



OPEN

## Extracellular vesicle-associated repetitive element DNAs as candidate osteosarcoma biomarkers

Linda Cambier<sup>1</sup>, Kevin Stachelek<sup>1,2</sup>, Martin Triska<sup>1</sup>, Rima Jubran<sup>3</sup>, Manyu Huang<sup>4</sup>, Wuyin Li<sup>4</sup>, Jianying Zhang<sup>5</sup>, Jitian Li<sup>4,5</sup>✉ & David Cobrinik<sup>1,2,6,7</sup>✉

Osteosarcoma (OS) is the most common malignant bone tumor in children and young adults. Despite that high-risk factors have been identified, no test for early detection is available. This study aimed to identify circulating nucleic acid sequences associated with serum extracellular vesicle (EV) preparations at the time of OS diagnosis, as a step towards an OS early detection assay. Sequencing of small nucleic acids extracted from serum EV preparations revealed increased representation of diverse repetitive element sequences in OS patient versus control sera. Analysis of a validation cohort using qPCR of PEG-precipitated EV preparations revealed the over-representation of *HSATI*, *HSATII*, *LINE1-P1*, and *Charlie 3* at the DNA but not RNA level, with receiver operating characteristic (ROC) area under the curve (AUC)  $\geq 0.90$ . *HSATI* and *HSATII* DNAs co-purified with EVs prepared by precipitation and size exclusion chromatography but not by exosome immunocapture, indicative of packaging in a non-exosomal complex. The consistent over-representation of EV-associated repetitive element DNA sequences suggests their potential utility as biomarkers for OS and perhaps other cancers.

Osteosarcoma (OS) is the most common malignant bone tumor of children, adolescents, and young adults, representing approximately 1% of newly diagnosed cancers in adults, and 3–5% in children<sup>1,2</sup>. With current treatment regimens, patients with non-metastatic OS have 5-year survival rates above 65% whereas the ~25% of patients presenting with metastases have a 5-year survival of less than 20%<sup>3,4</sup>. As such, early detection of OS prior to metastasis could significantly improve outcomes.

Early detection is especially needed in individuals who are predisposed to OS either genetically or through iatrogenic exposures. OS occurs at increased rates in several monogenic hereditary cancer syndromes such as retinoblastoma (*RBI*)<sup>5</sup>, Li-Fraumeni syndrome (*TP53*), Bloom syndrome (*RECQL2*), Werner syndrome (*RECQL3*), and Rothmund-Thomson syndrome (*RECQL4*). OS also occurs with increased frequency in children exposed to radiation or alkylating agents, in Diamond-Blackfan anaemia patients and in adults with bone disorders such as Paget's Disease. Combined genetic predisposition and exposure to DNA damaging agents confers particularly high risk; for example, relative risk for children with hereditary retinoblastoma increased from ~69 without such treatments to ~302 for radiotherapy and ~539 for radiotherapy plus chemotherapy in the largest treatment-stratified analysis<sup>5</sup>.

The need for OS biomarkers is reflected in a large yet inconclusive literature. Early studies focusing on bone markers such as alkaline phosphatase showed highly variable increases in OS patients<sup>6</sup>. Later proteomic studies revealed two as-yet uncharacterized OS-associated proteins<sup>7</sup> whereas studies of miRNAs showed variable

<sup>1</sup>The Vision Center and The Saban Research Institute, Children's Hospital Los Angeles, 4650 Sunset Blvd, MS163, Los Angeles, CA 90027, USA. <sup>2</sup>Cancer Biology and Genomics Program, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA. <sup>3</sup>Division of Hematology/Oncology, Children's Hospital Los Angeles, Los Angeles, CA 90027, USA. <sup>4</sup>Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital)/Henan Provincial Orthopedic Institute, 82 Qiming Road, Luoyang 471002, Henan, China. <sup>5</sup>Department of Biological Science, The University of Texas at El Paso, El Paso, TX 79968, USA. <sup>6</sup>Department of Ophthalmology and Roski Eye Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA. <sup>7</sup>Department of Biochemistry and Molecular Medicine and Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA. ✉email: jitianlee@hotmail.com; dcobrinik@chla.usc.edu

results<sup>8–10</sup>. A recent study identified 56 miRs that were upregulated in pre-treatment OS patient plasma<sup>11</sup>; however, among the top candidates (*miR-21*, *miR-221*, and *miR-106a*), levels increased by only ~2.4–8-fold and sensitivity was at best ~85%. An alternative approach is to detect aneuploidy via cell-free DNA (cfDNA) whole genome sequencing, yet at present this has limited sensitivity<sup>12</sup> due to the dilution of tumor with non-tumor cfDNA. Currently no biomarkers have been shown to reliably detect naïve pre-symptomatic OS in predisposed individuals<sup>13</sup>.

In this study, we aimed to identify circulating biomarkers that distinguish OS patients from healthy controls, as a step towards a liquid biopsy for early OS detection. Liquid biopsies may detect circulating tumor components including cfDNA, tumor cells, and extracellular vesicles (EVs)<sup>14–19</sup>, a category that includes exosomes, shedding vesicles, microparticles, retroviral-like particles, ectosomes, microvesicles, oncosomes, and apoptotic bodies<sup>20,21</sup>. EVs are released by most if not all cells<sup>22</sup> and carry components of their cell of origin such as proteins, lipids, metabolites, and various types of RNA<sup>23,24</sup>. Among the different types of EVs, exosomes and oncosomes are more highly produced by cancer cells than by normal cells, are often present at increased levels at cancer diagnosis, may further increase during tumor progression<sup>15</sup>, and carry cargo that reflects metastatic progression and treatment response<sup>25,26</sup>. Moreover, EV preparations may contain exosomal as well as non-exosomal tumor components. Thus, we aimed to identify OS biomarkers in serum-derived EV preparations.

To identify EV-associated OS biomarkers we compared the abundance of nucleic acid sequences in OS patient versus control serum EV preparations. Specifically, we sequenced small nucleic acids extracted from EV preparations and examined differential representation of unique as well as repetitive element sequences which are often produced and may be released by cancer cells<sup>27</sup>, including by OS cells<sup>28</sup>. We then evaluated whether the same sequences were differentially represented in different patient cohorts and by different EV isolation and analytic methods. Through these approaches we identified circulating EV-associated repetitive element DNA sequences that were more abundant in OS sera compared to healthy sera in two patient cohorts.

## Results

**Over-representation of repetitive element sequences in OS EV preparations.** To identify OS biomarkers, we compared nucleic acid sequences associated with EV preparations from sera of OS patients and healthy controls. Initial analyses were performed on a discovery cohort of treatment-naïve OS patients from Children's Hospital Los Angeles (CHLA) and Henan Luoyang Orthopedic Hospital (HLOH), comprised of males and females between 5 and 29 years old and presenting with different OS types. Control cohorts were comprised of healthy siblings of hereditary retinoblastoma patients who had not developed retinoblastoma (hereditary retinoblastoma controls; HRCs) and unrelated approximately age-matched healthy individuals (healthy controls; HC) (Table 1a). EV preparations were made with the commercial ExoQuick kit based on polyethylene glycol (PEG) precipitation, with recognition that EV as well as non-EV components are isolated<sup>29,30</sup>. Nanoparticle tracking analysis of each sample revealed similar size distributions of control and OS EVs between 50 and 150 nm, which is characteristic of exosomes (Fig. 1a,b). EV concentrations were not significantly higher in sera from OS patients compared to controls (Fig. 1c). Likewise, EV concentrations were similar in OS patient serum from USA (CHLA) and China (HLOH) and for different patient ages, genders, and OS types (Fig. 1c–e).

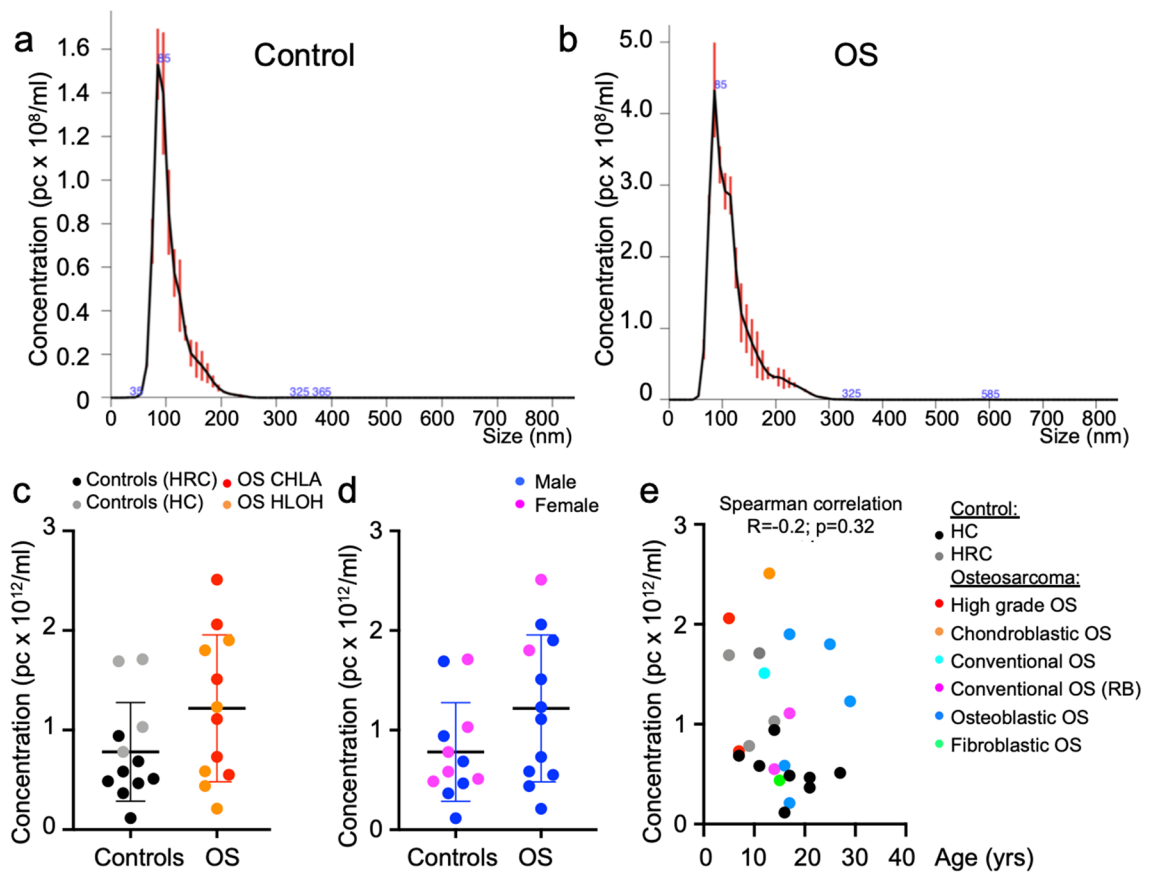
To detect differentially represented EV-associated RNA and DNA sequences, nucleic acids were extracted from OS and control EV preparations using SeraMir small RNA enrichment kit (SBI) without DNase treatment and a sequencing library was built by addition of a 5'-RNA adapter and a 3'-DNA adapter followed by PCR amplification and sequencing. Comparison of uniquely mapped sequences using DESeq2<sup>31</sup> identified 107 significantly over-represented genes and 587 significantly under-represented genes ( $\geq$  twofold change,  $p_{\text{adj}} < 0.05$ ) in OS samples (Supplementary Fig. 1). However, the over-represented sequences had poor OS sensitivity and specificity (not shown). As a far greater proportion of genes were under-represented in OS samples, we considered whether our analysis of uniquely mapped sequences ignored potentially relevant over-represented non-uniquely mapped sequences. Thus, we evaluated the differential representation of the major repetitive element categories after aligning reads to RepeatMasker<sup>32</sup>. This indicated that OS serum EV preparations had greater representation of almost all repetitive element categories including the most abundant LINE1, LTR/ERV,  $\alpha$ -satellite and SINE/Alu categories but not the low complexity ribosomal RNA and scRNA sequences (Supplementary Fig. 2A, B). The most highly represented sequence was the LINE1 family member *L1P1* (Supplementary Fig. 2C).

