

SERS analysis of serum for detection of early and locally advanced breast cancer

Silvia Cervo^{a,b}, Elena Mansutti^{c,d}, Greta Del Mistro^d, Riccardo Spizzo^c, Alfonso Colombatti^{c,e}, Agostino Steffan^{a,b}, Valter Sergo^d and Alois Bonifacio^{d,*}

^a CRO-Biobank, CRO Aviano, National CancerInstitute, Via Franco Gallini 2 - 33081 – Aviano (PN), Italy

^b ClinicalCancerPathology, CRO Aviano, National CancerInstitute, Via Franco Gallini 2 - 33081 – Aviano (PN), Italy

^c Division of Experim. Oncology 2, CRO Aviano, National CancerInstitute, Via Franco Gallini 2 - 33081 – Aviano (PN), Italy

^d Dept. of Engineering and Architecture, University of Trieste, Via Valerio 6a - 34127 – Trieste (TS), Italy

^e Centre of Excellence MAT Dept. of Medical and Biological Science, University of Udine, P.le M. Kolbe 4 - 33010 - Udine, Italy.

* Corresponding Author: Alois Bonifacio, tel.: +39-040-558-3768, fax: +39-040-572044. e-mail: abonifacio@units.it.

ABSTRACT

In this contribution we investigated whether Surface-Enhanced Raman Scattering (SERS) of serum can be a candidate method for detecting “luminal A” breast cancer (BC) at different stages. We selected three groups of participants aged over 50 years: 20 healthy women, 20 women with early localized small BC and 20 women affected by BC with lymph node involvement. SERS revealed clear spectral differences between these three groups. A predictive model using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) was developed based on spectral data, and its performance was estimated with cross-validation. PCA-LDA of SERS spectra could distinguish healthy from BC subjects (sensitivity: 92%; specificity: 85%), as well as subjects with BC at different stages, with a promising diagnostic performance (sensitivity and specificity $\geq 80\%$; overall accuracy: 84%). Our data suggest that SERS spectroscopy of serum, combined with multivariate data analysis,

represents a minimally invasive, easy to use and fast approach to discriminate healthy from BC subjects, and even to distinguish BC at different clinical stages.

KEYWORDS: Raman, SERS, breast cancer, serum, nanoparticles, diagnosis

INTRODUCTION

Breast cancer (BC) is the most common cause of cancer death in women worldwide [1], as well as the most common noncutaneous cancer in European women [2, 3]. In Europe, BC is more frequently diagnosed in women aged over 50 years, with 49% of cases in the age range 50-69 and 31% of cases in the range 70-99 [4]. At diagnosis, early-stage tumors are more frequent than advanced-stage ones: tumors confined to breast (N0M0) represent 49.9% of diagnosed BC, while 32.8% of cases are locally advanced tumors, i.e. tumors spread to regional lymph nodes (N+M0) [4]. Metastatic tumors spread to distant organs (M1) are rarer, counting only for 5% of cases [4].

BC is a molecularly heterogeneous disease; indeed, gene expression profiling of BC has identified two biologically distinct estrogen receptor (ER)-positive subtypes of BC: luminal A and luminal B. Luminal A subtype is defined as being ER positive, HER2 (Human Epidermal growth factor Receptor 2) negative, and Ki67 low ($\leq 14\%$ cells positive), whereas luminal B subtype is ER positive, often negative, and Ki67 high ($\geq 14\%$ cells positive) [5, 6]. HER2 is an oncogene associated to BC; in recent years this protein has become an important biomarker and target of therapy for almost one third of BC patients [7]. The Ki67 protein is a nuclear marker of cell proliferation associated with worse outcomes [8]. Biological subtyping by immunohistochemical determination of these four biomarkers (i.e. ER, PR, HER2, and Ki67) has statistically significant value in identifying high-risk women with hormone receptor-positive BC, independently of standard clinicopathological parameters (including age at diagnosis, tumor size, grade, lymphovascular invasion, and axillary lymph node status) [6]. Luminal A tumors have lower proliferation and better prognosis than luminal B tumors [9, 10] and are more common [6].

