

UNRAVELLING THE METABOLIC PROGRESSION OF BREAST CANCER CELLS TO BONE METASTASIS BY COUPLING RAMAN SPECTROSCOPY AND A NOVEL USE OF MCR-ALS ALGORITHM

Monica Marro ^{a,*}, Claudia Nieva ^b, Anna de Juan ^c, and Angels Sierra ^{d,e,**}

^a ICFO- ICFO-Institut de Ciències Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels, Barcelona, Spain

^b IDIBELL-Institut d'Investigació Biomèdica de Bellvitge, Av. Castelldefels, Km 2.7, 08907 L'Hospitalet de Llobregat, Barcelona, Spain

^c Department of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona. Diagonal 645, 08028 Barcelona, Spain

^d Molecular and Translational Oncology Laboratory, Biomedical Research Center CELLEX-CRBC, Institut d'Investigacions Biomèdiques August Pi i Sunyer-IDIBAPS, Centre de Recerca Biomèdica CELLEX, 08036 Barcelona, Spain

^e Faculty of Sciences, Universitat de Vic-Universitat Central de Catalunya, 08500 Vic.

Email: *: monica.marro.san@gmail.com, monica.marro@icfo.eu, www.icfo.eu; **: asierrajim@gmail.com; masierra@clinic.ub.es; <http://www.hospitalclinic.org/>. <http://www.idibaps.org/>

ABSTRACT

Raman spectroscopy (RS) has shown promise as a tool to reveal biochemical changes that occur in cancer processes at the cellular level. However, when analysing clinical samples, RS requires improvements to be able to resolve biological components from the spectra. We compared the strengths of Multivariate Curve Resolution (MCR) versus Principal Component Analysis (PCA) to deconvolve meaningful biological components, formed by distinct mixtures of biological molecules, from a set of mixed spectra. We exploited the flexibility of the MCR algorithm to easily accommodate different initial estimates and constraints. We demonstrate the ability of MCR to resolve undesired background signals from the RS that can be subtracted to obtain clearer cancer cell spectra. We used two triple negative breast cancer cell lines, MDA-MB 231 and MDA-MB 435, to illustrate the insights obtained by RS that infer the metabolic changes required for metastasis progression. Our results show that increased levels of amino acids and lower levels of mitochondrial signals are attributes of bone metastatic cells, whereas lung metastasis tropism is characterized by high lipid and mitochondria levels. Therefore, we propose a method based on the MCR algorithm to achieve unique biochemical insights into the molecular progression of cancer cells using RS.

INTRODUCTION

Metastatic cancer cells are characterized by reprogramming of the cellular metabolism that leads to increased uptake of glucose for use as an anabolic and catabolic substrate, thus satisfying the increased demands for macromolecules and energy for proliferation¹. Recent investigations into the mechanisms that underlie the Warburg effect, also known as aerobic glycolysis, suggested that mitochondrial uncoupling can promote aerobic glycolysis, which may represent a shift to the oxidative metabolism of non-glucose carbon sources². Understanding and analysing mitochondrial uncoupling and the metabolic changes observed in cancer cells are important because they may be associated with increased resistance to chemotherapeutic insults³. Moreover, increased mitochondrial oxidative metabolism may be a key driver of tumour cell metastasis⁴. We have previously studied the biochemical characteristics of breast cancer metastatic cells by Raman spectroscopy (RS)⁵, a technique based on the inelastic dispersion of monochromatic radiation that provides specific vibrational signatures of chemical bonds and therefore can be used to determine the chemical composition of cells⁶. RS has been used in the past decade to study

1 biological samples with the help of statistical analysis. Various techniques have been used to find groups and clusters in the RS data
2 set, such as Principal Component Analysis (PCA)^{5,7,8}, K-means clustering^{9,10}, and Neural networks¹¹. Partial-Least-Squares
3 Discriminant Analysis (PLS-DA) allows the construction of discriminatory models to distinguish different types of samples and thus
4 it has been used to differentiate breast cancer metastatic cells from non-metastatic cells without a basal-like phenotype⁵. Two-
5 dimensional correlation has been used to gain insights into the Raman bands and band evolution such as broadening, shifts, etc.¹² to
6 study dynamic processes. However, the multivariate techniques mentioned above do not provide sufficient information to extract the
7 molecular picture of the process being analysed and thus do not exploit all the capabilities of RS. Unmixing techniques such as
8 Vertex Component Analysis (VCA)^{10,13,14} has been used to analyse Raman spectral data in an effort to resolve more “pure” signals.
9 Although VCA and N-FINDR have been often used, these algorithms assume the presence of “pure” pixels in the data set (which in
10 the case of Raman spectra of biomedical samples is very complicated due to the highly mixed molecular content present). Second,
11 they need to use necessarily non-negativity and closure constrains to work, which is not always possible if Raman spectra are
12 transformed, e.g., into derivative spectra, or when pixel-to-pixel high variations of total signal exist. Within the group of unmixing
13 algorithms, Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS), is a very flexible algorithm in which many
14 constraints can be easily included, like non-negativity, selectivity or local rank^{15,25}, information of presence/absence of components
15 in different data blocks in multiset structures, (so-called correspondence among species²²). However, all constraints can be
16 optionally applied in MCR and none of them is compulsory. In this paper, we used non-negativity in the concentration and spectral
17 direction and the selectivity constraint linked to spectra of background contributions in the MCR algorithm in order to resolve more
18 easily meaningful spectral profiles of biological components from the mixed Raman spectra of cancer cells.

19 There are various drawbacks that need to be overcome to fully exploit RS for diagnostic purposes¹². First, cells are grown on the
20 surface of glass materials and the focus of the laser needs to be close to or past the surface window. Thus part of the confocal
21 volume analysed includes a portion of the glass, which consequently produces undesired background signals in the RS. Second,
22 samples contain stains and are prepared on cover slips that give a very high signal and disturb the inherent cellular RS. Bonnier et al.
23 ¹⁷ tried to remove the effect of the background signals in the spectra by growing the cells in collagen matrices on the top of
24 coverslips but this requires sample processing and modification of already established biomedical protocols. Finally, in applications
25 for in-vivo RS analysis which involve fiber optical probes, high background signals from silica perturb the Raman spectra and
26 consequently reduce the ability to perform stable and robust statistical models. Therefore, further development is essential to reduce
27 background spectral signals in these cases^{18, 19, 20}.

28 In this study, we present a powerful statistical methodology to remove undesired background signals and extract meaningful
29 molecular components from Raman spectra. The method is based on the use of MCR-ALS, a very flexible and powerful algorithm
30 that has been applied to several chemical problems²¹ Recently, it has been applied to biomedical RS data^{22, 23, 24, 25}. Specifically, we
31 first compared the ability of PCA and MCR analysis applied to the same RS data to resolve molecular spectral contributions.
32 Second, the use of the flexibility of the MCR-ALS algorithm was used to include specific constrains (non-negativity and selectivity
33 constraint) to extract the quartz and water background signal from the spectral data and remove it from the RS. Both procedures
34 have been used to gain insights into the metabolic changes that breast cancer cells undergo while progressing with different
35 metastasis tropisms²⁶.

