Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Cancer Res. Author manuscript; available in PMC 2018 July 15.

Published in final edited form as:

Cancer Res. 2018 January 15; 78(2): 326–337. doi:10.1158/0008-5472.CAN-17-0576.

Comparative transcriptome analysis quantifies immune cell transcript levels, metastatic progression and survival in osteosarcoma

Milcah C. Scott[1,2,3], **Nuri A. Temiz**[1,14], **Anne E. Sarver**[4,8], **Rebecca S. LaRue**[1,5], **Susan K. Rathe**[1], **Jyotika Varshney**[1], **Natalie K. Wolf**[1,6], **Branden S. Moriarity**[1,7,8,9], **Timothy D. O'Brien**[1,2,10,11], **Logan G. Spector**[1,8], **David A. Largaespada**[1,6,7,8,9,†], **Jaime F. Modiano**[1,2,3,11,12,13,†], **Subbaya Subramanian**[1,4,†], and **Aaron L. Sarver**[1,14,*]

[1]Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA

[2]Animal Cancer Care and Research Program, University of Minnesota, St. Paul, Minnesota 55108, USA

[3]Department of Veterinary Clinical Sciences, University of Minnesota College of Veterinary Medicine, St. Paul, Minnesota 55108, USA

[4]Department of Surgery, University of Minnesota School of Medicine, Minneapolis, MN 55455, USA

[5]Department of Medicine, University of Minnesota School of Medicine, Minneapolis, Minnesota 55455, USA

[6]Department of Genetics, Cell Biology and Development, University of Minnesota School of Medicine, Minneapolis, Minnesota 55455, USA

[7]Brain Tumor Program, University of Minnesota, Minneapolis, Minnesota 55455, USA

[8]Department of Pediatrics, University of Minnesota School of Medicine, Minneapolis, Minnesota 55455, USA

[9]Center for Genome Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA

[10]Department of Veterinary Population Medicine, University of Minnesota College of Veterinary Medicine, St. Paul, Minnesota 55108, USA

[11]Stem Cell Institute University of Minnesota, Minneapolis, Minnesota 55455, USA

[12]Department of Laboratory Medicine and Pathology, University of Minnesota School of Medicine, Minneapolis, Minnesota 55455, USA

[13]Center for Immunology, University of Minnesota, Minneapolis, Minnesota 55455, USA

[14]Institute for Health Informatics, University of Minnesota, Minneapolis, Minnesota 55455, USA

Abstract

[*]Corresponding author: sarver@umn.edu.
[†]equal contribution

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Overall survival of patients with osteosarcoma (OS) has improved little in the past three decades and better models for study are needed. OS is common in large dog breeds and is genetically inducible in mice, making the disease ideal for comparative genomic analyses across species. Understanding the level of conservation of inter-tumor transcriptional variation across species and how it is associated with progression to metastasis will enable us to more efficiently develop effective strategies to manage OS and improve therapy. In this study, transcriptional profiles of OS tumors and cell lines derived from humans (n=49), mice (n=103) and dogs (n=34) were generated using RNA-sequencing. Conserved inter- tumor transcriptional variation was present in tumor sets from all three species and comprised gene clusters associated with cell cycle and mitosis and with the presence or absence of immune cells. Further, we developed a novel Gene Cluster Expression Summary Score (GCESS) to quantify inter-tumor transcriptional variation and demonstrated that these GCESS values associated with patient outcome. Human OS tumors with GCESS values suggesting decreased immune cell presence were associated with metastasis and poor survival. We validated these results in an independent human OS tumor cohort and in 15 different tumor data sets obtained from The Cancer Genome Atlas (TCGA). Our results suggest that quantification of immune cell absence and tumor cell proliferation may better inform therapeutic decisions and improve overall survival for OS patients.

Introduction

Osteosarcoma (OS) is the most common primary bone malignancy, accounting for ~2% of childhood cancers. The total number of new human cases diagnosed each year in the United States is low, ~400-600 annually, which presents a significant challenge to studying this rare but deadly cancer (1). The relative 5-year survival rates for all OS (~60%) and for metastatic OS (~20%) have not significantly changed since the mid-1980s (1–7). OS in dogs is much more common, with an overall incidence rate 30-50 times higher than humans, and a lifetime risk up to 10% in larger, pre-disposed breeds(8). OS tumors can also be generated in wildtype and predisposed mice via tissue specific mobilization of the T2/ONC transposon by the Sleeping Beauty Transposase (9, 10).

While in humans OS is primarily a pediatric disease, in dogs the majority of cases occur in older dogs with an average onset at 7.5 years (8). Unfortunately ~80% of dogs die as a result of metastasis within one year of diagnosis (11). Human dog and mouse OS share many clinical and molecular features and insight gained from one species may be translatable to the others (9–12).

Identifying and understanding molecular mechanisms of OS have lagged behind other cancer types, partly due to the small number of human tumors available for study (12, 13). Additionally, OS is characterized by complex karyotypes with highly variable structural and numerical chromosomal aberrations in humans, dogs and mice (9, 14–16). Sequencing studies have identified recurring genetic alterations in OS, some of which are common to both human and dog OS (12, 14, 17–19). However, understanding of these genetic alterations has not led to improved outcomes for OS patients. There remains a critical need to develop diagnostic tests that can identify risk of OS progression at an early, pre-metastatic stage.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The goal of this study was to assess transcriptional variation in OS tumor samples, identify common patterns of expression across different species, and evaluate their association with metastases and clinical outcomes. To accomplish these goals, we developed the Gene Cluster Expression Summary Score (GCESS) technique to identify and quantify transcriptional patterns present across datasets and across species. We then use the GCESS method to derive associations between metastatic progression, outcome and transcriptional patterns in OS and extend these results to a wide range of human tumors. Based solely on our genome wide method, we show that the loss of immune cell transcripts as well as increases in cell cycle transcripts are associated with poor outcomes in OS and extend these results to a wide range of human tumor types.

Methods

Biospecimen collection and processing

Human and dog biospecimens were collected from newly diagnosed OS patients prior to treatment with cytotoxic chemotherapy drugs. Specimens were obtained under protocols approved by either the University of Minnesota's Institutional Review Board or Institutional Animal Care and Use Committee (protocol numbers 0802A27363, 1101A94713, 1312-31131A) or the University of Colorado Institutional Review Board or Institutional Animal Care and Use Committee (AMC 635040202, AMC 200201jm, AMC 2002141jm, 02905603(01)1F, COMIRB 06-1008).

Human samples

Human patient OS samples (n=44) and normal bone samples (n=3) were obtained from the University of Minnesota Biological Materials Procurement Network (UMN BioNet) or the Cooperative Human Tissue Network (CHTN), both of which follow standardized patient consent protocols. Samples had been de-identified and only a limited amount of patient information was provided. Saos-2, U-2 OS, MG- 63, and 143B human OS cells line purchased from American Type Culture Collection (Manassas, VA, USA) and authenticated by the University of Arizona Genetics Core using short tandem repeat profiling, as well as an osteoblast cell line, which was a gift from Dr. Richard G. Gorlick (Albert Einstein College of Medicine, NY, USA), were also sequenced. Supplemental Table 1 summarizes the available metadata characteristics for the samples used in this study.

Dog samples

Dog OS samples (total n=31) were obtained from dogs with naturally occurring primary appendicular tumors, recruited between 1999 and 2012. The majority of the samples were from Rottweilers and Golden Retrievers. Specimens were obtained with owner consent under approved protocols as previously described (13). Also included in this study were two cell lines (OSCA-8 and OSCA-78) derived from primary OS tumor as previously described (20), and one dog osteoblast sample, CnOB, purchased from Cell Applications (San Diego, CA, USA). The OSCA cell lines are available for distribution through Kerafast, Inc. (Boston, MA).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Mouse samples

Mouse OS samples (n=92 tumor, n=11 cell line) were derived from previously established experimental models. Included in this study were 25 OS samples from mice with somatic induction of $Trp53^{R270H}$ expression in Osx1-expressing cells (9), 67 samples from Sleeping Beauty transposon accelerated OS in wildtype and $Trp53^{R270H}$ mice (10), and 11 cell lines established from mouse OS tumors(9).

