Whole exome sequencing and gene expression analysis of canine osteosarcomas identify mutant TP53 and enriched immune pathways associated with longer survival

Sunetra Das
Colorado State University
Rupa Idate
Colorado State University
Daniel Regan
Colorado State University
Jared Fowles
Colorado State University
Susan Lana
Colorado State University
Douglas Thamm
Colorado State University
Daniel Gustafson
Colorado State University
Dawn Duval (dawn.duval@colostate.edu)
Colorado State University https://orcid.org/0000-0002-5310-0078

Article

Keywords: osteosarcoma, oncology, cancer genes

Posted Date: January 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-135094/v1

Version of Record: A version of this preprint was published at Communications Biology on October 11th, 2021. See the published version at https://doi.org/10.1038/s42003-021-02683-0.

Whole exome sequencing and gene expression analysis of canine osteosarcomas identify mutant *TP53* and enriched immune pathways associated with longer survival.

Authors

Sunetra Das[1,2], Rupa Idate[1,2], Daniel P. Regan[2,4,5], Jared S. Fowles[1,2,3], Susan E. Lana[1,2], Douglas H. Thamm[1,2,5], Daniel L. Gustafson[1,2,5], Dawn L. Duval[1,2,5]

Affiliation

[1]Department of Clinical Sciences, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, 80523, USA

[2]Flint Animal Cancer Center, Colorado State University, Fort Collins, CO, 80523, USA

[3]Cell and Molecular Biology Graduate Program, Colorado State University, Fort Collins, CO, 80523, USA

[4]Department of Microbiology, Immunology, & Pathology, Colorado State University, Fort Collins, CO, 80523, USA

[5]University of Colorado Cancer Center, Anschutz Medical Campus, Aurora, CO, 80045, USA

Corresponding author

Dawn Duval

300 West Drake Road, Campus Delivery 1620, Fort Collins, CO 80523

970-297-4064

Dawn.Duval@colostate.edu

ABSTRACT

Osteosarcoma affects about 2.8% of dogs with cancer, with a one-year survival rate of approximately 45%. The purpose of this study was to characterize the mutation and expression profiles of canine osteosarcoma associated with outcome in samples from dogs treated by amputation and chemotherapy.

The number of somatic variants identified across 26 samples ranged from 145 to 2,697 with top recurrent mutations observed in *TP53* (82% of the samples) and *SETD2* (22%). Additionally, 47 cancer genes were identified with copy number variations in at least 58% of the samples. We observed transcriptional down-regulation of myogenesis genes and up-regulation of extracellular matrix genes in tumors compared to normal bone. Patients with longer disease-free intervals (DFI) showed increased transcript levels of anti-tumor immune response genes. Wild-type/NULL *TP53* mutation status and high pre-treatment blood monocyte counts were associated with a shorter DFI. Immune cell infiltration was quantified via immunohistochemistry and gene expression profiling. CD3+ and MAC387+ myeloid cell quantifications were not significantly associated with outcome. Expression of immune related genes *PDL-1*, *CD160* and *ICOS* were correlated with T-cell abundance.

Overall, the association of gene expression and mutation profiles to outcome provides insights into pathogenesis and therapeutic interventions in osteosarcoma patients.

BACKGROUND

Osteosarcoma (OSA) is characterized as a neoplasm of bone (mesodermal origin) and occurs spontaneously in a wide spectrum of mammals including humans and dogs [1]. Most

commonly, OSAs arise in the metaphysis of long bones in both dogs and humans, and histologically these malignant tissues produce an extracellular matrix called tumor osteoid. This malignant cancer is the most common type of bone cancer in children, adolescents, and to a lesser extent in the aging population (> 60 yrs old), but not in adults. In comparison, 80% of OSA occurs in dogs >7 years of age and rarely in juveniles (6-8%) [2, 3]. In a study conducted in Norway, the authors concluded that OSA risk was breed dependent with increased incidence in large and giant breed dogs [4]. According to the Surveillance, Epidemiology, and End Results (SEER; https://seer.cancer.gov/) database, the five-year survival rate in humans is 66% but is only 27% in patients with measurable metastatic disease at diagnosis. Current treatments in humans include limb sparing surgery and neoadjuvant or adjuvant chemotherapy (using doxorubicin, cisplatin, methotrexate, and ifosfamide) [5]. This has led to an increase in the five-year survival rate from 20% to 70%. Osteosarcoma in dogs is typically treated by amputation of the affected limb followed by chemotherapy with doxorubicin and/or platinum-based therapies. A 2014 study of 470 dogs treated for OSA with amputation and chemotherapy found that the median disease-free interval (DFI) was 291 days, and was not statistically different based on type of chemotherapy administered [6].

The discovery of genomic modifications that lead to malignancies and the subsequent targeting of these genes or pathways is possible due to major advancements in sequencing technology and resources for computational analysis. Recent articles have been published detailing the genomic drivers of OSA in both humans and dogs [7-12]. The key discovery in human OSA is that these cancers bear large numbers of structural and copy number variations, with little to

no presence of activating mutations in oncogenes and infrequent point mutations in protein coding genes. One of the first papers to conduct whole genome sequencing in human OSA identified *TP53* structural variants (SV) and single nucleotide variants (SNV) in 95% of their samples [7]. The other tumor suppressor genes with recurrent somatic variants were *RB1*, *ATRX*, and *DLG2* in 29% - 53% of the patients. Perry et al., in 2014 reported similar variants; however, the authors identified the PI3K/MTOR pathway as a therapeutic target, as this pathway was altered in 24% of patients [8]. Further studies have used whole genome and RNA-sequencing technologies in patient tumors and patient-derived xenografts to demonstrate that genes with somatic copy number alterations can be targeted to reduce tumor burden [10]. To date, there have been two reports of genome and exome wide variant analyses in canine OSA [11, 12]. In addition to recurrent *TP53* point mutations and copy number variations, these studies identified two additional recurrently mutated genes: *SETD2* (histone lysine methyltransferase) and *DMD* (dystrophin) that have not been previously identified in human OSA. However, it is not clear if these genes represent cancer drivers in dogs. Similar to human OSA, the short variant mutational burden was very low in comparison to structural (SV) and copy number variants (CNV) in canine bone tumors.

Apart from surgical intervention and chemotherapy, immunotherapy is emerging as an alternative form of treatment for many forms of cancer. Although, there is relatively little success of immunotherapy in OSA patients to date, based on the tumor microenvironment profile it is suggested that OSA might be receptive to immune-therapies [13]. Recent articles on the immuno-genomic landscape in human OSA have sought to identify prognostic markers and

genomic targets for immune therapy [14, 15]. It has been shown that expression of *PD-L1* in human OSA was significantly associated with number of immune infiltrates such as T cells, dendritic cells, and natural killer cells [16]. However, the relatively low numbers of these infiltrates could be one reason for low success in treating OSA patients with immunotherapy [15, 17].

In this paper, we have conducted multi-platform analysis of 26 canine OSA samples, including whole exome sequencing, microarray analysis, and immune cell profiling. Similar to previous published work, we report a prevalence of copy number variants in comparison to short variants (SNVs and INDELs). The top two recurrently mutated cancer genes with short variants were *TP53* and *SETD2*. Using GISTIC2 to identify CNV, we identified a large portion of genes in deleted regions of the genome and fewer amplified regions. Additionally, we identified differentially expressed genes between tumors and normal metaphyseal bone based on Affymetrix Canine 2.0 microarrays. The observed variant and gene expression data were correlated with patient outcome data following treatment with limb amputation and doxorubicin and/or platinum-based therapies. The disease-free interval (DFI) was used to categorize the patients in the first (DFI <123 days) and fourth (DFI >395 days) quartiles. Tumors from the first quartile group were enriched for genes in immune-related pathways. In summary, the current work explores the relationship between the canine OSA mutational spectrum and associated changes in gene expression to identify pathways that contribute to cancer progression and therapeutic sensitivity.

RESULTS

Profiling the genomic landscape of canine osteosarcoma

The whole exome sequencing data from 26 primary tumors and 26 matched normals were bioinformatically analyzed to identify somatic short variants, i.e., single nucleotide variants (SNVs) and insertion and deletions (INDELs), and copy number variations (CNVs). In addition to 26 matched normal samples, 17 additional normal samples were sequenced to generate a panel of normals for identification of somatic variants (**Additional File 1; Table S1**). To further filter out germline mutations, the SNVs and INDELs reported from 722 dogs was used as a germline resource during variant calling [18]. The median depth of sequencing for normal and tumor samples was 247X (range: 88X – 578X) and 295X (range: 126X – 453X), respectively (**Additional File 2, Figure S1**). The total number of somatic short variants identified across 26 primary tumor samples ranged from 145 (T-1247) to 2,697 (T-153) (**Fig. 1A**). Of these variants, 6.9% (T-1272) to 25.9% (T-554) were annotated as protein coding variants. The protein-coding mutations per megabase ranged from 0.25 (T-1272) to 7.39 (T-153) (**Additional File 2, Figure S2**). There was no significant correlation between disease free interval (in days) and mutations per megabase (Hazard Ratio: 1.095, $p = 0.5$). Within the protein somatic variants identified as SNVs and INDELS, an average of 80% were missense mutations (range 64% to 92%) (**Fig. 1B**). Overall, there were 739 deleterious and 889 tolerated missense mutations as identified by SIFT scoring. We also sequenced a single metastatic tumor sample (M-1166). This sample had a total of 746 somatic variants of which 10.9% were located within protein-coding regions of genes. In comparison, the corresponding primary tumor (T-1166) had 908 somatic short variants and 14.6% of these variants were located in the gene coding regions.

The CNVs were analyzed using Sequenza and GISTIC to identify significantly amplified and deleted regions. The median number of genes with significant copy number alterations was 1,468 with a range from 749 (T- 1247) to 1,630 (T-153). The majority of CNVs among the 26 samples were gene deletions with the percentage of significantly deleted genes ranging from 81% (T-1247) to 90% (T-458) of CNVs (**Additional File 2, Figure S3**). The number of genes with CNVs in the metastatic tumor, M-1166, was 1,356, of which 91.5% were located in significantly deleted regions of chromosomes. In comparison, the primary tumor, T-1166, had 1,156 CNV genes with 98.6% located within the significantly deleted regions of the genome. Additionally, there was no significant correlation between DFI and total CNVs for 26 samples (Hazard Ratio: 1.001, p=0.4).

The distribution of six types of single nucleotide substitutions (C·G → A·T, C·G → G·C, C·G → T·A, T·A → A·T, T·A → C·G, and T·A → G·C) for all 26 samples revealed C·G → T·A transition mutations as the most frequent substitution (**Additional File 2, Figure S4A**). There are 30 COSMIC mutational signatures (v2, https://cancer.sanger.ac.uk/cosmic/signatures_v2.tt) that were developed based on the distribution of the six single nucleotide substitutions within 96 trinucleotide contexts and pathogenesis [19]. To further delineate the mutational signatures of individual samples, the non-negative matrix factorization (NMF) approach was used, and three top-ranked *de novo* signatures (Signatures A, B, and C) were generated from the frequency of those 96 types of SNVs within trinucleotide context among the 26 osteosarcoma samples (**Fig. 1C**). Signature A was most similar to COSMIC signature 9 (cosine similarity=0.72) which is characterized by mutations attributed to DNA polymerase η. DNA Polymerase η is recruited to

sites lacking bases during the repair of U:G mismatches generated by activation-induced cytidine deaminase through somatic hypermutation and is often found in chronic lymphocytic leukemias and malignant B-cell lymphomas [20]. Signature B was similar to COSMIC signature 1 (cosine similarity=0.89) which is characterized by spontaneous deamination of 5-methylcytosine resulting in C→T transitions. This signature correlates with age of cancer onset in humans. Signature C was most similar to COSMIC signature 5 which is characterized by transcriptional strand bias for T>C substitutions at ApTpN context and is found in most cancer types. The relative contribution of trinucleotide context of *de novo* signatures attributed to each sample indicated that 38%, 24%, 38% of samples resembled A, B, and C, respectively (**Fig. 1D**). Additionally, we have compared distribution of the trinucleotide context of 26 tumor samples to the 30 known COMIC signatures. Similar to the *de novo* analyses, the majority of the OSA samples were most similar to either signature 1 or 5; however, there was one sample with highest similarity to signature 9 (T-1247). The hierarchical clustering of 30 COSMIC mutational signatures suggests that the profile of tumors with signature 9 was similar to signature 5. There were four samples with highest cosine similarity for signature 6 (T-1192, T-907, T-992, T-999); and one each most similar to signature 17 (T-276) and signature 19 (T-1272) (**Additional File 2, Figure S4B**).

Short somatic variants with SNVs and INDELs

As mentioned earlier 6.9% - 25.9% of somatic short variants were identified in protein coding genes. A total of 1,579 protein coding genes had a variant in at least one of the 26 samples for a total of 1,934 protein coding variants (**Additional File 1, Table S2**). Although there was a range

of 321 (T-153) to 11 (T-1272) genes with variants across 26 samples, only 14 genes were recurrently mutated in at least 15% of the samples (**Additional File 2, Figure S5**). There were 1,100 genes that were mutated in only one sample and 129 genes mutated in a maximum of two samples. Using DAVID (https://david.ncifcrf.gov/), we functionally characterized the genes with protein coding variants to identify enriched pathways and domains (**Additional File 1, Table S3**). The significantly enriched (Benjamini FDR <0.05) KEGG pathways included ECM-receptor interaction (cfa04512), and Focal adhesion (cfa04510). There were two signal transduction pathways enriched at p-value <0.05, PI3K-Akt signaling (cfa04151) and Calcium signaling (cfa04020), in addition to cell cycle (cfa04110) and pathways in cancer (cfa05200) KEGG terms. Some significantly enriched PFAM domains and Interpro classifications (Benjamini FDR <0.1) were Laminin G, Spectrin repeat, Dynein heavy chain, Fibronectin type III, Laminin B, and Immunoglobulin domains (see **Additional File S1, Table S3** for details on genes).

The protein coding somatic variants were also filtered for known cancer genes using the curated dataset from the COSMIC Cancer Gene Census (v91, 570 genes). On average across the 26 samples, known cancer genes with short variants represented 5.9% (±3.3%) of coding mutations. There were 62 cancer genes with at least one variant among the 26 samples. The top four genes (mutated in at least in 12% of samples) with short variant mutations were *TP53*, *SETD2* (SET domain containing 2, histone lysine methyltransferase), *HSP90AA1* (heat shock protein 90kDa alpha, member A1), and *DNMT3A* (DNA-methyltransferase 3A) (**Fig. 2A**). The variants identified in *TP53* were primarily located within the DNA binding domain and are considered to be driver mutations in human cancers (**Additional file 2, Figure S6, Additional**

File 1, Table S4). *SETD2* mutations included 3 samples with both frameshift and stop-gained mutations and three samples with either a missense (variant =S1658P in T-134), frameshifts (variants =P1158Lfs in T-544 and T-399; L124Yfs in T-25; Y1033Ifs inT-1087), or stop-gained mutations (variant =R396* in T-544 and T-399; Q1546* in T-25; Q1431* in T-149). Homologous frameshift and stop gained mutations of the *SETD2* tumor suppressor gene in human cancers are considered to be likely oncogenic. The missense mutation, however, identified here in *SETD2* is not considered to be a cancer hotspot as reported by pan-cancer analysis in cBioPortal. The *HSP90AA1* gene was mutated in 12% of the samples and all three samples carried the same missense mutation (A149D in T-346, T-153, T-25). Sample T-346 also carried an additional stop-gained mutation (E530*). The only other recurrently mutated gene, *DNMT3A,* carried two different missense mutations in three samples (N597S in T-346 and T-856; W738R in T-1023). Of the 62 cancer genes, 58 genes were mutated in only one or two of the samples, limiting further analysis regarding their impact on clinical outcome. By binning the tumors based upon the *TP53* mutation with the highest allelic frequency, we identified missense *TP53* mutations in 65% of the tumors, with frameshift or stop-gained (null) mutations identified in an additional 23%, and wildtype (WT) *TP53* present in 15% of tumors (**Additional File S1, Table S4**). The next most recurrently mutated gene, *SETD2,* was mutated in approximately 23% of canine OSA samples. Finally, the primary tumor T-1166 had 3 different *TP53* variants (D13Afs*4 at 4.5% allelic frequency, L184P at 16.67%, G234E at 7.5%) while the subsequent metastatic lesion, M-1166 had only the L184P variant (68.2%). Coding sequence variants in the cancer genes *RANBP2* (RAN binding protein 2), *DGCR8* (DiGeorge syndrome critical region gene 8), and *FAM135B* (family with sequence similarity 135 member B) were also lost in the metastatic lesion, while

variants in *DEK* (DEK proto-oncogene), *FBXO11* (F-box protein 11), *NOTCH2* (neurogenic locus notch homolog protein 2), *XPC* (xeroderma pigmentosum, complementation group C), and *ZNF521* (zinc finger protein 521) were gained.

