Journal of Raman Spectroscopy

WILEY

Journal of RAMAN SPECTROSCOPY

RESEARCH ARTICLE OPEN ACCESS

Discriminating Chemoresistant and Chemosensitive Tubo-Ovarian High-Grade Serous Carcinoma Cells With Raman Microscopy

Elina Harju[1] | Parisa Movahedi[2] | Teemu Tomberg[1] | Sara J. Fraser-Miller[3,4] | Francesco Piga[1] | Jukka Saarinen[1] | Kathleen J. Sircombe[5] | Jacopo Zini[1] | Jukka Heikkonen[2] | Anna Laury[6] | Sarah Hook[5] | Keith C. Gordon[3] | Tapio Pahikkala[2] | Clare J. Strachan[1]

[1]Division of Pharmaceutical Chemistry and Technology, Faculty of Pharmacy, University of Helsinki, Helsinki, Finland | [2]Department of Computing, University of Turku, Turku, Finland | [3]The Dodd-Walls Centre for Photonic and Quantum Technologies—Te Whai Ao and Department of Chemistry, University of Otago, Dunedin, New Zealand | [4]College of Science and Engineering, Flinders University, Adelaide, South Australia, Australia | [5]School of Pharmacy, University of Otago, Dunedin, New Zealand | [6]Department of Pathology, University of Helsinki, Helsinki, Finland

Correspondence: Elina Harju (elina.a.harju@helsinki.fi)

Received: 12 March 2026 | **Revised:** 21 April 2026 | **Accepted:** 24 April 2026

ABSTRACT

Chemoresistance is a major obstacle to effective cancer treatment, particularly in tubo-ovarian high-grade serous carcinoma (HGSC), the most lethal gynaecological malignancy. Predicting patient-specific chemoresistance remains challenging due to tumour heterogeneity and the lack of reliable biomarkers. Raman spectroscopy, a label-free technique that provides biochemical insights into cells and tissues without the need for specific biomarkers, has been extensively applied in cancer research, but its full potential for detecting subtle biochemical changes linked to chemoresistance in HGSC at the single-cell and subcellular levels remains underexplored. Another critical challenge is the estimation of classification performance on future data with cross-validation (CV) in the presence of batch effects. In this study, we demonstrated that confocal Raman microscopy combined with multivariate analysis can discriminate between cisplatin-resistant (TYK-nu-CP.r) and cisplatin-sensitive (TYK-nu) HGSC cell lines with 78% accuracy without batch correction. After batch correction, the accuracy improved to 84%. Feature importance analysis suggested that the separation was linked to a higher level of lipid unsaturation and elevated glutathione levels in the chemosensitive cell line. Additionally, we proposed a new CV-based area under the receiver operating characteristic curve (AUC) estimator that accounts for the batch effects better than the popularly used leave-one-batch-out CV. Together, these results show that with careful data processing, accounting for biases and batch effects, Raman microscopy enables reliable detection of chemoresistance at a cellular level and can provide insights into the molecular basis of chemoresistance. This study suggests that Raman microscopy holds promise as a tool for predicting chemoresistance in HGSC and guiding personalised treatment strategies.

1 | Introduction

Chemotherapy is a cornerstone of cancer treatment. However, chemoresistance is a substantial cause of treatment failure and ultimately patient death. Furthermore, resistance to chemotherapy agents varies substantially between patients and is also affected by prior treatment. Despite such diversity in chemoresistance, chemotherapy treatment is generally based on standardised treatment regimens. Tools that would allow the prediction of patient-specific chemoresistance

Elina Harju and Parisa Movahedi contributed equally.

would allow more effective and personalised chemotherapy treatment [1].

One form of cancer for which the above situation is highly relevant is ovarian cancer. Ovarian cancer is the deadliest cancer among all gynaecological malignancies [2]. Tubo-ovarian high-grade serous carcinoma (HGSC) constitutes the majority of cases and is also the most aggressive. Platinum-based chemotherapy is the standard treatment for HGSC, often following cytoreductive surgery [3]. Resistance to platinum-based chemotherapy is common and a significant contributor to mortality in HGSC patients, and it may be intrinsic (existing prior to treatment) or acquired during the course of chemotherapy. Platinum drug resistance in tumour cells is believed to occur via several mechanisms [4], and elucidating the precise nature of these mechanisms is challenging and an enormous field of ongoing research [5–8]. Furthermore, HGSC tumours exhibit a large degree of intratumoral heterogeneity that contributes to the development of chemoresistance [9, 10], which underscores the need for spatially resolved analytical methods in HGSC research.

Raman spectroscopy has much potential for individualised chemoresistance screening prior to chemotherapy treatments, as it is not reliant on a single biomarker, or even a precise understanding of the molecular and cellular mechanisms [11]. It encompasses a group of label-free techniques (either spatially resolved or not) that provide biochemical information on cells and tissues. Non-spatially resolved Raman methods analyse bulk samples fast and are well-suited for situations where information on average biochemical composition is sufficient. In contrast, spatially resolved Raman microscopy techniques—including spontaneous confocal Raman microscopy and coherent methods such as coherent anti-Stokes Raman scattering (CARS) and stimulated Raman scattering (SRS)—enable biochemical imaging at subcellular resolution. Coherent techniques, particularly SRS, are especially advantageous for imaging applications due to their faster acquisition speeds and higher spatial resolution, which make them well-suited for studying even submicron-sized cellular structures such as lipid droplets (LDs) [12]. Although coherent Raman excels in imaging applications, spontaneous Raman is generally able to generate more information-rich spectra, potentially making it more suitable for detecting subtle molecular changes associated with chemoresistance.

The use of Raman spectroscopy in cancer research has been extensively studied since the 1990s, particularly for distinguishing malignant from healthy tissue [13]. Characterisation of chemoresistance or, more generally, the effect of chemotherapy drugs on cells is a more recent application [14–19], and presumably a more subtle and challenging task, given that it involves differentiating between phenotypically more similar cells and tissues. Despite the potential of Raman-based approaches for this task, only a couple of studies have investigated their use in differentiating between chemoresistant and chemosensitive ovarian cancer tissues [20, 21], and to our knowledge, only two studies have applied Raman microscopy to differentiate between single chemoresistant and sensitive ovarian cancer cells [17, 22]. In the study of Moradi et al., they measured spectra of diffraction-limited volumes at arbitrary locations within cells, and although they managed to differentiate between the cell lines and attributed this to relative increases in proteins and glutathione (GSH) in the cisplatin-resistant cells, the subcellular regions contributing to the observed spectral separation were not investigated. Very recently, Li et al. investigated two pairs of chemoresistant and chemosensitive ovarian cancer cell lines and reported very high classification performance [22]. Although their results are promising, the available methodological information is limited, making it difficult to evaluate the broader applicability of the findings. Moreover, neither of these studies involved an HGSC model, which is the most common and lethal ovarian cancer subtype.

Several methodological concerns remain underexplored in the existing literature on Raman microscopy and cancer. Sampling details are often sparse, even though Raman microscopy typically probes areas smaller than an individual cell, which has important implications for diagnostic development. Furthermore, even though the importance of proper model evaluation is increasingly acknowledged in the field, it is still not uncommon to report near-perfect classification performance based on small datasets [23]. In experimental domains, data are frequently affected by batch effects, which introduce systematic differences between samples measured at different timepoints or under varying experimental conditions. To ensure models capture true biological signals rather than experimental artifacts, evaluation strategies must explicitly account for these variations. A key focus of this work was to propose a new cross-validation (CV) strategy for performance estimation. We propose a novel batch-aware CV strategy for area under the receiver operating characteristic curve (AUC) estimation, particularly in settings where test samples from opposite classes originate from different batches.

In this study, we used confocal Raman microscopy with line mapping and partial least squares discriminant analysis (PLS-DA), linear support vector machines (SVMs) and logistic regression to distinguish between the HGSC cell line TYK-nu and its chemoresistant subline TYK-nu-CP.r at the single-cell level. Furthermore, we aimed to identify which subcellular regions contribute to the molecular differences driving chemoresistance. To this end, we compared the analysis of average cell spectra with that of spectra from specific cellular regions. Finally, to further explore the role of LDs in these differences, we employed SRS microscopy to directly analyse LD composition and its relationship to the chemoresistance classification results.

2 | Materials and Methods

2.1 | Cell Lines

For the spontaneous Raman analyses, human ovarian cancer cells TYK-nu and its more chemoresistant subline TYK-nu-CP.r (lots 02252019 and 08022018, Japanese collection of research bioresources cell bank, Japan) were cultured at the University of Otago, New Zealand. Cells were cultured in MEM with Glutamax (Gibco, Thermo Fisher Scientific, USA) supplemented with 100-μg/mL streptomycin, 100-U/mL penicillin (Pen-Strep, Gibco, Thermo Fisher Scientific, USA) and 10% foetal bovine serum (Moregate Biotech, New Zealand) at 37°C in a humidified 5% CO_2 environment. Once the cells achieved the logarithmic

growth phase (approximately 70%–90% confluency), they were prepared for subculturing by detaching them with 0.05% trypsin–EDTA (Gibco, Thermo Fisher Scientific, USA). To confirm differences in cisplatin resistance between the cell lines, a cytotoxicity assay was conducted (Figure S1).

For the Raman spectroscopy analysis, eight samples from each cell line were prepared from different cell culturing flasks, within a range of four passages. The number of cells analysed in each sample for each measurement batch is reported in Table S1. For each sample, 150,000 cells were seeded onto a Raman-grade calcium fluoride substrate (CaF_2, Crystran LTD, UK) within a petri dish and left to incubate for 2 days before fixation. Post-incubation, both cell lines showed similar confluency on the substrates without visible morphological differences or attachment issues between cells cultured in flasks and on CaF_2 substrates.

Besides the samples seeded on CaF_2, separate samples were prepared with the intention of collecting reference spectra of subcellular regions. For this purpose, the cells were seeded on glass-bottomed petri dishes (MatTek Corporation, USA). Fixation of all samples was carried out in the same manner; after the incubation, the cells were first washed with phosphate-buffered saline (PBS, Gibco, Thermo Fisher Scientific, USA) and then fixed with 4% paraformaldehyde (PFA, Image-iT, Thermo Fisher Scientific, USA) at room temperature for 10 min. After removing the PFA, the cells were washed with PBS three times and then kept in PBS during the measurements.

