UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

SINGLE CELL METABOLOMICS USING MASS SPECTROMETRY: DEVICES, METHODS AND APPLICATIONS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

RENMENG LIU
Norman, Oklahoma
2019

SINGLE CELL METABOLOMICS USING MASS SPECTROMETRY: DEVICES, METHODS AND APPLICATIONS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY

Dr. Zhibo Yang, Chair

Dr. Shaorong Liu

Dr. Anthony Burgett

Dr. Yihan Shao

Dr. Laura Bartley

Acknowledgement

I would like to thank my advisor, Dr. Zhibo Yang, for his guidance throughout my graduate study. Not only his diligence, humility and determination have a significant impact on my attitude towards research, but also have shaped myself to who I am today and who I will be in the future. Without him, many moments that should have been exhilarating lost their color, and many memories that should have been everlasting quickly faded.

I would like to thank everyone in my advisory committee, including Drs. Laura Bartley, Shaorong Liu, Anthony Burgett and Yihan Shao. Not only they have provided invaluable suggestions and feedback throughout my graduate study, but also their characters as leading experts encourage me to make continuous contributions in my field of study as well as the human society at large.

My acknowledgement extends to my collaborators including Drs. Ning Pan, Genwei Zhang, Mei Sun, Laura-Isobel McCall, Ann Chen, Jiannong Li, Yanlin Zhu and Yunpeng Lan for your contributions towards the fabrication of the T-probe, the construction of machine learning models, the development of statistical and machine learning approaches to enhance data interpretation, and the experimental validation.

In addition, I would like to thank everyone in the Yang lab for their generous advice and help whenever needed.

Last but not least, I would like to thank my parents for their support and sacrifice during my graduate study.

Table of Contents

Abstract

Cells are basic functional components of eukaryotic organisms containing rich biological and physiological information. To investigate the nature of cells, a variety of fundamental and mechanistic studies of the cell functions, cellular metabolisms, metabolomic pathways, cell-cell interactions, and the interaction between cells and the surrounding microenvironment, to name a few, were conducted in recent years. Besides considerable achievements, cell-to-cell heterogeneity was recognized as the intrinsic property of cells, that is, individual cells possess unique biological traits different from other cells, even compared with their neighbors of the same genotype. Unfortunately, such cell-to-cell heterogeneity is masked by conventional analysis at the cell population level, which generally results in an averaged signal of all cells analyzed. Therefore, analytical methods that well appreciate the cell heterogeneity are desired to investigate cells at a higher resolution, that is, at the single cell level.

However, several major challenges exist in conducting single cell-level studies. First, the volume of a single cell is small (i.e., ~10 μm in diameter of a mammalian cell). Due to such dimension, the analyte within a single cell is extremely limited. Second, it is difficult to handle single cell samples due to the fragile nature of single cells. Third, some cellular species are prone to rapid turnover, and therefore, a departure from near-native biological microenvironment. Therefore, in this work, an integrated and multifunctional device, the T-probe, was designed and fabricated to conduct online and in situ analysis of live single cells at ambient condition in chapter 2. After the data acquisition, a comprehensive and generalized data analysis workflow was proposed to conduct single cell metabolomics and reveal the underlying biological principals beneath the complex raw data matrix in chapter 3. The initial data acquisition and the metabolomic analysis

were further combined with more advanced data analysis methods such machine learning (ML) algorithms using artificial intelligence to handle big data to learn the underlying patterns of cellular metabolisms, and ultimately facilitating decision making. In chapter 4 and 5, different ML models were constructed based on the acquired datasets of single cells, and evaluated in terms of the predictive accuracy of an unknown single cell possessing drug resistance. Rapid and reliable predictions of cells possessing primary or drug-induced resistance were demonstrated using established ML models on the testing datasets, and the robustness of the model was validated through accurate predictions of cells on a different batch. Such results suggested promising potential of the ML models towards future point-of-care (POC) prognostic assays.

Chapter 1: Introduction

Cell, as the fundamental component of eukaryotic living creatures, carries rich information. Among reported analytical approaches such as mass spectrometry (MS),[1-3] nuclear magnetic resonance (NMR),[4,5] Raman spectroscopy,[6,7] optical imaging,[8,9] and hybrid methods[10-12] that study cellular properties and investigate the biological nature of cells, MS gains unique advantages for its high sensitivity,[13] mass resolution[14] and throughput[15] in interrogations of cellular contents. Therefore, a variety of MS-based techniques were proposed and applied towards different biological systems to illustrate biological mechanisms,[16,17] identify unknown species,[18,19] and contribute to translational applications.[20,21]

Among all related research, the study of cellular metabolites that are downstream products of upstream regulations by genes,[22] transcriptomes[23] and proteins,[24] becomes an intriguing field, metabolomics. In recent years, efforts and contributions were devoted in metabolomic studies to enhance the understanding of cellular metabolism.[25,26] However, most of those were conducted at the populational level, resulting in an averaged signal of the entire cohort. As the cell heterogeneity that refers to the inherent difference between each cell due to intrinsic and extrinsic factors[27,28] was recognized as the key factor in a variety of biological process such as differentiation,[29] metastasis,[30] and drug resistance,[31] the investigation of cells at the single cell resolution is desired.

To date, a collection of MS-based techniques have been developed in the analysis of single cell samples, with their working mechanisms illustrated and applications demonstrated in recently published review articles.[32-36] However, most of those articles heavily focused on the technical aspects in data acquisition (i.e., sample preparation and MS detection), rather than the methodologies in the metabolomic data analysis process that are of equal importance. Regardless of the MS platform (i.e., Orbitrap, QToF, FT-ICR, Ion Trap, etc.) employed to acquire the metabolomic information from single cells, the generated raw data set is complex. Therefore, efficient data analysis approaches are imperative to analyze and interpret the data set, and further gain a profound understanding of the cellular metabolism. Although in this current era, single cell metabolomics is still in its infancy and there is a lack of a standard workflow widely accepted to process the single cell metabolomic data for all labs, majority of data analysis approaches reported were derived from the standard workflow used in the conventional liquid chromatography-MS (LC-MS) metabolomics, and therefore, they share some similarities. In this review, we focus on the discussion of data analysis approaches including data pretreatment, multivariate analysis, univariate analysis and advanced data analysis to gain a biological perspective beneath the raw single cell metabolomic data. Especially, we provide an in-depth overview of the details in the single cell metabolomics data processing.

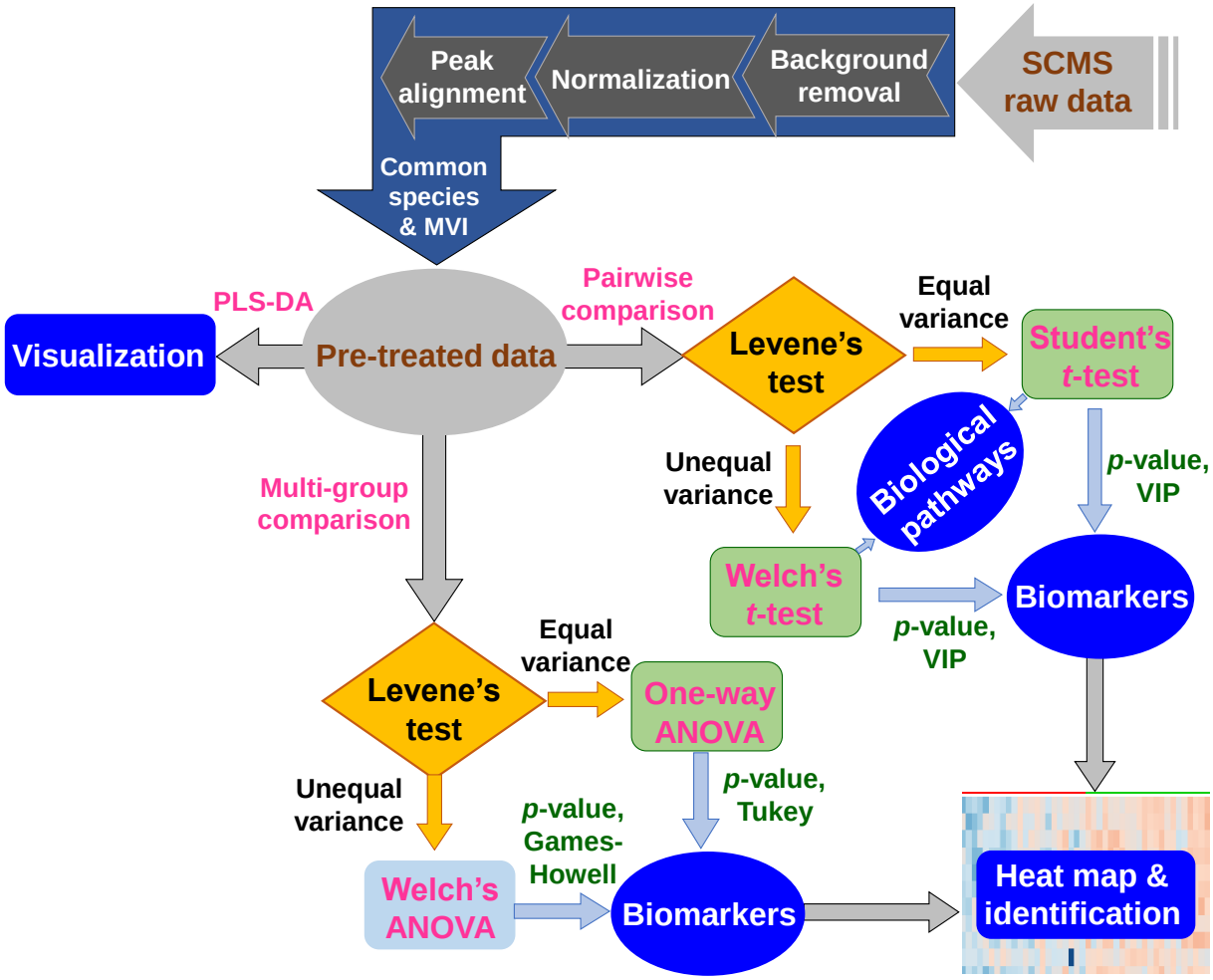

Figure 1.1. Workflow of SCMS data analysis consisting of data pre-treatment, multivariate analysis, and univariate analysis. This generalized procedure can be coupled to raw datasets obtained from broader types of SCMS platforms for single cell metabolomic analysis. Figure reproduced from Elsevier.

Data pretreatment. Single cell metabolomic data pre-treatment is the starting and essential step to extract useful biological information from the raw data and facilitate downstream statistical analysis. Referring to a recently published single cell metabolomic data analysis workflow[37] as a guideline, the data pretreatment generally includes background removal, peak normalization, data alignment and common species selection. First, raw data acquired by a variety of MS platforms were accessed by vendor-specific software (e.g., Xcalibur,[37,38] MassLynx,[39] Compass

Data Analysis,[40] etc.), followed by removal of MS peaks corresponding to the background (i.e., solvent, ambient culture medium, and other matrix). Such step could be conducted using either commercially available software or customized software. In addition, instrument noise that is considered random could be removed by eliminating MS peaks of lower abundance. By removing the background and instrument noise, MS signals corresponding to endogenous cellular metabolites were readily retained. Second, peak normalization was conducted to normalize the MS signals harvested from each individual cell to eliminate the difference of the absolute amount of analyte contained in each cell. Generally, the normalization was conducted by normalizing all peaks to the total ion current (TIC) in modern single cell metabolomic data analysis procedures. However, other normalization methods (i.e., normalizing to the base peak) could be suitable in particular cases. Third, MS peaks were aligned among all measured cells, followed by the selection of the missing value threshold (%) to include a portion of the processed data containing an acceptable level of non-zero values for downstream analysis. Although the selection criterion for the missing value threshold was not standard, adopting an appropriate level (i.e., < 50%)[38,41] is critical for convincing results. Last, selected variables (metabolites) along with their relative intensities were subjected to mean-centering and scaling (e.g., Auto or Pareto) to smooth the data distribution prior to downstream statistical analysis.

Multivariate analysis. Based on the pretreated single cell metabolomic data that contain multiple variables (metabolic species represented by their *m/z* values and the corresponding relative ion intensities), multivariate analysis was routinely performed to visualize such high dimensional data in lower (i.e., 2D or 3D) dimensional space. A variety of multivariate approaches were reported in previous publications to reduce the data dimensionality, and facilitate visualization.

In general, those approaches could be classified into two categories, unsupervised (i.e., requiring no prior knowledge of the data) and supervised (i.e., requiring some prior knowledge such as the grouping attribute of the data). For example, principle component analysis (PCA) and *t*-distributed stochastic neighbor embedding (*t*-sne) are two popular unsupervised methods that project high dimensional data in low dimensional space while retaining their features in the original space. PCA decomposes the original dataset into a series of orthogonal components (i.e., principle components), and projects them to the lower-dimensional space. The corresponding data points on the score plot represent the integral cellular profiles, and the difference among groups (e.g., different cell types) could be intuitively visualized based on the 95% confidence ellipses.[42] Despite the score plot, the loading plot contains rich information that can be used to distinguish cell subpopulations[43] or determine biomarker species characteristic to a certain condition.[44] Additionally, the distribution of the data points within a group could be used to represent the cell-to-cell heterogeneity in recent publications.[45-47] On the other hand, *t*-sne utilizes a non-linear function to retain the similarities between adjacent data points in the lower-dimensional space. Such dimension reduction method has been conducted to visualize cellular profiles and determine cell subpopulations.[48] However, the underlying mathematical principal for these methods is different (i.e., linear *vs.* non-linear model), therefore, the observed distribution of data points on the resulting score plot is different. In present studies, both methods have been widely applied and they provide complimentary information with respect to the cellular metabolomic profiles as well as similarities between cells under multiple conditions. On the contrary, supervised methods such as partial least squares-discriminant analysis (PLS-DA) and orthogonal partial least squares-discriminant analysis (OPLS-DA) require input of prior knowledge to the data analysis, and usually they outperformed the unsupervised methods in

clustering.[17,38,49] However, such models tend to be overfitted. Hence, careful model evaluation through cross-validation and permutation tests is essential. Again, the corresponding loading plot can be used to further reveal biomarker species that significantly contributed to the discrimination of different cell groups,[49-51] and the variable importance of projection (VIP) values could be used to indicate metabolic biomarkers.[37,38,52]

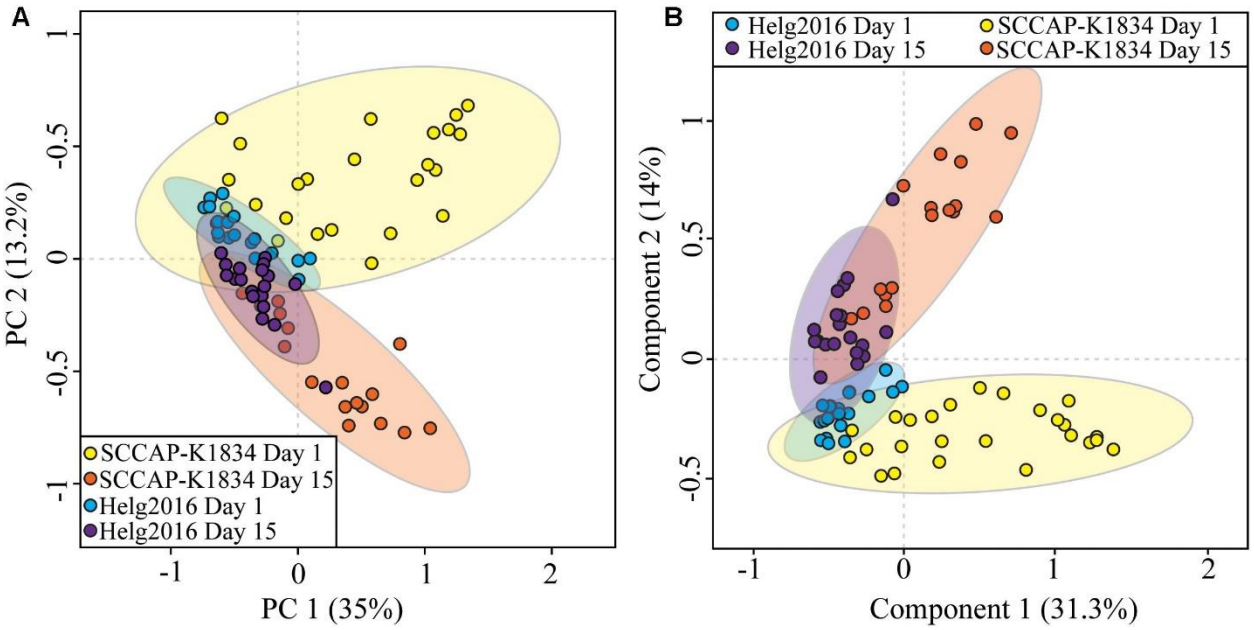

Figure 1.2. (A) PCA and (B) PLS-DA score plots of cellular profiles of *Coscinodiscus granii* strain SCCAP-K1834 and isolate Helg2016 recovered from early (Day 1) and late (Day 15) growth stages. An evident discrimination between SCCAP-K1834 Day 1 and SCCAP-K1834 Day 15 can be observed only in the PLS-DA plot. Figure reproduced from Frontiers.

Univariate analysis. The primary purpose of univariate analysis is to discover cellular species with significant change in abundance (i.e., metabolic biomarkers) among a series of groups compared. Standard statistical approaches including *t*-test and analysis of variance (ANOVA) were frequently used between two-group and among multi-group comparison, respectively. However, due to the unique nature of the single cell metabolomic data, the data distribution may be distorted, and the homogeneity of variance may be disturbed among groups to be compared.

Therefore, more rigorous statistical approaches should be conducted. For example, data transformation (i.e., generalized log transformation) was conducted to correct and approximate the data distribution to Gaussian or near-Gaussian, followed by the Levene's test to determine the type of statistical analysis (i.e., parametric or non-parametric) to be performed. Based on the testing statistic of the Levene's test, Student's or Welch's *t*-test can be performed using each variable in a two-group comparison setting; whereas one-way or Welch's ANOVA can be performed in a multi-group comparison setting. In case the data distribution is highly skewed, a non-parametric Mann-Whitley *U* test (for two groups)[48,53] or Kruskal-Wallis *H* test (for multiple groups)[54] can be performed. A confidence level of 95% ($\alpha = 0.05$) was constantly reported for single cell metabolomic studies, and the false discover rate (FDR) correction was generally adopted when performing multiple statistical tests to reduce the false discover rate.[52]

Pathway topology analysis. Similar to conventional LC-MS, the pathway topology analysis can be conducted by computing significantly altered or enriched pathways using the relative intensities of identified metabolites. Those significantly altered pathways revealed at the single cell level could be further mapped against a global biological system (e.g., zebrafish)[55] to generate a holistic picture of the metabolic activities.[56,57] More recently, bioinformatics tools such as Mummichog was released to compute for altered pathways using only the full MS data acquired from single cells,[37] paving a new avenue for future shotgun metabolomics/lipidomics studies at the single cell level.

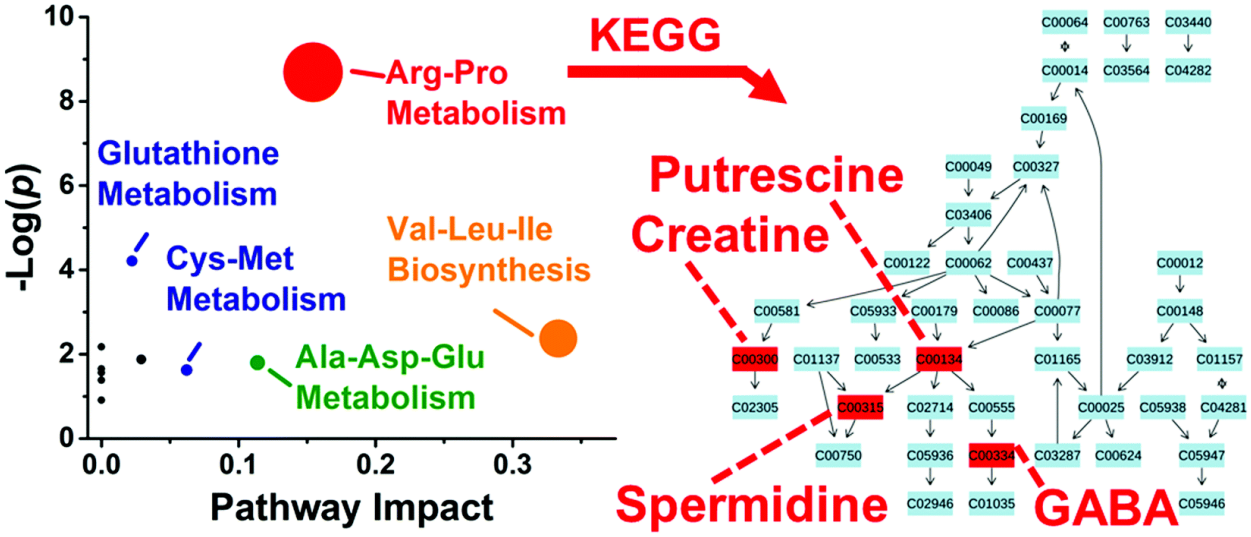

Figure 1.3. Pathway analysis for metabolites with asymmetric distribution between left and right D1 blastomeres. Correlation between *p* value from pathway enrichment analysis and pathway impact from pathway topology analysis in MetaboAnalyst identified arginine–proline as the most represented pathway (left panel). Identified metabolites underlying this pathway were mapped against KEGG using *Danio rerio* (zebrafish) as the model organism (right panel). Reproduced from The Royal Society of Chemistry.

Advanced data analysis. Despite conventional multivariate and univariate data analysis methods, advanced methods have emerged and been incorporated into the single cell metabolomic data analysis workflow. For example, machine learning (ML) algorithms that utilize artificial intelligence to handle big data have been introduced not only in LC-MS metabolomic research,[58] also recently in the single cell metabolomic studies.[37,44] With the introduction of different ML methods, subtle patterns or trends not intuitive to the human eyes can be revealed and discerned. For example, random forest (RF), logistic regression (LR) and artificial neural network (ANN) were used to construct ML models capable of predicting the attribute (i.e., primary drug resistance) of unknown single cells based on the cellular metabolomic profiles rapidly with high accuracy.[44] Those reported models showed promising potentials to be applied towards future point-of-care diagnostic assays.

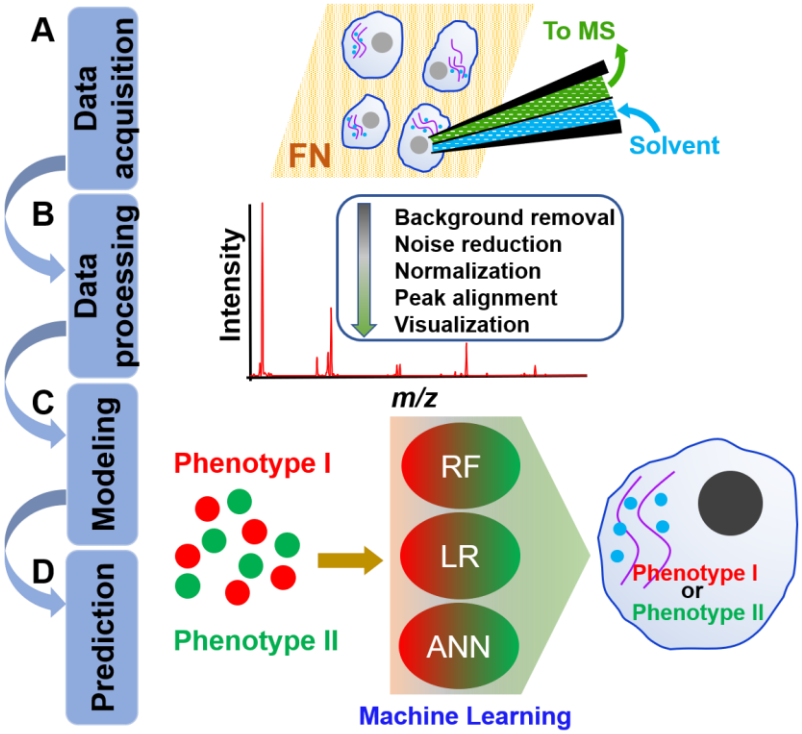

Figure 1.4. Workflow of the combined single cell mass spectrometry experiments and machine learning (ML) data analysis methods. (A) MS measurements of single cells using the single-probe SCMS technique. (B) A comprehensive data processing approach to extract metabolomic information from raw SCMS datasets and visualize cellular profiles in low dimensional space. (C) ML models built on cells with two different phenotypes (with or without CAM-DR). (D) Rapid and reliable prediction of drug-resistant phenotypes at the single cell level. Reproduced by permission of The Royal Society of Chemistry.