While these analyses revealed consistent over-representation of repetitive element sequences in OS serum EV preparations, the identities of the most over-represented elements were uncertain since programs that are not specifically designed for repetitive element detection may erroneously map repetitive element reads<sup>33</sup>. To more accurately define the differential repetitive element representation, sequences were evaluated with TETranscripts, which maps repetitive element sequences more accurately and quantitatively than non-dedicated programs<sup>34</sup>. Using default settings with reads mapped to the GRCh38 genome, TETranscripts confirmed that OS samples had far more significantly under-represented than over-represented sequences in comparison to control samples (Fig. 2a) and showed that most of the under-represented sequences were single-copy genes (Fig. 2b). In contrast, a vast proportion of repetitive element sequences were over-represented in OS EV preparations (Fig. 2c) albeit with only 19 significantly over-represented versus four significantly under-represented (Fig. 2d). Among significantly over-represented repeat elements, Human Satellite I (*HSAT1*) had the highest fold change ( $\log_2$  (6.14),  $p_{\text{adj}} = 0.007$ ) (Fig. 2d and Table 2, Analysis 1).

Because GRCh38 contains numerous alternative assemblies that are enriched for repetitive elements that might siphon repetitive element reads, adds synthetic centromeric repeat sequences, and hard-masks certain centromeric and genomic repeat arrays<sup>35</sup>, we considered whether these features might affect the ability to detect differential representation of unique or repetitive element sequences. To address this possibility, we re-performed TETranscripts analysis with reads aligned to hg19, which lacks the GRCh38 alternative assemblies. This identified

Source	Sample	Gender	Age (yrs)	OS type	
<b>a. Discovery cohort (nucleic acid sequencing)</b>					
<i>Osteosarcomas</i>					
CHLA	OS-C1	M	7	High grade, extensive necrosis	
	OS-C2	F	13	Chondroblastic, necrosis	
	OS-C3	M	12	Conventional	
	OS-C4	M	5	High grade, small cell variant	
	OS-C7	M	17	Conventional secondary to RB	
	OS-C8	M	14	Conventional secondary to RB	
HLOH	OS-H6	M	16	Osteoblastic	
	OS-H7	M	17	Osteoblastic	
	OS-H11	F	25	Osteoblastic	
	OS-H19	M	15	Fibroblastic	
	OS-H20	M	17	Osteoblastic	
	OS-H27	M	29	Osteoblastic	
<i>Controls</i>					
CHLA	HC1	F	17		
	HRC1	M	5		
	HRC2	F	14		
	HRC3	F	11		
	HRC4	F	9		
HLOH	HH12	M	10		
	HH17	M	14		
	HH21	M	21		
	HH23	M	16		
	HH25	M	27		
	HH28	F	21		
	HH29	F	11		
<b>b. Validation cohort (qPCR)</b>					
<i>Osteosarcomas</i>					
CHLA	OS-C1	M	7	High grade, extensive necrosis	
	OS-C3	F	12	Conventional	
	OS-C5	M	13	Conventional	
	OS-C9	M	19	Conventional	
HLOH	OS-H21	M	43	High grade, extensive necrosis	
	OS-H24	M	46	Osteoblastic	
	OS-H25	F	16	Osteoblastic	
	OS-H34	M	14	Osteoblastic	
<i>Controls</i>					
CHLA	HC2	F	38		
	HC3	M	38		
IR	HI1	M	20		
	HI2	M	19		
	HI3	F	20		
	HI4	F	18		
	HI5	M	20		
	HI6	M	19		
	HI7	M	20		
	HI8	F	18		
	HI9	F	19		
	HI10	F	20		

**Table 1.** Osteosarcoma patient and control donor characteristics. OS-C, osteosarcoma from CHLA; OS-H, osteosarcoma from HLOH; HC, healthy control from CHLA; HH, Healthy control from HLOH; HI, healthy control from Innovative Research Inc.; HRC, hereditary retinoblastoma sibling control.

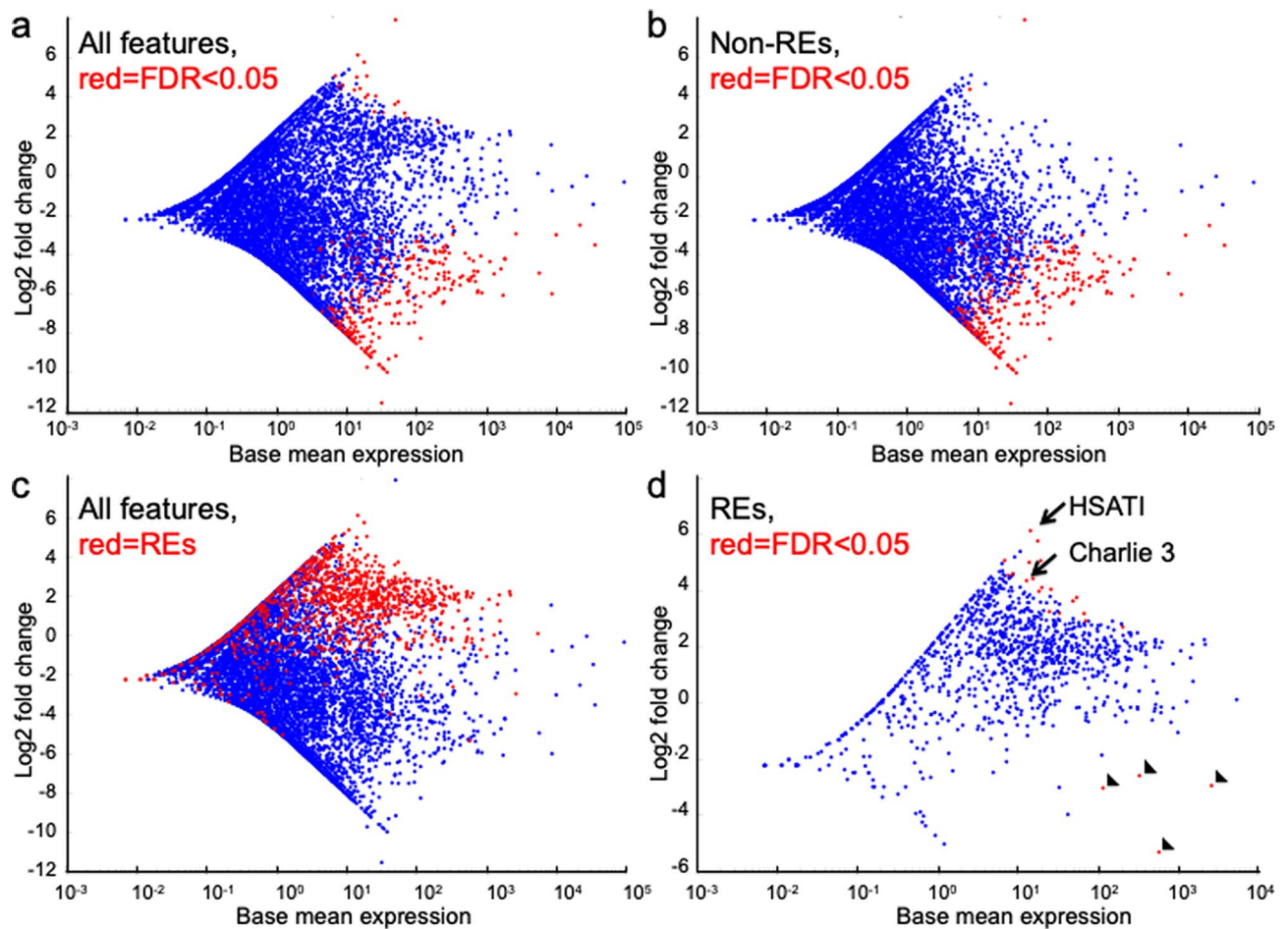


**Figure 1.** Patient EV preparation characteristics. (a,b) Representative control (a) and OS (b) particle size distributions in preparations used for nucleic acid extraction and sequencing, as defined by nanoparticle tracking (NanoSight). (c,d) Dot plots representing EV concentration of preparations from sera of OS patients ( $n = 12$ ) and controls from hereditary retinoblastoma sibling (HRC,  $n = 4$ ), and unrelated healthy (HC,  $n = 8$ ) and color-coded by source (c) and sex (d). Lines represent mean and standard deviation. Groups were compared using two-tailed, unpaired, Mann Whitney U test;  $*p < 0.05$ . (e) Dot plots of EV concentration versus donor age and color coded according to OS type. Spearman's correlation ( $r$ ) between EV concentration and donor age was not significant ( $p = 0.32$ ).

15 significantly over-represented repeat elements, of which Human Satellite II (*HSATII*) had highest fold change ( $\log_2(2.73)$ ,  $p_{\text{adj}} = 0.002$ ), and one significantly under-represented element in OS versus control sequences (Table 2, Analysis 2 and Supplementary Fig. 3). The significantly differentially represented repetitive elements identified using hg19 had little overlap with those identified when mapping to GRCh38.

**Validation of over-representation of repeat elements in OS EV preparations.** We next examined whether the increased representation of repetitive elements was evident in a validation set of mostly distinct samples. The validation cohort consisted of treatment-naïve OS patients from CHLA and HLOH including males and females between 7 and 46 years old and presenting with various OS types as well as approximately age-matched healthy individuals (Table 1b). The validation cohort was independent of the discovery cohort except for re-analysis of OS1 and OS3, which were the only samples with a sufficient quantity to re-test. To assess the repetitive element over-representation, EV-associated nucleic acids were isolated and evaluated using methods that differed from the discovery cohort analyses: EVs were isolated by PEG6000 precipitation<sup>36</sup> instead of Exo-Quick, nucleic acids were extracted using the miRNeasy micro RNA extraction kit (Qiagen) instead of SeraMir, and repetitive elements were examined by reverse transcription and quantitative PCR (RT-qPCR) instead of sequencing. Similar to the discovery cohort, EV concentrations were not significantly higher in sera from OS patients compared to controls (data not shown).

