# High-Throughput Screening of Alternative Micro-Metastasis-Specific Gene Predictors of Circulating Osteosarcoma Cells

**Pattaralawan Sittiju**
Chiang Mai University Faculty of Medicine
**Parunya Chaiyawat**
Chiang Mai University Faculty of Medicine
**Dumnoensun Pruksakorn**
Chiang Mai University Faculty of Medicine
**Jeerawan Klangjorhor**
Chiang Mai University Faculty of Medicine
**Weerinrada Wongrin**
Chiang Mai University Faculty of Science
**Phichayut Phinyo**
Chiang Mai University Faculty of Medicine
**Rawikant Kamolphiwong**
Prince of Songkla University Faculty of Medicine
**Areerak Phanphaisarn**
Chiang Mai University Faculty of Medicine
**Pimpisa Teeyakasem**
Chiang Mai University Faculty of Medicine
**Prachya Kongtawelert**
Chiang Mai University Faculty of Medicine
**Peraphan Pothacharoen** ( peraphan.pothacharoen@gmail.com )
Chiang Mai University Faculty of Medicine https://orcid.org/0000-0002-8560-553X

**Research**

**Keywords:** Osteosarcoma, Circulating tumor cells, Biomarker, mRNA expression, qRT-PCR

**Posted Date:** May 17th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-513080/v1

## Abstract

# Background

Current techniques to identify circulating-tumor cells (CTCs) in osteosarcoma (OS), which are an indication of a poor prognosis in cases of intermediate levels of metastasis, are complicated and time-consuming. This study investigated the efficacy of quantitative reverse transcription PCR (qRT-PCR), a molecular technique that is available in most laboratories, for detection of CTCs in buffy coat samples of OS patients and healthy donors.

# Methods

Previously published reports on data-reviewing and retrieval of data by calculation of differential gene expression from the Gene Expression Omnibus (GEO) database repository were reviewed identify candidate genes. Following analysis of the expression of the candidate genes identified a diagnostic model for detection of specific gene expression was derived using binary logistic regression with a multivariable fractional polynomial (MFP) algorithm.

# Results

A model incorporating *VIM*, *ezrin*, *COL1A2*, and *PLS3* exhibited an outstanding discriminative ability as determined by the receiver operating characteristic curve (AUC = 0.9896, 95%CI 0.9695, 1.000). At the probability cut-off value 0.2943, the sensitivity and the specificity of the model for detection of OS were 100% (95%CI 94.8, 100.0) and 96.49% (95%CI 87.9, 99.6), respectively.

# Conclusion

The qRT-PCR can identify the existence of OS circulating cells by detection of potential candidate genes (*VIM, Ezrin, COL1A2* and *PLS3*). Thus, these genes are worthy to be considered diagnostic biomarkers and alternative micro-metastasis predictors for OS.

## Background

Osteosarcoma (OS), although relatively rare, is the most common primary malignancy of bone, with a worldwide incidence of 3.4 per million people per year[1]. OS is found to occur predominantly in the second decade of life and in the [1, 2]. Forty percent of OS patients whose tumor is found to have spread from the primary site to secondary sites during or after diagnosis have a poor response to treatment and poor recovery even when combination therapies are employed[3].

For metastatic OS, magnetic resonance imaging (MRI), computed tomography (CT) scans and positron emission tomography (PET) scans are the standard methods for bone metastasis lesion diagnosis and follow-up. PET-CT is more sensitive for bone metastasis detection than scintigraphy which is currently the standard method[4, 5]. Even so, pulmonary nodules smaller than 5–9 mm are still undetectable by PET/PET-CT[1, 6, 7]. The sensitivity for evaluation of bone metastasis is increased when scintigraphy is combined with PET-CT[5].

Liquid biopsy is an alternative technique for predicting metastasis which represents a promising approach for diagnostic, prognostic, and personalized therapeutic purposes. Among liquid biopsy biomarkers, circulating tumor cells (CTCs) represent a promising avenue for identifying cancer metastasis. The challenge for CTC detection is the low incidence in circulating blood, about 1–10 cells per $10^6$ white blood cells[8]. There are several clinically significant CTC separation techniques which not only enrich the CTCs but are also able to purify and identify CTCs, e.g., immunomagnetic enrichment and microfluidic immunocapture. Owing to their novelty and the fact that the methods are complicated, these techniques are not widely used[9]. Cancer-specific mRNA analysis is one promising approach for tracing cancer cells in blood, but specific mRNA markers for OS are not widely established[10, 11]. Candidate gene tumor markers have generally been obtained from and identified in the by-products of previous studies, thus they might not represent robust markers for some cancers. Comparative expression analysis using information from biodata resources is a new pregenital approach for identification tumor specific makers. Among sources of bioinformatic data, the Gene Expression Omnibus (GEO) has been widely adopted for identifying tumor-specific genes[12, 13].

To predict the micro-metastasis of OS, we proposed a simple and inexpensive method to identify novel OS-specific genes using comparative gene expression analyses from a gene expression database and to determine the candidate gene expression in OS cell lines by quantitative reverse transcription PCR (qRT-PCR). Expression of the candidate genes was demonstrated in the buffy coat of 73 OS patients and 79 healthy donors. The analytical model in this study used *VIM* (vimentin-encoding gene), *ezrin*, *COL1A2*, and *PLS3* for OS diagnosis and for metastasis prediction markers using a simple CTC detection method which provides both high efficacy and reliability.

# Methods

# Patients

Ethylenediamine tetraacetic acid (EDTA) whole blood and tumor tissue samples had were retrospectively collected from 62 stage IIB and 11 stage III patients obtained during diagnostic procedures conducted between 2012 and 2020 at Maharaj Nakorn Chiang Mai Hospital. Residual anonymous EDTA buffy coat (500 µl) was also drawn from 79 healthy individuals during donor screening procedures and the residual buffy coat (120 ml) was obtained during blood component preparation by the Blood Bank Section of Maharaj Nakorn Chiang Mai Hospital. All blood and tissue samples were collected after receipt of

approval by the Research Ethics Committee Faculty of Medicine Chiang Mai University (ORT-2557-02717 and ORT-2562-06549). The overview of the study and the workflow of methods is shown in Fig. 1.

# Bioinformatic analysis

Affymetrix HG-U133Plus2.0 DNA microarray (Platform GPL570) of OS cell lines (GSE70414, GSE30807, GSE37552, GSE18947, GSE16089, GSE7454, GSE41828, GSE46493, GSE41445, and GSE55957), primary OS cells (GSE85537) and whole blood of healthy people (GSE93272) which represented buffy coat cell composition were retrieved from GEO. All data was from cells not treated with any agent or vector. Accession codes are given in Table S3. The gene expression analysis to identify candidate genes compared 1) OS cell lines with healthy whole blood samples and 2) primary OS cells with healthy whole blood samples.

The robust multi-array average (RMA) algorithm through a custom brainarray chip description file (CDF, ENTREZG, V19) was used to calculate the quantile normalization background adjustment and summarized as previously described[12]. For investigation of differential gene expression, *P* values were calculated with Linear Models for Microarray (limma) data in R. *P* values with a $\log_2$ Expression Ratio (ER) greater than 2 was a set as the cut-off for initial selection of candidate genes[14].