Copy number variants

In comparison to short variants, there were significantly more genes with somatic CNVs (**Additional File 1; Table S5**). The copy number analysis was conducted using the Sequenza tool to calculate gain or loss in comparison to matched normal and significantly amplified and deleted genes were identified by GISTC2.0 [21]. A total of 1,662 genes were significantly altered across 26 samples. The median number of genes with significant amplifications and deletions was 169 (range: 114: T-1166 to 205: T-346) and 1,295 (range: 607: T-1247 to 1460: T-458), respectively (**Additional File 2; Figure S3**). We analyzed the functional significance of these genes using Enricher KEGG pathways category (https://amp.pharm.mssm.edu/Enrichr/). KEGG pathways that were enriched at a q-value of 0.1 included MAPK signaling, Autophagy, PI3K-AKT, p53 signaling and FOXO signaling (**Additional File 1; Table S6A**).

A total of 47 cancer genes (COSMIC Cancer Gene Census, v91) were identified to have significantly deleted or amplified copy number aberrations (**Fig. 2B**). Six cancer genes were significantly amplified, including, *CCND3* (cyclin D3), *CUX1* (cut like homeobox 1), *EXT1* (exostosin glycosyltransferase 1), *MYC* (MYC proto-oncogene), *RAD21* (RAD21 cohesin complex component), and *TFEB* (transcription factor EB). Most of the remaining 41 genes were significantly deleted. *FOXO1* was considered to be both significantly deleted and amplified in individual samples. We also functionally annotated the significantly amplified and deleted

genes separately to identify pathways that might be altered by mis-regulation of these genes. The two KEGG pathways enriched for the genes that were significantly amplified were cell cycle and cellular senescence (adjusted p-value<0.05; **Additional File 1, Table S6B**). The top signaling KEGG pathways enriched for the significantly deleted genes were p53, FoxO, PI3K-Akt, MAPK, ErbB, TNF, mTOR, Wnt, Hippo, and JAK-STAT pathways (**Additional File 1, Table S6C**).

Correlation of copy number variation index with gene expression

The functional effect of copy number variations was evaluated through correlations with gene expression using Pearson correlations. We correlated the gene amplitude values obtained from Gistic2.0 "all_data_by_gene.txt" output to Z-scored gene expression values obtained from the Affymetrix Canine 2.0 microarray data. Of the 1,662 genes with recurrent copy number alterations, 1,471 genes had gene expression values from microarray data for all 26 samples. There were 256 genes with significant correlations between CNV amplitude and expression (**Additional File 1 Table S7**). The 12 cancer genes (COSMIC v91) that had significant correlation were *BRCA2*, *BUB1B*, *CCND1*, *CCND3*, *FLCN*, *FOXO1*, *MAP2K4*, *NCOR1*, *PTEN*, *RABEP1*, *SDHAF2*, and *TFEB*. There was a linear relationship between copy number changes and expression levels of corresponding gene transcripts, which included significantly amplified genes (CCND3 and TFEB) as well as significantly deleted genes including BRCA2, BUB1B, PTEN (**Additional File 2, Figure S7**). These data suggest that although there might be a plethora of CNV changes in the genome, the functional impact as evaluated by gene expression is far more limited.

Over-representation of extracellular matrix components in canine osteosarcoma samples

Using microarray analysis, we compared the transcriptome profiles of 26 canine OSA samples and eight normal bone samples. Following RMA (robust multi-array average) normalization of microarray data from tumor and normal bone samples, the expression data were processed to identify differentially expressed genes using limma [22]. The number of up- and down-regulated probes (adjusted p-value <0.05 and $\log_2$ fold change >±2) were 201 and 721, respectively (**Additional File 2; Figure S8**). These probes were collapsed to retain probes with maximum variance for each gene and those expression values were plotted as a heatmap in **Fig. 3A**. There were 585 differentially expressed genes (DEGs) between normal bone and OSA samples.

The DEGs were analyzed using pre-ranked Gene Set Enrichment Analysis (GSEA) to identify enriched up- and down-regulated pathways and GO terms in the tumor samples (**Additional File 1; Table S8**). There were 187 pathways or terms enriched at an FDR q-value of 0.05, using the following MSigDB Collections: Hallmark, Canonical pathways, GO biological process, GO molecular function, oncogenic signatures, immunologic signatures, and chemical and genetic perturbations. The results identified enrichment of multiple gene sets associated with extracellular matrix (ECM), cell proliferation, epidermal to mesenchymal transition, glycolysis, and metastasis genes (**Fig. 3B-C, Additional File 1 Table S8**). These genes were significantly upregulated in the tumor samples when compared to normal bone (positive NES score). The genes downregulated in tumors were enriched for functional pathways (negative NES score) including: myogenesis, innate and humoral immune response, paracrine hedgehog signaling, and drug transport (**Fig. 3D, Additional File 1; Table S8**). The enrichment analyses also showed

upregulation in tumors of genes that are increased in response to the loss of RB1 (Oncogenic signature: RB_P107_DN.V1_UP). Similar to human OSA, there is loss of RB1 in canine tumors, as evidenced by down-regulation of *RB1* gene expression (Affymetrix Canine 2.0 microarray), although copy number loss was not observed.

Immune response pathways were enriched in osteosarcomas from dogs with a longer disease-free interval

For identification of pathways enriched in tumors from long disease-free interval (DFI) and short DFI patients, we eliminated two of the 26 samples (T-474 and T-1192) due to lack of follow up and sorted the remaining samples into more highly defined DFI bins based on the 25^{th} (short DFI, <123 days) and 75^{th} (long DFI >395 days) percentiles of DFI across the 26 samples. This placed six samples in each of these two bins. The gene expression within these two bins was analyzed via GSEA to identify enriched pathways. Using the same MSigDB collections used for tumor vs normal analyses, 83 pathways/terms were enriched in long DFI patients. Of these 83, 80 gene sets were enriched in long DFI patients and only three gene sets were enriched in short DFI patients (**Table 1 and Additional File 1; Table S9**). The three gene sets that were enriched in short DFI patients were 1) KRAS.50_UP.V1_UP: genes upregulated in epithelial cell lines due to over-expression of KRAS (**Additional File 2; Figure S9A**); 2) BCAT_GDS748_UP: genes upregulated in kidney fibroblast cell line as a result of constitutively active β-catenin (CTNNB1); and 3) SINGH_NFE2L2_TARGETS: electrophile and drug detoxification genes that are down-regulated in *Nrf2* knocked-down cell lines (**Additional File 2; Figure S9B**). These selected drug detoxification genes are upregulated in short DFI patient tumors, suggesting that these

patients' tumors might be more efficient in clearing chemotherapy drugs. The Hallmark gene sets upregulated in long DFI patient tumors included several immune response datasets: Interferon alpha and gamma response, inflammatory response, and allograft rejection along with ECM gene sets (**Additional File 1; Table S9, Additional File 2; Figure S10**).

Table 1. Summarization of gene sets and pathways that are enriched in short or long DFI patients. See Additional File 1, Table S9 for details on enriched gene sets.

Biological process or gene sets	Number of gene sets	Average NES	Enrichment sample set
Immune system	36	-1.88	Long DFI patients
Extracellular matrix	5	-2.04	Long DFI patients
BCAT_GDS748_UP	1	1.79	Short DFI patients
KRAS.50_UP.V1_UP	1	1.92	Short DFI patients
SINGH_NFE2L2_TARGETS	1	1.98	Short DFI patients

Immune cell profiling in osteosarcoma

To analyze the immune infiltrate in these tumors, those samples with available tissue blocks were labeled with the pan-T cell marker CD3 and the myeloid cell marker MAC387 (S100A9) via immunohistochemistry (**Fig. 4A-D**). Whole slide images were captured and immune cell density as a percentage of total tumor area was determined using ImageJ software. The percent tumor area positive for CD3+ T cells ranged from 0.002% (T-25, DFI – 372 days) to 4.87% (T-29C, DFI – 1197 days as of July 23rd, 2020) (**Fig. 4E**). The percent tumor area positive for MAC387+ myeloid cells ranged from 0% (T-2013, DFI – 216 days) to 3.7% (T-74, DFI – 406 days) (**Fig. 4F**). Survival analysis using DFI as the time event showed no statistically significant difference when the

patients were stratified by high and low levels of either CD3 or MAC387 staining (**Additional File 2, Figure S10**).

We also used the gene expression data for more comprehensive immunogenomic profiling of the immune infiltrates in the tumor samples, given the lack of available canine-specific antibodies for immune phenotyping. This analysis was performed using three different methods: a) gene expression profiling for immune cell types as previously described, b) CIBERSORT tool, and c) ESTIMATE R package [23-25]. As described in the Methods section, the gene profiles in Rooney *et al.* were used to categorize immune cell subsets using signature gene sets for each cell type and immune responses (**Additional File 1, Table S10**). Using Pearson correlations, we created a matrix to identify significant associations of immune cell scores (derived from gene expression), DFI, mutations per megabase and immunostaining quantification of T cells (IHC T-cells) and macrophages (IHC-Macs). Statistically significant positive correlations were observed between IHC T-cells and gene expression scores for cytolytic activity, co-stimulation T-cell, co-inhibition T-cell, CD8+ T-cell, MHC class I, and NK cell. However, IHC-Macs did not correlate with the gene expression score for macrophages (**Fig. 5A**). Interestingly, the macrophage score was inversely correlated with CD4+ regulatory T-cells, but positively correlated with neutrophils and MHC Class I. This positive correlation with neutrophils is not unexpected, as MAC387 is ubiquitously expressed by all myeloid cells in the dog, including neutrophils [26]. In addition, the IHC-Macs score was positively correlated with co-inhibition T-cells and co-stimulation APCs. Neither DFI nor tumor mutational burden

(mutations per MB) were significantly correlated with these IHC staining scores or immune cell gene expression signatures (**Figure 5A**).

We also took the 85 genes used for generating immune signatures and correlated them with both IHC staining for T-cell and macrophage infiltration in tumors (**Table 2).** There were 18 genes that significantly correlated with T-cell infiltration at a false discovery rate of 0.05, including the known co-inhibitory immune checkpoint molecules *PDL1* (*CD274*), *CD160* (ligand for Herpesvirus entry mediator), as well as the cytotoxic T cell co-receptor *CD8A* (cell surface marker for cytotoxic T-lymphocytes) (**Table 2A**). Only one gene (*TNFSF4*) was negatively correlated with T-cell infiltration. Five genes, including *PRF1*, *SLAMF1*, OX40, TNFRSF9 and CXCR3 were positively correlated with MAC387 staining. (**Table 2B**). The cytolytic activity (derived as geometric mean of perforin and granzyme A transcript expression) in OS tumors was significantly correlated with 21 immune response genes that are markers for T-cells, macrophages, MHC Class I, dendritic cells, Type I and II IFN Response, B cells, and natural killer cells (**Additional file 1, Table S11**).

Table 2. Correlation of immune gene expression and immunohistochemical staining of T-cell and macrophage infiltration in the tumors.

2A. Gene expression correlation with CD3+ T cells in tumors.

Gene name	Associated immune cell type	Correlation coefficient	p-value	q-value
CD8A	CD8+ T cells	0.94	0	0
CD160	Co-inhibition, T cell	0.86	0	0
CD274 or PDL1	Co-inhibition, APC; Co-inhibition, T cell	0.86	0	0
GZMA	Cytolytic Activity	0.75	0.000	0.003
CD2	Co-stimulation, T cell	0.7	0.001	0.010
ICOS	Co-stimulation, T cell	0.68	0.002	0.015
CLEC5A	Macrophages	0.67	0.002	0.016
TNFRSF4	Co-stimulation, T cell	0.65	0.003	0.020

KLRF1	NK cells	0.65	0.003	0.020
AHR	Type II IFN Response	0.64	0.003	0.021
ISG20	Type I IFN Response	0.63	0.004	0.022
HAVCR2	Co-inhibition, T cell	0.63	0.004	0.023
IRF8	pDCs	0.6	0.007	0.031
CXCR3	pDCs	0.59	0.008	0.035
TAP1	MHC Class I	0.58	0.009	0.036
CD79B	B cells	0.58	0.010	0.039
KDM6B	Neutrophils	0.56	0.012	0.046
TNFSF4	Co-stimulation, APC	-0.61	0.006	0.028

2B. Gene expression correlation with MAC387+ myeloid cells in tumors.

Gene name	Associated immune cell type	Correlation coefficient	p-value	q-value
PRF1	Cytolytic Activity	0.74	0.000	0.008
SLAMF1	Co-stimulation, APC; Co-stimulation, T cell	0.71	0.000	0.011
CD40 or OX40	Co-stimulation, APC	0.68	0.001	0.020
TNFRSF9	Co-stimulation, T cell	0.65	0.002	0.028
CXCR3	pDCs	0.62	0.003	0.042

To further delineate the immune cell profiling with a larger dataset of genes, we used the CIBERSORT deconvolution tool [24, 27]. The results of this analysis indicate a higher prevalence of both resting/uncommitted M0 and alternatively activated M2-polarized macrophages in all the tumor samples and a lower prevalence of M1-polarized pro-inflammatory macrophages (**Fig. 5B**). However, there was no correlation between the abundance of these three types of macrophages and clinical outcome. The immunohistochemical staining for CD3+ T cells positively correlated with the CIBERSORT score for four different cell types including: CD8 T cells (R^2: 0.91), activated mast cells (R^2: 0.98), plasma cells (R^2: 0.77), and gamma delta T cells

(R^2: 0.73). Significant correlation of MAC387 positive cells with CIBERSORT immune scores were limited to follicular helper T cells (R^2: 0.67).

A third method used in this study to determine levels of immune infiltrates was the ESTIMATE algorithm [25]. The ESTIMATE immune score ranged from -83.61 (T-1087) 2110.4 (T-74) (Fig.6C). ESTIMATE scores significantly correlated with IHC staining for both T cells (Pearson R^2 – 0.58, p value – 0.008) and macrophages (Pearson R^2 – 0.43, p value – 0.04). However, the ESTIMATE scores did not correlate with genomic parameters like mutational burden, and number of deleted and amplified genes. Additionally, there was no association between ESTIMATE immune score and DFI.

Association of TP53 mutation status with clinical outcome

We observed that missense mutations in *TP53* were associated with a longer disease-free interval (DFI) than either the *TP53* WT or null tumors following treatment by amputation of the affected limb and chemotherapy with doxorubicin and/or carboplatin (**Figure 6A**). The median DFIs for patients with mutant and WT/NULL *TP53* were 296 and 95 days, respectively [HR (95% CI) = 0.21 (0.08 to 0.58), p = 0.002]. Along with *TP53* mutation status, we also evaluated six clinical co-variates that included age at diagnosis, tumor location (humerus versus other locations), sex, pre-treatment peripheral blood monocyte count, serum alkaline phosphatase levels and body weight to identify their association with DFI via univariate Cox proportional hazards (COXPH) regression analysis. Two additional variables used in this analysis were tumor immunohistochemical staining scores for CD3+ T cells and MAC387+ myeloid cells. Significantly

increased risk was associated with two of the nine covariates, including *TP53* mutation status and tumor location (**Table 3A**). Although overall survival among these 26 canine OS patients was not significantly reduced by increased numbers of peripheral blood monocytes (**Fig. 6B**), using five co-variates, *TP53* mutation status, tumor location, peripheral monocyte count, and alkaline phosphatase level (univariate COXPH p <0.2), the final model for forward stepwise COXPH regression analysis indicated that wildtype/Null *TP53* mutation status and elevated peripheral blood monocyte counts were predictive of a shorter DFI **(Table 3B)**.

Table 3. Univariate and stepwise cox proportional hazard modelling.

Table 3A. Results from univariate COXPH regression analysis using six clinical and one genomic parameter. The factors with a p-value <0.2 were used for the stepwise multivariate model.

Variable	Group	Median survival time	Percent survival - 1 year	Percent survival - 2 years	HR	95% CI	p value
TP53 mutation status	Mut	296	47	12	0.21	0.08 to 0.58	0.002
	WT/NULL	95	14	0			
Tumor location - Humerus	Yes	114	12.5	0	4.05	1.51 to 10.84	0.005
	No	372	50	13			
Pre-treatment peripheral blood monocytes (>or< 0.04*10^3 cells/µl)	High	216	20	0	2.12	0.87 to 8.14	0.091
	Low	309	50	14			
Alkaline phosphatase levels (>or< 140 units/liter)	High	216	29	0	1.93	0.74 to 4.99	0.176
	Normal	246	35	6			
IHC CD3 staining	NA	NA	NA	NA	0.56	0.23 to 1.33	0.191
Age	NA	NA	NA	NA	1.13	0.93 to 1.36	0.224
Sex	Male	296	42	8	0.75	0.33 to 1.72	0.504
	Female	142	33	8			
Weight	NA	NA	NA	NA	1	0.97 to 1.04	0.846
IHC MAC387 staining	NA	NA	NA	NA	1.02	0.61 to 1.69	0.954

Table 3B. Results from stepwise COXPH multivariate analysis using four parameters for subset selection. The two significant co-variates reported here were associated with poor prognosis.