2.2 | Confocal Raman Microscopy

The confocal Raman measurements were performed using a WiTec Alpha 300R+ confocal Raman microscope (WITec GmbH, Germany). To ensure consistency, all cell samples were measured within 8 h after fixation. Measurements were performed with an excitation wavelength of 532 nm (Coherent, Santa Clara, USA), and a 63× NA 1.0 water immersion objective (W Plan-Apochromat, Zeiss, Germany) was used to focus the light on the sample and to collect the signal. A grating with 600 g/mm and a blaze wavelength of 500 nm was used for spectral dispersion and detected with a charge-coupled device (CCD, Andor iDus 401, Oxford Instruments, Abingdon, UK). The measured spectral region was ~−60 to 3785 cm^{-1}. To minimise the day-to-day spectral variation, the Raman shift was calibrated with the silicon 520 cm^{-1} peak at the beginning of each measurement day. Spectra were acquired with WITec Control Plus software with a laser power of 15 mW, and the exposure time was 10 s with three accumulations.

2.3 | Spectral Preprocessing and Data Analysis

Raman data analysis was performed on both individual point spectra and average line map spectra and in both cases, all spectra underwent a similar preprocessing procedure. First, cosmic ray spikes were removed using two despiking algorithms, the Whitaker-Hayes [24] (*cosmicdd*) and Barton-Hennelly [25] (*cosmicmp*) functions, available in the Spectrapepper Python toolbox [26]. Spike removal was performed in multiple spectral segments with parameters specifically adjusted for each segment. The spectra were then smoothed using a Savitzky–Golay filter with a window length of 3 and a polynomial order of 2. Subsequently, a constant baseline correction (the minimum intensity value) was applied to the spectra. For the average spectra approach, line-averaged spectra were calculated at this stage.

Aqueous buffer subtraction from cell spectra involves several conceptual challenges, as both bulk water and biomolecule-associated water contribute to the OH-stretch region [27], and the optimal strategy for handling these contributions in the context of mitigating physical inhomogeneity has not been systematically investigated in the literature. Here, an average background spectrum of the PBS medium was fitted to each cell spectrum in the range of 3085–3153 cm^{-1} using non-negative least squares regression and subsequently subtracted. To prevent over-subtraction, a small constant (0.05) was subtracted from the residuals during the optimisation process. After subtraction, the resulting cell spectra were visually consistent with those reported by Lang et al., where only the bulk water contribution was removed [27]. After the buffer subtraction, the adaptive iteratively reweighted penalised least squares (airPLS) baseline correction was applied with the smoothing parameter of 1,000,000 and penalty order of 2, and the spectra were SNV-normalised. To account for technical variability across experimental days, we applied day-by-day batch correction using PyComBat, a Python implementation of the ComBat algorithm [28]. The final wavenumber range used for analysis was 430–3025 cm^{-1} with the silent region (1790–2780 cm^{-1}) excluded.

The average line map spectra for each cell were classified using PLS-DA, linear SVM and logistic regression, implemented with scikit-learn [29]. Linear models are well suited for Raman spectroscopy due to their robustness, simplicity and ability to handle high-dimensional spectral data. They perform especially well when sample sizes are limited, where more complex models, such as deep learning, may overfit [30]. Linear models also provide clear weightings for each wavenumber, making it easier to interpret results and link spectral features to underlying chemical or biological differences. Model performance was evaluated under three CV strategies (leave-one-batch-out, leave-one-cell-out and leave-two-batches-out), which are explained in depth in Section 3.

2.4 | Stimulated Raman Scattering Analysis of Lipid Droplets

For the SRS analyses, the same TYK-nu and TYK-nu-CP.r cell lines (obtained from the Institute of Biomedicine and FICAN West Cancer Centre, the University of Turku) were cultured at the University of Helsinki. Cell culturing, cytotoxicity assay and sample preparation were done similarly as described earlier, with minor modifications primarily in material suppliers (see section "Cell culturing for SRS cell samples" in the Supporting Information).

The samples were analysed using an in-house-built SRS microscope, which has been described in detail elsewhere [31]. Briefly, the system comprises an inverted Olympus FV3000 confocal laser scanning microscope (Olympus, Japan), coupled with a

dual-output femtosecond laser source (InSight X3+, Spectra-Physics, MKS Instruments, USA). This laser provides a tunable pump beam and a fixed Stokes beam at 1045 nm. The beams are passed through a spectral focusing, timing and recombination unit (SF-TRU, USA) before coupling to the FV3000 microscope. The light was focused on samples using a UPLSAPO 60xW objective (Olympus, Japan), and the signal was collected in transmission mode using an oil immersion condenser (Leica Microsystems, Germany) dipped in PBS containing fixed cells in a petri dish. Detection was carried out using an SRS detection system (APE Angewandte Physik & Elektronik Berlin, Germany). Measurements were done on the CH stretching vibrational region, using two pump laser wavelengths: 805 nm (2820–2945 cm^{-1}) and 795 nm (2950–3040 cm^{-1}). The step size in these hyperspectral SRS images was 5 cm^{-1}, resulting in SRS images with 45 wavenumbers. The images were 512 × 512–1024 × 1024 pixels in size, and the pixel size was 0.104 μm. Four replicate samples were analysed for each of the two cell lines, with a minimum of 10 images captured at different sites of each sample.

LD spectra were extracted and preprocessed from SRS images using MATLAB, following a multi-step procedure. First, noise reduction was applied to the SRS images using block-matching and 4D filtering (BM4D) [32]. Principal component analysis (PCA) was then performed, and PC3 was identified to include features corresponding to LDs. Circular features representing LDs were detected in the PC3 images using the *imfindcircles* function, and an LD mask was created based on this detection. Next, background subtraction was performed to account for local cellular background spectra specific to each LD. To minimise potential edge effects, an 8-pixel buffer was added to the radius of each LD in the mask. The *inpaintCoherent* function was then applied to replace the LD regions in the hyperspectral image with interpolated values derived from surrounding pixels. These interpolated background spectra were subtracted from the average spectra of the corresponding LDs using the *lsqlin* function for linear least-squares fitting. After this, the spectra were further processed by linear baseline correction, and the peak intensity ratio at 3010 cm^{-1}/2845 cm^{-1} was calculated for each spectrum.

3 | Results and Discussion

3.1 | Method Optimisation for Confocal Raman Microscopy

For Raman analysis of adherent cells, typical substrate options include glass, quartz and CaF_2. For this study, which aimed to resolve subtle spectral differences between two variants of the same cell line, we used Raman-grade CaF_2 because it results in the least spectral interference. However, its refractive index is close to that of the aqueous buffer, resulting in poor intracellular contrast in brightfield images of the confocal Raman microscope (Figure S2) and thereby influencing our sampling strategy.

It is well documented that Raman microscopy enables detection of biomolecular heterogeneity and subcellular structures in cells [33, 34]. The CH stretching region (2800–3000 cm^{-1}) contains features primarily derived from lipids, proteins and nucleic acids, and it alone provides sufficient contrast to differentiate between major cellular compartments, here categorised as the nucleus (all intranuclear structures) and cytoplasm (all non-nuclear intracellular regions) [35]. This was also demonstrated in our study, as evidenced in Figure 1, which shows single-cell hyperspectral Raman maps, generated using PCA on the CH stretching region. The cytoplasm contains a higher concentration of lipids than the nucleus, with these lipids having a characteristic signal attributed to CH_2 functional groups at 2845 cm^{-1}. In contrast, the nucleus, which contains fewer methylene groups and more vibrations originating from proteins and nucleic acids, exhibits a different spectral profile in the CH stretching region.

Previous studies suggest that cell-averaged spectra can produce robust classification models [34, 36, 37], and therefore, this approach was investigated first. The acquisition of representative Raman spectra that accurately reflect a cell's entire chemical composition is a topic of active discussion in the Raman spectroscopy community [34]. As noted by Schie et al., using a large focal spot to cover a larger area of the cell is one option, but the reduced confocality results in an increased background contribution in the spectra. Conversely, full-cell mapping of a large number of cells and samples with a small spot size is impractical due to the slow acquisition speed. As a result, we opted for a middle ground and a more feasible approach by collecting line maps across the geometric centre of cells, determined through visual estimation. This approach, inspired by previous studies [14, 38], provided a practical compromise between representativeness of the spectra and acquisition speed, albeit with a tendency to overrepresent nuclear spectra relative to cytoplasmic spectra. Another advantage of this approach is that it enabled the analysis of individual subcellular spectra, as described later.

Upon examining the two cell lines under a standard light microscope, some morphological differences were evident (Figure S3). Cells from the chemosensitive line typically appeared thinner and more spread out with irregular morphology, whereas those from the chemoresistant line were generally thicker and displayed a spindle-like form. Both cell morphologies were present in each cell line but typically differed in their prevalence. Cells are typically thicker at the central/nucleus area [33], and the centre part yields spectra with a relatively high signal-to-noise ratio. Conversely, signals close to the edges, from the thin spread-out cytoplasmic regions, tend to be weaker and dominated by the features originating from the aqueous buffer (Figure S2). Consequently, comparing Raman spectra of two morphologically different cell lines poses challenges, as the average spectral differences between cells could be enormous and reflect physical inhomogeneity rather than subtle differences in biochemical composition.

