SCIENTIFIC REPORTS

natureresearch

Check for updates

OPEN

SERRS multiplexing with multivalent nanostructures for the identification and enumeration of epithelial and mesenchymal cells

Lucio Litti [1], Andrea Colusso[1], Marcella Pinto[1], Erlis Ruli[3], Alessia Scarsi[1], Laura Ventura[3], Giuseppe Toffoli[4], Marco Colombatti[2], Giulio Fracasso[2] & Moreno Meneghetti[1]

Liquid biopsy represents a new frontier of cancer diagnosis and prognosis, which allows the isolation of tumor cells released in the blood stream. The extremely low abundance of these cells needs appropriate methodologies for their identification and enumeration. Herein we present a new protocol based on surface enhanced resonance Raman scattering (SERRS) gold multivalent nanostructures to identify and enumerate tumor cells with epithelial and mesenchimal markers. The validation of the protocol is obtained with spiked samples of peripheral blood mononuclear cells (PBMC). Gold nanostructures are functionalized with SERRS labels and with antibodies to link the tumor cells. Three types of such nanosystems were simultaneously used and the protocol allows obtaining the identification of all individual tumor cells with the help of a Random Forest ensemble learning method.

The identification of circulating biomarkers in different body fluids is an important step toward non invasive diagnosis for monitoring the dynamical evolution of tumors and for their treatment[1]. The metastatic spread of tumor cells from the primary mass is considered a leading cause of cancer-related death[2]. These cells represent a heterogeneous population of cells that are able to survive in the bloodstream and to initiate metastatic growth of new tumors[3–5]. Some of them have been validated as prognostic markers for metastasis of different types of tumors[6]. Recent experiments also suggest the importance of the presence of cell with mesenchymal markers, and not only with epithelial markers, in the blood[7–9].

The capture and analysis of tumor cells in the blood are challenging tasks because of their low abundance[10] and relevant obstacles are found today for technologies to be used in clinical diagnostics[11–15].

Nanotechnologies can help to face the challenges for the detection and analysis of these type of cells[16]. Assays based on highly specific and sensitive surface-enhanced Raman spectroscopy (SERS) technology are very efficient to detect these cells[17–20]. One of the crucial advantage of SERS over other techniques (usually fluorescence-based assays) is its sharp fingerprint-like spectral pattern, which are clearly different from interfering signals of the complex biological environment[19,21]. The narrow SERS signals[22–28] make possible a simple multiplexing approach with only one exciting wavelength in resonance with the localized plasmon resonance of the nanostructured system. This allows obtaining the SERS signals of molecules used as SERS reporters, which strongly interact with the plasmonic system[29–31]. Moreover, SERS signals do not suffer from bleaching, which is usually present in laser induced fluorescence. SERS signal intensities can be further increased, obtaining the surface-enhanced resonance Raman scattering (SERRS), when the laser is in resonance both with the localized surface plasmon resonance of the plasmonic nanostructures, and with the reporter molecules[31–33]. It is worth mentioning that the presence of a plasmonic structure usually suppresses fluorescence signals of molecules close to the nanoparticles (NPs), making

[1]Department of Chemical Science, University of Padova, via Marzolo 1, 35131 Padua, Italy. [2]Department of Medicine, University of Verona, P.le L.A. Scuro, 37134 Verona, Italy. [3]Department of Statistical Sciences, University of Padova, via Battisti 241, 35121 Padua, Italy. [4]SOC Farmacologia Sperimentale e Clinica, Centro di Riferimento Oncologico, Via Franco Gallini 2, 33081 Aviano, Italy. ✉email: giulio.fracasso@univr.it; moreno.meneghetti@unipd.it

background signals negligible. We have already shown that the opportunities given by the SERRS approach can be successfully exploited for specifically identifying prostatic tumor cells expressing prostate specific membrane antigen (PSMA) and prostate stem cell antigen (PSCA)[26,31] or other antigens using engineered peptides[25,34,35]. The functionalization of plasmonic nanostructures with a large number of antibodies (Abs) increases the targeting activity with respect to isolated Abs[36,37]. This interesting possibility can be exploited also using different types of Abs on the same nanostructure, which allows to further increase its avidity[38].

SERS signals for identification and enumeration of tumor cells were exploited by some laboratories without[17,23,39] or following an enrichment step[40–44]. Multiplexing approaches[45], coupling SERS and fluorescence[46] were also considered for cells identification[47].

In the present work, we address the possibility of identifying and enumerating tumor cells with epithelial and/ or mesenchymal markers, using SERRS multivalent nanostructures and the Random Forest ensemble learning method. Abs recognizing EpCAM and E-Cadherin (E-Cad) were used on the same nanostructure for epithelial cells identification, whereas CD44 and N-Cadherin (N-Cad) were used for the mesenchymal ones[48]. The choice of using more than one Ab on a single nanostructure allowed enhancing the avidity of the nanostructure for a cell phenotype and to simplify usual analysis in which a single Ab was used. A third type of nanostructure was used to identify PBMCs, with an Ab recognizing CD45, an antigen broadly expressed on white blood cells. The third type of nanostructures was introduced because it was not possible to completely deplete PBMCs from the spiked samples used for the experiments. Each type of nanostructure was functionalized with a different SERRS reporter molecule to permit their identification in a multiplexing approach. SERRS signals recorded for individual cells were analyzed with a Random Forest ensemble learning method[49], which allowed identifying the tumor cell phenotypes, also giving the level of identification.

All the needed functionalizations of the nanostructures were greatly simplified by using gold nanoparticles synthesized by laser ablation in solution, which do not need surfactants or other stabilizing molecules[50–52]. The naked surfaces of the nanoparticles simplify their full functionalization because it does not require control over the ligand exchange. Furthermore, the naked surfaces favor intense SERRS signals because of the strong interaction with the reporter molecules.

To produce spiked samples, we used LNCaP prostatic tumor cells, as model for the epithelial phenotype and U251 cells as model for the mesenchymal phenotype.

The present approach allows further analyses on the tumor cells after their identification, because they are always individually available on a glass slide. This is also an important step, which is not possible with other approaches like those based, for example, on PCR protocols.

