UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

EXPLORING CELL HETEROGENEITY: APPLICATIONS OF SINGLE-PROBE SINGLE CELL MASS SPECTROMETRY FOR SUBPOPULATION ANALYSIS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DORTOR OF PHILOSOPHY

By

TRA D. NGUYEN
Norman, Oklahoma
2023

EXPLORING CELL HETEROGENEITY: APPLICATIONS OF SINGLE-PROBE SINGLE CELL MASS SPECTROMETRY FOR SUBPOPULATION ANALYSIS

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY THE COMMITTEE CONSISTING OF

Dr. Zhibo Yang, Chair

Dr. Rakhi Rajan

Dr. Luca Fornelli

Dr. Yuchen Qiu

Dr. Chongle Pan

© Copyright by TRA D. NGUYEN 2023

All Rights Reserved.

ACKNOWLEDGEMENTS

First and foremost, I express my deepest gratitude to my incredible family, whose support has been the backbone of my 5-year Ph.D. journey. To my mom, your constant listening and encouragement carried me through overwhelming stress, pushing me to persevere in my studies. Dad, your faith in my abilities meant the world to me; it provided the strength I needed to keep going. Your belief in me truly made all the difference. My dear sister, your optimism and willingness to assist with my experiments were invaluable. Your support and honesty, both in academics and in life, kept me grounded and motivated. To Yunpeng Lan and Michael Ho, your support sustained me through my entire academic journey. I owe my current position to the encouragement of my beloved family.

I would like to express my heartfelt appreciation to my Ph.D. advisor, Dr. Zhibo Yang, for his support and guidance throughout my journey. Getting a Ph.D. degree was challenging, but working under your mentorship transformed those challenges into valuable learning experiences. Your patience, understanding, and encouragement plays an important role in shaping my research and academic growth from day one. I am also deeply grateful to my committee members whose invaluable guidance and suggestions steered my research in the right direction. Your support significantly contributed to the successful completion of my study. I am honestly thankful for your time, dedication, and assistance during crucial milestones in my journey, such as my preliminary/general exam, annual evaluations, and now, my dissertation defense.

Furthermore, I would like to express my sincere appreciation to all collaborators, including Dr. Rakhi Rajan, Dr. Martin-Paul Agbaga, Dr. Laura-Isobel McCall, Dr. Y. Ann Chen, and Dr. Yihan Shao, along with my colleagues from my lab. Special gratitude goes to my lab partner and co-first author, Yunpeng Lan, for your help with most of my experiments and our collaborative efforts. I am also grateful to Zongkai Peng for aiding in mass spectrometry calibration and experiments, and Dan Chen for assisting with data analysis. I extend my thanks to Amit Singh, Shakya Munige, Deepti Bhusal and others for allowing me to access instruments and running my last few experiments. My heartfelt appreciation goes to colleagues from collaborating labs, Dr. Jiannong Li, Dr. Songyuan Yao, Dr. Neeraj K. Chauhan, and Eniola Adewunmi, for their collaborative projects, research insights, and motivating me to successfully complete my study.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
MS	Mass Spectrometry
SCMS	Single Cell Mass Spectrometry
LC	Liquid Chromatography
PC	Phosphatidylcholines
PBS	Phosphate buffer serum
FBS	Fetal bovine serum

Abstract

Mass spectrometry (MS) has become an indispensable tool for transformed metabolomics studies, whereas exploring single-cell metabolomic profiles remains a challenge due to limited techniques and suitable algorithms. This dissertation delves into cell heterogeneity, including method development and applications to infectious disease, with a focus on Chagas disease caused by Trypanosoma cruzi. Using the effective Single-probe SCMS technique combined with a fixation method that can safely decontaminate samples, we examined individual cell responses during parasite infection. Our findings unveiled significant differences in cell metabolism, even in neighboring uninfected cells, shedding light on the broader impact of infection. This pioneering study, utilizing bioanalytical SCMS, offers versatile tools to understand infectious diseases and the complexities of cell behavior in diseases.

Additionally, this dissertation tackles these challenges by merging the Single-probe single-cell MS (SCMS) technique with SinCHet-MS, a specialized bioinformatics software package. This combination allowed us to understand cell diversity, quantify cell subgroups, and identify key metabolites representing cells in subpopulations. Testing this approach with melanoma cancer cell lines revealed new subgroups after drug treatment, showcasing the potential for in-depth exploration of cell diversity and marker identification. This label-free method enhances our comprehension of cell metabolism in diseases and therapeutic responses.

Furthermore, this research pioneers a novel approach by integrating CRISPR-Cas9

gene editing with the Single-probe SCMS metabolomics, focusing on *FASN*-knockout cells in the human cell model HEK293T. This innovative strategy provides valuable insights into gene-metabolite interactions at the single-cell level. By combining advanced SCMS techniques with gene editing, this dissertation opens new avenues for understanding gene editing efficiency and the complex relationship between genes and cell metabolomics. These integrated methods advance our understanding of cell diversity in cancer, infectious diseases, and gene therapy, offering a fresh perspective for future research and therapeutic intervention.

Chapter 1 : Introduction

1.1 Cell heterogeneity

Due to genetic and phenotypic variances, nearly all biological systems are heterogenous.[1] Understanding cell heterogeneity is critical for studies of cell biology and human diseases, especially in cancers. Cell heterogeneity in infectious diseases opens new avenues for understanding host cell responses to parasitic infections. This diversity within cell populations arises from stochastic processes in transcription, translation, and metabolism, and comprehending this heterogeneity is fundamental for unraveling fundamental cell biology and human diseases. For instance, tumors exhibit heterogeneous distributions of malignant cells shaped by intrinsic and extrinsic factors, influencing disease progression, drug resistance, and tumor relapse. Understanding cell heterogeneity is essential for advancing our knowledge of disease evolution and management strategies.

In the context of infectious diseases such as Chagas disease caused by *T. cruzi* infection, metabolic deregulation in host cells is well-documented. *T. cruzi* infection results in a major deregulation of lipid and glucose metabolism in the host.[2, 3] Metabolic alterations proportional to CD severity were observed in the heart during experimental *T. cruzi* infection.[4, 5] Differential spatial distribution of metabolic alterations in experimentally-infected animals reflects sites of Chagas disease tropism.[6-8] However, all of these reported studies have been performed using traditional metabolomic, gene

expression or functional studies from extracts and lysates prepared from cell population, which masks cellular-level heterogeneity and cellular-level spatiality. Traditional studies, using metabolomic, gene expression, or functional analyses from cell population extracts, have highlighted metabolic alterations in response to infection. However, these approaches mask cellular-level heterogeneity and spatial distribution, missing crucial information about individual cell responses. Investigating infectious diseases from a single-cell perspective provides valuable insights into the diverse host cell reactions and the spatial distribution of metabolic changes during infection. This approach could revolutionize our understanding of infectious diseases, paving the way for targeted therapeutic interventions at the cellular level.

Cell heterogeneity plays a vital role in promoting cancer progression, metastasis, and development of drug resistance.[9, 10] During the process of metastasis, cancer cells become heterogeneous during their metastasis,[9] leading to formation of distinct subpopulations with diverse physiological properties. This heterogeneity within metastatic cells often serves as a major challenges in cancer therapy, contributing tremendously to treatment failure.[11] Detecting metastatic cancer cells, particularly at early stage, is crucial as it allows for timely intervention, providing appropriate treatment for halting disease progression.[12] However, the study of these rare metastatic cells within the heterogeneous cancer cell population has been challenging due to the lack of sensitive techniques.[9] Addressing this challenge is important for developing targeted therapies and personalized treatments, offering a promising approach to improve cancer outcomes by studying cell heterogeneity using single cell technologies.

1.2 Single cell mass spectrometry

Single cell technologies are capable of profiling individual cancer cells and uncovering rare and hidden subpopulations of cells. Mass spectrometry (MS) is a powerful analytical technique to perform single cell analysis. With the development of modern MS techniques, single cell MS (SCMS) methods have been established to sensitively detect and accurately identify metabolites in single cells. SCMS techniques are composed of two major types: vacuum-based and ambient-based techniques.

1.2.1 Vacuum-based techniques

Single Cell Mass Spectrometry (SCMS) techniques have undergone significant advancements, leading to the development of various methods employing distinct sampling and ionization techniques for investigating single cell metabolomics.[13] Among these, vacuum-based techniques, such as Matrix-Assisted Laser Adsorption/Ionization-Mass Spectrometry (MALDI-MS)[14, 15] and Secondary Ion Mass Spectrometry (SIMS)[16], have emerged as prominent technology. One of the key features of vacuum-based SCMS methods is their ability to operate in vacuum conditions, enabling rapid and sensitive analysis of samples at the single-cell level. These techniques offer distinct advantages, notably the relatively higher throughput they provide. For instance, Matrix-Enhanced (ME)-SIMS can analyze up to 2000 cells,[16] MALDI can process 1544 cells,[17] and Fourier-Transform Ion Cyclotron Resonance (FT-ICR) has the capability to analyze 717 cells.[18] This higher throughput makes vacuum-based SCMS techniques invaluable for studying diverse cellular

populations and unraveling the intricacies of single cell metabolomics.

However, vacuum-based SCMS methods also come with their set of challenges. The vacuum environment required for these techniques demands meticulous sample preparation and handling, which can be time-consuming and technically challenging. Moreover, the sensitivity and accuracy of these methods are highly dependent on the quality of sample preparation, making standardization and optimization crucial for obtaining reliable results. Despite these challenges, the advantages offered by vacuum-based SCMS, such as the ability to analyze a substantial number of cells, make them indispensable tools in the field of single cell metabolomics. Researchers continue to explore innovative approaches and improvements in vacuum-based SCMS techniques to overcome limitations and harness their full potential in deciphering the complex metabolic landscapes of individual cells.

1.2.2 Ambient techniques

Ambient-based techniques, such as live single-cell video-MS,[19] probe ESI-MS,[20] LAESI MS,[21] nano-DESI MS,[22] and the Single-probe,[23] have transformed mass spectrometry by enabling the study of live cells in their natural environments. Unlike traditional methods, these techniques allow real-time observation of cellular processes, capturing dynamic changes in metabolism and biomolecular interactions. This capability provides invaluable insights into cellular behavior, offering a deeper understanding of responses to stimuli, drug treatments, and environmental changes.

However, the application of ambient-based methods necessitates careful sample handling to minimize perturbations to live cell metabolism. Due to the delicate nature of live cell analysis, these techniques often yield relatively smaller numbers of cells per experiment, such as 140 cells from nanoPOTS,[24] 108 cells from Single-probe SCMS,[25] 32 cells from microprobe CE-ESI-MS,[26] and 15 cells from nano-DESI MS.[22] This limitation can be challenging when studying heterogeneous cell populations or rare cell types, impacting the scalability and applicability of these techniques in certain research contexts. Despite these challenges, the ability to study live cells in near-native environments offers unique advantages, providing critical insights into cellular heterogeneity, drug responses, and disease mechanisms. As technology advances, addressing these limitations will likely lead to further refinements in ambient-based techniques, enhancing their utility and expanding our understanding of live cell biology.

1.3 Single-probe single cell mass spectrometry

Our group has developed multiple SCMS techniques for analysis of live single cells in ambient environment (i.e., room temperature and atmospheric pressure). Among them, the Single-probe SCMS has been routinely used in our studies.[23,25, 27-30] In my studies, the Single-probe SCMS method has been used to profile the difference in molecular characteristics and to study biological pathways associated with both primary and metastatic cancer cells. Briefly, the Single-probe is a home-built device that can be coupled to a mass spectrometer for microscale sampling (e.g., from single cells and tissue slices) and ionization, followed by MS analysis. The Single-probe tip is small

enough for insertion into single cells to collect intracellular analytes, which are immediately ionized by MS analysis.[23] We have utilized this technique in different single cell studies such as investigating the differences in drug resistance,[31, 32] quantifying anticancer drugs in single cells,[33-36] comparing metabolites in cancer stem cells and non-stem cancer cells,[37] and determining the influence of environment on algal cell metabolites.[38] In addition, the Single-probe device has been utilized for MS imaging studies to acquire the spatial distribution of molecules on tissue slices[39-43] as well as to analyze secreted metabolites inside multicellular spheroids.[44]

1.4 Single cell metabolomics

Single cell metabolomics has emerged as a powerful approach in understanding the intricacies of cellular function and heterogeneity. Due to the rapid turnover rate of cell metabolites (i.e., products of cell metabolism), metabolomics studies of live single cells can reveal status and molecular features of rare cells that cannot be studied using traditional bulk analysis. Metabolites are smaller molecules (<1.5 kDa), including sugars, lipids, amino acids, etc.[18,19] Metabolites reflect cell status and unveil functions of associated metabolic pathways.

The significance of single cell metabolomics lies in its ability to uncover the phenotypic variations from one cell to another, pave the way for advance research in cell heterogeneity that exists within cell populations. By studying metabolites at the single cell level, scientists can discover unique behavior of rare cells and hidden subpopulations that are often not able to analyze using conventional bulk analysis.[20]

This precision is especially crucial when investigating heterogeneous cell populations or identifying specific cell types that play pivotal roles in disease progression or therapeutic responses.

In metabolomic studies, MS stands out as the cornerstone technology. Its exceptional sensitivity, broad molecular coverage, and robust structural identification capabilities make it an indispensable tool in the realm of single cell metabolomics. MS enables researchers to unravel the intricate metabolic landscapes of individual cells, providing a deeper understanding of cellular heterogeneity and paving the way for advancements in fields such as disease research, drug discovery, and personalized medicine.

1.5 Subpopulation analysis

Studying cellular subpopulations is pivotal in understanding the intricate variations that exist within heterogeneous cell populations, offering valuable insights into disease progression, drug responses, and therapeutic strategies. However, until recently, there has been a gap in research methodologies capable of utilizing overall metabolomic profiles of single cells to quantify changes in cell heterogeneity and identify associated subpopulations. The challenge stemmed from the lack of appropriate metrics to quantify cell heterogeneity using single-cell metabolomics profiling data and the absence of unbiased data analysis methods for identifying subpopulations without prior knowledge.

To address these limitations, a groundbreaking approach named SinCHet-MS was introduced. This innovative method combines experimental techniques of Single-cell Mass Spectrometry (SCMS) with a sophisticated bioinformatics tool, providing a solution to the complexities of studying cell heterogeneity and subpopulations. SinCHet-MS facilitates a systematic and quantitative analysis of subpopulations, leveraging all detected ions to discern metabolomic heterogeneity changes induced by drug treatments. The technique enables the identification and visualization of distinct subpopulations based on their metabolomic features, even revealing new subpopulations emerging in response to drug-sensitive primary melanoma cancer cells.

In essence, SinCHet-MS marks a significant milestone in the field of metabolomics, offering a label-free method to study subpopulations that is different from traditional targeted approaches. This method provides researchers with a powerful tool to explore single-cell metabolomic datasets from various instrument platforms. By gaining a profound understanding of cellular metabolism at both the cell heterogeneity and subpopulational resolution levels, SinCHet-MS opens doors to unprecedented insights into the complexities of cellular behavior and disease mechanisms, ultimately shaping the future of personalized medicine and drug development.

Chapter 2 : Single-Cell Mass Spectrometry Enables Insight into Heterogeneity in Infectious Disease

2.1 Introduction

Cell heterogeneity commonly presents in nearly all biological systems. In addition to the genetic variation, cellular heterogeneity can be induced by nongenetic mechanisms, i.e., cells possessing similar genotypes but actually expressing morphological and phenotypical differences.[45, 46] Although cell heterogeneity has been reported in human diseases, such as cancer, diabetes, and chronic and age-related diseases[47], it is largely understudied in infectious disease. For the first time, this study will pave the way to study the heterogeneity that presents in infection with *Trypanosoma cruzi* (*T. cruzi*) at the single-cell level.

T. cruzi is a protozoan parasite causing Chagas disease (CD), which is an understudied tropical disease with severe cardiac and gastrointestinal symptoms. At the cellular level, *T. cruzi* trypomastigotes invade host cells and differentiate into amastigotes, which can proliferate, differentiate back into trypomastigotes, and then escape the host cells. These newly produced trypomastigotes can then invade new cells and continue this cycle of damage.[48] *T. cruzi* infection results in a major deregulation of lipid and glucose metabolism in the host cells.[2, 49] Metabolic alterations proportional to CD severity were observed in the heart during experimental *T. cruzi* infection.[4, 5] Differential spatial distribution of metabolic alterations in experimentally-infected animals reflects sites of Chagas disease tropism.[6-8] However, all of these reported

studies have been performed using traditional metabolomic, gene expression, or functional studies from extracts and lysates prepared from cell populations or infected tissues, which masks cellular-level heterogeneity and cellular-level spatiality.

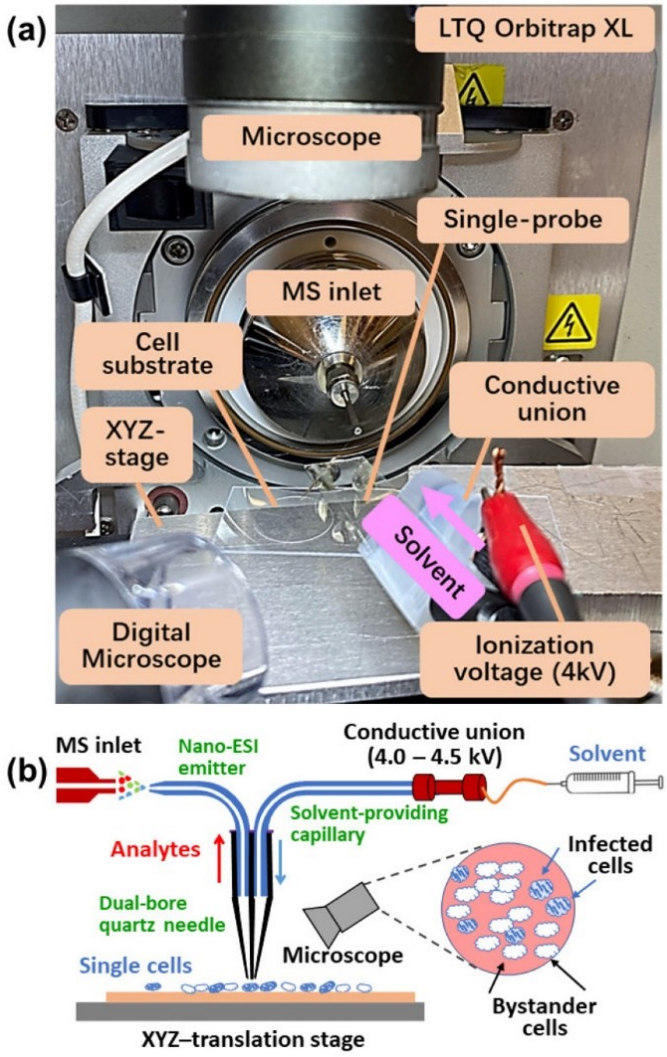


Figure 2-1 (a) Single-probe single-cell mass spectrometry (SCMS) setup. (b) Schematic of the working mechanisms of the experimental setup.

2.2 Methods

2.2.1 Sample preparation

2.2.1.1. Parasite culture

Beta-galactosidase-expressing *T. cruzi* strain Tulahuen (clone C4) were obtained through BEI Resources, NIAID, NIH[52] and maintained in mouse C2C12 myoblasts by once-weekly passaging. Trypomastigotes were collected from culture supernatant and used for infections.

2.2.1.2. Cell culture

HeLa cells were cultivated in DMEM cell culture medium (Corning) supplemented with 10% iron-supplemented calf serum (HyClone) and 1% penicillin-streptomycin (Gibco) in 5% CO_2 at 37 ºC. C2C12 cells were maintained in DMEM media supplemented with 5% iron-supplemented calf serum (HyClone) and 1% penicillin-streptomycin (Invitrogen), in 5% CO_2 and at 37 ºC, as previously described[30].

2.2.1.3. Cell infection and staining

HeLa cells were infected at a host:parasite ratio of 1:10. Two days post-infection, cells were washed with ice-cold PBS and fixed with 0.7% glutaraldehyde for 5 min, fixing and killing the parasites. Cells were then rinsed three times with PBS for 4 min. Cells were then stained overnight with 1 mg/mL of X-Gal in PBS containing 2 mM $MgCl_2$, 4.98 mM potassium ferricyanide and 5.76 mM potassium ferrocyanide,[53] pH 7.3 at 37°C.

2.2.2 Single-probe SCMS techniques

The single-probe SCMS setup includes a Single-probe, a digital microscope, a digital camera, a computer-controlled XYZ-translation stage system (CONEX-MFACC,

Newport Co., Irvine, CA, USA) and a Thermo LTQ Orbitrap XL mass spectrometer (Thermo Scientific, Waltham, MA, USA). The fabrication of the Single-probe and the SCMS set-up were detailed in details in our previous studies.[8,33,36–39,40] Briefly, the Single-probe was fabricated using a laser-pulled (P-2000 Micropipette Laser Puller, Sutter Instrument Co., Novato, CA) dual-bore quartz tubing (outer diameter (OD) 50 μm; inner diameter (ID) 127 μm, Friedrich & Dimmock, Inc., Millville, NJ, USA) embedded with a fused silica capillary (OD 105 μm; ID 40 μm, Polymicro Technologies, Phoenix, AZ, USA) in one channel and a nano-ESI emitter, which is produced from the same fused silica capillary, in another channel. The three parts were sealed using UV curing resin (Light Cure Bonding Adhesive, Prime-Dent, Chicago, Il, USA).

Glass coverslips containing cells were washed three times with fresh DMEM and placed on the XYZ-stage system of the Single-probe SCMS set-up for data acquisition. The targeted single cells were selected for analysis by precisely moving the stage system guided by the microscope. The sampling solvent (50% acetonitrile/50% methanol (v/v)) with 0.1% formic acid) was continuously delivered through the fused silica capillary to extract cellular contents followed by ionization via the nano-ESI emitter and real-time MS analysis. MS experiments were conducted under the following parameters: 200 nL/min flow rate; mass resolution, 60,000; +4.5 kV ionization voltage; 1 microscan; 100 ms max injection time. MS/MS experiments were conducted under the following parameters: 200 nL/min flow rate; mass resolution 60,000; +4.5 kV ionization voltage; 3 microscan; 500 ms max injection time. Collision energies were included in supporting information (Figure S4).

2.2.3. SCMS data analysis

SCMS data pretreatment was conducted following our established protocols [38,33]. MS data were exported with peaks (*m/z* values and relative intensities) generated by Thermo Xcalibur Qual Browser 3.0 (Thermo Scientific, Waltham, MA, USA). The exported raw data was subjected to background and noise subtraction in which all peaks with relative intensity < $3x10^3$ are removed. Background signals derived from organic solvent, cell culture medium and parasite medium were subtracted using an in-house R script as described in our prior works[35,54]. Normalization of ion intensities to the total ion chromatogram (TIC) was subsequently performed. The normalized data was uploaded to *Geena2* online software (http://proteomics.hsanmartino.it/geena2/geena2_ssi_norm.php)[55] for peak alignment with a mass tolerance of 10 ppm and the aligned *m/z* values are subjected for comparison. *Geena2* parameters were as follows: analysis range from 150 to 1500 *m*/z, maximum number of isotopic replicas: 3, maximum delta between isotopic peaks: 0.01 Da, maximum delta for aligning replicates: 0.01 Da and maximum delta for aligning average spectra: 0.01 Da. After performing peak alignment, missing values (50%) were removed using an in-house Python script (SI Supporting File 1).

Pretreated SCMS data were then imported to Metaboanalyst 5.0 [56,57,58,59,60] to perform principal component analysis (PCA), and hierarchical clustering. Random Forest analysis[40] was used to identify misclassified and correctly classified adjacent uninfected cells in comparison to infected cells using an in-house R script (with 500

trees and 7 predictors) (SI Supporting File 2). Then, the one-way analysis of variance (ANOVA)[61] was performed with an adjusted p-value cutoff of 0.05 using False Discovery Rate (FDR) correction. The ANOVA test revealed the most significant metabolites that were detected in three groups of cells (infected cells, correctly classified and mis-classified bystander cells). The hierarchical clustering heatmap[62] was generated using Ward's minimum variance clustering method and Euclidean distance method, from normalized data with autoscale features standardization. To minimize the technical variance [8,63], two replicates were performed for comparison under similar experimental conditions. Boxplots display median, upper and lower quartiles, with whiskers extending to largest and lowest datasets and outliers beyond the whiskers represented as dots. Annotations were generated as follows from the combined two replicates' ANOVA test results. 1) Via LC-MS/MS to obtain MS/MS spectra (see below for parameters). 2) Via SC-MS/MS (see above). . Annotations were generated from the resulting MS/MS spectra by spectral comparison to data deposited in METLIN (https://metlin.scripps.edu)[64], HMDB (http://www.hmdb.ca)[65] and GNPS (https://gnps.ucsd.edu/, see Table S3 for parameters)[66,67].

2.2.4. LC-MS/MS analysis

Metabolites were extracted from uninfected and infected HeLa cells using a two-step extraction with 50% methanol followed by 3:1 dichloromethane-methanol (all Fisher Optima LC-MS grade). Extracts were resuspended in 50% methanol, as in our prior work.[30] LC analysis was performed on a Thermo Vanquish LC equipped with a

1.7 µm Kinetex C18 50 x 2.1 mm column, 100 Å pore size, protected by a SecurityGuard ULTRA C18 Guard Cartridge (Phenomenex). Injection volume was 5 µL. Auto-injector was washed with 10% methanol at a rate of 10 µL/s for 2 seconds. LC gradient was composed of mobile phase A (water + 0.1% formic acid) and mobile phase B (acetonitrile + 0.1% formic acid) at a flow rate of 0.5 mL/min (Table S1). The autosampler was maintained at 10 ℃ and the column compartment at 40 ℃.

MS data were acquired on a Thermo Fisher Q-Exactive Plus hybrid quadrupole orbitrap mass spectrometer operating in positive parallel reaction monitoring (PRM mode, Table S2). Instrument calibration was performed using Thermo Fisher Calmix. All samples underwent a 12.5 minute runtime elution gradient as follows: start at 5% solvent B for one minute, gradual increase to 100% solvent B for eight minutes, hold at 100% solvent B for two minutes, drop to 5% solvent B for 30 seconds, and hold at 5% solvent B for one minute. Full PRM parameters were: scan range set to 100-1,500 *m/z*, default charge state was 1, resolution was 17,500, AGC target set to 2e5, maximum IT was 54 ms, isolation window set to 1 *m/z*, and normalized collision energy increased from 20-60%. MS source parameters were as previously described in [30].

Raw data files were converted to mzXML format using MSConvert[68].

2.2.5. Data availability

LC-MS data has been deposited in MassIVE, accession number MSV000087656. SCMS data has been deposited in MassIVE accession number MSV000089503.

2.3 Results and discussion

In the current study, we focused on metabolomics of single cells infected by *T. cruzi*, due to the crucial role of metabolism in CD.[8, 50] The experiments were conducted using the Single-probe SCMS technique to analyze HeLa cells, which were used as the model system in three different groups: T. cruzi-infected, bystander (i.e., uninfected cells that are adjacent to infected cells), and control cells (no parasite exposure). Our results revealed striking bystander effects of infection, including metabolic pathways commonly perturbed in infected cells and bystander cells. These results help improve our understanding of host pathways of CD pathogenesis and may help develop new treatments to address late-stage disease that cannot be cured by antiparasitic agents. Furthermore, our approach is compatible with biosafety protocols and thus should have broad applicability to other intracellular pathogenic agents.

During chronic T. cruzi infection, only a minority of cells are infected.[51, 52] Although parasite persistence is required for disease progression[53], CD symptoms can nevertheless develop even with low parasite load that may be spatially disconnected from sites of tissue damage.[54, 55] SCMS analyses of infected and uninfected cells in the same culture plate, in comparison to control wells, can deconvolute direct effects of T. cruzi infection from bystander effects of infection. HeLa cells were used as a model and infected with beta-galactosidase-expressing T. cruzi.[56] Cells were fixed by glutaraldehyde to kill the parasites and ensure biosafety. The fixed cells were stained by X-gal, enabling us to differentiate parasite-containing cells from bystander cells. The

invasion of amastigote stage of parasites can be observed in bright-field microscopic images (**Figure 2-2**a), and the infected cells containing amastigotes can be clearly distinguished from the bystander cells. These observations match with previous publications regarding this parasite strain.[57, 58] SCMS measurements were performed not only on these infected and bystander cells, but also on control cells from a separate, uninfected culture well.

PCA (principal component analysis) showed that the fixation and staining processes had no significant influence on the overall cellular metabolite profiles (**Figure 2-2b**, S4a). This conclusion was further confirmed ($p = 0.49$ from permutation test) by PLS-DA (partial least squares discriminant analysis) (Figure S4b and S4c). As expected, parasite-containing cells have different overall metabolite compositions compared with bystander cells. However, strikingly, both cell types differed in terms of overall metabolome from control and stained cells (both are uninfected). This finding supports bystander effects of T. cruzi infection on the overall cellular metabolome and provides a metabolic mechanism to explain the development of Chagas disease lesions at sites with low parasite burden.[59]

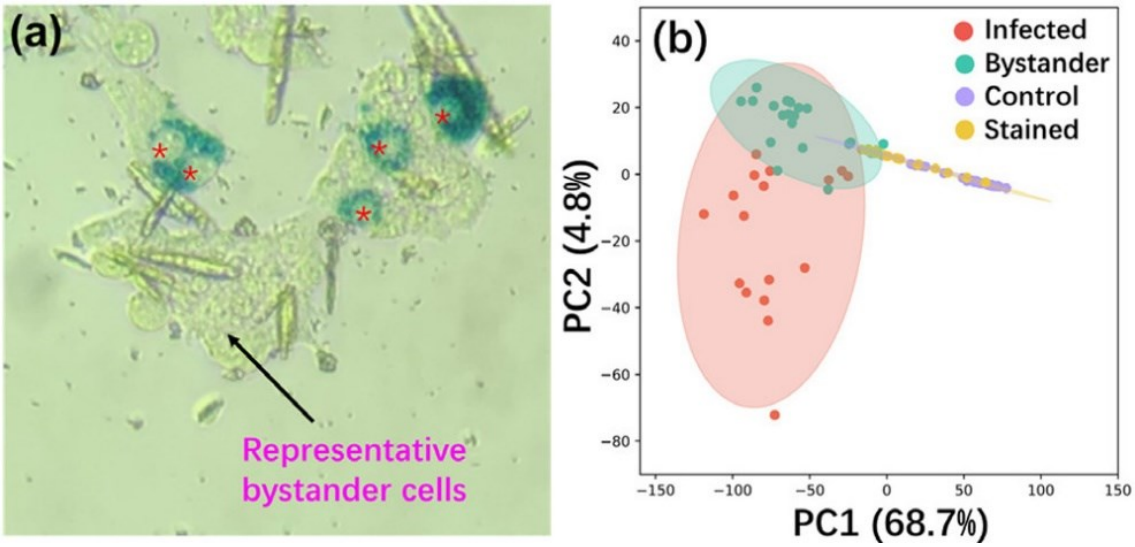


Figure 2-2 Influence of the optimized fixation and staining processes on the overall profiles of cellular metabolites in HeLa cells infected by T. cruzi. (a) HeLa cells were infected with beta-galactosidase-expressing *T. cruzi,* fixed by glutaraldehyde, and stained by X-gal. Infected cells (with stained parasites, circled) can be differentiated from unstained bystander cells (adjacent uninfected cells) by light microscopy. (b) PCA results. Without parasite infection, cells have comparable profiles of metabolites without (control) and with (stained) the fixation and staining processes. Cells exposed to parasites (infected and bystander cells) present significantly different metabolite profiles than unexposed cells (control and stained).