RNA extraction from frozen tumor tissue and cell lines

Isolation of total RNA from tissues, avoiding areas of necrosis, and from cell lines was performed according to the recommended protocol for Ambion's TotalRNA kit from Thermo Fisher Scientific (Denver, CO, USA). Samples were quantified using fluorescence by RiboGreen dye (Thermo Fisher Scientific). RNA integrity was assessed using capillary electrophoresis (RIN > 6.5) with the Agilent 2100 BioAnalyzer system (Agilent Technologies, Santa Clara, CA, USA).

RNA-Sequencing (RNA-Seq)

Sequencing libraries of each sample were prepared using the TruSeq Library Preparation Kit (Illumina, San Diego, CA). Paired end sequencing (30-40 million reads per sample) was done at the University of Minnesota Genomics Center (UMGC) on a High Seq 2000 (Illumina, San Diego, CA, USA). The raw FASTQ files are available at the NCBI Sequence Read Archive and linked to from Gene Expression Omnibus SuperSeries GSE87686.

Publically available human data

RNA-Seq FASTQ files and outcome related metadata for 35 additional human OS samples (HOS2) were obtained from dbGap:phs000699.v1.p1 (http://www.ncbi.nlm.nih.gov/gap) (21). RNA-Seq files were also obtained from 25 additional human OS samples (HOS3) available from previously published studies (9, 15, 17).

FASTQ files from the HOS2 cohort were mapped to the reference genome and FPKM values were calculated using the same protocols as the samples in this study, while FASTQ files from the HOS3 cohort were mapped to the reference genome and RSEM expression values were calculated using the TCGA RNA-Sequencing protocol (22, 23). In this study, the samples from both of these cohorts were used for calculating pairwise Pearson correlation coefficients and were included in the transcriptome analyses described below. The HOS2 cohort included survival data and information on presence of metastasis at diagnosis so this cohort was used for survival analysis.

GEO dataset series GSE21257, which consisted of genome-wide expression data of 53 human OS tumors produced from the Illumina human-6 v2.0 expression array was downloaded for this study. Patient outcome data, including survival and metastasis, was included in this study.

The Cancer Genome Atlas (TCGA) data portal, (http://tcga-data.nci.nih.gov/tcga) was used to download survival times, death events and RNA- Seq data for 5582 tumors as described in Supplemental Table 2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RNA-Seq workflow

Briefly, mapping was carried out using Tophat (24), Samtools (25) and Cufflinks (26) to generate FPKM values using reference genomes as described in supplementary methods. To minimize the effects of dividing FPKM values by numbers close to 0 and stochastic noise, 0.1 was added to each FPKM value (27, 28). The FPKM files are available at GEO GSE87686. Mapping statistics are summarized in Supplemental Table 3.

Transcriptome analyses

Due to the high correlations between human OS datasets from multiple sources, the human datasets were combined for further analyses (9, 15, 17, 21). The ~8,000 most variable genes across the full same-species datasets were identified for clustering by species (standard deviation cutoffs: human > 0.93, mouse > 0.72, and dog > 0.79). Cluster 3.0 (C Clustering Library 1.52) was used to log2 transform and gene-mean-center the data and then perform hierarchical average linkage clustering using the Pearson similarity metric. Clustering data were visualized in Java TreeView (version 1.1.6r4). OptiType, a precision HLA typing tool (29) was used to identify human OS samples derived from the same patient.

Systematic identification of gene clusters

Gene clusters with a dendrogram node correlation > 0.60 and at least 60 individual genes were identified in each of the species datasets. The 0.6 Pearson correlation cutoff value was chosen, as it is a widely accepted conservative confidence threshold. The minimum cluster size of 60 was chosen to ensure that only larger transcriptional patterns were identified. Permuted and random datasets were used to show that these thresholds would not identify clusters in artificial datasets that do not contain meaningful transcription patterns. Gene clusters representing batch effects from the combination of different human samples were removed from further analyses.

Generation of control datasets

For each species, a random dataset and a permuted dataset were generated as controls. Random datasets were generated by randomly selecting values between −2 and 2 to replace the actual (mean-centered) values. Permuted datasets were generated by randomly reordering the values for each gene.

Gene Cluster Expression Summary Score (GCESS) calculation

The GCESS is defined as the sum of expression values (log_2-transformed and mean centered) of all genes in a particular defined cluster for a single sample. The GCESS quantifies transcriptional variation between tumors. It takes many correlated individual transcript data points and condenses them into a single value.

Pathway analysis

The Ingenuity Pathway Analysis (IPA) suite (Qiagen, Redwood City, CA, USA) was used to identify pathways associated with gene clusters.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Statistics

Statistical significance was calculated using the log-rank test or by Fisher's exact test depending on analysis and a $p < 0.05$ was considered significant. Kaplan-Meier (KM) survival plots were generated using the 'survival' package in R (Version 0.98.1103)(30, 31). The GCESS values were used to rank the tumors into quartile groups and the quartile groups were systematically tested for association with outcome.

Histopathology, Immunohistochemistry and additional statistical information provided in Supplementary Methods.

Results

OS tumors show a common transcriptional profile across species that is distinct from other types of tumors

To better understand the OS transcriptome, we performed RNA-Seq analysis on mRNA libraries generated from human, dog, and mouse tumors and cell lines (Table 1). We hypothesized that despite the highly complex karyotypes associated with OS, tumors would maintain common transcriptional characteristics between species that were distinct from other types of tumors. To test this hypothesis, we calculated pairwise Pearson correlation coefficients within and between our OS cohort (HOS1) and two independent human OS tumor cohorts (HOS2, HOS3) recently published by other groups (9, 15, 17, 21) (Fig 1). Strong correlations were observed within each human cohort (average intra-cohort correlations: HOS1 0.62, HOS2 0.67, HOS3 0.66). Strong correlations were also observed between our cohort and the previously published cohorts (average inter-cohort correlations: HOS1-HOS2 0.61, HOS1-HOS3 0.53). In contrast, strong correlations were not observed between HOS1 and other human tumors obtained from TCGA (cervical 0.26, colorectal 0.41, glioblastoma 0.30, leukemia 0.24, prostate adenocarcinoma 0.34, and thyroid cancer 0.17), indicating that there are common transcriptional events that define OS pathology (Fig 1A).

To establish the cross-species relevance of the dog (DOS) (Fig 1B) and mouse (MOS) (Fig 1C) OS tumors we calculated both inter- and intra-species correlations. Similar to humans, average intra-species correlations were high in both dog (0.76) and mouse (0.68) samples. Dog and mouse OS samples were both correlated with the human OS tumors (HOS1-DOS 0.48, HOS1-MOS 0.52), but not with other human cancers (DOS-cervical 0.18, MOS-cervical 0.14, DOS-colorectal 0.31, MOS-colorectal 0.28, DOS-glioblastoma 0.21, MOS-glioblastoma 0.25, DOS-leukemia 0.14, MOS- leukemia 0.14, DOS-prostate adenocarcinoma 0.24, MOS-prostate adenocarcinoma 0.21, DOS-thyroid cancer 0.13, MOS-thyroid cancer 0.12), validating a cross-species analysis strategy to identify common transcriptional components in the development and progression of OS.

