Supplementary methods

HLA genotyping and expression in osteosarcoma cell lines

High-resolution HLA genotyping of the osteosarcoma cell lines was performed by the HLA diagnostics laboratory (LUMC, Leiden, the Netherlands) by next generation sequencing. For this, DNA was isolated from 1 million cells by Chemagen DNA isolation robot (14-820-0470 – 14-820-0472, Perkin Elmer Janus). Subsequently, HLA genes were amplified using the NGS-go-MX11-3 kit (7971864, Genome Diagnostics) according to manufacturer's instructions and sequenced on the Illumina MiniSeq platform. Cell surface expression of HLA class I was determined by flowcytometry analysis on the osteosarcoma cell lines U2OS, SAOS2 and OHS. JY cells were used as reference for positive HLA class I expression. Cell suspensions obtained after trypsinization with 0.25% Trypsin/EDTA (sc-391060 2.5% Trypsin/EDTA, Santa Cruz Biotechnologies Int, USA) were incubated with antibodies against HLA-ABC (fluorochrome: Alexa Fluor 647, 311414, Biolegend, USA), beta-2 microglobulin (B2M) (fluorochrome: PE, 316305, Biolegend, USA) and the Live/Dead™ Fixable Near-IR Dead stain kit (APC-Cy7, L34976, Invitrogen™ Life Technologies Corporation, Thermo Fisher Scientific, USA). Cells were washed with FACS buffer (PBS/1%BSA/0.1% Tween20). Flow cytometry was performed on a LSR-II Flow Cytometer (BD Biosciences). Results were analyzed in FlowJo.

HLA-peptide elution

Approximately 500 million cells were lysed in 50 mL lysis buffer (50 mM Tris-Cl pH 8.0, 150 mM NaCl, 5 mM EDTA, 0.5% Zwittergent 3-12 (N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate) and protease inhibitor (Complete, Roche Applied Science)) for 2 h at 0°C [1]. Lysates were successively centrifuged for 10 min at 2500 × g and for 45 min at 31,000 x g to remove nuclei and other insoluble material, respectively. Next, lysates were passed through 1 mL CL-4B Sepharose column to preclear the lysate. The cleared lysates were passed through a 1 mL column containing 2.5 mg anti- pan class I (w6-32) IgG coupled to protein A Sepharose [1]. The antibody columns were washed with 3 mL of lysis buffer, 3 mL of low salt buffer (20 mM Tris-Cl pH 8.0, 120 mM NaCl), 1 mL of high salt buffer (20 mM Tris-Cl pH 8.0, 1 M NaCl), and finally with 3 mL of low salt buffer. Peptides were eluted with 5 mL of 10% acetic acid and purified on a 10kDa filter (Microcon, Millipore). The filtrate was diluted with 10 mL of 0.1% TFA and purified by SPE (Oasis HLB, Waters) using 20% and 30% acetonitrile in 0.1% TFA to elute the peptides.

Peptides were lyophilized, dissolved in 95/3/0.1 v/v/v water/acetonitrile/formic acid and subsequently analyzed by on-line C18 nanoHPLC MS/MS with a system consisting of an Ultimate3000nano gradient HPLC system (Thermo, Bremen, Germany), and an Exploris480 mass