To ensure a meaningful spectral analysis between the two cell sublines, both types of morphologies were sampled in roughly equal amounts, rather than relying on random sampling. Sampling involved selecting approximately 10 cells from each sample. For each selected cell, a line map comprising 20 spectra was systematically measured across the geometric central portion of the cell, as described earlier. Measurements were conducted on eight cell samples per cell line, with each sample considered as a distinct biological replicate. One set comprising two samples—one from each cell line—was measured on the

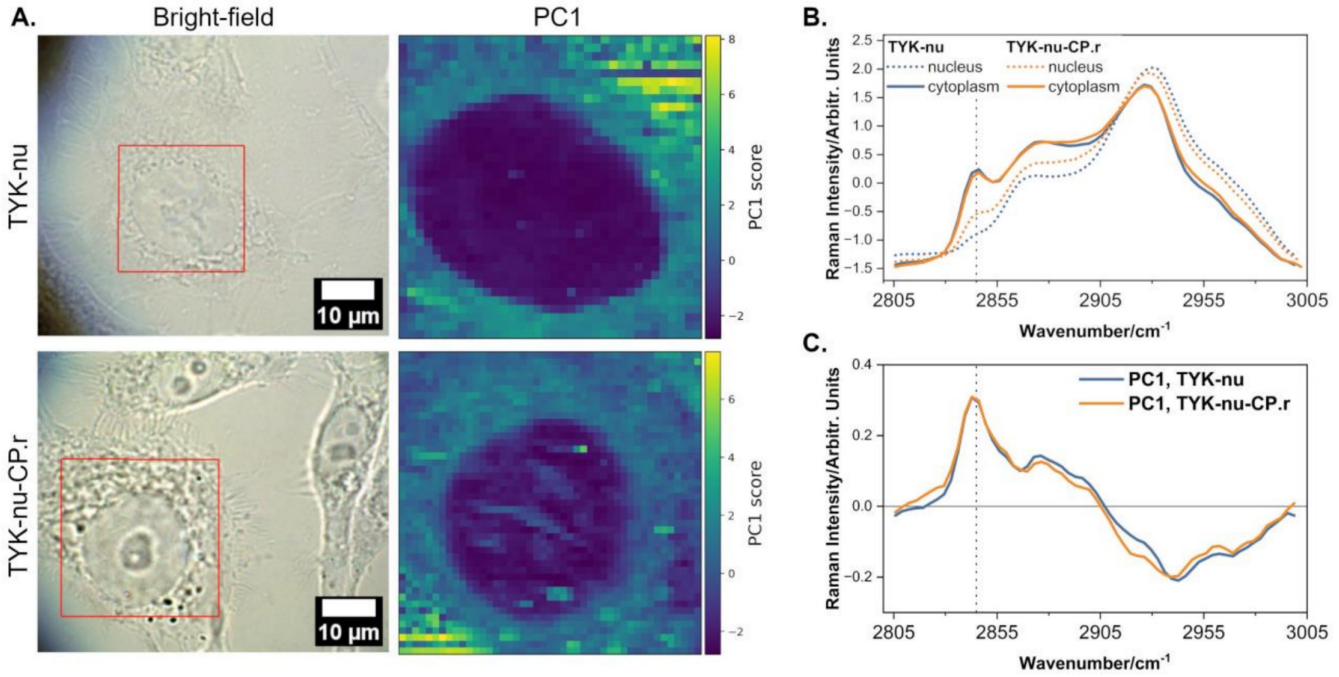

FIGURE 1 | Confocal Raman mapping and PCA analysis of the CH region of single TYK-nu and TYK-nu-CP.r cells, measured on glass substrate. (A) Left: Brightfield images of cells. Red rectangles indicate the regions where Raman maps were acquired with a step size of 1 µm. Right: Corresponding PC1 score maps obtained from PCA analysis of the CH regions of the Raman map insets, showing clear contrast between nuclear and cytoplasmic regions. (B) Mean Raman spectra of nuclear and cytoplasmic regions for both cell maps and (C) the corresponding PC1 loading plots; in (B) and (C), the vertical dashed line marks the Raman shift at 2845 cm^{-1}, typical for lipid vibrations. Both maps were recorded with an integration time of 3 s per pixel and a laser power of 15 mW.

same day, using the same experimental parameters. This was done in order to minimise any day-to-day inconsistencies or artifacts that could unintentionally introduce variance between the two cell lines in the collected data. To further minimise the impact of physical inhomogeneity on data analysis, the bulk water contribution was subtracted from the spectra.

3.2 | Classification Performance on Average Line Map Spectra

The aim of classification performance evaluation is to estimate, from a finite sample of data, how well the classifier under consideration succeeds on future data. For this purpose, we used binary classification accuracy and area under receiver operating characteristic (ROC) curve (AUC). Binary classification accuracy is the probability of correct classification on future data. AUC can be interpreted as the probability of predicting a larger value for a randomly drawn member of the positive class than that for the negative class plus 0.5 times the probability of predicting a tie, i.e.,

$$P\left(f(X) > f\left(X'\right)\right) + 0.5P\left(f(X) = f\left(X'\right)\right)$$

where X and X' are, respectively, randomly drawn members of positive and negative classes and f denotes a classifier that outputs a real-valued confidence level for the datum being a member of the positive class. In the literature, this is sometimes referred to as the Mann–Whitney parameter, since it is the distributional counterpart of the well-known Mann–Whitney statistic [39]. Indeed, the Mann–Whitney statistic is used as an AUC estimator from a finite sample of data. The statistic averages over every pair of positive and negative class members in the sample and counts one whenever the order of their predictions is correct, half whenever they are tied, and zero otherwise.

A batch effect refers to systematic, non-biological bias in experimental data that arises when samples processed in different groups (batches) show technical differences unrelated to the underlying biological signal [40, 41]. These unintentional, systematic errors typically come from inconsistencies in experimental conditions, such as samples being processed on different days or different runs of the instruments. To quantify the batch effects on classification performance in more detail, we considered two variants of AUC that we call intra-batch AUC and inter-batch AUC. Intra-batch AUC corresponds to the same probability as the standard AUC, but is conditional on the two data being from the same batch. Inter-batch AUC is AUC conditional on the two data being from different batches. These are different quantities that provide answers to separate practical research questions. The former is useful when two classes must be separated within future data obtained as a single batch. The latter corresponds to the more difficult problem of separating the classes with different batch effects.

When a learning algorithm is used for training a classifier from data, the CV is used for estimating the classifier's accuracy and the AUC variants on future data. CV estimators also account for the randomness in the training data in addition to that of the test data. For example, the expected classification accuracy of a learning algorithm is the probability of correctly classifying a

datum, given that a certain amount of training data drawn from the same distribution is used for training the classifier with the learning algorithm under consideration. The CV estimators typically average over several splits of the available sample of data into training and test parts.

To account for batch effects on learning algorithms' prediction performance in terms of classification accuracy, we considered two separate CV estimators. The first assumes that the future data is from the same batches as the training data, and it is called leave-one-cell-out cross-validation (LOCO-CV) (Figure 2A). The second reflects the more realistic setup of future data being from a new batch unseen during the training. In the literature, the latter is popularly called the leave-one-batch-out cross-validation (LOBO-CV), illustrated in Figure 2B [42].

Regarding AUC, the LOBO-CV estimator matches intra-batch AUC on a future batch not seen during the training phase, as the AUC values for each of its rounds are obtained from a single batch. To estimate inter-batch AUC on future data, we employed what we call leave-two-batches-out cross-validation (LTBO-CV). In each of its rounds, two batches are simultaneously left out from training, so that the members of the positive class (TYK-nu-CP.r) are from the first held-out batch and the members of the negative class (TYK-nu) are taken from the second, as shown in Figure 2C. The results from each possible pair of positive–negative batches are averaged to form the overall CV estimate. In addition to the two CV pipelines described for AUC estimation, we also applied LOCO-CV as a comparative baseline where samples from the same batch can appear in both the training and test sets, Figure 2A.

The average line map spectra for each Raman cell were classified using three supervised learning algorithms (PLS-DA, linear SVM and logistic regression), and model performance was evaluated with the three different CV strategies described above. Since hyperparameter tuning was not the focus of this study, default settings were used where applicable, including for logistic regression and SVM. However, the default hyperparameters performed poorly for PLS-DA, so nested CV was employed to select an appropriate number of components (see Supporting Information for detailed description). The performance of the

