Raman Spectroscopy reveals that biochemical composition of breast microcalcifications correlates with histopathological features

Renzo Vanna [1], Carlo Morasso [1], Beatrice Marcinnò [2], Francesca Piccotti [1], Emanuele Torti [2], Davide Altamura [3], Sara Albasini [1], Manuela Agozzino [4], Laura Villani [4], Luca Sorrentino[5], Oliver Bunk [6], Francesco Leporati [2], Cinzia Giannini [3], Fabio Corsi [5,7*]

[1] Nanomedicine and Molecular Imaging Lab, Istituti Clinici Scientifici Maugeri IRCCS, via Maugeri 4, 27100 Pavia, Italy

[2] Custom Computing and Processing Systems Laboratory, Dipartimento di Ingegneria Industriale e dell'Informazione, University of Pavia, via Ferrata 5, 27100 Pavia, Italy

[3] Institute of Crystallography, National Research Council, via Giovanni Amendola 122/O, 70126 Bari, Italy

[4] Pathology Unit, Istituti Clinici Scientifici Maugeri IRCCS, via Maugeri 4, 27100 Pavia, Italy

[5] Department of Biomedical and Clinical Sciences "Luigi Sacco", University of Milan, via G. B. Grassi 74, 20157 Milan, Italy

[6] Paul Scherrer Institut, 5232 Villigen-PSI, Switzerland

[7] Breast Unit, Department of Surgery, Istituti Clinici Scientifici Maugeri IRCCS, via Maugeri 4, 27100 Pavia, Italy

Running title:
Raman characterization of breast microcalcifications

***Corresponding author:** Fabio Corsi, MD, Breast Unit, Department of Surgery, Istituti Clinici Scientifici Maugeri IRCCS, via Maugeri 4, 27100 Pavia, Italy; telephone: +39 (0)382 592 277; e-mail: fabio.corsi@icsmaugeri.it

Disclosures: All authors declare that there is no conflict of interest regarding the publication of this article.

Abstract

Breast microcalcifications (MC) are a common mammographic finding. MC are considered suspicious signs of breast cancer and a breast biopsy is required, however, cancer is diagnosed in only a few patients. Reducing unnecessary biopsies and rapid characterization of breast MC are unmet clinical needs. In this study, 473 MC detected on breast biopsy specimens from 56 patients were characterized entirely by Raman mapping, and confirmed by X-ray scattering.

MC from malignant samples were generally more homogeneous, more crystalline and characterized by a less substituted crystal lattice compared to benign samples. There were significant differences in Raman features corresponding to the phosphate and carbonate bands between the benign and malignant groups. In addition to the heterogeneous composition, the presence of whitlockite specifically emerged as marker of benignity in benign MC.

The whole Raman signature of each MC was then used to build a classification model that distinguishes MC according to their overall biochemical composition. After validation, MC found in benign and malignant samples were correctly recognized with 93.5% sensitivity and 80.6% specificity. Finally, MC identified in malignant biopsies, but located outside the lesion, reported as well malignant features in 65% of in situ and 98% of invasive cancer cases, respectively, suggesting that the local microenvironment influences MC features. This study confirms that the composition and structural features of MC correlate with breast pathology, and indicates new diagnostic potentialities based on MC assessment.

INTRODUCTION

The clinical presentation of breast cancer has profoundly changed in recent years, thanks to the wide adoption of screening mammography. Therefore, a shift toward early breast cancer has been observed. In this context, breast microcalcifications (MC) are a common finding on mammography and they even increased following the evolution of digital imaging techniques, but only a small proportion of them reveals a breast cancer (1,2). This results in a positive predictive value (PPV) for biopsy with MC ranging between 20% and 40% (3) and it contributes to the overall high false positive rates of mammography, accounting for $2.8–4 billion per year in the United States only (4,5). Nowadays, the description of MC in clinics is only based on morphological features in mammograms or by their general appearance on stained histological samples (6,7). MC are classified as Type I and Type II. Type I MC (calcium oxalate, CO) is an uncommon type of mineralized component, mainly associated with benign lesions; Type II MC, composed by calcium phosphate (mainly by hydroxyapatite (HA)), are the most frequently observed, both in benign and malignant samples (6,8). Several computational efforts have been attempted to improve the assessment of MC during mammography using deep learning and other computer-aided tools (9) but these are not able to discriminant their molecular composition. Only recently, a contrast-phase X-ray mammography approach reported the possibility to study MC composition and to distinguish Type I and Type II MC (10), but what is missing is the assessment of Type II MC malignancy.

Raman spectroscopy (RS) is a non-destructive approach able to define the biochemical composition of samples in label-free (11) and it is particularly adequate to study minerals

components in biological samples (12). In addition, RS is compatible with real-time in-vivo diagnostic evaluations (13–16). Haka et al. reported for the first time that RS is able to recognize biochemical differences between benign and malignant MCs (74 and 16 calcifications, respectively) by carrying out single acquisitions at selected sites inside the calcified lesion (8). Subsequently, Raman-based tools were proposed in order to study MC in tissue. Stone, Matousek and collaborators reported both space-offset RS (SORS) and transmission-RS approaches potentially able to study MC without penetrate the tissue, and tests were performed on phantom samples (17,18). Fitzmaurice, Barman and collaborators developed a RS-based optical fiber probe potentially able to investigate MC penetrating the tissue in a semi-invasive approach (19,20). The optical probe was tested on *ex-vivo* tissue samples confirming the potentiality of RS-based approaches to assess MC. In parallel to RS, Stone and collaborators used infrared spectroscopy on histological samples reporting a decrease of carbonate content of HA passing from benign to invasive samples (21). Even considering the mentioned literature, the current knowledge about MC composition is still limited. No previous Raman-based studies investigated the detailed composition of MC by studying the mineral composition over their entire surface; previous RS-based studied considered a relative low number of calcified regions and the investigation of MC associated with invasive carcinoma was scarce or absent. Here, we report the currently most extended biochemical characterization of breast MC thanks to the use of a Raman mapping approach on core biopsies. The aim of the present study is to clarify the association between MC composition and the pathological features of breast lesions and the surrounding tissue, validating some specific evidences by X-ray scattering approaches. Moreover, MC surrounding malignant lesions were investigated to

understand the influence of nearby cancerous tissue on their biochemical features.

MATERIALS AND METHODS

Samples

Fifty-six patients affected by suspicious breast MC on screening mammography, undergoing a core biopsy and treated at the Breast Unit of ICS Maugeri (Pavia, Italy) from 2018 to 2019, were included. Patients with mass-like lesions or previous breast surgery were excluded from the study. All patients signed a written informed consent before inclusion in the study, which was authorized by the Ethical Committee of the Institution (protocol 2281/2018 CE), which approved the study in compliance with the Declaration of Helsinki. In details, according to the B-categories defined by the UK National Health Service Breast Screening Program (NHSBSP) (22), 6 patients reported normal tissue or minimal changes (B1), 9 patients reported benign lesions (B2), 8 patients reported lesions of uncertain malignancy (B3), 17 reported *in-situ* carcinoma (B5a) and 16 reported invasive carcinoma (B5b) (**Table 1**). Detailed characteristics of patients are reported in **Table S1**.

Tissue preparation

Tissue slices were generated from formalin fixed paraffin embedded (FFPE) tissue blocks. For each patient, a 10 μm slice was microtomed and mounted on mirrored stainless-steel slides specific for Raman measurements (Renishaw plc, Wotton-under-Edge, UK). In parallel, for each tissue sample, a contiguous 6 μm tissue slice was colored using standard hematoxylin eosin stain for standard diagnostic evaluation. To correctly assess the concordance of histopathology with specific Raman features, MC located

inside *in-situ* or invasive carcinoma were primarily considered in B5a and B5b samples, respectively, and defined among “representative” MC. MC detected in samples reporting cancer but locally surrounded by tissue that could be categorized with a lower B-category (i.e. non-cancerous tissue in B5a samples and both non-cancerous tissue or *in situ* carcinoma in B5b samples) were analyzed separately. The 10 µm slices selected for Raman analyses were then deparaffinized using a simple protocol optimized starting from a previously reported method (23). Briefly, tissue slices mounted onto mirrored steel slides were dewaxed by two baths of hexane 95% (Merck KGaA, Darmstadt, Germany), two baths of ethanol absolute (Merck KGaA, Darmstadt, Germany) and a final bath of ethanol 95%, repeating these steps three times, followed by air drying for 2 h.

Raman spectroscopy

The experimental and data analysis workflow is illustrated in **Figure 1**. A commercial confocal Raman microscope (InVia Reflex, Renishaw plc, Wotton-under-Edge, UK) was used to perform Raman mapping acquisitions. A 785 nm laser source with round shape spot, powered with around 90 mW, was coupled with a N-Plan 100x (NA 0.75, WD 0.37) Leica objective and with a 1200 l/mm grating, centred around 1250 cm^{-1} thus obtaining a spectral range between 700 and 1760 cm^{-1}. The final power was then filtered to reach around 20 mW on the sample. The detector is a CCD (1024x256 pixels) sensitive between 400-1060 nm and cooled at -70 °C. The instrument was daily aligned and intensity calibrated using automated procedures implemented in the instrument start-up process. The wavelength shift calibration was periodically performed using multiple standards (polystyrene, paracetamol, silica) and daily checked by an automatic procedure using the silica band at 520 cm^{-1}.

Raman mapping measurements of each MC identified by the pathologist on the contiguous haematoxylin-eosin slice were performed. All MC with diameter between 15 and 1200 μm were selected. A squared region was centred on each MC defining a step size (mapping size) between 1 and 15 μm (on average 6 μm), depending on MC size. The focus was pre-set for each MC in the middle of the calcified part. For each step the acquisition time was of 3 sec with a single repetition.

Data processing of Raman data

Data analysis were performed using the commercial software WiRe (Renishaw plc, Wotton-under-Edge, UK) or MATLAB (MathWorks, Natick, MA, USA) or OriginPro2019 (Originlab Corporation, Wellesley Hills, MA, USA). The data analysis included 1) pre-processing; 2) spectral classification and 3) statistical/multivariate analysis.

