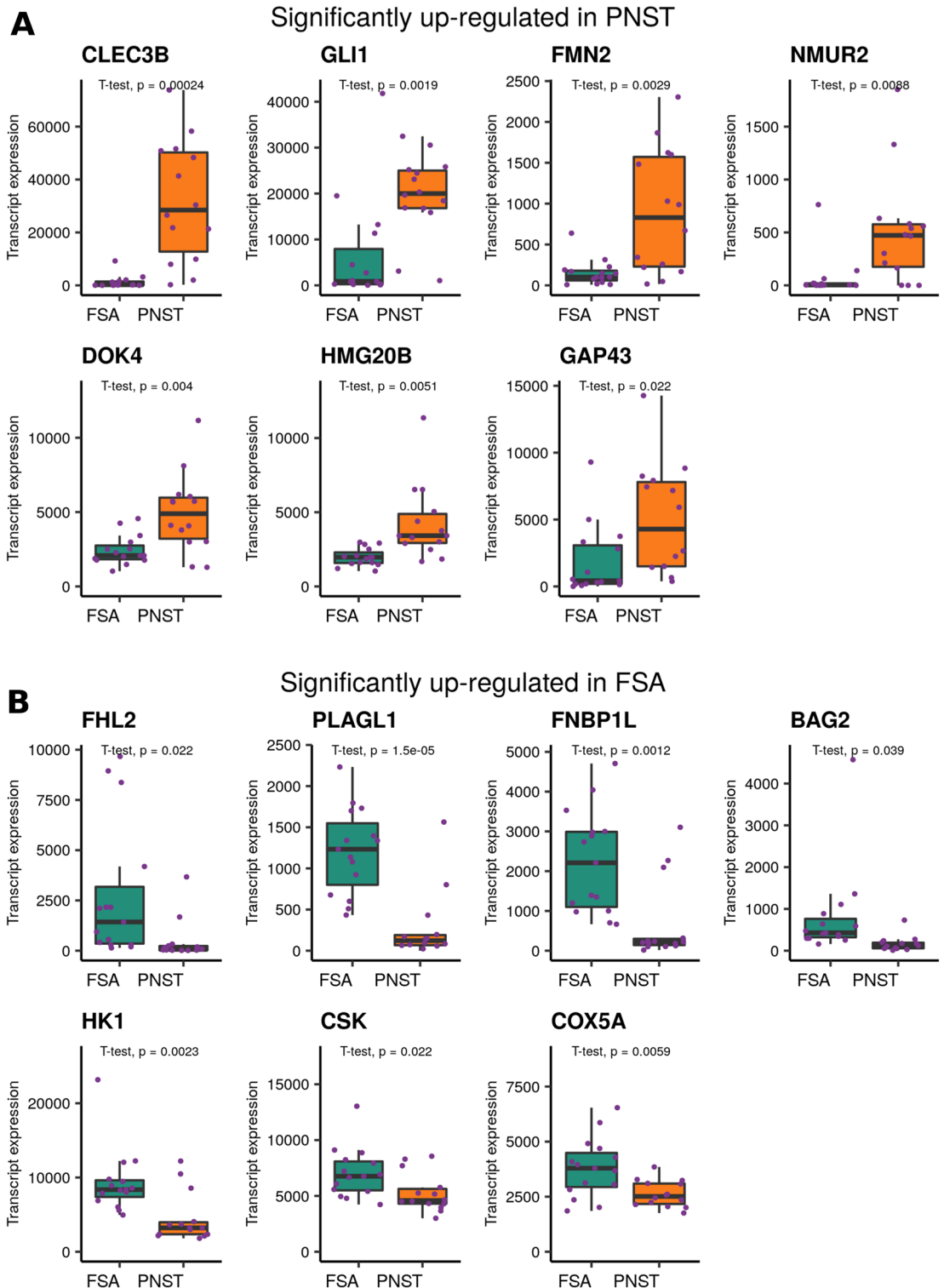


Figure 5. Expression levels of previously identified PNST and FSA marker genes. **(A)** Transcript expression boxplots of seven previously identified PNST marker genes. These genes had significantly higher transcript expression in canine PNST samples when compared to FSA samples. **(B)** Transcript expression boxplots of previously identified FSA marker genes with significantly higher expression in FSA samples in comparison to PNST samples. The bar in boxplots represents median transcript expression. Expression levels plotted here are normalized count data. The sample T1923 (UPS) was included with in the FSA group.

samples (n = 15) (Fig. 5B). The FSA group in this analysis also included one histologically similar UPS sample (T1923).

Immunogenomic profiling. The RNAseq transcriptome data was used to generate an immunogenomic profile of the STS tumors and compared to immunohistochemical staining in STS tumors (n = 27) with the pan-T cell marker CD3. The median percent tumor area staining positive for CD3 + T cells was 0.135%, ranging from 0.006 in T479 (H2 cluster) to 7.557% in T1923 (H4 cluster). This high immune cell infiltrate in T1923, along with a mixed morphology of fibroblast-like and Iba-1/CD204 staining histiocytoid-like tumor cells is consistent with the diagnosis of “undifferentiated pleomorphic sarcoma (UPS)”. There was a significant difference in CD3 + T cell infiltrates across the 4 clusters of tumors (Kruskal-Wallis test, p-value 0.024) with elevated staining in cluster H4 samples compared to H1 cluster samples (Benjamini-Hochberg adjusted p-value 0.022), but not the H2 and H3 clusters (Fig. 6A, Supplementary Fig. S16). Using Pearson correlations, we identified 319 and 8 genes positively and negatively correlated with the CD3 + immune score, respectively (Supplementary Table S11). The eight genes negatively correlated to CD3 staining were: *LMNA*, *RGS12*, *KIF7*, *LPAR2*, *LRRC4B*, *OBSL1*, *NAT14*, and *MAMDC4*. Of the 319 genes positively correlated to the IHC immune score, *IDO1*, *TNFRSF14*, *BTLA*, *TIGIT*, *PDCD1LG2*, *TNFRSF4*, *PDCD1*, *ICOS*, *CTLA4* and *CD274* are actionable immune checkpoint molecule targets approved or under clinical evaluation in human immuno-oncology studies. Enriched KEGG pathways linked with these genes included Cytokine-cytokine receptor interaction, T cell receptor signaling, PD-L1 expression and PD-1 checkpoint pathway in cancer, and B cell receptor signaling (Supplementary Fig. S17).