Factors selected in stepwise model	HR	95% CI	p value

TP53 mutation status (WT/NULL)	6.39	2.12 to 19.209	0.0009
Monocyte count (>0.04 *10^3 cells/µl)	2.83	1.09 to 7.33	0.032

419

420

421 **DISCUSSION**

422 The current whole exome sequencing and microarray data analyses of 26 canine OSA samples
423 combines the identification of somatic variants that could drive cancer with an analysis of genes
424 with altered expression in those tumors compared to normal bone. Additionally, the clinical
425 outcome data was used to identify the pathways and somatic variants that were modulated
426 based on disease-free intervals (DFI). The mutational burden in OSA for short somatic variants
427 is usually low in both canine and pediatric cancers. The protein coding mutations per Mb in
428 most of the 26 canine OSA samples were similar to those reported in human OSA (less than 2
429 mutations per Mb, across 42 samples) [28]. There was only one sample (T-153) with a high
430 mutational burden, greater than 10 mutations per Mb, which could be considered
431 hypermutable as per criteria established by Gröbner et al. 2018. In this study, neither
432 mutational burden or number of CNVs were correlated with DFI, suggesting that the absolute
433 numbers of tumor mutations were not the primary predictors of disease progression or
434 response to therapy in canine OSA.

435

436 Similar to previous WES and WGS studies of canine OSA [11, 12], we identified C>T transitions
437 as the primary point mutation type in these canine tumors. Within the context of the 96
438 trinucleotide combinations, we identified 3 *de novo* signatures with similarity to COSMIC
439 signatures 9, 1, and 5. Sakthikumar *et al.* identified COSMIC signatures 1 and 17 and found a

greater representation of COSMIC signature 1, associated with aging, in Rottweilers and Greyhounds, while COSMIC signature 17 and a novel signature were more common in Golden Retrievers. Like the Gardner et al. study, we identified high similarity to COSMIC signature 17 in only one sample (**Additional Figure S4B**). This sample was from a female mixed breed dog that fell within the DFI <100-day group, had 1.91 protein coding mutations per megabase, and is *TP53* wildtype. Given the rarity of this signature in our samples, this sample was binned as a signature A tumor with overall similarity to COSMIC signature 9. Another tumor, sample 1247, with a high contribution of signature A was also *TP53* WT, but had a DFI >300 days. While only identified in the WGS analysis, the Gardner et al study also identified a group of tumors with the COSMIC signature 9.

Similar to human OSA variant types, the samples in this study had more copy number variants than SNVs and INDELs [7]. Over 80% of the samples had at least one *TP53* short variant and/or CNV which is similar to previous reports in both human and canine studies [8, 11, 12]. We identified missense *TP53* mutations in 70% (18) of the canine OSAs with nonsense or frameshift variants identified in 15% (4) of tumors and the remaining 15% (4) being *TP53* wildtype. In comparison, Chen *et al.* identified *TP53* pathway mutations in each of 20 tumors examined [7]. This study, with only 2 somatic missense mutations, one frameshift variant, and 55% (11/20) bearing structural variants in the first intron, suggests that for human OSA, TP53 missense variants are comparatively rare. One of the surprising results from the current study was the correlation of increased DFI and tumors bearing missense mutations in *TP53*. Similar to the current association with outcome, a previous study showed that there was no significant

462 difference in DFI between tumors with *TP53* mutant and wildtype variants [29], but did identify
463 a longer overall survival in dogs with wildtype *TP53*. However, that study reported *TP53*
464 variants in only 40.7% of cases, included both missense mutations and frameshift or nonsense
465 mutations in the *TP53* mutant category for DFI analysis, and not all the dogs included in these
466 analyses received chemotherapy. In addition, survival times can be confounded by the use of
467 euthanasia in companion animals, making DFI a potentially more biologically relevant measure
468 of outcome. In contrast, each of the dogs included in this study received 1 to 5 doses of
469 doxorubicin and/or 1 to 5 doses of carboplatin. A targeted exome sequencing study in human
470 small-cell lung cancer identified *TP53* mutations in 54% of patients, where it was reported that
471 the patients with mutant TP53 had longer relapse free intervals as compared to patients with
472 wild type *TP53* [30]. Similarly, *TP53* mutant human cancers including breast, are significantly
473 more likely to achieve pathological complete responses to chemotherapy [31-35]. Examination
474 of this phenomenon has shown that tumors from mice with the murine Tp53 R172H variant
475 exhibit greater sensitivity to doxorubicin, and fail to exit the cell cycle following treatment,
476 resulting in aberrant mitosis, and cell death [36]. Gain of function *TP53* mutants, including
477 R175H, R248W, and R273H, inactivate the ATM-dependent DNA damage response leading to
478 chromosomal translocations and a defective G2/M checkpoint, and improved treatment
479 responses [37]. This suggests that a similar phenomenon may occur in *TP53* mutant canine OSA.
480 The multivariate COXPH analysis with *TP53* and relevant clinical factors indicated that both high
481 numbers of pre-treatment blood monocytes and wild type/null *TP53* may be poor prognostic
482 markers in canine OSA treated by amputation and doxorubicin and/or carboplatin. The higher

pre-treatment monocyte counts were previously reported to be associated with a shorter DFI in OSA patients [38].

The second most frequently mutated gene with short variants was *SETD2.* This gene is a potential tumor suppressor and is mutated in several human solid tumors [39], but not in human OSA. Although the majority of the *SETD2* variants resulted in frameshifts, mutation status did not impact transcript expression based on Canine 2.0 Affymetrix microarray analysis in this study. The significance of these alterations at the transcript and protein level is unknown at this time.

As reported in the results, a majority of genes with short variants were mutated in only one of the samples, suggesting that these are passenger mutations. In addition, we did not observe much overlap between the less frequently mutated genes in our study and the two other canine studies [11, 12]. The functional annotation of genes with short somatic variants showed enrichment of PI3K-AKT signaling and cell cycle pathways. This is similar to functional annotation of somatic variants previously identified in pediatric OSA [8].

As mentioned earlier, copy number variations are more prevalent in both canine and human OSA. Perry et al. 2014, identified 850 genes that had significant copy number alterations in pediatric OSA. Upon comparing the CNV genes with those (1,662) identified here, we found 77 genes that overlapped, including *MYC*, *CCND3,* and *TFEB.* However, there were only 38 CNV genes that were common between this study and another canine whole exome sequencing

study [11]. This could be due to variation across canine breeds for OSA samples and/or differences in the protocols used for bioinformatics analyses.

We also identified significant correlations between the transcript levels of 257 genes (13 cancer genes and 244 additional genes) and altered copy number variation amplitude, showing that changes in gene copy number effects the transcription of these genes. This suggests that only about 15% of CNV genes might have altered expression that is biologically relevant. Despite CNVs being identified in both human samples and the current study, the amplification of *MYC* was not significantly correlated with increased expression within the Canine 2.0 microarray data. Similarly, array CGH analysis has identified recurrent *MYC* amplifications in canine OSA that were confirmed by fluorescent *in situ* hybridization but were not associated with elevations in *MYC* transcript expression [40]. Like the functional annotation of small somatic variants, pathway analysis of genes with CNV showed enrichment of MAPK signaling pathway, Autophagy, PI3K-Akt signaling pathway and p53 signaling pathway. All of these signaling pathways are known to be altered in both canine and human OSA.

The differential expression analysis between normal metaphyseal bone and tumor samples identified dysregulated pathways in canine OSA. Although the samples used here were primary tumors, several pathways that pertain to metastatic cancer were identified. One of the major gene sets up-regulated in the tumors included extracellular matrix components which included integrins, collagen, proteoglycan, and glycoprotein genes. ECM dysregulation leads to progression of osteosarcoma and promotes metastasis [41]. When an OSA cell line was cultured

526 in the presence of ECM, the cells exhibited doxorubicin resistance [42] and *TP53* protein levels
527 were decreased, suggesting a mechanism for drug resistance.
528 We also observed downregulation of several gene sets including myogenesis, immune
529 response, and drug transport in OSA compared to normal samples. A
530 component of the myogenesis gene set, *Duchenne muscular dystrophy* or *dystrophin* (*DMD*),
531 was shown to have major structural rearrangements resulting in deletion in about 50% of
532 canine OSA samples [12] analyzed using whole genome sequencing. Although these
533 rearrangements could not be detected in our WES analysis, the average expression of tumor
534 *DMD* was significantly lower than in normal metaphyseal bone samples (Student T-test p value
535 = 0.006). Loss of DMD in *mdx* mice is associated with reduced life span and these mice are
536 prone to spontaneously develop rhabdomyosarcomas [43]. However, the effect of DMD
537 deletion in OSA is still unknown. The other major player in pediatric OSA development is
538 deletion of *RB1*, which occurs in 29% of patients [7]. Although we did observe copy number
539 variation within the chromosomal region of *RB1*, significant transcript loss was not observed.
540 However, the loss of *RB1* in tumor samples may have led to identification of an enriched gene
541 set (RB_P107_DN.V1_UP) which is comprised of genes that are upregulated due to loss of RB1.
542 Based on the core enriched genes identified as RB_P107_DN.V1_UP components, canine OSA
543 had an upregulation of genes that dysregulate cell migration (*SLIT3, slit guidance ligand 3*),
544 extracellular matrix remodeling (*COL6A1, COL5A1, COL16A1, PCOLCE*), cell proliferation and
545 differentiation (*FSTL1, follistatin like 1*) and protein folding (*FKBP10, FKBP prolyl isomerase 10*).
546

The availability of outcome data associated with each canine patient allowed for the analysis and identification of pathways enriched in patients in the 25th (short) and 75th (long) percentiles of DFI. In other words, comparison between patients that took less than 123 days or more than 395 days to develop metastasis identified pathways associated with outcome. One of the three gene sets upregulated in short DFI patients included oncogenic signature gene set, KRAS.50_UP.V1_UP, which includes upregulated genes due to over-expression of oncogenic KRAS. Although we did not identify a mutant variant of KRAS in the tumor samples, significant upregulation of transcription factors downstream of KRAS signaling like *HEY1* and *ETV1* was observed in short DFI patients. ETV1 is known to promote metastasis in several cancer types [44-46], hence elevated levels of *ETV1* in short DFI patients might play a role in accelerating metastasis. In gastric cancer cell lines, upregulated ETV1 induces epithelial to mesenchymal transition and increased invasiveness [47]. Similarly, mouse models of KRAS-driven pancreatic cancer exhibit elevated expression of *Hey1* associated with tumor progression [48, 49]. Although we did not observe an increase in mesenchymal markers like VIM (vimentin) and CDH2 (N-cadherin), possibly due to the small sample size (N=6), the EMT gene set was enriched in tumor samples when compared to normal bone samples. Additionally, another gene set comprised of drug detoxification genes (ABC transporters: ABCC1, ABCC2 and NRF2 pathway genes: GSR, NQO1, HMOX1, GCLC, ME1) was upregulated in short DFI patients. The *ABCC* genes are associated with multi-drug resistance [50] and the NRF2 pathway genes play a role in reducing oxidative stress, thus providing defense against exogenous and endogenous stressors and a survival advantage for cancer cells [51]. Drugs such as AEM1 and ML385 that disrupt the

activated Nrf2 pathway, have been tested in pre-clinical cancer studies [52]. Taken together, the study of NRF2 pathway in aggressive OSA can provide us with new avenues of treatment.

Clinical trials over the past decade have dramatically increased the relevance of cancer immunotherapy, demonstrating immune checkpoint inhibitors (ICI) to be an effective treatment modality in multiple human tumor types [53-57]. Despite a high degree of genomic instability in OS, this unfortunately has not translated to observable clinical benefit of ICI therapy in these patients, underscoring an immediate need for additional pre-clinical models that allow investigation of novel therapeutic combinations that could this paradigm [7, 58-60]. The negative results of ICI trials in OS are surprising, given the prior clinical success of the macrophage-activating innate immune stimulant Liposomal-Muramyl TriPeptide-PhosphatidylEthanolamine (L-MTP-PE) immunotherapy in this disease. L-MTP-PE was developed to stimulate the antitumor effects of monocytes and macrophages, and has resulted in longer overall survival in both canine and human osteosarcoma patients in combination with chemotherapy [61, 62]. Due to the extensive comparative similarities between canine and human OS [63, 64]; it is very likely that clinical studies in dogs with OS also hold significant potential to inform novel combination approaches for human immunotherapy studies in this tumor type. To this end, we sought to characterize the immune landscape of canine OS as a necessary prerequisite to informed immunotherapy trial design in this species. In this study, we have identified several immune response related gene sets, primarily associated with an effector\cytotoxic T cell response, which are upregulated or enriched in patients with long DFI, or stated another way, decreased immune response was associated with poor outcome. Similar

590 to our findings, a cross-species (human, dog, and mouse) transcriptome comparison conducted
591 by Scott et al. 2018, revealed a common gene cluster profiles across the three species
592 modulated in osteosarcoma tumors and cell lines [65]. Similar to the findings in this study, the
593 authors identified an association between lower transcript levels of immune-related genes and
594 poor clinical outcome. Their data further suggests that there is an association between
595 increased immune cell infiltration, namely T cells, that may be indicative of more effective
596 immune surveillance associated with prolonged impedance of progression to metastasis.
597 Due to a lack of comprehensive immunophenotyping antibody reagents in dogs, multiple
598 methods were used to comprehensively assess tumor immune cell infiltration in this study,
599 including: 1) quantitative IHC analysis of intra-tumor T cell and myeloid cell densities, and 2)
600 transcriptomic profiling of previously identified immune cell subsets by two separate methods
601 [23, 24]. The IHC staining scores for CD3+ T cells significantly correlated with transcriptomic
602 scores for CD8 T-cells and cytolytic activity, but surprisingly, there was no correlation between
603 mutational burden and the composition of the immune landscape of our canine OSA tumors,
604 specifically with regard to T cell density and cytolytic activity. However, these findings are
605 consistent with those observed in a recently published study on the immune landscape of OSA
606 human patients [15], and suggest that, in contrast to other tumor types, the degree of
607 mutational burden is likely to be a poor indicator for the presence of pre-existing anti-tumor
608 immunity or immune therapy response in OSA.
609
610 Importantly, the significant positive correlation between transcriptomic CD8 T cell and cytolytic
611 activity with CD3 immunolabeling suggests that when present, T cell infiltrates in canine OSA

tumors are primarily of an effector CD8 cytotoxic T cell phenotype, similar to their human counterparts. These data also demonstrate that while comprehensive immunophenotyping reagents for prospective immunotherapy trial design and correlative studies in dogs may be limited, CD3 IHC is a feasible and cost-effective surrogate for baseline anti-tumor immune response assessment in dogs. Furthermore, we observed significant associations between CD3 T cell infiltration and co-inhibitory immune checkpoint expression, including PD-L1, suggesting that when immune responses are present, similar mechanisms of T cell immune exhaustion and adaptive immune resistance are at play between human and canine tumors. While our study did not find any association between increased cytolytic activity score and prolonged disease-free interval in dogs with OSA, this could possibly be due to low sample size, as previous studies demonstrating this association in human patients were done in high-dimensional data [23, 66]. Additionally, Wu et al. 2020 reported that the ESTIMATE immune score is negatively correlated with the total number of deleted genes in human OS patients [15]. However, in our data we did not observe this association. This could be due to the method for identification of deleted genes in our study (whole exome sequencing) when compared to the human OSA study (whole genome sequencing). Nonetheless, taken together, results of our gene set enrichment analysis and immunogenomic profiling suggest the presence of both distinct subsets of so-called immunologically “cold” and “hot” tumors in canine OSA, and utilization of similar mechanisms of adaptive immune resistance in these tumors, providing an important foundation and rationale for designing novel combination immunotherapy studies in dogs as a translational strategy to improve solid tumor immunotherapy.

CONCLUSIONS

This study analyzed the genomic and transcriptomic profiles of 26 tumors from canine OSA patients and associated them with outcome data. Similar to other canine OSA studies, we have identified *TP53* and *SETD2* as the most recurrently mutated genes. We found that in dogs treated with doxorubicin and/or carboplatin, wild type/NULL *TP53* mutation status was associated with short DFI. The transcriptome profile of normal and tumor samples identified significant upregulation of extracellular matrix genes while myogenesis was downregulated in tumors. Additionally, there was decreased expression of immune response genes and higher expression of KRAS and NFE2L2 targets in tumors from patients with a short DFI. The immune-genomic profiling of tumors showed association of immune checkpoint genes and T cell infiltration, but no association with genomic criteria such as mutational burden.

