



Raman Spectroscopy Analysis of The Temporal Heterogeneity in Lung Cell Carcinogenesis Induced by Benzo(a)pyrene*

ZHOU Hai-Tao^{1)**,**}, YAO Wei^{1)**}, CUI Cao-Zhe¹⁾, ZHOU Xiao-Tong¹⁾,

LIANG Xi-Long³⁾, QIN Cheng-Bing²⁾, XIAO Lian-Tuan²⁾, WU Zhi-Fang^{1)**,**}, LI Si-Jin^{1)**,**}

¹⁾Collaborative Innovation Center of Molecular Imaging Precision Medical, Department of Nuclear Medicine, The First Hospital of Shanxi Medical University, Shanxi Medical University, Taiyuan 030001, China;

²⁾Collaborative Innovation Center of Extreme Optics, Shanxi University, Institute of Laser Spectroscopy, State Key Laboratory of Quantum Optics and Quantum Optics Devices, Taiyuan 030006, China;

³⁾Central Laboratory, Taiyuan Central Hospital of Shanxi Medical University, Taiyuan 030001, China)

Abstract Objective Temporal heterogeneity in lung cancer presents as fluctuations in the biological characteristics, genomic mutations, proliferation rates, and chemotherapeutic responses of tumor cells over time, posing a significant barrier to effective treatment. The complexity of this temporal variance, coupled with the spatial diversity of lung cancer, presents formidable challenges for research. This article will pave the way for new avenues in lung cancer research, aiding in a deeper understanding of the temporal heterogeneity of lung cancer, thereby enhancing the cure rate for lung cancer. **Methods** Raman spectroscopy emerges as a powerful tool for real-time surveillance of biomolecular composition changes in lung cancer at the cellular scale, thus shedding light on the disease's temporal heterogeneity. In our investigation, we harnessed Raman spectroscopic microscopy alongside multivariate statistical analysis to scrutinize the biomolecular alterations in human lung epithelial cells across various timeframes after benzo(a)pyrene exposure. **Results** Our findings indicated a temporal reduction in nucleic acids, lipids, proteins, and carotenoids, coinciding with a rise in glucose concentration. These patterns suggest that benzo(a)pyrene induces structural damage to the genetic material, accelerates lipid peroxidation, disrupts protein metabolism, curtails carotenoid production, and alters glucose metabolic pathways. Employing Raman spectroscopy enabled us to monitor the biomolecular dynamics within lung cancer cells in a real-time, non-invasive, and non-destructive manner, facilitating the elucidation of pivotal molecular features. **Conclusion** This research enhances the comprehension of lung cancer progression and supports the development of personalized therapeutic approaches, which may improve the clinical outcomes for patients.

Key words temporal heterogeneity, lung cancer, Raman spectroscopy, multivariate statistical analysis, benzo(a)pyrene exposure

DOI: 10.16476/j.pibb.2023.0447

Cancer, with lung cancer as the deadliest form, significantly impacts global health, reflected in a sobering five-year survival rate of merely 22%^[1-2]. Lung cancer accounts for more deaths annually than colon and liver cancers combined. It is predominantly linked to smoking, which is responsible for 81% of lung cancer fatalities, with second-hand smoke leading to an additional 3% of these deaths^[3-4]. Benzo(a)pyrene (B(a)P) from cigarette smoke, identified as the first carcinogen, is often used in toxicology as a surrogate for polycyclic aromatic hydrocarbons.

* This work was supported by grants from The National Natural Science Foundation of China (62005150, 81971655, 62127817, 62222509, U22A6008, U22A2091).

** These authors contributed equally to this work.

*** Corresponding author.

ZHOU Hai-Tao. Tel: 86-351-4639425, E-mail: zht@sxmu.edu.cn

WU Zhi-Fang. Tel: 86-351-4639425, E-mail: wuzhifang01@163.com

LI Si-Jin. Tel: 86-351-4639425, E-mail: lisjnm123@163.com

Received: November 23, 2023 Accepted: January 8, 2024

Recognized as a group I carcinogen by the International Agency for Research on Cancer, B(a)P's prolonged exposure is notorious for DNA damage and mutagenesis, initiating tumorigenesis^[5].

Late-stage diagnoses are common in lung cancer, leading to missed early therapeutic intervention windows and resulting in dismal prognoses and treatment outcomes^[6-8]. Such challenges are compounded by lung cancer's temporal heterogeneity—a variance in biological and genomic traits over time, complicating effective treatment strategies. This heterogeneity, particularly at the cellular level, often evades clinical detection, necessitating a technique for the collection of temporal molecular information.

Biopsy-based assessments of temporal heterogeneity in lung carcinogenesis are limited by the invasive nature of tissue sampling and potential misrepresentations of tumor biochemistry^[9-10]. Non-recurring biopsies fail to capture the evolving mutational landscape, which is essential for understanding carcinogenesis and developing tailored therapies. Raman spectroscopy, leveraging the principle of light scattering, facilitates the sensitive detection of biochemical shifts at a single-cell level without the need for labels^[11-12]. It offers a direct, non-destructive assessment of cellular macromolecules, such as nucleic acids, proteins, and lipids, which are crucial for studying molecular alterations during carcinogenesis^[13-15]. Hence, Raman spectroscopy is emerging as an indispensable tool for investigating the temporal heterogeneity in lung cancer.

To address gaps in current research, namely the lack of prolonged and dynamic biochemical monitoring, our study harnesses Raman spectroscopy for in-depth analysis of B(a)P-induced changes in human lung epithelial cells. Through multivariate analysis, we aim to decode the cellular and biochemical narratives of lung cancer progression, thereby offering fresh perspectives on cancer heterogeneity, the underpinnings of carcinogenesis, and paving the way for precise cancer diagnostics and therapeutics.

1 Materials and methods

1.1 Human lung epithelial cells

A human lung epithelial cell line, BEAS-2B, was purchased from Shanghai Zhong Qiao Xin Zhou Biotechnology Co., Ltd. (Shanghai, China). The cells

were cultured in Dulbecco's modified Eagle medium (DMEM) (Gibco, CA, USA), supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. The cell cultures were maintained in an incubator at 37°C, in 5% CO₂, and with 95% relative humidity. Cells were cultured in 25 cm² flasks and passaged every 3 d, with an inoculum of approximately 5×10⁵ cells.

1.2 Chemical carcinogen

B(a)P (Beijing Solarbio Science & Technology Co., Ltd., Beijing, China), as a yellow-green powder, was dissolved in dimethyl sulfoxide at a stock solution concentration of 2.5 g/L. Next, 10 µl of the stock solution was added to 5 ml of DMEM to yield a concentration of 5 mg/L.

1.3 Study groups

Experiment groups were defined as 70% of the area of cells growing against the wall of the culture flask. To these, we added B(a)P stock solution to yield the staining concentration of 5 mg/L. The mixture was homogenized and then added to the incubator; the cells were passaged and cultured after 24 h. When the area of cell growth against the wall again reached 70%, the same dose of B(a)P master mix was added to maintain the staining concentration of 5 mg/L, and the cells continued to pass through the culture after 24 h. After 5 cycles of B(a)P staining, we named the cells passages in rising order from “BBP1” to “BBP35”. Solvent control groups, in which the same volume of dimethyl sulfoxide as the B(a)P master mix was added to the medium to achieve a staining concentration of 2%, were stained in the same way as the B(a)P experimental groups, and the first passaged cells were recorded as BDP1. Untreated BEAS-2B cells were included as a blank control group.