Evaluation of technical and biological variance. Due to the cell heterogeneity, fluctuation of MS signals and the destructive nature of some sampling methods, the technical (i.e., difference of measured signals induced by extrinsic factors such as instrumental stability, random error, etc.) and biological (i.e., difference of measured signals induced by the biological nature of the sample) variance were usually evaluated as an experimental validation. To date, multiple evaluations were conducted using different sampling devices on a variety of biological systems, and consistently, they demonstrated the dominance of biological variance among single cell metabolomic datasets.[41,51,53]

Reference

(1) Gstaiger, M.; Aebersold, R. *Nat Rev Genet* **2009**, *10*, 617.
(2) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198.
(3) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom Rev* **2007**, *26*, 51-78.
(4) Viant, M. R.; Rosenblum, E. S.; Tjeerdema, R. S. *Environmental Science & Technology* **2003**, *37*, 4982-4989.
(5) Yee, A.; Chang, X.; Pineda-Lucena, A.; Wu, B.; Semesi, A.; Le, B.; Ramelot, T.; Lee, G. M.; Bhattacharyya, S.; Gutierrez, P.; Denisov, A.; Lee, C.-H.; Cort, J. R.; Kozlov, G.; Liao, J.; Finak, G.; Chen, L.; Wishart, D.; Lee, W.; McIntosh, L. P., et al. *Proc Natl Acad Sci U S A* **2002**, *99*, 1825.
(6) Cherney, D. P.; Ekman, D. R.; Dix, D. J.; Collette, T. W. *Anal Chem* **2007**, *79*, 7324-7332.
(7) Smith, R.; Wright, K. L.; Ashton, L. *Analyst* **2016**, *141*, 3590-3600.
(8) Stender, A. S.; Marchuk, K.; Liu, C.; Sander, S.; Meyer, M. W.; Smith, E. A.; Neupane, B.; Wang, G.; Li, J.; Cheng, J.-X.; Huang, B.; Fang, N. *Chem Rev* **2013**, *113*, 2469-2527.
(9) Park, G. K.; Hoseok, s.; Kim, G. S.; Hwang, N. S.; Choi, H. S. *Applied spectroscopy reviews* **2018**, *53*, 360-375.
(10) Zhang, D.; Xie, Y.; Mrozek, M. F.; Ortiz, C.; Davisson, V. J.; Ben-Amotz, D. *Anal Chem* **2003**, *75*, 5703-5709.
(11) Marshall, D. D.; Powers, R. *Prog. Nucl. Magn. Reson. Spectrosc.* **2017**, *100*, 1-16.
(12) Deng, C.; Xiong, X.; Krutchinsky, A. N. *Molecular & cellular proteomics : MCP* **2009**, *8*, 1413-1423.
(13) Cifani, P.; Kentsis, A. *Molecular & Cellular Proteomics* **2017**, *16*, 2006.
(14) Rathahao-Paris, E.; Alves, S.; Junot, C.; Tabet, J.-C. *Metabolomics* **2015**, *12*, 10.
(15) Kempa, E. E.; Hollywood, K. A.; Smith, C. A.; Barran, P. E. *Analyst* **2019**, *144*, 872-891.
(16) Zagorevskii, D.; Song, M.; Breneman, C.; Yuan, Y.; Fuchs, T.; Gates, K. S.; Greenlief, C. M. *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 881-892.
(17) Sun, M.; Yang, Z. *Anal Chem* **2019**, *91*, 2384-2391.
(18) Tacheny, A.; Dieu, M.; Arnould, T.; Renard, P. *J Proteomics* **2013**, *94*, 89-109.
(19) Lynn, K.-S.; Cheng, M.-L.; Chen, Y.-R.; Hsu, C.; Chen, A.; Lih, T. M.; Chang, H.-Y.; Huang, C.-j.; Shiao, M.-S.; Pan, W.-H.; Sung, T.-Y.; Hsu, W.-L. *Anal Chem* **2015**, *87*, 2143-2151.
(20) Zhang, J. L.; Rector, J.; Lin, J. Q.; Young, J. H.; Sans, M.; Katta, N.; Giese, N.; Yu, W. D.; Nagi, C.; Suliburk, J.; Liu, J. S.; Bensussan, A.; DeHoog, R. J.; Garza, K. Y.; Ludolph, B.; Sorace, A. G.; Syed, A.; Zahedivash, A.; Milner, T. E.; Eberlin, L. S. *Sci Transl Med* **2017**, *9*.
(21) Ferreira, C. R.; Yannell, K. E.; Jarmusch, A. K.; Pirro, V.; Ouyang, Z.; Cooks, R. G. *Clin Chem* **2016**, *62*, 99.
(22) Kjærbølling, I.; Vesth, T. C.; Frisvad, J. C.; Nybo, J. L.; Theobald, S.; Kuo, A.; Bowyer, P.; Matsuda, Y.; Mondo, S.; Lyhne, E. K.; Kogle, M. E.; Clum, A.; Lipzen, A.; Salamov, A.; Ngan, C. Y.; Daum, C.; Chiniquy, J.; Barry, K.; LaButti, K.; Haridas, S., et al. *Proc Natl Acad Sci U S A* **2018**, *115*, E753.
(23) Hoppe, A. *Metabolites* **2012**, *2*, 614-631.
(24) Akpunarlieva, S.; Weidt, S.; Lamasudin, D.; Naula, C.; Henderson, D.; Barrett, M.; Burgess, K.; Burchmore, R. *J Proteomics* **2017**, *155*, 85-98.
(25) Lei, Z.; Huhman, D. V.; Sumner, L. W. *J Biol Chem* **2011**, *286*, 25435-25442.

(26) Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W. *Mol Biosyst* **2012**, *8*, 470-481.
(27) Marjanovic, N. D.; Weinberg, R. A.; Chaffer, C. L. *Clin Chem* **2013**, *59*, 168.
(28) Sun, X.; Yu, Q. *Acta pharmacologica Sinica* **2015**, *36*, 1219-1227.
(29) Gerlach, C.; Rohr, J. C.; Perié, L.; van Rooij, N.; van Heijst, J. W. J.; Velds, A.; Urbanus, J.; Naik, S. H.; Jacobs, H.; Beltman, J. B.; de Boer, R. J.; Schumacher, T. N. M. *Science* **2013**, *340*, 635.
(30) Lawson, D. A.; Kessenbrock, K.; Davis, R. T.; Pervolarakis, N.; Werb, Z. *Nature Cell Biology* **2018**, *20*, 1349-1360.
(31) Hoey, T. *Sci Transl Med* **2010**, *2*, 28ps19.
(32) Yang, Y. Y.; Huang, Y. Y.; Wu, J. H.; Liu, N.; Deng, J. W.; Luan, T. G. *Trac-Trend Anal Chem* **2017**, *90*, 14-26.
(33) Duncan, K. D.; Fyrestam, J.; Lanekoff, I. *Analyst* **2019**, *144*, 782-793.
(34) Ali, A.; Abouleila, Y.; Shimizu, Y.; Hiyama, E.; Emara, S.; Mashaghi, A.; Hankemeier, T. *TrAC, Trends Anal. Chem.* **2019**.
(35) Zhang, L.; Vertes, A. *Angew. Chem. Int. Ed.* **2018**, *57*, 4466-4477.
(36) DeLaney, K.; Sauer, C. S.; Vu, N. Q.; Li, L. J. *Molecules* **2019**, *24*, 42.
(37) Liu, R.; Zhang, G.; Sun, M.; Pan, X.; Yang, Z. *Anal. Chim. Acta* **2019**.
(38) Baumeister, T. U. H.; Vallet, M.; Kaftan, F.; Svatoš, A.; Pohnert, G. *Front Plant Sci* **2019**, *10*, 172.
(39) Zhang, L.; Foreman, D. P.; Grant, P. A.; Shrestha, B.; Moody, S. A.; Villiers, F.; Kwak, J. M.; Vertes, A. *Analyst* **2014**, *139*, 5079-5085.
(40) Onjiko, R. M.; Moody, S. A.; Nemes, P. *Proc Natl Acad Sci U S A* **2015**, *112*, 6545.
(41) Onjiko, R. M.; Portero, E. P.; Moody, S. A.; Nemes, P. *Anal Chem* **2017**, *89*, 7069-7076.
(42) Do, T. D.; Comi, T. J.; Dunham, S. J. B.; Rubakhin, S. S.; Sweedler, J. V. *Anal Chem* **2017**, *89*, 3078-3086.
(43) Do, T. D.; Ellis, J. F.; Neumann, E. K.; Comi, T. J.; Tillmaand, E. G.; Lenhart, A. E.; Rubakhin, S. S.; Sweedler, J. V. *ChemPhysChem* **2018**, *19*, 1180-1191.
(44) Liu, R.; Zhang, G.; Yang, Z. *Chem. Commun.* **2019**, *55*, 616-619.
(45) Huang, Q.; Mao, S.; Khan, M.; Zhou, L.; Lin, J.-M. *Chem. Commun.* **2018**, *54*, 2595-2598.
(46) Zhang, Y.; Jin, L.; Xu, J.; Yu, Y.; Shen, L.; Gao, J.; Ye, A. *Analyst* **2018**, *143*, 164-174.
(47) Standke, S. J.; Colby, D. H.; Bensen, R. C.; Burgett, A. W. G.; Yang, Z. *Anal Chem* **2019**, *91*, 1738-1742.
(48) Neumann, E. K.; Comi, T. J.; Rubakhin, S. S.; Sweedler, J. V. *Angew. Chem. Int. Ed.* **2019**, *0*.
(49) Liu, R.; Pan, N.; Zhu, Y.; Yang, Z. *Anal Chem* **2018**, *90*, 11078-11085.
(50) Stolee, J. A.; Shrestha, B.; Mengistu, G.; Vertes, A. *Angew. Chem. Int. Ed.* **2012**, *51*, 10386-10389.
(51) Stopka, S. A.; Khattar, R.; Agtuca, B. J.; Anderton, C. R.; Paša-Tolić, L.; Stacey, G.; Vertes, A. *Front Plant Sci* **2018**, *9*, 1646.
(52) Onjiko, R. M.; Plotnick, D. O.; Moody, S. A.; Nemes, P. *Analytical Methods* **2017**, *9*, 4964-4970.
(53) Zhang, L.; Sevinsky, C. J.; Davis, B. M.; Vertes, A. *Anal Chem* **2018**, *90*, 4626-4634.
(54) Wang, R.; Zhao, H.; Zhang, X.; Zhao, X.; Song, Z.; Ouyang, J. *Anal Chem* **2019**, *91*, 3667-3674.
(55) Onjiko, R. M.; Morris, S. E.; Moody, S. A.; Nemes, P. *Analyst* **2016**, *141*, 3648-3656.

(56) Portero, E. P.; Nemes, P. *Analyst* **2019**, *144*, 892.
(57) Sun, M.; Yang, Z. B.; Wawrik, B. *Front Plant Sci* **2018**, *9*, 571.
(58) Ma, C.; Ren, Y.; Yang, J.; Ren, Z.; Yang, H.; Liu, S. *Anal Chem* **2018**, *90*, 10881-10888.

Chapter 2: T-Probe: An Integrated Microscale Device for Online In Situ Single Cell Analysis and Metabolic Profiling Using Mass Spectrometry

2.1. Introduction

Cell, as a basic component of living organisms, contains rich biological information such as expression of genes,[1] interaction with other cells,[2] and metabolic response to altered microenvironment.[3] Numerous techniques have been dedicatedly developed to study cellular constituents and reveal the biological principles of cellular metabolism.[4-7] Mass spectrometry (MS), a rapidly developing technique to analyze ion species with high accuracy and sensitivity, stands out as a powerful tool to achieve efficient and reliable analysis of cell extractions.[8-10] Conventional MS-based methodologies, such as the analysis of cell lysate using LC (liquid chromatography)-MS, have been broadly adopted in studies of populations of cells, and consequently resulted in averaged chemical information from cell populations analyzed.[11-14] In fact, cells are unique individuals that present heterogeneity due to intrinsic (e.g., genetic information) and extrinsic (e.g., surrounding microenvironment) factors.[15] To overcome drawbacks of the conventional methods and obtain biological signature of individual cells, a number of MS based techniques have been developed to conduct single cell MS (SCMS) analysis. Depending on the sampling mechanisms, these methods can be generally classified as ion beam-based, laser-based, probe-based, and other methods. Ion beam-based SCMS methods primarily refer to secondary ion MS (SIMS), which provides high special resolution, allowing for subcellular analysis.[16-17] Laser-based SCMS methods include but are not limited to matrix assisted laser desorption/ionization (MALDI) MS, laser ablation electrospray ionization (LAESI) MS, and laser desorption/ionization on porous silicon (DIOS) MS. MALDI MS is a popular method to achieve

tissue level and single cell level analyses with high resolution and throughput.[18-20] LAESI MS, as an ambient SCMS method that allows for studies of live cells with minimum sample preparation, has been successfully applied to analyze a large variety of cell lines *in situ*.[21-23] DIOS MS is an alternative ambient SCMS method, which allows cells to be cultured directly on porous silicon substrates followed by laser ablation and MS detection.[24] Probe-based ambient SCMS methods include live single cell video MS,[25-26] probe electrospray ionization (PESI) MS,[27] pressure probe electrospray ionization (PPESI) MS,[28] induced nano-electrospray ionization (InESI) MS,[29] and the Single-probe MS methods.[30-31] In addition to these non-separative methods, separation techniques, such as capillary electrophoresis (CE) and ion mobility separation (IMS), have been coupled with MS to study left-right patterning of single blastomeres in frog embryos,[32-33] and obtain metabolic signatures of single human hepatocytes.[34]

Despite broad applications reported, the majority of existing SCMS methods are classified as offline SCMS methods, because they generally involve multi-step, offline sample preparation procedures. However, cellular metabolites shift immediately (in seconds) upon changes of surrounding microenvironment,[9] and therefore, rapid real-time SCMS techniques are obligatory to analyze live single cells. We have previously developed a probe-based SCMS method, the Single-probe MS, and demonstrated its applications in single cell analysis,[30, 35] MS imaging of brain tissues,[36-37] and MS analysis of live multicellular tumor spheroids.[38]

Here, we report the design and application of a novel miniaturized, integrated sampling and ionization device, the T-probe, that can be coupled to MS for online *in situ* live single cell analysis and metabolic profiling. As shown in Figure 2.1A, the T-probe has three major components: a sampling probe, a solvent-providing capillary, and a nano-ESI emitter. These capillaries are sandwiched by two polycarbonate (PC) substrates engraved with T-shaped grooves, and bound

through thermal binding. When the T-probe is used to conduct SCMS experiments, the solvent-providing capillary delivers sampling solvent (e.g., methanol (50%)/water (50%) with 0.1% formic acid) from a syringe. A DC voltage is applied to the conductive union and transmitted though the solvent to provide an ionization voltage at the nano-ESI emitter (aperture size ~20 μm). Using the optimized solvent flow rate (e.g., 0.2 μL/min) and ionization voltage (e.g., ~4 kV), a stable electrospray can be observed at the nano-ESI emitter without forming solvent droplets at the sampling probe tip. Our experiments demonstrate that a suction force is generated at the sampling probe tip to withdraw cellular contents. At the T-junction, the solvent is mixed with cellular species, and the mixture is subsequently delivered to the nano-ESI emitter for ionization and MS detection (Figure 2.1B). Although the exact mechanism of generating such suction force is unclear, this self-aspiration process is likely due to the continuous consumption of solution inside the nano-ESI emitter, leading to a liquid flow towards the nano-ESI emitter through the capillary action. Switching off the ionization voltage immediately resulted in a paused liquid suction from the sampling probe, until the ionization voltage was turned on again. Similar phenomena of solvent self-aspiration have been observed in other ambient MS methods such as liquid microjunction surface sampling probe (LMJ-SSP),[39] nano-desorption electrospray ionization (nano-DESI),[40] and the Single-probe.[30] Under the optimized experimental conditions, the T-probe has no efflux of solvent from the sampling probe during the analysis of a target cell, minimizing the perturbation of the microenvironment of nearby cells to be analyzed. The perturbation of cell's microenvironment in some other sampling methodologies, in which the solvent, heat, or matrices is needed, can potentially alter the native biological states of live cells. In addition, the sampling probe (with one channel) of the T-probe can be further minimized (e.g., <1 μm), and potentially used for subcellular-level analysis of mammalian cells. To provide additional evidence and shed

light on the sampling process of the T-probe, we carried out experiments using a prepared solution, and estimated the fraction of liquid introduced from the sampling probe to that from the solvent-providing capillary.

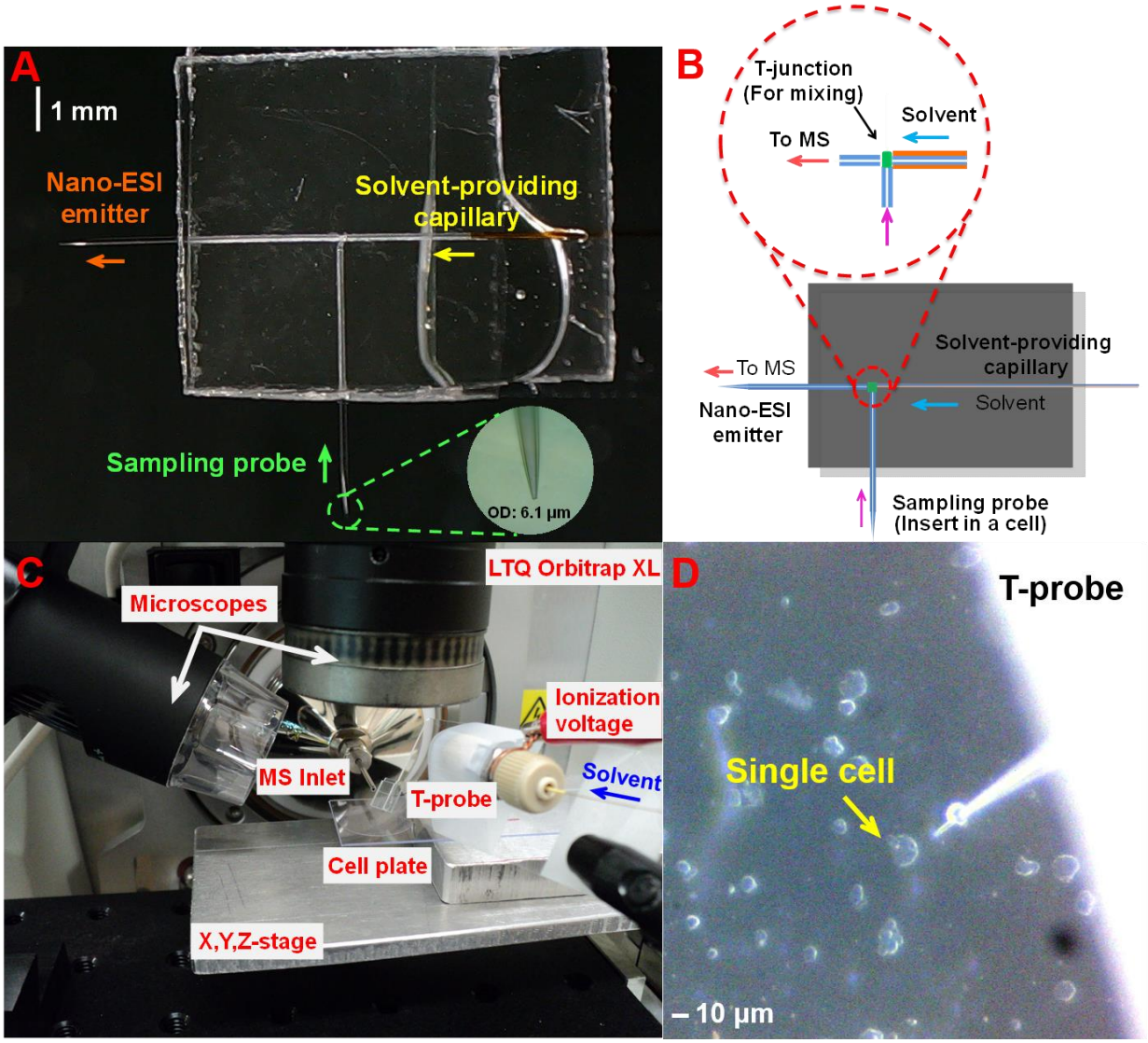

Figure 2.1. Utilizing the T-probe for the SCMS experiments. (A) Photo of a T-probe. Inset: a zoomed-in photo of the sampling probe tip. (B) Illustration of working mechanism and fluid flow directions in the T-probe. (C) Components of the T-probe SCMS experimental setup. (D) A photo illustrating the insertion of the T-probe tip into a cell.

Coupled to a mass spectrometer, the T-probe allows for online *in situ* MS analysis of live single cells in real time with minimal sample preparation. Due to its high sensitivity, the T-probe was applied to the detection of a variety of species present in single cells such as endogenous cellular metabolites, environmental stimulus (i.e., anticancer drug), and xenobiotic metabolites. Among those, we performed online MS/MS analysis for abundant species of interest to confirm their chemical identities. In addition, we established a comprehensive MS data analysis approach, including multivariate and univariate methods, to perform metabolic profiling at single cell level. As a result, we illustrated cellular response to drug treatment and discovered cellular species sensitive to such microenvironmental alternation.

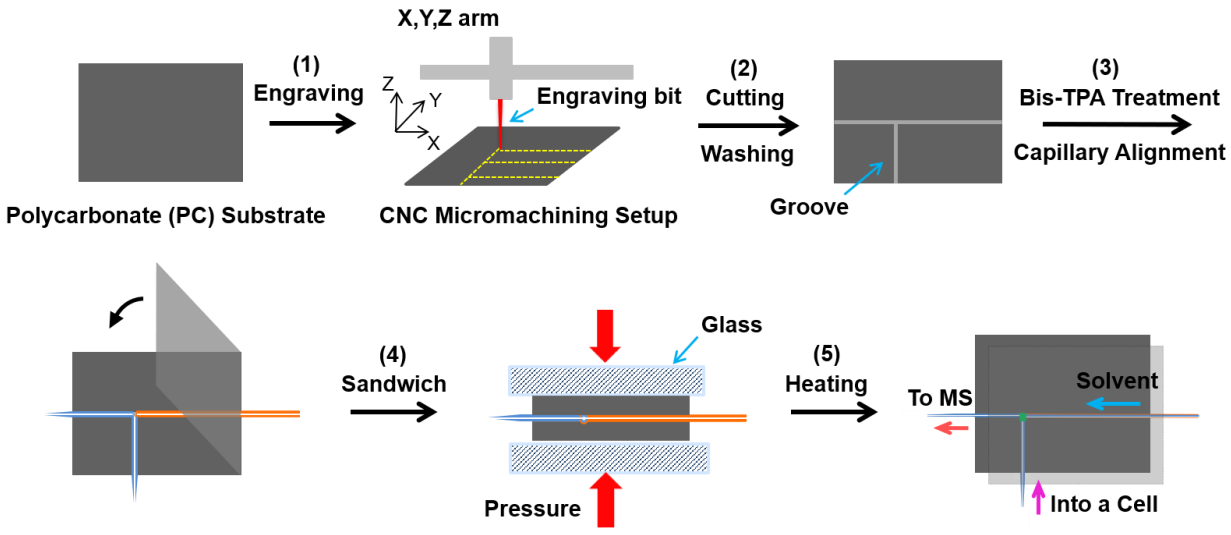

Figure 2.2. Semi-automated fabrication workflow of the T-probe. (1) The T-shaped grooves are engraved on a polycarbonate (PC) substrate using computer numeric control (CNC) micromachining. (2) PC slides are cut and (3) subjected to surface silconization treatment. (4)

Three capillaries are positioned within grooves and sandwiched by a pristine PC piece. (5) The sandwiched assembly is heated in an oven for thermal binding.

2.2. Fabrication of the T-probe

We developed a semi-automated fabrication protocol to manufacture T-probes with high reproducibility and efficiency as demonstrated in Figure 2.2. The fabrication workflow includes five major steps. (1) We engraved multiple sets of T-shaped grooves (width ~150 μm) on one PC substrate using a computer numeric control (CNC) micro-engraver. (2) A mini electric cutting saw was used to cut the engraved PC substrate into individual pieces (each piece contains one set of T-grooves). The pristine PC substrates (i.e., not engraved) were also cut into pieces with identical dimensions. Both engraved and pristine PC substrates were ultrasonically cleaned in *iso*-propanol to remove any residual debris within the grooves and contamination on surfaces. (3) Those cleaned PC substrates were dried and then treated with Bis[3-(trimethoxysilyl)propyl]amine (Bis-TPA), a siliconization reagent, to increase hydrophilicity of PC surfaces and enhance PC–PC binding in the following thermal binding step.[41] (4) Three capillaries (i.e., a solvent-providing capillary, a sampling probe, and a nano-ESI emitter) were produced from the same type of fused silica capillary (O.D. = 150 μm, I.D. = 75 μm). Among them, the sampling probe and the nano-ESI emitter were pulled using a laser micro pipette puller (Model P-2000, Sutter Instrument, Novato, CA). To fabricate a T-probe, these three capillaries were carefully positioned in grooves on an engraved PC substrate. The solvent-providing capillary and the nano-ESI emitter were aligned in-line, whereas the sampling probe was vertically placed so that they formed a T-junction. A pristine PC slide was used to sandwich these three capillaries to form a unit. To retain the positions of capillaries and PC slides, we used two glass slides and applied pressure using two paper clippers.

(5) Lastly, we kept the sandwiched assembly in an oven at 110 °C for 30 min, and two PC slides were bound to form a T-probe through thermal binding. With enhanced PC–PC binding strength, solvent leakage from T-probes was rarely observed during the following SCMS experiments. Moreover, the employment of CNC micromachining promotes our manufacture reproducibility and efficiency. These standardized T-probe fabrication protocols can be readily adopted by other labs or industrial manufacturing facilities. In addition, the relevant techniques (e.g., micromachining, hydrophobic treatment PC surface, and thermal binding process) are valuable for studies using plastic microfluidic devices.

