


RESEARCH

Open Access



Characterization of cancer-associated adipocytes by Raman spectroscopy and trajectory inference

Nicolas Goffin¹, Emilie Buache¹, Nathalie Lalun², Marion Fernandes¹, Ines Miguel¹, Catherine Muller³, Charlotte Vaysse^{3,4}, Landry Blanc³, Cyril Gobinet¹ and Olivier Piot^{1*} 

*Correspondence:
olivier.piot@univ-reims.fr

¹ Université de Reims
Champagne-Ardenne, BioSpect
UR 7506, Reims, France

² INSERM, Université de Reims
Champagne-Ardenne, P3Cell,
UMR-S 1250, Reims, France

³ Institut de Pharmacologie et
de Biologie Structurale (IPBS),
Université Toulouse III - Paul
Sabatier (UT3), Toulouse, France

⁴ Département de Chirurgie
Gynécologique oncologique,
CHU-Toulouse, Institut
Universitaire du Cancer de
Toulouse-Oncopole, Toulouse,
France

Abstract

Cancer-associated adipocytes (CAAs) have emerged as pivotal players in various cancers, particularly in such as breast cancer, significantly influencing their progression and therapy resistance. Understanding the adipocytes/cancer cells crosstalk is crucial for effective treatment strategies. Raman spectroscopy, a label-free optical technique, offers potential for characterizing biological samples by providing chemical-specific information. In this study, we used Raman spectroscopy and Trajectory Inference methods, specifically the Partition-based graph abstraction algorithm, to investigate the interactions between 3T3-L1 differentiated adipocytes and MDA-MB-231 breast cancer cells in a 2D co-culture model. We demonstrate the existence of subpopulations of adipocytes and the molecular changes associated with CAAs phenotype. This work contributes to understanding the role of CAAs in breast cancer progression and may guide the development of targeted therapies disrupting this interaction.

Keywords: Cancer-associated adipocytes, Raman spectroscopy, Trajectory inference, Breast cancer

Introduction

Breast cancer is the most diagnosed cancer in women and the second leading cause of cancer-related deaths globally [1]. Obesity not only increases the probability of cancer development but also heightens the risk of cancer recurrence and mortality [2]. The connection between obesity and cancer is multifaceted and complex. Adipocytes, a major component of the tumoral microenvironment (TME) of breast cancers, have drawn considerable attention for their ubiquitous roles in cancer pathophysiology. Cancer-associated adipocytes (CAAs) are adipocytes situated at the tumor invasive front that have undergone phenotypic and functional changes in response to signals from cancer cells, acquiring a pro-tumoral phenotype that supports cancer progression [3]. The CAAs are characterized by the loss of adipocyte markers like PPAR γ , and they have reduced lipid content, giving them a fibroblast-like appearance [3]. CAAs secrete a variety of factors, including cytokines, chemokines, growth factors, and adipokines, which can enhance the proliferation, survival, and migration of breast cancer cells. Additionally, CAAs

undergo metabolic reprogramming, leading to increased lipolysis and fatty acid release. These fatty acids can be used by cancer cells as an energy source, promoting their survival, proliferation, and migration [4]. CAAs also contribute to the remodeling of the extracellular matrix through the production of matrix metalloproteinases and other proteases, which facilitate cancer cell invasion and metastasis [5]. Furthermore, CAAs have been shown to induce a cancer stem cell phenotype in breast cancer cells, which is associated with increased tumor-initiating capacity, resistance to therapy, and metastasis [6]. Finally, CAAs can influence the immune response in the TME, creating an immunosuppressive environment that promotes tumor progression [7].

Despite the growing body of evidence implicating adipocytes in cancer progression, the precise mechanisms underlying CAAs formation remain poorly understood. Targeting the CAAs-cancer cells crosstalk and their metabolic interdependence may provide innovative and more effective treatments aimed at disrupting this interplay.

One promising technology that can address these needs is Raman spectroscopy, a non-invasive, and optical label-free analytical method that has shown great potential for investigating the complex molecular landscape of cancer [8–10]. Its high chemical specificity, coupled with the ability to perform *ex vivo* and *in vivo* measurements without the need for tissue processing, renders Raman spectroscopy an attractive technique for cancer research. Over the past few decades, Raman spectroscopy has been successfully applied to investigate various aspects of cancer, including the identification of tumor subtypes [11], probing the tumor microenvironment [12], and monitoring treatment response [13, 14].

In this study, we applied Raman spectroscopy to a 2D co-culture model of 3T3-L1 adipocytes and MDA-MB-231 breast cancer cells. The 3T3-L1 cell line, a murine pre-adipocyte cell line widely used in adipogenesis studies [15–17], can differentiate into mature adipocytes upon exposure to appropriate stimuli [18]. The MDA-MB-231 cell line, a triple-negative human breast cancer cell line, has been extensively studied for its aggressive and invasive characteristics [19]. This 2D co-culture model facilitated the formation of a contact zone between adipocytes and cancer cells, allowing us to create a controlled system to study their interactions, mimicking their proximity in the TME.

However, the cellular variability poses a significant obstacle for studying adipocytes/cancer cells interaction [20–22]. In the co-culture model used in this work, the differentiation process itself contributes to the heterogeneity in 3T3-L1 adipocyte cultures, as not all cells within the population undergo adipogenesis uniformly or synchronously [23, 24]. The resulting adipocytes may display a broad range of sizes, lipid droplets (LDs) number and sizes, and morphology. Moreover, the adipocyte delipidation, induced by breast cancer cells, is also a heterogeneous process [25, 26], depending on the distance between the two cell types and possibly other factors. These various sources of heterogeneity confound and complicate the data processing. Thus, a comprehensive understanding of this heterogeneity necessitates adapted data processing tools.

Recent advancements in single-cell transcriptomics (scRNAseq) have led to the development of specialized tools, known as Trajectory Inference (TI) methods [27], to reconstruct cellular trajectories [28] and identify subpopulations [29] by analyzing snapshots of cellular variability. These methods leverage the underlying principle that biological processes occur primarily asynchronously within cells. From transcriptome

measurements conducted on heterogeneous cellular samples, these tools can digitally reorder cellular states by analyzing the global topology of the multivariate dataset.

Trajectory Inference (TI) methods consist of three key components: dimensionality reduction, trajectory construction, and pseudotime ordering [30]. Dimensionality reduction simplifies complex data by reducing dimensions while preserving essential information, making it easier to visualize and interpret. Linear methods like Principal Component Analysis (PCA) are commonly used, but for more complex, non-linear relationships, manifold learning approaches such as UMAP (Uniform Manifold Approximation and Projection) [31] and t-SNE (t-Distributed Stochastic Neighbor Embedding) [32] are preferred [30, 33]. UMAP is widely used in scRNA-seq and was proven to be particularly effective for exploring complex transcriptomic datasets by revealing patterns such as distinct cell populations [34]. It constructs a high-dimensional graph of data points, with edges reflecting connection likelihood based on proximity, and optimizes this graph into a low-dimensional space that preserves both global and local data relationships [31]. UMAP's effectiveness in visualizing and interpreting complex biological data is demonstrated across various fields, including spectroscopy, where it has been used to analyze Raman microspectroscopy data, distinguish between tissue types and disease states [35], study bacterial antibiotic resistance [36], or extracellular vesicles [37].

Trajectory construction involves mapping the progression of cellular states by connecting data points to reflect their evolutionary paths [38]. Different methods can be employed to achieve this; each with its own strengths suited to specific aspects of cellular transitions. These include principal curve methods for modelling smooth trajectories in high-dimensional spaces [39], clustering-based approaches for connecting grouped cells, graph-based techniques [40], and probabilistic models that account for the stochastic nature of state transitions [41].

Pseudotime algorithms estimate the temporal order of cell states from single-cell data, creating a pseudotime that reflects the progression of cells along the trajectories. These algorithms employ different methods to calculate pseudotime. Some, like TSCAN and Waterfall [42, 43], use the minimum spanning tree algorithm to project cells along edges, while others, such as Slingshot [39], use smooth curves. Diffusion Pseudotime (DPT) calculates pseudotime using diffusion maps and random-walk-based distances from a selected root, typically representing the earliest step of the biological process [44].

We have previously demonstrated the possibility of employing these methods on Raman data for unraveling dynamic biological processes [45]. We specifically employed the partition-based graph abstraction (PAGA) algorithm [40] to construct spectral trajectories, both on an artificial dataset and on Raman spectra of lipid droplets (LD) acquired during adipocyte differentiation. In the present study, our objectives are to (1) highlight subpopulations of adipocytes based on their Raman signature, (2) identify the molecular changes associated with CAAs formation, and (3) show the possibility of understanding the potential interactions between cancer cells and adipocytes.