**Sample preparation.**

**Healthy peripheral blood mononuclear cells (PBMCs)**:

Buffy coat samples from the blood component separation process were diluted with phosphate buffered saline (1:1). Gradient centrifugation using Lymphoprep™ (STEMCELL Technologies, Canada) was employed for PBMCs isolation. The PBMCs were collected and counted under a light microscope with a hemocytometer.

**Buffy coat**: EDTA whole blood specimens were centrifuged at 1,600g for 15 minutes, the buffy coat layer between packed red blood cells and plasma was collected and stored at -80 $^{o}$C. The cryopreserved buffy coat samples from OS patients and healthy donors were lysed with lysis buffer RA1, (Macherey Nagel, Düren, Germany).

**Cell lines and primary cells**

Human OS cell lines HOS-MNNG and U2OS were obtained from the Cell Line Service (Eppelheim, Baden-Württemberg, Germany). HOS-143B, MG63 and Saos-2 were obtained from the American Type Culture Collection (ATCC, Manassas, VA). Primary OS cells were generated and extracted following previously described protocol[15, 16]. The MNNG-HOS cells were grown in the Roswell Park Memorial Institute (RPMI). The MG63, and primary cells were grown in Dulbecco's modified Eagle's medium (DMEM). U2OS and Saos-2 cells were cultured in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 medium (DMEM-F12). HOS-143B was cultured in DMEM-Brdu. All cell lines and primary cells were cultured in 10% (v/v) fetal bovine serum and maintained in a humidified atmosphere of 37°C with 5% $CO_2$ [15].

# Spiking assay

Spiking assay was conducted to select potential genes which can distinguish candidate gene expression between the spiked samples and non-spiked samples in a minimum number of OS cells. The spiking assay was performed by spiking various numbers of each cell line (0-$10^4$ cells) into normal PBMCs ($1.25x10^5$-$2x10^6$ cells) and assessing candidate gene expression by qRT-PCR.

# qRT-PCR

Total RNA was extracted with an illustra RNAspin Mini Kit (GE Healthcare Europe GmbH, Freiburg, Germany) and 20 µg of cDNAs were generated by iScriptTM (Bio-Rad, Hercules, CA, USA). The PCR reactions were performed with an Applied Biosystems 7500/7500 Fast Real-Time PCR system using SensiFAST™ SYBR® Lo-ROX (Bio-Rad, Hercules, CA, USA) for 45 cycles. Each cycle was performed as follows: 5 seconds at 95°C, 10 seconds at 60°C and then 30 seconds at 72°C. The RNA levels of *CD45* and candidate genes (*COL1A2, GJA1, PLS3, COL5A2, COL3A1, CDR1, COL2A1, EGR1, Ezrin* and *VIM*) from both bioinformatic analysis and previous publications were normalized with beta-actin (ACTB) as a housekeeping gene using the 2 (-ΔΔC (T)) method. Primer sequences are listed in Table S1.

# Statistical analysis

Statistical analysis of candidate genes relative to expression was performed using SPSS 22.0 (IBM Corp., Armonk, NY, USA), Stata 16 (StataCorp, College station, Texas, USA), and Prism 8.4.3 (GraphPad, La Jolla, CA, USA). The data are shown as mean ± standard deviation (SD). The significance of difference between means of OS and healthy donors was determined using the Mann-Whitney U-Test for ordinal or continuous data which is not normally distributed. *P* values less than 0.05 were considered statistically significant.

Candidate genes were selected by a retrospective study. Non-parametric regression and multivariable modeling were constructed with both OS patients and healthy donors using fractional polynomials. We explored the shape of the association between relative gene expression and log odds of osteosarcoma using locally weighted scatter plot smoothing (LOWESS) and fractional polynomial plots. A diagnostic model for prediction of OS and OS metastasis was derived using binary logistic regression with a multivariable fractional polynomial (MFP) algorithm for fitting of continuous determinants based on the actual shape of their association with the predicted endpoints[17, 18]. The *p*-value cut-off was set at 0.2 to exclude gene expression with non-significant contribution from the equation model. Model discriminative ability was measured as the area under the receiver operating characteristic curve (ROC). Predicted probabilities of OS and OS metastasis were calculated using the model. Cut-off points for diagnosis of OS and OS metastasis were established based on sensitivity and specificity.

# Results

# Identification of circulating osteosarcoma cell specific candidate genes using bioinformatic data analysis.

To identify OS-specific genes which express in blood circulating cells, publicly available microarray gene expression datasets from GEO of OS cell lines (n = 29), primary OS cells (n = 3) and healthy whole blood samples (n = 36) representing OS circulating cells and blood cells in buffy coat samples were selected for the bioinformatic analysis. Datasets of samples which had been administered any agent vectors were excluded. Quantile normalization background adjustment and summarization were calculated using a robust multi-array average (RMA) algorithm and a custom brainarray chip description file (CDF, ENTREZG, V19). As described earlier, probe sets of genes and adjusted *P* values were calculated using a limma package availably in R to compare gene expression.

From the GEO data, there were 20,188 and 20,186 genes which had been reported in OS cell lines and primary OS cells, respectively. After calculation, we found significant up-regulation of 1,426 and 1,899 genes in OS cell lines and primary OS cells, respectively ($p$ value < 0.001) when compared to the healthy whole blood cells with $\log_2$ expression ratio (ER) > 2 (Fig. 2A and 2B, respectively). These sets of genes were considered to be upregulating genes. Among the upregulating genes, only the eight genes, *COL1A2, GJA1, PLS3, COL5A2, COL3A1, CDR1, COL2A1*, and *EGR1*, which presented a 500-fold change of expression in OS cell lines or primary OS cells compared to healthy whole blood samples were considered as novel OS-specific genes (Fig. S1).

# Candidate gene selection

To select potential candidate genes which could distinguish the difference between the spiked sample and non-spiked sample in a minimum number of OS cells, the OS cells were spiked into normal PBMCs. The qRT-PCR was performed to evaluate the expression of the novel OS-specific genes in all spiked samples. To determine the cycle threshold (Ct) value of each candidate (Y axis) and *CD45*, a common leukocyte antigen, (X axis) were converted to $\log_2$-Ct and plotted on quantification curves. The genes which were able to distinguish between samples with and without OS were designated as potential candidate genes.

The qRT-PCR analysis provided fifty relative curves of candidate gene expression (Y-axis) and *CD45* expression (X-axis) from the set of OS cells (HOS-MNNG, U2OS, HOS-143B, MG63, and Saos-2) spiked PBMC [Fig. S2-6].

The 50 curve patterns were categorized into three groups: 1) completely separated, 2) partially separated, and 3) non-separated patterns. The genes which exhibited a completely or partially separated pattern were considered candidate genes. The genes which exhibited a non-separation pattern were considered to be unacceptable genes. As shown in Table S2, the candidate genes were *EGR1, PLS3, VIM, COL1A2, COL3A1, COL5A2* and *Ezrin* and the unacceptable genes were *COL2A1*, *CDR1 and GJA1*.