2.3. SCMS Experiments Using the T-probe

The experimental setup for SCMS analysis using the T-probe was adapted from our Single-probe SCMS setup. The T-probe was attached to a manual XYZ-manipulator for position adjustment. The nano-ESI emitter was ~2 mm away from the mass spectrometer inlet, and the solvent-providing capillary was connected to a syringe through a conductive union, where the ionization voltage was applied. A glass coverslip containing live HeLa cells and culture medium was placed onto a motorized XYZ-translational stage system controlled by a LabView package.[42] Two microscopes were utilized to monitor the working status of the T-probe: a lateral microscope focusing on the nano-ESI emitter of the T-probe to monitor its electrospray conditions, and a stereo microscope focusing on the sampling probe tip to monitor the cell sampling process. During SCMS experiments, the sampling solvent was continuously delivered through the solvent-providing capillary, and stable signals of solvent ions were obtained without solvent dripping from the sampling probe by carefully adjusting the solvent flow rate and ionization voltage. We then selected a target single cell by moving the motorized XYZ-translational stage so that both the cell

and the sampling probe tip were clearly observed under the stereo microscope. By gradually lifting the stage (at an increment of 0.1 μm), the sampling probe tip was first submerged into culture medium above cells, and then penetrated the cell membrane with a continuous suction. Cellular contents along with the surrounding culture medium were subsequently withdrawn into the sampling probe and immediately ionized by nano-ESI emitter for MS analysis. Correspondingly, we observed dramatic changes of mass spectra profiles from the solvent background through culture medium to the mixed cellular constituents and culture medium. Upon finishing the measurement of a cell, we lowered down the translational stage to remove the sampling probe tip out of the cell. Meanwhile, due to the loss of cytoplasmic contents, a significant change of cell morphology was observed under the microscope, and the cell was unlikely viable. To eliminate residues (carry-overs) from the sampled cell, we rinsed the T-probe by gently flushing it with the sampling solvent while monitoring the MS signal change. Such process ensured cleanness of the T-probe prior to the analysis of the next cell. A Thermo LTQ Orbitrap XL mass spectrometer was employed in our SCMS data acquisition process for high sensitivity and mass resolution. The experimental mass analyze parameters are listed as follows: mass resolution 60,000 at *m/z* 400, mass range 200–1500, +4 kV ionization voltage, 1 microscan, 100 ms max injection time and AGC (automatic gain control) on.

2.4. SCMS Data Analysis

Generally, ion signals corresponding to cellular metabolites were extracted from the raw data, whereas background signals (from exogenous species such as cell culture medium and sampling solvent) and instrument noise were removed. We carried out Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA)[43] in combination with univariate analysis (e.g., two-sample *t*-

test) to study changes of cellular profiles induced by drug treatment. We determined cellular species with significant abundance change after drug treatment, which are of potential interest of clinical and pharmaceutical studies. Lastly, by referring to available online metabolome database such as METLIN (https://metlin.scripps.edu)[44] and HMDB (http://www.hmdb.ca),[45] we accomplished tentative labeling for detected species, and further identified (through online MS/MS at single cell level) abundant species by comparing their fragmentation patterns with *in silico* data from online database.

2.5. Estimation of Mixing Ratio at the T-junction

To gain a better understanding of the working mechanism of the T-probe, we estimated the volume mixing ratio of cellular contents withdrawn by the sampling probe to the sampling solvent delivered from the solvent-providing capillary. Considering the viscosity of cell cytoplasm (2–3 cP),[46] we prepared 250 nM deuterated irinotecan (d10-irinotecan) solution in *iso*-propanol (viscosity = 2.86 cP) in a vial located underneath the T-probe, whereas regular irinotecan with the same concentration was prepared in the sampling solvent (methanol (50%)/water (50%) with 0.1% formic acid). In the following experiment, the sampling solution was continuously delivered through the solvent-providing capillary at the flow rate used in our SCMS experiment (0.2 μL/min). Upon applying the ionization voltage, the protonated irinotecan ([irinotecan + H]$^+$) was observed. We then submerged the sampling probe tip into the vial containing d10-irinotecan, and observed both [irinotecan + H]$^+$ and [d10-irinotecan + H]$^+$. The volume mixing fraction was estimated by calculating the relative ion intensities of [irinotecan + H]$^+$ to [d10-irinotecan + H]$^+$. Such process was repeated for 5 times, and we found 41.8% ± 1.4% (mean ± standard deviation) volume of solution was drawn through the sampling probe. Our experimental results indicate that

a sufficient suction is produced at the tip of the sampling probe, and a dilution of extracted cellular contents with the sampling solvent occurs at the T-junction, which promotes detection sensitivity.[47] However, it is worth noting that the volume mixing ratio may vary among individual T-probes in SCMS experiments, and further studies, such as computational fluid dynamics (CFD) simulations, need to be carried out in future work.

2.6. Limit of Detections (LODs) of the T-probe

To evaluate the sensitivity of the T-probe in MS experiments, we used both the standard nano-ESI ion source and the T-probe device to measure the prepared solutions and determine their LODs, which are defined as the minimum concentrations allowing for MS detection with obvious ion signal intensities (i.e., signal-to-noise (S/N) > 3). We selected four standard compounds relevant to our SCMS experiments, including cellular lipids 1-oleoyl-2-palmitoyl-sn-glycero-3-phosphocholine (PC(18:1/16:0)) and 1,3-dihexadecanoyl-2-(9Z-octadecenoyl)-glycerol (TG(16:0/18:1/16:0)), anticancer drug irinotecan, and a peptide example leucine encephalin, to conduct LODs determination. Our results (Table 2.1) indicate that the T-probe device exhibits comparable LODs as the standard nano-ESI method. In addition, the LOD of the T-probe is comparable to that of the Single-probe (e.g., the LOD of PC(18:1/16:0) is 5 nM).[31]

Table 2.1. LODs (nM) of standard compounds detected by nano-ESI and the T-probe.

Compound	Nano-ESI	T-probe
Irinotecan	0.1	0.1
Leucine Enkephalin	0.8	1.0

PC (18:1/16:0)	5.0	10.0
TG (16:0/18:1/16:0)	0.1	0.1

2.7. SCMS Experiments of Cells under Control Condition

The T-probe was used to obtain cellular profiles of randomly selected HeLa cells under control condition (without drug treatment). The ion signals of cellular species were observed within seconds upon the insertion of the sampling probe tip into a cell. Typically, the ion signals from one cell could last for > 15 s before a rapid decrease, providing sufficient time to conduct online MS/MS analysis of abundant species. Due to the robustness of the T-probe, multiple cells can be consecutively measured using one T-probe without changing the device, which paves the way towards high throughput SCMS analysis. We tentatively labelled the detected cellular species based on their accurate *m/z* values. According to the tentative labeling, a broad range of cellular metabolites, including organic acids, cholesteryl esters (CE), phosphatidylcholines (PC), phosphatidylethanolamides (PE), and triglycerides (TG) in their protonated, sodiated, and potassiated forms (Figure 2.3A) were present in single cells. It is worth noting that these tentative labels are based on the accurate *m/z* measurement. They are not regarded as identifications, and therefore, can be only used as the reference to screen for molecules of interest. Compared with previously published studies of single HeLa cells,[30-31, 48] our current results show a comparable number of lipid species at a higher mass range (i.e., *m/z* 600–900). While at a lower mass range (i.e., *m/z* 200–600), our method provides more coverage of cellular metabolites (e.g. organic acids) than those obtained using other methods (e.g., the Single-probe MS and MALDI-MS). For ions with high abundances, such as $[PC(34:1) + Na]^+$ (*m/z* 782.5676) and $[PC(36:2) + Na]^+$ (*m/z*

808.5832), online *in situ* MS/MS spectra were obtained though collision-induced dissociation (CID) tandem mass spectrometry (MS/MS) to confirm their molecular structures at single cell level. Due to the extremely small volume of one HeLa cell (1.2–4.3 pL)[49] and the influence of matrix effect on ionization (e.g., culture medium extracted along with cellular species suppresses the ionization efficiency), it is very challenging to acquire a large number of abundant cellular species for MS/MS analysis. Improvement of the current technologies, such as combining with microseparation (e.g., capillary electrophoresis) or post-ionization separation (e.g., ion mobility separation) methods, is needed to enhance both the coverage and identification of metabolites.

2.8. SCMS Experiments of Cells under Drug Treatment Condition

The SCMS experiments were also carried out using cells treated with irinotecan (7-ethyl-10-[4-(1-piperidino)-1-piperidino]-carbonyloxy camptothecin), which is known as a *Topoisomerase I* (*Topo-I*) inhibitor that damages the replication and transcription processes of single strand DNAs and ultimately induces cell apoptosis.[50] In our experiments, HeLa cells were treated with irinotecan under a series of concentrations (i.e., 10 μM, 1 μM, 100 nM, and 10 nM) for 1 h, and then rinsed using fresh cell culture medium to remove drug molecules in extracellular medium. We performed SCMS analysis and detected the protonated species, $[\text{irinotecan} + \text{H}]^+$ (*m/z* 587.2870), in all drug treated single cells (Figure 2.3D). Furthermore, the detection of this drug target was confirmed by online MS/MS analysis at single cell level. As we decreased the treatment concentration of irinotecan, the ion intensity of this drug target markedly decreased. For example, $[\text{irinotecan} + \text{H}]^+$ detected from cells treated with 10 nM irinotecan presented a low signal intensity (S/N ~4) that is slightly above its LOD (Figure 2.3D). We then explored cellular xenobiotic activities in response to drug stimulus. Particularly, we detected related drug metabolites including decarboxyl-irinotecan (*m/z* 543.2971), dehydro-irinotecan (*m/z* 585.2713), and hydroxyl-irinotecan (*m/z*

603.2819) from single cells. Our findings are in good agreement with previous results, as those metabolites were reported in other studies of multicellular spheroids treated using the same drug compound.[38, 51] It is worth noting that these drug metabolites were observed from single cells treated using 10 μM irinotecan, but they were not detected from cells under lower treatment concentrations (i.e., 1 μM, 100 nM, and 10 nM). The missing drug metabolites from single cells under low-concentration treatment conditions are likely attributed to multiple factors, including limited uptake of drug molecules with lower initial dosage,[52] insufficient treatment time to develop xenobiotic activities,[53] and matrix effect (i.e., due to mixed cellular metabolites and species in the culture medium) in MS measurements.[54]

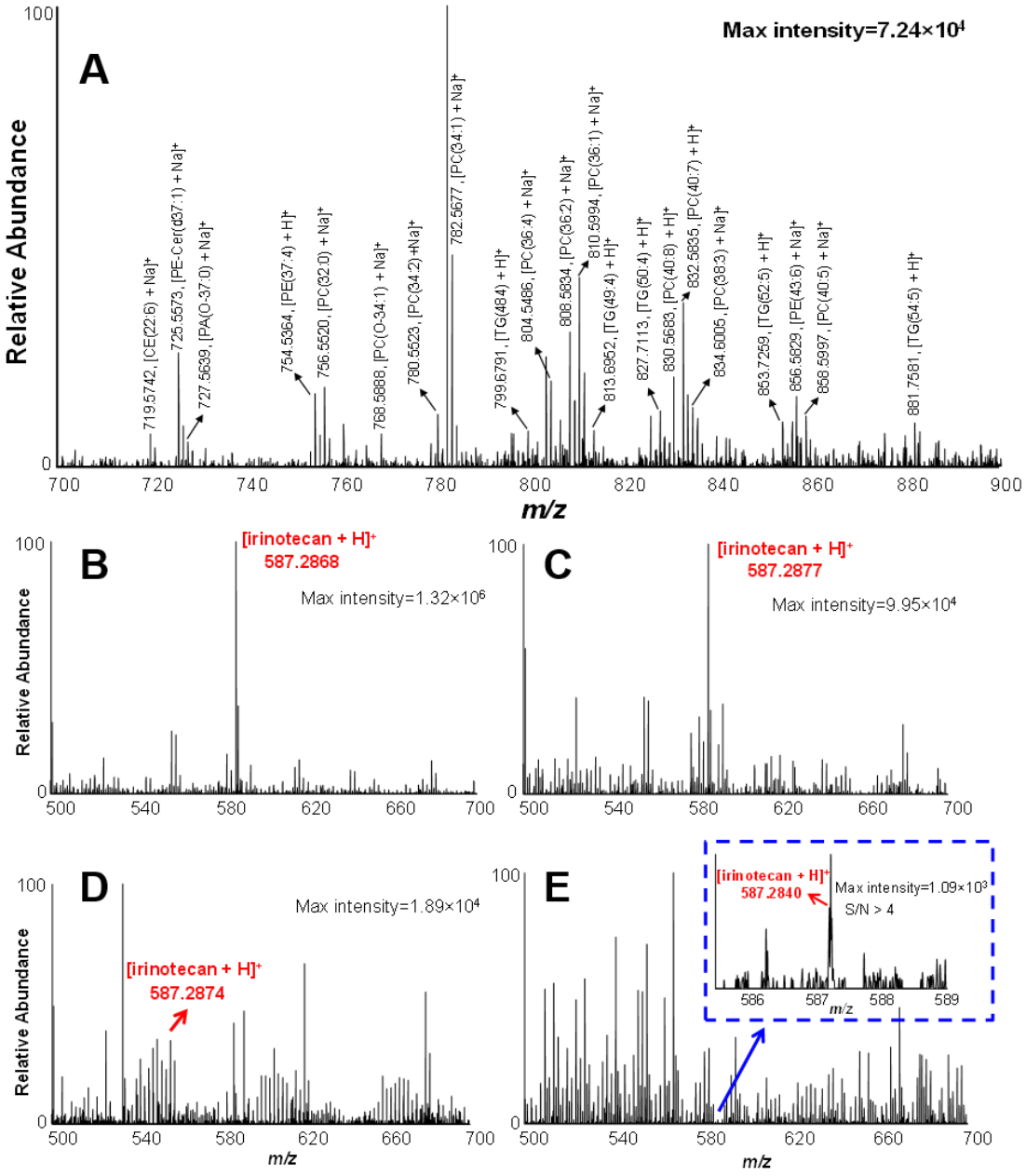

Figure 2.3. Mass spectra obtained from SCMS experiments using the T-probe device. (A) A zoomed-in view of tentatively labeled cellular species in a mass range of *m/z* = 700–900. (B–E) Metabolic profiles of single cells treated for 1 h using irinotecan at concentrations: (B) 10 µM, (C) 1 µM, (D) 100 nM, and (E) 10 nM. Drug target is highlighted in red.

2.9. Changes of Metabolic Profiles after Drug Treatment

We conducted SCMS data analysis to gain biological insights into cellular response to the altered microenvironment due to drug treatment. Generally, we obtained SCMS data sets from both the control (9 cells) and the drug treatment group (11 cells, treated using 1 μM irinotecan for 1 h), and performed multivariate and univariate analyses. Typically, we applied OPLS-DA to illustrate group discrimination and visualize high dimensional cellular profiles in lower dimensional space, as shown on the OPLS-DA score plot (Figure 2.4A). The statistical rationale behind this approach is that the between-group variation (i.e., differences of cellular profile between control and treatment group) is represented by the predictive component T[1], whereas the within-group variation (rendered by cell-to-cell heterogeneity) is represented by the orthogonal component To[1].[55] A decent separation between cells from two groups can be achieved, indicating cellular profiles were significantly altered by drug treatment. The visualized separation obtained from the OPLS-DA model was well explained by the variation of each orthogonal component ($R^2X(cum) = 0.798$, $R^2Y(cum) = 0.973$) with high predictive ability ($Q^2(cum) = 0.701$). We further performed permutation tests[56] for the established model to evaluate the potential for model overfitting. Our results suggest that irinotecan treatment can induce a rapid change of cellular profiles (e.g., within 1 h upon treatment),[57] and our model has no overfitting.

In addition to the shifted metabolic profiles, we further discovered a number of cellular species with both high variable magnitude (p[1]) and high reliability (p(corr)[1]) in the corresponding loading plot (S-plot, Figure 2.4B). These species located on both "wings" of the loading plot have major contributions to the observed groups separation,[58] and therefore, they are species sensitive to drug treatment and can be regarded as potential biomarkers reflecting the treatment efficacy.[59] We performed two-sample *t*-test for each metabolite and we found 17 species with testing *p*-values

< 0.05 (as marked in Figure 2.4B), indicating statistically significant change of abundance due to drug treatment. Among those, the protonated pyridoxamine-5'-phosphate ($[PMP + H]^+$, *m/z* 249.0640, shown in Figure 2.3B) and sodiated nicotinamide adenine dinucleotide ($[NAD + Na]^+$, *m/z* 686.0989, shown in Figure 2.3C) were detected along with other common cellular species. They were tentatively assigned with relatively high confidence using the combined information such as accurate mass, isotopic distribution, and database search. PMP is a precursor of pyridoxal-5'-phosphate (PLP), a coenzyme regulating a variety of intracellular pathways, and it was reported as a biomarker of cancer risk.[60] NAD, as a coenzyme that maintains energy homeostasis and mediates cell signaling pathways, was reported to be a pharmaceutical target for cancer treatment.[61-62] Our findings suggest that cellular metabolites, such as PMP and NAD, can be closely monitored upon irinotecan treatment at single cell level, and they may be potential biomarkers to indicate drug efficacy (inset of Figure 2.4B).

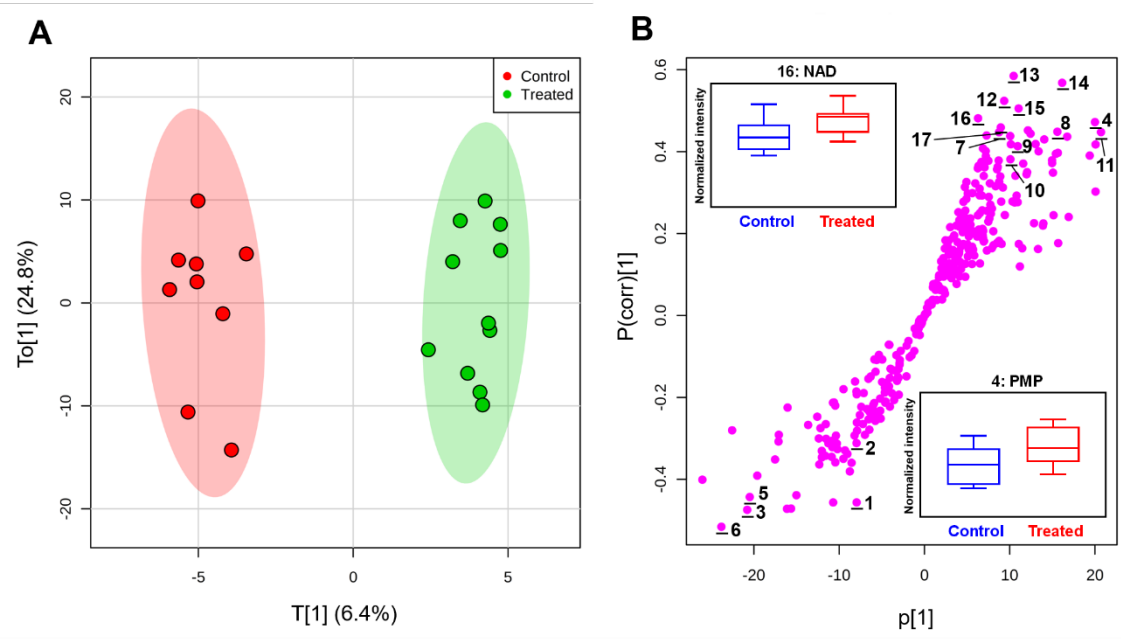

Figure 2.4. Multivariate analysis of metabolic profiles of single cells using OPLS-DA. (A) Score plot of OPLS-DA model shows group separation between the control and drug treated single cells

(T[1]) as well as intrinsic cell-to-cell heterogeneity (To[1]). (B) Loading plot (S-plot) of OPLS-DA model indicates cellular species significantly changed by drug treatment (primarily located on two "wings" of the S-plot). These species were revealed using two sample *t*-test and numbered as 1–17. Inset: box plots of two tentatively assigned species (PMP and NAD) present significant abundance changes (*t*-test p-value < 0.05) upon drug treatment.

2.10. Conclusion

We designed and fabricated a microscale multifunctional device, the T-probe, to facilitate online *in situ* SCMS analysis and metabolic profiling at single cell level in ambient conditions. We demonstrated high sensitivity of the T-probe in SCMS analysis, and interrogated cellular contents using our in-house developed platform. Metabolic profiles of single HeLa cells under control and a series of drug treatment conditions were obtained and compared. In addition, online MS/MS analyses were performed to identify abundant species of interest (i.e., cellular metabolites, drug target). We observed that metabolic profiles of cells were significantly changed by the treatment of anticancer drug irinotecan, and discovered cellular species sensitive to drug treatment using statistical analyses. Those species, among which are PMP and NAD, can be potentially used as early prognostic indicators for treatment efficacy, and arouse a great interest in biological and pharmaceutical studies. In the future, the T-probe can be potentially coupled with broader types of ambient MS instruments to study fundamental cell biology, perform single cell metabolomics, and develop novel chemotherapies at single cell or even subcellular level.

Reference

(1) Gawad, C.; Koh, W.; Quake, S. R., Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **2016,** *17* (3), 175-188.

(2) Folch, A.; Toner, M., Microengineering of Cellular Interactions. *Annu. Rev. Biomed. Eng.* **2000,** *2* (1), 227-256.

(3) Bakker, E.; Qattan, M.; Mutti, L.; Demonacos, C.; Krstic-Demonacos, M., The role of microenvironment and immunity in drug response in leukemia. *BBA - Mol. Cell Res.* **2016,** *1863* (3), 414-426.

(4) Azegrouz, H.; Karemore, G.; Torres, A.; Alaíz, C. M.; Gonzalez, A. M.; Nevado, P.; Salmerón, A.; Pellinen, T.; del Pozo, M. A.; Dorronsoro, J. R.; Montoya, M. C., Cell-Based Fuzzy Metrics Enhance High-Content Screening (HCS) Assay Robustness. *J. Biomol. Screen.* **2013,** *18* (10), 1270-1283.

(5) Liras, M.; Simoncelli, S.; Rivas-Aravena, A.; Garcia, O.; Scaiano, J. C.; Alarcon, E. I.; Aspee, A., Nitroxide amide-BODIPY probe behavior in fibroblasts analyzed by advanced fluorescence microscopy. *Org. Biomol. Chem.* **2016,** *14* (17), 4023-4026.

(6) Santana-Filho, A. P. d.; Jacomasso, T.; Riter, D. S.; Barison, A.; Iacomini, M.; Winnischofer, S. M. B.; Sassaki, G. L., NMR metabolic fingerprints of murine melanocyte and melanoma cell lines: application to biomarker discovery. *Sci. Rep.* **2017,** *7*, 42324.

(7) Nehme, H.; Nehme, R.; Lafite, P.; Duverger, E.; Routier, S.; Morin, P., Electrophoretically mediated microanalysis for in-capillary electrical cell lysis and fast enzyme quantification by capillary electrophoresis. *Anal. Bioanal. Chem.* **2013,** *405* (28), 9159-67.

(8) Walker, B. N.; Stolee, J. A.; Vertes, A., Nanophotonic Ionization for Ultratrace and Single-Cell Analysis by Mass Spectrometry. *Anal. Chem.* **2012,** *84* (18), 7756-7762.

(9) Zenobi, R., Single-Cell Metabolomics: Analytical and Biological Perspectives. *Science* **2013,** *342* (6163), 1243259.

(10) Amantonico, A.; Urban, P. L.; Zenobi, R., Analytical techniques for single-cell metabolomics: state of the art and trends. *Anal. Bioanal. Chem.* **2010,** *398* (6), 2493-2504.

(11) Wang, Y. N.; Park, D.; Galermo, A. G.; Gao, D.; Liu, H. X.; Lebrilla, C. B., Changes in cellular glycosylation of leukemia cells upon treatment with acridone derivatives yield insight into drug action. *Proteomics* **2016,** *16* (23), 2977-2988.

(12) Jimmerson, L. C.; Bushman, L. R.; Ray, M. L.; Anderson, P. L.; Kiser, J. J., A LC-MS/MS Method for Quantifying Adenosine, Guanosine and Inosine Nucleotides in Human Cells. *Pharm. Res.* **2017,** *34* (1), 73-83.

(13) Gu, L. Q.; Hou, P. Y.; Zhang, R. W.; Liu, Z. Y.; Bi, K. S.; Chen, X. H., An analytical strategy to investigate Semen Strychni nephrotoxicity based on simultaneous HILIC-ESI-MS/MS detection of Semen Strychni alkaloids, tyrosine and tyramine in HEK 293t cell lysates. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2016,** *1033*, 157-165.

(14) van Kampen, J. J. A.; Burgers, P. C.; Gruters, R. A.; Osterhaus, A. D. M. E.; de Groot, R.; Luider, T. M.; Volmer, D. A., Quantitative Analysis of Antiretroviral Drugs in Lysates of Peripheral Blood Mononuclear Cells Using MALDI-Triple Quadrupole Mass Spectrometry. *Anal. Chem.* **2008,** *80* (13), 4969-4975.

(15) Marjanovic, N. D.; Weinberg, R. A.; Chaffer, C. L., Cell Plasticity and Heterogeneity in Cancer. *Clin. Chem.* **2013,** *59* (1), 168-179.