Methods

Cell culture

The triple negative (ER⁻/PR⁻/HER2⁻) human breast carcinoma cell line MDA-MB-231 (HTB-26) was obtained from the American Type Culture Collection (ATCC). The

murine pre-adipocyte cell line 3T3-L1 was kindly provided by MC. Rio (IGBMC, Illkirch, France). MDA-MB-231 and 3T3-L1 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 4.5 g/L glucose, 10% fetal calf serum (FCS) (Dutscher, S1810-500) and 1% penicillin-streptomycin (Invitrogen, 15140). All cell cultures were maintained at 37 °C in 5% CO₂ (v/v). Cells were routinely passaged at pre-confluency using 0.05% trypsin, 0.53 mM EDTA (Invitrogen, 25300) and screened for the absence of mycoplasma using PCR methods.

In vitro differentiation of 3T3-L1 preadipocytes

In brief, 2-day post confluent cells (designated day 0) were treated with DMEM (4.5 g/L glucose), 10% FCS, 1% penicillin-streptomycin, 0.5 mmol/L 3-isobutyl-1-methylxanthine (IBMX), 0.5 µmol/L dexamethasone, and 10 µg/ml insulin (all Sigma-Aldrich) to induce differentiation. After 2 days, medium was replaced by DMEM (4.5 g/L glucose), 10% FCS, 1% penicillin-streptomycin, and 10 µg/ml insulin, and further exchanged every 2 days. From day 12 after initiating differentiation, 3T3-L1 adipocytes were co-cultured with MDA-MB-231 human breast cancer cells.

Co-culture model

For 2D direct co-culture between murine 3T3-L1 adipocytes and MDA-MB-231 human breast cancer cells, the CytoSelect™ 24-well Cell Co-culture System (Cell Biolabs, San Diego, CA, USA) was used following the manufacturer's protocol. 3T3-L1 cells were seeded on a 0.16 mm-thick glass slide, around the insert of the co-culture system (70,000 cells per well), and the first step was dedicated to adipocyte differentiation. After 12 days, the insert is removed, resulting in a well-defined cell-free zone (with a diameter of 8 mm) and allowing the seeding of MDA-MB-231 cells (100,000 cells per well). The culture medium remains a mixture of DMEM (4.5 g/L glucose), 10% FCS, and 1% penicillin-streptomycin, to which 10 µg/mL insulin is added to maintain adipocyte differentiation. Once the invasion front is formed (24 hours after breast cancer cells seeding), the co-culture is maintained for 3 days. The adipocytes distant from the invasive front are considered as control adipocytes, and those present at the invasive front are referred to as CAAs.

Oil red O staining

To assess the adipocyte phenotype, the cell cultures were fixed in 4% paraformaldehyde (Electron Microscopy Sciences) for 15 minutes at room temperature, followed by three rinses with distilled water. The fixed cells were then stained with 0.5% Oil-Red-O (Sigma-Aldrich) for 20 minutes, washed with distilled water, and counterstained with hematoxylin (Sigma-Aldrich).

Raman spectroscopy

For Raman analysis, cells were fixed with 4% paraformaldehyde and left in the water during acquisitions. Raman measurements were carried out at room temperature using a HORIBA Xplora spectrometer equipped with a 532 nm laser diode for excitation. The spectral range of 550–3100 cm⁻¹ was covered with a resolution of approximately 3 cm⁻¹, employing a 900 gr/mm grating. The slit size was set at 200 µm, and the pinhole

diameter was 500 μm . A long working distance $\times 100$ water immersion objective (Olympus LUMPlanFI) on the attached Olympus BX51 microscope provided a lateral resolution of around 0.5 μm and a depth of field of about 3 μm . The laser power at the sample was adjusted to 3 mW. Spectral images were collected with a spatial step of 1 μm in X and Y directions. Each spectrum was obtained by recording two accumulations with an automatic exposure time to achieve 20,000 counts. This acquisition mode effectively prevented signal saturation for lipids while ensuring adequate signal intensity for proteins with longer acquisition times. Prior to data collection, the system was calibrated and aligned using silicon and neon references. The cell images are obtained by manually delineating the adipocytes. A total of 79,552 spectra were acquired from 51 control adipocytes and 102 CAAs.

Raman data pre-processing

The Raman dataset was normalized by the Extended Multiplicative Signal Correction (EMSC) method integrating neutralization of the water and glass spectral interferences [46]. The data pre-processing was carried out under Orange software (v.3.29.3) using the spectroscopy package (v.0.6).

Dimensionality reduction

The analysis was performed using the Scanpy [47] package (v1.7.2), a Python package designed specifically for single-cell RNA sequencing (scRNA-seq) datasets. It is also available within the Dynverse [27] package, an R-based package. Detailed information about the trajectory inference analysis can be found in reference publications [30, 48].

To begin the analysis, the Raman data matrix was transformed into an “Anndata” file format (<https://github.com/theislab/anndata>), which served as the input for subsequent analysis. First, a dimensionality reduction was applied using the UMAP algorithm. UMAP works through two distinct steps. Initially, a KNN graph is created based on proximity between spectra, typically measured using Euclidean distance. In this graph, each spectrum is represented by a node connected by edges to a neighborhood of its most similar spectra. Instead of having binary connections (either connected or not), UMAP assigns probabilities to the edges based on the distances between points. These probabilities represent the likelihood of connection or relationship between the points. This approach enables UMAP to capture complex and non-linear relationships in the data. In the Scanpy package, this step was performed using the `scanpy.api.pp.neighbors` function where the “`n_neighbors`” parameter is introduced. This parameter represents the number of neighboring data points considered when constructing the neighborhood graph. It controls the balance between preserving local and global structures (higher values capturing more global information and lower values focusing on local details). Here, a default value of 25 nearest neighbors is used.

Secondly, UMAP optimizes the low-dimensional embedding by minimizing cross-entropy between the distribution of distances between datapoints in the original high-dimensional and low-dimensional spaces (typically two dimensions). Stochastic gradient descent is used to iteratively adjust the positions of data points in the low-dimensional space, preserving topological relationships from the high-dimensional data. In Scanpy, this optimization step was done with the `scanpy.tl.umap` function.

Construction of trajectories with PAGA

PAGA constructs a statistical model to capture the connectivity between partitions of kNN graphs [40]. The kNN graph created previously using UMAP can be reused for this purpose. The partitions, representing distinct groups of spectra, are generated in an unsupervised manner using clustering tools. Here, K-Means was performed directly on UMAP reduced data. The `sc.tl.paga` function uses the PAGA tool to evaluate the level of connectivity among clusters (nodes) in the K-nearest neighbor (KNN) graph. We obtain a connectivity matrix that represents all the links existing between different clusters. We used the default connectivity cutoff of 0.1 to eliminate weaker connections between clusters. The final PAGA paths are obtained by applying a Minimum Spanning Tree (MST) to this connectivity matrix. This process simplifies connections between nodes minimizing the cumulative weight of the edges.

Segmentation of spectral images

The segmentation was performed by manually outlining the cells using the 'hyper-spectral' widget present in the Orange Spectroscopy package (Fig. S1).

Spatial organisation of the clustered images

Spatial analysis was conducted using the Squidpy package (v1.3.0). This approach, described by Schapiro et al. [49], allows for the assessment of spatial proximity between spectral clusters. In the first step, we used pixel coordinates of clustered spectral images to compute a graph of spatial neighbors. Here, we created the graph by selecting the 10 closest neighbors to each pixel. This number is intentionally small to consider only one cell at a time. This step is performed by applying the function `sq.gr.spatial_neighbors` on the `adata` file.

The next step involves estimating the spatial proximity between each possible couple of clusters. Firstly, the number of connections between a pair of clusters i and j is calculated and denoted x_{ij} . In parallel, a similar number is calculated from a random configuration of the two clusters, and repeated 1000 times. The average number of connections is given by μ_{ij} , and its standard deviations by σ_{ij} . Finally, for each pair of clusters i and j , we calculate a z-score as follows:

$$Z_{ij} = \frac{x_{ij} - \mu_{ij}}{\sigma_{ij}} \quad (1)$$

The z-score quantifies the spatial proximity between two clusters. A hierarchically clustered heatmap is then used to visualize the spatial organisation of the clustered images.

Pseudotime ordering

Pseudotime analysis allows ordering individual spectra along PAGA paths based on their spectral similarities. The pseudotime calculation was performed using the diffusion pseudotime algorithm [44], available in the Scanpy package (`scanpy.tl.dpt`). Pseudotime calculation involves selecting a starting point, typically a data point that represents the beginning of the studied process. Once the starting point is selected, it

is then possible to calculate the pseudotime value for each point, which determines its position along the trajectory from the starting point. Pseudotime provides directionality to the trajectories.

Identification and visualization of evolving spectral features

Highlighting of biochemical changes along the trajectories involves identifying spectral variables using a feature selection method such as Random Forest, which is a multivariate approach available in the *Dynverse* package. To explore how Raman features of interest change over pseudotime, a heatmap visualization is employed. This heatmap displays the evolution of the selected variables as a function of pseudotime. By observing the heatmap, patterns and trends in the data can be identified, highlighting how the variables change along the trajectories. Additionally, a dendrogram can be constructed to group the spectral variables based on their evolution profiles. This dendrogram provides a hierarchical clustering of the variables, allowing for the identification of related variables with similar patterns of change over pseudotime.