PCA results showed that a subset of bystander cells was particularly similar to (i.e., overlapped with) infected cells from the same culture plate (**Figure 2-2b**, **Figure 2-3a**). Indeed, random forest machine learning algorithms mis-classified 16 out of 53 bystander cells as infected (**Table 2-1**). In contrast, 62 out of 68 infected cells were correctly classified. It is worth noting that a large portion of control cells was misclassified as stained cells and vice versa, supporting that fixation and staining processes have no significant influence on cell metabolites. We then manually regrouped the bystander cells into correctly classified and mis-classified subgroups and conducted PCA. We observed a high degree of similarity between the mis-classified and infected cells (**Figure 2-3b**). Similar trends were observed from results from

hierarchical clustering of metabolites (Figure S1). To determine metabolites with significantly different abundances among the infected and two bystander groups (correctly classified and mis-classified infected cells), we performed ANOVA (with False Discovery Rate (FDR) correction and adjusted p-value ≤ 0.05) (Table S4). We obtained 16 ions from all groups possessing strikingly similar patterns for both mis-classified bystander cells and infected cells across two independent experimental replicates (e.g., lower levels of m/z 267.0620, 322.886, and 359.025 compared to correctly-classified bystander cells) (Table S4, Figure S2).

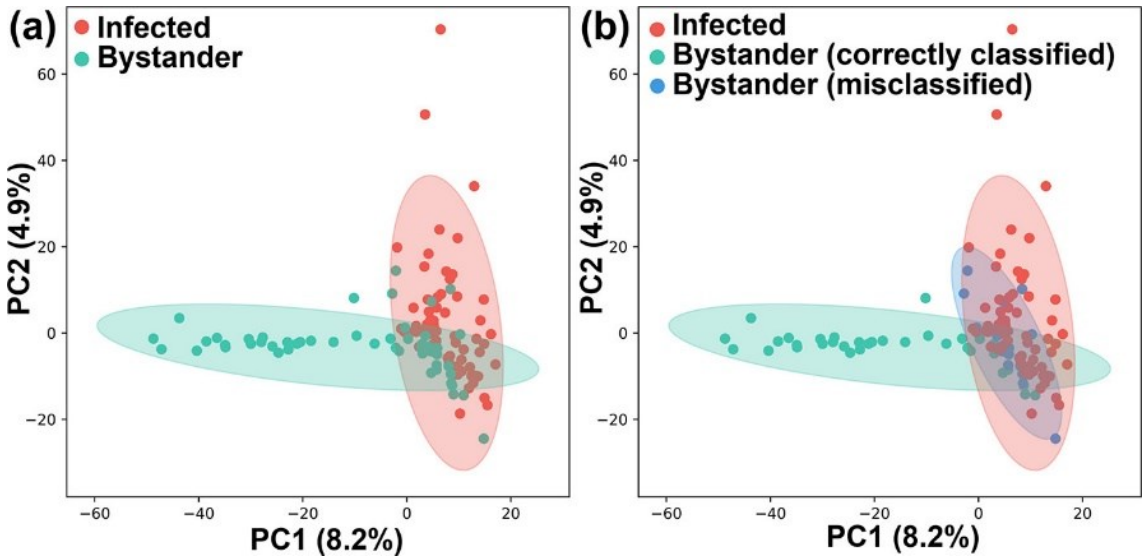


Figure 2-3 Impact of *T. cruzi* infection on the metabolome of bystander uninfected cells. (a) PCA of SCMS data highlighting metabolic overlap between T. cruzi infected cells and a subset of bystander cells. (b) PCA analysis of SCMS data as in (a), colored based on random forest classifier prediction. Mis-classified uninfected bystander cells have similar overall metabolomes to infected cells.

Table 2-1 Random Forest classification. 17 mis-classified bystander cells classified as infected cells and 36 correctly classified bystander cells, out of a total of 53 bystander cells.

Predicted	Control	Stained	Infected	Bystander	Classification

Correct					error
Control	48	12	1	0	0.213
Stained	25	13	0	0	0.658
Infected	0	1	62	5	0.014
Bystander	3	1	16	33	0.32

To annotate these ions, we performed MS/MS of both single cells (using the Single-probe SCMS method) and cell lysate (using LC-MS/MS). Similar to our previous studies[30, 60], some species could only be detected in the SCMS experiments, likely due to multiple reasons (e.g., differences in sample preparation methods, matrix compositions, and stabilities of molecules during sample preparation) (Table S4). As expected in untargeted metabolomics[61], most metabolite features could not be annotated (Table S4). Among all annotatable metabolites, m/z 756.547 was annotated as PC(34:3), LPC(34:4), or PC(O-34:4) (Table S4, **Figure 2-4**). This lipid significantly differed in abundance between cell groups ($p = 2.33x10-4$ using ANOVA test with False Discovery Rate correction (Figure 4a). It is interesting to note that, similar to infected cells, mis-classified bystander cells also contain high abundances of this species (Figure 4a). Other infection-elevated metabolites were also annotated as glycerophosphocholines (GPCs), including m/z 768.583, 780.5460, 782.5630, 808.5770 and 810.5940 (**Figure *2-1***Table S4, Figure S3). This observation concurs with our prior findings of infection-elevated GPCs in heart tissue in proportion to disease severity and in the infected

esophagus and large intestine, in mice across multiple infection timepoints and parasite strains.[4, 6, 7, 49] While confidently assigning a parasite vs host origin to these GPCs is challenging, very long-chain GPCs and lysoglycerophosphocholines (Lyso-GPCs) are elevated in isolated amastigote-stage T. cruzi compared to host cells.[62] These findings may support further re-development of therapeutics targeting phosphatidylcholine metabolism, such as miltefosine, currently in clinical use for the related parasite Leishmania, but in this case to target the metabolic consequences of infection on the host.[63, 64]

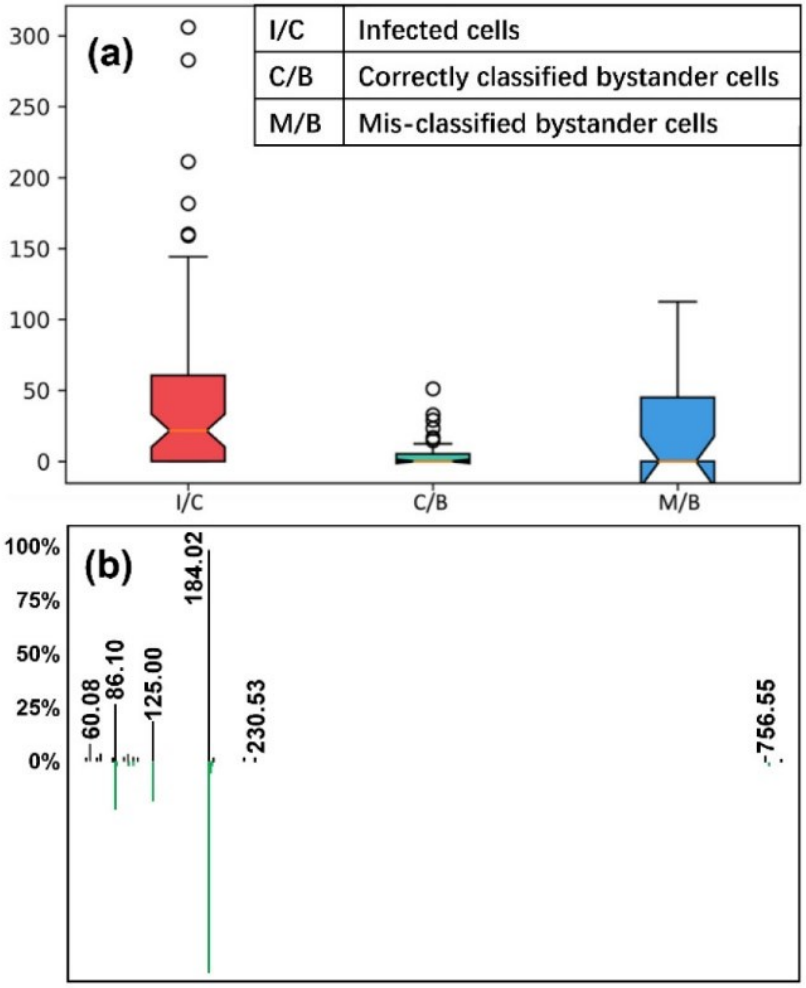


Figure 2-4 Representative glycerophosphocholine (*m/z* 756.547) differentiating

between cell groups. (a) Normalized intensity of PC(34:3) in three different cell types ($p = 0.000233$ using ANOVA test with FDR correction). (b) LC-MS/MS mirror plot supporting PC annotation. Green, reference library MS/MS spectrum for 1-Oleoyl-2-palmitoyl-sn-glycero-3-phosphocholine (PC 34:1). Black, experimental MS/MS spectrum for *m/z* 756.547. *M/z* 756.547 is smaller by 4.03 to 1-Oleoyl-2-palmitoyl-sn-glycero-3-phosphocholine.

2.4 Conclusions

In conclusion, we used the Single-probe SCMS technique for metabolomics studies of cells with heterogeneous infection by T. cruzi at the single-cell level. This represents, to the best of our knowledge, the first implementation of single-cell metabolomics in mammalian-infectious disease. We discovered that necessary cell fixation (to kill the parasites) and staining (to illustrate T. cruzi infection) have no significant influence on the overall cell metabolome (**Figure 2-2b**, S4). Our results demonstrate for the first-time bystander effects of T. cruzi on infection-adjacent uninfected cells (**Figure 2-2b**, **Figure 2-3**, **Figure 2-4**). Although our current studies cannot fully explain the mechanisms of the bystander effects, it is very likely that the uneven infection was due to the heterogeneity of host cells. The bystander cells may belong to a subpopulation of host cells containing lower levels of glucose, which is needed to support parasites' replication internally. It has been reported that T. cruzi amastigotes transport extracellular glucose to fuel their own metabolism and replicate in the host cytosol.[65] In addition, nutritional deficiencies in the host cells will lead the failure of T. cruzi infection.[66]

Our results provide a significant insight into CD pathogenesis, explaining lesion development in sites that do not contain parasites.[54, 55] This has major implications for

CD treatment, indicating that killing parasites alone may not be sufficient. Our results may explain the failure of Benznidazole Evaluation for Interrupting Trypanosomiasis (BENEFIT) clinical trial[67], and pave the way for future work to assess the role of metabolic heterogeneity in CD pathogenesis, tissue resilience, parasite dormancy and antiparasitic susceptibility.

2.5 Acknowledgements

The author would like to thank the funding sources which comprise start-up funds from the University of Oklahoma (to Dr. Laura-Isobel McCall). Dr. McCall. holds an Investigators in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund. The author's Ph.D. advisor, Dr. Zhibo Yang, is supported by the National Institutes of Health (Grant R01GM116116), National Science Foundation (Grant OCE-1634630), and Research Council of the University of Oklahoma Norman Campus. This work is adapted with permission from *Analytical Chemistry*.[68] Copyright 2022.

2.5 Author contributions

Conceptualization, Z.Y., L.-I.M.; methodology, Z.Y., L.-I.M., **T.D.N.**, Y.L., R.L.; sample preparation, S.S.K., L.-I.M., **T.D.N.**, Y.L.; single cell MS and MS/MS analysis, **T.D.N.**, Y.L.; LC-MS/MS analysis, J.J.H.; data analysis, **T.D.N.**, Y.L.; R and Python script preparation, Y.L.; resources, Z.Y., L.-I.M.; writing—original draft preparation, **T.D.N.**, Y.L.; writing—review and editing, L.-I.M., Z.Y., **T.D.N.**, Y.L.

Chapter 3 : Quantifying Cell Heterogeneity and Subpopulations Using Single Cell Metabolomics

3.1 Introduction

It has been well accepted that nearly all biological systems are heterogenous[1] due to genetic and phenotypic variances. Even within the isogenic cell populations, cell-to-cell heterogeneity is prevalent, because of stochastic processes in transcription, translation, and metabolism.[69] Uncovering cell heterogeneity is critical for studying fundamental cell biology and human diseases. For example, tumors contain heterogeneous distributions of malignant cells with varied physiological and biological properties.[70] Such cell-to-cell heterogeneity was reported as a result of intrinsic and extrinsic factors,[71] and recognized to play a key function in diseases evolution, drug resistance, and tumor relapse.[72] In particular, cell heterogeneity reflects the effectiveness of cancer treatment and management, because an escape of a small subpopulation of cells, such as circulating tumor cells and cancer stem cells, from drug treatment can cause disease remission.[73]

To date, a variety of single cell studies using different approaches (e.g., flow cytometry,[74] image-based signaling marker colonization,[75] single cell genomics,[75, 76] single cell transcriptomics,[77, 78] single cell western blotting,[79] and single cell metabolomics[80]) have revealed the coexistence of multiple cell subpopulations in the same environment. Among various platforms, single cell RNA-seq quantification[81, 82

has gained most attraction, likely due to the availability of the instruments (e.g., 10X Genomics[83]) and analytical software packages.[84-87] Although the transcriptomic profiling is informative and powerful, the downstream proteomic or metabolomic responses are still unclear. As metabolites can rapidly and accurately reflect cell status and functions, single cell metabolomics is a promising approach to uncovering cell heterogeneity. Among all analytical techniques, single cell mass spectrometry (SCMS) has become the most popular tool for single cell metabolomics studies.[14, 23, 88-94]

Heterogeneous cells could be grouped into subpopulations with similar biological traits[95, 96] (e.g., morphology,[97] surface marker expression level,[98] and intracellular metabolism[80]). Several SCMS metabolomics studies have been performed to group cells into different sub-groups based on individual characteristic metabolites.[99,100] However, to the best of our knowledge, no methodologies have been reported to use the overall metabolomic profiles of single cells to quantify the changes of cell heterogeneity and the associated cell subpopulations. The absence of relevant work is likely due to two major reasons: (1) the lack of the metrics to quantify cell heterogeneity using the single cell metabolomics profiling data and (2) suitable data analysis approaches that can determine cell subpopulations with minimum artificial bias without prior knowledge of specific subpopulations. In this regard, we report a comprehensive method combining SCMS experimental method with a novel bioinformatics tool to address these challenges.

Metastatic melanoma cancer cell lines have higher drug resistance than primary

melanoma cancer cell lines.[101] In this proof-of-concept study, we used two cancer cell lines as models: the primary melanoma (i.e., drug-sensitive cell line) WM115 (Figure S1A) and metastatic melanoma (i.e., drug-resistant cell line) WM266-4 (Figure S1B). Previous studies reported differential expression of genes and global proteins in these two melanoma cell lines to unveil various proteins that are associated with the drug-resistant phenotype.[102] The Single-probe SCMS experimental technique,[23, 25, 27, 28, 30, 37] a homebuilt method for real-time *in situ* data acquisition of live single cells, was combined with a novel bioinformatics tool, SinCHet-MS, for quantitative analysis of cell subpopulations. Briefly, we cultured cancer cells under normal conditions. Cells were attached to glass cover slips during incubation, and then treated by 1 μM vemurafenib, an anticancer drug for melanoma therapy, for 48 h, and then analyzed using the Single-probe SCMS method (Figure 1). Although batch-to-batch variation is commonly recognized in conventional metabolomic studies (e.g., using liquid chromatography (LC)-MS),[103, 104] it is under-appreciated in most SCMS metabolomic analyses. To accurately evaluate the cellular response to microenvironmental stimulus (vemurafenib), we examined the batch-to-batch variation, which is potentially introduced by minor difference in sample preparation and fluctuations of instrument conditions on different days, and thus separating it from biological variance. We performed experiments for both the control (untreated) and treated cells within a batch, and repeated the experiments on a different day for both cell lines (Table S1). We then performed data pre-treatment, including noise removal, background reduction, peak alignment, and ion intensity normalization, prior to analyses using SinCHet-MS (Figure

2).[105] Built on SinCHet,[106] a computational toolbox with a graphical user interface (GUI) for analyzing single cell mRNA expression and methylation data, we developed SinCHet-MS to analyze SCMS data by devising three crucial features: batch correction, a novel d-statistic for determining default cell subpopulation resolution for further investigation, and the sGF score (Subpopulation Generalized Fisher Product Score) for prioritizing biomarkers defining cell subpopulations.

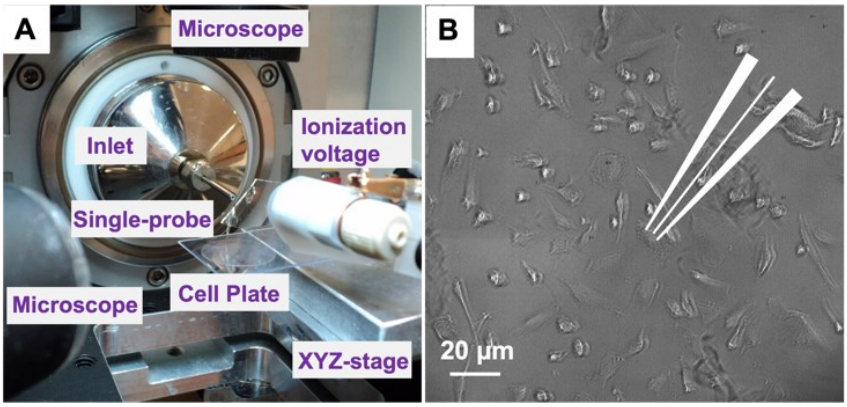


Figure 3-1 (A) Setup of the Single-probe SCMS experiment. (B) Analyzing a single cell guided by high-resolution microscopes.

3.2 Methods

3.2.1 Cell culture and sample preparation

Human melanoma cell lines, WM-115 and WM-266-4 cells (generously provided by Dr. Yinsheng Wang at the University of California, Riverside) were classified as the primary and metastatic cell lines, respectively, as established from the same melanoma patient. Cells were subcultured every three (WM-266-4) to five (WM-115) days in Dulbecco's Modified Eagle Medium (Santa Cruz Biotechnology Inc., Dallas, TX) supplemented with 10% fetal bovine serum (FBS, Life Technologies, Grand Island, NY, USA) and 1% penicillin-streptomycin (Life Technologies, Grand Island, NY, USA).

Cells were maintained in a cell incubator (HeraCell) at 37 °C in a humidified environment containing 5% CO_2. When cells reached ~ 80% confluence, they were rinsed twice using phosphate buffered saline (PBS) solution followed by trypsinization in the incubator for detachment. Trypsinization was quenched and the cell suspension were then transferred onto a glass cover slip (diameter = 18 mm, VWR International). After overnight incubation, cells were attached to the coverslip, and then transferred to the XYZ-translational stage system (MFA-CC, Newport Co., Irvine, CA, USA) for SCMS experiments. To conduct Vemurafenib treatment, 500 μM Vemurafenib stock solution in dimethyl sulfoxide (DMSO) (>99.9%, MilliporeSigma Co. St. Louis, MO, USA) was prepared and diluted in the complete culture medium at a final concentration of 1 μM. Cells after overnight culture were then treated with 1 μM Vemurafenib solution for a duration of 48 h and maintained in the incubator, followed by washing with fresh culture medium (without FBS) twice to remove residual drug molecules prior to SCMS analysis.

3.2.2 SCMS experiments

The Single-probe SCMS experiments were performed following our previously published protocols[23], and only a brief description is provided here. A Single-probe was fabricated by embedding a solvent-providing fused silica capillary (O.D. 105μm; I.D. 40 μm, Polymicro Technologies, Phoenix, AZ) , a nano-ESI emitter (produced from the same fused silica capillary using a butane micro torch) into a dual-bore quartz needle (produced from qual-bore quartz tubing (O.D. 500 μm; I.D. 127 μm, Friedrich &

Dimmock, Millville, NJ) using a laser micropipette puller (Sutter P-2000, Sutter Instrument, Novato, CA)). The Single-probe device was coupled to a LTQ Orbitrap XL mass spectrometer (Thermofisher Scientific, San Jose, CA) (Figure 1A). Cells were cultured, attached onto glass coverslip, treated by anticancer drug, and then rinsed by fresh culture medium (no fetal bovine serum). The glass coverslip containing cells was placed onto the XYZ-translational stage (step size = 0.1 μm). Guided by a digital microscope (Shenzhen D&F Co., China), a target cell was selected and penetrated by gradually moving the stage (Figure 1B). Cellular metabolites were extracted by the liquid junction (acetonitrile with 0.1% formic acid) formed at the tip of the Single-probe, and immediately ionized and analyzed. WM-115 and WM-266-4 cells prepared on the same day were randomly selected and analyzed with the MS analysis parameters listed as follows: ionization voltage +4.5 kV, mass range 150–2000, mass resolution 60,000 at m/z 400, 1 microscan, 100 ms max injection time, and automatic gain control (AGC) on.

3.2.3 SCMS Data Pre-treatment

The obtained SCMS raw datasets of all single cells were accessed using Xcalibur 5.0 (Thermofisher Scientific). Detection of single cells was confirmed from the mass spectra (including detected ions and their intensities) of common cellular species (e.g., PC (34:1); *m/z* 782.57). Background ions from the sampling solvent and culture medium were subtracted, and the instrument noise (i.e., ions with intensities < 10^3) was removed using an in-house developed software.[105] We normalized ion intensities of each cellular metabolite to the total ion current (TIC) and submitted the datasets to

Geena2[107] for peak alignment and isotope grouping. The pre-treated data were submitted to MetaboAnalyst 5.0[108] to select commonly detected species with 50% missing value (i.e., metabolites that can be detected in > 50% of all measured single cells). All datasets were subjected to log2 transformation prior to the downstream analysis.

3.2.4 SinCHet-MS

Built on SinCHet,[106] a bioinformatics toolbox for performing heterogeneity analysis of single cell transcriptomes, we introduced a new tool, SinCHet-MS, for analyzing SCMS data (Figure 2). Details on how to run SinChet-MS were described in the manual (Supporting Information). There are five panels: (1) Input/Output, (2) Data Processing, (3) Heterogeneity Analyses, (4) Subpopulation, and (5) Biomarkers. In "Heterogeneity Analyses", hierarchical cluster analyses were performed to group cells into subpopulations based on the similarities of metabolites' profiles. Clustergram and heatmaps are available for visual examination. The following three novel features of SinCHet-MS:

(1) Batch effect evaluation and removal. In the panel of "Data Processing", we include a function to evaluate and remove potential batch effects. First, principal component analysis (PCA), an unsupervised dimension reduction method, in "Data Processing" is used to quickly examine potential batch effect and then determine if undesirable batch or technical effect is observed. For instance, if data from different batches are clearly separated in PCA plot, especially when the separation of different

batches of cells is observed, say, along the first PC, which means the difference due to batch difference explains the largest variability of the data, while the experimental effect of research interest (e.g., treatment effect) explains less variability of the dataset. In this case, the user could remove this observed undesirable batch effect by selecting Yes from the drop-down option (YES/NO) for debatching. COMBAT,[109] a commonly used debatching method (based on empirical Bayes frameworks), often used in transcriptomic studies, was implemented to remove potential batch effects here.

(2) The *d* statistic. In the 'Subpopulation' panel, we introduce a d statistic to determine the default number of clusters for further investigation of subpopulations. The *d* statistic was modified from the *D* statistic defined previously[46]. Briefly, the *D* statistic, quantifies the overall change of heterogeneity before and after treatment, is defined as the areas under the Shannon Profiles (SPs) between two conditions. The *d* statistic is defined as the difference of the Shannon index (H) between two conditions of research interest at the minimum number of clusters with significance estimated using permutation (Equations S1 and S2 in the Supporting Information). It is worth noting that these two statistic methods are different: the D statistic can be perceived as the heterogeneity difference between two conditions considering all possible clustering resolutions, whereas our novel *d* statistic is defined as the difference of the Shannon index (H) between two conditions at a given clustering resolution of research interest. The default clustering resolution of SinCHet-MS is determined using the minimum number of clusters with the *d* statistic differs significantly between two conditions. SinCHet-MS has the flexibility for users to explore alternative clustering resolutions.

The statistical significance is estimated from the permutation test, similar to that reported in our previous publication[106] and described in the Supplemental Equation S2 (Supporting Information).

(3) Subpopulation_Generalized Fisher Product Score (sGF). As implemented in 'BioMarkers' panel, this function can be used to prioritize biomarkers. sGF was devised to summarize the overall difference among cell subpopulations for each metabolite, with the consideration of *p*-values from multiple comparison tests and fold change of pairwise comparison between the subpopulation of interest and any other subpopulation (Supplemental equations S3 and S4 in the Supporting Information). Those metabolites with Benjamini-Hochberg false discover rate (FDR) adjusted *p*-values < 0.05 were regarded as the subpopulational biomarkers.

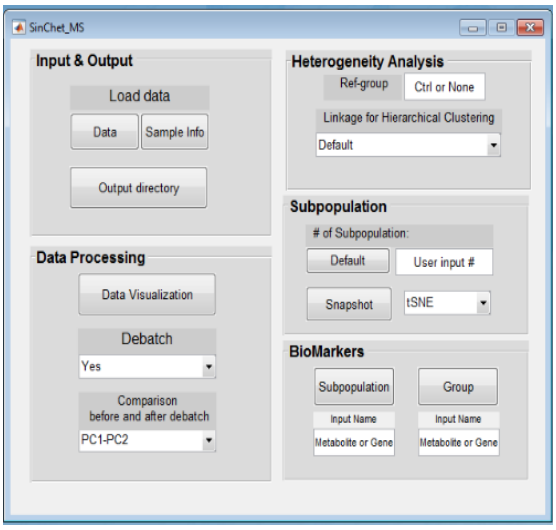


Figure 3-2 The main Graphic User Interface (GUI) of the SinCHet-MS software package. This GUI integrates functions of batch correction, determination and visualization of cell subpopulations, and prioritization of subpopulation diagnostic features.

3.3 Results and Discussion

In total, we analyzed 75 and 128 cells for WM115 and WM266-4, respectively. The representative mass spectra of single cells (before and after vemurafenib treatment) are shown in Figure S2. Rich metabolomic information can be observed in a mass range of *m/z* 650–950, which encompassed a variety of lipids, along with background ions (e.g., *m/z* = 493.25). Numerous species were detected in each cell. To extract the essential information from all mass spectra, experimental data need to be carefully treated and analyzed.

3.3.1. Batch correction for SCMS datasets

Batch effect may be present in SCMS experiments. Debatching can enhance statistical power by enabling concurrent data analysis across multiple batches obtained under the same experimental condition. We used PCA to visualize cellular metabolomic profiles (for all cells) to evaluate potential batch effect: a significantly different PCA grouping between two batches (i.e., from the same cell line with the same treatment conditions) indicates an evident batch effect. To remove the observed technical batch effect, we performed batch correction, which is integrated in SinCHet-MS based on COMBAT.[109]

Our experimental results indicate that for the sensitive cell line WM115, no significant batch effect was observed (data not shown). For the resistant cell line WM266-4, there is observed batch effect even though the biological differences are larger than the batch effect (Figure 3A). Figure 3B indicated that the debatch functions

of SinChet-MS can remove the batch effect for further analysis. However, group separation due to treatment effect was not observed in the first four PC dimensions before and after debatching (Figures 3A vs 3B; Figure S3A vs S3B). Similar trends can also be observed using box plot (Fig. S4). Such minimum change in cellular profiles agrees with earlier publications reporting higher drug resistance of metastatic melanoma, which corresponds with the hardness of medical therapies for melanoma cancer.[24] As a proof-of-concept study, we demonstrated batch correction of SCMS data, which were acquired from the same passage of cells but with different time of experiments. The batch effect originated from different cell passages needs to be investigated in future studies.

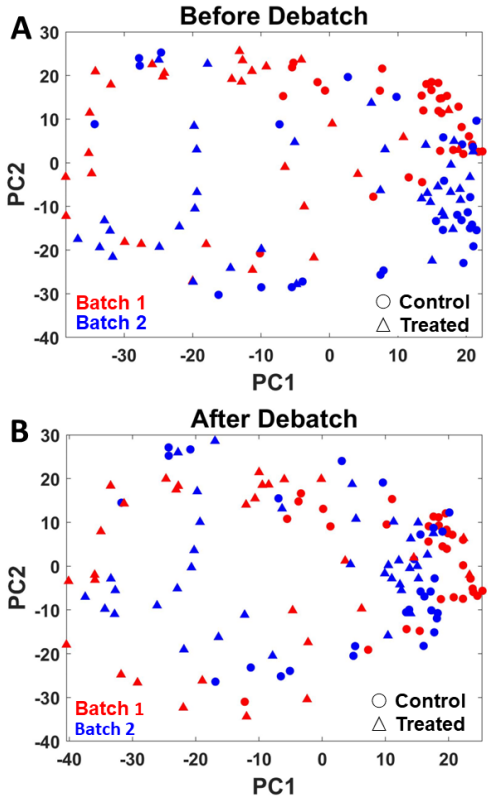


Figure 3-3 Batch correction. PCA plots of WM266-4 cell lines (A) before and (B) after COMBAT in the PC1 and PC2 dimensions. (Symbols represent control (○) and treatment (Δ) groups, and colors represent batch 1 (red) and batch 2 (blue) experiments.)