1.4 Cell counting kit-8 (CCK-8)

We collected BEAS-2B, BBP5, BBP15, BBP25, BBP35, BDP5, BDP15, BDP25, and BDP35 cells at the logarithmic growth stage. All cells were seeded in 96-well plates at a density of 2 000 cells per 100 µl in each well. After 24 h and 72 h of cell adhesion, 10 µl of CCK-8 reagent (Dojindo, Kyushu Island, Japan) was added to each well, and the absorbance (*A*) values of the cells were obtained after 4 h. Using Equation (1) to calculate the mean 24 h and 72 h viability, the ratio of 72 h to 24 h viability was taken to represent the multiplicative index of 72 h relative to 24 h to assess the growth speed. The test was repeated

three times.

$$\text{Mean survival rate} = \frac{A_{\text{treated groups}} - A_{\text{blank groups}}}{A_{\text{control groups}} - A_{\text{blank groups}}} \quad (1)$$

$A_{\text{treated groups}}$: absorbance value of cells in wells with CCK-8 solution.

$A_{\text{control groups}}$: absorbance values of cells and medium only in the wells.

$A_{\text{blank groups}}$: absorbance values of only the medium in the wells.

1.5 Cell cloning experiments

We collected BEAS-2B, BBP5, BBP15, BBP25, BBP35, BDP5, BDP15, BDP25, and BDP35 cells at the logarithmic growth stage. All cells were inoculated in 6-well plates with 200 cells/2 ml per well and incubated for 2 weeks at 37°C in a cell incubator. Subsequently, the cells were treated with 4% paraformaldehyde and crystalline violet staining solution to count the number of cell clones. The ability of the cells to become tumor-forming *in vitro* was indicated by the clonogenic rate. The test was repeated three times.

$$\begin{aligned} & \text{Cell clone formation rate} \\ &= \frac{\text{Number of cell clones formed}}{\text{Number of cells inoculated}} \quad (2) \end{aligned}$$

1.6 Cell scratching experiments

We collected BEAS-2B, BBP5, BBP15, BBP25, BBP35, BDP5, BDP15, BDP25, and BDP35 cells at the logarithmic growth stage. All cells were seeded in 6-well plates at a density of 4×10^5 cells/2 ml in each well. Once the cells reached 90%–100% confluency, the cell monolayer was scratched using a 1 ml pipette tip. This was performed at 0, 12, 24, 36, and 48 h, and the migration capacity was expressed in terms of cell migration rate. The test was repeated three times.

$$\begin{aligned} & \text{Cell migration index} \\ &= \frac{\text{Initial scratch area} - \text{Corresponding time point area}}{\text{Initial scratch area}} \quad (3) \end{aligned}$$

1.7 Cell smears

We collected cell suspensions of BEAS-2B, BBP5, BBP15, BBP25, and BBP35 cells at the logarithmic growth stage, which were centrifuged at 300g for 5 min to remove the supernatant. The cell pellet was resuspended in 3 ml of PBS, and the mixture was centrifuged at 300g for 5 min to remove the supernatant. This was repeated twice. Finally, we resuspended the cells with 1 ml of PBS and

centrifuged the mixture at 600g for 10 min to remove the supernatant. A 10 μ l cell suspension was obtained and dropped onto a slide. We placed the slide with the cell suspension flat on the table and used coverslip tilted at 45° to slowly and evenly push the cell suspension away to form a cell smear.

1.8 Spectral data acquisition

Raman data on cell smears were measured using the MStarter 100 Microspectral Scanning Test System (Metatest Corporation, Nanjing, China). The device was equipped with a 532 nm laser (1 mW) that allows spectral scanning through a high-precision piezoelectric translation stage. Data on the cells in the smear were collected after calibration with a Raman peak of 520 cm^{-1} *in silico* before measurement. A microscope (Metatest Corporation, Nanjing, China) was set to 50 \times magnification with the laser focused on the surface of the cells; it was used for an acquisition time of 10 s. The spectral range was within 500–3 500 cm^{-1} ; data from 100 cells at three locations per cell were collected from different generations. The data were averaged to represent the spectral data for that generation of cells.

1.9 Statistical analysis

In our study, the data were analyzed using GraphPad Prism 9 (GraphPad Software, CA, USA). This software is specifically designed for scientific research and offers a wide range of analytical tools. The differences between groups were analyzed using the Student's *t*-test and one-way analysis of variance (ANOVA). This approach helps in determining whether there are any statistically significant differences between the means of these groups. In our analysis, we considered a *P* value (probability value) of less than 0.05 as the threshold for statistical significance. This means that if the *P* value calculated from the one-way ANOVA is below 0.05, we interpret it as a significant difference in the data among the groups we compared, thus providing credibility to our findings. The experiments were performed in triplicate, and the data were presented as mean \pm standard deviation (SD).

1.10 Raman data pre-processing

Spectra are easily influenced by fluorescence background, laser power fluctuations, sample variation, and other factors, making it necessary to eliminate interference for subsequent multivariate analysis. Particularly in biological samples, the

signals of natural biofluorescence clusters interfere with the Raman signal. Therefore, in this study, all raw Raman data were pre-processed by performing Savitzky-Golay filter smoothing, vector normalization, and baseline calibration using OriginPro 2021 (OriginLab, MA, USA). A Raman spectral range of 600–1 800 cm^{-1} was selected as it is where the biochemical information was most notably altered^[16].

1.11 Multivariate statistical analysis

In this study, principal component analysis (PCA) and linear discriminant analysis (LDA), two typical multivariate statistical methods, were employed (see MATLAB code IRootLab for processing details)^[17-18]. PCA determined the principal components (PCs) and extracted key features to facilitate data visualization. PCs were the eigenvectors of the data covariance matrix, and PC1 was generated based on the maximum variance of the data. The percentage of the variance of each PC explained the variance of data. The mathematical representation of PCA involves several linear algebra concepts, but at its core, the process can be represented by the equation:

$$PC_{\text{new}} = X_{\text{standardized}} \times V \quad (4)$$

PC_{new} is the matrix of the transformed data (principal components), $X_{\text{standardized}}$ is the standardized version of the original data, V is the matrix containing the eigenvectors (principal component directions). However, PCA was an unsupervised method of data dimensionality reduction that treated the entire dataset as a category and could only identify the overall variance of the dataset. PCA had a low ability to identify the dataset and made it difficult to analyze the differences in biological information among groups. Therefore, following PCA, the supervised dimensionality reduction method, LDA, was used. LDA enabled variance to be maximized for variables in different categories and minimized in the same type, forming tighter small clusters, and building better classification models.

2 Results and discussion

2.1 Cell characterization

We observed the cellular malignancy phenotype (Figure 1). As the number of cell generations increased, the morphology of the cells in the experimental groups treated with B(a)P gradually

changed, exhibiting blurred edges, poorly defined contours, observable small triangular shapes, and increased nucleoplasmic ratios, in contrast to the cells in the blank control and solvent control groups. The proliferation index of the cells in the experimental group gradually increased over 72 h; the clonogenic rate of the cells increased over two weeks; the migration index and the slope of the linear fit of the cells increased over 48 h. These characteristics were consistent with those of cancer cells, demonstrating the increasing degree of malignant carcinogenesis of the cells.