Data pre-processing started with cosmic rays removal performed by WiRe (nearest neighbour and width of features algorithms). Next step was the background correction used to reduce fluorescence signals that may compromise the study of Raman spectra. Background signal was removed in MATLAB using baseline correction by fitting and subtracting a polynomial function of the 10^{th} order to each spectrum. Then, data were normalized in MATLAB by unit vector function in order to neutralize spectra intensity differences. The normalized data were then filtered with a moving average filter with a window size equal to five in order to reduce spectral noise.

Spectral classification was carried out to assign specific biochemical features (i.e. to identify specific components) to each map data point (i.e. for each single spectrum acquired). For each spectrum, three different indices (i.e. monotonicity of the spectra, correlation with reference spectra and specific peaks positions) were evaluated and used

to perform spectral classification (24). Spectra described by a monotonic function are those for which the function preserve the given order. Concerning Raman spectra, it is known that the fluorescence background is a monotonically decreasing (or increasing) function (25). Monotonicity was here mainly used to support the identification of spectra with high fluorescence (typical of necrotic regions, see below) and for the identification of spectra related to optical support (not monotonic functions). The reference spectra of inorganic components (hydroxyapatite (HA), whitlockite (WIT), amorphous calcium carbonate (aCaCa), crystalline calcium carbonate (calcite), calcium oxalate (CaO)) are reported in **Figure 1**. Reference spectra were produced by collecting Raman signals from real samples and by confirming their nature by X-ray scattering analysis and/or by data reported in literature. In details, at least 25 spectra from at least 2 samples were recorded and averaged to produce each reference spectrum. For calcium phosphate (both HA and WIT), after confirming their nature by data in literature (8,26), 50 µm contiguous slices, from the same samples, were studied by X-ray analysis that finally confirmed their nature (**Figure S1** and **Figure S2**). For calcium oxalate (CaO), crystalline calcium carbonate (calcite) and amorphous calcium carbonate (*a*CaCa) reference spectra were confirmed thanks to available data from literature (from (27) and (8) for CaO; from RRUFF id X050034 (28) and from (29) for calcite; from (29) and (30) for *a*CaCa). For these three cases, the scarce number and the small size of MC presenting these components did not permit to perform accurate X-ray scattering analysis. Together with mineralized components, also references of tissue surrounding MC, necrotic regions (presenting typical fluorescence spectra) and stainless-steel microscope slide were recorded and used in order to isolate pure signals from the calcified area only. Peak

positions, used as third criterion, were extracted, when possible, from above described reference spectra. The three mentioned indices were fused with the majority vote system in order to produce a thematic (false-colour) map which contains a classification based only on spectral information. Spatial classification was taken into account adopting the K-means clustering algorithm which is specific to detecting the borders (31). The algorithm is based on a random initialization of the groups; different executions of this algorithm produce different clusters, depending from the initial centroids. Therefore, in order to produce reproducible results, the clustering is repeated ten times with different initializations. The results of ten clustering phases are then fused together in order to obtain a classification false-colour map. The classification map produced by the K-means clustering has not a biochemical meaning, since K-means is a non-supervised classification method. The biochemical information was obtained by fusing the K-means clustering with the spectral classification map produced before. For each K-means cluster, the correspondent area was classified by the spectral classification and averaged. As result, for each map-dataset (for each MC), each map-point was assigned to a specific pre-selected component and, for each component, an average Raman signal was produced and extracted.

Statistical analysis and multivariate analysis were then conducted on the calcified components only. If not specified, variables were reported as means (± standard deviations) or median with range of values or as absolute numbers and percentages. Continuous variables were compared using non-parametric Wilcoxon-Mann-Whitney/ Krustal-Wallis test for variables with non-normal distribution. Statistical significance level was set at $p<0.05$ (two tailed).

For multivariate analysis, for each MC, a single Raman spectrum, calculated by averaging only signals from Type II MC (i.e. HA and WIT components) was than considered. Principal component analysis (PCA) was performed obtaining 380 principal components (PCs). A linear discriminant analysis (LDA) classification model was built using the first fourteen PCs as training data, using pure benign (i.e. B1 and B2 MC) and pure malignant (i.e. representative B5a and B5b MC) and setting prior probabilities proportional to the group size. The number of PCs to be utilized was selected in order to represent about 90% of dataset variability and in order to exclude PCs associated to noise and/or small artefacts. The PCA–LDA classification model was validated by the leave-one-out cross validation. B3 MC (uncertain malignancy), B5a MC detected outside *in-situ* carcinoma or B5a MC detected outside invasive carcinoma were used as test data. ROC curve, with relative AUC, was automatically calculated by OriginLab using as input the canonical variable 1 emerging from the PCA-LDA classification. From the ROC curve the optimal threshold (cut-off point) was obtained. This was used to produce confusion matrices and relative diagnostic performances.

Small and wide-angle X-ray scattering (SAXS and WAXS), X-ray radiography (XTM)

The XTM microscopies map more (darker) or less (lighter) X-ray absorbing portions of the biopsy; SAXS microscopies display the abundance and D-spacing of collagen; WAXS microscopies map the abundance and lattice parameters of the HA structure of MC. To confirm the crystal phase composition of MC, 50 µm slices, contiguous to samples used for RS, were deparaffinized using the protocol reported above and analyzed using XTM, SAXS and WAXS microscopy data were collected at the cSAXS beamline of the Swiss Light Source (SLS) in Villigen, Switzerland (32). The

monochromatic X-ray beam had a wave-length of 0.091216 nm corresponding to an energy of 13.6 keV. It was focused down to about 20 µm (vertical) and 40 µm (horizontal) by a bent monochromator crystal and a bent mirror. Sample to detector distances were set to 308 mm (WAXS) and 7089 mm (SAXS). Collection time per point was 0.3 s (WAXS and SAXS). A Pilatus 2 M detector (33) was used for recording SAXS/WAXS; the intensity of the transmitted direct beam was measured by means of a point detector in the beam stop, to obtain the X-ray transmission microscopy (XTM) (radiography). The samples were raster-scanned through the focused X-ray beam in a continuous line-scan mode, i.e. with the sample moving at constant velocity while the detectors record data at constant frequency. The resulting spatial distance between data points of the raster scan is 40 µm. Depending on the investigated area, data have been recorded from 6'731 to 61'875 positions on the sample in each raster scan.

Data processing - SAXS and WAXS

The SAXS and WAXS data have been azimuthally integrated within 16 angular sectors. The resulting integrated data were screened through a statistical signal classification approach (34) to reduce the large number of WAXS and SAXS data to few characteristic profiles. The characteristic WAXS profiles were fitted by a whole profile Rietveld approach (35) implemented in the FullProf program (36). The crystallographic unit cell parameters (a and c), the unit cell volume (V) and the crystalline domain size along several crystallographic directions were determined, gaining the trend of these parameters for increasing malignancy. For validating the representativeness of the characteristic WAXS profiles selected using signal classification, we explored also the entire WAXS dataset across all investigated sample areas. For each pixel of the WAXS

microscopy, the peak position of the (002) reflection as an isolated single diffraction peak was fitted with a Gaussian to extract peak position, amplitude and width, and in this way their variations have been mapped across the whole area. Then, we evaluated the mean and dispersion of: the c parameter, by means of the Bragg law applied to the peak position; the crystalline domain size along the [002] crystallographic direction, extracted with the Scherrer formula from the peak width. The statistical evaluation of c value and domain size along the [002] direction mapped across the whole area, were found to be nicely in agreement with the same values extracted from the Rietveld fitting of the characteristic WAXS profiles determined in a signal classification approach. Both the c value and the crystalline domain size along the [002] direction were found to increase with the malignancy level.

In case of the SAXS data, the characteristic profiles have been used to define regions of interest for an analysis of characteristic peaks in the scattering signal at each point of each raster scan (37). In particular, two main peaks were identified and interpreted as the 6th order of collagen fibers with 63.0nm D-spacing and as the 7th order of collagen fibers with 61.8 nm D-spacing. We monitored their variation across the investigated area and related them to WAXS and XTM microscopies.

RESULTS

A total of 474 MC from 56 patients were mapped by Raman spectroscopy as above described. Out of these, 65 MC were detected in normal tissue (B1), 67 in samples with benign lesion (B2), 61 in samples with uncertain malignant features (B3), 97 in samples with *in situ* carcinoma (B5a) and 184 in samples with invasive carcinoma (B5b). (**Table 1** and **Table S1**). Only 59 and 63 MC derived from pure B5a and B5b, respectively, and primarily considered as representative MC. The remaining were locally surrounded by tissue categorized within a lower diagnostic category and studied separately.

Benign MC contain specific components and are more heterogeneous than malignant ones

A Raman mapping approach was optimized in order to characterize and automatically identify all inorganic components contained in MC at micrometric scale, **Figure 1.** Out of 315 representative MC, 273 (86%) contains HA, the most common form of calcium phosphate defining Type II MC (**Figure 2A**). Overall, HA represents also the most abundant component (>74%) of MC (**Figure 2B**). Considering different diagnosis, benign calcifications (B1, B2) are more heterogeneous than uncertain malignant (B3) or malignant MC (B5a, B5b). In particular, 33 and 40 MC found in B1 and B2 samples, respectively, reported spectral features of WIT (38) (**Figure 2A,B**), a particular magnesium-containing crystal phase of calcium phosphate, as also confirmed by wide-angle X-ray scattering (WAXS) performed on the selected tissue specimens (**Figure S2).** In RS data, WIT was not found as unique inorganic component but always co-localized with HA and it constitutes, overall, 2.7% and 6.2% of B1 and B2 MC, respectively. Some benign MC (i.e. 7 and 25 MC in B1 and B2 lesion, respectively) also show the presence of

a few signals (<2%) of *a*CaC. In addition, only 14 MC from B1 samples contain CaO as sole component (near 100%) and these calcifications correspond to Type I MC.