Additionally, CibersortX gene expression deconvolution method was used to characterize tumor infiltrating immune cells²¹. The most abundant cell types infiltrating STS tumors were naïve B cells (n = 12), macrophages M2 (n = 5), macrophages M0 (n = 4), dendritic cells resting (n = 3), Plasma cells (n = 2, T cells CD4 memory resting (n = 2), and mast cells activated (n = 1) (Fig. 6B). Additionally, consistent with CD3 + T cell immunostaining results, samples from cluster H4 had a higher fraction of CD8 T cells compared to samples in the remaining clusters.

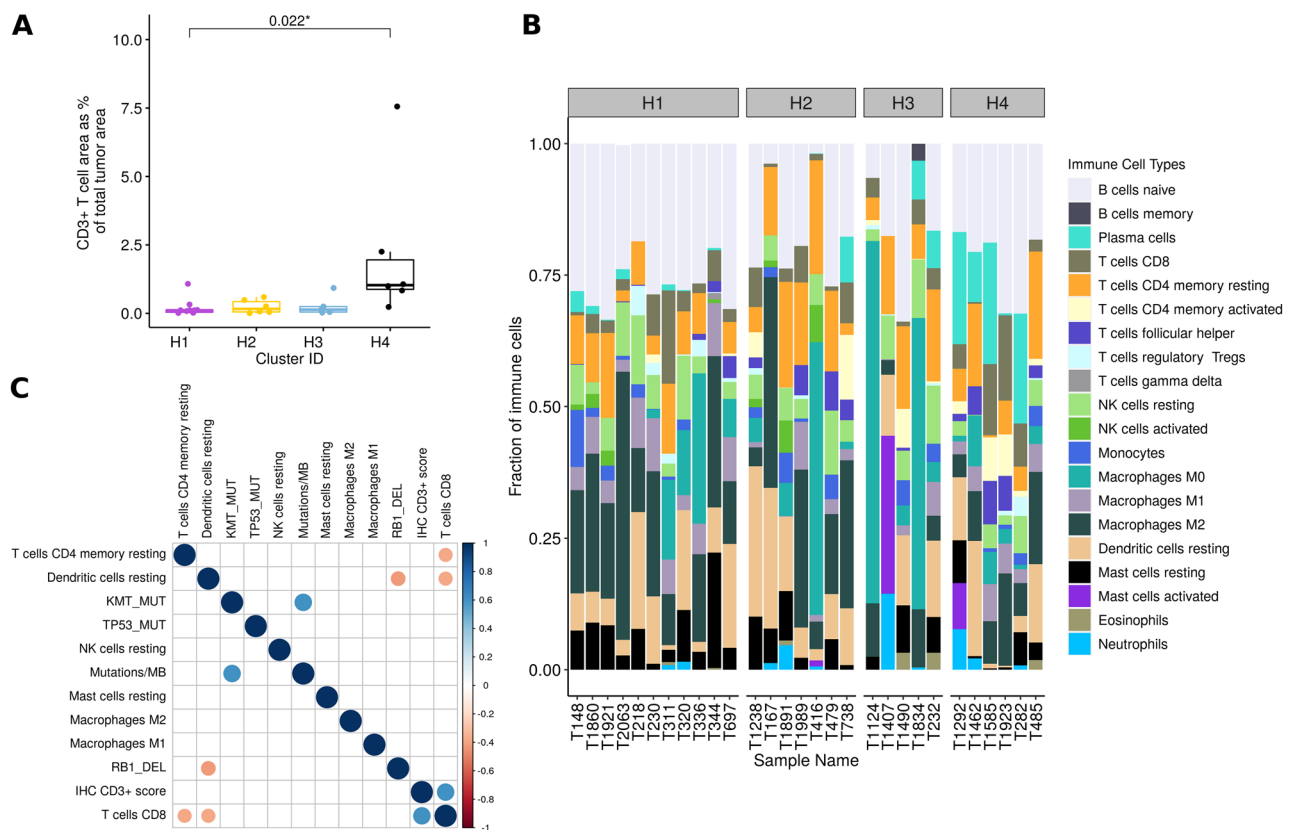


Figure 6. Immunogenomic profile of STS tumors. **(A)** Barplot showing distribution of CD3 + T-cell density in each cluster, as determined by immunohistochemistry. The samples in the H4 cluster had a significantly higher density of intra-tumoral CD3 + T cells. **(B)** Deconvolution of tumor microenvironment using CIBERSORT. The relative abundance of 20 cell types in each sample were plotted based on gene expression profiles. **(C)** Correlation matrix of immune cell type scores generated using CIBERSORT, genomic variants, and immunohistochemical derived CD3 + T-cell density (IHC CD3 + score). Only the significant Pearson correlation coefficients were plotted. The CD3 + T cell IHC scores of 29 STS samples positively correlated with CIBERSORT relative frequencies of CD8 T-cells.

To identify significant monotonic associations between the immune cell score, IHC score, mutations/MB, and genomic variables, Pearson rank correlation coefficients were computed (Fig. 6C). The genomic variables included samples with and without KMT gene mutations (KM_MUT), TP53 mutations (TP53_MUT, RB1 deletions (RB1_DEL). The CD3 + immune score positively correlated with gene expression scores for T cell CD8. There was an inverse correlation between the gene expression score for dendritic cell resting and scores for T cells CD8 and RB1_DEL. Increased mutational burden (mutations/MB) significantly correlated with *KMT_MUT* but not with any of the immune scores.

Discussion

This study represents the first comprehensive genomic analysis of a cohort of 29 canine STS reporting on the whole exome and RNA-sequencing profiles of fibrosarcoma and peripheral nerve sheath tumor (PNST) subtypes. From the transcriptome analysis, we identified 4 clusters (H1–H4) with distinct enriched biological features. We maintained this 4-cluster framework for downstream comparisons. Using WES and RNAseq gene fusion analysis we identified putative driver mutations associated with each of these STS subtypes, evaluated CNVs associated with transcript changes, and performed immune profiling of these clusters. Overall, this analysis begins to identify critical similarities and differences in the potential genetic drivers between these canine soft tissue sarcoma subtypes and their human counterparts.