spectrometer (Thermo). Fractions were injected onto a cartridge precolumn (300 µm × 5 mm, C18 PepMap, 5 um, 100 A), and eluted via a homemade analytical nano-HPLC column (50 cm × 75 µm; Reprosil-Pur C18-AQ 1.9 µm, 120 A, Dr. Maisch, Ammerbuch, Germany). The gradient was run from 2% to 40% solvent B (20/80/0.1 water/acetonitrile/formic acid (FA) v/v) in 120 min. The nano-HPLC column was drawn to a tip of ∼5 µm and acted as the electrospray needle of the MS source. The mass spectrometer was operated in data-dependent MS/MS mode for a cycle of 20 MS/MS scans, with a HCD collision energy at 30 V and recording of the MS2 spectrum in the orbitrap, with a quadrupole isolation width of 1.2 Da. In the master scan (MS1) the resolution was 60,000, the scan range 300-1500, at standard AGC target at a maximum fill time of 50 ms. A lock mass correction on the background ion m/z=445.12 was used. Precursors were dynamically excluded after N=1 with an exclusion duration of 45 s, and with a precursor range of 20 ppm. Charge states 1-3 were included. Singly charged precursors were selected from the m/z 800-1400 range, doubly charged precursors were selected from the m/z 400-800 range, and triply charged precursors were selected from the m/z 300-600 range. For MS2 the first mass was set to 110 Da, and the MS2 scan resolution was 30,000 at an AGC target of 100% with a maximum fill time of 100 ms.

Data analysis for protein identification

In a post-analysis process, raw data were first converted to peak lists using Proteome Discoverer version 2.1 (Thermo Electron), and then submitted to the Uniprot Homo sapiens minimal database (20596 entries), (supplemented with specific sequences from highly expressed CTAs in osteosarcoma metastases) using Mascot v. 2.2.07 (www.matrixscience.com) for protein identification. Mascot searches were done with 10 ppm and 0.02 Da deviation for precursor and fragment mass, respectively, and no enzyme was specified. Methionine oxidation and cysteinylation of cysteine were set as variable modifications. The false discovery rate was set < 1% and, in addition, peptides with mascot ion scores <35 were discarded.

1. Hassan C, Kester MG, de Ru AH, Hombrink P, Drijfhout JW, Nijveen H*, et al.*, The human leukocyte antigen-presented ligandome of B lymphocytes*. Mol Cell Proteomics* **2013**. 12(7): p. 1829-43.

Supplementary Tables

Table S1. clinical information

	Discovery cohort		Validation cohort
TMA	TMA-1		TMA-2
Applied technologies	RNA-seq	IMC, IHC (MAGEA3)	IHC (MAGEA3)
Patients	7		91
Total number of samples	23	28	150
Primary biopsies	5	6	79
Primary resections	5	6	47
Local recurrences	3	4	
Metastatic lesions	10	12	24
Location of primary tumor			
Femur	2		46
Tibia			27
Humerus	2		10
Fibula	1		5
Radius	1		
Ulna			1
Costa			1
Hand			1
Os Ilium	1		
Sex			
M	4		49
F	3		42
Age at diagnosis			
Age range (years)	14-30		5-53
Average	20		18
Median	20		18
Histotype			
Osteoblastic conventional OS	4		59
Chondroblastic conventional OS	1		9
Fibroblastic conventional OS	1		3
Teleangiectatic OS	1		
Unknown			20
Chemotherapy response			
Good	2		
Poor	5		
Outcome			
Alive	4		
Dead of disease	3		

Table S2. Patient samples used in RNAseq and IMC

Patient	RNAseq and IMC*	RNAseq only*	IMC only*
P1	P1Pb, P1Pr, P1R1		
P2	P2R1, P2R2, P2M1	P2Pb	P2Pr
P3	P3Pb, P3M1		P3Pr, P3M2
P4	P4Pb, P4Pr, P4M1, P4M2, P4M3, P4M4, P4M5		
P5	P5Pb, P5Pr, P5M1		
P6	P6M1	P6Pr	P6Pb
P7	P7Pr, P7M1		P7Pb, P7R1, P7M2

*Sample IDs are generated by Patient ID, followed by an abbreviation of the specific sample: Pb, Primary biopsy; Pr, primary resection; R, recurrence; M, metastasis. In the event of multiple recurrences or metastases, the number of the occurrence is specified in order of diagnosis.