36 We used two triple negative human breast cancer cell lines, MDA-MB-231 and its osteotropic cell line B02 subclone and MDA-
37 MB-435 (435P) with tropism to metastasize in the lung and its bone metastatic variant 435-B. We carried out RS and MCR-ALS
38 analysis to study biochemical differences between metastasis tropisms. The results showed that an increase in amino acids and a

1 decrease in mitochondrial signals were attributes of bone metastatic cells, whereas lung metastasis tropism was characterized by
2 high lipid and mitochondria (cytochrome C and RNA) levels. We demonstrate that RS coupled with the MR-ALS algorithm resolve
3 meaningful molecular components during metastatic progression of cancer cells, giving more biochemical information than other
4 commonly used mathematical techniques. Furthermore, we propose exploiting the flexibility of the MCR-ALS algorithm to enhance
5 and expand the strengths of RS for clinical diagnosis, and demonstrate the successful subtraction of undesired background signals
6 (that are unavoidable in clinical practice), allowing the extraction of more valuable information from RS.
7
8
9

10 11 12 **MATERIALS AND METHODS**

13 **Cell lines used**

14 We used the breast cancer bone metastatic cell line MDA-MB-231 (231P), originally obtained from the European Type Culture
15 Collection (ECACC 92020424), and B02 (kindly provided by Philippe Clezardin, INSERM, UMR 664, IFR62, Laennec School of
16 Medicine, F-69372, Lyon, France) established from bone metastases caused by MDA-MB-231 after six *in vivo/in vitro* passages in
17 nude mice using a heart injection model²⁷. The proprietary 435-B cells were obtained from primary cultures of bone metastasis
18 induced by MDA-MB 435 (435P) cells inoculated in the intramammary fat path²⁸. The proprietary B02/PRDX2 cells^{29, 30, 31} were
19 previously obtained from highly bone metastatic BO2/GFP cells overexpressing PRDX2 (BO2/PRDX2) through the pBIL vector
20 system.
21
22

23 Cells were cultured under standard conditions in DMEM/F12 medium supplemented with 5% horse serum, 1 mM pyruvate, 2 mM
24 L-glutamine in 5% CO₂-95% air at 37°C in a humidified incubator.
25

26 **Raman spectroscopy (RS)**

27 For RS analysis $3 \cdot 10^5$ cells were used and measurements were carried out by seeding cells over a quartz crystal (ESCO products,
28 Oak Ridge, NJ) in six-well plates (Becton Dickinson, NJ). After 24 h, cells were fixed with 4% cold paraformaldehyde (PFA) in
29 PBS1x for 15 min, washed with PBS 1x and maintained in the same solution at 4°C until the spectroscopic analysis.
30

31 A total of 225 spectra were obtained from the cell cytoplasm (few μm away from the nucleus, see the inset in Fig 2b), from both
32 the parental cell lines and their corresponding metastatic variants.
33

34 The Raman system used was an inVia Renishaw (Apply Innovation, Gloucestershire, UK) and comprises a 532 nm laser that
35 supplies an excitation beam of about 10 mW power, which is focused onto the sample via a microscope with a 60x objective
36 (Edmund, York, UK) using a backscattered configuration. The laser spot size in the sample is 0.8 μm . The Raman spectrum is
37 recorded on a deep depletion charge-coupled device (CCD) detector (Renishaw RenCam). The recorded Raman spectrum is
38 digitalized and displayed on a personal computer using Renishaw WiRE software, which allows the experimental parameters to be
39 set. The spectra were background subtracted with a custom-written Labview program using an established method³².
40
41
42
43
44
45
46
47

48 **Statistical analysis**

49 First, an exploration of the spectral dataset was performed using PCA. The number of components selected was the number in
50 which the cumulative variance explained was more than 99%. Two independent analyses (PCA and MCR) were performed on the
51 same dataset, which contained the spectra of the cell lines in rows. Additionally, to improve background component resolution five
52 spectra from the PBS solution on the same quartz substrate and at the same focal distance as the cell measurements and far away
53 from the cells were added. These spectra were used to include selectivity constraints to the MCR algorithm indicating the
54 presence/absence of molecular components in some of the measured spectra. For the MCR-ALS algorithm, the protocol followed
55 was: first, the initial estimates were selected using a purest variable selection method algorithm, which selects the purest rows
56 (spectra) in the dataset; second, the iterative least squares calculation started with the addition of constraints (non-negativity of
57
58
59
60

spectra and concentration matrices and selectivity constraint for concentrations linked to spectra with only background

contributions). Finally, the quality parameters from the MCR model were calculated: Lack of fit:

$$LOF = 100 \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m e_{i,j}^2}{\sum_{i=1}^n \sum_{j=1}^m d_{i,j}^2}}$$
 and

$$R^2 = 100 \frac{\sum_{i=1}^n \sum_{j=1}^m d_{i,j}^2 - \sum_{i=1}^n \sum_{j=1}^m e_{i,j}^2}{\sum_{i=1}^n \sum_{j=1}^m d_{i,j}^2}$$

the parameter: where $e_{i,j} = d_{i,j} - \hat{d}_{i,j}$ and $d_{i,j}$ is the original data matrix and $\hat{d}_{i,j}$ the calculated by means of the MCR-ALS algorithm.

For PCA and MCR analysis two different Matlab toolboxes were used: PLS toolbox (from Eigenvector Research) and MCR-ALS toolbox (from University of Barcelona).

RESULTS AND DISCUSSION

Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) is able to resolve more meaningful molecular components than Principal Component Analysis (PCA) on Raman spectra (RS)

First, a PCA was performed on the spectral data matrix containing 135 spectra from the cytoplasm area (avoiding the nucleus) of 231, B02, 435 and 435-B cells and five more from PBS, from which the principal component loadings were plotted (Fig. 1A). These results were compared with those obtained from the MCR-ALS algorithm subjected to specific constraints (non-negativity of spectra and concentrations, and selectivity constraint) (Fig. 1B, 1C) and applied to the same spectral dataset. We demonstrated that when using MCR-ALS more chemically meaningful molecular components are extracted. Furthermore, we were able to separately resolve two components containing quartz and water signals.

The results of PCA (Fig. 1A) showed that the signals of different substrates were mixed in among the loadings of the principal components. The contributions of quartz signal^{17, 33} (mainly bands around 800 and 1050 cm^{-1}) were found in PC3, together with water signal and other cellular substrates. Indeed, the presence of positive and negative bands in the principal component loadings made the assignment of each component to a molecular RS contribution difficult. In contrast, MCR decomposed four meaningful components (Fig. 1B) that could be assigned to cellular lipids (component 1), quartz signal (component 2), water signal³¹ (component 3) and proteins and other cellular components (component 4). The variance explained by PCA and MCR models was 99.74% and 99.66%, respectively.

MCR spectra provided more chemically meaningful results than PCA, providing two components, 2 (quartz) and 3 (water), that did not correspond to cellular molecular components^{17, 33, 34}. Indeed, the higher ability of MCR to resolve pure signals relied on the flexibility of the MCR-ALS algorithm. First, the spectral data matrix was constructed by displaying all cell RS in rows and adding five more rows containing the signal from PBS, acquired at the same focal distance as for the cells on the quartz substrate. Second, a purest variable selection algorithm was used to find the purest spectra in the dataset in order to provide initial estimates of the MCR-ALS algorithm. As a result the four most dissimilar spectra found were: two signals from PBS solution and two from different cell lines. Once the initial estimates were selected, the iterative least square calculation was initialized by placing constraints on the algorithm. First, we included non-negativity in the resolved spectra and concentration profiles. Second, a matrix showing the presence/absence of molecular components such as water, quartz and cellular spectra was encoded (Fig. 1c). In the rows where there was absence of cellular spectra, zeros were included to the components from Quartz and Water that we wanted to resolve. The above process yielded four chemically and physically meaningful components.