OS tumors show common transcriptional variation patterns across three species

We hypothesized that patterns of transcriptional variation would be conserved in OS tumors across species. To systematically assess transcriptional variations across human, dog, and mouse OS tumors each dataset was first analyzed independently using average linkage

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

clustering. Tightly correlated and large gene clusters representing patterns of inter-tumor transcriptional variation were defined as having both a dendrogram node correlation greater than 0.6 and a minimum of 60 genes. A total of nine human, five dog, and 11 mouse gene clusters were identified (Fig 2A, Supplemental Table 4). To ensure that these clusters were not artifacts, similarly sized random datasets and shuffled permutations of the real data were generated and clustered. No clusters passed the threshold criteria in either the random or permuted datasets, validating that the clusters observed in the real RNA- Seq data represent true transcriptional variation (Supplemental Figs 1-9).

To determine whether the identified gene clusters represented sets of genes common across the 3 species, pairwise percent overlaps were calculated between each gene cluster from a single species and all clusters in the other 2 species. The resulting percentage overlap values were clustered for each species pair. Several clusters of genes showed clear overlap across species (Fig 2B-D, Supplemental Table 4), indicating conserved patterns of transcriptional variation in OS tumor samples.

To describe the potential biological significance of these common gene clusters, enriched functional annotations were identified using IPA software (Qiagen). The most conserved cluster across species was composed of transcripts that were independently highly enriched in genes associated with cell cycle and mitotic functions. The next two highly conserved clusters across species were each independently enriched in transcripts relating to immune cell functions. (Fig 2D, Supplemental Table 5). A gene cluster enriched for muscle differentiation genes was observed in the human and mouse tumors but it was not present in the dog tumors.

We next asked if the genes in the two immune cell gene clusters represented a broad spectrum of immune cells or were enriched for particular leukocytes. We compared these immune cell gene clusters to a gene signature matrix (LM22), which consists of 547 genes and was created to identify genes unique to 11 different subtype populations of immune cells (32). Overall, our human immune cell annotated gene clusters were highly enriched in the LM22 gene transcripts (Supplemental Table 4). Using the LM22 annotations to determine the types of active leukocytes present (as represented by their unique transcripts), we found genes representing monocytes to be most enriched in human cluster-1 and genes representing T cells to be most enriched in human gene cluster-8 (Supplemental Table 4). Results from assessing enrichment of the LM22 genes in the dog and mouse immune cell gene clusters also supported these results (Supplemental Table 4).

The Gene Cluster Expression Summary Score (GCESS) technique quantifies tumor transcriptional variation

The GCESS was developed to reduce the high dimensionality of expression profiles generating a normalized and easily comparable value per sample for use in further association analyses. The GCESS is defined as the sum of expression values (log2-transformed and mean centered) of all genes in a particular gene cluster for a single sample. A negative GCESS indicates relative under-expression of the group of genes in that sample compared to all of the samples in the analysis set, a positive GCESS indicates overexpression, and a GCESS close to 0 (zero) indicates mean expression. The GCESS

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

summarizes the relative transcript levels of many correlated genes into a single value. This value is calculated for each cluster in each tumor. This allows for tumors to be rank ordered by summary score, which is based on the observed transcriptional data. Multiple summary scores were generated for each tumor sample, allowing for the independent comparison of the impact of each identified gene cluster, thereby achieving an unbiased dimensional reduction.

To better characterize the conserved OS tumor transcriptional variation, GCESS values were calculated for the identified for all identified gene clusters in each sample (Fig 3A and Supplemental Table 6). In addition to human, dog, and mouse OS tumor samples, we also analyzed tumor-derived cell lines, normal human bone cells, and dog osteoblasts. The distribution of GCESS values (Fig 3B-D) shows distinct patterns for cells/cell lines, normal bone, and tumor samples.

Tumor-derived cell lines had higher cell cycle GCESS values compared to tumor tissues, while normal human bone had cell cycle GCESS values corresponding to the lower range of GCESS values obtained from tumor samples (Fig 3B). Dog osteoblasts had an extremely low cell cycle GCESS. These results are consistent with the upregulation of cell cycle genes in a subset of highly proliferating OS tumors (13).

Tumor-derived cell lines had lower immune GCESS values compared to the majority of tumor tissues in all three species datasets. The immune cell GCESS of the normal human bone sample was within the middle of the range seen in human tumor samples. Dog osteoblasts had GCESS values corresponding to the tumor-derived cell lines (Fig 3C-D). These results are consistent with the absence of immune cells in cell lines and the variable presence of immune cells in tumor-derived tissues from all 3 species.

To validate the variable presence of immune cells indicated by GCESS immune cluster scores, we evaluated 10 FFPE sections derived from canine OS tumors, which were also sequenced, for the presence of T cells and macrophages within tumor stroma by immunohistochemistry. (Supplemental Figure 10A-D and Supplemental Table 7). Immunohistochemistry staining supported the transcriptome derived GCESS data for both MAC387 and CD3 staining. Stromal MAC387 was not observed in the 3 tumors with the lowest GCESS scores for the immune cell cluster-1 and occasional MAC387 positive cells were observed for the 3 tumors with the highest immune cluster-1 GCESS scores. One of the middle range tumors showed the presence of MAC387 positive staining cells within the stroma while 3 others did not. For CD3, only the sample with a positive GCESS immune cluster-2 score showed T-cells present within the tumor stroma. Of note, the next four tumors sorted by immune cluster-2 GCESS score all showed MAC387 positive cells within the tumor stroma.

Minimization of multiple testing errors

We hypothesized that patterns of transcriptional variation present in tumors should indicate associations with patient outcome and that methodological data analyses improvements would be necessary to reveal these associations. Associations between OS patient outcomes and individual gene transcript levels are commonly calculated using the Kaplan-Meier (KM)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

estimator, a non-parametric statistic that estimates the survival function based on censored lifetime data. KM survival analyses were calculated using groups defined by individual transcript levels for our dog and human HOS2 datasets. Potentially significant events were present following comparison of all transcripts ($p < 0.00001$) (Supplemental Fig 11) used in clustering. To assess whether these associations were false positives or not, similarly sized random datasets and shuffled permutations of the real data were also analyzed. Random and shuffled permutations led to similarly "significant events" as were observed in the real data due to multiple testing of ~8-9000 individual tests, indicating that methodological improvements were necessary (Supplemental Fig 11).

Two routes are typically used to minimize the effects of multiple testing in statistical analyses. The first entails increasing sample numbers in order to surpass multiple testing corrections via increased statistical power. This was not feasible in this case. A second approach, which we used here, was to drastically decrease the number of tests performed (dimensionality reduction) by testing the gene cluster GCESS values rather than all individual transcript values. The GCESS was used to rank the tumors into quartile (Q) groups and systematically compare Q1- (lowest GCESSs) vs-Q234, Q12-vs-Q34, and Q123-vs-Q4. (Fig. 3A) Using this approach, associations between outcome and immune cell GCESS could be systematically examined without prior knowledge of the type of association present. Examination of randomly generated and real, but permuted datasets using the full pipeline did not identify gene clusters, limiting association analyses to gene clusters defined by non-random transcript patterns.

Systematic examination of GCESSs with tumor outcomes identifies poor patient outcome association with high expression of cell cycle cluster genes and low expression of immune cell cluster genes

Applying the GCESS approach to the dog data revealed an association between increased transcript levels of cell cycle genes and decreased time to death. Significantly worse outcomes were associated with the higher GCESS groups in both the Q123-vs-Q4 and Q12-vs-Q34 comparisons. Similarly, the human outcome analysis revealed that the highest cell cycle GCESSs (Q4) also had a strong trend towards decreased patient survival times (Q123-vs-Q4). Analyzing the human Immune-1 GCESSs demonstrated that human patients with the lowest immune cell cluster GCESSs had significantly shorter times to death (Q1-vs-Q234). The Immune-2 GCESSs also showed a trend towards association between low GCESSs and decreased time to death (Fig 3, Supplemental Figs 12-15).

