METABOLIC DIVERSITY IN HUMAN NON-SMALL CELL LUNG CANCER CELLS

APPROVED BY SUPERVISORY COMMITTEE

_____
Ralph J. DeBerardinis, M.D., Ph.D. (Mentor)

_____
John Minna, M.D.

_____
Michael White, Ph.D.

_____
Joseph Garcia, Ph.D., M.D.

DEDICATION

I dedicate my work to all the people that have or are currently fighting against cancer.
You are the reason I am doing this.

METABOLIC DIVERSITY IN HUMAN NON-SMALL CELL LUNG CANCER CELLS

by

PEI - HSUAN CHEN

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

December, 2015

ACKNOWLEDGEMENTS

I thank:

my Ph.D. mentor, Dr. Ralph J. DeBerardinis, for helping me to grow as a scientist;

my lab mates and friends, for making my time in UTSW wonderful;

my parents, Chun-Hung Chen and Hui-Mei Wu, for their support and encouragement.

my husband, Carlos Rodrigo Gil del Alcazar, and our son.

I love you so much.

METABOLIC DIVERSITY IN HUMAN NON-SMALL CELL LUNG CANCER CELLS

PEI - HSUAN CHEN

The University of Texas Southwestern Medical Center at Dallas, 2015

Ralph J. DeBerardinis, M.D., Ph.D.

ABSTRACT

Cancer cells display oncogene-driven rewiring of metabolism to produce energy and macromolecules for growth. Inhibition of growth-promoting metabolic pathways may prove to be a useful therapeutic strategy in cancer. However, neither the full breadth of cancer cell metabolic diversity, nor the complement of mechanisms by which tumor mutations elicit metabolic reprogramming, are known. We set out to characterize cell-autonomous metabolic heterogeneity in non-small cell lung cancer (NSCLC) and to use orthogonal high-content data sets to understand the mechanisms by which metabolic phenotypes are established in lung cancer. A major goal is to understand whether these metabolic phenotypes predict therapeutic liabilities to novel metabolic inhibitors, targeted therapies, or conventional chemotherapeutic agents.

We used a highly annotated panel of more than 80 NSCLC cell lines to develop the most comprehensive database of cancer cell metabolism to date. These cell lines were analyzed for a set of ~100 metabolic parameters derived from nutrient utilization, nutrient addiction, and isotope labeling patterns following culture with ^{13}C-glucose or ^{13}C-glutamine. Orthogonal data sets included analysis of the genome, epigenome, transcriptome and proteome, as well as sensitivity to over 40 chemotherapeutic agents. Several cell lines were also subjected to high-throughput chemical compound and genome-wide siRNA screens.

NSCLC cell lines display a surprising degree of cell-autonomous metabolic heterogeneity in culture. Many canonical hallmarks of cancer cell metabolism, including the Warburg effect, were observed to span at least a 10-fold range among cell lines grown under identical conditions. Affinity propagation clustering using metabolic features alone produced families that were largely distinct from clusters based solely on gene expression. Nevertheless, databases of metabolic features and orthogonal data sets could be cross-queried to identify robust, novel relationships connecting metabolic preferences to oncogenotypes, transcriptomic phenotypes and therapeutic responses.

Focused metabolic assays can produce a highly informative view of the metabolic phenotyping among large panels of cell lines. NSCLC cell metabolism is highly heterogeneous in every parameter so far assessed. Functional metabolic families describe an unparalleled view of the connections between genetics, drug sensitivity and cell-autonomous metabolism in NSCLC.

TABLE OF CONTENTS

PUBLICATIONS

DeNicola GM, Chen PH, Mullarky E, Sudderth JA, Wu D, Tang H, Xie Y, Asara JM, Huffman KE, Wistuba II, Minna JD, DeBerardinis RJ and Cantley LC. NRF2 regulates serine biosynthesis in non-small cell lung cancer. Nature Genetics. (2015, under revision)

Mullen AR, Wheaton WW, Jin ES, Chen PH, Sullivan LB, Cheng T, Yang YF, Linehan WM, Chandel NS, DeBerardinis RJ. Reductive carboxylation supports growth in tumor cells with defective mitochondria. Nature. 481:385-8 (2011)

LIST OF FIGURES

LIST OF ABBREVIATIONS

[U-^{13}C]: Molecule in which all carbons are ^{13}C

3PG: 3-Phosphoglycerate

3PP: 3-phosphohydroxypyruvate

5-FU: 5-fluorouracil

ALK: anaplastic lymphoma kinase

ATP: Adenosine Triphosphate

AAA: Amino acid analyzer

BSA: Bovine serum albumin

CoA: Coenzyme A

DHFR: dihydrofolate reductase

EGFR: epidermal growth factor receptor

FDG: Fluorodeoxyglucose (^{18}F)

FACS: Fluorescence Activated Cell Sorting

FBS: Fetal bovine serum

GARFT: glycinamide ribonucleotide formyltransferase

GC: Gas Chromatography

Glc: Glucose

Gln: Glutamine

Glu: Glutamate

Gly: Glycine

HIF: hypoxia inducible factor

HPLC: High-pressure liquid chromatography

KEGG: Kyoto encyclopedia of genes and genomes

Lac: Lactate

LDH: Lactate deyhydrogenase

NAD: Nicotinamide adenine dinucleotide

M+n: full parent ion for a given metabolite with n ^{13}C carbons incorporated

Mal: Malate

MID: Mass isotopomer distribution

Min: minutes

MS: Mass spectrometry

NMR: Nuclear magnetic resonance

NSCLC: Non-small cell lung cancer

OAA: Oxaloacetate

PBS: Phosphate-buffered saline

PC: Pyruvate carboxylase

PDH: Pyruvate dehydrogenase

PEM: Pemetrexed

PHGDH: phosphoglycerate dehydrogenase

PSAT: Phosphoserine aminotransferase

PSPH: Phosphoserine Phosphatase

Ser: serine

SHMT: Serine hydroxymethyltransferase

TCA: Tricarboxylic acid

TS: Thymidylate synthase

^{x}C: Carbon atom with x atomic mass unit

CHAPTER ONE
BACKGROUND AND OBJECTIVE

Introduction

Metabolism is the set of enzyme-catalyzed transformations within living cells. One of the emerging hallmarks of malignant cell is an altered metabolic state that enables growth and proliferation of the cancer cells (Hanahan and Weinberg, 2011). The discovery of cancer-associated mutations in genes encoding key metabolic enzymes has provided a link between altered metabolism and cancer. However, cancer is not just one disease, but a collection of disorders. Different cancer cells, tissues and organs may develop diverse strategies for metabolic reprogramming. Even though targeting cancer metabolism has been suggested as a therapeutic strategy, neither the full breadth of cancer cell metabolic diversity, nor the complement of mechanisms by which tumor mutations elicit metabolic reprogramming, are known.

The goal of this project is to obtain an integrated view of cancer genetics and metabolism, using Non-Small Cell Lung Cancer (NSCLC) as a model. In this introductory chapter, I briefly discuss glycolysis, glutaminolysis and the tricarboxylic acid (TCA) cycle with its role in energy generation and macromolecular synthesis. I then focus on stable isotope tracing and its role in metabolic phenotyping. I then briefly introduce NSCLC and end with discussing my main goal and the outline of the subsequent chapters.

Glucose and glutamine fuel TCA cycle and biosynthesis

Uncontrolled growth is a universal trait of tumor cells, which requires profound changes in cellular metabolism to sustain the additional energy and biosynthetic precursor demands of proliferation (Zhao et al., 2013). The Warburg effect is defined by an increased utilization of glucose via glycolysis as a cellular resource, and is a common phenotype of cancerous cells. The enhanced glucose uptake observed in cancer cells has been used to detect and image cancers via PET, which detect 2-18F-2-deoxyglucose as a result of their rapid uptake of glucose (DeBerardinis et al., 2007).

Glycolysis is the set of reactions that begins with glucose and yields two molecules of pyruvate in addition to energy. Despite the dramatic upregulation of glycolysis in many cancer cells, this process alone is insufficient to provide cell proliferation and must be supplied by additional metabolic processes such as glutaminolysis (Mazurek and Eigenbrodt, 2003). Glycolysis and glutaminolysis provide than just energy generation. Intermediates produced through glycolysis and glutaminolysis are necessary to produce the basic building blocks of cell growth. For example, several TCA cycle metabolites such as oxaloacetate (OAA) and α-ketoglutarate are maintained at low cellular concentration and must be replenished via anaplerosis (Deberardinis et al., 2008b) to the TCA cycle or the cycle will cease to function. Furthermore, the carbon and nitrogen from glucose and

glutamine supply diverse biosynthesis pathways like fatty acids, nucleic acids, and proteins (Nelson D.L. Cox M.M, 2005).

Genomics and metabolomics studies

Although metabolic reprogramming is thought to be essential for rapid cancer cell proliferation, a systematic characterization of the metabolic pathways active in transformed cells is lacking, and the contribution of these pathways in promoting rapid cancer cell proliferation remains unclear (Hsu and Sabatini, 2008). A standard reductionist approach was used to discover the regulatory mechanisms of how oncogenes and tumor suppressors regulate cancer metabolism. It is usually done by analyzing the metabolic differences between nontransformed cells and the cells transformed by the addition of chemical or oncogene. However, with this approach we can only study the effect of over a few genes a time. Due to the interconnected nature of metabolic pathways and the pleiotropic effects mediated by oncogenic signals, a systems approach is required to elucidate the mechanisms of such transforming events (Gaglio et al., 2011).

Many systems-biology technologies are used for cancer research, like gene expression microarrays, serial analysis of gene expression, two-dimensional differential gel electrophoresis, protein chips and antibody-based arrays. However, challenges remains such as knowing that the statistical solution is correct, complete or accurate and

avoiding the trap of self-fulfilling prophecy. Using incomplete biological knowledge to guide class identification can lead to incorrect class assignment or incorrect data interpretation. Avoiding the trap of self-fulfilling prophecy is a challenge for which incomplete knowledge of gene function and cellular context can lead to the creation of incorrect signaling links in network building (Clarke et al., 2008). Metabolomics is the newest layer of omics data that is rapidly gaining attention. It seeks to quantify the metabolites in the metabolome, and to use this data to improve cancer prognosis and therapeutics (Shajahan-Haq et al., 2015).

Although the human metabolome contains more than 15,000 metabolites, the bulk of metabolism in cancer cells flows through a relatively small set of core pathways providing essentially all of the energy and precursors needed for survival and growth. Thus while metabolomics seeks to catalog as many metabolites as possible, few studies have systematically analyzed activity of the pathways themselves. For example, citrate is an essential precursor for fatty acids, sterols, and other molecules, and is formed by condensing OAA and acetyl-CoA. Many independent pathways exist to produce OAA and acetyl-CoA, each subject to various levels of control. Simply quantifying citrate abundance cannot provide any information about which pathways are active or how they are regulated.