1 Unlike MCR, PCA found the orthogonal directions in the dataset with the maximum variance, but no additional chemical
2 information could be added. Consequently, in the case of complicated biological Raman datasets, PCA could be used for an
3 exploratory analysis of the dataset to find for instance outliers or groups or to understand the chemical structure of the data as a first
4 approach, but not meaningful molecular spectral components can be unmixed using this method. This problem has been mentioned
5 in previous cellular studies³⁵ and in studies involving image segmentation³³.
6
7
8

10 **Subtraction of the background signals by MCR allowed the resolution of metabolites in metastatic breast cancer cells**

11 In order to clean and remove quartz and water signals from the original spectral dataset, direct subtraction of both signals (MCR
12 loadings) multiplied by the concentration matrix (MCR scores) of those components for each original spectrum was performed. This
13 led to greater visibility of some spectral changes inherent to the cellular spectra, making it easier to build a statistical model with
14 chemically meaningful outputs. The results are shown in Figure 2, where the plots of spectra before (Fig. 2A) and after (Fig. 2B) the
15 subtraction of quartz and water signals are compared. Furthermore, by using this method, statistical analysis of cell RS is improved
16 as long as the background is removed. Thus, based on the clean cell RS, a new MCR model was calculated to resolve pure
17 molecular RS involved in the different biochemical components of metastatic progression (Fig. 3).
18

19 Recently, various applications of MCR to Raman images have detected the contribution of background signals in the resolved
20 components^{36, 37, 38, 39}. However, to our knowledge this is the first time that the MCR-ALS algorithm has been proposed as a tool to
21 subtract the background signal contributions in the Raman analysis of cells when using standard sample preparation. For clinical
22 relevancy and acceptance of RS by the clinical community, it is important that the sample preparation protocols necessary for RS
23 are coincident with current practice. Nowadays this is a challenging issue and researchers are working to overcome this limitation.
24 For instance, a study compared Raman signals of unprocessed tissue preparations and dewaxed tissue sections, demonstrating that
25 tissue processing has a significant impact on the extracellular structure, in particular on the lipidic structure⁴⁰. Also, in the clinical
26 applications of RS related with endoscopy, background signals due to the use of fiber optics have been reported as one of the main
27 issues^{41, 19}. Furthermore, the use of routine stained tissue sections or cells for RS analysis has been limited due to background signals
28 that interfere with the Raman spectra⁴². Therefore, to prevent the loss of critical cellular information, RS should be compatible, as
29 much as possible, with conventional sample preparation, without the need for further processing. This challenge might be overcome
30 by resolving non-inherent signals from current cytological preparations from the cell spectra. In this way, everyday techniques could
31 be performed in parallel with Raman spectroscopy analysis using the same samples. Moreover, the use of stored samples might
32 permit retrospective studies based on prognosis and patient follow-up.
33

34 It is important to note that from a matrix with very similar cell RS the MCR-ALS algorithm permitted the resolution of three
35 meaningful molecular components. Indeed, RS were assigned to almost pure molecules. Therefore, this analytical method extracted
36 more biochemical information than other commonly used techniques, such as PCA or cluster analysis, providing information about
37 the biochemical changes occurring in metastatic cancer cells (Fig. 3). The concentrations of each MCR component for each cell
38 variant measured were plotted (Fig. 3 right) with their corresponding spectra (Fig. 3 left). In particular component 1 was assigned to
39 cell lipids⁴³. Raman bands in component 2 have been related to amino acids and cell proteins^{44, 45}. Finally, component 3 contained
40 bands previously assigned to mitochondria, containing bands from cytochrome C⁴⁶⁻⁵⁴ and DNA or nucleotides⁵⁵. As mitochondria
41 are the organelles in the cytoplasm that contain the highest concentrations of DNA and cytochrome C, this led to the conclusion that
42 component 3 could be used as a measure of the number of mitochondria present in breast cancer metastatic cells. Table 1 shows the
43 main assignments of Raman bands from each component.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 When comparing the three main biochemical components that differentiate cells, we observed that component 1, assigned to cell
2 lipids, was inversely proportional to capacity for bone metastasis such that its value decreased as long as the bone tropism increased.
3 This is in agreement with the described "low-lipid" phenotype that characterized breast cancer cells compared with normal breast
4 epithelial cells, and negatively correlated with invasiveness and metastatic progression⁵⁶. The lowest level of component 1 was
5 found in B02 cells and the highest level was found in 435P cells, which showed the greatest lung metastatic tropism. Moreover, we
6 were able to distinguish two different populations belonging to 435P cells, one with lower lipid content that was clearly enriched in
7 the 435-B bone metastatic variant (Fig. 3A). These clear differences in lipid level between the two triple negative (231P and 435P)
8 breast cancer metastatic cell lines might be due to the lung metastatic background that characterizes 435P cells, in contrast to 231P
9 cells that have greater bone metastatic ability. Therefore, the low fat content is associated with bone metastatic ability.

10 Component 2, which was assigned to protein content, increased in parallel with increasing bone metastasis tropism. The lowest level
11 of component 2 was found in 435P cells, and the highest level was found in the B02 bone metastatic variant. Moreover, 435-B had a
12 higher amino acid content than 435P (Fig. 3B).

13 Component 3, which contain bands from cytochrome C⁴⁶⁻⁵⁴ and DNA or nucleotide bands⁵⁵, was inversely related to bone metastatic
14 ability. Thus, similarly to lipid content, scores in Component 3 increased in 435P cells with regard to bone metastatic ones (Fig.
15 3C). In summary, the MCR-ALS algorithm applied to a set of cytological RS provided insights into the subtle metabolic changes
16 that breast cancer cells undergo on the path towards bone metastasis tropism, as revealed by the different profiles obtained in the
17 analysis of 231P cells, with tropism to metastasize in bone, and 435P, with tropism to metastasize in lung. Moreover, the respective
18 bone metastatic variants, B02 and 435-B, were classified according to their corresponding phenotype. These results are consistent
19 with reports demonstrating that the mechanism used to cope with the cumulative reactive oxygen species differs between lung and
20 bone tissues, with the lower oxygen gradient in bone marrow inducing fewer metabolic free radical challenges⁵⁴.

21 Mitochondria are an important cell metabolic hub because they orchestrate the biosynthesis of lipids, proteins, and nucleic acids and
22 maintain redox homeostasis via the generation of NADH and NADPH⁵. Although in normal cells the TCA cycle is mostly involved
23 in glucose oxidation, the mitochondrial metabolism of cancer cells is generally uncoupled from glycolysis⁵⁸. It has been reported
24 that mitochondrial glucose oxidation may be incompatible with the survival of some cancer cells⁵⁹. Due to genetic and
25 environmental cues, such as hypoxia, the enzyme responsible for the entry of glucose-derived carbons into the mitochondria, the
26 pyruvate dehydrogenase complex (PDH), is frequently inactivated in cancer⁶⁰. This mechanism might be particularly important in
27 metastatic cells in an hypoxic organ like bones. Furthermore, the homeostatic mechanism of the antioxidant system might prevent
28 collapse by redirecting the glycolytic flux into the pentose phosphate pathway⁵⁸. All together, these results clearly indicate that a
29 metabolic switch in breast cancer cells might be used to fit with the microenvironment at the metastatic loci.