Association between GCESS and patient outcome validated in independent cohort

If the conserved tumor transcriptional variation and association between GCESS and patient outcome revealed by this methodology are generally observable in OS tumors, they should also be observable in older array hybridization based data. To validate this hypothesis, we analyzed GSE21257, an Illumina mRNA array dataset, which contained genome wide expression data for 53 tumors from patients with known outcome data, including survival and metastasis (33). Following a strategy similar to the one used for the RNA-Seq data, 14 highly correlated gene clusters were identified. Gene overlap analyses comparing clusters derived from array and RNA- Seq data sets identified gene clusters from the array which

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

corresponded to the RNA- Seq defined Immune-1, Immune-2, Cell Cycle and muscle transcript clusters (Fig 4A). KM survival analyses utilizing sample GCESSs generated from each of these array gene clusters again showed that patients whose tumors had low immune cell GCESSs were more likely to succumb to OS (Fig 4B-D, Supplemental Figs 16-17).

A KM-analysis examining the time to metastasis was also performed. Low immune cell GCESSs or high cell cycle GCESSs were strongly associated with faster progression to metastasis ($p = 0.0001$ and $p = 0.02$ respectively) (Supplemental Figs 18-20). These results independently validate the initial human data set findings as well as the general utility of the methodology described independent of experimental platform.

Low Immune-2 GCESS associated with metastasis in Human and Mouse samples

Following necropsy, many of the mice had observable metastases. Scores from samples with Metastases had lower Immune-2 scores than samples where metastatic tumors were not observed. In both Human datasets, Immune-2 scores were lower in samples where Metastases were present at diagnosis. This difference became significant when patients with metastasis in less than one year were combined with patients where metastasis was observed at diagnosis. The significance became even stronger when patients which showed metastasis at any point were compared to patients where metastasis was not observed. These findings indicate that the Immune-2 score has prognostic potential for determining the likelihood of a tumor metastasizing in human patients (Fig 4E).

OS-derived cell cycle and immune cell GCESSs correlate with poor clinical outcomes across multiple tumor types

We hypothesized that the patterns of transcriptional variations we found to be associated with outcomes in OS may also be relevant across many different types of tumors. To determine if increased cell cycle transcripts and decreased immune cell transcripts are each also associated with poor clinical outcome measures across different tumor types, GCESSs were calculated for each of the TCGA tumor datasets (using the genes comprising the respective human OS clusters to identify each tumor's cell cycle and immune cluster), which were then subjected to KM-survival analysis. High cell cycle GCESSs were significantly associated with poor survival outcomes in Kidney Renal Clear Cell Carcinoma (KIRC), and clear trends were apparent in Liver Hepatocellular Carcinoma (LIHC), Lung Adenocarcinoma (LUAD), Pancreatic adenocarcinoma (PAAD), Head and Neck Squamous Cell Carcinoma (HNSC), and Cutaneous Melanoma (SKCM) (Table 2). Low immune cell GCESSs from Human OS gene cluster 1 were significantly associated with poor survival outcomes in SKCM, and clear trends were apparent in LUAD, Colon Adenocarcinoma (COAD), and Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (CESC) (Table 2). Low immune cell GCESSs derived from Human OS gene cluster 8 were significantly associated with poor survival outcomes in SKCM and clear trends were visible in, HNSC, LUAD, LIHC, COAD, CESC and Breast Invasive Carcinoma (BRCA). These results indicate that the survival associations between cell cycle and immune transcript expression levels observed in OS are also present in a wide range of tumor types and that the GCESS methodology is capable of observing these associations in datasets where improved sample power exists.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Discussion

Many diverse genetic events, including a catalog of rare events, have been reported to lead to OS formation, progression and metastasis. Our data indicate that loss of immune cell infiltration and increased levels of cell cycle transcripts are two specific transcriptional prognostic biomarkers for metastasis, and overall poor survival of OS patients. These transcriptional signatures also had prognostic utility across many types of human tumors, suggesting they are common transcriptional markers of pathological progression. Further, these two transcriptional patterns appear to be independent (Supplemental Table 6, Supplemental Figure 21).

Tumor transcription can be conceptualized as resulting from loss of control of a series of independent transcriptional modules (gene clusters). The GCESS technique described in this paper generates a single meaningful value for each module in a tumor. The GCESS value can then be used for phenotype association discovery, thereby minimizing the multiple testing risks in datasets underpowered for meaningful genome-wide association analyses. These types of errors are clearly described for SNP association studies but remain prevalent in genome-wide studies utilizing large numbers of parallel analyses. Empirical testing of single gene based strategies to associate tumor transcript level with outcome revealed that for every 20 random transcripts tested, 1 false positive prognostic transcript was likely to be identified (using an uncorrected $p < 0.05$). If 1,000 transcripts were tested, then this would result in 50 likely false-positives. Many genome-wide analyses of RNA-Seq or array data routinely involve testing thousands of transcripts, which would potentially result in hundreds of false positives. This may explain why many transcription-based prognostic tests fail to be reproducible in independent cohorts.

Our previous work identified increased levels of cell cycle transcripts in cell lines generated from OS tumors with worse outcomes (13). We have now extended this work to OS tumors and provide a novel methodology for identifying transcriptionally related gene clusters, even if they are not the primary component present in the dataset, and testing their association with a variety of outcomes while minimizing errors from multiple testing. These results are highly consistent with the CINSARC signature (Complexity INdex in SARComas) predictive of metastasis free survival. Many of the genes identified encode for proteins involved in cell cycle/mitosis, cytokinesis, mitotic checkpoint, and DNA damage repair (34, 35).

Our data also indicate that cell cycle mediated events have more prognostic potential for dog disease progression compared to human disease. We speculate this is largely due to differences in therapeutic regimens and overall commitment to therapy, however it may also represent a fundamental difference between human and dog immune responses. Another potential confounding factor is the general characteristic of dog OS to progress much faster than typically observed in either normal or late-onset human OS. Dogs with OS may not survive long enough for the role of the immune cells to become observable. In specific dog breeds, the immune system may also have a reduced capability to recognize and respond to tumor cells.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Transcriptional profiles derived from grossly dissected tumors have variable types and quantities of cells present, including stromal and immune cells in addition to cancer cells. Our GCESS technique provides a straightforward method to indicate the relative abundance of immune cells in the tumor tissue sample, which can also be generally applied to any component present in a transcriptional dataset. Importantly, we compared results from our method with a previously published tool (ESTIMATE) which infers tumor purity (36). We found that in the naturally occurring human and dog OS tumor samples there was a strong correlation between our immune cell GCESS value and the ESTIMATE immune score.

The GCESS method provides a useful tool to distinguish between osteosarcoma tumor sub-types that seem to have distinctly different likelihoods of metastatic progression as a result of the presence of decreased numbers of immune cells within the tumor stroma. Supporting our conclusion is the lack of expression of these genes in OS cell lines, high correlation to the ESTIMATE scores, validation in multiple independent datasets, and a growing body of literature on the potential, variable immunogenicity in cancers such as OS.

As a genomically chaotic disease, conventional wisdom suggests OS must provide an abundance of antigens that typically would result in increased recognition by the host immune system, yet somehow OS tumors and their metastases have proven capable of escaping immune surveillance. Our results reinforce the theory that immune cell infiltration and monitoring is a normal process in bone tissue, and suggests that disruption of this process in a subset of tumors is associated with increased tumor mortality and progression to metastasis (Supplemental Figure 22). Determining whether the observed decrease in immune cells is intrinsically controlled by the tumor, or extrinsically by the immune system, or both, is an important, future research question in OS and other cancers.

Analyses of the TCGA data showed that the association between decreased immune cell GCESS and outcome was strongest in Cutaneous Melanoma (SKCM), a data set containing a large number of metastatic samples. Importantly, melanoma patients respond to immune checkpoint blockade therapy (37), and our results indicate that this same approach may have clinical utility in some OS patients. Furthermore, these results support an association between absence of immune cells, metastasis, and clinical outcome. Where immune cell outcome associations were seen in the TCGA the effects were more significant using the T cell enriched cluster relative to the monocyte enriched cluster. On the contrary, in our OS datasets the monocyte enriched gene cluster was more significant (lower p values) than the T cell enriched gene cluster.

