bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1

2 Kaposi's Sarcoma Herpesvirus Is Associated with Osteosarcoma in
3 Xinjiang Uyghur Population

4

5 (Running title: KSHV and Osteosarcoma)

6

7 Qian Chen [a,¶], Jiangtao Chen [b,¶], Yuqing Li [a,¶,#], Dawei Liu [c], Yan Zeng [d], Zheng Tian[b],
8 Akbar Yunus [b], Yong Yang [b], Jie Lu [e], Xinghua Song [b,##,*] and Yan Yuan [a,e,*]
9
10 [a] Institute of Human Virology, Zhongshan School of Medicine, Sun Yat-Sen University,
11 Guangzhou, China;
12 [b] First Affiliated Hospital, Xinjiang Medical University, Urumqi, Xinjiang Uyghur
13 Autonomous Region, China;
14 [c] Department of Pathology, The First Affiliated Hospital, Sun Yat-sen University,
15 Guangzhou, China;
16 [d] Shihezi University, Shihezi, Xinjiang Uyghur Autonomous Region, China;
17 [e] Department of Basic and Translational Sciences, University of Pennsylvania School of
18 Dental Medicine, Philadelphia, PA 19104, USA
19
20 [¶] Equal Contribution.
21
22 [#] Present address: Department of Urological Surgery, Affiliated Luohu Hospital,
23 Shenzhen University, Shenzhen, China;
24 [##] Present address: The Orthopedic Department, the First Affiliated Hospital of Jinan
25 University, Guangzhou, China
26
27 *Corresponding authors. Mailing Address: Department of Basic and Translational
28 Sciences, University of Pennsylvania School of Dental Medicine, Philadelphia, PA 19104
29 Email: yuan2@upenn.edu; Phone: (215) 573-7556; Fax: (215) 898-8385.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1

2 Keywords: Kaposi's sarcoma-associated herpesvirus (KSHV), Human herpesvirus type 8
3 (HHV-8), Osteosarcoma, Uyghur ethnic population, Sero-epidemiology, viral-associated
4 endemic cancer

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 ## Abstract

2 Osteosarcoma is the most common malignant tumor of bone predominately affecting
3 adolescents and young adults. Viral etiology of osteosarcoma has been proposed more
4 than a half-century ago but never been proven by identifying any virus authentically
5 associated with human osteosarcoma. The Uyghur ethnic population in Xinjiang China
6 has an unusually high prevalence of Kaposi's sarcoma-associated herpesvirus (KSHV)
7 infection and elevated incidence of osteosarcoma. In the current study, we explored the
8 possible association of KSHV infection and osteosarcoma occurrence. Our
9 seroepidemiological study revealed that KSHV prevalence was significantly elevated in
10 osteosarcoma patients versus the general population in the Xinjiang Uyghur population
11 (OR, 10.23; 95%CI, 4.25, 18.89). The KSHV DNA genome and viral latent nuclear
12 antigen LANA were detected in most osteosarcoma tumor cells. Gene expression
13 profiling analysis showed that KSHV positive osteosarcoma represents a distinct subtype
14 of osteosarcomas with viral gene-driven signaling pathways that are important for
15 osteosarcoma development. We conclude that KSHV infection is a risk factor for
16 osteosarcoma and KSHV is associated with some osteosarcomas, representing a newly
17 identified viral-associated endemic cancer.

18 ## Significance

19 Viral etiology of osteosarcoma was proposed previously but has never been proven by
20 identifying any virus that is authentically associated with human osteosarcoma. The
21 current study revealed an association of human osteosarcoma with KSHV infection in
22 Uyghur osteosarcoma patients. First, this study provides the first evidence that supports

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 the possible viral etiology of human osteosarcoma. The gene expression profiling study
2 showed that KSHV-positive osteosarcoma represents a distinct subtype of osteosarcomas,
3 which is of diagnostic, prognostic and therapeutic significance. Second, KSHV-
4 associated osteosarcomas preferentially occur in children and young adults, predicting
5 that KSHV-positive children in KSHV endemic region may be at great risk for
6 osteosarcoma. Third, the finding extended the range of human cancers associated with
7 viruses.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 # Introduction

2 Osteosarcoma is the most common malignant tumor of bone with an incidence of
3 approximately three cases per million annually worldwide, predominately affecting
4 adolescents and young adults with a major peak between 10 and 14 years old and a
5 second smaller peak in the geriatric population (1, 2). Osteosarcomas have a rather
6 heterogeneous genetic profile and lack any consistent unifying event that leads to its
7 pathogenesis. It may result from oncogenic events sustained by cells in the differentiation
8 lineage hierarchy from mesenchymal stem cells through to osteoblast. Still, it may also
9 suggest that osteosarcoma could arise with multiple etiologies and different pathogenesis.
10 Concerning the etiology of osteosarcoma, three primary potential etiological agents,
11 namely chemical agents, physical agents, and viruses, are believed to associate with
12 osteosarcoma. Numerous chemicals are known to induce osteosarcoma, such as beryllium
13 and methylcholanthrene (3, 4). Radiation exposure is also related to osteosarcoma
14 development (5, 6). The viral etiology for osteosarcoma has been suggested since Finkel
15 isolated viruses from mice that produced similar tumors when injected into newborn mice
16 (7). There was also evidence of a bone tumor virus in the human disease as injection of
17 cell-free extracts of human bone cancer into newborn Syrian hamsters induced various
18 mesenchymal tumors including osteosarcoma (8). However, viral etiology has never been
19 proven by the identification of any virus that is authentically associated with human
20 osteosarcoma. The current study aimed to explore if any human osteosarcoma is
21 associated with a viral infection.

22 Kaposi's sarcoma-associated herpesvirus (KSHV), also termed human herpesvirus
23 type 8 (HHV-8), was first identified in Kaposi's sarcoma (KS) lesions in 1994 (9). KSHV

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 can be found in almost 100% of KS lesions, regardless of their source or clinical subtype
2 (i.e., classic, AIDS-associated, African endemic, or post-transplant KS). Additionally,
3 KSHV is also associated with two lymphoproliferative diseases, namely primary effusion
4 lymphoma (PEL) and multicentric Castleman's disease (MCD) (10 – 12). We and others
5 showed that human mesenchymal stem cells (MSCs) are highly susceptible to KSHV
6 infection and infection promotes multi-lineage (osteogenic, adipogenic and angiogenic)
7 differentiation (13 – 15). Several lines of evidence suggest that the KSHV-infection of
8 MSCs leads to KS through a mesenchymal-to-endothelial transition (MEndT) process
9 (13, 16).

10 Given that osteosarcomas originate from mesenchymal stem cells or their
11 immediate lineage progenitors (17, 18) and KSHV can effectively infect MSCs and drive
12 osteogenic differentiation (13), the question was raised if KSHV infection of MSCs
13 contributes to osteosarcoma development. Towards this end, we investigated the
14 association of KSHV infection and osteosarcoma occurrence in the ethnic
15 Uyghur population in Xinjiang Uyghur autonomous region of China, where there is a
16 high seroprevalence of KSHV, a high incidence of KS and a high occurrence of
17 osteosarcoma among the Uyghur population (19– 21). We found that the seroprevalence
18 of KSHV in Uyghur osteosarcoma patients is significantly higher than that of the Uyghur
19 general population. KSHV genomic DNA and viral latent nuclear antigen (LANA) were
20 detected in osteosarcoma tumors of most of the patients who were KSHV seropositive.
21 Furthermore, gene expression profile analysis of osteosarcoma clinical samples
22 demonstrated that KSHV infection regulates the genes and signaling pathways essential
23 for osteosarcoma development. These results revealed a strong association between
24 KSHV infection and osteosarcoma.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Results**

2 **Sero-epidemiological evidence for association of KSHV infection with osteosarcoma**
3 **in Xinjiang Uyghur ethnic population**

4 Xinjiang, China is an endemic area for Kaposi's sarcoma (KS) and classic KS is
5 prevalent among the ethnic Uyghur population (19). The seroprevalence of KSHV among
6 the Xinjiang Uyghur population is unusually high, ranging from 20.7 to 40.4% compared
7 to that of the general population of China (11.3%) (20). It is striking that although the
8 Uyghurs account for 45% of the total Xinjiang population (according to the 2010 Census
9 of Xinjiang), 68% of osteosarcoma patients diagnosed in our hospital (the First Affiliated
10 Hospital, Xinjiang Medical University) happened to be Uyghurs. It raised a question if
11 KSHV infection is associated with at least some osteosarcomas. To obtain
12 epidemiological evidence regarding this, we compared the KSHV seroprevalence
13 between osteosarcoma patients and the general population. A diagnostic enzyme-linked
14 immunosorbent assay (ELISA) was developed with three recombinant proteins of KSHV,
15 namely LANA, ORF65 and K8.1, and used to examine sera of 21 Uyghur osteosarcoma
16 patients and 327 control individuals (the general Uyghur population). Seventeen of 21
17 (81%) Uyghur osteosarcoma patients were found seropositive to at least one of the
18 KSHV antigens (*SI Appendix*, Table S1). The serum samples of osteosarcoma patients
19 were also examined in a blinded fashion by an immunofluorescence assay with BCBL-1
20 cells and LANA nuclear staining with distinctive punctate dots was detected with sera of
21 17 out of 21 patients, consistent with the results of the ELISA (Fig. 1). In contrast, the
22 control group sera showed a seropositive rate of 29.4% (Table 2; *SI Appendix*, Table S2),
23 consistent with the data previously reported (19, 20). Odds ratio (OR = 10,23, 95%CI:

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 4.25, 18.89) and P-value (P < 0.0001) indicate that KSHV infection is a risk factor for
2 osteosarcoma occurrence in Xinjiang Uygur population.

3 **The KSHV genomic DNA and LANA protein can be detected in most of the**
4 **osteosarcoma tumors of the KSHV seropositive patients**

5 Surgical specimens of osteosarcoma tumors and adjacent normal tissue samples
6 were obtained from 17 of these 21 patients. These included 14 cases of KSHV
7 seropositive and three seronegative patients. These samples were examined in a blinded
8 fashion for the presence of the KSHV genome in these tumors (the examiners were
9 unaware of patient identities and sample types). Nested PCR was employed to detect
10 KSHV genomic DNA using five sets of primers targeting viral genes K5, ORF25,
11 ORF26, ORF37 and ORF73 (LANA) (Fig. 2A). Quantitative real-time PCR was also
12 performed using ORF73 specific primers in a blinded fashion (Fig. 2B). With 100%
13 consistency, both assays showed that the KSHV DNA genome was detected in 12 out of
14 14 tumors from the KSHV-seropositive patients, and the other two, namely P14 and P16,
15 had tumors in which KSHV DNA is absent or below the detection of the PCR. The
16 KSHV genome was not detected in the tumors from three KSHV-seronegative patients.
17 No KSHV DNA sequence was detected in adjacent normal tissues of all cases, regardless
18 of their KSHV serological status. Two sets of primers for the EBV genome were included
19 in the Nested PCR assay. No EBV DNA was detected in any of these tumors (Fig. 2A).
20 Osteosarcoma samples of five patients were subjected to immunohistochemical
21 analysis for KSHV latent nuclear antigen (LANA). P1, P11, P20 and P21 osteosarcomas
22 exhibited intense LANA staining in spindle- or cigar-shaped osteosarcoma cells (Fig.
23 2C), indicating that these osteosarcoma cells carry latently infected KSHV. The

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 osteosarcoma of P16 did not express LANA, which was consistent with the fact that the
2 KSHV genome was not detected in the tumor of this patient.