30 **RS supported by MCR-ALS algorithm dissects metabolic functions of metastatic breast cancer cells**

31 Since breast cancer bone metastatic cells might be characterized by a clear increase in protein content and a lower lipid content than
32 other breast cancer cells, we analyzed B02 cells overexpressing PRDX2 (B02/PRDX2 cells) that exhibited a low bone metastatic
33 burden *in vivo*²⁹. PRDX2 is a highly efficient redox protein that neutralizes hydrogen peroxide, resulting in protection of cells from
34 oxidative damage and in regulation of peroxide-mediated signal transduction events⁶². The low metastatic burden of B02/PRDX2
35 cells in bones suggests that PRDX2 might interfere in the preferential bone metastatic cells' metabolism with deleterious
36 consequences for carcinoma cells that try to adapt to the bone microenvironment. This model was included in the study to better
37 assess the sensitivity of RS-MCR to differentiate small molecular changes that induce important functional cell changes.

1 The model described above was tested in an independent matrix containing the RS of BO2, 231P and B02/PRDX2 cells. First, the
2 signals from quartz and water were subtracted following the method described before and then an MCR was built to confirm the
3 molecular behavior of the different cell lines and specifically the role of PRDX2 up-regulation Fig. 4.

4
5
6 MCR analysis of these experiments revealed that lipid levels were higher in B02/PRDX2 cells as reflected in Component 1 levels
7 (Fig. 4 A) and strongly correlated with the above observation that the lipid content of cells was inversely related to the bone
8 metastasis potential (Fig. 4 A). In fact, as previously demonstrated⁶³, the preferential bone metastatic metabolism of B02 cells can be
9 disrupted by overexpressing PRDX2, which induces low glucose dependence. These results suggest a metabolic shift from glucose
10 oxidation to fatty acid oxidation that would resist metabolic insults in an hypoxic and hypoglycemic tissue like bones.

11
12
13 Moreover, Component 2 (Fig. 4 B) (related to amino acids) was found at lower levels in B02/PRDX2 cells in comparison to B02
14 cells, but at similar levels to those in MDA-MB-231 cells, supporting the idea that lower amino acid content is associated with
15 higher bone tropism. In addition, Component 3 was found at higher levels in B02/PRDX2 cells, coinciding with the fact that
16 cytochrome C was previously found to be inversely associated with bone tropism, increased from B02 to B02/PRDX2 cells that is
17 not able to metastasize in bones (Fig. 4 C). This is in agreement with the results showed in Fig 3 C, where we defined bone
18 metastasis tropism, and which all together supported the biological use of MCR applications to sensitively analyze the fine
19 biochemical changes that occur in bone metastasis progression. Moreover, a PLS-DA analysis to discriminate B02 cells from
20 MDA-MB-435P was performed (Fig. 5) that achieved a specificity of 100% and specificity of 90%. Cross-validation was performed
21 using 10 samples that were excluded from the dataset. This corroborates our hypothesis and highlighted that RS is able to assess and
22 identify bone tropism in breast cancer cells

23
24
25 Despite advances in treatment, 20% to 30% of patients with early breast cancers will experience relapse with distant metastatic
26 disease⁶³. Triple negative tumours have the highest metastatic rate with regard to luminal and HER2 positive tumors, with bone and
27 lung being the main target of breast cancer metastasis. Transcriptomic analysis of a variety of cell lines has identified the genes that
28 mediate metastasis to bone or lungs^{64, 65, 66}, but in clinical settings breast cancer subtyping is based on pathological characteristics⁶⁷.

29
30
31 With the development of metabolomics we have learnt that metabolic variability and flexibility enable tumour cell progression by
32 generating ATP as an energy source⁶⁸. Three major components of the metabolic transformation of cancer cells have been
33 described⁶⁹: aerobic glycolysis, mitochondrial metabolism, and lipid synthesis.

34
35
36 We provide data demonstrating that RS coupled with the MCR-ALS algorithm is a powerful method that can be used in a clever and
37 flexible way to reveal the underlying molecular information stored in RS, allowing a further step forward in the analysis of
38 biological RS relative to the multivariate techniques already used to analyse metabolic changes in cancer cells. Moreover, RS
39 provides a unique possibility for real-time and non-intensive-processing of tissue material for the detection of normal, precancerous
40 and cancerous breast tissues⁷⁰. We propose the use of the MCR-ALS algorithm to enhance and expand the strengths of RS for
41 clinical diagnosis.

42 CONCLUSION

43
44
45 We propose a new methodology for extracting the useful molecular information encoded in RS of biological samples. By using the
46 flexible MCR-ALS algorithm meaningful molecular components can be extracted, thanks to the prior subtraction of background
47 signals that perturb the inherent cell RS and including specific constrains indicating the presence/absence of molecular components.
48 This methodology opens up new possibilities for the clinical diagnostic use of RS, since it permits the removal of signals from
49 substrates or chemicals used in cytological techniques. It also provides a rapid, reliable and label free method with which to
50

1 disentangle, in cancer cytology, the biochemical components involved in metastasis progression, and as such could improve
2 prognosis and early prevention.
3

4 We applied this methodology to study the molecular differences between primary breast cancer cell lines and their metastatic
5 variants in bone. We identified biochemically meaningful molecular components that play a role in the metastatic ability of breast
6 cancer cells. Specifically, the protein component was found at higher levels whereas mitochondria and lipids bands were found at
7 lower levels in bone metastatic cell variants (B02) in relation to their corresponding parental cell lines (231P and 345P). Thus, a
8 ratio of these three components might be used to validate this data in clinical samples. In addition, a PLS-DA was performed
9 showing high sensitivity and specificity in the discrimination of B02 and MDA-MB-435P cells.
10
11
12
13
14
15

16 ACKNOWLEDGMENTS AND FUNDING

17
18 This study was supported by grants from the Spanish Ministry of Health and Consumer Affairs FIS-PII4/00336, both from the
19 I+D+I National Plan with financial support from ISCIII-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo
20 Regional (FEDER), and by grants from 2014 SGR 530 and from the Fundació Privada Cellex.
21
22

23
24 Dr. Marro acknowledges financial support from the Spanish Ministry of Economy and Competitiveness through the “Severo Ochoa”
25 program for Centres of Excellence in R&D (SEV-2015-0522), from Fundació Privada Cellex, Fundació Mig-Puig, from
26 Generalitat de Catalunya through the CERCA program and from Laserlab-Europe (EU-H2020 654148). This research has been
27 partially conducted at ICFO’s Super Resolution Light Microscopy and Nanoscopy Facility.
28
29

30 The authors are grateful to Toffa Evans from Assessorament Lingüístic i Terminologia Universitat de Barcelona for language advice
31 and to the COST action BM1401, Raman4Clinics, <https://www.raman4clinics.eu/raman4clinics-a-european-cost-action/>.
32
33
34

35 FIGURES

36
37 **Figure 1. MCR-ALS analysis gives more chemically meaningful components than PCA.** PCA (A) and MCR-ALS (B) loadings
38 of the matrix spectral dataset containing the Raman spectra (RS) obtained from the cytoplasm of cancer cell lines 231P, B02, 435P
39 and 435B and PBS. (C) MCR-ALS constraints were applied to the data: non-negativity and selectivity constraint to resolve quartz
40 and PBS spectra.
41
42

43 **Figure 2. Subtraction of background signals in cellular RS using MCR-ALS algorithm:** RS before (A) and after the subtraction
44 of quartz and water signals (B). MCR loadings from Figure 1B were used to resolve the meaningful molecular contributions. The
45 figure inset in (B) shows an image of a MDA-MB-231 cell indicating with green arrows the areas in which the Raman spectra was
46 obtained. A Raman image shows the lipid distribution (Component 1) in the cytoplasm of the cell.
47
48

49 **Figure 3. Using MCR-ALS analysis of Raman spectra from 231P, B02, 435P and 435B cell lines to delineate bone metastasis**
50 **ability.** An improved MCR-ALS model was built using RS after the subtraction of quartz and water signals. Three components were
51 resolved and their respective MCR loadings (left) and scores (right) are plotted. Red triangles: B02, blue squares: 231P, blue
52 crosses: 435B and white diamonds: 435P. Components 1 (A), 2 (B), and 3 (C) have bands compatible with cellular lipids, amino
53 acids and mitochondria-related molecules, respectively. The variance explained for each component is shown as a %.
54
55

56
57 **Figure 4. MCR analysis of Raman spectra from 231P, B02 (the high bone metastatic cell variant) and B02-PRDX2 that over-**
58 **express PRDX2 and have low bone metastasis ability *in vivo* (Stresing, 2011).** Three components were resolved and their
59 respective MCR loadings (left) and scores (right) are plotted. Red triangles: B02, blue squares: 231P, green stars: B02-PRDX2.
60

Components 1 (A), 2 (B), and 3 (C) have bands compatible with cellular lipids, amino acids and mitochondria-related molecules, respectively. The variance explained for each component is shown as a %.