# Evaluation of the OS diagnostic potential of the genes

The samples from OS patients (n = 73) and healthy donors (n = 79) (clinical characteristics shown in Table 1) were evaluated for expression of the candidate genes using qRT-PCR; the Ct values of each of the candidate genes were normalized with *CD45* gene expression. The analysis demonstrated that the relative expression of *EGR1 (*P = 0.0010), *PLS3* (*P* = 0.0038), *and VIM* (*P* < 0.0001) in OS patients was significantly higher than in healthy donors (Fig. 3A). Even though *COL1A2, COL3A1, COL5A2* and *Ezrin* showed no significant difference between OS patients and healthy donors overall, these four genes did present a stronger difference of expression in patients and donors who were younger than 25 years old with borderline statistical significance (*P* values of *COL3A1, COL5A2 and Ezrin* were 0.0926, 0.0669 and 0.0628, respectively) with the exception of *COL1A2* (*P* = 0.1707) (Fig. 3A-B).

Table 1
Clinical characteristics of OS patients and healthy donors

| **Parameters** | | **OS patients (n = 73)** | **Healthy donors (n = 79)** |
|---|---|---|---|
| **Median age [range]** | Childhoods and adolescents | 14 [5–24], (71.2%) | 22 [18–24], (46.8%) |
| | Adults | 57 [25–75], (28.8) | 42 [25–55], (53.2%) |
| **Gender** | Male | 36 (49.3%) | 52 (65.8%) |
| | Female | 37 (50.7%) | 37 (34.2%) |
| **Enneking stage** | IIB | 62 (84.9%) | - |
| | III | 11 (15.1%) | - |
| **Tumor location** | Femur | 35 (47.9%) | - |
| | Tibia | 15 (20.5%) | - |
| | Other | 23 (31.5%) | - |
| **Metastasis** | Bone | 3 (4.1%) | - |
| | Lung | 6 (8.2%) | - |
| | Bone and Lung | 2 (2.7%) | - |
| | None | 62 (84.9%) | |

OS, osteosarcoma.

Table 2
Multivariable fractional polynomial logistic regression model for CTCs detection of osteosarcoma

| Candidate gene | Covariate transformation | | β | 95% CI | *P* |
|---|---|---|---|---|---|
| | df | FP term after MFP transformation | | | |
| *Intercept* | - | - | 3.611 | 1.303, 5.919 | - |
| *COL1A2* | 4 | $(COL1A2 + 0.0099997520446777)^{-1-27.29695355}$ | 0.020 | 1.006, 1.035 | 0.005 |
| *COL1A2* | 4 | $(COL1A2 + 0.0099997520446777)^{-1} \times \ln(COL1A2) + 90.26488644$ | 0.002 | 1.001, 1.004 | 0.005 |
| *PLS3* | 1 | *PLS3*-22.18452404 | 0.068 | 0.988, 1.161 | 0.097 |
| *Ezrin* | 2 | ln (*Ezrin*/1000) + 1.766007255 | -1.552 | 0.080, 0.563 | 0.002 |
| *VIM* | 2 | ln (*VIM*/10000) + 1.474027931 | 2.225 | 2.996, 28.602 | < 0.0001 |

*COL1A2*, Collagen Type I Alpha 2 Chain; *PLS3*, Plastin-3; *VIM*, Vimentin; df, degree of freedom; CI, confidence interval.

The association between the relative expression of each candidate gene and the log odds of OS was non-linear (Fig. S7). In the MFP algorithm, *COL1A2, PLS3, Ezrin*, and *VIM*, which showed significant contribution to the model, were included in the diagnostic model, while *COL3A1*, *COL5A2* and *EGR1* were excluded as their *P* values were less than 0.2 (data not shown). Following to the diagnostic model, the probability of OS was calculated as follows:

$$\text{Probability of Osteosarcoma} = \frac{exp^{lp}}{1+exp^{lp}} \qquad (3.3.)$$

where *lp* is the linear predictor yielded from the formula:

$$\begin{aligned} linear\ predictor\ (lp) = constant &+ 2.83\ (COL1A2\ FP\ term) + 2.81\ (COL1A2\ FP\ term) + \\ &1.66\ (PLS3\ FP\ term) + -3.11\ (Ezrin\ FP\ term) + \\ &3.87\ (vimentin\ FP\ term) \end{aligned} \qquad (3.3.)$$

Each candidate gene term is referenced in Fig. 2.

The ROC curve analysis was performed on the expression of *COL1A2, PLS3, Ezrin* and *VIM* to examine the diagnostic performance of the model for identifying OS in samples from healthy donors. All OS samples were positive with two false positives in the healthy donor samples at the probability cutoff of 0.2943. The sensitivity was 100% (95%CI 94.8, 100.0) and the specificity was 96.49% (95%CI 87.9, 99.6), with an area under the ROC curve of 0.9896 (95%CI: 0.9695, 1.0000) (Table 3).

Table 3
Diagnostic and metastatic prediction accuracy (n = 152)

| Clinical character (n) | | | Probability cut-off point | AUC (95% CI) | Sensitivity (95%CI) | Specificity (95%CI) | LHR+ (95%CI) |
|---|---|---|---|---|---|---|---|
| UD | Normal | OS | 0.2943 | 0.9896 | 100% | 96.49% | 28.50 |
| 26 | 57 | 69 | | (0.9695, 1.0000) | (94.8, 100.0) | (87.9, 99.6) | (7.30, 111.19) |
| UD | Non-metastasis (Normal and IIB) | Metastasis (III) | 0.8243 | 0.7257 | 100% | 54.78% | 2.2115 |
| | | | | (0.621, 0.8302) | (71.5, 100.0) | (45.2, 64.1) | (1.81, 2.70) |
| 26 | 115 | 11 | | | | | |

OS, osteosarcoma; UD, undetectable; AUC, area under the ROC curve; CI, confidence interval; LHR+, positive likelihood ratio; LHR-, negative likelihood ratio.

Based on the derived MFP model, we further evaluated the ability of *COL1A2, PLS3, Ezrin* and *VIM* to predict metastatic OS. All stage III OS samples were positive at the cut-off point 0.8243 with 100% sensitivitiy (95%CI 71.5, 100.0), 54.78% specificitity (95%CI 45.2, 64.1) and 0.7257 area under the ROC curve (95%CI 0.6212, 0.8302) (Table 3).

**Expression of candidate genes from bioinformatic data and previous studies in OS cell lines and primary cells compared to heathy donor PBMCs.**

In the bioinformatic analysis, the obtained gene expression data was from OS cell lines and primary OS cells which did not originate directly from circulating OS cells. To further explore whether the four candidate genes (*COL1A2, PLS3, Ezrin and VIM*) were highly specific to clinical OS tumor origin, we analyzed mRNA expression of candidate genes in primary osteosarcomas from OS patients (n = 24) and compared them to PBMC from healthy donors (n = 3) by qRT-PCR. The results showed that the expression of *COL1A2*, *PLS3* and *VIM* in the primary cells was significantly higher than in normal PBMCs while *Ezrin* expression was non-significantly different ($P < 0.05$) (Fig. S8).

# Discussion

The CTCs are invasive circulating metastatic cells migrating from primary or metastatic sites of tumors. These cells, which survive in circulating blood, function as metastasis precursors and may also colonize, forming secondary tumors and causing refractory and recurrent cancer[19–21]. Thus, any molecular predictors which demonstrate a high specificity to OS and a strong correlation to metastasis might be applied as CTC detection tools.