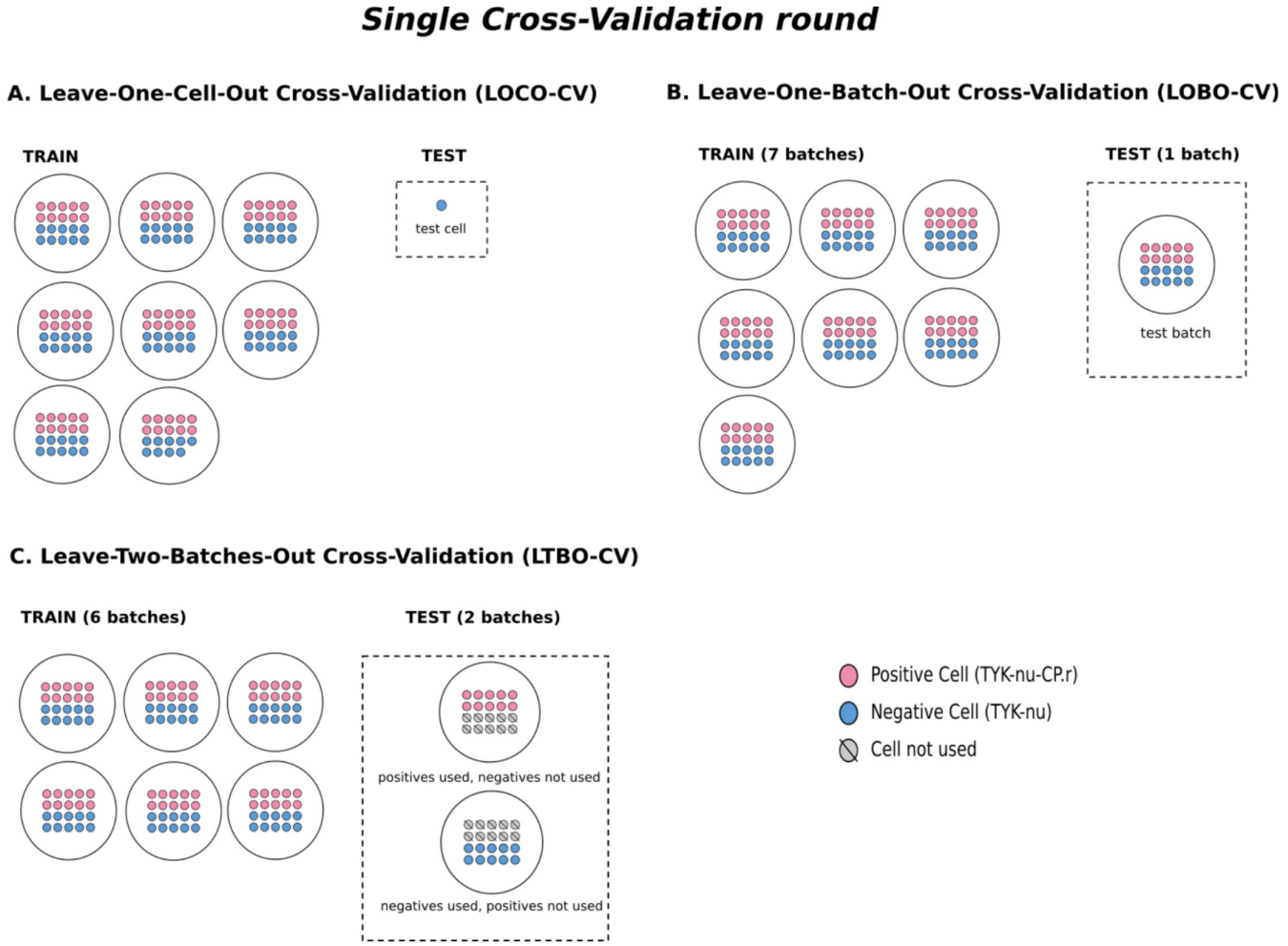

FIGURE 2 | Overview of the three cross-validation strategies used in the study. Each of the Sections A, B and C illustrates a single CV round: (A) leave-one-cell-out, where one individual cell is held out for testing in each round; (B) leave-one-batch-out, where all cells from one batch (day) are used for test in each round; (C) leave-two-batches-out, where two batches (days) are excluded, and the test set contains only the positive cells from one batch and the negative samples from the other (for every left-out batch pair, two rounds are built so that each batch contributes once as the source of positives and once as the source of negatives).

three classifiers was rather similar, as indicated by the AUCs in Figure 3 and the accuracies in Figure 4.

Out of the three CV strategies, LOCO-CV showed the highest classification accuracy of 88% (PLS-DA) (Figure 4A). This CV strategy is similar to that used by Moradi et al. [17], who applied Raman microscopy and PCA-LDA to distinguish between chemosensitive and chemoresistant endometrioid ovarian adenocarcinoma cell lines (A2780s vs. A2780cp) and reported a classification accuracy of 82.5%. The goal of our CV was to assess the generalisability of the models to predict unknown cell samples across different passages (biological replicates) and measurement days (batches). Therefore, we consider the LOBO-CV as the primary evaluation method. Using LOBO-CV, we obtained AUCs ranging from 92%–94% and classification accuracies of 75%–78% (Figures 3A and 4A).

To further analyse the changes that batch effects can cause on CV-based prediction performance estimates, batch correction was applied as an initial preprocessing step prior to model training. The PyCombat algorithm was applied, which adjusts the day-specific mean and variance shifts while preserving biological variability through empirical Bayes shrinkage, ensuring that model performance reflected true class separation rather than batch effects. Batch correction improved classification accuracies (Figure 4B) and LTBO-CV AUC (Figure 3B). For the PLS-DA model, LOBO-CV classification accuracy increased to 84% following batch correction.

The improvement in classification after batch correction suggests that day-to-day variation was present in the data, which, uncorrected, could mask the true biochemical signal and reduce generalisability. After batch correction, the performance can improve because the data becomes more comparable across measurement days. In our case, accuracy benefits when correcting day-wise variation reduced threshold-related misclassifications, whereas inter-day AUC improved when the underlying ordering between positive and negative samples became more consistent across days because, with LTBO-CV, positive cells from one batch are compared with negative cells from another batch.

All samples were prepared and measured by a single operator within 2 weeks, with no optical realignments, grating changes, or other system modifications. Daily wavenumber calibration was performed using the Si 520 cm^{-1} peak by fine-tuning the

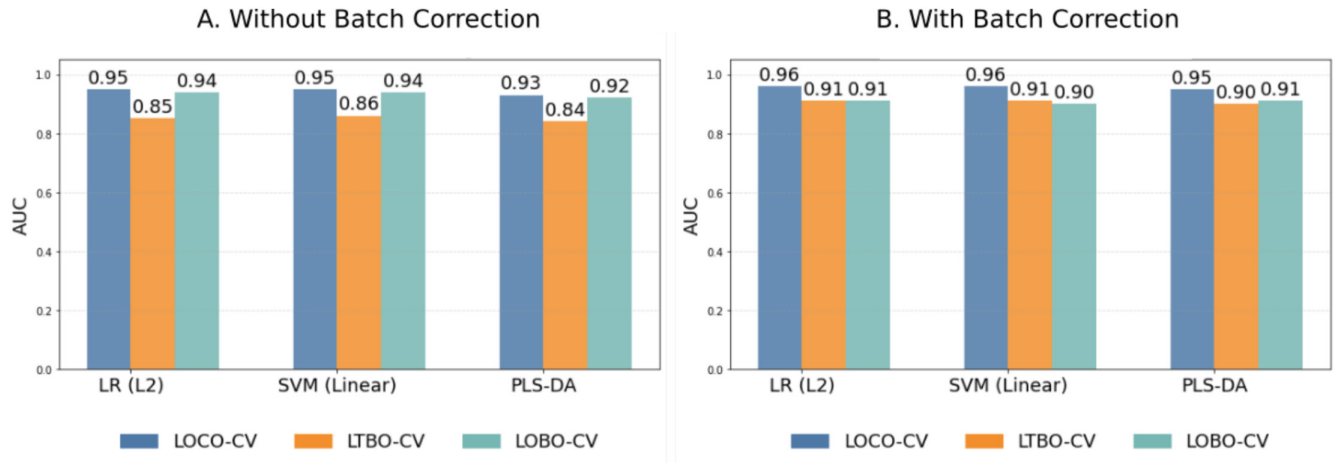

FIGURE 3 | AUCs for leave-one-cell-out (LOCO-CV), leave-two-batches-out (LTBO-CV) and leave-one-batch-out (LOBO-CV), across three classifiers (logistic regression [LR], SVM and PLS-DA) with Column (A) without batch correction and Column (B) with batch correction.

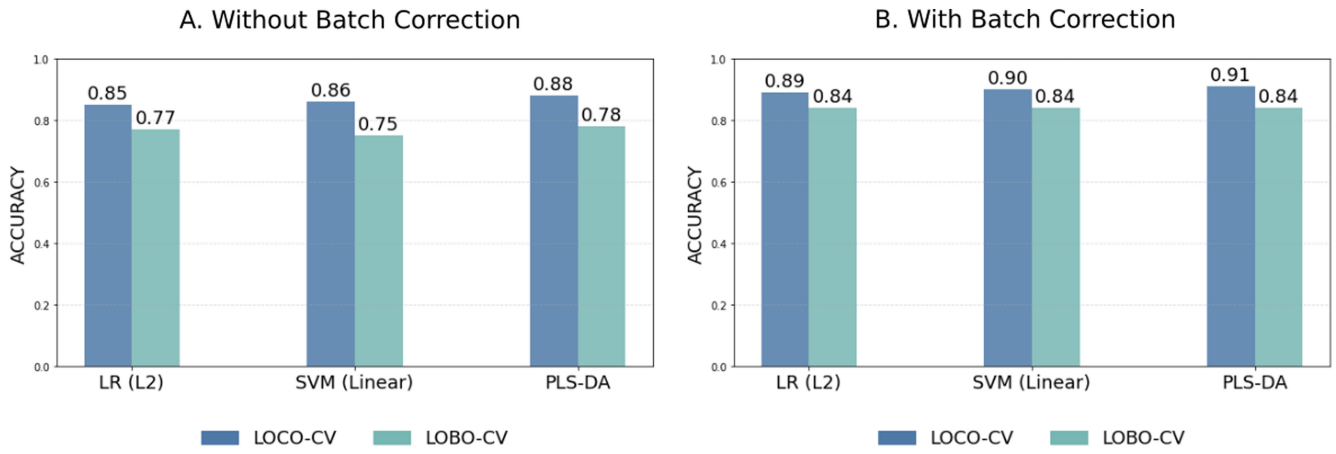

FIGURE 4 | Classification accuracies for leave-one-cell-out (LOCO-CV) and leave-one-batch-out (LOBO-CV), across three classifiers (logistic regression [LR], SVM and PLS-DA) with Column (A) without batch correction and Column (B) with batch correction.

offset, assuming linear grating dispersion. A more rigorous calibration using a standard covering the full spectral range (fingerprint and CH regions) could have improved precision. However, even careful wavenumber or intensity calibration cannot fully remove time-dependent spectral variation. Part of the apparent day-to-day variation may also reflect biological differences between replicates or small technical inconsistencies in their handling. Although statistical batch correction methods are rarely used in vibrational spectroscopy, their application was appropriate here because measurement days were independent of the biological classes, allowing removal of unwanted day effects without affecting class-related variation, and demonstrating the potential value of such approaches in this field.

3.3 | Feature Importance Analysis for Classification of Average Line Map Spectra

To explore the biochemical basis underlying the classification, we analysed the spectral features that contributed most to the classification. In our case, the spectral differences between classes were not obvious upon visual inspection of their average spectra (Figure 5). Therefore, we analysed the top 25 sets of features that contributed most significantly to class separation, as determined by the classifier coefficients. To do this, we trained the classifiers on the entire dataset and identified the features with the highest coefficients. Also, to assess the variability and stability of feature selection across different CV rounds in the LOBO-CV, we averaged the coefficients (feature importance scores) obtained from each round and chose the features that had been selected in at least 80% of the CV rounds. This approach allowed us to highlight features that were consistently important for classification across all CV rounds. Among the classifiers, PLS-DA produced the largest number of consistently selected features. The results for all three classifiers are presented in Table S2 and Figure S4.