METHODS

Sample processing - Bone tumors and matched normals

Tumors from previously untreated dogs with OSA were collected along with blood, peripheral blood mononuclear cells (PBMC) or stroma as matched normal samples **(Additional File 1; Table S1A**). The samples were flash frozen and kept at -80^{o} C until processed for whole exome sequencing and microarray analysis. Following the manufacturer's protocol for TRIzol (Invitrogen), genomic DNA was extracted from 26 primary tumors, one metastatic tumor, and 26 matched canine blood, PBMC or stromal samples, and DNA was cleaned using DNeasy or QiaAMP DNA Blood mini kits (Qiagen). RNA was extracted from the same 26 tumor samples and 8 normal canine bone samples using TRIzol (Invitrogen) followed by RNeasy cleanup (Qiagen)

for microarray analysis **(Additional File 1; Table S1B**). Both RNA and DNA were quantified on a NanoDrop Microvolume Spectrophotometer and quality was assessed by TapeStation or Bioanalyzer (Agilent).

The whole exome DNA library was created and genomic exonic regions were captured using the Agilent SureSelect XT All Exon Canine V2 (part number: 931198, Santa Clara, CA) capture kit. This capture encompasses 43.45 Mb of canine exonic regions. The SureSelectXT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library kit was used to create the genomic DNA library. The resultant library was sequenced on an Illumina HiSEQ4000 sequencer generating 151 bp paired end reads.
RNA samples with a RIN value >8 were analyzed on GeneChip Canine 2.0 Genome Arrays (Affymetrix) in five separate batches.

Whole exome sequence data processing

Mapping Illumina reads with BWA

The Illumina reads were processed to identify somatic single nucleotide variants (SNVs), insertions and deletions (INDELs), and copy number variations (CNVs). The 150 bp reads were trimmed to eliminate low-quality (phred score <20) and adapter sequences by using the Trimmomatic tool (v0.36) [67]. Both the raw and trimmed/high-quality reads were assessed by FastQC (v0.11.5) [68]. The reads were then mapped against the canine genome (CanFam3.1) using the BWA (v0.7.15-r1140) tool [69]. The binary alignment files (BAM) were processed to mark duplicates and recalibrate bases prior to variant calling as per GATK (v4.1.2.0) best-practices [70].

679 *Indel and SNV calling using Mutect2 and post-processing of somatic variants*

680 The short somatic variants, SNVs and INDELs, were called from the BAM files using Mutect2. In
681 addition to using a matched normal sample for each tumor, we created a panel of normals
682 using 43 in-house samples (**Additional File 1; Table S1A**). We also used 90 million population
683 variants that were called from 722 dogs as the germline resource option within Mutect2 for
684 calling somatic variants [18]. The variants were filtered using filterMutectCalls and variants
685 with a PASS notation in the FILTER column were characterized as somatic variants. The variants
686 were annotated using Ensembl Variant Effect Predictor (VEP, Ensembl version 99) and the VCF
687 file was converted to MAF (Mutation annotation file) format using the perl code: vcf2maf.pl
688 (https://github.com/mskcc/vcf2maf).

689

690 The mutational signature of each sample was deduced by using the MutationalPatterns R
691 package (v1.12.0) [71]. Using the mut_mat function the count matrix of 96 trinucleotide
692 changes was derived from sample VCF files. To estimate the factorization rank, the non-
693 negative matrix factorization (NMF) algorithm (NMF R package) was used. The *de novo*
694 mutational signature was derived using the extract_signature function within the
695 MutationalPatterns package. This resulted in decomposition of the mutation count matrix into
696 three top ranked signatures and estimated the relative contribution within each sample. The
697 sample signatures were compared to 30 known COSMIC signatures by calculating the cosine
698 similarity between the NMF signatures and known cancer signatures.

699

The somatic variants were also processed to bin protein coding and cancer gene variants. The gene variants with HGVS.p variable notations were selected as protein coding genes. From this list, COSMIC (version 91) was used to identify variants within known cancer genes [72]. The oncoplots were plotted using R package ComplexHeatmap (v2.2.0). Pathway analysis of the mutated genes was conducted using DAVID Functional Annotation Bioinformatics Microarray Analysis (v6.8) tool and Enrichr to functionally annotate genes against Gene Ontology and KEGG databases [73, 74].

Copy number variant calling using Sequenza

The allele-specific copy number variants were assessed using the Sequenza (v2.1.9999b1) tool [75]. Briefly, the sequenza-utils bam2seqz function was used along with paired BAM files (tumor and normal) to extract loci with A (major) and B (minor) allele frequencies. The sequenza R package used the output of bam2seqz function for GC content normalization of tumor normal depth ratio, allele-specific segmentation using the copynumber R package, model fitting to infer cellularity and ploidy parameters, and copy number profiles of tumors. The resulting segmentation file was annotated using the default options in the GISTIC2.0 tool to identify genes in the recurrently copy number altered regions of the tumor genome. A q-value cut-off of 0.01 was used to generate the final list of amplification and deletion peaks. The genes associated with these peaks were identified and cancer genes (COSMIC v91) were selected for plotting and cross-species comparison.

Microarray data processing

Normalization and differential expression analysis using RMA and limma

Following the manufacturer's standard protocol, RNA from 26 tumor and 8 normal bone samples were transcribed to cDNA, labelled and hybridized onto GeneChip™ Canine Genome 2.0 Array (ThemoFisher Scientific, catalog number: 900727), and scanned on an Affymetrix Scanner 3000. The data files (CEL) were processed in R for normalization of probes and differential expression of genes. Using simpleaffy R package (v2.62.0), the CEL files for both normal and tumor samples were normalized by Robust Multi-Array Average (RMA) method, which generated $\log_2$ expression values for 43,035 probes. The differential expression analyses between normal and tumor expression were carried out using limma (v3.42.2), which used the Empirical Bayes method for model fitting [22]. The significantly differentially expressed genes (DEGs) were selected using a q-value (false discovery rate corrected p-value) of <0.05 and $\log_2$ fold change of >2 for upregulated and <(-2) for downregulated genes, respectively.

GSEA analysis

Identification of enriched gene sets and pathways were conducted using the Gene Set Enrichment Analysis (GSEA, v3.0) tool [76]. The genes sets used for this analysis were part of MSigDB database, which included Hallmark (50 gene sets), Canonical pathways (2199 gene sets), GO biological process (7350 gene sets), GO molecular function (1645 gene sets), oncogenic signatures (189 gene sets), immunologic signatures (browse 4872 gene sets), and chemical and genetic perturbations (3302 gene sets) [77]. Functional annotation of differentially expressed genes from tumor and normal bone comparison was done using the GSEAPreranked tool within GSEA program (FDR <0.05). In addition, we ran GSEA analysis on

samples (n=6) in the 25th (T-856, T-907, T-1087, T-458, T-276, T-1097) and 75th (T-74, T-1166, T-346, T-399, T-29C, T-1246) percentile of DFI. Transcript expression of 15,563 genes were used as input for this analysis and enriched genes sets were identified in short (25th percentile) and long (75th percentile) DFI patients (FDR <0.05). The visualization of core enriched gene expression from selected significant gene sets were plotted as heatmaps using pheatmap (v1.0.12) R package [78].

Immune cell characterization

Immunohistochemistry

Archived, formalin-fixed, paraffin embedded (FFPE) tissue samples were obtained from the Colorado State University Flint Animal Cancer Center Tissue Archive. Available paraffin blocks were routinely processed for hematoxylin and eosin (H&E) staining, as well as immunohistochemistry. The H&E stained slides were evaluated by a board-certified pathologist (DPR) to confirm diagnosis and the presence of adequate viable tumor sample for IHC analysis. Immunohistochemistry was performed via routine, automated methods on the Leica Bond Max autostainer (Leica Biosystems Inc.), with the following panel of previously published canine cross-reactive antibodies: rabbit polyclonal anti-human CD3 (pan T lymphocyte marker; Dako, clone A0452), and monoclonal mouse anti-human Myeloid/Histiocyte antigen (MAC387; monocytes/macrophages; Dako, clone MAC387). Antigen retrieval was performed using Leica Epitope Retrieval 2 (Tris-EDTA buffer, pH 9) for 20 min. Detection was performed with PowerVision IHC detection systems (Leica Biosystems, Inc.), using a polymeric alkaline phosphatase anti-mouse IgG (MAC387) or anti-rabbit IgG (CD3) and Fast Red chromogen.

Whole slide brightfield 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 open-source ImageJ software (National Institutes of Health). Parent images were converted to gray scale .tiff images for analysis. Tumor tissue regions-of-interest (ROIs) were segmented from adjacent normal tissue by manual outlining in ImageJ in blinded fashion by a board-certified veterinary pathologist. Following determination of the ROI, positively labeled infiltrating immune cells were counted using the color deconvolution algorithm. Briefly, a positive pixel threshold for all immune cell markers was determined visually by a veterinary pathologist using appropriate isotype-stained control slides. Images were subjected to color deconvolution, followed by global, automated application of this intensity threshold to all images. Following automated image analysis, positive pixel masks of each image were blindly evaluated by a pathologist to ensure thresholding accuracy. Data was analyzed and the number of infiltrating immune cells was expressed as a percentage of total tumor tissue area.

Gene expression profiling of immune cells

In the last few years, multiple groups have characterized immune cell types using gene expression profile. Using the gene sets for immune cell types provided by Rooney et al 2015, we have calculated a score for immune cell types in each sample [23]. A score was calculated as geometric mean of gene expression for signature genes of each cell type (**Additional File 1, Table S10)**. These scores were used to assess associations with DFI, mutations per megabase

(tumor burden), and immune staining quantification data for T-cells and myeloid cells using Pearson correlation. The significance of multiple correlations was corrected for false discovery rates using Benjamini-Hochberg method (R package fdrtool, v1.2.15). The immune cell profiling for tumor microenvironment was also carried out by using CIBERSORT tool [24]. Using transcript expression data as input, deconvolution of 22 functionally defined immune cell types was carried out [27]. A third method to calculate immune cell infiltration was done by using the ESTIMATE (v1.0.13) R package for each tumor samples [25].

Survival analyses and graphical plots

The association of *TP53* mutation status, monocyte counts, and IHC staining scores with patient DFI was assessed using Kaplan-Meier survival analyses (R packages used: survival, v.3.2-3 and survminer, v.0.4.8). The univariate Cox proportional hazard regression (R package: survival, v.3.2-3) was used to identify significant association of clinical parameters and *TP53* mutation status to patient outcome. A p value cut-off of 0.2 was used to select factors for running the stepwise algorithm. The stepwise forward multivariate COXPH regression was carried out by R package My.stepwise (v0.1.0). The significance level for model entry and exit was 0.1. The final model was comprised of factors what were significantly associated with outcome at a p-value less than 0.05.

Data processing and availability

The whole exome data was processed on the RMACC Summit supercomputer [79]. The Illumina raw fastq files were submitted to NCBI Short Read Archive (SRA) database. The canine osteosarcoma and normal samples have been submitted to Bioproject PRJNA613479 and

PRJNA503860, respectively. The microarray data can be downloaded from Gene Expression Omnibus (GEO) database using accession number GSE76127. The tools, databases, and their versions, in addition to the bioinformatic pipeline used in this study can be accessed here: https://github.com/sdas2019/Canine-Osteosarcoma-Whole-Exome-Sequencing-Pipeline.

Figure legends.

Fig. 1. Mutational signature profile in canine osteosarcoma. A. Distribution of total short variants (SNVs and INDELs) across 26 samples. The samples were sorted in ascending order of disease-free interval (DFI) and binned in three discrete groups. B. Distribution of mutation types as identified by VEP across 26 samples. Missense mutations were the most common type of short variant in this dataset. C. Mutation signature analyses showing the relative distribution of six single nucleotide changes in 96 different trinucleotide contexts across three *de novo* signatures. D. Relative contribution of three predominant signatures across 26 samples.

Fig. 2. Oncoplot of cancer census genes with variants in canine osteosarcoma samples. A. The genes with short variants (SNVs and INDELs) are plotted along with sample characteristics like disease free interval (DFI) and tumor location. The cancer genes were categorized as oncogenes and tumor suppressor genes in accordance to the Cancer Gene Census. B. The genes with recurrent copy number variations (amplifications and deletions) were plotted. The samples were sorted in ascending order of DFI in both plots. Gene names colored red indicate significant correlation of copy number amplitude and transcript expression.

Fig.3. Graphical representation of differentially expressed genes (DEG). A. Heatmap of DEGs between normal bone (N-) and canine osteosarcoma samples (T-). B-D. The heatmap of core-enriched genes associated with three significantly enriched Hallmark gene sets in tumor samples in comparison to normal bone, as identified by GSEA analysis.

Fig.4. Association of TP53 mutation status and clinical parameters with outcome. A. Kaplan Meier plots for disease free interval (DFI) in dogs with osteosarcoma stratified by TP53 mutation status. The dogs with wild type/null TP53 status had a shorter DFI compared to dogs with TP53 missense mutations. B. Kaplan Meier plots for dogs with osteosarcoma stratified by pre-treatment monocyte count (> or < $0.04*10^3$ cells/µl). Higher monocyte counts were associated with short DFI patients.

Fig 5. Representative figures of immunohistochemical staining of canine osteosarcoma samples. A-B. These samples were stained for CD3+ T-cell infiltration in A (low levels – T-1023) and B (high levels – T-1162). C-D. These samples were stained with the MAC387 antibody against S100A9 to distinguish macrophage infiltration. C represents low levels (T-1162) and D represents high levels (T-856). E-F. Quantification of immunohistochemical labelling of T-cell (CD3+ cells) and Macrophages (MAC387 cells) across 19 and 22 samples, respectively.

Fig.6. Profiling of immune cell infiltration using gene expression data. A. Pearson correlation matrix of immune cell expression score and immune staining levels of T-cell (IHC percent CD3+ cells) and macrophages (IHC percent MAC387 cells). We also correlated mutations per megabases (MB) and disease-free interval (DFI) with other variables. Only the significant (FDR <0.05) correlation values are plotted. B. Quantification of 22 immune cell types using CIBERSORT. The samples are sorted by DFI. The most prevalent cell type in the tumor microenvironment were M0 and M2 macrophages. C. Immune score quantification using ESTIMATE R package.

References

1. Withrow SJ, Khanna C: **Bridging the gap between experimental animals and humans in osteosarcoma.** *Cancer Treat Res* 2009, **152:**439-446.
2. Makielski KM, Mills LJ, Sarver AL, Henson MS, Spector LG, Naik S, Modiano JF: **Risk Factors for Development of Canine and Human Osteosarcoma: A Comparative Review.** *Vet Sci* 2019, **6**.
3. Gustafson DL, Duval DL, Regan DP, Thamm DH: **Canine sarcomas as a surrogate for the human disease.** *Pharmacol Ther* 2018, **188:**80-96.
4. Anfinsen KP, Grotmol T, Bruland OS, Jonasdottir TJ: **Breed-specific incidence rates of canine primary bone tumors--a population based survey of dogs in Norway.** *Can J Vet Res* 2011, **75:**209-215.
5. Longhi A, Errani C, De Paolis M, Mercuri M, Bacci G: **Primary bone osteosarcoma in the pediatric age: state of the art.** *Cancer Treat Rev* 2006, **32:**423-436.
6. Selmic L, Lafferty M, Kamstock D, Garner A, Ehrhart N, Worley D, Withrow S, Lana S: **Outcome and prognostic factors for osteosarcoma of the maxilla, mandible, or calvarium in dogs: 183 cases (1986-2012).** *Journal of the American Veterinary Medical Association* 2014, **245**.
7. Chen X, Bahrami A, Pappo A, Easton J, Dalton J, Hedlund E, Ellison D, Shurtleff S, Wu G, Wei L, et al: **Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma.** *Cell Rep* 2014, **7:**104-112.
8. Perry JA, Kiezun A, Tonzi P, Van Allen EM, Carter SL, Baca SC, Cowley GS, Bhatt AS, Rheinbay E, Pedamallu CS, et al: **Complementary genomic approaches highlight the**

PI3K/mTOR pathway as a common vulnerability in osteosarcoma. *Proc Natl Acad Sci U S A* 2014, **111:**E5564-5573.