3.3.2. Quantitative analysis of cell subpopulations and heterogeneity differences

Cellular subpopulations can be reflected from variances of metabolite expression levels and their associated metabolic noise distributions. Previous studies have investigated cell subpopulations by fitting the relative abundances of certain metabolites using probability functions such as normal, lognormal, and gamma distributions.[93, 110] These pioneering studies have demonstrated the possibilities of using the combined SCMS experiments and statistical data analyses to reveal subpopulations of cells. However, only a limited number of metabolites were selected for analyses, whereas the majority of mass spectra information was not efficiently used. Our previous studies, in which machine learning and SCMS experiments were combined to predict phenotypes of cells, indicate that analyses involving all detected species produced higher reliability than those using selected metabolic biomarkers, which may result in information loss.[25, 30]

For the first time, we performed systematic and quantitative analysis of the changes of subpopulations. In the current study, all detected ions (after pretreatment) were utilized for cell heterogeneity analyses. From SinChet-MS, our results showed that metabolomic heterogeneity was significantly changed by drug treatment in both cell lines ($D = 83.2$, $p < 0.001$ for WM115 (Figure S5A); $D = 54.2$, $p < 0.001$ for WM266-4 (Figure S5B)). When there is no different D (with statistical significance between conditions (i.e., $p \geq 0.05$)), the default number of cell clusters for user investigation is determined using the minimum value of the change points derived from the multivariate

adaptive regression splines (MARS) model[106]. In this study, the number of subpopulations is determined using the *d* statistics with heterogeneity significantly different between the control and treatment groups (d = 0.13, $p < 0.001$ for WM115 (Figure 5A); d = 0.02, $p < 0.001$ for WM266-4 (Figure 5B)). The composition of cell subpopulation before and after the treatment were examined and visualized in hierarchical heatmap and pie chart (Figure 4). For the sensitive cell line WM115, there was only one population in the control group; however, a new subpopulation emerged after treatment (Figure 4A and 4B). For WM266-4 cells, the number of subpopulation (two) was unchanged, but their relative abundances were altered: ~7:3 and ~3:7 in the control and treatment groups, respectively (Figure 4C and 4D). The subpopulations of these two cell lines can also be intuitively visualized using tSNE (t-distributed stochastic neighbor embedding) without detailed quantitative information (Figure S6). These findings agreed with published studies reporting an increase in cell heterogeneity upon drug treatment for cancer cell line[24]. In addition, as shown in the representative single cell MS spectra corresponding to each subpopulation (Figure S7), the metabolomic features (*m/z* 650–950) can be visually differentiated.

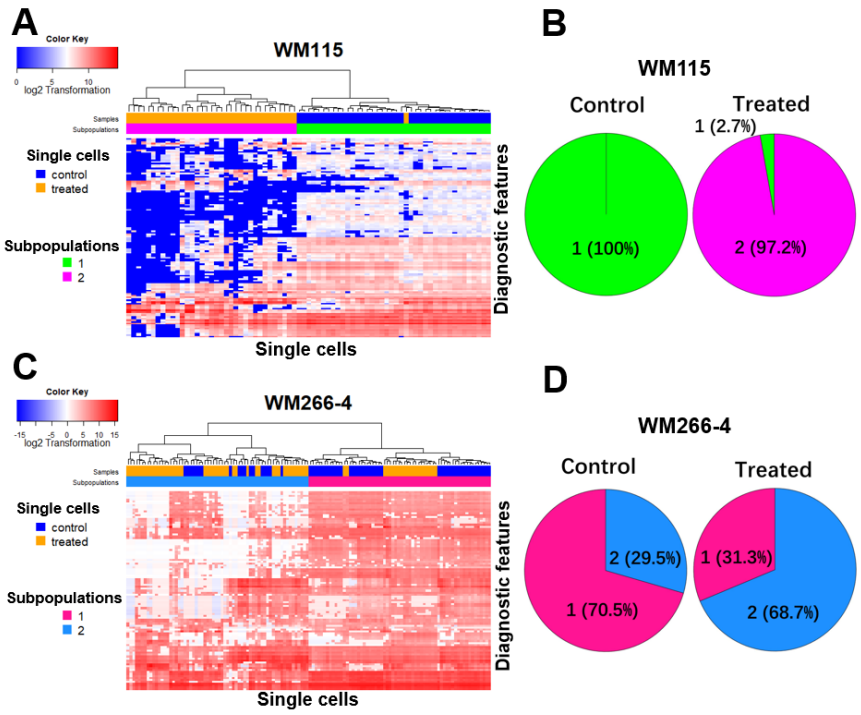


Figure 3-4 Visualization of subpopulation compositions of control and drug treated single cells before and after drug treatment for (A & B) WM115 and (C & D) WM266-4 cell lines using hierarchical heat map (left column) and pie chart (right column). The determination of cell subpopulation is based on the minimum number of clusters where d statistic indicates significantly different heterogeneity found between control and treatment.

3.3.3. Identification of Subpopulational Biomarkers

As metabolites reflect cell metabolism, cells in different subgroups likely possess different metabolomic features. For the first time, we identified and prioritized biomarkers for different subpopulations. SinCHet-MS provides three types of GF scores (Generalized Fisher scores): 1) subpopulation GF score (sGF), 2) grouped GF score (gGF), and 3) the widely used GF score as previously described.[106] In the current study, sGF was utilized to prioritize subpopulation diagnostic features. The top three species with the highest sGF, representing significant contributions to the subpopulation discrimination, were visualized among all subpopulations in both cell lines (Figure 5).

The diagnostic features with high loadings for the first two PCs were also displayed in the loading plot for each cell line (Figure S8). It is worth noting that, unlike metabolomic biomarkers discovered in conventional LC-MS[111, 112] and other single cell metabolomic studies,[90, 113, 114] the subpopulation biomarkers reported here were based on differences of metabolites' abundances among cell subpopulations, rather than different treatment conditions or types of cells. Based on multiple comparison among subpopulations (FDR < 0.05), we prioritized 95 and 67 subpopulational diagnostic features from WM115 and WM266-4 cells, respectively (Table S2–S3). Further, we conducted MS/MS to identify these diagnostic features using single cells and cell lysates (as detailed in the Supporting Information). As a result, a majority of identified subpopulation biomarkers are cellular lipids (e.g., phosphatidylcholine, phosphatidylethanolamine, diacylglycerol, triacylglycerol) that are related to cellular signal transduction,[115] and their compositions are sensitive to cells' ambient microenvironment.[116] Through the current proof-of-concept studies, these identified subpopulational biomarkers are likely associated with the emergence, expansion, or reduction of certain cell subpopulations due to change of microenvironment (i.e., drug treatment). However, the correlation between cell subpopulations and drug mechanisms are still to be understood.

3.3.4. Evaluation of Technical and Biological Variation of SCMS Datasets

We considered the influence of SCMS technical variation on our analyses. Our previous studies indicated that the technical variation of the Single-probe SCMS method is ~20%, which is determined from the relative standard deviation (RSD) of the ion intensities of standard compound solutions measured using different Single-probe devices.[30] The combined biological and technical variance (RSD = 95%~110%) of cells induced by drug treatment was estimated from the RSD of intensities of the top-121 (WM115) and top-103 (WM266-4) ions due to drug treatment. Similarly, the combined biological and technical variance between different subpopulations (RSD = 75~98%) was also estimated from those 95 (WM115) and 67 (WM266- 4) metabolite biomarkers of subpopulations. Our results indicate that the technical variance is significantly less than the combined biological and technical variance in all cases, indicating the reliability of discovered biomarkers using our method.

3.3.5. Limitations of SCMS Datasets

One limitation for this proof-of-concept study is that the sample size is small for the analysis, primarily due to the limited throughput of the Single-probe SCMS technique. We analyzed a total of 75 WM115 cells (i.e., 31 and 44 cells from batch 1 and 2, respectively) 128 WM266-4 cells (i.e., 62 and 66 cells from batch 1 and 2, respectively). To evaluate if our biological findings shown above are reproducible, we performed the combined analysis (by pooling the data from both batches) and compared the results with those per-batch analyses, performed using the data from each batch separately. The observed agreement (Figure S9 and S10) indicated that findings from

per-batch analysis are similar to those from the combined analyses. For sensitive cell line WM115, two subpopulations were identified in batch 1 and batch 2, respectively. Those two subpopulations represent the control and treated groups of cells separately (Figure S9A and S9B), which are the same as those two subpopulations identified in the combined analyses (Figure 4A). In the combined analysis, only one cell from the treatment group was clustered differently from the rest of the cells in the treatment group (Figure 4A). In addition, there are statistically significant correlations of sGF score, which were used to prioritize the subpopulation diagnostic features, between each batch and the combined analysis (Figure S9C). For the resistant cell line WM266-4, even though the proportions of two subpopulations in batch 1 and batch 2 are different (Figure S10A and S10B), their proportions changed by treatment are similar to findings in the combined analysis (Figure 4C and 4D). Furthermore, there are statistically significant correlations of sGF score of subpopulation biomarkers between analyses by each batch and results from the combined analysis (Figure S10C). In summary, per-batch analysis generated similar conclusions drawn from the combined analyses. Although the sample size is small, but the observations made between batches were similar.

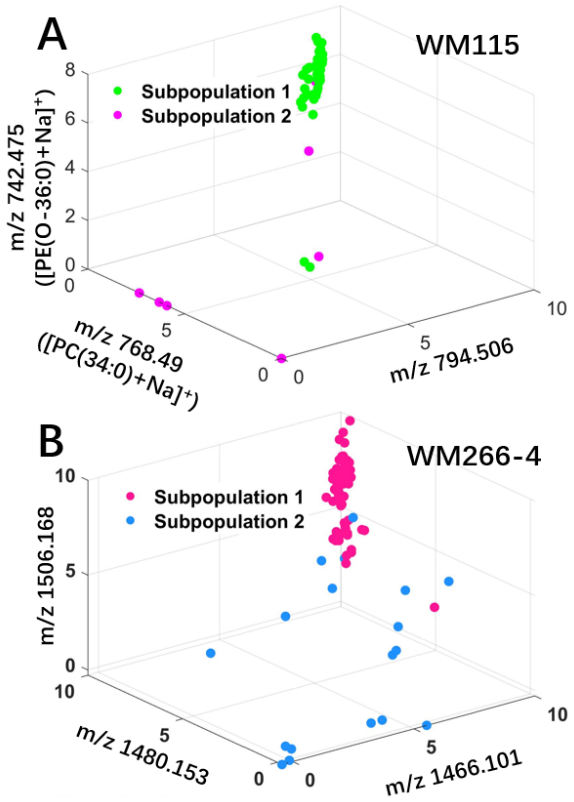


Figure 3-5 Relative abundances of top-three subpopulation diagnostic features (with top-three highest sGF scores) for (A) WM115 and (B) WM266-4 cell lines. Annotated species were identified through MS/MS analyses.

3.4 Conclusions

In conclusion, for the first time from a metabolomics perspective, we reported a combined experimental and bioinformatic method to reveal changes of cell heterogeneity and quantify subpopulation compositions. Cellular metabolomic profiles of drug-sensitive and drug-resistant cancer cells were measured using the Single-probe SCMS techniques, and experimental data were subjected to batch correction prior to downstream analysis. Using comprehensive statistical analyses, we revealed that the subpopulations were evidently changed, and a new subpopulation emerged in drug-sensitive primary melanoma cancer cells (WM115) treated with vemurafenib. The

emergence of new subpopulations was not clearly observed in the drug-resistant cell line (WM266-4); however, it was evident that proportional change between subpopulations occurred. Although the exact correlation between the determined cell subpopulations and specific cellular biophysical properties of each subpopulation is currently unclear, our technique provides a new label-free method, which is different from traditional targeted approaches (isotope tracing,[117] fluorescent labeling,[118] etc.), to study subpopulations. In addition, we integrated multiple functions for SCMS metabolomic studies, including the batch correction, visualization of cellular metabolomic profiles, comparison of cell heterogeneity, determination of subpopulations, and prioritization of subpopulational biomarkers, in a package with a user-friendly GUI (SinCHet-MS, freely available for non-profit academic use at http://lab.moffitt.org/chen/software/). SinCHet-MS could be applied to analyze single cell metabolomic datasets obtained from different instrument platforms. Profound understanding of cellular metabolism can be gained not only from the cell heterogeneity perspective, but also at the subpopulational resolution.

3.5 Acknowledgements

The author would like to thank Dr. Yinsheng Wang (University of California, Riverside) for generously providing the cell lines. Besides, the author would like to collaborator, Dr. Y. Ann Chen, and Dr. Jiannong Li (Moffitt Cancer Center, Tampa, FL), Dr. Zhibo Yang (my Ph.D. advisor), Dr. Renmeng Liu and my co-first author (Yunpeng

Lan) for all their help with this project. This work is adapted with permission from *Analytical Chemistry.*[119] Copyright 2023.

3.6 Author contributions

Conceptualization, Z.Y., Y.A.C., R.L., J.L.; methodology, Z.Y., Y.A.C., R.L., J.L., **T.D.N.**, Y.L.; single cell MS and MS/MS analysis, **T.D.N.**, Y.L.; MS data analysis, **T.D.N.**, Y.L.; Python script preparation, Y.L.; SinCHet-MS software development and bioinformatic analysis, Y.A.C, J.L.; resources, Z.Y., Y.A.C; writing—manuscript preparation, Z.Y., Y.A.C., R.L., J.L, **T.D.N.**, Y.L.

Chapter 4 : Using Single Cell Mass Spectrometry to Evaluate CRISPR-Cas9 Gene Editing Results

4.1 Introduction

Cas9 protein[120], which belongs to the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-CRISPR-associated (Cas) system, is a revolutionary and powerful tool in the field of molecular biology and genomic engineering.[121, 122] CRISPR-Cas9 system comprises a Cas endonuclease (Cas9) and a guide RNA.[123] A Cas9 enzyme, which acts as a molecular scissor, cuts the DNA at a specific site targeted by a single guide RNA (sgRNA) molecule.[124] Once Cas9 introduces a double-strand DNA (dsDNA) break in the host cell's genome, it can be repaired by the cell's natural DNA damage repair machinery.[124, 125] CRISPR-Cas9 has been widely used to alter the genome of mammalian cells which allow gene editing by knocking-in or knocking-out of a host genome rapidly and efficiently.[126] One of CRISPR-Cas9 gene knockout techniques is to deliver plasmids into the cells via transfection to express the Cas9 enzyme and the sgRNA.[127] As an established gene editing technique, CRISPR-Cas9 can add and remove nucleotides, thus altering genomic sequences.[128] This technology offer several advantages in gene editing, including its precision in enabling highly targeted modifications in the genome,[129] its versatility across a wide range of cell types and organisms (e.g., *in vitro* cell models, *in vivo* animal models, human cells, cancer cells, crop plant, fungi),[129, 130] and its application in developing genetically modified crops that are more resistant to diseases, pests or have enhanced nutritional value in

biotechnology and agriculture research.[131, 132] Despite these advantages, the CRISPR-Cas9 system also presents several drawbacks. These include variance in editing efficiency[133] and off-target effect,[134] where the Cas9 nuclease binds and cleaves DNA at off-target sites in the genome.[135] As a result, CRISPR-Cas9 knockout often suffers from low editing efficiencies and creates heterogenous populations of cells.[136] Bulk analysis methods are not able to investigate heterogeneous population resulting from off-target effects, establishing new analytical methods is essential to evaluate the gene editing efficiency, and measure off-target effects in single cells.[137, 138]

Single cell technologies (e.g., single cell genomics[139] and transcriptomics[140]) are capable of profiling individual cells and uncovering rare and hidden subpopulations of cells.[9] Conventional bulk analysis ignores the unique behavior resulting from cell-to-cell variations, specifically cellular metabolism.[1, 141] Due to the rapid turnover rate of cell metabolites, which are final products of cell metabolism, metabolomics studies of live single cells can reveal status and molecular features of rare cells that cannot be studied using traditional bulk analysis. Mass spectrometry (MS) is a powerful analytical technique to study single cell metabolomics. Various methods for single-cell mass spectrometry have been developed, employing different sampling and ionization techniques. These techniques fall into two major categories: vacuum-based and ambient-based methods. Vacuum-based techniques, such as matrix-assisted laser adsorption/ionization-mass spectrometry (MALDI-MS)[142] and secondary ion mass spectrometry (SIMS),[143] require samples to be analyzed in vacuum conditions. On the other hand, ambient-based techniques, including live single-cell mass spectrometry,[19]

probe electrospray ionization MS (probe ESI-MS),[144] laser ablation electrospray ionization MS (LAESI MS),[21] nanospray desorption electrospray ionization MS (nano-DESI MS).[22] We have developed multiple ambient-based MS techniques, in which we fabricated our own sampling and ionization devices, such as the Single-probe,[23] micropipette capillary,[145] and T-probe,[146] to couple with mass spectrometry for live single cells in an ambient environment (i.e., room temperature and atmospheric pressure). The Single-probe SCMS technique has been routinely used in our studies,[23,25, 27-30] and this novel method can be potentially utilized to investigate the cell-to-cell variation in cell metabolism resulting from altered genes.

The Single-probe is a microscale device that can be coupled to a mass spectrometer for sampling and ionization in single-cell metabolomics studies. This method has been applied to investigate cell heterogeneity in parasite infection in Chagas disease,[68] study influence of drug treatment on cell metabolism,[119] quantify anticancer drugs in individual cells,[15, 33] compare metabolites in cancer stem cells and non-stem cancer cells,[37] and assess the influence of the environment on algal cell metabolites.[27] Additionally, the Single-probe device has been used in previous mass spectrometry imaging studies to determine the spatial distribution of molecules on tissue slices[42] and analyze secreted metabolites from multi-cellular spheroids.[43, 44]

Our long-term goal is to develop a novel approach of integrating CRISPR-Cas9 gene editing with Single-probe SCMS metabolomics to rapidly evaluate the gene editing results. The central hypothesis is that modifying a functional gene can induce alteration in cell functions and metabolism. Specifically, fatty acid synthase (*FASN*)

regulates fatty acid synthesis, thereby implying the fatty acids and lipids profiles in *FASN*-knockout cells will differ from those in control cells. The research plan is novel as it marked the integration of Single-probe SCMS metabolomics and CRISPR-Cas9 gene editing techniques. This research approach presents a novel strategy to investigate gene editing efficiency and establish a direct correlation between cell metabolomics, lipidomics and genes at the single-cell level. The innovation lies in using SCMS metabolomics as a method to assess successful target gene editing in individual cells. Moreover, this novel approach expedites the determination of gene editing efficiency without conducting sequencing and allow for the investigation of off-target effects within the CRISPR-Cas9 system.

In this study, fatty acid synthase (*FASN*) is selected as a target gene for knockout in HEK293T, a human embryonic kidney cell model. Because *FASN*-knockout inhibits the *de novo* synthesis of fatty acids and mediates the lipid metabolism, the changes in fatty acids and lipids are expected to be observed in successfully gene edited cells. *FASN* overexpression is associated with nonalcoholic fatty liver disease (NAFLD), nonalcoholic steatohepatitis (NASH), and many types of cancers.[147] In particular, *FASN* knockouts can decrease metastatic potential in multiple cancers such as breast,[148] prostate[149], and colorectal[150] cancers. *FASN*-knockout cells are generated through the use of CRISPR-Cas9 Knock-out (KO) Plasmids (Santa Cruz Biotechnology, Dallas, TX), containing *cas9* and sgRNA genes. The sgRNAs are typically about 100 nucleotides (nt) long.[151] Each sgRNA contains its unique 20 nt guide region that directs the protein to cut a specific site within the targeted DNA (Figure 1). In the *cas9* gene

knockout process, the guide RNA sequence present in sgRNA that is bound to *cas9* to guides *cas9* to the *FASN* locus in the genome. Targeting of the right DNA activates *cas9* to induce a site-specific double-strand break (DSB) in *FASN* DNA of the target cells. This precise DSB formation disrupts production of the *FASN* protein in edited cells, leading to its loss of function. As a result, *FASN*-knockout cells are generated.

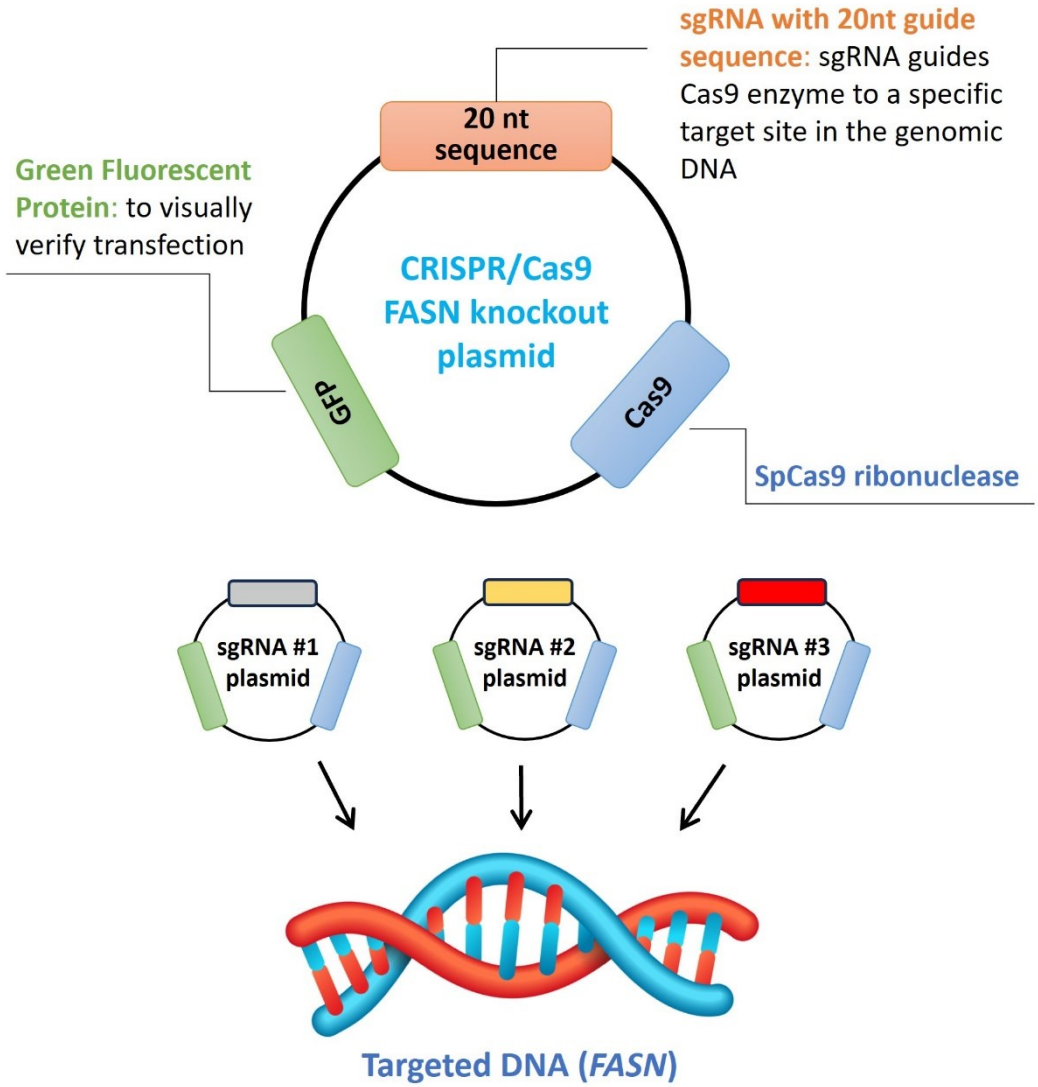


Figure 4-1 CRISPR-Cas9 Knockout (KO) plasmid (#sc-400440-KO-2) (Santa Cruz Biotechnology, Dallas, TX) comprises three distinct plasmids designed to improve the identification and cleavage of *FASN* gene to maximize *FASN*-knockout efficiency. Each plasmid contains its unique 20nt guide region, which cuts at a different site within the targeted DNA (*FASN*).

4.2 Methods

4.2.1 Sample preparation

4.2.1.1. Cell culture

Human embryonic kidney cells, HEK293T cells, were purchased from ATCC (Manassas, Virginia). HEK293T cells were sub-cultured in T-75 flasks every three days in Dulbecco's Modified Eagle Medium (DMEM) cell culture medium (Gibco) supplemented with 10% fetal bovine serum (FBS, Life Technologies, Grand Island, NY, USA) and 1% penicillin-streptomycin (PS, Life Technologies, Grand Island, NY, USA) in 5% CO_2 incubator at 37 ^oC (HeraCell, Heraeus, Germany).

4.2.1.2. Plasmid transformation and isolation

Chemically competent E. coli DH5α cells (Invitrogen, Waltham, MA) were used for the plasmid transformation. The plasmid containing both cas9 and single-guide RNA (sgRNA) expression cassettes was added to the competent cells and incubated for 30 mins on the ice. This was followed by heat shock for 45 seconds at 42 °C in the water bath and placing them back to the ice for 2 mins (PMID: 2265755; 18997900)[152]. Heat shock is used to temporarily form pores in the cell membrane, allowing transfer of exogenous DNA into the cell. After the heat shock, the transformed cells were incubated in LB media (or SOC media) for 1 hour at 37 °C for recovery. The transformed cells were then pelleted and resuspended in the same recovery media and plated 50 uL on LB agar plate (with ampicillin, 100 ug/ml) followed by overnight incubation at 37 °C.

Transformants grown on the plates were confirmed after extracting the plasmids from single colonies and sequencing the plasmid through nanopore sequencing (Plasmidsaurus, Eugene, OR).[153] Oxford nanopore sequencing were carried out for different isolated plasmid constructs and successfully obtained the distinct sequences of three distinct sgRNA plasmids (sgRNA #1: GATCGGTGGCGTGCACCCGC; sgRNA #2: TGGTGATTGCCGGCATGTCC; sgRNA #3: GGATGGTGGCGTACACCCGC), which were confirmed with the sequences of FASN CRISPR/Cas9 KO Plasmid (#sc-400440-KO-2) provided by the vendor (Santa Cruz Biotechnology, Dallas, TX).

4.2.1.3. Lipofectamine Transfection

Three different isolated plasmids (sgRNA #1, #2, and #3) were prepared to transiently transfect HEK293T cells using Lipofectamine 2000.[154] HEK293T cells (0.3 x 10^6 cells) were plated in a 6-well plate and cells were transfected at 70 – 80% confluency. Reduced serum medium (Opti-MEM, Gibco) was substituted with complete cultured medium prior to transfection. A ratio of 1:20 Lipofectamine 2000: opti-MEM medium (v/v) was mixed and incubated for 30 minutes to allow complex formation. After that, plasmid DNA was added to the Lipofectamine-containing mixture at a ratio of 1 µg plasmid DNA: 5 µL Lipofectamine and incubated for 30 minutes to allow the formation of DNA-Lipofectamine complex. The DNA-Lipofectamine complex was then added to the cells and incubated. After 24 hours, complete cultured medium (with 10% FBS and 1% Penicillin-Streptomycin) was replaced to supplement nutrients to the transfected cells. Over the remaining 48 hours, the cells express the

sgRNA plasmids which can be observed as fluorescent cells using Green Fluorescent Protein as a marker. At 72 hours, the cells' microscopic images were captured to monitor the successfully transfected cells. (Figure 4). Then, cells were washed with PBS, quenched with trypsin, resuspended in new complete cultured medium prior to sorting.

4.2.1.4. GFP-positive cells sorting

Transfection efficiency was observed at 72 hours post transfection. To isolate successfully transfected (GFP-positive) cells for both SCMS and Western blot experiments, fluorescence-activated cell sorting (FACS, BD FACSAria III Cell Sorter, BD Biosciences, San Jose, CA, USA) was performed under sterile conditions to obtain the GFP-positive cells. Regular HEK293T cells (no GFP-expression) were used as negative control to gate GFP-positive cells.

4.2.2 Single-probe SCMS

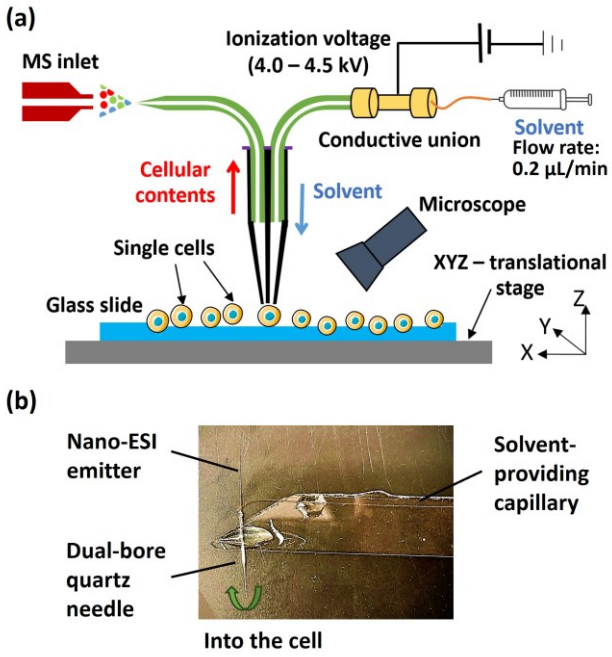


Figure 4-2 (a) Diagram of Single-probe SCMS experiment set-up. (b) Real-life image of the Single-probe, which were made of three major components, including nano-ESI emitter, solvent-providing capillary, and dual-bore quartz needle.

4.2.2.1. Single-probe fabrication

The single-probe comprises three major components, including a nano-ESI emitter, a solvent proving capillary, and a dual-bore quartz needle. The Single-probe SCMS setup consists of a Single-probe, a digital microscope, a digital camera, a computer-controlled XYZ-translation stage system (CONEX-MFACC, Newport Co., Irvine, CA, USA), and a Thermo LTQ Orbitrap XL mass spectrometer (Thermo Scientific, Waltham, MA, USA). Details about the fabrication of the Single-probe and the SCMS setup can be found in our previous studies.[23, 68, 155] In summary, the Single-probe was created using laser-pulled dual-bore quartz tubing (O.D. 500 µm; I.D. 127 µm, Friedrich & Dimmock, Inc., Millville, NJ, USA) embedded with a fused silica capillary (O.D. 105 µm; I.D. 40 µm, Polymicro Technologies, Phoenix, AZ, USA) in one channel and a

nano-ESI emitter, made from the same fused silica capillary, in another channel. These components were sealed using UV curing resin (Light Cure Bonding Adhesive, Prime-Dent, Chicago, Il, USA).