All classifiers yielded quite similar patterns in feature importance. In all three classifiers, the Raman shifts 842, 1069–1079, 1146, 1408–1425, 1708 and 3011 cm^{-1} were among the top 25 ranked features and showed high consistency across CV rounds (Figure 5). The Raman shifts at 842, 1146 and 1708 cm^{-1} were positively associated with TYK-nu-CP.r. Based on previous assignments, 842 and 1146 cm^{-1} are tentatively attributed to glucose [43, 44], whereas the feature at 1708 cm^{-1} corresponds to the shoulder of the Amide I band [45]. The region at 1408–1425 cm^{-1}, which was positively associated with TYK-nu, is typically associated with CH_2/CH_3 bending modes, common in lipids and proteins [46]. One of the most consistently selected and highly ranked spectral features was 3011 cm^{-1}, which showed a strong positive association with TYK-nu, especially in the logistic regression and SVM classification models. This band is commonly attributed to = C–H stretching vibration of unsaturated lipids [47], suggesting increased lipid unsaturation in TYK-nu as a key differentiator between the two cell lines. Additionally, features in the range of 1069–1079 cm^{-1} were associated with TYK-nu. This region is tentatively attributed to C–C stretching of lipids [47]. As each spectral peak originates from vibrational modes of specific functional groups, no individual peak can be uniquely assigned to a single biochemical compound, and therefore, our

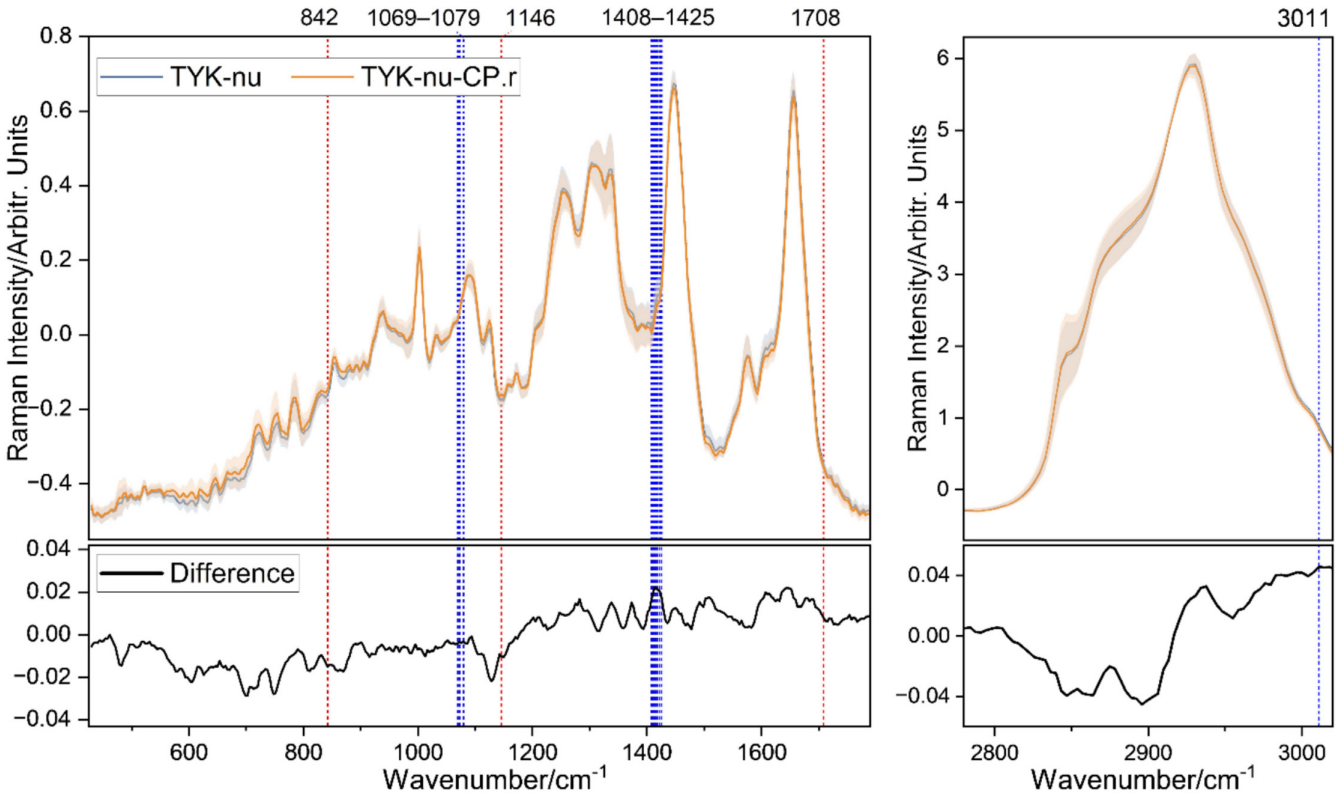

FIGURE 5 | Fingerprint and CH stretching region average spectra of each cell line ± standard deviation, together with the corresponding difference spectrum (TYK-nu – TYK-nu-CP.r). The vertical lines indicate the important wavenumbers contributing to the separation in all three classifiers. Red lines represent Raman shifts positively associated with TYK-nu-CP.r, whereas blue lines represent those positively associated with TYK-nu.

tentative peak assignments represent the most plausible major contributors to these spectral features.

An earlier Raman microscopy study reported elevated GSH levels in chemoresistant ovarian cancer cells [17]. GSH (both the reduced and oxidised forms) in aqueous solution exhibits relatively strong Raman bands at approximately 920 and 1410 cm^{-1}, although overall GSH is not a particularly strong Raman scatterer and it does not benefit from resonance enhancement with 532-nm excitation [48]. Reduced GSH is characterised by a sharp Raman band at 513 cm^{-1} and oxidised GSH approximately at 2570 cm^{-1} [48, 49]. In our analysis, the Raman shifts at 1408–1425 cm^{-1} emerged as one of the most highly ranked and consistently selected features in all the classifiers, showing a strong positive association with TYK-nu. Additionally, the bands at 513 and 918 cm^{-1} were also positively associated with TYK-nu, although they were not highly ranked. Moreover, when the classification region was extended to include the area beyond 2550 cm^{-1}, the feature at 2570 cm^{-1} exhibited a relatively weak association with TYK-nu (results not shown). Overall, our findings suggest that elevated GSH levels are unexpectedly associated with the chemosensitive TYK-nu rather than its chemoresistant subline. To further investigate this surprising result, we performed a GSH well-plate assay, which revealed that the chemosensitive subline contained approximately 4.6 times more GSH than the chemoresistant counterpart (Figure S5).

Substantial experimental evidence supports a functional role for GSH in mediating chemoresistance, particularly to platinum-based agents [50, 51]. Numerous studies have associated elevated intracellular GSH levels with platinum resistance, and it has been proposed that GSH can directly conjugate cisplatin, neutralising its cytotoxic effects (among other mechanisms) [52–55]. The mechanistic plausibility of this model is strong and well demonstrated in cell-based systems, with several studies demonstrating that a reduction in the levels of GSH can sensitise ovarian cancer cells to cisplatin [51, 56]. However, the relationship between intracellular GSH content and platinum resistance remains controversial. For example, although Mistry et al. reported a strong positive correlation ($r = 0.91$) between cisplatin IC_{50} values and intracellular GSH levels across eight ovarian cancer cell lines, closer inspection of the plotted data suggests that the relationship is non-uniform and may be weak or even inverse at low IC_{50} values [56]. What is particularly relevant to our findings is the work of Criscuolo et al., who reported higher GSH levels in chemosensitive ovarian cancer cells compared with resistant counterparts [57]. They hypothesised that cisplatin-GSH adduct formation contributes to cisplatin cytotoxicity, and that platinum-resistant cells evade this pathway by downregulating GSH while upregulating compensatory antioxidant mechanisms.

Moreover, there is evidence suggesting that dynamic changes in GSH during cisplatin exposure may provide more insights into chemoresistance than baseline GSH alone. For example, Jamali et al. observed no significant difference in baseline GSH levels between cisplatin-sensitive and resistant A2780 cells but noted a time-dependent increase in GSH in the resistant variant