3 **Gene expression profiling reveals that KSHV-positive osteosarcoma represents a**
4 **distinct subtype of osteosarcomas.**

5 The gene expression profiles of osteosarcoma samples were characterized to reveal
6 possible role of KSHV in osteosarcoma development. Total RNAs were purified from six
7 KSHV-positive (P1, P5, P9, P11, P20 and P21) and four KSHV-negative osteosarcomas
8 (P2, P3, P14 and P16) along with related adjacent normal tissues (except P2 that lacks its
9 adjacent normal tissue) and subjected to RNA-seq analysis. RNA-seq reads were first
10 mapped to the KSHV genome (GQ994935.1) (*SI Appendix*, Table S3) and visualized on a
11 linear scale to provide an overview of highly expressed regions of the genome. At first
12 glance, six KSHV-positive osteosarcomas exhibit two distinct types of viral gene
13 expression pattern: (i) P1, P11 and P20 expressed a high level of PAN RNA; (ii) P5, P9,
14 and P21 expressed a relatively low level of PAN RNA, but a high level of K2 (vIL-6). In
15 addition, the expression of ORF4 (KCP), ORF45 and ORF50 (RTA) were also detected
16 in these tumors (Fig. 3A). Therefore, KSHV-positive osteosarcomas can be grouped into
17 two classes based on the viral gene expression profiles, namely the PAN class and the
18 vIL-6 class. The complete range of read depths across the KSHV genome is visualized on
19 a log scale (*SI Appendix*, Fig. S1).

20 Then the RNA-seq reads were aligned to the human genome (hg19/GRCh37) and
21 the FPKMs (Fragments Per Kilobase Million) were subjected to unsupervised clustering
22 and differential expression analyses. Clustering of cell samples (X-axis) and genes (Y-
23 axis) were performed by hierarchical clustering with average linkage method and

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 Euclidean distance metric (Fig. 3B). The convergence and divergence among these
2 osteosarcomas and their adjacent normal tissues were determined by linkage distance
3 based on the Pearson correlation coefficient and the principal component analysis (PCA).
4 Results show that the gene expression profiles of all ten osteosarcomas are distinct from
5 the adjacent normal tissues. Four KSHV-negative osteosarcomas are very similar in their
6 gene expression profiles but distantly categorized from KSHV-positive tumors ($p<0.05$)
7 (Fig. 3C and D), suggesting that KSHV-positive osteosarcomas arose by different
8 pathogenesis than that of non-viral osteosarcomas. Furthermore, KSHV-positive
9 osteosarcomas were divided into two categories visualized by PCA on their cellular gene
10 expression profiles (Fig. 3D). Interestingly, these two categories respectively correspond
11 to the PAN and the vIL-6 classes, suggesting that the different viral gene expression
12 pattern leads to divergent cell reprogramming through either directly affecting host gene
13 expression or changing the environment of host cells.

14 Differentially expressed genes (DEGs) of osteosarcomas versus their adjacent
15 normal tissues were subjected to a Gene Ontology (GO) analysis to reveal specific and
16 significant associations with specific GO terms. KSHV-positive and -negative
17 osteosarcomas share some common characteristics in gene expression but exhibit more
18 diversity (*SI Appendix*, Fig. S2A). For instance, all categories of osteosarcoma exhibit
19 enriched "multicellular organism development", "angiogenesis" and "Extracellular matrix
20 organization" Biological Process categories. However, "inflammatory response" and
21 "positive regulation of cell migration" are unique to KSHV-positive osteosarcomas, while
22 "DNA repair" and "response to organic substance" terms are only seen in KSHV-negative
23 osteosarcomas (*SI Appendix*, Fig S2 B – D). Besides, common and unique terms of
24 "Biological Process" or "Molecular Function" are also found between two KSHV-

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 positive classes (PAN-class and vIL6-class) (*SI Appendix*, Fig. S2). It is worth noting that
2 "inflammatory response" is significantly enriched in KSHV-positive osteosarcoma, both
3 PAN-class (2.73% DEGs involved, $p = 0.006$) and vIL6-class (2.78% DEGs involved, p
4 $= 0.0021$), but not in KSHV-negative osteosarcomas (*SI Appendix*, Fig S2. B – D).
5 Furthermore, "positive regulation of interferon-gamma production" is uniquely enriched
6 in vIL6-class (0.67% DEGs involved, $p = 0.024$). These data provide additional evidence
7 of the virus burden in KSHV-positive osteosarcomas and suggest elevated inflammation
8 in KSHV-associated osteosarcoma. Overall, our gene expression profiling analysis
9 showed that KSHV-associated osteosarcoma represents a distinct subtype of
10 osteosarcomas with viral gene-driven signaling pathways that are important for
11 osteosarcoma development. KSHV-positive osteosarcoma is a newly identified viral-
12 associated endemic cancer.

13 ## Discussion

14 Numerous agents have been implicated in osteosarcoma etiology, including
15 chemical, radiation, and virus (3 – 8, 22). Although osteosarcoma can be induced by a
16 virus in experimental animals (7), whether any human osteosarcoma is caused by virus
17 has been a longstanding and intriguing question but remain elusive. In the current study,
18 we found an association of human osteosarcoma with KSHV in the Xinjiang Uyghur
19 population that has an unusually high prevalence of KSHV infection. First, a serological
20 study for KSHV prevalence in Uyghur osteosarcoma patients versus the general Uyghur
21 population provided epidemiological evidence that KSHV infection is a risk factor for
22 osteosarcoma. Second, the KSHV genome and viral latent protein LANA were detected
23 in most osteosarcoma tumors of KSHV-positive patients. Third, gene expression profiling

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

analysis showed that KSHV-positive osteosarcoma represents a distinct subtype of osteosarcoma. Taken together, our results demonstrated that KSHV-positive osteosarcoma is a newly identified viral-associated endemic cancer, presenting the first evidence that human osteosarcoma is associated with an oncogenic virus.

Another important phenomenon revealed in this study is that KSHV-associated osteosarcomas preferentially occur in children and young adults. 86% osteosarcomas of the patients aged under 30 years are KSHV-associated, in contrast to that none of the patients above 30 was found to have KSHV-positive osteosarcoma (P-value = 0.024, Table 1). It has been reported that KSHV infection occurs in early childhood in the Xinjiang Uyghur population (23). Thus, KSHV-positive children may be at significant risk for osteosarcoma, which appears to be true in the Uyghur population. Interestingly, two older adult patients (P14 and P16, aged 53 and 39, respectively) had KSHV-negative osteosarcomas despite being KSHV seropositive. This observation suggests that KSHV-positive osteosarcoma is mainly associated with pediatric patients. In adults, osteosarcoma may, therefore, arise through non-viral etiology even in KSHV-seropositive individuals.

Osteosarcoma is not a single disease but a collection of neoplasms with different etiologies, sharing a histological hallmark of osseous matrix production in association with malignant cells. Our study showed that KSHV-positive and -negative osteosarcomas exhibit distinct gene expression profiles, implicating that osteosarcomas caused by biological (virus), chemical (carcinogens), or physical (radiation) agents can be categorized into different subtypes based on their gene expression profiles and specific GO enrichment. Furthermore, KSHV-positive osteosarcomas can be classified into two distinct classes (the PAN class and the vIL-6 class) based on their viral and cellular gene

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 expression profiles. Further studies are warranted to determine if these gene signature-
2 defined categories are linked to certain clinical features and osteosarcoma manifestation,
3 which will be of great diagnostic or prognostic significance. On the other hand, it is also
4 important to identify gene signatures for pan-osteosarcomas. For example, TGF-β1 and 3,
5 as well as their receptor TGF-βRII, were found to be up-regulated in all osteosarcomas
6 regardless of KSHV infection status. It is consistent with previous observations that high-
7 grade osteosarcomas have a significantly higher expression of TGF-β1, which may
8 influence the aggressive clinical behavior of the sarcoma (24), and that TGF-β3 is
9 associated with disease progression (25). In addition, "multicellular organism
10 development" and "extracellular matrix (ECM) organization" are also found at the top of
11 the significant GO term list for all osteosarcoma samples and present a common feature
12 for all osteosarcomas.

13 There is geographic variation in the prevalence of KSHV and associated diseases
14 (26). The prevalence of KSHV-associated osteosarcoma, as well as the ratio of
15 osteosarcomas with different etiologies, may vary in different regions upon the causative
16 virus prevalence and ethnic background. In North America and the Han population of
17 China (the ethnic majority of China) where the KSHV prevalence is low, KSHV-
18 associated osteosarcomas are believed to be rare. We collected nine osteosarcoma patients
19 of the Han, and only one was KSHV-seropositive (11%) (*SI Appendix*, Table S4). We
20 analyzed this Han KSHV-positive osteosarcoma (HP2) along with a KSHV-negative
21 tumor (HP9) for their gene expression profiles. Result showed that HP2 osteosarcoma is a
22 PAN-class tumor and the biology of the HP2 osteosarcoma is similar to Uyghur tumors of
23 the same category (P1, P11 and P20), revealed RNA-seq principle component analysis

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 (PCA) (*SI Appendix*, Fig. S3). Although the sample size for osteosarcomas of Han
2 patients is too small to draw any conclusion, the trend suggests that (i) KSHV-associated
3 osteosarcoma may not be the majority of this cancer in the Han population; (ii) once
4 being infected by KSHV at an early age, children of other races can also develop KSHV-
5 associated osteosarcoma. KSHV prevalence is known to be high in specific areas such as
6 Eastern African (Uganda, Cameroon, Democratic Republic of Congo, Tanzania and
7 Zambia). It is very intriguing to know if the occurrence of childhood osteosarcoma is
8 high in this area. Since many developing countries do not have cancer registries or many
9 childhood cancers are not diagnosed, there is a lack of epidemiology study for childhood
10 osteosarcoma in these countries. However, a simulation-based study of global childhood
11 cancer suggested that pediatric osteosarcoma in Eastern Africa is several-fold higher than
12 in Southern and Northern Africa in parallel with the incidence of Kaposi's sarcoma (27).
13 On the other hand, the possibility also exists that KSHV-associated osteosarcoma is
14 unique in the Xinjiang Uyghur population, analogous to EBV-associated Burkitt's
15 lymphoma, which is common in African children, and EBV-caused nasopharyngeal
16 carcinoma, which mainly occurs in Southern China. Further investigation is warranted to
17 identify the risk factors associated with the area or population in hope that strategies can
18 be developed to reduce osteosarcoma occurrence such as preventing children from KSHV
19 infection at an early age.

20 Materials and Methods

21 Patients and participants

22 The study included patients with osteosarcoma (age from 5 to 53 years-old, median 15

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 years-old) who were admitted and treated in the period of 2016 – 2019 in the Division of
2 Orthopedic Oncology, the First Affiliated Hospital of Xinjiang Medical University,
3 Urumqi, China. Among the patients, the majority is of Uyghur ethnicity (21 Uyghurs, 1
4 Kazakh, 9 Hans). All patients underwent a core needle biopsy or surgical biopsy for
5 diagnosis. Blood samples were also collected. Fifteen Uyghur osteosarcoma patients
6 received 4 – 10 weeks neo-adjuvant chemotherapy (Doxorubicin, Cisplatin and
7 Ifosfamide) and six were not treated with chemotherapy prior to surgery. Then extensive
8 resection or radical surgery was performed. Surgically removed osteosarcoma tumors and
9 adjacent normal tissues were collected from these patients. Demographic and clinical data
10 were listed in Table 1 and *SI Appendix* Table S1. Sera were also collected from 327
11 healthy Uyghur donors who came to the hospital for annual physical examination as
12 reference (*SI Appendix,* Table S2). The human sample collection and the use of clinical
13 samples in the research were approved by the Institutional Review Boards of the Xinjiang
14 Medical University, the First Affiliated Hospital (Approval No. 2018-112903) and Sun
15 Yet-sen University (Approval No. 2015-028). Informed consent was obtained from all
16 osteosarcoma patients (or their guardians) and healthy donors.