Results and discussion

Nanosystem synthesis.   A schematic representation of the synthesis of the nanostructures used in this work is reported in Fig. 1a. Gold nanoparticles (AuNPs) were obtained by the laser ablation synthesis in solution (LASiS). This synthesis is performed directly in water and the nanoparticles do not need to be stabilized with molecules, like citrate or thiols, because they are produced with native surface charges which, by coulomb interaction, stabilize the colloidal solution[50]. We obtained AuNPs with nanomolar concentrations in water by focusing 9 ns pulses of a Nd:YAG pulsed laser at 1064 nm on the surface of a pure gold plate under water. It was already shown that AuNPs have dimensions of about 25 nm and a ζ-potential of about − 30 mV[31]. According to a previous defined protocol[31], a centrifugation of the colloidal solution allowed to obtain a controlled aggregation of the AuNPs with clusters containing, on average, 10–20 AuNPs and dimensions of the order of 100–200 nm (see Supplementary Fig. S1). The plasmonic properties and the strong SERS enhancement of the aggregated AuNPs were already analyzed in a recent work[53]. The aggregated AuNPs can be easily re-dispersed with a soni-cating bath. A tiny amount of a thiolated dye was linked to the AuNPs to label them with a SERS signal. From the dye library, that we already considered in previous works[31,54], Malachite Green, Texas Red and Nile Blue were chosen in the present case as reporter molecules for the three type of nanostructures (see Fig. 1a). Since these dyes show absorptions close to or in resonance with the excitation laser at 632.8 nm, used for the SERRS measurements, the nanostructures showed high intensity signals.

The SERRS nanostructures were then functionalized with antibodies (Abs) as targeting agents. To obtain a better link with the gold nanostructures, Abs were first functionalized with 2-iminothiolane to introduce, on average, about 1–2 new thiols per antibody[31]. This procedure was found effective to preserve the affinity and specificity of the antibodies and to obtain a total binding of about 300 Abs/SERRS nanostructure.

The following nanostructures were obtained, also inspired by the work of Bulfoni et al.[55] NS-1 (see Fig. 1) were functionalized for the mesenchymal cells identification with anti-CD44 and anti-N-Cad Abs, and with Malachite Green as SERRS reporter. NS-2 were functionalized for epithelial cells identification with anti-EpCAM and anti-E-Cad Abs, and with Texas Red as SERRS reporter. NS-3 were functionalized, to recognize PBMCs, with anti-CD45 Ab and with Nile Blue as SERRS reporter. The UV–Vis-NIR extinction and Raman spectra of the three different nanostructures are reported in Fig. 1b. Compared to the pristine AuNPs, the NS-i show plasmon bands shifted to the near infrared region, characteristic of aggregated nanostructures[53], providing a good resonance with the laser excitation at 632.8 nm used for the SERRS measurements. The different SERRS spectra of the three nanostructures (see Fig. 1c) allow easy multiplexing measurements, namely the simultaneous recording of the three types of SERRS signals in a single spectrum using only one laser excitation. Only a partial cell targeting activities of NS-1 and NS-2, were observed with standard flow cytometry (see Supplementary Fig. S2). This can derive from the fluorescence quenching induced by the presence of the plasmonic nanostructures, which reduces the reliability of these measurements. For this reason these data have not been used as reference points for the Random Forest analysis (see below).
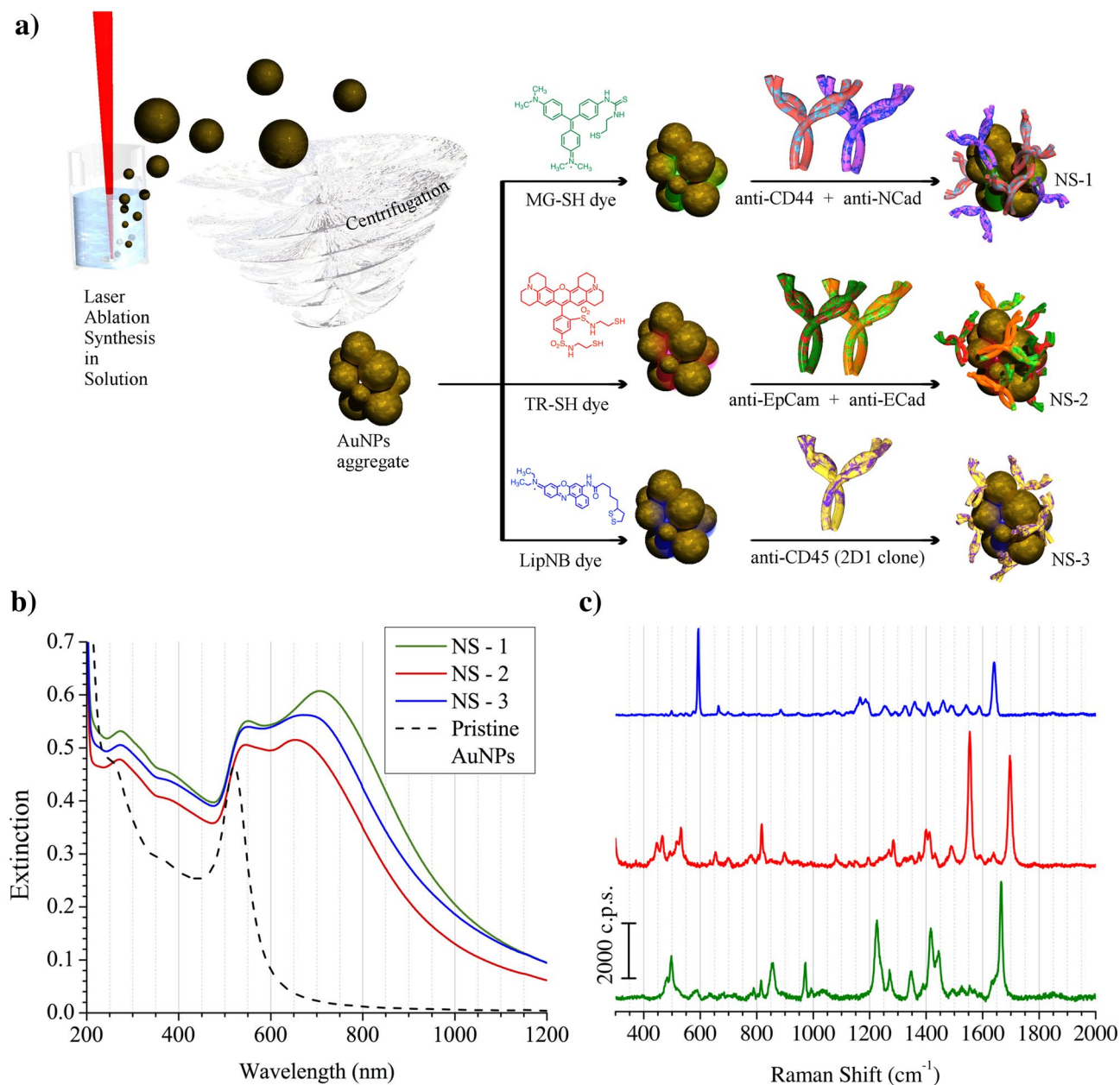
Figure 1. (**a**) Schematic representation of the synthesis of NS-1, NS-2 and NS-3: LASiS, aggregation by centrifugation, functionalization with a SERRS reporter molecules and then with the Abs. (**b**) UV–Vis-NIR extinction spectra of the three nanostructures, showing the characteristic plasmon resonances of aggregated NPs in the NIR. The extinction spectrum of the not aggregated NP, as obtained by the LASiS synthesis, is reported with a dashed line. (**c**) SERS spectra of the same nanostructures (NS-1, NS-2 and NS-3 from below) in quartz cuvette exciting at 632.8 nm.