RT-qPCR was used to analyze four representative repetitive element categories including the *HSATI* and *HSATII* satellite sequences that were most differentially over-represented in Tetranscripts Analyses 1 and 2 (Table 2), the *LINE1 P1* family member (*L1P1*) that showed the highest fold change in the RepeatMasker analysis (Supplementary Fig. 2), and *Charlie 3*, another over-represented repetitive element with a significant  $\log_2$  fold change of 4.36 (20.5-fold increase) in Tetranscripts Analysis 1 (Table 2, Fig. 2d). RT-qPCR reactions yielded the predicted product sizes for *HSATI* (406 bp), *L1P1* (83 bp), and *Charlie 3* (104 bp). RT-qPCR of *HSATII* yielded prominent products of 85, 134, 183 and 281 bp, in agreement with *HSATII* genomic structure (Supplementary



**Figure 2.** Over-representation of repetitive elements in OS EV-associated sequences. MA plot for sequence features differentially represented in control and OS serum EV preparations as defined by Tetrascripts analysis 1. (a) Differentially represented single-copy genes and repetitive elements (REs), significantly differentially represented in red. (b) Differentially represented single-copy genes, significantly differentially represented in red. (c) Differentially represented single-copy genes in blue and REs in red. (d) Differentially represented REs, significantly differentially represented in red. Arrows, the significantly over-represented *HSATI* and *Charlie 3*. Arrowheads, significantly under-represented REs. Significantly differentially represented defined by FDR < 0.05, Wald test.

Fig. 4), instead of a reported ~200 bp amplicon found by RT-PCR with the same primers in an OS cell line<sup>37</sup>. The *HSATI* and *Charlie 3* products were confirmed to represent the predicted sequences by Sanger sequencing.

For each sample, RT-qPCR was performed on the same proportion of total EV nucleic acid extracted from the same serum volume and was normalized against a spike-in RNA. The analyses confirmed the over-representation of *HSATI*, *HSATII*, *L1P1*, and *Charlie 3* sequences in OS EV preparations (12.42-fold,  $p = 0.0040$ ; 3.33-fold,  $p = 0.062$ ; 3.56-fold,  $p = 0.016$ ; 12.6-fold,  $p = 0.0007$ ; respectively) (Fig. 3a). In contrast, the single-copy gene *HECDT2*, chosen on the basis of a 4.67 log<sub>2</sub> fold change in Tetrascripts analysis, did not show a significant difference (Supplementary Fig. 5A and B), and was deemed to have been spuriously identified, possibly due to biases in the library construction method<sup>38</sup>. The over-representation of *HSATI*, *HSATII*, *L1P1*, and *Charlie 3* sequences was similar after removal of the OS1 and OS3 samples that were also used in the discovery cohort (Supplementary Fig. 6A). Thus, the over-representation of repetitive element sequences initially detected by nucleic acid sequencing was confirmed in a validation cohort using RT-qPCR.

**Over-representation of repetitive element DNA but not RNA in OS EV preparations.** As our nucleic acid isolation and analysis methods could detect RNA as well as DNA sequences, we examined the nucleic acid origin of the over-represented repetitive element sequences by performing qPCR without reverse transcription. With this approach, the *HSATI*, *HSATII*, *L1P1* and *Charlie 3* amplification signals were significantly over-represented in OS versus control EVs (22.18-fold,  $p = 0.0015$ ; 3.7-fold,  $p < 0.0001$ ; 2.86-fold,  $p = 0.0015$ ; 10.29-fold,  $p = 0.011$ ; respectively) (Fig. 3b), indicative of differential representation of repetitive element DNA, rather than RNA. Evaluation of the sensitivity and specificity by Receiver Operating Characteristic (ROC) curves yielded area under the curves (AUCs) of 0.86 or greater for each repetitive element (Fig. 3c). As for the RT-qPCR analyses, results were similar after removal of OS1 and OS3 (Supplementary Fig. 6B and C).

Analysis 1 (mapped to GRCh38)	log <sub>2</sub> FC	p.adj	Analysis 2 (mapped to hg19)	log <sub>2</sub> FC	p.adj
<b><i>HSATI:Satellite:Satellite</i></b>	<b>6.14</b>	<b>0.007</b>	<b><i>HSATII:Satellite:Satellite</i></b>	<b>2.73</b>	<b>0.002</b>
LTR85a:Gypsy:LTR	5.77	0.003	L1MA6:L1:LINE	2.65	0.041
LTR75:ERV.L:LTR	5.07	0.042	MSTC:ERV.L-MaLR:LTR	2.52	0.001
LTR16E2:ERV.L:LTR	5.07	0.001	(GAATG)n:Satellite:Satellite	2.51	0.035
LTR16A1:ERV.L:LTR	5.01	0.014	Charlie4z:hAT-Charlie:DNA	2.37	0.011
Tigger8:TcMar-Tigger:DNA	4.60	0.025	Harlequin-int:ERV1:LTR	2.28	0.032
MLT1H2-int:ERV.L-MaLR:LTR	4.54	0.016	L1M3:L1:LINE	2.05	0.038
BLACKJACK:hAT-Blackjack:DNA	4.43	0.029	L1MB8:L1:LINE	2.03	0.005
<b><i>Charlie3:hAT-Charlie:DNA</i></b>	<b>4.36</b>	<b>0.030</b>	L1PB4:L1:LINE	2.00	0.021
MER113:hAT-Charlie:DNA	4.11	0.019	L1MC3:L1:LINE	1.90	0.046
MLT1A1-int:ERV.L-MaLR:LTR	4.05	0.015	L1PB2:L1:LINE	1.89	0.026
L1MCb:L1:LINE	3.98	0.037	L1MEc:L1:LINE	1.88	0.047
MLT2B2:ERV.L:LTR	3.75	0.018	MLT1A0:ERV.L-MaLR:LTR	1.81	0.016
HSAT4:centr:Satellite	3.63	0.025	L1MB7:L1:LINE	1.80	0.028
LTR32:ERV.L:LTR	3.25	0.042	L1MDa:L1:LINE	1.45	0.041
Charlie4z:hAT-Charlie:DNA	3.20	0.034	L1PA12:L1:LINE	-2.27	0.002
CER:Satellite:Satellite	3.18	0.021			
MLT1E2:ERV.L-MaLR:LTR	2.92	0.029			
L1MC4a:L1:LINE	2.71	0.040			
MER74A:ERV.L:LTR	-2.61	0.041			
L1PA16:L1:LINE	-2.96	0.029			
FAM:Alu:SINE	-3.05	0.015			
FordPrefect:hAT-Tip100:DNA	-5.34	<0.001			

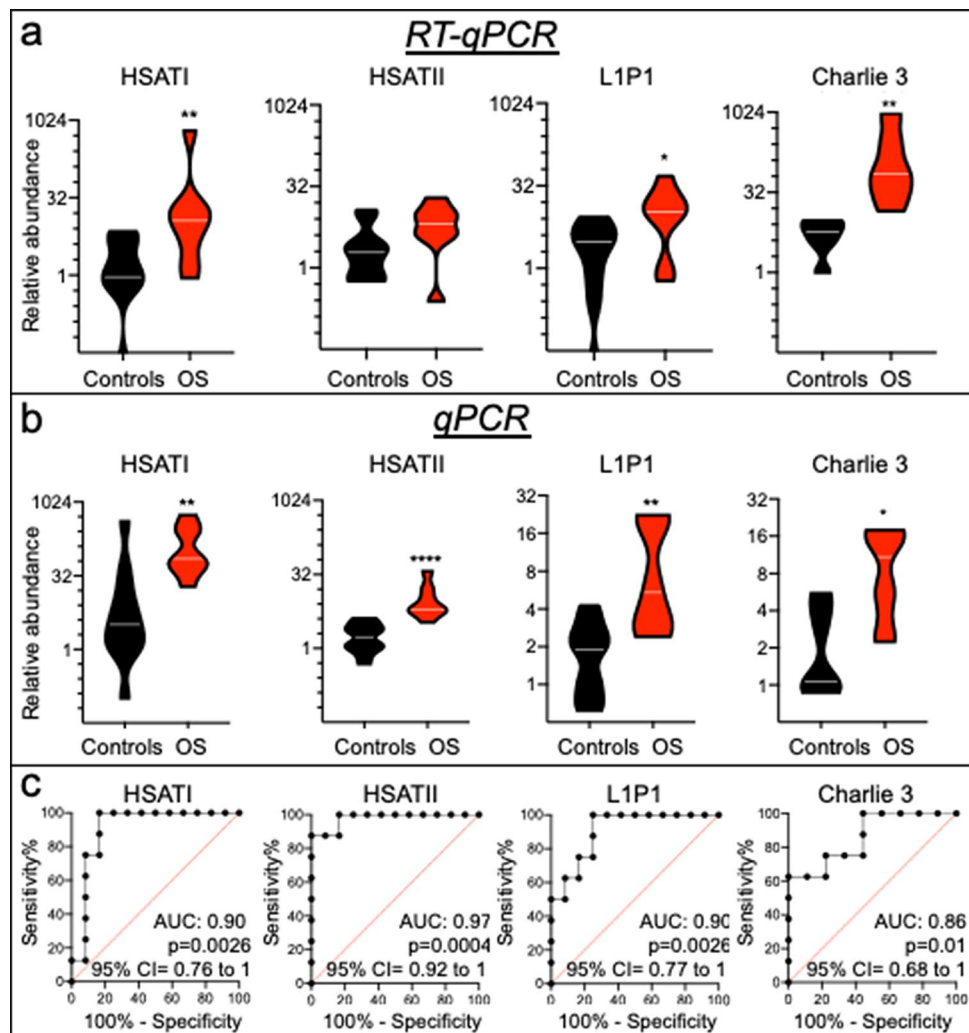
**Table 2.** Repeat elements significantly differentially represented in OS versus control EV preparations ( $p$ .adj < 0.05; elements examined in the validation cohort in bold italics).

*HSATI*, *HSATII*, *LIP1* and *Charlie 3* DNAs were similarly increased in OS samples from USA (CHLA) and China (HLOH) and in high-grade as well as non-high grade OS samples (Supplementary Fig. 7). In contrast, qPCR of *IL17RA* and *ZNF3* sequences, which are within 1 Mb of *HSATI* on chromosome 22 (chr22q11.21) and within 40 Mb of *HSATII* on chromosome 7 (chr7q22.1), respectively, did not show a significant difference between OS and control EVs (Supplementary Fig. 8A and B). Thus, major repetitive elements DNAs (*HSATI*, *HSATII*, *LIP1*, *Charlie 3*) were over-represented, whereas single copy genes were not over-represented, in OS EV preparations.

To further evaluate the abundance of repetitive element DNAs and control for possible artefactual generation of RT-independent products, PEG-precipitated EV preparations from four OS and four control sera were treated with DNase I or RNase A prior to nucleic acid extraction. RNase A treatments were performed in 1 M NaCl in order to cleave single-stranded RNA as well as in the absence of NaCl in order to cleave single-stranded and double-stranded RNA and RNA strands in RNA-DNA hybrids<sup>39</sup>. After these treatments, nucleic acids were extracted with the miRNeasy Micro kit and *HSATI* and *HSATII* abundance were assessed by qPCR. In these analyses, DNase I treatment eliminated 97–99% of *HSATI* and 80–99% of *HSATII* signals in both OS and control samples, whereas RNase A treatments had no significant effect (Fig. 4a). Bioanalyzer assessments revealed that DNase treatment slightly decreased the amount of nucleic acid whereas RNase A eliminated most but not all of the nucleic acids (Fig. 4b), with the majority of the remaining nucleic acid likely representing protected EV RNA. The repetitive elements' sensitivity to DNase I prior to nucleic acid extraction implied that the repetitive element DNA sequences were not sequestered inside of EVs.