Table S3. lineage markers used for immune cell phenotypes

Cell type	Lineage markers
Immune cells	CD45+
T cells	CD3+
B cells	CD20+
Innate lymphoid cells	CD3- CD7+
Dendritic cells	CD11c+ CD68- | CD14- CD68- HLA-DR+
Monocytes	CD11c- CD14+ CD68-
Macrophages	CD68+
Granulocytes	CD15+

Table S4. Primers used for MAGEA3, GAPDH and HPRT expression

Gene	Forward primer	Reverse primer
MAGEA3	CGGTGAGGAGGCAAGGTTCT	TCTCCTTCAGTGCTCCTCCG
GAPDH	GAGTCAACGGATTTGGTCGT	TTGATTTTGGAGGGATCTCG
HPRT1	CATTATGCTGAGGATTTGGAAAGG	CTTGAGCACACAGAGGGCTACA

Table S5. Significantly expressed genes between primary and first occurring metastasis (padj < 0.05).

Supplementary_table_S5.xlsx

Table S6. HLA types of osteosarcoma cell lines and samples

Cell line	*MAGEA3*	HLA-A1	HLA-A2	HLA-B1	HLA-B2	HLA-C1	HLA-C2
SAOS2	+	02:01	24:02	13:02	44:27	06:02	07:04
U2OS	+	02:01	32:01	44:02	44:27	05:01	07:04
OHS	+	02:01	02:01	15:01	44:05	02:02	03:04
MNNG	-	02:11	02:11	52:01	52:01	12:02	12:02
Patient	*MAGEA3* in metastasis						
P1	-	03:01	68:01	07:02	07:02	07:02	07:02
P2	+	02:01	32:01	55:01	15:01	03:04	01:02
P3	+	24:02	26:01	37:01	14:01	06:02	08:02
P4	+	02:01	03:01	27:05	40:01	03:04	02:02
P5	-	68:01	26:01	55:01	15:01	03:03	03:03
P6	+	02:01	23:01	45:01	15:03	02:10	16:01
P7	-	02:01	68:01	57:03	51:01	07:01	15:02

Table S7. MAGEA peptides presented on HLA class I detected by peptidomics in osteosarcoma cell lines