Tumor cells capture and targeting.   The protocol for the identification and enumeration of tumor cells in peripheral blood was applied to spiked samples in which a fixed number of tumor cells (about some tenths per ml, see below) were added to PBMCs (i.e. white blood cells, WBC) from healthy donors, separated by the whole blood by a standard density gradient centrifugation (see "Methods" below). This step, namely samples in which tumor cells and PBMCs are present after red blood cells of a whole blood sample are discarded, is the important step for their identification[13–15,55]. The spiked samples were obtained with a number of WBC corresponding to the amount present in 7.5 ml of blood (about 7×10^7 PBMCs), which is considered a reference volume for such analysis[56]. Although this could be considered a limiting aspect of the preparation of the samples, it allowed focusing on the validation of the protocol for the identification and enumeration the tumor cells after their fixing on a slide (see below). The separation of the tumor cells in peripheral blood of a patient would follow the same protocol, and the tumor cells would be present together with PBMC. However, in this case an unknown number of tumor cells would be present, which is a problem for the validation of the protocol. The identification of low
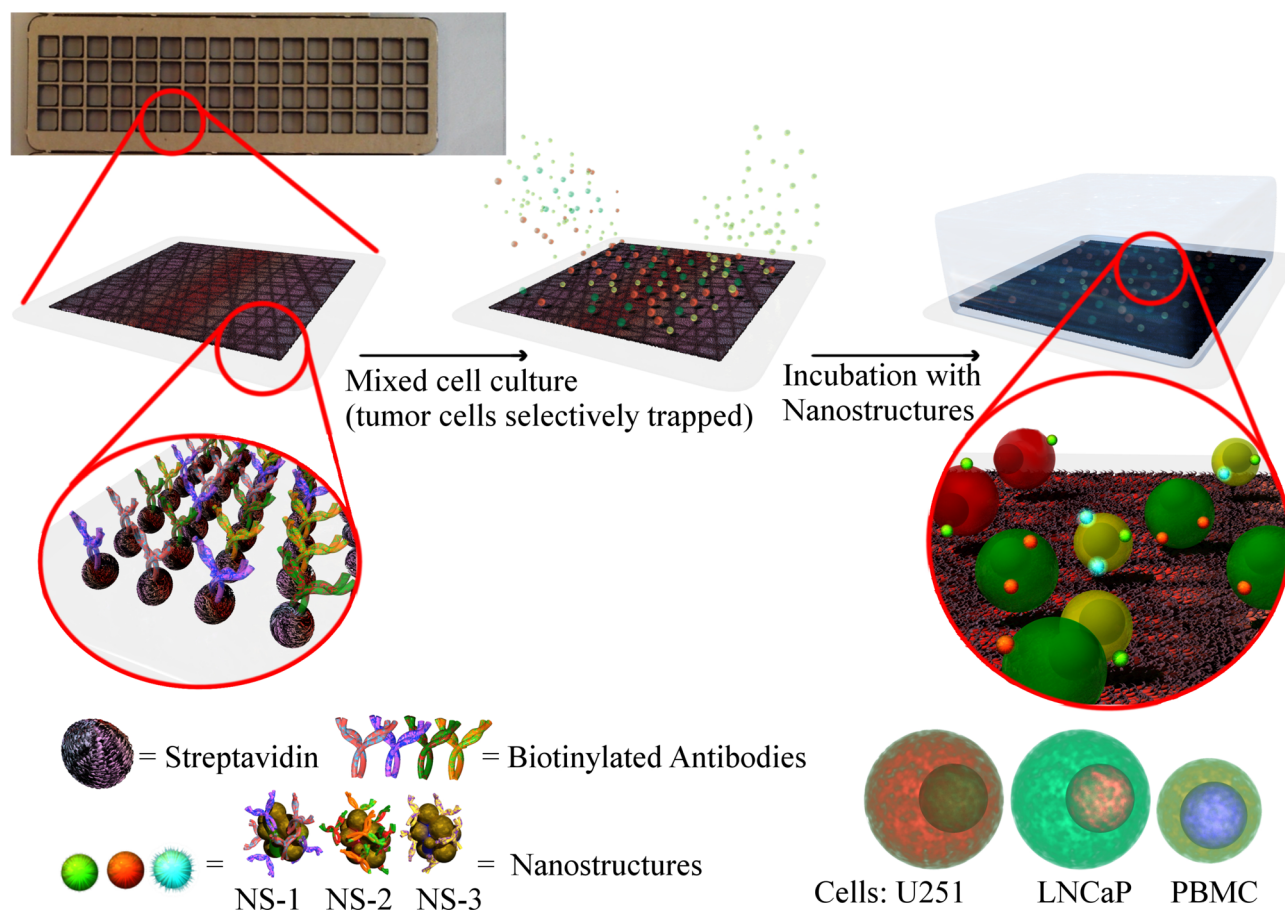
Figure 2. Schematic representation of the capture chamber, the tumor cells trapping and their incubation with a mixing of the three nanostructures (NS-1, NS-2 and NS-3), selectively recognizing the antigens of interest.

amount of tumor cells required a pre-analytical step in which the depletion of the large amount of WBC was obtained with an immunomagnetic sorting using anti-CD45-commercial magnetic beads (anti-human CD45 magnetic beads from Milteny) and following the manufacturer's protocol. Depletion of the order of 99.99% was obtained, which, however, means that thousands of PBMCs continued to be present in the samples and that, therefore, their identification were always needed. After this step the tumor cells were trapped on a glass slide activated with avidin, and functionalized with the Abs also used for functionalizing NS-1 and NS-2. For the immobilization on a glass slide, Abs were first biotinylated to exploit the high biotin-avidin affinity and stability. The avidin activated glass slide was also divided into small chambers of 3.5×3.5 mm with an adhesive silicon mask to reduce the incubation chamber where the cells were trapped and incubated (see Fig. 2).