Mutational signatures using relative frequencies of single nucleotide substitutions have been used to elucidate the underlying cancer etiology¹⁸. Although mathematical association of mutational spectrum rates to various signatures is possible using WES data, mechanistic understanding of cancer causation needs experimental validation²². In this study, a majority (52%) of canine STS exhibited the SBS5-like signature which is an age-related or clock-like human signature. In comparison, approximately 53% of human STS had signatures attributed to SBS5, while an additional 37% were attributed to another clock-like signature, SBS1²³. In this human data set, 2 tumors with the highest mutational burden were associated with the SBS6 defective mismatch repair signature. Eleven of the 29 canine STS (38%) were designated with the SBS6-like signature. There was no difference in the age at diagnosis or TMB between the different signatures, nor was there a correlation between age at diagnosis and mutations/MB for the age-related signature SBS5. Of the 3 tumors bearing the tobacco chewing signature, SBS29 (high C > A substitutions), 2 were localized to the mouth, suggesting that a similar environmental exposure could be involved. Interestingly, both tumors were in the H4 cluster with elevated immune system-related genes which might link this signature to an inflammatory response.

We identified recurrent somatic mutations in *TP53* and *KMT2D* in 21% of the tumors. There was a positive correlation between mutations in KMT family genes and mutational burden, where *KMT2D* mutated tumors in the H1 cluster had two of the highest mutational rates. Loss of *KMT2D* function in human tumors has been associated with high mutational burden and immune checkpoint blockade response²⁴. Previously, Sanger sequencing of amplified exons 5–8 of *TP53* revealed missense mutations in 20% of canine STS samples, with 3 of 7 variants identified in PNST samples and no variants were identified among 3 fibrosarcomas tested⁸. Southern blot analysis identified *MDM2* amplification in 3 of 7 PNST. The previous WES study of 10 STS tumors⁷, including 7 PNSTs, did not identify recurrent cancer gene variants with *MDM4* amplification in one tumor as the only *TP53* pathway variant. Similarly, among the 14 histologically identified PNST tumors, we did not identify any recurrent mutations. An overall look at the P53/cell cycle pathway revealed that 66% of the STS tumors had dysregulation of *TP53*, *MDM2*, *CDKN2A* or *RB1* (Supplementary Fig. S12). Six tumors had variants in *TP53* with 5 missense mutations in FSAs, and a single PNST with a stop-gained variant. CNV evaluation identified *RB1* loss in 10 of the 29 tumors, but only modest reductions in transcript expression. Eight tumors had < 10% of average *CDKN2A* transcript levels; seven tumors were FSAs with 5 grouped in H2 cluster and significant *CDKN2A* homodeletion was identified in 4 FSAs. As seen in most human cancers, loss of *CDKN2A* transcript expression was generally exclusive of *RB1* loss, while three tumors had both *TP53* variants and *RB1* loss and one tumor had *MDM2* overexpression with *RB1* loss²⁵. Finally, SBDS ribosome maturation factor (*SBDS*) CNV loss was identified in 45% of the tumors. Germline mutations in this gene lead to the Shwachman-Diamond syndrome and increased cancer incidence possibly due to SBDS regulation of *TP53*²⁶. Of the 10 tumors without pathway modifications, 6 were in the PNST H1 cluster (54%), indicating that for this cluster other pathways are involved in pathogenesis.

Given the prevalence of gene fusions in subsets of human soft tissue sarcomas, we analyzed the RNAseq data to identify fusions that may play a role in tumor progression. Notable, among these fusions are in-frame fusions of the collagen genes (exon 6 *COL1A2*, exon 20 *COL1A2* and exon 35 *COL3A1*) with exon 2 of *PDGFB* in tumors of the H2 fibrosarcoma cluster which was associated with extracellular matrix and G-protein coupled receptor signaling pathways. Similar fusions of *COL1A1*-*PDGFB* have been identified in two human STS subtypes: dermatofibrosarcoma protuberans (DFSP) and giant cell fibroblastoma; and increased expression of *PDGFB* is considered predictive of therapeutic response to PDGFR inhibitors¹⁴, although resistance subsequently develops^{27,28}. However, we found that cultured STS cells bearing a *COL3A1*-*PDGFB* fusion were only responsive to the PDGFR inhibitor, imatinib, at concentrations in the molar range, similar to a cell line without the *PDGFB* fusion or other targets for this multi-target tyrosine kinase. Since canine PDGFRA/B are 96% and 89% identical to the human proteins, lack of response does not seem to be due to species specificity. In fact, similar results have been observed in cultures of human DFSP cells²⁹, suggesting that therapeutic responses in these tumors may rely on inhibition of tumor-stroma interactions³⁰. Finally, losses of *CDKN2A* in these cell lines or *CDKN2A*/*RB1* in these tumors in the H2 group may also contribute to tumor development and modulate drug responses.

Malignant PNST in humans often develops subsequent to neurofibromatosis type I in which inactivation of the NF1 tumor suppressor gene (TSG) results in activation of the MAP kinase signaling pathway³¹. Consequently, MEK inhibitors have been approved for treatment of manifestations of this syndrome. In addition, a variety of other sporadic human tumors are also associated with inactivating mutations in NF1 including glioblastoma,

lung adenocarcinoma, acute myeloid leukemia, and ovarian and breast cancers³². However, only one *NF1* variant was identified in an H2 cluster sample histologically identified as a PNST. Similarly, the human Legius syndrome, caused by germline mutations in *SPRED1*, which can limit binding to *NF1*, has similar but milder neurofibromatosis type I symptoms³³ and an increased incidence in pediatric leukemias with elevated ERK signaling. Here we provide the first report of a *SPRED1* fusion gene in a canine FSA sample. *SPRED1* mutations are observed in about 2% of human cancers with *SPRED1* loss attributed as a driver in 37% of mucosal melanomas³⁴.

In a previous study, we utilized a MAP Kinase pathway activity (MPAS) score developed to predict MEK1/2 inhibitor sensitivity in human cancers across a panel of canine cancer cell lines^{20,35}. In the current study, MPAS scores > 0 were identified in 8 tumors. Plotted by hierarchical clusters, we found that none of the H1 PNST tumors had positive MPAS scores. In contrast, 8 tumors within the H2/H3/H4 FSA clusters had positive MPAS scores (Supplementary Fig. S18). These included tumors carrying oncogenic PDGFB fusions, the *SPRED1*-*FAM98B* fusion, EGFR T724M, and HRAS G13R. The other 2 tumors with positive MPAS values had no obvious MAP kinase activators. Surprisingly, the single tumor bearing *NF1* A699Ifs*3 did not have a positive MPAS score, however, the allelic frequency of this variant was only 0.1, so loss of this TSG may have been insufficient to promote measurable MAP kinase pathway activation compared to oncogenic activators. This same tumor also had a low frequency frameshift mutation in *MAP3K1* G394Kfs*28. Overall, this analysis indicates that, unlike human malignant PNST, inactivation of *NF1* and the resultant MAP kinase pathway activation are not characteristics of canine PNST, while some canine FSA exhibit MAP kinase activation in response to oncogenic drivers and may be amenable to treatment with targeted inhibitors.