Assessment of spectra dispersion on the UMAP embedding

The raw UMAP representations (Fig. 2a) do not permit to see if the data distribution is homogeneous or not. To assess more quantitatively the spectra distribution on the UMAP representation, we utilize a Gaussian kernel density estimation. This is achieved by employing the `sc.tl.embedding_density` function in *Scanpy*. To compare control adipocytes and CAAs, the densities were calculated for each condition separately.

Results

Highlighting of adipocyte subpopulations using raman spectroscopy and PAGA

Figure 1 illustrates the methodology employed in this study to characterize the cellular co-culture model using Raman spectroscopy and PAGA. The top left panel of Fig. 1 presents the co-culture model used in the experiment. The 2D co-culture system offers a framework for observing direct interactions between two distinct types of cells within a single well. This approach permits modeling an invasive front of mammary cancer cells in proximity to CAAs that exhibit known morphological characteristics. Indeed, they have a typically smaller and more elongated form, contrasting markedly with the unaffected adipocytes found at the glass slide periphery, which are subsequently considered as control adipocytes. Figure S2 illustrates the distinct morphological characteristics typical of both control adipocytes and CAA phenotypes.

Raman images were obtained from control adipocytes and CAAs (top right panel in Fig. 1). Upon normalization by EMSC, the Raman spectra follow the process outlined in Fig. 1, bottom panel. After dimensionality reduction via Uniform Manifold Approximation and Projection (UMAP) and segmentation by K-means, the Partition-based Graph Abstraction (PAGA) process achieves the construction of a simplified representation of the dataset. Subsequently, each point is organized in accordance with the computation of pseudotime, indicating the relative distance of each point from a chosen point defined as the process origin. In the final stage, the points are arranged along trajectories, which enables a continuous depiction of the spectral feature progression along various paths.

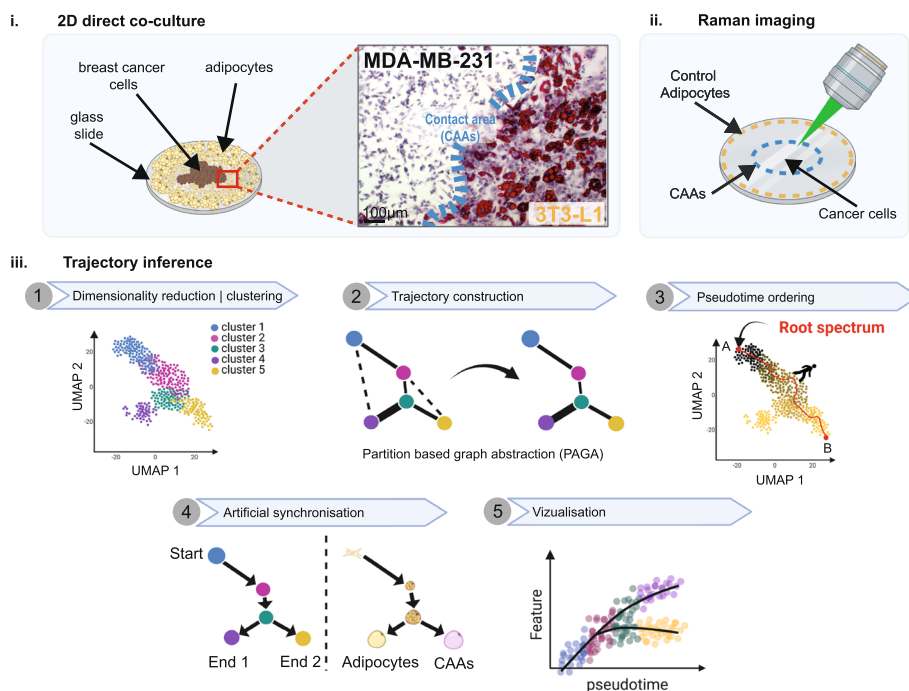


Fig. 1 Characterization of Adipocyte-Cancer Cell interactions using Raman spectroscopy and trajectory inference. (i) The 2D co-culture model enables the formation of an interaction zone between murine 3T3-L1 adipocytes and MDA-MB-231 human breast cancer cells. The visible image, stained with Oil Red O, provides a close-up of the interaction zone. Adipocytes located at the periphery of the slide, distant from the contact area, are designated as control adipocytes, while those within the contact area are classified as cancer-associated adipocytes (CAAs). (ii) Raman imaging differentiates between control adipocytes and CAAs by capturing high-resolution spectral data, revealing distinct biochemical profiles. (iii) Trajectory inference from Raman data: Dimensionality reduction using UMAP and clustering with K-means identified distinct spectral clusters. The Partition-based Graph Abstraction (PAGA) constructed a graph that reflects the connectivity between clusters, preserving their topological structure. Pseudotime analysis aligned the spectra along inferred trajectories, revealing molecular transitions linked to both intra- and intercellular heterogeneity, as well as the formation of cancer-associated adipocytes (CAAs). This method captures spectral heterogeneity and dynamic changes, providing valuable insights into the cellular interactions within the co-culture system

Figure 2a shows the two-dimensional space after dimensionality reduction using UMAP where points are color-coded based on the culture conditions. Orange datapoints represent control adipocytes, while blue datapoints denote CAAs. As demonstrated in Fig. 2a, both conditions globally overlap within the UMAP representation, indicating that no region appears to be specific to one condition over another. The same observations were obtained using other more conventional processing methods, such as PCA or t-SNE (Fig. S3).

Figure 2b shows the UMAP embedding segmented by K-means clustering ($k=20$). The resulting mean cluster spectra are presented in Fig. S4. Based on the clustering results, PAGA provides a weighted graph representation, as illustrated in Fig. 2c. This representation is subsequently simplified by applying the MST algorithm (Fig. 2d), a technique used in graph theory, that aims to interconnect all nodes within a graph in a manner that minimizes the cumulative weight of the edges. The data points are then color-coded according to the paths they are assigned. Finally, high-confidence paths in the PAGA graph correspond to 6 terminal states and 4 branching points, which can be highlighted.

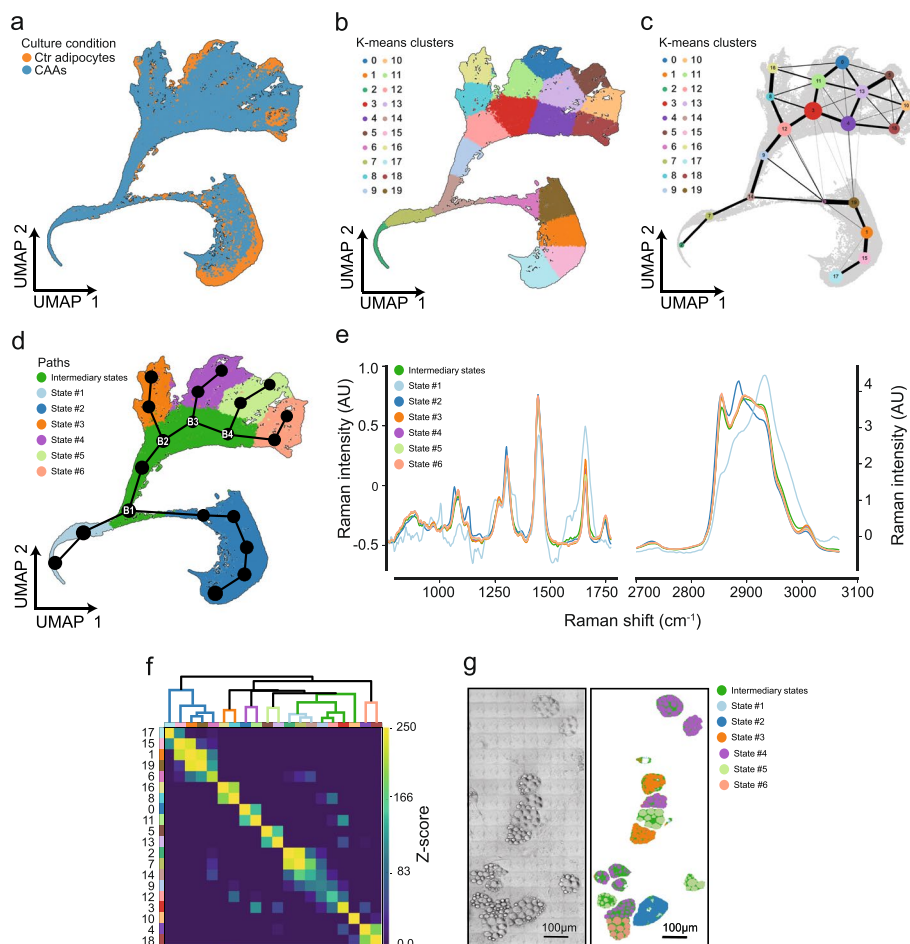


Fig. 2 Topological trajectory construction using PAGA analysis. **a** Spectral dataset after dimensionality reduction with UMAP. Data points are color-coded based on the adipocyte status: control adipocytes in orange and CAAs in blue. **b** Clustering results using K-means ($k=20$), visualized on a 2D UMAP reduced space, with spectra color-coded according to their assigned cluster. **c** PAGA graph overlaying the UMAP visualization. Within this representation, nodes correspond to clusters, while the edges depict inter-cluster connections. Node size reflects the number of members in a cluster, and edge thickness indicates the degree of connectivity, with thicker edges representing stronger connections. **d** The weighted PAGA graph is denoised through the application of a minimum spanning tree. Spectra are color-coded based on their assigned paths. **e** Averaged spectra of the various paths emphasized by PAGA. **f** Heatmap illustrating the spatial organization of clusters. High Z-scores indicate close spatial relationships between two clusters, while low scores suggest spatial separation between them. By applying the trajectory color-code to branches, hierarchical clustering reveals spatial relationships among clusters within the same trajectory. **g** Overlay of visible adipocyte images with the color-coded representation, determined by pixels assignment to the associated pathways emphasized by PAGA, this overlay reveals the presence of distinct adipocyte subpopulations