17 **Reagents and antibodies**

18 Cell culture medium (MEM-alpha), streptomycin, penicillin, TRIzol reagent were
19 purchased from Invitrogen. Nonessential amino acids, glutamine, β-glycerophosphate,
20 dexamethasone, Alizarin Red S, paraformaldehyde were obtained from Sigma-Aldrich.
21 Antibody against LANA (ab4103) antibody was purchased from Abcam. Alexa Fluor 555
22 goat anti-Human IgG (H+L) antibody (A21433) was purchased from Invitrogen. Alexa
23 Fluor 555 goat anti-Rat IgG antibody (A21434) was purchased from Life Technologies.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 Hematoxylin (G1004) and eosin (G1002) were obtained from Servicebio technology.
2 HRP labeled goat anti-rabbit IgG (SP-D2) and DAB reaction kit (DAB-1031) were
3 purchased from Maxim Biotechnologies.

4 **Expression and purification of KSHV K8.1, ORF65, LANA proteins in *E. coli***

5 cDNAs of KSHV K8.1, ORF65 and LANA were cloned into the pET-28a vector with a
6 hexahistidine (6xHis) tagged at the N-terminus. *E.coli* Rosetta cells were transformed
7 with each of the recombinant plasmids and cultured in LB medium contain 30μg/mL
8 Kanamycin. The expression of these proteins was induced with 1mM isopropyl β-D-
9 thiogalactoside (IPTG) when the optical density of culture reached OD of 0.6. For K8.1,
10 induced culture was collected after 4 hours cultivation at 16°C, while for LANA and
11 ORF65, cultures were collected after 4 hours cultivation at 37°C. Cells were
12 ultrasonicated in lysis buffer containing PMSF without DTT and proteins were purified
13 with Ni-NTA column chromatography. Protein concentrations were determined by BCA
14 protein assay kit (Thermo Fisher).

15 **Enzyme-Linked Immunosorbent Assay (ELISA)**

16 Purified K8.1, ORF65, and LANA proteins (100 μl, 5 μg/mL) were respectively coated
17 on ELISA plates (Jet Biofil) in coating buffer (0.1 M $NaHCO_3$, pH9.6 –10.0) at 4°C
18 overnight. Plates were saturated with blocking buffer (5% dried skimmed milk in PBS
19 containing 0.1% Tween 20). Each serum in a series of dilutions (1:50 to 1:1600) reacted
20 with coated plates at 37°C for 90 min. After washing with PBST three times, a
21 peroxidase-conjugated anti-human IgG antibody (1:3000 dilution) was added and
22 incubated at 37°C for 30min. Tetramethyl-benzidine (TMB) and hydrogen peroxide

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 substrates were dispensed and incubated in the dark at 37°C for 15 min. The plates were
2 read at 450 nm (OD450) using a microplate reader (BioTek) with the cut-off for
3 seropositivity of OD450>0.5. Before this ELISA system was used to analyze
4 osteosarcoma patient sera, the system had been verified with 21 Uyghur sera with known
5 KSHV serological status (LANA, ORF65m K8.1 seropositivity) (28) with 98.4%
6 consistency and accuracy. Each osteosarcoma patient serum was tested three times in a
7 blinded fashion. An ELISA titer of >1:100 was considered to be positive. To ensure inter-
8 assay comparability, we used a "highly positive" serum from the previous assay as a
9 positive control and a negative serum from a healthy donor as a negative control.

10 **Immunofluorescence Assay (IFA)**

11 BCBL-1 (KSHV latently infected) and BJAB (KSHV-negative) cells were fixed with 3.6%
12 formaldehyde in PBS and permeabilized with 0.1% Triton X-100. Cells were pre-
13 incubated with 1% BSA and then reacted with patient sera in two-fold serial dilutions.
14 AIDS-KS patient serum and an anti-LANA antibody (ab4103, Abcam) were included as
15 positive controls. Alexa Fluor555 goat anti-human IgG (1:500 dilution) was used as the
16 secondary antibody (Alexa Fluor555 goat anti-Rat IgG for anti-LANA control). Cells
17 were visualized under a Zeiss Observer.Z1 fluorescence microscope. The assay was done
18 in a blinded manner, and the serum yielding the nuclear punctate pattern in > 1:64
19 dilution was considered positive.

20 **Polymerase chain reaction (PCR) analyses**

21 Total DNA was extracted from each tumor or adjacent normal tissue sample using a
22 HiPure Tissue DNA Mini Kit (Magen). Two hundred nanogram of each DNA was

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 subjected to nested polymerase chain reaction (PCR). The oligonucleotide primers for the
2 first and second-round PCR for KSHV K5, ORF25, ORF26, ORF37, ORF73 (LANA),
3 and EBV LMP1, EBNA1 were listed in *SI Appendix* Table S5. PCR was performed in a
4 50 μl-volume reaction [0.4 μM primers and 2xPrimeSTAR HS (Premix), 5% DMSO] as
5 follows: denaturing at 95°C for 5 min, 25 cycles of reaction (95°C for 30 sec, 55–62°C
6 for 30 sec, 72°C for 30 sec) and final elongation of 72°C for 7 min. PCR was carried out
7 in a blinded fashion (the examiner was unaware of patient identities and sample types).
8 Each sample was tested three times independently.

9 **Real-time polymerase chain reaction (PCR) and quantitative RT-PCR analyses**

10 Two hundred nanogram purified tissue DNA was subjected to quantitative real-time PCR
11 with KSHV ORF73 primers and SYBR Green Master Mix (ThermoFisher) on
12 LightCycler 480II (Roche). LANA standard DNA in serial dilutions were analyzed
13 simultaneously with osteosarcoma DNAs for a standard curve. KSHV genomic DNA
14 copy numbers of initial specimens were calculated according to the standard curve. The
15 lowest standard DNA copy number used to construct the standard curve (10 DNA copies)
16 still exhibited positive peak and was in the linear range in the standard curve, 10 copies is
17 set to be the sensitive line for our assay. Ct value cutoff is 35 cycles. For the
18 quantification of RNA, total RNA was isolated using TRIzol reagent (Invitrogen). RNA
19 concentration was measured by Nanodrop 2000 (ThermoFisher). Five hundred nanogram
20 RNA was reverse-transcribed into cDNA followed by real-time PCR with SYBR Green
21 Master Mix and specific primers on LightCycler 480II. The primer sequences used are
22 listed in *SI Appendix* Table S5.

23 **Immunohistochemical (IHC) analysis**

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 Osteosarcoma clinical samples were fixed in 4% Paraformaldehyde (PFA) and
2 decalcified in 10% EDTA/0.2% PFA in PBS in microwave decalcifying apparatus.
3 Complete decalcification was verified by X-ray images. Tumor samples, including
4 Osteosarcoma, Kaposi's sarcoma and Lymphangioma, were impregnated in paraffin and
5 sections were subjected to hematoxylin and eosin (H&E) and immunohistochemical (IHC)
6 staining. For IHC, after removal of endogenous peroxidase with 3% H_2O_2 and rinsing in
7 PBS, sections were incubated with an antibody against LANA (ab4103, Abcam) in 1:100
8 dilution at 4^oC overnight. A goat anti-rabbit HRP secondary antibody (DAB-1031,
9 Maxim) was used, followed by metal enhanced DAB colorimetric detection. Then
10 sections were counterstained with hematoxylin.

11 **RNA sequencing (RNA-seq)**

12 Osteosarcoma tumors and adjacent normal tissues were lysed and total RNA was
13 extracted with TRIzol reagent. DNA contamination was eliminated by DNase I treatment.
14 RNA purity was determined using the Qubit 3.0 Fluorometer (Life Technologies). For
15 each sample, 1-3 μg RNA was used to generate sequencing libraries with NEBNext
16 Ultra™ RNA Library Prep Kit for Illumina (#E7530L, NEB, USA) with poly(A)
17 Magnetic Isolation Module (NEB#7490). Fragmentation was carried out using divalent
18 cations under elevated temperature in NEBNext first strand synthesis reaction buffer.
19 First-strand cDNA was synthesized using random hexamer primer and RNase H. Second
20 strand cDNA was synthesized using DNA polymerase I and RNase H. The library
21 fragments were purified with QiaQuick PCR kits. Index codes were added to attribute
22 sequences to each sample. The libraries were sequenced on an Illumina Hiseq 4000
23 platform and 150 bp paired-end reads were generated.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Data process**

2 Raw data in fastq format were filtered, mapped and analyzed as before.13 Briefly, the
3 short reads were aligned to the KSHV reference genome (version GQ994935.1) and the
4 human reference genome (version hg19/GRCh37). The quality of raw data was viewed
5 by multiQC. The number of clean tags mapped to each gene was counted by FPKM
6 (Fragments per Kilo bases per Million reads). Corrected p-value (q-value) < 0.05 and
7 |log2 (fold change) | > 1 were set as threshold for significantly different expression.

8 **Data analysis**

9 Gene Ontology (GO) analysis was performed with GO tools
10 (http://www.geneontology.org/), and GO terms with p-value < 0.05 were considered as
11 significantly enriched gene sets. Venn diagrams were drawn by Venny 2.0 online
12 (http://bioinfogp.cnb.csic.es/tools/venny/index.html). Unsupervised clustering of samples
13 (X-axis) was performed with the average linkage method and Euclidean distance metric,
14 respectively. Linkage distance was performed by calculating the Pearson correlation
15 coefficient with Normal control and then subtracted by 1. GSEA was performed using
16 GSEA-3.0 software between osteosarcoma tumor and adjacent normal tissue samples
17 with 2000 of "Geneset" permutations type and default values for other parameters. FPKM
18 and RPKM values of RNA expression were used in this analysis. KEGG, Reactome, and
19 hallmark gene sets were used in this analysis
20 (http://software.broadinstitute.org/gsea/msigdb/collections.jsp). All differentially
21 expressed pathways with FDR q-value < 0.1 were kept for subsequent analysis.

22 **Statistical analyses**

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Data were analyzed with SPSS 25.0 program. Fisher's exact test was used to analyze baseline data. Odds Ratio (OR) was calculated to estimate the association between KSHV infection and osteosarcoma occurrence. Woolf's method was applied to calculate a 95% confidence interval. P values <0.05 were considered significant (*$P<0.05$, **$P<0.01$, ***$P<0.001$).

Data availability

RNA-seq data obtained in this study are available in NCBI GEO (Gene Expression Omnibus) database, accession: GSE126209.

Acknowledgments

We thank Yuan Lab members for discussion, constructive suggestions, and participation in blinded serological and quantitative PCR experiments. The research reported in this publication was supported by the Natural Science Foundation of China (81530069).

Declaration of Interests

All authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. T. A. Darmon, W. G. Ward, A. Stewart, Osteosarcoma, chondrosarcoma, and Ewing's sarcoma: National Cancer Data Base Report. *Clin. Orthop. Relat. Res.* **459**, 40–47 (2007).
2. L. Mirabello, R. J. Troisi, A. R. Savage, International osteosarcoma incidence patterns in children and adolescence, middle age, and elderly persons. *Int. J. Cancer* **125**, 229–234 (2009).
3. L. U. Gardner, H. F. Heslington, Osteosarcoma from intravenous beryllium compounds in rabbits. *Fed. Proc.* **5**(1 pt 2)**,** 221 (1946).
4. M. F. Stanton, R. Blackwell, Primary tumors of bone and lung in rats following local deposition of cupric-chelated N-hydroxy-2-acetyl-aminofluorene. *Cancer Res.* **27,** 1000–1006 (1967).
5. P. M. Hatfield, M. D. Schulz, Postirradiation sarcoma. Including 5 cases after X-ray therapy of breast carcinoma. *Radiology* **96,** 593–602 (1970).
6. M. Arlen et al., Radiation-induced sarcoma of bone. *Cancer* **28,** 1087–1099 (1971).
7. M. P. Finkel, B. O. Biskis, P. B. Jinkins, Virus induction of osteosarcoma in mice. *Science* **151,** 698–701 (1966).
8. M. P. Finkel, B. O. Biskis, C. Farrell, Osteosarcomas appearing in Syrian hamsters after treatment with extracts of human osteosarcomas. *Proc. Natl. Acad. Sci. U.S.A.* **60,** 1223–1230 (1968).
9. Y. Chang et al., Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266,** 1865–1869 (1994).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 10 K. Antman, Y. Chang, Kaposi's sarcoma. *N. Engl. J. Med.* **342,** 1027–1038 (2000).