Figure 5: A specificity of 100% and specificity of 90% is achieved with the use of PLS-DA analysis to discriminate BO2 cells from MDA-MB-435P. This corroborates that RS is able to assess and identify the bone tropism of breast cancer cells.

TABLE

Table 1: Raman Spectra band assignments of MCR components obtained from the MCR-ALS analysis in Fig. 3. Assignments based on references⁴³⁻⁵⁵.

Component 1		Component 2		Component 3	
Band (cm ⁻¹)	Assignment	Band (cm ⁻¹)	Assignment	Band (cm ⁻¹)	Assignment
836	Saccharide	693	Amino acid methionine	750	Cytochrome C
1057	Triacylglycerols	710	Amino acid methionine	777	Cytosine, Thymine
1075	Triacylglycerols	822	C-C stretch	804	Phosphoric acid
1119	Triacylglycerols	997	Phenylalanine	996	Phenylalanine
1260	Fatty acid, Cholesterol, Triacylglycerols, membrane lipids	1026	Phenylalanine	1082	Phospholipids, nucleic acid
1295	Fatty acid	1082	Glutamic acid	1120	Cytochrome C
1442	Fatty acid, Cholesterol	1122	C-N proteins	1203	Nucleic acid
1660	Unsaturated fatty acid	1201	Trypt, aromatic C-N and amide III	1240	RNA
		1263	Amide III	1312	Cytochrome C
		1293	Alanine	1333	Adenine, Guanine
		1336	C-H, Amide III, Trypt, Glycine	1445	
		1442	CH ₂	1570	Cytochrome C, mitochondria
		1654	Amide I	1585	Cytochrome C, mitochondria

REFERENCES

- [1] Samudio, I.; Fiegl, M.; Andreeff, M. *Cancer R* **2009**, *69*, 2163-2166.
- [2] DeBerardinis, R.J.; Mancuso, A.; Daikhin, E.; Nissim, I.; Yudkoff, M.; Wehrli, S.; Thomson C. B. *PNAS* **2007**, *104*, 19345-19350.
- [3] Farnie, G.; Sotgia, F.; Lisanti, M. *Oncotarget* **2015**, *6*, 30472-30486.
- [4] Sotgia, F.; Martinez-Outschoorn, U. E.; Lisanti, M. *BMC Medicine* **2011**, *9*, 1-5.
- [5] Nieva, C.; Marro, M.; Santana-Codina, N.; Rao, S.; Petrov, D.; Sierra, A. *PlosOne* **2012**, *7*, 1-10. <https://doi.org/10.1371/journal.pone.0046456>
- [6] Fenn, M. B.; Xanthopoulos, P.; Pyrgiotakis, G.; Grobmyer, S. R.; Pardalos, P. M.; Hench, L. L. *Adv. Opt. Techn.* **2011**, *2011*, 1-20. doi:10.1155/2011/213783
- [7] Chan, J.; Taylor, D.; Zwerdling, T.; Lane, S.; Ihara, K. *Biophys. J.* **2006**, *90*, 648-656.
- [8] Huang, W.; Griffiths, R.; Thompson, I.; Bailey, M.; Whiteley, A. *Anal. Chem.* **2004**, *76*, 4452-4458.
- [9] Hedegaard, M.; Krafft, C.; Ditzel, H.; Johansen, L.; Hassing, S.; Popp, J. *Anal. Chem.* **2010**, *82*, 2797-2802.
- [10] Hedegaard, M.; Matthaues, C.; Hassing, S.; Krafft, C.; Diem, M.; J. Popp *Theor. Chem. Acc.* **2011**, *130*, 1249-1260.
- [11] Harris, A.; Garg, M.; Yang, X.; Fisher, S.; Kirkham, J.; Smith, D.; Martin, D.; High, A. *Head Neck Onc.* **2009**, *38*, 1-6.
- [12] Balint, S.; Rao, S.; Marro, M.; Miskovsky, P.; Petrov, D. *J. Raman Spectrosc.* **2010**, *42*, 1215-1221
- [13] Bioucas-Dias, J.; Plaza, A.; Dobigeon, N.; Parente, M.; QuianDu, Gader, P.; Chanussot, J. *J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2012**, *5*, 354-379.
- [14] Winter, M. *SPIE* **1999**, *3753*, 266-275.
- [15] Tauler, R.; Smilde, A.; Kowalski, B. *J. Chemometrics* **1995**, *9*, 31-58.
- [16] Tu, Q.; Chang, C. *Nanomedicine* **2012**, *8*, 545-558.
- [17] Bonnier, F.; Knief, P.; Meade, A.; Dorney, J.; Bhattacharya, K. *Proc SPIE* **2010**, 80870F.
- [18] Kallaway, C.; Almond, L. M.; Barr, H.; Wood, J.; Hutchings, J.; Kendall, C.; Stone, N. *Photodiagnosis Photodyn. Ther.* **2013**, *10*, 207-219.
- [19] Pence, I.; Mahadevan-Jansen, A. *Chem Soc Rev.* **2016**, *45*, 1958-1979.
- [20] Kong, K.; Kendall, C.; Stone, N.; Notingher I. *Adv. Drug Deliv. Rev.* **2015**, *89*, 121-134.
- [21] Azzouz, T.; Tauler, R. *Talanta* **2008**, *74*, 1201-1210
- [22] Marro, M.; Nieva, C.; Sanz-Pamplona, R.; Sierra, A. *BBA Mol. Cell. Research* **2014**, *1843*, 1785-1795
- [23] Patel, I.; Trevisan, J.; Evans, G.; Llabjani, V.; Martin-Hirsch, P.L.; Stringfellow, H. F.; Martin, F. L. *Analyst*, **2011**, *136*, 4950
- [24] Marro, M.; Taubes, A.; Abernathy, A.; Balint, S.; Moreno, B.; Sanchez-Dalmau, B.; Martinez-Lapiscina, E. H.; Amat-Roldan, I.; Petrov, D.; Villoslada, P. *J Biophotonics*, **2014**, *7*, 724-734.
- [25] Olmos, V.; Benitez, L.; Marro, M.; Loza-Alvarez, P.; Piña, B.; Tauler, R.; de Juan, A. *TrAC* **2017**, *94*, 130-140.
- [26] Minn, A. J.; Kang, Y.; Serganova, I.; Gupta, G. P.; Giri, D. D.; Doubrovina, M.; Ponomarev, V.; Gerald, W. L.; Blasberg, R.; Massague, J. *J Clin Invest* **2005**, *115*, 44-55.
- [27] Peyruchaud, O.; Serre, C. M.; NicAmhlaioibh, R.; Fournier, P.; Clezardin, P.; *J Biol Chem* **2003**; *278*: 45826-45832.
- [28] Mendez, O.; Fernandez, Y.; Peinado, M.A.; Moreno, V.; Sierra, A. *Clin Exp Metastasis* **2005**; *22*, 297-307.
- [29] Stresing, V.; Baltziskueta, E.; Rubio, N.; Blanco, J.; Arriba, M. C.; Valls, J.; Janier, M.; Clézardin, P.; Sanz-Pamplona, R.; Nieva, C.; Marro, M.; Petrov, D.; Sierra, A. *Oncogene* **2013**, *32*, 724-735.
- [30] Boucharaba, A.; Serre, C.M.; Gres, S.; Saulnier-Blache, J.S.; Bordet, J.C.; Guglielmi, J.; Clézardin, P.; Peyruchaud, O. *J Clin Invest* **2004**, *114*, 1714-1725.
- [31] Román, I.; Vilalta, M.; Rodriguez, J.; Matthies, A. M.; Srouji, S.; Livne, E.; Hubbell, J. A.; Rubio, N.; Blanco, J. *Biomaterials* **2007**, *28*, 2718-2728.