The gold standard for screening programs of BC is mammography in women above 50 years and breast ultrasound in women below 50 years [11]. However, despite substantial increase in the number of cases of early-stage BC detected, screening mammography has only marginally reduced the rate at which women present with an advanced cancer (8% in the US) [12]. This imbalance suggests

that: i) there is substantial overdiagnosis (i.e. the diagnosis of a "cancer" that would otherwise not go on to cause symptoms or death), and consequently overtreatment, accounting for nearly a third of all newly diagnosed BCs; ii) screening has only a minimal effect on the rate of death from BC [12]. These results imply that mammography screening entails a substantial risk of detecting tumors that would not have become symptomatic during a woman's lifetime if no screening had taken place, but which are nevertheless treated. BC is thus an ideal setting to evaluate alternative screening systems which can discriminate between aggressive and non-aggressive cancer, to reduce overtreatment without decreasing survival rates.

There is a growing interest for the application of Surface-enhanced Raman scattering (SERS) in diagnostics [13–17]. In this technique, detection is achieved upon adsorption of the analytes of interest onto a nanostructured metal surface (i.e. a SERS substrate) having adequate plasmonic properties, followed by illumination with a laser light, and analysis of its inelastic scattering using a spectrograph. Aqueous colloidal dispersions of Au or Ag nanoparticles are often used as SERS substrates, and they are mixed together with the analytes solution. The bands in the SERS spectrum thus obtained are related to the molecular vibrations of the chemical species adsorbed on the substrate. Since such a spectrum (like Raman or IR spectra) is unique to each molecular structure, it is often considered as a kind of “molecular fingerprint”, which is sensitive to changes in the chemical composition of the sample, and could thus detect differences between samples belonging to diseased and healthy subjects.

Recently, several groups reported SERS spectra of blood plasma or serum [16, 18–22], often with diagnostic purposes. We also reported that using metal nanoparticles as SERS substrates in the absence of aggregating agents, filtration of proteins from serum or plasma is important to rapidly obtain intense, repeatable SERS spectra of these biofluids [19]. Furthermore, that study suggested that serum should be preferred over plasma, because of the possible spectral interference of anticoagulants, and established experimental conditions for obtaining reproducible SERS spectra of serum.

In the present study, we investigated whether SERS spectroscopy can discriminate serum from healthy subjects from that of BC patients, and whether it can further distinguish between BC at different stages, i.e. with or without lymph node involvement. However, diagnostically useful information, in SERS spectra of a complex chemical mixture such as serum, is often hidden in a wealth of spectral data. To extract such meaningful information and use it for diagnostic purposes, we used multivariate statistical methods, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) [23, 24], developing a predictive model and estimating its performance with cross-validation.

MATERIALS AND METHODS

Cohort of samples

Since SERS spectra depend on metabolites which can be influenced by several factors, our strategy was to strictly select homogeneous groups of subjects in order to reduce these factors. We aimed to focus on the most widespread conditions, i.e. luminal A histotype BC in menopausal patients [4], grouping tumors with different prognosis according to stage: i) early localized small tumors (pT1N0M0): better prognosis; ii) advanced tumors with lymph node involvement (pTxN+M0): worse prognosis.

We thus selected 3 groups of participants: 20 healthy women (age range 61.2 ± 8.1); 20 women with luminal A BC, stage: pT1N0 (age range 61.6 ± 9.0); and 20 women with luminal A BC, stage: pTxN+ (age range 61.8 ± 8.2). We selected patients over 50 years since this is the most frequent age for BC diagnosis and a target age for mammary screening [4]. Serum samples were collected at diagnosis, so that no therapies could affect the results of our analysis. All participants of the three groups were matched for variables related to BC development (Table S1): i) age (individual matching up to 4 age difference); ii) menopausal status (all patients are in menopause); iii) Body Mass Index (BMI; frequency matching); iv) smoking habits (frequency of smokers or ex-smokers vs. frequency of non-smokers).

All serum samples of the 60 subjects were prepared and stored by CRO-Biobank (the biobank of the CRO National Cancer Institute, Aviano, Italy), according to Cervo et al. [25]. Briefly, blood samples were collected in Serum Z tubes (Monovette®, Sarstedt). After collection, blood tubes were promptly placed in ice and centrifuged at 2608 g for 10 minutes at room temperature to separate serum. Aliquots of serum were then stored at -80°C .

All participants provided written informed consent to participate. The CRO-Biobank project has been approved by the CRO Institutional Ethics Committee.

Serum processing

Serum aliquots were thawed at room temperature and immediately filtered (Amicon Ultra 3K Centrifugal filter devices, cutoff 3 kDa) by centrifugation at 14000 g, 4°C for 75 minutes to enhance SERS spectrum, as described in [19]. All samples were stored at -80°C until SERS analysis.