Supplementary_table_S7.xlsx

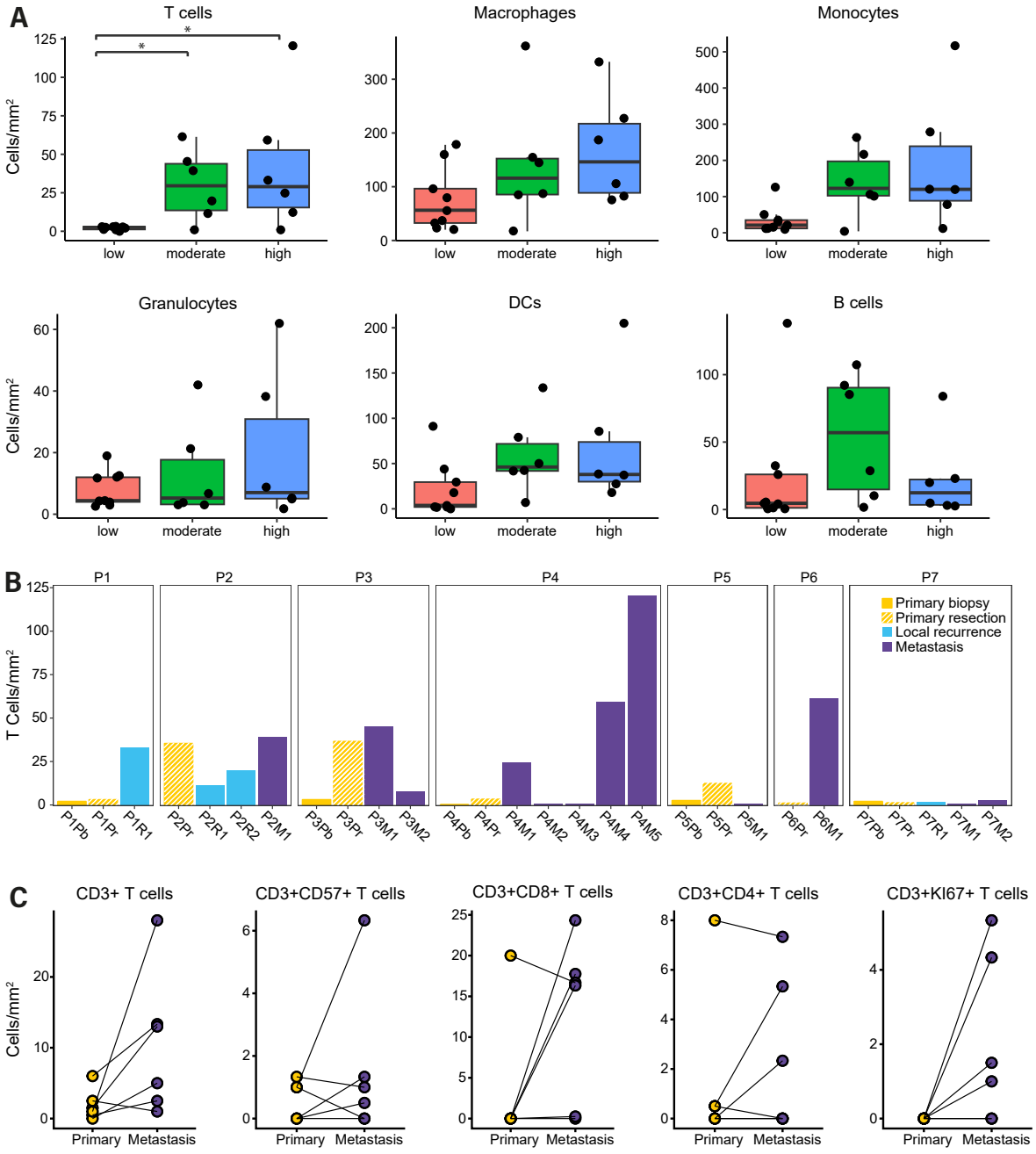

Figure S1. Comparison of immune cell lineage frequencies in osteosarcoma by IMC. (A) Number of immune cells per main phenotype grouped by ICR score (Kruskall-Wallis post-hoc Dunn test, * = p.adj value < 0.05). (B) Cumulative number of T cells counted based on CD3 signal ordered by serial samples per patient. (C) Number of T cells in primary biopsies and first occurring metastases, separated by (from left to right): CD3+ T cells, CD3+CD57+ exhausted T cells, CD3+CD8+ cytotoxic T cells, CD3+CD4+ CD4 helper T cells and CD3+KI67+ proliferating T cells.