4.2.2.2. SCMS analysis of *FASN* knockout cells and controls

During the experiment, glass coverslips containing cells were washed with fresh DMEM and placed on the XYZ-stage system of the Single-probe SCMS setup for data collection. The single cells were chosen for analysis by precisely moving the stage system under the guidance of the microscope. Four groups of cells (HEK293T controls, GFP-positive (sgRNA#1), GFP-positive (sgRNA#2), and GFP-positive (sgRNA #3) cells) were prepared for SCMS experiments. GFP-positive cells were sorted using Fluorescent activated cell sorter (FACS) flow cytometry and recovered for 12-24 hours prior to SCMS analysis. Each individual cell was carefully analyzed using the Single probe with an optimized tip size of ~9 μm. The sampling solvent (acetonitrile with 0.1% formic acid) was continuously delivered through the fused silica solvent-providing capillary to extract small molecules. The extracted cellular contents were then ionized via the nano-ESI emitter and analyzed in real-time SCMS analysis. The MS experiments were conducted in both positive ion mode (ESI+) and negative ion mode (ESI-) with an *m/z* range of 150-2000, and 50-900, respectively. Other MS parameters were described with a flow rate of 200 nL/min; a mass resolution of 60,000; a 4.5 kV ionization voltage; 1 microscan; and a maximum injection time of 100 ms. MS/MS experiments followed these parameters: CID, HCD mode; 200 nL/min flow rate; 60,000 mass resolution; 4.5 kV ionization voltage; 3 microscans; and 500ms maximum

injection time. Normalized collision energies ranged from 15-30 NCE, and 30-45 NCE for CID, and HCD mode, respectively.

4.2.2.3. SCMS data analysis

Our pretreatment of SCMS data adhered to established protocols.[23] Initially, MS data peaks (m/z values and relative intensities) were exported using Thermo Xcalibur Qual Browser 3.0 (Thermo Scientific, Waltham, MA, USA). Noise subtraction involved eliminating peaks with relative intensity below $3x10^3$ in positive ion mode, and below $3x10^2$ in negative ion mode. Background signals from organic solvent and cell culture medium were removed using a customized R script, as detailed in previous studies.[105] Ion intensities were normalized to total ion current (TIC) and underwent peak alignment (with a 10 ppm mass tolerance) using *Geena2* online tool (http://proteomics.hsanmartino.it/geena2/geena2_ssi_norm.php)[107] with specified parameters. Geena2 parameters were as follows: analysis range from 150 to 2000 m/z, maximum number of isotopic replicas: 3, maximum delta between isotopic peaks: 0.01 Da, maximum delta for aligning replicates: 0.01 Da and maximum delta for aligning average spectra: 0.01 Da. After performing peak alignment, missing values (50%) were removed using an in-house Python script (SI Supporting File).

Subsequently, the pretreated data underwent statistical analysis, including principal component analysis (PCA), ANOVA, student *t-test*, and hierarchical clustering, conducted using Metaboanalyst 5.0.[156] Significantly different ion abundances among cell groups were determined using one-way analysis of variance (ANOVA)[157] (for >2 cell groups) or student *t*-test[119, 158] (for 2 cell groups) with an adjusted p-value of 0.05

using False Discovery Rate (FDR) correction. Hierarchical clustering heatmaps[119] were generated using Ward's minimum variance clustering method and Euclidean distance. Hierarchical clustering heatmaps[119] were generated using Ward's minimum variance clustering method and Euclidean distance. Technical variance[30, 158] was minimized by conducting two replicates under similar experimental conditions. Boxplots depict the median along with the upper and lower quartiles, and their whiskers extend to the highest and lowest quartiles. The resulting SCMS/MS and LC-MS/MS data were annotated using databases such as METLIN, HMDB, and GNPS.

4.2.3 Western Blotting analysis

HEK293T cells were collected and centrifuged to form a cell pellet before preparing the cell lysate. The cell pellet was lysed in 1x RIPA lysis buffer (25 mM Tris-HCl, 150 mM NaCl, 1% NP40, 1% sodium deoxycholate, 0.1% SDS), containing 1 mM phenylmethanesulfonyl fluoride (PMSF) protease inhibitor cocktail (Sigma-Aldrich, Burlington, MA). Cells were lysed through sonication using an Ultrasonic processor at an amplitude of 40 for approximately 20-30 pulses. The lysate samples were then centrifuged, and total protein concentration was determined by the Pierce BCA assay (Thermo Fisher Scientific, Rockford, IL). 20ug of protein samples were resolved by electrophoresis on 8% SDS-PAGE gels and transferred onto nitrocellulose membranes. The membranes were blocked in Licor blocking buffer (Licor, Lincoln, NE) at room temperature for 1 hour and then incubated with primary antibodies at 4oC overnight. Then next day, blot was rinsed in 1x Tween-TBS thrice and incubated in anti-

rabbit IR 650 or anti-mouse IR 800 fluorescent secondary antibody (Licor, Lincoln, NE) at 1:25,000 dilution for 1 h at room temperature. The primary antibodies used were as follows: anti-FASN (Cell signaling, Danvers, MA) Rabbit mAb at 1:1000, anti-GAPDH (Proteintech, Rosemont, IL).). Protein expression was visualized using Azure C600 Imager (Dublin, CA). Statistical analysis of immunoblotting results was performed using one-way ANOVA with Tukey's multiple comparison test was used to compare means across genotypes (GraphPad Prism version 9). The threshold for statistical significance was set to $p<0.05$.

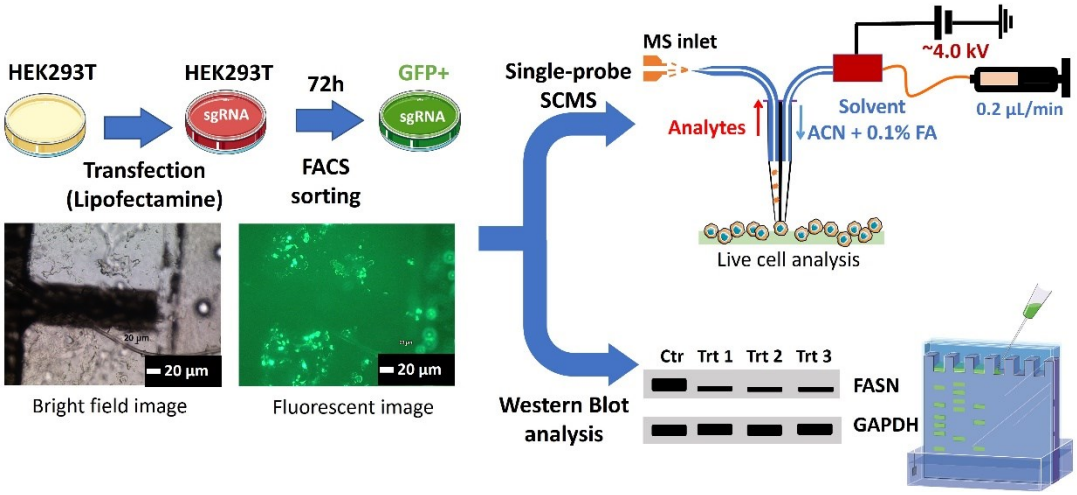


Figure 4-3 The working principle of assessing CRISPR-Cas9 gene editing result, specifically assessing the efficiency of knocking out *FASN* in HEK293T cells, by combining Single-probe SCMS with Western Blot analysis. The bright field and fluorescent images of transfected cells with sgRNA #1 plasmid were captured at 72-hour post transfection.

Table 4-1 Cells analyzed by the Single-probe SCMS technique.

No. of detected single cells	Control (HEK293T)	GFP+ cells (sgRNA #1)	GFP+ cells (sgRNA #2)	GFP+ cells (sgRNA #3)
ESI+	45	41	42	41
ESI-	58	49	50	57

4.3 Results and discussions

4.3.1. Transfection efficiency of HEK293T cells

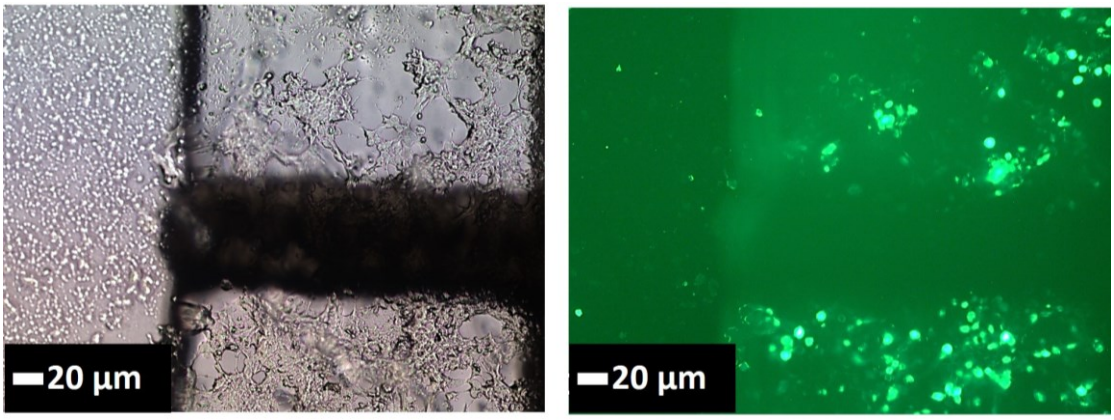


Figure 4-4 Representative microscopic images were captured for transfected cells with sgRNA #2 plasmid at 72-hour post transfection with their (a) bright field image and (b) fluorescent image using Fluorescence Microscope (with 10X lens).

Lipofectamine transfection facilitated the transient transfection of three distinct sgRNA plasmids (sgRNA #1, #2, and #3) into HEK293T cells, creating three different cells groups post-transfection. Figure 4 illustrates the assessment of transfection efficiency 72 hours after transfection. Successfully transfected cells expressed the green fluorescent protein (GFP) present in the plasmid as a marker. In contrast, cells that did not take in the *cas9*-sgRNA plasmid remained non-fluorescent. Bright field and fluorescent images were captured using a fluorescence microscope (10X objective lens). ImageJ software was employed for rapid counting of green, fluorescent cells, total cells, and estimation of the percentage of successfully transfected cells. Transfected (GFP-positive) cells were isolated through cell sorting, as GFP expression did not necessarily guarantee *FASN* gene knockout. Three groups of GFP-positive cells (transfected cells

with sgRNA #1, #2, and #3) and regular HEK293T cells were subjected to perform Single-probe SCMS metabolomics and Western blot experiments to monitor the alterations in cell metabolism and *FASN* protein expression resulting from *FASN* knockout.

4.3.2. Overall metabolic profiles of GFP-positive cell

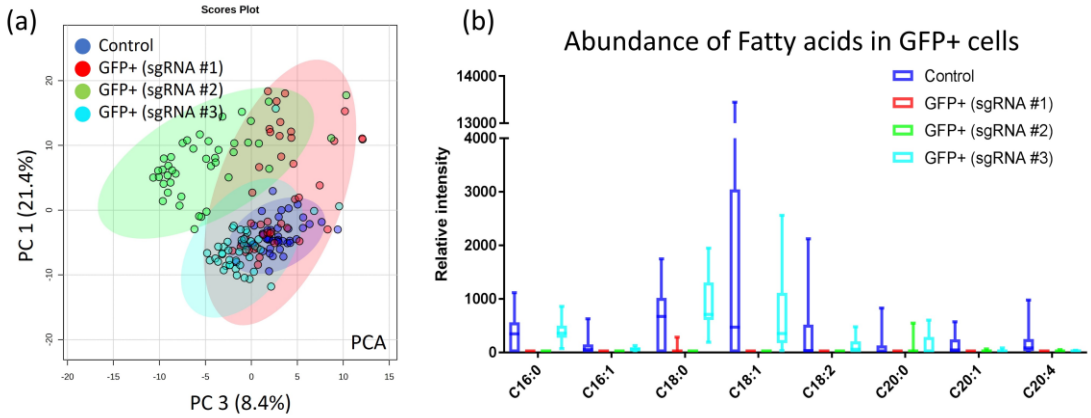


Figure 4-5 Influence of CRISPR-Cas9 *FASN* knockout process on the overall cell metabolism. (a) PCA results demonstrated that GFP-positive cells have significantly altered their metabolic profiles due to *FASN* knockout (using SCMS results obtained from ESI+ mode) (b) Boxplots indicated that the abundance of most common fatty acids had drastically decreased in GFP-positive cells (using SCMS results obtained from ESI- mode). Fatty acids (C18:0 and C18:1) profiles are statistically significant with p-value<0.05.

PCA reveals global metabolomic profiles between GFP-positive (transfected) cells from the three plasmids (sgRNA #1, #2, and #3) and controls (Figure 4a). The metabolomic profiles of GFP-positive cells are significantly different from those of control cells. Knocking out the *FASN* gene led to substantial alterations in metabolomic profiles, particularly evident in the decreased abundances of fatty acids in GFP-positive cells (Figure 4b).[159] Long-chain fatty acids (LCFAs) have chain-lengths of

C11-20, of which C16 and C18 LCFAs are the most abundant fatty acids (FAs) in cells. The FAs longer than C20 (C>20) are classified as very long chain FAs and are less abundant than LCFAs.[160] The *FASN* knockout aims to disrupt the gene responsible for fatty acid synthesis, potentially affecting the production of these fatty acids and altering cellular metabolism.[161] Notably, *FASN* knockout downregulated the relative abundance of long chain fatty acids (LCFAs) in cells, specifically most abundant fatty acids (saturated FA, C18:0) and (unsaturated FA, C18:1) ($p < 0.05$). This downregulation highlights *FASN* 's crucial role in cellular lipid metabolism. The observed metabolic shifts underscore the impact of *FASN* knockout on the overall cellular metabolic profile, offering valuable insights into the intricate relationship between *FASN* gene expression and cellular metabolism.

4.3.3. Subpopulation analysis of GFP-positive cells

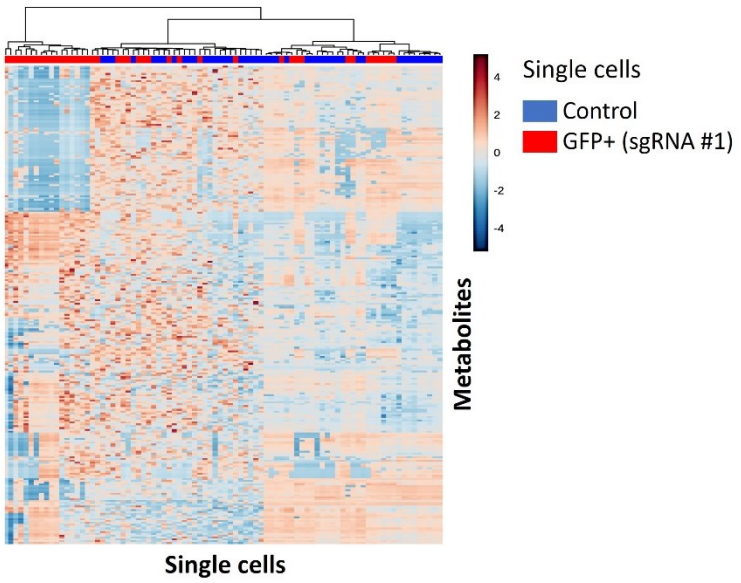


Figure 4-6 Subpopulation analysis was performed representatively for a pairwise comparison of GFP-positive cells (expressing sgRNA #1 plasmid) and controls. The analysis was developed based upon the hierarchical clustering analysis.

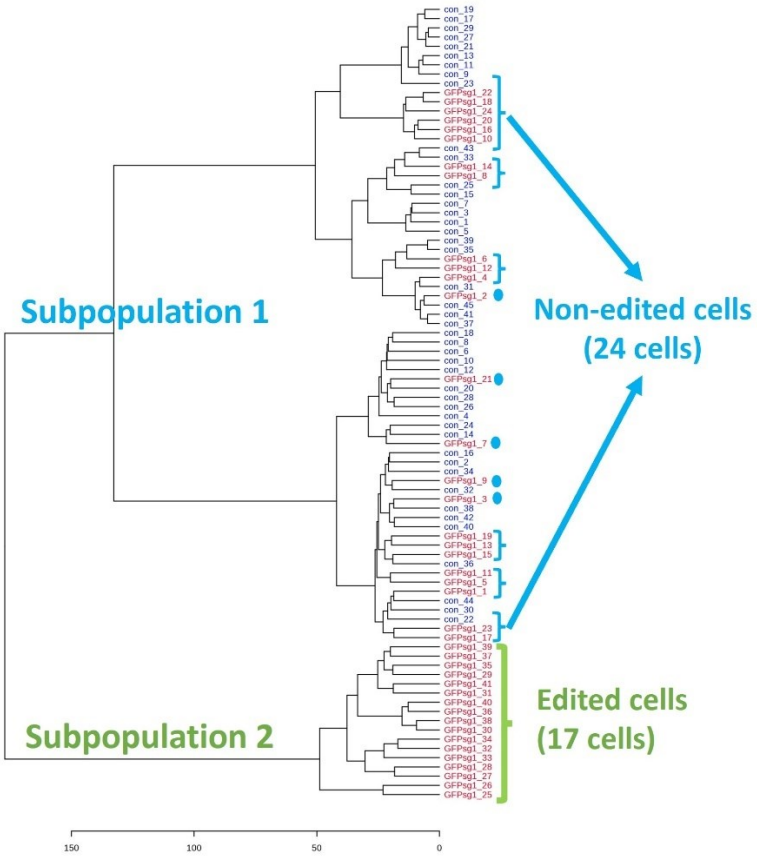


Figure 4-7 Subpopulation analysis identities were revealed with dendrogram classification to distinguish the metabolomic profiles of non-edited cells from the edited cells.

Utilizing the hierarchical clustering analysis (Figure 5a) and dendrogram classification (Figure 5b) developed in previous project,[119] subpopulation analysis was conducted revealing two major clusters in Figure 5. Subpopulation analysis determined based upon the minimum number of clusters.[119] The hierarchical heatmap in Figure 5a compares the overall metabolomic profiles of GFP-positive cells (expressing sgRNA #1 plasmid) with their controls. Interestingly, some of GFP-positive cells clustered with controls, suggesting that they share similar metabolites as the control, highlighting that all GFP-positive cells are not *FASN* knockouts. This indicates that even though the

plasmid is taken up by the GFP-positive cells, the target FASN gene was not edited, likely due to multiple factors such as faulty expression of sgRNA, Cas9, and issues in accessing the *FASN* genomic region.

To further distinguish edited (*FASN* -knockout cells) from non-edited cells among GFP-positive cells, dendrogram classification (**Figure 4-7**) was performed. Subpopulation 1 showed a mix of GFP-positive cells and controls, indicating unsuccessful *FASN* knockout and generating non-edited cells, leading to similar metabolic profiles. In contrast, subpopulation 2 exclusively consisted of GFP-positive cells, which were distinctly different from the controls, confirming successful *FASN* knockout in this subpopulation. As a result, 17 cells from subpopulation 2 were identified as successfully edited, constituting 41% of the total cells using sgRNA #1 plasmid. This percentage offers valuable insights for assessing the *FASN* knockout efficiency using the three distinct plasmids (sgRNA #1, #2, and #3), indicating which plasmid yields the most favorable editing results.

4.3.4. Evaluating gene editing efficiency using SCMS subpopulation analysis and Western blotting analysis

Table 2-2. Gene editing efficiency determined from SCMS experiments in both ESI+ and ESI- ion mode.

SCMS results	sgRNA #1	sgRNA #2	sgRNA #3
ESI+	41%	97%	68%
ESI-	53%	44%	49%

Table 3-3. Gene editing efficiency determined from *FASN* protein expression level.*

Replicates	sgRNA #1	sgRNA #2	sgRNA #3
1	52%	35%	42%
2	72%	47%	36%
3	44%	38%	53%
4	59%	37%	39%
Average	57%	39%	43%
Standard Deviation	11.89%	5.31%	7.42%

*Western blotting experiments were conducted using of GFP-positive (transfected) cells by sgRNA #1, #2, and #3.

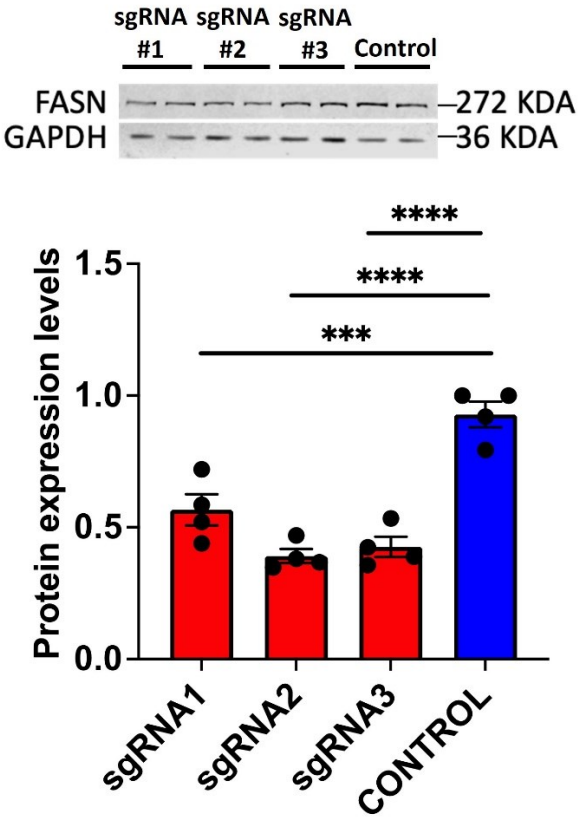


Figure 4-8 Western blot analysis of *FASN* protein expression levels in GFP-positive cells. Cells transfected by sgRNA #1, #2 and #3 show significantly lower FANS compared to control cells (***: $p < 0.05$; ****: $p < 0.01$).

Using SCMS results obtained from both ESI+ and ESI- modes, the percentage of successfully edited cells using each sgRNA plasmid was estimated (**Table 4-2**). A

significant difference in the percentage of successfully edited cells have been observed between ESI+ and ESI- experiments because ESI+ results were from a different batch of samples. ESI- experiments were conducted at a different time. Besides, ESI+ experiments can detect a variety of lipids, such as glycerophospholipids and other membrane lipids (e.g., diacylglyceride and triacylglycerides) in cells. In our studies, ESI+ results show a comprehensive metabolite change which does not only explain the specific focus on *FASN* knockout's impact on fatty acid abundance, but it also includes the change of other small molecules in cells. *FASN* knockout changed fatty acid levels, which can be demonstrated by ESI- ion mode results, allowing us to estimate the editing efficiency. In **Table 4-3**, the *FASN* protein expression levels of GFP-positive cells were reported as 57, 39, and 43% for sgRNA #1, #2, and #3, respectively. By comparing the editing efficiency across Single-probe SCMS results and Western blotting analysis, the ESI- results (53% for sgRNA #1, 44% for sgRNA #2, and 49% for sgRNA #3) share a similar trend with the average percentage estimated from Western blotting results. In Western blot analysis, pairwise comparison of each individual sgRNA plasmid with controls are shown to have their *FASN* protein expression level to be statistically significant ($p < 0.05$) (**Figure 4-8**). As a result, subpopulation analysis was conducted based upon Single-probe SCMS data which successfully estimates the efficiency of gene editing. Western blot analysis is capable of cross validating the efficiency of *FASN* knockout.

4.4 Conclusions

In summary, our study delves into the intricate connection between gene editing and cellular metabolism after *FASN* knockout in HEK293T cells. We found that transfection efficiency is influenced by experimental conditions, especially the choice of sgRNA plasmids. Through single cell mass spectrometry (SCMS) analysis, global metabolic profiles of GFP-positive (transfected) cells following *FASN* knockout have been altered compared to controls, which established a direct link between gene editing and metabolic shifts. SCMS metabolomics analyses distinguished different subpopulations within GFP-positive cells and revealed comprehensive metabolic alterations, including variations in fatty acids, phosphatidylcholines (PCs), diacylglycerides (DGs), and triacylglycerides (TGs). Specifically, the decrease in fatty acid synthesis underscored the impact of *FASN* knockout on cellular lipid metabolism. The proportion of successfully edited cells directly mirrored the efficiency of our gene editing approach. SCMS metabolomics data can be used to estimate the efficiency of gene editing. In addition, Western Blot analysis confirms the effective sgRNA-mediated *FASN* knockout in human cells, and cross validate the *FASN* knockout efficiency with percentage obtained from SCMS results. Besides, genomic DNA sequencing is conducted since it has been widely used as an approval method to obtain indel percentage and accurately determine the editing efficiency, which provides clear evident and validates our novel approach.

4.5 Acknowledgements

The author would like to thank our collaborators, Dr. Rakhi Rajan and Dr. Martin-Paul Agbaga and for their help with the gene editing experiments. The author would like to thank all funding sources, including NIH (R01 EY030513), NSF (MCB-1716423), OCAST (HR20-103), Research Council of the University of Oklahoma Norman Campus, NIH (R01AI177469), NSF (CMI-2305182), and OU Dodge Family College of Art & Science (DFCAS Dissertation Fellowship) (to T.D.N.).

4.5 Author contributions

Conceptualization, Z.Y., R.R., M.-P. A. and **T.D.N.**; methodology, Z.Y., R.R., M.-P. A.; SCMS analysis, **T.D.N.**; CRISPR/Cas9 gene editing, E.A, N.K.C, **T.D.N.**, L.M.; FACS sorting, **T.D.N.**, Z.P.; resources, Z.Y., R.R., M.-P. A., **T.D.N.**; writing—original draft preparation, **T.D.N.**; writing—review and editing, Z.Y., R.R., M.-P. A., **T.D.N.**, E.A., N.K.C.

Chapter 5 Conclusions and Future Directions of Applications of Single-Probe Single Cell Mass Spectrometry for Subpopulation Analysis

5.1 Summary of work

In this dissertation, the applications of Single-probe SCMS have been explored to investigate cell heterogeneity and unveil hidden subpopulations. First, the Single-probe SCMS was employed in conjunction with biosafety standard protocols to examine the heterogeneity in the metabolic response of cells infected with parasites in Chagas' disease. With two subpopulations of cells, namely *T. cruzi*-infected cells and bystander uninfected cells, statistical analysis revealed a concealed subgroup of bystander cells that have been misclassified and shared majority of a metabolomic profiles with *T. cruzi*-infected cells. In addition, Single-probe SCMS demonstrated for the first-time bystander effect of *T. cruzi* on infection-adjacent uninfected cells. This work represents, to the best of our knowledge, the first implementation of single cell metabolomics in mammalian-infectious disease.

Subsequently, Single-probe SCMS metabolomics research was integrated with a bioinformatic tool, SinCHet-MS, to subtract batch effect, identify cell subpopulations, and prioritize metabolite biomarkers. This analysis extended beyond drug-sensitive cells, encompassing observations of changes in the composition of cell subgroups in drug-resistant cells. Using comprehensive statistical analysis in SinCHet-MS, significant changes in subpopulation have been identified, particularly the emergence

of a new subpopulation in drug-sensitive primary melanoma cancer cells (WM115) treated with vemurafenib. While the drug-resistant cell line (WM266-4) did not exhibit clear emergence of new subpopulations, proportional changes between subpopulations were evident. Our technique offers a label-free method distinct from traditional approaches and sheds light on cell subpopulations' dynamics. Additionally, our SCMS metabolomic studies integrated various functions through SinCHet-MS (freely available at http://lab.moffitt.org/chen/software/). This approach enhances our understanding of cellular metabolism not only from a cell heterogeneity perspective but also at the sub populational resolution.

Having successfully developed subpopulation analysis utilizing global metabolomic profiles of cells, Single-probe SCMS was employed to rapidly assess CRISPR-Cas9 gene editing results based on subpopulation analysis. Given the off-target effects of CRISPR-Cas9, subpopulation analysis proved effective in distinguishing edited cells from non-edited cells within a heterogeneous population. Initially, Single-probe SCMS analysis revealed altered global metabolic profiles in GFP-positive (transfected) cells following *FASN* knockout compared to controls, establishing a direct link between gene editing and metabolic shifts. The observed decrease in fatty acid synthesis underscored the impact of *FASN* knockout on cellular lipid metabolism. Then, SCMS results distinguished subpopulations within GFP-positive cells, as not all of GFP-positive cells had successfully undergone *FASN* gene knockout. This discovery of concealed subgroups of cells within the heterogenous GFP-positive cell population provides a rapid method to estimate the efficiency of CRISPR-

Cas9 gene editing. In addition, Western Blot analysis and genomic DNA sequencing validated the effectiveness of sgRNA-mediated *FASN* knockout, cross-validating the efficiency with the percentage obtained from SCMS results. In summary, Single-probe is a powerful analytical sampling and ionization device coupled with mass spectrometry to conduct single-cell metabolomics and lipidomics studies in mammalian-infectious disease, cancers, and genome engineering research. The results obtained from Single-probe SCMS significantly contribute to the discovery of hidden subgroups of cells using unique metabolomic features.

5.2 Future directions

Future research endeavors could focus on utilizing Single-probe SCMS to have a better understanding of the metabolic responses of host cells in mammalian-infectious disease. Investigating a broader range of infectious agents, and host response could provide more insights into pathogenesis. Additionally, exploring real-time dynamics of metabolic changes during infection and treatment could unveil critical insights for targeted therapeutic interventions. The infection time could be increased to observe the progress of *Trypanosoma cruzi (T. cruzi)* infection in cells. Benznidazole is in the frontline drug used against *T. cruzi*, the causative agent of Chagas disease.[162] The influence of Benznidazole will alter cell metabolism of *T. cruzi*-infected cells and bystander uninfected cells, which will contribute the treatment of Chagas disease. Due to the advantage of sampling single cells using Single-probe SCMS, microscopic pictures could be captured prior to SCMS analysis to measure the distance between *T.*

cruzi infected cells and their bystander uninfected cells. It is possible that *T. cruzi* altered the cells metabolism of not only infected cells, but also adjacent bystander cells, especially those in close proximity. For those bystander cells that lie far away from *T. cruzi*-infected cells, their metabolomic profiles are expected to be very different compared to those within closer proximity. Integrating advanced analytical techniques, such as mass spectrometry imaging (MSI) (e.g., DESI-MSI, Single-probe MSI) , may further enhance our ability to visualize and characterize localized metabolic changes within infected cells and tissues.