To assess whether OS serum EV preparations might also have an increased abundance of repetitive element RNAs, PEG-precipitated EVs were prepared, treated with DNase I, and the remaining nucleic acids extracted and examined by RT-qPCR. In these samples, no amplification signal was detected for *HSATI* or *Charlie 3*, while *HSATII* and *LIP1* products were reduced ~ 32–64-fold compared to non-DNase I treated samples and showed no significant difference in control and OS samples (Fig. 4c and data not shown). Thus, *HSATI*, *HSATII*, *LIP1*, and *Charlie 3* DNAs were over-abundant in OS compared to control EV preparations whereas their RNAs were either undetectable or present in similar quantities.

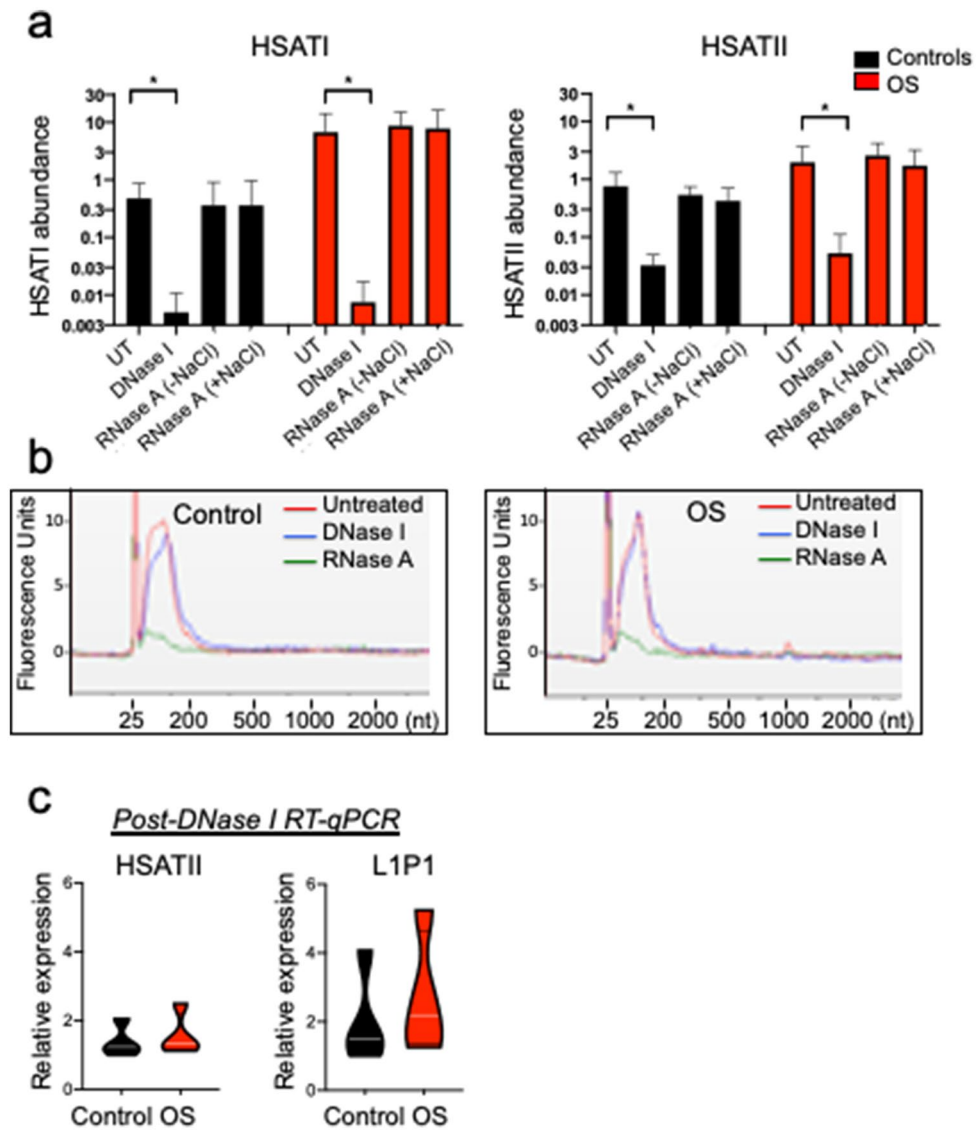
**Co-purification of OS-associated repetitive element DNAs with EVs in size exclusion chromatography but not exosome immunoaffinity capture.** To further evaluate if repetitive element DNAs that were more abundant in OS patient PEG-precipitations (here termed 'OS-associated' repetitive element DNAs) are associated with EVs, we examined whether they co-purified with EVs prepared by size exclusion chromatography (SEC) and exosome immunoaffinity capture. SEC yields more pure EV populations<sup>40,41</sup> with lower protein contamination compared to PEG precipitation<sup>42,43</sup>, whereas exosome immunoaffinity capture uses well-characterized surface markers CD9 or CD81 to highly purify intact exosomes<sup>41,44</sup>.



**Figure 3.** Over-representation of repetitive elements in OS compared to control EV preparations in a validation cohort. **(a)** Violin plots representing relative abundance of *HSATI*, *HSATII*, *L1P1* and *Charlie 3* by RT-qPCR of control (n=6–11) and OS (n=7–8) serum EV preparations. RT-qPCR was normalized against *C. elegans* external spike-in miR-39–3p RNA. White lines represent median. **(b)** Violin plots representing relative abundance of *HSATI*, *HSATII*, *L1P1* and *Charlie 3* DNA by qPCR, in the absence of reverse transcription, in control (n=12) and OS (n=8) serum EV preparations. qPCR was performed on equal proportions of nucleic acid extracted from 200  $\mu$ l of OS and control serum. White lines represent median. **(c)** Diagnostic value of *HSATI*, *HSATII*, *L1P1* and *Charlie 3* DNA sequences in serum OS preparations. ROC curves were generated using data in **(b)**. Groups were compared using two-tailed, unpaired, Mann Whitney U test; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

We first examined if repetitive element DNAs co-purify with SEC-isolated EVs from two control and two OS samples. Nanoparticle tracking analyses revealed that the control and OS EVs both eluted from size exclusion columns solely in fractions 6 and 7 (Fig. 5a,b) and had similar size profiles (Fig. 5c,d). qPCR analyses revealed that the OS EV fractions had more abundant *HSATI* and *HSATII* DNA (Fig. 5e), as observed with the PEG-isolated EVs of the same samples (Fig. 5f). Similarly, *HSATI* and *HSATII* levels were higher in the same OS versus control sera when normalized to EV concentration (Supplementary Fig. 9A and B). Furthermore, nucleic acids obtained from arbitrarily selected non-EV fractions from one control and one OS fractionation (Fig. 5g,h) revealed no detectable *HSATI* and only minimal *HSATII*, which was not higher in OS samples (Fig. 5i,j). Thus, *HSATI* and *HSATII* DNAs co-purified with EVs in SEC with a greater abundance in OS compared to control sera, similar to that of PEG preparations.

We next assessed whether OS-associated repetitive element DNAs co-purified with EVs in exosome CD9 or CD81 immunoaffinity capture. In pilot studies, we confirmed that our immunocapture approach enriched for exosomes by single particle interferometric reflectance imaging sensing (SP-IRIS) using an ExoView instrument<sup>45</sup>. SP-IRIS analyses showed that a similar number of EV particles eluted from CD9 immunoaffinity capture from



**Figure 4.** Repetitive element sensitivity to DNase in EV preparations. **(a)** Abundance of *HSATI* and *HSATII* sequences determined by qPCR in control (n=4) and OS (n=4) serum EV preparations isolated with PEG and either untreated (UT) or pretreated with DNase I or RNase A with or without NaCl prior to nucleic acid extraction. Treated groups were compared to untreated groups using unpaired Kruskal–Wallis test with uncorrected Dunn’s test where each comparison stands alone; \*,  $p < 0.05$ . Error bars represent standard deviation of biological replicates. **(b)** Bioanalyzer electropherograms of equal proportions of nucleic acids prepared from 200  $\mu$ l of representative control and OS EV preparations that were untreated or treated with DNase I or RNase A prior to nucleic acid extraction. **(c)** Violin plots representing relative abundance of *HSATII*, *L1P1* and *Charlie 3* by RT-qPCR in control (n=4) and OS (n=4) serum EV preparations pre-treated with DNase I. qPCR was normalized against *C. elegans* external spike-in miR-39–3p RNA. White lines represent median.

control (21,006 particles, n=1) and OS sera (21,264  $\pm$  374 (SEM) particles, n=2), that the eluted particles could be re-immunocaptured on the microarray-based solid phase chip coated with antibodies to exosomal surface markers (Fig. 5k,l), and that the re-captured particles were from 50 to 80 nm diameter (characteristic of exosomes) (Fig. 5m,n) and expressed various combinations of exosomal markers CD81, CD63, and CD9, similar to the PEG precipitated EV preparations of the same samples (Supplementary Fig. 10). Nanoparticle tracking analyses of the CD9 and CD81 immunoaffinity capture eluates showed particle size distributions similar to that of PEG precipitations but larger than reported by SP-IRIS (Fig. 5o,p, Supplementary Fig. 11A and B), as expected<sup>46</sup>. However, in contrast to PEG- and SEC-isolated EVs, CD9 and CD81 immunoaffinity captured EV preparations showed no significant difference in *HSATI* and *HSATII* DNA abundance between control and OS samples (Fig. 5q,r and Supplementary Fig. 11C). Thus, the OS-associated *HSATI* and *HSATII* DNAs failed to co-purify with EVs during immunocapture in contrast to their co-purification with EVs isolated via PEG precipitation or



SEC. It is inferred that OS-associated *HSATI* and *HSATII* DNAs either fail to bind CD9+ or CD81+ exosomes or dissociate from such exosomes under immunocapture conditions.

**Enrichment of human satellite sequences in EV-associated DNA but not in total cfDNA in OS patient sera.** Our finding that repetitive element DNAs were increased in OS patient PEG and SEC EV preparations yet not tightly bound to CD9+ or CD81+ exosomes raised the possibility that repetitive element DNAs might be more abundant in total cfDNA of OS patients and were proportionately present as contaminants in OS and control EV nucleic acid preparations. To examine this possibility, nucleic acids were extracted directly from equal volumes of OS and control sera using the same miRNeasy micro RNA extraction kit as used for EV preparations and the repetitive element abundance was examined by qPCR. This revealed that *LIP1* and *Charlie 3* were significantly more abundant whereas *HSATI* and *HSATII* were present at similar levels in OS and control cfDNA samples (Fig. 6a). Importantly, omitting the PEG EV preparation step eliminated the diagnostic sensitivity of *HSATI* and *HSATII* ( $AUC = <0.72$ ) while reducing that of *LIP1* ( $AUC = 0.81$ ) and not affecting that of *Charlie 3* ( $AUC = 0.85$ ) (Fig. 6b) as compared to the AUC values obtained after PEG precipitation (Fig. 3c). These data imply that EV enrichment by PEG or SEC was required in order to detect the increased representation of *HSATI* and *HSATII* and to increase the sensitivity for *LIP1* in OS patient versus control sera.