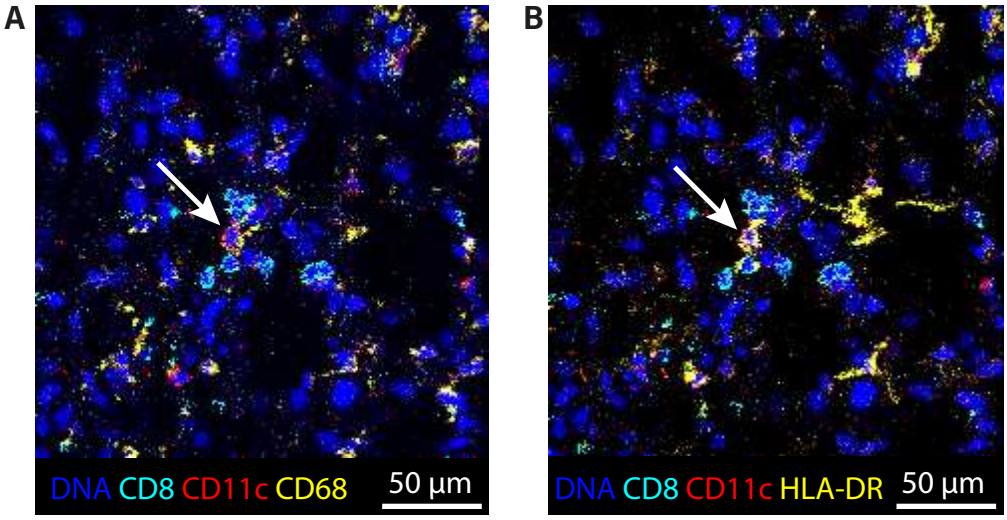

Figure S2. Expression of HLA-DR on CD11c macophages. (A) IMC image of an osteosarcoma metastasis sample displaying a CD11c+ CD68+ macrophage that is (B) HLA-DR positive, next to CD8 T cells.