In this study, the differences in gene expression between OS cell lines or primary OS cells *vs* healthy donor cells identified significant upregulation of 8 novel OS-specific genes (*COL1A2*, *GJA1*, *PLS3*, *COL5A2*, *COL3A1*, *CDR1*, *COL2A1*, and *EGR1*). Most of them, including *COL1A2, GJA1, COL5A2, EGR1* and

*COL3A*1, have been previously reported as upregulated genes or the translated proteins of those genes associated with OS progression in OS tissues when compared to normal tissues[22–26]. In addition, high expression of *COL2A1*, *CDR1* and *PLS3* has also been found to be related to tumor progression in several types of tumors[27–31].

The previously reported OS markers *Ezrin* and *VIM* were also analyzed for gene expression in retrospective samples. Ezrin, a cross-linker protein, plays an essential role in many metastatic phenotypes of cancers including pediatric sarcomas, OS and rhabdomyosarcoma[32]. The expression level of Ezrin was high in OS circulating cells, especially in OS metastatic stage III in the Enneking staging system[11]. Vimentin, a mesenchymal marker, has recently been reported to be an indicator of epithelial-to-mesenchymal transition (EMT) associated with migration and metastasis in various cancers as well[33, 34]. Vimentin has also been reported to be highly expressed in human OS tumor tissue[35].

The CTC enrichment process is necessary to discard blood components which might [36]. There are several techniques to enrich CTCs from fresh whole blood including Ficoll-Hypaque density gradient centrifugation, filter-based methods, magnetic bead based CD45 negative and vimentin positive selection[10, 11, 37] which need fresh whole blood or an affinity column for isolating CTCs. The samples for gene expression analysis in our study were enriched by buffy coat preparation. The total RNA was extracted from the − 80$^{o}$C frozen buffy coats without preservatives since the frozen cells would otherwise be lysed by ice crystallization after thawing leading to injury[38]. To avoid losing CTC-total RNAs and to reduce the number of enrichment steps, total RNA from all buffy coats was extracted immediately after thawing. Other genetic components which are not released by CTCs, especially leukocyte RNAs, were main interfering factors which were not able to be discarded in this study. Previous publications have demonstrated that leucocyte common antigen (CD45) expression level is related to a number of leukocytes[39, 40]. Accordingly, the gene expression level of each OS cell line spiked PBMC sample was presented as a relative quantification curve between OS-specific gene and *CD45*.

We expected a correlative relationship between the expression of OS-specific genes and the number of OS cells, i.e., an expression pattern of the relative quantification curve of the specific OS genes (n = 8) and the two evaluated metastatic genes. The results demonstrated that only seven genes (*COL1A2, COL3A1, COL5A2, EGR1, PLS3, Ezrin* and *VIM*) out of the 10 genes tested showed a completely or partially separated pattern for at least one of the 5 OS cell lines (Table S1 and Fig. S2-6). *CDR1, COL2A1* and *GJA1* were not highly expressed among the five OS cells (SaOS-2, MNNG, MG63, U2OS and 143B) resulting in interference when spiked OS cell lines had high numbers of PBMCs, causing the level of expression of those genes to be an unreliable indicator.

Measurement of candidate genes expression was also performed in clinical samples, including OS and healthy buffy coats. The expression of *EGR1*, *PLS3* and *VIM* showed statistically significant differences between the two groups ($P < 0.05$) but not *COL1A2*, *COL3A1*, *Ezrin* and *COL5A2* (Fig. 3A). Not surprisingly, OS exhibits a high heterogeneity and complexity of genomic and expression level between patients[41].

We further narrowed the samples to patients and donors age younger than 25, then reanalyzed the expression of seven candidate genes. The results indicated that for *COL1A2*, *COL3A1*, *Ezrin* and *COL5A2*, all gene expression differences between OS patients and donors were noticeably increased while the differences for *COL3A1*, *Ezrin* and *COL5A2* bordered on statistical significance ($P < 0.1$).

The efficiency of each candidate gene in the prediction of OS was evaluated with binary logistic regression with the MFP algorithm using relative expression data. The model that included *VIM, Ezrin, COL1A2* and *PLS3* performed the best in terms of ability to discriminate OS samples from healthy donor samples. We identified the OS probability cut-off point at 0.2943. In further clinical studies, patients with a higher possibility of OS than the cut-off point will require confirmation of results with standard diagnostic tests. Patients who present with an OS probability higher than 0.8243 are suspect for metastasis occurrence; in those cases, follow-up with high sensitivity micro-metastasis tests such as bone scans or PET scans is appropriate. In this study, the model identified all OS patients as positive; there were two false positive samples from normal buffy coats at a probability cut-off value of 0.2943. The same model also exhibited the ability to predict OS metastasis at a probability cut-off value of 0.8243 with all positive results in metastatic OS (III) samples and some false positive results in non-metastatic OS and normal samples. We suggest that patients with positive results should be informed and that metastasis should be confirmed using other clinical tools.

The origin of OS is known to be mesenchymal stem cells. Most OS is malignancy developing from osteoblast cells with genetic and epigenetic mutation accumulation[42]. Due to the limited number of samples, it was not possible to isolate circulating OS cells from frozen buffy coat. To confirm whether candidate genes were specific to OS cells, the expression of *COL1A2*, *EGR1*, *PLS3*, *Ezrin* and *VIM* was evaluated in primary OS cells with qRT-PCR and compared to PBMCs from healthy donors. Among the four candidate genes, the expression of *COL1A2*, *PLS3* and *VIM* in OS cells was significantly higher than PBMC ($P < 0.05$) but not *Ezrin* [Fig. S8].

In agreement with previous studies, our qRT-PCR results demonstrated that *VIM* normally expresses in OS during diagnosis and consistently expresses in both buffy coats and primary cells[35, 42, 43]. Expression of *COL1A2* has been found in human osteoblast lineages[44] and has been shown to be related to migration, invasion, proliferation, and metastasis promotion in various cancers including OS, with significantly higher expression in buffy coats and primary cells[23, 45, 46].

The roles of PLS3 on actin (F-actin) formation in normal bone have been previously described. Oncology studies have demonstrated that *PLS3* is a novel CTC marker for prognosis in breast and colorectal cancer[31, 47]. In our study, a significantly higher expression level of *PLS3* was found in OS patients when compared to healthy donors. Thus, *PLS3* could be a novel liquid biopsy marker for prognostic prediction in OS.

On the other hand, ezrin, which has been reported to be a typical EMT marker, was borderline significantly highly expressed in frozen OS buffy coat but not in primary OS cells. Due to the fact that there was no difference in *Ezrin* expression between OS primary cells and PBMCs, we suggest that the population of

clusters of tumor cells with the potential to become CTCs in the tumor population was low. However, *Ezrin* expression in primary cells might not be an indication of CTC clusters. High expression of *Ezrin*, both in RNA and in protein levels in OS patients, was positively correlated with metastatic stage and OS recurrence[48, 49].