9. Kovac M, Blattmann C, Ribi S, Smida J, Mueller NS, Engert F, Castro-Giner F, Weischenfeldt J, Kovacova M, Krieg A, et al: **Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency.** *Nat Commun* 2015, **6:**8940.
10. Sayles LC, Breese MR, Koehne AL, Leung SG, Lee AG, Liu HY, Spillinger A, Shah AT, Tanasa B, Straessler K, et al: **Genome-Informed Targeted Therapy for Osteosarcoma.** *Cancer Discov* 2019, **9:**46-63.
11. Sakthikumar S, Elvers I, Kim J, Arendt ML, Thomas R, Turner-Maier J, Swofford R, Johnson J, Schumacher SE, Alfoldi J, et al: **SETD2 Is Recurrently Mutated in Whole-Exome Sequenced Canine Osteosarcoma.** *Cancer Res* 2018, **78:**3421-3431.
12. Gardner HL, Sivaprakasam K, Briones N, Zismann V, Perdigones N, Drenner K, Facista S, Richholt R, Liang W, Aldrich J, et al: **Canine osteosarcoma genome sequencing identifies recurrent mutations in DMD and the histone methyltransferase gene SETD2.** *Commun Biol* 2019, **2:**266.
13. Wedekind M, Wagner L, Cripe T: **Immunotherapy for osteosarcoma: Where do we go from here?** *Pediatric blood & cancer* 2018, **65**.
14. Yu Y, Zhang H, Ren T, Huang Y, Liang X, Wang W, Niu J, Han Y, Guo W: **Development of a prognostic gene signature based on an immunogenomic infiltration analysis of osteosarcoma.** *Journal of cellular and molecular medicine* 2020, **24:**11230-11242.
15. Wu C, Beird H, Andrew LJ, Advani S, Mitra A, Cao S, Reuben A, Ingram D, Wang W, Ju Z, et al: **Immuno-genomic landscape of osteosarcoma.** *Nature communications* 2020, **11**.
16. Koirala P, Roth M, J G, Piperdi S, Chinai J, Geller D, Hoang B, Park A, Fremed M, Zang X, Gorlick R: **Immune infiltration and PD-L1 expression in the tumor microenvironment are prognostic in osteosarcoma.** *Scientific reports* 2016, **6**.
17. Wu C, Livingston J: **Genomics and the Immune Landscape of Osteosarcoma.** In *Advances in experimental medicine and biology. Volume* 1258. Edited by Eugenie S. Kleinerman RG: Adv Exp Med Biol; 2020: 21-36
18. Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, Decker B, Parker HG, Ostrander EA: **Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology.** *Nat Commun* 2019, **10:**1489.
19. Alexandrov LB, Stratton MR: **Mutational signatures: the patterns of somatic mutations hidden in cancer genomes.** *Curr Opin Genet Dev* 2014, **24:**52-60.
20. Zhao Y, Gregory M, Biertümpfel C, Hua Y, Hanaoka F, Yang W: **Mechanism of somatic hypermutation at the WA motif by human DNA polymerase η.** *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**.
21. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G: **GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers.** In *Genome Biol. Volume* 12; 2011: R41
22. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **Limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015, **43:**e47.

23. Rooney M, Shukla S, Wu C, Getz G, Hacohen N: **Molecular and genetic properties of tumors associated with local immune cytolytic activity.** *Cell* 2015, **160**.
24. Newman A, Liu C, Green M, Gentles A, Feng W, Xu Y, Hoang C, Diehn M, Alizadeh A: **Robust enumeration of cell subsets from tissue expression profiles.** *Nature methods* 2015, **12**.
25. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird P, Levine D, et al: **Inferring tumour purity and stromal and immune cell admixture from expression data.** *Nature communications* 2013, **4**.
26. Vanherberghen M, Day M, Delvaux F, Gabriel A, Clercx C, Peeters D: **An immunohistochemical study of the inflammatory infiltrate associated with nasal carcinoma in dogs and cats.** *Journal of comparative pathology* 2009, **141**.
27. Chen B, Khodadoust M, Liu C, Newman A, Alizadeh A: **Profiling Tumor Infiltrating Immune Cells with CIBERSORT.** *Methods in molecular biology (Clifton, NJ)* 2018, **1711**.
28. Gröbner S, Worst B, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva V, Johann P, Balasubramanian G, Segura-Wang M, Brabetz S, et al: **The Landscape of Genomic Alterations Across Childhood Cancers.** *Nature* 2018, **555**.
29. Kirpensteijn J, Kik M, Teske E, Rutteman G: **TP53 gene mutations in canine osteosarcoma.** *Veterinary surgery : VS* 2008, **37**.
30. Yokouchi H, Nishihara H, Harada T, Yamazaki S, Kikuchi H, Oizumi S, Uramoto H, Tanaka F, Harada M, Akie K, et al: **Detection of somatic TP53 mutation in surgically resected small-cell lung cancer by targeted exome sequencing: association with longer relapse-free survival.** *Heliyon* 2020, **6**.
31. Esserman L, Berry D, Cheang M, Yau C, Perou C, Carey L, DeMichele A, Gray J, Conway-Dorsey K, Lenburg M, et al: **Chemotherapy response and recurrence-free survival in neoadjuvant breast cancer depends on biomarker profiles: results from the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657).** *Breast cancer research and treatment* 2012, **132**.
32. Wang Y, Y X, Chen J, Ouyang T, Li J, Wang T, Fan Z, Fan T, Lin B, Xie Y: **TP53 mutations are associated with higher rates of pathologic complete response to anthracycline/cyclophosphamide-based neoadjuvant chemotherapy in operable primary breast cancer.** *International journal of cancer* 2016, **138**.
33. Chen M, Zhu Y, Xu J, Wang L, Liu C, Ji Z, Lu P: **Value of TP53 status for predicting response to neoadjuvant chemotherapy in breast cancer: a meta-analysis.** *PloS one* 2012, **7**.
34. Bertheau P, Plassa F, Espié M, Turpin E, de Roquancourt A, Marty M, Lerebours F, Beuzard Y, Janin A, de Thé H: **Effect of mutated TP53 on response of advanced breast cancers to high-dose chemotherapy.** *Lancet (London, England)* 2002, **360**.
35. Bertheau P, Turpin E, Rickman D, Espié M, de Reyniès A, Feugeas J, Plassa L, Soliman H, Varna M, de Roquancourt A, et al: **Exquisite sensitivity of TP53 mutant and basal breast cancers to a dose-dense epirubicin-cyclophosphamide regimen.** *PLoS medicine* 2007, **4**.
36. Jackson J, Pant V, Li Q, Chang L, Quintás-Cardama A, Garza D, Tavana O, Yang P, Manshouri T, Li Y, et al: **p53-mediated senescence impairs the apoptotic response to chemotherapy and clinical outcome in breast cancer.** *Cancer cell* 2012, **21**.

37. Liu D, Song H, Xu Y: **A common gain of function of p53 cancer mutants in inducing genetic instability.** *Oncogene* 2010, **29**.
38. Sottnik J, Rao S, Lafferty M, Thamm D, Morley P, Withrow S, Dow S: **Association of blood monocyte and lymphocyte count and disease-free interval in dogs with osteosarcoma.** *Journal of veterinary internal medicine* 2010, **24:**1439-1444.
39. Chen R, Zhao W, Fang C, Yang X, Ji M: **Histone Methyltransferase SETD2: A Potential Tumor Suppressor in Solid Cancers.** *Journal of Cancer* 2020, **11**.
40. Angstadt AY, Motsinger-Reif A, Thomas R, Kisseberth WC, Guillermo Couto C, Duval DL, Nielsen DM, Modiano JF, Breen M: **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* 2011, **50:**859-874.
41. Cui J, Dean D, Hornicek FJ, Chen Z, Duan Z: **The role of extracelluar matrix in osteosarcoma progression and metastasis.** *Journal of Experimental & Clinical Cancer Research* 2020, **39:**1-11.
42. Harisi R, Dudas J, Nagy-Olah J, Timar F, Szendroi M, Jeney A: **Extracellular matrix induces doxorubicin-resistance in human osteosarcoma cells by suppression of p53 function.** *Cancer biology & therapy* 2007, **6**.
43. Chamberlain J, Metzger J, Reyes M, Townsend D, Faulkner J: **Dystrophin-deficient mdx mice display a reduced life span and are susceptible to spontaneous rhabdomyosarcoma.** *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2007, **21**.
44. Heeg S, Das K, Reicher tM, Bakir B, Takano S, Caspers J, Aiello N, Wu K, Neesse A, Maitra A, et al: **ETS-Transcription Factor ETV1 Regulates Stromal Expansion and Metastasis in Pancreatic Cancer.** *Gastroenterology* 2016, **151**.
45. de Launoit Y, Chotteau-Lelievre A, Beaudoin C, Coutte L, Netzer S, Brenner C, Huvent I, Baert J: **The PEA3 group of ETS-related transcription factors. Role in breast cancer metastasis.** *Advances in experimental medicine and biology* 2000, **480**.
46. Chi P, Chen Y, Zhang L, Guo X, Wongvipat J, Shamu T, Fletcher JA, Dewell S, Maki RG, Zheng D, et al: **ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumours.** *Nature* 2010, **467:**849-853.
47. Li Z, Zhang L, Ma Z, Yang M, Tang J, Fu Y, Mao Y, Hong X, Zhang Y: **ETV1 induces epithelial to mesenchymal transition in human gastric cancer cells through the upregulation of Snail expression.** *Oncology reports* 2013, **30**.
48. Bailey J, Alsina J, Rasheed Z, McAllister F, Fu Y, Plentz R, Zhang H, Pasricha P, Bardeesy N, Matsui W, et al: **DCLK1 marks a morphologically distinct subpopulation of cells with stem cell properties in preinvasive pancreatic cancer.** *Gastroenterology* 2014, **146**.
49. Maniati E, Bossard M, Cook N, Candido J, Emami-Shahr iN, Nedospasov S, Balkwill F, Tuveson D, Hagemann T: **Crosstalk between the canonical NF-κB and Notch signaling pathways inhibits Pparγ expression and promotes pancreatic cancer progression in mice.** *The Journal of clinical investigation* 2011, **121**.
50. Slot A, Molinski S, Cole S: **Mammalian multidrug-resistance proteins (MRPs).** *Essays in biochemistry* 2011, **50:**179-207.

51. DeNicola G, Karreth F, Humpton T, Gopinathan A, Wei C, Frese K, Mangal D, Yu K, Yeo C, Calhoun E, et al: **Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis.** *Nature* 2011, **475**.
52. Panieri E, Saso L: **Potential Applications of NRF2 Inhibitors in Cancer Therapy.** *Oxidative medicine and cellular longevity* 2019, **2019**.
53. Robert C, Ribas A, Schachter J, Arance A, Grob J, Mortier L, Daud A, Carlino M, McNeil C, Lotem M, et al: **Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study.** *The Lancet Oncology* 2019, **20**.
54. Motzer R, Escudier B, McDermott D, George S, Hammers H, Srinivas S, Tykodi S, Sosman J, Procopio G, Plimack E, et al: **Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma.** *The New England journal of medicine* 2015, **373**.
55. Bellmunt J, de Wit R, Vaughn D, Fradet Y, Lee J, Fong L, Vogelzang N, Climent M, Petrylak D, Choueiri T, et al: **Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma.** *The New England journal of medicine* 2017, **376**.
56. Le D, Uram J, Wang H, Bartlett B, Kemberling H, Eyring A, Skora A, Luber B, Azad N, Laheru D, et al: **PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.** *The New England journal of medicine* 2015, **372**.
57. Ferris R, Blumenschein G, Fayette J, Guigay J, Colevas A, Licitra L, Harrington K, Kasper S, Vokes E, Even C, et al: **Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck.** *The New England journal of medicine* 2016, **375**.
58. Behjati S, Tarpey P, Haase K, Ye H, Young M, Alexandrov L, Farndon S, Collord G, Wedge D, Martincorena I, et al: **Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma.** *Nature communications* 2017, **8**.
59. Tawbi H, Burgess M, Bolejack V, Van Tine B, Schuetze S, Hu J, D'Angelo S, Attia S, Riedel R, Priebat D, et al: **Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial.** *The Lancet Oncology* 2017, **18**.
60. D'Angelo S, Mahoney M, Van Tine B, Atkins J, Milhem M, Jahagirdar B, Antonescu C, Horvath E, Tap W, Schwartz G, Streicher H: **Nivolumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2 trials.** *The Lancet Oncology* 2018, **19**.
61. Meyers P: **Muramyl Tripeptide-Phosphatidyl Ethanolamine Encapsulated in Liposomes (L-MTP-PE) in the Treatment of Osteosarcoma.** In *Current Advances in Osteosarcoma. Volume* 1257: Adv Exp Med Biol; 2020: 133-139
62. Meyers P, Schwartz C, Krailo M, Healey J, Bernstein M, Ferguson W, Gebhardt M, Goorin A, Harris M, Kleinerman E, et al: **Osteosarcoma: the addition of muramyl tripeptide to chemotherapy improves overall survival--a report from the Children's Oncology Group.** *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008, **26**.
63. Paoloni M, Davis S, Lana S, Withrow S, Sangiorgi L, Picci P, Hewitt S, Triche T, Meltzer P, Khanna C: **Canine tumor cross-species genomics uncovers targets linked to osteosarcoma progression.** *BMC genomics* 2009, **10**.
64. Fan T, Khanna C: **Comparative Aspects of Osteosarcoma Pathogenesis in Humans and Dogs.** *Veterinary sciences* 2015, **2**.

65. Scott M, Temiz N, Sarver A, LaRue R, Rathe S, Varshney J, Wolf N, Moriarity B, O'Brien T, Spector L, et al: **Comparative Transcriptome Analysis Quantifies Immune Cell Transcript Levels, Metastatic Progression, and Survival in Osteosarcoma.** *Cancer research* 2018, **78**.
66. Zaravinos A, Roufas C, Nagara M, de Lucas Moreno B, Oblovatskaya M, Efstathiades C, Dimopoulos C, Ayiomamitis G: **Cytolytic activity correlates with the mutational burden and deregulated expression of immune checkpoints in colorectal cancer.** *Journal of experimental & clinical cancer research : CR* 2019, **38:**364.
67. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics* 2014, **30:**2114-2120.
68. Andrews S: **FastQC A Quality Control tool for High Throughput Sequence Data (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).** 2010.
69. Li H, Durbin R: **Fast and accurate long-read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2010, **26:**589-595.
70. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al: **From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.** *Curr Protoc Bioinformatics* 2013, **43:**11.10.11-33.
71. Blokzijl F, Janssen R, van Boxtel R, Cuppen E: **MutationalPatterns: comprehensive genome-wide analysis of mutational processes.** *Genome medicine* 2018, **10**.
72. Tate J, Bamford S, Jubb H, Sondka Z, Beare D, Bindal N, Boutselakis H, Cole C, Creatore C, Dawson E, et al: **COSMIC: the Catalogue Of Somatic Mutations In Cancer.** *Nucleic acids research* 2019, **47**.
73. Huang da W, Sherman BT, Zheng X, Yang J, Imamichi T, Stephens R, Lempicki RA: **Extracting biological meaning from large gene lists with DAVID.** *Curr Protoc Bioinformatics* 2009, **Chapter 13:**Unit 13.11.
74. Chen E, Tan C, Kou Y, Duan Q, Wang Z, Meirelles G, Clark N, Ma'ayan A: **Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.** *BMC bioinformatics* 2013, **14**.
75. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, Szallasi Z, Eklund AC: **Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data.** *Ann Oncol* 2015, **26:**64-70.
76. Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert B, Gillette M, Paulovich A, Pomeroy S, Golub T, Lander E, Mesirov J: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**.
77. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, JP M: **Molecular signatures database (MSigDB) 3.0.** *Bioinformatics (Oxford, England)* 2011, **27**.
78. **R: The R Project for Statistical Computing.** [https://www.r-project.org/]
79. Anderson J, Burns PJ, Milroy D, Ruprecht P, Hauser T, Siegel HJ: **Deploying RMACC Summit: An HPC Resource for the Rocky Mountain Region.** In *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact*; *07/09/2017*. ACM; 2017: 8.

Figures.