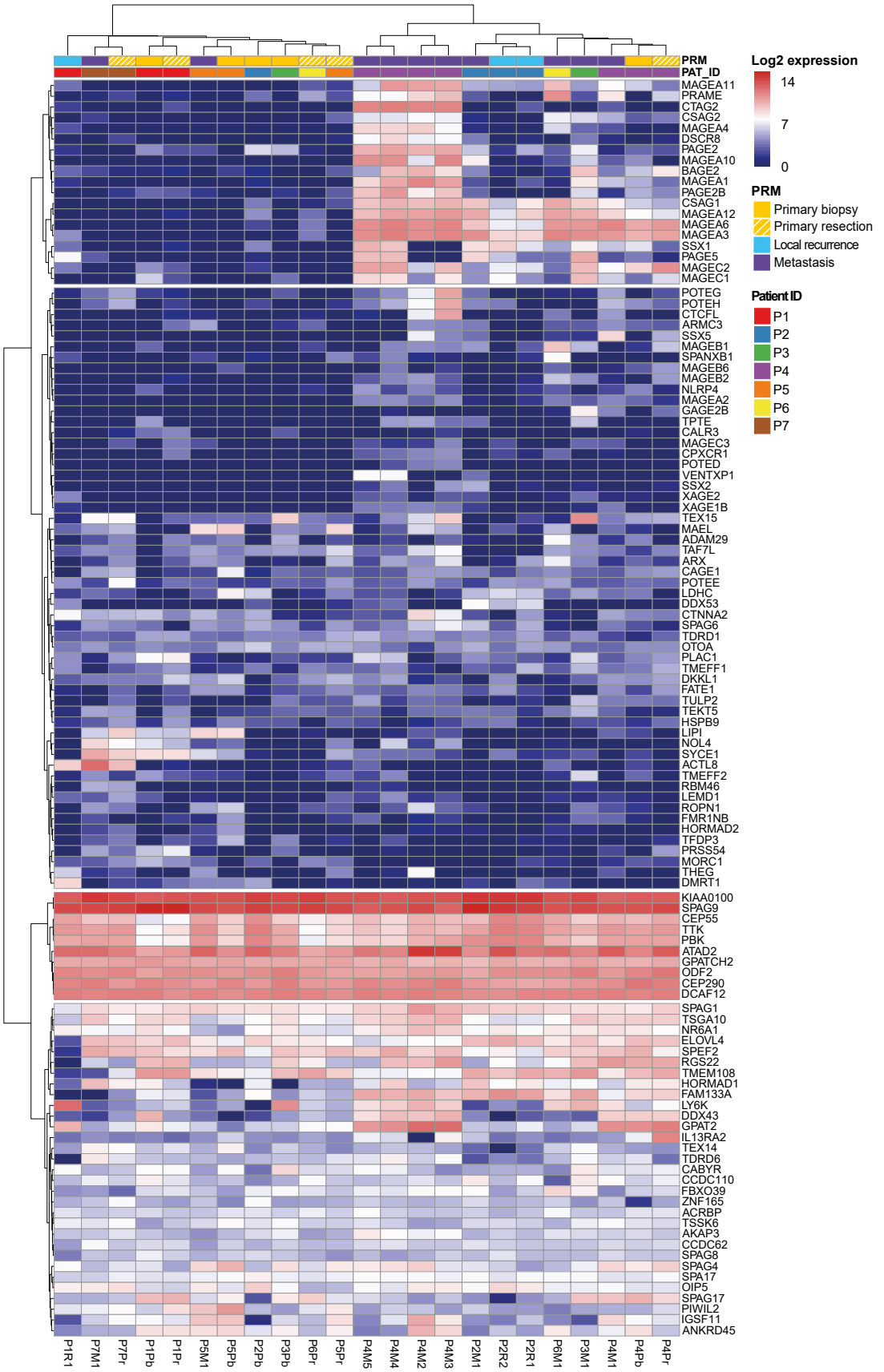

Figure S3. Expression of all CTAs. Unsupervised clustering of normalized log2 CTA gene expression. Rows on the heatmap are split by clustering (k=4).

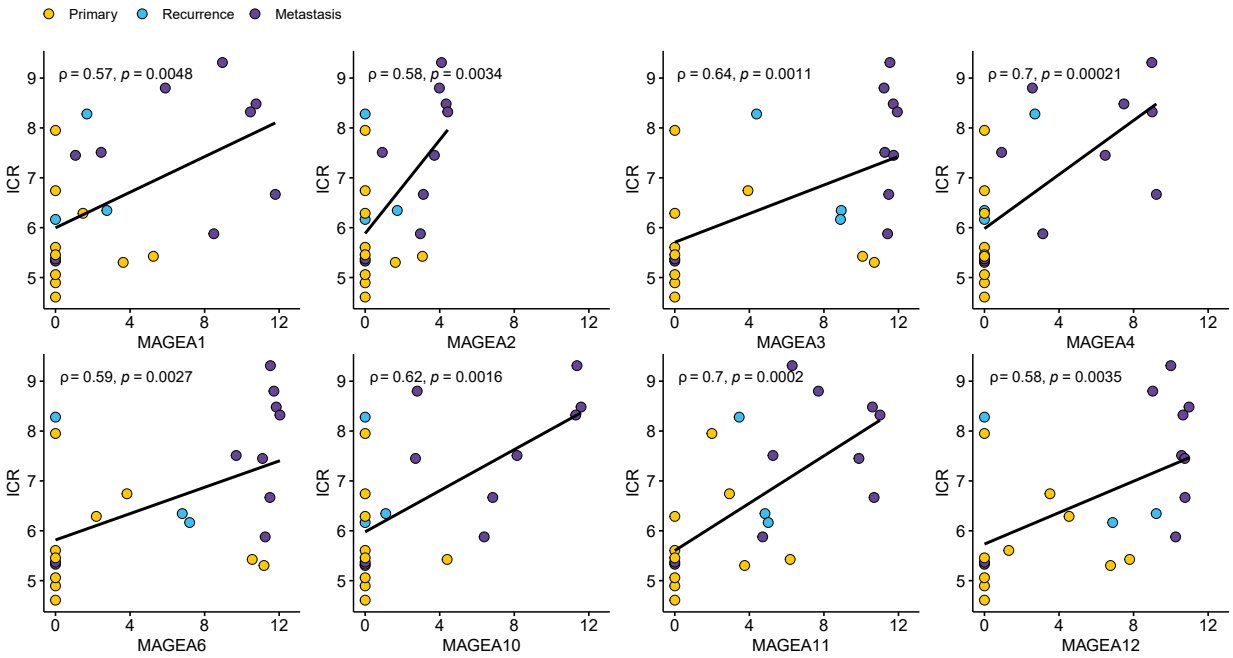

Figure S4. Correlations between MAGEA gene expression and ICR scores. MAGEA gene expression are in log2 normalized counts and the ICR scores are calculated by taking the mean of log2 normalized counts of the 20 genes in the ICR signature. Correlations are calculated with Spearman rank correlation.