## Discussion

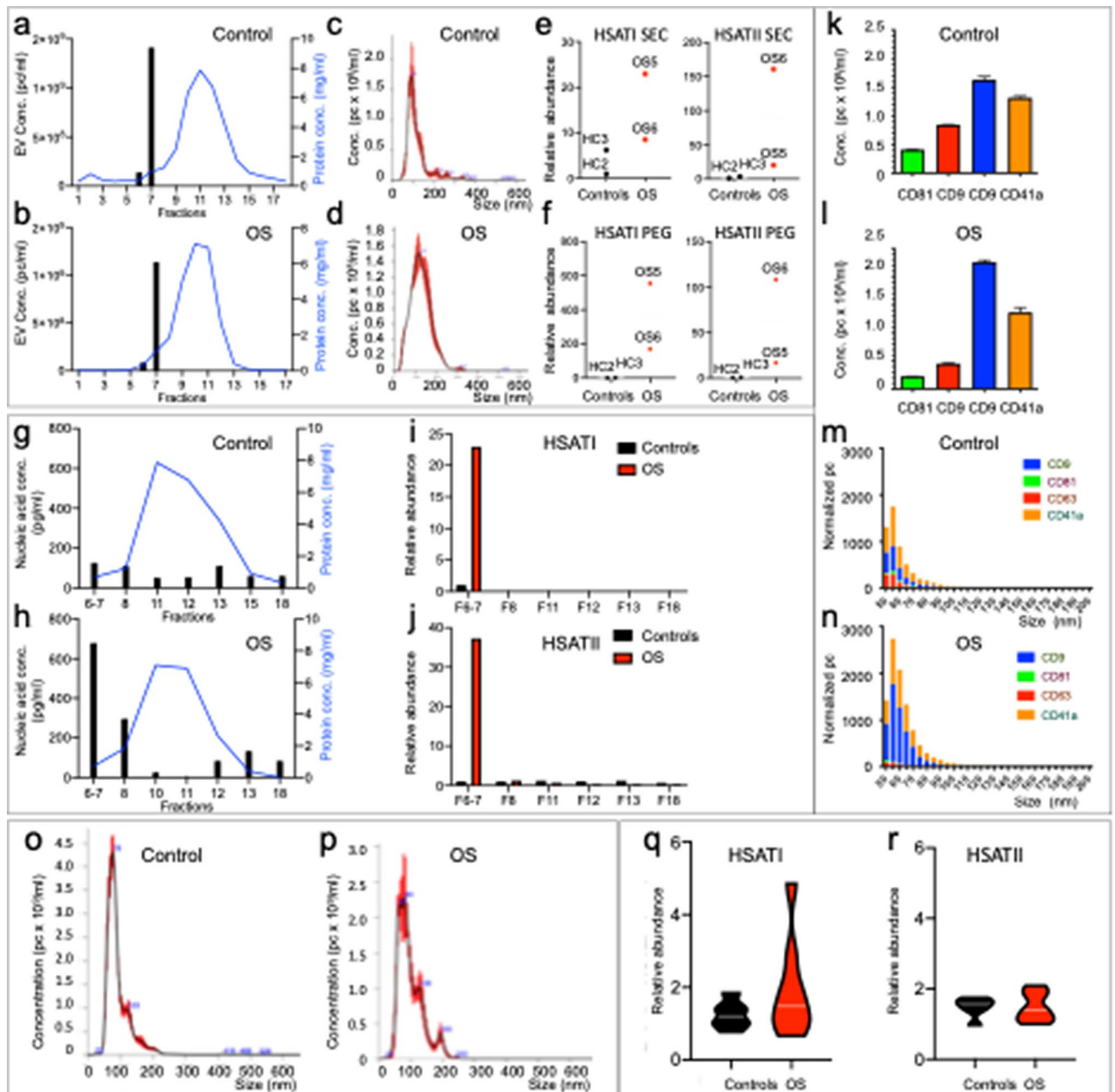
Sensitive biomarkers are needed to detect incipient OS tumors and enable life-saving interventions in predisposed individuals. Prior studies identified a variety of potential OS biomarkers yet none had sufficient sensitivity to enable reliable OS detection<sup>12,13</sup>. To identify new OS biomarkers, we investigated nucleic acid sequences associated with circulating EVs in OS patients. This revealed an over-representation of diverse repetitive element sequences, among which human satellites *HSATI* and *HSATII* were the most significantly increased upon mapping to the GRCh38 and hg19 genome builds. The over-represented repetitive element sequences were confirmed in a validation cohort and found to reflect repetitive element DNAs that co-purified with circulating EVs but were not tightly bound to CD9+ or CD81+ exosomes. *HSATI* and *HSATII* were distinguished from other repetitive elements in that they were enriched in serum EV preparations but not in total cfDNA, implying that they segregated into distinct complexes in the circulation of OS patients.

Our detection of increased repetitive element DNAs in OS patient sera was enabled by a novel screening approach. First, in the discovery cohort, we prepared serum EVs using a precipitation method that concentrates exosomes as well as other EVs and non-vesicular constituents<sup>29,30</sup>, which enlarged the population of biomarker candidates. Second, we did not treat EVs or nucleic acid preparations with DNase, which enabled isolation of DNAs as well as RNAs and further diversified the potential biomarker pool. Third, we isolated nucleic acids using small RNA preparation kits that also captured repetitive element DNAs, and we built our sequencing library by direct ligation of adapters to extracted DNAs as well as RNAs, using an activity with properties similar to T4 RNA ligase<sup>47</sup>, which allowed the discovery of differentially represented DNA as well as RNA species. Finally, we evaluated repetitive element sequences that include diverse satellite and non-satellite categories that may comprise more than two-thirds of the human genome<sup>48</sup>. Repetitive elements are often ignored in human sequencing studies because of the complexity involved in properly aligning short sequencing reads to highly repetitive regions as well as poor understanding of their functional relevance<sup>33</sup>. However, the paucity of over-represented single copy sequences in OS patient EV preparations prompted us to consider whether repetitive element sequences might be over-represented.

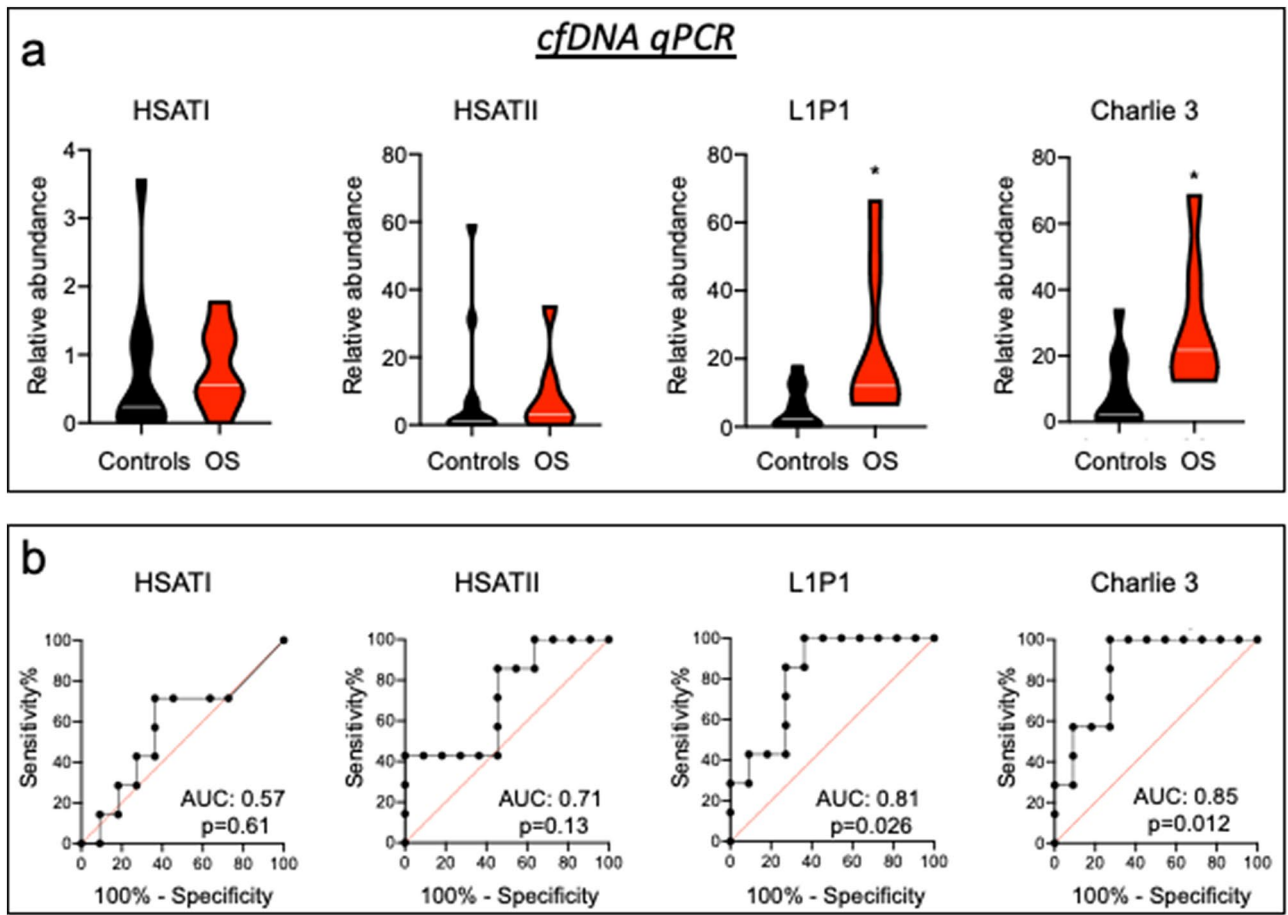
To examine differential repetitive element sequence representation, we initially used RepeatMasker to align sequence reads against the Repbase library of known repeats<sup>32</sup>. This revealed an over-representation of all repetitive element categories in OS serum EV preparations, with the *LINE1* family member *LIP1* as the most significantly over-represented species (Supplementary Fig. 2). To more accurately identify differentially represented repetitive elements, we used Tetranscripts, which assigns both uniquely and ambiguously mapped reads to all possible gene and transposable element-derived transcripts in order to statistically infer the correct gene or transposable element abundances<sup>34</sup>. Tetranscripts analyses confirmed that repetitive elements were over-represented in OS serum EV preparations and identified *HSATI*, *HSATII* and *Charlie 3*, among others, as significantly over-represented (Table 2). Different elements were identified when reads were aligned to GRCh38 or to hg19, likely due to the presence of alternative repetitive-element-enriched sequence assemblies in GRCh38<sup>35</sup>.

We validated the increased abundance of selected repetitive element sequences in a second patient cohort. In the validation set, we observed their increased abundance via qPCR, without reverse-transcription and in a DNase-sensitive manner, implying differential representation of repetitive element DNAs. Repetitive element DNA sequences were similarly increased in OS samples from USA (CHLA) and China (HLOH) and did not correlate with the OS grade, suggesting that these elements are produced independently of OS type. EV-associated repetitive element DNAs showed a high sensitivity and specificity for sera of patients with an OS diagnosis, with a significant  $AUC > 0.9$  for *HSATI*, *HSATII* and *LIP1*. However, the sensitivity was diminished by omitting the EV preparation step, particularly for *HSATI* and *HSATII* (Fig. 6), suggesting that these OS-associated repetitive element DNAs are segregated from bulk cfDNA.