Fig. 1

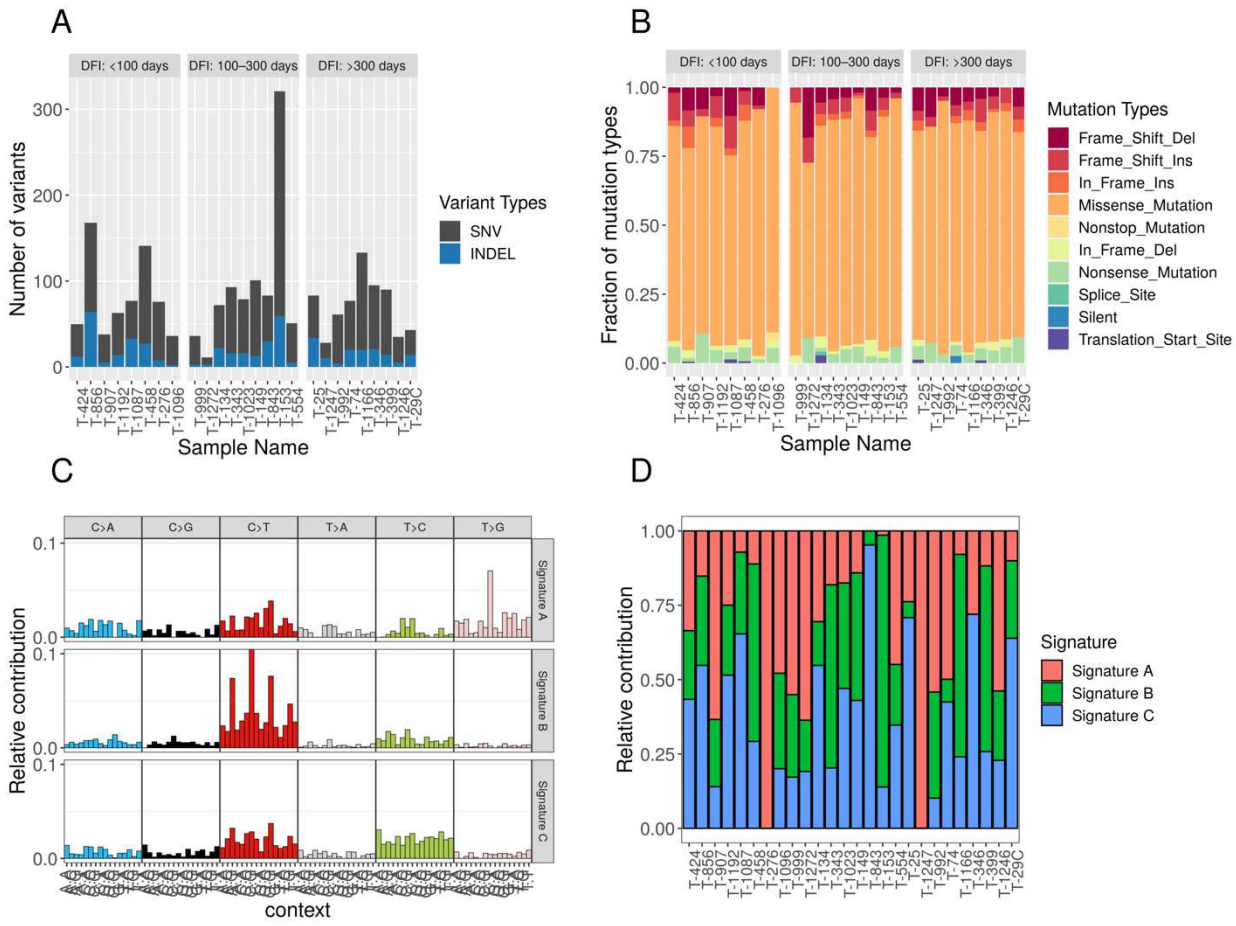


Fig. 1. Mutational signature profile in canine osteosarcoma. A. Distribution of total short variants (SNVs and INDELs) across 26 samples. The samples were sorted in ascending order of disease-free interval (DFI) and binned in three discrete groups. B. Distribution of mutation types as identified by VEP across 26 samples. Missense mutations were the most common type of short variant in this dataset. C. Mutation signature analyses showing the relative distribution of six single nucleotide changes in 96 different trinucleotide contexts across three *de novo* signatures. D. Relative contribution of three predominant signatures across 26 samples.

Fig. 2

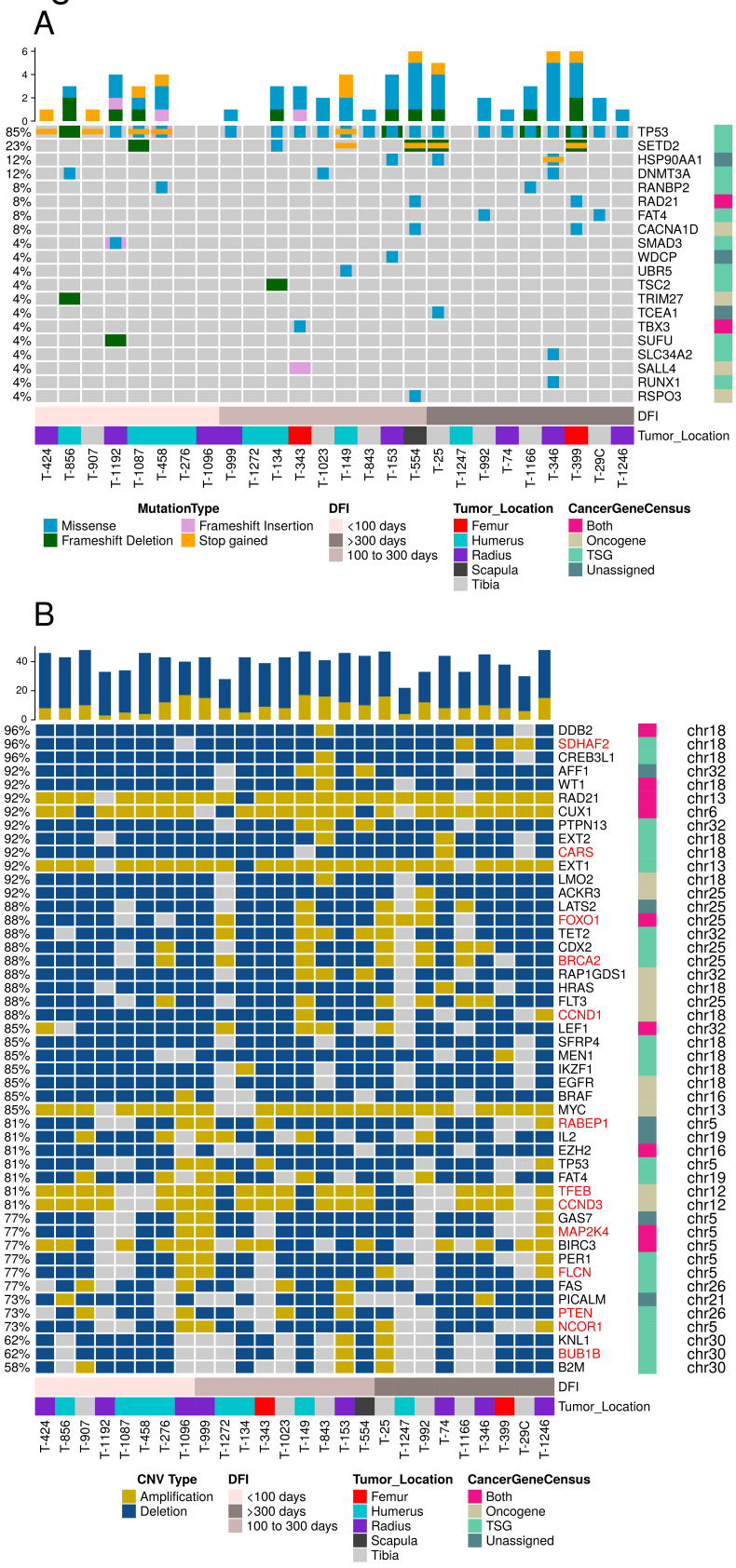
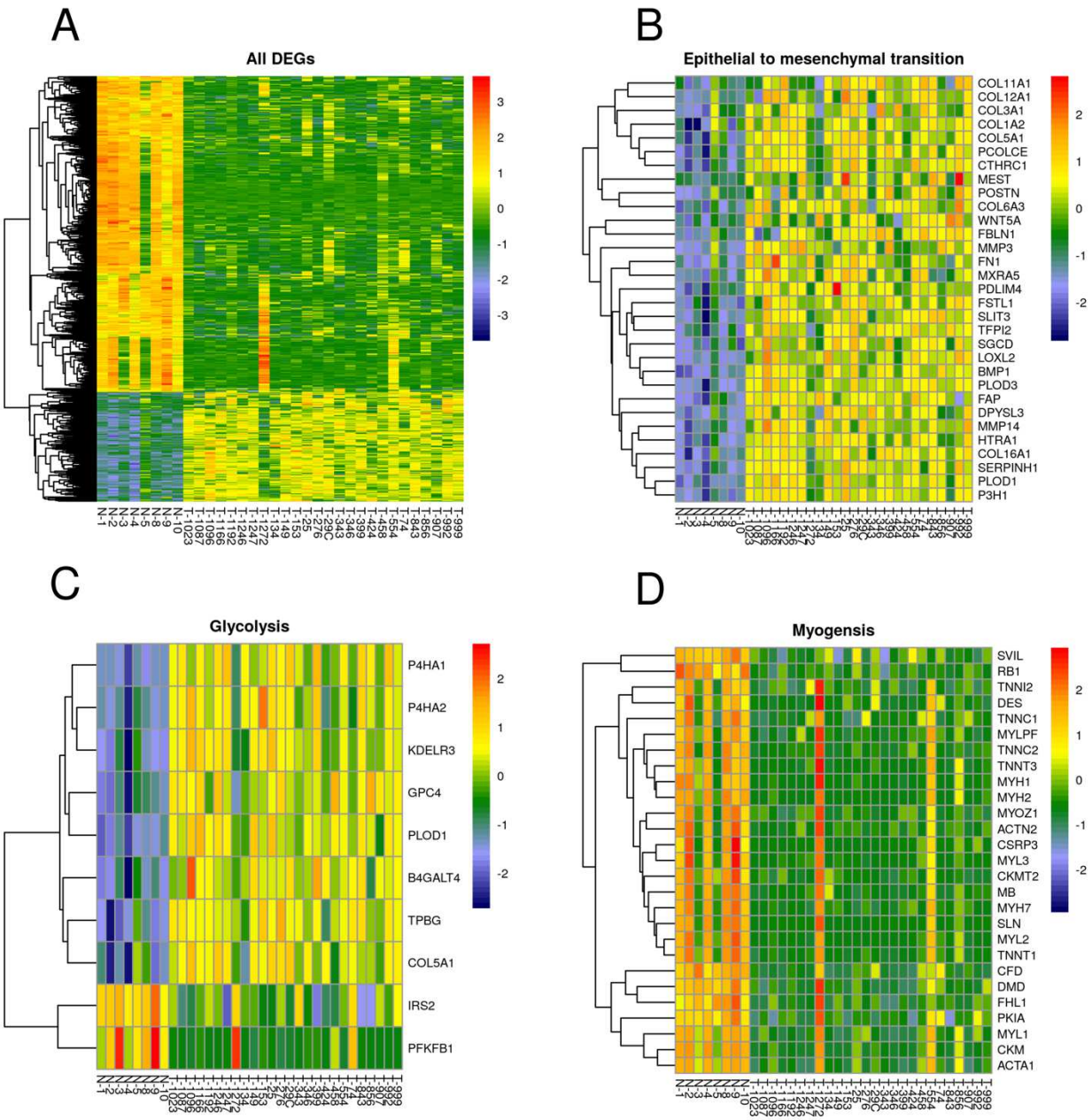


Fig. 2. Oncoplot of cancer census genes with variants in canine osteosarcoma samples. A. The genes with short variants (SNVs and INDELs) are plotted along with sample characteristics like disease free interval (DFI) and tumor location. The cancer genes were categorized as oncogenes and tumor suppressor genes in accordance to the Cancer Gene Census. B. The genes with recurrent copy number variations (amplifications and deletions) were plotted. The samples were sorted in ascending order of DFI in both plots. Gene names colored red indicate significant correlation of copy number amplitude and transcript expression.

Fig. 3

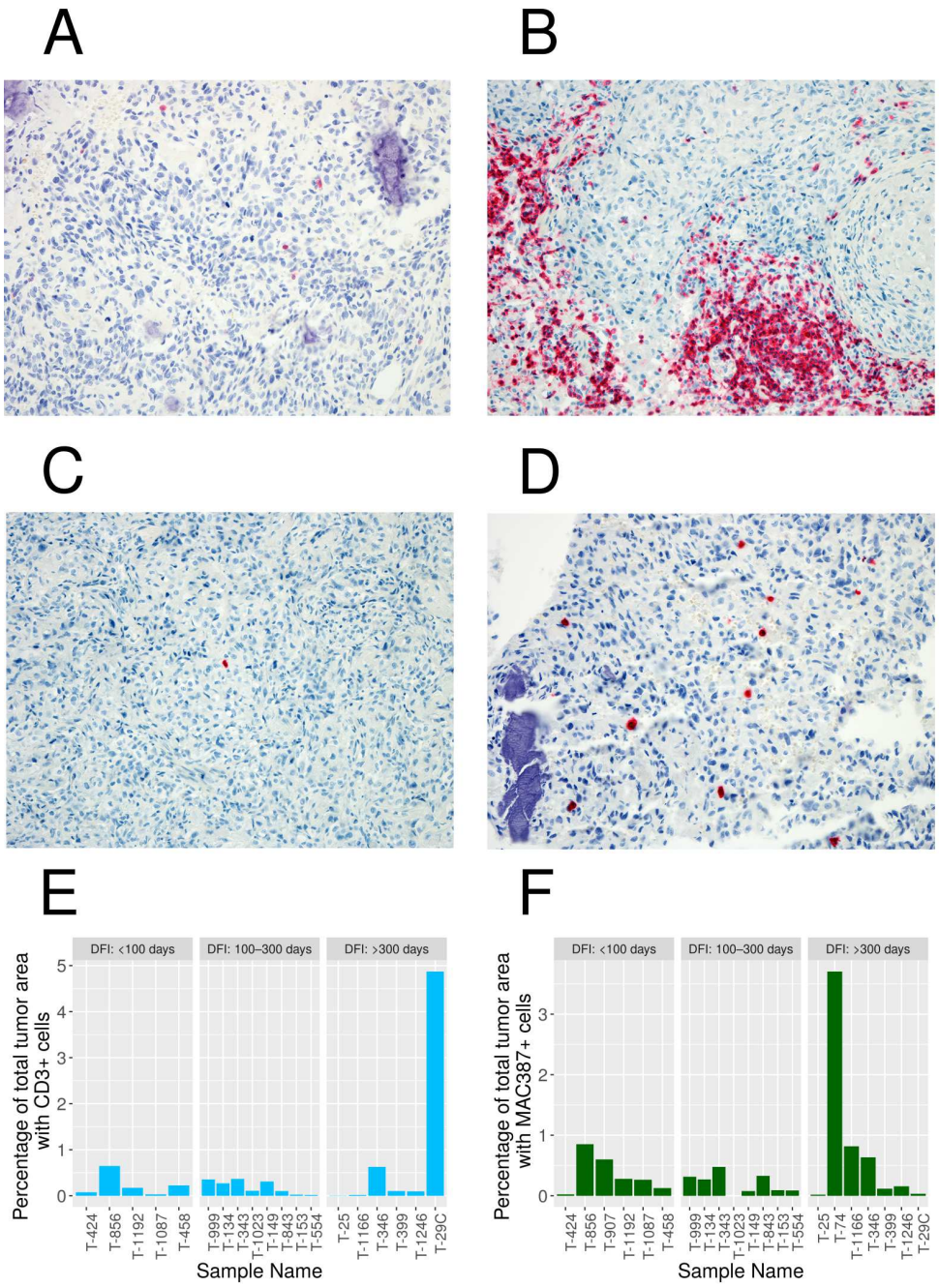


1154

1155 Fig.3. Graphical representation of differentially expressed genes (DEG). A. Heatmap of DEGs
1156 between normal bone (N-) and canine osteosarcoma samples (T-). B-D. The heatmap of core-
1157 enriched genes associated with three significantly enriched Hallmark gene sets in tumor
1158 samples in comparison to normal bone, as identified by GSEA analysis.