Averaged spectra representing the spectral variability associated with the different paths identified by PAGA are displayed in Fig. 2e. These spectra present a lipidic profile except for State #1, which displays characteristic protein vibrations.

To discern the biological meaning associated with the trajectories emphasized by PAGA, an investigation of the spatial organization of clusters on spectral images was done (Fig. 2f). Here we visualize pairwise cluster neighborhood scores ordered by hierarchical clustering of the Z-scores. For instance, we observe a strong spatial neighboring

between clusters #17, #15, #1, #19, and #6. Interestingly, these clusters all reside within a single trajectory, corresponding to State #2 (blue branch of the dendrogram at the top of the heatmap). Each dendrogram is color-coded based on the trajectory to which the clusters belong, suggesting a spatial coexistence of points within a trajectory.

Figure 2g shows a side-by-side display of a visible image of adipocytes and a reconstructed Raman image, with color coding based on pixel assignment to pathways identified by PAGA. This representation confirms that spectra of a cell correspond almost exclusively to one single trajectory. This observation indicates, therefore, that the trajectories represent distinct cellular states. Moreover, Fig. S5 confirms this result; indeed, when the spectra of an image of a unique cell are projected onto the UMAP representation, the points are not scattered randomly but follow a single trajectory. Overall, using Raman spectroscopy associated with PAGA, we identified, from the spectral heterogeneity captured in the cell Raman images, distinct subpopulations of adipocytes, CAAs, and control adipocytes indiscriminately.

Biochemical bases of the cellular state distinction

The tree-like structure constructed previously does not present any ordering of the data distribution. However, by ordering individual spectra along trajectories, pseudotime analysis allows the reconstruction of intricate spectral variations, showing the evolution of Raman intensities along the different trajectories as a function of pseudotime value. Pseudotime calculation involves selecting a starting point, typically a spectrum that signifies the beginning of the biological process under study. In our case, a marker often used to track adipocyte differentiation is the increase in lipid content. Indeed, during the initial phase of 3T3-L1 differentiation, cells exhibit a fibroblast-like phenotype with small LDs; thus, the protein signal appears favored over the lipid signal. Conversely, the final stage of differentiation is characterized by adipocytes with large LDs, which result in spectra dominated by lipidic contributions. Based on these considerations, we selected the root as the spectrum exhibiting maximum protein signals, specifically in the C-H stretching mode region of proteins at 2930 cm^{-1} . The root is represented by a red dot on Fig. 3a. It is located on the branch “State #1” (Fig. 2e). Once the starting point is selected, it is then possible to calculate the pseudotime value for each point, which allows computing how far it is along the trajectory from the starting point. Figure 3a shows the projection of pseudotime value on the UMAP embedding. From Fig. 2d and the pseudotime values, it is then possible to construct a directed graph, leading from the starting point to 5 cellular states (Fig. 3b). Figure 3c illustrates the pseudotime values according to each cellular state, with states #5 and #6 presenting the most advanced values.

For a global visualization of the evolving Raman features involved in the distinctive states, we constructed a heatmap (Fig. 3d) depicting the intensities of the most significant wavenumbers against the pseudotime. These Raman features were identified by employing the Random Forest algorithm as a supervised variable selection method. Certain vibrations correspond to lipid signals such as C=C stretching ($1660\text{--}1665\text{ cm}^{-1}$) indicative of lipid unsaturation, symmetric CH_2 stretching (2848 cm^{-1}) linked to carbon chain length, and antisymmetric CH_2 stretching (2880 cm^{-1}) vibrations associated with the acyl chain order, which are indicators of the trans conformation within the hydrocarbon chain. Protein signals are also noteworthy, such as the aromatic and aliphatic ν

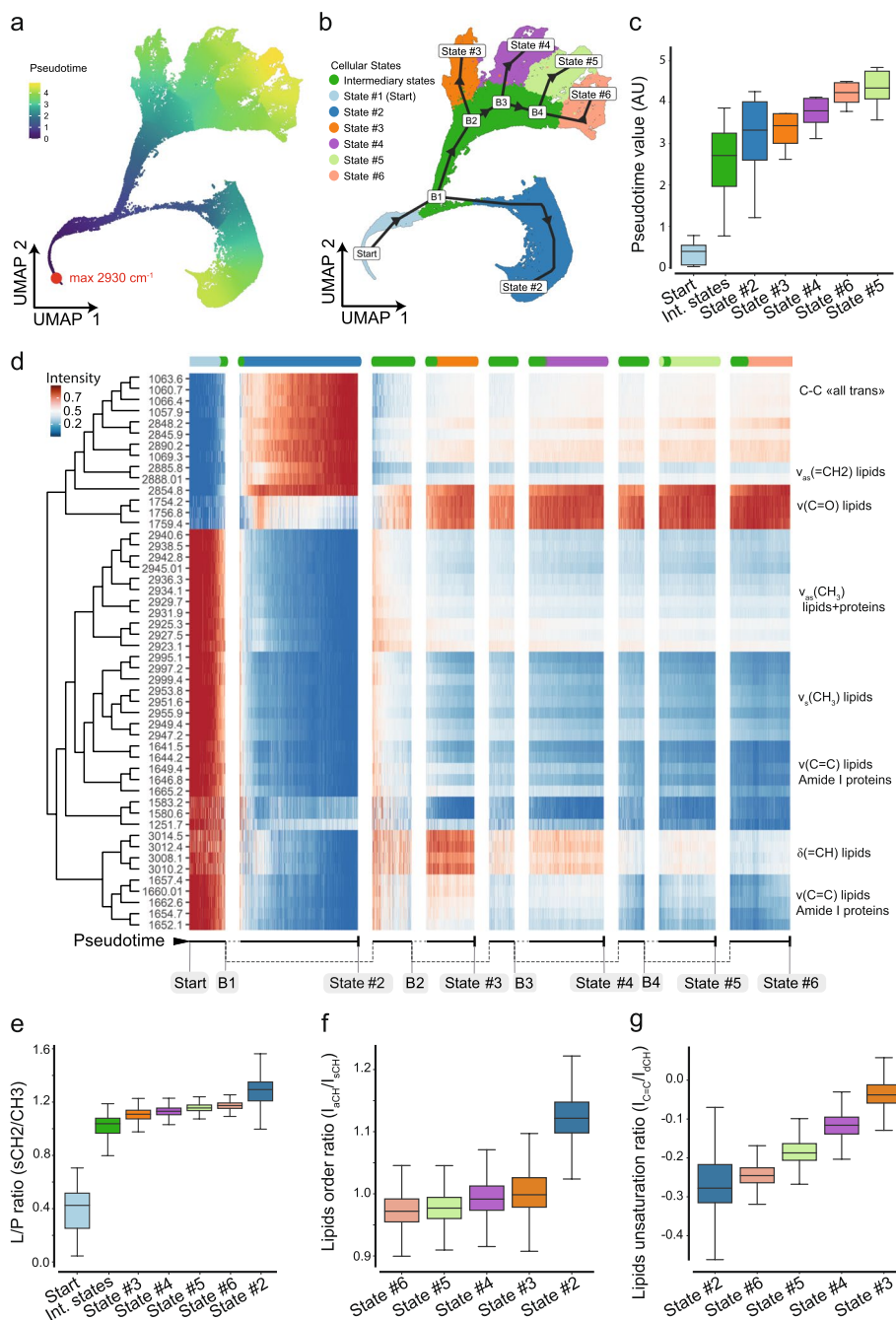


Fig. 3 Ordering of spectral data thanks to pseudotime and identification of evolving features along the different paths. **a** UMAP representation where points are color-coded according to their pseudotime values, representing the distance from the root. Early-stage spectra are depicted in blue, while later-stage spectra are shown in yellow. The red point corresponds to the root spectrum, characterized by the strongest protein signal, specifically in the C-H stretching mode region of proteins at 2930 cm^{-1} . **b** Directed graph constructed based on the evolution of pseudotime values, representing the progression from the initial point to five distinct cellular states. **c** Heatmap representation showing intensity variations of the 50 most informative Raman wavenumbers determined by Random Forest feature selection. **d** Quantification of pseudotime value as a function of cellular states. **e** Quantification of the ratio between lipids and proteins signal based on the intensity ratio between the bands at 2852 and 2934 cm^{-1} corresponding to symmetric CH_2 and CH_3 vibrations. **f** Quantification of lipid ordering based on the intensity ratio between the bands at 2880 and 2852 cm^{-1} corresponding to antisymmetric and symmetric CH_2 vibrations, respectively. **g** Quantification of the unsaturation degree based on the intensity ratio between the bands at 1656 and 1442 cm^{-1} corresponding to $\text{C}=\text{C}$ stretching mode and CH_2 scissoring, respectively

(CH) vibration around 2930 cm^{-1} , or the Amide I band (1650 cm^{-1}). This analysis allows unmasking the pivotal molecular vibrations at play in the biochemical heterogeneity of the biological process.