2 11 L. Giffin, B. Damania, KSHV: pathways to tumorigenesis and persistent infection.
3 *Adv. Virus Res.* **88,** 111–159 (2014).

4 12 J. N. Martin, The epidemiology of KSHV and its association with malignant disease.
5 In *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis,* Chapter 54,
6 (Cambridge University Press, Cambridge, 2007).

7 13 Y. Li et al., Evidence for Kaposi's sarcoma originating from Mesenchymal stem cell
8 through KSHV-induced Mesenchymal-to-Endothelial Transition. *Cancer Res.* **78,**
9 230–245 (2018).

10 14 C. H. Parsons, B. Szomju, D. H. Kedes, Susceptibility of human fetal mesenchymal
11 stem cells to Kaposi sarcoma-associated herpesvirus. *Blood* **104,** 2736–2738
12 (2004).

13 15 M. S. Lee et al., Human Mesenchymal Stem Cells of Diverse Origins Support
14 Persistent Infection with Kaposi's Sarcoma-Associated Herpesvirus and Manifest
15 Distinct Angiogenic, Invasive, and Transforming Phenotypes. *MBio* **7,** e02109-15
16 (2016).

17 16 J. Naipauer et al., PDGFRA defines the mesenchymal stem cell Kaposi's sarcoma
18 progenitors by enabling KSHV oncogenesis in an angiogenic environment. *PLoS*
19 *Pathog.* ***15,*** e1008221 (2019).

20 17 A. J. Mutsaers, C. R. Walkley, Cell of origin in osteosarcoma: mesenchymal stem
21 cells or osteoblast committed cells? *Bone* **62,** 56–63 (2014).

22 18 A. Abarrategi et al., Osteosarcoma: Cells-of-Origin, Cancer Stem Cells, and

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 Targeted Therapies. *Stem Cells Int.* **2016,** 3631764 (2016).

2 19 P. Dilnur et al., Classic type of Kaposi's sarcoma and human herpesvirus 8 infection
3 in Xinjiang, China. *Pathol. Int.* **51,** 845–52 (2001).

4 20 T. Zhang et al., Human Herpesvirus 8 Seroprevalence, China. *Emerg. Infect. Dis.*
5 **18,** 150–152 (2012).

6 21 Y. Yang, J. Chen, H. Chu, T. Wang, X. Song, Epidemiological Features Analysis of
7 Uighur, Han and Kazak Patients with Osteosarcoma in Xinjiang Area. *J. Pract.*
8 *Orthop.* **24,** 130–133 (2018).

9 22 B. Fuchs, D. J. Pritchard, Etiology of osteosarcoma. *Clin. Orthop. Relat. Res*. **397,**
10 40–52 (2002).

11 23 Y. Cao et al. High prevalence of early childhood infection by Kaposi's sarcoma-
12 associated herpesvirus in a minority population in China. *Clin. Microbiol. Infect.* **20,**
13 475–481 (2014).

14 24 A. Franchi et al. Expression of transforming growth factor beta isoforms in
15 osteosarcoma variants: association of TGF beta 1 with high-grade osteosarcomas. *J.*
16 *Pathol.* **185,** 284–289 (1998).

17 25 P. Kloen et al., Expression of transforming growth factor-beta (TGF-beta) isoforms
18 in osteosarcomas: TGF-beta3 is related to disease progression. *Cancer* **80**: 2230–
19 2239 (1997).

20 26 S. de Sanjose et al., Geographic Variation in the Prevalence of Kaposi Sarcoma–
21 Associated Herpesvirus and Risk Factors for Transmission. *J. Infect. Dis.* **10,** 1449–
22 1456 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 27 Z. J. Ward, M. Yeh, N. Bhakta, A. L. Frazier, R. Atun, Estimating the total incidence
2 of global childhood cancer: a simulation-based analysis. *Lancet Oncol.* **20,** 483–
3 493 (2019).

4 28 J. Zheng, et al. Prevalence of Kaposi's sarcoma-associated herpesvirus in Uygur and
5 Han populations from the Urumqi and Kashgar regions of Xinjiang, China. *Virol Sin.*
6 **32,** 396-403 (2017).

7 29 D. H. Kedes et al. The seroepidemiology of human herpesvirus 8 (Kaposi's
8 sarcoma-associated herpesvirus): distribution of infection in KS risk groups and
9 evidence for sexual transmission. *Nat. Med.* **2,** 918–924 (1996).

10

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 # Figure Legends

2 **Fig. 1. The presence of antibodies specific to KSHV in sera of osteosarcoma patients.**
3 Serum samples from 21 Uyghur osteosarcoma patients were examined for KSHV
4 seropositivity with KSHV-negative BJAB and KSHV latently infected BCBL-1 cells.
5 Serum from a Kaposi's sarcoma patient and an anti-LANA antibody were included as
6 positive controls. Sera and antibody were diluted in a two-fold serial fashion and reacted
7 with the cells on slides (the dilution for each serum was listed under the patient identity
8 number on the left). No serum was found to react with BJAB cells, but some
9 osteosarcoma sera reacted with BCBL-1 cells showing bright punctate staining in the
10 nuclei, a typical LANA staining reported previously (29). The nuclear punctate staining
11 pattern can also be seen in BCBL-1 cells staining with an anti-LANA antibody, which
12 was illustrated in the last row of the figure.

13 **Fig. 2. KSHV genomic DNA and latent nuclear antigen (LANA) in osteosarcoma**
14 **tumors.** (A) Detection of KSHV genomic DNA in osteosarcoma tumors using a nested
15 PCR approach. Total DNA was isolated from osteosarcoma tumors and adjacent normal
16 tissues. Two hundred nanogram of each sample was subjected to nested PCR with
17 primers specific to ORFs K5, 25, 26, 37, and 73 (LANA). Primers for EBV LMP1 and
18 EBNA1 are included as a reference. (B) Detection of KSHV genomic DNA in
19 osteosarcoma tumors by quantitative real-time PCR with a pair of primers specific to
20 ORF73 (LANA). The absolute quantification of the KSHV DNA genome (copy number)
21 of each clinical specimen is illustrated. A dashed line shows the limit of detection of the
22 quantitative PCR. (C) Osteosarcoma tumors were subjected to immunohistochemical
23 staining with an antibody against KSHV nuclear antigen LANA. An AIDS-KS tumor

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 sample and a lymphangioma sample were included as controls.

2 **Fig. 3. KSHV-positive and -negative osteosarcomas represent distinct subtypes of**
3 **osteosarcoma on gene expression profiles.** (A) KSHV transcriptome of six KSHV-
4 positive osteosarcomas (P1, P5, P9, P11, P20, and P21). RNA-seq reads were aligned to
5 the complete KSHV reference genome (version GQ994935.1) and visualized on a linear
6 scale. The Y-axis represents the number of reads aligned to each nucleotide position of
7 the KSHV genome. KSHV genome positions are indicated at the bottom of each panel.
8 Total reads mapped to the KSHV genome of each clinical sample are listed on the right.
9 (B) The RNA-seq reads of ten osteosarcomas as well as their adjacent normal tissue
10 samples were mapped to the human reference genome (version hg19/GRCh37).
11 Unsupervised clustering of osteosarcoma samples (X-axis) and genes (Y-axis) were
12 performed by the average linkage method. (C) Linkage distance to normal tissue was
13 determined by the Pearson correlation coefficient. (D) The first two principal components
14 of these data were identified and shown in multiple dimensional scaling (MDS) plot.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Baseline Characteristics of Uyghur Osteosarcoma Patients.

Characteristics	Patient KSHV Sero-status (N = 21)			Patient KSHV Genomic DNA (N = 17)		
	Number of Subjects (%)	KSHV Sero-positivity (%)	P Value	Number of Subjects (%)	KSHV DNA-positivity (%)	P Value
Gender						
Male	9 (42.9)	9 (100.0)	0.104	8 (47.1)	6(75.0)	1.00
Female	12 (57.1)	8 (66.7)		9 (52.9)	6 (66.7)	
Age group						
5-30 years	18 (85.7)	15 (83.3)	0.489	14 (82.4)	12 (85.7)	0.024
>30 years	3 (14.3)	2 (66.7)		3 (17.6)	0 (0)	
Diagnosis						
CIMOS [a], Fibroblastic	3 (14.2)	2 (66.6)	0.962	2 (11.8)	1 (50.0)	0.478
CIMOS, Osteoblastic	6 (28.6)	5 (83.3)		4 (23.6)	4 (100.0)	
CIMOS, Chondroblastic	2 (9.5)	2 (100.0)		2 (11.8)	2 (100.0)	
CIMOS, Myxoid	1 (4.8)	1 (100.0)		0 (0)	0 (0)	
CIMOS, Round Cell	1 (4.8)	1 (100.0)		1 (5.9)	1 (100.0)	
CIMOS, Mixed	5 (23.8)	3 (60.0)		5 (29.4)	2 (40.0)	
CIMOS, NOS [b]	1 (4.8)	1 (100.0)		1 (5.9)	1 (100.0)	
WDIOS [c]	2 (9.5)	2 (100.0)		2 (11.8)	1 (50.0)	
Tumor location						
Femur	11 (52.3)	9 (81.8)	0.421	8 (47.0)	4 (50.0)	0.109
Tibia	5 (23.8)	5 (100.0)		5 (29.4)	5 (100.0)	
Humerus	2 (9.5)	2 (100.0)		2 (11.8)	2 (100.0)	
Pelvis	1 (4.8)	0 (0)		1 (5.9)	0 (0)	
Clacivle	1 (4.8)	1 (100.0)		1 (5.9)	1 (100.0)	
Scapula	1 (4.8)	1 (100.0)		0 (0)	0 (0)	
Chemotherapy						
Neo-adjuvant Chemotherapy	15 (71.4)	12 (80.0)	1.00	12 (70.6)	8 (66.7)	1.00
Never	6 (28.6)	5 (83.3)		5 (29.4)	4 (80.0)	
Total	21 (100.0)	17 (81.0)		17 (100.0)	12 (70.6)	

[a] CIMOS: Conventional intramedullary osteosarcoma; [b] NOS: No otherwise specified; [c] WDIOS: Well-differentiated intramedullary osteosarcoma.