Future research directions should center on refining and expanding the integration of Single-Probe SCMS with the bioinformatic tool SinChet-MS. Investigating novel applications for SinChet-MS in diverse cellular contexts and cell models can provide valuable insights into cell subpopulations and metabolite biomarkers. Exploring the integration of other bioinformatics tools or machine learning algorithms could further enhance the accuracy and efficiency of subpopulation identification. Additionally, extending the methodology to analyze multi-omic data, such as integrating genomics and proteomics, could offer a more holistic understanding of cellular processes and dysregulations.

Future direction endeavors should concentrate on optimizing and expanding the use of Single-Probe SCMS for the rapid evaluation of CRISPR-Cas9 gene editing efficiency. Investigating the applicability of this approach to different cell types, tissues, and gene targets will broaden its utility. Single cell omics research should be integrated knowing that CRISPR-Cas9 has a capability to knock-out and knock-in

functional genes. Changing a functional gene will result in alteration in cell function, change in cell metabolism, protein expression and genomic expression. Single cell proteomics and genomics are great research approaches that will contribute the overview of how particular functional gene*,* *FASN*, can regulate the fatty acid synthesis in cells. As a result, *FASN* knockout will downregulate fatty acids*.* *FASN* knockout also results in a significant reduction in *FASN* protein expression level. Incorporating real-time monitoring capabilities during the gene editing process could provide valuable insights into the temporal dynamics of metabolic changes associated with CRISPR-Cas9 activity. Furthermore, exploring the potential integration of Single-Probe SCMS with emerging gene editing technologies, such as base editing or prime editing, can contribute to advancing the precision and efficiency of genome engineering strategies.

References

(1) Altschuler, S. J.; Wu, L. F. Cellular heterogeneity: do differences make a difference? *Cell* **2010**, *141* (4), pp 559-563.

(2) D'Avila, H.; Freire-de-Lima, C. G.; Roque, N. R.; Teixeira, L.; Barja-Fidalgo, C.; Silva, A. R.; Melo, R. C.; Dosreis, G. A.; Castro-Faria-Neto, H. C.; Bozza, P. T. Host cell lipid bodies triggered by Trypanosoma cruzi infection and enhanced by the uptake of apoptotic cells are associated with prostaglandin E_2 generation and increased parasite growth. *J Infect Dis* **2011**, *204* (6), pp 951-961.

(3) Nagajyothi, F.; Desruisseaux, M. S.; Weiss, L. M.; Chua, S.; Albanese, C.; Machado, F. S.; Esper, L.; Lisanti, M. P.; Teixeira, M. M.; Scherer, P. E.; et al. Chagas disease, adipose tissue and the metabolic syndrome. *Mem Inst Oswaldo Cruz* **2009**, *104 Suppl 1* (Suppl 1), pp 219-225.

(4) Hoffman, K.; Liu, Z.; Hossain, E.; Bottazzi, M. E.; Hotez, P. J.; Jones, K. M.; McCall, L. I. Alterations to the Cardiac Metabolome Induced by Chronic T. cruzi Infection Relate to the Degree of Cardiac Pathology. *ACS Infect Dis* **2021**, *7* (6), pp 1638-1649.

(5) McCall, L. I.; Morton, J. T.; Bernatchez, J. A.; de Siqueira-Neto, J. L.; Knight, R.; Dorrestein, P. C.; McKerrow, J. H. Mass Spectrometry-Based Chemical Cartography of a Cardiac Parasitic Infection. *Anal Chem* **2017**, *89* (19), pp 10414-10421.

(6) Dean, D. A.; Gautham, G.; Siqueira-Neto, J. L.; McKerrow, J. H.; Dorrestein, P. C.; McCall, L.-I. Spatial metabolomics identifies localized chemical changes in heart tissue during chronic cardiac Chagas Disease. *PLOS Neglected Tropical Diseases* **2021**, *15* (10), p e0009819.

(7) Hossain, E.; Khanam, S.; Dean, D. A.; Wu, C.; Lostracco-Johnson, S.; Thomas, D.; Kane, S. S.; Parab, A. R.; Flores, K.; Katemauswa, M.; et al. Mapping of host-parasite-microbiome interactions reveals metabolic determinants of tropism and tolerance in Chagas disease. *Sci Adv* **2020**, *6* (30), p eaaz2015.

(8) Liu, Z.; Ulrich vonBargen, R.; McCall, L. I. Central role of metabolism in Trypanosoma cruzi tropism and Chagas disease pathogenesis. *Curr Opin Microbiol* **2021**, *63*, pp 204-209.

(9) Lawson, D. A.; Kessenbrock, K.; Davis, R. T.; Pervolarakis, N.; Werb, Z. Tumour heterogeneity and metastasis at single-cell resolution. *Nature Cell Biology* **2018**, *20* (12), pp 1349-1360.

(10) Dagogo-Jack, I.; Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology* **2018**, *15* (2), pp 81-94.

(11) Qian, C.-N.; Mei, Y.; Zhang, J. Cancer metastasis: issues and challenges. *Chin J Cancer* **2017**, *36* (1), pp 38-38.

(12) van Zijl, F.; Krupitza, G.; Mikulits, W. Initial steps of metastasis: cell invasion and endothelial transmigration. *Mutat Res* **2011**, *728* (1-2), pp 23-34.

(13) Liu, Q.; Martínez-Jarquín, S.; Zenobi, R. Recent Advances in Single-Cell Metabolomics Based on Mass Spectrometry. *CCS Chemistry* **2023**, *5* (2), pp 310-324.

(14) Amantonico, A.; Urban, P. L.; Fagerer, S. R.; Balabin, R. M.; Zenobi, R. Single-Cell MALDI-MS as an Analytical Tool for Studying Intrapopulation Metabolic Heterogeneity of Unicellular Organisms. *Analytical Chemistry* **2010**, *82* (17), pp 7394-7400.

(15) Bensen, R. C.; Standke, S. J.; Colby, D. H.; Kothapalli, N. R.; Le-McClain, A. T.; Patten, M. A.; Tripathi, A.; Heinlen, J. E.; Yang, Z.; Burgett, A. W. G. Single Cell Mass Spectrometry Quantification of Anticancer Drugs: Proof of Concept in Cancer Patients. *ACS Pharmacology & Translational Science* **2021**, *4* (1), pp 96-100.

(16) Do, T. D.; Comi, T. J.; Dunham, S. J. B.; Rubakhin, S. S.; Sweedler, J. V. Single Cell Profiling Using Ionic Liquid Matrix-Enhanced Secondary Ion Mass Spectrometry for Neuronal Cell Type Differentiation. *Analytical Chemistry* **2017**, *89* (5), pp 3078-3086.
(17) Xie, Y. R.; Castro, D. C.; Bell, S. E.; Rubakhin, S. S.; Sweedler, J. V. Single-Cell Classification Using Mass Spectrometry through Interpretable Machine Learning. *Anal Chem* **2020**, *92* (13), pp 9338-9347.
(18) Xie, Y. R.; Chari, V. K.; Castro, D. C.; Grant, R.; Rubakhin, S. S.; Sweedler, J. V. Data-Driven and Machine Learning-Based Framework for Image-Guided Single-Cell Mass Spectrometry. *J Proteome Res* **2023**, *22* (2), pp 491-500.
(19) Masujima, T. Live Single-cell Mass Spectrometry. *Analytical Sciences* **2009**, *25* (8), pp 953-960.
(20) Gong, X.; Zhao, Y.; Cai, S.; Fu, S.; Yang, C.; Zhang, S.; Zhang, X. Single cell analysis with probe ESI-mass spectrometry: detection of metabolites at cellular and subcellular levels. *Anal Chem* **2014**, *86* (8), pp 3809-3816.
(21) Shrestha, B.; Vertes, A. In Situ Metabolic Profiling of Single Cells by Laser Ablation Electrospray Ionization Mass Spectrometry. *Analytical Chemistry* **2009**, *81* (20), pp 8265-8271.
(22) Bergman, H.-M.; Lanekoff, I. Profiling and quantifying endogenous molecules in single cells using nano-DESI MS. *Analyst* **2017**, *142* (19), pp 3639-3647.
(23) Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R.; Burgett, A. W. G.; Yang, Z. The Single-Probe: A Miniaturized Multifunctional Device for Single Cell Mass Spectrometry Analysis. *Analytical Chemistry* **2014**, *86* (19), pp 9376-9380.
(24) Zhu, Y.; Piehowski, P. D.; Zhao, R.; Chen, J.; Shen, Y.; Moore, R. J.; Shukla, A. K.; Petyuk, V. A.; Campbell-Thompson, M.; Mathews, C. E.; et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. *Nat Commun* **2018**, *9* (1), p 882.
(25) Liu, R.; Zhang, G.; Yang, Z. Towards rapid prediction of drug-resistant cancer cell phenotypes: single cell mass spectrometry combined with machine learning. *Chemical Communications* **2019**, *55* (5), pp 616-619.
(26) Portero, E. P.; Nemes, P. Dual cationic-anionic profiling of metabolites in a single identified cell in a live Xenopus laevis embryo by microprobe CE-ESI-MS. *Analyst* **2019**, *144* (3), pp 892-900.
(27) Sun, M.; Yang, Z.; Wawrik, B. Metabolomic Fingerprints of Individual Algal Cells Using the Single-Probe Mass Spectrometry Technique. *Frontiers in Plant Science* **2018**, *9*.
(28) 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. *Analytical Chemistry* **2016**, *88* (13), pp 6812-6819.
(29) Sun, M.; Yang, Z. Metabolomic Studies of Live Single Cancer Stem Cells Using Mass Spectrometry. *Analytical Chemistry* **2019**, *91* (3), pp 2384-2391.
(30) Liu, R.; Sun, M.; Zhang, G.; Lan, Y.; Yang, Z. Towards early monitoring of chemotherapy-induced drug resistance based on single cell metabolomics: Combining single-probe mass spectrometry with machine learning. *Analytica chimica acta* **2019**, *1092*, pp 42-48.
(31) Liu, R.; Zhang, G.; Yang, Z. Towards rapid prediction of drug-resistant cancer cell phenotypes: single cell mass spectrometry combined with machine learning. *Chem Commun (Camb)* **2019**, *55* (5), pp 616-619.
(32) Liu, R. M.; Sun, M.; Zhang, G. W.; Lan, Y. P.; Yang, Z. B. Towards early monitoring of chemotherapy-induced drug resistance based on single cell metabolomics: Combining single-probe mass spectrometry with machine learning. *Analytica Chimica Acta* **2019**, *1092*, pp 42-48.

(33) Pan, N.; Standke, S. J.; Kothapalli, N. R.; Sun, M.; Bensen, R. C.; Burgett, A. W. G.; Yang, Z. B. Quantification of Drug Molecules in Live Single Cells Using the Single-Probe Mass Spectrometry Technique. *Analytical Chemistry* **2019**, *91* (14), pp 9018-9024.
(34) Roberts, B. L.; Severance, Z. C.; Bensen, R. C.; Lei, A. T.; Kothapalli, N. R.; Nunez, J. I.; Ma, H. Y.; Wu, S.; Standke, S. J.; Yang, Z. B.; et al. Transient Compound Treatment Induces a Multigenerational Reduction of Oxysterol-Binding Protein (OSBP) Levels and Prophylactic Antiviral Activity. *Acs Chem Biol* **2019**, *14* (2), pp 276-287.
(35) Liu, M.; Zhang, Y.; Yang, J.; Cui, X.; Zhou, Z.; Zhan, H.; Ding, K.; Tian, X.; Yang, Z.; Fung, K. A.; et al. ZIP4 Increases Expression of Transcription Factor ZEB1 to Promote Integrin α3β1 Signaling and Inhibit Expression of the Gemcitabine Transporter ENT1 in Pancreatic Cancer Cells. *Gastroenterology* **2020**, *158* (3), pp 679-692 e671.
(36) Bensen, R. C.; Standke, S. J.; Colby, D. H.; Kothapalli, N. R.; Le-McClain, A. T.; Patten, M. A.; Tripathi, A.; Heinlen, J. E.; Yang, Z. B.; Burgett, A. W. G. Single Cell Mass Spectrometry Quantification of Anticancer Drugs: Proof of Concept in Cancer Patients. *Acs Pharmacol Transl* **2021**, *4* (1), pp 96-100.
(37) Sun, M.; Yang, Z. Metabolomic Studies of Live Single Cancer Stem Cells Using Mass Spectrometry. *Anal Chem* **2019**, *91* (3), pp 2384-2391.
(38) Sun, M.; Yang, Z. B.; Wawrik, B. Metabolomic Fingerprints of Individual Algal Cells Using the Single-Probe Mass Spectrometry Technique. *Frontiers in Plant Science* **2018**, *9*, p 10.
(39) Rao, W.; Pan, N.; Yang, Z. High Resolution Tissue Imaging Using the Single-probe Mass Spectrometry under Ambient Conditions. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (6), pp 986-993.
(40) 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), pp 124-134.
(41) Rao, W.; Pan, N.; Yang, Z. Applications of the Single-probe: Mass Spectrometry Imaging and Single Cell Analysis under Ambient Conditions. *Journal of visualized experiments : JoVE* **2016**, (112), p e53911.
(42) Tian, X.; Xie, B. E.; Zou, Z.; Jiao, Y.; Lin, L. E.; Chen, C. L.; Hsu, C. C.; Peng, J. M.; Yang, Z. B. Multimodal Imaging of Amyloid Plaques: Fusion of the Single-Probe Mass Spectrometry Image and Fluorescence Microscopy Image. *Anal. Chem.* **2019**, *91* (20), pp 12882-12889.
(43) Tian, X.; Zhang, G.; Zou, Z.; Yang, Z. Anticancer Drug Affects Metabolomic Profiles in Multicellular Spheroids: Studies Using Mass Spectrometry Imaging Combined with Machine Learning. *Anal Chem* **2019**, *91* (9), pp 5802-5809.
(44) Sun, M.; Tian, X.; Yang, Z. Microscale Mass Spectrometry Analysis of Extracellular Metabolites in Live Multicellular Tumor Spheroids. *Anal Chem* **2017**, *89* (17), pp 9069-9076.
(45) Xin, X.; Wang, H.; Han, L.; Wang, M.; Fang, H.; Hao, Y.; Li, J.; Zhang, H.; Zheng, C.; Shen, C. Single-Cell Analysis of the Impact of Host Cell Heterogeneity on Infection with Foot-and-Mouth Disease Virus. *J Virol* **2018**, *92* (9).
(46) Raj, A.; Rifkin, S. A.; Andersen, E.; van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **2010**, *463* (7283), pp 913-918.
(47) Kamies, R.; Martinez-Jimenez, C. P. Advances of single-cell genomics and epigenomics in human disease: where are we now? *Mammalian Genome* **2020**, *31* (5), pp 170-180.
(48) Fernandes, M. C.; Andrews, N. W. Host cell invasion by Trypanosoma cruzi: a unique strategy that promotes persistence. *FEMS Microbiol Rev* **2012**, *36* (3), pp 734-747.

(49) Nagajyothi, F.; Desruisseaux, M. S.; Weiss, L. M.; Chua, S.; Albanese, C.; Machado, F. S.; Esper, L.; Lisanti, M. P.; Teixeira, M. M.; Scherer, P. E.; et al. Chagas disease, adipose tissue and the metabolic syndrome. *Memórias do Instituto Oswaldo Cruz* **2009**, *104*, pp 219-225.
(50) Dumoulin, P. C.; Burleigh, B. A. Metabolic flexibility in Trypanosoma cruzi amastigotes: implications for persistence and drug sensitivity. *Curr Opin Microbiol* **2021**, *63*, pp 244-249.
(51) Khan, A. A.; Langston, H. C.; Costa, F. C.; Olmo, F.; Taylor, M. C.; McCann, C. J.; Kelly, J. M.; Lewis, M. D. Local association of Trypanosoma cruzi chronic infection foci and enteric neuropathic lesions at the tissue micro-domain scale. *PLOS Pathogens* **2021**, *17* (8), p e1009864.
(52) Ward, A. I.; Lewis, M. D.; Khan, A. A.; McCann, C. J.; Francisco, A. F.; Jayawardhana, S.; Taylor, M. C.; Kelly, J. M.; Burleigh, B. In Vivo Analysis of Trypanosoma cruzi Persistence Foci at Single-Cell Resolution. *mBio* **2020**, *11* (4), pp e01242-01220.
(53) Tarleton, R. L. Chagas disease: a role for autoimmunity? *Trends Parasitol* **2003**, *19* (10), pp 447-451.
(54) Marcon, G. E.; de Albuquerque, D. M.; Batista, A. M.; Andrade, P. D.; Almeida, E. A.; Guariento, M. E.; Teixeira, M. A.; Costa, S. C. Trypanosoma cruzi: parasite persistence in tissues in chronic chagasic Brazilian patients. *Mem Inst Oswaldo Cruz* **2011**, *106* (1), pp 85-91.
(55) Ward, A. I.; Lewis, M. D.; Taylor, M. C.; Kelly, J. M. Incomplete recruitment of protective T cells facilitates Trypanosoma cruzi persistence in the mouse colon. *bioRxiv* **2021**, p 2021.2002.2026.433032.
(56) Buckner, F. S.; Verlinde, C. L.; La Flamme, A. C.; Van Voorhis, W. C. Efficient technique for screening drugs for activity against Trypanosoma cruzi using parasites expressing beta-galactosidase. *Antimicrob Agents Chemother* **1996**, *40* (11), pp 2592-2597.
(57) Buckner, F. S.; Verlinde, C. L.; Flamme, A. C. L.; Voorhis, W. C. V. Efficient technique for screening drugs for activity against Trypanosoma cruzi using parasites expressing beta-galactosidase. *Antimicrobial Agents and Chemotherapy* **1996**, *40* (11), pp 2592-2597.
(58) Buckner, F. S.; Wilson, A. J.; Van Voorhis, W. C. Detection of live Trypanosoma cruzi in tissues of infected mice by using histochemical stain for beta-galactosidase. *Infect Immun* **1999**, *67* (1), pp 403-409.
(59) Higuchi Mde, L.; De Brito, T.; Martins Reis, M.; Barbosa, A.; Bellotti, G.; Pereira-Barreto, A. C.; Pileggi, F. Correlation between Trypanosoma cruzi parasitism and myocardial inflammatory infiltrate in human chronic chagasic myocarditis: Light microscopy and immunohistochemical findings. *Cardiovasc Pathol* **1993**, *2* (2), pp 101-106.
(60) Liu, R.; Yang, Z. Single cell metabolomics using mass spectrometry: Techniques and data analysis. *Anal Chim Acta* **2021**, *1143*, pp 124-134.
(61) Silva, R. R. d.; Dorrestein, P. C.; Quinn, R. A. Illuminating the dark matter in metabolomics. *Proceedings of the National Academy of Sciences* **2015**, *112* (41), pp 12549-12550.
(62) Gazos-Lopes, F.; Martin, J. L.; Dumoulin, P. C.; Burleigh, B. A. Host triacylglycerols shape the lipidome of intracellular trypanosomes and modulate their growth. *PLOS Pathogens* **2017**, *13* (12), p e1006800.
(63) Vakil, N. H.; Fujinami, N.; Shah, P. J. Pharmacotherapy for Leishmaniasis in the United States: Focus on Miltefosine. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* **2015**, *35* (5), pp 536-545.
(64) Sales Junior, P. A.; Molina, I.; Fonseca Murta, S. M.; Sánchez-Montalvá, A.; Salvador, F.; Corrêa-Oliveira, R.; Carneiro, C. M. Experimental and Clinical Treatment of Chagas Disease: A Review. *Am*

J Trop Med Hyg **2017**, *97* (5), pp 1289-1303.
(65) Shah-Simpson, S.; Lentini, G.; Dumoulin, P. C.; Burleigh, B. A. Modulation of host central carbon metabolism and in situ glucose uptake by intracellular Trypanosoma cruzi amastigotes. *PLoS Pathog* **2017**, *13* (11), p e1006747.
(66) Malafaia, G.; Talvani, A. Nutritional Status Driving Infection by Trypanosoma cruzi: Lessons from Experimental Animals. *J Trop Med* **2011**, *2011*, p 981879.
(67) Morillo, C. A.; Marin-Neto, J. A.; Avezum, A.; Sosa-Estani, S.; Rassi, A., Jr.; Rosas, F.; Villena, E.; Quiroz, R.; Bonilla, R.; Britto, C.; et al. Randomized Trial of Benznidazole for Chronic Chagas' Cardiomyopathy. *N Engl J Med* **2015**, *373* (14), pp 1295-1306.
(68) Nguyen, T. D.; Lan, Y.; Kane, S. S.; Haffner, J. J.; Liu, R.; McCall, L.-I.; Yang, Z. Single-Cell Mass Spectrometry Enables Insight into Heterogeneity in Infectious Disease. *Analytical Chemistry* **2022**, *94* (30), pp 10567-10572.
(69) Barh, D.; Azevedo, V. *Single-cell omics. Volume 2 : technological advances and applications : (Applications in biomedicine and agriculture)*; Academic Press, an imprint of Elsevier, 2019.
(70) Dagogo-Jack, I.; Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* **2018**, *15* (2), pp 81-94.
(71) Sun, X.-x.; Yu, Q. Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. *Acta Pharmacol Sin* **2015**, *36* (10), pp 1219-1227.
(72) Stanta, G.; Bonin, S. Overview on Clinical Relevance of Intra-Tumor Heterogeneity. *Front Med (Lausanne)* **2018**, *5*, p 85.
(73) Qian, M.; Wang, D. C.; Chen, H.; Cheng, Y. Detection of single cell heterogeneity in cancer. *Semin Cell Dev Biol* **2017**, *64*, pp 143-149.
(74) Wilmore, J. R.; Jones, D. D.; Allman, D. Protocol for improved resolution of plasma cell subpopulations by flow cytometry. *Eur J Immunol* **2017**, *47* (8), pp 1386-1388.
(75) Slack, M. D.; Martinez, E. D.; Wu, L. F.; Altschuler, S. J. Characterizing heterogeneous cellular responses to perturbations. *Proc Natl Acad Sci U S A* **2008**, *105* (49), pp 19306-19311.
(76) Kashtan, N.; Roggensack, S. E.; Rodrigue, S.; Thompson, J. W.; Biller, S. J.; Coe, A.; Ding, H.; Marttinen, P.; Malmstrom, R. R.; Stocker, R.; et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. *Science* **2014**, *344* (6182), pp 416-420.
(77) Nguyen, Q. H.; Pervolarakis, N.; Blake, K.; Ma, D.; Davis, R. T.; James, N.; Phung, A. T.; Willey, E.; Kumar, R.; Jabart, E.; et al. Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat Commun* **2018**, *9* (1), pp 2028-2028.
(78) Buettner, F.; Natarajan, K. N.; Casale, F. P.; Proserpio, V.; Scialdone, A.; Theis, F. J.; Teichmann, S. A.; Marioni, J. C.; Stegle, O. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* **2015**, *33* (2), pp 155-160.
(79) Kang, C. C.; Lin, J. M.; Xu, Z.; Kumar, S.; Herr, A. E. Single-cell Western blotting after whole-cell imaging to assess cancer chemotherapeutic response. *Anal Chem* **2014**, *86* (20), pp 10429-10436.
(80) 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), pp 8790-8794.
(81) Chen, G.; Ning, B.; Shi, T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Frontiers in Genetics* **2019**, *10*.

(82) Hwang, B.; Lee, J. H.; Bang, D. Author Correction: Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* **2021**, *53* (5), p 1005.
(83) Weisenfeld, N. I.; Kumar, V.; Shah, P.; Church, D. M.; Jaffe, D. B. Corrigendum: Direct determination of diploid genome sequences. *Genome Res* **2018**, *28* (4), p 606.601.
(84) Butler, A.; Hoffman, P.; Smibert, P.; Papalexi, E.; Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **2018**, *36* (5), pp 411-420.
(85) Li, H.; Courtois, E. T.; Sengupta, D.; Tan, Y.; Chen, K. H.; Goh, J. J. L.; Kong, S. L.; Chua, C.; Hon, L. K.; Tan, W. S.; et al. Author Correction: Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nature Genetics* **2018**, *50* (12), pp 1754-1754.
(86) Ji, Z.; Ji, H. TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Research* **2016**, *44* (13), pp e117-e117.
(87) Smalley, I.; Chen, Z.; Phadke, M.; Li, J.; Yu, X.; Wyatt, C.; Evernden, B.; Messina, J. L.; Sarnaik, A.; Sondak, V. K.; et al. Single-Cell Characterization of the Immune Microenvironment of Melanoma Brain and Leptomeningeal Metastases. *Clin Cancer Res* **2021**, *27* (14), pp 4109-4125.
(88) Agüi-Gonzalez, P.; Jähne, S.; Phan, N. T. N. SIMS imaging in neurobiology and cell biology. *Journal of Analytical Atomic Spectrometry* **2019**, *34* (7), pp 1355-1368.
(89) Chen, Y.; Li, G.; Yuan, S.; Pan, Y.; Liu, Y.; Huang, G. Ultrafast Microelectrophoresis: Behind Direct Mass Spectrometry Measurements of Proteins and Metabolites in Living Cell/Cells. *Analytical Chemistry* **2019**, *91* (16), pp 10441-10447.
(90) Liu, R.; Pan, N.; Zhu, Y.; Yang, Z. T-probe: an integrated microscale device for online in situ single cell analysis and metabolic profiling using mass spectrometry. *Analytical chemistry* **2018**, *90* (18), pp 11078-11085.
(91) Nakashima, T.; Wada, H.; Morita, S.; Erra-Balsells, R.; Hiraoka, K.; Nonami, H. Single-Cell Metabolite Profiling of Stalk and Glandular Cells of Intact Trichomes with Internal Electrode Capillary Pressure Probe Electrospray Ionization Mass Spectrometry. *Analytical Chemistry* **2016**, *88* (6), pp 3049-3057.
(92) Passarelli, M. K.; Newman, C. F.; Marshall, P. S.; West, A.; Gilmore, I. S.; Bunch, J.; Alexander, M. R.; Dollery, C. T. Single-Cell Analysis: Visualizing Pharmaceutical and Metabolite Uptake in Cells with Label-Free 3D Mass Spectrometry Imaging. *Anal Chem* **2015**, *87* (13), pp 6696-6702.
(93) Stopka, S. A.; Khattar, R.; Agtuca, B. J.; Anderton, C. R.; Pasa-Tolic, L.; Stacey, G.; Vertes, A. Metabolic Noise and Distinct Subpopulations Observed by Single Cell LAESI Mass Spectrometry of Plant Cells in situ. *Frontiers in Plant Science* **2018**, *9*.
(94) Yin, R.; Prabhakaran, V.; Laskin, J. Quantitative extraction and mass spectrometry analysis at a single-cell level. *Analytical chemistry* **2018**, *90* (13), pp 7937-7945.
(95) Pevsner-Fischer, M.; Levin, S.; Zipori, D. The origins of mesenchymal stromal cell heterogeneity. *Stem Cell Rev Rep* **2011**, *7* (3), pp 560-568.
(96) Auerbach, R.; Plendl, J.; Kusha, B. Endothelial Cell Heterogeneity and Differentiation. In *Angiogenesis in Health and Disease*, Maragoudakis, M. E., Gullino, P., Lelkes, P. I. Eds.; Springer US, 1992; pp 55-62.
(97) Marklein, R. A.; Klinker, M. W.; Drake, K. A.; Polikowsky, H. G.; Lessey-Morillon, E. C.; Bauer, S. R. Morphological profiling using machine learning reveals emergent subpopulations of interferon-γ-stimulated mesenchymal stromal cells that predict immunosuppression. *Cytotherapy* **2019**, *21*

(1), pp 17-31.
(98) Wang, H.; Sridhar, B.; Leinwand, L. A.; Anseth, K. S. Characterization of Cell Subpopulations Expressing Progenitor Cell Markers in Porcine Cardiac Valves. *PLOS ONE* **2013**, *8* (7), p e69667.
(99) Zhang, L.; Sevinsky, C. J.; Davis, B. M.; Vertes, A. Single-Cell Mass Spectrometry of Subpopulations Selected by Fluorescence Microscopy. *Analytical Chemistry* **2018**, *90* (7), pp 4626-4634.
(100) Stopka, S. A.; Khattar, R.; Agtuca, B. J.; Anderton, C. R.; Paša-Tolić, L.; Stacey, G.; Vertes, A. Metabolic Noise and Distinct Subpopulations Observed by Single Cell LAESI Mass Spectrometry of Plant Cells in situ. *Front Plant Sci* **2018**, *9*, p 1646.
(101) Smalley, K. S. M.; Haass, N. K.; Brafford, P. A.; Lioni, M.; Flaherty, K. T.; Herlyn, M. Multiple signaling pathways must be targeted to overcome drug resistance in cell lines derived from melanoma metastases. *Molecular Cancer Therapeutics* **2006**, *5* (5), pp 1136-1144.
(102) Helmbach, H.; Rossmann, E.; Kern, M. A.; Schadendorf, D. Drug-resistance in human melanoma. *Int J Cancer* **2001**, *93* (5), pp 617-622.
(103) Wehrens, R.; Hageman, J. A.; van Eeuwijk, F.; Kooke, R.; Flood, P. J.; Wijnker, E.; Keurentjes, J. J.; Lommen, A.; van Eekelen, H. D.; Hall, R. D.; et al. Improved batch correction in untargeted MS-based metabolomics. *Metabolomics* **2016**, *12*, p 88.
(104) Salerno, S., Jr.; Mehrmohamadi, M.; Liberti, M. V.; Wan, M.; Wells, M. T.; Booth, J. G.; Locasale, J. W. RRmix: A method for simultaneous batch effect correction and analysis of metabolomics data in the absence of internal standards. *PLOS ONE* **2017**, *12* (6), p e0179530.
(105) Liu, R.; Zhang, G.; Sun, M.; Pan, X.; Yang, Z. Integrating a generalized data analysis workflow with the Single-probe mass spectrometry experiment for single cell metabolomics. *Anal Chim Acta* **2019**, *1064*, pp 71-79.
(106) Li, J.; Smalley, I.; Schell, M. J.; Smalley, K. S. M.; Chen, Y. A. SinCHet: a MATLAB toolbox for single cell heterogeneity analysis in cancer. *Bioinformatics* **2017**, *33* (18), pp 2951-2953.
(107) Romano, P.; Profumo, A.; Rocco, M.; Mangerini, R.; Ferri, F.; Facchiano, A. Geena 2, improved automated analysis of MALDI/TOF mass spectra. *BMC Bioinformatics* **2016**, *17* (4), p 61.
(108) Pang, Z.; Chong, J.; Zhou, G.; de Lima Morais, D. A.; Chang, L.; Barrette, M.; Gauthier, C.; Jacques, P.-É.; Li, S.; Xia, J. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research* **2021**, *49* (W1), pp W388-W396.
(109) Johnson, W. E.; Li, C.; Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **2006**, *8* (1), pp 118-127.
(110) Zhang, L. W.; Sevinsky, C. J.; Davis, B. M.; Vertes, A. Single-Cell Mass Spectrometry of Subpopulations Selected by Fluorescence Microscopy. *Anal. Chem.* **2018**, *90* (7), pp 4626-4634.
(111) Beger, R. D. A review of applications of metabolomics in cancer. *Metabolites* **2013**, *3* (3), pp 552-574.
(112) Zhang, Y.; Liu, P.; Li, Y.; Zhang, A.-H. Exploration of metabolite signatures using high-throughput mass spectrometry coupled with multivariate data analysis. *RSC Advances* **2017**, *7* (11), pp 6780-6787.
(113) 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), pp 6545-6550.
(114) 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. *Proceedings of the National Academy of Sciences* **2017**, *114* (10), pp 2586-2591.