Stable isotope tracing analysis

To understand and quantify pathway actually, researchers use stable isotopes to trace pathways connecting extracellular nutrients to intracellular metabolites, including citrate. Cancer cells are cultured in medium containing ^{13}C labeled nutrients, then metabolites are extracted and analyzed by mass spectrometry (MS) to analyze the mass isotopomer distribution (MID) in each metabolite. In other words, the fraction of each metabolite pool that contains 0, 1, 2,…n ^{13}C atoms, where n is the total number of carbons in the molecule. Specific information about which pathway led to the formation of the metabolite is encoded within the MID. MID of citrate is particularly useful because this metabolite is uniquely positioned at the intersection between oxidative metabolism typical of quiescent cells and rapid utilization of TCA cycle intermediates for biosynthesis/growth in cancer cells (Parlo and Coleman, 1984). These isotope tracing methods previously demonstrated that glucose and glutamine supply essentially all of the core metabolites used by cancer cells (DeBerardinis et al., 2007) and that oncogenes regulate specific pathways of glucose and glutamine utilization (Cheng et al., 2011; Wise et al., 2008; Yang et al., 2009).

Stable isotope tracing uses isotopically labeled carbon substrates, which, in this project, are mainly glucose and glutamine. After cells are incubated with specific tracers for a given period of time, metabolites are extracted from the cells and analyzed. In the GC-MS analysis, cellular metabolites are first separated between mobile and stationary phases. Based on the different affinity of the metabolite, the metabolite is eluted from the column into an ion mass detector at a certain time point. The mass detector then

fragments each metabolite and determines an m/z (mass/charge) ratio for each ion entering the detector at a specific time. Metabolites whose carbons are derived from heavy glucose or glutamine have an increased m/z ratio. The identity of the metabolite is determined based on the analysis of a pure standard on the GC-MS. By examining the MIDs of various metabolites when cells are cultured with heavy glucose or glutamine, we can understand which metabolic pathways might be active. Our lab used isotope tracing to identify the induction of a novel pathway involving reversal of the TCA cycle in cancer cells lacking the metabolic tumor suppressor, fumarate hydratase (Mullen et al., 2012). Pure metabolomics without isotope tracing would have missed this novel pathway.

Non-small-cell lung cancer

Lung cancer is the leading cause of cancer-related mortality worldwide and is linked to 28% of all cancer-related deaths in the United States, with an estimated 160,000 deaths per year for the USA alone (Jemal et al., 2010). Despite advances in traditional therapeutic strategies involving surgery, ionizing radiation therapy, and chemotherapy, the 5-year survival remains less than 20%, and has changed little over the past four decades (Rapp et al., 1988). There are 2 major types of lung cancer, small cell lung cancer (SCLC) or non-small cell lung cancer (NSCLC). About 85% of lung cancers are NSCLC. There are three major subtypes of NSCLC: squamous cell carcinoma,

adenocarcinoma and large cell carcinoma. Cells in these subtypes differ in size, shape, and grow and spread in different ways.

In the past decade, a deeper understanding of the molecular abnormalities present in NSCLC that define disease subsets has been achieved, particularly in adenocarcinoma (Gower et al., 2014). The discovery of oncogenic drivers such as EGFR, ALK, RAS, MET and HER2 led to the design of therapies targeting tumors harboring specific gene alterations that cause aberrant signaling and proliferation (Pao and Chmielecki, 2010). Currently, testing for EGFR mutations and ALK fusions, regardless of sex, race, smoking history, or other risk factors, are prioritized over other molecular predictive tests (Lindeman et al., 2013). This paradigm for targeted therapies in oncology allows the right patient to receive the most active therapy, while sparing those that are unlikely to benefit from the cost and potential morbidity associated with unresponsive therapeutic interventions. However, despite some clinical success of targeted therapies in NSCLC, responses typically last less than a year as tumors inevitably acquire resistance (Hrustanovic et al., 2013; Niederst and Engelman, 2013).

An improved understanding of potential vulnerabilities of lung cancer is still urgently needed. For example, little is known about specific susceptibilities that may derive from the loss of classical tumor suppressor genes like LKB1 (Whang et al., 2015), which is inactivated in 20-30% of lung adenocarcinoma patients (Sanchez-Cespedes et al., 2002). Recently, it was reported that LKB1 deficiency enhances sensitivity to energetic stress induced by Erlotinib treatment (Whang et al., 2015) and elicits a redox Imbalance to

modulate therapeutic response (Li et al., 2015). The metabolic changes of NSCLC were also investigated by infusing uniformly labeled ^{13}C-glucose into human lung cancer patients. This revealed that both glycolysis and oxidative metabolism of pyruvate in the TCA cycle were apparent in the tumor, and that the abundance of metabolites labeled by these pathways was generally higher in the tumors than in the lung (Fan et al., 2009).

Statement of purpose

Cancer cells display oncogene-driven rewiring of metabolism to produce energy and macromolecules for growth. Targeting cancer metabolism has been suggested as a therapeutic strategy (Deberardinis et al., 2008b; Hsu and Sabatini, 2008). However, little is known about the heterogeneity at the cell metabolic level or the complement of mechanisms by which tumor mutations elicit metabolic reprogramming. In this project, we set out to characterize cell-autonomous metabolic heterogeneity in NSCLC and to use orthogonal high-content data sets to understand the mechanisms by which metabolic phenotypes are established in lung cancer.

In order to close the gap, we used a highly annotated panel of more than 80 NSCLC cell lines to develop the most comprehensive database of cancer cell metabolism to date. Then we compared it with orthogonal data sets including analysis of the genome, epigenome, transcriptome and proteome, as well as sensitivity to over 40 chemotherapeutic agents. There are three major questions that we want to answer by

this project. First, how much metabolic diversity exists among cancer cell lines grown under identical conditions? Second, how do metabolic features cross-correlate with each other? Third, how well does the metabolic phenotype correlate with other datasets?

Briefly, each chapter provides evidence for each corresponding major question. In chapter three, I will describe how I performed the highly informative metabolic assays and reveal the surprising heterogeneity from the data set. Then chapter four will indicate the correlations among a set of ~100 metabolic parameters derived from nutrient utilization, nutrient addiction, and isotope labeling patterns. Chapters five and six provide examples of novel mechanisms discovered by comparing metabolic profile data with orthogonal data sets. Together, this project provides a unique tool and new avenues for diagnosis, therapy, and biomarker detection in NSCLC.

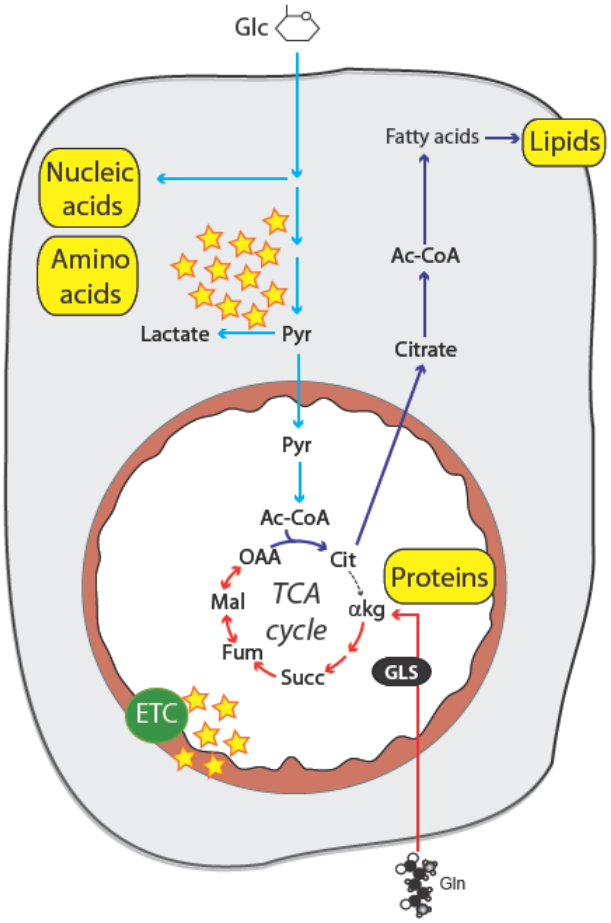


Fig 1.1 Glucose and glutamine fuel the TCA cycle and biosynthesis

The schematic shows glucose entering the TCA cycle through the glycolysis and provides Ac-CoA, then Ac-CoA combines with OAA to generate citrate. Citrate then transfers outside mitochondria and process multiple biogenesis including lipid generation. Carbon derived from glutamine enters the TCA cycle as α-ketoglutarate to replace the OAA pool, which is known as glutamine anaplerosis. Thus, glucose and glutamine are major nutrients to fuel the TCA cycle and biosynthesis pathways. Products of cataplerotic reactions are depicted in yellow. Abbreviations: α-KG, α-ketoglutarate; GLN, glutamine; GLS, glutaminase; Cit, citrate; Ac-CoA, acetyl-coA; Glc, glucose; Pyr, pyruvate; OAA, oxaloacetate.

CHAPTER TWO
MATERIALS AND METHODS

Cell culture

The Non-Small Cell Lung Cancer cell lines came from Dr. John Minna's lab at UTSW. All cell lines were cultured in Roswell Park Memorial Institute medium (RPMI) supplemented with 5% FBS (Sigma). Nutrient deprivation and metabolic labeling experiments used RPMI with dialyzed FBS (Yang et al., 2009).

Stable Isotopes labeling analysis

Stable isotope tracing of TCA cycle intermediates was performed as described previously (Cheng et al., 2011). Briefly, dishes of 80–90% confluent cells were rinsed twice in PBS, then overlaid with medium containing the isotopically enriched nutrient, and cultured for 6 hours. For analysis of intracellular metabolites by GC/MS, cells labeled in 6-cm dishes were rinsed in ice-cold normal saline and lysed with three freeze-thaw cycles in cold 50% methanol. The lysates were centrifuged to remove precipitated protein, a standard (50 nmols of sodium 2-oxobutyrate) was added, and the samples were evaporated and derivatized by trimethylsilylation (Tri-Sil HTP reagent, Thermo Scientific). Three microliters of the derivatized material was injected onto an Agilent 6970 gas chromatograph equipped with a fused silica capillary GC column and networked to an Agilent 5973 mass selective detector. Retention times of all

metabolites of interest were validated using pure standards. The abundance of the following ions was monitored: m/z 245-249 for fumarate, m/z 335-339 for malate, m/z 219-222 for lactate, m/z 306-309 for serine, m/z 276-278 for glycine, and m/z 465–471 for citrate. The measured distribution of mass isotopomers was corrected for natural abundance of ^{13}C (Fernandez et al., 1996). All isotopes were obtained from Cambridge Isotope Laboratories

Nutrient utilization rates

To measure metabolic rates, one million cells were plated into 6-cm dishes and cultured until 90% confluent. At time 0, the cells were rinsed in PBS, fed with 1.5 mL of the test medium (RPMI with 10 mM glucose, 2 mM glutamine and dialyzed FBS), and cultured. End-point experiments proceeded for 7 hours, then the medium was collected and analyzed for metabolite abundance. Concentrations of glucose, lactate, glutamine, and glutamate were determined from 0.6-mL aliquots of medium using an automated electrochemical analyzer (BioProfile Basic-4 analyzer; NOVA). Metabolic rates were determined by normalizing absolute changes in metabolite abundances to protein content.