In addition, a Raman ratiometric analysis also offers interpretable information within a biological context. Figure 3e shows the measurement of the ratio between lipid and protein signals based on the intensity ratio between the bands at 2852 and 2934 cm^{-1} corresponding to symmetric CH_2 and CH_3 stretching vibrations. The starting point is rationally marked by a low lipid/protein ratio. The evolution is rapid from the start to the intermediate states, and then it grows slowly, indicating a minimal change in the lipid/protein ratio until the terminal states. Figure 3f shows the quantification of lipid ordering (only for final states rich in lipids) based on the intensity ratio between the bands 2880 and 2852 cm^{-1} corresponding to antisymmetric and symmetric CH_2 vibrations, respectively. These results show that State #2 corresponds to lipids that have undergone a phase change. Finally, Fig. 3g shows the quantification of the unsaturation degree based on the intensity ratio between the bands at 1656 and 1442 cm^{-1} corresponding to $\text{C}=\text{C}$ stretching and CH_2 scissoring modes, respectively. This ratio is a current spectral marker to characterize the lipid content.[50] These results demonstrate that, firstly, State #2 has the lowest degree of unsaturation, while State #3 has the highest.

After this characterization of the spectral features associated with the biochemical heterogeneity observed within both control adipocytes and CAAs, we will investigate the spectral differences between these two phenotypes.

Highlighting of differences between control adipocytes and CAAs

We have demonstrated in Fig. 2a that the two cell populations overlap on the UMAP plots, indicating that the global spectral variability is equivalent in both populations. Nevertheless, the dispersion of points within the UMAP embedding could potentially vary between the two populations. By employing Gaussian kernel density estimation, we can compute for each population the corresponding spectral density on the UMAP embeddings, as illustrated in Fig. 4a. The results clearly show that spectral distribution differs between the two conditions. Specifically, it's noteworthy to see that in the case of control adipocytes, there's a higher density found at the trajectory endpoints, at the expense of the beginning and intermediate area of trajectories. Conversely, CAAs demonstrate a high degree of density at the beginning and middle stages. These results can further be confirmed by calculating the median pseudotime for each condition. Indeed, as shown in Fig. 4b, CAAs possess a significantly reduced pseudotime in comparison to control adipocytes. This observation indicates that CAAs follow shorter spectral trajectories. Furthermore, Fig. 4c shows that the CAAs are smaller than the control adipocytes, based on the number of pixels that constitute a cell. In parallel, in Fig. 4d, we also quantified the respective proportions of spectra in the different cellular states. Firstly, the findings verify the previous results by showing that states associated with the initial and intermediate trajectory states are more commonly found in CAAs. Secondly, CAAs exhibit an increased population of State #3, which represents the more unsaturated cellular state. This contrasts with the decreasing populations of States #4, #5, and #6, which correspond to more saturated states. Lastly, a significant reduction is noticed in the population of State #2, which is about three times lower in CAAs.

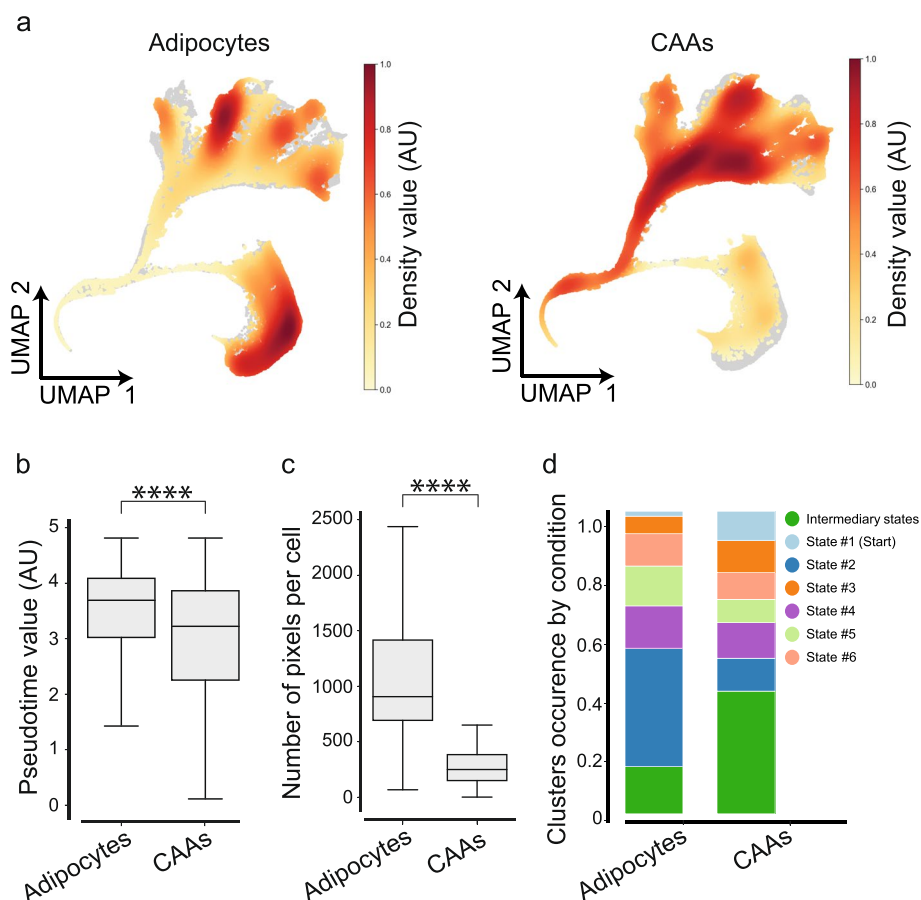


Fig. 4 Analyses of spectral differences between control adipocytes and CAAs. **a** UMAP embedding with points colored according to the Gaussian kernel density estimation of spectra in the reduced space. This measurement was conducted for control adipocytes and CAAs separately, the grey color represents the outline of the data distribution for the other condition. **b** Quantitative analysis of pseudotime value. **c** Quantitative analysis of the number of pixels composing a cell. **d** Relative occurrence of the different cellular states for each condition. For **b** and **c**, p -values were determined by two-tailed student's t -test. **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$

Comparing the breast cancer cells at the invasive front or far from CAAs

It is of interest to also examine cancer cells to elucidate the potential crosstalk and interaction between cancer cells and adipocytes in the co-culture system. In this case, we compared spectra from the lipid droplets of MDA-MB-231 cells located at the contact zone and those located in the central region of the well, which can be considered as controls (Fig. 5a). These spectra were projected onto the UMAP embedding constructed from adipocytes spectra. First, Fig. 5b reveals that spectra from MDA-MB-231 cells totally overlap with those from adipocytes, highlighting spectral similarities in the LDs of both cell types. Similarly, to previous analyses, we conducted a quantitative examination of the pseudotime values between the two tumoral cell conditions. Figure 5c demonstrates that the median pseudotime is higher for MDA-MB-231 cells located at the contact zone. Additionally, we carried out a comparison of the Lipids/Proteins ratios as depicted in Fig. 5d. The results show that the MDA-MB-231 cells situated at the contact zone contain a higher lipid content, which is coherent with the increased pseudotime