(115) Sillence, D. *Lipid rafts : properties, controversies and roles in signal transduction*; 2014.
(116) Rowinsky, E. K. Signal events: Cell signal transduction and its inhibition in cancer. *Oncologist* **2003**, *8 Suppl 3*, pp 5-17.
(117) Jang, C.; Chen, L.; Rabinowitz, J. D. Metabolomics and Isotope Tracing. *Cell* **2018**, *173* (4), pp 822-837.
(118) Li, Q.; Chen, P.; Fan, Y.; Wang, X.; Xu, K.; Li, L.; Tang, B. Multicolor Fluorescence Detection-Based Microfluidic Device for Single-Cell Metabolomics: Simultaneous Quantitation of Multiple Small Molecules in Primary Liver Cells. *Anal Chem* **2016**, *88* (17), pp 8610-8616.
(119) Liu, R.; Li, J.; Lan, Y.; Nguyen, T. D.; Chen, Y. A.; Yang, Z. Quantifying Cell Heterogeneity and Subpopulations Using Single Cell Metabolomics. *Analytical Chemistry* **2023**, *95* (18), pp 7127-7133.
(120) Liu, X.; Hussain, M.; Dai, J.; Li, Y.; Zhang, L.; Yang, J.; Ali, Z.; He, N.; Tang, Y. Programmable Biosensors Based on RNA-Guided CRISPR/Cas Endonuclease. *Biological Procedures Online* **2022**, *24* (1), p 2.
(121) Tyagi, S.; Kumar, R.; Das, A.; Won, S. Y.; Shukla, P. CRISPR-Cas9 system: A genome-editing tool with endless possibilities. *Journal of Biotechnology* **2020**, *319*, pp 36-53.
(122) Hryhorowicz, M.; Lipiński, D.; Zeyland, J.; Słomski, R. CRISPR/Cas9 Immune System as a Tool for Genome Engineering. *Archivum Immunologiae et Therapiae Experimentalis* **2017**, *65* (3), pp 233-240.
(123) Redman, M.; King, A.; Watson, C.; King, D. What is CRISPR/Cas9? *Arch Dis Child Educ Pract Ed* **2016**, *101* (4), pp 213-215.
(124) Eid, A.; Mahfouz, M. M. Genome editing: the road of CRISPR/Cas9 from bench to clinic. *Experimental & Molecular Medicine* **2016**, *48* (10), pp e265-e265.
(125) Bibikova, M.; Carroll, D.; Segal, D. J.; Trautman, J. K.; Smith, J.; Kim, Y.-G.; Chandrasegaran, S. Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Molecular and Cellular Biology* **2001**, *21* (1), pp 289-297.
(126) Gupta, S. K.; Shukla, P. Gene editing for cell engineering: trends and applications. *Critical Reviews in Biotechnology* **2017**, *37* (5), pp 672-684.
(127) Fajrial, A. K.; He, Q. Q.; Wirusanti, N. I.; Slansky, J. E.; Ding, X. A review of emerging physical transfection methods for CRISPR/Cas9-mediated gene editing. *Theranostics* **2020**, *10* (12), pp 5532-5549.
(128) Ceasar, S. A.; Rajan, V.; Prykhozhij, S. V.; Berman, J. N.; Ignacimuthu, S. Insert, remove or replace: A highly advanced genome editing system using CRISPR/Cas9. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2016**, *1863* (9), pp 2333-2344.
(129) Ma, Y.; Zhang, L.; Huang, X. Genome modification by CRISPR/Cas9. *The FEBS Journal* **2014**, *281* (23), pp 5186-5193.
(130) Doudna, J. A.; Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014**, *346* (6213), p 1258096.
(131) Gupta, S.; Kumar, A.; Patel, R.; Kumar, V. Genetically modified crop regulations: scope and opportunity using the CRISPR-Cas9 genome editing approach. *Molecular Biology Reports* **2021**, *48* (5), pp 4851-4863.
(132) Eş, I.; Gavahian, M.; Marti-Quijal, F. J.; Lorenzo, J. M.; Mousavi Khaneghah, A.; Tsatsanis, C.; Kampranis, S. C.; Barba, F. J. The application of the CRISPR-Cas9 genome editing machinery in food and agricultural science: Current status, future perspectives, and associated challenges. *Biotechnology Advances* **2019**, *37* (3), pp 410-421.

(133) Zhang, S.; Wu, S.; Hu, C.; Yang, Q.; Dong, T.; Sheng, O.; Deng, G.; He, W.; Dou, T.; Li, C.; et al. Increased mutation efficiency of CRISPR/Cas9 genome editing in banana by optimized construct. *PeerJ* **2022**, *10*, p e12664.
(134) Manghwar, H.; Li, B.; Ding, X.; Hussain, A.; Lindsey, K.; Zhang, X.; Jin, S. CRISPR/Cas Systems in Genome Editing: Methodologies and Tools for sgRNA Design, Off-Target Evaluation, and Strategies to Mitigate Off-Target Effects. *Advanced Science* **2020**, *7* (6), p 1902312.
(135) Tycko, J.; Myer, Vic E.; Hsu, Patrick D. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Molecular Cell* **2016**, *63* (3), pp 355-370.
(136) Modell, A. E.; Lim, D.; Nguyen, T. M.; Sreekanth, V.; Choudhary, A. CRISPR-based therapeutics: current challenges and future applications. *Trends Pharmacol Sci* **2022**, *43* (2), pp 151-161.
(137) Veres, A.; Gosis, B. S.; Ding, Q.; Collins, R.; Ragavendran, A.; Brand, H.; Erdin, S.; Cowan, C. A.; Talkowski, M. E.; Musunuru, K. Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* **2014**, *15* (1), pp 27-30.
(138) Liu, W.; Li, L.; Jiang, J.; Wu, M.; Lin, P. Applications and challenges of CRISPR-Cas gene-editing to disease treatment in clinics. *Precis Clin Med* **2021**, *4* (3), pp 179-191.
(139) Lim, B.; Lin, Y.; Navin, N. Advancing Cancer Research and Medicine with Single-Cell Genomics. *Cancer Cell* **2020**, *37* (4), pp 456-470.
(140) Puram, S. V.; Tirosh, I.; Parikh, A. S.; Patel, A. P.; Yizhak, K.; Gillespie, S.; Rodman, C.; Luo, C. L.; Mroz, E. A.; Emerick, K. S.; et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **2017**, *171* (7), pp 1611-1624.e1624.
(141) Evers, T. M. J.; Hochane, M.; Tans, S. J.; Heeren, R. M. A.; Semrau, S.; Nemes, P.; Mashaghi, A. Deciphering Metabolic Heterogeneity by Single-Cell Analysis. *Analytical Chemistry* **2019**, *91* (21), pp 13314-13323.
(142) Boggio, K. J.; Obasuyi, E.; Sugino, K.; Nelson, S. B.; Agar, N. Y. R.; Agar, J. N. Recent advances in single-cell MALDI mass spectrometry imaging and potential clinical impact. *Expert Review of Proteomics* **2011**, *8* (5), pp 591-604.
(143) Lanni, E. J.; Rubakhin, S. S.; Sweedler, J. V. Mass spectrometry imaging and profiling of single cells. *J Proteomics* **2012**, *75* (16), pp 5036-5051.
(144) Gong, X.; Zhao, Y.; Cai, S.; Fu, S.; Yang, C.; Zhang, S.; Zhang, X. Single Cell Analysis with Probe ESI-Mass Spectrometry: Detection of Metabolites at Cellular and Subcellular Levels. *Analytical Chemistry* **2014**, *86* (8), pp 3809-3816.
(145) Zhu, Y.; Wang, W.; Yang, Z. Combining Mass Spectrometry with Paternò–Büchi Reaction to Determine Double-Bond Positions in Lipids at the Single-Cell Level. *Analytical Chemistry* **2020**, *92* (16), pp 11380-11387.
(146) Liu, R.; Pan, N.; Zhu, Y.; Yang, Z. T-Probe: An Integrated Microscale Device for Online In Situ Single Cell Analysis and Metabolic Profiling Using Mass Spectrometry. *Anal Chem* **2018**, *90* (18), pp 11078-11085.
(147) Fhu, C. W.; Ali, A. Fatty Acid Synthase: An Emerging Target in Cancer. *Molecules* **2020**, *25* (17), p 3935.
(148) Gonzalez-Salinas, F.; Rojo, R.; Martinez-Amador, C.; Herrera-Gamboa, J.; Trevino, V. Transcriptomic and cellular analyses of CRISPR/Cas9-mediated edition of FASN show inhibition of aggressive characteristics in breast cancer cells. *Biochemical and Biophysical Research Communications* **2020**, *529* (2), pp 321-327.

(149) Bastos, D. C.; Ribeiro, C. F.; Ahearn, T.; Nascimento, J.; Pakula, H.; Clohessy, J.; Mucci, L.; Roberts, T.; Zanata, S. M.; Zadra, G.; et al. Genetic ablation of FASN attenuates the invasive potential of prostate cancer driven by Pten loss. *The Journal of Pathology* **2021**, *253* (3), pp 292-303.
(150) Lu, T.; Sun, L.; Wang, Z.; Zhang, Y.; He, Z.; Xu, C. Fatty acid synthase enhances colorectal cancer cell proliferation and metastasis via regulating AMPK/mTOR pathway. *Onco Targets Ther* **2019**, *12*, pp 3339-3347.
(151) Hendel, A.; Bak, R. O.; Clark, J. T.; Kennedy, A. B.; Ryan, D. E.; Roy, S.; Steinfeld, I.; Lunstad, B. D.; Kaiser, R. J.; Wilkens, A. B.; et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol* **2015**, *33* (9), pp 985-989.
(152) Inoue, H.; Nojima, H.; Okayama, H. High efficiency transformation of Escherichia coli with plasmids. *Gene* **1990**, *96* (1), pp 23-28.
(153) Brown, S. D.; Dreolini, L.; Wilson, J. F.; Balasundaram, M.; Holt, R. A. Complete sequence verification of plasmid DNA using the Oxford Nanopore Technologies' MinION device. *BMC Bioinformatics* **2023**, *24* (1), p 116.
(154) Jones, J.; Nivitchanyong, T.; Giblin, C.; Ciccarone, V.; Judd, D.; Gorfien, S.; Krag, S. S.; Betenbaugh, M. J. Optimization of tetracycline-responsive recombinant protein production and effect on cell growth and ER stress in mammalian cells. *Biotechnol Bioeng* **2005**, *91* (6), pp 722-732.
(155) Renmeng Liu, J. L., Yunpeng Lan, Tra D. Nguyen, Y. Ann Chen, Zhibo Yang. Quantifying Cell Heterogeneity and Subpopulations Using Single Cell Metabolomics. *submitted* **2022**.
(156) Pang, Z.; Chong, J.; Zhou, G.; de Lima Morais, D. A.; Chang, L.; Barrette, M.; Gauthier, C.; Jacques, P.; Li, S.; Xia, J. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res* **2021**, *49* (W1), pp W388-w396.
(157) Quirk, T. J.; Quirk, T. J. One-way analysis of variance (ANOVA). *Excel 2007 for Educational and Psychological Statistics: A Guide to Solving Practical Problems* **2012**, pp 163-179.
(158) Liu, R.; Yang, Z. Single cell metabolomics using mass spectrometry: Techniques and data analysis. *Analytica Chimica Acta* **2021**, *1143*, pp 124-134.
(159) Ameer, F.; Scandiuzzi, L.; Hasnain, S.; Kalbacher, H.; Zaidi, N. De novo lipogenesis in health and disease. *Metabolism* **2014**, *63* (7), pp 895-902.
(160) Sassa, T.; Kihara, A. Metabolism of very long-chain Fatty acids: genes and pathophysiology. *Biomol Ther (Seoul)* **2014**, *22* (2), pp 83-92.
(161) Fhu, C. W.; Ali, A. Fatty Acid Synthase: An Emerging Target in Cancer. *Molecules* **2020**, *25* (17).
(162) Mejia, A. M.; Hall, B. S.; Taylor, M. C.; Gómez-Palacio, A.; Wilkinson, S. R.; Triana-Chávez, O.; Kelly, J. M. Benznidazole-resistance in Trypanosoma cruzi is a readily acquired trait that can arise independently in a single population. *J Infect Dis* **2012**, *206* (2), pp 220-228.

Appendix A

Chapter 2

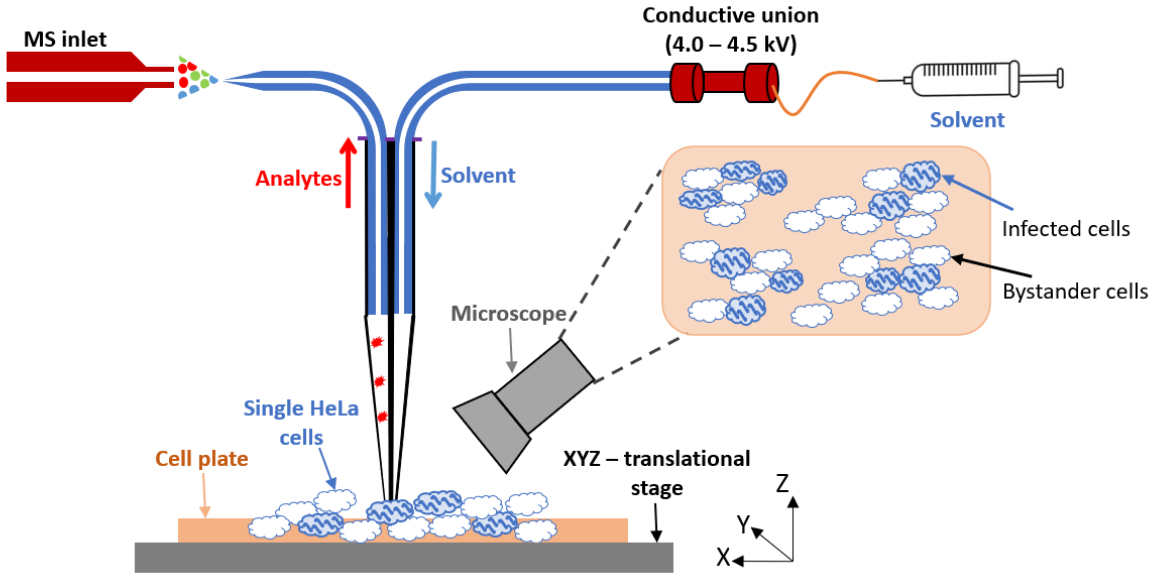


Figure S1. The schematic of the working mechanisms of the Single-probe SCMS experimental set-up.

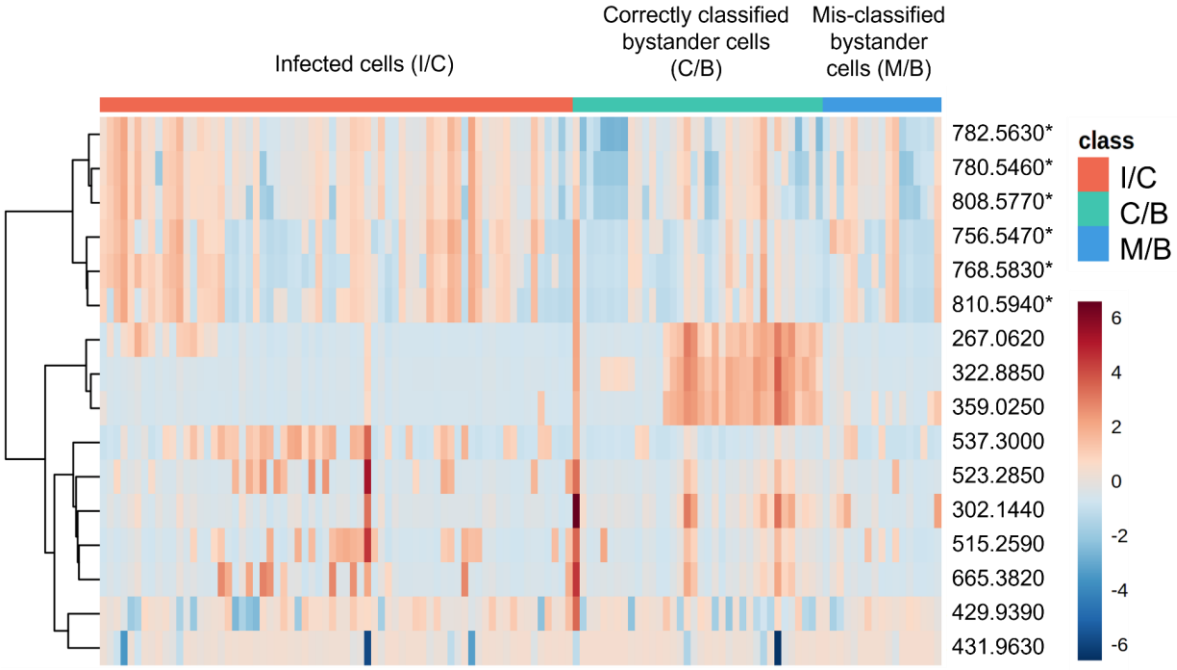


Figure S2: Hierarchical Clustering of metabolite features differing between infected cells (I/C), correctly classified bystander cells (C/B) and mis-classified bystander cells (M/B). The annotated features are marked (*).

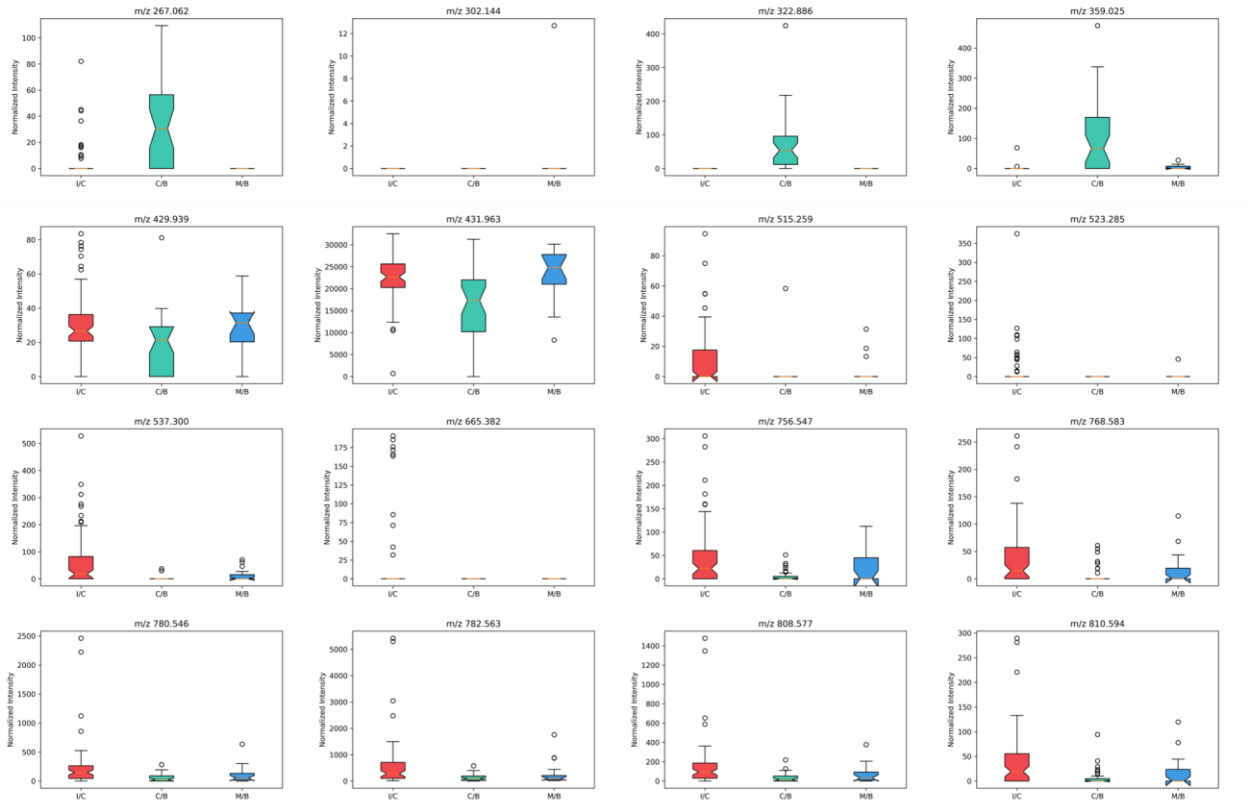
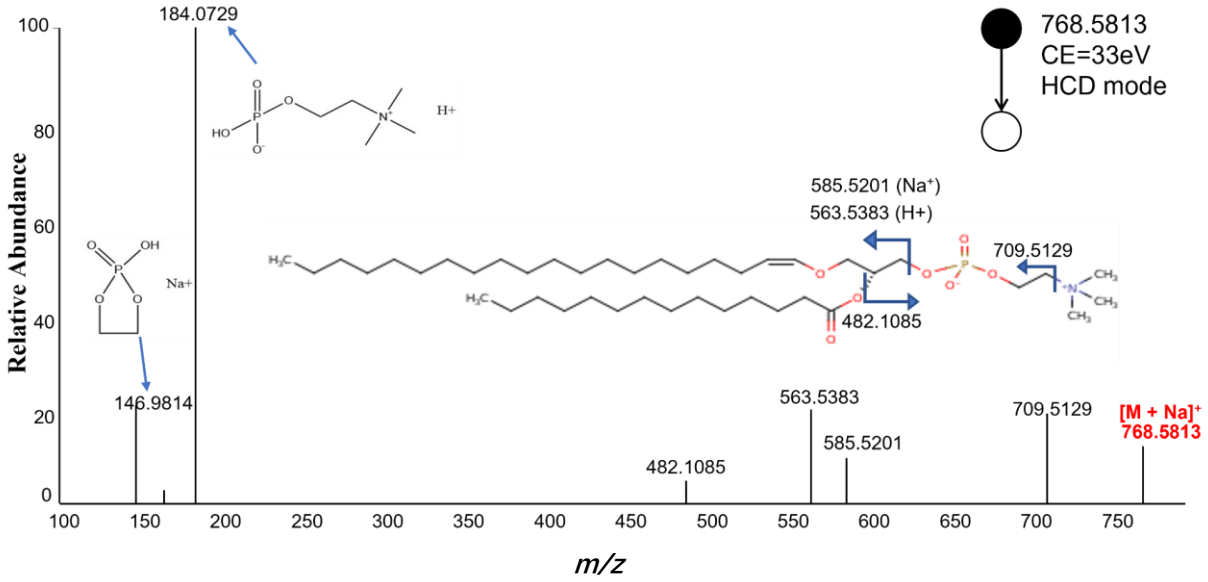


Figure S3. Box plots for 16 metabolites showed comparable behavior in infected cells (I/C) and in mis-classified bystander cells (M/B), as determined by ANOVA test with an adjusted p-value ≤ 0.05.



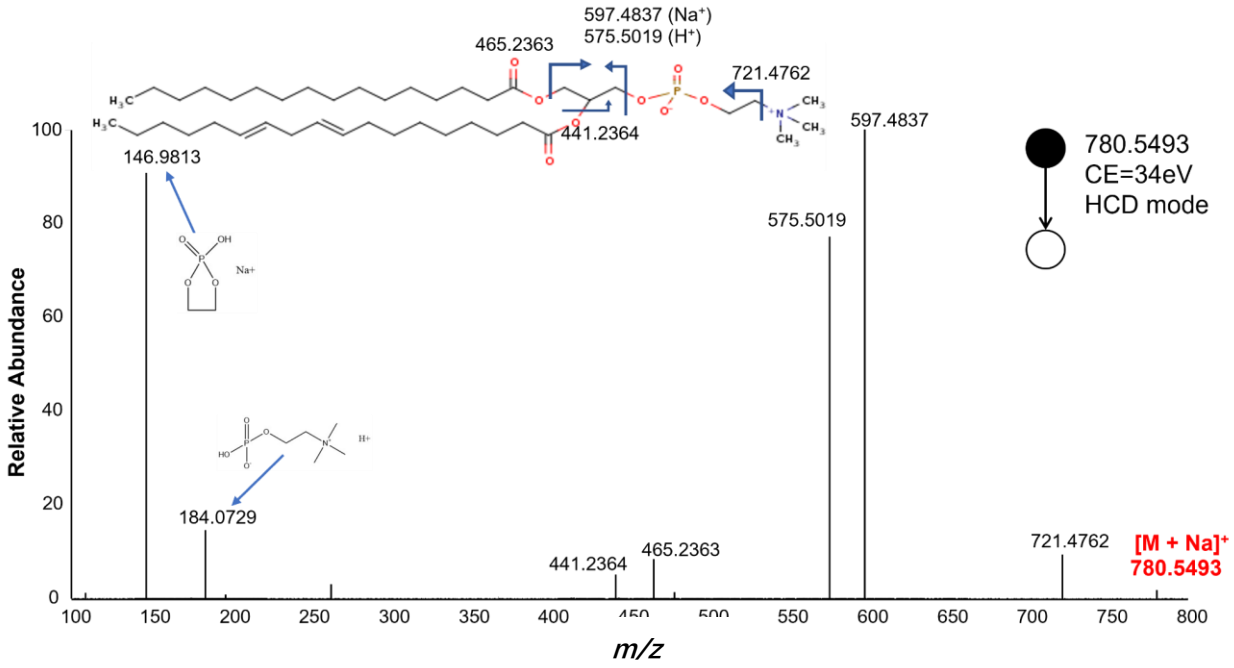
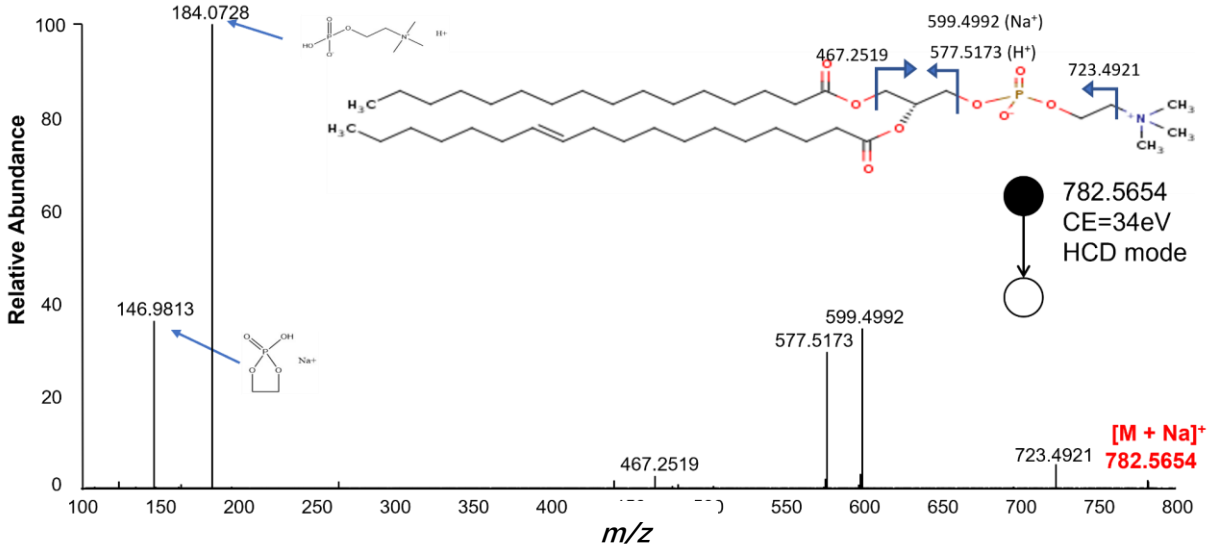


Figure S4. Annotated MS/MS spectra of $[PC(P\text{-}20:0/14:0)+Na]^+$ at *m/z* 768.5813 (upper panel) and $[PC(16:0/18:2)+Na]^+$ at *m/z* 780.5493 (lower panel), acquired from individual HeLa cells. CE: collision energy.