- [32] Mikhailyuk, I.; Razzhivin, A. *Instr. Exp. Tech* **2003**, 46, 765-769.
- [33] Read, D. Springer 2010
- [34] <http://lid.phys.msu.su/englishVersion/Science/water.html>
- [35] Bonnier, F.; Byrne, H. *Analyst* **2012**, 137, 322-332.
- [36] Piqueras S.; Duponchel, D.; Tauler, R.; Juan, A. *Anal Chim Acta* **2011**, 705, 182-192.
- [37] Huang, C.; Ando, M.; Hamaguchi, H.; Shigeto, S. *Anal Chem* **2012**, 84, 5661-5668.
- [38] Hsu, J.; Hsieh, P.; Hsu, H.; Shigeto, S. *Sci Rep* **2015**, 5, 17541.
- [39] Noothalapati, H.; Iwasaki, K.; Yamamoto, T. *Anal Sci* **2017**, 33, 15-22.
- [40] Ali, S.; Bonnier, F.; Tfiayli, A.; Lambkin, H.; Flynn, K. **2012**, *J. Biomed. Opt* 18, 061202.
- [41] Kallaway, C.; Almond, L. M.; Barr, H.; Wood, J.; Hutchings, J.; Kendall, C.; Stone, N. *Photodiagnosis Photodyn Ther* **2013**, 10, 207-219.
- [42] Andronie, L.; Miresan, V.; Coroian, A.; Pop, I.; Raducu, A.; Rotaru, A.; Coroian, C. O. *ProEnvironment* **2015**, 8, 590-600.
- [43] Krafft, C.; Neudert, L.; Simat, T.; Salzer, R. *Spectrochim Acta A* **2005**, 61, 1529-1535.
- [44] Zhu, G.; Zhu, X.; Fan, Q.; Wan, X. *Spectrochimica Acta Part A* **2011**, 78, 1187-1195.
- [45] Notingher I., *Sensors* 2007, 7, 1343-1358.
- [46] Tang, H.; Yao, H.; Wang, G.; Li, Y.; Feng, M. *Optics Express* **2007**, 15, 12708-12716.
- [47] Miljkovic, M.; Chernenko, T.; Romeo, M.; Bird, B.; Matthaus, C.; Diem, M. *Analyst* **2010**, 135, 2002-2013.
- [48] Adar, F.; Erecinska, M.; *Biochemistry* 1978, 17, 5484-5488.
- [49] Matthaus, C.; Chernenko, T.; Newmark, J.; Warner, C.; Diem, M. *Biophys J* **2007**, 93, 668-673.
- [50] Fujita, K.; Smith, I. *Mol cells* 2008, 28, 530-535.
- [51] Okada, M.; Smith, N.; Palonpon, A.; Endo, H.; Kawata, S. *PNAS* **2012**, 109, 28-32.
- [52] Hamada, K.; Fujita, K.; Smith, N.; Kobayashi, M.; Inouye, Y.; Kawat, S. *J Biom Opt* **2008**, 13, 0440271- 0440274
- [53] Chiu, L.; Ando, M.; Hamaguchi, H. *J Raman Spect* **2010**, 41, 2-3.
- [54] Neugebauer, U.; Marz, A.; Henkel, T.; Schmitt, M.; Popp, J. *Anal Bioanal Chem* **2012**, 404, 2819-2829.
- [55] Movasaghi, Z.; Rehman, S.; Rehman, I. *Applied Spectrosc Rev* **2007**, 42, 493-541.
- [56] Wrighta, H. J.; Houb, J.; Xua, B.; Corteza, M.; Potmac, E. O.; Trombergb, B. J.; Razorenova, O. V. *PNAS* **2017**, 114, 6556-6565.
- [57] Jang, Y.; Sharkis, S. *Blood* **2007**, 110, 3056-3063.
- [58] Sciacovelli, M.; Gaude, E.; Hilvo, M.; Frezza, C. *Methods Enzymol* **2014**, 542, 1-23.
- [59] Bonnet, S.; Archer, S.L.; Allalunis-Turner, J.; *Cancer Cell* **2007**, 11, 37-51.
- [60] Sutendra, G.; Michelakakis, E. D. *Front Onc* **2013**, 3, 1-11.
- [61] Ralser, M.; Wamelink, M.M.; Latkolik, S.; Jansen, E.E.; Lehrach, H.; Jakobs, C. *Nat Biotechnol* **2009**, 27,604-605.
- [62] Graves, J.A.; Metukuri, M.; Scott, D.; Rothermund, K.; Prochownik E.V. *J Biol Chem* **2009**, 284: 6520-6529.
- [63] Kennecke, H.; Yerushalmi, R.; Woods, R.; Cheang, M.; Voduc, D.; Speers, C. H.; Nielsen, T. O.; Gelmon, K. *J Clin Onc* **2010**, 28, 3271-7277
- [64] Minn, A. J.; Gupta, G.P.; Siegel, P.M; Bos, P.D.; Shu, W.; Giri, D.; Viale, A.; Olshen, A.B.; Gerald, W.L.; Massagué, J. *Nature* **2005** 436, 518-524.
- [65] Minn, A. J; Gupta, G. P.; Padua, D.; Bos, P.; Nguyen, D.X.; Nuyten, D.; Kreike, B.; Zhang, Y.; Wang, Y.; Ishwaran, H.; Foekens, J. A.; Vijver, M.; Massague, J. *PNAS* **2007**, 104, 6740-5.
- [66] Gupta, G.; Nguyen, D.X.; Chiang, A. C.; Bos, P. D.; Kim, J. Y.; Nadal, C. Gomis, R. R.; Manova-Todorova, K.; Massagué, J. *Nature* **2007**, 446, 765-770.
- [67] Hug, J.; Hanson, J.; Cheang, M. C.; Nielsen, T. O.; Perou, C. M.; Dumontet, C.; Reed, J.; Krajewska, M.; Treilleux, I.; Rupin, M.; Magherini, E.; Mackey, J.; Martin, M.; Vogel, C. *J Clin oncology* **2009**, 27, 1168-1176.
- [68] Martinez-Outschoorn, U. E.; Peiris-Pages, M.; Pestell, R. G.; Sotgia, F.; Lisanti, M.P. *Nat. Rev. Clin. Oncol* **2017**, 14, 11-31.
- [69] Chowdary, M. V. P; Kumar, K. K.; Kurien, J.; Mathew, S.; Krishna, C. M. *Biopolymers* **2006** 83, 556-569.
- [70] Haka, A.; Volynskaya, Z.; Gardecki, J. A.; Nazemi, J.; Lyons, J.; Hicks, D.; Fitzmaurice, M.; Dasari, R. R.; Crowe, J. P.; Feld, M. S. *Cancer Research* **2006**, 66, 3317-3322.