SERS substrates preparation and characterization

The SERS spectra were acquired using hydroxylamine-reduced Ag nanoparticles (h-AgNP), prepared by the Leopold-Lendl method [26]. Briefly, 0.017 g of AgNO₃ were dissolved in 90 mL milliQ water, to which a solution consisting of 0.021 g of hydroxylamine hydrochloride and 18 mg of NaOH dissolved in 9.5 mL of milliQ water was rapidly added. After few seconds silver colloids were obtained which were stable for 1 week (stored in dark at RT).

All chemicals were purchased from Sigma-Aldrich and used as received. All glassware used for h-AgNP preparation was carefully cleaned with strong acids (concentrated HNO₃ for Ag colloids) and thoroughly rinsed with milliQ water. For all cleaning procedures and preparation of solutions, milliQ water was used. To check repeatability between different preparations, h-AgNP colloids were characterized by UV-visible absorption spectroscopy after each preparation (Fig. S1 in Supplementary Information), using a Lambda 20bio UV-Vis spectrometer (Perkin-Elmer, Monza, Italy). The extinction band maxima were at 410 nm, consistent with the values previously reported in literature [27], and characteristic of an average nanoparticle diameter of about 23 nm.

SERS instrumentation

SERS spectra were collected using an InVia Raman microscope (Renishaw plc, Wotton-under-Edge, UK) equipped with high-power near infrared diode laser (Toptica Photonics AG, Germany) emitting at 785 nm and delivering 120 mW of laser power at the sample. For data acquisition, the laser was focused on the sample via a 10x microscope objective (N.A. 0.25). A 1800 l/mm grating yielded a spectral resolution of 4 cm⁻¹. A thermoelectrically cooled charge coupled device (CCD) camera was used for detection. The spectrograph was calibrated using the lines of a Ne lamp, and the calibration was checked prior to each measurement using the 520 cm⁻¹ band of a silicon reference sample. Data were acquired using the software WiRE 3.4 (Renishaw).

Sample preparation for SERS measurements

Filtered serum aliquots were mixed with h-AgNP (prepared as described above and working as SERS substrate) in a 1.5 mL polypropylene tube, with a 1:9 biofluid-substrate ratio for a total volume of 50 μL (i.e. 4 μL + 45 μL). The mixtures obtained by adding the sample to the h-AgNP were rapidly transferred on a vacuum-UV quality CaF₂ microscope slide (Crystal GmbH, Berlin, Germany), and placed under the Raman microscope for spectral acquisition. CaF₂ is devoid of Raman bands in the spectral region above 320 cm⁻¹, and thus its use avoids any possible spectral interference due to the

substrate onto which the liquid sample is deposited. The laser was then focused on the top of the drop, and SERS spectra were immediately acquired with an exposure time of 10 s.

Data preprocessing

All data preprocessing was performed within the R software environment for statistical computing and graphics [28]. In particular, data import and export, preprocessing and visualization were performed with the *hyperSpec* package [29] for R. The preprocessing consisted of three steps: i) baseline correction ii) smoothing interpolation to project the spectra on an evenly spaced wavenumber axis (using the function *spc.loess* from package *hyperSpec*), and iii) intensity vector-normalization. For the baseline correction, a linear baseline was fit automatically to the whole spectral range and was subtracted from each spectrum of the dataset using function *modpolyfit* from package *baseline* [30], and peak picking was performed using *detectPeaks* from package *MALDIquant* [31].

Multivariate analysis of SERS spectra

All data analysis was performed within the R software environment for statistical computing and graphics [28]. LDA was combined with PCA, into an approach commonly called PCA-LDA (or PCLDA), to build a classification model for SERS spectra of serum. PCA was performed to extract the most meaningful information out of SERS spectral data, “compacting” the information contained in the spectra into a limited set of principal components (PCs) and thus reducing the number of variables to be used in LDA. The first 12 PCs, explaining the 99% of spectral variance (see Fig. S2 in Supplementary Information), were used in LDA (which was performed using the *lda* function of package *MASS* [32]). We stress the fact that, in this approach, PCA is only used to reduce the dimension of spectral data, and not for an exploratory analysis of the spectral differences between the different groups. The PCA-LDA model was then validated using the “Leave-One-Out” (LOO) cross validation technique: since there was one spectrum per patient, the validation corresponds to a “leave-one-patient-out”. It should be noted that for each cross validation step, PCA was carried out each time on the training set of data only (all sample but one, as expected in the LOO technique), while the PCs for the test sample were calculated, on the basis of the PCA done on the training set, using the *preProcess* function of package *caret* [33]. In other words, the training and test set were kept independent from each other with respect to PCA. The Receiver Operator Curve (ROC) was generated by changing the thresholds to determine incorrect and correct classifications for all samples, using the *ROCR* package [34].