Nutrient dependence assay and relative survival assay

To monitor proliferation, cells were seeded at 5,000/well in 48-well plates. The next day, cells were replenished with 0.5 ml test medium (complete RPMI, RPMI withdraw

glucose, or RPMI withdraw glutamine). After 1, 3 and 5 days, 0.25 ml water was added to each well and frozen at −80°C for 2 hr. Cells were then warmed to room temperature, and 0.5 ml of 0.1 µg/ml Hoechst 33258 in TNE buffer (2 M NaCl, 10 mM Tris-HCl [pH 7.4], and 1 mM EDTA) was added. The plate was incubated in the dark at room temperature, and O.D. at 350 nm was measured in plate reader (Dong et al., 2010).

Protein expression

Whole-cell lysates were prepared in RIPA buffer and quantified using the BCA protein assay (Thermo Scientific). Protein was separated by SDS/PAGE, transferred to a PVDF membrane, and probed with antibodies. The membrane was blocked overnight at 4°C in PBS with Tween 20 (PBST) containing 5% milk and subsequently probed with primary antibodies overnight at 4°C. All antibodies were purchased from Cell signaling Technology except Cyclophilin B (Proteintech).

Quantitative reverse transcriptase-polymerase chain reaction

RNA was harvested from one million cells using TRIzol Reagent (Invitrogen) or High Pure RNA Isolation Kit (Roche Applied Science) following the manufacturer's user manual. cDNAs were synthesized with the iSCRIPT cDNA synthesis kit (BioRad). Gene-specific primers were designed using the on-line program D-LUX™ Designer Software from Invitrogen or published papers. Primer sequences were show on Table2.1. 10 µl qPCR reactions contained 25 ng of cDNA, each primer at 150 nM, and 5 µl of 2X SYBR

Green PCR master mix (Applied Biosystems). Amplifications were performed on a CFX384 Touch Real-Time PCR Detection System (Bio-Rad). Results of qPCR were evaluated by the comparative Ct method (user bulletin No.2, Perkin Elmer Life Sciences) using ribosomal protein L13a (RPL13A) as the control gene.

^{14}C-Palmitate oxidation assay

Fatty acid oxidation was measured by the production of $^{14}CO_2$ from ^{14}C-palmitate as described previously (Fediuc et al., 2006) with a few modifications. Briefly, cells were incubated for 6 hours in 24-well plate with Krebs buffer containing ^{14}C-palmitic acid (10% BSA: ^{14}C-Palmitate= 3:1) and nonlabeled palmitate. Each working well was set with Micro-Bridges, which had a piece of chromatography paper wetted with 2M NaOH to trap the CO_2 produced during the incubation period. Subsequently, 50 µl of 20% trichloroacetic acid was added to the cells, which were then incubated for an additional 1 hour at 37°C. Finally, the pieces of paper were transferred to scintillation vials for radioactivity counting.

RNA Interference

Vectors for RNAi were obtained from commercial libraries based on the pGIPZ (shRNAmir) or pLKO.1 (shRNA) backbones (Open Biosystems). The MigCD8t control and LKB1 overexpression vectors were kindly provided by Dr. Russell Jones from McGill University. Lentiviral particles were produced by cotransfecting 293T cells with the

lentiviral construct, pCMVΔR8.91, and pMD2.G using Fugene HD (Roche). Virus-containing supernatant was collected 2 d post transfection and used to infect cells. LKB1 stable cell lines were selected by CD8t markers through flow cytometry (Nicolas Loof, the Moody Foundation Flow Cytometry Facility of UTSW). For other stable cell lines, Puromycin (1 μg/mL) was added 2 days after infection, and selection was continued for 10 d before any experiments. Stably infected pools with adequate silencing were maintained in 1 μg/mL puromycin.

As for siRNA transfection, small interfering RNAs were obtained from Integrated DNA Technologies (IDT) and reconstituted in water. Cells were transfected with PolyJet reagent (SignaGen) as manufacturer protocols, and protein abundance was examined after 72 h using western blot. Sequences for RNA interference are listed on Table 2.2.

BrdU staining

Cells were cultured with 10uM BrdU for one hour, and fixed in 70% ethanol at -20°C. After washed with phosphate/citric buffer (40ml Na2HPO4 with 4ml 0.1M citric acid), cells were incubated with BrdU antibody for one hour, and proceed to PI staining. For PI/RNase staining, cells were incubated with 0.5 mL of PI/RNAse staining buffer (BD Pharmingen) at 25°C for 15 minutes, then analyzed with FACS.

Free amino acid concentration quantitation

Consumption and secretion of amino acids were measured by HPLC using an amino acid analyzer (Hitachi; L8900). One million cells were cultured in 6cm dishes and switched to RPMI medium with dialyzed FBS for 7 hours. The starting and used medium samples were collected and mixed with same amount of 1M sulfosalicylic acid, followed by centrifugation at 10000 rpm for 15 min at 4°C. The supernatant phase was collected and run in the machine with AEC standard (1:10 by volume).

Soft agar colony formation assay

1000 cells were planted in 12-well plates using 0.33% Noble agar (Difco). After 2 weeks, the cultures were stained with 0.05% crystal violet in 20% methanol. Colonies >200 μm in diameter were counted.

***In vivo* tumor growth**

Animal procedures were performed with the approval of the UT Southwestern IACUC. Cells were suspended in RPMI, mixed 1:1 with Matrigel (Becton Dickinson), and 0.1 million cells were implanted subcutaneously into 6- to 8-week-old male NCr nude mice. Mice received 10 days i.p. injections of normal saline or Pemetrexed (200 mg/Kg). Tumor size was measured with electronic calipers every 3 days. Tumor volume was calculated according to the formula $[(\text{length}) \times (\text{width})^2]/2$. Pemetrexed disodium was purchased from Medchem Express.

Statistical Analysis

Data are presented as mean ± S.D., and Student's t test (two-tailed distribution, two-sample unequal variance) was used to calculate p values. Statistical significance is displayed as $*p < 0.05$ or $**p < 0.01$ unless specified otherwise. The tests were performed using Microsoft Excel.

name	primer
ADORA2B -L	AATGAAAGCTGCTGCCTTGT
ADORA2B -R	GCAAAAATCCCCACAATCAT
GSS -L	TGGCTGGGACTAAGAAGGTG
GSS -R	TTCAGGGCCTGTACCATTTC
CXCL8 -L	CTGCGCCAACACAGAAATTA
CXCL8 -R	ATTGCATCTGGCAACCCTAC
SLC2A1 -L	TGATGCTCAGGCTTGAAATG
SLC2A1 -R	AAATGAACACAGGGCAGCTT
PHGDH -F	ATCTCTCACGGGGGTTGTG
PHGDH -R	AGGCTCGCATCAGTGTCC
PSAT1 -F	ACTTCCTGTCCAAGCCAGTGGA
PSAT1 -R	CTGCACCTTGTATTCCAGGACC
PSPH -F	ACTTCCTGTCCAAGCCAGTGGA
PSPH -R	CTGCACCTTGTATTCCAGGACC
SHMT1 -F	TGAACACTGCCATGTGGTGACC
SHMT1 -R	TCTTTGCCAGTCTTGGGATCC
SHMT2 -F	GCCTCATTGACTACAACCAGCTG
SHMT2 -R	ATGTCTGCCAGCAGGTGTGCTT
NRF2 -F	GAGAGCCCAGTCTTCATTGC
NRF2 -R	TTGGCTTCTGGACTTGGAAC
NQO1 -F	TGAAGAAGAAAGGATGGGAGGT
NQO1 -R	GGCCTTCTTTATAAGCCAGAACA
ATF4 -F	GTTCTCCAGCGACAAGGCTA
ATF4 -R	GCATCCAAGTCGAAACTCCTT
bACTIN -F	CAACCGCGAGAAGATGACC
bACTIN -R	ATCACGATGCCAGTGGTACG

Table 2.1 primers for qPCR

name	siRNA
SHMT1-A	CTGCTGTAAATCAGAAGTGTA
SHMT1-B	CTGACGGAGCTGGGCTACAAA
SHMT1-C	CTCCCGTAATCAGGAAGCCAA
SHMT2-A	CAGGCGCAGCAAATTCAATTT
SHMT2-B	CCGGGAGATCCCTTACACATT
SHMT2-C	CCCAGCCAACCTGGCCGTCTA

Table 2.2 sequences for RNA interference

CHAPTER THREE

Discover the metabolic diversity by systematic analysis of glucose and glutamine metabolism in NSCLC cells

Introduction

In order to reveal the intrinsic diversity of NSCLC cell metabolism, I performed three metabolic assays that are the most systematic, informative and appropriate to apply on large scale analysis. First assay is the stable isotope labeling analysis. Since glucose and glutamine are the two major carbon sources feeding in TCA cycle, I used uniformly-labeled ^{13}C glucose or glutamine in cell culture medium to replace the unlabeled ^{12}C glucose or glutamine. Then by analyzing cellular metabolites using GC-MS, I will be able to determine the metabolic pathways used by different cell lines. The second assay is to characterize the major nutrient uptake and secretion rate of each cell line. Most cells consume glucose and glutamine as the major nutrients *in vitro*, and secrete out lactate and glutamate as byproducts. This nutrient utilization rate assay will profile the content of these four important metabolites in cell medium as bulk metabolic phenotypes. These data can later be used as constraints to estimate intracellular metabolic features. Nutrient dependence assay is the third assay I performed and it provides critical cell survival information in the absence of glucose or glutamine.

Using these three assays, I was able to extract 100 metabolic features as Table 3.1. It includes the full range of isotopomer analysis in critical metabolites such as citrate,

malate, fumarate, lactate, serine and glycine, which can be used for tracing multiple metabolic pathways. Also the nutrient utilization rates and nutrient dependence assay data provide basic metabolic characterization of the cells, such as how much major nutrient the cells need, how addictive the cells depend on the nutrients, and the growth rate. Each assay measures a different view of cell metabolism, and together, the 100 features show an informative, quantitative and broad view of cancer metabolism profiling.

This is the first large scale metabolic profiling project in NSCLC cell lines, and I was able to establish and reveal remarkable diversity and correlations by using 81 cell lines. In this chapter I will describe how I perform the assays, and the autonomous metabolic diversity revealed by the results.

Data validation shows a well-established, consistent and repeatable system

To optimize the stable isotope labeling assay for large scale experiments, I needed to first establish how much ^{13}C labeled tracer to use, and how long to treat cells with the tracers. In the beginning, two cell lines (H460 and Hcc366) with dramatically different growth rates were used to demonstrate that cells show substantial ^{13}C enrichment with 6 hour stable isotope labeling, no matter how fast the cell grows. Then I expand the scale to 8 cell lines to optimize experimental conditions, and the first thing is to choose a ^{13}C-glucose and ^{13}C-glutamine concentration for labeling test.