The tumor cells were captured within these chambers and fixed to improve their retention for the following steps, namely for the incubation with the NS-i and washings. The efficiency of the capture and fixing steps for LNCaP and U251 cells were determined by counting the cells on the slide after the final washing. The efficiency of this procedure was found to be $84 \pm 7\%$.

Incubation of the cells was obtained with 0.5 nM solutions of NS-1, NS-2 and NS-3. The fixed cells were incubated with a mix of the three types of nanostructures for two hours at room temperature and than washed with PBS to remove all unbound nanostructures.

SERRS spectra acquisition and tumor cells identification. SERRS spectra were acquired for each cell present in a chamber with a μ-Raman spectrophotometer using the 632.8 nm laser line. Spatial coordinates of all the cells present in a chamber were determined using an image recorded with the camera of the instrument and a $20\times$ objective. The real area of each cell was measured, using the image pixels, for the following analysis. Then the μ-Raman spectrophotometer collected the SERRS spectra for each individual cell.

The threshold for a positive targeting activity of the three nanostructures was obtained with the Pearson coefficients calculated by comparing the spectrum collected for each cell with the three spectra characteristic of the three SERRS reporters.

These data, together with the dimensions of the cells, were used for a Random Forest (RF) analysis. RF is a well known learning method for supervised classification problems and is an ensemble learner, that is, a method that generates classifiers and aggregates their results[49]. The RF analysis makes possible to assign to each cell a percentage of attribution to the three classes: PBMC, U251 (mesenchymal cell) or LNCaP (epithelial cell). The dominant attribution identifies the type of cell.

Reference samples, with only PBMCs, U251 or LNCaP cells were used for calibrating the RF analysis. Supplementary Figs. S3–S5 show examples of the spectra obtained for the different type of cells, after incubation with the three types of NS-i, mixed together, and washing to remove unbound nanostructures.

Pearson correlation coefficients for all the cells of the three reference nanostructures showed that NS-2 SERRS signals well identified LNCaP cells, as expected, whereas NS-1 and NS-3 SERRS signals were found both for U251 and PBMCs, although with different relative abundances. This result was not surprising considering that both monocyte and lymphocyte show a positive CD44 and CD45 expressions. A large amount of cells into the well was constituted by PBMCs since the complete depletion of PBMCs with the immuno-magnetic sorting was unworkable. For a better identification of the U251 cells, another information was used in the RF analysis to discard PBMCs. A simple analysis of the dimension of the cells showed that large part of the PBMCs had surface area smaller than 30 μm^2, whereas for LNCap and U251 the area was larger than 30 μm^2 with an upper limit of 180 μm^2 (see Supplementary Fig. S6). Therefore, RF was instructed to consider only cells with surface larger than 30 and smaller than 180 μm^2. Clearly, these limits can be modified although they are appropriate also for other tumor cells[57].

Incubating three samples of a single type of cells with all the three nanostructures, the RF analysis allowed estimating the errors in the identification of the three types of cells (see Supplementary Table S1). One found that the best classified cells were the epithelial ones (LNCaP, 93% of cells identified), but also the mesenchymal ones (U251, 80% of cells identified) were satisfactorily identified.

A calibration of the RF analysis was then possible with a theoretical mixing of cells randomly chosen from the three samples with single type cells. In this case we knew the exact number of each type of cells present in the samples and they were compared with the RF output. RF gave a percentage of attribution to the identification of cells. 33% of percentage of attribution for the three types of cells, indicates uncertainty in the identification. The results obtained with the theoretical mixing showed (see Supplementary Fig. S7) that 60% of percentage of attribution identifies both LNCaP and U251, recovering the tumor cells of the theoretical mixing.

After the RF model calibration, two real spiked samples (Mix1 and Mix2) were constructed with virtually the same number of tumor cells. Nominally, they contained 24 ± 1 per ml of LNCaP and 29 ± 1 per ml of U251 mixed with a PBMCs amount similar to their number present in whole blood, namely about 10^7 per ml. Uncertainties in the number of LNCaP and U251 cells were determined with ten replicas of the dilution needed for counting the cells.

The two samples (Mix1 and Mix2) were incubated, washed and the SERRS signals recorded for each cells as above explained. The data were then analyzed with the RF model.

In Fig. 3b,c the RF attribution result for Mix1 and Mix 2 are presented. Using 60% of percentage of attribution, previously determined with the theoretical mixing, RF recovered 19 ± 1 LNCaP and 25 ± 5 U251 in Mix 1, and 21 ± 1 LNCaP and 32 ± 6 U251 in Mix 2. The results are summarized in Table 1. One should recall that the recovered amounts are 84% of the initial population because of the capturing and fixing steps (see above). Considering the uncertainties reported above, the enumeration of LNCaP is very satisfactory, as well as that of U251 in Mix1, whereas an overestimation is found for the U251 cells in Mix2 (32 with respect to 24), which, however, is within the estimated errors for this type of cells.

Considering a percentage of attribution higher than 60% (see Fig. 3b,c) a better cell identification can be obtained, although for a smaller number of cells. This can be useful for example for further analyses that can be made on single cells for studying other types of markers useful for a more complete understanding of a tumor stage.

Conclusions

Tumor cells in peripheral blood are growing up as important elements for liquid biopsy due to their early cancer diagnostic potential. For this reason, accurate and specific assays are needed for their identification, characterization and enumeration. In the present approach surface enhanced resonance Raman scattering (SERRS) signals from plasmonic multivalent nanostructures and a Random Forest ensemble learning analysis allowed identifying the tumor cells of spiked samples. In particular, for their importance in a tumor progression analysis, mesenchymal and epithelial cells were considered and identified at the same time. The final identification is given as a percentage of attribution for each single cell. The results show that the present protocol could be efficient for circulating tumor cells (CTCs) identification and enumeration suggesting that it could be used for an early detection of a tumor, its recurrence and a precise and fast monitoring of a response to therapy.

Methods

Synthesis of gold nanoparticles (AuNPs) by laser ablation synthesis in solution (LASiS). As previously reported[31], AuNPs were synthesized using 9 ns pulses at 1064 nm of a Nd:YAG laser. The laser beam was focused with a lens (10 cm focal length) on a pure Au target at the bottom of a cell filled with 50 ml of 10^{-5} M NaCl aqueous solution. Fluence of 1 J/cm² was used for obtaining nanoparticles with average diameters of 25 nm and a ζ-potential of about − 30 mV.