RESULTS AND DISCUSSION

Average SERS spectra of serum samples of healthy subjects (control group) and of patients with a diagnosis of a Luminal A BC are shown in Fig. 1, along with their difference. Both spectra present features already reported by several groups [16, 18, 20, 21, 35], using various substrates, and in particular the bands due to uric acid and hypoxanthine [18, 19, 36].

The spectra from the two groups of subjects seem very similar, but the difference spectrum clearly reveals dissimilarities, where positive and negative peaks correspond, respectively, to metabolites present in higher and lower relative amounts in BC patients than in the control group. Some positive peaks, and in particular those at 721, 1093, 1324 and 1444 cm^{-1} , can be assigned to hypoxanthine [18, 19], which thus appears to be present in higher amounts in the serum samples of the BC group with respect to the control group. Besides the high number of metabolites present in serum (several thousands, reported so far [37]), spectral interpretation is further complicated by surface selection rules, which can cause SERS spectra to look very different from the corresponding Raman spectra, since the relative intensity of SERS bands depend on the adsorption geometry of the metabolite onto the surface of the nanoparticles. We are currently making efforts to identify the metabolites corresponding to the other positive and negative peaks. However, the scope of the present study is to assess whether the differences between serum SERS spectra of BC patients and of the control group are sufficient to differentiate the two groups for diagnostic purposes.

A separation between the two groups on the basis of SERS spectra was achieved by applying a PCA-LDA approach (Fig. S3). A validation of the PCA-LDA model (using LOO cross validation) yielded a confusion table (Table S2) from which sensitivity, specificity and accuracy were calculated (Table 1). Serum of subjects with BC can be distinguished from those of healthy controls with an overall accuracy of 90%, a sensitivity of 92% and a specificity of 85%, which are better than those achieved by mammography alone (sensitivity: 67.8%, specificity: 75%) and by mammography combined with clinical breast examination (sensitivity: 77.4%, specificity: 72%) [11]. The good performance of SERS combined with PCA-LDA as a diagnostic test for BC is also evidenced by the ROC curve (Fig. 2), having an AUC of 0.97.

Besides the ability to simply distinguish BC patients from the controls, we also assessed the capability of SERS to classify different types of BC, and in particular those with lymph node involvement from those without any involvement. The PCA-LDA applied to the same data but

considering three classes (i.e. pT1N0, pTxN+ and controls) instead of two (i.e. BC and controls) yielded promising results. Average SERS spectra of serum samples of the pT1N0 and the pTxN+ subjects are shown in Fig. 3, along with their difference. Both spectra present features in common with the spectra in Fig. 1, but clear differences are present between the two groups. Such dissimilarities, as inferred from the difference spectrum (Fig. 3, in grey), are not the same as those found in Fig. 1, suggesting that SERS spectra contain information useful to distinguish between the two stages of BC. In spite of the still limited biochemical interpretation of SERS bands in serum spectra, these preliminary results obtained from a limited number of subjects are encouraging, since PCA-LDA of SERS spectra can separate the three groups (Fig. 4), with a promising diagnostic performance. The validated PCA-LDA model yielded both sensitivities and specificities above 80%, and an overall accuracy of 84% (Table 2).

Interestingly, the method seems better at detecting early (pT1N0) than locally advanced (pTxN+) cancers. Moreover, separation is essentially achieved by the first discriminant function (Fig. 4), where the pTxN+, unexpectedly, appears to be “in between” the pT1N0 group and the controls. In other words, SERS spectra of serum from locally advanced tumors look more similar to spectra from the healthy subjects than those of the early cancers. This observation is very intriguing, as it suggests that those metabolites detected by SERS, which are discriminating between these two groups of patients, might not depend on the tumor burden (i.e. the total amount of tumor cells distributed throughout the body) *per se* but rather on the host. Such a situation is reminiscent of the immune response of the host which is more active in early compared to locally advanced tumors [38]. Otherwise, our results may suggest that the host is less "aware" of BC tumors which disseminated locally than of BC tumor which did not disseminate.