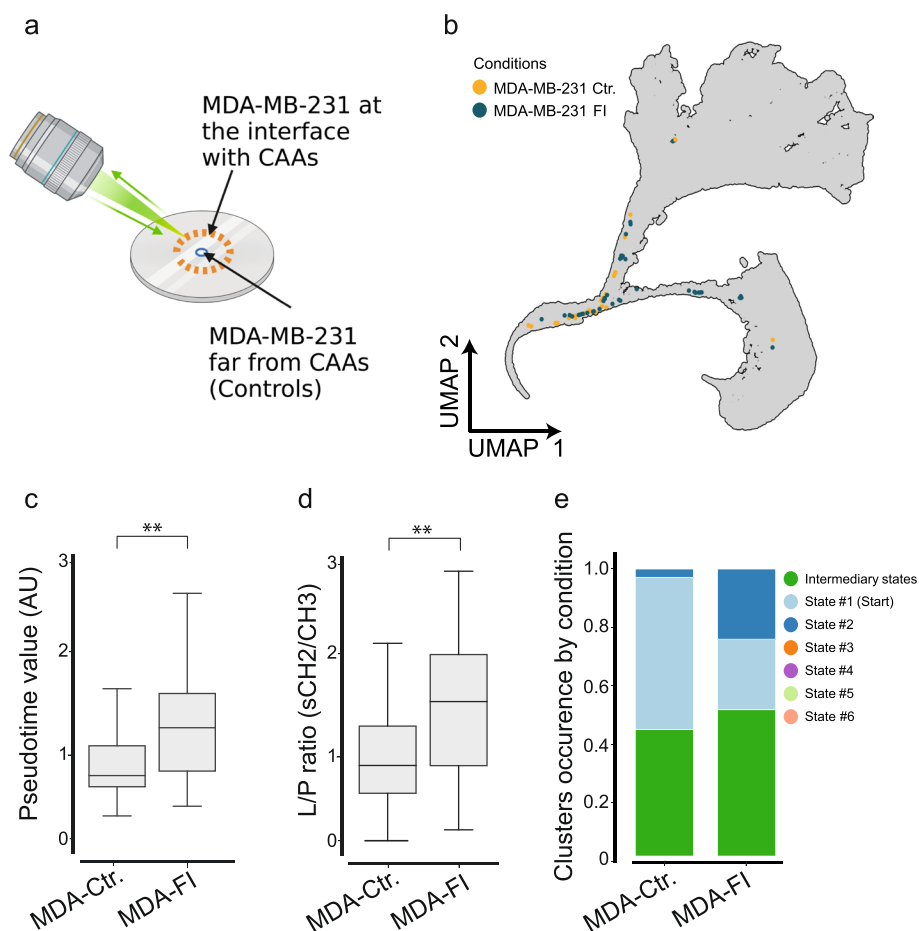


Fig. 5 Analysis of spectral differences of LDs between MDA-MB-231 cells at the invasive front (FI) and those distant from CAAs (Ctr.). **a** Schematic representation of the localization of the two conditions. **b** Projection of tumoral cells spectra on the 2D UMAP constructed from adipocytes data. **c** Quantitative assessment of pseudotime value. **d** Quantitative assessment of lipids/proteins ratio. **e** Relative occurrence of the different cellular states for each condition. For **c** and **d**, p -values were determined by two-tailed student's t -test. **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$

level. This observation indicates that cancer cells located in proximity to CAAs accumulate a greater amount of lipids. We also conducted a quantitative examination of the different states occupied by each of the two types of cancer cells (Fig. 5e). The results reveal that both cell conditions of MDA-MB-231 exhibit spectral profiles corresponding to different states, namely the states #1 and #2 but also the intermediary state. Moreover, in contrast to MDA-MB-231 far from adipocytes, MDA-MB-231 cells at the invasive front exhibit a higher abundance of spectra corresponding to more advanced states (intermediary state and State #2) and fewer spectra corresponding to the initial state.

Discussion

In recent years, there has been growing interest in understanding the role of adipocytes in cancer progression, particularly in the context of breast cancer. Adipocytes are a major component of the TME of breast cancers and have been shown to interact with cancer cells in various ways. However, the heterogeneity of adipocytes and their

potential impact on cancer progression remain poorly understood. In this study, we used Raman spectroscopy and trajectory inference methods to investigate the interactions between 3T3-L1 differentiated adipocytes and MDA-MB-231 breast cancer cells in a 2D co-culture model with cell/cell contact. We employed the Partition-based graph abstraction algorithm (PAGA) for spectral trajectory construction on Raman spectral images of cells.

Trajectory construction is dependent on the choice of dimensionality reduction and clustering algorithms. Various algorithms are available, but there is no consensus on the optimal choice of methods and their parameters. We benchmarked three well-known dimensionality reduction methods: PCA, t-SNE, and UMAP. The results are available Fig. S2. The choice of using UMAP over other methods, such as PCA or t-SNE, is based on several arguments. First, UMAP can highlight fine details while preserving both local and global structures of multidimensional datasets [31]. UMAP also incorporates negative sampling strategies to handle noise and outliers, reducing their influence during the optimization step and resulting in a more robust and accurate low-dimensional representation [31]. Importantly, UMAP exhibits reduced disconnected structures compared to t-SNE, making it well-suited for representing continuous phenomena. Concerning clustering, we benchmarked two well-known algorithms with different configurations and principles, namely Leiden [51] (graph-based method) and K-means, to assess their ability to reconstruct relevant trajectories (Fig. S6 and S6). For our dataset, the visual comparison of these results led to select K-means clustering with $k=20$ clusters to obtain smooth and non-overlapping trajectories spanning the UMAP data distribution.

Based on these selected parameters for trajectory construction, we identified distinct adipocyte subpopulations, each displaying unique spectral signatures indicative of varying lipid compositions irrespective of droplet size. The existence of such adipocyte subpopulations has been reported in 3T3-L1 cell line with immunolabeling methods. For example, Shigmatsu et al. [52] found, using that different subpopulations of adipocytes respond differently to insulin, suggesting that the signaling pathways involved in glucose uptake and lipid metabolism may be regulated differently in these subpopulations. Loo et al. [23] highlight four subpopulations based on their adiponectin, a key player in lipid metabolism, and LDs levels. In our work using label-free Raman spectroscopy associated with PAGA, we identified distinct subpopulations based on their levels of lipid unsaturation. Previous studies using various vibrational techniques like Raman spectroscopy [50, 53–56], Coherent Anti-Stokes Raman Spectroscopy (CARS) [57, 58] or Stimulated Raman Spectroscopy (SRS) [59, 60], combined with dimensionality reduction like UMAP [61], t-SNE [62], or Autoencoder [11], highlighted the unsaturation variability of lipid package but not clearly identified adipocyte subpopulations. Using TI processing, we were able to highlight different subpopulations.

Our findings did not reveal discriminant Raman profiles between CAAs and control adipocytes. However, when comparing the two conditions, we noticed variations in the occurrence of each adipocytes sub-population. This observation is consistent with Tang et al. [20], who found with single nucleus transcriptomic approach no specific adipocyte population unique to either breast cancer adipose tissues or normal breast adipocytes but did notice a variation in the frequency of certain adipocyte subpopulations.

In CAAs, we noted a higher prevalence of spectra linked to the intermediate state, suggesting overall smaller droplets. This trend aligns with the lower pseudotime values found in CAAs, indicating they might move backward in the trajectories compared to control adipocyte populations, redirecting towards the root in the initial State #1. This observation could reflect the delipidation of adipocytes, a well-known process, that has been previously demonstrated by Dirat et al. [3]. In addition, our results show that CAAs are predominantly composed of adipocytes subpopulations marked by an unsaturated lipid profile. This observation is in line with the study of You et al. [63] in mouse and human breast cancer tissue using Raman spectroscopy, where comparing bulk adipose tissue located near and distant from the tumor, but without revealing adipocyte subpopulations. Furthermore, Guryleva et al. [64] found that the fatty acid desaturase 2 (FADS2), which adds unsaturation to the acyl chains of fatty acids, is upregulated in breast adipose tissue compared to normal tissues. Similarly, it has been shown that the stearoyl-CoA desaturase-1 (SCD1) is overexpressed in breast cancer and is associated with poor prognosis [65, 66].

The reprogramming of energy metabolism is acknowledged as a cancer hallmark that plays a significant role in the cancer progression. In addition, metabolic rewiring is seen as one of the mechanisms by which adipocytes can impact the progression of breast cancer [67, 68]. Indeed, in the local area, tumors located near adipocytes (such as breast, prostate cancer, and melanoma) establish a metabolic symbiosis with adipocytes, which results in the alteration of the cancer cell's metabolic program from glycolysis to lipid-dependent energy production. Moreover, studies on the bidirectional interactions between adipocytes and cancer cells indicate that adipocytes provide fatty acids to cancer cells to generate energy, ultimately giving cancer cells a competitive edge through lipid oxidation. In this context, we took advantage of the 2D direct contact co-culture model to conduct preliminary measurements on MDA-MB-231 breast cancer cells. We showed that when in contact with adipocytes, cancer cells exhibited a higher abundance of spectra corresponding to more advanced states and lipid-rich profiles.

While Raman spectroscopy has proven to be a powerful tool for providing label-free, in situ insights into the biochemical composition of cells, it offers only one facet of the information needed to fully understand the complexity of CAAs. Raman spectroscopy is a very effective tool in identifying chemical bonds and associated molecular information, such as lipid unsaturation levels. However, it does not provide the full molecular, or functional context necessary to comprehensively characterize CAAs. To address these limitations, it would be very beneficial to be able to couple Raman spectroscopy with other complementary techniques, such as flow cytometry, immunofluorescence, and single-cell RNA sequencing (scRNA-seq), to gain a more holistic understanding of the CAAs phenotype.