Our RNA sequencing data validated 47% of the previously identified markers for canine FSA^{4,5}. This could be due to variability in analyses techniques and/or histological characterization of canine PNSTs. Similar to our study, a recent RNAseq study of canine STS subtypes, also showed elevated expression of *GLI* and *CLECB1* in PNST samples³⁶, but failed to confirm elevation of S100 and vimentin in PNST samples. Further studies at both transcript and protein levels are required to determine specific molecular markers for PNST.

With respect to the STS tumor micro-environment, the infiltrating CD3 + T-cell density as assessed in each tumor by IHC correlated with in silico quantified CD8 + T-cell abundance. Further, expression of several current actionable co-stimulatory or co-inhibitory immune checkpoint targets, including PDL1 and CTLA4, positively correlated with the number of infiltrating T-cells. These data are consistent with observations in murine pre-clinical models and human tumors wherein upregulation of inhibitory immune checkpoint molecule expression on CD8 + tumor-infiltrating lymphocytes (TILs) is a critical mechanism by which cancer cells can escape the antitumor immune response^{37–39}. Moreover, these data also suggest that, like human cancer patients, high T-cell infiltration of canine soft tissue sarcoma tumors as a determinant of probability of clinical response to cancer immunotherapies warrants further investigation^{40,41}. However, another marker of immunotherapy, tumor mutational burden, did not correlate with T cell density in this study. While several studies have shown improved responses to immune checkpoint inhibitors in tumors with high mutational burden^{42,43}, a recent pan-cancer study revealed no significant association between mutational burden and CD8 + T-cell infiltration; however, tumor histo-type specific associations do exist⁴⁴. Interestingly, most abundant tumor microenvironment cell type was B cells in 41% of STS canine samples. A recent gene expression profiling study of 608 human STS tumors identified B cells as the strongest prognostic factor associated with patient immunotherapy response and survival⁴⁵. A similar categorization of canine patients with STS can help classify cohorts that may positively respond to immunotherapy.

In summary, this study shows STS tumors are diverse with respect to transcriptomic, genomic and immunogenomic profiling. The H1 cluster of PNSTs, largely comprised of grade I and II tumors, had few variants, with 6 of 11 tumors having no identified putative SNV mutational drivers. However, copy number losses of *RB1* were observed in 4 of the 11 H1 tumors. The H2 cluster was notable for the presence of *PDGFB* gene fusions in 3 of the 7 tumors. Of the 6 tumors bearing *TP53* driver mutations, 50% were from cluster H3 (n = 5). The H4 cluster exhibited elevated immune activation, T-cell infiltration, and a lack of clock-like COSMIC signatures. MAP kinase pathway activation was only observed in clusters H2-H4 with 90% of the grade III tumors. Although prognosis for this cancer is improved by surgical intervention, overall survival of dogs with grade III tumors is poor compared to dogs with grade I tumors⁴⁶. Thus, the molecular and genetic driver characterization of canine STS subtypes may aid in diagnosis and the identification of novel therapies that improve the outcomes for canine and human patients with recurrent or metastatic disease.

Materials and methods

Tumor sample selection and categorization. Archived primary tumor samples with a diagnosis of soft tissue sarcoma were obtained from the Colorado State University (CSU) Flint Animal Cancer Center's tissue archive. Soft tissue sarcoma tumor and normal samples were collected and archived from dogs presenting to the Flint Animal Cancer Center during standard treatment in compliance with the CSU Institutional Animal Care and Use Committee protocol (KP 652) after owner consent was obtained. No experimental procedures were conducted on these dogs and this study had no impact on their treatment. Archived tumors samples were reviewed by a board-certified veterinary anatomic pathologist (DR) and selected based on 1) quality of the formalin-fixed paraffin embedded tissue, assessed by hematoxylin and eosin (H&E)-stained slide evaluation confirming adequate overall biopsy size and a high degree of viable tumor tissue, 2) availability of a corresponding fresh frozen tumor biopsy sample for RNA extraction, and 3) availability of normal tissue or whole blood for isolation of normal genomic DNA. Tumors were histologically graded as previously describe^{47,48} and given a diagnosis corresponding to one of two broad histologic subtypes based on patterns of cell growth: 1) fibrosarcoma, based on the presence of a solid or fasciculated growth pattern, or 2) malignant nerve sheath tumor/perivascular wall tumor, based on the predominance of any mixture of palisading, Antoni A or Antoni B, or concentric whorling

growth patterns. Additionally, those samples occurring in the oral cavity or diagnosed as pleomorphic/undifferentiated sarcoma were further screened and confirmed to be non-melanocytic or non-histiocytic in origin based on negative immunoreactivity for melanocytic (Melan A and PNL2) and histiocytic (Iba1 and CD204) antigens, respectively. Grading and diagnosis was performed in a blinded fashion without any a priori knowledge of transcriptomic clustering.

Canine tumor sample preparations. Soft tissue sarcoma (STS) tumors ($n = 29$) were collected along with matched normal stroma ($n = 29$) prior to treatment and stored at -80°C . Following the manufacturer's protocol for TRIzol (Invitrogen, Catalogue #15596026), genomic DNA was extracted from tumors and matched normals and RNA was extracted from tumor samples only. The DNA and RNA samples were purified using DNeasy or QiaAMP DNA Blood mini kits and RNeasy (Qiagen, Catalogue #69504, 51104, 74004), respectively, quantified on a NanoDrop Microvolume Spectrophotometer and quality was assessed by TapeStation or Bioanalyzer (Agilent). The total RNA was processed using Universal Plus™ mRNA-Seq library preparation kits with NuQuant® (product number: 0520-A01) and the resulting cDNA library was sequenced on Illumina NovaSeq6000 to generate 150 bp paired end reads. The SureSelectXT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library kit was used to create genomic DNA libraries that were also sequenced on a NovaSeq6000 generating 150 bp paired-end reads.

RNAseq pipeline. The quality of reads was assessed by FastQC resource (v 0.11.5, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Low quality reads and adaptors were eliminated from Illumina RNAseq reads with Trimmomatic (v0.36)⁴⁹. The paired end reads (45.8–64.5 million) were mapped against CanFam3.1 (Ensembl version 99) with STAR (v 2.6.1a)⁵⁰, quantified using HTSeq-count (v1.16.5)⁵¹, normalized with DESeq2 (v1.26.0) median of ratios method⁵², and log-transformed for plotting and downstream analyses. STAR aligner mapped 94% of trimmed reads to CanFam3.1 genome.