In conclusion, our data suggest that SERS spectroscopy of serum, in combination with multivariate data analysis, represents a promising approach to discriminate healthy subjects from subjects having BC with high sensitivity and specificity, and even to distinguish BC at different clinical stages, a feature which cannot be achieved with the gold standard method (mammography) currently used for screening programs. It should be stressed that, when compared with available diagnostic methods, SERS has the advantage of being minimally invasive, easy to use, fast and portable, all characteristics compatible with a point-of-care technology, ideal for screening purposes.

ACKNOWLEDGEMENTS

AB would like to thank Dr. Claudia Beleites for sharing her ideas on LDA applied to vibrational spectra.

REFERENCES

1. Key TJ, Verkasalo PK, Banks E (2001) Epidemiology of breast cancer. *Lancet Oncol* 2:133–140. doi: 10.1016/S1470-2045(00)00254-0
2. Parkin D., Bray F., Devesa S. (2001) Cancer burden in the year 2000. The global picture. *Eur J Cancer* 37:4–66. doi: 10.1016/S0959-8049(01)00267-2
3. Tyczynski J, Bray F, Parkin D ENCR Cancer Fact Sheet. *Eur Netw Cancer Regist* 2:1–4.
4. Allemani C, Minicozzi P, Berrino F, Bastiaannet E, Gavin A, Galceran J, Amejjide A, Siesling S, Mangone L, Ardanaz E, Hédelin G, Mateos A, Micheli A, Sant M, the EURO CARE Working Group (2013) Predictions of survival up to 10 years after diagnosis for European women with breast cancer in 2000-2002. *Int J Cancer* 132:2404–2412. doi: 10.1002/ijc.27895
5. Feeley LP, Mulligan AM, Pinnaduwege D, Bull SB, Andrulis IL (2014) Distinguishing luminal breast cancer subtypes by Ki67, progesterone receptor or TP53 status provides prognostic information. *Mod Pathol* 27:554–561. doi: 10.1038/modpathol.2013.153
6. Cheang MCU, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS, Perou CM, Ellis MJ, Nielsen TO (2009) Ki67 Index, HER2 Status, and Prognosis of Patients With Luminal B Breast Cancer. *JNCI J Natl Cancer Inst* 101:736–750. doi: 10.1093/jnci/djp082
7. Mitri Z, Constantine T, O’Regan R (2012) The HER2 Receptor in Breast Cancer: Pathophysiology, Clinical Use, and New Advances in Therapy. *Chemother Res Pract* 2012:1–7. doi: 10.1155/2012/743193
8. De Azambuja E, Cardoso F, de Castro G, Colozza M, Mano MS, Durbecq V, Sotiriou C, Larsimont D, Piccart-Gebhart MJ, Paesmans M (2007) Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients. *Br J Cancer* 96:1504–1513. doi: 10.1038/sj.bjc.6603756
9. Correa Geyer F, Reis-Filho JS (2009) Microarray-based Gene Expression Profiling as a Clinical Tool for Breast Cancer Management: Are We There Yet? *Int J Surg Pathol* 17:285–302. doi: 10.1177/1066896908328577
10. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, Desmedt C, Ignatiadis M, Sengstag T, Schutz F, Goldstein DR, Piccart M, Delorenzi M (2008) Meta-analysis of gene-

expression profiles in breast cancer: toward a unified understanding of breast cancer sub-typing and prognosis signatures. *Breast Cancer Res* 10:R65. doi: 10.1186/bcr2124