At present the physical state of the repetitive element DNAs that enables their enrichment from OS patient sera is unknown. Prior studies detected repetitive element DNAs in growth medium of human OS cells, yet their physical state was not characterized<sup>49</sup>. In serum EV preparations, the OS-associated repetitive element DNAs were sensitive to DNase I, indicating that they were not a protected EV cargo (Fig. 4a), in contrast to DNAs in cultured cancer cell-derived microvesicles<sup>27</sup>. As repetitive element DNAs co-purified with EVs in PEG precipitation and SEC but not in CD9 or CD81 immunoaffinity capture, they might associate with exosomes too weakly to survive immunocapture, associate with CD9- and CD81-negative exosomes, associate with non-exosome EVs,



**Figure 5.** Co-purification of OS-associated repetitive element DNAs with EVs in size exclusion chromatography but not exosome immunoaffinity capture. (a,b) Representative protein concentration (blue line) and EV concentration (black bars) elution profiles from control (a) and OS (b) serum separated by size exclusion chromatography (SEC). (c,d) Representative size distribution of control (c) and OS (d) serum EV particles in pooled fractions 6 and 7 analyzed by nanoparticle-tracking. (e,f) Relative abundance of *HSATI* and *HSATII* DNA in two control and two OS SEC (e) and PEG (f) EV fractions as defined by qPCR. (g,h) Representative protein concentration (blue line) and nucleic acid concentration (black bars) of pooled EV fraction (F6-7) and selected non-EV fractions (F8, F11, F12, F13 and F18) from the same representative control (g) and OS (h) SEC separations as in (a) and (b). (i,j) Abundance of *HSATI* (i) and *HSATII* (j) in pooled EV fractions 6-7 and non-EV fractions on one control and one OS SEC analysis as evaluated by qPCR. (k-n) SP-IRIS analyses by ExoView of EVs isolated by CD9-immunoaffinity capture. (k,l) Representative concentration of control (k) and OS (l) fluorescent EV particles immunocaptured on the CD9, CD81 and CD63 antibody spots. Results depict the mean of the measurement of triplicate spots  $\pm$  SEM, subtracted for IgG spot values and adjusted by dilution factor. (m,n) Representative size distribution of control (m) and OS (n) label-free EV particles immunocaptured on the CD9, CD81 and CD63 antibody spots. Results depict the mean of the measurement of triplicate spots  $\pm$  SEM, subtracted for IgG spot values. (o,p) Size distribution and particle number of control (o) and OS (p) EVs isolated by CD9 immunoaffinity capture and analyzed by nanoparticle-tracking. (q,r) Violin plots representing abundance of *HSATI* (q) and *HSATII* (r) of control (n=6) and OS (n=8) immunoaffinity capture of CD9-positive exosomes evaluated by qPCR. White lines represent median. Groups were compared using two-tailed, unpaired, Mann Whitney U test; \* $p < 0.05$ . Similar CD81 immunoaffinity capture results in Supplementary Fig. 11.



**Figure 6.** Human satellite sequences not enriched in total cfDNA in OS patient sera. (a) Violin plots representing relative abundance of *HSATI*, *HSATII*, *L1P1* and *Charlie 3* by qPCR of control ( $n = 11$ ) and OS ( $n = 7$ ) whole serum. White lines represent median. (b) Diagnostic value of *HSATI*, *HSATII*, *L1P1* and *Charlie 3* in OS patient whole serum. ROC curves were generated using data in (a). Groups were compared using two-tailed, unpaired, Mann Whitney U test;  $*p < 0.05$ .

or participate in large non-vesicular complexes that co-purify with EVs in the PEG and SEC methods. Despite prior reports that DNA is present in cancer exosomes or small EVs<sup>27,50–52</sup>, a recent reassessment demonstrated that small vesicles are not vehicles of active DNA release and that double stranded DNA was associated with non-vesicular entities that are extruded from cancer cells<sup>53</sup>. Similar to our results, this reassessment did not observe DNA associated with immunoaffinity capture of CD81-positive exosomes, further supporting that exosomes do not contain DNA or tightly associate with other particles that contain DNA. Understanding the biogenesis of repetitive element DNAs that are enriched in OS patient EV preparations could provide insight into OS pathogenesis as well as a tool with which to detect asymptomatic OS or other conditions in which EV-associated repetitive element DNAs are elevated.

A limitation of this study is that it is based on two relatively small patient cohorts and examined a mixture of EV-associated DNAs and RNAs in the discovery cohort. Thus, the sensitivity of the repetitive element DNA analyses for OS and the specificity for OS versus other cancers may be refined by defining differential DNA representation in a larger series of OS as well as other cancers and non-cancerous conditions. Nevertheless, the results suggest that EV-associated repetitive element DNAs are among the most sensitive markers of newly diagnosed OS identified to date, with ROC curve AUCs of 0.90 for *HSATI* and 0.97 for *HSATII* (Fig. 3c). By comparison, in prior OS biomarker analyses, circulating miRNAs had at best ROC curve AUCs of 0.833–0.955, yet were significantly elevated in only a subset of many similar miRNA screens<sup>11,54–56</sup>. Likewise, a deep sequencing approach detected circulating tumor DNA aneuploidy in only 50% of treatment-naïve OS patients<sup>12</sup>. Still, in this study, the increased repetitive element DNAs were observed in sera of already-diagnosed OS patients, and further refinements are likely to be required in order to detect repetitive element DNAs produced by smaller, incipient tumors prior to their clinical appearance.

A final question raised by our findings is whether similar EV-associated repetitive element DNAs are increased in the circulation of patients with other cancers. Notably, centromeric and pericentric repetitive element RNA sequences, particularly alpha satellites and satellite II and III sequences, were reported to be overexpressed in testicular, liver, ovarian, and lung cancers compared to corresponding normal tissues<sup>57</sup>. The pericentric human satellite II (*HSATII*) RNA was reported to be the most differentially expressed satellite subfamily in pancreatic

cancer tissue and was also overexpressed in lung, kidney, ovarian, colon and prostate cancers<sup>58,59</sup>. Moreover, *HSATII* was one of the six most up-regulated satellite sequences in a study comparing fresh bone and OS samples by RNA-seq<sup>28</sup>. *LINE-1* was also overexpressed in pancreatic and prostate tumor samples<sup>60,61</sup>. However, although *LINE1* and other repetitive element RNAs were detected in cancer cell-derived EVs in culture<sup>27</sup>, their up-regulation has not been reported for circulating cell-free RNA in cancer patients. Our detection of the enriched DNA templates of these repetitive element RNAs in the OS patient circulation raises the possibility that circulating repetitive element DNAs may also be enriched in additional cancer types.

## Materials and methods

**Patients and samples.** This study was reviewed and approved by the institutional review board at Children's Hospital Los Angeles (approval no. CCI-13-00223) and at Henan Luoyang Orthopedic Hospital (approval no. 2015-01). All participants gave a written informed consent. Parents/Legally authorized persons gave informed consent on behalf of the all minors and subjects above 14 years old gave assent. All analyses were conducted in accordance with relevant guidelines and regulations. Blood samples were collected during a clinically indicated venipuncture from previously untreated patients with primary diagnosis of OS and from volunteer subjects with no known medical conditions, i.e. healthy controls. Control sera for the validation cohort were obtained from local volunteer subjects and from Innovative Research Inc. (Novi, MI, USA). Blood was drawn in serum separator collection tubes (SST), clotting was allowed for 30 min at room temperature in vertical position and then tubes were centrifuged at 1000g for 10 min at 4 °C. Serum was collected, immediately aliquoted, and stored at -80 °C.

**Discovery cohort: EV isolation, nucleic acid extraction, and sequencing.** Serum EVs were isolated using ExoQuick (System Biosciences Inc. (SBI), Mountain View California, USA) and aliquots frozen. One aliquot was used for NTA analyses and on confirmation of high EV purity aliquots were thawed and nucleic acid extracted using SeraMir (SBI) without DNase treatment, according to manufacturer instructions. The sequencing library was constructed using TailorMix miRNA Sample Preparation (SeqMatic) with a selection of small nucleic acids from 140 to 300 bases. 5'-RNA adapters and 3'-DNA adapters (SeqMatic, personal communication) were directly ligated to nucleic acid substrates, followed by PCR amplification. Libraries were sequenced to generate single-end 50 bp reads on MiSeq 500 platform (Illumina).

**Validation cohort: EV isolation and nucleic acid extraction.** Serum was cleared by centrifugation at 3000 × g for 15 min at 4 °C. For polyethylene glycol (PEG) precipitation, 50–200 µl of cleared serum was combined with an equal volume of freshly prepared 16% PEG 6000 (Sigma-Aldrich) in 1 M NaCl, to give a final concentration of 8%, incubated for 30 min on ice, centrifuged in a tabletop microfuge at 16,000 × g for 2 min at room temperature (Eppendorf, model 5424 R using an FA-45-24-11 fixed angle rotor) and the pellet resuspended in a volume of PBS equal to that of the starting serum volume. For size exclusion chromatography (SEC), ~300 µl of cleared supernatant was centrifuged at 10,000 g for 30 min at 4 °C in a fixed angle rotor and loaded onto a glass Econo-column (Bio-Rad, 10 cm height, 1.5 cm diameter) packed with Sephacryl S-300 High Resolution (GE Healthcare) and pre-washed with 0.32% Sodium Citrate in PBS. The cleared serum was allowed to enter the resin by gravity flow and eluate collected in 20 fractions of 15 drops (~500 µl) on a Model 2110 Fraction Collector (Bio-Rad). For each fraction, the protein concentration and the presence of EVs was characterized by Bradford method (Bio-Rad) and nanoparticle tracking analysis (see below), respectively. EV fractions were concentrated on a 100 kDa Amicon ultra centrifugal filter (Millipore) from 2 × 500 µl to a final volume of ~100 µl. Immunoaffinity capture of CD81+ or CD9+ EVs was carried out using the Exo-Flow Exosomes Purification Kit (SBI, Mountain View California). Briefly, 200 µl of cleared serum was precipitated with 200 µl of 16% PEG 6000 as above, and the pellet re-suspended in 200 µl of PBS. 50 µl of this EV preparation were incubated in 20 µl of anti-CD81 or anti-CD9 pre-coated magnetic beads (9.1 µm) on a rotating rack at 4 °C overnight. CD81+ or CD9+ EVs were eluted from the beads in the Exosome Elution Buffer at 25 °C for 30 min.

Nucleic acids were extracted from 20 to 200 µl of EV preparations (or from 50 µl of serum) using miRNeasy Micro kit (Qiagen) and suspended in 14 µl of RNase/DNase-free H<sub>2</sub>O (depending the initial volumes of serum) according to the manufacturer's instruction. For samples intended for reverse-transcription, a spike-in control (*C. elegans* miR-39-3p) miRNA mimic (Qiagen) was added ( $1.6 \times 10^9$  copies) after the lysis step. Nucleic acid size and concentration were analyzed on an RNA Pico 6000 chip using an Agilent Bioanalyzer (Agilent, Palo Alto, CA, USA), equipped with Expert 2100 software, which generated an electrophoretic profile and the corresponding 'pseudo' gel of the sample. After separation, nucleic acid sizes were normalized to a 25 bp RNA marker. Samples showing nucleic acids of > 200 bp were eliminated from the study.