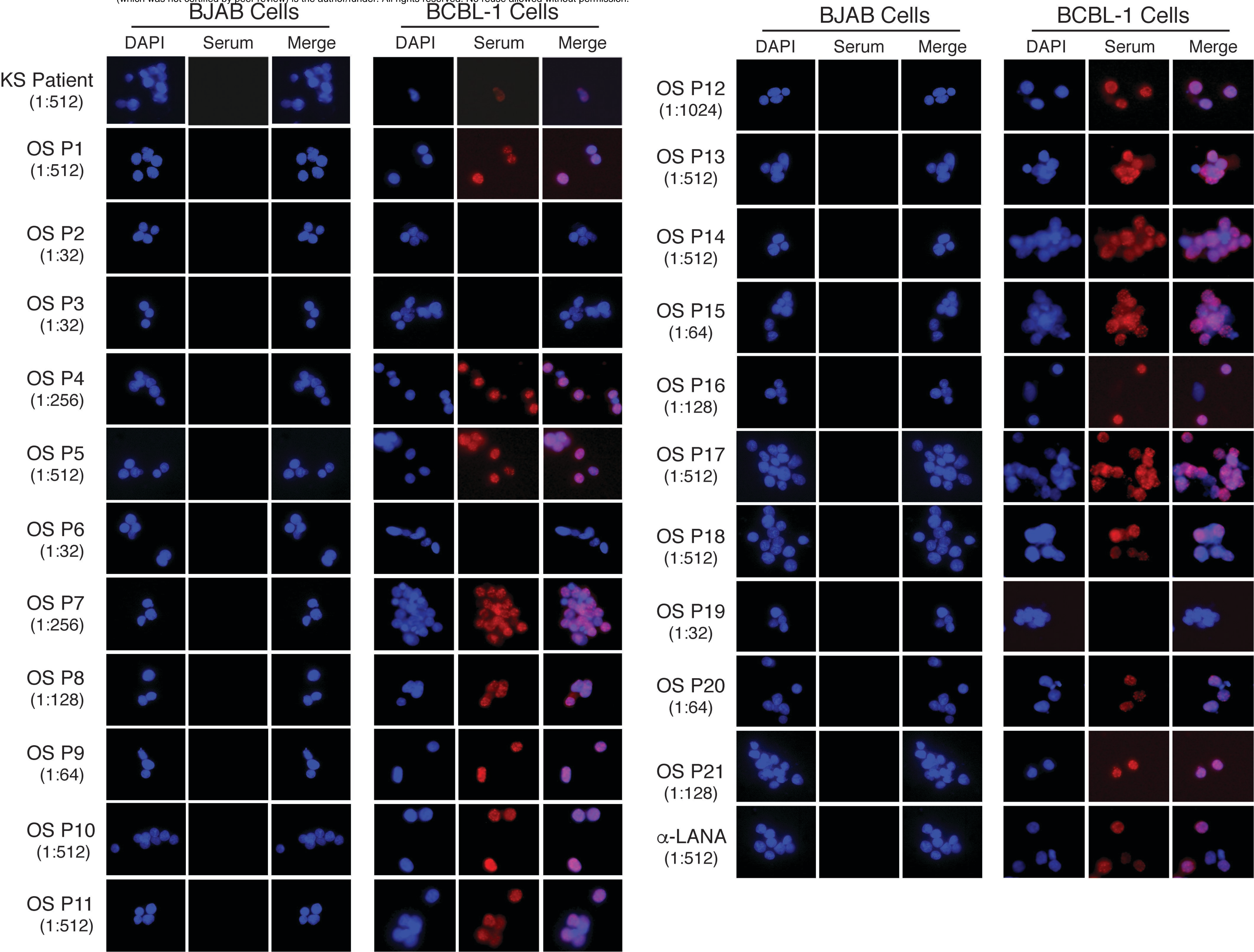
bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

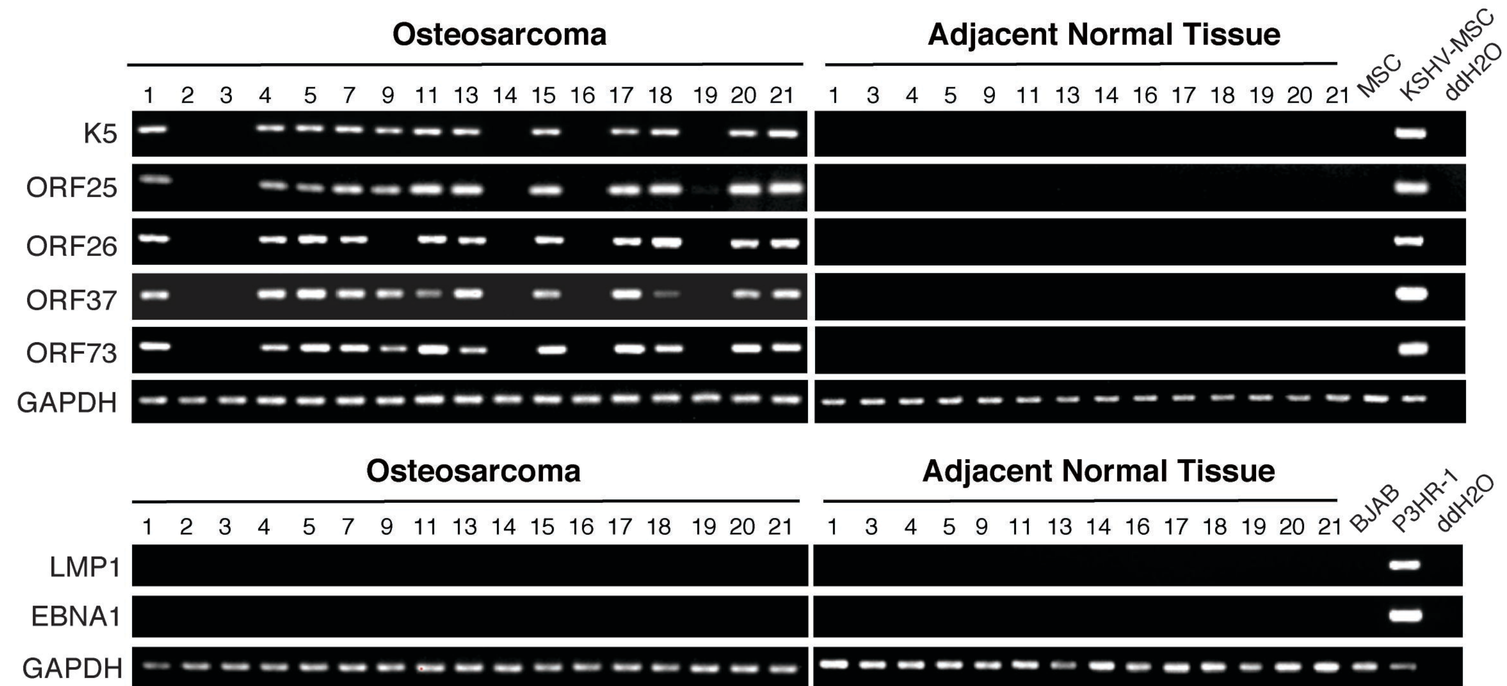
Table 2. Risk of KSHV Infection with Osteosarcoma Occurrence.

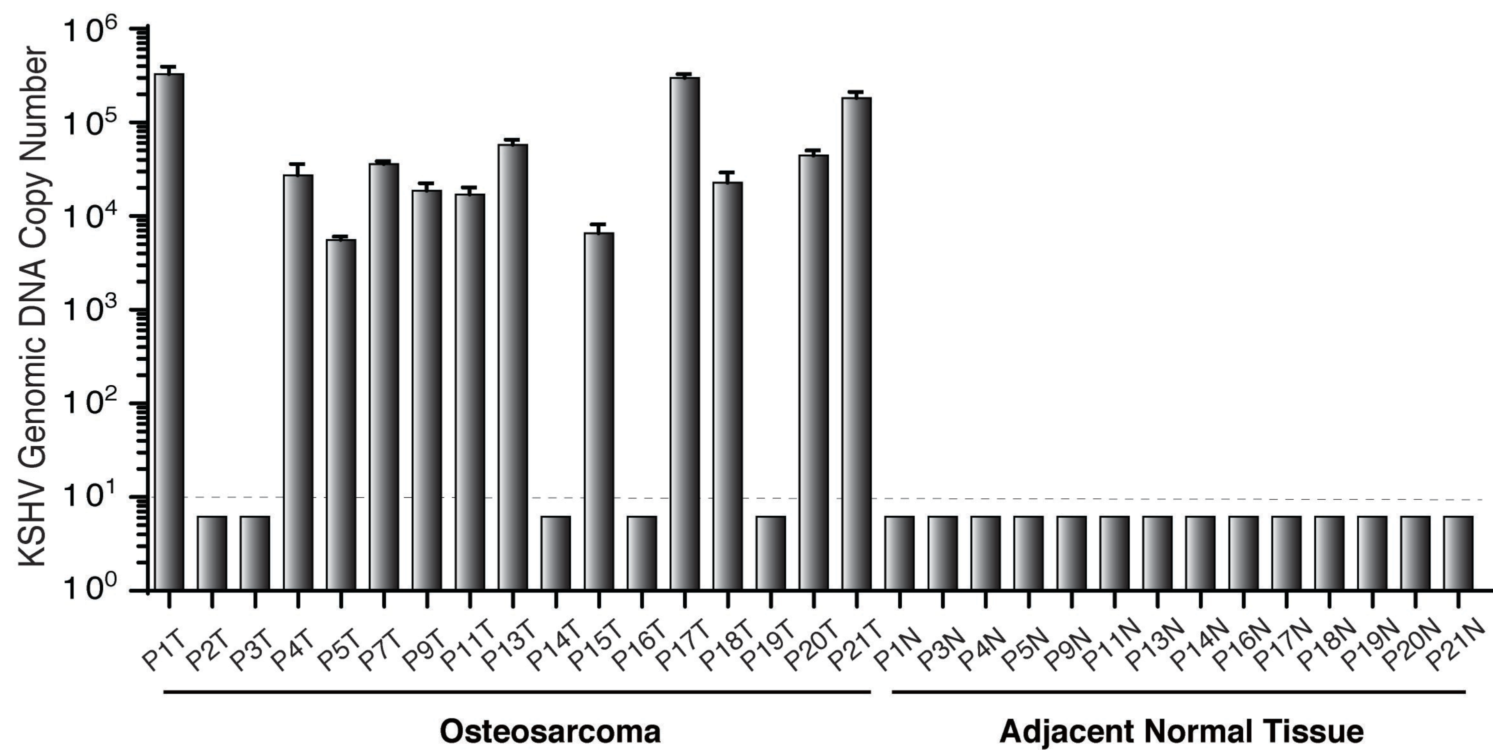
	Osteosarcoma (%) (N = 21)	Control Group (%) [a] (N = 327)	Odds Ratio (95% CI)	P Value
KSHV-positive	17 (81.0)	96 (29.4)	10.23 (4.25-18.89)	< 0.0001
KSHV-negative	4 (19.0)	231 (70.6)		
Total	21 (100.0)	327 (100.0)		

[a] Control group includes 327 Uyghur donors who came to the hospital for annual physical examination.
KSHV prevalence in Uyghur osteosarcoma patients = 81.0%
KSHV prevalence in Uyghur general population = 29.4%

(which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.