Conversely, malignant (B5a, B5b) and B3 lesions, generally exhibit homogeneous MC, containing almost only HA, representing ≥97% of their overall composition (**Figure 2A, B**). A single B3 sample MC presented Type I MC (n=3) containing pure CaO. Among the whole dataset, a few MC from both benign and malignant samples reported a highly crystalline form of calcium carbonate (calcite), never reported before in breast MC and apparently not associated with pathology from our data. Finally, 17 MC identified by the pathologists did not show any specific signal of mineralized components and were likely amorphous material attributable to necrotic regions.

The chemical features of MC are associated with the diagnostic classification

Type II MC were mainly investigated here considering that Type I MC are uncommon presentations, rarely observed in our samples. The mean Raman spectra of Type II MC from different diagnostic categories revealed some differences around the major signals; in particular, around the 960 cm^{-1} band, related to calcium phosphate vibrational modes, around the 1070-1090 cm^{-1} band, related to calcium carbonate content, and around the 1450 cm^{-1} band, mainly related to protein and lipids (here referred as “organic matrix”) content (**Figure 3A-F, Figure S3**). In details, a broadening of the phosphate band is observed in benign samples (**Figure 3A,D)** and this is mainly associated with a more disordered and more substituted phosphate crystal lattice (39,40). The broadening extent is similar in normal (B1) and benign (B2) MC (p=0.43) but significantly higher (p=7.90*10^{-10}) if compared with MC found in lesion of uncertain malignancy (B3), *in-situ* (B5a) and

invasive lesions (B5b). In turn, B3, B5a and B5b exhibit only minor variation of broadening values (p = 0.032), with main contribution by differences between B3 and B5b samples (p = 0.011) (**Table S2**). In parallel, observing the phosphate peak, a statistically significant shift toward higher wavenumbers (here referred as "red-shift") is also observed in benign MC (p=0.025). This is mainly derived by the presence of WIT in benign MC, characterized by maximum intensity of the phosphate band around 970 cm-1 (38) (**Figure 3, Figure S4, Table S3**).

Also the carbonate band reveals higher intensity and broadening in benign samples in the spectral region between 1070-1090 cm^{-1}, showing significant differences when comparing benign (B1, B2) and non-benign (B3, B5a, B5b) subtypes (p=2.86*10^{-9}) (**Figure 3B,E**)**.** No significant differences can be observed when comparing B1 and B2 (p = 0.51) and a certain variance characterize B3, B5a and B5b (p = 0.001), especially due to differences between B5a and B5b (**Figure 3B,E**)**.** Finally, the intensity of the 1450 cm-1 band, corresponding to the organic matrix, also changes among diagnostic categories, but differences do not correlate with the pathological status (**Figure 3C,F**).

The crystalline lattice of MC is more ordered and expands with increasing malignancy

In order to better understand some biochemical and structural features emerging from Raman data, crystallographic evaluations by WAXS were performed on randomly selected samples, focusing on Type II MC, and on HA structures. **Figure 4A** shows the HA crystalline domain parameter and the HA crystallographic unit cell "*c"* parameter increasing when passing from benign (B1, B2) to carcinoma samples (B5a, B5b) (overall diffractograms in **Figure S1**). This behaviour is coherent with the broadening of

phosphate Raman peak (**Figure 4A,B**) passing from malignant to benign samples. Both these data indicate that the crystallinity and homogeneity of HA produced by malignant MC is significantly higher if compared with benign samples. In particular, the increasing of the "*c*" parameter indicates that the unit-cell of HA in MC in malignant samples is generally elongated in comparison to the one in benign samples.

Finally, SAXS data show that collagen is co-located with MC and that it has longer inter-molecular D-spacing of 63.0 nm in comparison to 61.8 nm of collagen found outside the MC (**Figure S5, S6**). These findings reveal that HA nanocrystals are spatially correlated to the larger D-spacing of the collagen fibres, which can be interpreted as a hint towards crystallization nucleating on collagen fibres. A weak tendency towards an overall increase of the intramolecular spacing of collagen is found passing from benign to malignant MC (**Figure S6**) and this is interpreted speculatively as a tighter binding of collagen to malignant MC.

The multivariate analysis of MC Raman maps allows for the accurate classification of different histological subtypes

Multivariate analysis was applied to extract all potential pathological biomarkers from the complexity of Raman data thus improving the limited and subjective analysis of single spectrometric variables. First, principal component analysis (PCA) was performed and the first fourteen principal components (**Figure S7**), representing 90% of the variability of the entire dataset, were extracted and used as variables for further classification. A classification model was made by LDA firstly using only representative MC from pure benign samples (i.e. B1 and B2) and from pure malignant samples (B5a and B5b) thus

producing a Raman canonical correlation coefficient (canonical variable 1) further used as unique classification variable (**Figure 5A-C**). After excluding a correlation between age and the biochemical composition of MC (R^2=0.020, slope *p* =0.33) (**Figure S8**), a receiver operating characteristic (ROC) curve (**Figure 6**) was calculated and used to find the optimal threshold point (i.e. -0.445 of canonical variable) for the classification. Out of 108 malignant MC, 103 were correctly classified giving 95.4% sensitivity and 94.6% negative predictive value (NPV) (**Table S4**). Among MC incorrectly assigned as benign, 4 were from ductal carcinoma in situ (DCIS), 1 from invasive ductal carcinoma (IDC). Out of 103 benign MC, 15 were wrongly assigned as malignant giving 85.4% of specificity and 87.3% PPV. Out of benign MC wrongly classified as malignant 10 were from normal tissue, 1 from fibroadenoma (FAD), 2 from fibrocystic change (FIB), 2 from usual ductal hyperplasia (UDH). Overall, the model gives 90.5% accuracy. These results were validated through leave-one-out cross validation which virtually assigns all MC with an unknown diagnosis before to be classified. This gave 93.5% sensitivity and 80.6% specificity with 87.2% of overall accuracy (**Table S5**).

The classification model was then used to investigate the pathological fingerprint of MC found in lesion of uncertain malignancy (B3). Out of 53 B3 MC, 44 (29 from flat epithelial atypia tissue and 15 from ADH) were assigned as malignant and 9 MC (7 ADH, 1 flat epithelial atypia and 1 from papillary lesion) were recognized as having benign features (**Figure 5B**, **Table S6**). Moreover, also a correlation between the Raman data and mammographic evaluation according to the Breast Imaging-Reporting and Data System (BI-RADS) (26) classification was observed (**Figure S9).** Finally, considering that MC size can be very heterogeneous, as showed in **Figure S10** (mean: 14511 μm^2; SD: 27229

μm^2; IQR 25-75%: 11700 μm^2), the correlation between MC size and malignant features, as revealed by Raman mapping (i.e. Canonical variable 1), has been investigated showing a weak correlation (-0,1894; p = 0.0002 (Spearman's correlation)), suggesting that larger MC are slightly associated with lower malignancy.

Most of MC found in locally healthy tissue neighbouring carcinoma lesions show malignant features

116 MC (28 and 88 for B5a and B5b, respectively) were excluded from the main data analysis because locally surrounded by non-cancerous tissue in B5a samples, or because surrounded by non-cancerous tissue or *in-situ* carcinoma in B5b samples. When data from B5a samples were used as test dataset, out of 28 MC found locally benign, 18 showed malignant features and were classified as malignant, the remaining 10 MC were classified as benign (**Figure 5C, Table S6**). Out of 88 MC from B5b detected outside the invasive carcinoma, only 2 MC (locally B1) were classified as benign but the remaining 80 were classified as malignant (**Figure 5C, Table S6**). These include 29 MC situated in *in-situ* carcinoma (B5a), 23 MC surrounded by B3 tissue, but also 34 MC surrounded by B1 tissue.

DISCUSSION

The rapid and accurate characterization of breast MC is currently an unmet clinical need. MC are considered as suspicious signs of a breast lesion, and histopathology is mandatory to assess the nature of such lesions. Up to now, MC have been traditionally seen by mammography as simple bystanders of cancer and only characterized by descriptive criteria (i.e. morphology and spatial distribution), according to the Breast Imaging-Reporting and Data System (BI-RADS) (41). Recently, research teams have begun to show the potential of further spectral analysis using Raman and IR spectroscopy (8,17,21,40,42), but this paper advances the concept significantly.

Here we investigated breast MC from different diagnostic categories by studying the whole area of each MC using a Raman mapping approach on a relevant number of MC. This allowed for the first time to describe all inorganic components contained in MC and to correlate it with the pathological status. Previously reported Raman characterizations have been made on a significantly lower number of MC and by performing single acquisitions at selected sites inside the calcified lesion (8) or single spectra on fresh biopsies (40), thus obtaining partial information. Only recently the spatial composition of MC has been investigated by Raman mapping on the whole MC surface, although on a single DCIS specimen, with explorative aims (26).

The first evidence emerging from this study is that MC from pure benign lesions (i.e. B1 and B2) are largely heterogeneous if compared with both lesions of uncertain malignant potential (B3) and with carcinoma lesions (B5a, B5b). First, in benign samples both Type I (calcium oxalate) and Type II (hydroxyapatite) MC were observed, confirming what previously reported; also confirming that Type I MC are uncommon findings (8). Second,

in most of Type II MC identified in benign samples, hydroxyapatite is not the only component but also whitlockite and aCaCa were observed, (in 55% and 24% of benign MC, respectively). Whitlockite is a crystal phase of calcium phosphate, sporadically detected in human tissues where magnesium partially substitutes calcium if compared with hydroxyapatite (43). In this study, investigating 56 sample and 315 representative MC, whitlockite overall represents 4.5% of benign calcifications (B1, B2), 0.32% of B3 samples, and the 0.22% of cancerous calcification (B5a, B5b). These results were further confirmed when the mean spectrum of each MC was considered. Recently, whitlockite was identified by X-ray in a previous study reporting more whitlockite in malignant (1.18%) then in benign (0.46%) samples (44). This discrepancy could be explained by a lower number of samples (57 MC from 15 biopsies) included in the mentioned study and by the lower spatial resolution of X-ray measures (10 µm), if compared with an average mapping size of 6 µm used here for Raman mapping.