Pericyte marker genes. To distinguish between PWT and PNST at a molecular level, we used literature searches to catalog genes that are used as markers for pericytes^{16,53,54}. We also used the gene ontology terms, vascular associated smooth muscle cell proliferation (GO:1990874) and pericyte cell differentiation (GO:1904238) to curate additional markers for pericytes.

Clustering tumor samples. Unsupervised hierarchical clustering was used to group 29 soft tissue sarcomas into distinct clusters. A subset of genes ($n = 1926$) with mean expression $\log_2 > 2$ and mean variance $\log_2 > 5$ was used to cluster the tumor samples. The optimal number of hierarchical clusters was determined using the elbow method¹⁷. Spearman correlation and Ward's minimum variance method (Ward.D2) was used to calculate distance and cluster STS samples, respectively.

Differential expression. We ran four sets of pair-wise comparisons to identify differentially expressed genes (DEGs) between samples from each cluster and combined samples from the other clusters (e.g. H2 vs H1, H3, H4) using DESeq2 in R environment⁵². Briefly, the raw counts for all genes were filtered to retain genes with at least 4 reads in at least 4 samples which resulted in 18,201 genes that were processed for downstream analysis. Following normalization of raw counts, the DEGs were identified using DESeq function. The DEGs with fold change of $\log_2 > 2$ and adjusted p-value < 0.05 were used as input for identifying pathways that were up-regulated in each cluster. The functional annotation of genes was done using g:Profiler R package (v0.2.1) and sources included KEGG, Reactome, WikiPathways, and Gene Ontology databases^{55,56}. The pathways with false discovery rate of < 0.05 were considered enriched.

Whole exome sequencing pipeline. The WES Illumina reads were processed to eliminate low quality reads and adapters using Trimmomatic and mapped against CanFam3.1 genome (https://may2021.archive.ensembl.org/Canis_lupus_familiaris/Info/Index) via BWA (v 0.7.15)⁵⁷.

Somatic simple variant identification. Duplicate reads were marked with Picard tools (v1.119) and somatic variants were called using GATK Mutect2 resource (v4.1.2.0)⁵⁸. Germline variants were eliminated using a panel of normals created from 118 normal samples following the GATK pipeline. Additionally, ~90 million population variants, called from 722 dogs, were used as the germline resource option within Mutect2 when calling somatic variants⁵⁹. The remaining variants were processed using the filterMutectCalls GATK function and variants with a PASS notation in the FILTER column were characterized as somatic variants. The VCF files with somatic variants were converted to MAF (Mutation Annotation File) format (https://docs.gdc.cancer.gov/Data/File_Formats/MAF_Format/) using the perl code: vcf2maf.pl (<https://github.com/mskcc/vcf2maf>). Each variant was mapped to only one of all possible gene transcripts using the “canonical” isoform from Ensembl database (v99). The variants within cancer genes, as curated in COSMIC (v94) and OncoKB databases¹⁹ were extracted from the list of somatic variants and putative drivers were identified as variants homologous to human cancer gene variants considered oncogenic/likely oncogenic. Pathway analysis of the mutated genes was conducted using g:profiler R package using hypergeometric tests to determine significant pathway/gene ontology terms. The mutational signatures of each sample were deduced by using the MutationalPatterns R package⁶⁰.

Somatic copy number variant identification. The allele-specific copy number variant (CNV) genes were identified using Sequenza (v2.1.9999b1)⁶¹. Allele frequencies were used to calculate tumor normal depth

ratio, allele-specific segmentation using the copynumber R package. The resulting segmentation file was annotated and genes with recurrent CNVs were identified with GISTIC2.0 (v2.0.23; q -value ≤ 0.05). Genes with amplitude changes significantly correlated with transcript expression levels (Pearson correlation) were identified and cancer genes (OncoKB) were selected for plotting and cross-species comparison.

Homozygous deletion identification. Genes with homodeletion in 29 STS tumor samples (Supplementary Fig. S13), were identified using VarScan2 (v2.4.3) from the whole exome sequence data⁶². The candidate homozygous deletions were called using *copyCaller* function and by specifying the option *-output-homdel-file*.

Cell line validation and drug sensitivity Assays. Canine cell lines (C2, STSA-1, and FACC-19-CSTS36) were confirmed to be canine, mycoplasma free, and STR fingerprinted using PCR-based protocols (Supplementary Table S12)^{63–65}. The cell lines were plated on 96-well plates at 2000 cells/well in DMEM with 10% FBS. Following adherence for 24 h, cells were treated with serial dilutions of imatinib (Gleevec, 100 mM to 1.7 nM, 3–5 wells/concentration) or DMSO vehicle. After 72 h, cell confluence was monitored on the IncuCyte ZOOM Live-Cell Analysis System (Essen BioScience). Cell number was normalized to time zero readings and fit to 4 parameter non-linear curves of log dose versus percent of control and LD50 values were interpolated as the dose at which cell number was 50% of control.

Immunogenomic profile. In silico scores for 22 of immune cell types were extracted and deconvoluted using CIBERSORT to cross reference against immunohistochemical staining of tumor-infiltrating CD3 + T-cells²¹. For immunohistochemical and H&E staining, archived, formalin-fixed, paraffin embedded (FFPE) tumor samples with a diagnosis of soft tissue sarcoma were obtained from either the Colorado State University Flint Animal Cancer Center Tissue Archive or Veterinary Diagnostic Laboratory, and were sectioned at five microns. H&E-stained slides were evaluated by a board-certified pathologist (DPR) to confirm diagnosis and the presence of adequate viable tumor tissue for analysis. Immunolabeling was performed via routine, automated methods on the Leica Bond Max autostainer (Leica Biosystems Inc.) as previously described, using a mouse monoclonal anti-human CD3 (pan T lymphocyte marker; Leica, clone LN10, ready-to-use format). Antigen retrieval was performed using Leica Epitope Retrieval 2 (Tris-EDTA buffer, pH 9) for 20 min. Primary antibody incubation was carried out at room temperature for 30 min, followed by detection with PowerVision IHC detection systems (Leica Biosystems, Inc.), using a polymeric horseradish peroxidase anti-mouse IgG and DAB chromogen. Slides were counterstained with hematoxylin. Whole slide images of IHC stained slides were digitally captured using an Olympus IX83 microscope at 10X magnification and fixed exposure times for all samples. Quantitative image analysis was performed using ImageJ software (National Institutes of Health, NIH) as previously described⁶⁶. Briefly, whole slide images were converted to gray scale, and tumor tissue regions-of-interest were segmented from adjacent normal tissue or section artifacts by manual outlining in ImageJ by a veterinary pathologist. Positively labeled infiltrating immune cells were counted using the ImageJ color deconvolution plugin (Fast Red, Fast Blue, and DAB vector), with a minimum positive particle size threshold of 10 pixels and a positive pixel threshold for all immune cell markers set based on two standard deviations above the mean of appropriate isotype-stained control slides and confirmed visually by a veterinary pathologist. Images were subjected to global, automated application of this intensity and size threshold to all images. Following image analysis, positive pixel masks of each image were evaluated by a pathologist to ensure thresholding and counting accuracy. Data was analyzed and the number of infiltrating immune cells was expressed as total positive cells per mm² of tumor tissue area.