Synthesis of thiolated antibodies. Anti-EpCAM (HEA-125 clone, Antibodies on-line), anti-ECadherin (67A4 clone, BioLegend), anti-NCadherin (8C11 clone, eBiosciene), anti-CD44 (IM7 clone, eBioscience) and anti-CD45 (2D1 clone, EXBIO) antibodies were used as targeting agents and functionalized with free thiol groups following the same experimental procedure already reported[31]. Briefly, 150 μl of 1 M solution of NaHCO₃ were mixed with a solution of 2-iminothiolane hydrochloride (0.5 g/l) and 0.7 mg of one antibody. The reaction proceeded for two hours at room temperature and then overnight at 4 °C. The product was than purified by centrifugal ultrafiltration (5000 Da, Vivaspin 500 Sartorius) and washed for three times with a PBS (phosphate
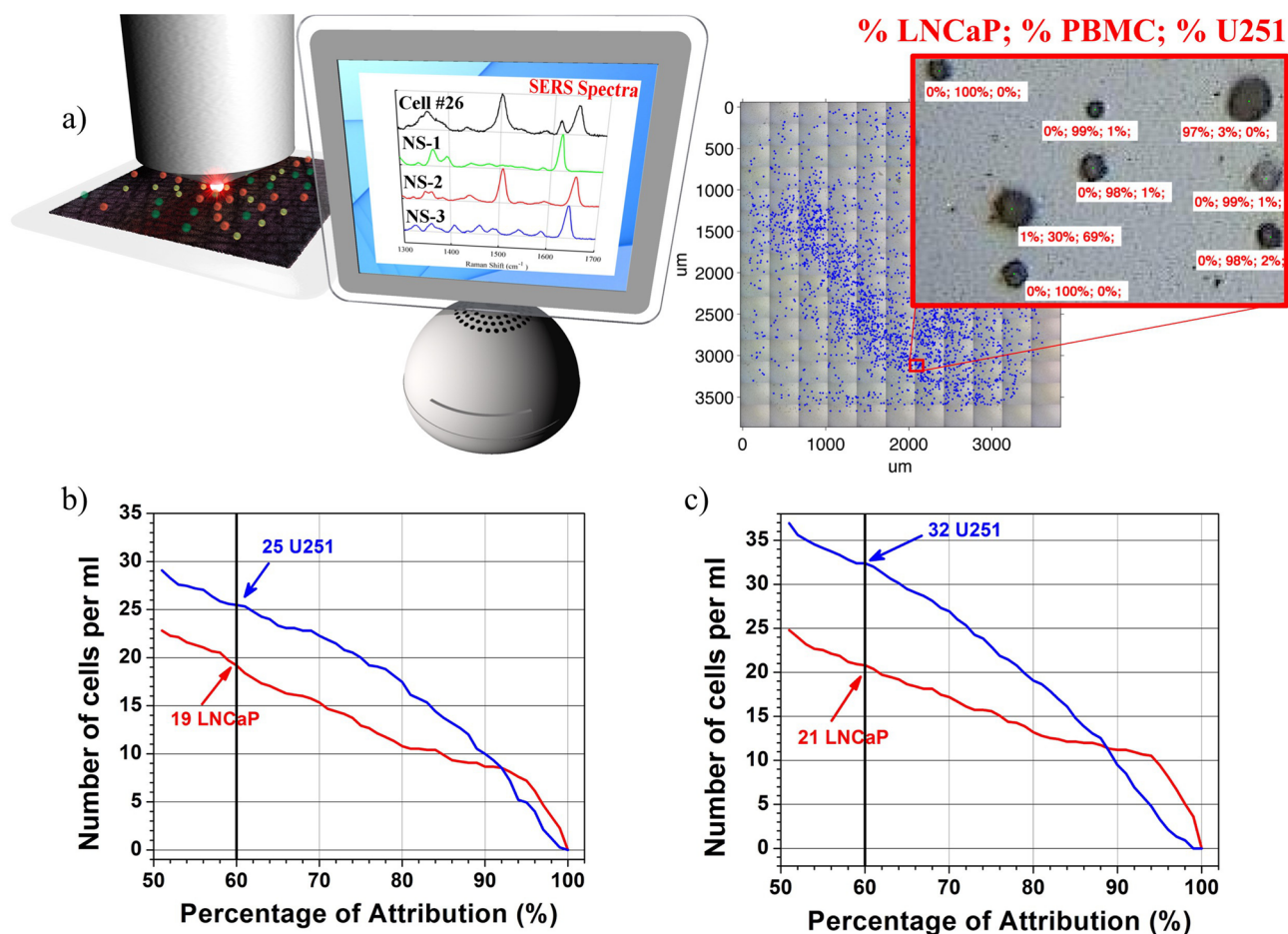
Figure 3. (**a**) Schematic representation of μ-Raman recording of the SERRS spectra of cells in a chamber slide. Thousands of cells were analyzed one-by-one and their SERRS spectra compared with reference ones. A Random Forest ensemble learning method was used to classify each cell as LNCaP, PBMC or U251. (**b**) and (**c**) The number of cells identified as a function of percentage of attribution given by RF for Mix1 and Mix 2 respectively.

buffer saline, 50 mM PO_4^{3-} and 150 mM NaCl at pH 7.4 added with EDTA 4 mM) solution. The functionalization grade was determined through a titration with 2,4-dithiodipyridine, resulting between 1 and 2 thiols per antibody.

Per ml	Mix1	Mix2
Spiked LNCaP	24 ± 1	24 ± 1
Spiked U251	29 ± 1	29 ± 1
Fixed LNCaP	20 ± 2	20 ± 2
Fixed U251	24 ± 2	24 ± 2
Identified LNCaP[a]	19 ± 1	21 ± 1
Identified U251[a]	25 ± 5	32 ± 6

Table 1. Number of spiked, fixed and identified tumor cells in Mix1 and Mix2. [a]Based on a 60% percentage of attribution.