In parallel, the finding of both amorphous (aCaCa) and crystalline calcium carbonate (calcite) as isolated components in MC is a new evidence. aCaCa was mostly found in benign samples, always localized with whitlockite, and this could be due to the fact that magnesium, contained in whitlockite, is known to stabilize aCaCa (45). Calcite is the stable and crystalline form of CaCa that can naturally form if not stabilized by active processes (46) and has been find in both benign and malignant samples only as single very small crystals of pure calcite. The amount of data is not sufficient to explain an association with pathology.

When the overall Raman signals originating from single Type II MC were extracted to verify their correlation with diagnosis, the first evident feature was the broadening and

red-shift of the 960cm^{-1} phosphate band of calcium phosphate in B1 and B2 samples. The broadening of the phosphate band is normally explained by the increase of carbonate content into hydroxyapatite, resulting in the alteration of the symmetry of the crystal structure (39). The red-shift of the phosphate band is due to the change of the crystal composition as in the case of whitlockite, due to the presence of magnesium (38). In our data, both contributions have been seen and this is explained by the co-presence of (carbonated) apatite and of whitlockite in benign MC. Previous RS studies on real samples, only mentioned peak broadening and this could be due to single point acquisition, not able to recapitulate the whole MC composition (8,40). Noteworthy, when 960 cm^{-1} peak broadening (FWHM) is compared among diagnostic categories, B3, B5a and B5b samples cluster together with only minor differences (p = 0.032). In parallel, also the band of carbonate changes among diagnostic categories, decreasing in its overall intensity passing from benign to malignant samples. Also in this case, there is a possible contribution derived by the presence of whitlockite in benign samples, producing band broadening and a weak shoulder at higher Raman shift, and this was not revealed in previous studies describing carbonate as indicator of benignity (8,21,40). As a final point, organic matrix (i.e. proteins and lipids) has been described as MC component but its contribution did not show any statistically significant correlation with pathology as previously reported (8,40).

Spectral data were also confirmed by WAXS acquisitions on randomly selected samples, showing that malignant samples (B5a and B5b) exhibit an increase of crystallinity, in accordance with previous findings (44). In connection to the observation in RS that WIT is more abundant in benign than in malignant MC, one could speculate that malignant MC

contain less Mg, thus exhibiting a more well-ordered HA crystalline lattice. In healthy conditions, biomineralization involves hydroxyapatite with partial substitution of Ca ions (47). Smaller cations, such as Mg ions, lead to a cell contraction compared to the stoichiometric crystal (48). Therefore, the observed trend of the "*c*" parameter can be correlated to an altered metabolism of Mg ions, which changes with the malignancy level. Indeed, neoplastic cells, i.e. for B5 samples, are avid of Mg cations (49), which therefore might reduce the Mg quantity available for MC, justifying the lattice expansion, compared to what expected in benign biomineralization. In addition, the observed trend of the "*c*" parameter, which increases passing from benign to malignant MC, may be also associated with carbonate substitution in HA structures (50). This is coherent with relatively higher levels of carbonate Raman signals in benign samples, reported in this and in previous studies (21,40). In particular, lowering of the "*c"* parameter occurs when carbonate mainly substitutes the hydroxyl site of HA (called "A-Type substitution") and this has also been observed in a very recent X-ray diffraction study on MC (51). SAXS data exhibit slightly larger D-spacing of collagen in MC found in malignant samples compared to benign ones, in agreement with SAXS data collected on tissue (52). This weak trend can be speculatively attributed to a tighter binding of MC to collagen with increasing malignancy.

A general overview of results coming from the detailed characterization of breast MC suggests that malignant microcalcifications are more homogeneous, more crystalline, less substituted (by Mg and carbonate) and more tightly bound to collagen, if compared with benign ones. In addition, chemical and structural MC from lesion of uncertain malignancy (B3), are general similar to pure malignant ones (B5a and B5b). Studies performed on

tissue samples and *in vitro* on cultured cells demonstrated that MC formation is a cell-active process influenced by the microenvironment and by the overexpression of bone matrix proteins (i.e. osteonectin (OSN) and osteopontin (OPN)) (53,54). Both these studies reported that active processes of MC formation are significantly more represented in case of malignancy. In parallel, a retrospective study on patients referred for needle-guided biopsy, reported that the formation of new MC significantly correlates with high probability of ductal invasive carcinoma (55). These data suggest that malignant lesions are more active in the formation of MC. In this regard, we have seen a weak correlation between MC size and malignancy features indicating that, smaller MC are lightly associated with higher malignancy. This could be speculatively associated with the smaller size of new forming MC during carcinoma growth. Starting from these assumptions, the crystallinity and homogeneity of malignant MC could originate from a faster and active process stimulated by cancer and by its microenvironment. On the contrary, the heterogeneity and low crystallinity of benign MC could be explained by a slower and less regulated mechanism of mineralization, thus permitting both loss of crystal stability and/or the intercalation of external components (i.e. carbonate, magnesium) from the surrounding tissue.

In order to verify the accuracy of the presented Raman-based MC characterization, a multivariate approach was applied, including the use of a LDA-based model to automatically verify the potential diagnostic performance of the proposed approach. The model was built using pure benign (B1, B2) and pure malignant (B5a, B5b) categories; including only MC detected inside carcinoma region in case of malignant MC. The results are promising, showing 93.5% sensitivity and 80.6% specificity with an NPV of 92.2%

after cross-validation. The same classification model was used to investigate the grade of malignancy of B3, according to the biochemical composition, revealing that 39 (83%) of 53 B3 MC detected show malignant features, mostly by FEA subtype. This data is in agreement with the fact that FEA represents a direct pre-cancerous lesion possibly leading to DCIS in up to 18.6% of cases (56).

In addition, we investigated whether all MC detected in malignant biopsy samples, but outsides the specific cancer lesion, exhibit the features of the local “benign” (or B3 or DCIS) surrounding tissue, or the malignant features of the carcinoma nearby. As described above, 64% of locally benign (or B3) MC found around DCIS samples showed malignant features. In invasive carcinoma samples, 86 (98%) of 88 MC found outside the invasive region (including 34 (94%) of 36 locally classified as B1) were classified as malignant. This is interesting data showing that the biochemical composition and structural features of MC are influenced by the tumour even if MC are not directly surrounded or in close contact with cancer cells. As a consequence we can assume that the tumour environment, and probably the metabolism of breast tissue in the presence of malignancy (especially if invasive carcinoma), influence a relatively extended tissue region around the malignancy, as also recently suggested (57). This evidence may also help the translation of Raman-based optical probes or Raman-based non-invasive tools potentially compatible with *in vivo* approaches (17,19,58). In particular, these data suggest that the investigation of a relatively large region (500-2000 μm) of suspected tissue containing MC could correctly inform about the malignancy even if some locally benign MC regions surrounding the lesion are inevitably probed due to low-resolution of *in vivo* configuration.

In conclusion, the present study reports new detailed information about MC composition and demonstrats that Raman-based approaches can provide a direct and reliable description of breast lesions thanks to the study of MC. These evidences may play an important role in the development of new tools for the assessment of suspected lesions presenting MC. For example, SORS and transmission-Raman approaches are able to illuminate a circumscribed portion of sample at a certain depth in a non-invasive manner and they were already tested on phantom samples containing calcified components (17,58). Similarly, Raman needle probes are able to perform accurate measurements into the tissue in a semi-invasive modality (15,16) and a proof-of-concept was also tested to study breast MC *ex-vivo* (19,42). Moreover, preliminary experiments using phase-contrast X-ray imaging approaches demonstrated the possibility to study the crystalline structure of MC (10). The mentioned tools have not yet been tested or transferred in clinics also due to the lack of detailed information about MC composition. The data reported here may be step forward to this direction.

Acknowledgements

The proposal ' Correlative Imaging of scanning micro Raman and SAXS/WAXS Scanning Microscopies [...]' (E181100214) and thereby access to the cSAXS beamline of the Swiss Light Source has been supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 731019 (EUSMI) (C.Giannini). We thank Andreas Menzel and Xavier Donath for providing assistance for the X-ray measurements. We thank also Luciana Russo and Chiara Guerra for technical support in samples preparation, and Daniele Baldassarra for his help in collecting Raman data.

REFERENCES

1. Gajdos C, Ian Tartter P, Bleiweiss IJ, Hermann G, de Csepel J, Estabrook A, et al. Mammographic Appearance of Nonpalpable Breast Cancer Reflects Pathologic Characteristics. Ann Surg. 2002;235:246–51.

2. Lakhani SR. WHO Classification of Tumours of the Breast. International Agency for Research on Cancer; 2012.

3. Wilkinson L, Thomas V, Sharma N. Microcalcification on mammography: approaches to interpretation and biopsy. BJR. 2016;90:20160594.

4. Ong M-S, Mandl KD. National Expenditure For False-Positive Mammograms And Breast Cancer Overdiagnoses Estimated At $4 Billion A Year. Health Affairs. 2015;34:576–83.

5. Vlahiotis A, Griffin B, Stavros AT, Margolis J. Analysis of utilization patterns and associated costs of the breast imaging and diagnostic procedures after screening mammography. Clinicoecon Outcomes Res. 2018;10:157–67.

6. Frappart L, Remy I, Lin HC, Bremond A, Raudrant D, Grousson B, et al. Different types of microcalcifications observed in breast pathology. Vichows Archiv A Pathol Anat. 1987;410:179–87.

7. Morgan MP, Cooke MM, McCarthy GM. Microcalcifications associated with breast cancer: an epiphenomenon or biologically significant feature of selected tumors? J Mammary Gland Biol Neoplasia. 2005;10:181–7.

8. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. Identifying Microcalcifications in Benign and Malignant Breast Lesions by Probing Differences in Their Chemical Composition Using Raman Spectroscopy. Cancer Res. 2002;62:5375–80.

9. Yassin NIR, Omran S, El Houby EMF, Allam H. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. Computer Methods and Programs in Biomedicine. 2018;156:25–45.

10. Wang Z, Hauser N, Singer G, Trippel M, Kubik-Huch RA, Schneider CW, et al. Non-invasive classification of microcalcifications with phase-contrast X-ray mammography. Nature Communications. 2014;5:3797.

11. I. Ellis D, P. Cowcher D, Ashton L, O'Hagan S, Goodacre R. Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool. Analyst. 2013;138:3871–84.

12. Mandair GS, Morris MD. Contributions of Raman spectroscopy to the understanding of bone strength. Bonekey Rep [Internet]. 2015 [cited 2019 Dec 21];4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4296861/

13. Jermyn M, Mok K, Mercier J, Desroches J, Pichette J, Saint-Arnaud K, et al. Intraoperative

brain cancer detection with Raman spectroscopy in humans. Science Translational Medicine. 2015;7:274ra19-274ra19.

14. Matousek P, Draper ERC, Goodship AE, Clark IP, Ronayne KL, Parker AW. Noninvasive Raman Spectroscopy of Human Tissue In Vivo. Appl Spectrosc. 2006;60:758–63.

15. Day JCC, Bennett R, Smith B, Kendall C, Hutchings J, Meaden GM, et al. A miniature confocal Raman probe for endoscopic use. Phys Med Biol. 2009;54:7077–87.

16. Hattori Y, Komachi Y, Asakura T, Shimosegawa T, Kanai G-I, Tashiro H, et al. In vivo Raman study of the living rat esophagus and stomach using a micro-Raman probe under an endoscope. Appl Spectrosc. 2007;61:579–84.

17. Stone N, Matousek P. Advanced Transmission Raman Spectroscopy: A Promising Tool for Breast Disease Diagnosis. Cancer Res. 2008;68:4424–30.

18. Stone N, Baker R, Rogers K, Parker AW, Matousek P. Subsurface probing of calcifications with spatially offset Raman spectroscopy (SORS): future possibilities for the diagnosis of breast cancer. Analyst. 2007;132:899–905.

19. Saha A, Barman I, Dingari NC, McGee S, Volynskaya Z, Galindo LH, et al. Raman spectroscopy: A real-time tool for identifying microcalcifications during stereotactic breast core needle biopsies. Biomedical Optics Express. 2011;2:2792–803.

20. Saha A, Barman I, Dingari NC, Galindo LH, Sattar A, Liu W, et al. Precision of Raman Spectroscopy Measurements in Detection of Microcalcifications in Breast Needle Biopsies. Anal Chem. 2012;84:6715–22.

21. Baker R, Rogers KD, Shepherd N, Stone N. New relationships between breast microcalcifications and cancer. Br J Cancer. 2010;103:1034–9.

22. Ellis IO, Humphreys S, Michell M, Pinder SE, Wells CA, Zakhour HD, et al. Best Practice No 179. Guidelines for breast needle core biopsy handling and reporting in breast screening assessment. J Clin Pathol. 2004;57:897–902.

23. Faoláin EÓ, Hunter MB, Byrne JM, Kelehan P, Lambkin HA, Byrne HJ, et al. Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents. Journal of Histochemistry and Cytochemistry. 2005;53:121–9.

24. Torti E, Marcinnò B, Vanna R, Morasso C, Piccotti F, Villani L, et al. Automatic and Unsupervised Identification of Specific Biochemical Features from Raman Mapping Data. 2019 22nd Euromicro Conference on Digital System Design (DSD). :464–9.

25. Bigio IJ, Fantini S. Quantitative Biomedical Optics: Theory, Methods, and Applications. Cambridge University Press; 2016.

26. Kunitake JAMR, Choi S, Nguyen KX, Lee MM, He F, Sudilovsky D, et al. Correlative imaging reveals physiochemical heterogeneity of microcalcifications in human breast carcinomas. Journal of Structural Biology. 2018;202:25–34.

27. Kodati VR, Tomasi GE, Turumin JL, Tu AT. Raman Spectroscopic Identification of Calcium-Oxalate-Type Kidney Stone. Appl Spectrosc, AS. 1990;44:1408–11.

28. Lafuente B, Downs RT, Yang H, Stone N. The power of databases: The RRUFF project. Highlights in Mineralogical Crystallography. 2016;1–29.

29. Wehrmeister U, Jacob DE, Soldati AL, Loges N, Häger T, Hofmeister W. Amorphous, nanocrystalline and crystalline calcium carbonates in biological materials. Journal of Raman Spectroscopy. 2011;42:926–35.

30. Tlili MM, Amor MB, Gabrielli C, Joiret S, Maurin G, Rousseau P. Characterization of CaCO3 hydrates by micro-Raman spectroscopy. Journal of Raman Spectroscopy. 2002;33:10–6.

31. Torti E, Florimbi G, Castelli F, Ortega S, Fabelo H, Callicó GM, et al. Parallel K-Means Clustering for Brain Cancer Detection Using Hyperspectral Images. Electronics. 2018;7:283.

32. Bunk O, Bech M, Jensen TH, Feidenhans\textquotesinglel R, Binderup T, Menzel A, et al. Multimodal x-ray scatter imaging. New J Phys. 2009;11:123016.

33. Kraft P, Bergamaschi A, Bronnimann C, Dinapoli R, Eikenberry EF, Graafsma H, et al. Characterization and Calibration of PILATUS Detectors. IEEE Transactions on Nuclear Science. 2009;56:758–64.

34. Lutz-Bueno V, Arboleda C, Leu L, Blunt MJ, Busch A, Georgiadis A, et al. Model-free classification of X-ray scattering signals applied to image segmentation. J Appl Crystallogr. 2018;51:1378–86.

35. Rietveld HM. A profile refinement method for nuclear and magnetic structures. Journal of Applied Crystallography. 1969;2:65–71.

36. Rodríguez-Carvajal J. Recent advances in magnetic structure determination by neutron powder diffraction. Physica B: Condensed Matter. 1993;192:55–69.

37. Giannini C, Ladisa M, Lutz-Bueno V, Terzi A, Ramella M, Fusaro L, et al. X-ray scanning microscopies of microcalcifications in abdominal aortic and popliteal artery aneurysms. IUCrJ. 2019;6:267–76.

38. Jolliff BL, Hughes JM, Freeman JJ, Zeigler RA. Crystal chemistry of lunar merrillite and comparison to other meteoritic and planetary suites of whitlockite and merrillite. American Mineralogist. 2006;91:1583–95.

39. Mul FFM de, Otto C, Greve J, Arends J, Bosch JJT. Calculation of the Raman line broadening on carbonation in synthetic hydroxyapatite. Journal of Raman Spectroscopy. 1988;19:13–21.

40. Sathyavathi R, Saha A, Soares JS, Spegazzini N, McGee S, Rao Dasari R, et al. Raman spectroscopic sensing of carbonate intercalation in breast microcalcifications at stereotactic biopsy. Scientific Reports. 2015;5:9907.

41. D'Orsi C, Sickles E, Mendelson E, Morris E. ACR BI-RADS® Atlas, Breast Imaging

Reporting and Data System. Reston, VA, American College of Radiology; 2013.

42. Barman I, Dingari NC, Saha A, McGee S, Galindo LH, Liu W, et al. Application of Raman Spectroscopy to Identify Microcalcifications and Underlying Breast Lesions at Stereotactic Core Needle Biopsy. Cancer Res. 2013;73:3206–15.

43. Lagier R, Baud C-A. Magnesium Whitlockite, a Calcium Phosphate Crystal of Special Interest in Pathology. Pathology - Research and Practice. 2003;199:329–35.

44. Scott R, Stone N, Kendall C, Geraki K, Rogers K. Relationships between pathology and crystal structure in breast calcifications: an *in situ* X-ray diffraction study in histological sections. npj Breast Cancer. 2016;2:16029.

45. Loste E, Wilson RM, Seshadri R, Meldrum FC. The role of magnesium in stabilising amorphous calcium carbonate and controlling calcite morphologies. Journal of Crystal Growth. 2003;254:206–18.

46. Beniash E, Aizenberg J, Addadi L, Weiner S. Amorphous calcium carbonate transforms into calcite during sea urchin larval spicule growth. Proceedings of the Royal Society of London Series B: Biological Sciences. 1997;264:461–5.

47. Boanini E, Gazzano M, Bigi A. Ionic substitutions in calcium phosphates synthesized at low temperature. Acta Biomaterialia. 2010;6:1882–94.

48. Laurencin D, Almora-Barrios N, de Leeuw NH, Gervais C, Bonhomme C, Mauri F, et al. Magnesium incorporation into hydroxyapatite. Biomaterials. 2011;32:1826–37.

49. Castiglioni S, Maier JAM. Magnesium and cancer: a dangerous liason. Magnes Res. 2011;24:S92-100.

50. Šupová M. Substituted hydroxyapatites for biomedical applications: A review. Ceramics International. 2015;41:9203–31.

51. Gosling S, Scott R, Greenwood C, Bouzy P, Nallala J, Lyburn ID, et al. Calcification Microstructure Reflects Breast Tissue Microenvironment. J Mammary Gland Biol Neoplasia. 2019;24:333–42.

52. Fernández M, Keyriläinen J, Serimaa R, Torkkeli M, Karjalainen-Lindsberg M-L, Tenhunen M, et al. Small-angle x-ray scattering studies of human breast tissue samples. Phys Med Biol. 2002;47:577–592.

53. Cox RF, Hernandez-Santana A, Ramdass S, McMahon G, Harmey JH, Morgan MP. Microcalcifications in breast cancer: novel insights into the molecular mechanism and functional consequence of mammary mineralisation. British Journal of Cancer. 2012;106:525–37.