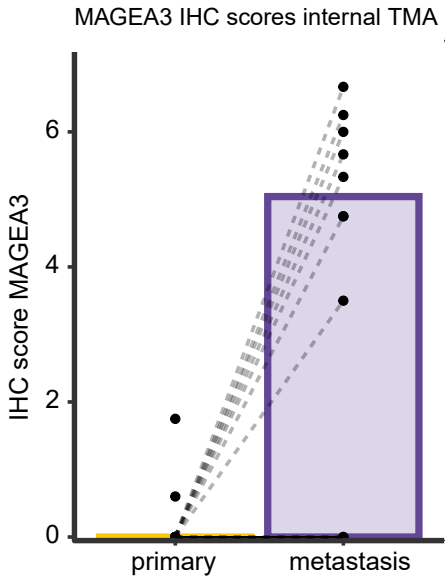

Figure S5. MAGEA3 IHC scores TMA-1. Barplot of IHC scores grouped on primary and metastasis. Bars represent medians, dashed lines connect primary biopsies with metastases from the same patient.

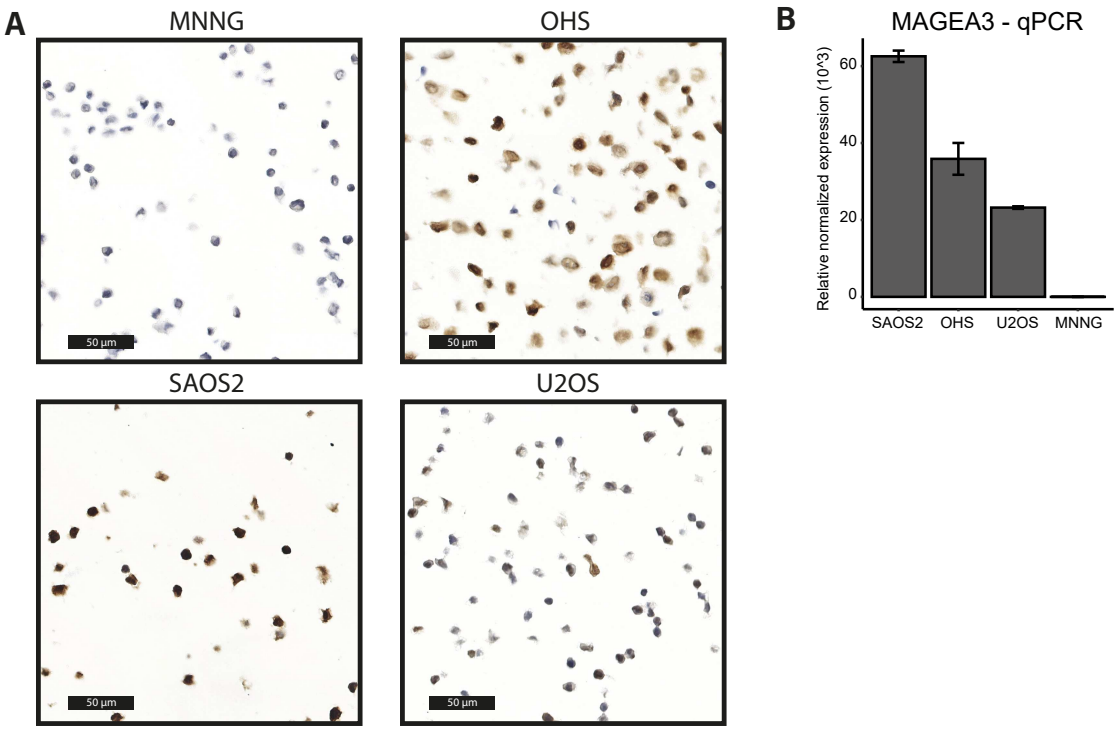

Figure S6. MAGEA3 expression in osteosarcoma cell lines. (A) Anti-MAGEA3 IHC staining in osteosarcoma cell lines. (B) Relative normalized expression of MAGEA3 in osteosarcoma cell lines measured by qPCR.