Comparative expression analysis in this study did not find significant differences in *VIM* or *Ezrin* between OS primary cells and PBMCs, although they have previously been reported as OS candidate genes[11, 50]. These gene expression unrepresented in whole tumor population, this might be definitive for OS cells in circulating blood. Comparative expression analysis of genes between single circulating tumor cells and other circulating blood cells should be further investigated. This could potentially improve the detection of OS circulating cells by identifying more precise predictors.

# Conclusions

This study demonstrated the feasibility of using *VIM*, *Ezrin*, *COL1A2* and *PLS3* as potential candidate biomarkers to detect CTCs in OS diagnostic and metastatic monitoring using simple molecular techniques (qRT-PCR). This gene set identifies the existence of OS cells in circulation which could be use as diagnostic and metastatic prediction tools for OS. The analysis of the gene expression by qRT-PCR, simple and high throughput methods, is suitable for a variety of sample types including fresh, preserved, and non-enriched CTCs. Further studies should be designed to validate the use of these four predictors for monitoring disease progression, prediction of therapeutic response as well as tumor recurrence in a large sample.

# Abbreviations

OS, Osteosarcoma; CTCs, Circulating tumor cells; qRT-PCR, Real-time quantitative.

# Declarations


### Acknowledgements

The authors thank the Blood Bank Section, Maharaj Nakorn Chiang Mai Hospital for supporting all steps related to obtaining donor buffy coats.


### Authors' contributions

Peraphan Pothacharoen*: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision.

Pattaralawan Sittiju: Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration.

Parunya Chaiyawat: Conceptualization, Methodology, Writing - Review & Editing, Visualization.

Dumnoensun Pruksakorn: Conceptualization, Methodology, Resources, Writing - Review & Editing.

Jeerawan Klangjorhor: Conceptualization, Methodology, Writing - Review & Editing.

Weerinrada Wongrin: Methodology, Formal analysis.

Phichayut Phinyo: Methodology, Formal analysis, Writing - Review & Editing.

Rawikant Kamolphiwong: Methodology, Formal analysis, Investigation, Writing - Review & Editing.

Areerak Phanphaisarn: Formal analysis, Visualization, Data Curation.

Pimpisa Teeyakasem: Investigation, Resources, Data Curation.

Prachya Kongtawelert: Resources.

## Availability of data and materials

All public datasets were obtained from GEO (https://www.ncbi.nlm.nih.gov/geo)

## Ethics approval and consent to participate

This article was approved the Research Ethics Committee Faculty of Medicine Chiang Mai University (ORT-2557-02717 and ORT-2562-06549).

## Consent for publication

Written informed consent was obtained from all osteosarcoma patients.

## Competing interests

The authors declare that they have no conflicts of **interest.**

**Funding:** This research was funded by The Royal Golden Jubilee Ph.D. Program, grant No. PHD/0191/2559, a PhD Placement for Scholars 2017/18 grant, ID 4.C.CM/59/L.1.O.XX.

## Author details

[1] Thailand Excellence Center for Tissue Engineering and Stem Cells, Department of Biochemistry, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200, Thailand

[2] Musculoskeletal Science and Translational Research Center, Department of Orthopedics, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

[3] Center of Multidisciplinary Technology for Advanced Medicine (CMUTEAM), Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

[4] Department of Orthopedics, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

[5] Biomedical Engineering Institute, Chiang Mai University, Chiang Mai, Thailand

[6] Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, 50200 Thailand

[7] Center for Clinical Epidemiology and Clinical Statistics, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

[8] Department of Family Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

[9] Biomedical Sciences, Faculty of Medicine, Prince of Songkla University

# References

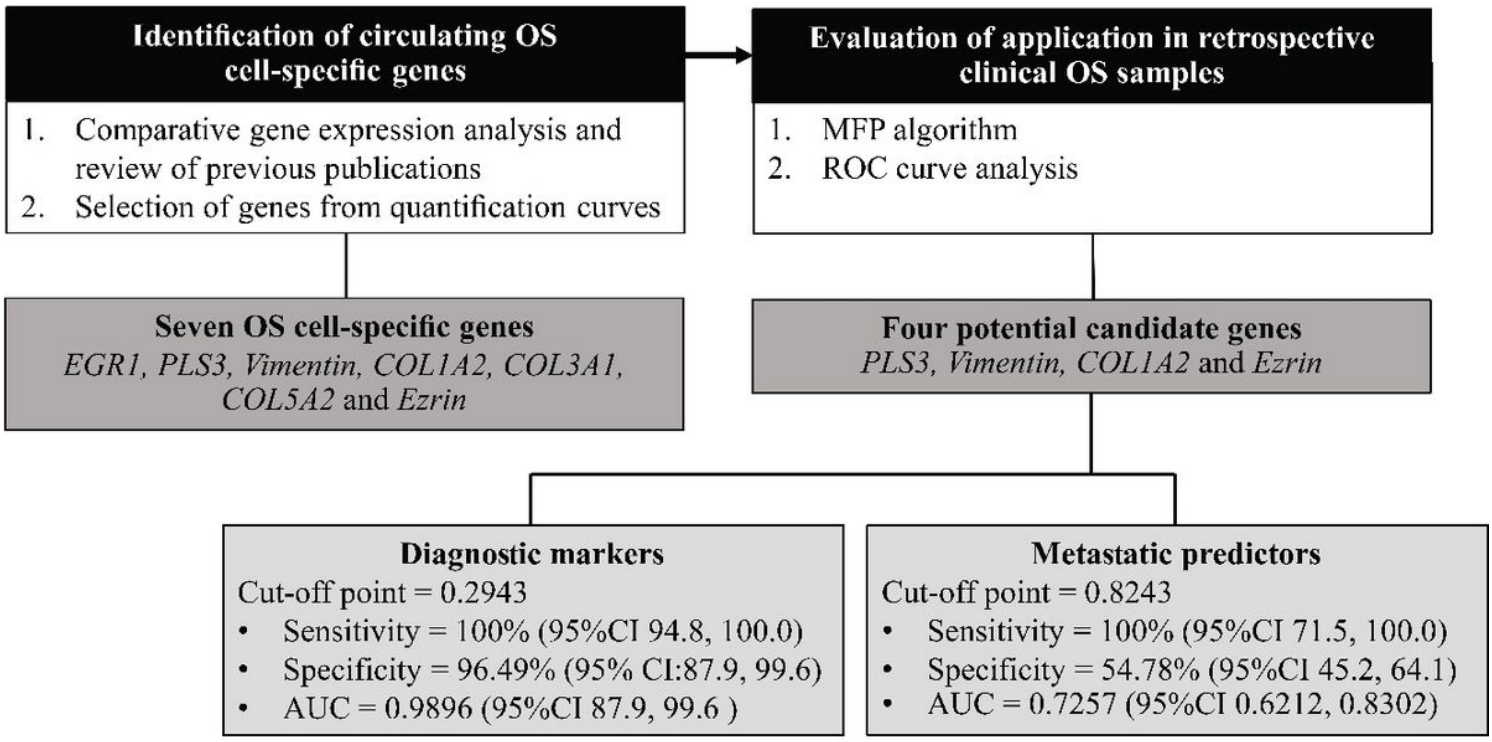
1. Misaghi A, Goldin A, Awad M, Kulidjian AA: Osteosarcoma: a comprehensive review. Sicot j 4,12(2018)
2. Moore DD, Luu HH: Osteosarcoma. Cancer Treat Res 162,65-92(2014)
3. Marko TA, Diessner BJ, Spector LG: Prevalence of Metastasis at Diagnosis of Osteosarcoma: An International Comparison. Pediatr Blood Cancer 63(6),1006-1011(2016)
4. Franzius C, Sciuk J, Daldrup-Link HE, Jürgens H, Schober O: FDG-PET for detection of osseous metastases from malignant primary bone tumours: comparison with bone scintigraphy. Eur J Nucl Med 27(9),1305-1311(2000)
5. Hurley C, McCarville MB, Shulkin BL, Mao S, Wu J, Navid F, Daw NC, Pappo AS, Bishop MW: Comparison of (18) F-FDG-PET-CT and Bone Scintigraphy for Evaluation of Osseous Metastases in Newly Diagnosed and Recurrent Osteosarcoma. Pediatr Blood Cancer 63(8),1381-1386(2016)
6. Brenner W, Bohuslavizki KH, Eary JF: PET imaging of osteosarcoma. J Nucl Med 44(6),930-942(2003)
7. Kaste SC: Imaging pediatric bone sarcomas. Radiol Clin North Am 49(4),749-765, vi-vii(2011)
8. Raimondi L, De Luca A, Costa V, Amodio N, Carina V, Bellavia D, Tassone P, Pagani S, Fini M, Alessandro R *et al*: Circulating biomarkers in osteosarcoma: new translational tools for diagnosis and treatment. Oncotarget 8(59),100831-100851(2017)
9. Bankó P, Lee SY, Nagygyörgy V, Zrínyi M, Chae CH, Cho DH, Telekes A: Technologies for circulating tumor cell separation from whole blood. Journal of Hematology & Oncology 12(1),48(2019)
10. Wong IH, Chan AT, Johnson PJ: Quantitative analysis of circulating tumor cells in peripheral blood of osteosarcoma patients using osteoblast-specific messenger RNA markers: a pilot study. Clin Cancer Res 6(6),2183-2188(2000)
11. Zhong GX, Feng SD, Shen R, Wu ZY, Chen F, Zhu X: The clinical significance of the Ezrin gene and circulating tumor cells in osteosarcoma. Onco Targets Ther 10,527-533(2017)