Synthesis of NS-1, NS-2 and NS-3. For NS-1, 8 ml of laser ablated AuNPs with 2.5 nM concentration, were aggregated by centrifugation at 25,000 RCF and re-dispersed in 1 ml of bidistilled water by sonication. 100 μl of a 50 μM hydro-alcoholic solution of thiolated Malachite Green[31] (MG-SH) were added. The unbounded dye was removed by centrifugation. Then, 0.04 and 0.1 mg of thiolated anti-CD44 and anti-N-Cad were added. The reaction proceeded for 5 h at room temperature and overnight at 4 °C. The excess of antibodies was removed by centrifugation and the nanostructures re-dispersed in 1 ml of PBS. 100 μl of 0.5 g/l mPEG5000-

SH (α-methoxy-ω-mercaptopolyethylenglycol, 5 kDa, Rapp Polymere) was added to complete the coverage of the gold surfaces. After 3 h of mixing at room temperature the nanostructures were purified by centrifugation. Finally, they were re-dispersed in 250 μl of PBS also adding 7 mg of BSA (Bovine Serum Albumin, Sigma Aldrich). The presence of antibodies on the surface of nanostructure was verified by the spectrophotometric analysis of supernatant after the reaction. The decrease of the characteristic absorption band at 280 nm of antibodies was used for determining the reaction yield.

The synthesis of NS-2 and NS-3 were performed following the same procedures, but changing both the SERRS reporter and the targeting Abs.

To construct NS-2 we used 100 μl of a 50 μM solution of thiolated Texas Red[31] (TR-SH) as SERRS reporter, 0.05 mg of anti-EpCAM and 0.09 mg of anti-E-Cad, whereas, for NS-3, 100 μl of a 50 μM solution of thiolated Nile Blue[54] (NB), and 0.14 mg of anti-CD45.

UV–Vis-NIR and SERRS spectra for the three nanostructures are reported in Fig. 1b,c. TEM images are reported in Supplementary Fig. S1.

Cell cultures. LNCaP (human prostate cancer cells) and U251 (human malignant glioblastoma cells) were obtained from American Type Culture Collection (ATCC, Rockville, USA). The cells were maintained at 37 °C in a humidified atmosphere containing 5% CO_2 and 90% of humidity. LNCaP cells were cultured in RPMI 1640 Medium supplemented with 2 mM L-Glutamine and 0.01 M HEPES. U251 cells were cultured in Dulbecco Modified Eagle's Medium (DMEM) containing 3.7 g/l sodium bicarbonate, 4.5 g/l glucose and supplemented with Glutamine and HEPES. Moreover, all the media were supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Invitrogen, NY, USA) and antibiotics (0.1 mg/ml streptomycin and 100 units/ml penicillin G, Sigma-Aldrich, St Louis, MO, USA). Before the experiments, cells were detached with PBS-EDTA to obtain a monocellular suspension. The cell dilutions applied to create spiked samples were counted ten times by a Neubauer chamber.

Blood was obtained from healthy donor and the study was approved by CRO (Centro di Riferimento Oncologico, Aviano, Italy) Ethical Committee (code: IRB-18-2011) and all participants gave their informed consent. Moreover, all experiments were performed in accordance with relevant guidelines and regulations. PBMCs were obtained from whole blood by a standard density gradient centrifugation according to the supplier guidelines (Ficoll-Paque, GE Healthcare)[58].

Incubation and trapping. To create the chamber for samples trapping/reading we used avidin-coated glass slide 75 × 25 mm (TRIDIA BA Slides, SurModics). To reduce the working chamber area to a 3.5 × 3.5 mm^2 well we applied an adhesive multiwell gasket (FlexWell 64, GRACE BIO-LABS). After rehydration, the well surface was activated by incubation with a cocktail of biotinylated Abs (i.e. anti-EpCAM, anti-E-Cad, anti-N-Cad and anti-CD44 antigens) to trap the desired cell populations. We used around 0.25 μg of each Abs per chamber, with an incubation times of 30 min at room temperature. After three washing steps, cell samples (i.e. spiked samples, PBMCs only and tumor cells only) were incubated at room temperature for two hours. Then cells were fixed for one hour with PFA/Sucrose 4%, neutralized with Tris 10 mM pH = 8.0, washed with PBS and quenched with FBS. Staining was performed at room temperature for one hour using the targeting nanosystems preincubated in PBS and non-fat dry milk 5%. Finally, the chamber was washed with PBS and prepared for SERRS analysis.

μ-Raman measurements. SERRS spectra were recorded with the μRaman instrument inVia Renishaw, equipped with a motorized stage and with excitation at 632.8 nm using 3 mW measured on sample plate. A 20 × LEICA objective and 1 s acquisition time were used. Since cell mixtures were fixed on the bottom of 3.5 × 3.5 mm^2 wells, the entire surface was considered. Raman spectra were recorded for each single cell. Pattern recognition algorithms were used for cells identification in the chambers. The full colored image of a chamber was transformed into a binary image and a morphological operator detected specific cell patterns. Cells dimensions were also obtained by summing the image pixels of each identified cell. The centroid of each cell was recorded as xy coordinates and passed to the instrument for the SERRS measurements. All algorithms were developed in Matlab with graphical interfaces.

Data analysis. For building the Random Forest (RF) classifier we used the RandomForest package of R[59,60]. We used the Pearson correlation coefficients to compare the recorded spectra of each cell with the SERRS signals of NS-1, NS-2 and NS-3. We considered only cells with diameter in the range 30–180 $μm^2$. The pairwise Mann–Whitney test with the HOLM correction for multiplicity was applied in order to detect significant differences in terms of cell diameter between LNCaP, PBMC and U251. The RF classifier was built using 5000 trees, with the number of variables tried at each split, i.e. the mtry parameter, fixed at the default value, here equal to 3. The out of bag (OOB) error rate of the final RF model was 11.41% whereas the class error rates were: 6.6%, 11.3% and 20.1% for LNCaP, PBMC and U251, respectively. Adding interactions and squares or cubes of the features did not help improving the OOB error rate or the class error rates.