Conclusion

Overall, our study contributes to the growing body of research on the role of adipocytes in cancer progression and underscores the importance of understanding the heterogeneity of adipocyte subpopulations within TME. Using Raman spectroscopy and trajectory inference methods, we have highlighted the complex interactions between adipocytes and breast cancer cells, providing new insights into the metabolic reprogramming that

characterizes the CAA phenotype. Future research should continue to explore these avenues, by combining Raman spectroscopy with flow cytometry, immunofluorescence, and scRNA-seq to fully elucidate the complex biology of CAAs. This will not only advance our understanding of the TME but also pave the way for the development of more targeted and effective cancer therapies.

Abbreviations

3T3-L1	A murine pre-adipocyte cell line used in adipogenesis studies.
ATCC	American type culture collection
CAA	Cancer-associated adipocyte
CH ₂	Methylene group
CH ₃	Methyl group
CO ₂	Carbon dioxide
DMEM	Dulbecco's modified eagle medium
EDTA	Ethylenediaminetetraacetic acid
EMSC	Extended multiplicative signal correction
ER	Estrogen receptor
FADS2	Fatty acid desaturase 2
FCS	Fetal calf serum
HER2	Human epidermal growth factor receptor 2
IBMX	3-isobutyl-1-methylxanthine
IF	Invasive front
KNN	K-nearest neighbors
LD	Lipid droplet
MDA-MB-231	A triple-negative human breast cancer cell line
MST	Minimum spanning tree
PAGA	Partition-based graph abstraction
PCA	Principal component analysis
PPAR γ	Peroxisome proliferator-activated receptor gamma
PR	Progesterone receptor
scRNAseq	Single-Cell RNA sequencing
SCD1	Stearoyl-CoA desaturase-1
TME	Tumor microenvironment
t-SNE	t-Distributed stochastic neighbor embedding
UMAP	Uniform manifold approximation and projection

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s43074-024-00146-3>.

Additional file 1. The supplementary information is available in the 'Supplementary Information' file.

Acknowledgements

We thank the ITMO Cancer and ITMO Technologies pour la Santé coordinated by AVIESAN (National Alliance for Life Sciences & Health), the European Union, the Ligue Nationale contre le Cancer, Conférence de Coordination Inter Régionale du Grand Est (CCIR-GE) and the Grand Est region for their financial support. Special thanks to the Graduate School NANO-PHOT for their framework support. The authors thank the PICT-URCATech platform for technical support.

Authors' contributions

N.G. performed the experiments (Cell culture, Raman acquisition, data processing and writing of the paper); N.G., E.B., C.G., C.M., C.V., L.B., and O.P. performed the study concept, design, and writing of the paper; E.B. and O.P. provided the funding.

Funding

This study was possible thanks to the financial support of the ITMO Cancer and ITMO Technologies pour la Santé coordinated by AVIESAN (National Alliance for Life Sciences & Health), within the framework of the Cancer Plan (France). We thank the European Union and the Grand Est region for financing the doctoral scholarship of Nicolas Goffin. Europe is engaged in the Grand-Est region with FEDER.

Availability of data and materials

Codes and data associated with this study are available on demand.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors agree with the publication.

Competing interests

The authors declare no competing interests.

Received: 17 June 2024 Revised: 30 August 2024 Accepted: 21 September 2024

Published online: 23 October 2024

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin.* 2023;73(1):17–48.
2. Pati S, Irfan W, Jameel A, Ahmed S, Shahid RK. Obesity and cancer: a current overview of epidemiology, pathogenesis, outcomes, and management. *Cancers.* 2023;15(2):485.
3. Dirat B, Bochet L, Dabek M, Daviaud D, Dauvillier S, Majed B, et al. Cancer-associated adipocytes exhibit an activated phenotype and contribute to breast cancer invasion. *Cancer Res.* 2011;71(7):2455–65.
4. Wu Q, Li B, Li Z, Li J, Sun S, Sun S. Cancer-associated adipocytes: key players in breast cancer progression. *J Hematol Oncol.* 2019;12:1–15.
5. Wu C, Dong S, Huang R, Chen X. Cancer-associated adipocytes and breast cancer: intertwining in the tumor micro-environment and challenges for cancer therapy. *Cancers.* 2023;15(3):726.
6. Brock CK, Hebert KL, Artiles M, Wright MK, Cheng T, Windsor GO, et al. A role for adipocytes and adipose stem cells in the breast tumor microenvironment and regenerative medicine. *Front Physiol.* 2021;12:751239.
7. Wu Q, Li B, Li J, Sun S, Yuan J, Sun S. Cancer-associated adipocytes as immunomodulators in cancer. *Biomark Res.* 2021;9:1–21.
8. Kouri MA, Spyratou E, Karnachoriti M, Kalatzis D, Danias N, Arkadopoulos N, et al. Raman spectroscopy: a personalized decision-making tool on clinicians' hands for in situ cancer diagnosis and surgery guidance. *Cancers.* 2022;14(5):1144.
9. Sbroscia M, Di Gioacchino M, Ascenzi P, Crucitti P, di Masi A, Giovannoni I, et al. Thyroid cancer diagnosis by Raman spectroscopy. *Sci Rep.* 2020;10(1):13342.
10. Hanna K, Krzoska E, Shaaban AM, Muirhead D, Abu-Eid R, Speirs V. Raman spectroscopy: Current applications in breast cancer diagnosis, challenges and future prospects. *Br J Cancer.* 2022;126(8):1125–39.
11. He C, Zhu S, Wu X, Zhou J, Chen Y, Qian X, et al. Accurate tumor subtype detection with raman spectroscopy via variational autoencoder and machine learning. *ACS Omega.* 2022;7(12):10458–68.
12. Paidi SK, Rodríguez Troncoso J, Raj P, Monterroso Diaz P, Ivers JD, Lee DE, et al. Raman spectroscopy and machine learning reveals early tumor microenvironmental changes induced by immunotherapy. *Cancer Res.* 2021;81(22):5745–55.
13. Paidi SK, Diaz PM, Dadgar S, Jenkins SV, Quick CM, Griffin RJ, et al. Label-free Raman spectroscopy reveals signatures of radiation resistance in the tumor microenvironment. *Cancer Res.* 2019;79(8):2054–64.
14. Torre-Gutiérrez LDL, Martínez-Zérega B, Oseguera-Galindo D, Aguilar-Lemarroy A, Jave-Suárez L, Torres-González L, et al. Breast cancer chemotherapy treatment monitoring based on serum sample Raman spectroscopy. *Lasers Med Sci.* 2022;37(9):3649–59.
15. Cave E, Crowther NJ. The use of 3T3-L1 murine preadipocytes as a model of adipogenesis. *Pre-Clin Models Tech Protoc.* 2019;263–72.
16. Ralston JC, Mutch DM. SCD1 inhibition during 3T3-L1 adipocyte differentiation remodels triacylglycerol, diacylglycerol and phospholipid fatty acid composition. *Prostaglandins Leukot Essent Fat Acids.* 2015;98:29–37.
17. Rizzatti V, Boschi F, Pedrotti M, Zoico E, Sbarbati A, Zamboni M. Lipid droplets characterization in adipocyte differentiated 3T3-L1 cells: size and optical density distribution. *Eur J Histochem EJH.* 2013;57(3):159–62.
18. Zebisch K, Voigt V, Wabitsch M, Brandsch M. Protocol for effective differentiation of 3T3-L1 cells to adipocytes. *Anal Biochem.* 2012;425(1):88–90.
19. Huang Z, Yu P, Tang J. Characterization of triple-negative breast cancer MDA-MB-231 cell spheroid model. *Oncotargets Ther.* 2020;5395–405.
20. Tang L, Li T, Xie J, Huo Y, Ye J. Diversity and heterogeneity in human breast cancer adipose tissue revealed at single-nucleus resolution. *Front Immunol.* 2023;14:1158027.
21. Ferrero R, Rainer P, Deplancke B. Toward a consensus view of mammalian adipocyte stem and progenitor cell heterogeneity. *Trends Cell Biol.* 2020;30(12):937–50.
22. Kassotis CD, Hoffman K, Völker J, Pu Y, Veiga-Lopez A, Kim SM, et al. Reproducibility of adipogenic responses to metabolism disrupting chemicals in the 3T3-L1 pre-adipocyte model system: An interlaboratory study. *Toxicology.* 2021;461:152900.
23. Loo LH, Lin HJ, Singh DK, Lyons KM, Altschuler SJ, Wu LF. Heterogeneity in the physiological states and pharmacological responses of differentiating 3T3-L1 preadipocytes. *J Cell Biol.* 2009;187(3):375.
24. Le TT, Cheng JX. Single-cell profiling reveals the origin of phenotypic variability in adipogenesis. *PLoS One.* 2009;4(4):e5189.
25. Zhu Q, Zhu Y, Hepler C, Zhang Q, Park J, Gliniak C, et al. Adipocyte mesenchymal transition contributes to mammary tumor progression. *Cell Rep.* 2022;40(11).
26. Lee J, Hong BS, Ryu HS, Lee HB, Lee M, Park IA, et al. Transition into inflammatory cancer-associated adipocytes in breast cancer microenvironment requires microRNA regulatory mechanism. *PLoS One.* 2017;12(3):e0174126.
27. Saelens W, Cannoodt R, Todorov H, Saey Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol.* 2019;37(5):547–54.