2.2 Raman spectra

After treatment, smear samples of the BEAS-2B cells and 5th, 15th, 25th, and 35th generation cells were prepared for Raman spectroscopy observations. The MStarter 100 Microspectral Scanning Test System (Figure 2a) was used as follows: a 532 nm laser interacted with the samples through a confocal scanning microscope to produce Raman scattered light. This was filtered using a Notch Filter to remove stray light and produce the Raman spectrum. Different cell sites vary owing to slight differences in biological structure resulting in slightly different Raman spectra. Therefore, we analyzed 3 selected locations on a single cell and collected data from 100 cells, allowing us to obtain an average Raman spectrum for each generation. Considering the average size of BEAS-2B cells ranges from 20 μm to 30 μm , we acknowledge that obtaining Raman spectra from only three locations within a cell might not comprehensively represent the entire cell. However, these locations were strategically chosen to include key cellular components, such as the nucleus and cytoplasm, to ensure that the spectra captured critical aspects of cellular biochemistry. The Raman spectra of the cells were obtained after data pre-processing (Figure 2b). The Raman spectra of the BEAS-2B cells and those of different generations after staining showed similar trends and peak spectral shapes. It was difficult to visually distinguish the differences between the different Raman spectra.

2.3 Multivariate statistical analysis

The combined PCA and LDA method offered marked advantages in dealing with multiple variables by enabling the maximum quantification of spectral differences among different categories and effectively extracting biomolecular information on the carcinogenic process of B(a)P. We first imported the

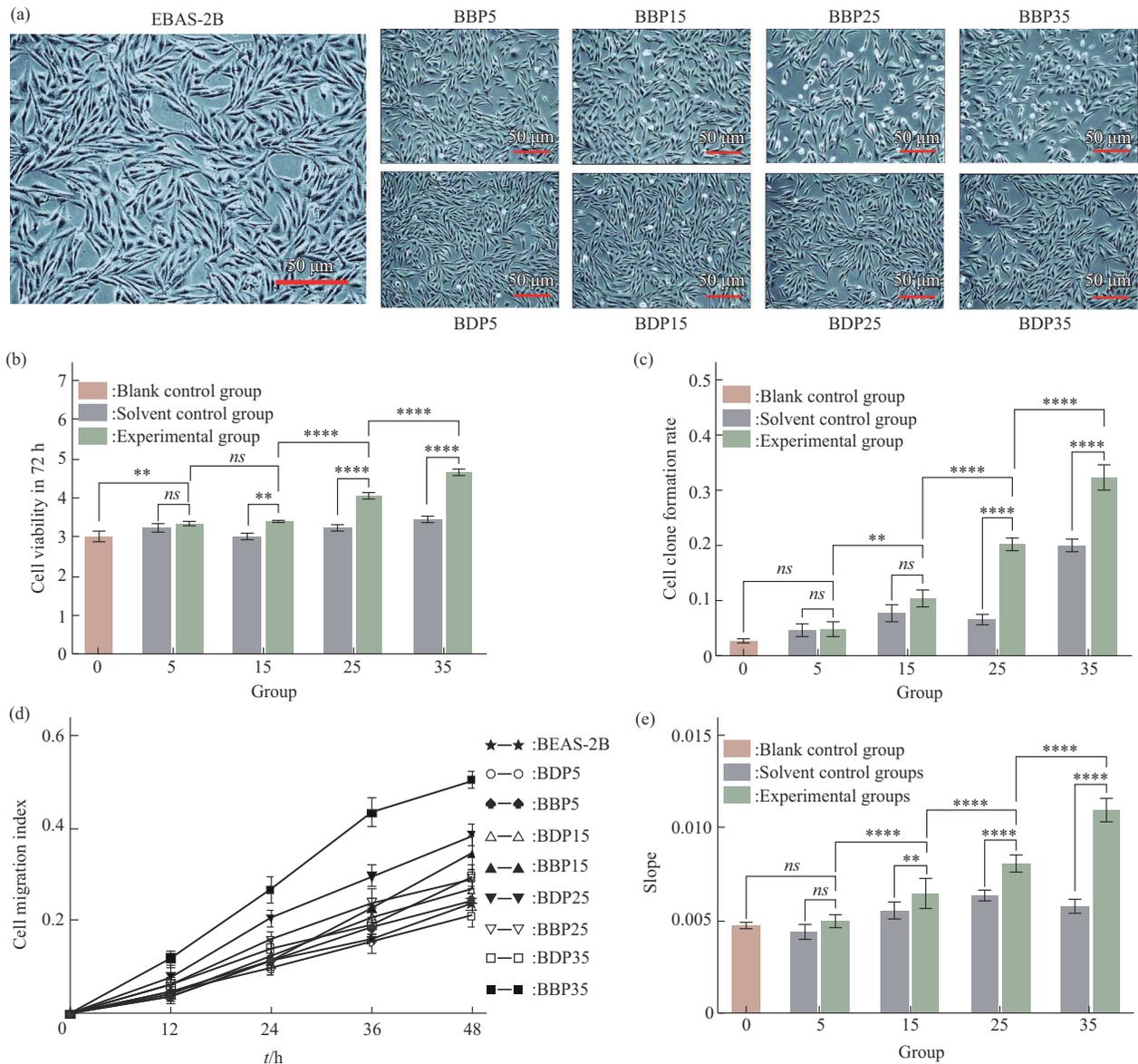


Fig. 1 Characterization of cell malignancy

(a) Cell shape; (b) cell proliferation index; (c) cell clonogenic rate; (d) cell migration index; (e) slope of the linear fit of cell scratch change. Experimental data are expressed as $mean \pm SD$, *ns* indicates no significance, ** $P < 0.01$, **** $P < 0.0001$.

pre-processed Raman spectral data into the MATLAB code IRootLab package and plotted the covariance matrix of BEAS-2B and BBP35 (Figure 3a). The red spots in Figure 3a indicate the main differential spectral regions for cells of different generations, including 1332 cm^{-1} , 1417 cm^{-1} , 1443 cm^{-1} , and 1608 cm^{-1} , suggesting that the age of lung cancer cells was related to structural alterations in nucleic acids, lipids, and proteins. Second, we used PCA on the Raman spectra and selected the 10 PCs with the highest LDA percentage (Figure 3b). The top 10 PCs contained more than 92% of the information of the

original variables and could replace the original spectral data. We used small icons of different shapes and colors to indicate the different categories to which they belonged, plotted the 2D scores for the first two PCs, and drew 95% confidence ellipses, which showed that the 2D score plot clustered as the number of cell generations increased after staining (Figure 3c). As the number of cell passages increased, the 95% confidence ellipse gradually decreased, and clustering became apparent. This is because normal cells, under the influence of B(a)P, gradually undergo DNA damage and genetic mutations, leading to the