**Particle size and concentration measurement by nanoparticle tracking analysis.** EV preparations were analyzed by nanoparticle tracking using a NanoSight NS300 (Malvern, Worcestershire, U.K.) configured with a high sensitivity sCMOS camera (OrcaFlash2.8, Hamamatsu C11440, NanoSight Ltd). In brief, each sample was mixed by vortexing, and subsequently diluted in particle-free PBS to obtain a concentration within the recommended measurement range ( $10^8$ – $10^9$  particles/mL), corresponding to dilutions from 1:100 to 1:500. After optimization, settings were kept constant between measurements. Ambient temperature was recorded manually and did not exceed 25 °C. Approximately 20–40 particles were in the field of view for each measurement. Three videos of 30 s duration were recorded for each sample. Experiment videos were analyzed using NTA 3.2 Dev Build 3.2.16 software (Malvern).

**Single particle interferometric reflectance imaging sensing (SP-IRIS).** EVs from PEG and immunocapture preparations were analyzed on ExoView R100 platform (Nanoview Biosciences, MA). Briefly, EVs within these preparations were immunocaptured on a multiplexed microarray chip with CD9, CD81, CD63, and CD41a antibody spots, as well as negative control IgG antibody spots to determine the level of non-specific binding, and then probed for CD9, CD81, CD63, and CD41a surface markers with respective additional fluorescent antibodies. EVs from PEG preparation and eluted EVs from immunoaffinity were diluted in solution A (Nanoview Biosciences, MA). The samples were incubated on the ExoView Tetraspanin Chip (EV-TC-TTS-01) placed in a sealed 24-well plate for 16 h at room temperature. The chips were then washed three times in 1 ml PBST for 3 min each on an orbital shaker. Then, chips were incubated with ExoView Tetraspanin Labeling ABs (EV-TC-AB-01) that consist of anti-CD81 Alexa-555, anti-CD63 Alexa-488, and anti-CD9 Alexa-647. The antibodies were diluted 1:5000 in PBST with 2% BSA. The chips were incubated with 250  $\mu$ l of the labeling solution for 2 h. The chips were then washed once in PBST, three times in PBS followed by a rinse in filtered deionized water and dried. Immunocaptured EVs on the microarray chip were imaged on a single EV-basis with the ExoView R100 reader using the nScan2 2.9 acquisition software. The data were then analyzed using the NanoViewer 2.9 software (Nanoview Biosciences, MA) that counts and sizes fluorescent nanoparticles immunocaptured on the antibody spots. For exosome analysis the size window was selected to include particle sizes from 50 to 200 nm.

**Reverse transcription (RT) and qPCR.** Equal volumes of nucleic acid prepared as above were reverse-transcribed using iScript cDNA Synthesis Kit (Bio-Rad) in 20  $\mu$ l volume according to the manufacturer's protocol. 0.5  $\mu$ l of the samples produced with or without the RT step were analyzed in 10  $\mu$ l qPCR reactions with iQ Green Supermix (Bio-Rad) on an ABI 7900 Fast Real-Time PCR System (Applied Biosystems) with the following cycling parameters: 94  $^{\circ}$ C, 30 s; 59  $^{\circ}$ C, 15 s; 68  $^{\circ}$ C, 25 s for 35 cycles. Relative sequence abundance was determined by the  $\Delta\Delta$ Ct method. In most PCR runs, a negative control with no nucleic acid template was added and never generated PCR product. PCR primers were designed manually to have a melting temperature of 58  $^{\circ}$ C and to generate amplicons of ~100 bp or as previously described for *HSATI*<sup>62</sup> and *HSATII*<sup>37</sup> (Supplementary Fig. 4) and obtained from Integrated DNA Technologies:

*HSATI* F: 5'-TAATGTGTGGGCTTGGGATT-3', *HSATI* R: 5'-TGCATATGGAAAATACAGAGGCTA-3' (amplicon: 406 bp); *HSATII* F: 5'-ATTCGATTCCATTCGATGATGATTCC-3', *HSATII* R: 5'-GGAACCGAA TGAATCCTCATTGAATG-3' (prominent amplicons; 85, 134, 183 and 281 bp); *LIP1-orf2* F: 5'-ATCAGA GAATACTACAAACACCTCTAC-3', *LIP1-orf2* R: 5'-AGAGTGTATGTGTCGAGGAAT-3' (amplicon: 83 bp); *Charlie 3* F: 5'-ACAAAAGCACTGAAAAGCCTGC-3', *Charlie 3* R: 5'-TCCAGTCTACTCCGTAATCTCGT-3' (amplicon: 104 bp); *HECDT2* F: 5'-TGTGAAAGACTTTCAGGAAGATGTAGAAAAA-3', *HECDT2* R: 5'-GAG GGAGTGGCATCTTTCTTAAATG-3' (amplicon: 131 bp). For *ZNF3* and *IL17RA*, TaqMan primers were used (*IL17RA*: Hs01285262\_cn and *ZNF3*: Hs03631848\_cn (Applied Biosystems) and PCR reactions were performed according the manufacturer's instructions.

**DNase I and RNase A treatment of EV preparations.** Intact EV preparations were treated with DNase I (Qiagen) in RDD buffer for 15 min at room temperature and then inactivated for 10 min at 70  $^{\circ}$ C. Intact EV preparations were treated with RNase A (Thermo Scientific) at final concentration 0.4  $\mu$ g/ $\mu$ l with or without NaCl at final concentration 1 M for 10 min at 37  $^{\circ}$ C<sup>39</sup> and inactivated by RNase inhibitor (Takara) at final concentration 2u/ $\mu$ l.

**RNA-Seq data processing, alignment and analysis.** Fastq files were aligned to GRch38 (for analysis 1) or hg19 (for analysis 2) using STAR using the parameters recommended for TEtranscripts (i.e. allowing for up to 100 alignments per read)<sup>34</sup>, and the resulting BAM files were processed using TEtranscripts to quantify both non-repetitive element and repetitive element abundance.

**Statistical analysis.** Groups were compared using two-tailed, unpaired, Mann Whitney U test (\* $p$ <0.05, \*\* $p$ <0.01, \*\*\* $p$ <0.001, \*\*\*\* $p$ <0.0001). All analyses were performed using Prism 8 software (GraphPad).

Received: 1 May 2020; Accepted: 29 October 2020

Published online: 08 January 2021

## References

- Damron, T. A., Ward, W. G. & Stewart, A. Osteosarcoma, chondrosarcoma, and Ewing's sarcoma: national cancer data base report. *Clin. Orthop. Relat. Res.* **459**, 40–47. <https://doi.org/10.1097/BLO.0b013e318059b8c9> (2007).
- Ottaviani, G. & Jaffe, N. The epidemiology of osteosarcoma. *Cancer Treat Res.* **152**, 3–13. [https://doi.org/10.1007/978-1-4419-0284-9\\_1](https://doi.org/10.1007/978-1-4419-0284-9_1) (2009).
- Dorfman, H. D. & Czerniak, B. Bone cancers. *Cancer* **75**, 203–210. [https://doi.org/10.1002/1097-0142\(19950101\)75:1+%3c203::aid-cnrcr2820751308%3e3.0.co;2-v](https://doi.org/10.1002/1097-0142(19950101)75:1+%3c203::aid-cnrcr2820751308%3e3.0.co;2-v) (1995).
- Misaghi, A., Goldin, A., Awad, M. & Kulidjian, A. A. Osteosarcoma: a comprehensive review. *SICOT J.* **4**, 12. <https://doi.org/10.1051/sicotj/2017028> (2018).
- Kleinerman, R. A. *et al.* Risk of new cancers after radiotherapy in long-term survivors of retinoblastoma: an extended follow-up. *J. Clin. Oncol.* **23**, 2272–2279. <https://doi.org/10.1200/JCO.2005.05.054> (2005).
- Zumarraga, J. P., Baptista, A. M., Rosa, L. P., Caiero, M. T. & Camargo, O. P. Serum values of alkaline phosphatase and lactate dehydrogenase in osteosarcoma. *Acta Orthop. Bras.* **24**, 142–146. <https://doi.org/10.1590/1413-785220162403157033> (2016).