C

H&E

KSHV-LANA

200X 400X 200X 400X

AIDS-Kaposi's Sarcoma

Osteosarcoma P1

Osteosarcoma P11

Osteosarcoma P16

Osteosarcoma P20

Osteosarcoma P21

Lymphangioma

A

ORF4 ORF6 ORF11 K2 K4 K7 PAN ORF21 ORF40 ORF41 ORF45 ORF46 ORF50 K8 K8.1 ORF57 K9 ORF58 ORF59 K11 K12 T0.7 ORF73

Reads Aligned to the KSHV Genome (GQ994935.1)

Depth

OS-P1: 223,133

OS-P5: 9,871

OS-P9: 18,035

OS-P11: 1,461

OS-P20: 10,031,401

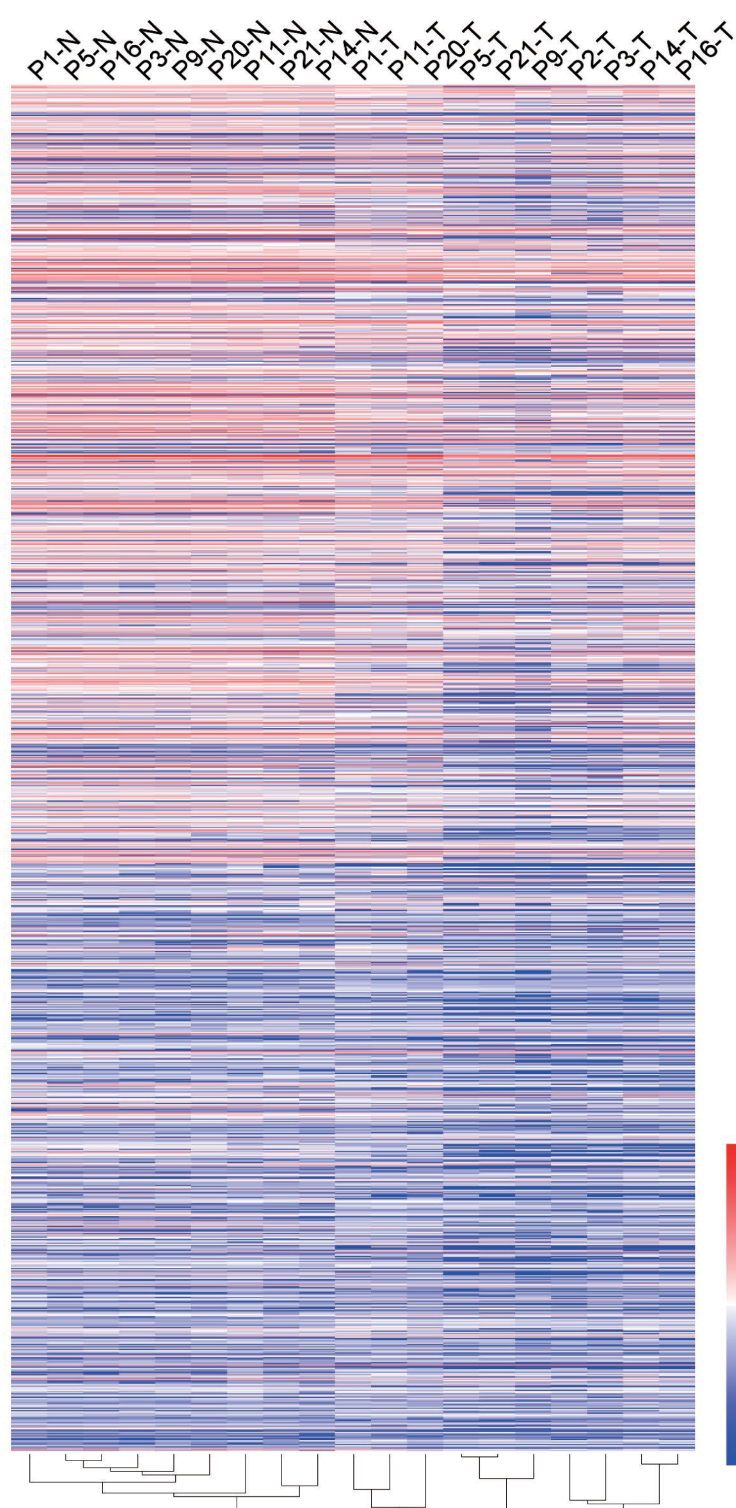
OS-P21: 19,127

0 10K 20K 30K 40K 50K 60K 70K 80K 90K 100K 110K 120K 130K 140K

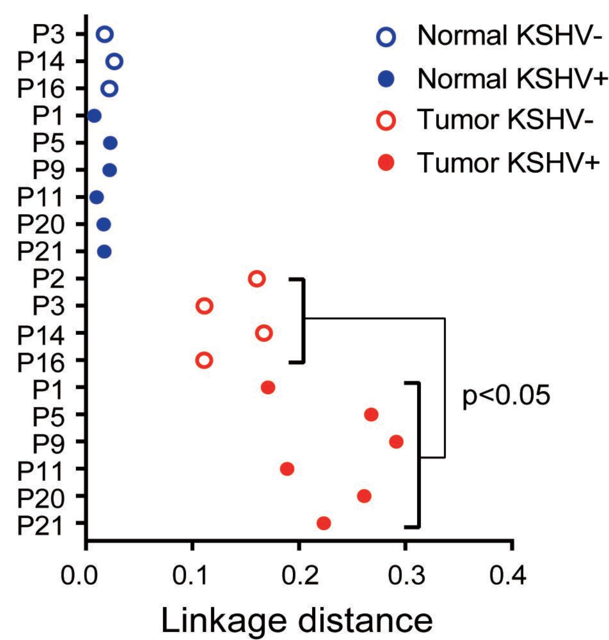
Position at KSHV Genome (0 to 137,969 bp)

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

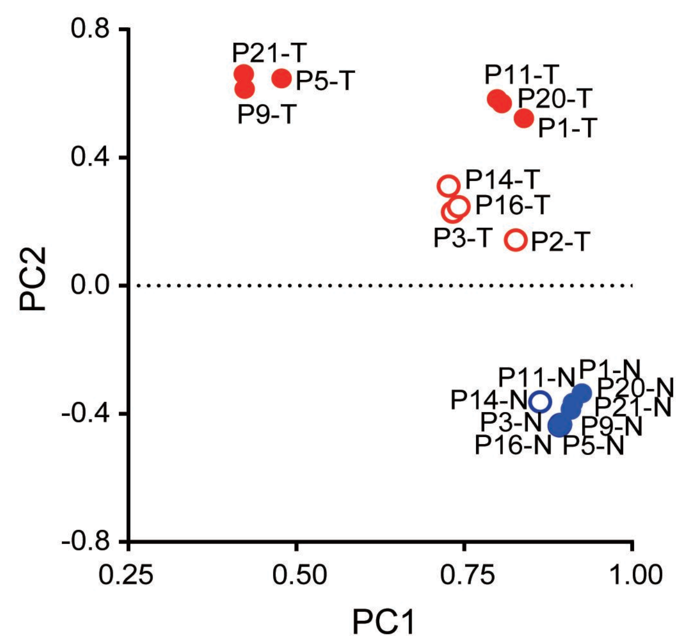
B



C



D



bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Information for

Kaposi's Sarcoma Herpesvirus Is Associated with Osteosarcoma in Xinjiang Uyghur Population

Qian Chen, Jiangtao Chen, Yuqing Li, Dawei Liu, Yan Zeng, Zheng Tian, Akbar Yunus, Yong Yang, Jie Lu, Xinghua Song* and Yan Yuan*

*Corresponding author: Yan Yuan
Email: yuan2@upenn.edu

This PDF file includes:

Table S1. Demographic Characterization and KSHV Positivity of 21 Uyghur Osteosarcoma Patients.

Table S2. Baseline Characteristics of 327 Control Participants.

Table S3. Osteosarcoma RNA-seq Raw Reads and Alignment to the KSHV and Human Genomes.

Table S4. Demographic Characterization and KSHV Positivity of 9 Han Osteosarcoma Patients.

Table S5. Oligonucleotides Used in the Study.

Figure S1. KSHV Transcriptome of Six KSHV-Positive Osteosarcoma in Log Scale.

Figure S2. Gene Ontology (GO) Analysis of KSHV-positive and KSHV-negative Osteosarcomas.

Figure S3. Comparison of Han Osteosarcoma with Six Uyghur Osteosarcomas on KSHV Transcription Profile.

Figure S4. KSHV-positive and -negative Osteosarcomas, Regardless of Han and Uyghur Patients, represent distinct subtypes of osteosarcoma on Gene Expression Profile.

Supporting Information – Table S1. Demographic Characterization and KSHV Positivity of 21 Uyghur Osteosarcoma Patients.

Patient Code	Sex	Age	Diagnosis	Tumor Location	Ethnicity	Serology ELISA			IFA	KSHV Genome in Tumor
						K8.1	ORF65	LANA	KSHV	
P1	M	28	CIMOS, NOS	Left humerus	Uyghur	-	+	+	+	+
P2	F	15	CIMOS, Mixed	Left pelvis	Uyghur	-	-	-	-	-
P3	F	46	CIMOS, Fibroblastic	Left femur	Uyghur	-	-	-	-	-
P4	F	8	CIMOS, Osteoblastic	Right tibia	Uyghur	+	-	+	+	+
P5	M	22	CIMOS, Chondroblastic	Right tibia	Uyghur	+	+	+	+	+
P6	F	17	CIMOS, Osteoblastic	Right scapula	Uyghur	-	-	-	-	N.S.
P7	F	30	CIMOS, Mixed	Right clacivle	Uyghur	-	+	+	+	+
P8	F	27	CIMOS, Osteoblastic	Left femur	Uyghur	-	-	+	+	N.S.
P9	M	10	CIMOS, Osteoblastic	Left femur	Uyghur	-	-	+	+	+
P10	M	20	CIMOS, Fibroblastic	Right femur	Uyghur	-	+	+	+	N.S.
P11	F	10	CIMOS, Chondroblastic	Left tibia	Uyghur	-	-	+	+	+
P12	F	13	CIMOS, Myxoid	Left femur	Uyghur	+	+	+	+	N.S.

P13	F	11	CIMOS, Round cell	Left femur	Uyghur	+	+	+	+	+
P14	M	53	WDIOS	Left femur	Uyghur	+	+	+	+	-
P15	F	28	WDIOS	Left femur	Uyghur	-	-	+	+	+
P16	M	39	CIMOS, Mixed	Left femur	Uyghur	-	+	+	+	-
P17	F	9	CIMOS, Osteoblastic	Right tibia	Uyghur	+	+	+	+	+
P18	M	9	CIMOS, Fibroblastic	Left humerus	Uyghur	+	+	+	+	+
P19	F	14	CIMOS, Mixed	Right femur	Uyghur	-	-	-	-	-
P20	M	15	CIMOS, Osteoblastic	Left femur	Uyghur	-	-	+	+	+
P21	M	15	CIMOS, Mixed	Left tibia	Uyghur	+	-	+	+	+

CIMOS: Conventional intramedullary osteosarcoma; WDIOS: Well-differentiated intramedullary osteosarcoma; NOS: No otherwise specified; +: Positive; - : Negative; F: Female; M : Male; N.S. : No Sample.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supporting Information – Table S2. Baseline Characteristics of 327 Control Participants (General Uyghur Population)

Characteristics	Number of subjects (%)	KSHV sero-positivity (%)	P Value
Gender			
Male	142 (43.4)	46 (32.4)	0.291
Female	185 (56.6)	50 (27.0)	
Age group			
5-30 years	114 (34.9)	28 (24.6)	0.164
>30 years	213 (65.1)	68 (31.9)	
Smoking tobacco			
Never	229 (70.0)	65 (28.4)	0.555
Ever	98 (30.0)	31 (31.6)	
Drinking Alcohol			
No	189 (57.8)	53 (28.0)	0.541
Yes	138 (42.2)	43 (31.2)	
Total	327 (100.0)	96 (29.4)	

Age (Years) Mean ± SD): 38.37 ± 13.40; Minimum – maximum: 9 – 86.

Supporting Information – Table S3. Osteosarcoma RNA-seq Raw Reads and Alignment to the KSHV and Human Genomes.