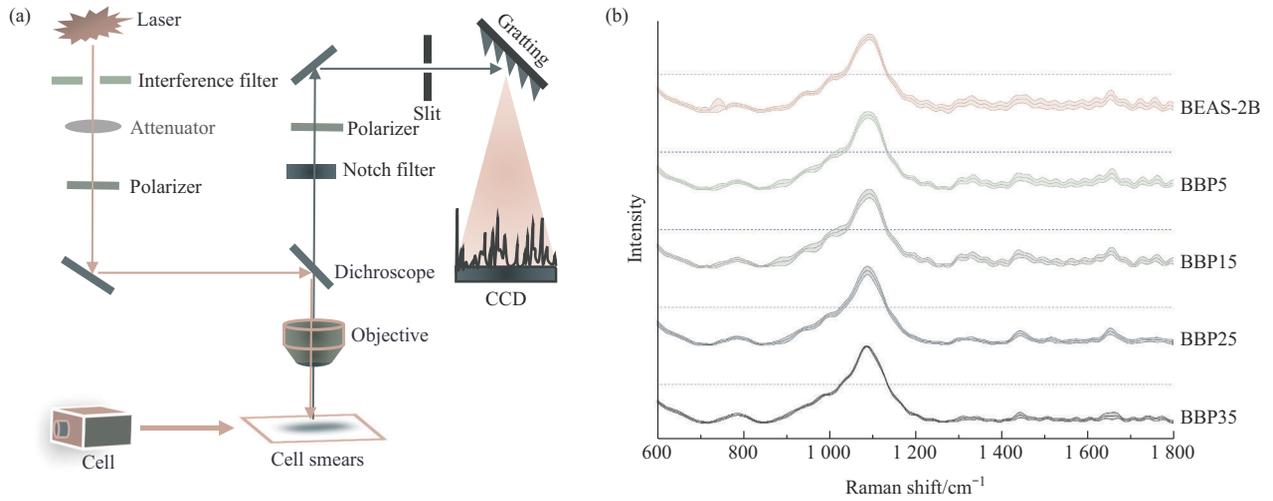


Fig. 2 Raman optical path diagram and spectrogram

(a) Raman optical path of the measured cell samples. (b) The average and standard deviation of Raman spectra of cells from BEAS-2B and after staining for the 5th (BBP5), 15th (BBP15), 25th (BBP25), and 35th (BBP35) generations.

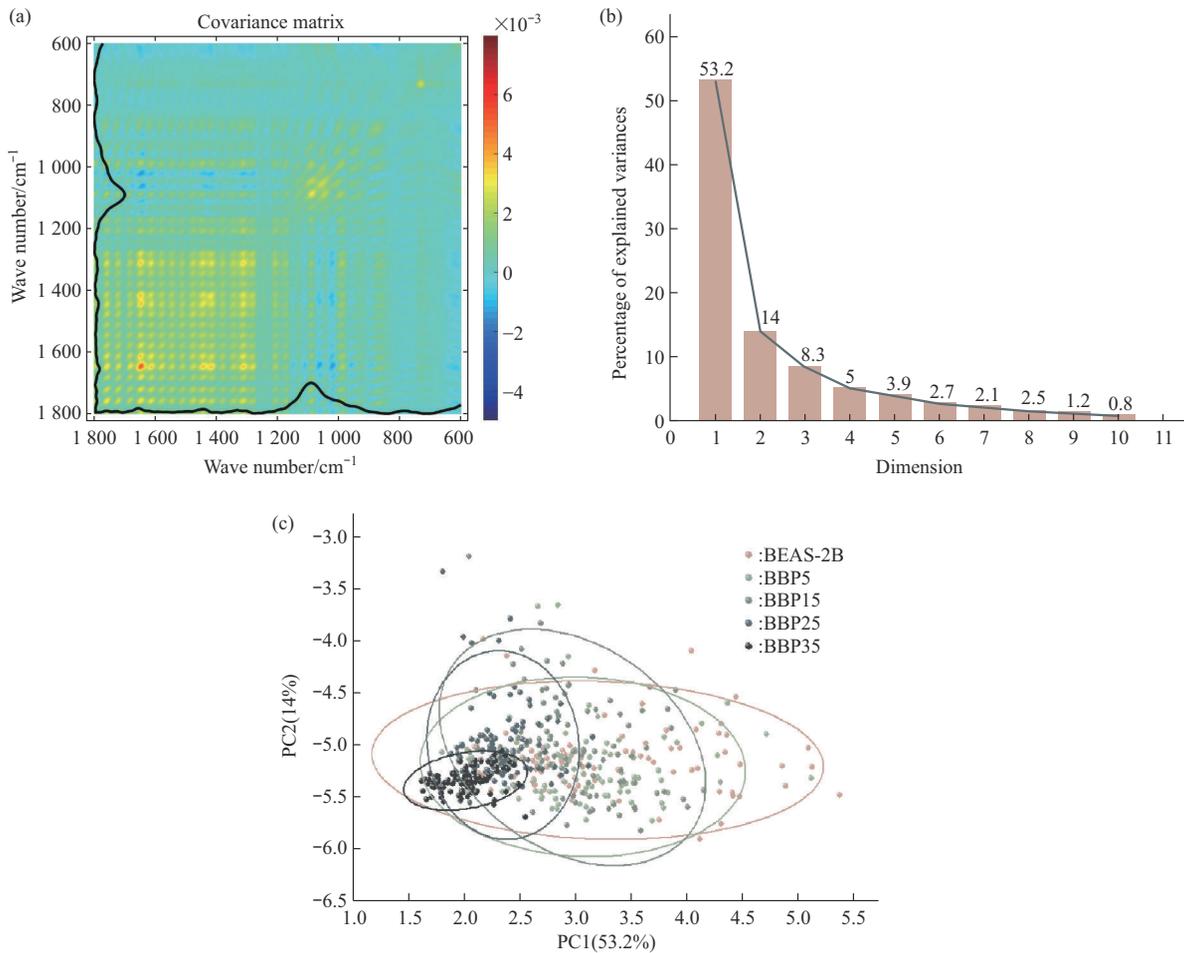


Fig. 3 The related maps of principal component analysis (PCA)

(a) Map of covariance matrices for BEAS-2B and post-staining 35 generations (BBP35). (b) Column plot of the percentage of the top ten principal components (PCs). (c) 2D PCA score plots for BEAS-2B and the 5th (BBP5), 15th (BBP15), 25th (BBP25), and 35th (BBP35) generations post-staining. PC1 vs. PC2 as axes, with 95% confidence ellipses outlined for each generation of cells.

selection of certain cells adapted to specific survival conditions, forming clusters where these cells may exhibit characteristics better suited for the tumor microenvironment.

We extracted the top 10 PCs for LDA. Linear discriminants (LDs) were new variables generated by LDA after the dimensionality reduction of the data, with LD1 making the largest contribution. We compared 1D score plots of LD1 for cells from different periods, and the plots show clustering similar to that in PCA (Figure 4a). The loadings in LD indicated the degree of inter-group variability, with larger absolute values indicating greater inter-group variability in the corresponding Raman spectra. We screened the top 25 waves with the highest load values in LD1 (Figure 4b) (Table 1 contains the

detailed wave numbers), and attributed them primarily to nuclear acids and lipids, and secondarily to proteins, carotenoids, and glucose. These 5 substances were also major cellular components. We selected the Raman peaks with the highest load values for these five substances. The larger loading values of $1\ 332\ \text{cm}^{-1}$ and $1\ 301\ \text{cm}^{-1}$ represented nucleic acids and lipids, the smaller loading values of $1\ 005\ \text{cm}^{-1}$ and $1\ 514\ \text{cm}^{-1}$ represented proteins and carotenoids, respectively. The Raman peak at $913\ \text{cm}^{-1}$ represents the C—C stretching and C—OH bending modes in glucose, while the peak at $1\ 125\ \text{cm}^{-1}$ signifies the C—O stretching mode of glucose. Thus, B(a)P had a greater effect on the structure of BEAS-2B nucleic acids and lipids, and a lesser effect on the structure of proteins, carotenoids, and glucose.