(16) Lanekoff, I.; Kurczy, M. E.; Adams, K. L.; Malm, J.; Karlsson, R.; Sjovall, P.; Ewing, A. G., An in situ fracture device to image lipids in single cells using ToF-SIMS. *Surf. Interface Anal.* **2011,** *43* (1-2), 257-260.

(17) Chandra, S.; Smith, D. R.; Morrison, G. H., Peer Reviewed: A Subcellular Imaging by Dynamic SIMS Ion Microscopy. *Anal. Chem.* **2000,** *72* (3), 104 A-114 A.

(18) Xie, W.; Gao, D.; Jin, F.; Jiang, Y.; Liu, H., Study of Phospholipids in Single Cells Using an Integrated Microfluidic Device Combined with Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal. Chem.* **2015,** *87* (14), 7052-7059.

(19) Rubakhin, S. S.; Garden, R. W.; Fuller, R. R.; Sweedler, J. V., Measuring the peptides in individual organelles with mass spectrometry. *Nat. Biotechnol.* **2000,** *18* (2), 172-175.

(20) Ibáñez, A. J.; Fagerer, S. R.; Schmidt, A. M.; Urban, P. L.; Jefimovs, K.; Geiger, P.; Dechant, R.; Heinemann, M.; Zenobi, R., Mass spectrometry-based metabolomics of single yeast cells. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (22), 8790.

(21) Stolee, J. A.; Vertes, A., Toward Single-Cell Analysis by Plume Collimation in Laser Ablation Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2013,** *85* (7), 3592-3598.

(22) Shrestha, B.; Vertes, A., In Situ Metabolic Profiling of Single Cells by Laser Ablation Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2009,** *81* (20), 8265-8271.

(23) Shrestha, B.; Nemes, P.; Vertes, A., Ablation and analysis of small cell populations and single cells by consecutive laser pulses. *Appl. Phys. A Mater. Sci. Process.* **2010,** *101* (1), 121-126.

(24) Kruse, R. A.; Rubakhin, S. S.; Romanova, E. V.; Bohn, P. W.; Sweedler, J. V., Direct assay of Aplysia tissues and cells with laser desorption/ionization mass spectrometry on porous silicon. *J. Mass Spectrom.* **2001,** *36* (12), 1317-1322.

(25) Mizuno, H.; Tsuyama, N.; Harada, T.; Masujima, T., Live single-cell video-mass spectrometry for cellular and subcellular molecular detection and cell classification. *J. Mass Spectrom.* **2008,** *43* (12), 1692-1700.

(26) Tejedor, M. L.; Mizuno, H.; Tsuyama, N.; Harada, T.; Masujima, T., Direct Single-Cell Molecular Analysis of Plant Tissues by Video Mass Spectrometry. *Anal. Sci.* **2009,** *25* (9), 1053-1055.

(27) Gong, X. Y.; Zhao, Y. Y.; Cai, S. Q.; Fu, S. J.; Yang, C. D.; Zhang, S. C.; Zhang, X. R., Single Cell Analysis with Probe ESI-Mass Spectrometry: Detection of Metabolites at Cellular and Subcellular Levels. *Anal. Chem.* **2014,** *86* (8), 3809-3816.

(28) Gholipour, Y.; Erra-Balsells, R.; Hiraoka, K.; Nonami, H., Living cell manipulation, manageable sampling, and shotgun picoliter electrospray mass spectrometry for profiling metabolites. *Anal. Biochem.* **2013,** *433* (1), 70-78.

(29) Zhu, H.; Zou, G.; Wang, N.; Zhuang, M.; Xiong, W.; Huang, G., Single-neuron identification of chemical constituents, physiological changes, and metabolism using mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114* (10), 2586.

(30) Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R. M.; Burgett, A. W. G.; Yang, Z. B., The Single-Probe: A Miniaturized Multifunctional Device for Single Cell Mass Spectrometry Analysis. *Anal. Chem.* **2014,** *86* (19), 9376-9380.

(31) Pan, N.; Rao, W.; Standke, S. J.; Yang, Z., Using Dicationic Ion-Pairing Compounds To Enhance the Single Cell Mass Spectrometry Analysis Using the Single-Probe: A Microscale Sampling and Ionization Device. *Anal. Chem.* **2016,** *88* (13), 6812-6819.

(32) Onjiko, R. M.; Moody, S. A.; Nemes, P., Single-cell mass spectrometry reveals small molecules that affect cell fates in the 16-cell embryo. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112* (21), 6545-6550.

(33) Onjiko, R. M.; Morris, S. E.; Moody, S. A.; Nemes, P., Single-cell mass spectrometry with multi-solvent extraction identifies metabolic differences between left and right blastomeres in the 8-cell frog (Xenopus) embryo. *Analyst* **2016,** *141* (12), 3648-3656.

(34) Zhang, L. W.; Vertes, A., Energy Charge, Redox State, and Metabolite Turnover in Single Human Hepatocytes Revealed by Capillary Microsampling Mass Spectrometry. *Anal. Chem.* **2015,** *87* (20), 10397-10405.

(35) Sun, M.; Yang, Z. B.; Wawrik, B., Metabolomic Fingerprints of Individual Algal Cells Using the Single-Probe Mass Spectrometry Technique. *Front. Plant Sci.* **2018,** *9*, 571.

(36) Rao, W.; Pan, N.; Tian, X.; Yang, Z. B., High-Resolution Ambient MS Imaging of Negative Ions in Positive Ion Mode: Using Dicationic Reagents with the Single-Probe. *J. Am. Soc. Mass Spectrom.* **2016,** *27* (1), 124-134.

(37) Rao, W.; Pan, N.; Yang, Z., Applications of the Single-probe: Mass Spectrometry Imaging and Single Cell Analysis under Ambient Conditions. *J. Vis. Exp.* **2016,** (112), e53911.

(38) Sun, M.; Tian, X.; Yang, Z., Microscale Mass Spectrometry Analysis of Extracellular Metabolites in Live Multicellular Tumor Spheroids. *Anal. Chem.* **2017,** *89* (17), 9069-9076.

(39) ElNaggar, M. S.; Barbier, C.; Van Berkel, G. J., Liquid Microjunction Surface Sampling Probe Fluid Dynamics: Computational and Experimental Analysis of Coaxial Intercapillary Positioning Effects on Sample Manipulation. *J. Am. Soc. Mass Spectrom.* **2011,** *22* (7), 1157.

(40) Roach, P. J.; Laskin, J.; Laskin, A., Nanospray desorption electrospray ionization: an ambient method for liquid-extraction surface sampling in mass spectrometry. *Analyst* **2010,** *135* (9), 2233-2236.

(41) Jang, M.; Park, C. K.; Lee, N. Y., Modification of polycarbonate with hydrophilic/hydrophobic coatings for the fabrication of microdevices. *Sens. Actuators B: Chem.* **2014,** *193* (Supplement C), 599-607.

(42) Lanekoff, I.; Heath, B. S.; Liyu, A.; Thomas, M.; Carson, J. P.; Laskin, J., Automated Platform for High-Resolution Tissue Imaging Using Nanospray Desorption Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2012,** *84* (19), 8351-8356.

(43) Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J., OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* **2006,** *20* (8-10), 341-351.

(44) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN - A metabolite mass spectral database. *Ther. Drug Monit.* **2005,** *27* (6), 747-751.

(45) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A., HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res.* **2013,** *41* (Database issue), D801-D807.

(46) Mastro, A. M.; Babich, M. A.; Taylor, W. D.; Keith, A. D., Diffusion of a Small Molecule in the Cytoplasm of Mammalian-Cells. *Proc. Natl. Acad. Sci. U. S. A.* **1984,** *81* (11), 3414-3418.

(47) Tang, K.; Page, J. S.; Smith, R. D., Charge competition and the linear dynamic range of detection in electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2004,** *15* (10), 1416-1423.

(48) Schober, Y.; Guenther, S.; Spengler, B.; Römpp, A., Single Cell Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging. *Anal. Chem.* **2012,** *84* (15), 6293-6297.

(49) Guillaume-Gentil, O.; Grindberg, R. V.; Kooger, R.; Dorwling-Carter, L.; Martinez, V.; Ossola, D.; Pilhofer, M.; Zambelli, T.; Vorholt, J. A., Tunable Single-Cell Extraction for Molecular Analyses. *Cell* **2016,** *166* (2), 506-516.

(50) Wallin, A.; Svanvik, J.; Holmlund, B.; Ferreud, L.; Sun, X. F., Anticancer effect of SN-38 on colon cancer cell lines with different metastatic potential. *Oncol. Rep.* **2008,** *19* (6), 1493-1498.

(51) Liu, X.; Weaver, E. M.; Hummon, A. B., Evaluation of Therapeutics in Three-Dimensional Cell Culture Systems by MALDI Imaging Mass Spectrometry. *Anal. Chem.* **2013,** *85* (13), 6295-6302.

(52) Fujita, D.; Saito, Y.; Nakanishi, T.; Tamai, I., Organic Anion Transporting Polypeptide (OATP)2B1 Contributes to Gastrointestinal Toxicity of Anticancer Drug SN-38, Active Metabolite of Irinotecan Hydrochloride. *Drug Metab. Dispos.* **2016,** *44* (1), 1.

(53) Takeba, Y.; Kumai, T.; Matsumoto, N.; Nakaya, S.; Tsuzuki, Y.; Yanagida, Y.; Kobayashi, S., Irinotecan Activates p53 With Its Active Metabolite, Resulting in Human Hepatocellular Carcinoma Apoptosis. *J. Pharmacol. Sci.* **2007,** *104* (3), 232-242.

(54) Zhang, X.-C.; Wei, Z.-W.; Gong, X.-Y.; Si, X.-Y.; Zhao, Y.-Y.; Yang, C.-D.; Zhang, S.-C.; Zhang, X.-R., Integrated Droplet-Based Microextraction with ESI-MS for Removal of Matrix Interference in Single-Cell Analysis. *Sci. Rep.* **2016,** *6*, 24730.

(55) Lundmark, A.; Davanian, H.; Båge, T.; Johannsen, G.; Koro, C.; Lundeberg, J.; Yucel-Lindberg, T., Transcriptome analysis reveals mucin 4 to be highly associated with periodontitis and identifies pleckstrin as a link to systemic diseases. *Sci. Rep.* **2015,** *5*, 18475.

(56) Triba, M. N.; Le Moyec, L.; Amathieu, R.; Goossens, C.; Bouchemal, N.; Nahon, P.; Rutledge, D. N.; Savarin, P., PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol. Biosyst.* **2015,** *11* (1), 13-19.

(57) Wang, J.; Hwang, K.; Braas, D.; Dooraghi, A.; Nathanson, D.; Campbell, D. O.; Gu, Y.; Sandberg, T.; Mischel, P.; Radu, C.; Chatziioannou, A. F.; Phelps, M. E.; Christofk, H.; Heath, J. R., Fast Metabolic Response to Drug Intervention Through Analysis on a Miniaturized, Highly Integrated Molecular Imaging System. *J. Nucl. Med.* **2013,** *54* (10), 1820-1824.

(58) Stolee, J. A.; Shrestha, B.; Mengistu, G.; Vertes, A., Observation of Subcellular Metabolite Gradients in Single Cells by Laser Ablation Electrospray Ionization Mass Spectrometry. *Angew. Chem. Int. Ed.* **2012,** *51* (41), 10386-10389.

(59) Luies, L.; du Preez, I.; Loots, D. T., The role of metabolomics in tuberculosis treatment research. *Biomark. Med.* **2017,** *11* (11), 1017-1029.

(60) Galluzzi, L.; Vacchelli, E.; Michels, J.; Garcia, P.; Kepp, O.; Senovilla, L.; Vitale, I.; Kroemer, G., Effects of vitamin B6 metabolism on oncogenesis, tumor progression and therapeutic responses. *Oncogene* **2013,** *32* (42), 4995-5004.

(61) Cantó, C.; Menzies, K.; Auwerx, J., NAD(+) metabolism and the control of energy homeostasis - a balancing act between mitochondria and the nucleus. *Cell Metabolism* **2015,** *22* (1), 31-53.

(62) Kennedy, B. E.; Sharif, T.; Martell, E.; Dai, C.; Kim, Y.; Lee, P. W. K.; Gujar, S. A., NAD+ salvage pathway in cancer metabolism and therapy. *Pharmacol. Res.* **2016,** *114*, 274-283.

Chapter 3: Integrating a Generalized Data Analysis Workflow with the Single-probe Mass Spectrometry Experiment for Single Cell Metabolomics

3.1. Introduction

Cell, as a fundamental component of living organisms, regulates cellular metabolic activities through a variety of biological pathways.[1,2] In recent years, a tremendous number of metabolites that participate in rapid and subtle biological and physiological activities[3,4] were intensively investigated to gain a profound perspective towards the dynamic nature of the cell.[5,6] Metabolomics, serving as a bridge between cellular metabolism and phenotypes,[7,8] becomes an increasingly intriguing research field where modern instrumentation and methodologies are involved.[9,10] Among all techniques for metabolomics studies, mass spectrometry (MS) based approaches possess considerable advantages over others for providing a large amount of molecular information from complex samples. MS methods are widely used in metabolomic studies due to their high sensitivity to detect low-abundance cellular metabolites,[11] high mass resolution to resolve isobaric species,[12] flexible capabilities to be coupled to versatile chromatographic separations to enhance metabolite coverage,[13,14] and wide selections of orthogonal yet compatible analytical methods to discriminant isomers.[15,16] To date, the majority of current MS based cell metabolomic studies are carried out by analyzing cell lysates prepared from a large cohort of cells, and consequently, leading to an accumulative result of populations analyzed.[17] However, each cell is an individually functional unit that is encoded with heterogeneous genomic information, and presents diverse biological status in different microenvironment.[18] Single cell MS (SCMS), as an emerging field of study, appreciates such cell-to-cell heterogeneity masked by conventional liquid chromatography-mass spectrometry

(LC-MS) methods through interrogating cellular contents of individual cells. A number of SCMS techniques have been dedicatedly developed, and they are roughly classified as ion-beam based,[19] laser based,[20-23] probe based,[24-28] and other techniques.[29-32] Typically, they have been employed for distinguishing cellular fingerprints, identifying intracellular metabolites, and discovering new biological mechanisms through single cell metabolomic analysis (i.e., single cell metabolomics).[33,34] However, to the best of our knowledge, the majority of reported single cell metabolomic studies rely on the non-specialized software, which is either vendor-specific (MassLynx,[35,36] Compass Data Analysis,[29,37] etc.) or derived from conventional LC-MS analysis (e.g., Decon2LS),[38] to process the datasets. Therefore, further efforts are needed to establish the standardized data analysis procedure for the single cell metabolomic analysis of data obtained from broader types of MS instruments. On the other hand, cells are sensitive to their surrounding microenvironment, and cellular metabolites have rapid turnover rate upon subtle changes,[39] which adds another layer of complexity to single cell metabolomics in native status. Facing those challenges, it is imperative to develop a comprehensive single cell metabolomics approach consisting of SCMS experiments and a generalized pipeline for SCMS metabolomic data analysis. Ultimately, a fully developed single cell metabolomics method can be used to capture metabolomic signatures of individual cells, identify metabolic phenotypes, and disclose underlying biological principles of live single cells.

In our single cell metabolomics approach, we used a miniaturized multifunctional sampling device, the Single-probe,[27,40-45] coupled to MS to analyze live single cells in ambient conditions, followed by multivariate and univariate data analysis. We selected human cervical cancer cell line, HeLa, as our model system, to demonstrate changes of metabolomic profile of each cell upon exposure to external stimuli (i.e., anticancer drugs). Specifically, two types of mitotic

inhibitors, paclitaxel (taxol) and vinblastine, were selected for a series of time- and concentration-dependent treatments. Both taxol and vinblastine inhibit cell mitotic process in G2/M phase by either stabilizing (taxol) or destabilizing (vinblastine) microtubules, and ultimately induce cell apoptosis.[46,47] Although both drug compounds share similarities such as the binding target (microtubules) and IC_{50} values,[48] their influence on the cellular metabolism needs to be further understood at the single cell level.

Similar to the untargeted LC-MS metabolomics data handling procedures, our SCMS metabolomic data processing aims to discriminate metabolic phenotypes, discover phenotypic biomarkers (i.e., characteristic species closely related to specific phenotypes), and unveil related biological pathways. However, due to the nature of cell heterogeneity, each cell may have a different response to drug treatment resulting in varied metabolomic profiles. Therefore, it is impractical to directly apply the conventional metabolomic data analysis procedure to SCMS datasets, as the underlying assumption (i.e., homogeneity of variance) of a variety of statistical tests are challenged. Here, we developed a comprehensive approach to SCMS metabolomics studies by performing data pre-treatment, visualization, statistical analysis, machine learning, and pathway enrichment analysis (Figure 3.1).

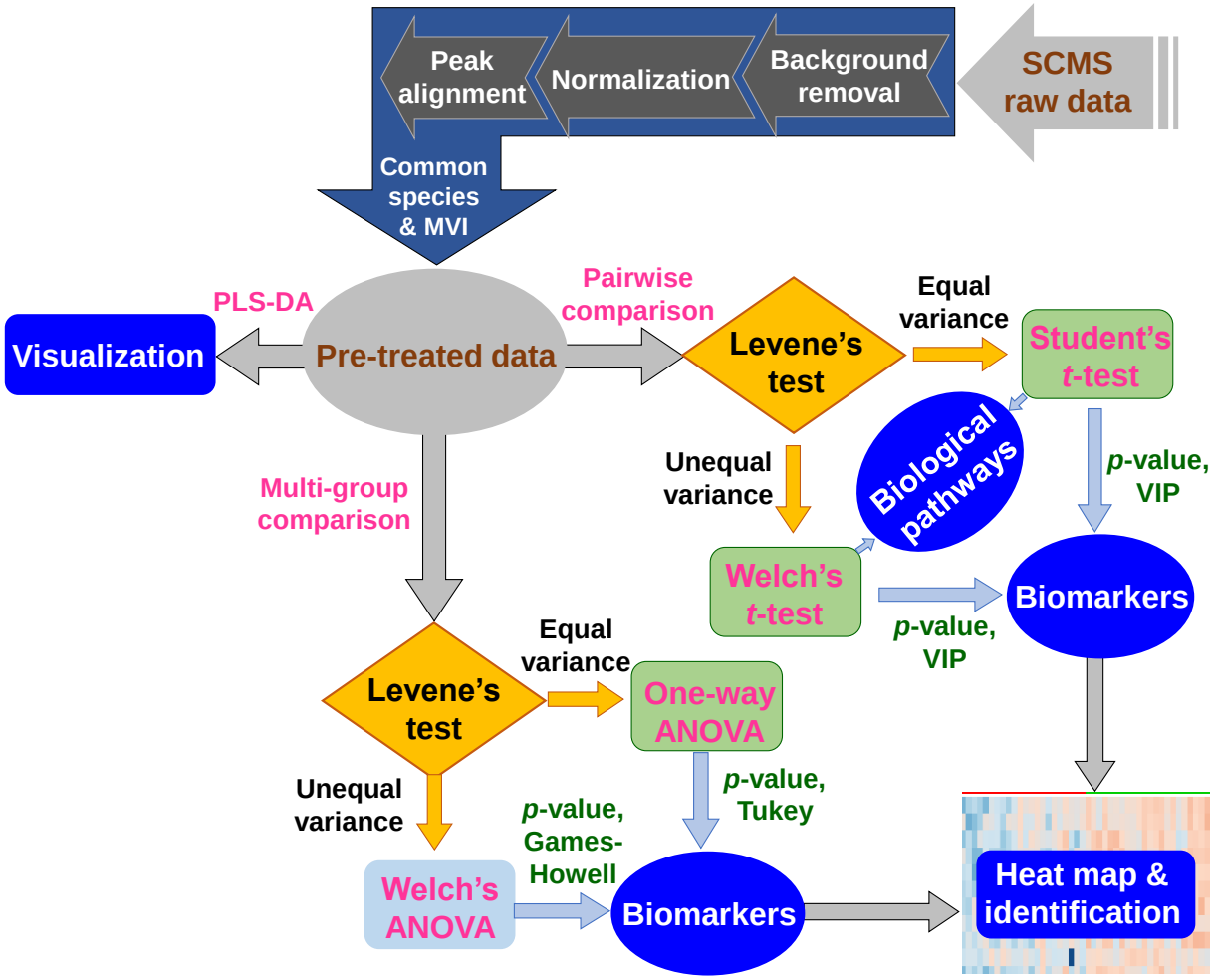

Figure 3.1. Workflow of SCMS data analysis consisting of data pre-treatment, multivariate analysis, and univariate analysis. This generalized procedure can be coupled to raw datasets obtained from broader types of SCMS platforms for single cell metabolomic analysis.

3.2. SCMS data acquisition

Detailed fabrication procedure and working mechanisms of the Single-probe device are provided in our previous publications.[27,40] To conduct the SCMS experiment, the Single-probe device is coupled to a Thermo LTQ Orbitrap XL mass spectrometer. The tip (size < 10 μm) of the Single-probe is inserted into a target cell to extract cellular contents through a liquid junction at the probe tip, and then the extracted mixture is driven towards the nano-ESI emitter for

immediate ionization and MS detection (Figure 3.2B). Cell selection and penetration are precisely controlled by our in-house built XYZ-translational stage system (Figure 3.2A), and these processes are visualized using a stereo microscope (Figure 3.2C). The experimental MS parameters are listed as follows: ionization voltage +4.5 kV, mass range 150–1500 *m/z* (mass-to-charge ratio), mass resolution 60,000 at *m/z* 400, 1 microscan, and 100 ms max injection time and automatic gain control (AGC) on.

Table 3.1. HeLa cells in the control and treatment groups for SCMS experiments.

Condition	Drug	Concentration (μM)	Time (h)	Group Name	Number of Cells
Control	N/A	N/A	N/A	Control	23
Treatment	Taxol	0.1	2	TaxA	25
		0.1	6	TaxB	28
		1.0	2	TaxC	22
	Vinblastine	0.1	2	VinA	28
		0.1	6	VinB	23
		1.0	2	VinC	24

To study changes of metabolomic profiles of cancer cells induced by taxol and vinblastine, we cultured HeLa cells under normal condition (control), and treated them using a series of drug treatment conditions (Table 3.1). Individual cells in both control and treatments groups were randomly selected for analysis using the Single-probe SCMS technique. We carefully designed our treatment conditions allowing for sufficient cellular metabolomic changes to be detected, while minimizing other factors (environmental perturbations, mutations, etc.) that

could potentially interfere with phenotypic identification. In our experiments, 22–28 cells were sampled from the control group and each of those drug treatment groups.

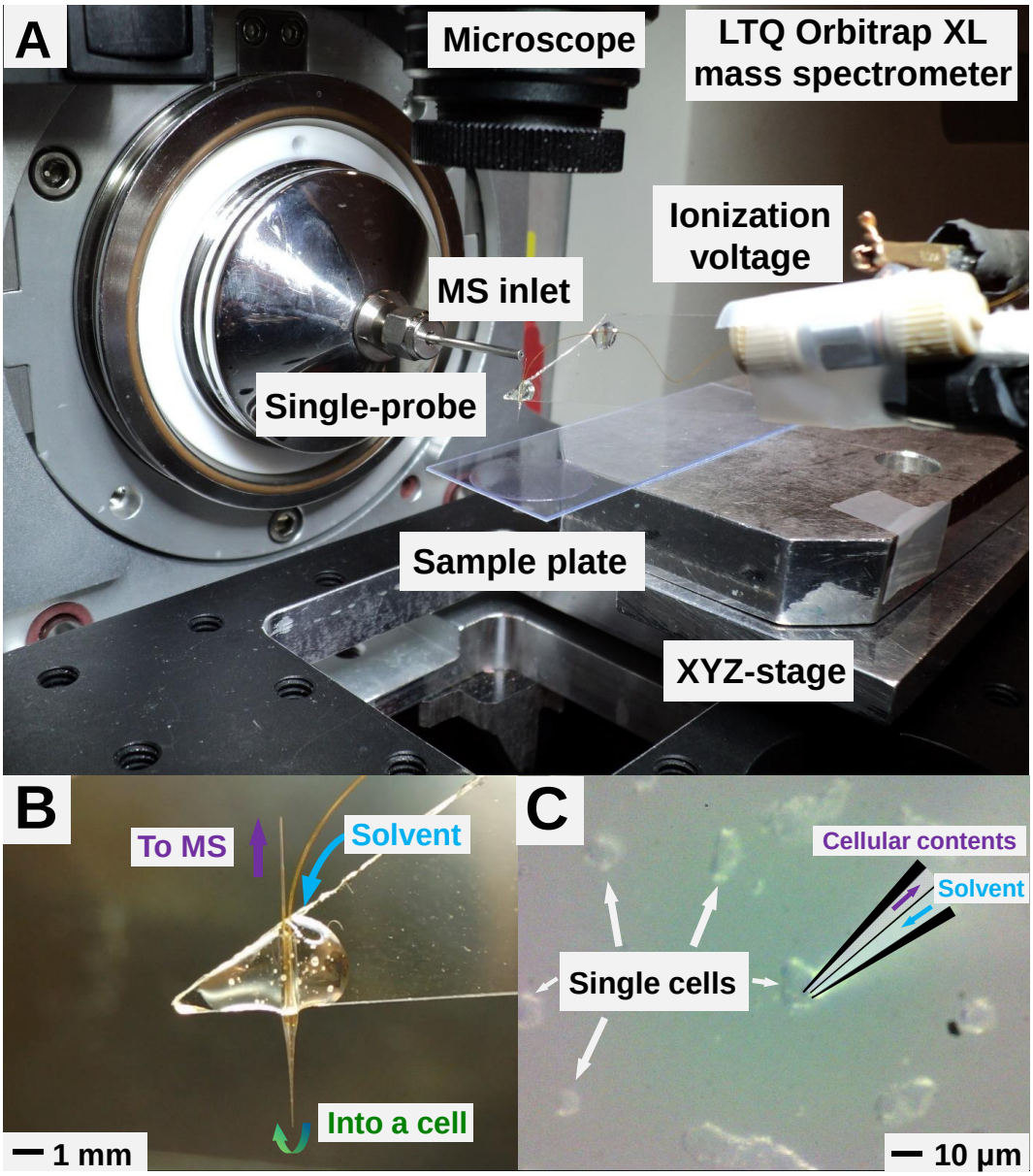

Figure 3.2. Experimental setup of the Single-probe SCMS system. (A) Key components of the in-house developed Single-probe SCMS platform. (B) A zoomed-in photo of the Single-probe and illustration of its working mechanism. (C) The insertion of the Single-probe tip into a single cell monitored using a high-resolution digital stereo microscope during a SCMS experiment.