54. Bellahcène A, Castronovo V. Increased expression of osteonectin and osteopontin, two bone matrix proteins, in human breast cancer. Am J Pathol. 1995;146:95–100.

55. Lev-Toaff AS, Feig SA, Saitas VL, Finkel GC, Schwartz GF. Stability of malignant breast

microcalcifications. Radiology [Internet]. 1994 [cited 2019 Oct 3]; Available from: https://pubs.rsna.org/doi/abs/10.1148/radiology.192.1.8208928

56. Rudin AV, Hoskin TL, Fahy A, Farrell AM, Nassar A, Ghosh K, et al. Flat Epithelial Atypia on Core Biopsy and Upgrade to Cancer: a Systematic Review and Meta-Analysis. Ann Surg Oncol. 2017;24:3549–58.

57. Vidavsky N, Kunitake JAMR, Chiou AE, Northrup PA, Porri TJ, Ling L, et al. Studying biomineralization pathways in a 3D culture model of breast cancer microcalcifications. Biomaterials. 2018;179:71–82.

58. Stone N, Baker R, Rogers K, Parker AW, Matousek P. Subsurface probing of calcifications with spatially offset Raman spectroscopy (SORS): future possibilities for the diagnosis of breast cancer. Analyst. 2007;132:899–905.

TABLES

Table 1. Number of MC detected from different subjects. Samples were classified according to the B-categories defined by the UK National Health Service Breast Screening Programme (NHSBSP) (22) and according to histological classification. "MC outside the lesion" were MC detected in the tissue surrounding B5a or B5b lesions and categorized with a lower diagnostic malignancy. "Representative MC" were MC detected inside the malignant lesion specified by the diagnostic category (B5a or B5b).

B Category	Histological classification	Patients	Total MC	MC found outside the lesion	Representative MC
B1 "normal tissue"	NOR	6	65	-	65
Total		**6**	**65**	**-**	**65**
B2 "benign"	FAD	4	30	-	30
	FIB	2	8	-	8
	FNE	1	11	-	11
	UDH	2	18		18
Total		**9**	**67**	**-**	**67**
B3 "of uncertain malignancy"	PAP	2	2	-	2
	FEA (DIN1A)	2	32	-	32
	ADH (DIN1B)	4	27	-	27
Total		**8**	**61**	**-**	**61**
B5a "carcinoma in situ"	DCIS DIN1C	3	19	*11*	8
	DCIS DIN2	8	32	*8*	24
	DCIS DIN3	6	46	*19*	27
Total		**17**	**97**	***38***	**59**
B5b "invasive carcinoma"	IDC G2	12	133	*92*	41
	IDC G3	1	2	*2*	0
	ILC G2	1	8	*8*	0
	ILC G3	1	33	*19*	14
	IMC	1	8	*0*	8
Total		**16**	**184**	***121***	**63**
Grand Total		**56**	**474**	***159***	**315**

Abbreviations: normal tissue (NOR); fibroadenoma (FAD); fibrocystic change (FIB); fat necrosis (FNE); usual ductal hyperplasia (UDH); papillary lesion (PAP); flat epithelial atypia (FEA); atypical ductal hyperplasia (ADH); ductal carcinoma in situ (DCIS); invasive ductal carcinoma (IDC); invasive lobular carcinoma (ILC); invasive mucinous carcinoma (IMC).

FIGURE LEGENDS

Figure 1. Experimental workflow and components. Overview of the experimental workflow (left) and components detected in each Raman map and considered in this study (right). The Raman map reported as example was obtained from the analysis of a B5a samples and reports the only presence of HA (green), surrounded by tissue (gray).

Figure 2. Detailed composition of representative MC (n=315). **A,** Number of MC exhibiting at least one pixel of the components described by the legend, for each diagnostic category. **B,** Overall composition of MC, calculated considering altogether all MC belonging to each diagnostic category. Components: calcium phosphate in the hydroxyapatite form (HA), calcium phosphate in the whitlockite form (WIT); amorphous calcium carbonate (aCaC); crystalline calcium carbonate (calcite); calcium oxalate (CaO); MC not showing specific Raman spectra associated to mineralized components (none).

Figure 3. Vibrational features of representative Type I MC (n=264). Average phosphate band (**A**), average carbonate band (**B**) and average organic matrix band (**C**) of each of the diagnostic categories, including standard deviation (shaded area). In **A**, the intensity of spectra is shifted for clarity. **D,** box-plot reporting the broadening (FWHM) of the phosphate band. **E,** box-plot reporting the intensity (area) of the carbonate band, between 1070-1090 cm^{-1}. **F,** box-plot reporting the intensity (area) of organic matrix band, between 1425-1475 cm^{-1}. Data are shown as box ad whiskers. Each data point represents a single MC. Each box represents the 25th to 75th percentiles (interquartile range [IQR]). Dots

inside the box is the mean, lines inside the boxes represent the median. The whiskers represent the lowest and highest values within the boxes ±1.5x the IQR. All *p* values are reported in **Table S2**. "n.s." means that not significant difference was observed ($p > 0.05$ (two tailed)).

Figure 4. Crystalline feature of HA from selected representative samples. **A,** Variation of the crystalline domain size along the [002] direction. **B,** Variation of the crystallographic unit cell parameter *c* along the [002] direction. Filled circle represent results of Rietveld fits of the *few* patterns extracted from the entire dataset by a statistical signal classification approach (32); empty circles represent the main value resulting from a single-peak fit of the (002) reflection at 2θ=15.29 deg for *all* the patterns collected in the entire explored sample area of a sample. (The analysis reported in **Fig. S2** is not included in these trends as the sample was a mixture of two crystalline phases, i.e. HA and WIT.)

Figure 5. Results from the LDA-classification model. **A,** box-plot showing the canonical variable score emerging from the LDA-based classification of MC from pure benign (B1, B2) and pure malignant (B5a, B5b) samples (n=211). **B,** box plot reporting the canonical variable score of MC found in B3 samples and used as test dataset (n=53). **C,** box plot reporting the canonical variable score of MC found outside carcinoma lesions in B5a (n=28) and B5b samples (n=88), used as test dataset.. Data are shown as box ad whiskers. Each data point represents a single MC. Each box represents the 25th to 75th percentiles (interquartile range [IQR]). Dots inside the box is the mean, lines inside the boxes represent the median. The whiskers represent the lowest and highest values within

the boxes ±1.5x the IQR. The dashed lines refer to the optimal cut-off value (-0.445) defined by the ROC.

Figure 6: ROC curve. The curve was obtained from the post probability assignment produced by the LDA-based classification model and considering only pure benign (B1,B2) and pure malignant MC (B5a,B5b) (n=211).