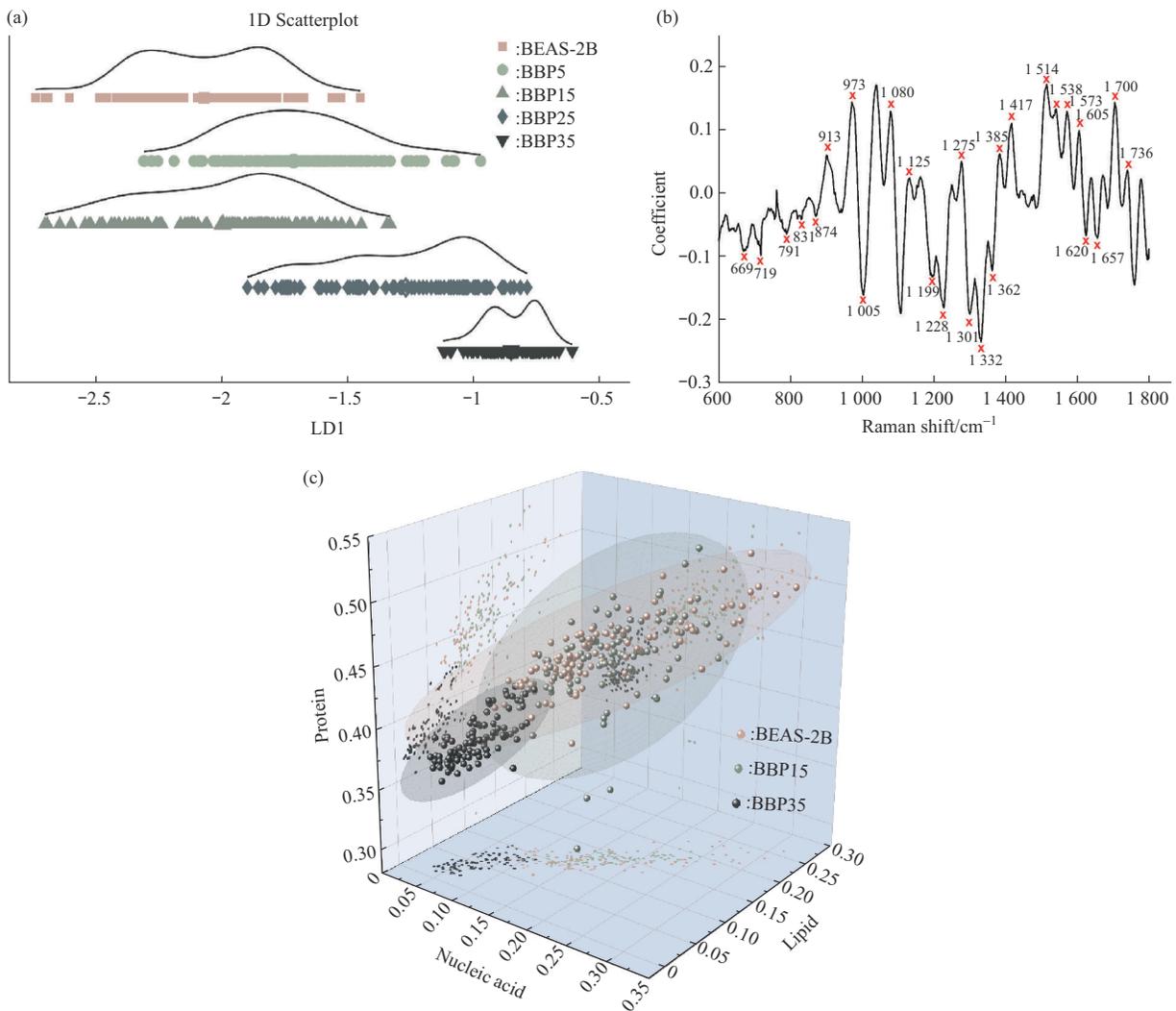


Fig. 4 Related linear discriminant analysis (LDA) maps

(a) 1D score plots of LD1 after LDA. (b) Loading plots corresponding to the score plots (x indicates the number of 25 characteristic Raman waves selected). (c) 3D scatter plots of BEAS-2B, and the 15th (BBP15) and 35th (BBP35) generations post-staining.

Table 1 Raman peak positions and mode description

Peak position/cm ⁻¹	Mode description
669	T, G, DNA/RNA
719	$\delta(\text{CCN}^+)$, lipids (choline)
791	$\rho(\text{NH}_3^+)$, lysine
831	O—P—O stretching, DNA
874	Ring breathing, proteins (tyrosine, proline), glycogen
913	C—C stretching, C—OH bending, glucose
973	$\rho(\text{CH}_3)$, $\delta(\text{CCH})$ olefinic, proteins
1 005	Symmetric stretching (ring breathing) mode, phenylalanine
1 080	C—N stretching, chain C—C stretching, C—O stretching, proteins, lipids, carbohydrates
1 125	C—O stretching, glucose
1 199	Tryptophan, phenylalanine
1 228	Asymmetric phosphate stretching
1 275	$\delta(\text{=CH})$, amide III
1 301	CH ₂ twisting, lipids
1 332	Nucleic acids
1 362	CH ₂ and CH ₃ symmetric deformation, proteins
1 385	CH ₂ and CH ₃ symmetric deformation, proteins
1 417	Nucleic acids
1 514	C=C stretching, carotenoid
1 538	C=C stretching, carotenoid
1 573	Phenylalanine, tyrosine
1 605	Tyrosine
1 620	NH ₂ bending, urea, tryptophan, phenylalanine, heme
1 657	Amide I, lipids (unsaturated fatty acid)
1 700	C=O stretching, amino acids aspartic, glutamic acid
1 736	C=O ester, lipids

PCA-LDA allows the screening of three distinctly different biomolecules; nucleic acids, lipids, and proteins, and the 3D scatter plots of BEAS-2B, BBP15, and BBP35 are plotted with the corresponding Raman intensities of the biomolecules, with 95% confidence ellipses outlined for each category, in Figure 4c. In the 3D scatterplot, the confidence ellipses for BEAS-2B, BBP15, and BBP35 had overlapping sections, with approximately 90% of BBP15 and 60% of BBP35 located in the confidence ellipse for BEAS-2B and approximately 40% of BBP35 located in the confidence ellipse of BBP15. BBP15 and BBP35 both originated from cells after B (a)P treatment of BEAS-2B, but their confidence ellipses did not completely overlap, reflecting temporal heterogeneity across cell generations after treatment.