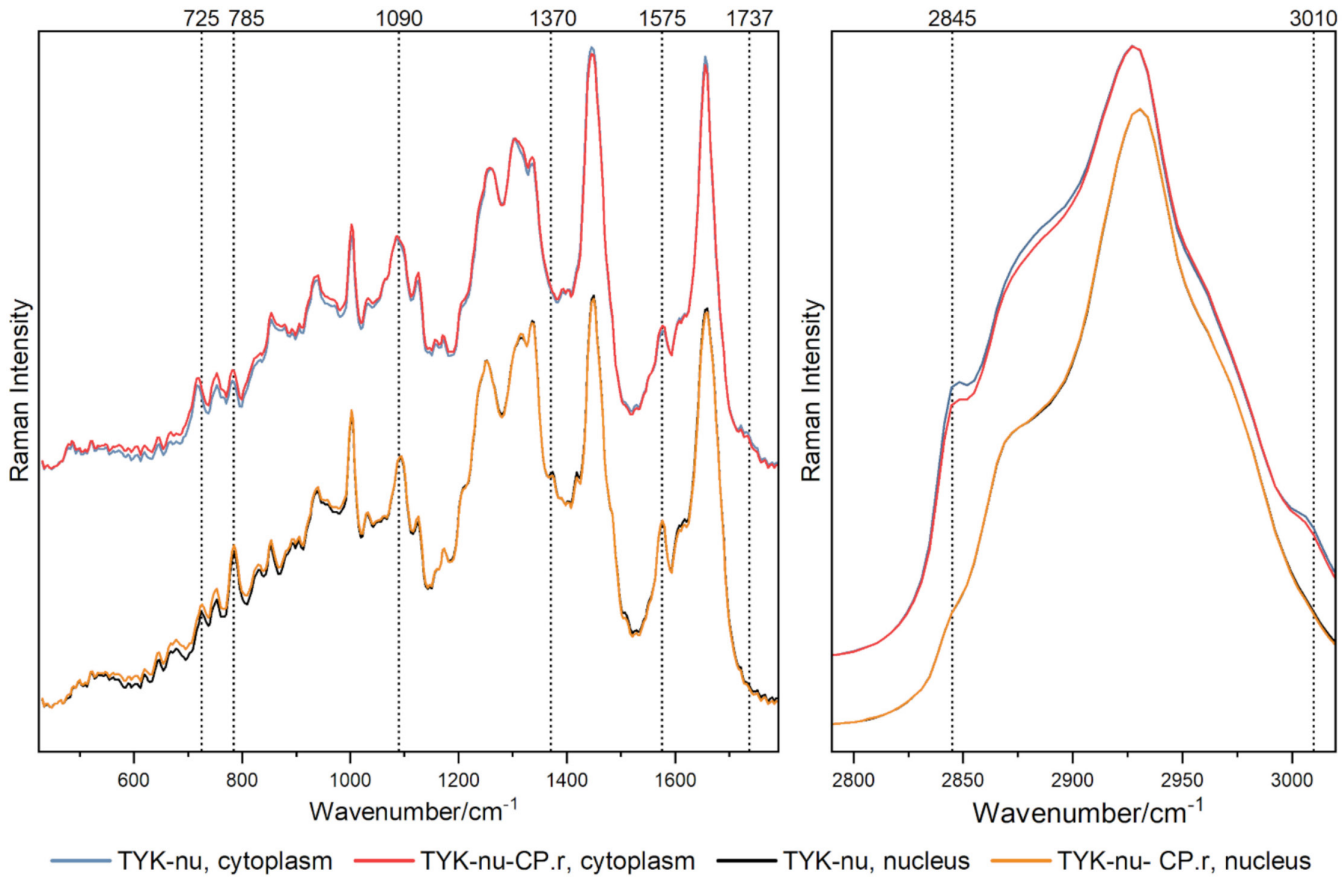

FIGURE 6 | Average Raman spectra of nucleus and cytoplasm of TYK-nu and TYK-nu-CP.r cells in the fingerprint and CH stretching regions. Spectra are offset for visual clarity. Vertical lines indicate wavenumbers corresponding to molecular vibrations associated with nucleic acids or lipids.

upon cisplatin exposure [58]. Taken together, our findings further highlight that elevated intracellular baseline GSH is not a universal feature of platinum resistance in ovarian cancer cells. Instead, the role of GSH may be cell line-dependent, with the dynamic changes in GSH levels (and the ratio of the oxidised and reduced forms) upon cisplatin exposure providing more comprehensive insights than baseline levels alone.

3.4 | Classification Performance on Individual Spectra

To understand the subcellular basis for the separation between the average spectra of the two cell lines, all individual Raman spectra were classified into either nucleus or cytoplasm clusters, and then, classification on individual point spectra was performed. To build a subcellular classification model applicable to both cell lines, a reference data set was collected. Due to difficulties in visualising cells on CaF_2, these reference data were collected from cells seeded on glass, which provided better contrast under brightfield microscopy, allowing spectra to be collected from specific subcellular regions.

Line maps were collected either from the cytoplasm or the nucleus areas of the cells, as shown in Figure S6. Two samples per cell line were analysed, with a minimum of 10 cells per sample, resulting in 210 cytoplasm spectra and 200 nucleus spectra. A linear SVM classifier was trained and evaluated using leave-one-cell-out CV, achieving a classification accuracy of 94%. Since neither glass nor CaF_2 substrates produce background signals in the CH region, this region was used for the classification task. Before the final classification task, poor-quality spectra were excluded based on a signal-to-noise ratio threshold.

The average cytoplasmic and nuclear spectra for TYK-nu and TYK-nu-CP.r are shown in Figure 6. These spectra are consistent with previous studies distinguishing nuclear and cytoplasmic regions, with clear differences such as increased nucleic acid-associated bands at 725, 785, 1090, 1370 and 1575 cm^{-1} in the spectra of the nuclei and stronger lipid-associated bands at 1737, 2845 and 3010 cm^{-1} in spectra of the cytoplasmic regions [33, 34, 38, 46].

The classification accuracies with logistic regression using spectra from only the nucleus or cytoplasm were 65% and 70%, respectively (LOBO-CV). When using individual spectra encompassing both subcellular regions, the classification accuracy was 70%, indicating comparable performance. These accuracies are lower than the classification accuracy obtained using average line map spectra, likely due to the reduction of random noise in the line-map data achieved through spectral averaging. Moreover, it is important to note that both the actual data and

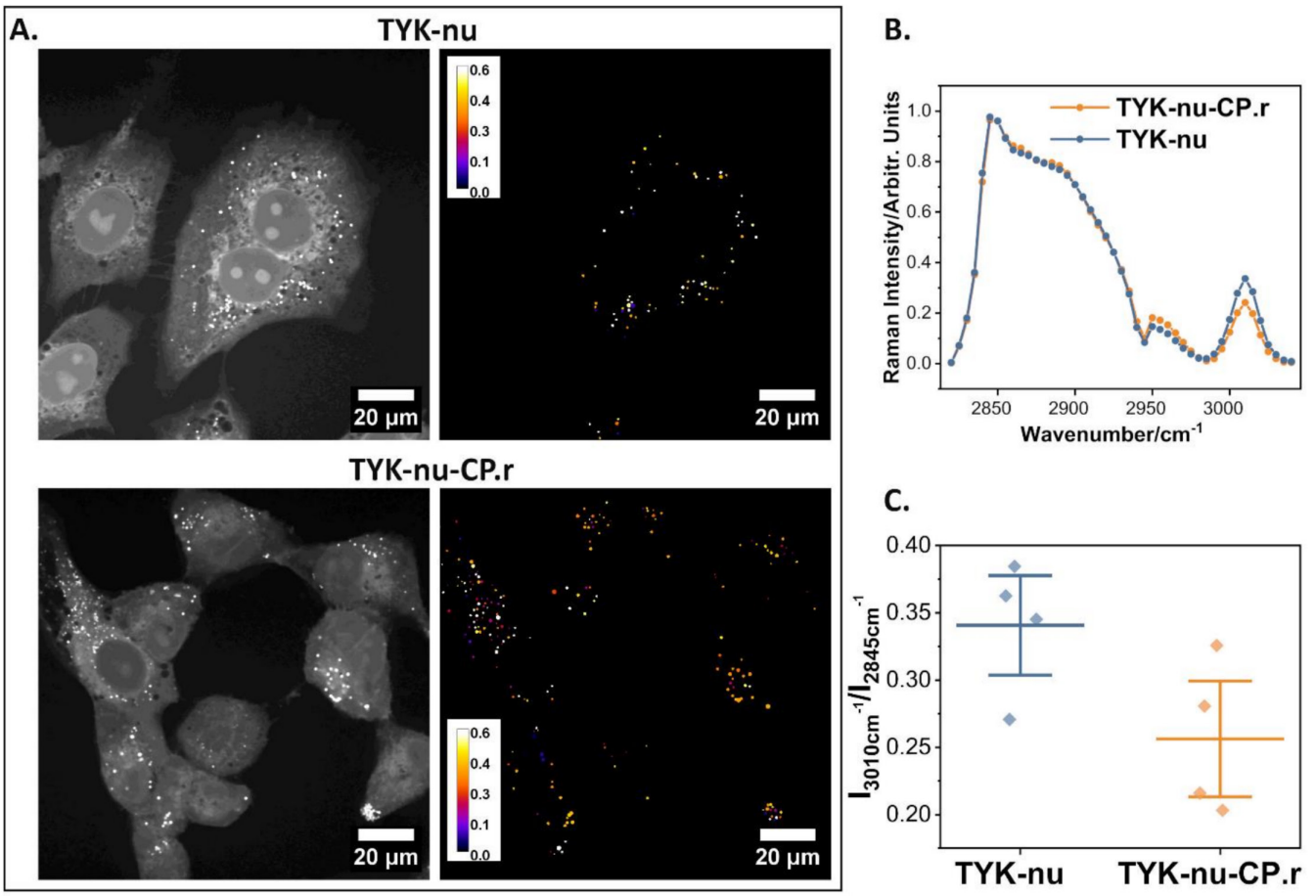

FIGURE 7 | SRS microscopy of LDs in TYK-nu and TYK-nu-CP.r cells. (A) Left: average SRS images of all spectral points; right: images showing lipid droplets coloured by the 3010/2845 cm^{-1} peak intensity ratio. (B) Average lipid droplet spectra. (C) Scatter plot of 3010/2845 cm^{-1} peak intensity ratios. Each dot represents a single replicate cell culture dish, and the bars indicate means ± standard error of the mean; $p = 0.0341$.

the reference data used to build the classifier were likely affected by some level of spectral mixing, as nuclei spectra can contain signals from the surrounding cytoplasm, which could have contributed to the lower classification performance.

Based on these results, we cannot confidently determine whether the cytoplasm or nucleus provides more biochemical information critical for classification. However, the feature importance analysis of the average line map analysis suggested unsaturated lipids as a potential key contributor to the separation between the two cell lines. Since cell membranes and cytoplasmic LDs are rich in unsaturated lipids, this may provide a plausible explanation for the slightly better classification performance observed with cytoplasmic spectra compared with nuclear spectra.

3.5 | Stimulated Raman Scattering Microscopy Analysis of Lipid Droplets

To test the hypothesis that TYK-nu cells have higher LD unsaturation, we performed SRS microscopy. SRS was employed here due to its ability to image LDs with high spatial resolution, especially in the z-dimension, when compared with confocal Raman microscopy based on spontaneous Raman scattering. Compared with spontaneous confocal Raman microscopy, SRS also enables faster acquisition of chemical maps, making it ideal for analysing large sample areas with numerous LDs. The CH stretching region (2820–3040 cm^{-1}) was chosen for SRS imaging due to its well-established use in LD analysis [12, 59]. Additionally, this region allowed us to directly investigate the = C–H stretching band at ~3010 cm^{-1}, which was one of the most consistently selected and highly ranked spectral features in the feature importance analysis of our confocal Raman microscopy data.

A total of 8548 LDs were segmented from 101 SRS images, and their spectra were extracted and analysed (Figure 7). The intensity ratio of the = C–H stretch (3010 cm^{-1}) to the CH_2 symmetric stretch (2845 cm^{-1}) in the resulting spectra was used as a relative measure of lipid unsaturation. The results showed a higher level of unsaturation ($n = 4$, $p = 0.0341$) in LDs from TYK-nu cells compared with TYK-nu-CP.r, as indicated by increased relative intensity of the = C–H stretching band. Interestingly, this result contrasts with a previous study that reported higher LD unsaturation in ovarian cancer stem cells (CSCs) compared with non-CSCs, where CSCs are generally associated with therapy resistance [59]. Additionally, differences in LD size distribution were observed between the two cell lines, with TYK-nu-CP.r cells appearing to have a higher proportion of larger LDs. Size distributions and peak intensity ratios for each size category are shown in Figure S7 in the Supporting Information.