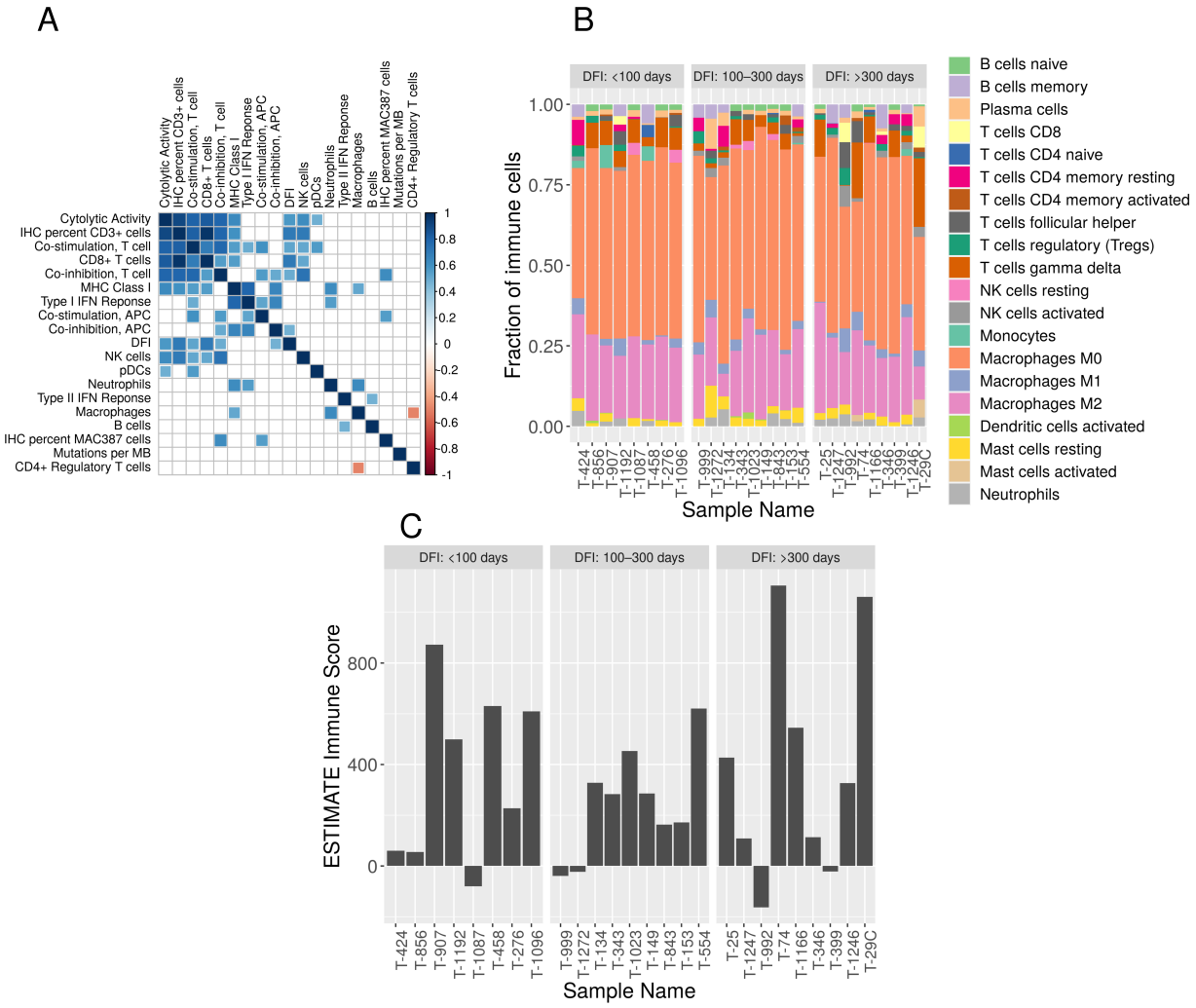
1159

Fig. 4



1179 Fig.4. Association of TP53 mutation status and clinical parameters with outcome. A. Kaplan
1180 Meier plots for disease free interval (DFI) in dogs with osteosarcoma stratified by TP53
1181 mutation status. The dogs with wild type/null TP53 status had a shorter DFI compared to dogs
1182 with TP53 missense mutations. B. Kaplan Meier plots for dogs with osteosarcoma stratified by
1183 pre-treatment monocyte count (> or < $0.04*10^3$ cells/µl). Higher monocyte counts were
1184 associated with short DFI patients.

Fig. 5



1185

1186 Fig 5. Representative figures of immunohistochemical staining of canine osteosarcoma samples.
1187 A-B. These samples were stained for CD3+ T-cell infiltration in A (low levels – T-1023) and B
1188 (high levels – T-1162). C-D. These samples were stained with the MAC387 antibody against
1189 S100A9 to distinguish macrophage infiltration. C represents low levels (T-1162) and D
1190 represents high levels (T-856). E-F. Quantification of immunohistochemical labelling of T-cell
1191 (CD3+ cells) and Macrophages (MAC387 cells) across 19 and 22 samples, respectively.

1192

Fig. 6

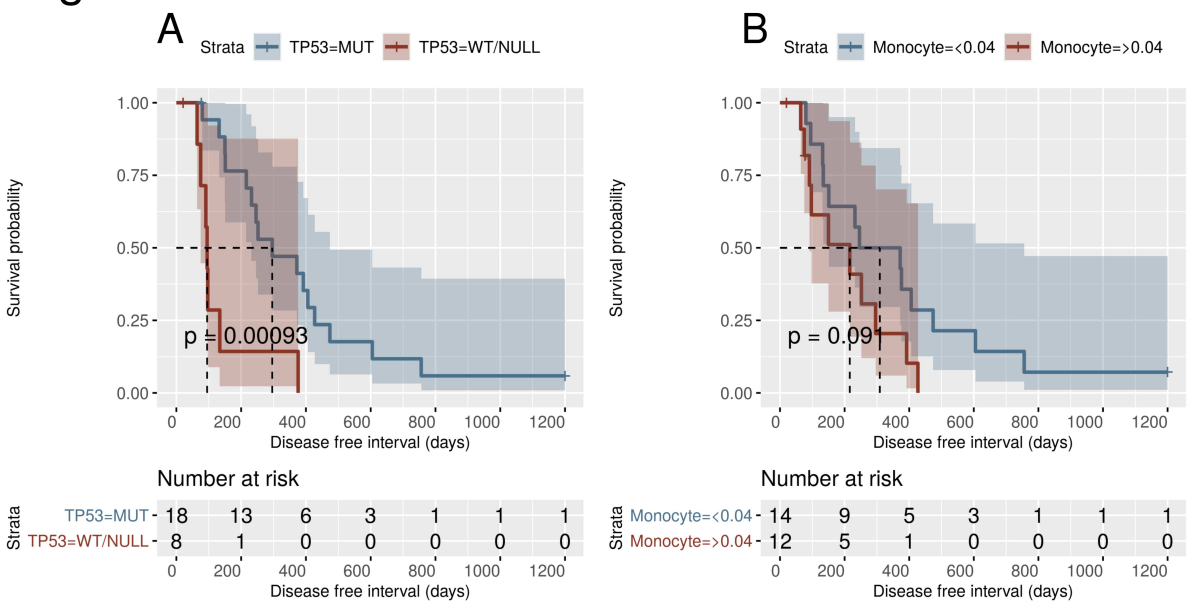


Fig.6. Profiling of immune cell infiltration using gene expression data. A. Pearson correlation matrix of immune cell expression score and immune staining levels of T-cell (IHC percent CD3+ cells) and macrophages (IHC percent MAC387 cells). We also correlated mutations per megabases (MB) and disease-free interval (DFI) with other variables. Only the significant (FDR <0.05) correlation values are plotted. B. Quantification of 22 immune cell types using CIBERSORT. The samples are sorted by DFI. The most prevalent cell type in the tumor microenvironment were M0 and M2 macrophages. C. Immune score quantification using ESTIMATE R package.

DECLARATIONS

Ethics approval and consent to participate

The animal study was reviewed and approved by Colorado State University Animal Care and Use Committee. Written informed consent was obtained from the owners for the participation of their animals in this study.

Funding

Funding was provided by Anschutz Foundation (DLD, DLG and DHT), P30 CA046934 (University of Colorado Cancer Center Support Grant, Genomics and Microarray Shared Resource and DL Gustafson). Funding was also provided by COXEN Retrospective Grant: Morris Animal Foundation D13CA-044 (PI: DL Gustafson) and COXEN Prospective Grant: Morris Animal Foundation D16CA-003 (PI: DL Gustafson). Additional funding includes The National Institutes of Health, Office of the Director, award number K01ODO22982 (PI: DP Regan) and National Center for Advancing Translational Sciences, award number L30 TR002126 (PI: DP Regan).

Acknowledgements

We thank Tyler Eike for IT support and software installation. We would also like to thank Keegan Collins for prepping tumor samples, Liza Pfaff for prepping normal bone samples, and Irene Mok for her help with sample curation. The authors gratefully acknowledge the efforts of Mary Lafferty for collection of clinical outcome data.

Competing interests

There are no competing interests.

Availability of data and materials

The Illumina raw fastq files were submitted to NCBI Short Read Archive (SRA) database. The canine osteosarcoma and normal samples have been submitted to Bioproject PRJNA613479 and PRJNA503860, respectively. The microarray data can be downloaded from Gene Expression Omnibus (GEO) database using accession number GSE76127. The tools, databases, and their versions, in addition to the bioinformatic pipeline used in this study can be accessed here: https://github.com/sdas2019/Canine-Osteosarcoma-Whole-Exome-Sequencing-Pipeline.

Author contribution

Sunetra Das: Contributed towards design of the work, analysis and interpretation of data, drafting of the manuscript.

Rupa Idate: Contributed towards sample preparation used for whole exome sequencing.

Daniel P. Regan: Contributed towards immunohistochemical sample processing and drafting the manuscript.

Jared S. Fowles: Contributed towards microarray sample processing.

Susan E. Lana: Contributed towards curations of tumor samples and is the Director of FACC tissue archive.

Douglas H. Thamm: Contributed towards analysis and interpretation of survival analysis and revision of the draft.

Daniel L. Gustafson: Contributed towards providing resources for gene expression analysis and interpretation.

Dawn L. Duval: Contributed towards conception and design of the work, interpretation of data, and drafting and revisions of the manuscript.

Figures

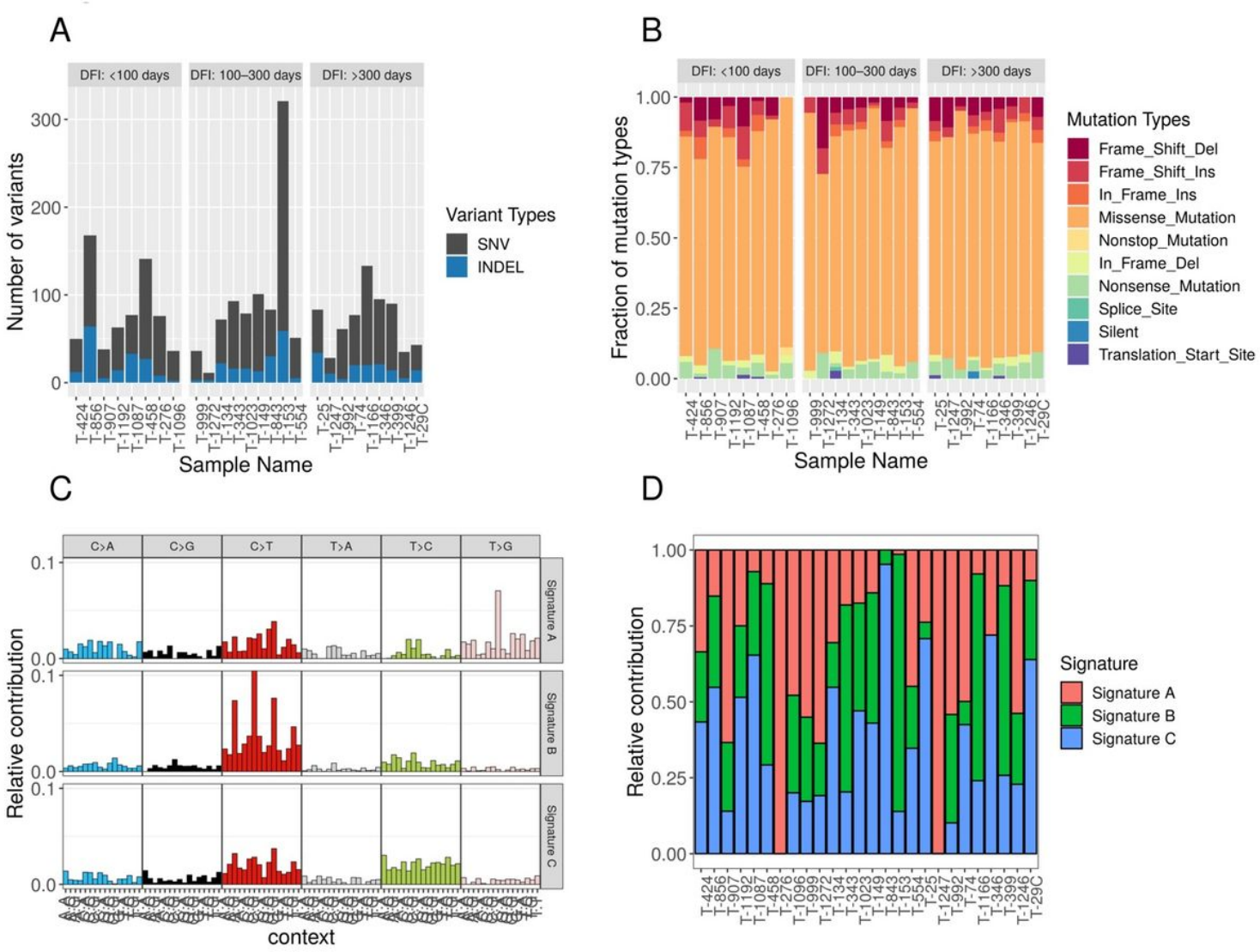


Figure 1

Mutational signature profile in canine osteosarcoma. A. Distribution of total short variants (SNVs and INDELs) across 26 samples. The samples were sorted in ascending order of disease-free interval (DFI) and binned in three discrete groups. B. Distribution of mutation types as identified by VEP across 26 samples. Missense mutations were the most common type of short variant in this dataset. C. Mutation signature analyses showing the relative distribution of six single nucleotide changes in 96 different trinucleotide contexts across three de novo signatures. D. Relative contribution of three predominant signatures across 26 samples.

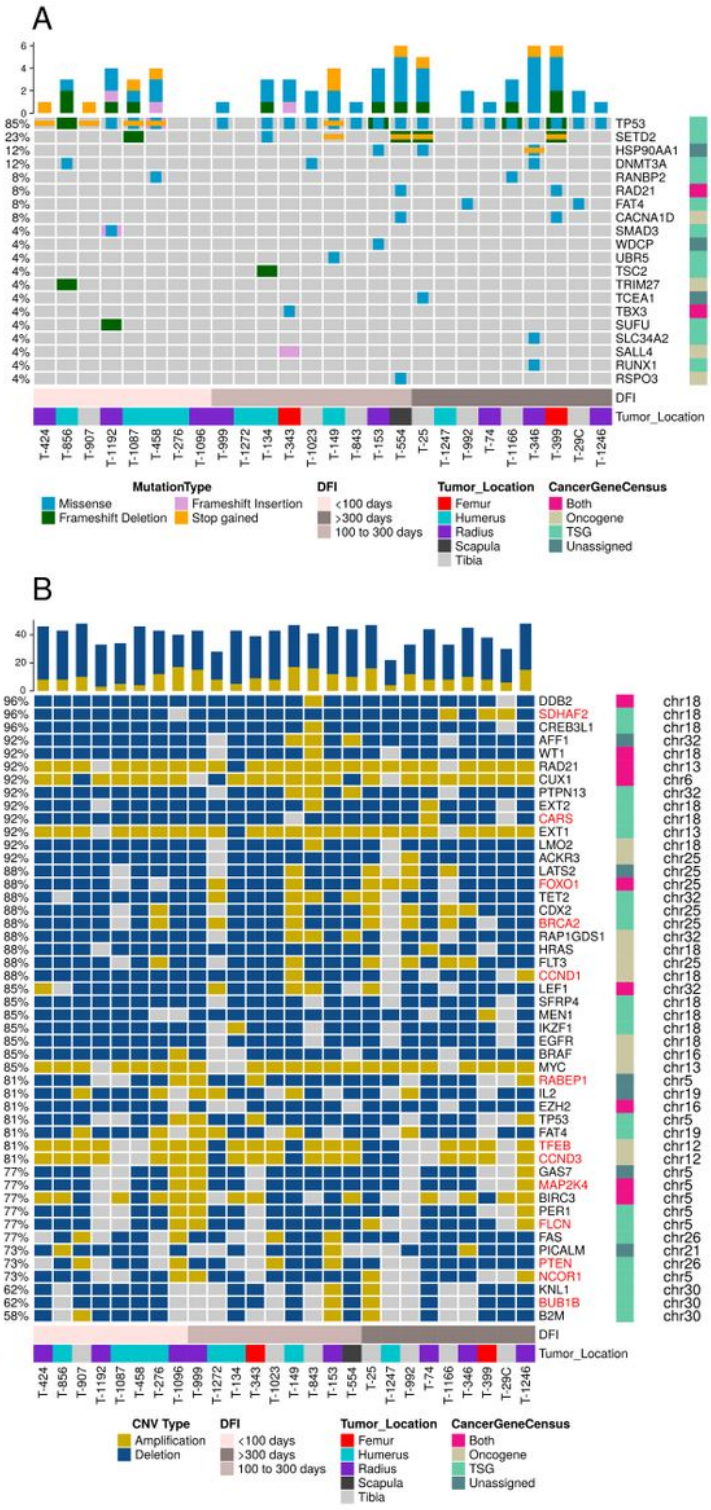


Figure 2

Oncoplot of cancer census genes with variants in canine osteosarcoma samples. A. The genes with short variants (SNVs and INDELs) are plotted along with sample characteristics like disease free interval (DFI) and tumor location. The cancer genes were categorized as oncogenes and tumor suppressor genes in accordance to the Cancer Gene Census. B. The genes with recurrent copy number variations (amplifications and deletions) were plotted. The samples were sorted in ascending order of DFI in both

plots. Gene names colored red indicate significant correlation of copy number amplitude and transcript expression.