MPAS score assessment. The MAPK Pathway Activation Score (MPAS) quantifies relative MAPK activity using gene expression data. The expression of 10 downstream MAPK pathway targets (CCND1, DUSP4, DUSP6, EPHA2, EPHA4, ETV4, ETV5, PHLDA1, SPRY2, SPRY4) were used to calculate the score $MPAS = (\text{sum of } z\text{-score normalized expression for MPAS genes}) / (\sqrt{10})$. MPAS score has been shown to correlate with sensitivity to MAPK inhibitors in both cancer cell lines and cancer patients³⁵.

Statistical analysis and data potting. All statistical analysis were done using R statistical tool (v3.6.1). Significantly different groups of tumors with T-cell infiltration were analyzed using non-parametric Kruskal-Wallis test. For correlation analysis, Pearson correlation coefficient were calculated and FDRtool R package was used for correcting multiple testing of p-values. The oncoplots and heatmaps were plotted using ComplexHeatmap R package⁶⁷. Barcharts and boxplots were plotted using R package ggplot2⁶⁸. The lollipop plots were created by Lollipops software (v1.3.2) using canine uniprot ID⁶⁹.

Omics data processing. All the analyses were run on Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University.

Consent to participate. Dog samples were archived during clinical treatment by the Flint Animal Cancer Center Biorepository with owner consent and the Colorado State University Institutional Animal Care and Use Committee protocol (KP 652) approved the study.

Data availability

The tumor and matched normal Illumina reads from WES were submitted to SRA database under Bioproject ID PRJNA882864. The RNAseq sequencing reads as well as raw count data for CanFam3.1 genes from 29 samples were submitted to GEO database (accession number: GSE214154). The list of commands, tools, R packages and their versions are available at <https://github.com/sdas2019/Canine-soft-tissue-sarcoma-omics-pipeline>.