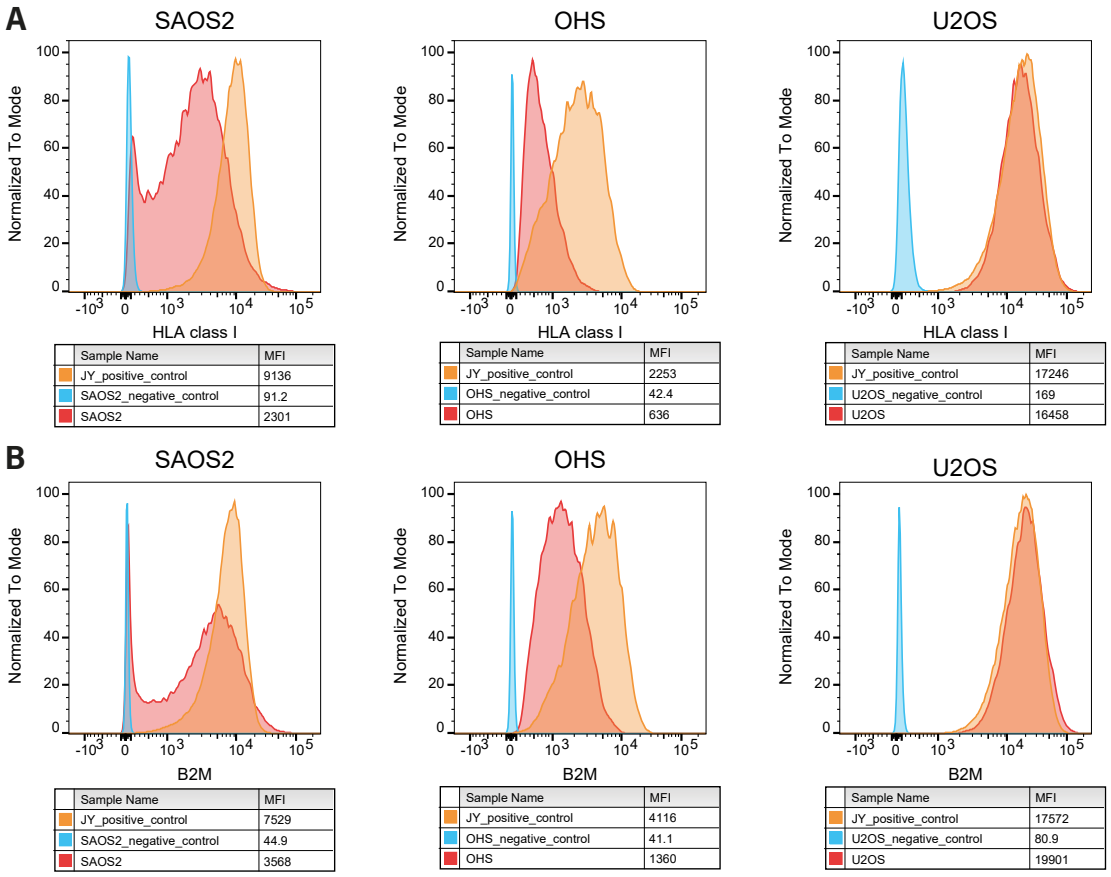

Figure S7. MAGEA3 and HLA expression in osteosarcoma cell lines. (A) Expression of HLA class I and (B) Beta-2 microglobulin (B2M) in MAGEA3 expressing cell lines, measured by flow cytometry. Corresponding Median Fluorescence Intensities (MFI) are depicted below the histograms. For positive and negative controls, JY cells and unstained SAOS2/OHS/U2OS cells were used, respectively.