Sample	Patient ID	Total Reads	Aligned to KSHV Transcriptome	Aligned to Human Transcriptome
Osteosarcoma	P1	63,953,856	223,133	60,491,393
	P2	69,061,060	1	65,452,961
	P3	81,498,617	8	79,602,603
	P5	72,040,573	9,871	68,864,981
	P9	70,845,288	18,035	66,470,038
	P11	60,649,859	1,461	57,958,356
	P14	75,958,617	3	74,449,359
	P16	75,675,293	5	74,067,075
	P20	97,844,318	10,031,401	83,058,763
	P21	98,122,848	19,127	94,631,752
	HP2	43,313,298	1,646,210	39,515,009
	HP9	45,779,360	234	42,938,247
Adjacent Normal Tissue	P1	65,287,890	2,717	62,538,711
	P3	77,402,745	3	75,688,331
	P5	79,508,710	5	77,772,027
	P9	72,239,269	7	70,439,146
	P11	59,988,497	40	57,712,755
	P14	77,691,776	66	76,247,965
	P16	95,230,519	2	93,522,293
	P20	86,440,986	28	84,727,078
	P21	96,247,969	112	94,370,220
	HP2	47,348,598	2	45,446,160
	HP9	37,135,030	6	35,434,755

Supporting Information – Table S4. Demographic Characterization and KSHV Positivity of 9 Han Osteosarcoma Patients.

Patient Code	Sex	Age	Diagnosis	Tumor Location	Ethnicity	Serology ELISA				KSHV Genome in Tumor
						K8.1	ORF65	LANA	KSHV	
HP1	M	52	Conventional	Left hip joint	Han	-	-	-	-	-
HP2	F	5	Conventional	Right humerus	Han	+	+	+	+	+
HP3	F	37	Telangiectatic	Right pubis	Han	-	-	-	-	-
HP4	M	33	Conventional	Left ilium	Han	-	-	-	-	N.S.
HP5	M	22	Conventional	Right femur	Han	-	-	-	-	N.S.
HP6	F	34	Conventional	Right fibula	Han	-	-	-	-	-
HP7	M	39	Conventional	Left femur	Han	-	-	-	-	N.S.
HP8	M	20	Conventional	Left ilium	Han	-	-	-	-	-
HP9	M	45	Conventional	Left femur	Han	-	-	-	-	-

+: Positive; - : Negative; F: Female; M : Male; N.S. : No Sample.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

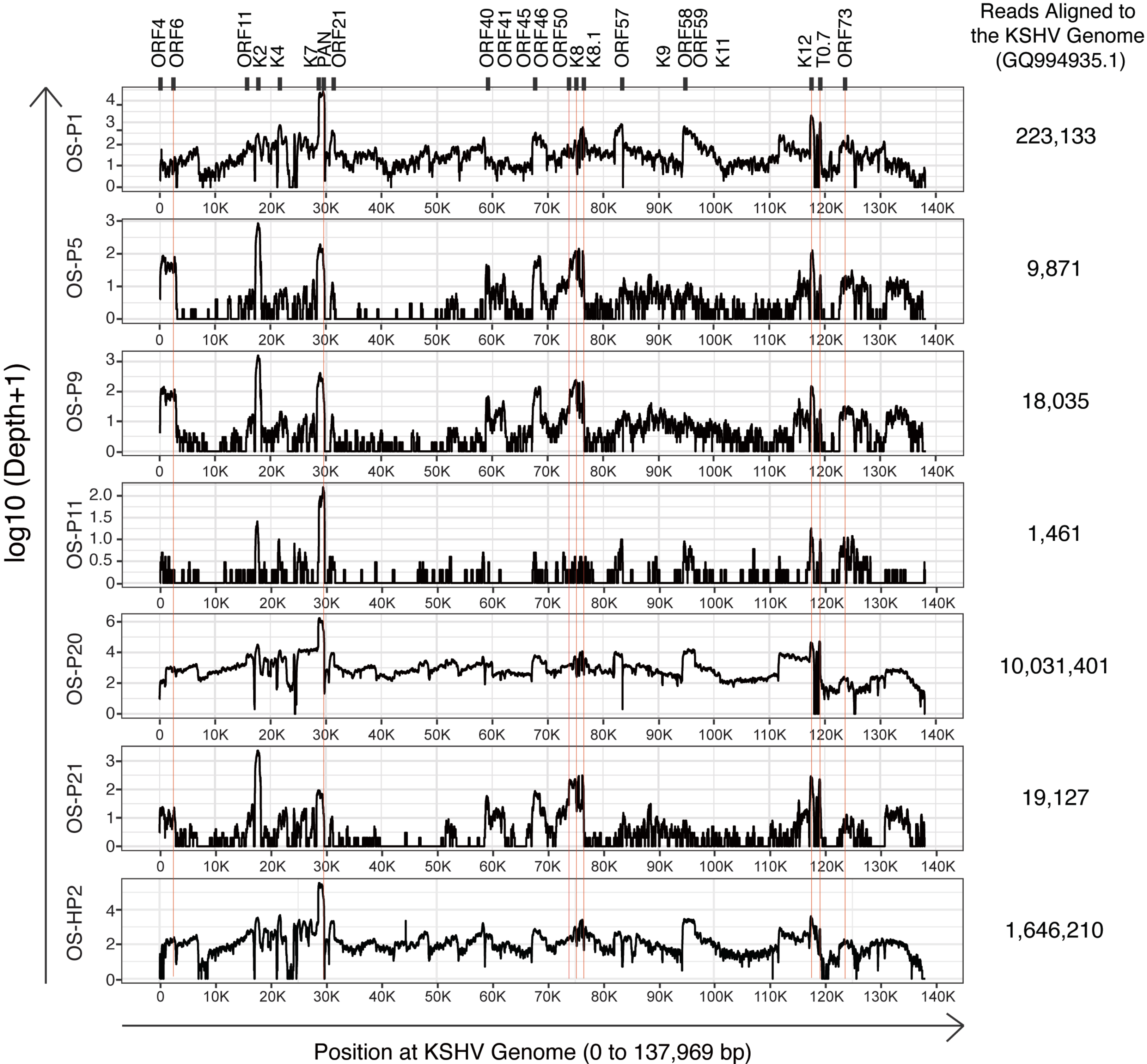
Supporting Information – Table S5. Oligonucleotides used in the study.

Target gene	Species	Primer Sequence (5'→3')	Usage
LANA	KSHV	F: ACTATGGAAGATTGTAGGTT R: ACCAGTCGCCCATAACTTAT	qPCR
LANA#1	KSHV	F: TTATGTCATTTCCTGTGGAGAGT R: GTGGATTACCCTGTTGTTAG	Cloning
LANA#2	KSHV	F: AACAACAGGGTAATCCACTT R: GGAAACAAAACGTTGAGCAT	Cloning
LANA#3	KSHV	F: CGTTTTGTTTCCATCGCCC R: ATGGCGCCCCCGGGAAT	Cloning
ORF65	KSHV	F: CTATTTCTTTTTGCCAGAGGG R: ATGTCCAACTTTAAGGTGAGAG	Cloning
K8.1	KSHV	F: ATGAGTTCCACACAGATTCGCA R: CTGAAAAGGCTGATATTAAGGCAT	Cloning
K5	KSHV	F: TCAACCGTTGTTTTTTGGATGAT R: ATGGCGTCCAAGGACGTAGAAGA	Nested PCR(Outer)
K5	KSHV	F: CATCTCCGGCCACAGGTTAA R: TACCGGAGAGCTGGATGT	Nested PCR(Inner)
ORF25	KSHV	F: GAGAAAGCTATTCTACTATGTTTT R: CAAAACCAGGGTGCATGCGGTGT	Nested PCR(Outer)
ORF25	KSHV	F: CCTGACGCTGACTTACAAC R: TGCCTGCAGAATGTCCT	Nested PCR(Inner)
ORF26	KSHV	F: AGTGGGACAGCAACACCCAGCTA R: TTAGCGTGGGGAATACCAACAGG	Nested PCR(Outer)
ORF26	KSHV	F: GTGCTCGAATCCAACGGATT R: TCCGTGTTGTCTACGTCCAGA	Nested PCR(Inner)
ORF37	KSHV	F: CCAGTGAATCTGATTACTTGGTG R: CTACGGGCTGTGAGGGACGTTTG	Nested PCR(Outer)
ORF37	KSHV	F: TCGGTGGCGATGCTTTAGAC R: TGAAGCAGACGATGCTTTGC	Nested PCR(Inner)
LANA	KSHV	F: TTATGTCATTTCCTGTGGAGAGT R: AGGAGATAATACACCAGACGATG	Nested PCR(Outer)
LANA	KSHV	F: CGCGAATACCGCTATGTACTCA R: GGAACGCGCCTCATACGA	Nested PCR(Inner)

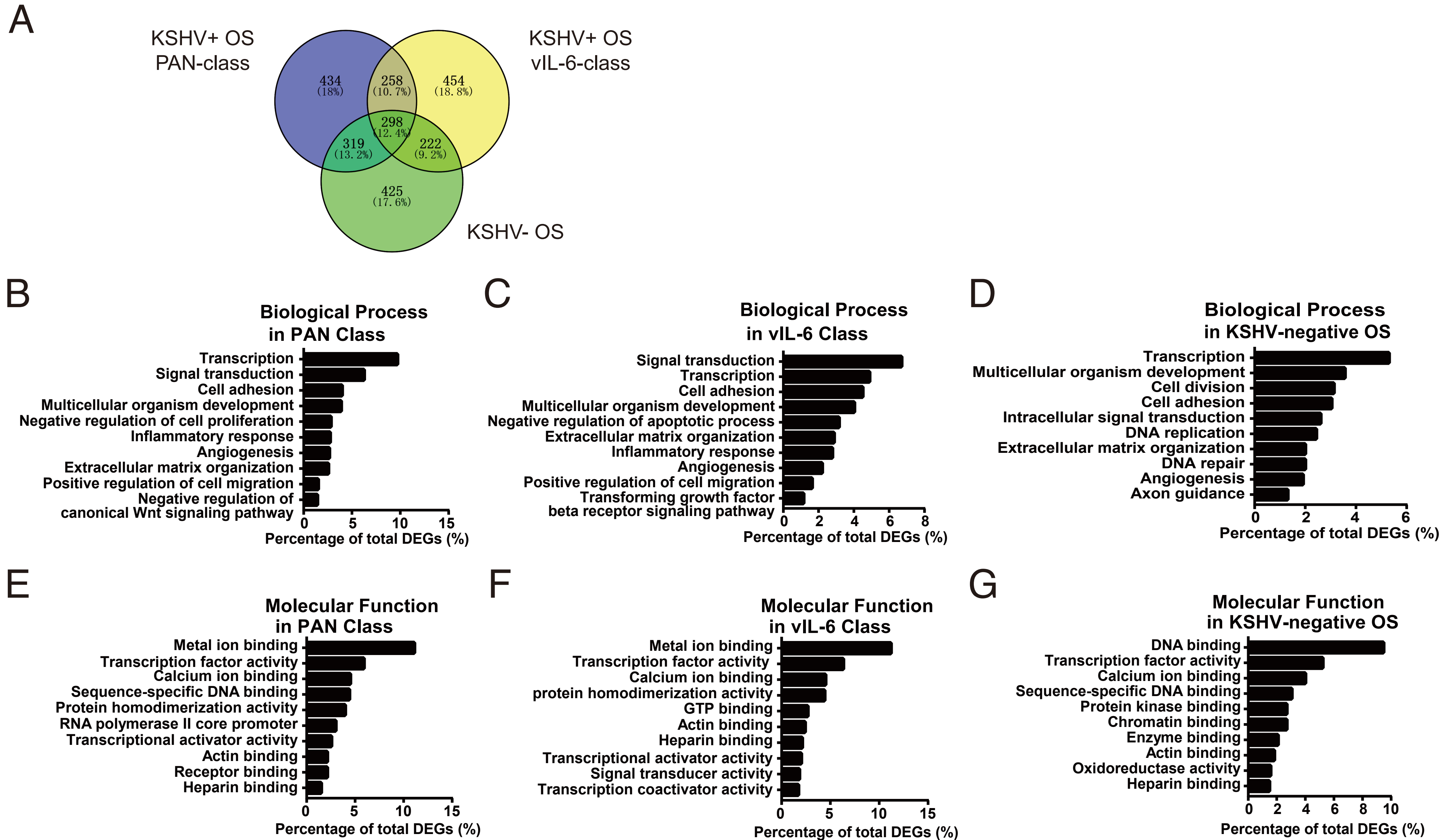
bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