Received: 1 March 2023; Accepted: 19 June 2023

Published online: 27 June 2023

References

- Gustafson, D. L., Duval, D. L., Regan, D. P. & Thamm, D. H. Canine sarcomas as a surrogate for the human disease. *Pharmacol Ther* **188**, 80–96. <https://doi.org/10.1016/j.pharmthera.2018.01.012> (2018).
- Forrest, L. J., Chun, R., Adams, W. M., Cooley, A. J. & Vail, D. M. Postoperative radiotherapy for canine soft tissue sarcoma. *J Vet Intern Med* **14**, 578–582. [https://doi.org/10.1892/0891-6640\(2000\)014%3c0578:prfcst%3e2.3.co;2](https://doi.org/10.1892/0891-6640(2000)014%3c0578:prfcst%3e2.3.co;2) (2000).
- Chijiwa, K., Uchida, K. & Tateyama, S. Immunohistochemical evaluation of canine peripheral nerve sheath tumors and other soft tissue sarcomas. *Vet Pathol* **41**, 307–318. <https://doi.org/10.1354/vp.41-4-307> (2004).
- Klopfleisch, R., Meyer, A., Lenze, D., Hummel, M. & Gruber, A. D. Canine cutaneous peripheral nerve sheath tumours versus fibrosarcomas can be differentiated by neuroectodermal marker genes in their transcriptome. *J Comp Pathol* **148**, 197–205. <https://doi.org/10.1016/j.jcpa.2012.06.004> (2013).
- Meyer, A. & Klopfleisch, R. Multiple polymerase chain reaction markers for the differentiation of canine cutaneous peripheral nerve sheath tumours versus canine fibrosarcomas. *J Comp Pathol* **150**, 198–203. <https://doi.org/10.1016/j.jcpa.2013.08.006> (2014).
- Boerkamp, K. M. *et al.* Unclassified sarcomas: A study to improve classification in a cohort of Golden Retriever dogs. *J. Vet. Diagn. Invest.* **28**, 623–631. <https://doi.org/10.1177/1040638716660130> (2016).
- Roberts, N. J. *et al.* Intratumoral injection of Clostridium novyi-NT spores induces antitumor responses. *Sci. Transl. Med.* **6**, 249ra111. <https://doi.org/10.1126/scitranslmed.3008982> (2014).
- Nasir, L., Rutteman, G. R., Reid, S. W., Schulze, C. & Argyle, D. J. Analysis of p53 mutational events and MDM2 amplification in canine soft-tissue sarcomas. *Cancer Lett.* **174**, 83–89. [https://doi.org/10.1016/s0304-3835\(01\)00637-1](https://doi.org/10.1016/s0304-3835(01)00637-1) (2001).
- Evans, D. G. *et al.* Malignant peripheral nerve sheath tumours in neurofibromatosis 1. *J. Med. Genet.* **39**, 311–314. <https://doi.org/10.1136/jmg.39.5.311> (2002).
- Brohl, A. S., Kahen, E., Yoder, S. J., Teer, J. K. & Reed, D. R. The genomic landscape of malignant peripheral nerve sheath tumors: Diverse drivers of Ras pathway activation. *Sci. Rep.* **7**, 14992. <https://doi.org/10.1038/s41598-017-15183-1> (2017).
- Aguirre-Hernandez, J. *et al.* Disruption of chromosome 11 in canine fibrosarcomas highlights an unusual variability of CDKN2B in dogs. *BMC Vet. Res.* **5**, 27. <https://doi.org/10.1186/1746-6148-5-27> (2009).
- Lannon, C. L. & Sorensen, P. H. ETV6-NTRK3: A chimeric protein tyrosine kinase with transformation activity in multiple cell lineages. *Semin. Cancer Biol.* **15**, 215–223. <https://doi.org/10.1016/j.semcancer.2005.01.003> (2005).
- Arbajian, E. *et al.* In-depth genetic analysis of sclerosing Epithelioid fibrosarcoma reveals recurrent genomic alterations and potential treatment targets. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **23**, 7426–7434. <https://doi.org/10.1158/1078-0432.CCR-17-1856> (2017).
- Mertens, F., Antonescu, C. R. & Mitelman, F. Gene fusions in soft tissue tumors: Recurrent and overlapping pathogenetic themes. *Genes Chromosom. Cancer* **55**, 291–310. <https://doi.org/10.1002/gcc.22335> (2016).
- Gambo, A. C., Gronchi, A. & Cardona, K. Soft-tissue sarcoma in adults: An update on the current state of histiotype-specific management in an era of personalized medicine. *CA Cancer J. Clin.* **70**, 200–229. <https://doi.org/10.3322/caac.21605> (2020).
- Chang, L., Scott, M. A., Meyers, C. A. & James, A. W. Pericytes in sarcomas and other mesenchymal tumors. *Adv. Exp. Med. Biol.* **1147**, 109–124. https://doi.org/10.1007/978-3-030-16908-4_4 (2019).
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. & Satoto, B. D. Integration K-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conf. Ser. Mater. Sci. Eng.* <https://doi.org/10.1016/j.patcog.2012.07.021> (2018).
- Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101. <https://doi.org/10.1038/s41586-020-1943-3> (2020).
- Chakravarty, D. *et al.* OncoKB: A precision oncology knowledge base. *JCO Precis. Oncol.* <https://doi.org/10.1200/PO.17.00011> (2017).
- Das, S., Idate, R., Cronise, K., Gustafson, D. & Duval, D. Identifying candidate druggable targets in canine cancer cell lines using whole-exome sequencing. *Mol. Cancer Ther.* <https://doi.org/10.1158/1535-7163.MCT-18-1346> (2019).
- Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782. <https://doi.org/10.1038/s41587-019-0114-2> (2019).
- Koh, G., Degasperi, A., Zou, X., Momen, S. & Nik-Zainal, S. Mutational signatures: Emerging concepts, caveats and clinical applications. *Nat. Rev. Cancer* **21**, 619–637. <https://doi.org/10.1038/s41568-021-00377-7> (2021).
- Abeshouse, A. *et al.* Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell* **171**, 950–965. <https://doi.org/10.1016/j.cell.2017.10.014> (2017).
- Wang, G. *et al.* CRISPR-GEMM pooled mutagenic screening identifies KMT2D as a major modulator of immune checkpoint blockade. *Cancer Discov.* **10**, 1912–1933. <https://doi.org/10.1158/2159-8290.CD-19-1448> (2020).
- Knudsen, E. S. *et al.* Pan-cancer molecular analysis of the RB tumor suppressor pathway. *Commun. Biol.* **3**, 158. <https://doi.org/10.1038/s42003-020-0873-9> (2020).
- Hao, Q. *et al.* Dual regulation of p53 by the ribosome maturation factor SBDS. *Cell Death Dis.* **11**, 197. <https://doi.org/10.1038/s41419-020-2393-4> (2020).
- Rutkowski, P. *et al.* Long-term results of treatment of advanced dermatofibrosarcoma protuberans (DFSP) with imatinib mesylate - The impact of fibrosarcomatous transformation. *Eur. J. Surg. Oncol.* **43**, 1134–1141. <https://doi.org/10.1016/j.ejso.2017.03.011> (2017).
- Stacchiotti, S. *et al.* Efficacy and biological activity of imatinib in metastatic dermatofibrosarcoma protuberans (DFSP). *Clin. Cancer Res.* **22**, 837–846. <https://doi.org/10.1158/1078-0432.CCR-15-1243> (2016).
- Oyama, R. *et al.* Establishment of novel patient-derived models of dermatofibrosarcoma protuberans: Two cell lines, NCC-DFSP1-C1 and NCC-DFSP2-C1. *Vitro cell. Dev. Biol. Anim.* **55**, 62–73. <https://doi.org/10.1007/s11626-018-0305-z> (2019).
- Shah, S. & Sizemore, G. M. Diverse roles of tumor-stromal PDGFB-to-PDGFRbeta signaling in breast cancer growth and metastasis. *Adv. Cancer Res.* **154**, 93–140. <https://doi.org/10.1016/bs.acr.2022.01.003> (2022).
- de Blank, P. M. K. *et al.* MEK inhibitors for neurofibromatosis type 1 manifestations: Clinical evidence and consensus. *Neuro Oncol.* **24**, 1845–1856. <https://doi.org/10.1093/neuonc/noac165> (2022).
- Philpott, C., Tovell, H., Frayling, I. M., Cooper, D. N. & Upadhyaya, M. The NF1 somatic mutational landscape in sporadic human cancers. *Hum. Genomics* **11**, 13. <https://doi.org/10.1186/s40246-017-0109-3> (2017).