11. Alberts D, Hess LM (2008) *Fundamentals of Cancer Prevention*. Springer Science & Business Media
12. Bleyer A, Welch HG (2012) Effect of Three Decades of Screening Mammography on Breast-Cancer Incidence. *N Engl J Med* 367:1998–2005. doi: 10.1056/NEJMoa1206809
13. McAughtrie S, Faulds K, Graham D (2014) Surface enhanced Raman spectroscopy (SERS): Potential applications for disease detection and treatment. *J Photochem Photobiol C Photochem Rev* 21:40–53. doi: 10.1016/j.jphotochemrev.2014.09.002
14. Nima ZA, Biswas A, Bayer IS, Hardcastle FD, Perry D, Ghosh A, Dervishi E, Biris AS (2014) Applications of surface-enhanced Raman scattering in advanced bio-medical technologies and diagnostics*. *Drug Metab Rev* 46:155–175. doi: 10.3109/03602532.2013.873451
15. Driscoll AJ, Harpster MH, Johnson PA (2013) The development of surface-enhanced Raman scattering as a detection modality for portable in vitro diagnostics: progress and challenges. *Phys Chem Chem Phys* 15:20415. doi: 10.1039/c3cp52334a
16. Chen R, Lin J, Feng S, Huang Z, Chen G, Wang, Jing, Li, Yongzeng, Zeng, Haishan (2012) Applications of SERS Spectroscopy for Blood Analysis. In: Ghomi, Mahmoud (ed) *Adv. Biomed. Spectrosc.* IOS Press, Amsterdam, pp 72–105
17. Kho KW, Fu CY, Dinis US, Olivo M (2011) Clinical SERS: are we there yet? *J Biophotonics* 4:667–684. doi: 10.1002/jbio.201100047
18. Premasiri WR, Lee JC, Ziegler LD (2012) Surface-Enhanced Raman Scattering of Whole Human Blood, Blood Plasma, and Red Blood Cells: Cellular Processes and Bioanalytical Sensing. *J Phys Chem B* 116:9376–9386. doi: 10.1021/jp304932g
19. Bonifacio A, Dalla Marta S, Spizzo R, Cervo S, Steffan A, Colombatti A, Sergio V (2014) Surface-enhanced Raman spectroscopy of blood plasma and serum using Ag and Au nanoparticles: a systematic study. *Anal Bioanal Chem* 406:2355–2365. doi: 10.1007/s00216-014-7622-1
20. Li SX, Zhang YJ, Zeng QY, Li LF, Guo ZY, Liu ZM, Xiong HL, Liu SH (2014) Potential of cancer screening with serum surface-enhanced Raman spectroscopy and a support vector machine. *Laser Phys Lett* 11:065603. doi: 10.1088/1612-2011/11/6/065603
21. Lin D, Pan J, Huang H, Chen G, Qiu S, Shi H, Chen W, Yu Y, Feng S, Chen R (2014) Label-free blood plasma test based on surface-enhanced Raman scattering for tumor stages detection in nasopharyngeal cancer. *Sci Rep*. doi: 10.1038/srep04751
22. Casella M, Lucotti A, Tommasini M, Bedoni M, Forvi E, Gramatica F, Zerbi G (2011) Raman and SERS recognition of β -carotene and haemoglobin fingerprints in human whole blood. *Spectrochim Acta A Mol Biomol Spectrosc* 79:915–919. doi: 10.1016/j.saa.2011.03.048

23. Brereton RG (2003) *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. John Wiley & Sons
24. Varmuza K, Filzmoser P (2009) *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press
25. Cervo S, De Paoli P, Perin T, Canzonieri V, Steffan A (2015) Cost effective organization of an Institutional Human Cancer Biobank in a clinical setting: CRO-Biobank experience toward harmonization. *Br Med J*. (in press) doi: 10.5301/jbm.5000138
26. Leopold N, Lendl B (2003) A New Method for Fast Preparation of Highly Surface-Enhanced Raman Scattering (SERS) Active Silver Colloids at Room Temperature by Reduction of Silver Nitrate with Hydroxylamine Hydrochloride. *J Phys Chem B* 107:5723–5727. doi: 10.1021/jp027460u
27. Larmour IA, Faulds K, Graham D (2012) SERS activity and stability of the most frequently used silver colloids. *J Raman Spectrosc* 43:202–206. doi: 10.1002/jrs.3038
28. R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
29. Beleites C, Sergo V hyperSpec: a package to handle hyperspectral data sets in R. R package version 0.98-20140220. <http://hyperspec.r-forge.r-project.org>
30. Liland KH, Mevik B-H (2015) baseline: Baseline Correction of Spectra. R package version 1.1-2. <http://CRAN.R-project.org/package=baseline>
31. Gibb S, Strimmer K (2012) MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics* 28:2270–2271. doi: 10.1093/bioinformatics/bts447
32. Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*. Springer
33. Kuhn M. Contributions from Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, Team the RC, Benesty M, Lescarbeau R, Ziem A, Scrucca and L (2015) caret: Classification and Regression Training. R package version 6.0-37. <http://CRAN.R-project.org/package=caret>
34. Sing T, Sander O, Beerwinkler N, Lengauer T (2005) ROCR: visualizing classifier performance in R. *Bioinformatics* 21:3940–3941. doi: 10.1093/bioinformatics/bti623
35. Liu R, Zi X, Kang Y, Si M, Wu Y (2011) Surface-enhanced Raman scattering study of human serum on PVA-Ag nanofilm prepared by using electrostatic self-assembly. *J Raman Spectrosc* 42:137–144. doi: 10.1002/jrs.2665
36. Hu P, Zheng X-S, Zong C, Li M-H, Zhang L-Y, Li W, Ren B (2014) Drop-coating deposition and surface-enhanced Raman spectroscopies (DCDRS and SERS) provide complementary information of whole human tears: DCDRS and SERS provide complementary information of whole human tears. *J Raman Spectrosc* 45:565–573. doi: 10.1002/jrs.4499

37. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, Young N, Xia J, Knox C, Dong E, Huang P, Hollander Z, Pedersen TL, Smith SR, Bamforth F, Greiner R, McManus B, Newman JW, Goodfriend T, Wishart DS (2011) The Human Serum Metabolome. PLoS ONE 6:e16957. doi: 10.1371/journal.pone.0016957
38. Corthay A (2014) Does the Immune System Naturally Protect Against Cancer? Front Immunol. doi: 10.3389/fimmu.2014.00197

FIGURES AND TABLES

FIGURES

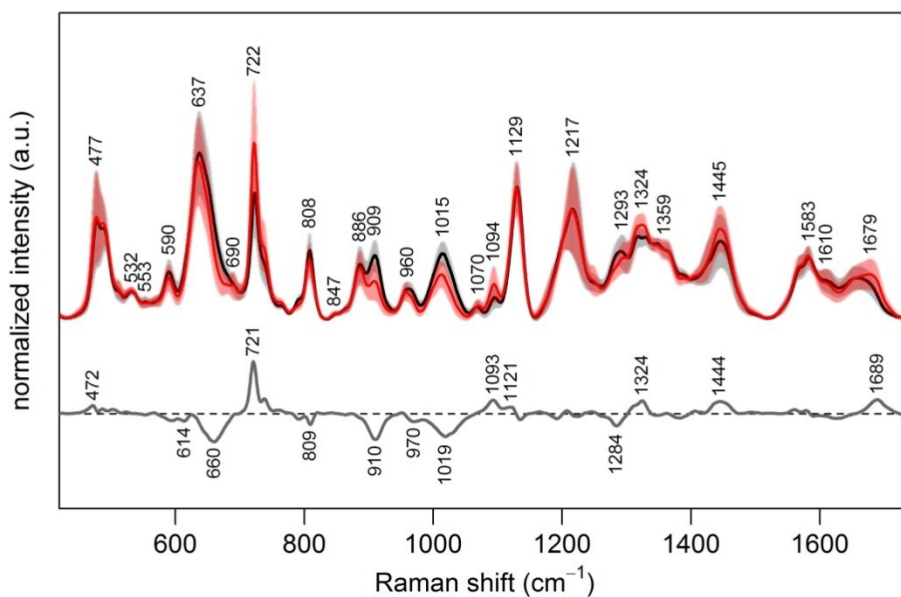


Fig. 1. Average normalized SERS spectra of serum samples of the BC group (red, $n = 40$) and of the Control group (black, $n = 20$), together with their difference spectra (BC – Control, grey). Intensity standard deviations are reported as pink and grey shaded areas.

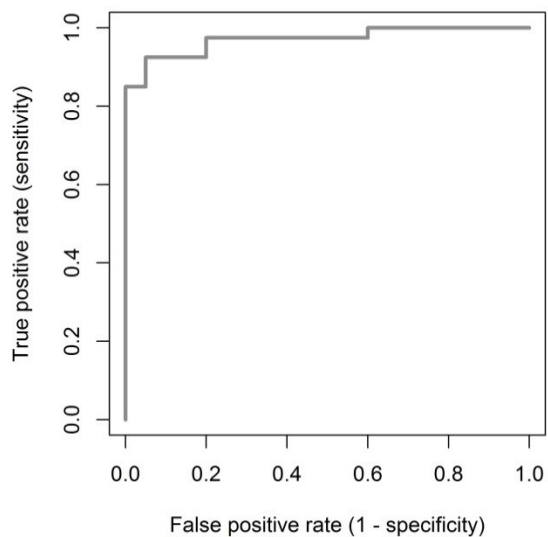


Fig. 2. Receiver Operating Characteristic (ROC) curve of discrimination results for serum SERS spectra out of the PCA-LDA-based spectral classification with LOO cross-validation method. The integration area under the ROC curve (Area Under Curve, AUC) is 0.97.