We conclude that using multi-species datasets provides unique opportunities and insights to identify biologically meaningful gene signatures and to identify aspects of the immune response that can be manipulated therapeutically to improve the quality of life and outcomes of children with bone cancer, as well as patients with other sarcomas and solid tumors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Acknowledgments

The authors would like to thank John Garbe for technical assistance in the RNA-Seq data processing pipeline, Rebecca S. Larue for running the RNA-Seq data processing pipeline, Colleen Forster and the Clinical and Translational Science Institute's Bionet laboratory for optimization and performance of immunohistochemistry, the Minnesota Supercomputing Institute for computing time and storage space, and the University of Minnesota Genomics Center for sequencing services.

Grant Support

This work was generously supported by the Zach Sobiech Osteosarcoma Fund, Keren Wyckoff Rein in Sarcoma Foundation, grants to AL Sarver NCI(CA211249), JF Modiano NCI(CA208529) and Morris Animal Foundation(D13CA-032), DA Largaespada NCI(CA113636) and ACS (#123939), AE Sarver NIH(CA099936), S Subramanian ACS(RSG-13-381-01) and the Masonic Cancer Center Comprehensive Cancer Center Support Grant (CA077598).

References

1. Mirabello L, Troisi RJ, Savage SA. Osteosarcoma incidence and survival rates from 1973 to 2004: data from the Surveillance, Epidemiology, and End Results Program. Cancer. 2009; 115(7):1531–43. [PubMed: 19197972]
2. Damron TA, Ward WG, Stewart A. Osteosarcoma, chondrosarcoma, and Ewing's sarcoma: National Cancer Data Base Report. Clinical orthopaedics and related research. 2007; 459:40–7. [PubMed: 17414166]
3. Aljubran AH, Griffin A, Pintilie M, Blackstein M. Osteosarcoma in adolescents and adults: survival analysis with and without lung metastases. Annals of Oncology. 2009; 20(6):1136–41. [PubMed: 19153114]
4. Janeway KA, Maki RG. New strategies in sarcoma therapy: linking biology and novel agents. Clinical cancer research : an official journal of the American Association for Cancer Research. 2012; 18(21):5837–44. [PubMed: 22929804]
5. Kansara M, Thomas DM. Molecular pathogenesis of osteosarcoma. DNA and cell biology. 2007; 26(1):1–18. [PubMed: 17263592]
6. Marulanda GA, Henderson ER, Johnson DA, Letson GD, Cheong D. Orthopedic surgery options for the treatment of primary osteosarcoma. Cancer control : journal of the Moffitt Cancer Center. 2008; 15(1):13–20. [PubMed: 18094657]
7. Allison DC, Carney SC, Ahlmann ER, Hendifar A, Chawla S, Fedenko A, et al. A meta-analysis of osteosarcoma outcomes in the modern medical era. Sarcoma. 2012; 2012:704872. [PubMed: 22550423]
8. 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. Canadian journal of veterinary research = Revue canadienne de recherche veterinaire. 2011; 75(3):209–15. [PubMed: 22210997]
9. Moriarity BS, Otto GM, Rahrmann EP, Rathe SK, Wolf NK, Weg MT, et al. A Sleeping Beauty forward genetic screen identifies new genes and pathways driving osteosarcoma development and metastasis. Nature Genetics. 2015; 47(6):615–24. [PubMed: 25961939]
10. Temiz NA, Moriarity BS, Wolf NK, Riordan JD, Dupuy AJ, Largaespada DA, et al. RNA sequencing of Sleeping Beauty transposon-induced tumors detects transposon-RNA fusions in forward genetic cancer screens. Genome Research. 2016; 26(1):119–29. [PubMed: 26553456]
11. Varshney J, Scott M, Largaespada D, Subramanian S. Understanding the Osteosarcoma Pathobiology: A Comparative Oncology Approach. Veterinary Sciences. 2016; 3(1):3.
12. Fenger JM, London CA, Kisseberth WC. Canine osteosarcoma: a naturally occurring disease to inform pediatric oncology. ILAR journal/National Research Council, Institute of Laboratory Animal Resources. 2014; 55(1):69–85.
13. Scott MC, Sarver AL, Gavin KJ, Thayanithy V, Getzy DM, Newman RA, et al. Molecular subtypes of osteosarcoma identified by reducing tumor heterogeneity through an interspecies comparative approach. Bone. 2011; 49(3):356–67. [PubMed: 21621658]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

14. Angstadt AY, Thayanithy V, Subramanian S, Modiano JF, Breen M. A genome- wide approach to comparative oncology: high-resolution oligonucleotide aCGH of canine and human osteosarcoma pinpoints shared microaberrations. Cancer genetics. 2012; 205(11):572–87. [PubMed: 23137772]
15. Man TK, Lu XY, Jaeweon K, Perlaky L, Harris CP, Shah S, et al. Genome-wide array comparative genomic hybridization analysis reveals distinct amplifications in osteosarcoma. BMC cancer. 2004; 4:45. [PubMed: 15298715]
16. Xiong Y, Wu S, Du Q, Wang A, Wang Z. Integrated analysis of gene expression and genomic aberration data in osteosarcoma (OS). Cancer gene therapy. 2015; 22(11):524–9. [PubMed: 26427513]
17. Chen X, Bahrami A, Pappo A, Easton J, Dalton J, Hedlund E, et al. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. Cell reports. 2014; 7(1):104–12. [PubMed: 24703847]
18. Kuijjer ML, Rydbeck H, Kresse SH, Buddingh EP, Lid AB, Roelofs H, et al. Identification of osteosarcoma driver genes by integrative analysis of copy number and gene expression data. Genes, chromosomes & cancer. 2012; 51(7):696–706. [PubMed: 22454324]
19. Joseph CG, Hwang H, Jiao Y, Wood LD, Kinde I, Wu J, et al. Exomic analysis of myxoid liposarcomas, synovial sarcomas, and osteosarcomas. Genes, chromosomes & cancer. 2014; 53(1): 15–24. [PubMed: 24190505]
20. Scott MC, Sarver AL, Tomiyasu H, Cornax I, Van Etten J, Varshney J, et al. Aberrant RB-E2F Transcriptional Regulation Defines Molecular Phenotypes of Osteosarcoma. The Journal of biological chemistry. 2015; 290(47):28070–83. [PubMed: 26378234]
21. Perry JA, Kiezun A, Tonzi P, Van Allen EM, Carter SL, Baca SC, et al. Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111(51):E5564–73. [PubMed: 25512523]
22. Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, et al. MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Research. 2010; 38(18):e178. [PubMed: 20802226]
23. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. [PubMed: 21816040]
24. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics (Oxford, England). 2009; 25(9):1105–11.
25. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009; 25(16):2078–9.
26. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols. 2012; 7(3):562–78. [PubMed: 22383036]
27. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nature biotechnology. 2014; 32(9):903–14.
28. Sarver AL. Toward Understanding the Informatics and Statistical Aspects of Micro-RNA Profiling. Journal of Cardiovascular Translational Research. 2010; 3(3):204–11. [PubMed: 20560041]
29. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics (Oxford, England). 2014; 30(23): 3310–6.
30. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2017. URL https://www.R-project.org/
31. Therneau, TM. A Package for Survival Analysis in S version 2.38. 2015. Available from: http://CRAN.R-project.org/package=survival
32. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature methods. 2015; 12(5):453–7. [PubMed: 25822800]
33. Buddingh EP, Kuijjer ML, Duim RA, Burger H, Agelopoulos K, Myklebost O, et al. Tumor-infiltrating macrophages are associated with metastasis suppression in high-grade osteosarcoma: a rationale for treatment with macrophage activating agents. Clinical cancer research : an official