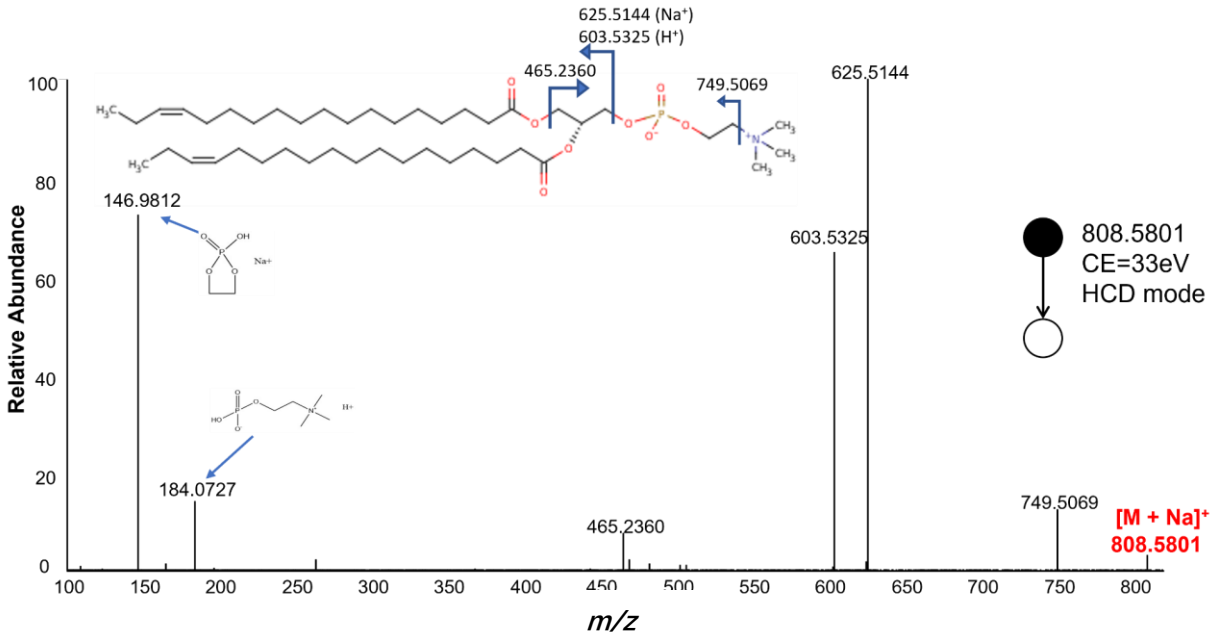


Figure S5. Annotated MS/MS spectra of [PC(16:0/18:1)+Na]$^+$ at *m/z* 782.5654 (upper panel) and [PC(18:1/18:1)+Na]$^+$ at *m/z* 808.5801 (lower panel), acquired from individual HeLa cells. CE: collision energy.

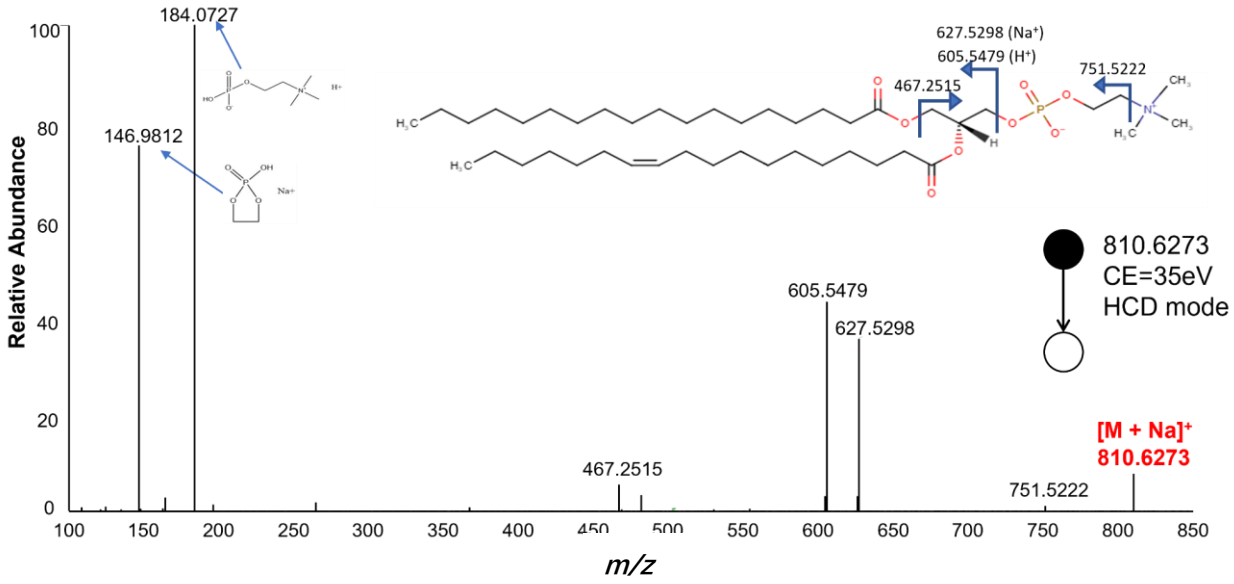


Figure S6. Annotated MS/MS spectrum of [PC(18:0/18:1)+Na]$^+$ at *m/z* 810.6273 from individual HeLa cells. CE: collision energy.

Chapter 3

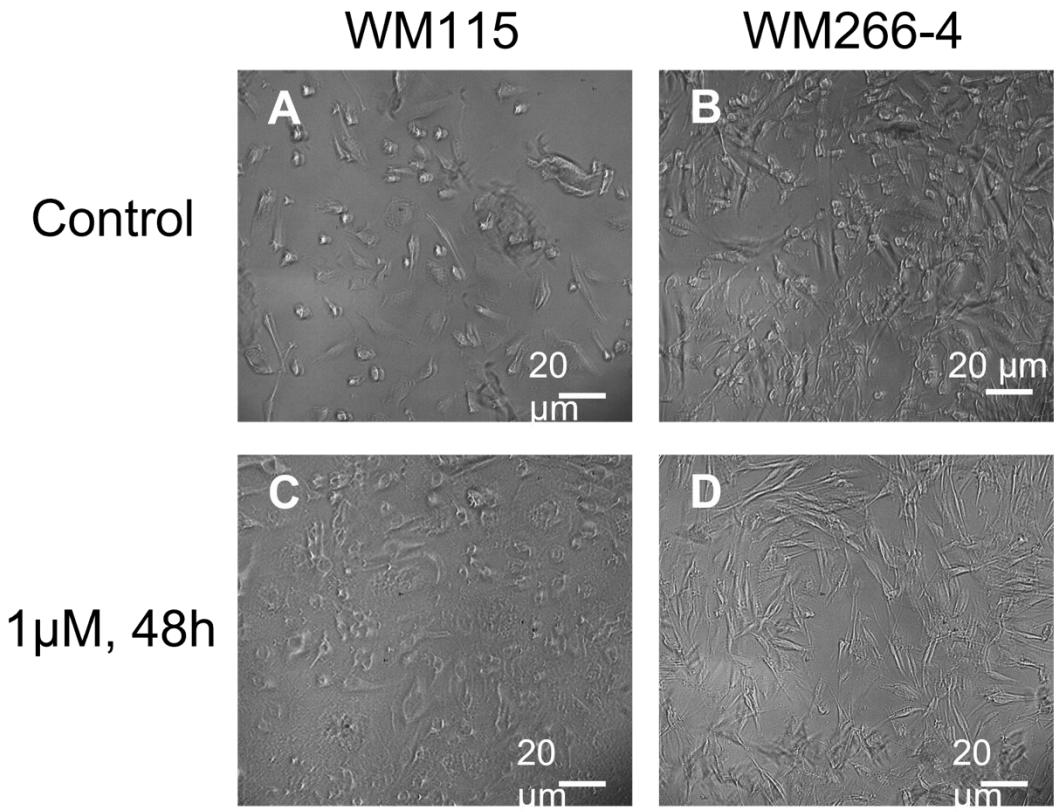
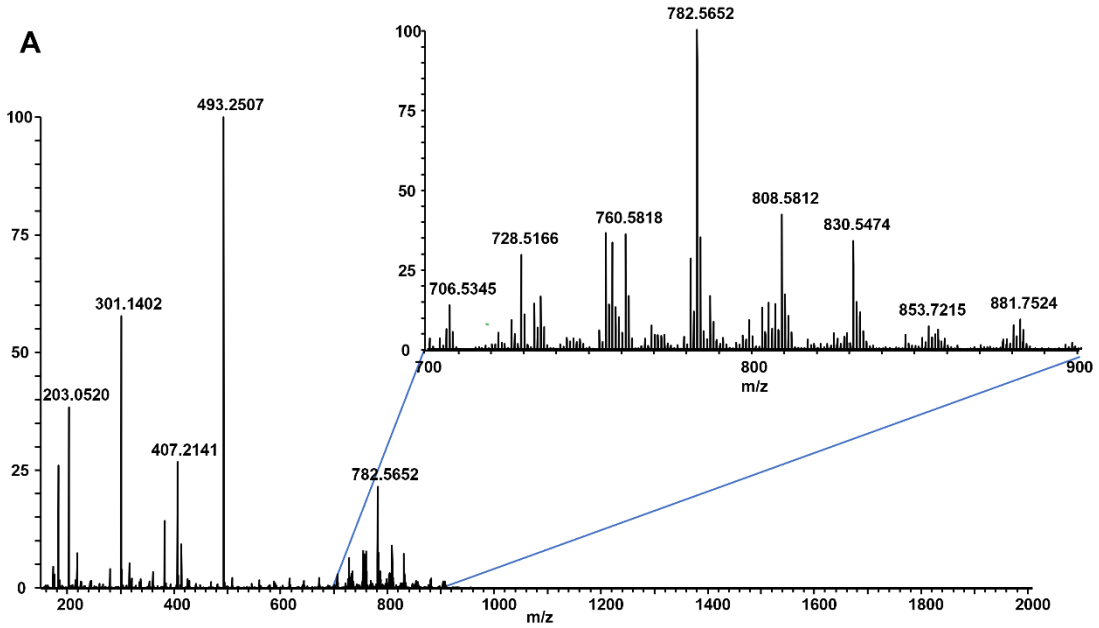
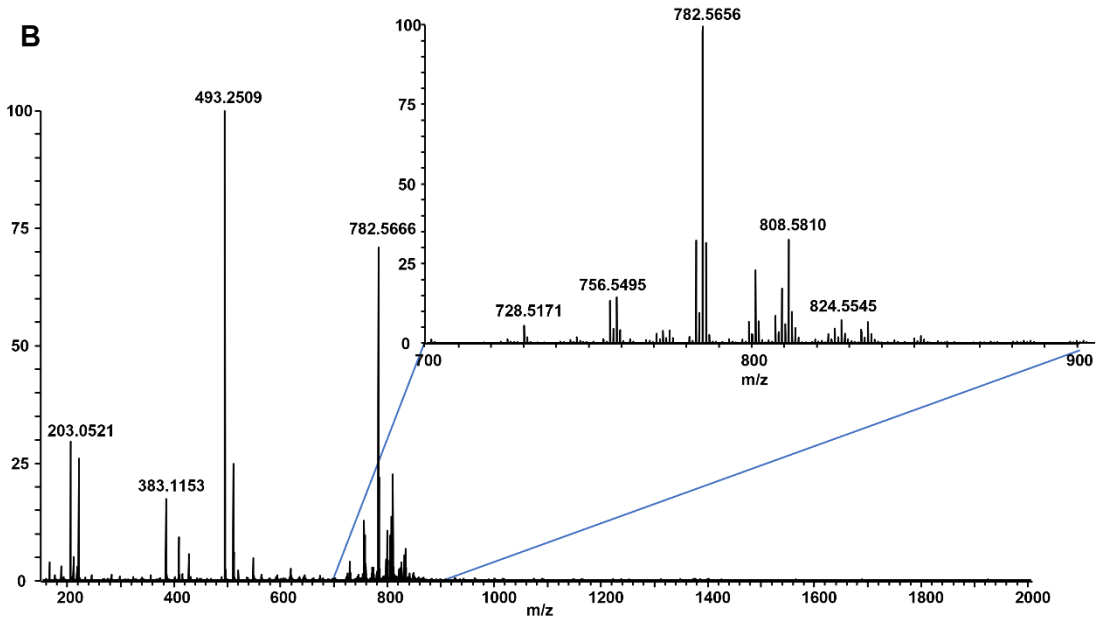


Figure S1. Photos of WM115 (A and C) and WM266-4 (B and D) cells before (A and B) and after treatment (C and D) (1 μM for 48 h).


A
493.2507
301.1402
203.0520
407.2141
782.5652
782.5652
808.5812
760.5818
830.5474
728.5166
706.5345
853.7215
881.7524
m/z


B
493.2509
203.0521
383.1153
782.5666
782.5656
808.5810
756.5495
728.5171
824.5545
m/z

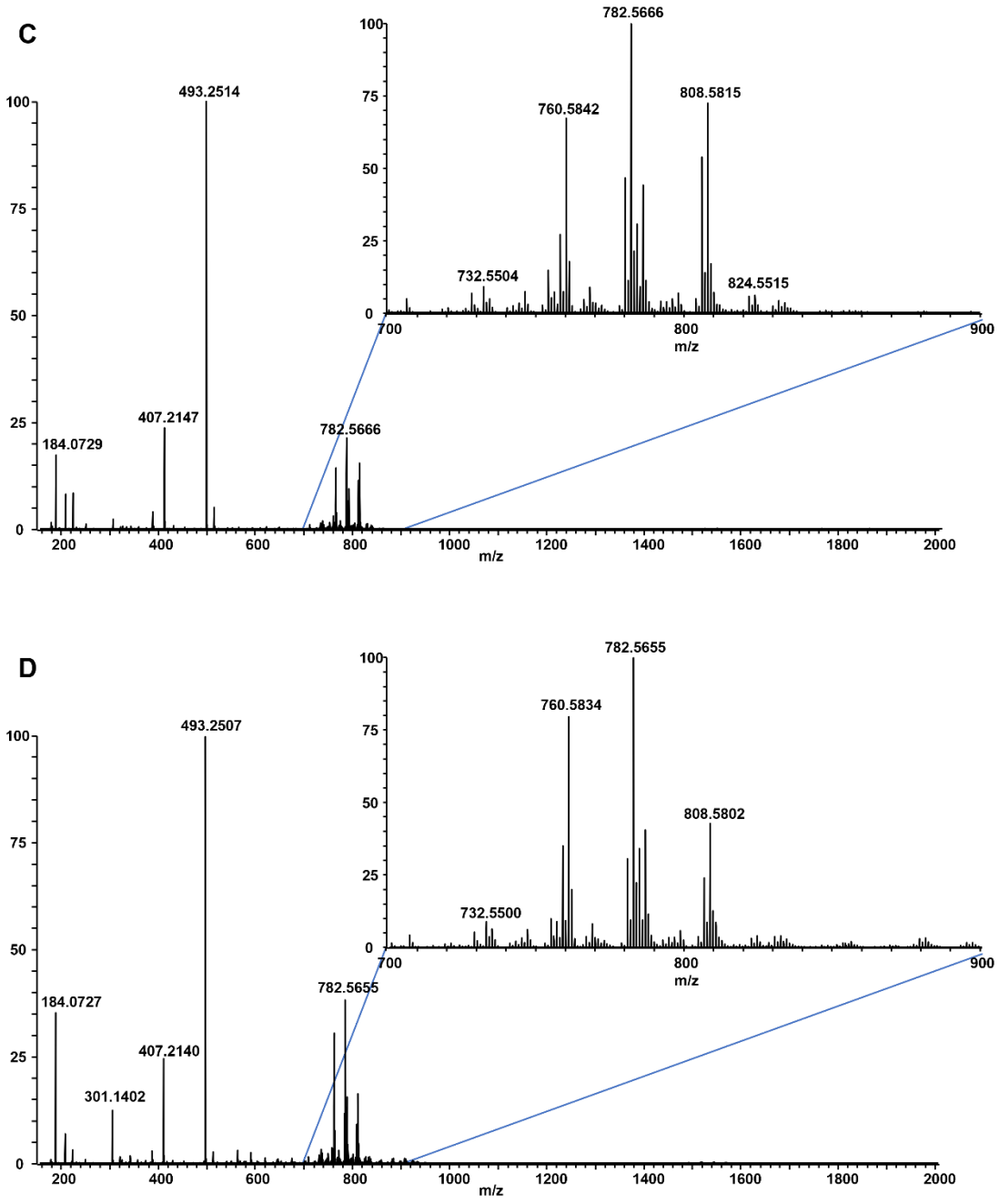


Figure S2. Representative mass spectra of single WM115 (A-B) and WM266-4 (C-D) cells before (A and C) and after treatment (B and D) (1 μM for 48 h). The zoomed-in mass spectra illustrate *m/z* regions with abundant cellular species. Major lipid species including PC 34:1 and PC 36:2 are present in both cell lines before and after treatment with different relative intensity.

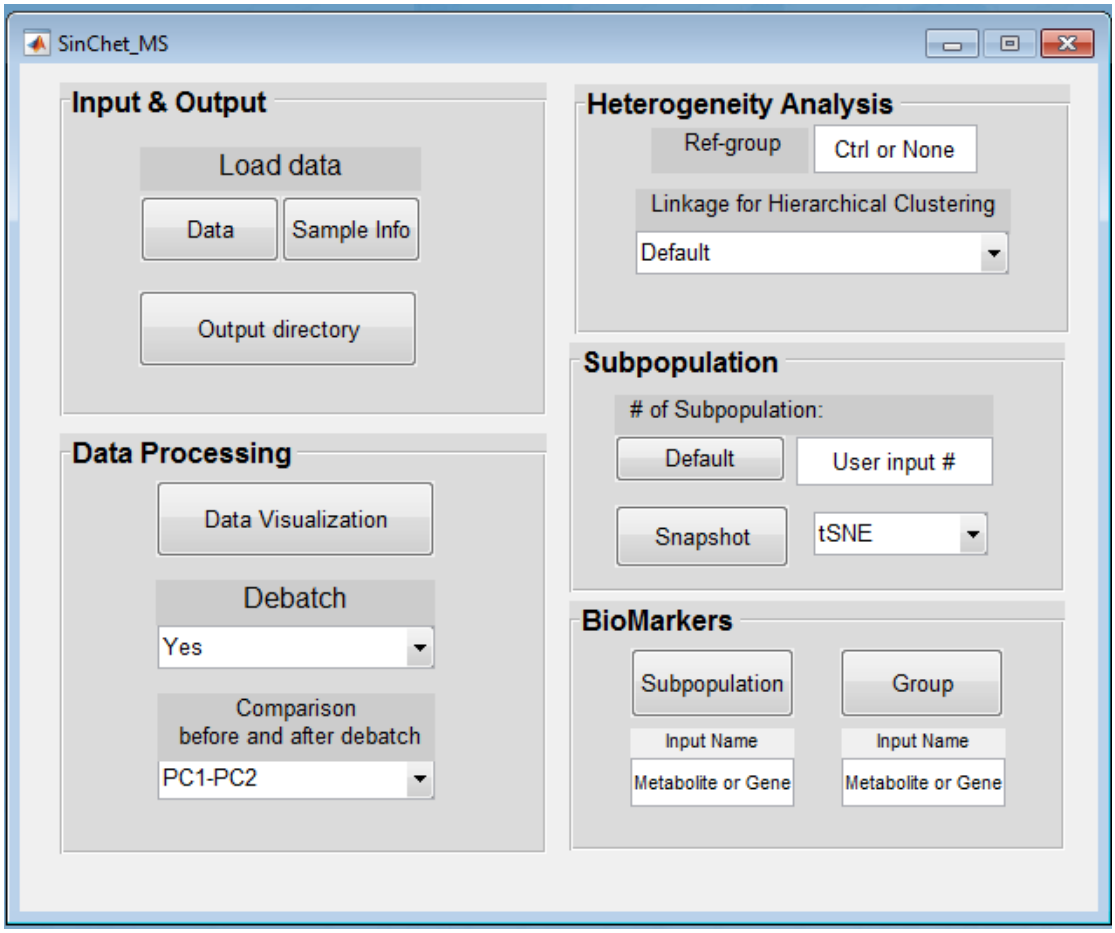


Figure S3. The main Graphic User Interface (GUI) of the SinCHet-MS software package. It integrates functions of debatching, determination and visualization of cell subpopulations and prioritization of subpopulational biomarkers.

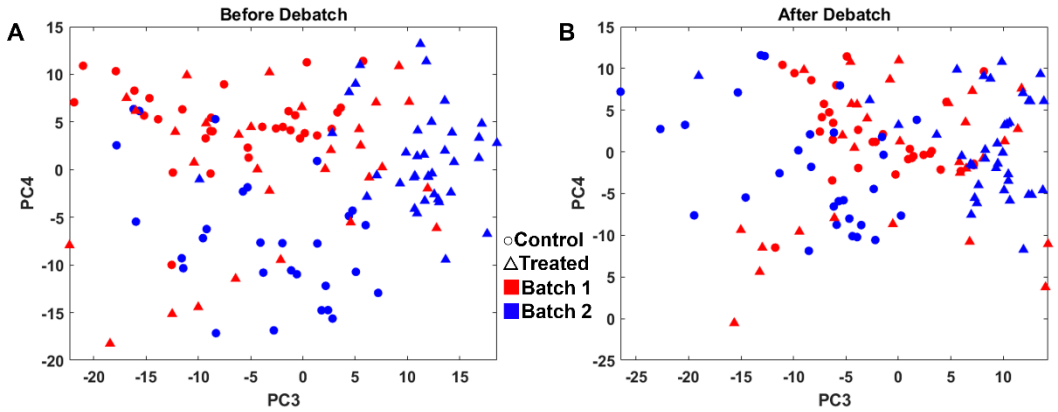


Figure S4. PCA score plots of WM266-4 cells (A) before and (B) after debatching in the PC3 and PC4 dimensions. The shapes of the symbol represent control (○) and treatment (△), and the colors of the symbol represent batch 1 (■) and batch 2 (■).

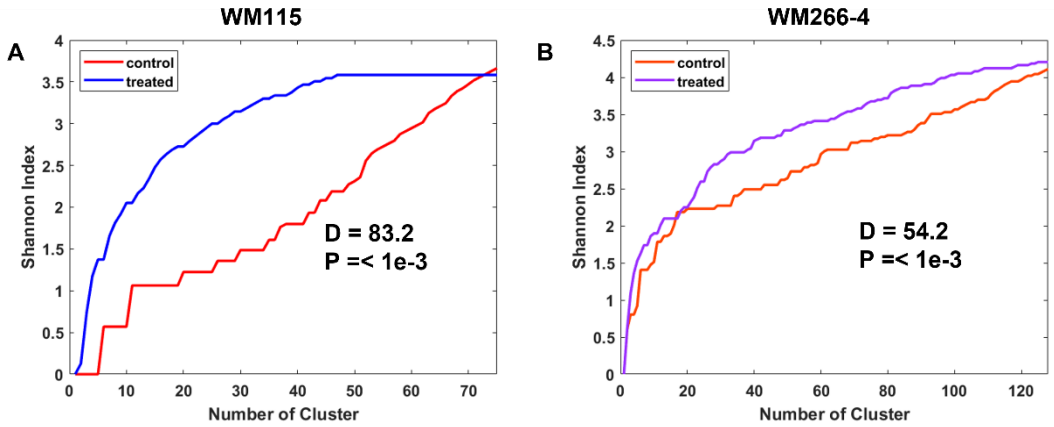


Figure S5. Change of cell heterogeneity before and after drug treatment for the (A) WM115 and (B) WM266-4 cell lines using d-statistics.

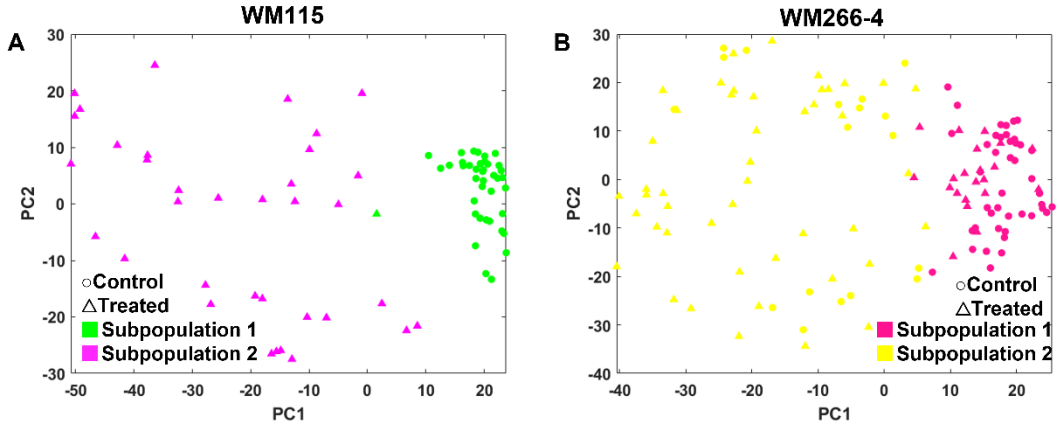
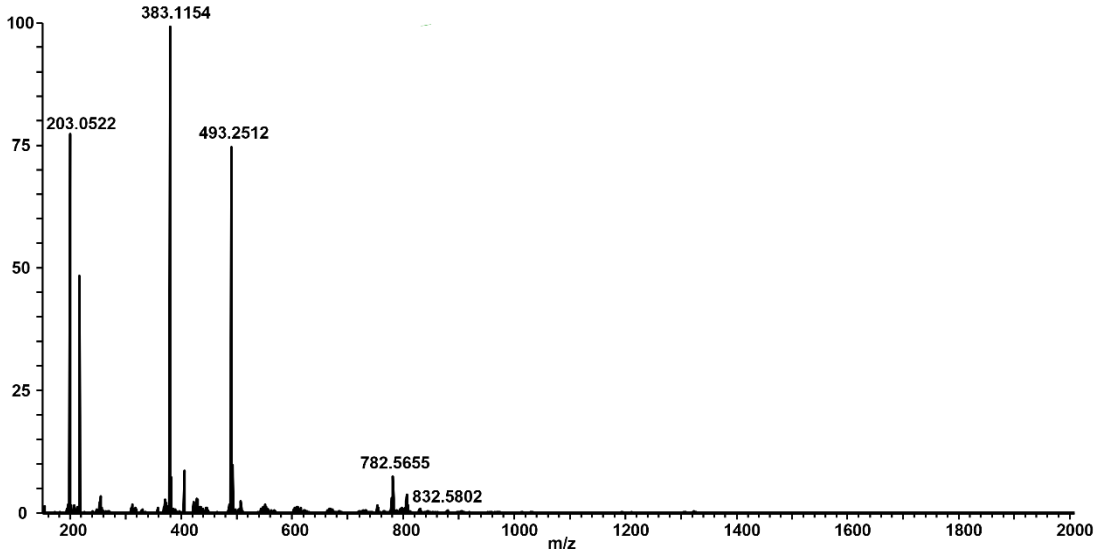
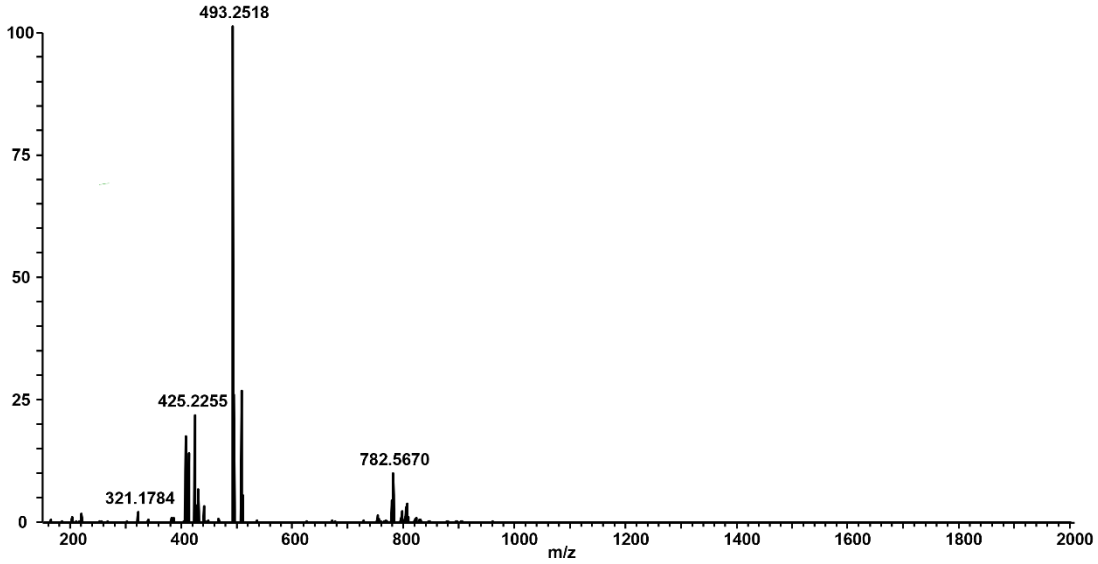


Figure S6. PCA score plots of (A) WM115 and (B) WM266-4 cells classified by subpopulations. The shapes of the symbol represent control (○) and treatment (△), and the colors of the symbol represent batch 1 (■) and batch 2 (■).