In regular RPMI medium, L-Glutamine is 2mM and D-Glucose is around 10mM. Usually we use the same concentration for ^{13}C labeled tracers to replace the unlabeled nutrient in growth medium. However, some cell lines consumed all the glutamine within 24 hours (data not shown) so I increased ^{13}C-glutamine labeling concentration to 4mM. For the stable isotope labeling assay, the carbon enrichment patterns changed with labeling time. As the TCA cycle progresses, cells have different turnover rate and distribute the ^{13}C-labeled tracer to divergent pathways. Most cell lines show significant carbon enrichment by 6 hours, then achieve steady-state enrichment pattern after 24 hours. I performed 6 hour and 24 hour labeling for all 81 cell lines.

Another important point of my assay is that all the biological replicates were performed in separate dishes on different days. That means, for every cell line, the ^{13}C-glucose and ^{13}C-glutamine labeling assays were processed in separate dishes and repeated on at least three different days to get the final three biological replicates. In most cell lines, the three biological replicates were collected within 2 weeks to one month, and the results remained highly consistent from dish to dish. In order to further ensure the data is repeatable and consistent, I re-tested the same cell line after several months, or tested cell lines that came from different sources. Fig 3.1a shows the similarity of the data sets, indicating the system is well-established and repeatable. In next chapter I will also describe the strong correlations among the data sets and use it as an evidence of this highly reliable system.

The following part of this chapter will show that how the metabolic phenotyping data reveal autonomous metabolic diversity of NSCLC cells. However, before showing the complete data set, we also analyzed the relative contribution of technical variation and biological variation on the complexity of the data. Fig 3.1b shows the standard deviation from replicates and standard deviation between cell lines are distinctively separated and without any overlap, indicating that the metabolic activity differences between cell lines came from real biology diversity, not only technique instabilities.

Stable isotope labeling assays disclosed astonish diversity between NSCLC cell lines

Stable isotope labeling involves the use of non-radioactive isotopes that can act as tracers. They can be taken by cells in multiple ways such as carbon source (^{13}C) or nitrogen source (^{15}N), and used to generate products like amino acids, nucleic acids, lipids, and other metabolites. These non-radioactive isotope tracers can be uniformly-labeled or specifically labeled on single/multiple positions. For example, the [U-$^{13}C_6$] glucose means all the 6 carbons of glucose are replaced by ^{13}C in this compound, while the [3,4-$^{13}C_2$] glucose have labeled ^{13}C at the third and fourth carbon positions of glucose. The isotopomers can be identified through equipment like mass spectrometry (MS), and used to analyze how many labeled tracers are incorporated into target metabolites.

Our lab uses stable isotopes to trace pathways connecting extracellular nutrients to intracellular metabolites. The best example is how glucose and glutamine metabolism contribute to TCA cycle. As the two most abundant nutrients *in vitro*, glucose is known to be a carbon source of TCA cycle metabolites, while the importance of glutamine anaplerosis came out much later (DeBerardinis et al., 2007). However, the full detail of how cells incorporate these two nutrients to TCA cycle remains unclear, while researchers kept discovering novel pathways through these years (Mullen et al., 2012). By systematically applying ^{13}C-glucose or ^{13}C-glutamine in large scale NSCLC cell line, I hoped to establish a broad view of how cell lines use both nutrients to fill TCA cycle. In this chapter I will focus on the diversity, and discuss the correlations in next chapter.

Appendix online file shows the carbon enrichment fractions of ^{13}C-glucose or ^{13}C-glutamine labeled metabolites such as Citrate, Malate, Fumarate, Serine, Glycine and Lactate. Each metabolite has different mass isotopomer distributions (MID) based on the carbon numbers. For example, Citrate MID is the most informative marker of metabolic flux because it converges carbon sources from glucose and glutamine pathways. Analyzing citrate MID could provide a relatively simple assess went of central metabolism. Figure 3.2 shows there are seven possible mass isotopologues for citrate from m+0 to m+6. M+0 means all carbons came from nature unlabeled ^{12}C, while m+6 means all six carbons came from ^{13}C labeled nutrient. The schematic shows the pathway and illustrates the common citrate labeling pattern. When using uniformly labeled glucose to label cells, ^{13}C glucose generates ^{13}C-labeled acetyl-coA through glycolysis

and enters TCA cycle as a two-carbon source. The ^{13}C acetyl-coA combines with unlabeled ^{12}C OAA in the first round and generates the major m+2 pattern of citrate from uniformly labeled glucose. In the other hand, uniformly labeled glutamine generates ^{13}C-labeled OAA as a four-carbon source for citrate, and forms the major m+4 pattern.

In order to show a clear view of the MIDs, we made the citrate MID data into dot plot and show as Fig 3.4. Although all cells used glucose and glutamine to produce citrate, the contributions of these two nutrients varied considerably (Fig 3.3), as did the specific pathways used. Each dot presents a single cell line and the enrichment diversity of glucose or glutamine labeled citrate, in 6 or 24 hour labeling time. For individual isotopic labeling patterns, differences of more than 20-fold were observed across the panel. The dot plots of other metabolites also show dramatic diversity between cell lines. (Fig 3.4 Fum, Mal, Ser)

We also use an unsupervised analysis to group the isotope labeling data. Fig 3.5 is the affinity propagation clustering (APC) of all carbon enrichment features, based on 6 hour or 24 hour data set. This cluster demonstrates the relationships among the metabolic features. This APC cluster reinforces the data set because there are several carbon enrichment features that are known to have close correlations stay in the same APC group. Fig 3.6 and 3.7 are the affinity propagation clustering of all NSCLC lines we use, based on their 6 hour or 24 hour carbon enrichment features. This is a novel way to

divide cell lines, and shows that cell lines belongs to different groups even with their intrinsic metabolic features.

Nutrient utilization rates data reflex how cells take central metabolic pathways differently

The second metabolic assay I performed is the nutrient utilization rates, which shows how cells consumption or secretion the major nutrients, such as glucose, glutamine, lactate and glutamate. Cells use glucose and glutamine for energy generation and biosynthesis, and secrete out lactate and glutamate as major byproducts which contribute to other metabolic pathways. These bulk metabolic rates have been reported to be under control of oncogenic signaling (Cheng et al., 2011), and therefore will provide additional, quantifiable markers of the metabolic phenotype.

In Fig 3.8a, individual cell lines were ranked for glucose utilization rates as black columns, and the lactate secretion rates were displayed in the same order as grey columns. The highly positive correlation of both was shown as dot plot also ($r=0.6$, $p=3.2e^{-09}$). Glutamine and glutamate are shown in Fig 3.8b as the low to high ranking and correlation ($r=0.62$, $p=6.5e^{-10}$). Among the cell lines, differences of more than 10-fold were observed across the panel. It is surprising that cells utilize these bulk metabolites distinctively, indicating the cells must take and use the central metabolic pathways differently.

Nutrient dependence assay data shows diverse addiction rate to nutrients

The third metabolic assay is to determine the cells addiction to glucose and glutamine. Cells were cultured in glucose or glutamine deprivation mediums for 1, 3 and 5 days, then DNA assays were used to determine the relative DNA contents compared to cells grow in complete medium. In Fig 3.9, individual cell lines were ranked for 3 day glucose deprivation as black columns, and the glutamine deprivation were displayed as corresponding grey columns. Across the panel, differences of more than 10-fold were observed. The cell growth rate can be examined from this assay by normalizing the DNA contents in complete medium on day 1 and day 3.

Discussion

In this chapter, I described a systematic analysis of glucose and glutamine metabolism in NSCLC cell lines, and discovered the highly variable metabolic phenotypes among them. This is the first large scale metabolism phenotyping project on NSCLC *in vitro*, proving that the core set of metabolic parameters, including ^{13}C enrichment and nutrient utilization, can be performed in large panels of cell lines.

With all the assays I performed, these NSCLC cells contain a substantial amount of cell-autonomous metabolic diversity although all the cell lines were cultured under

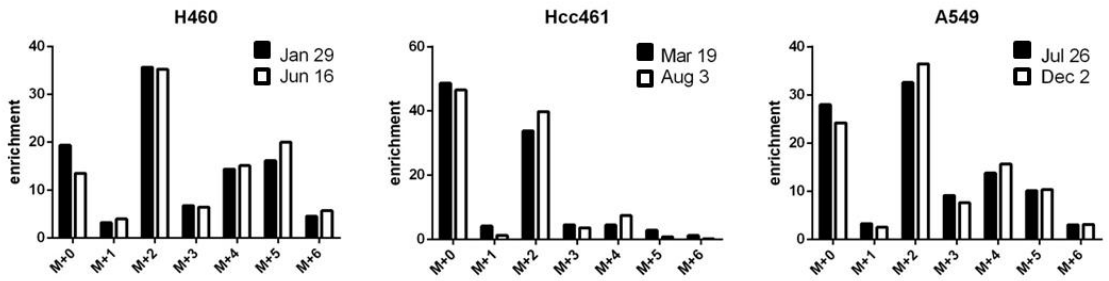
identical conditions. This is the first systematic evidence to indicate the existing of intrinsic cell metabolic diversities. Even the cell lines are separated from patients for many decades and adapted to cell culture condition, they still remain the autonomous metabolic diversity. This evidence makes cell culture a more confident mode for researchers to study the different genes and pathways.

Moreover, the stable isotope labeling assay data provides evidence for two questions that remain debating in cancer metabolism field. First, do cancer cells decrease TCA cycle activity in order to achieve Warburg effect? Second, does glutaminolysis play an important role in cancer cells as glycolysis? From our data, the answer to first question is no, and yes for the second question. Table 3.2 and Table 3.3 clearly show that all the NSCLC cell lines took both ^{13}C-glucose and ^{13}C-glutamine labeling, and the ^{13}C tracers can be detected in extensive TCA cycle metabolites. This result indicates that cells not only take the two nutrients into TCA cycle, but also use them to generate broadly metabolites through the panel. In other word, the TCA cycle remains active in cancer cells; also, both glutaminolysis and glycolysis are crucial for NSCLC cell lines.

In the ^{13}C enrichment data set, many major patterns could be explained with existed knowledge. For example I described the citrate m+2 from glucose, m+4 from glutamine in previous paragraph, and the lactate m+3 from glucose indicates that three ^{13}C carbons are transferred from glucose to lactate through glycolysis. However, there are many other heterogeneous within cell lines and many unexplained MID patterns though the panel. With further study, these heterogeneous metabolic phenotyping may provide

a foundation to uncover novel pathways, or interesting cell-autonomous metabolism in NSCLC.

a.



b.

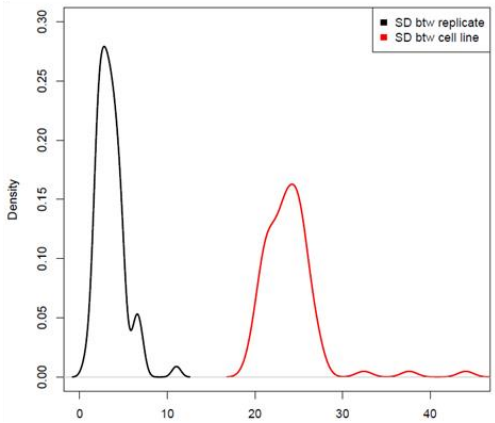


Fig 3.1 Data validation shows a consistent and repeatable system

(a) Stable isotopes labeling assay data on separate date. Cells were cultured in the medium with [u-^{13}C] glucose for 6 hours, and the carbon enrichment data of citrate were analysis. Notice that the performing dates were several months apart but the data remained consistent. (b) The standard deviation from biological replicates and standard deviation among cell lines are distinctively separated.