Figure 1

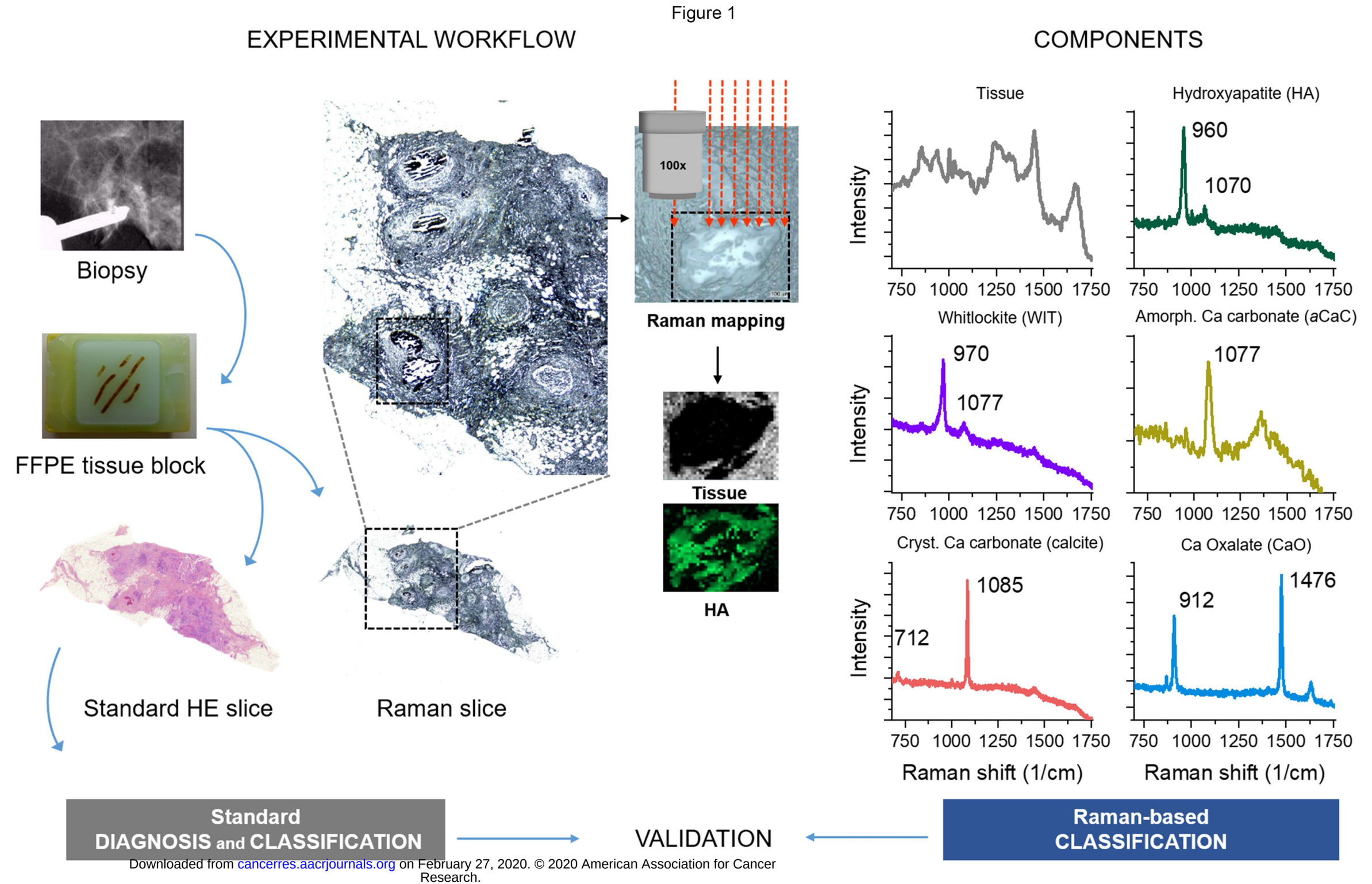
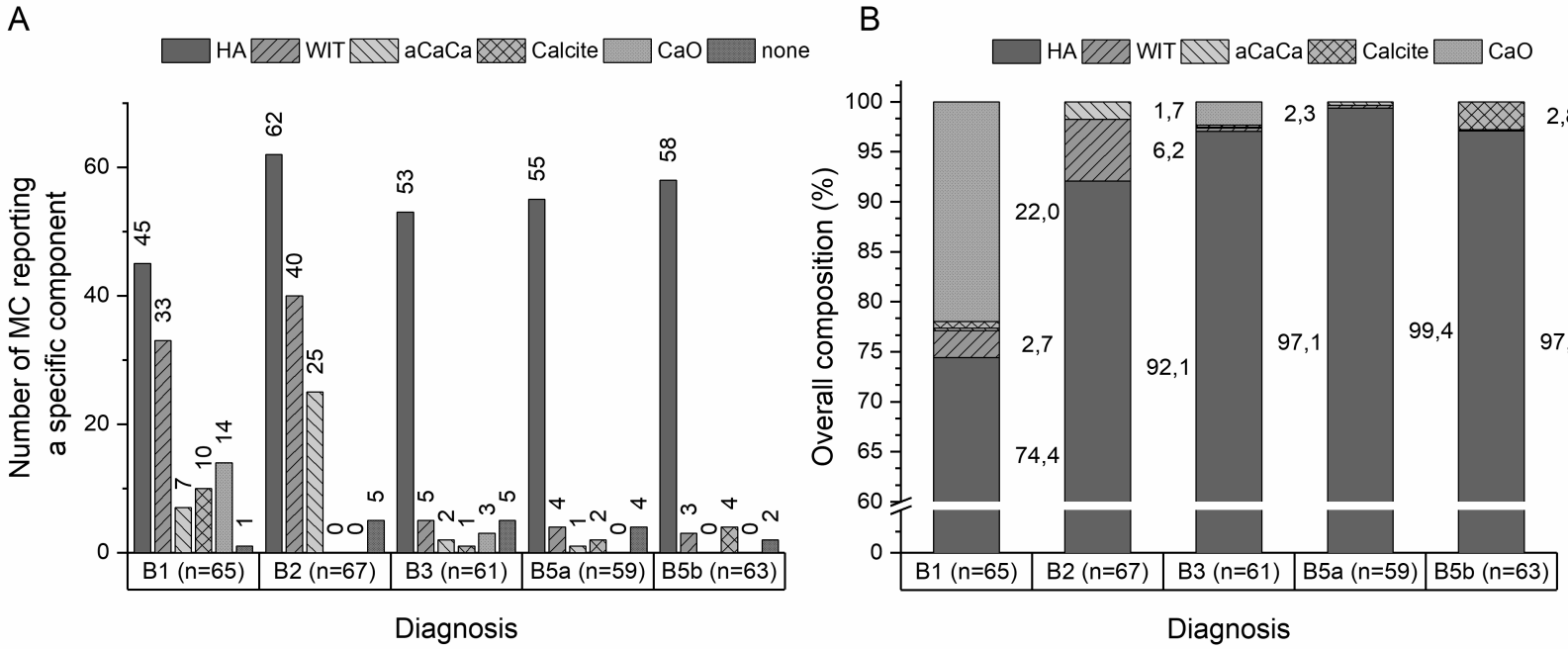
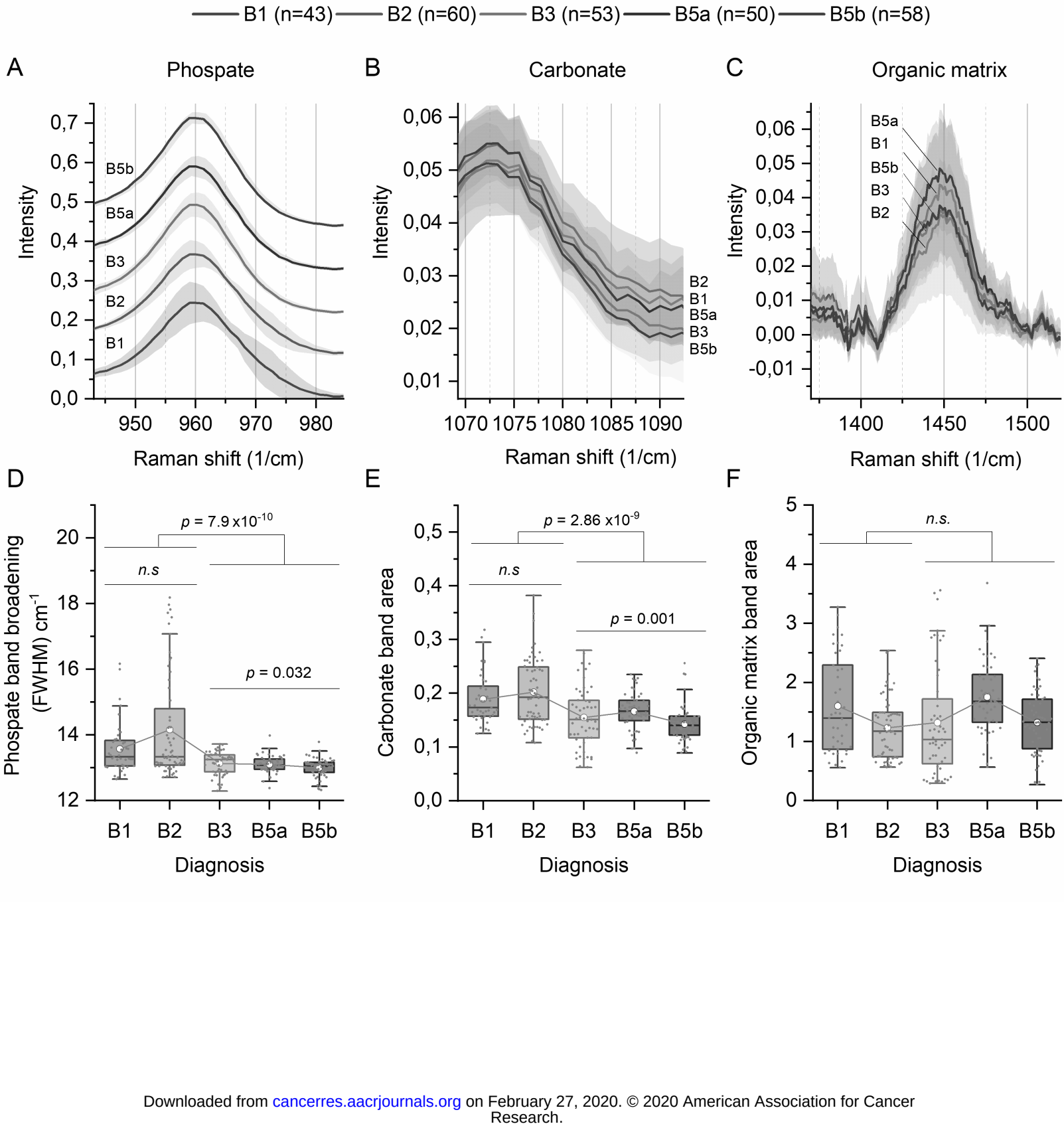


Figure 2



Author Manuscript Published OnlineFirst on February 24, 2020; DOI: 10.1158/0008-5472.CAN-19-3204
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 3



Author Manuscript Published OnlineFirst on February 24, 2020; DOI: 10.1158/0008-5472.CAN-19-3204
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 4