33. Pasmant, E. *et al.* SPRED1 germline mutations caused a neurofibromatosis type 1 overlapping phenotype. *J. Med. Genet.* **46**, 425–430. <https://doi.org/10.1136/jmg.2008.065243> (2009).
34. Ablain, J. *et al.* Human tumor genomics and zebrafish modeling identify SPRED1 loss as a driver of mucosal melanoma. *Science* **362**, 1055–1060. <https://doi.org/10.1126/science.aau6509> (2018).
35. Wagle, M. C. *et al.* A transcriptional MAPK pathway activity score (MPAS) is a clinically relevant biomarker in multiple cancer types. *NPJ Precis. Oncol.* **2**, 7. <https://doi.org/10.1038/s41698-018-0051-4> (2018).
36. Lam, L. *et al.* Comparative whole transcriptome analysis of gene expression in three canine soft tissue sarcoma types. *PloS One* **17**, e0273705. <https://doi.org/10.1371/journal.pone.0273705> (2022).
37. Li, H. *et al.* Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. *Cell* **176**, 775–789 e718. <https://doi.org/10.1016/j.cell.2018.11.043> (2019).
38. Thommen, D. S. *et al.* A transcriptionally and functionally distinct PD-1(+) CD8(+) T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nat. Med.* **24**, 994–1004. <https://doi.org/10.1038/s41591-018-0057-z> (2018).
39. Fehlings, M. *et al.* Checkpoint blockade immunotherapy reshapes the high-dimensional phenotypic heterogeneity of murine intratumoural neoantigen-specific CD8(+) T cells. *Nat. Commun.* **8**, 562. <https://doi.org/10.1038/s41467-017-00627-z> (2017).
40. Fridman, W. H., Pages, F., Sautes-Fridman, C. & Galon, J. The immune contexture in human tumours: Impact on clinical outcome. *Nat. Rev. Cancer* **12**, 298–306. <https://doi.org/10.1038/nrc3245> (2012).
41. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571. <https://doi.org/10.1038/nature13954> (2014).
42. Yarchoan, M., Hopkins, A. & Jaffee, E. M. Tumor mutational burden and response rate to PD-1 inhibition. *N. Engl. J. Med.* **377**, 2500–2501. <https://doi.org/10.1056/NEJMc1713444> (2017).
43. Maleki Vareki, S. High and low mutational burden tumors versus immunologically hot and cold tumors and response to immune checkpoint inhibitors. *J. Immunother. Cancer* **6**, 157. <https://doi.org/10.1186/s40425-018-0479-7> (2018).
44. Varn, F. S., Wang, Y., Mullins, D. W., Fiering, S. & Cheng, C. Systematic pan-cancer analysis reveals immune cell interactions in the tumor microenvironment. *Can. Res.* **77**, 1271–1282. <https://doi.org/10.1158/0008-5472.CAN-16-2490> (2017).
45. Petitprez, F. *et al.* B cells are associated with survival and immunotherapy response in sarcoma. *Nature* **577**, 556–560. <https://doi.org/10.1038/s41586-019-1906-8> (2020).
46. Paulos, P. E. *et al.* Outcome with surgical treatment of canine soft tissue sarcoma in the region of the ischiatic tuberosity: A veterinary society of surgical oncology retrospective study. *Vet. Comp. Oncol.* **20**, 669–678. <https://doi.org/10.1111/vco.12821> (2022).
47. Kuntz, C. A. *et al.* Prognostic factors for surgical treatment of soft-tissue sarcomas in dogs: 75 cases (1986–1996). *J. Am. Vet. Med. Assoc.* **211**, 1147–1151 (1997).
48. McSparran, K. D. Histologic grade predicts recurrence for marginally excised canine subcutaneous soft tissue sarcomas. *Vet. Pathol.* **46**, 928–933. <https://doi.org/10.1354/vp.08-VP-0277-M-FL> (2009).
49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30**, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> (2014).
50. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21. <https://doi.org/10.1093/bioinformatics/bts635> (2013).
51. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169. <https://doi.org/10.1093/bioinformatics/btu638> (2015).
52. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8> (2014).
53. Mentzel, T., Dei Tos, A. P., Sapi, Z. & Kutzner, H. Myopericytoma of skin and soft tissues: Clinicopathologic and immunohistochemical study of 54 cases. *Am. J. Surg. Pathol.* **30**, 104–113. <https://doi.org/10.1097/01.pas.0000178091.54147.b1> (2006).
54. Meguro, S. *et al.* Phenotypic characterization of perivascular myoid cell neoplasms, using myosin 1B, a newly identified human pericyte marker. *Hum. Pathol.* **62**, 187–198. <https://doi.org/10.1016/j.humpath.2016.12.024> (2017).
55. Raudvere, U. *et al.* g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198. <https://doi.org/10.1093/nar/gkz369> (2019).
56. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
57. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **26**, 589–595. <https://doi.org/10.1093/bioinformatics/btp698> (2010).
58. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11–10. <https://doi.org/10.1002/0471250953.bi1110s43> (2013).
59. Plassais, J. *et al.* Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* **10**, 1489. <https://doi.org/10.1038/s41467-019-09373-w> (2019).
60. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: Comprehensive genome-wide analysis of mutational processes. *Genome Med.* <https://doi.org/10.1186/s13073-018-0539-0> (2018).
61. Favero, F. *et al.* Sequenza: Allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70. <https://doi.org/10.1093/annonc/mdu479> (2015).
62. Koboldt, D. C. *et al.* VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics (Oxford, England)* **25**, 2283–2285. <https://doi.org/10.1093/bioinformatics/btp373> (2009).
63. Cooper, J. K. *et al.* Species identification in cell culture: A two-pronged molecular approach. *Vitro Cell. Dev. Biol. Anim.* **43**, 344–351. <https://doi.org/10.1007/s11626-007-9060-2> (2007).
64. Uphoff, C. C. & Drexler, H. G. Detection of mycoplasma contamination in cell cultures. *Curr. Protoc. Mol. Biol.* **106**, 28. <https://doi.org/10.1002/0471142727.mb2804s106> (2014).
65. O'Donoghue, L. E., Rivest, J. P. & Duval, D. L. Polymerase chain reaction-based species verification and microsatellite analysis for canine cell line validation. *J. Vet. Diagn. Investig.* **23**, 780–785. <https://doi.org/10.1177/1040638711408064> (2011).
66. Regan, D. P., Escaffi, A., Coy, J., Kurihara, J. & Dow, S. W. Role of monocyte recruitment in hemangiosarcoma metastasis in dogs. *Vet. Comp. Oncol.* **15**, 1309–1322. <https://doi.org/10.1111/vco.12272> (2017).
67. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics (Oxford, England)* **32**, 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313> (2016).
68. Wickham, H. *Ggplot2: Elegant graphics for data analysis* (Springer, 2009).
69. Jay, J. J. & Brouwer, C. Lollipops in the clinic: Information dense mutation plots for precision medicine. *PloS One* **11**, e0160519. <https://doi.org/10.1371/journal.pone.0160519> (2016).

Acknowledgements

We would like to thank Irene Mok and Lindsay Carroll for Tissue Archive curation. We acknowledge the help of Marina Schlaepfer and Rujuta Idate in screening variants for putative driver mutations.

Author contributions

S.D. and D.L.D. have made substantial contributions to the conception, design of the work, interpretation of data, drafted the work and substantively revised it. R.I. contributed towards the sample acquisition and processing. S.L. substantially contributed towards sample acquisition. D.L.D., S.L., and D.P.R. contributed towards funding for this project. D.P.R. contributed towards writing and revising drafts.

Funding

Funding was provided by Anschutz Foundation (PI: DLD, SEL). Additional funding includes The National Institutes of Health, Office of the Director, award number K01ODO22982 (PI: DPR) and National Center for Advancing Translational Sciences, award number L30 TR002126 (PI: DPR) and P30 CA046934 (University of Colorado Cancer Center Support Grant, Genomics and Microarray Shard Resource).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-37266-y>.

Correspondence and requests for materials should be addressed to D.L.D.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2023