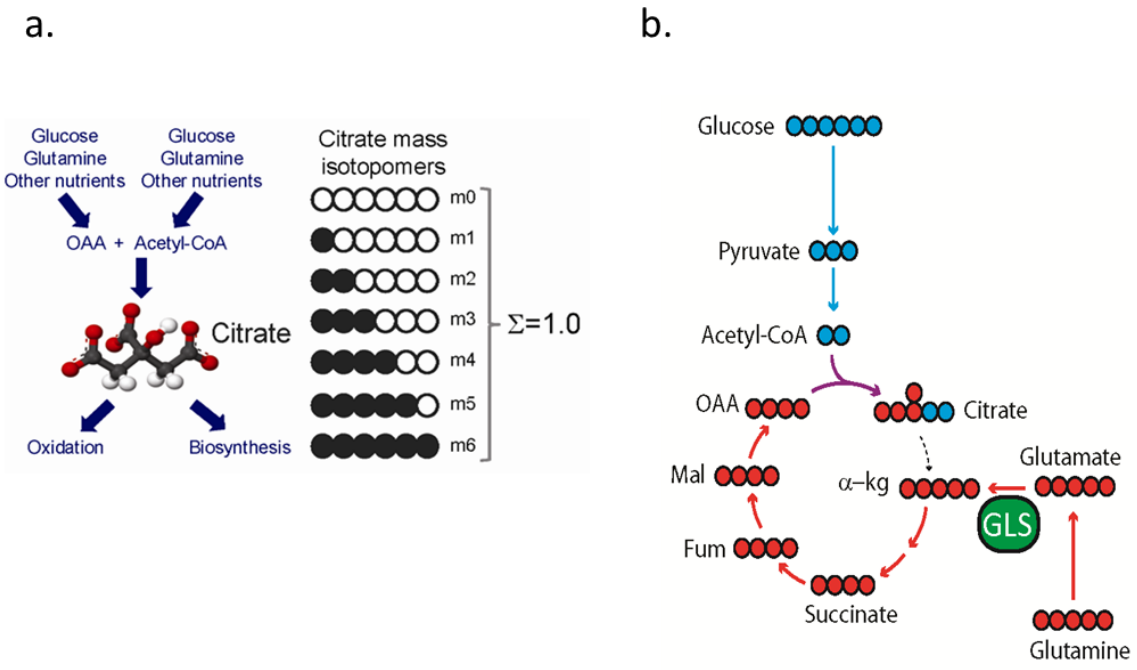


Fig 3.2 Tracing metabolic flux using stable isotopes

In stable isotope tracing analysis, cells are cultured in medium containing ^{13}C labeled nutrients, then metabolites are extracted and analyzed by mass spectrometry to analyze the mass isotopomer distribution (MID) in each metabolite. (a) Citrate MID. Citrate contains six carbon positions, each of which could be occupied by ^{12}C or ^{13}C; thus there are seven possible mass isotopomers of citrate (m0 to m6) after culture with ^{13}C-labeled nutrients. MID involves measuring the fractional contribution of each isotopomer, which by definition sum to 1.0. (b) Example of a metabolic phenotype analyzed by citrate MID. In this example, glucose supplied most of the carbon for acetyl-CoA (blue), whereas glutamine supplied oxaloacetate (red). Therefore, most of the citrate molecules in the population contain two carbons from glucose and four from glutamine. The contribution of unlabeled substrates is detected by suppression of ^{13}C enrichment in citrate.

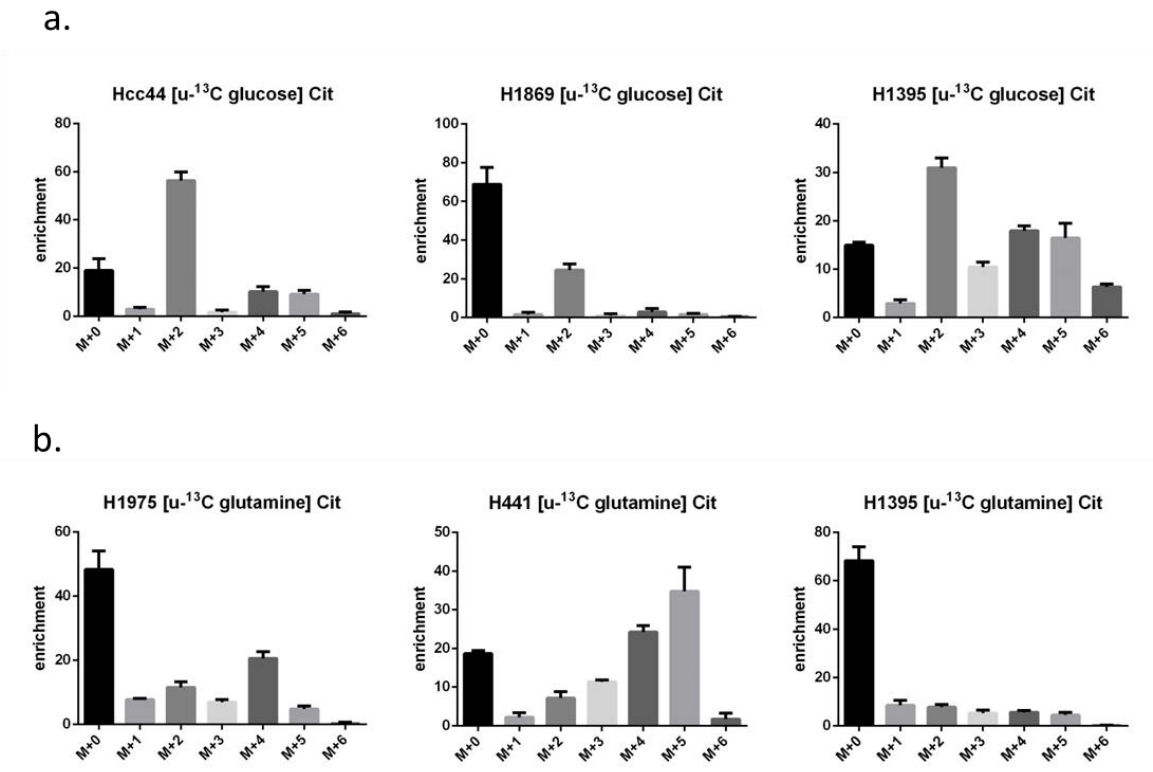
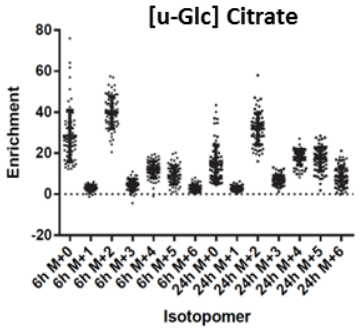
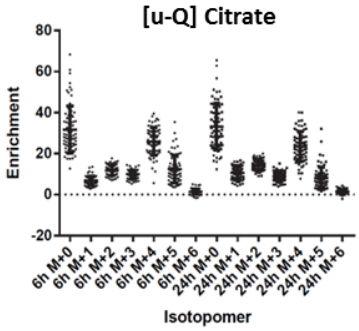


Fig 3.3 Heterogeneous MIDs revealed by stable isotopes labeling

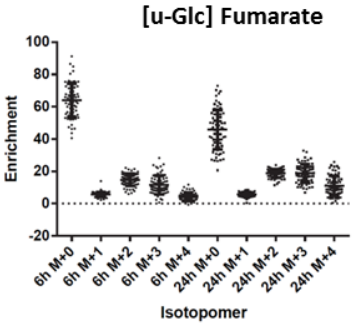
Different cell lines show distinctive citrate MID patterns by labeling with ^{13}C-glucose (a) or ^{13}C-glutamine (b) for 6 hours.

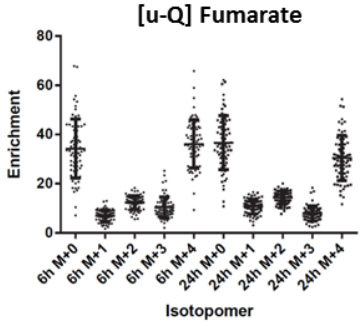
a.


[u-Glc] Citrate
Enrichment
Isotopomer
6h M+0
6h M+1
6h M+2
6h M+3
6h M+4
6h M+5
6h M+6
24h M+0
24h M+1
24h M+2
24h M+3
24h M+4
24h M+5
24h M+6

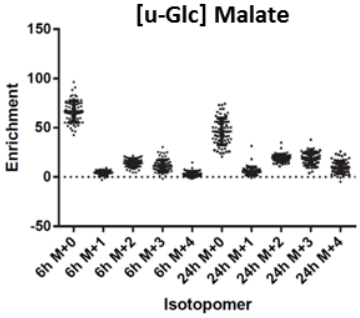

[u-Q] Citrate
Enrichment
Isotopomer
6h M+0
6h M+1
6h M+2
6h M+3
6h M+4
6h M+5
6h M+6
24h M+0
24h M+1
24h M+2
24h M+3
24h M+4
24h M+5
24h M+6

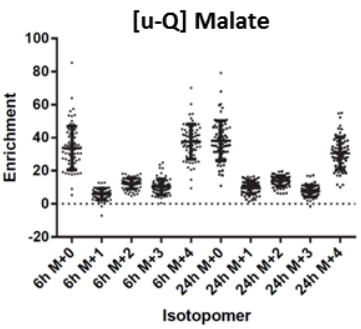
b.


[u-Glc] Fumarate
Enrichment
Isotopomer
6h M+0
6h M+1
6h M+2
6h M+3
6h M+4
24h M+0
24h M+1
24h M+2
24h M+3
24h M+4

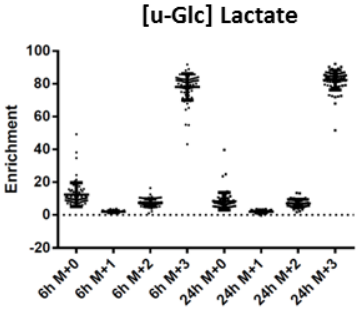

[u-Q] Fumarate
Enrichment
Isotopomer
6h M+0
6h M+1
6h M+2
6h M+3
6h M+4
24h M+0
24h M+1
24h M+2
24h M+3
24h M+4

c.


[u-Glc] Malate
Enrichment
Isotopomer
6h M+0
6h M+1
6h M+2
6h M+3
6h M+4
24h M+0
24h M+1
24h M+2
24h M+3
24h M+4


[u-Q] Malate
Enrichment
Isotopomer
6h M+0
6h M+1
6h M+2
6h M+3
6h M+4
24h M+0
24h M+1
24h M+2
24h M+3
24h M+4

d.