3.3. SCMS data pre-treatment

Following online data acquisition, we performed a generalized comprehensive SCMS metabolomics data analysis, including multivariate and univariate analysis, to gain biological insights into raw data matrices (Figure 3.1). Particularly, to preserve metabolomic information of endogenous species from single cells while avoiding interference with other species (exogenous species from the sampling environment, detection noise, etc.), we conducted data pre-treatment that can be generally divided into three consecutive steps.

3.3.1. Generation of metabolomic peak list

The acquired raw data files (.raw) from our SCMS experiments were accessed using Xcalibur 3.0 (Thermo Fisher Scientific). A common cellular species with relatively high ion intensity, PC (34:1),[27] was selected as an indicator of successful MS detection of cellular contents from individual cells. We exported an averaged MS spectrum from each cell containing all detected peaks (i.e., *m/z* values) along with their corresponding ion intensities as the metabolic peak list. Similar lists of metabolites can be generated from all other major types of MS platforms, including quadrupole time-of-flight (qTOF), fourier transform-ion cyclotron resonance (FT-ICR), and Orbitraps, using vendor-specific software.

3.3.2. MS background removal

Under our experimental conditions, a raw file typically consists of more than 6,000 distinct peaks, which can be attributed to endogenous species (i.e., cellular metabolites), exogeneous species (i.e., from surrounding matrix such as cell culture medium and sampling solvent), and instrument noise. Based on the data obtained from 10 randomly selected cells, we estimated the total ion current (TIC) of exogenous species and noise are ~11 fold higher than that of

endogenous species. This result could be attributed to both the extremely limited amount of analytes within a single cell (in picoliter range)[39,49] and the reduced detection sensitivity due to the matrix effect.[50] Because only the relative abundances of cellular metabolites from single cells were used in the downstream analysis, we excluded interfering ion as described below. First, we removed ion signals of exogenous species (i.e., background ion signals), which were detected from cell culture medium and the sampling solvent used in SCMS experiments. Second, we filtered out instrument noise, which may result in false positive discovery and unnecessary computational burden to data analysis. Instrument noise accounts for ~20–40% of the total number of peaks detected in SCMS experiments, and it was removed by eliminating ions with evidently lower ion intensities ($< 10^3$). Removing background and noise greatly reduced the dimensionality of SCMS data matrices and preserved the molecular information of endogenous cellular metabolites. Lastly, we normalized the ion intensity of each metabolite to TIC prior to the following data processing steps. It is worth noting that our background removal method is similar to those used in prevalent LC-MS metabolomics data analysis software (e.g., MZmine 2). However, discriminating instrument noise from low-abundance MS peaks of cellular metabolites is challenging. Other advanced noise removal algorithms, such as repetition rate filtering (RRF) that has been demonstrated effective in shotgun lipidomics,[51] can be incorporated in future studies.

3.3.3. Peak alignment and common species determination

The SCMS datasets obtained from the previous step were submitted to Geena 2[52] for MS peak alignment. We then utilized MetaboAnalyst[53,54] to determine the common species, which are defined as cellular species that can be frequently detected from measured cells in each group. Here, we referred to the standard 80% rule (i.e., excluding species with > 20% missing values

from all measured cells), a broadly accepted rule for feature selection in untargeted LC-MS metabolomics research,[55] as the criterion to determine common species. In addition, a missing value imputation (MVI) algorithm, K-nearest neighbor (KNN),[56] was employed to eliminate missing values and reduce false positive results in our analysis. Using the above data pre-treatment procedures, we promptly reduced the size of our datasets while retaining the essential metabolomic information from individual cells. However, this 80% rule may eliminate rare cells, which can be critical for a variety of biological mechanisms present in a large population of cells.[57,58] To include more measured cells for SCMS data analysis, a lenient missing value threshold can be employed. However, applying a loose missing value threshold may reduce the statistical power, introduce bias, and increase computing demand.[59]

3.4. SCMS data visualization

To evaluate the differences of metabolomic profiles of single cells among all groups, we conducted the dimensionality reduction of pre-treated SCMS datasets, which facilitates the visualization of high-dimensional data matrices in a low-dimensional space through multivariate analysis. Here, we employed Partial Least Square-Discriminant Analysis (PLS-DA), a supervised method, to achieve phenotypic separation when the within-group variation (i.e., variation of cellular metabolite abundance within the control and each of the treatment groups) is pronounced.[60] To evaluate the quality of PLS-DA models and avoid data overfitting, the explained variation (R^2) and the predictive ability (Q^2) were calculated through a 10-fold cross validation procedure.[61] PLS-DA models with $Q^2 > 0.5$ were considered to be robust,[62] and they were further analyzed using permutation tests[63] to identify significantly separated phenotypes. We performed 2,000 permutation tests for each model, and a small statistic *p*-value (< 0.05) indicated a significant phenotypic discrimination.

3.5. Discovery of phenotypic biomarkers

To study changes of metabolomic profiles of single cells induced by microenvironmental alternation (i.e., drug treatment) and discover phenotypic biomarkers, we utilized statistical methods to process SCMS metabolomic datasets after data pre-treatment.

3.5.1. Biomarkers from pairwise group comparison

To discover phenotypic biomarkers corresponding to a particular treatment condition, we utilized the pre-treated SCMS datasets from the control group and that treatment group for the PLS-DA. We then calculated Variable Importance in Projection (VIP) scores for all cellular species, and selected those with VIP scores >1.2, representing major contributions to group discrimination,[64] as biomarker candidates. These biomarker candidates were subsequently subjected to unpaired two-sample *t*-test for the comparison of abundances. Due to cell heterogeneity, cellular response to the drug treatment is different. Therefore, Levene's test was conducted prior to *t*-test to evaluate the homogeneity of variance of each metabolite, allowing us to determine which type of *t*-test to be performed. Depending on the results from Levene's test, we performed Student's *t*-test (data with equal within-group variance) or Welch's *t*-test (data with unequal within-group variance). Cellular species with both VIP scores > 1.2 and *p*-value (from *t*-test) < 0.05 were marked as potential biomarkers related to the examined phenotypes.

3.5.2. Biomarkers from multi-group comparison

To obtain common biomarkers reflecting the influence of drug molecules on cellular metabolism, we compared the ion abundance of detected species in the control and all treatment groups for each drug compound. First, similar to the pairwise group comparison in the previous step, we conducted Levene's test to evaluate the homogeneity of variance for each metabolite

among multiple examined groups. Second, to determine if there are statistically significant differences of metabolites among all groups, we used one-way (data with equal within-group variance) or Welch's (data with unequal within-group variance) Analysis of Variance (ANOVA). Third, to rigorously select biomarkers (i.e., metabolites with significant abundance change), we then performed two types of *post hoc* tests for metabolites with p-value < 0.05 (from ANOVA): Tukey's HSD (Honestly Significant Difference) and Games-Howell tests for one-way ANOVA and Welch's ANOVA tests, respectively. Cellular species with p-values < 0.05 (from both the ANOVA and the corresponding *post hoc* tests) among all examined groups were highlighted as biomarkers. The above procedures have been previously used in LC-MS metabolomics studies to discover biomarkers corresponding to dyslipidemia progression.[65] All above statistical analyses were performed in R environment with functions available in Metabox, a toolbox for metabolomic studies.[66]

3.5.3. Tentative assignment and identification of biomarkers

To identify discovered biomarkers, their accurate *m/z* values were compared with those registered at online metabolome database, METLIN[67] and HMDB[68]. Moreover, we performed online MS/MS analysis of biomarkers with relatively higher abundance at the single cell level, whereas conventional LC-MS/MS experiments were also carried out as a complimentary approach to molecular identification at the population level.

3.6. Potential biological pathways

Mummichog, a program for data analysis in untargeted metabolomics studies,[69] was used in the current work to address potential biological pathways involved in drug treatment at the single cell level. Unlike many other available programs, Mummichog only utilizes the information of

accurate *m/z* values, rather than identified metabolites, to perform pathway enrichment analysis.[70] Required inputs, such as *m/z* values, *t*-test *p*-values, and fold change of all cellular species, were fulfilled based on results from the data analysis as described in the previous steps, and Mummichog was operated using default settings.

3.7. Metabolic response to drug treatment

To study cellular metabolic response and visualize phenotypic separation induced by drug treatment, we constructed PLS-DA models for SCMS datasets collected from the control and each of those drug treatment groups. As shown in Figure 3.3, a data point represents the metabolomic profile of a single cell, and the cell-to-cell heterogeneity can be reflected by the distribution of data points within a group.[71,72] The phenotypic separation can be visualized by the distance of data points between two groups. Generally, the first PLS-DA component explains more than 25% of variance (i.e., Component 1 > 25%) in all score plots, and significant phenotypic discrimination ($p < 0.0035$) between two groups is further demonstrated through permutation tests. However, the overlapped regions can still be observed between the control and a "shorter" treatment time condition (i.e., TaxA, VinA, TaxC, or VinC) for both drugs. In contrast, complete group separation (no overlapped region) can be observed between the control and a "longer" treatment time condition (TaxB or VinB). This trend is also visually reflected on PLS-DA score plots containing multiple groups, in which a complete phenotypic separation is only observed between the control and "longer" treatment time condition. From a biological perspective, though cellular xenobiotic activity was reported to be both time- and concentration-dependent,[73] our SCMS results demonstrate that treatment time has a more significant influence on cells' metabolomic profiles, at least at early treatment stage (e.g., treatment time < 6 h). To validate our SCMS results, we prepared lysates using cells, which were treated under the same

conditions as those in the SCMS experiments, for LC-MS analysis. We further conducted principal component analysis (PCA) of the LC-MS results, and obtained similar trends observed in the SCMS studies: longer treatment time resulted in more evident changes of cellular metabolomic profiles. In addition, we compared the number of metabolites detected using the LC-MS and SCMS approaches. As shown in the Venn diagram, 230 cellular metabolites can only be detected in the SCMS datasets; these metabolites are likely to have rapid turnover rates, and therefore could be potentially lost during the lengthy LC-MS sample preparation process. On the other hand, due to significantly larger amounts of cellular species contained in the cell lysate and chromatographic separation (i.e., minimized matrix effect) in the LC-MS measurement, more metabolites were detected in the LC-MS (1612) than SCMS measurements (340). Thus, traditional LC-MS measurements can provide complementary information to our novel SCMS studies.

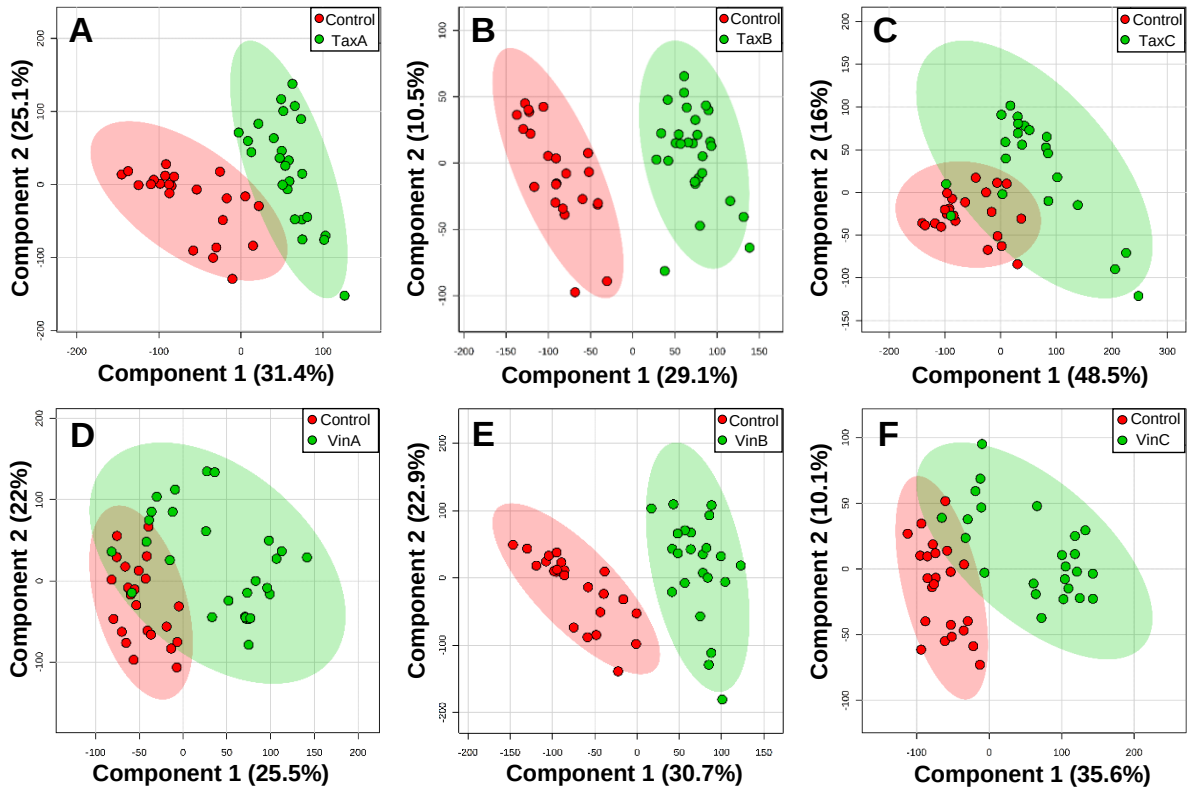

Figure 3. PLS-DA score plots in 2D space for phenotypic discrimination between the control and a drug treatment group, including (A) TaxA, (B) TaxB, (C) TaxC, (D) VinA, (E) VinB, and (F) VinC. Each data point represents the metabolomic profile of an individual cell, and the ellipse highlights the 95% confidence region.

Despite significant phenotypic separation demonstrated by permutation tests for each PLD-DA analysis, certain types of uncertainties are regarded as "noise",[74] including cell heterogeneity and technical variation (e.g., the sampling process, ionization stability, and instrument condition in SCMS experiments), and they may interfere with phenotypic separation in the SCMS data. To evaluate the influence of such "noise" on our data analysis, we employed a well-established machine learning algorithm in metabolomics studies, random forest,[75] to perform phenotypic classification. Each pair of pre-treated SCMS datasets were subjected to the classification, and

results are summarized in the confusion matrices and pie plots (Figure 3.4). Among them, cells in the control and a "shorter" time treatment condition yielded a low misclassification rate (6–11%, Figures 4A, 4C–4D, and 4F), implying a minor interference of such "noise" on the phenotypic separation. More interestingly, the misclassification rate is even lower (0 and 2%, Figures 4B and 4E) in the pair of datasets from the control and a "longer" treatment time condition, agreeing with the complete separation observed in the PLS-DA score plots (Figure 3.3). To our best knowledge, this is the first report of employing random forest as an alternative approach to evaluate the influence of the "noise"[74] on phenotypic separation in single cell metabolomics studies.

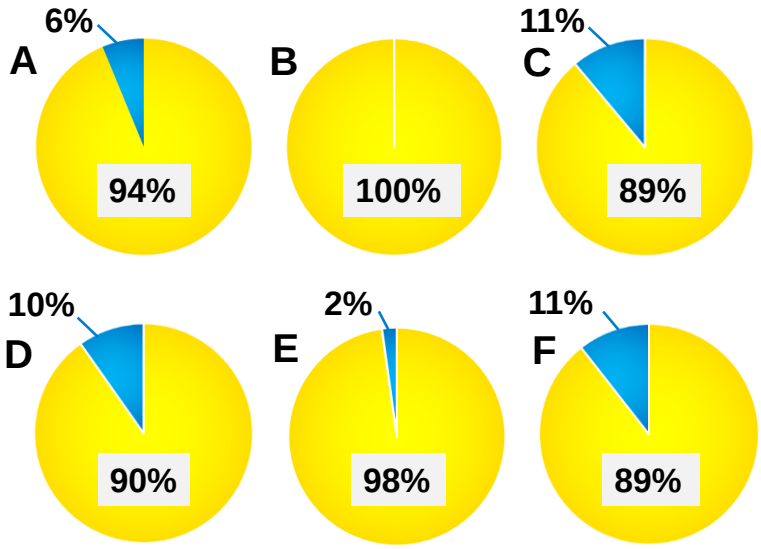

Figure 4. Cellular species correctly classified (yellow) and misclassified (blue) by random forest classification between the control and a drug treatment group, including (A) TaxA, (B) TaxB, (C) TaxC, (D) VinA, (E) VinB, and (F) VinC.

3.8. Study of phenotypic biomarkers

As previously reported, the magnitude of the abundance change of cellular metabolites represents the degree of difference between phenotypes.[76] Therefore, metabolites with significant abundances change after drug treatment are suitable candidates for phenotypic biomarkers, and they may arise biological interest and further suggest related cellular xenobiotic activities. Through our biomarker selection criteria as described above, we discovered a variety of phenotypic biomarkers corresponding to multiple treatment conditions, followed by tentative assignment of those biomarkers based on accurate mass. To further confirm the chemical identities of tentatively labeled biomarkers, we performed MS/MS analysis at the single cell level for those with relatively higher ion abundances. Six phenotypic biomarkers, i.e., $[PC(16:0) + Na]^+$ (*m/z* 518.3194), $[PC(18:0) + Na]^+$ (*m/z* 546.3506), $[SM(34:1) + Na]^+$ (*m/z* 725.5539), $[PC(32:1) + Na]^+$ (*m/z* 754.5345), $[PC(34:1) + Na]^+$ (*m/z* 782.5660), and $[PC(36:2) + Na]^+$ (*m/z* 808.5813), were identified from single cells. As a complimentary approach to enhance the biomarker identification, LC-MS/MS was also utilized to analyze cell lysates. In addition to the above six identified biomarkers, three more identifications, i.e., $[PC(34:1) + H]^+$ (*m/z* 760.5860), $[PC(34:2) + Na]^+$ (*m/z* 780.5561), and $[PC(36:3) + Na]^+$ (*m/z* 806.5643), were obtained. The majority of the identified biomarkers are phospholipids, which are related to the regulation of cell signal transduction in response to external stimuli.[77] Heat maps were constructed to intuitively visualize the relative abundances of the discovered biomarkers of cells in each treatment group (Figures S9 and S10). The overall color clusters (i.e., red and blue) matched well with cell attributes, although slight color variations can be observed for each biomarker among multiple cells likely due to cell heterogeneity.

3.9. Potential biological pathways

As biomarkers are tightly related to biological pathways regulating cellular metabolism,[78] we used the biomarkers discovered from our SCMS studies to unveil potential biological pathways related to metabolomic response to the drug treatment. We found two pathways, biopterin metabolism (p-value = 0.025) and glycerophospholipid metabolism (p-value = 0.041), were significantly enriched by taxol treatment. Other two pathways, bile acid biosynthesis (p-value = 0.021) and *de novo* fatty acid biosynthesis (p-value = 0.043), were significantly enriched by vinblastine treatment. The altered biopterin metabolism may be attributed to enzymatic activities related to guanosine triphosphate (GTP) cyclohydrolase I, which regulates biopterin metabolism[79] and is sensitive to drug treatment.[80] Glycerophospholipid metabolism involves a variety of phospholipids, which are responsible for cellular signal transduction[77] sensitive to surrounding microenvironment.[81] Bile acids are cell signaling molecules that are closely related to the regulation of energy and metabolic homeostasis,[82] and our SCMS results achieved good agreement with previous publications reporting a suppressed bile acid metabolism following vinblastine treatment.[83] *De novo* fatty acid biosynthesis has been reported to be suppressed upon drug treatment.[84] Our results, from the perspective of single cell metabolomics, suggest that those biological pathways may be significantly influenced by the treatment of mitotic inhibitors. It is worth noting that MS experimental conditions, such as the solvent composition[85], ionization polarity[86], and instrument type and tuning, can affect the detection sensitivity of different classes of species. Because our SCMS measurements were conducted under the same condition (i.e., using acidified acetonitrile sampling solvent, positive ion mode, and only one model of mass spectrometer), metabolites detected in the current study are likely within a limited coverage range. Therefore, more comprehensive experimental conditions can be used in future studies for

broader coverage of metabolites, and ultimately enhance the statistical power in pathway enrichment analysis.

3.10. Conclusion

We performed live single cell metabolomics studies using the Single-probe SCMS experiments in combination with a generalized comprehensive data analysis procedure. Cellular response to two mitotic inhibitors, taxol and vinblastine, were investigated and compared under multiple treatment conditions. Through the visualization using PLS-DA and the following permutation tests, our SCMS metabolomics results showed a rapid emergence of new phenotypes upon drug treatment. Similar trends were observed from traditional LC-MS experiments utilizing lysates prepared from population cells treated under the same conditions. Phenotypic biomarkers corresponding to two or multiple treatment conditions were discovered through statistical tests, with some of those further identified at both single cell and population levels. Based on the information of discovered biomarkers, potential biological pathways related to drug treatment were unveiled using the pathway enrichment analysis. Our methodology holds a promising potential to be readily coupled to other SCMS datasets produced from broader types of MS based analytical approaches to implement metabolomics at the single cell level, and ultimately gain insights into biological principles that regulate cellular metabolism.

Reference

(1) McKnight, S. L. *Science* **2010**, *330*, 1338.
(2) Metallo, C. M.; Heiden, M. G. V. *Molecular cell* **2013**, *49*, 388-398.
(3) Stopka, S. A.; Mansour, T. R.; Shrestha, B.; Maréchal, É.; Falconet, D.; Vertes, A. *Analytica Chimica Acta* **2016**, *902*, 1-7.
(4) Knolhoff, A. M.; Nautiyal, K. M.; Nemes, P.; Kalachikov, S.; Morozova, I.; Silver, R.; Sweedler, J. V. *Anal Chem* **2013**, *85*, 3136-3143.
(5) Chen, W. W.; Freinkman, E.; Wang, T.; Birsoy, K.; Sabatini, D. M. *Cell* **2016**, *166*, 1324-1337.e1311.
(6) You, M.; Litke, J. L.; Jaffrey, S. R. *Proceedings of the National Academy of Sciences* **2015**, *112*, E2756-E2765.
(7) Newgard, C. B. *Cell Metabolism* **2017**, *25*, 43-56.
(8) Patti, G. J.; Yanes, O.; Siuzdak, G. *Nat Rev Mol Cell Biol* **2012**, *13*, 263-269.
(9) Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X. *Analyst* **2012**, *137*, 293-300.
(10) Nicholson, J. K.; Lindon, J. C. *Nature* **2008**, *455*, 1054-1056.
(11) Si, X.; Xiong, X.; Zhang, S.; Fang, X.; Zhang, X. *Anal Chem* **2017**, *89*, 2275-2281.
(12) Ghaste, M.; Mistrik, R.; Shulaev, V. *Int J Mol Sci* **2016**, *17*, 816.
(13) Madji Hounoum, B.; Blasco, H.; Nadal-Desbarats, L.; Diémé, B.; Montigny, F.; Andres, C. R.; Emond, P.; Mavel, S. *Anal Bioanal Chem* **2015**, *407*, 8861-8872.
(14) Cuykx, M.; Negreira, N.; Beirnaert, C.; Van den Eede, N.; Rodrigues, R.; Vanhaecke, T.; Laukens, K.; Covaci, A. *Journal of Chromatography A* **2017**, *1487*, 168-178.
(15) Zhang, W.; Hankemeier, T.; Ramautar, R. *Curr Opin Biotech* **2017**, *43*, 1-7.
(16) Mairinger, T.; Causon, T. J.; Hann, S. *Curr. Opin. Chem. Biol.* **2018**, *42*, 9-15.
(17) Halama, A. *Archives of Biochemistry and Biophysics* **2014**, *564*, 100-109.
(18) Walker, B. N.; Antonakos, C.; Retterer, S. T.; Vertes, A. *Angewandte Chemie International Edition* **2013**, *52*, 3650-3653.
(19) Waki, M.; Ide, Y.; Ishizaki, I.; Nagata, Y.; Masaki, N.; Sugiyama, E.; Kurabe, N.; Nicolaescu, D.; Yamazaki, F.; Hayasaka, T.; Ikegami, K.; Kondo, T.; Shibata, K.; Hiraide, T.; Taki, Y.; Ogura, H.; Shiiya, N.; Sanada, N.; Setou, M. *Biochimie* **2014**, *107*, 73-77.
(20) Xie, W. Y.; Gao, D.; Jin, F.; Jiang, Y. Y.; Liu, H. X. *Anal Chem* **2015**, *87*, 7052-7059.
(21) Schober, Y.; Guenther, S.; Spengler, B.; Römpp, A. *Anal Chem* **2012**, *84*, 6293-6297.
(22) Shrestha, B.; Vertes, A. *Anal Chem* **2009**, *81*, 8265-8271.
(23) Ibáñez, A. J.; Fagerer, S. R.; Schmidt, A. M.; Urban, P. L.; Jefimovs, K.; Geiger, P.; Dechant, R.; Heinemann, M.; Zenobi, R. *Proc Natl Acad Sci U S A* **2013**, *110*, 8790.
(24) Mizuno, H.; Tsuyama, N.; Harada, T.; Masujima, T. *Journal of Mass Spectrometry* **2008**, *43*, 1692-1700.
(25) Gong, X.; Zhao, Y.; Cai, S.; Fu, S.; Yang, C.; Zhang, S.; Zhang, X. *Anal Chem* **2014**, *86*, 3809-3816.
(26) Zhu, H. Y.; Zou, G. C.; Wang, N.; Zhuang, M. H.; Xiong, W.; Huang, G. M. *P Natl Acad Sci USA* **2017**, *114*, 2586-2591.
(27) Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R. M.; Burgett, A. W. G.; Yang, Z. B. *Anal Chem* **2014**, *86*, 9376-9380.
(28) Liu, R.; Pan, N.; Zhu, Y.; Yang, Z. *Anal Chem* **2018**, *90*, 11078-11085.
(29) Onjiko, R. M.; Portero, E. P.; Moody, S. A.; Nemes, P. *Anal Chem* **2017**, *89*, 7069-7076.