12. Baldauf MC, Orth MF, Dallmayer M, Marchetto A, Gerke JS, Rubio RA, Kiran MM, Musa J, Knott MML, Ohmura S *et al*: Robust diagnosis of Ewing sarcoma by immunohistochemical detection of super-enhancer-driven EWSR1-ETS targets. Oncotarget 9(2),1587-1601(2018)
13. Rui X, Shao S, Wang L, Leng J: Identification of recurrence marker associated with immune infiltration in prostate cancer with radical resection and build prognostic nomogram. BMC Cancer 19(1),1179(2019)
14. Dong S, Huo H, Mao Y, Li X, Dong L: A risk score model for the prediction of osteosarcoma metastasis. FEBS Open Bio 9(3),519-526(2019)
15. Chaiyawat P, Pruksakorn D, Pipatwattana P, Phanphaisarn A, Teeyakasem P, Klangjorhor J, Settakorn J: Endoplasmic reticulum protein 29 (ERp29) as a novel prognostic marker and tumor suppressor in osteosarcoma. Journal of Bone Oncology 16,100233(2019)
16. Jonsson KB, Frost A, Nilsson O, Ljunghall S, Ljunggren O: Three isolation techniques for primary culture of human osteoblast-like cells: a comparison. Acta Orthop Scand 70(4),365-373(1999)
17. Dupont WD: Review of Multivariable Model-building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modeling Continuous Variables, by Royston and Sauerbrei. The Stata Journal 10(2),297-302(2010)
18. Zhang Z: Multivariable fractional polynomial method for regression model. Ann Transl Med 4(9),174-174(2016)
19. Heller G, McCormack R, Kheoh T, Molina A, Smith MR, Dreicer R, Saad F, de Wit R, Aftab DT, Hirmand M *et al*: Circulating Tumor Cell Number as a Response Measure of Prolonged Survival for Metastatic Castration-Resistant Prostate Cancer: A Comparison With Prostate-Specific Antigen Across Five Randomized Phase III Clinical Trials. J Clin Oncol 36(6),572-580(2018)
20. Kim M-Y, Oskarsson T, Acharyya S, Nguyen DX, Zhang XHF, Norton L, Massagué J: Tumor Self-Seeding by Circulating Cancer Cells. Cell 139(7),1315-1326(2009)
21. Wu Z-J, Tan J-C, Qin X, Liu B, Yuan Z-C: Significance of circulating tumor cells in osteosarcoma patients treated by neoadjuvant chemotherapy and surgery. Cancer Manag Res 10,3333-3339(2018)
22. Meng S, Fan X, Zhang J, An R, Li S: GJA1 Expression and Its Prognostic Value in Cervical Cancer. BioMed Research International 2020,8827920(2020)
23. Wu D, Chen K, Bai Y, Zhu X, Chen Z, Wang C, Zhao Y, Li M: Screening of diagnostic markers for osteosarcoma. Mol Med Rep 10(5),2415-2420(2014)
24. Diao C, Xi Y, Xiao T: Identification and analysis of key genes in osteosarcoma using bioinformatics. Oncol Lett 15(3),2789-2794(2018)
25. Matsunoshita Y, Ijiri K, Ishidou Y, Nagano S, Yamamoto T, Nagao H, Komiya S, Setoguchi T: Suppression of Osteosarcoma Cell Invasion by Chemotherapy Is Mediated by Urokinase Plasminogen Activator Activity via Up-Regulation of EGR1. PLOS ONE 6(1),e16234(2011)
26. Li Y, Du W, Han J, Ge J: LAMP3 promotes the invasion of osteosarcoma cells via SPP1 signaling. Mol Med Rep 16(5),5947-5953(2017)