[u-Glc] Lactate
Enrichment
Isotopomer
6h M+0
6h M+1
6h M+2
6h M+3
24h M+0
24h M+1
24h M+2
24h M+3

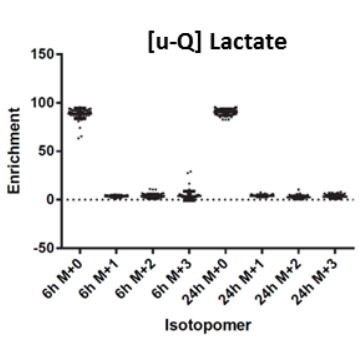
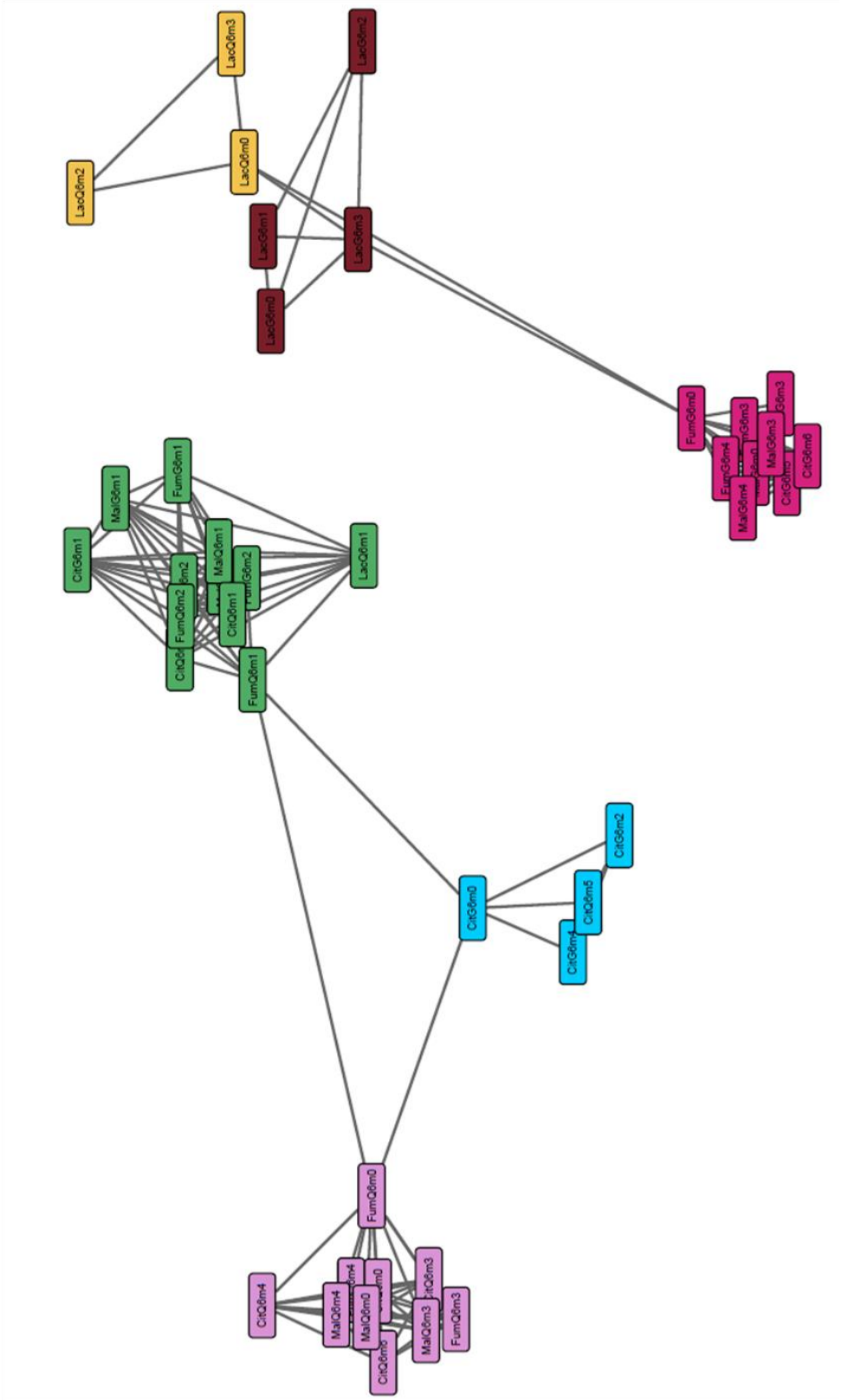

[u-Q] Lactate
Enrichment
Isotopomer
6h M+0
6h M+1
6h M+2
6h M+3
24h M+0
24h M+1
24h M+2
24h M+3

Fig 3.4 Carbon enrichment diversity in metabolites among NSCLC cell lines

The scatterplots show the isotope-derived carbon enrichment diversity of metabolites labeled with glucose (Glc) or glutamine (Q), in 6 or 24 hour. Each dot presents a single cell line. (a) Citrate (b) Fumarate (c) Malate (d) Lactate.

a.


LacQ6m3
LacG6m2
LacQ6m0
LacQ6m2
LacG6m1
LacG6m3
LacG6m0
FumG6m0
MalG6m3
CitG6m0
MalG6m4
FumG6m1
MalG6m1
CitG6m1
LacQ6m1
MalQ6m1
FumG6m2
FumQ6m2
CitQ6m1
FumQ6m1
CitG6m2
CitG6m0
CitQ6m5
CitG6m1
FumQ6m0
CitQ6m4
MalQ6m4
MalQ6m0
MalQ6m3
FumQ6m3
CitQ6m0

b.

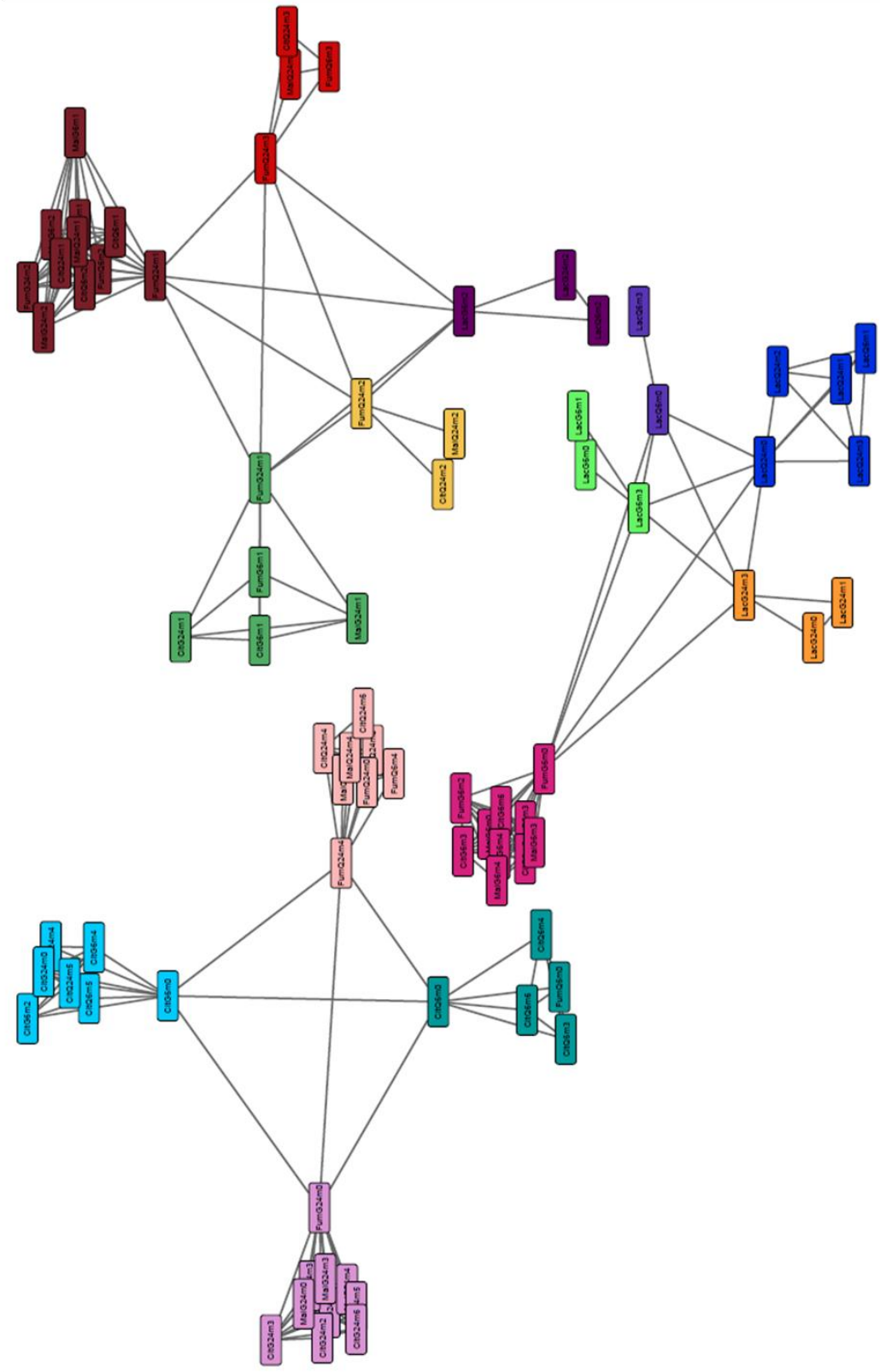
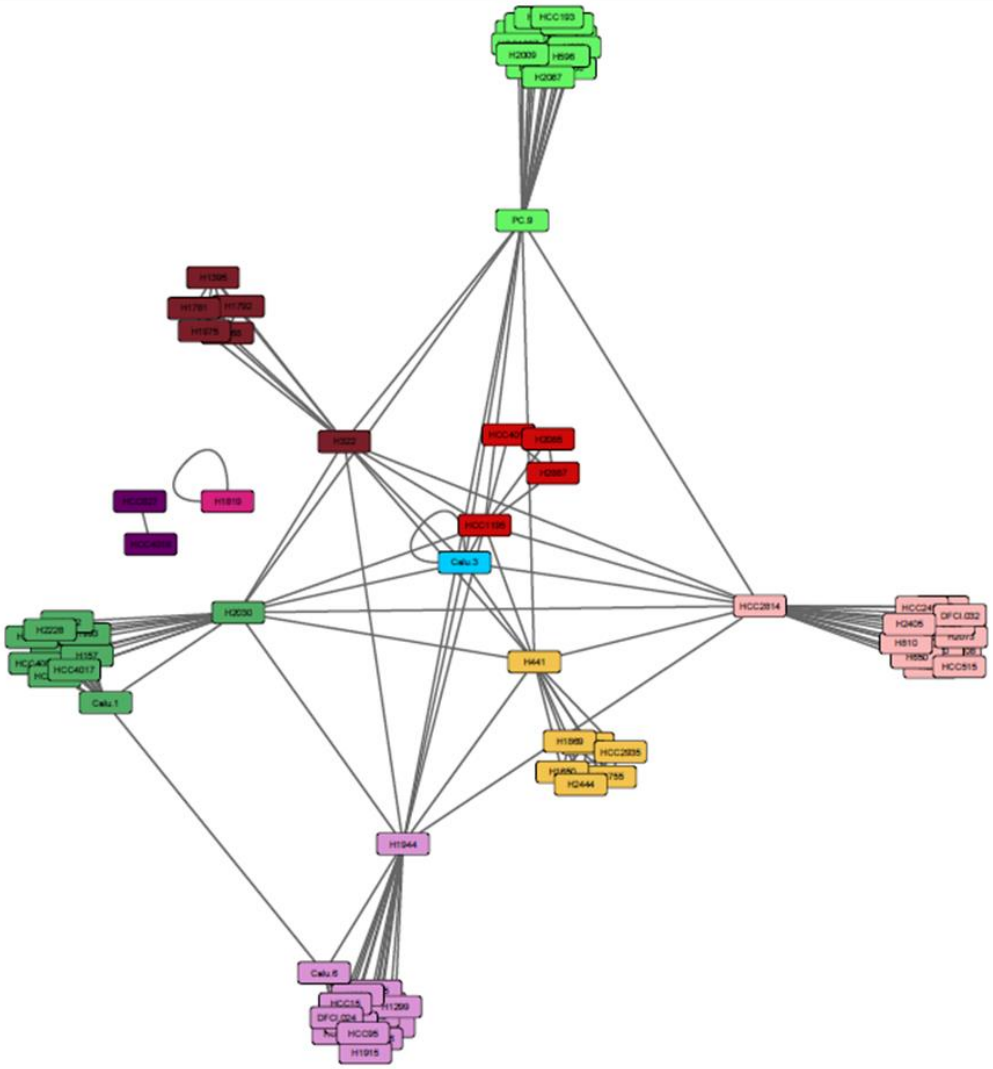