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

journal of the American Association for Cancer Research. 2011; 17(8):2110–9. [PubMed: 21372215]

34. Chibon F, Lagarde P, Salas S, Perot G, Brouste V, Tirode F, et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. Nature medicine. 2010; 16(7):781–7.
35. Lesluyes T, Perot G, Largeau MR, Brulard C, Lagarde P, Dapremont V, et al. RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. European journal of cancer (Oxford, England: 1990). 2016; 57:104–11.
36. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4
37. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer. 2012; 12(4):252–64. [PubMed: 22437870]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

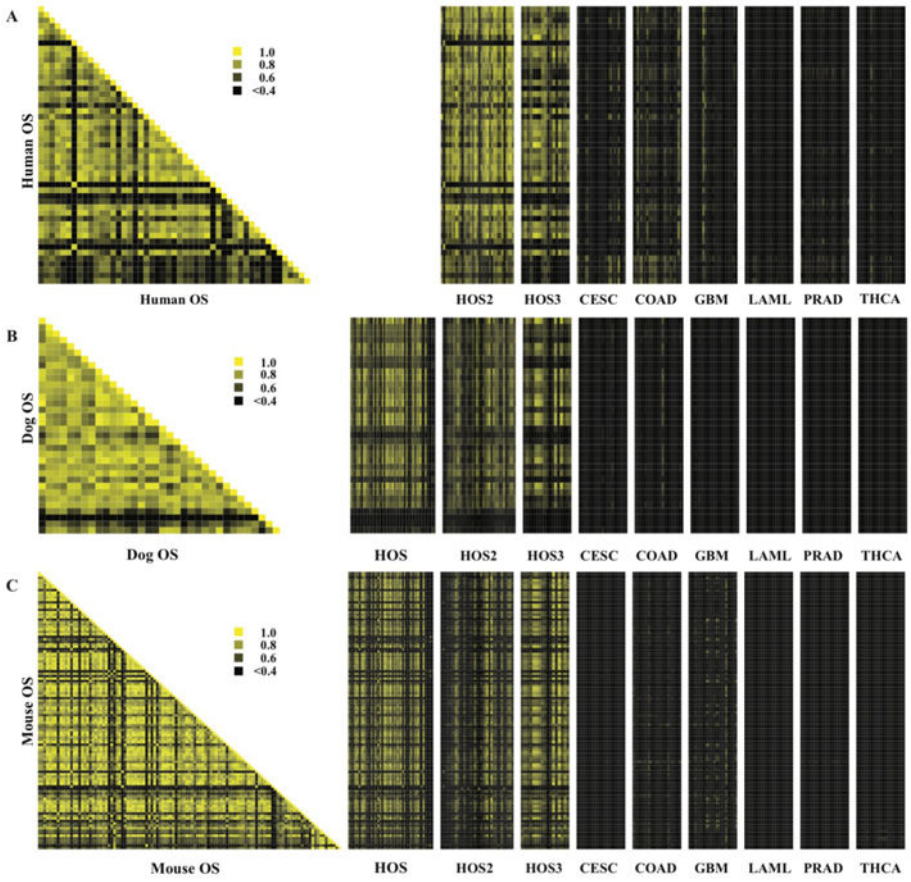

Fig 1. OS transcriptome profiles across 3 species are more similar to each other than to other types of human tumors

Pairwise Pearson correlations were calculated between (**A**) Human (HOS1) (**B**) Dog (DOS) and (**C**) Mouse (MOS) OS data sets, publicly available OS datasets (HOS2 and HOS3) and sets of 25 tumors from the TCGA representing a wide range of tumor types using 12,062 genes common in all 3 species.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

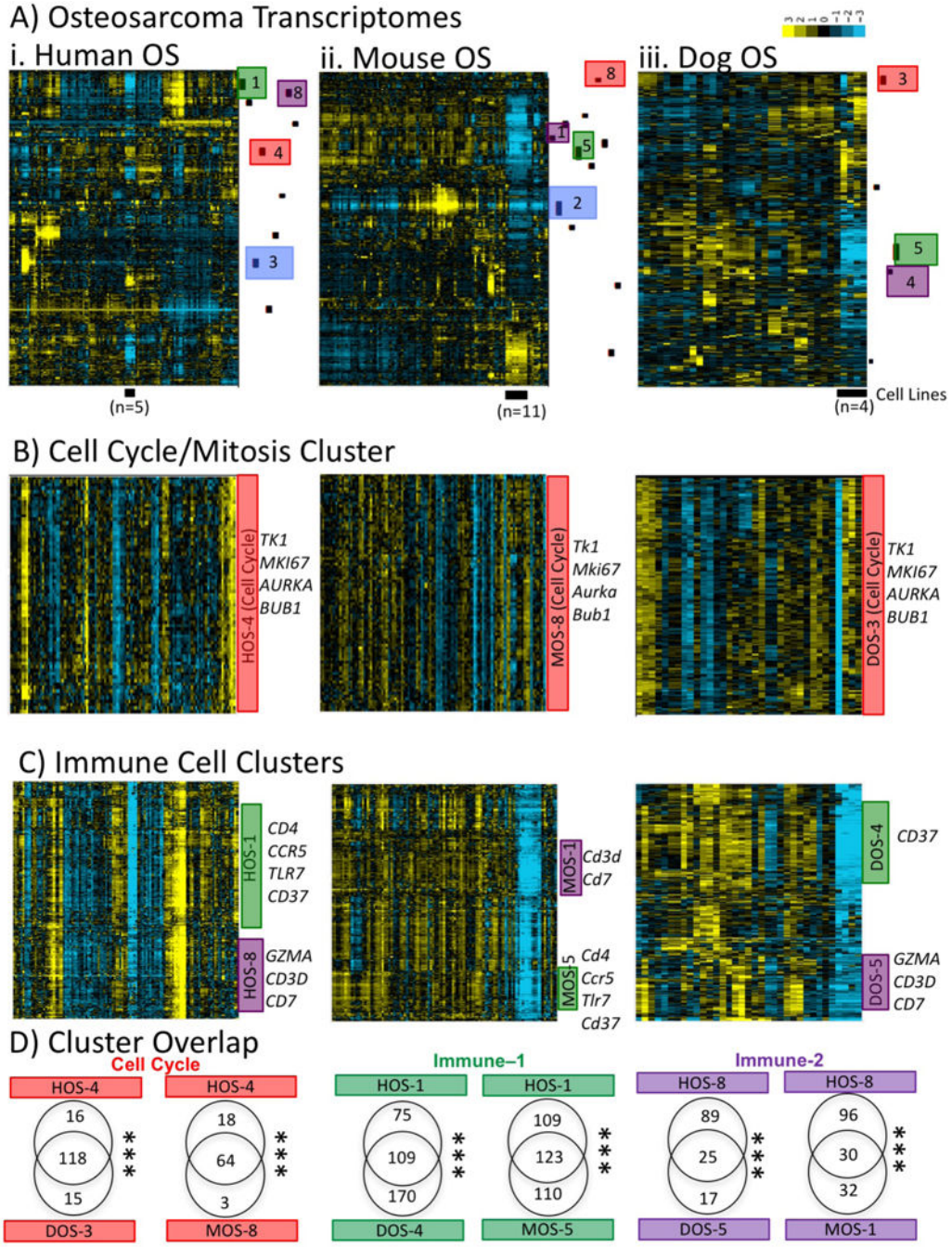

Fig 2. OS transcriptome profiles show common inter-tumor transcriptional variation across human, mouse, and dog samples