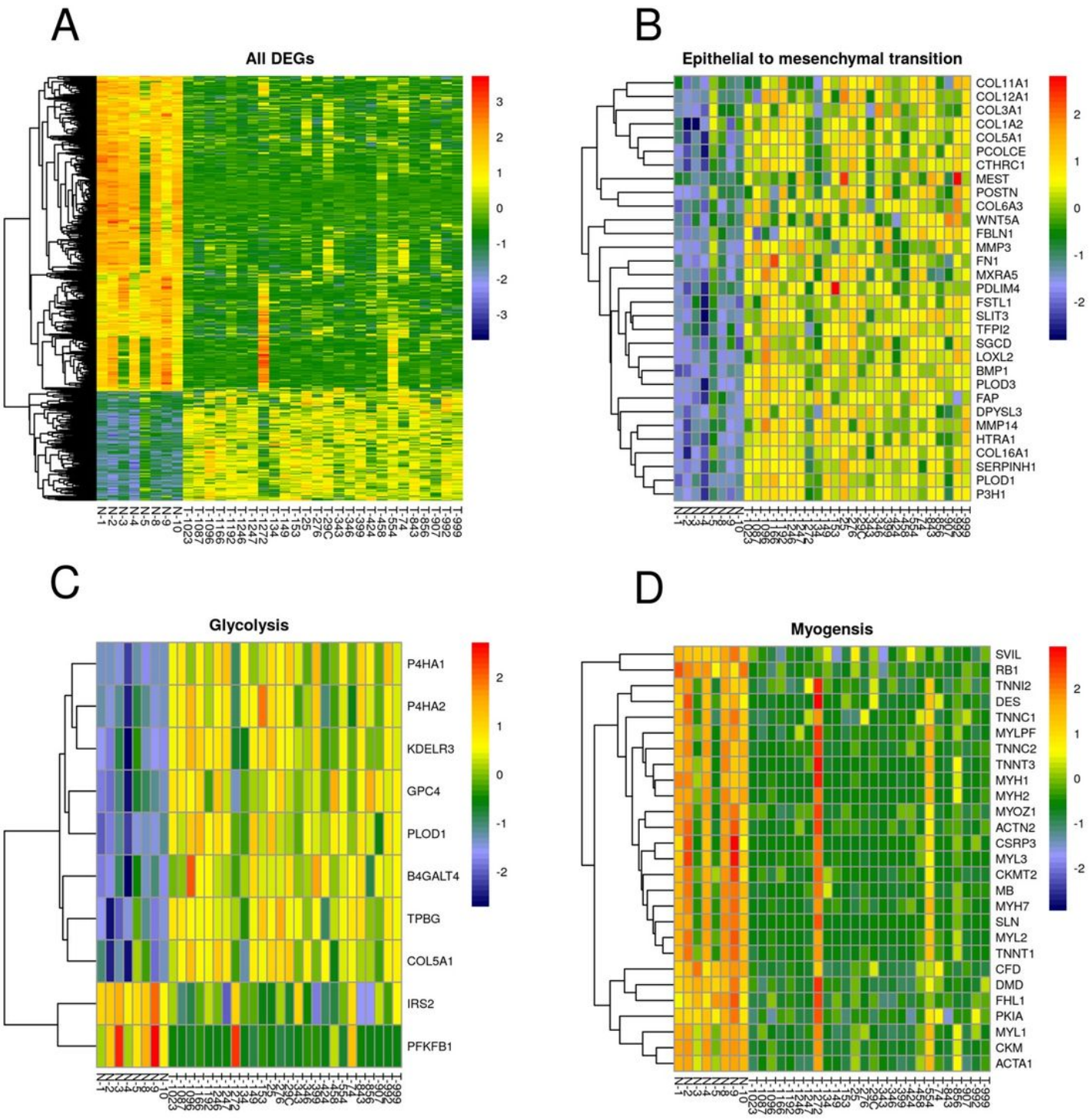


Figure 3

Graphical representation of differentially expressed genes (DEG). A. Heatmap of DEGs between normal bone (N-) and canine osteosarcoma samples (T-). B-D. The heatmap of core enriched genes associated

with three significantly enriched Hallmark gene sets in tumor samples in comparison to normal bone, as identified by GSEA analysis.

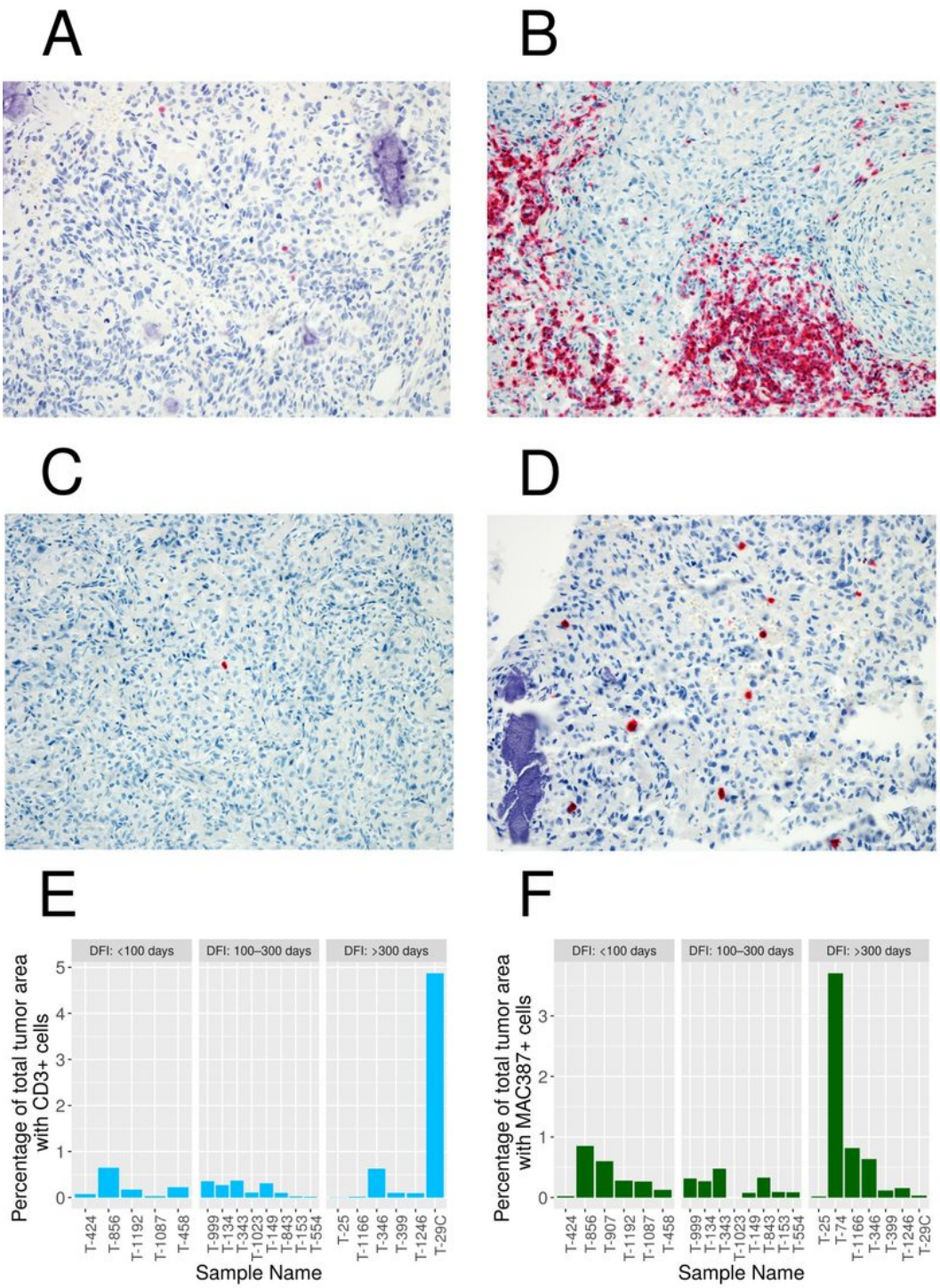


Figure 4

Association of TP53 mutation status and clinical parameters with outcome. A. Kaplan Meier plots for disease free interval (DFI) in dogs with osteosarcoma stratified by TP53 mutation status. The dogs with wild type/null TP53 status had a shorter DFI compared to dogs with TP53 missense mutations. B. Kaplan

Meier plots for dogs with osteosarcoma stratified by pre-treatment monocyte count (> or < 0.04*103 cells/µl). Higher monocyte counts were associated with short DFI patients.

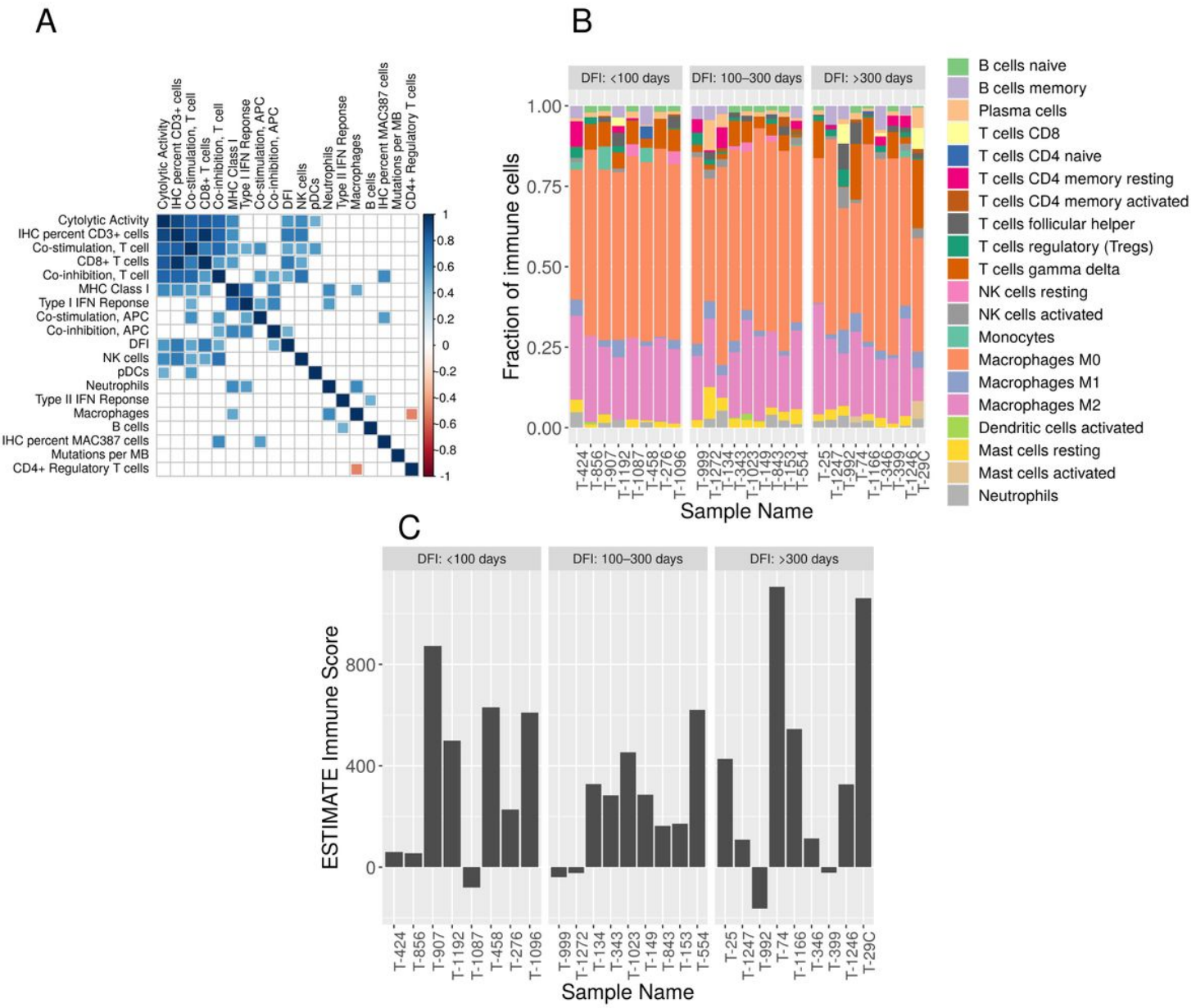


Figure 5

Representative figures of immunohistochemical staining of canine osteosarcoma samples. A-B. These samples were stained for CD3+ T-cell infiltration in A (low levels – T-1023) and B (high levels – T-1162). C-D. These samples were stained with the MAC387 antibody against S100A9 to distinguish macrophage infiltration. C represents low levels (T-1162) and D represents high levels (T-856). E-F. Quantification of immunohistochemical labelling of T-cell (CD3+ cells) and Macrophages (MAC387 cells) across 19 and 22 samples, respectively.

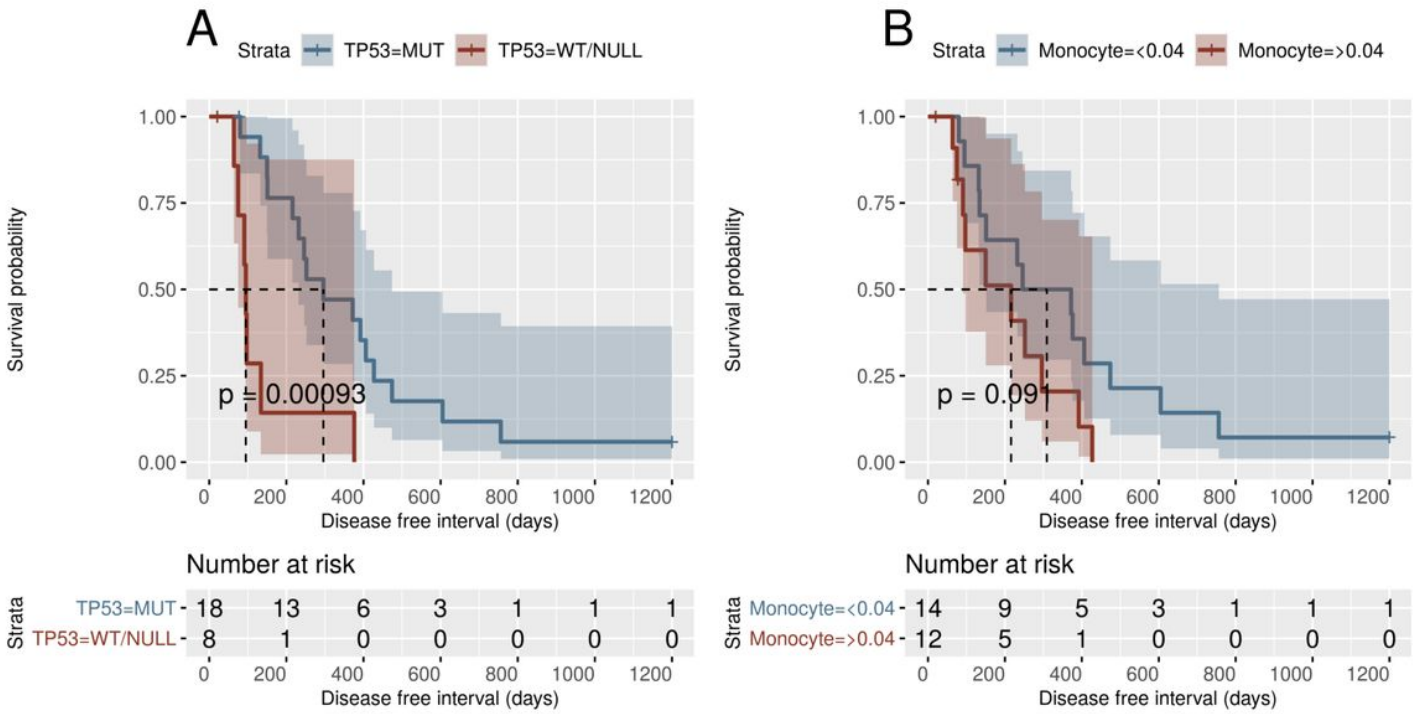


Figure 6

Profiling of immune cell infiltration using gene expression data. A. Pearson correlation matrix of immune cell expression score and immune staining levels of T-cell (IHC percent CD3+ cells) and macrophages (IHC percent MAC387 cells). We also correlated mutations per megabases (MB) and disease-free interval (DFI) with other variables. Only the significant (FDR <0.05) correlation values are plotted. B. Quantification of 22 immune cell types using CIBERSORT. The samples are sorted by DFI. The most prevalent cell type in the tumor microenvironment were M0 and M2 macrophages. C. Immune score quantification using ESTIMATE R package.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.xlsx
- AdditionalFile1.pdf
- AdditionalFile2.pdf
- nrreportingsummary.pdf