Fig 3.5 Affinity propagation clustering of carbon enrichment features

The affinity propagation clustering (APC) divides carbon enrichment features to multiple metabolic families, revealing the relationships among the metabolic features. (a) APC including 6 hours stable isotopes labeling data. (b) APC including both 6 and 24 hours stable isotopes labeling data.

a.



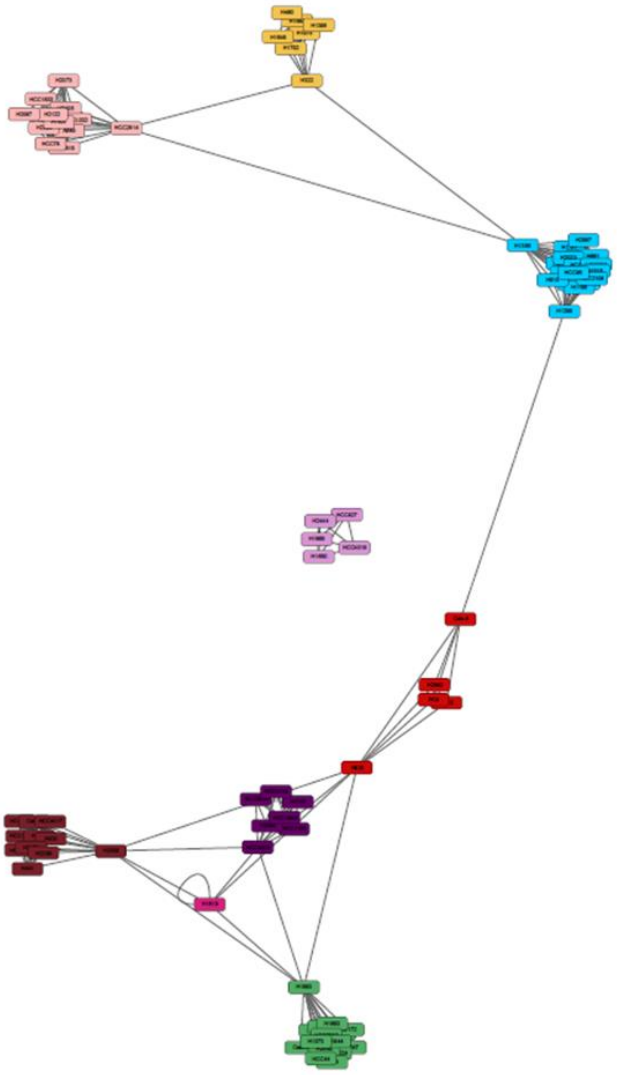
b.

group1	group2	group3	group4	group5	group6	group7	group8	group9	group10
Calu.3	H1819	Calu.6	Calu.1	H1395	H1650	H2085	A549	HCC4018	H1373
		DFCI.024	H157	H1568	H1755	H2887	DFCI.032	HCC827	H1693
		H1155	H1993	H1781	H1869	HCC1195	H1355		H2009
		H1299	H2030	H1792	H2444	HCC4019	H1437		H2087
		H1703	H2126	H1975	H441		H2023		H2170
		H1915	H2172	H322	HCC2935		H2073		H2258
		H1944	H2228		HCC461		H2405		H2291
		H2122	HCC1588				H460		H2347
		H23	HCC2302				H650		H2882
		H647	HCC4006				H810		H522
		H661	HCC4017				H838		H596
		HCC15	HCC78				HCC2108		H820
		HCC366					HCC2450		H920
		HCC44					HCC2814		HCC1833
		HCC95					HCC515		HCC1897
									HCC193
									PC.9

Fig 3.6 Affinity propagation clustering of NSCLC cell lines

The affinity propagation clustering (APC) divides NSCLC cell lines to multiple function families based on their 6 hours stable isotopes labeling data, revealing the relationships among the cell lines. (a) APC diagram. (b) The cell line lists in each family.

a.



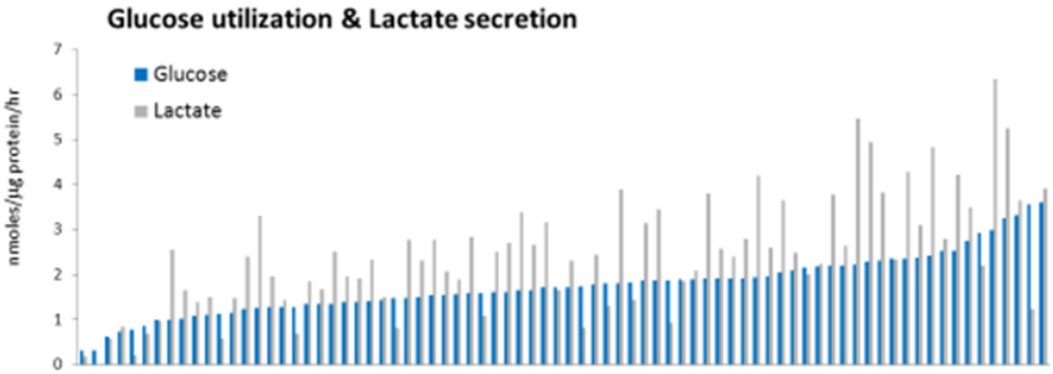
b.

group1	group2	group3	group4	group5	group6	group7	group8	group9	group10	group11
H1299	H1819	H1650	Calu.3	H1395	Calu.1	Calu.6	H2087	A549	H1755	HCC4018
H1355		H1869	DFCI.024	H1568	H2009	H2882	HCC1833	DFCI.032	H2085	
H1781		H2444	H1155	H1975	H2170	H522		H1437	H2228	
H1915		HCC827	H1373	H322	H2172	H920		H1792	HCC1195	
H2023			H157	H460	H2258	PC.9		H2073	HCC1588	
H2126			H1693		H2291			H2122	HCC4006	
H2887			H1703		H2347			H2405	HCC4019	
H650			H1944		H441			HCC2814		
H810			H1993		H820			HCC515		
H838			H2030		HCC1897			HCC78		
HCC2108			H23		HCC193					
HCC2302			H596		HCC2935					
HCC2450			H647		HCC461					
HCC95			H661							
			HCC15							
			HCC366							
			HCC4017							
			HCC44							

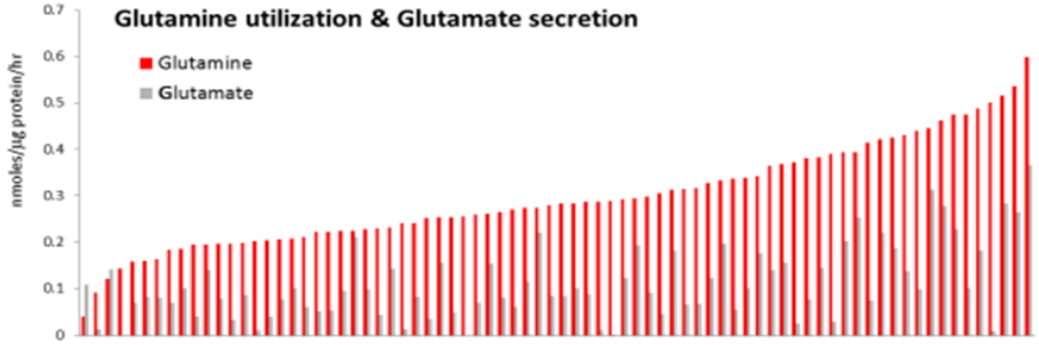
Fig 3.7 Affinity propagation clustering of NSCLC cell lines

The affinity propagation clustering (APC) divides NSCLC cell lines to multiple function families based on their 6 and 24 hours stable isotopes labeling data, revealing the relationships among the cell lines. (a) APC diagram. (b) The cell line lists in each family.

a.


Glucose utilization & Lactate secretion
Glucose
Lactate
nmoles/µg protein/hr
0
1
2
3
4
5
6
7

b.


Glutamine utilization & Glutamate secretion
Glutamine
Glutamate
nmoles/µg protein/hr
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

c.

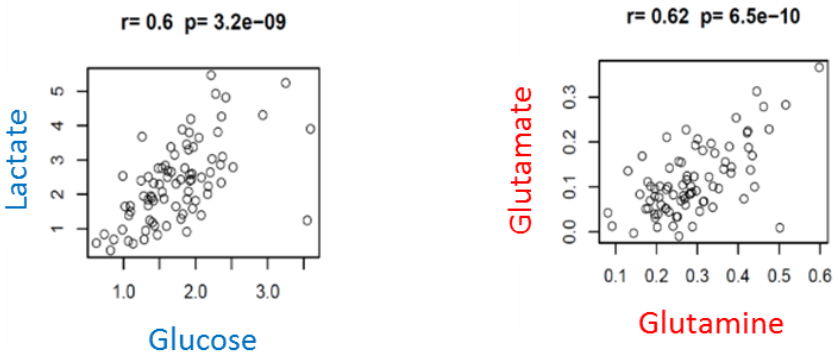

r= 0.6 p= 3.2e−09
Lactate
Glucose
1.0
2.0
3.0
1
2
3
4
5
r= 0.62 p= 6.5e−10
Glutamate
Glutamine
0.0
0.1
0.2
0.3
0.1
0.2
0.3
0.4
0.5
0.6

Fig 3.8 Diversity of nutrient utilization rates

For each cell line, the glucose, lactate, glutamine, and glutamate concentrations in culture medium were measured with a chemical analyzer (NOVA Biomedical BioProfile Basic4), then normalized by total protein content and cell culture time to calculate rates of utilization/secretion in nmoles/ug protein/hr. (a) Glucose utilization and lactate secretion. Individual cell lines were ranked (low to high) for glucose utilization rates, and the lactate secretion rates were displayed in the same order. (b) Glutamine utilization and glutamate secretion. Individual cell lines were ranked (low to high) for glutamine utilization rates, and the glutamate secretion rates were displayed in the same order. (c) The correlation between glucose utilization and lactate secretion (blue), and correlation between glutamine utilization and glutamate secretion (red).