References

1. Di Meo, A., Bartlett, J., Cheng, Y. F., Pasic, M. D. & Yousef, G. M. Liquid biopsy: A step forward towards precision medicine in urologic malignancies. *Mol. Cancer* https://doi.org/10.1186/s12943-017-0644-5 (2017).
2. Plaks, V., Koopman, C. D. & Werb, Z. Circulating tumor cells. *Science* **341**, 1186–1188. https://doi.org/10.1126/science.1235226 (2013).
3. Ashworth, T. R. A case of cancer in which cells similar to those in the tumours were seen in the blood after death. *Aust. Med. J.* **14**, 146–149 (1869).
4. Clare, S. E. *et al.* Prognostic significance of occult lymph node metastases in node-negative breast cancer. *Ann. Surg. Oncol.* **4**, 447–451 (1997).
5. Braun, S. *et al.* Cytokeratin-positive cells in the bone marrow and survival of patients with stage I, II, or III breast cancer. *N. Engl. J. Med.* **342**, 525–533 (2000).
6. Cabel, L. *et al.* Circulating tumor cells: Clinical validity and utility. *Int. J. Clin. Oncol.* **22**, 421–430. https://doi.org/10.1007/s10147-017-1105-2 (2017).
7. Liu, S. *et al.* Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Rep.* **2**, 78–91 (2014).
8. Hong, Y. P. & Zhang, Q. Phenotype of circulating tumor cell: Face-off between epithelial and mesenchymal masks. *Tumor Biol.* **37**, 5663–5674. https://doi.org/10.1007/s13277-016-4796-5 (2016).
9. Yu, M. *et al.* Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**, 580–584. https://doi.org/10.1126/science.1228522 (2013).
10. Alunni-Fabbroni, M. & Sandri, M. T. Circulating tumour cells in clinical practice: Methods of detection and possible characterization. *Methods* **50**, 289–297 (2010).
11. Alix-Panabieres, C. & Pantel, K. OPINION challenges in circulating tumour cell research. *Nat. Rev. Cancer* **14**, 623–631. https://doi.org/10.1038/nrc3820 (2014).
12. Shen, Z. Y., Wu, A. G. & Chen, X. Y. Current detection technologies for circulating tumor cells. *Chem. Soc. Rev.* **46**, 2038–2056. https://doi.org/10.1039/c6cs00803h (2017).
13. Cristofanilli, M. *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N. Engl. J. Med.* **351**, 781–791. https://doi.org/10.1056/NEJMoa040766 (2004).
14. Cohen, S. J. *et al.* Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *J. Clin. Oncol.* **26**, 3213–3221. https://doi.org/10.1200/jco.2007.15.8923 (2008).
15. de Bono, J. S. *et al.* Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin. Cancer Res.* **14**, 6302–6309. https://doi.org/10.1158/1078-0432.ccr-08-0872 (2008).
16. Yoon, H. J., Kozminsky, M. & Nagrath, S. Emerging role of nanomaterials in circulating tumor cell isolation and analysis. *ACS Nano* **8**, 1995–2017 (2014).
17. Wang, X. *et al.* Detection of circulating tumor cells in human peripheral blood using surface-enhanced Raman scattering nanoparticles. *Can. Res.* **71**, 1526–1532. https://doi.org/10.1158/0008-5472.CAN-10-3069 (2011).
18. Dreaden, E. C., Alkilany, A. M., Huang, X., Murphy, C. J. & El-Sayed, M. A. The golden age: Gold nanoparticles for biomedicine. *Chem. Soc. Rev.* **41**, 2740–2779 (2012).
19. Lane, L. A., Qian, X. & Nie, S. SERS nanoparticles in medicine: From label-free detection to spectroscopic tagging. *Chem. Rev.* **115**, 10489–10529 (2015).
20. Feliu, N. *et al.* SERS quantification and characterization of proteins and other biomolecules. *Langmuir* **33**, 9711–9730. https://doi.org/10.1021/acs.langmuir.7b01567 (2017).
21. Ru Le, E. & Etchegoin, P. *Principles of Surface-Enhanced Raman Spectroscopy: And Related Plasmonic Effects* (Elsevier, Amsterdam, 2008).
22. Liu, R. Y. *et al.* Click-functionalized SERS nanoprobes with improved labeling efficiency and capability for cancer cell imaging. *ACS Appl. Mater. Interfaces.* **9**, 38222–38229. https://doi.org/10.1021/acsami.7b10409 (2017).
23. Wu, X. *et al.* Improved SERS-active nanoparticles with various shapes for CTC detection without enrichment process with super-sensitivity and high specificity. *ACS Appl. Mater. Interfaces.* **8**, 19928–19938. https://doi.org/10.1021/acsami.6b07205 (2016).
24. Zhang, P., Zhang, R., Gao, M. & Zhang, X. Novel nitrocellulose membrane substrate for efficient analysis of circulating tumor cells coupled with surface-enhanced Raman scattering imaging. *ACS Appl. Mater. Interfaces.* **6**, 370–376. https://doi.org/10.1021/am404406c (2014).
25. Biscaglia, F. *et al.* Enhanced EGFR targeting activity of plasmonic nanostructures with engineered GE11 peptide. *Adv. Healthc. Mater.* https://doi.org/10.1002/adhm.201700596 (2017).
26. Bertorelle, F. *et al.* Safe core-satellite magneto-plasmonic nanostructures for efficient targeting and photothermal treatment of tumor cells. *Nanoscale* **10**, 976–984. https://doi.org/10.1039/c7nr07844g (2018).
27. Litti, L. *et al.* A SERRS/MRI multimodal contrast agent based on naked Au nanoparticles functionalized with a Gd(III) loaded PEG polymer for tumor imaging and localized hyperthermia. *Nanoscale* **10**, 1272–1278. https://doi.org/10.1039/c7nr07398d (2018).
28. Jimenez de Aberasturi, D., Henriksen-Lacey, M., Litti, L., Langer, J. & Liz-Marzán, L. M. Using SERS tags to image the three-dimensional structure of complex cell models. *Adv. Funct. Mater.* https://doi.org/10.1002/adfm.201909655 (2020).
29. Wu, J. *et al.* Bioorthogonal SERS nanoprobes for mulitplex spectroscopic detection, tumor cell targeting, and tissue imaging. *Chem. A Eur. J.* **21**, 12914–12918 (2015).
30. Wang, Y. W., Kang, S., Khan, A., Bao, P. Q. & Liu, J. T. C. In vivo multiplexed molecular imaging of esophageal cancer via spectral endoscopy of topically applied SERS nanoparticles. *Biomed. Opt. Express* **6**, 3714–3723 (2015).
31. Meneghetti, M. *et al.* Plasmonic nanostructures for SERRS multiplexed identification of tumor-associated antigens. *Small* **8**, 3733–3738. https://doi.org/10.1002/smll.201201196 (2012).
32. Amendola, V. & Meneghetti, M. Exploring how to increase the brightness of surface-enhanced raman spectroscopy nanolabels: The effect of the Raman-active molecules and of the label size. *Adv. Funct. Mater.* **22**, 353–360 (2012).
33. dos Santos, D. P., Temperini, M. L. A. & Brolo, A. G. Mapping the energy distribution of SERRS hot spots from anti-stokes to stokes intensity ratios. *J. Am. Chem. Soc.* **134**, 13492–13500 (2012).
34. Biscaglia, F. *et al.* Gold nanoparticle aggregates functionalized with cyclic RGD peptides for targeting and imaging of colorectal cancer cells. *ACS Appl. Nano Mater.* **2**, 6436–6444. https://doi.org/10.1021/acsanm.9b01392 (2019).
35. Biscaglia, F. *et al.* PreS1 peptide-functionalized gold nanostructures with SERRS tags for efficient liver cancer cell targeting. *Mater. Sci. Eng. C Mater. Biol. Appl.* https://doi.org/10.1016/j.msec.2019.109762 (2019).
36. Goldberg, M. S. Immunoengineering: How nanotechnology can enhance cancer immunotherapy. *Cell* **161**, 201–204 (2015).
37. Cheung, A. S. & Mooney, D. J. Engineered materials for cancer immunotherapy. *Nano Today* **10**, 511–531 (2015).
38. Zhou, Y. *et al.* Impact of single-chain Fv antibody fragment affinity on nanoparticle targeting of epidermal growth factor receptor-expressing tumor cells. *J. Mol. Biol.* **371**, 934–947 (2007).
39. Wu, X. X. *et al.* Improved SERS nanoparticles for direct detection of circulating tumor cells in the blood. *ACS Appl. Mater. Interfaces.* **7**, 9965–9971. https://doi.org/10.1021/acsami.5b02276 (2015).