(30) Zhang, L. W.; Foreman, D. P.; Grant, P. A.; Shrestha, B.; Moody, S. A.; Villiers, F.; Kwake, J. M.; Vertes, A. *Analyst* **2014**, *139*, 5079-5085.
(31) Comi, T. J.; Makurath, M. A.; Philip, M. C.; Rubakhin, S. S.; Sweedler, J. V. *Anal Chem* **2017**, *89*, 7765-7772.
(32) Herrmann, A. J.; Techritz, S.; Jakubowski, N.; Haase, A.; Luch, A.; Panne, U.; Mueller, L. *Analyst* **2017**, *142*, 1703-1710.
(33) Fujii, T.; Matsuda, S.; Tejedor, M. L.; Esaki, T.; Sakane, I.; Mizuno, H.; Tsuyama, N.; Masujima, T. *Nature Protocols* **2015**, *10*, 1445.
(34) Zhang, L.; Vertes, A. *Angew. Chem. Int. Ed.* **2018**, *57*, 4466-4477.
(35) Stolee, J. A.; Shrestha, B.; Mengistu, G.; Vertes, A. *Angew. Chem. Int. Ed.* **2012**, *51*, 10386-10389.
(36) Zhang, L.; Sevinsky, C. J.; Davis, B. M.; Vertes, A. *Anal Chem* **2018**, *90*, 4626-4634.
(37) Onjiko, R. M.; Moody, S. A.; Nemes, P. *Proc Natl Acad Sci U S A* **2015**, *112*, 6545.
(38) Bergman, H.-M.; Lanekoff, I. *Analyst* **2017**, *142*, 3639-3647.
(39) Zenobi, R. *Science* **2013**, *342*, 1243259.
(40) Pan, N.; Rao, W.; Standke, S. J.; Yang, Z. B. *Anal Chem* **2016**, *88*, 6812-6819.
(41) Rao, W.; Pan, N.; Yang, Z. *Journal of Visualized Experiments : JoVE* **2016**, *112*, 53911.
(42) Sun, M.; Yang, Z. B.; Wawrik, B. *Front Plant Sci* **2018**, *9*, 571.
(43) Liu, R.; Zhang, G.; Yang, Z. *Chem. Commun.* **2019**, *55*, 616-619.
(44) Sun, M.; Yang, Z. *Anal Chem* **2018**, *91*, 2384-2391.
(45) Standke, S. J.; Colby, D. H.; Bensen, R. C.; Burgett, A. W. G.; Yang, Z. *Anal Chem* **2019**, *91*, 1738-1742.
(46) Kim, K. S.; Cho, C. H.; Park, E. K.; Jung, M.-H.; Yoon, K.-S.; Park, H.-K. *PLOS ONE* **2012**, *7*, e30066.
(47) Lee, J.-W.; Park, S.; Kim, S. Y.; Um, S. H.; Moon, E.-Y. *Phytomedicine* **2016**, *23*, 705-713.
(48) Takara, K.; Obata, Y.; Yoshikawa, E.; Kitada, N.; Sakaeda, T.; Ohnishi, N.; Yokoyama, T. *Cancer Chemotherapy and Pharmacology* **2006**, *58*, 785-793.
(49) Heath, J. R.; Ribas, A.; Mischel, P. S. *Nat Rev Drug Discov* **2016**, *15*, 204-216.
(50) Zhang, X.-C.; Wei, Z.-W.; Gong, X.-Y.; Si, X.-Y.; Zhao, Y.-Y.; Yang, C.-D.; Zhang, S.-C.; Zhang, X.-R. *Sci Rep-Uk* **2016**, *6*, 24730.
(51) Schuhmann, K.; Thomas, H.; Ackerman, J. M.; Nagornov, K. O.; Tsybin, Y. O.; Shevchenko, A. *Anal Chem* **2017**, *89*, 7046-7052.
(52) Romano, P.; Profumo, A.; Rocco, M.; Mangerini, R.; Ferri, F.; Facchiano, A. *BMC Bioinformatics* **2016**, *17*, 61.
(53) Xia, J.; Wishart, D. S. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc., 2016.
(54) Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S. *Nucleic Acids Research* **2015**, *43*, W251-W257.
(55) Smilde, A. K.; van der Werf, M. J.; Bijlsma, S.; van der Werff-van der Vat, B. J. C.; Jellema, R. H. *Anal Chem* **2005**, *77*, 6729-6736.
(56) Di Guida, R.; Engel, J.; Allwood, J. W.; Weber, R. J. M.; Jones, M. R.; Sommer, U.; Viant, M. R.; Dunn, W. B. *Metabolomics* **2016**, *12*, 93.
(57) Ong, T.-H.; Kissick, D. J.; Jansson, E. T.; Comi, T. J.; Romanova, E. V.; Rubakhin, S. S.; Sweedler, J. V. *Anal Chem* **2015**, *87*, 7036-7042.
(58) Morrison, S. J.; Shah, N. M.; Anderson, D. J. *Cell* **1997**, *88*, 287-298.

(59) Do, K. T.; Wahl, S.; Raffler, J.; Molnos, S.; Laimighofer, M.; Adamski, J.; Suhre, K.; Strauch, K.; Peters, A.; Gieger, C.; Langenberg, C.; Stewart, I. D.; Theis, F. J.; Grallert, H.; Kastenmüller, G.; Krumsiek, J. *Metabolomics* **2018**, *14*, 128.
(60) Bradley, W.; Robert, P. *Current Metabolomics* **2013**, *1*, 92-107.
(61) Chai, T.; Cui, F.; Yin, Z.; Yang, Y.; Qiu, J.; Wang, C. *Scientific Reports* **2016**, *6*, 33481.
(62) Triba, M. N.; Le Moyec, L.; Amathieu, R.; Goossens, C.; Bouchemal, N.; Nahon, P.; Rutledge, D. N.; Savarin, P. *Mol Biosyst* **2015**, *11*, 13-19.
(63) Barberini, L.; Noto, A.; Saba, L.; Palmas, F.; Fanos, V.; Dessì, A.; Zavattoni, M.; Fattuoni, C.; Mussap, M. *Data in Brief* **2016**, *9*, 220-230.
(64) Kim, J.; Hu, Z. P.; Cai, L.; Li, K. L.; Choi, E.; Faubert, B.; Bezwada, D.; Rodriguez-Canales, J.; Villalobos, P.; Lin, Y. F.; Ni, M.; Huffman, K. E.; Girard, L.; Byers, L. A.; Unsal-Kacmaz, K.; Pena, C. G.; Heymach, J. V.; Wauters, E.; Vansteenkiste, J.; Castrillon, D. H., et al. *Nature* **2017**, *546*, 168-+.
(65) Suarez-Garcica, S.; Caimari, A.; del Bas, J. M.; Suarez, M.; Arola, L. *Scientific Reports* **2017**, *7*.
(66) Wanichthanarak, K.; Fan, S.; Grapov, D.; Barupal, D. K.; Fiehn, O. *PLOS ONE* **2017**, *12*, e0171046.
(67) Guijas, C.; Montenegro-Burke, J. R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A. E.; Wolan, D. W.; Spilker, M. E.; Benton, H. P.; Siuzdak, G. *Anal Chem* **2018**, *90*, 3156-3164.
(68) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F., et al. *Nucleic Acids Research* **2013**, *41*, D801-D807.
(69) Li, S.; Park, Y.; Duraisingham, S.; Strobel, F. H.; Khan, N.; Soltow, Q. A.; Jones, D. P.; Pulendran, B. *PLOS Computational Biology* **2013**, *9*, e1003123.
(70) Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. *Journal of The American Society for Mass Spectrometry* **2016**, *27*, 1897-1905.
(71) Huang, Q.; Mao, S.; Khan, M.; Zhou, L.; Lin, J.-M. *Chem. Commun.* **2018**, *54*, 2595-2598.
(72) Zhang, Y.; Jin, L.; Xu, J.; Yu, Y.; Shen, L.; Gao, J.; Ye, A. *Analyst* **2018**, *143*, 164-174.
(73) Wang, T.-H.; Wang, H.-S.; Ichijo, H.; Giannakakou, P.; Foster, J. S.; Fojo, T.; Wimalasena, J. *J Biol Chem* **1998**, *273*, 4928-4936.
(74) Balázsi, G.; van Oudenaarden, A.; Collins, James J. *Cell* **2011**, *144*, 910-925.
(75) Touw, W. G.; Bayjanov, J. R.; Overmars, L.; Backus, L.; Boekhorst, J.; Wels, M.; van Hijum, S. A. F. T. *Briefings in Bioinformatics* **2013**, *14*, 315-326.
(76) Vinaixa, M.; Samino, S.; Saez, I.; Duran, J.; Guinovart, J. J.; Yanes, O. *Metabolites* **2012**, *2*.
(77) Simons, K.; Toomre, D. *Nature Reviews Molecular Cell Biology* **2000**, *1*, 31.
(78) Sun, H.; Wang, B.; Wang, J.; Liu, H.; Liu, J. *Journal of Animal Science and Biotechnology* **2016**, *7*, 49.
(79) Kapatos, G. *IUBMB Life* **2013**, *65*, 323-333.
(80) Pickert, G.; Lim, H.-Y.; Weigert, A.; Häussler, A.; Myrczek, T.; Waldner, M.; Labocha, S.; Ferreirós, N.; Geisslinger, G.; Lötsch, J.; Becker, C.; Brüne, B.; Tegeder, I. *International Journal of Cancer* **2013**, *132*, 591-604.
(81) Rowinsky, E. K. *Oncologist* **2003**, *8*, 5-17.
(82) Qi, Y. P.; Jiang, C. T.; Cheng, J.; Krausz, K. W.; Li, T. G.; Ferrell, J. M.; Gonzalez, F. J.; Chiang, J. Y. L. *Bba-Mol Cell Biol L* **2015**, *1851*, 19-29.

(83) Gregory, D. H.; Vlahcevic, Z. R.; Prugh, M. F.; Swfll, L. *Gastroenterology*, *74*, 93-100.
(84) Mashima, T.; Seimiya, H.; Tsuruo, T. *Br J Cancer* **2009**, *100*, 1369-1372.
(85) Onjiko, R. M.; Morris, S. E.; Moody, S. A.; Nemes, P. *Analyst* **2016**, *141*, 3648-3656.
(86) Portero, E. P.; Nemes, P. *Analyst* **2019**.

Chapter 4: Towards Rapid Prediction of Drug-resistant Cancer Cell Phenotypes: Single Cell Mass Spectrometry Combined with Machine Learning

4.1. Introduction

Drug resistance, a phenomenon that renders tumor evasion of anticancer agents, is regarded as the major reason for chemotherapeutic failures.[1] In other words, a small population of cells capable of surviving from chemo-treatment through complex drug-resistant mechanisms, become immune to the original therapy, and eventually induce cancer relapse.[2] In general, there are two major types of drug resistance: primary and acquired. Primary resistance reduces the efficacy of chemotherapies before drug exposure, whereas acquired drug resistance develops afterwards.[3] Unfortunately, drug resistance cannot be monitored or evaluated in advance using common molecular imaging techniques, such as positron emission tomography, until accomplishing one or two chemo-treatment cycles in modern clinical practice,[4] resulting in ineffective treatment accompanied by serious toxicity for the patients. In addition, different tumor cells within the same histological region may respond differently to chemo-treatment due to intratumor heterogeneity.[5] However, conventional studies of drug resistance based on cell populations lack the ability to uncover biological information masked by such tumor cell heterogeneity. Herein, it is imperative to study drug resistance through interrogation and evaluation of individual cells using single-cell based methodologies. Mass spectrometry (MS) is a fast developing technique with broad applications in fundamental science and biomedical studies.[6,7] Recent development in MS allows for analysis of single cells with limited amount of analytes available (as low as in pL range for mammalian cells)[8] due to its extraordinary sensitivity, high accuracy,

and high throughput. To date, reported single cell MS (SCMS) techniques include but are not limited to secondary ion MS (SIMS),[9] matrix-assisted laser desorption/ionization (MALDI) MS,[10] laser ablation electrospray ionization (LAESI) MS,[11] live-single cell video-MS,[12] induced nanoESI MS,[13] the Single-probe MS,[14] and the T-probe MS.[15] Among these techniques, the Single-probe MS method stands out as an ambient technique to analyze live single cells of interest in situ and in real time with high efficiency and reliability.[14,16,17]

On the other hand, cell adhesion-mediated drug resistance (CAM-DR) was reported for myelogenous leukemia cells upon adhering to extracellular matrix (ECM), which coexists with those leukemic cells in the bone marrow, through integrin-ECM interaction.[18] Interestingly, this cell-ECM interaction confers reduced cell apoptosis upon exposure to cytotoxic drugs, and was recognized as one important form of primary drug resistance.[19] Despite the achievements of illustrating related biological mechanisms,[20,21] limited effort was contributed to predict such drug-resistant phenotype prior to any chemo-treatment, exposing patients to the risk of ineffective chemotherapy and associated toxicity. Limited studies in this area are likely due to a variety of factors, including 1) the lack of rapid and sensitive single cell analytical approaches that can simultaneously unveil phenotypical discrimination and intratumor heterogeneity, 2) the shortage of methods for systematic metabolomic analysis of single cells to reveal cellular metabolomic profiles associated with different phenotypes, and 3) the absence of advanced data mining methods towards rapid and reliable prediction.

To address those issues, we used the Single-probe SCMS technique to conduct metabolomic analysis at single cell level (i.e., single cell metabolomics) of cultured chronic

myelogenous leukemia (CML) cells (K-562) and obtain metabolomic information that is sensitive to upstream gene expression, protein regulation, and change of surrounding microenvironment.[22] Data analysis was conducted using machine leaning (ML) algorithms to mine the complex metabolomic datasets and unveil hidden biological patterns by performing clustering, regression, and prediction.[23] To the best of our knowledge, it is the first time to combine SCMS experiments with ML models for single cell metabolomics studies. Our approach provides a potential solution towards rapid and reliable prediction of drug-resistant cancer cell phenotypes (e.g. CAM-DR) based on cellular metabolomic profiles.

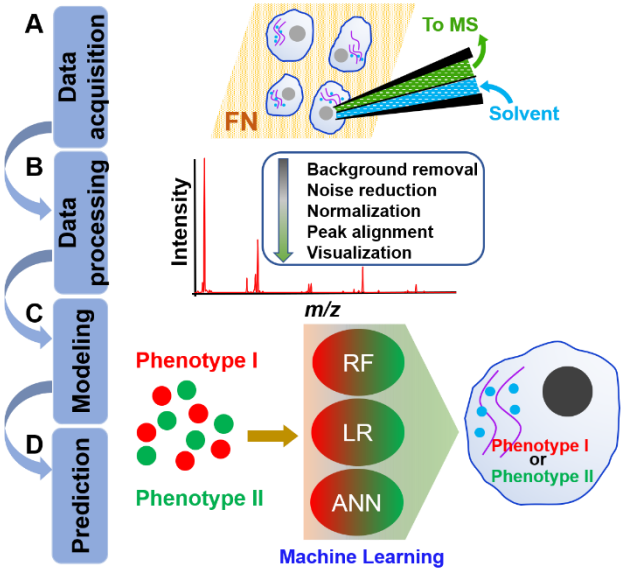

Figure 4.1. Workflow of the combined single cell mass spectrometry (SCMS) experiments and machine learning (ML) data analysis methods. (A) MS measurements of single cells using the Single-probe SCMS technique. (B) A comprehensive data processing approach to extract metabolomic information from raw SCMS datasets and visualize cellular profiles in low dimensional space. (C) ML models built on cells with two different phenotypes (with or without CAM-DR). (D) Rapid and reliable prediction of drug-resistant phenotypes at single cell level.

4.2. Methods

K-562 cell line was used as a model system to demonstrate our strategy as shown in Figure 4.1. As a well-established model, this cell line has been previously used to study the mechanism of CAM-DR in cancer cells.[18-20] We followed the published protocols to prepare two different phenotypes.[18] In brief, we first coated glass cover slips with fibronectin (FN), a major component of ECM,[24] and then allowed CML cells (K-562 cell line) to interact with FN in the cell culture plate. Cells that can adhere to FN (phenotype I) were reported to present CAM-DR compared with those suspended in the culture medium (phenotype II).[18] We prepared single cells of both phenotypes on the same type of glass cover slips. Using a hemocytometer, we estimated that 23.9% ± 5.3% of cells possessed CAM-DR in a typical experiment. We then utilized the Single-probe SCMS platform (Figure 4.2A) to interrogate individual cells and obtained their corresponding metabolomic profiles in real-time analysis (Figure 4.2B). We analyzed 100 and 108 single cells of phenotypes I and II, respectively. The raw MS data were subjected to pre-treatment, including background removal, noise reduction, peak normalization, and peak alignment. The endogenous cellular metabolites along with their relative ion intensities were subjected to downstream comprehensive analyses, including statistical analyses and ML predictions.

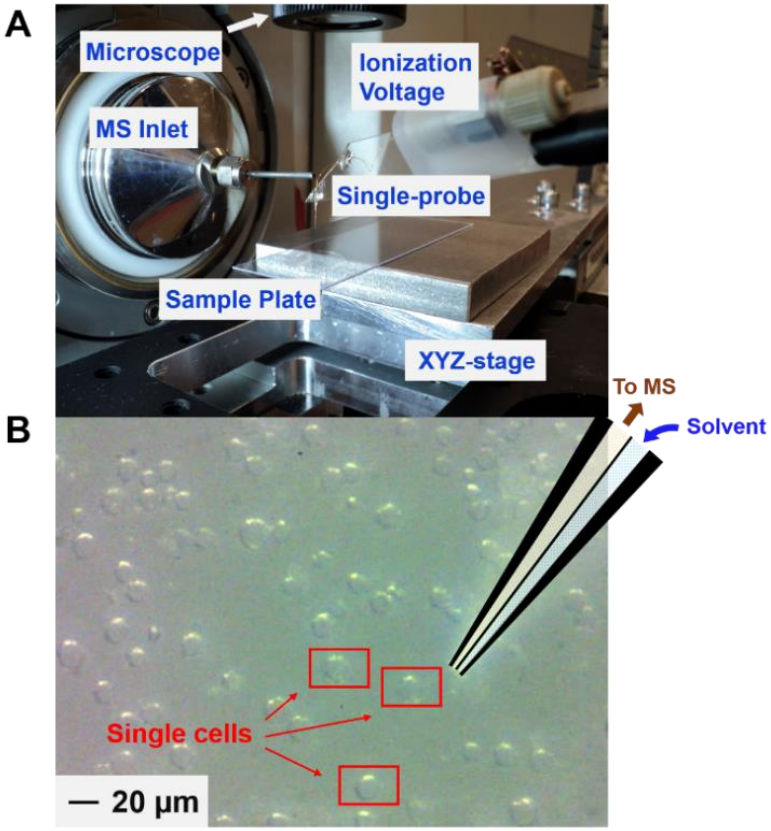

Figure 4.2. (A) Experimental setup of the SCMS platform, which is an integrated system including a Single-probe device, a Thermo Orbitrap XL mass spectrometer, two microscopes, and a motorized XYZ-stage system. (B) Individual leukemic cells located on the sample plate to be analyzed.

4.3. Results

To qualitatively evaluate and visualize the difference of metabolomic profiles between these two phenotypes, we analyzed the SCMS data using the *t*-distributed stochastic neighbor embedding (*t*-SNE), an algorithm for dimensionality reduction and visualization of data points in a non-linear fashion to achieve subtle group discrimination.[25,26] As shown on the *t*-SNE plot (Figure 4.3), an evident discrimination between these two phenotypes

can be intuitively observed, although some overlapped data points still exist likely due to cell heterogeneity. Our results suggest that the metabolomic profiles of two phenotypes are significantly different, which might be attributed to integrin-ECM interaction. With such evident discrimination, we further applied ML algorithms to establish models capable of predicting cellular phenotypes (i.e., CAM-DR or non-CAM-DR) based on the metabolomic profiles of cells.

In our study, we constructed ML models using random forest (RF), penalized logistic regression (LR), and artificial neural network (ANN) following SCMS data pre-treatment as described earlier. RF is an ensemble learning method based on multiple constructed decision trees and eventually outputs the averaged decision. Penalized LR builds nonlinear relationship between the response variable and independent variables through a logistic function, followed by minimizing the impact of less contributing variables. Both RF and penalized LR methods have been broadly applied to conventional metabolomic studies using liquid chromatography-mass spectrometry (LC-MS)[27,28] and single cell RNA-seq datasets.[29] ANN, as a fast-developing ML method, was inspired by the biological neural networks in animal brains. ANN optimizes parameters by learning from the prior knowledge, and the optimized model generates predictions through connected units and nodes. ANN has been previously applied to sorting single cells based on measured biomechanical properties,[30] and prediction of patient survival though genomics data.[31] Here, we further expanded the applications of those three methods to the analysis of single cell metabolomics datasets obtained from the Single-probe SCMS technique. Specifically, we applied RF, penalized LR (i.e., elastic net LR), and ANN to our pre-treated single cell metabolomics datasets, evaluated the predictive accuracy of each ML model, and recorded

the demanded computing time under each experimental condition. We performed model construction, evaluation, and k-fold validation for each ML model.

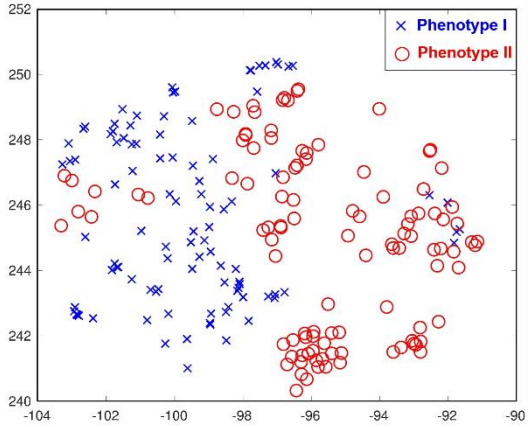

Figure 4.3. Visualization of cellular metabolomic profiles in two-dimensional space using t-distributed stochastic neighbor embedding (t-SNE). Phenotypic discrimination between two types of cells (phenotype I and II) is evident.