2.4 Quantitative biomolecular analysis

Based on the Raman intensities corresponding to

the 5 wave numbers selected, we plotted the changes in the content of biomacromolecules with increasing cell generations (Figure 5). The nucleic acid and protein content did not change between the 5th and 15th generations after treatment and decreased in the 25th and 35th generations, with an overall decreasing trend. The content of lipids did not vary by the 5th generation but increased in the 15th generation and decreased in the 25th and 35th generations. Collectively, the trend was to increase and then decrease. The carotenoid content did not change in the 5th generation, decreased in the 15th generation, did not change in the 25th generation, and continued to decrease in the 35th generation, with an overall decreasing trend. The glucose content remained unchanged in the 5th, 15th, and 25th generations after toxicity, and increased in the 35th generation, with an upward trend. The ratios of protein to nucleic acid and protein to lipid did not change in the 5th and 15th

generations after the treatment, but increased in the 25th and 35th generations, showing an overall upward trend. The ratio of nucleic acid to lipid did not change between the 5th, 15th, and 25th generations but increased in the 35th generation, showing an overall

upward trend. As the number of cell passages increases, the content of various biomolecules varies over time, producing considerable variability and reflecting temporal heterogeneity in the lung cell carcinogenesis process.

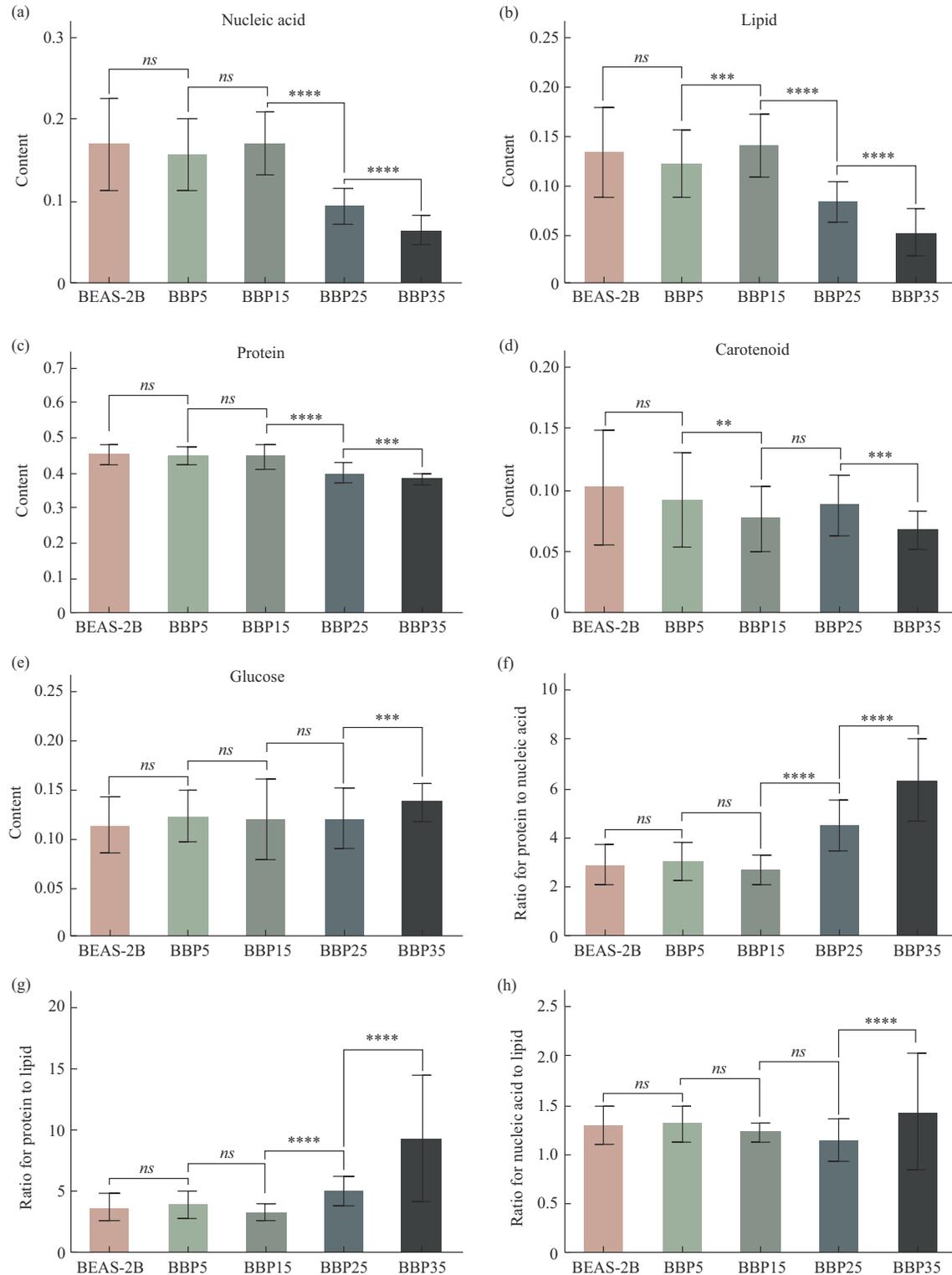


Fig. 5 Column chart for quantitative analysis of biological macromolecules

(a) Nucleic acid; (b) lipid; (c) protein; (d) carotenoid; (e) glucose; (f) protein to nucleic acid ratio; (g) protein to lipid ratio; (h) nucleic acid to lipid ratio. All data are expressed as *mean*±*SD*, *ns* indicates no significance, ***P*<0.01, ****P*<0.001, *****P*< 0.000 1.

Human lung epithelial BEAS-2B cells, as well as cells from the 5th, 15th, 25th, and 35th generations post-B(a)P treatment, were used to determine the degree of cellular carcinogenesis by characterizing their cell morphology and measuring their proliferative, invasive, and tumorigenic capacities. Raman spectroscopy data from BEAS-2B cells and the 5th, 15th, 25th, and 35th generations after B(a)P-induced carcinogenesis were obtained using the MStarter 100 Microspectral Scanning Test System. Data pre-processing and multivariate statistical analyses enabled the exploration of biomolecular changes arising from temporal heterogeneity during single-cell lung carcinogenesis. LDs were the new characteristic variables arising from PCA-LDA, with LD1 contributing the most to the differences between groups. LD1 presented distinct Raman wave numbers, primarily attributed to nucleic acids and lipids and secondarily to proteins, carotenoids, and glucose. The 3D scatter plot of nucleic acids, lipids, and proteins showed temporal heterogeneity throughout the BEAS-2B carcinogenesis process. Screening the content of the 5 biomacromolecules in relation to the degree of cell malignancy indicated that the content of these biomacromolecules varied at different stages during B(a)P-induced carcinogenesis. Moreover, metabolic dysfunctions promoted the formation of tumor cells.

B(a)P, a highly lipophilic chemical carcinogen, readily enters cells through the cell membrane and binds to aryl hydrocarbon receptor (AHR), such as cytochrome P450, thereby metabolically activating and damaging DNA molecules' structure, resulting in the formation of DNA adducts^[19-20]. Additionally, B(a)P can be metabolized to o-quinones by dehydrogenases or reductases, resulting in redox cycling and the formation of large amounts of reactive oxygen species, leading to oxidative DNA damage and elevated lipid peroxidation^[21]. Using Raman spectroscopy, we directly observed the structure of nucleic acids and lipid molecules and assessed their content over time. The content of nucleic acids and lipids showed a clear downward trend after the 15th generation. However, the lipid content increased at the 15th generation, probably related to the accelerated input of human fatty acid transporter protein 1 mediated by AHR signaling^[22]. Excess fatty acids can synthesize other biologically active metabolites; thus, the initial increase in lipid content is followed by a downward trend^[23-24]. After the 25th generation, there

was an upward trend in the nucleic acid to lipid ratio. The findings revealed that, after the 15th generation, the structures of nucleic acid and lipid molecules were severely disrupted, disrupting nucleotide and lipid metabolism. After the 25th generation, the nucleoplasmic ratio of the cells increased considerably, and the disturbance to lipid metabolism intensified. The changes in nucleic acid and lipid content were the most obvious, indicating that these are the main biomolecules affected by B(a)P, which constantly promotes the production of tumor cells.