27. Koks S, Reimann E, Maasalu K, Koks G, Xuan DH, Prans E, Martson A: Whole transcriptome analysis of osteosarcoma. The FASEB Journal 30(S1),515.516-515.516(2016)
28. Kuijjer ML, Namløs HM, Hauben EI, Machado I, Kresse SH, Serra M, Llombart-Bosch A, Hogendoorn PCW, Meza-Zepeda LA, Myklebost O *et al*: mRNA expression profiles of primary high-grade central osteosarcoma are preserved in cell lines and xenografts. BMC Medical Genomics 4(1),66(2011)
29. Talluri B, Amar K, Saul M, Shireen T, Konjufca V, Ma J, Ha T, Chowdhury F: COL2A1 Is a Novel Biomarker of Melanoma Tumor Repopulating Cells. Biomedicines 8,360(2020)
30. Totland C, Kråkenes T, Mazengia K, Haugen M, Vedeler C: Expression of the onconeural protein CDR1 in cerebellum and ovarian cancer. Oncotarget 9(35),23975-23986(2018)
31. Kujawski R, Przybyłowska-Sygut K, Mik M, Lewandowski M, Trzciński R, Berut M, Dziki Ł, Majsterek I, Dziki A: Expression of the PLS3 Gene in Circulating Cells in Patients with Colorectal Cancer. Pol Przegl Chir 87(2),59-64(2015)
32. Briggs JW, Ren L, Nguyen R, Chakrabarti K, Cassavaugh J, Rahim S, Bulut G, Zhou M, Veenstra TD, Chen Q *et al*: The ezrin metastatic phenotype is associated with the initiation of protein translation. Neoplasia 14(4),297-310(2012)
33. Habel N, Stefanovska B, Carène D, Patiño-Garcia A, Lecanda F, Fromigué O: CYR61 triggers osteosarcoma metastatic spreading via an IGF1Rβ-dependent EMT-like process. BMC Cancer 19(1),62(2019)
34. Satelli A, Li S: Vimentin in cancer and its potential as a molecular target for cancer therapy. Cell Mol Life Sci 68(18),3033-3046(2011)
35. Muro-Cacho CA: The Role of Immunohistochemistry in the Diagnosis of Primary Tumors of the Bone. Cancer Control 5(6),561-569(1998)
36. Ferreira MM, Ramani VC, Jeffrey SS: Circulating tumor cell technologies. Molecular Oncology 10(3),374-394(2016)
37. Arun Satelli AMJJCMDXXDRIDDNSKETVRJALESK: Universal Marker and Detection Tool for Human Sarcoma Circulating Tumor Cells. Cancer Research 74(6),1645-1650(2014)
38. Shaik SM, Devireddy RV: Heat and Mass Transfer Models and Measurements for Low-Temperature Storage of Biological Systems. In: Handbook of Thermal Science and Engineering. Edited by Kulacki FA. Cham: Springer International Publishing, 1-39(2017)
39. Obermayr E, Maritschnegg E, Agreiter C, Pecha N, Speiser P, Helmy-Bader S, Danzinger S, Krainer M, Singer C, Zeillinger R: Efficient leukocyte depletion by a novel microfluidic platform enables the molecular detection and characterization of circulating tumor cells. Oncotarget 9(1),812-823(2018)
40. Zavridou M, Mastoraki S, Strati A, Koutsodontis G, Klinakis A, Psyrri A, Lianidou E: Direct comparison of size-dependent versus EpCAM-dependent CTC enrichment at the gene expression and DNA methylation level in head and neck squamous cell carcinoma. Scientific Reports 10(1),6551(2020)
41. Poos K, Smida J, Maugg D, Eckstein G, Baumhoer D, Nathrath M, Korsching E: Genomic Heterogeneity of Osteosarcoma - Shift from Single Candidates to Functional Modules. PLOS ONE 10(4),e0123082(2015)

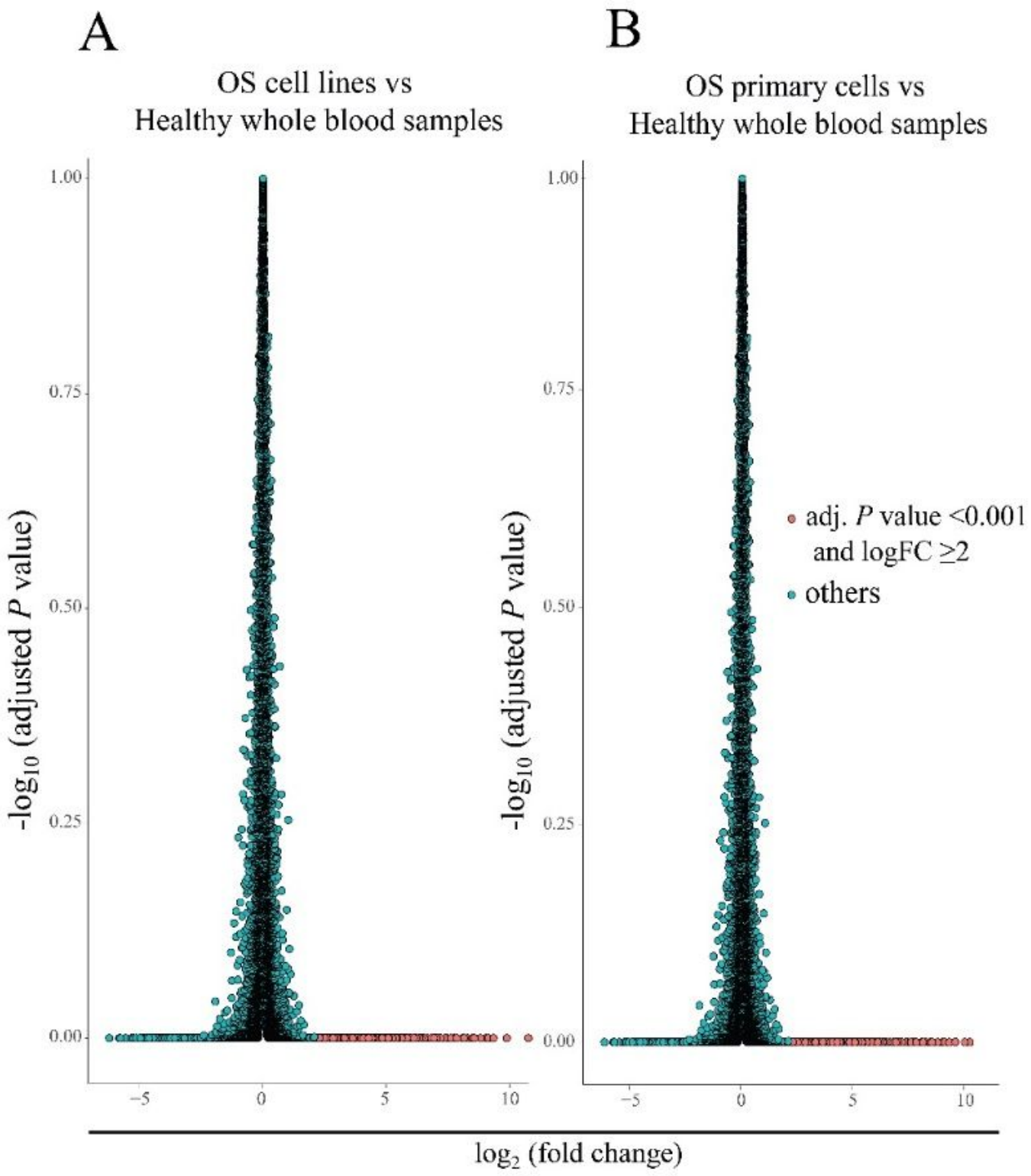
42. de Azevedo JWV, Fernandes TAAdM, Fernandes JV, de Azevedo JCV, Lanza DCF, Bezerra CM, Andrade VS, de Araújo JMG, Fernandes JV: Biology and pathogenesis of human osteosarcoma (Review). Oncol Lett 19(2),1099-1116(2020)
43. Liu Y, Feng X, Zhang Y, Jiang H, Cai X, Yan X, Huang Z, Mo F, Yang W, Yang C *et al*: Establishment and characterization of a novel osteosarcoma cell line: CHOS. Journal of Orthopaedic Research 34(12),2116-2125(2016)
44. Calabrese G, Bennett BJ, Orozco L, Kang HM, Eskin E, Dombret C, De Backer O, Lusis AJ, Farber CR: Systems Genetic Analysis of Osteoblast-Lineage Cells. PLOS Genetics 8(12),e1003150(2012)
45. Li J, Ding Y, Li A: Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. World J Surg Oncol 14(1),297(2016)
46. Yu Y, Liu D, Liu Z, Li S, Ge Y, Sun W, Liu B: The inhibitory effects of COL1A2 on colorectal cancer cell proliferation, migration, and invasion. J Cancer 9(16),2953-2962(2018)
47. Yokobori T, Iinuma H, Shimamura T, Imoto S, Sugimachi K, Ishii H, Iwatsuki M, Ota D, Ohkuma M, Iwaya T *et al*: Plastin3 Is a Novel Marker for Circulating Tumor Cells Undergoing the Epithelial-Mesenchymal Transition and Is Associated with Colorectal Cancer Prognosis. Cancer research 73(2013)
48. Wang YF, Shen JN, Xie XB, Wang J, Huang G: Expression change of ezrin as a prognostic factor in primary osteosarcoma. Medical Oncology 28(1),636-643(2011)
49. Zhang C, Hou W-h, Ding X-x, Wang W: Correlation of ezrin expression level and clinical significance for patients with bone and soft tissue sarcomas : a meta-analysis. In*: 2018*.
50. Amantini C, Morelli MB, Nabissi M, Piva F, Marinelli O, Maggi F, Bianchi F, Bittoni A, Berardi R, Giampieri R *et al*: Expression Profiling of Circulating Tumor Cells in Pancreatic Ductal Adenocarcinoma Patients: Biomarkers Predicting Overall Survival. Frontiers in Oncology 9(874) (2019)