7. Gu, J. *et al.* Identification of osteosarcoma-related specific proteins in serum samples using surface-enhanced laser desorption/ionization-time-of-flight mass spectrometry. *J. Immunol. Res.* **2014**, 649075. <https://doi.org/10.1155/2014/649075> (2014).
8. Bottani, M., Banfi, G. & Lombardi, G. Circulating miRNAs as diagnostic and prognostic biomarkers in common solid tumors: focus on lung, breast, prostate cancers, and osteosarcoma. *J. Clin. Med.* **8**, 1661. <https://doi.org/10.3390/jcm8101661> (2019).
9. Raimondi, L. *et al.* Circulating biomarkers in osteosarcoma: new translational tools for diagnosis and treatment. *Oncotarget* **8**, 100831–100851. <https://doi.org/10.18632/oncotarget.19852> (2017).
10. Viera, G. M. *et al.* miRNA signatures in childhood sarcomas and their clinical implications. *Clin. Transl. Oncol.* **21**, 1583–1623. <https://doi.org/10.1007/s12094-019-02104-z> (2019).
11. Nakka, M. *et al.* Biomarker significance of plasma and tumor miR-21, miR-221, and miR-106a in osteosarcoma. *Oncotarget* **8**, 96738–96752. <https://doi.org/10.18632/oncotarget.18236> (2017).
12. Klega, K. *et al.* Detection of somatic structural variants enables quantification and characterization of circulating tumor DNA in children with solid tumors. *JCO Precis. Oncol.* <https://doi.org/10.1200/PO.17.00285> (2018).
13. Zamborsky, R., Kokavec, M., Harsanyi, S. & Danisovic, L. Identification of prognostic and predictive osteosarcoma biomarkers. *Med. Sci. (Basel)* **7**, 28. <https://doi.org/10.3390/medsci7020028> (2019).
14. De Rubis, G., Rajeev Krishnan, S. & Bebawy, M. Liquid biopsies in cancer diagnosis, monitoring, and prognosis. *Trends Pharmacol. Sci.* **40**, 172–186. <https://doi.org/10.1016/j.tips.2019.01.006> (2019).
15. Kalluri, R. The biology and function of exosomes in cancer. *J. Clin. Invest.* **126**, 1208–1215. <https://doi.org/10.1172/JCI81135> (2016).
16. Logozzi, M. *et al.* High levels of exosomes expressing CD63 and caveolin-1 in plasma of melanoma patients. *PLoS ONE* **4**, e5219. <https://doi.org/10.1371/journal.pone.0005219> (2009).
17. O'Brien, K. *et al.* Exosomes from triple-negative breast cancer cells can transfer phenotypic traits representing their cells of origin to secondary cells. *Eur. J. Cancer* **49**, 1845–1859. <https://doi.org/10.1016/j.ejca.2013.01.017> (2013).
18. Kahlert, C. & Kalluri, R. Exosomes in tumor microenvironment influence cancer progression and metastasis. *J. Mol. Med. (Berl.)* **91**, 431–437. <https://doi.org/10.1007/s00109-013-1020-6> (2013).
19. Jalalian, S. H., Ramezani, M., Jalalian, S. A., Abnous, K. & Taghdisi, S. M. Exosomes, new biomarkers in early cancer detection. *Anal. Biochem.* **571**, 1–13. <https://doi.org/10.1016/j.ab.2019.02.013> (2019).
20. Lotvall, J. *et al.* Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. *J. Extracell. Vesicles* **3**, 26913. <https://doi.org/10.3402/jev.v3.26913> (2014).
21. Cocucci, E. & Meldolesi, J. Ectosomes and exosomes: shedding the confusion between extracellular vesicles. *Trends Cell Biol.* **25**, 364–372. <https://doi.org/10.1016/j.tcb.2015.01.004> (2015).
22. Ludwig, A. K. & Giebel, B. Exosomes: small vesicles participating in intercellular communication. *Int. J. Biochem. Cell Biol.* **44**, 11–15. <https://doi.org/10.1016/j.biocel.2011.10.005> (2012).
23. Valadi, H. *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* **9**, 654–659. <https://doi.org/10.1038/ncb1596> (2007).
24. Haraszti, R. A. *et al.* High-resolution proteomic and lipidomic analysis of exosomes and microvesicles from different cell sources. *J. Extracell. Vesicles* **5**, 32570. <https://doi.org/10.3402/jev.v5.32570> (2016).
25. Corrado, C. *et al.* Exosomes as intercellular signaling organelles involved in health and disease: basic science and clinical applications. *Int. J. Mol. Sci.* **14**, 5338–5366. <https://doi.org/10.3390/ijms14035338> (2013).
26. Whiteside, T. L. Tumor-derived exosomes and their role in cancer progression. *Adv. Clin. Chem.* **74**, 103–141. <https://doi.org/10.1016/bs.acc.2015.12.005> (2016).
27. Balaj, L. *et al.* Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat. Commun.* **2**, 180. <https://doi.org/10.1038/ncomms1180> (2011).
28. Ho, X. D. *et al.* Analysis of the expression of repetitive DNA elements in osteosarcoma. *Front. Genet.* **8**, 193. <https://doi.org/10.3389/fgene.2017.00193> (2017).
29. Lobb, R. J. *et al.* Optimized exosome isolation protocol for cell culture supernatant and human plasma. *J. Extracell. Vesicles* **4**, 27031. <https://doi.org/10.3402/jev.v4.27031> (2015).
30. Tang, Y. T. *et al.* Comparison of isolation methods of exosomes and exosomal RNA from cell culture medium and serum. *Int. J. Mol. Med.* **40**, 834–844. <https://doi.org/10.3892/ijmm.2017.3080> (2017).
31. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8> (2014).
32. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **4**, 10. <https://doi.org/10.1002/0471250953.bi0410s25> (2009).
33. Slotkin, R. K. The case for not masking away repetitive DNA. *Mob. DNA* **9**, 15. <https://doi.org/10.1186/s13100-018-0120-9> (2018).
34. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TETranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593–3599. <https://doi.org/10.1093/bioinformatics/btv422> (2015).
35. Church, D. M. *et al.* Extending reference assembly models. *Genome Biol.* **16**, 13. <https://doi.org/10.1186/s13059-015-0587-3> (2015).
36. Rider, M. A., Hurwitz, S. N. & Meckes, D. G. Jr. ExtraPEG: a polyethylene glycol-based method for enrichment of extracellular vesicles. *Sci. Rep.* **6**, 23978. <https://doi.org/10.1038/srep23978> (2016).
37. Hall, L. L. *et al.* Demethylated HSATII DNA and HSATII RNA foci sequester PRC1 and MeCP2 into cancer-specific nuclear bodies. *Cell Rep.* **18**, 2943–2956. <https://doi.org/10.1016/j.celrep.2017.02.072> (2017).
38. Fuchs, R. T., Sun, Z., Zhuang, F. & Robb, G. B. Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS ONE* **10**, e0126049. <https://doi.org/10.1371/journal.pone.0126049> (2015).
39. Ausubel, M. *et al.* Current protocols in molecular biology. Volumes 1 and 2. John Wiley & Sons, Inc., Media, PA, 1988. *Mol. Reprod. Dev.* **1**, 146–146. <https://doi.org/10.1002/mrd.1080010210> (1988).
40. Nordin, J. Z. *et al.* Ultrafiltration with size-exclusion liquid chromatography for high yield isolation of extracellular vesicles preserving intact biophysical and functional properties. *Nanomedicine* **11**, 879–883. <https://doi.org/10.1016/j.nano.2015.01.003> (2015).
41. Taylor, D. D. & Shah, S. Methods of isolating extracellular vesicles impact down-stream analyses of their cargoes. *Methods* **87**, 3–10. <https://doi.org/10.1016/j.ymeth.2015.02.019> (2015).
42. Andreu, Z. *et al.* Comparative analysis of EV isolation procedures for miRNAs detection in serum samples. *J. Extracell. Vesicles* **5**, 31655. <https://doi.org/10.3402/jev.v5.31655> (2016).
43. Konoshenko, M. Y., Lekchnov, E. A., Vlassov, A. V. & Laktionov, P. P. Isolation of extracellular vesicles: general methodologies and latest trends. *Biomed. Res. Int.* **2018**, 8545347. <https://doi.org/10.1155/2018/8545347> (2018).
44. Greening, D. W., Xu, R., Ji, H., Tauro, B. J. & Simpson, R. J. A protocol for exosome isolation and characterization: evaluation of ultracentrifugation, density-gradient separation, and immunoaffinity capture methods. *Methods Mol. Biol.* **1295**, 179–209. [https://doi.org/10.1007/978-1-4939-2550-6\\_15](https://doi.org/10.1007/978-1-4939-2550-6_15) (2015).
45. Daaboul, G. G. *et al.* Digital detection of exosomes by interferometric imaging. *Sci. Rep.* **6**, 37246. <https://doi.org/10.1038/srep37246> (2016).
46. Bachurski, D. *et al.* Extracellular vesicle measurements with nanoparticle tracking analysis—an accuracy and repeatability comparison between NanoSight NS300 and ZetaView. *J. Extracell. Vesicles* **8**, 1596016. <https://doi.org/10.1080/20013078.2019.1596016> (2019).

47. Higgins, N. P., Geballe, A. P. & Cozzarelli, N. R. Addition of oligonucleotides to the 5'-terminus of DNA by T4 RNA ligase. *Nucl. Acids Res.* **6**, 1013–1024. <https://doi.org/10.1093/nar/6.3.1013> (1979).
48. de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**, e1002384. <https://doi.org/10.1371/journal.pgen.1002384> (2011).
49. Bronkhorst, A. J. *et al.* Sequence analysis of cell-free DNA derived from cultured human bone osteosarcoma (143B) cells. *Tumour Biol.* **40**, 1010428318801190. <https://doi.org/10.1177/1010428318801190> (2018).
50. Montermini, L. *et al.* Inhibition of oncogenic epidermal growth factor receptor kinase triggers release of exosome-like extracellular vesicles and impacts their phosphoprotein and DNA content. *J. Biol. Chem.* **290**, 24534–24546. <https://doi.org/10.1074/jbc.M115.679217> (2015).
51. Zhang, H. *et al.* Identification of distinct nanoparticles and subsets of extracellular vesicles by asymmetric flow field-flow fractionation. *Nat. Cell Biol.* **20**, 332–343. <https://doi.org/10.1038/s41556-018-0040-4> (2018).
52. Takahashi, A. *et al.* Exosomes maintain cellular homeostasis by excreting harmful DNA from cells. *Nat. Commun.* **8**, 15287. <https://doi.org/10.1038/ncomms15287> (2017).
53. Jepsen, D. K. *et al.* Reassessment of exosome composition. *Cell* **177**, 428–445. <https://doi.org/10.1016/j.cell.2019.02.029> (2019).
54. Ouyang, L. *et al.* A three-plasma miRNA signature serves as novel biomarkers for osteosarcoma. *Med. Oncol.* **30**, 340. <https://doi.org/10.1007/s12032-012-0340-7> (2013).
55. Yuan, J., Chen, L., Chen, X., Sun, W. & Zhou, X. Identification of serum microRNA-21 as a biomarker for chemosensitivity and prognosis in human osteosarcoma. *J. Int. Med. Res.* **40**, 2090–2097. <https://doi.org/10.1177/030006051204000606> (2012).
56. Yang, Z. *et al.* Serum microRNA-221 functions as a potential diagnostic and prognostic marker for patients with osteosarcoma. *Biomed. Pharmacother.* **75**, 153–158. <https://doi.org/10.1016/j.biopha.2015.07.018> (2015).
57. Eymery, A. *et al.* A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. *Nucl. Acids Res.* **37**, 6340–6354. <https://doi.org/10.1093/nar/gkp639> (2009).
58. Ting, D. T. *et al.* Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* **331**, 593–596. <https://doi.org/10.1126/science.1200801> (2011).
59. Bersani, F. *et al.* Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. *Proc. Natl. Acad. Sci. USA* **112**, 15148–15153. <https://doi.org/10.1073/pnas.1518008112> (2015).
60. Contreras-Galindo, R. *et al.* Human endogenous retrovirus K (HML-2) elements in the plasma of people with lymphoma and breast cancer. *J. Virol.* **82**, 9329–9336. <https://doi.org/10.1128/JVI.00646-08> (2008).
61. Criscione, S. W., Zhang, Y., Thompson, W., Sedivy, J. M. & Neretti, N. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genom.* **15**, 583. <https://doi.org/10.1186/1471-2164-15-583> (2014).
62. Babcock, M., Yatsenko, S., Stankiewicz, P., Lupski, J. R. & Morrow, B. E. AT-rich repeats associated with chromosome 22q11.2 rearrangement disorders shape human genome architecture on Yq12. *Genome Res.* **17**, 451–460. <https://doi.org/10.1101/gr.56515.07> (2007).

## Acknowledgements

This work was performed with the support of the Extracellular Vesicle Core and the SC2 Single Cell, Sequencing, and CyTOF Core at The Saban Research Institute of Children's Hospital Los Angeles. We thank Paolo Neviani, Petra Wise, and Muller Fabbri for assistance and advice with EV analyses, Kenny Gouin for bioinformatics advice, Narine Harutyunyan, Mohammed Elgendy, Wen Xie, and Xiaofei Qin for sample acquisition assistance, Sunhye Lee for initial RNA-Seq validation efforts, and Xiaowu Gai for helpful discussions and critical reading of the manuscript. This work was supported by an Innovation Grant from the Alex's Lemonade Stand Foundation, the Major Project of TCM research in Henan Province (No. 2018ZYZD01), the Major Project of Science and Technology in Henan Province (No.192102310437), Retinoblastoma International, and The Larry and Celia Moh Foundation.

## Author contributions

L.C. established hypotheses, designed studies, performed experiments, analyzed data, and wrote the paper. K.S. managed samples and performed RNA-seq and repetitive element analyses. M.T. performed RNA-seq analyses. R.J., J.L., M.Y., W.L., and J.Z. accrued, annotated, and provided samples. D.C. established hypotheses, designed studies, accrued samples, co-analyzed data, and co-wrote the paper.

## Competing interests

D.C. and L.C. are inventors on a U.S. provisional patent, Children's Hospital Los Angeles as applicant, relating to REPETITIVE ELEMENT DNAS.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-77398-z>.

**Correspondence** and requests for materials should be addressed to J.L. or D.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



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

© The Author(s) 2021