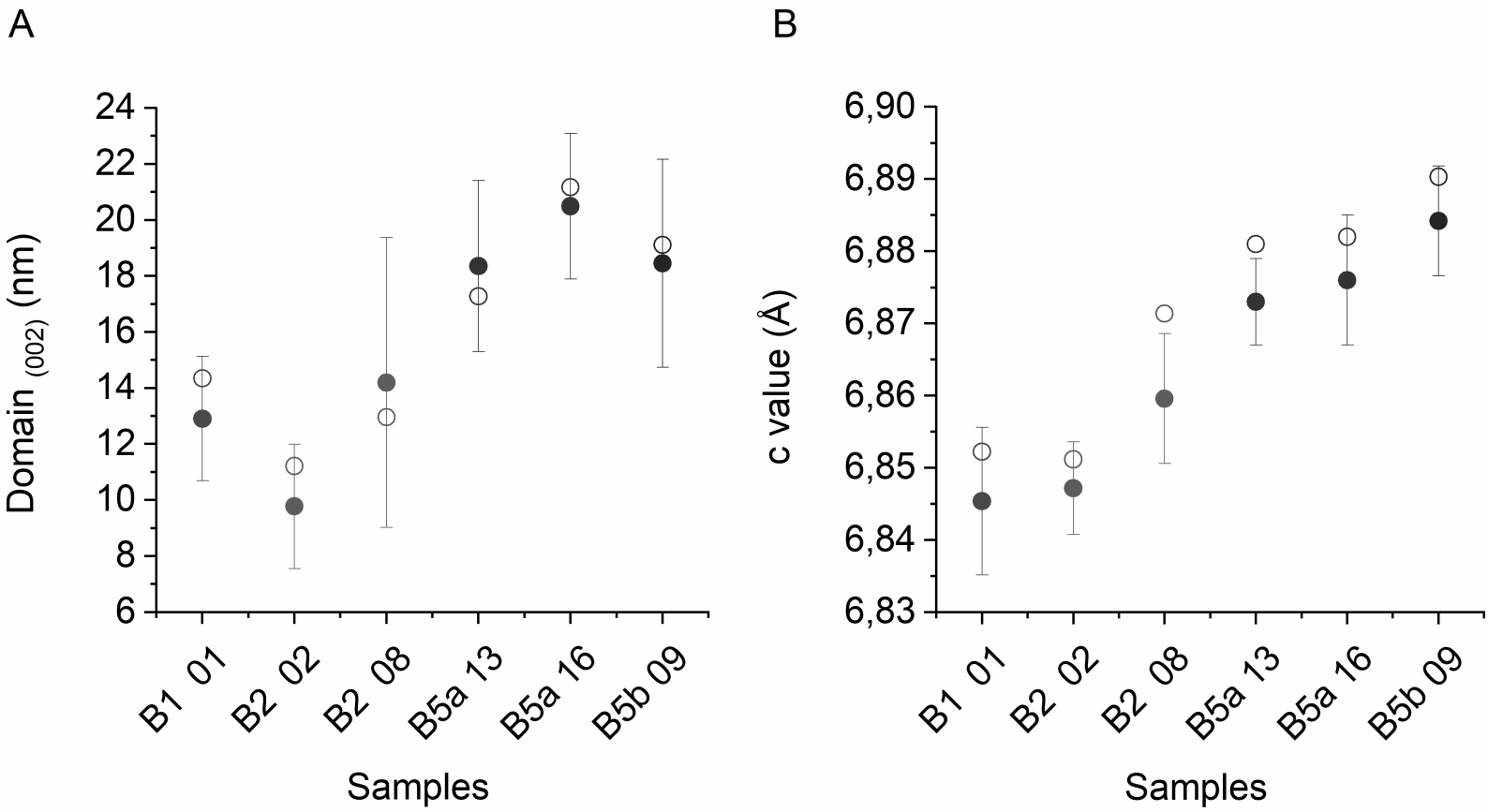


Figure 5

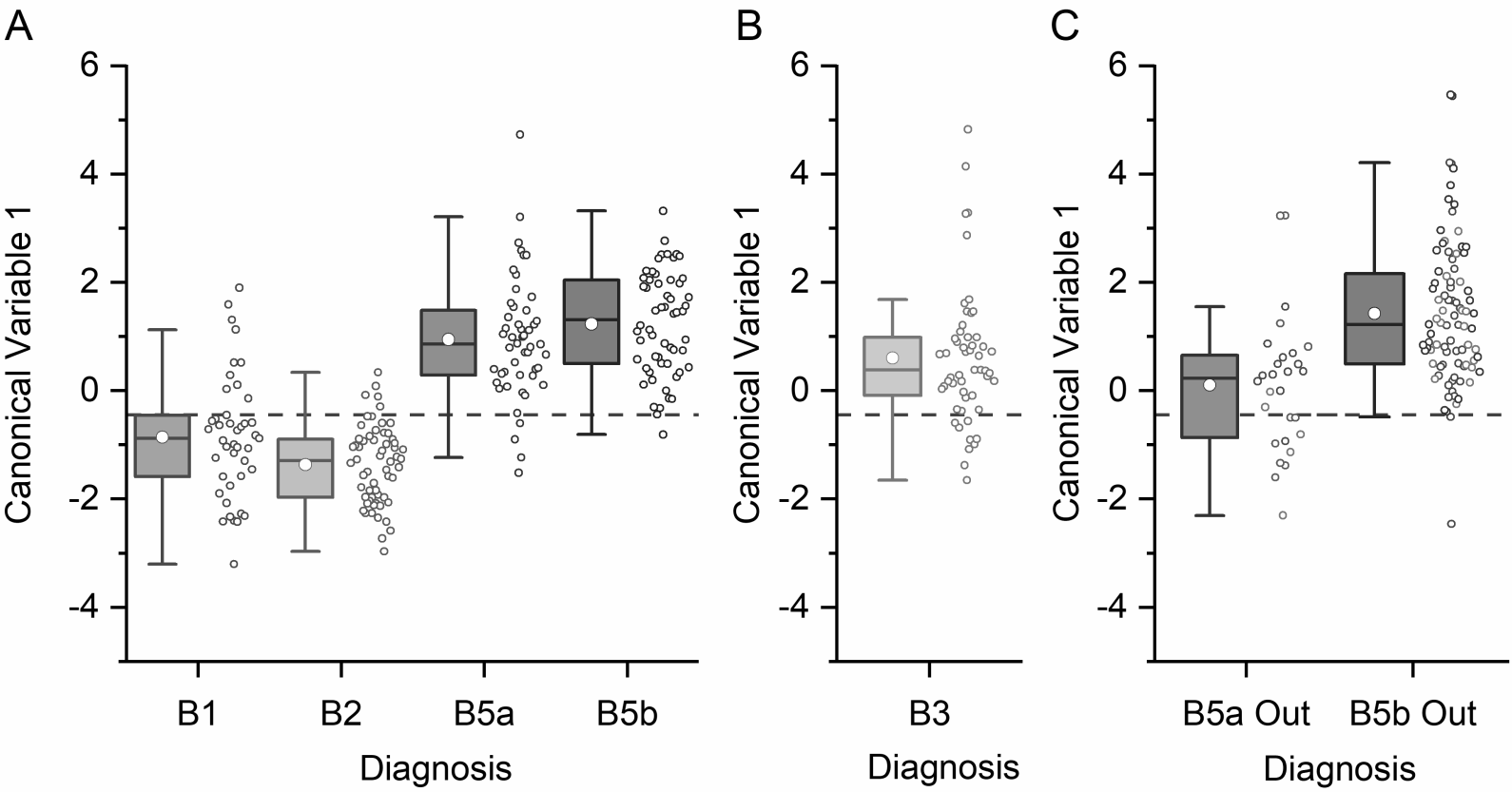
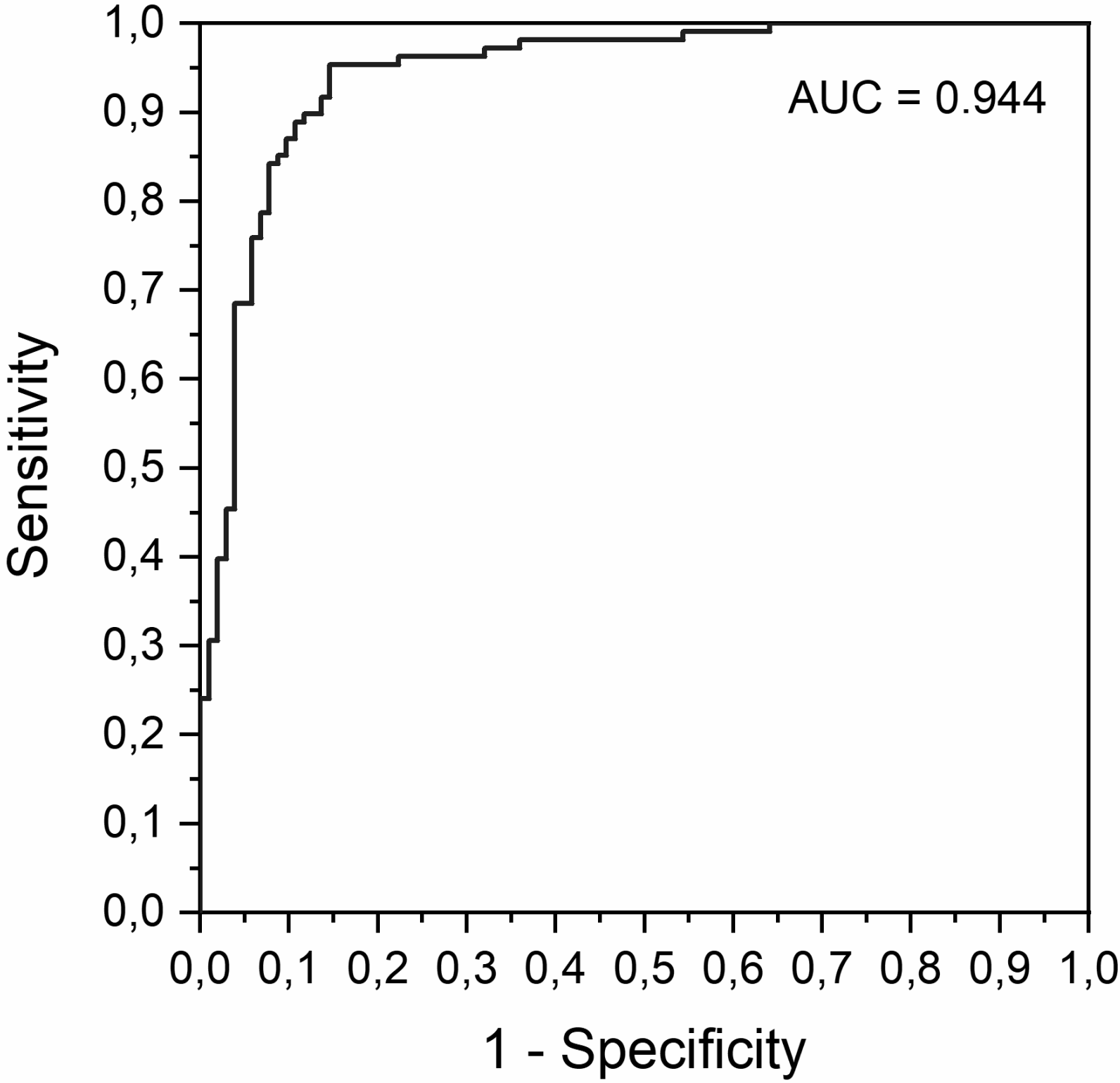


Figure 6



AACR American Association for Cancer Research

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Raman Spectroscopy reveals that biochemical composition of breast microcalcifications correlates with histopathological features

Renzo Vanna, Carlo Morasso, Beatrice Marcinnò, et al.

Cancer Res Published OnlineFirst February 24, 2020.