A) FPKM values derived from OS tumors and cell line data (indicated by black bars below heatmaps) were log transformed and mean centered within each species. Invariant genes were then removed leaving **(i)** human (n=9,190), **(ii)** mouse (n=8,051), and (**iii**) dog (n=8,003) genes which were used for unsupervised average linkage clustering. Transcripts with increased levels are shown in yellow while transcripts with decreased levels are shown in blue. Transcript level clusters with correlation > 0.60 and containing ≥ 60 genes were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

systematically identified and these clusters are visualized with a numbered black bar to the right of each of the heatmaps. Lists of each gene in each cluster are provided as Supplementary Table 4. Gene clusters observed in more than one species are surrounded by colored boxes and given a reference number. The “Cell Cycle” conserved transcript cluster is shown in red (HOS-4,MOS-8,DOS-3). The “Immune-1” transcript cluster is shown in green (HOS-1,MOS-4,DOS-4). The “Immune-2” transcript cluster is shown in Purple (HOS-8,MOS-1,DOS-5). A cluster composed of muscle transcripts only present in Human and Mouse data is shown in Blue (HOS-3, MOS-2). B) Blow up the conserved Cell Cycle clusters (HOS-4, MOS-4 and DOS-3) showing the location of representative genes. C) Blow up of the Immune-1 (HOS-1,MOS-4,DOS-4) and Immune-2 (HOS-8,MOS-1 and DOS-5) regions showing the location of representative genes. **D**) Venn diagrams showing the number of overlapping genes observed to be commonly present in both datasets. Fishers Exact Test results indicated that the observed overlap is highly unlikely to occur by random chance. Highly significant FET ($p < 10E-10$) are marked with ***.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

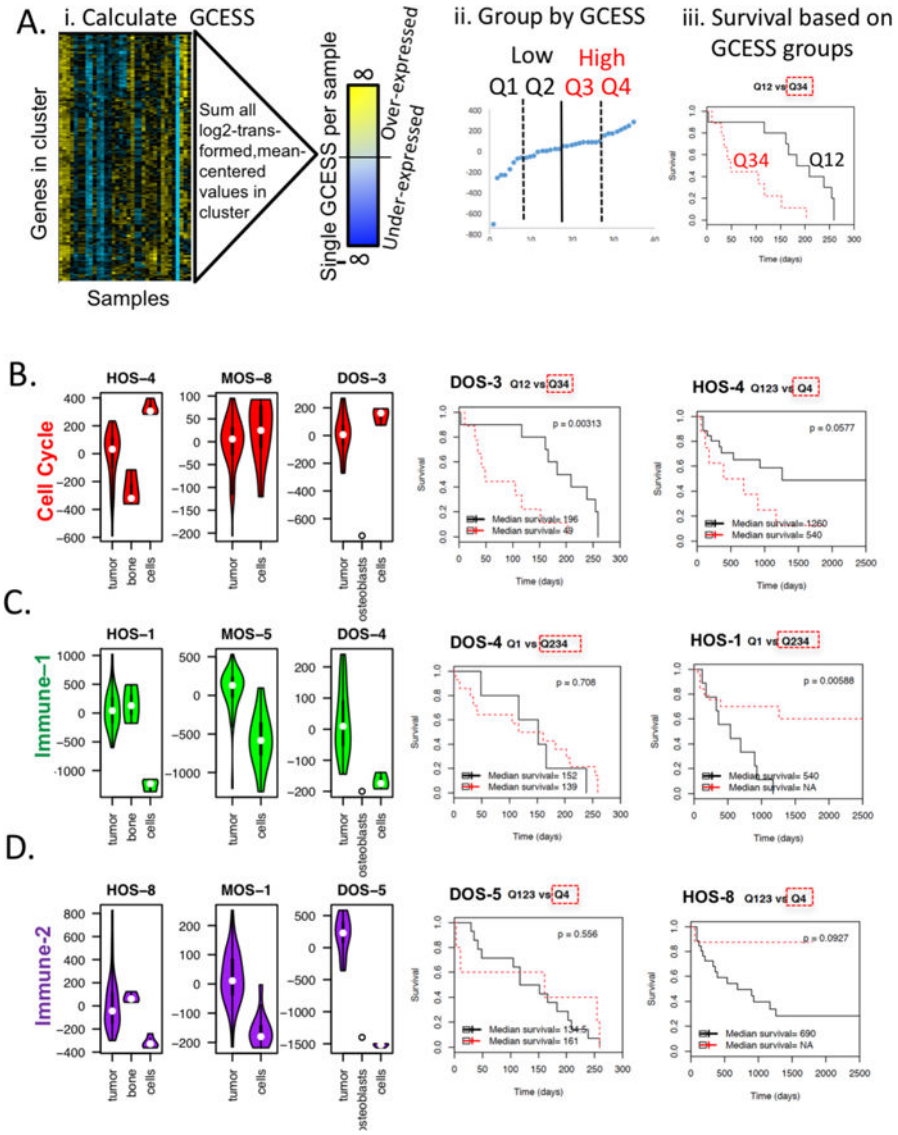

Fig 3. Gene Cluster Expression Summary Scores (GCESS) represent the relative amount of transcript present for each cluster of genes and identify correlation between high cell cycle or low immune cell GCESSs and poor survival

A) Overview of analyses method. **(i)** The log2-transformed and mean-centered values for each gene in a cluster are summed to generate a single score (GCESS value) for each sample. Samples with high relative expression levels have large positive GCESSs while samples with low relative levels have large negative GCESSs. **(ii.)** The GCESS scores can be used to separate the tumors into groups based on GCESS score Quartiles. **(iii.)** The GCESS Quartile based groups can then be examined for associations with outcome using Kaplin Meier Analyses. GCESSs for human (tumors, normal bone, OS cell lines), mouse (tumors, OS cell lines), and dog (tumors, osteoblasts, OS cell lines) samples were calculated and Violin Plots were generated for (**B**) the “cell cycle” cluster, (**C**) the “immune-1” cluster, and (**D**) the “immune-2” cluster. Samples were ranked by GCESS and divided into quartiles groups (Q1 = lowest GCESSs, Q4 = highest GCESSs). KM analyses were performed, using human (n=35) and dog (n=19) samples for which survival data was known, to determine

correlations between low/high GCESSs and survival. There were 17 (out of 35) death events in the human data (HOS2), and 19 (out of 19) death events in the dog data. (**B**) A significant association with shorter time to death was observed with a high cell cycle GCESS in the dog cohort, and a strong trend was also present in the human data. **(C)** In the human data, low Immune-1 GCESS was significantly associated with a worse survival and **(D)** a strong trend was present between low Immune-2 and worse survival