28. Chen B, Zhu L, Yang S, Su W. Unraveling the heterogeneity and ontogeny of dendritic cells using single-cell RNA sequencing. *Front Immunol.* 2021;12:711329.
29. Zhao J, Jaffe A, Li H, Lindenbaum O, Sefik E, Jackson R, et al. Detection of differentially abundant cell subpopulations in scRNA-seq data. *Proc Natl Acad Sci.* 2021;118(22):e2100293118.
30. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol.* 2019;15(6):e8746.
31. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. 2018;arXiv:1802.03426.
32. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res.* 2008;9(11):2579–605.
33. Moon KR, Stanley JS III, Burkhardt D, van Dijk D, Wolf G, Krishnaswamy S. Manifold learning-based methods for analyzing single-cell RNA-sequencing data. *Curr Opin Syst Biol.* 2018;7:36–46.
34. Xiang R, Wang W, Yang L, Wang S, Xu C, Chen X. A comparison for dimensionality reduction methods of single-cell RNA-seq data. *Front Genet.* 2021;12:646936.
35. Sigle M, Rohlfing AK, Kenny M, Scheuermann S, Sun N, Graebner U, et al. Translating genomic tools to Raman spectroscopy analysis enables high-dimensional tissue characterization on molecular resolution. *Nat Commun.* 2023;14:5799.
36. Yang KX, Xu F, Zhu L, Li H, Sun Q, Yan A, et al. An Isotope-Labeled Single-Cell Raman Spectroscopy Approach for Tracking the Physiological Evolution Trajectory of Bacteria toward Antibiotic Resistance. *Angew Chem Int Ed Engl.* 2023;62(14):e202217412.
37. Kazemzadeh M, Martinez-Calderon M, Otupiri R, Artuyants A, Lowe MM, Ning X, et al. Manifold learning enables interpretable analysis of Raman spectra from extracellular vesicle and other mixtures. *bioRxiv [Preprint].* 2023;2023.03.20.533481. <https://doi.org/10.1101/2023.03.20.533481>.
38. Deconinck L, Cannoodt R, Saelens W, Deplancke B, Saeys Y. Recent advances in trajectory inference from single-cell omics data. *Curr Opin Syst Biol.* 2021;27:100344.
39. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics.* 2018;19:477.
40. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Götting B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* 2019;20:1–9.
41. Setty M, Kisieliovas V, Levine JH, Gayoso A, Mazutis L, Pe'er D. Characterization of cell fate probabilities in single-cell data with Palantir. *Nat Biotechnol.* 2019;37:451–60.
42. Ji Z, Ji H. TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* 2016;44:e117.
43. Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, et al. Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. *Cell Stem Cell.* 2015;17(3):360–72.
44. Haghverdi L, Büttner M, Wolf FA, Büttner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods.* 2016;13:845–8.
45. Goffin N, Buache E, Charpentier C, Lehrter V, Morjani H, Gobinet C, et al. Trajectory inference for unraveling dynamic biological processes from Raman spectral data. *Anal Chem.* 2023;95(9):4395–403.
46. Afseth NK, Kohler A. Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. *Chemometr Intell Lab Syst.* 2012;117:92–9.
47. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 2018;19(1):15.
48. Tritschler S, Büttner M, Fischer DS, Lange M, Bergen V, Lickert H, et al. Concepts and limitations for learning developmental trajectories from single cell genomics. *Development.* 2019;146(12):dev170506.
49. Schapiro D, Jackson HW, Raghuraman S, Fischer JR, Zanotelli VRT, Schulz D, et al. histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat Methods.* 2017;14(9):873–6. <https://doi.org/10.1038/nmeth.4391>.
50. Tratwal J, Falgayrac G, During A, Bertheaume N, Bataclan C, Tavakol DN, et al. Raman microspectroscopy reveals unsaturation heterogeneity at the lipid droplet level and validates an in vitro model of bone marrow adipocyte subtypes. *Front Endocrinol.* 2022;13:1001210.
51. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports.* 2018;9.
52. Shigematsu S, Miller SL, Pessin JE. Differentiated 3T3L1 Adipocytes Are Composed of Heterogeneous Cell Populations with Distinct Receptor Tyrosine Kinase Signaling Properties*. *J Biol Chem.* 2001;276:15292–7.
53. Haka AS, Sue E, Zhang C, Bhardwaj P, Sterling J, Carpenter C, et al. Noninvasive Detection of Inflammatory Changes in White Adipose Tissue by Label-Free Raman Spectroscopy. *Anal Chem.* 2016;88(4):2140–8.
54. Mitchell A, Ashton L, Yang XB, Goodacre R, Smith A, Kirkham J. Detection of early stage changes associated with adipogenesis using Raman spectroscopy under aseptic conditions. *Cytometry.* 2015;87:1012–9.
55. Czamara K, Majka Z, Stanek E, Hachlica N, Kaczor A. Raman studies of the adipose tissue: Current state-of-art and future perspectives in diagnostics. *Prog Lipid Res.* 2022;87:101183.
56. Stanek E, Pacia MZ, Kaczor A, Czamara K. The distinct phenotype of primary adipocytes and adipocytes derived from stem cells of white adipose tissue as assessed by Raman and fluorescence imaging. *Cell Mol Life Sci CMLS.* 2022;79(7):383.
57. Rinia HA, Burger KNJ, Bonn M, Müller M. Quantitative label-free imaging of lipid composition and packing of individual cellular lipid droplets using multiplex CARS microscopy. *Biophys J.* 2008;95(10):4908–14.
58. Bonn M, Müller M, Rinia HA, Burger KNJ. Imaging of chemical and physical state of individual cellular lipid droplets using multiplex CARS microscopy. *J Raman Spectrosc.* 2009;40:763–9.
59. Zhang C, Huang KC, Rajwa B, Li J, Yang S, Lin H, et al. Stimulated Raman scattering flow cytometry for label-free single-particle analysis. *Optica.* 2017;4:103.
60. Suzuki Y, Kobayashi K, Wakisaka Y, Deng D, Tanaka S, Huang CJ, et al. Label-free chemical imaging flow cytometry by high-speed multicolor stimulated Raman scattering. *Proc Natl Acad Sci U S A.* 2019;116:15842–8.

61. Pavillon N, Smith NI. Non-invasive monitoring of T cell differentiation through Raman spectroscopy. *Sci Rep.* 2023;13(1):3129.
62. Hsu CC, Xu J, Brinkhof B, Wang H, Cui Z, Huang WE, et al. A single-cell Raman-based platform to identify developmental stages of human pluripotent stem cell-derived neurons. *Proc Natl Acad Sci U S A.* 2020;117:18412–23.
63. You S, Tu H, Zhao Y, Liu YZ, Chaney EJ, Marjanovic M, et al. Raman Spectroscopic Analysis Reveals Abnormal Fatty Acid Composition in Tumor Micro- and Macroenvironments in Human Breast and Rat Mammary Cancer. *Sci Rep.* 2016;6:32922.
64. Guryleva MV, Penzar DD, Chistyakov DV, Mironov AA, Favorov AV, Sergeeva MG. Investigation of the Role of PUFA Metabolism in Breast Cancer Using a Rank-Based Random Forest Algorithm. *Cancers.* 2022;14(19):4663.
65. Zhao T, Gao P, Li Y, Tian H, Ma D, Sun N, et al. Investigating the role of FADS family members in breast cancer based on bioinformatic analysis and experimental validation. *Front Immunol.* 2023;14:1074242.
66. Sen U, Coleman C, Sen T. Stearoyl coenzyme A desaturase-1: multitasker in cancer, metabolism, and ferroptosis. *Trends Cancer.* 2023;9(6):480–9.
67. Wang YY, Attané C, Milhas D, Dirat B, Dauvillier S, Guérard A, et al. Mammary adipocytes stimulate breast cancer invasion through metabolic remodeling of tumor cells. *JCI Insight.* 2017;2(4):e87489.
68. Balaban S, Shearer RF, Lee LS, van Geldermalsen M, Schreuder M, Shtein HC, et al. Adipocyte lipolysis links obesity to breast cancer growth: adipocyte-derived fatty acids drive breast cancer cell proliferation and migration. *Cancer Metab.* 2017;5:1.

Publisher's Note

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