# Figures



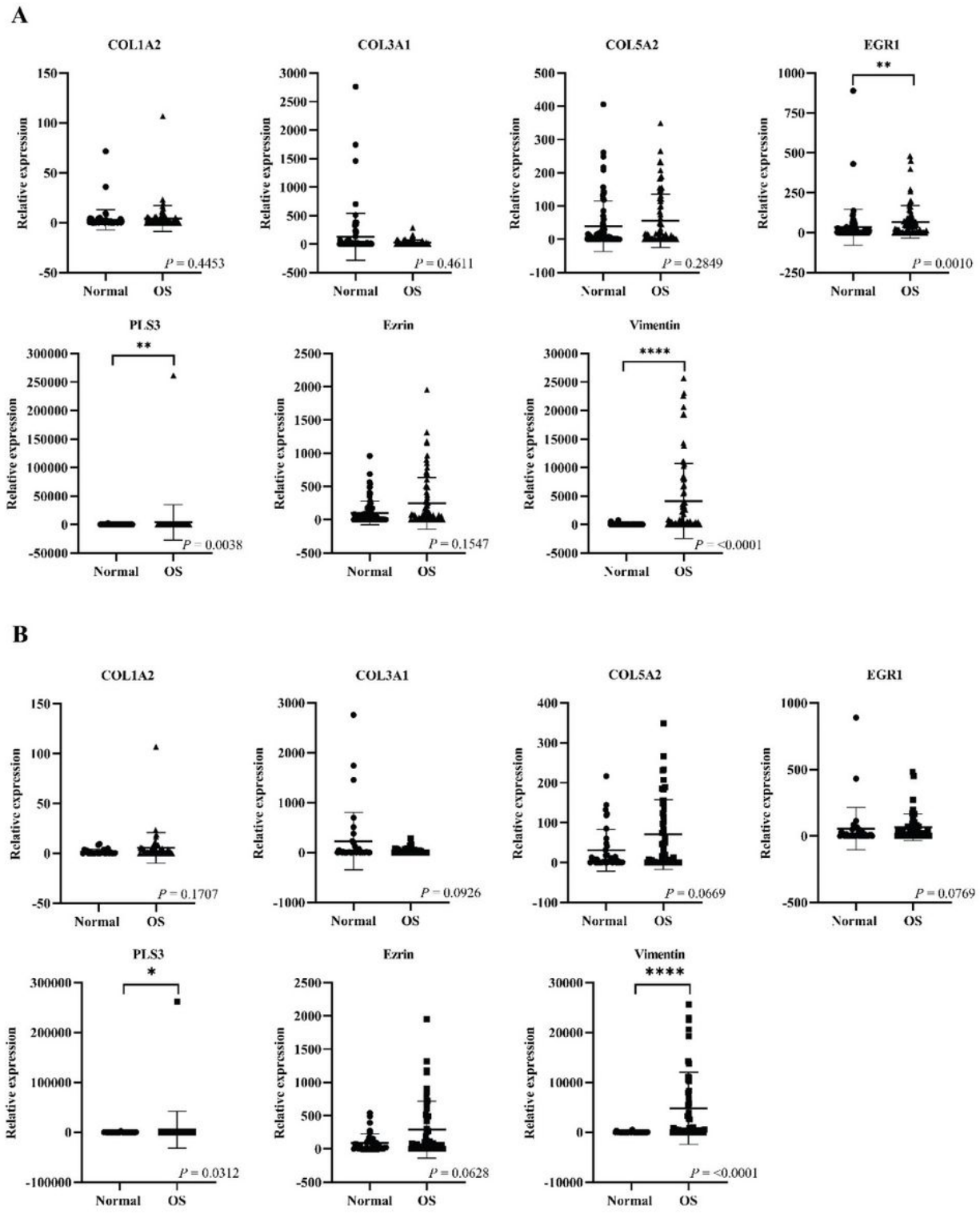

## Figure 1

Overview of the study and workflow of methods. OS, osteosarcoma; COL1A2, Collagen Type I Alpha 2 Chain; PLS3, Plastin-3; COL5A2, Collagen Type V Alpha 2 Chain; COL3A1, Collagen Type III Alpha 1 Chain; EGR1, Early Growth Response Protein 1; VIM, Vimentin; MFP; multivariable fractional polynomial, ROC; receiver operating characteristic curve; AUC, area under the ROC Curve; CI, confidence interval.




## Figure 2

Comparative expression analysis. Volcano plots showing (A) pairwise comparison of gene expression in osteosarcoma (OS) cells and (B) OS primary cells VS healthy donors. The genes represent in red had an expression ratio > 2 (log2) and adjusted P value < 0.01.




## Figure 3

Detection of circulating osteosarcoma (OS) cells in buffy coat of OS patients and healthy donors by qRT-PCR technique. (A) Data are shown as scatter plots of COL1A2, COL3A1, COL5A2, EGR1, PLS3, Ezrin and VIM relative expression by qRT-PCR. The expression of seven candidate genes in 73 OS patients and 79 healthy donors were quantified using qRT-PCR. (B) Only data from patients and healthy donors who were under 24 years of age are presented in the scatter plots. Each data expression was normalized on CD45

RNA level by the 2-ΔCt method. Each sample was analyzed in triplicate. Data are displayed as vertical scatter plots; bars represent mean ± SD. The Mann-Whitney U-Test test was used to determine the P-values *p<0.05; **, p<0.01; ***, p<0.001.

## Supplementary Files

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

- Supplementarydata.docx