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

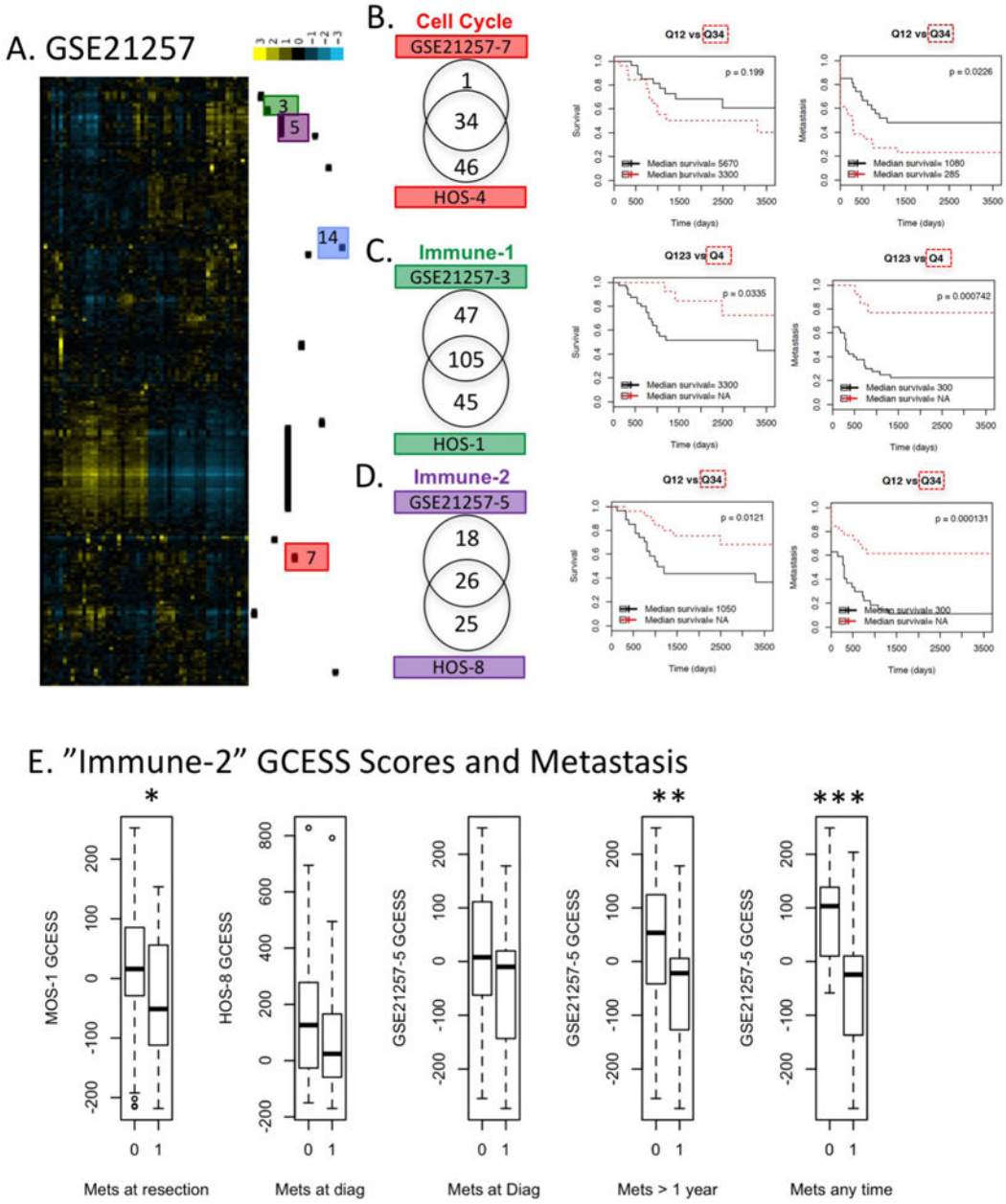

Fig 4. Replication of association between high cell cycle GCESS or low immune GCESS and poor survival outcomes using array data from independent cohort

Following a similar strategy to the one used for the RNA-Seq data, 14 strong and highly correlated clusters were identified in the **A)** GSE21257 dataset. Gene cluster overlap analyses comparing clusters derived from human array and human RNA-Seq data sets identified four clusters that corresponded to the conserved RNA-Seq clusters. The **B)** Cell cycle cluster is labeled in red, the **C)** Immune-1 cluster is labeled in green and the **D)** Immune-2 cluster is labeled in purple. Fisher Exact to assess the likelihood of observing the overlap by random chance indicated that the enrichment was highly significant ($p < 1E-10$) for the Cell Cycle and Immune clusters. KM analyses using GCESS groups using the approach outlined in Figure 3a showed that high levels of **B)** Cell cycle transcripts were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

associated with worse outcomes and significantly increased likelihood of tumor metastasis. Low levels of **(C)** Immune-1 and **(D)** Immune-2 transcripts were associated with significantly worse survival and significantly increased likelihood of tumor metastasis **(E)** Metastatic Samples have lower Immune-2 transcript levels in mouse and human samples. MOS-1 GCESS values were lower in tumors from mice where metastatic lesions were observed during necropsy ($p < 0.05$). Tumors from Human patients with metastases present at diagnoses showed a trend towards lower Immune-2 GCESS scores in both the RNA- SEQ data as well as the array data and this trend became increasingly significant in the array data when tumors where metastases were observed in the patient within one year ($p < 0.001$) or at any point ($p < 0.0001$) were included with the patients with metastases at diagnosis.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

RNA-Seq samples used for this study.

RNA-Seq samples used in this study					
Source	**Species**	**OS tumors**	**OS cell lines**	**Other**	**Total**
this study	Human	44	5	3 normal bone	52
this study	Mouse	92	11		103
this study	Dog	31	2	1 osteoblast cell line	34
Perry et al. (21)	Human	35			35
Previous Studies (9,15,17)	Human	25			25
					249

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cancer Res. Author manuscript; available in PMC 2018 July 15.

Table 2

Validation of GCESS methodology in TCGA cancer data sets

GCESSs were calculated for each of the TCGA tumor datasets (using the genes comprising the respective human OS clusters to identify each tumor's cell cycle and immune cell clusters) and KM-analyses were used to determine correlation with poor survival outcomes. (**A**) High cell cycle GCESSs (based on human OS cluster-4) were highly significantly associated (in red) with poor survival outcomes in KIRC, and clear trends were also apparent LIHC, LUAD, and PAAD. (**B**) Low Immune-1 GCESSs (based on human OS cluster-1 (Monocyte enriched)) were significantly associated (in grey) with poor survival outcomes in SKCM and clear trends were apparent in CESC, COAD, LUAD, and SKCM. (**C**) Low Immune-2 GCESSs (based on human OS cluster-8 (T-cell enriched)) were significantly associated (in grey) with poor survival outcomes in SKCM and clear trends were apparent in BRCA, CESC, COAD, HNSC, LIHC, LUAD, Rectum adenocarcinoma (READ), and SKCM.

HOS-4 "Cell Cycle"				**HOS-1 "Immune-1"**				**HOS-8 "Immune-2"**			
	q1 vs q234	**q12 vs** q34	**q123 vs** q4		**q1 vs** q234	**q12 vs** q34	**q123 vs** q4		**q1 vs** q234	**q12 vs** q34	**q123 vs** q4
BLCA	0.33	0.82	0.74	**BLCA**	0.72	0.81	0.64	**BLCA**	0.60	0.38	0.11
BRCA	0.94	0.11	0.71	**BRCA**	0.75	0.37	0.76	**BRCA**	0.04	0.07	0.28
CESE	0.41	0.27	0.55	**CESE**	0.27	0.31	0.01	**CESE**	0.07	0.02	0.008
COAD	0.41	0.31	0.79	**COAD**	0.01	0.15	0.01	**COAD**	0.04	0.97	0.44
GBM	0.95	0.97	0.22	**GBM**	0.29	0.19	0.47	**GBM**	0.84	0.56	0.18
HNSC	0.52	0.49	0.03	**HNSC**	0.56	0.43	0.15	**HNSC**	0.003	0.02	0.02
KIRC	0.17	1.26E-06	4.68E-15	**KIRC**	0.28	0.11	0.14	**KIRC**	0.05	0.03	0.20
LIHC	0.01	0.01	0.001	**LIHC**	0.09	0.25	0.78	**LIHC**	0.07	0.007	0.13
LUAD	0.002	0.01	0.02	**LUAD**	0.61	0.05	0.02	**LUAD**	0.20	0.09	0.008
LUSC	0.01	0.03	0.78	**LUSC**	0.21	0.30	0.11	**LUSC**	0.65	0.21	0.40
OV	0.78	0.73	0.88	**OV**	0.74	0.67	0.69	**OV**	0.44	0.17	0.19
PAAD	0.002	0.004	0.003	**PAAD**	0.08	0.87	0.52	**PAAD**	0.86	0.67	0.37
PRAD	0.37	0.06	0.77	**PRAD**	0.25	0.14	0.92	**PRAD**	0.94	0.20	0.92
READ	0.21	0.39	0.33	**READ**	0.66	0.64	0.85	**READ**	0.03	0.35	0.96
SKCM	0.11	0.04	0.83	**SKCM**	4.99E-03	2.21E-05	6.48E-04	**SKCM**	6.58E-05	4.43E-06	1.88E-04

Lower GCISS's → poor outcome

Higher GCESS's → poor outcome