B(a)P-induced carcinogenesis can change the structure of essential and non-essential amino acids. Phenylalanine, tryptophan, and lysine are essential amino acids. Phenylalanine can be metabolized by its hydroxylase into tyrosine, which is involved in glucose and lipid metabolism and is also a precursor for the antioxidant melanin^[25]. Kynurenine, a metabolite of tryptophan, is an endogenous agonist of AHR and influences carcinogenesis by B(a)P. Moreover, most tumor cells had high levels of tryptophanolytic enzymes, which are associated with low survival rates in patients with lung cancer^[26-27]. Lysine acetylation is one of the most studied modifications in tumor epigenetics, with lysine acetyltransferases playing a dual role in cancer development, acting as both oncogenic and pro-carcinogenic^[28-29]. Aspartic acid, glutamic acid, and proline are non-essential amino acids. Aspartic acid can be produced by asparagine metabolism, and the inhibition of asparagine synthesis reduces the survival and proliferation of tumor cells^[30]. Glutamate is essential for tumor cell survival, and its amide, glutamine, is a carbon and nitrogen source for synthesizing lipids and nucleotides^[31]. Proline is an intermediate in the urea cycle and tricarboxylic acid cycle. It could serve as a source of energy for tumor cells, and the overexpression of its metabolic enzyme (proline dehydrogenase) could promote cancer progression^[32-33]. In the case of phenylalanine, our Raman spectroscopy analysis observed a severe disruption to the molecular structure and decrease in content after the 15th generation; the protein to nucleic acid and protein to lipid ratios increased after the 15th generation, indicating a disturbance in protein metabolism at this generation. The content of protein molecules showed a decreasing trend, and metabolic disorders gradually deepened cellular malignancy.

Raman spectroscopy detected structural changes in carotenoids and glucose. Treatment of cells with B(a)P caused structural changes in carotenoids, characterized by C=C stretching vibrations. Carotenoids include carotenes and lutein, which are important sources of antioxidants and vitamin A, which are related to immune regulation and gene transcription^[34]. The content of carotenoids tended to decrease after the 5th and 25th generations, and the abnormal function of the carotenoids promoted carcinogenesis. Carotenoid content showed a decreasing trend, with an associated weakening of its antioxidant effect leading to a deepening of cellular malignancy. The content of glucose, one of the main sources of energy for the cells, tended to increase up to the 25th generation, with an increase in the metabolic capacity of the cells. It has been observed that increased cellular malignancy may occur with the reprogramming of glucose metabolism^[35-36].

3 Conclusion

We have demonstrated that B(a)P-induced carcinogenesis in human lung epithelial cells is a multi-stage, long-term, and complex process involving multiple biomolecular events and generating temporal heterogeneity. By using Raman spectroscopy combined with multivariate statistical analysis, we observed differences in the content of nucleic acids, lipids, proteins, carotenoids, and glucose at various stages post-treatment with B(a)P, which implied molecular changes related to the mechanism of lung cell carcinogenesis. As the number of cell generations increases following B(a)P treatment, the structure of genetic material becomes damaged, leading to nucleic acid instability, DNA damage, and the formation of DNA adducts. Cell metabolism was also altered, evidenced by lipid peroxidation, disruption of amino acid metabolism, inhibition of carotenoid secretion, and reprogramming of glucose metabolism. The cells continuously adapt to carcinogens, which results in increased carcinogenic capacity, malignancy, and heterogeneity, ultimately leading to low survival rates. This study enabled the real-time, non-invasive, and non-destructive observation of biomolecular changes in lung cell carcinogenesis *via* Raman spectroscopy and analysis of the characteristics of key molecules. The innovation of this study lies in the successful

construction of a lung cell carcinogenesis model using B(a)P, and the application of Raman spectroscopy technology in conjunction with multivariate statistical analysis to observe the temporal heterogeneity changes in lung cells during carcinogenesis. This approach provides a previously unexplored avenue for lung cancer research, facilitating in-depth exploration of temporal heterogeneity in lung cancer and offering a new direction for a more profound understanding and effective management of this complex disease.

References

- [1] Sung H, Ferlay J, Siegel R L, *et al.* Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 2021, **71**(3): 209-249
- [2] Leiter A, Veluswamy R R, Wisnivesky J P. The global burden of lung cancer: current status and future trends. *Nat Rev Clin Oncol*, 2023, **20**(9): 624-639
- [3] Islami F, Goding Sauer A, Miller K D, *et al.* Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the united states. *CA Cancer J Clin*, 2018, **68**(1): 31-54
- [4] Murray R L, O'dowd E. Smoking cessation and lung cancer: never too late to quit. *Lancet Public Health*, 2023, **8**(9): e664-e665
- [5] Porwisiak P, Werner M, Kryza M, *et al.* Modelling benzo (a) pyrene concentrations for different meteorological conditions-analysis of lung cancer cases and associated economic costs. *Environ Int*, 2023, **173**: 107863
- [6] Lim Z F, Ma P C. Emerging insights of tumor heterogeneity and drug resistance mechanisms in lung cancer targeted therapy. *J Hematol Oncol*, 2019, **12**(1): 134
- [7] Dagogo-Jack I, Shaw A T. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol*, 2018, **15**(2): 81-94
- [8] Megyesi Z, Gay C M, Popper H, *et al.* Clinical insights into small cell lung cancer: tumor heterogeneity, diagnosis, therapy, and future directions. *CA Cancer J Clin*, 2023, **73**(6): 620-652
- [9] Murugaesu N, Wilson G A, Birkbak N J, *et al.* Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov*, 2015, **5**(8): 821-831
- [10] Findlay J M, Castro-Giner F, Makino S, *et al.* Differential clonal evolution in oesophageal cancers in response to neo-adjuvant chemotherapy. *Nat Commun*, 2016, **7**: 11111
- [11] Wang X, Ren L, Diao Z, *et al.* Robust spontaneous raman flow cytometry for single-cell metabolic phenome profiling *via* pDEP-DLD-RFC. *Adv Sci*, 2023, **10**(16): e207497
- [12] Auner G W, Koya S K, Huang C, *et al.* Applications of raman spectroscopy in cancer diagnosis. *Cancer Metastasis Rev*, 2018, **37**(4): 691-717
- [13] Zheng Q, Li J, Yang L, *et al.* Raman spectroscopy as a potential diagnostic tool to analyse biochemical alterations in lung cancer. *Analyst*, 2020, **145**(2): 385-392