8

40. Sha, M. Y., Xu, H. X., Natan, M. J. & Cromer, R. Surface-enhanced raman scattering tags for rapid and homogeneous detection of circulating tumor cells in the presence of human whole blood. *J. Am. Chem. Soc.* **130**, 17214. https://doi.org/10.1021/ja804494m (2008).
41. Shi, W., Paproski, R. J., Moore, R. & Zemp, R. Detection of circulating tumor cells using targeted surface-enhanced Raman scattering nanoparticles and magnetic enrichment. *J. Biomed. Opt.* https://doi.org/10.1117/1.jbo.19.5.056014 (2014).
42. Bhana, S., Chaffin, E., Wang, Y. M., Mishra, S. R. & Huang, X. H. Capture and detection of cancer cells in whole blood with magnetic-optical nanoovals. *Nanomedicine* **9**, 593–606. https://doi.org/10.2217/nnm.13.77 (2014).
43. Sun, C. L., Zhang, R., Gao, M. X. & Zhang, X. M. A rapid and simple method for efficient capture and accurate discrimination of circulating tumor cells using aptamer conjugated magnetic beads and surface-enhanced Raman scattering imaging. *Anal. Bioanal. Chem.* **407**, 8883–8892. https://doi.org/10.1007/s00216-015-9049-8 (2015).
44. Pang, Y., Wang, C., Xiao, R. & Sun, Z. Dual-selective and dual-enhanced SERS nanoprobes strategy for circulating hepatocellular carcinoma cells detection. *Chem. Eur. J.* **24**, 7060–7067. https://doi.org/10.1002/chem.201801133 (2018).
45. Nima, Z. A. *et al.* Circulating tumor cell identification by functionalized silver-gold nanorods with multicolor, super-enhanced SERS and photothermal resonances. *Sci. Rep.* https://doi.org/10.1038/srep04752 (2014).
46. Wang, Z. Y. *et al.* SERS-fluorescence joint spectral encoded magnetic nanoprobes for multiplex cancer cell separation. *Adv. Healthc. Mater.* **3**, 1889–1897. https://doi.org/10.1002/adhm.201400092 (2014).
47. Tsao, S. C. H. *et al.* Characterising the phenotypic evolution of circulating tumour cells during treatment. *Nat. Commun.* https://doi.org/10.1038/s41467-018-03725-8 (2018).
48. Kalluri, R. & Weinberg, R. A. The basics of epithelial–mesenchymal transition. *J. Clin. Investig.* **119**, 1420–1428. https://doi.org/10.1172/JCI39104 (2009).
49. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
50. Amendola, V. & Meneghetti, M. Controlled size manipulation of free gold nanoparticles by laser irradiation and their facile bio-conjugation. *J. Mater. Chem.* **17**, 4705–4710 (2007).
51. Amendola, V., Polizzi, S. & Meneghetti, M. Free silver nanoparticles synthesized by laser ablation in organic solvents and their easy functionalization. *Langmuir* **23**, 6766–6770 (2007).
52. Amendola, V., Litti, L. & Meneghetti, M. LDI-MS assisted by chemical-free gold nanoparticles: Enhanced sensitivity and reduced background in the low-mass region. *Anal. Chem.* **85**, 11747–11754 (2013).
53. Litti, L. & Meneghetti, M. Predictions on the SERS enhancement factor of gold nanosphere aggregate samples. *Phys. Chem. Chem. Phys.* **21**, 15515–15522. https://doi.org/10.1039/c9cp02015b (2019).
54. Sciutto, G. *et al.* Alternative SERRS probes for the immunochemical localization of ovalbumin in paintings: An advanced mapping detection approach. *Analyst* **138**, 4532–4541 (2013).
55. Bulfoni, M. *et al.* In patients with metastatic breast cancer the identification of circulating tumor cells in epithelial-to-mesenchymal transition is associated with a poor prognosis. *Breast Cancer Res.* **18**, 1 (2016).
56. Allard, W. J. *et al.* Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin. Cancer Res.* **10**, 6897–6904. https://doi.org/10.1158/1078-0432.ccr-04-0378 (2004).
57. Hao, S. J., Wan, Y., Xia, Y. Q., Zou, X. & Zheng, S. Y. Size-based separation methods of circulating tumor cells. *Adv. Drug Deliv. Rev.* **125**, 3–20. https://doi.org/10.1016/j.addr.2018.01.002 (2018).
58. Böyum, A. Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g. *Scand. J. Clin. Lab. Investig. Suppl.* **97**, 77–89 (1968).
59. Liaw, A. & Wiener, M. Classification based on a forest of classification trees using random inputs. *R News* **2/2**, 43 (2002).
60. Team, R. C. Vol. https://www.R-project.org. (Vienna, Austria, 2016).

Acknowledgements

Author contributions

L.L., A.C., A.S. prepared the nanostructures and performed the SERRS experiments, M.P. carried out experiments with cells, E.R. and L.V. performed the R.F. calculations. G.T. and M.C. contributed to the research general approach, G.F. and M.M. conceived and designed the experimental design and setup. All authors provided contributions and comments to the paper writing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-72911-w.

Correspondence and requests for materials should be addressed to G.F. or M.M.

Reprints and permissions information is available at www.nature.com/reprints.

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