A


383.1154
203.0522
493.2512
782.5655
832.5802
100
75
50
25
0
200
400
600
800
1000
1200
1400
1600
1800
2000
m/z

B


493.2518
425.2255
782.5670
321.1784
100
75
50
25
0
200
400
600
800
1000
1200
1400
1600
1800
2000
m/z

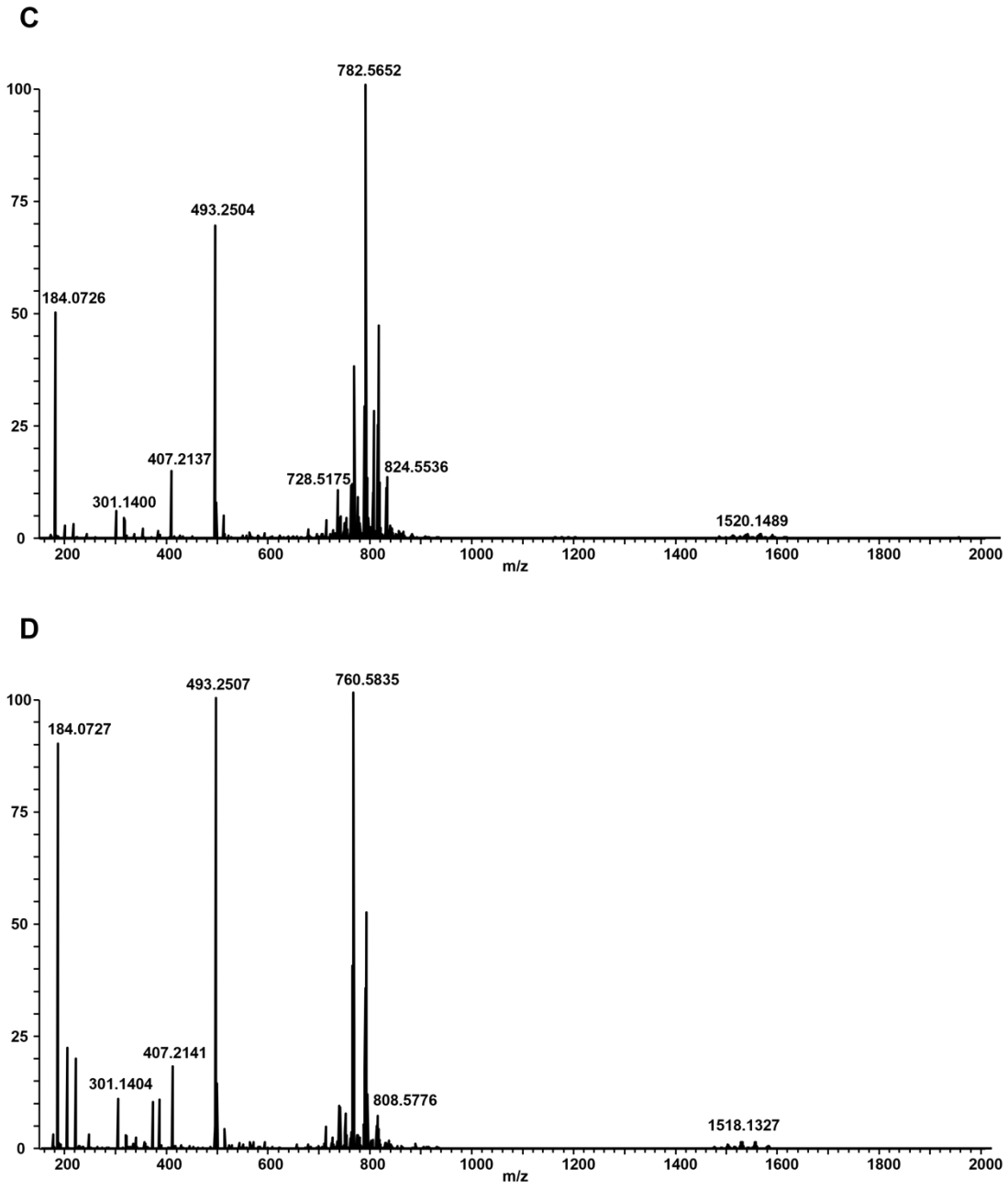


Figure S7. Representative mass spectra of single WM115 (A-B) and WM266-4 (C-D) cells with different subpopulation. (A) Subpopulation 1 of WM115; (B) Subpopulation 2 of WM115; (C) Subpopulation 1 of WM266-4; (D) Subpopulation 2 of WM266-4. Major lipid species including PC 34:1 and PC 36:2 are present in both cell lines before and after treatment with different relative intensity.

Chapter 4

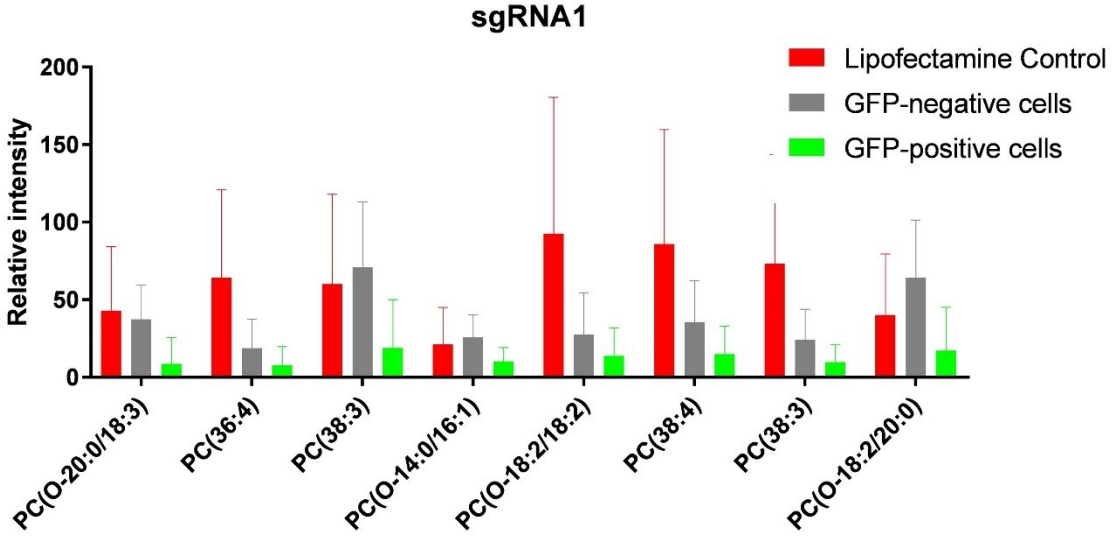


Figure S1. Phosphatidylcholines (PC) are decreased significantly in GFP-positive cells which indicates that knocking out FASN suppresses the production of PCs.

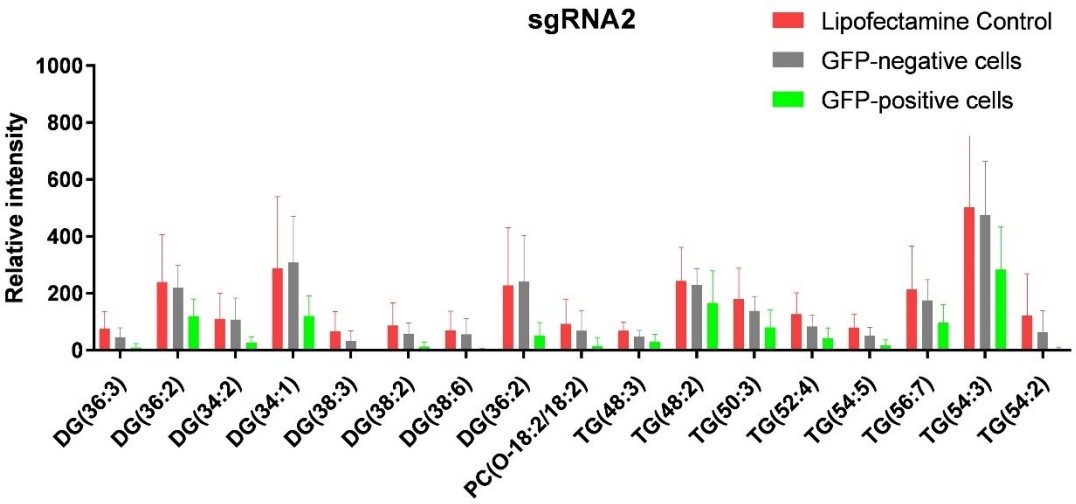


Figure S2. Diacyglycerides (DG) and Triacyglycerides (TG) are reduced in GFP-positive cells which indicate that knocking out FASN decreases the synthesis of DGs and TGs.

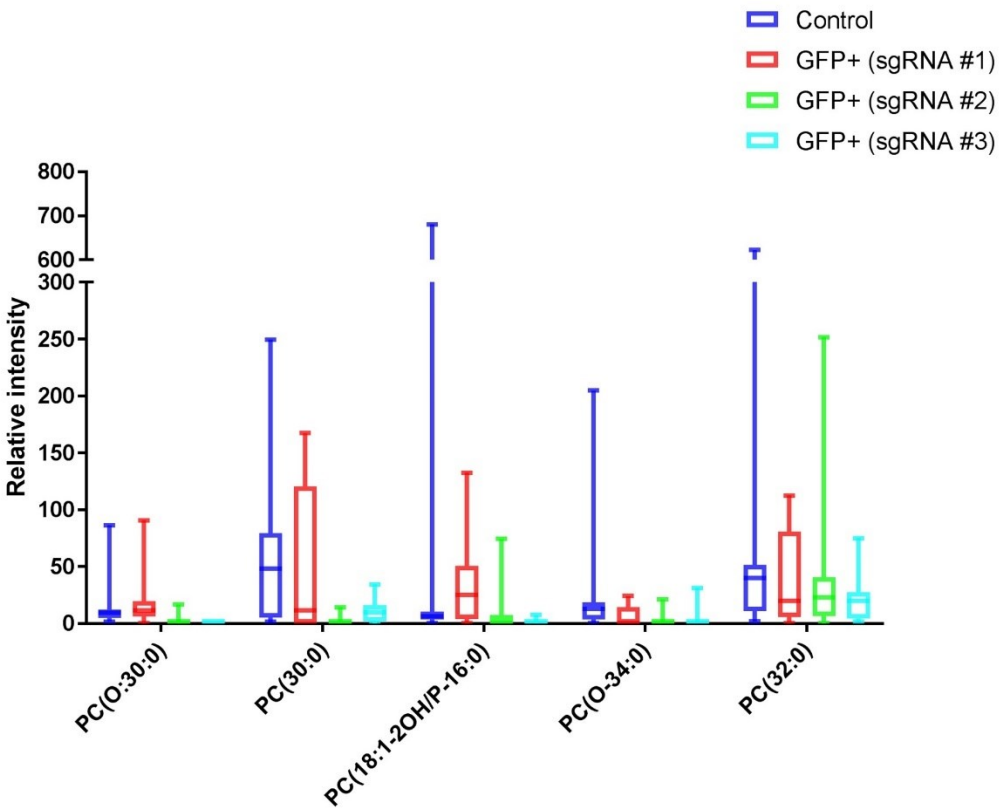


Figure S3. Glycerophospholipids are downregulated in GFP+ cells (transfected) with the expression of sgRNA #1, #2 and #3.

Appendix B

Chapter 2

Table S1. Parameters of the Thermo Fisher Q-Exactive Plus hybrid quadrupole orbitrap mass spectrometer used for Full MS/dd-MS^2 analysis.

Properties of Full MS/dd-MS^2	
General	
Runtime	0 to 12.5 min
Polarity	Positive
Default Charge	1
Inclusion	-
Exclusion	On
Full MS	
Resolution	70,000
AGC target	1 x 10^6
Scan range	70 to 1050 *m/z*
Maximum IT	246 ms
dd-MS^2	
Resolution	17,500
AGC target	2 x 10^5
Maximum IT	54 ms
Loop count	5
TopN	5
Isolation window	1.0 *m/z*
Fixed mass	-
(N)CE/stepped	NCE: 20, 40, 60

dd Settings	
Minimum AGC	8.00e3
Peptide match	Preferred
Exclude isotopes	on
Dynamic exclusion	10.0 s
ESI Ion Source	
ID	HESI
Sheath gas flow rate	35
Auxiliary gas flow rate	10
Sweep gas flow rate	0
Spray voltage	3.80 kV
S-lens RF level	50 V
Capillary temperature	320 ℃
Auxiliary gas temperature	350 ℃

Table S2. GNPS parameters used for annotation.

GNPS Search Single Spectrum	
Search Options	
Find Related Datasets	Do it
Select Databases to Search	All
Parent Mass Tolerance	0.02 Da
Ion Tolerance	0.02 Da
Min Matched Peaks	4
Score Threshold	0.7
Advanced Search Options	
Library Class	Bronze
Search Analogs	Do Search

Search Unclustered Data	Don’t Search
Top Hit Per spectrum	5
Maximum Analog Search Mass Difference	500.0
Advanced Filtering Options	
Filter StdDev Intensity	0.0
Minimum Peak Intensity	0.0
Min Peak Int	0.0
Filter Precursor Window	Filter
Filter Library	Filter Library
Filter peaks in 50 Da Window	Filter

Table S3. Metabolites differing between cell groups as determined by ANOVA (p-value <0.05, FDR-corrected)

m/z	Annotation	p value	FDR-corrected p value
267.0620	N/A	1.66E-10	4.34E-09
302.1440	N/A	0.001722	0.011115
322.885	N/A	2.01E-27	5.66E-24
359.0250	N/A	6.65E-16	3.99E-14
429.9390	N/A	0.001033	0.007073
431.9630	N/A	0.000185	0.001564
515.2590	N/A	0.005041	0.026922
523.2850	N/A	0.00346	0.019204
537.3000	N/A	1.60E-06	2.15E-05
665.3820	N/A	0.010602	0.048378
756.5470	PC(34:3) or LPC(34:4) or PC(O-34:4) (Library match to 1-Oleoyl-2-palmitoyl-sn-glycero-3-phosphocholine (PC 34:1)) (*)	0.000168	0.001429
768.583	$[PC(P\text{-}20:0/14:0)+Na]^+$ (**)	0.000742	0.005306
780.5460	$[PC(16:0/18:2)+Na]^+$ (**)	2.59E-07	4.06E-06
782.5630	$[PC(16:0/18:1)+Na]^+$ (**)	9.58E-08	1.63E-06
808.5770	$[PC(18:1/18:1)+Na]^+$ (**)	2.57E-06	3.31E-05
810.5940	$[PC(18:0/18:1)+Na]^+$ (**)	0.000114	0.001031

(*)*Features were annotated by GNPS (cosine score = 0.92; number of shared peaks = 5; mass difference to library reference =4.03) and supported by the annotated spectrum in Figure 4b.*

(**) *Features were annotated manually and supported by the annotated spectra in the Figure S4-S6.*

N/A: Metabolites were not annotatable

Chapter 3.

Table S1. Summary of the groups of cells used in the SCMS experiments.

Cell Lines	Batch	Treatment Condition*	Analyzed Cells
WM115	1	Control	13
		Treatment	18
	2	Control	26
		Treatment	18
WM266-4	3	Control	32
		Treatment	30
	4	Control	29
		Treatment	37

* 1 μM Vemurafenib was used treat cells for 48 h in the treatment groups.

Table S1. Subpopulational biomarkers between subpopulations 1 and 2 in WM115 cells.

m/z	*RSD*	Identification	PPM	$P_{post\ hoc}$*
189.987	128.1056	[2(3H)-Benzothiazolethione + Na]$^+$	23	0.03309
192.039	123.7538	[Phosphodimethylethanolamine + Na]$^+$	3	0.012697
203.052	103.7025	[D-Psicose + Na]$^+$	3	5.50E-05
225.034	129.3706	[Cysteinyl-Cysteine + H]$^+$	3	0.01681
296.065	150.745	[5-amino-1-(5-phosphonato-D-[(ribosyl)imidazol-3-ium + H]$^+$	3	9.37E-05
341.012	176.1595	[Bis(4-nitrophenyl) hydrogen phosphate + H]$^+$	14	0.000441
354.077	128.0285	[Didesethylflurazepam + Na]$^+$	3	0.040612
383.115	137.969	[Mollicellin B + H]$^+$	6	0.000359
405.097	173.8225	[Mollicellin B + Na]$^+$	6	0.008092
589.478	83.80356	[DG(32:1) + Na]$^+$ ‡	4	9.45E-08
605.452	116.7758	[DG(33:0) + Na]$^+$ ‡	18	8.01E-10
615.493	76.47161	[DG(14:0/0:0/20:2n6) + Na]$^+$	5	8.92E-07
617.509	73.18834	[DG(34:1) + Na]$^+$ ‡	4	5.66E-06
631.467	108.898	[DG(35:1) + Na]$^+$ ‡	16	8.01E-10
633.483	93.51578	[DG(35:0) + Na]$^+$ ‡	17	9.28E-10
641.509	84.88917	[DG(38:6) + H]$^+$ ‡	4	0.000501
643.524	75.0312	[DG(36:2) + Na]$^+$ †	4	0.000169
657.483	116.3986	[PA(P-16:0/18:2(9Z,12Z)) + H]$^+$	4	8.01E-10
659.498	98.13423	[PA(i-24:0/10:0) + H – H_2O]$^+$	5	9.15E-10
700.486	95.96191	[PE-NMe2(15:0/14:0) + Na]$^+$	4	8.01E-10

721.554	186.793	[DG(PGD2/0:0/i-20:0) + H]	10	0.000521
723.491	74.77151	[PA(20:2(11Z,14Z)/16:0) + Na]+	3	7.04E-06
726.501	71.44518	[PC(32:4) + H]+ ‡	8	8.01E-10
728.518	56.08359	[PC(30:0) + Na]+ †	2	1.70E-06
740.553	100.734	[PE(O-35:1) + Na]+ †	4	8.01E-10
742.475	109.1735	[PE(O-36:0) + Na]+ ‡	13	8.01E-10
742.533	80.06013	[PE(20:0/14:0) + Na]+	4	8.43E-10
742.569	89.43842	[PC(O-16:1(9Z)/18:2(9Z,12Z)) + H]+	7	8.43E-10
744.491	79.16645	[PE(P-18:1(9Z)/18:4(6Z,9Z,12Z,15Z)) + Na]+	4	9.53E-10
752.517	61.77599	[PC(34:5) + H]+ ‡	7	1.01E-08
754.534	50.81652	[PC(32:1) + Na]+ †‡	2	1.88E-05
756.549	51.33026	[PC(34:3) + H]+ ‡	6	0.000429
766.569	83.4998	[PE(O-37:2) + Na]+ †	4	3.97E-07
768.49	92.68474	[PC(32:1) + Na]+ ‡	33	8.01E-10
768.549	76.8444	[PC(16:1(9Z)/17:0) + Na]+	3	0.00706
768.585	47.3431	[PE(P-37:0) + Na]+ †	3	0.002989
770.507	69.87939	[PC(34:4) + H]+ ‡	34	1.01E-08
772.523	70.82742	[PE(36:4) + H]+ ‡	14	7.45E-07
776.516	96.64662	[PC(36:7) + H]+ ‡	8	8.01E-10
778.533	62.08769	[PC(34:3) + Na]+ †	3	1.23E-07
780.55	50.97458	[PC(34:2) + Na]+ †	1	0.000145
782.566	50.84661	[PC(34:1) + Na]+ †	1	1.05E-05
784.522	111.4313	[PC(36:3) + H]+ ‡	12	8.01E-10
788.516	100.5926	[PE(38:5) + H]+ ‡	5	0.007988
790.568	66.30461	[PC(O-18:2(9Z,12Z)/18:2(9Z,12Z)) + Na]+	5	1.05E-09
794.506	99.14211	[PC(18:3(9,11,15)-OH(13)/P-16:0) + K]+	5	8.01E-10
794.564	74.30683	[PC(36:1) + H]+ ‡	50	1.43E-06
794.601	130.933	[PC(O-18:1(9Z)/18:1(11Z)) + Na]+	3	0.000428
796.523	64.26495	[PC(34:2) + K]+ †	2	2.51E-08
796.58	76.08471	[PC(18:1(9Z)/17:0) + Na]+	3	8.81E-07
798.539	76.09119	[PE(P-18:1(9Z)/22:5(4Z,7Z,10Z,13Z,16Z)) + Na]+	2	4.37E-09
804.549	51.08599	[PC(36:4) + Na]+ †	2	0.00061
806.565	55.84285	[PC(38:6) + H]+ ‡	5	0.010579
808.582	49.38926	[PC(36:2) + Na]+ †‡	0	0.008852
810.597	46.10142	[PC(36:1) + Na]+ †	1	0.00692
816.585	61.71127	[PC(38:4) + Na]+ ‡	3	5.00E-07
818.601	102.5192	[PC(38:3) + H]+ ‡	3	8.01E-10
820.522	75.10244	[PE(39:4) + K]+ †	4	9.79E-10
822.539	70.32211	[PC(36:3) + K]+ †	2	7.34E-09
824.554	66.77384	[PC(36:2) + K]+ †	3	4.27E-08
825.691	73.80641	[SM(d41:0) + Na]+ ‡	11	1.50E-07
826.571	101.7181	[PE(42:3) + H]+ ‡	14	8.01E-10
828.71	89.96173	[PC(P-40:1) + H]+ ‡	31	3.93E-09

830.565	62.75414	[PC(40:8) + H]$^+$ ‡	5	2.83E-05
832.58	53.71502	[PC(38:4) + Na]$^+$ †	3	0.001517
834.597	77.88288	[PC(38:3) + Na]$^+$ †	1	1.08E-09
836.613	114.3903	[PE-NMe(20:2(11Z,14Z)/20:0) + Na]$^+$	1	0.000186
841.665	100.8033	[TG(49:1) + Na]$^+$ ‡	40	8.01E-10
843.681	93.34589	[PA(46:1 + H]$^+$ ‡	3	9.53E-10
846.539	79.29715			8.01E-10
848.554	82.36365			8.01E-10
851.707	69.3612	[TG(52:6) + H]$^+$ ‡	6	8.12E-06
853.722	50.85418	[TG(48:6) + H]$^+$ ‡	4	0.002183
856.741	73.06928	[PC(P-42:1) + H]$^+$ ‡	30	4.73E-07
858.596	85.44335	[PC(40:5) + Na]$^+$ ‡	3	5.34E-09
867.681	90.26346	[PA(46:0) + Na]$^+$ ‡	0	8.01E-10
869.697	80.84171	[PA(24:1(15Z)/24:1(15Z)) + H]$^+$	3	4.88E-09
871.713	92.1332	[PA(24:1(15Z)/24:0) + H]$^+$	2	1.95E-09
877.722	71.98826	[TG(16:0/14:0/22:4(7Z,10Z,13Z,16Z)) + Na]$^+$	4	0.003149
879.738	53.06125	[TG(14:1(9Z)/20:1(11Z)/18:1(9Z)) + Na]$^+$	4	0.043587
881.754	49.77383	[TG(52:2) + Na]$^+$ ‡	3	0.01081
893.696	85.14766	[PA(48:1) + Na]$^+$ ‡	1	1.61E-09
895.712	73.01709	[PA(48:0) + Na]$^+$ ‡	0	1.22E-07
897.728	75.97223	[TG(52:2) + K]$^+$	3	1.25E-07
899.744	124.2535	[TG(52:1) + K]$^+$	3	8.01E-10
909.786	101.0202	[TG(16:0/18:1(11Z)/20:1(11Z)) + Na]$^+$	2	0.001135
919.713	87.96139	[TG(54:5) + K]$^+$	2	4.77E-08
921.728	71.15654	[TG(54:4) + K]$^+$	3	5.35E-07
923.744	74.74833	[PA(a-25:0/a-25:0) + Na]$^+$	0	1.09E-07
925.76	112.0745	[TG(54:2) + K]$^+$	2	8.01E-10
947.744	114.6453	[TG(56:5) + K]$^+$	3	3.02E-09

*FDR adjusted *p*-value from *post hoc* pairwise comparison between subpopulation 1 and 2 under a familywise error rate.

†Biomarker identified at the population level.

‡Biomarker identified at the single cell level.

PC: phosphatidylcholine; PE: phosphatidylethanolamine; PA: Phosphatidic acids, SM: sphingomyelin; DG: diglyceride; TG: triglyceride.

Species in blue font color are tentatively labeled.

Table S2. Subpopulational biomarkers between subpopulations 1 and 2 in WM266-4 cells.

m/z	*RSD*	Identification	PPM	$P_{post\ hoc}^{*}$
174.013	159.4572			0.00298
176.065	184.2156	[Guanidinosuccinic acid + H]$^+$	9	0.001055
184.073	54.73049	[Phosphorylcholine + H]$^+$	5	2.97E-10
189.987	159.1209			0.030178
203.052	163.2219	[cis-Inositol + Na]$^+$	3	0.000611
219.026	134.2137	[cis-Inositol + K]$^+$	2	0.036753
226.95	134.136	[2,5-Dichloro-4-oxohex-2-enedioate + H]$^+$	4	0.045047
354.076	219.6756	[Sanguinarine + Na]$^+$	5	0.005098
383.115	202.9088	[Mollicellin B + H]$^+$	6	0.001485
650.434	111.3089			1.44E-07
672.416	130.9384			0.005258
678.501	99.25205	[PC(28:0) + H]$^+$ ‡	9	2.97E-10
692.553	88.21998	[PC(29:0) + H]$^+$ ‡	8	3.05E-10
703.573	159.5482	[SM(34:1) + H]$^+$ ‡	3	4.07E-08
704.516	77.16659	[PE(O-34:1) + H]$^+$ ‡	9	2.97E-10
706.534	80.57585	[PC(30:0) + H]$^+$ †	5	2.97E-10
718.568	74.11582	[PC(O-16:0/16:1(9Z)) + H]$^+$	9	2.97E-10
720.547	90.48607	[PE(34:0) + H]$^+$ ‡	9	3.37E-10
720.586	105.4016			2.97E-10
725.55	207.105	[SM(34:1) + Na]$^+$ ‡	9	0.000322
730.531	90.67068	[PE-NMe(20:2(11Z,14Z)/14:0) + H]$^+$	10	4.43E-09
731.599	131.3111	[SM(36:1) + H]$^+$ ‡	10	1.97E-05
732.55	76.242	[PE(35:1) + H]$^+$ †‡	5	2.97E-10
734.566	80.08571	[PC(32:0) + H]$^+$ ‡	5	2.97E-10
742.566	106.06	[PE(O-36:0) + Na]$^+$ ‡	11	0.01014
744.546	117.12	[PE(18:1(9Z)-O(12,13)/P-18:0) + H]$^+$	10	4.40E-10
744.584	73.40015	[PC(33:2) + H]$^+$ ‡	8	2.97E-10
746.562	65.493	[PC(16:1(9Z)/17:0) + H]$^+$	10	3.80E-10
746.579	86.20477	[PC(O-34:1) + H]$^+$ ‡	13	2.97E-10
746.602	82.15653	[PC(O-34:1) + H]$^+$ ‡	5	2.97E-10
748.616	101.1125	[PC(O-16:0/18:0) + H]$^+$	7	3.05E-10
758.566	63.44769	[PE(37:2) + H]$^+$ †	4	2.97E-10
760.582	65.28295	[PC(34:1) + H]$^+$ †	4	2.97E-10
762.584	66.62915	[PC(P-34:3) + Na]$^+$ ‡	22	8.37E-08
762.596	78.46355	[PE(22:0/15:0) + H]$^+$	6	1.17E-08
770.599	54.68057	[PC(O-34:0) + Na]$^+$ ‡	6	2.97E-10
772.577	83.68257	[PC(35:2) + H]$^+$ ‡	10	2.97E-10

772.615	71.29696	[PC(O-18:1(9Z)/18:1(11Z)) + H]$^+$	8	3.02E-10
774.593	84.97327	[PC(35:1) + H]$^+$ ‡	10	2.97E-10
774.631	96.38335	[PC(O-18:1(9Z)/18:0) + H]$^+$	8	2.97E-10
784.582	74.77287	[PC(36:2) + H]$^+$ ‡	1	2.97E-10
786.597	67.40736	[PE(39:2) + H]$^+$ †	4	2.97E-10
794.612	116.6464	[PC(O-36:1) + Na]$^+$ ‡	8	1.35E-06
796.614	69.19304	[PC(37:4) + H]$^+$ ‡	6	2.97E-10
811.661	141.1078	[SM(40:0) + Na]$^+$ ‡	10	0.000206
812.609	106.9853	[PC(37:4) + H]$^+$ ‡	9	1.00E-06
813.677	151.3288	[SM(42:2) + H]$^+$ ‡	9	0.001812
814.625	91.72509	[PC(38:2) + H]$^+$ ‡	9	2.97E-10
830.545	132.5015	[PC(40:8) + H]$^+$ ‡	11	0.00117
834.593	103.2547	[PC(38:3) + Na]$^+$ †	6	0.032411
1464.086	120.3208	[CL(18:0/18:0/18:1(9Z)/18:0) + H]$^+$	2	2.97E-10
1466.101	112.6823	[CL(i-16:0/i-17:0/i-17:0/22:0) + H]$^+$	3	2.97E-10
1478.138	125.9171			2.97E-10
1480.153	131.1825	[CL(i-16:0/i-17:0/i-17:0/23:0) + H]$^+$	22	2.97E-10
1490.102	109.0322	[CL(i-16:0/a-17:0/18:2(9Z,11Z)/23:0) + H]$^+$	2	2.97E-10
1492.117	101.9152			2.97E-10
1504.153	120.7291	[CL(i-16:0/a-17:0/18:2(9Z,11Z)/i-24:0) + H]$^+$	22	2.97E-10
1506.168	119.1866			2.97E-10
1516.117	104.9038	[CL(i-16:0/a-17:0/i-20:0/i-21:0) + Na]$^+$	1	2.97E-10
1518.132	99.1318	[CL(i-16:0/a-17:0/18:2(9Z,11Z)/25:0) + H]$^+$	3	2.97E-10
1520.147	104.2193			2.97E-10
1542.133	104.062			5.47E-10
1544.148	101.6704	[CL(i-16:0/a-17:0/i-21:0/i-22:0) + Na]$^+$	1	2.97E-10
1546.163	106.3662	[CL(16:0/18:2(9Z,11Z)/i-22:0/i-22:0) + H]$^+$	3	2.97E-10
1570.164	119.6729			2.97E-10

*FDR adjusted *p*-value from *post hoc* pairwise comparison between subpopulation 1 and 2 under a familywise error rate.

†Biomarker identified at the population level.

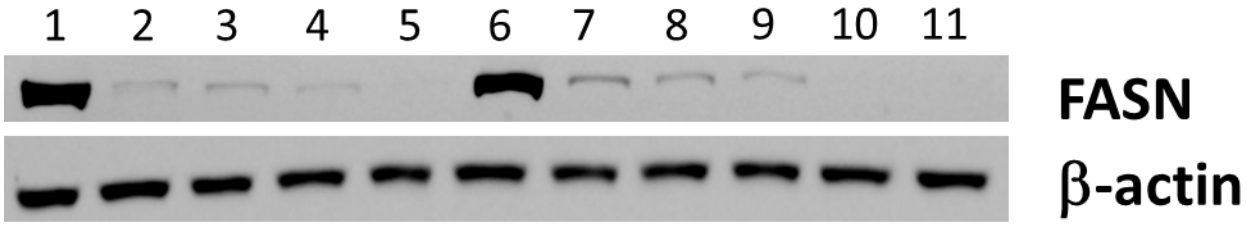
‡Biomarker identified at the single cell level.

PC: phosphatidylcholine; PE: phosphatidylethanolamine; PA: Phosphatidic acids; SM: sphingomyelin; DG: diglyceride; TG: triglyceride; CL: phospholipid cardiolipin.

Species in blue font color are tentatively labeled.

Chapter 4

Table S1. The Western blot analysis review the endogenous and exogenous fasn protein expression using sgRNA #1 and sgRNA #2 plasmid at different concentration.



Lane Sample

1. 1.0 ug FASN
2. 1.0 ug FASN + 1.5ug sgRNA1
3. 1.0 ug FASN + 2.0 ug sgRNA1
4. **1.0 ug FASN + 3.0 ug sgRNA1**
5. 3.0 ug sgRNA1 only
6. 1.0 ug FASN
7. 1.0 ug FASN + 1.5ug sgRNA2
8. 1.0 ug FASN + 2.0 ug sgRNA2
9. **1.0 ug FASN + 3.0 ug sgRNA2**
10. 3.0 ug sgRNA2 only
11. No DNA control