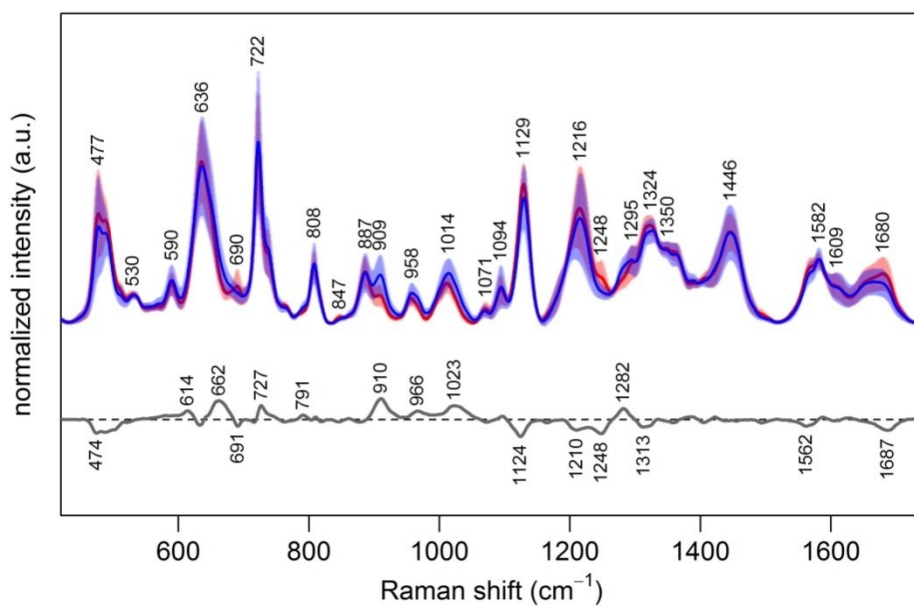


Fig. 3. Average normalized SERS spectra of serum samples of the pT1N0 group (without lymph node involvement, red, $n = 20$) and of the pTxN+ group (with lymph node involvement, blue, $n = 20$), together with their difference spectra (pT1N0 – pTxN+, grey). Intensity standard deviations are reported as shaded areas.

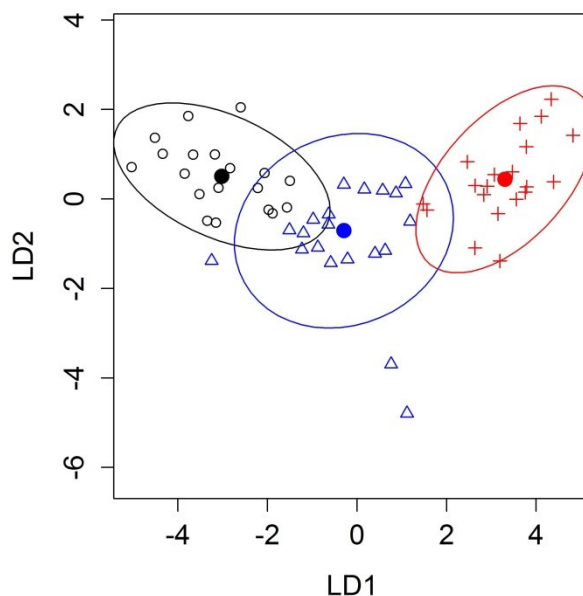


Fig. 4. Scatter plot of the scores of the first linear discriminant (LD1) against the second linear discriminant (LD2), describing the separation between the SERS spectra of the control group (black circles), the pTxN+ group (i.e. BC with lymph node involvement; blue triangles) and the pT1N0 group (i.e. BC, without lymph node involvement; red crosses) achieved by the PCA-LDA model. The ellipses mark the area that contains 95% of the samples (considering data with a bi-variate, normal distribution) for each group. Group centroids are indicated as colored filled circles. This plot is shown for descriptive (and not predictive) purposes only, since it was obtained from a non-validated PCA-LDA model, in which the training set was used as test set. For the predictive performance of the cross-validated model see Tables 2 and S3.

TABLES

	Control (%)	BC (%)
Sensitivity	85	92
Specificity	92	85
Accuracy	90	

Table 1. Diagnostic sensitivity and specificity of the LOO cross validated PCA-LDA classification algorithm, applied to SERS spectra of serum, for the classification between serum from Control patients and BC patients.

	Control	BC	
		pTxN+	pT1N0
Sensitivity	80	80	90
Specificity	95	85	95
Accuracy	84		

Table 2. Diagnostic sensitivity and specificity of the LOO cross validated PCA-LDA classification algorithm, applied to SERS spectra of serum, for the classification between serum from Control patients, BC patients with lymph node involvement (pTxN+) and BC patients without lymph node involvement (pT1N0).