4 | Conclusions

In this study, we have demonstrated that the TYK-nu HGSC cell line can be differentiated from its chemoresistant subline using confocal Raman microscopy combined with multivariate data analysis. This study emphasises the accounting for batch effects when estimating the classification performance with CV. In terms of AUC, the probability of predicting a larger positive class membership confidence score for the class members than for the non-members, this manifests as three separate statistical questions. Firstly, how well the classifier can distinguish a new member of the positive class from a new non-member when both are from the same batches from which the data used for training the classifier were obtained. Secondly, how well this can be done when they are from a new batch not seen during the training phase. Thirdly, how well the classifier distinguishes a new positive class member from a new batch from a new non-member from a different new batch.

In this study, a classification accuracy of 78% was achieved using average cell spectra without batch correction, whereas batch correction improved the accuracy to 84%. The analysis of individual point-spectra yielded an accuracy of 70%. This shows that cell lines could be distinguished moderately even at the subcellular level, though better results were obtained through averaging, likely due to noise reduction. Feature importance analysis of the average cell spectra suggested that the separation is contributed by elevated lipid unsaturation and GSH levels in TYK-nu cells. These results were further supported by colorimetric glutathione assay and SRS microscopy. These preliminary findings suggest that Raman microscopy holds promise as a label-free analytical tool for the prediction of chemoresistance in HGSC and guiding personalised treatment strategies, without the need for a specific biomarker.

Acknowledgements

The project has been financially supported by the Research Council of Finland (decision numbers 331837, 340182 and 345804). The Quantitative Chemically-Specific Imaging Infrastructure for Material and Life Sciences (qCSI), University of Helsinki, was utilised in this research. Tapio Pahikkala acknowledges the research environment provided by ELLIS Institute Finland. We thank Alexandra Correia, Andrej Tibenský and Perttu Mulari for helping with sample preparation. Open access publishing facilitated by Helsingin yliopisto, as part of the Wiley - FinELib agreement.

Funding

This work was supported by the Academy of Finland (331837, 340182 and 345804).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data and code may be obtained from the authors upon reasonable request.

References

1. E. Kozłowska, A. Färkkilä, T. Vallius, et al., "Mathematical Modeling Predicts Response to Chemotherapy and Drug Combinations in Ovarian Cancer," *Cancer Research* 78, no. 14 (2018): 4036–4044.

2. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer Statistics, 2018," *CA: A Cancer Journal for Clinicians* 68, no. 1 (2018): 7–30.

3. A. A. Wright, K. Bohlke, D. K. Armstrong, et al., "Neoadjuvant Chemotherapy for Newly Diagnosed, Advanced Ovarian Cancer: Society of Gynecologic Oncology and American Society of Clinical Oncology

Clinical Practice Guideline," *Journal of Clinical Oncology* 34, no. 28 (2016): 3460–3473.

4. L. Galluzzi, L. Senovilla, I. Vitale, et al., "Molecular Mechanisms of Cisplatin Resistance," *Oncogene* 31, no. 15 (2012): 1869–1883.

5. L. Jin, Y. Huo, Z. Zheng, et al., "Down-Regulation of Ras-Related Protein Rab 5C-Dependent Endocytosis and Glycolysis in Cisplatin-Resistant Ovarian Cancer Cell Lines," *Molecular & Cellular Proteomics* 13, no. 11 (2014): 3138–3151.

6. K. L. Lloyd, I. A. Cree, and R. S. Savage, "Prediction of Resistance to Chemotherapy in Ovarian Cancer: A Systematic Review," *BMC Cancer* 15 (2015): 1–32.

7. E. V. Nguyen, K. Huhtinen, Y. A. Goo, et al., "Hyper-Phosphorylation of Sequestosome-1 Distinguishes Resistance to Cisplatin in Patient Derived High Grade Serous Ovarian Cancer Cells," *Molecular & Cellular Proteomics* 16, no. 7 (2017): 1377–1392.

8. B. Davidson, "Biomarkers of Drug Resistance in Ovarian Cancer—An Update," *Expert Review of Molecular Diagnostics* 19, no. 6 (2019): 469–476.

9. R. F. Schwarz, C. K. Ng, S. L. Cooke, et al., "Spatial and Temporal Heterogeneity in High-Grade Serous Ovarian Cancer: A Phylogenetic Analysis," *PLoS Medicine* 12, no. 2 (2015): e1001789.

10. A. R. Laury, S. Zheng, N. Aho, et al., "Opening the Black Box: Spatial Transcriptomics and the Relevance of Artificial Intelligence–Detected Prognostic Regions in High-Grade Serous Carcinoma," *Modern Pathology* 37, no. 7 (2024): 100508.

11. Z. Farhane, H. Nawaz, F. Bonnier, and H. J. Byrne, "In Vitro Label-Free Screening of Chemotherapeutic Drugs Using Raman Microspectroscopy: Towards a New Paradigm of Spectralomics," *Journal of Biophotonics* 11, no. 3 (2018): e201700258.

12. Á. Szkalisity, L. Vanharanta, H. Saito, et al., "Nuclear Envelope-Associated Lipid Droplets Are Enriched in Cholesteryl Esters and Increase During Inflammatory Signaling," *EMBO Journal* 44, no. 10 (2025): 2774.

13. G. W. Auner, S. K. Koya, C. Huang, et al., "Applications of Raman Spectroscopy in Cancer Diagnosis," *Cancer and Metastasis Reviews* 37, no. 4 (2018): 691–717.

14. H. Nawaz, F. Bonnier, A. D. Meade, F. M. Lyng, and H. J. Byrne, "Comparison of Subcellular Responses for the Evaluation and Prediction of the Chemotherapeutic Response to Cisplatin in Lung Adenocarcinoma Using Raman Spectroscopy," *Analyst* 136, no. 12 (2011): 2450–2463.

15. H. Huang, H. Shi, S. Feng, et al., "Confocal Raman Spectroscopic Analysis of the Cytotoxic Response to Cisplatin in Nasopharyngeal Carcinoma Cells," *Analytical Methods* 5, no. 1 (2013): 260–266.

16. D. Franco, S. Trusso, E. Fazio, et al., "Raman Spectroscopy Differentiates Between Sensitive and Resistant Multiple Myeloma Cell Lines," *Spectrochimica Acta Part a: Molecular and Biomolecular Spectroscopy* 187 (2017): 15–22.

17. H. Moradi, A. Ahmad, D. Shepherdson, et al., "Raman Micro-Spectroscopy Applied to Treatment Resistant and Sensitive Human Ovarian Cancer Cells," *Journal of Biophotonics* 10, no. 10 (2017): 1327–1334.

18. M. K. Hammoud, H. K. Yosef, T. Lechtonen, et al., "Raman Micro-Spectroscopy Monitors Acquired Resistance to Targeted Cancer Therapy at the Cellular Level," *Scientific Reports* 8, no. 1 (2018): 15278.

19. G. Cutshaw, N. Hassan, S. Uthaman, et al., "Monitoring Metabolic Changes in Response to Chemotherapies in Cancer With Raman Spectroscopy and Metabolomics," *Analytical Chemistry* 95, no. 35 (2023): 13172–13184.

20. M. Kluz-Barłowska, T. Kluz, W. Paja, et al., "FT-Raman and FTIR Spectroscopy as a Tools Showing Marker of Platinum-Resistant Phenomena in Women Suffering From Ovarian Cancer," *Scientific Reports* 14, no. 1 (2024): 11025.

21. K. Honda, T. Hishiki, S. Yamamoto, et al., "On-Tissue Polysulfide Visualization by Surface-Enhanced Raman Spectroscopy Benefits Patients With Ovarian Cancer to Predict Post-Operative Chemosensitivity," *Redox Biology* 41 (2021): 101926.

22. Y.-J. Li, Y.-C. Sun, H.-R. Li, et al., "Machine Learning Assisted Raman Spectroscopy for the Classification of Ovarian Cancer Cells," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 351 (2026): 127465.

23. N. Blake, R. Gaifulina, L. D. Griffin, I. M. Bell, and G. M. Thomas, "Machine Learning of Raman Spectroscopy Data for Classifying Cancers: A Review of the Recent Literature," *Diagnostics* 12, no. 6 (2022): 1491.

24. D. A. Whitaker and K. Hayes, "A Simple Algorithm for Despiking Raman Spectra," *Chemometrics and Intelligent Laboratory Systems* 179 (2018): 82–84.

25. S. J. Barton and B. M. Hennelly, "An Algorithm for the Removal of Cosmic Ray Artifacts in Spectral Data Sets," *Applied Spectroscopy* 73, no. 8 (2019): 893–901.

26. E. Grau-Luque, F. Atlan, I. Becerril-Romero, A. Perez-Rodriguez, M. Guc, and V. Izquierdo-Roca, "Spectrapepper: A Python Toolbox for Advanced Analysis of Spectroscopic Data for Materials and Devices," *Journal of Open Source Software* 6, no. 67 (2021): 3781.

27. X. Lang, L. Shi, Z. Zhao, and W. Min, "Probing the Structure of Water in Individual Living Cells," *Nature Communications* 15, no. 1 (2024): 5271.

28. A. Behdenna, M. Colange, J. Haziza, et al., "pyComBat, a Python Tool for Batch Effects Correction in High-Throughput Molecular Data Using Empirical Bayes Methods," *BMC Bioinformatics* 24, no. 1 (2023): 459.

29. F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research* 12 (2011): 2825–2830.

30. S. Guo, R. Zhang, T. Wang, and J. Wang, "Comparative Study of Machine- and Deep-Learning Based Classification Algorithms for Biomedical Raman Spectroscopy (RS): Case Study of RS Based Pathogenic Microbe Identification," *Analytical Sciences* 40, no. 12 (2024): 2101–2109.