GAPDH	Human	F: CATCTTCCAGGAGCGAGATCCCT R: GGCCATGAGGTCCACCACCCTGT	Nested PCR(Outer)
GAPDH	Human	F: ACATCATCCCTGCCTCTAC R: TCAAAGGTGGAGGAGTGG	Nested PCR(Inner)
LMP1	EBV	F: GGTGTTCCTGCCTCTGCTGGCAT R: GGTGATCCACACCTTCCTACGCT	Nested PCR(Outer)
LMP1	EBV	F: ATGTGGTCCCCCGCTGGTATGC R: CATTACCATGTCATAGGCTTGCC	Nested PCR(Inner)
EBNA1	EBV	F: CGGGAGCGATAGAGCAGGGCCCC R: CTCACCCTCATCTCCATCACCTC	Nested PCR(Outer)
EBNA1	EBV	F: CATTGAGTCGTCTCCCCTTTGGAAT R: TCATAACAAGGTCCTTAATCGCATC	Nested PCR(Inner)

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

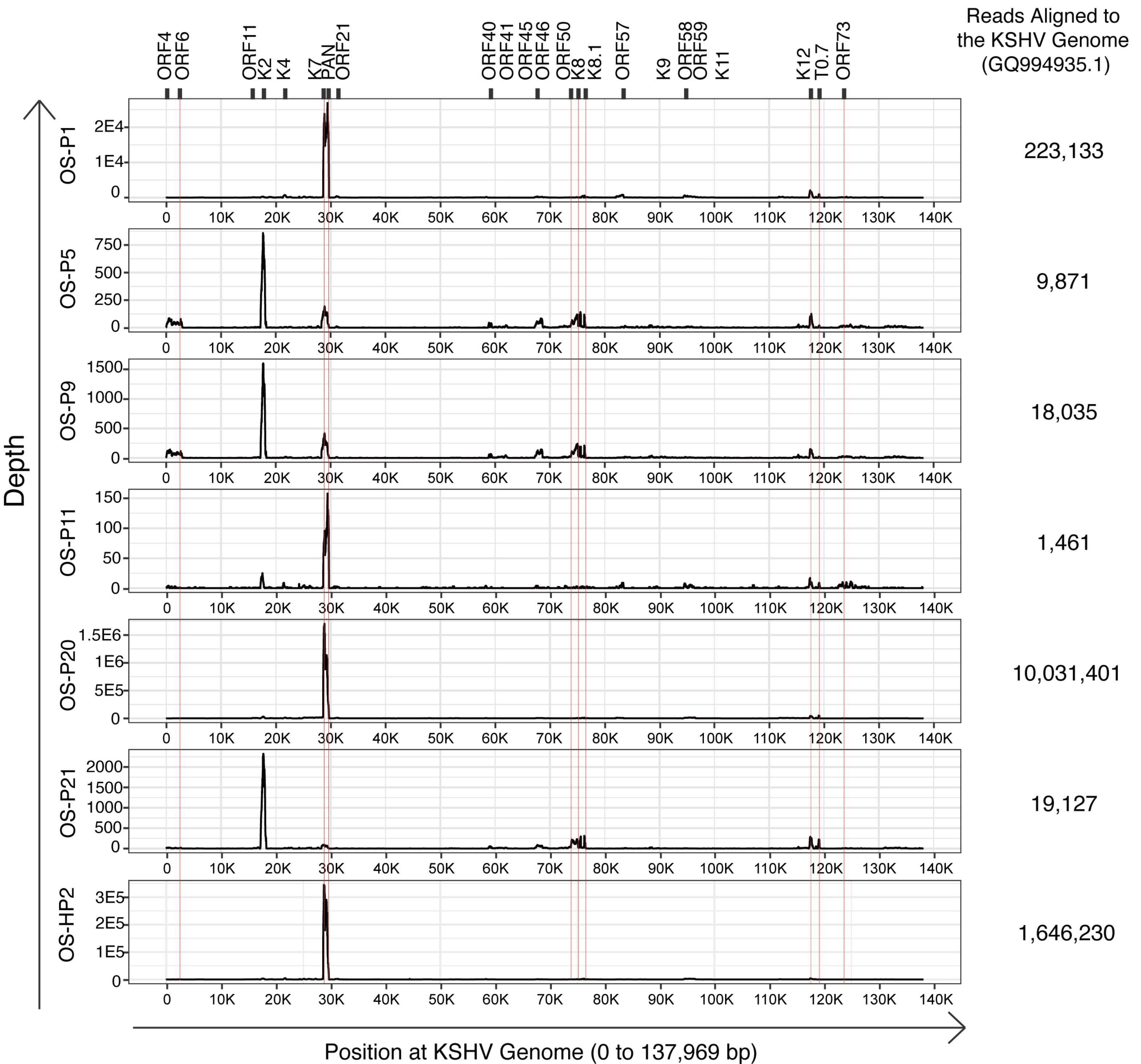


Supporting Information – Fig. S1. KSHV transcriptome of six KSHV-positive osteosarcomas on logarithmic scale. RNA-seq reads were aligned to the complete KSHV reference genome (version GQ994935.1). The Y-axis represents the number of reads aligned to each nucleotide position of the KSHV genome in a log-10 scale. Total reads mapped to the KSHV genome of each clinical sample are listed on the right.



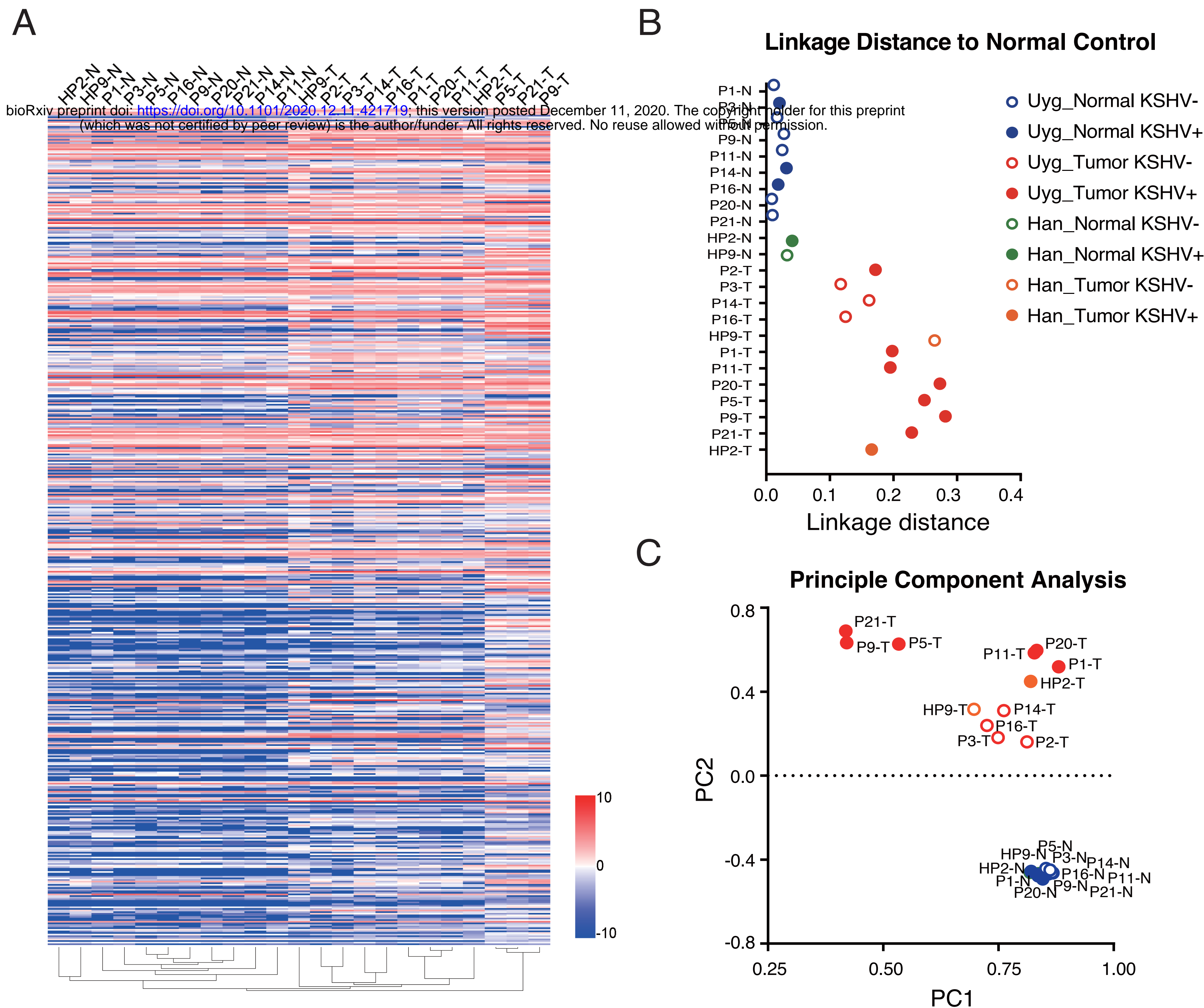
Supporting Information – Fig. S2. Gene ontology (GO) analysis of KSHV-positive (PAN and vIL-6 classes) and -negative osteosarcomas. (A) DEGs of KSHV-positive and -negative osteosarcomas vs. their adjacent normal tissues were analyzed and compared by Venny diagram. Numbers and percentage of DEGs unique for each of the three pairwise comparisons as well as common to two or three classes of osteosarcoma are listed. (B – D) The top 10 significantly enriched terms of biological process in DEGs of KSHV-positive osteosarcomas (PAN class, B), KSHV-positive osteosarcoma-s (vIL-6 class, C), and KSHV-negative osteosarcomas (D), vs. their adjacent normal tissues. (E – G) The top 10 significantly enriched terms of molecular function in DEGs of KSHV-positive osteosarcomas (PAN class, E), KSHV-positive osteosarcomas (vIL-6 class, F) and KSHV-negative osteosarcomas (G), vs. their adjacent normal tissues.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supporting Information – Fig. S3. Comparison of Han osteosarcoma (OS-HP2) with six Uyghur osteosarcomas (OS-P1, P5, P9, P11, P20 and P21) on KSHV transcription profile. RNA-seq reads were aligned to the complete KSHV reference genome (version GQ994935.1) and visualized on a linear scale. The Y-axis represents the number of reads aligned to each nucleotide position of the KSHV genome. KSHV genome positions are indicated at the bottom of each panel. Total reads mapped to the KSHV genome of each clinical sample are listed on the right.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.11.421719; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supporting Information – Fig. S4. KSHV-positive and -negative osteosarcomas, regardless of Han and Uyghur patients, represent distinct subtypes of osteosarcoma on gene expression profile. (A) The RNA-seq reads of two Han (HP2 and HP9) and ten Uyghur osteosarcomas as well as their adjacent normal tissue samples were mapped to the human reference genome (version hg19/GRCh37). Unsupervised clustering of osteosarcoma samples (X-axis) and genes (Y-axis) were performed by the average linkage method. (B) Linkage distance to normal tissue was determined by the Pearson correlation coefficient. (C) The first two principal components of these data were identified and shown in multiple dimensional scaling (MDS) plot.