TABLE OF CONTENTS

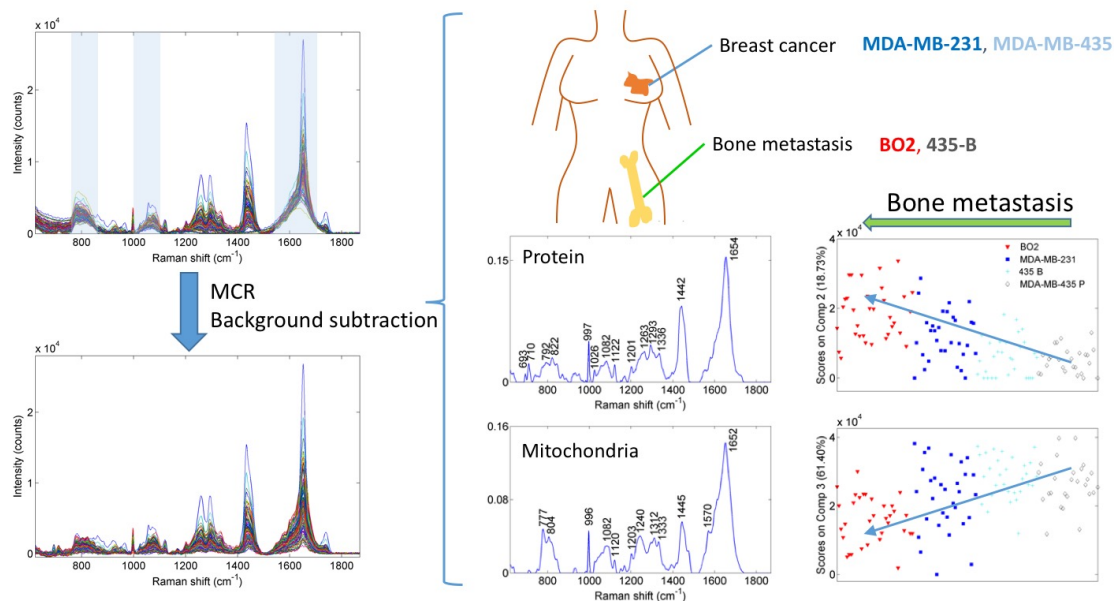


Figure 1

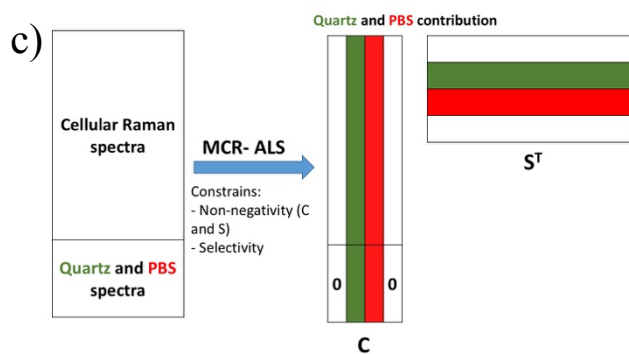
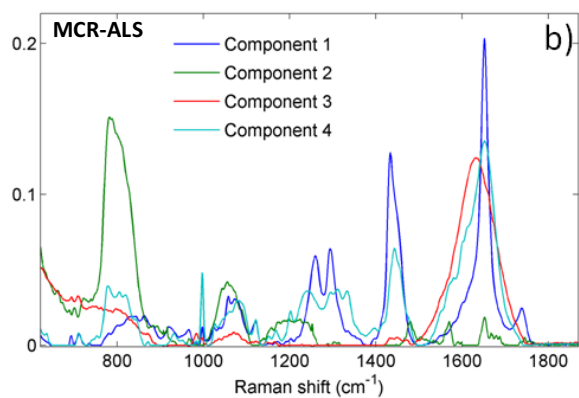
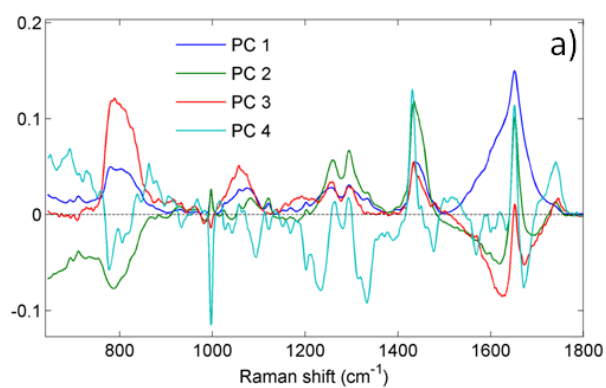