The pre-treated datasets were randomly shuffled with 80% cells being selected as the training set and the remaining 20% being selected as the testing set. The training set was used to construct and train ML models, whereas the testing set was used to evaluate the model performance. Due to tumor cell heterogeneity and experimental variation, single cell metabolomics datasets contain missing values (i.e., undetected cellular metabolites that were labelled as in 0 values in SCMS metabolomics datasets) in some SCMS measurements. Therefore, we evaluated the model performance according to different missing value threshold (MVT) as shown in Figure 4.4. For example, a dataset with 20% MVT contains variables (metabolites) that can be detected in at least 80% of all measured single cells. As the MVT increases, the number of variables increases accordingly (i.e., from 7 to 3232 as the MVT increases from 0% to 90%) in each ML model. A gradually

improved predictive accuracy was also observed in all three models (Figure 4.4A–C). Notably, a pronounced improvement was observed in predictive accuracy (from 77.1% ± 10.2% to the highest value of 94.8% ± 4.2%) in the RF model when the MVT was raised from 0% to 50%. However, further improvement of predictive accuracy was not observed with higher MVTs. Compared with the RF model, both penalized LR and ANN methods produced higher predictive accuracy when the MVT was below 40%, whereas comparable predictive accuracy was achieved as the MVT exceeded 40%. In addition, the highest predictive accuracy (i.e., 94.7% ± 1.8% and 96.2% ± 2.7%) can be obtained at 80% MVT for penalized LR and ANN models, respectively. Considering the trade-off between predictive accuracy and computing cost, which is a critical factor when handling larger sizes of data, we adopted ANN model with 40% MVT for rapid (~ 6 s) and reliable prediction (> 95% predictive accuracy) of drug-resistant phenotypes. We further demonstrated the predictive power of all ML models (with 40% MVT) in distinguishing two phenotypes using receiver operating characteristic (ROC) curve analysis[32] that examines the sensitivity and specificity of the model (Figure 4.4D–F). Consistently, the ANN model is superior in prediction with the area under the curve (AUC_{ANN}) = 0.9976 compared with the other two models (AUC_{RF} = 0.9542 and $AUC_{penalized\ LR}$ = 0.9884). To experimentally validate our method and evaluate the predictive accuracy of the ANN model, we conducted SCMS experiments and data pre-treatment for another batch of 31 single cells prepared on a different day, and utilized the trained ANN model to predict this new set of data. Our results show that the ANN model produced 87.1% ± 4.8% predictive accuracy, achieving comparable performance compared with our earlier results on the testing set.

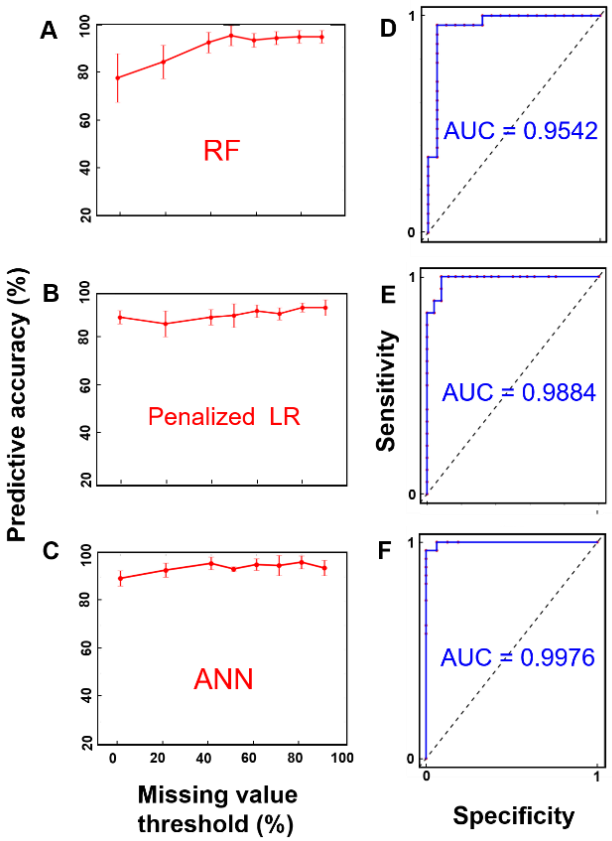

Figure 4.4. Evaluation of ML models. (A–C) Predictive accuracy of the random forest (RF), penalized logistic regression (LR), and artificial neural network (ANN) model were evaluated using different missing value thresholds (MVTs). (D–F) Evaluation of the predictive power of the corresponding RF, penalized LR, and ANN models using receiver operating characteristic (ROC) curve analysis. The area under the curve (AUC) were calculated for all three models.

4.4. Conclusion

In conclusion, we reported studies using the combined ambient SCMS technique (i.e., the Single-probe MS) and ML models to distinguish and predict drug-resistant phenotypes (e.g. CAM-DR) of live single cells through cellular metabolomic profiles for the first time.

Previous studies reported a number of prediction methods based on metabolic biomarkers (i.e., cellular species characteristic of specific disease, phenotype, etc.), including two-sample *t*-test,[33] analysis of variance (ANOVA),[28] loadings of principle component analysis (PCA),[34] and orthogonal partial least squares-discriminant analysis (OPLS-DA).[15] Compared with the above reported models, our method presents the following unique advantages: 1) SCMS based experiments allow for recognition of heterogeneous cells with different phenotypes. 2) Minimum sample preparation enables metabolomic signatures of live cells to be captured through online and in situ measurements. 3) Constructed ML models provide rapid results, which facilitates their potential translational applications towards future point-of-care (POC)[35] prognostic assays. 4) Because our methods utilized a variety of cellular metabolites other than metabolic biomarkers alone, the model predictive accuracy is significantly improved (p-value < 0.05, from Welch's one-tail t-test) compared with other models utilizing biomarkers discovered through two-sample *t*-test or PCA loading plot. As a complementary approach to identify discovered biomarkers at the population level, LC-MS/MS was performed and a total number of 28 biomarkers were identified. These results can potentially benefit future targeted studies, although not all discovered biomarkers can be identified possibly due to rapid metabolite turnover during cell lysis.[7] In addition, we validated our methods using cells prepared from different batches to obtain comparable results. Although the cultured CML cells were used as the model in the current study, our method can be potentially used towards future prediction and prognosis of patient derived samples. However, because the clinical samples are rather complex, additional procedures for sample preparation are necessary. For example, heterogenous cells obtained from bone marrow biopsy in clinic need to be firstly purified,

followed by enrichment of leukemic cells using standard protocols including centrifugation and flow cytometry analysis[36] prior to the SCMS experiments (~ 30 s/cell) and ML predictions of drug-resistant phenotypes.

Reference

(1) Gottesman, M. M. *Annu Rev Med* **2002**, *53*, 615-627.
(2) Garraway, L. A.; Janne, P. A. *Cancer Discov* **2012**, *2*, 214-226.
(3) Zahreddine, H.; Borden, K. L. B. *Front Pharmacol* **2013**, *4*.
(4) Lippert, T. H.; Ruoff, H. J.; Volm, M. *Int J Med Sci* **2011**, *8*, 245-253.
(5) Zhao, B.; Pritchard, J. R.; Lauffenburger, D. A.; Hemann, M. T. *Cancer Discov* **2014**, *4*, 166.
(6) Feng, X. J.; Liu, X.; Luo, Q. M.; Liu, B. F. *Mass Spectrom Rev* **2008**, *27*, 635-660.
(7) McLafferty, F. W. *Annual Review of Analytical Chemistry* **2011**, *4*, 1-22.
(8) Zhang, L.; Vertes, A. *Angew. Chem. Int. Ed.* **2017**, 4466-4477.
(9) Musat, N.; Foster, R.; Vagner, T.; Adam, B.; Kuypers, M. M. M. *Fems Microbiology Reviews* **2012**, *36*, 486-511.
(10) Ibáñez, A. J.; Fagerer, S. R.; Schmidt, A. M.; Urban, P. L.; Jefimovs, K.; Geiger, P.; Dechant, R.; Heinemann, M.; Zenobi, R. *Proc Natl Acad Sci U S A* **2013**, *110*, 8790.
(11) Shrestha, B.; Vertes, A. *Anal Chem* **2009**, *81*, 8265-8271.
(12) Masujima, T. *Analytical Sciences* **2009**, *25*, 953-960.
(13) Zhu, H.; Zou, G.; Wang, N.; Zhuang, M.; Xiong, W.; Huang, G. *Proc Natl Acad Sci U S A* **2017**, *114*, 2586.
(14) Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R. M.; Burgett, A. W. G.; Yang, Z. B. *Anal Chem* **2014**, *86*, 9376-9380.
(15) Liu, R.; Pan, N.; Zhu, Y.; Yang, Z. *Anal Chem* **2018**, *90*, 11078-11085.
(16) Pan, N.; Rao, W.; Standke, S. J.; Yang, Z. B. *Anal Chem* **2016**, *88*, 6812-6819.
(17) Rao, W.; Pan, N.; Yang, Z. *Journal of Visualized Experiments : JoVE* **2016**, *112*, 53911.
(18) Damiano, J. S.; Hazlehurst, L. A.; Dalton, W. S. *Leukemia* **2001**, *15*, 1232.
(19) Shain, K. H.; Dalton, W. S. *Mol Cancer Ther* **2001**, *1*, 69.
(20) Damiano, J. S.; Cress, A. E.; Hazlehurst, L. A.; Shtil, A. A.; Dalton, W. S. *Blood* **1999**, *93*, 1658-1667.
(21) Hazlehurst, L. A.; Dalton, W. S. *Cancer Metast Rev* **2001**, *20*, 43-50.
(22) Patti, G. J.; Yanes, O.; Siuzdak, G. *Nature Reviews Molecular Cell Biology* **2012**, *13*, 263.
(23) Grapov, D.; Fahrmann, J.; Wanichthanarak, K.; Khoomrung, S. *OMICS* **2018**.
(24) Harper, P. A.; Brown, P.; Juliano, R. L. *Journal of Cell Science* **1983**, *63*, 287.
(25) Do, T. D.; Comi, T. J.; Dunham, S. J. B.; Rubakhin, S. S.; Sweedler, J. V. *Anal Chem* **2017**, *89*, 3078-3086.
(26) Li, X.; Chen, W.; Chen, Y.; Zhang, X.; Gu, J.; Zhang, M. Q. *Nucleic Acids Research* **2017**, *45*, e166-e166.
(27) Xi, B.; Gu, H.; Baniasadi, H.; Raftery, D. *Methods in molecular biology (Clifton, N.J.)* **2014**, *1198*, 333-353.

(28) Grissa, D.; Pétéra, M.; Brandolini, M.; Napoli, A.; Comte, B.; Pujos-Guillot, E. *Frontiers in Molecular Biosciences* **2016**, *3*, 30.
(29) Pouyan, M. B.; Kostka, D. *Bioinformatics* **2018**, *34*, i79-i88.
(30) Darling, E. M.; Guilak, F. *Tissue Eng Pt A* **2008**, *14*, 1507-1515.
(31) Ching, T.; Zhu, X.; Garmire, L. X. *PLOS Computational Biology* **2018**, *14*, e1006076.
(32) Xia, J. G.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. *Metabolomics* **2013**, *9*, 280-299.
(33) Hinton, D. J.; Vázquez, M. S.; Geske, J. R.; Hitschfeld, M. J.; Ho, A. M. C.; Karpyak, V. M.; Biernacka, J. M.; Choi, D.-S. *Sci Rep-Uk* **2017**, *7*, 2496.
(34) Nemes, P.; Knolhoff, A. M.; Rubakhin, S. S.; Sweedler, J. V. *Anal Chem* **2011**, *83*, 6810-6817.
(35) Ferreira, C. R.; Yannell, K. E.; Jarmusch, A. K.; Pirro, V.; Ouyang, Z.; Cooks, R. G. *Clin Chem* **2016**, *62*, 99.
(36) Cloos, J.; Harris, J. R.; Janssen, J. J. W. M.; Kelder, A.; Huang, F.; Sijm, G.; Vonk, M.; Snel, A. N.; Scheick, J. R.; Scholten, W. J.; Carbaat-Ham, J.; Veldhuizen, D.; Hanekamp, D.; Oussoren-Brockhoff, Y. J. M.; Kaspers, G. J. L.; Schuurhuis, G. J.; Sasser, A. K.; Ossenkoppele, G. *JoVE* **2018**, e56386.

Chapter 5: Towards Early Monitoring of Chemotherapy-induced Drug Resistance Using Single Cell Metabolomics and Machine Learning

5.1. Introduction

Although broadly recognized as a treatment approach to cancer, chemotherapy suffers from poor outcomes in the clinic[1] due to the fact that the efficacy of a variety of chemotherapeutic agents is hindered by drug resistance through complex mechanisms (e.g., decreased drug uptake,[2] increased cellular detoxification,[3] oncogene mutation,[4] and other mechanisms[5,6]). The reduced chemotherapeutic efficacy renders treatment failure and relapse of malignancy. There are two general types of drug resistance: primary drug resistance, which denotes intrinsic resistance to chemotherapeutics prior to treatment due to genetic and epigenetic factors;[7] and acquired drug resistance, which refers to the one that adapts and evolves upon treatment pressure.[8] Recent evidence showed that the drug resistance observed following a chemotherapy may be due to a combination of both types.[5,9] Despite in-depth mechanistic studies conducted to fundamentally enhance our understanding of drug resistance, a variety of approaches have been reported to monitor chemotherapy-induced drug resistance. These methods include well-established procedures (e.g., clinically applicable imaging-assisted tissue biopsy[10,11]) and development-stage techniques (e.g., fluorescence-labeled[12] or nanoparticle-bound[13] drug monitoring, liquid biopsy,[14] resistance-related protein monitoring,[15] integral cell response monitoring,[16] and real-time monitoring using optofluidic chips[17]). However, three major limitations still exist. First, a vast majority of reported approaches evaluate drug resistance based on cell populations, whereas the molecular information of tumor cell heterogeneity, which plays a key role in cancer progression, is inevitably lost.[18] Second, depending on the type of drug resistance, most methods need a

lengthy monitoring period (e.g., one[19] to several months[20] after chemotherapy) for reliable results, exposing patients to ineffective chemo-treatment and accompanied toxicity.[21] Third, some methods require isolating tumor cells from their native biological microenvironment, leading to altered cellular metabolism and biophysical properties.[22] Thus, it is needed to develop new analytical approaches capable of monitoring early treatment-stage chemotherapy-induced drug resistance of single cancer cells in their native states.

To address these limitations, we have previously reported a method using single cell mass spectrometry (SCMS) experiments combined with machine learning (ML) data analysis to obtain metabolomic information of live individual cells, and predicted cells with primary drug resistance (i.e., cell adhesion-mediated drug resistance) before drug intervention.[23] In this current work, we further developed an analytical approach to monitor different degrees of chemotherapy-induced drug resistance of live cancer cells after drug exposure.

5.2. Experimental Section

5.2.1. Single-probe single cell metabolomics and cytotoxic assays. In our approach, we coupled the Single-probe,[24-29] a miniaturized and multifunctional device, to an in-house developed SCMS platform (Figure 5.1A) to perform metabolomic analysis at the single cell level (i.e., single cell metabolomics). Such online and in situ data acquisition allows for metabolomic signatures of individual cells to be captured in near native biological status with minimum metabolite turnover (Figure 5.1B).[30] Human colon cancer cell line, HCT-116, was selected as a model system to demonstrate our method. Particularly, we exposed HCT-116 cells to low-

concentration irinotecan (1 μM), a topoisomerase I (Topo I) inhibitor widely used in chemotherapies,[31] to mimic chemo-treatment, followed by sequentially monitoring the temporal change of cellular metabolomic profiles induced by drug exposure. Typically, SCMS experiments were performed using cells in three different groups: control (no drug exposure; n = 94), short-time drug exposure (1 μM for 10 days; n = 67), and long-time drug exposure (1 μM for 20 days; n = 70). Meanwhile, we measured the 50% inhibition concentration (IC_{50} at 72 h) of cells using cytotoxicity assay (i.e., MTT assay) to determine chemotherapy-induced drug resistance of cells at each time point (Figure 5.1C). We then referred to the resistance index (RI = $IC_{50\ \text{drug-resistant cell}}/IC_{50\ \text{parental cell}}$), a widely accepted standard to quantify drug resistance in clinical research,[32] as a reference of the degree of drug resistance at the population level.

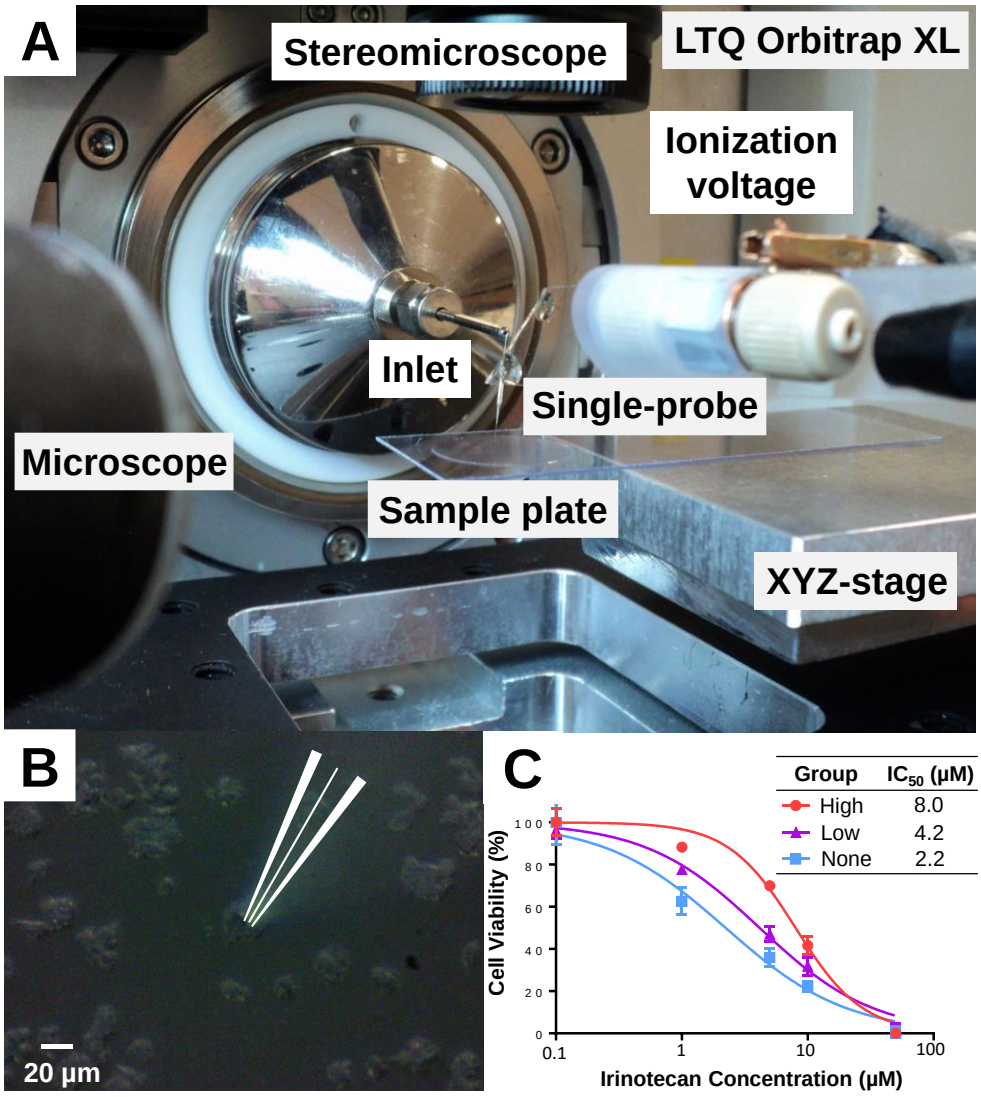

Figure 5.1. (A) SCMS data acquisition performed on an in-house developed SCMS platform with key components labeled. (B) Sampling a single cell using the Single-probe device. The contour of the Single-probe is drawn as visual guides. (C) Cytotoxicity measurements of HCT-116 cells exposed to irinotecan for 72 h.

5.2.2. SCMS metabolomics and statistical analysis. Obtained SCMS metabolomic datasets containing three groups of single cells (i.e., control, 10-day, and 20-day drug exposure) were

subjected to a comprehensive data pre-treatment procedure, including background subtraction, noise removal, intensity normalization, peak alignment, and selection of common species, to extract cellular metabolomic information from the raw data matrix. Pre-treated SCMS metabolomic datasets were subsequently subjected to dimensionality reduction using partial least square-discriminant analysis (PLS-DA),[33] and metabolomic profiles of individual cells were intuitively visualized in the 2D space. In addition, univariate analysis such as analysis of variance (ANOVA) was performed to compare the relative ion intensities of detected cellular species among three groups, and metabolic biomarkers that are significantly different in abundance among all examined groups were further discovered.

5.2.3. Machine learning. In addition to intuitive visualization of cellular metabolomic profiles using multivariate approach (i.e., PLS-DA), reliable mathematical models are imperative to monitor the chemotherapy-induced drug resistance by predicting the degree of possessed drug resistance (e.g., no, low, or high resistance) of single cells during chemo-treatment. Therefore, we applied ML methods to construct models that can learn the underlying patterns of the obtained SCMS datasets, and further predict drug-resistant cells based on their cellular metabolomic profiles. Three ML methods were utilized in our studies: random forest (RF) that outputs the most voted decision in the ensemble model,[34] artificial neural network (ANN) that optimizes model prediction by modifying the network configuration between nodes and layers,[35] and multinominal penalized (i.e., elastic net) logistic regression (LR) that predicts the categorical outcomes by maximizing the likelihood logistic function while minimizing less contributing variables[36]. Typically, 80% of single cells were randomly selected from the obtained SCMS datasets as the training set to construct a ML model, whereas the remaining cells were used as the testing set to evaluate the predictive accuracy. Each model was evaluated using five

independent predictions followed by a 5-fold cross-validation (CV) to avoid model bias.[37] The model construction, validation, and evaluation were performed using an in-house developed script.

5.3. Results and Discussion

5.3.1. Development of chemotherapy-induced drug resistance. Through MTT measured IC_{50} values of cells in three different groups (Figure 5.1C), we defined three levels of chemotherapy-induced drug resistance, i.e., none (RI = 1), low (RI = 1.9), and high (RI = 3.6), using cells in the control group (no drug exposure), 10-day, and 20-day drug exposure groups, respectively.[38] The increased drug resistance is likely associated with a variety of drug-resistant mechanisms,[31] and can catastrophically hinder the drug efficacy of irinotecan in chemotherapies. It is also worth noting that the MTT readouts can only be used as a reference, as the cell-to-cell heterogeneity was masked by such population-level measurements.

Table 1. Relative standard deviation (RSD) obtained from MS measurements of standard solution and single cells.

Origin	*m/z*	Species	RSD
Standard Solution*	760.5856	$[PC(34:1) + H]^+$	20.7%
	782.5676	$[PC(34:1) + Na]^+$	20.4%
SCMS Datasets**	756.5462	$[PC(32:0) + Na]^+$	68.9%
	760.5820	$[PC(34:1) + H]^+$	125.9%
	780.5457	$[PC(34:2) + Na]^+$	79.9%
	782.5616	$[PC(34:1) + Na]^+$	68.1%

808.5798	$[PC(36:2) + Na]^+$	75.2%
810.5947	$[PC(36:1) + Na]^+$	65.8%

*RSD was caculated based on 10 independent measurements using two different Single-probes.

**RSD was caculated based on measurements of 231 single cells possessing no, low and high ADR in the SCMS datasets.

5.3.2. Temporal Change of Cellular Metabolomic Profiles. To study the change of cellular metabolomic profiles induced by drug exposure, we performed PLS-DA[33] for the SCMS datasets, and evaluated the potential issues of model overfitting using a 10-fold CV[39] (i.e., through 10 iterative model construction and predictions). As shown in Figure 5.2, our PLS-DA model generated an excellent cross-validated predictive ability ($Q^2 > 0.9$), indicating no overfitting ($Q^2 > 0.5$).[40] In addition, 46% of total data variance can be explained by the first two dominating principal components. Despite cell heterogeneity reflected by the spatial distribution of data points,[41,42] the discrimination between the "none" and "high" groups is evident, representing a significant influence of long-time drug exposure on the cellular metabolism.[43] However, the "low" group partially overlaps with the other two groups. Such overlapped distribution is likely attributed to heterogeneous metabolic response of individual cells to the drug exposure at the early treatment stage. We further evaluated the influence of technical variance (i.e., fluctuation of ion intensities due to technical factors of SCMS experiments) and biological variance (i.e., variation in cellular metabolomic profiles due to cell heterogeneity and drug exposure) on the established PLS-DA model. First, we used two different Single-probes to sequentially measure a standard lipid solution (1 μM of PC(34:1)), and calculated the relative standard deviation (RSD) two adducts, $[PC(34:1) + H]^+$ (*m/z* 760.5856) and $[PC(34:1) + Na]^+$

(*m/z* 782.5676). The results represent the magnitude of technical variance[44,45] during the data acquisition using different probes. Second, we calculated the RSD of a variety of identified cellular species, including $[PC(34:1) + H]^+$ and $[PC(34:1) + Na]^+$, from the SCMS datasets, representing the combination of technical and biological variance.[44,45] Last, we compared the technical and biological variance as shown in Table 1. Evidently, the RSDs obtained from the standard solution (~20%) are significantly lower than those from the single cells (> 65%, i.e., > 3 fold of increment). Assuming that the technical variance is similar among all SCMS experiments, the biological variance among individual cells is significant, eliminating the possibility that the observed group discrimination in the PLS-DA score plot is merely due to technical variance. However, the relation between two major factors (i.e., cell heterogeneity and drug exposure) that could induce biological variance is complex.[42,46,47] Future experiments are needed to illustrate the underlying biological principals, however, these goals are beyond the scope of this study.