- [14] Traynor D, Behl I, O'dea D, *et al.* Raman spectral cytopathology for cancer diagnostic applications. *Nat Protoc*, 2021, **16**(7): 3716-3735
- [15] Wang J, Lin K, Hu H, *et al.* *In vitro* anticancer drug sensitivity sensing through single-cell Raman spectroscopy. *Biosensors (Basel)*, 2021, **11**(8): 286
- [16] Liu Y J, Kyne M, Wang C, *et al.* Data mining in raman imaging in a cellular biological system. *Comput Struct Biotechnol J*, 2020, **18**: 2920-2930
- [17] Trevisan J, Angelov P P, Scott A D, *et al.* Irootlab: a free and open-source matlab toolbox for vibrational biospectroscopy data analysis. *Bioinformatics*, 2013, **29**(8): 1095-1097
- [18] Martin F L, Kelly J G, Llabjani V, *et al.* Distinguishing cell types or populations based on the computational analysis of their infrared spectra. *Nat Protoc*, 2010, **5**(11): 1748-1760
- [19] Huang M, Blair I A, Penning T M. Identification of stable benzo[a]pyrene-7, 8-dione-DNA adducts in human lung cells. *Chem Res Toxicol*, 2013, **26**(5): 685-692
- [20] Kasala E R, Bodduluru L N, Barua C C, *et al.* Benzo(a)pyrene induced lung cancer: role of dietary phytochemicals in chemoprevention. *Pharmacol Rep*, 2015, **67**(5): 996-1009
- [21] Penning T M, Burczynski M E, Hung C F, *et al.* Dihydrodiol dehydrogenases and polycyclic aromatic hydrocarbon activation: generation of reactive and redox active o-quinones. *Chem Res Toxicol*, 1999, **12**(1): 1-18
- [22] Ye G, Gao H, Wang Z, *et al.* Pparalpha and ppargamma activation attenuates total free fatty acid and triglyceride accumulation in macrophages *via* the inhibition of fatp1 expression. *Cell Death Dis*, 2019, **10**(2): 39
- [23] Hankinson O. The role of ahr-inducible cytochrome p450s in metabolism of polyunsaturated fatty acids. *Drug Metab Rev*, 2016, **48**(3): 342-350
- [24] Yao L, Wang C, Zhang X, *et al.* Hyperhomocysteinemia activates the aryl hydrocarbon receptor/cd36 pathway to promote hepatic steatosis in mice. *Hepatology*, 2016, **64**(1): 92-105
- [25] Rzepka Z, Buszman E, Beberok A, *et al.* From tyrosine to melanin: Signaling pathways and factors regulating melanogenesis. *Postepy Hig Med Dosw (Online)*, 2016, **70**: 695-708
- [26] Ye G, Gao H, Wu Z, *et al.* Comprehensive metabolomics insights into benzo[a]pyrene-induced metabolic reprogramming related to h460 cell invasion and migration. *Sci Total Environ*, 2021, **774**: 145763
- [27] Lemos H, Huang L, Prendergast G C, *et al.* Immune control by amino acid catabolism during tumorigenesis and therapy. *Nat Rev Cancer*, 2019, **19**(3): 162-175
- [28] Subbaiah V K, Zhang Y, Rajagopalan D, *et al.* E3 ligase edd1/ubr5 is utilized by the hpv e6 oncogene to destabilize tumor suppressor tip60. *Oncogene*, 2016, **35**(16): 2062-2074
- [29] Tan K N, Avery V M, Carrasco-Pozo C. Metabolic roles of androgen receptor and tip60 in androgen-dependent prostate cancer. *Int J Mol Sci*, 2020, **21**(18): 6622
- [30] Deng L, Yao P, Li L, *et al.* P53-mediated control of aspartate-asparagine homeostasis dictates lkb1 activity and modulates cell survival. *Nat Commun*, 2020, **11**(1): 1755
- [31] Fernandez-De-Cossio-Diaz J, Vazquez A. Limits of aerobic metabolism in cancer cells. *Sci Rep*, 2017, **7**(1): 13488
- [32] Phang J M. Proline metabolism in cell regulation and cancer biology: recent advances and hypotheses. *Antioxid Redox Signal*, 2019, **30**(4): 635-649
- [33] Liu Y, Mao C, Wang M, *et al.* Cancer progression is mediated by proline catabolism in non-small cell lung cancer. *Oncogene*, 2020, **39**(11): 2358-2376
- [34] Eggersdorfer M, Wyss A. Carotenoids in human nutrition and health. *Arch Biochem Biophys*, 2018, **652**: 18-26
- [35] Li Z, Zhang H. Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. *Cell Mol Life Sci*, 2016, **73**(2): 377-392
- [36] Vaupel P, Schmidberger H, Mayer A. The warburg effect: essential part of metabolic reprogramming and central contributor to cancer progression. *Int J Radiat Biol*, 2019, **95**(7): 912-919

苯并(a)芘诱导的肺细胞癌变过程中时间异质性的拉曼光谱分析*

周海涛^{1)**,**} 姚伟^{1)**} 崔曹哲¹⁾ 周晓彤¹⁾ 梁喜龙³⁾ 秦成兵²⁾
肖连团²⁾ 武志芳^{1)**} 李思进^{1)**}

¹⁾ 山西医科大学第一医院核医学科, 山西医科大学分子影像精准诊疗省部共建协同创新中心, 太原 030001;

²⁾ 山西大学量子光学与光量子器件国家重点实验室, 激光光谱研究所, 极端光学省部共建协同创新中心, 太原 030006;

³⁾ 太原市中心医院, 中心实验室, 太原 030001)

摘要 目的 肺癌在生物学特性、基因组变异、增殖速度及化疗响应方面的时间异质性, 构成了对有效治疗的显著阻碍。肺癌时间异质性的复杂性, 结合其空间异质性, 为研究带来了极大挑战。本文将为肺癌研究开辟新的方向, 有助于更深入地理解肺癌的时间异质性, 从而提升对肺癌的治疗成功率。**方法** 应用拉曼光谱显微技术作为监测肺癌细胞生物分子组成实时变化的有力工具, 揭示了疾病的时间异质性。通过拉曼光谱与多元统计分析的结合, 对苯并(a)芘处理后人类肺上皮细胞的生物分子变化进行了细致观察。**结果** 随时间推移, 核酸、脂质、蛋白质及类胡萝卜素含量呈现下降趋势, 而葡萄糖浓度上升。这些变化模式暗示, 苯并(a)芘可导致遗传物质结构损伤、促进脂质过氧化、干扰蛋白质代谢、降低类胡萝卜素生成, 并改变葡萄糖代谢路径。运用拉曼光谱技术, 以实时、无侵入性、非破坏性的方式监控肺癌细胞内的生物分子动态, 进而阐明其关键分子特性。**结论** 本研究深化了对肺癌演进的认识, 并为发展个性化治疗策略提供支持, 助力提升患者的临床治疗效果。

关键词 时间异质性, 肺癌, 拉曼光谱, 多元统计分析, 苯并(a)芘暴露

中图分类号 O657.3, R734.2

DOI: 10.16476/j.pibb.2023.0447

* 国家自然科学基金 (62005150, 81971655, 62127817, 62222509, U22A6008, U22A2091) 资助项目。

** 并列第一作者。

*** 通讯联系人。

周海涛 Tel: 0351-4639425, E-mail: zht@sxmu.edu.cn

武志芳 Tel: 0351-4639425, E-mail: wuzhifang01@163.com

李思进 Tel: 0351-4639425, E-mail: lisjnm123@163.com

收稿日期: 2023-11-23, 接受日期: 2024-01-08