Figure 2

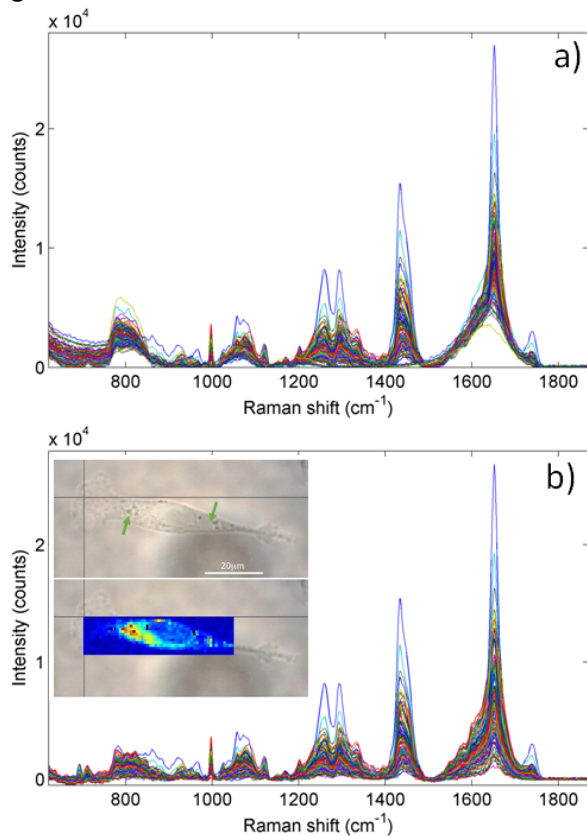


Figure 3

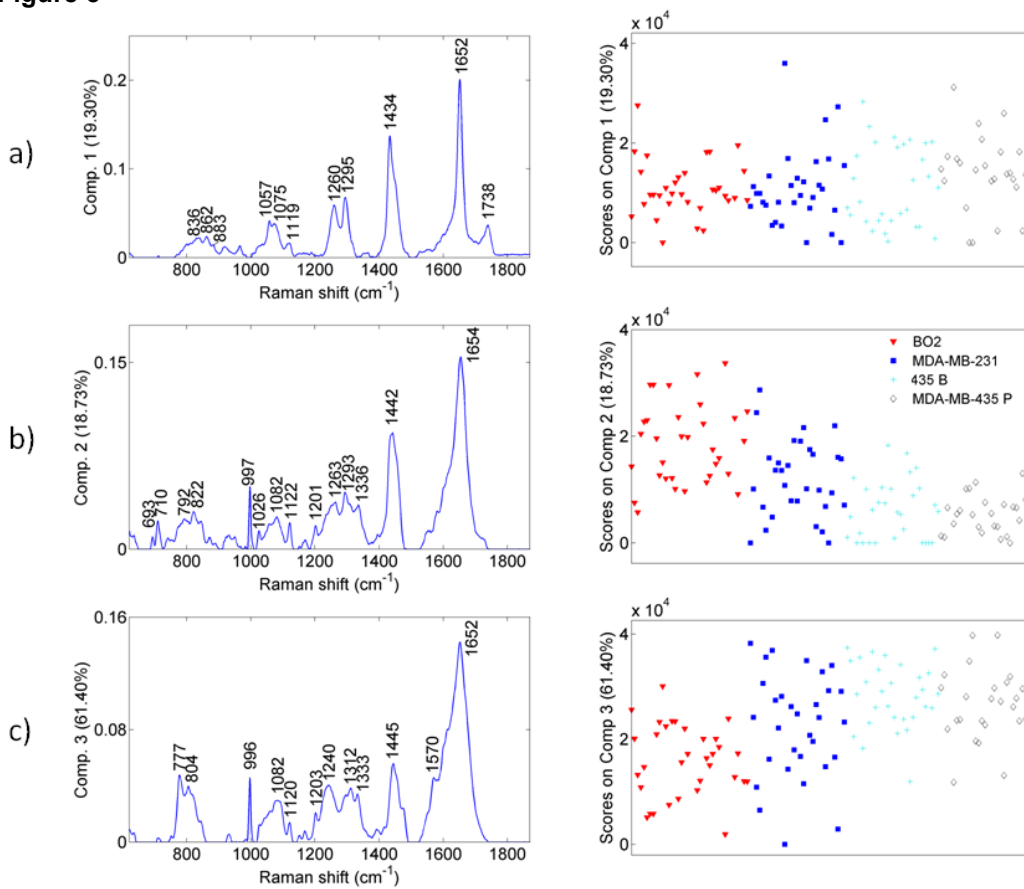


Figure 4

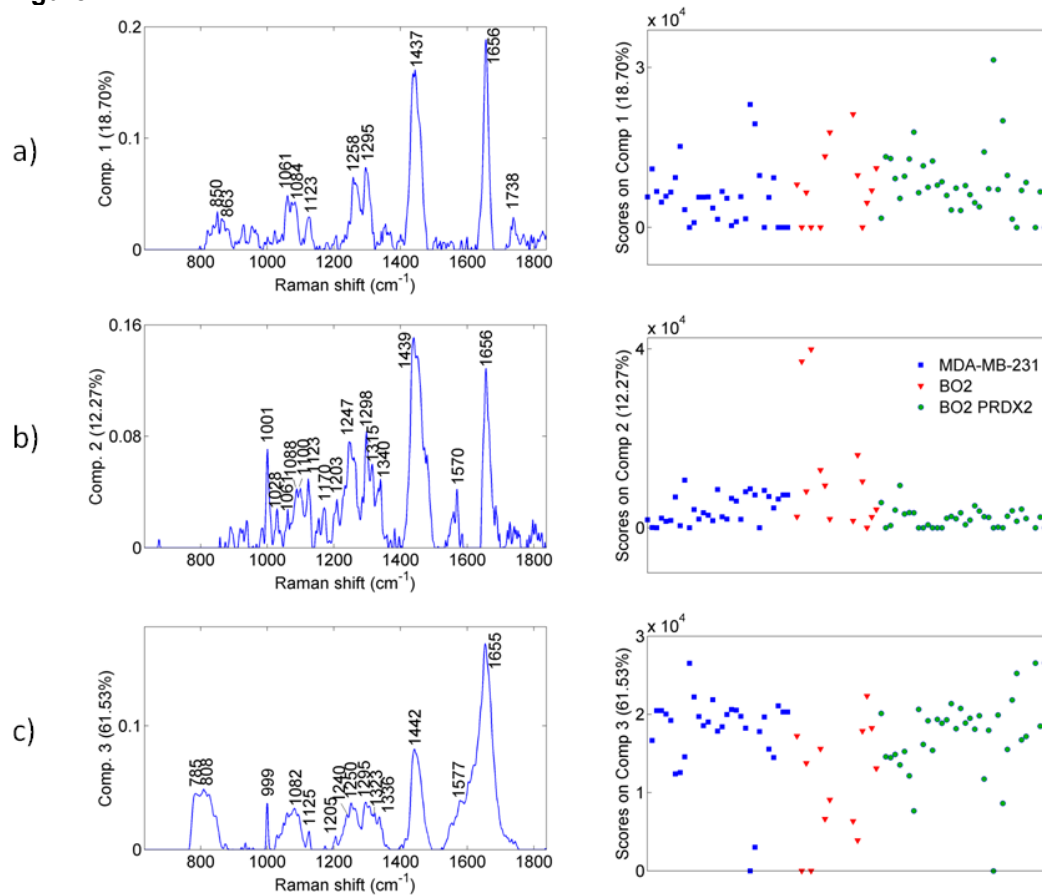
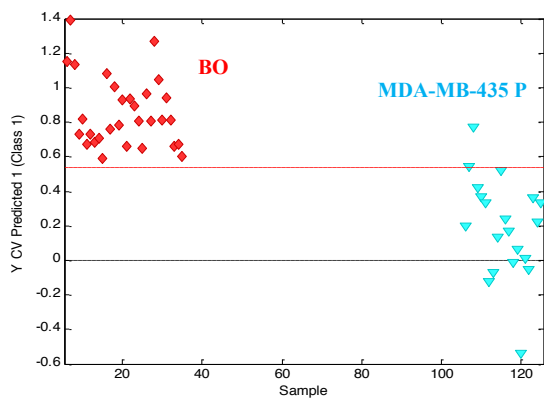
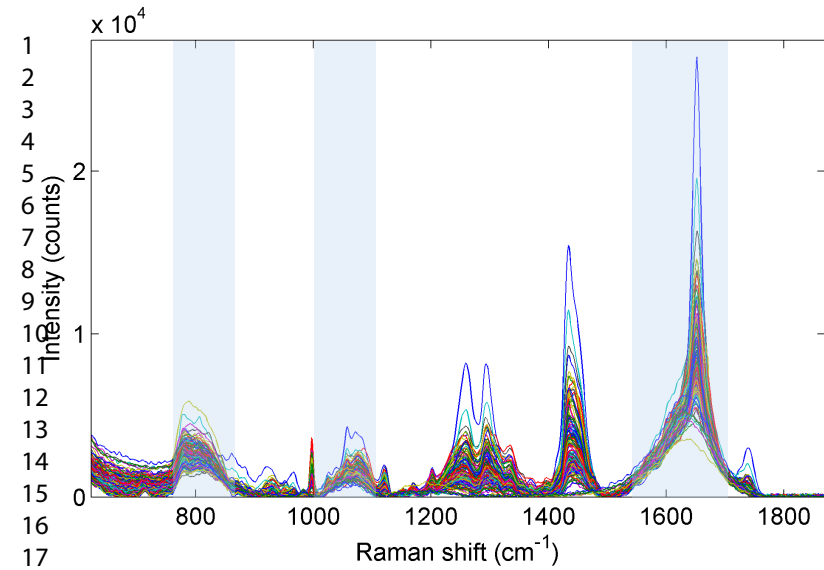


Figure 5





MCR
Background subtraction