31. T. Tomberg, A. Isomäki, M. Vainio, M. Metsälä, J. Saarinen, and C. Strachan, "Multimodal Nonlinear Optical Microscope for Material and Life Sciences," *Optics Continuum* 3, no. 10 (2024): 1918–1937.

32. M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "A Nonlocal Transform-Domain Filter for Volumetric Data Denoising and Reconstruction," *IEEE Transactions on Image Processing* 22, no. 1 (2013): 119–133.

33. Z. Farhane, F. Bonnier, A. Casey, A. Maguire, L. O'Neill, and H. J. Byrne, "Cellular Discrimination Using In Vitro Raman Micro Spectroscopy: The Role of the Nucleolus," *Analyst* 140, no. 17 (2015): 5908–5919.

34. I. W. Schie, R. Kiselev, C. Krafft, and J. Popp, "Rapid Acquisition of Mean Raman Spectra of Eukaryotic Cells for a Robust Single Cell Classification," *Analyst* 141, no. 23 (2016): 6387–6395.

35. T. Meyer, N. Bergner, A. Medyukhina, et al., "Interpreting CARS Images of Tissue Within the C–H-Stretching Region," *Journal of Biophotonics* 5, no. 10 (2012): 729–733.

36. T. Tolstik, C. Marquardt, C. Matthäus, et al., "Discrimination and Classification of Liver Cancer Cells and Proliferation States by Raman Spectroscopic Imaging," *Analyst* 139, no. 22 (2014): 6036–6043.

37. I. W. Schie and J. W. Chan, "Estimation of Spectra Sample Size for Characterizing Single Cells Using Micro-Raman Spectroscopy," *Journal of Raman Spectroscopy* 47, no. 4 (2016): 384–390.

38. J. M. Surmacki, B. J. Woodhams, A. Haslehurst, B. A. Ponder, and S. E. Bohndiek, "Raman Micro-Spectroscopy for Accurate Identification of Primary Human Bronchial Epithelial Cells," *Scientific Reports* 8, no. 1 (2018): 12604.

39. M. P. Fay and Y. Malinovsky, "Confidence Intervals of the Mann-Whitney Parameter That Are Compatible With the Wilcoxon-Mann-Whitney Test," *Statistics in Medicine* 37, no. 27 (2018): 3991–4006.

40. A. Scherer, *Batch Effects and Noise in Microarray Experiments: Sources and Solutions*, (Wiley Online Library, 2009).

41. C. Soneson, S. Gerster, and M. Delorenzi, "Batch Effect Confounding Leads to Strong Bias in Performance Estimates Obtained by Cross-Validation," *PLoS ONE* 9, no. 6 (2014): e100335.

42. S. Guo, J. Popp, and T. Bocklitz, "Chemometric Analysis in Raman Spectroscopy From Experimental Design to Machine Learning–Based Modeling," *Nature Protocols* 16, no. 12 (2021): 5426–5459.

43. E. Wiercigroch, E. Szafraniec, K. Czamara, et al., "Raman and Infrared Spectroscopy of Carbohydrates: A Review," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 185 (2017): 317–335.

44. M. Mathlouthi and D. V. Luu, "Laser-Raman Spectra of d-Glucose and Sucrose in Aqueous Solution," *Carbohydrate Research* 81, no. 2 (1980): 203–212.

45. A. Rygula, K. Majzner, K. M. Marzec, A. Kaczor, M. Pilarczyk, and M. Baranska, "Raman Spectroscopy of Proteins: A Review," *Journal of Raman Spectroscopy* 44, no. 8 (2013): 1061–1076.

46. J. de Gelder, K. Gussem, P. Vandenabeele, and L. Moens, "Reference Database of Raman Spectra of Biological Molecules," *Journal of Raman Spectroscopy* 38, no. 9 (2007): 1133–1147.

47. K. Czamara, K. Majzner, M. Z. Pacia, K. Kochan, A. Kaczor, and e. M. Baranska, "Raman Spectroscopy of Lipids: A Review," *Journal of Raman Spectroscopy* 46, no. 1 (2015): 4–20.

48. A. Domagała, S. Buda, M. Baranska, and G. Zając, "Glutathione and Its Structural Modifications Recognized by Raman Optical Activity and Circularly Polarized Luminescence," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 324 (2025): 124995.

49. Z. Ke, Z. Yu, and Q. Huang, "Assessment of Damage of Glutathione by Glow Discharge Plasma at the Gas–Solution Interface Through Raman Spectroscopy," *Plasma Processes and Polymers* 10, no. 2 (2013): 181–188.

50. A. Bansal and M. C. Simon, "Glutathione Metabolism in Cancer Progression and Treatment Resistance," *Journal of Cell Biology* 217, no. 7 (2018): 2291–2298.

51. P. Alarcon-Zapata, A. J. Perez, K. Toledo-Oñate, et al., "Metabolomics Profiling and Chemoresistance Mechanisms in Ovarian Cancer Cell Lines: Implications for Targeting Glutathione Pathway," *Life Sciences* 333 (2023): 122166.

52. A. K. Godwin, A. Meister, P. J. O'Dwyer, C. S. Huang, T. C. Hamilton, and M. E. Anderson, "High Resistance to Cisplatin in Human Ovarian Cancer Cell Lines Is Associated With Marked Increase of Glutathione Synthesis," *Proceedings of the National Academy of Sciences* 89, no. 7 (1992): 3070–3074.

53. R. Sarkhosh-Inanlou, M. Molaparast, A. Mohammadzadeh, and V. Shafiei-Irannejad, "Sanguinarine Enhances Cisplatin Sensitivity via Glutathione Depletion in Cisplatin-Resistant Ovarian Cancer (A2780) Cells," *Chemical Biology & Drug Design* 95, no. 2 (2020): 215–223.

54. G. M. Lewandowicz, P. Britt, A. W. Elgie, et al., "Cellular Glutathione Content, In Vitro Chemoresponse, and the Effect of BSO Modulation in Samples Derived From Patients With Advanced Ovarian Cancer," *Gynecologic Oncology* 85, no. 2 (2002): 298–304.

55. Z.-G. Zheng, H. Xu, S.-S. Suo, et al., "The Essential Role of H19 Contributing to Cisplatin Resistance by Regulating Glutathione Metabolism in High-Grade Serous Ovarian Cancer," *Scientific Reports* 6, no. 1 (2016): 26093.

56. P. Mistry, L. Kelland, G. Abel, S. Sidhar, and K. Harrap, "The Relationships Between Glutathione, Glutathione-S-Transferase and Cytotoxicity of Platinum Drugs and Melphalan in Eight Human Ovarian Carcinoma Cell Lines," *British Journal of Cancer* 64, no. 2 (1991): 215–220.

57. D. Criscuolo, R. Avolio, M. Parri, et al., "Decreased Levels of GSH Are Associated With Platinum Resistance in High-Grade Serous Ovarian Cancer," *Antioxidants* 11, no. 8 (2022): 1544.

58. B. Jamali, M. Nakhjavani, L. Hosseinzadeh, S. Amidi, N. Nikounezhad, and F. H. Shirazi, "Intracellular GSH Alterations and Its Relationship to Level of Resistance Following Exposure to Cisplatin in Cancer Cells," *Iranian Journal of Pharmaceutical Research: IJPR* 14, no. 2 (2015): 513.

59. J. Li, S. Condello, J. Thomes-Pepin, et al., "Lipid Desaturation Is a Metabolic Marker and Therapeutic Target of Ovarian Cancer Stem Cells," *Cell Stem Cell* 20, no. 3 (2017): 303–314.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** The cytotoxic responses of the TYK-nu and TYK-nu-CP.r ovarian cancer cell lines to varying concentrations of cisplatin in (a) University of Otago and (b) University of Helsinki. Viability is expressed as a percentage relative to the untreated control group, and the error bars represent the standard deviation of 3–4 replicates. **Table S1:** Number of cells analysed in each sample for each measurement batch for both cell lines. **Figure S2:** (a) Brightfield images of the same TYK-nu-CP.r cell (indicated by red rectangle) on calcium fluoride substrate immersed in PBS, imaged under two different lighting conditions (stage moved in between). In the centre of the field of view, the intracellular structures of the cells appeared low-contrast, but improved contrast was achievable by closing the field diaphragm, enhancing visibility in darker peripheral areas. (b) PCA analysis on the CH stretching spectral region (2800–3000 cm^{-1}) and (c) total integrated spectral intensity of hyperspectral Raman map of the TYK-nu-CP.r cell (background removed). Arrows indicate the pixels where spectra shown in d) were plotted. **Figure S3:** Microscopy images of TYK-nu and TYK-nu-CP.r cells cultured on calcium fluoride. Images were acquired using an Olympus IX53 inverted microscope equipped with an Olympus DP73 digital camera and Olympus TH4-200 transmitted light illumination (Olympus Corporation, Tokyo, Japan). Images were captured and analysed using cellSens Standard v1.8 software (Olympus Corporation, Tokyo, Japan). Red circles on the images mark examples of cells with 'spread out and irregular' morphology and blue circles mark cells considered as 'elongated and spindle-like.' In Raman analyses, both types of morphologies were sampled in roughly equal amounts. **Table S2:** Top 25 features ranked by descending coefficient magnitude for the partial least squares discriminant analysis (PLS-DA), support vector machine (SVM) and logistic regression (LR) classifiers with L2 regularisation. **Figure S4:** Top features selected at least 80% of the leave-one-batch-out CV, for three learning algorithms (logistic regression [LR], PLS-DA and SVM). Each box represents the distribution of the model coefficient for a single Raman shift across all folds. The middle line shows the median coefficient, the box shows the interquartile range, and the whiskers indicate the range of non-outlier values. Positive values correspond to features associated with TYKnuCPR and negative values with TYKnu. **Figure S5:** Relative glutathione concentrations in TYK-nu and TYK-nu-CP.r cells. Bars represent the mean and error bars indicate standard deviation ($n = 3$, $p < 0.0001$). **Figure S6:** Bright-field images showing representative line maps (10 spectra/line) collected from either the nucleus or cytoplasm regions of a cell seeded on glass. These data were used to classify spectra into nucleus and cytoplasm categories. **Figure S7:** (a) Size distributions of LDs and (b) average peak intensity ratios for each size category.