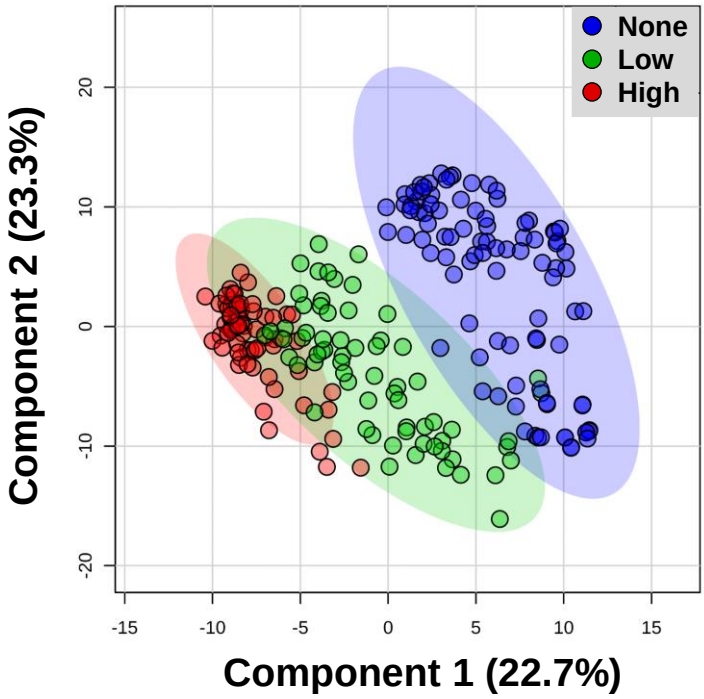

Figure 5.2. PLS-DA score plot of the SCMS datasets. Cellular metabolomic profiles corresponding to three groups of single cells possessing no ("None", colored in blue), low ("Low", colored in green) and high ("High", colored in red) drug resistance are intuitively visualized in 2D space.

5.3.3. Monitoring chemotherapy-induced drug resistance using ML models. RF, ANN, and penalized LR have been applied to conventional untargeted metabolomic research, such as identification of detected metabolites through liquid chromatography-mass spectrometry (LC-MS),[48] prediction of chromatographic retention time,[49] and evaluation of metabolic changes.[50] These approaches were employed in our study to monitor chemotherapy-induced drug resistance by predicting the degrees of drug resistance (i.e., "none", "low", or "high") of single cells based on established models. The predictive accuracy of each model was reported (Table 2) using the

confusion matrix. Our models generated excellent predictive accuracy (97.4% ± 1.8%, 97.4% ± 2.8% and 97.4% ± 2.3% for the RF, ANN and penalized LR model, respectively) compared with previously reported models in LC-MS metabolomic studies, in which relatively smaller numbers of measurements (e.g., ~100)[51,52] were performed. Less-than-perfect (100%) predictive accuracy was present in our models possibly due to heterogenous cellular response to the drug pressure. Our results demonstrated the predictive power of ML models, particularly for analysis of subtle differences in cellular metabolomic profiles among multiple groups (e.g., as shown in Figure 5.2). Notably, the predictive accuracies of those three models were comparable ($p > 0.99$ from Welch's two-tail *t*-test) on the testing set, indicating their capability and reliability to monitor early-stage chemotherapy-induced resistance at the single cell level.

5.3.4. Model Comparison. Metabolic biomarkers are of the great interest to conventional LC-MS metabolomic studies, and they were frequently discovered and used to monitor drug resistance.[53,54] Therefore, we carried out a systematic comparison of the performance (i.e., predictive accuracy) between models based on SCMS datasets and those utilizing metabolic biomarkers. A variety of criteria (e.g., statistical tests,[55] loading plots of multivariate analysis,[56,57] or variable importance[58]) have been generally used for biomarker discovery in other metabolomic research. In our studies, biomarkers selection was carried out using ANOVA, loadings of principal component analysis (PCA), and variable importance (VI) measured by mean decrease accuracy (MDA),[59] a value representing the contribution of a variable to the group separation. Different numbers of biomarkers were discovered using the above three methods: 24 from ANOVA (metabolites with ANOVA *p*-value < 0.05 and *post-hoc p*-value < 0.05 between each compared groups); 22 from PCA loadings (metabolites with highest PC1 and

PC2 loading scores); and 15 from VI (top-15 metabolites with the highest MDA obtained from the RF model constructed on SCMS datasets).[58] The number of variables in each dataset is different, with some metabolites mutually or exclusively discovered as biomarkers by the selection criterion (Figure 5.3). For example, seven metabolites were mutually regarded as biomarkers by all selection criteria, 20 metabolites were exclusively discovered under a specific selection criterion, and 41 metabolites in the SCMS datasets were not selected as biomarkers by any of the criterion.

Figure 5.3. Venn diagram of the number of variables (metabolites) in each dataset, including the SCMS dataset (brown) and other subsets consisting of dicovered metabolic biomarkers through ANOVA (red), PCA loadings (green) and variable importance (VI, blue).

We subsequently used these biomarkers discovered from different approaches and all metabolites included in the SCMS dataset to construct ML models using the RF, ANN and

penalized LR algorithms, and then evaluated the performance of these models (Table 2). Generally, all trained ML models exhibited excellent predictive accuracy on the testing set, despite that different groups of metabolites were utilized for model construction. Therefore, they all demonstrated the potential of our method towards prediction of drug-resistant cells in a rapid (< 30 s computing time in model construction) and reliable (> 94.9% predictive accuracy) fashion. On the other hand, the computing time for ML models based on biomarkers is less than the model based on SCMS datasets, likely due to fewer number of variables included.

Table 2. Predictive accuracy of RF, ANN and penalized LR models based on SCMS datasets and biomarkers discovered using different criteria.

Datasets	**Predictive Accuracy (RF)** *	**Error Rate$_{RF}$** **	**Predictive Accuracy (ANN)** *	**Error Rate$_{ANN}$** **	**Predictive Accuracy (LR)** *	**Error Rate$_{LR}$** **
SCMS	97.4% ± 1.8%	2.2%	97.4% ± 2.8%	2.1%	97.4% ± 2.3%	3.5%
ANOVA	97.4% ± 1.8%	3.0%	97.4% ± 1.8%	4.3%	98.7% ± 1.2%	3.0%
PCA Loadings	97.9% ± 1.5%	4.3%	95.7% ± 2.6%	5.7%	94.9% ± 2.4%	5.2%
VI	96.2% ± 1.0%	3.5%	98.3% ± 1.8%	1.7%	98.7% ± 1.9%	1.7%

*Predictive accuracy of single cells possessing no, low and high ADR was calculated from five independent predictions (average ± standard deviation).

**Error rate was estimated using a 5-fold CV in each model.

5.3.5. Multi-class ROC Analysis. Despite the predictive accuracy, receiver operating characteristic (ROC) analysis was routinely conducted for classification models in LC-MS metabolomic studies,[36] and the area under curve (AUC) was utilized to evaluate the classification capability, which is a complimentary measurement of performance in a diagnostic model.[52,60] In our study, we conducted multi-class ROC analysis,[61] a generalized form of traditional binary ROC analysis, to evaluate the classification capability of different ML models (Figure 5.4). Specifically, two pairs of data, i.e., one group of cells (e.g., "none") and the other group consisting of all remaining cells (e.g., pooled "Low" and "High"), were used to calculate an AUC. This type of calculation was repeated on the other two groups of cells (e.g., "Low" and "High"), and the final results were reported as an averaged AUC (i.e., $\overline{\text{AUC}}$). As a result, all models showed outstanding[62] classification capability ($\overline{\text{AUC}} > 0.99$) in predicting the cell attributes among all groups, agreeing with our earlier findings of the excellent model predictive accuracy.

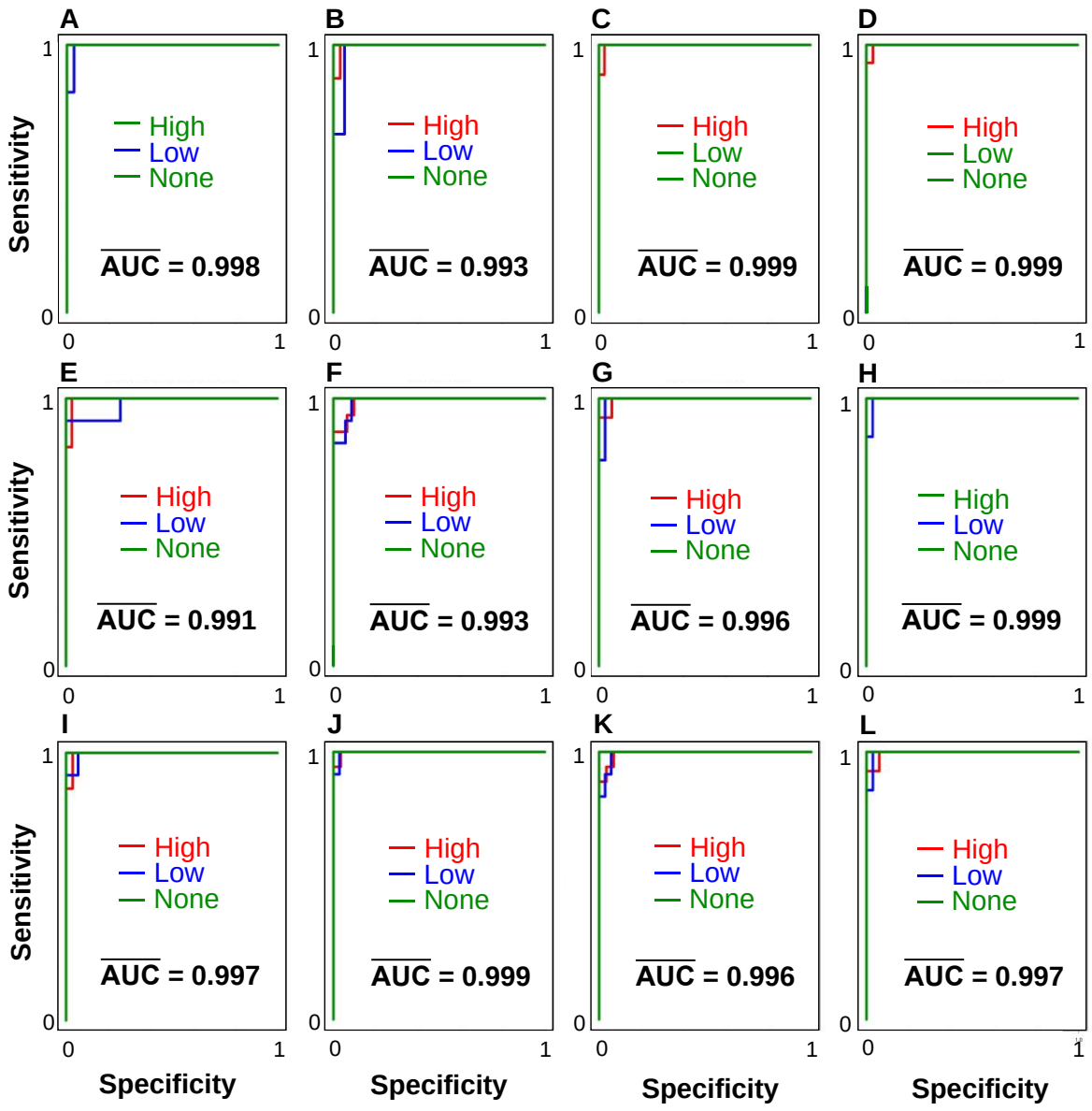

Figure 5.4. Multi-class ROC analysis of ML models including RF models constructed on the (A) SCMS datasets, (B) ANOVA subset, (C) PCA loadings subset, (D) variable importance (VI) subset; ANN models constructed on the (E) SCMS datasets, (F) ANOVA subset, (G) PCA loadings subset, (H) VI subset; and penalized LR models constructed on the (I) SCMS datasets, (J) ANOVA subset, (K) PCA loadings subset, (L) VI subset .The model classification ability is

represented by the averaged area under curve ($\overline{AUC}$) from three pairwise ROC analyses (e.g., "High" *vs.* pooled "Low" and "None", etc.) in each model.

5.3.6. Experimental Validation

Our ML models were trained and evaluated using a collection of SCMS data measured on different days (batches). To further evaluate the robustness of these established ML models, we performed an experimental validation using additional batches of single cells. Specifically, the established ML models were directly used to predict the attributes of 37 single cells prepared and measured on a different day (n_{None} = 12, n_{Low} = 14, n_{High} = 11). As a result (Table 3), accurate (86.5% ± 3.3%) and rapid (~0.5 s) prediction was achieved using the RF model constructed on the SCMS datasets. Therefore, our methods demonstrated their potential to be directly adopted by other labs or clinics to monitor the degree of chemotherapy-induced drug resistance with minimum effort. However, the predictive accuracy of the ANN and penalized LR models are generally lower than the RF model, likely due to the nature of these ML algorithms, and future studies are still needed. In addition, ML models based on biomarkers selected using different criteria showed lower predictive accuracy compared with the corresponding model based on SCMS datasets using each of the ML algorithm (i.e., RF, ANN, and penalized LR). Although the reason for such compromised predictive accuracy is complex, it is likely due to the loss of information during biomarker selection (i.e., exclusion of non-biomarkers in the dataset). Similar trends were observed in our previous studies of predicting the primary drug resistance of cells using ML models based on metabolic biomarkers alone.[23]

Table 3. Predictive accuracy of established RF, ANN and penalized LR models for additional batch of SCMS data.

Datasets	Predictive Accuracy (RF)*	Predictive Accuracy (ANN)*	Predictive Accuracy (LR)*
SCMS	86.5% ± 3.3%	75.1% ± 6.7%	65.9% ± 2.4%
ANOVA	68.1% ± 1.2%	64.3% ± 7.3%	64.9% ± 1.9%
PCA Loadings	68.7% ± 2.4%	68.7% ± 4.1%	63.2% ± 6.8%
VI	68.1% ± 1.2%	52.4% ± 3.1%	54.1% ± 3.3%

*Predictive accuracy of single cells possessing no, low and high ADR was calculated from five independent predictions (average ± standard deviation).

5.4. Conclusion

We reported an analytical approach combining single cell metabolomics with ML to monitor chemotherapy-induced drug resistance of cancer cells upon drug exposure. Three algorithms, including RF, ANN and penalized LR, were used to construct models based on datasets obtained through SCMS analysis of control and drug-exposed cancer cells. Following the model construction, we performed comprehensive evaluation such as 5-fold CV, predictive accuracy, and classification capability (i.e., ROC) to evaluate the performance of these ML models. Our results indicate that the RF model constructed using SCMS datasets could provide accurate (86.5% in predictive accuracy) and rapid (~0.5 s) predictions of individual cells possessing different levels of resistance (i.e., none, low, and high), and therefore, could be directly adopted by other labs for facile predictions. Although this current study was carried out using *in vitro* cancer cell lines, it can be potentially applied to early and real-time monitoring of chemotherapy-

induced drug resistance of patient samples, provided that efficient isolation of malignant cells from clinical specimen is performed[63] followed by SCMS measurements and ML predictions using an established model. Together with our previously reported method in predicting the primary drug resistance,[23] we demonstrated novel approaches based on single cell metabolomics that can be potentially used for future point-of-care (POC)[64] diagnostic assays in the clinic.

Reference

(1) Morgan, G.; Ward, R.; Barton, M. *Clinical Oncology* **2004**, *16*, 549-560.
(2) Pisco, A. O.; Jackson, D. A.; Huang, S. *Front Oncol* **2014**, *4*, 306-306.
(3) Syed, S. B.; Arya, H.; Fu, I. H.; Yeh, T.-K.; Periyasamy, L.; Hsieh, H.-P.; Coumar, M. S. *Sci Rep-Uk* **2017**, *7*, 7972.
(4) He, Y. J.; Meghani, K.; Caron, M.-C.; Yang, C.; Ronato, D. A.; Bian, J.; Sharma, A.; Moore, J.; Niraj, J.; Detappe, A.; Doench, J. G.; Legube, G.; Root, D. E.; D'Andrea, A. D.; Drané, P.; De, S.; Konstantinopoulos, P. A.; Masson, J.-Y.; Chowdhury, D. *Nature* **2018**, *563*, 522-526.
(5) Zahreddine, H.; Borden, K. *Front Pharmacol* **2013**, *4*, 28.
(6) Gottesman, M. M. *Annu Rev Med* **2002**, *53*, 615-627.
(7) Sharma, P.; Hu-Lieskovan, S.; Wargo, J. A.; Ribas, A. *Cell* **2017**, *168*, 707-723.
(8) Foo, J.; Michor, F. *Journal of theoretical biology* **2014**, *355*, 10-20.
(9) Yardley, D. A. *International Journal of Breast Cancer* **2013**, *2013*, 15.
(10) Ames, V.; Britton, P. D. *Insights into imaging* **2011**, *2*, 171-176.
(11) Fei, B.; Schuster, D. M. *AJR. American journal of roentgenology* **2017**, *209*, 255-269.
(12) Liang, X.-J.; Shen, D.-W.; Chen, K. G.; Wincovitch, S. M.; Garfield, S. H.; Gottesman, M. M. *J Cell Physiol* **2004**, *202*, 635-641.
(13) Lin, G.; Mi, P.; Chu, C.; Zhang, J.; Liu, G. *Adv Sci* **2016**, *3*, 1600134.
(14) Siravegna, G.; Mussolin, B.; Buscarino, M.; Corti, G.; Cassingena, A.; Crisafulli, G.; Ponzetti, A.; Cremolini, C.; Amatu, A.; Lauricella, C.; Lamba, S.; Hobor, S.; Avallone, A.; Valtorta, E.; Rospo, G.; Medico, E.; Motta, V.; Antoniotti, C.; Tatangelo, F.; Bellosillo, B., et al. *Nat Med* **2015**, *21*, 795.
(15) Paguio, M. F.; Cabrera, M.; Roepe, P. D. *Biochemistry-Us* **2009**, *48*, 9482-9491.
(16) Hammoud, M. K.; Yosef, H. K.; Lechtonen, T.; Aljakouch, K.; Schuler, M.; Alsaidi, W.; Daho, I.; Maghnouj, A.; Hahn, S.; El-Mashtoly, S. F.; Gerwert, K. *Sci Rep-Uk* **2018**, *8*, 15278.
(17) Liang, L.; Jin, Y. X.; Zhu, X. Q.; Zhou, F. L.; Yang, Y. *Lab on a Chip* **2018**, *18*, 1422-1429.
(18) Tellez-Gabriel, M.; Ory, B.; Lamoureux, F.; Heymann, M.-F.; Heymann, D. *Int J Mol Sci* **2016**, *17*, 2142.
(19) Lippert, T. H.; Ruoff, H. J.; Volm, M. *Int J Med Sci* **2011**, *8*, 245-253.
(20) Diaz Jr, L. A.; Williams, R. T.; Wu, J.; Kinde, I.; Hecht, J. R.; Berlin, J.; Allen, B.; Bozic, I.; Reiter, J. G.; Nowak, M. A.; Kinzler, K. W.; Oliner, K. S.; Vogelstein, B. *Nature* **2012**, *486*, 537.

(21) Fox, P.; Darley, A.; Furlong, E.; Miaskowski, C.; Patiraki, E.; Armes, J.; Ream, E.; Papadopoulou, C.; McCann, L.; Kearney, N.; Maguire, R. *European Journal of Oncology Nursing* **2017**, *26*, 63-82.
(22) Muir, A.; Danai, L. V.; Vander Heiden, M. G. *Disease Models & Mechanisms* **2018**, *11*, 035758.
(23) Liu, R.; Zhang, G.; Yang, Z. *Chem. Commun.* **2019**, *55*, 616-619.
(24) Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R. M.; Burgett, A. W. G.; Yang, Z. B. *Anal Chem* **2014**, *86*, 9376-9380.
(25) Pan, N.; Rao, W.; Standke, S. J.; Yang, Z. B. *Anal Chem* **2016**, *88*, 6812-6819.
(26) Rao, W.; Pan, N.; Yang, Z. *Journal of Visualized Experiments : JoVE* **2016**, *112*, 53911.
(27) Sun, M.; Yang, Z. B.; Wawrik, B. *Front Plant Sci* **2018**, *9*, 571.
(28) Sun, M.; Yang, Z. *Anal Chem* **2018**, *91*, 2384-2391.
(29) Standke, S. J.; Colby, D. H.; Bensen, R. C.; Burgett, A. W. G.; Yang, Z. *Anal Chem* **2019**, *91*, 1738-1742.
(30) Zenobi, R. *Science* **2013**, *342*, 1243259.
(31) Xu, Y.; Villalona-Calero, M. A. *Ann Oncol* **2002**, *13*, 1841-1851.
(32) Zhong, X.; Xiong, M.; Meng, X.; Gong, R. *J Exp Clin Canc Res* **2010**, *29*, 115.
(33) Worley, B.; Powers, R. *Current Metabolomics* **2013**, *1*, 92-107.
(34) Breiman, L. *Machine Learning* **2001**, *45*, 5-32.
(35) Zipser, D. *Neuroscience* **1992**, *47*, 853-862.
(36) Xi, B.; Gu, H.; Baniasadi, H.; Raftery, D. *Methods in molecular biology (Clifton, N.J.)* **2014**, *1198*, 333-353.
(37) Mitchell, M. W. *Open Journal of Statistics* **2011**, *1*, 205-211.
(38) McDermott, M.; Eustace, A.; Busschots, S.; Breen, L.; Clynes, M.; O'Donovan, N.; Stordal, B. *Front Oncol* **2014**, *4*, 40.
(39) Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A. *Metabolomics* **2008**, *4*, 81-89.
(40) Triba, M. N.; Le Moyec, L.; Amathieu, R.; Goossens, C.; Bouchemal, N.; Nahon, P.; Rutledge, D. N.; Savarin, P. *Mol Biosyst* **2015**, *11*, 13-19.
(41) Huang, Q.; Mao, S.; Khan, M.; Zhou, L.; Lin, J.-M. *Chem. Commun.* **2018**, *54*, 2595-2598.
(42) Zhang, Y.; Jin, L.; Xu, J.; Yu, Y.; Shen, L.; Gao, J.; Ye, A. *Analyst* **2018**, *143*, 164-174.
(43) Zheng, Y.; Zhou, J.; Tong, Y. *The Pharmacogenomics Journal* **2014**, *15*, 135.
(44) Zhang, L.; Sevinsky, C. J.; Davis, B. M.; Vertes, A. *Anal Chem* **2018**, *90*, 4626-4634.
(45) Stopka, S. A.; Khattar, R.; Agtuca, B. J.; Anderton, C. R.; Paša-Tolić, L.; Stacey, G.; Vertes, A. *Front Plant Sci* **2018**, *9*, 1646.
(46) Shah, A. T.; Diggins, K. E.; Walsh, A. J.; Irish, J. M.; Skala, M. C. *Neoplasia* **2015**, *17*, 862-870.
(47) Yip, C.; Davnall, F.; Kozarski, R.; Landau, D. B.; Cook, G. J. R.; Ross, P.; Mason, R.; Goh, V. *Diseases of the Esophagus* **2015**, *28*, 172-179.
(48) Samaraweera, M. A.; Hall, L. M.; Hill, D. W.; Grant, D. F. *Anal Chem* **2018**, *90*, 12752-12760.
(49) Miller, T. H.; Musenga, A.; Cowan, D. A.; Barron, L. P. *Anal Chem* **2013**, *85*, 10330-10337.
(50) Narath, S. H.; Mautner, S. I.; Svehlikova, E.; Schultes, B.; Pieber, T. R.; Sinner, F. M.; Gander, E.; Libiseller, G.; Schimek, M. G.; Sourij, H.; Magnes, C. *Plos One* **2016**, *11*, e0161425.
(51) Grissa, D.; Pétéra, M.; Brandolini, M.; Napoli, A.; Comte, B.; Pujos-Guillot, E. *Frontiers in Molecular Biosciences* **2016**, *3*, 30.

(52) Hinton, D. J.; Vázquez, M. S.; Geske, J. R.; Hitschfeld, M. J.; Ho, A. M. C.; Karpyak, V. M.; Biernacka, J. M.; Choi, D.-S. *Sci Rep-Uk* **2017**, *7*, 2496.
(53) Armitage, E. G.; Southam, A. D. *Metabolomics* **2016**, *12*, 146.
(54) Zong, L.; Pi, Z.; Liu, S.; Liu, Z.; Song, F. *Rsc Adv* **2018**, *8*, 15831-15841.
(55) Xiao, J. F.; Zhou, B.; Ressom, H. W. *Trends in analytical chemistry : TRAC* **2012**, *32*, 1-14.
(56) Nemes, P.; Knolhoff, A. M.; Rubakhin, S. S.; Sweedler, J. V. *Anal Chem* **2011**, *83*, 6810-6817.
(57) Liu, R.; Pan, N.; Zhu, Y.; Yang, Z. *Anal Chem* **2018**, *90*, 11078-11085.
(58) Viswan, A.; Singh, C.; Rai, R. K.; Azim, A.; Sinha, N.; Baronia, A. K. *Plos One* **2017**, *12*, e0187545.
(59) Hong, H.; Xiaoling, G.; Hua, Y. In *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2016, pp 219-224.
(60) Grund, B.; Sabin, C. *Current opinion in HIV and AIDS* **2010**, *5*, 473-479.
(61) Hand, D. J.; Till, R. J. *Machine Learning* **2001**, *45*, 171-186.
(62) Mandrekar, J. N. *Journal of Thoracic Oncology* **2010**, *5*, 1315-1316.
(63) Valente, M. J.; Henrique, R.; Costa, V. L.; Jerónimo, C.; Carvalho, F.; Bastos, M. L.; de Pinho, P. G.; Carvalho, M. *Plos One* **2011**, *6*, e19337.
(64) Ferreira, C. R.; Yannell, K. E.; Jarmusch, A. K.; Pirro, V.; Ouyang, Z.; Cooks, R. G. *Clin Chem* **2016**, *62*, 99.