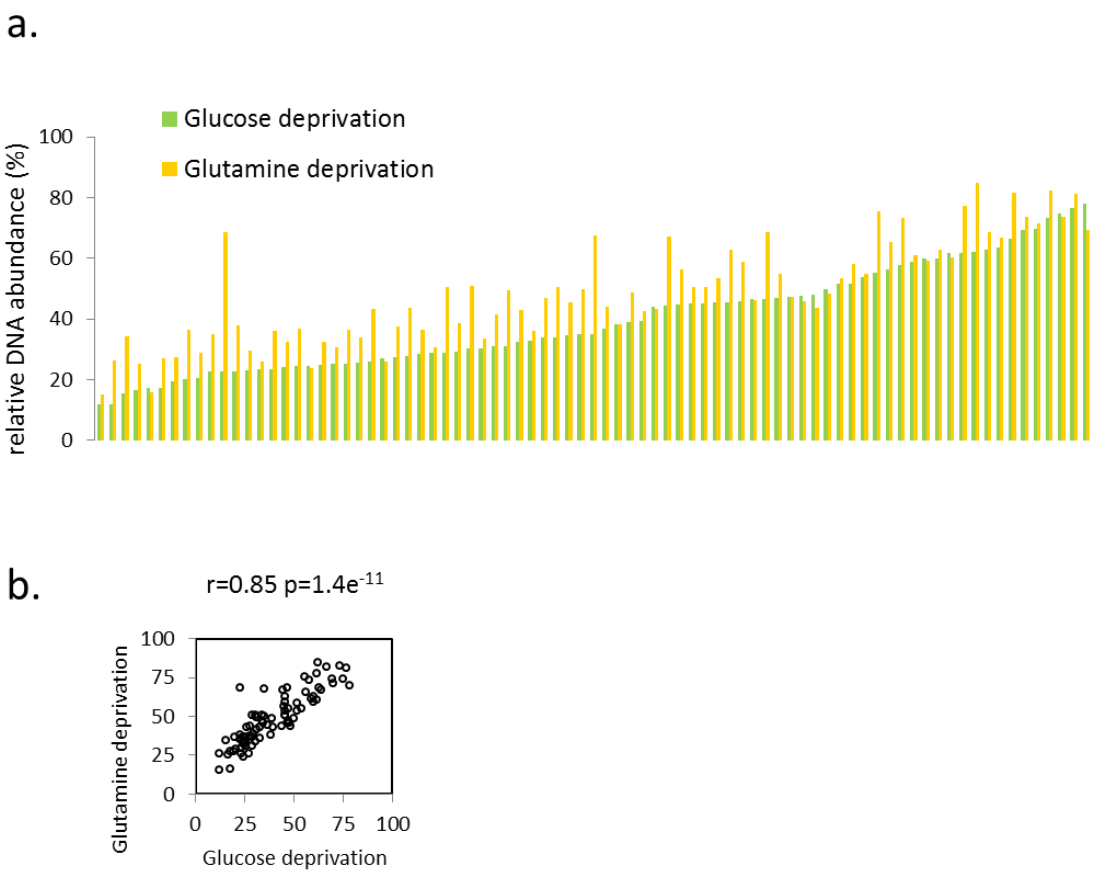


Fig 3.9 Diversity of nutrient dependence assay

Cell lines were cultured in media containing both glucose and glutamine, lacking glucose, or lacking glutamine. Dialyzed serum was used to ensure complete deprivation of each nutrient. After 1, 3, 5 days total DNA content was determined in each culture. DNA content in the glucose-free and glutamine-free cultures were normalized to DNA content in complete medium. (a) Results from the 3-day culture are shown for glucose deprivation (green column) and glutamine deprivation (yellow column) ranked low to high. (b) The correlation between these two parameters.

CHAPTER FOUR

Identify the correlation among cell-autonomous metabolic features

Introduction

Chapter three describes the cell-autonomous metabolic diversity of NSCLC cell as evidenced by stable isotope labeling, nutrient utilization and dependence assays, which poised further question such as: What useful information can be harnessed from these highly heterogeneous datasets? Do all NSCLC share common metabolic features? What association can we find between cells or different metabolic assays? Among the 100 metabolic features, which features positively or negatively correlate with each other? What are the underlying reasons behind the associations? And what can we learn from the correlations? Can we use them to explain, demonstrate or predict any metabolic activities?

The metabolic profiling datasets serve as a great tool. Combined with biostatistical power, it provides a platform for analyzing and answering the questions mentioned above. I want to acknowledge my lab mate Ling Cai, and Elizabeth McMillan from the Michael White lab for their help with bioinformatics, which made it possible to analyze the 100 metabolic features and determine the novel correlations revealed in this chapter.

Consensus cluster of carbon enrichment features reinforce the reliability of profiling data set

Clustering often is a first step in data analysis. Many different clustering methods have been developed to assign objects into groups, so that objects from the same cluster are more similar to each other than objects from different clusters (Gibbons and Roth, 2002). Consensus clustering, also called aggregation of clustering, yields a stable and robust final clustering that is in agreement with multiple clustering. It refers to the situation in which a number of different input clusters have been obtained for a particular dataset and it is desired to find a single clustering which is a better fit in some sense than the existing clusters (Abu-Jamous et al., 2013).

Fig 4.1 shows the consensus cluster of the 81 cell lines based on all their 6 hours or 24 hours carbon enrichment features. Even though all NSCLC cell lines grew under same culture condition, the cell lines segregated to different clusters based on their intrinsic metabolic differences. On the other hand, Fig 4.2 shows the consensus cluster between different carbon enrichment features on 6 hours and 24 hours. These features include the carbon enrichment fractions from ^{13}C-glucose or ^{13}C-glutamine labeled citrate, malate, fumarate and lactate. Based on the different carbon components in each molecule, there are 42 MID features. Features were grouped to four major clusters and the correlation between each other was shown by colors; red represents a positive

correlation whereas green represents a negative correlation. In following paragraphs I will further analyze the clustering, and indicate the important findings within them.

For the ^{13}C labeling data nomenclature, each labeling component has four components to the name: metabolite (Cit for citrate, Lac for lactate, Fum for fumarate and ser for serine), labeled nutrient (G for [U-^{13}C] glucose and Q for [U-^{13}C] glutamine), duration of labeling (6 hr or 24 hr) and isotopologue (m0, m1, m2, ... mn, where n = number of carbons in parent molecule). For example, SerG6m0 is fractional content of the m0 isotopologue after 6 hours of labeling in [U-^{13}C] glucose.

There are some carbon enrichment features that are known to be tightly associated and therefore cluster in the same vicinity (Fig 4.3). The most obvious and straight forward examples are the citrate carbon enrichment features. Citrate is a very abundant and informative metabolite in the TCA cycle. Citrate lies in the joint point between glucose and glutamine labeling pathways; therefore, the citrate MID patterns reflect the two tracers very well. When we supplement ^{13}C-glucose in culture medium, the carbon will be incorporated to citrate as a two carbon Acetyl-CoA form, resulted in a citrate Glc m+2 pattern. Conversely, when ^{13}C-glutamine is supplemented in culture medium, glutamine usually enters the TCA cycle as alpha-ketoglutarate to refill the four carbons molecule oxaloacetate, and contributes to citrate formation as the m+4 MID pattern. This pathway is known as the Glutamine-dependent anaplerosis pathway (DeBerardinis et al., 2008a). Therefore, we should see a strong correlation between Glc m+2 and glutamine m+4 in citrate. Not only the dot plot of all cell lines shows that (r=0.64,

$p=1.1e^{-10}$, Fig 4.3a), but also the consensus cluster reveals the same correlation (Fig 4.2, around middle group).

Another example of how different metabolic pathways affect carbon enrichment features is the Glutamine-dependent reductive carboxylation pathway, which contributes to the strong correlation between Glc m+0 and glutamine m+5 in citrate. In this pathway, ^{13}C-glutamine is the major source of carbon. Alpha-ketoglutarate is reductively carboxylated through the IDH reaction to generate citrate with a glutamine m+5 MID pattern, instead of the regular m+4 pattern generated in the conventional oxidation form of the TCA cycle. Glucose is not incorporated in this pathway so the major Cit MID from glucose is m+0, which results in a strong correlation in the dot plot ($r=0.75$, $p=4.4e^{-16}$, Fig 4.3b). Fumarate and Malate also get m+3 MID from glutamine instead of the usual m+4 pattern which results in tight clustering among cell lines. (Fig 4.2, bottom group).

Besides citrate, there are many other interesting associations in the consensus clustering. For example, the lactate m+3 from glucose clusters closely with lactate m+0 from glutamine, which is in agreement with the three ^{13}C carbons that are usually transferred from ^{13}C-glucose to lactate through glycolysis. In the ^{13}C glutamine labeling condition, most carbons came from unlabeled ^{12}C-glucose and made the lactate m+0 pattern. Moreover, in the cluster malate and fumarate are adjacent to each other because they can be reversibly converted by fumarase enzyme.

These strong correlations demonstrate the reliability of this platform to mine novel correlations. Additionally this result underscores the high quality of our data set and the robustness of the phenotype. Of technical importance and asserting to the stringency of this screen, assays were carried out in separate dishes, replicated on different days, and analyzed from distinct metabolites.

The full rank of Spearman correlation to target feature may identify unexpected relationships

The consensus cluster results focused on correlations between carbon enrichment features, and here we are going to show a broader view of the correlations between all 100 metabolic features that came from the three assays.

Spearman rank correlation (Spearman's rho) is a non-parametric test that is used to measure the degree of association between two variables. Rho is synonymous with the Spearman rank correlation coefficients, and it describes both the strength and the direction of the relationship. The rho value varies between +1 and -1, where 1/-1 is a perfect degree of association between the two variables and 0 is no degree of association. In our data set, rho values that exceed ± 0.2 were considered significant hits.

Ling Cai performed the Spearman rank correlation among the 100 features and put the detail together as appendix online file. It can be mined by selecting a feature of

interest and then ranking the correlations to other features based on the rho value. In this way, we were able to take a closer look of the full range of association of all features to one specific feature, and use it as proof of principle or uncover novel biological relationships. Fig 4.4 provides couple examples. As a proof of principle, when I ranked the full rho value correlated with citrate m+0 from glutamine (CitQ6m0), the most positive features were unlabeled glutamine m+0 from malate, fumarate and citrate; while the most negative features were glutamine labeled malate and fumarate. Similarly, when I ranked the associations of all features to serine m+3 from glucose (serG6m3), features with the highest rho values were labeled serine and glycine while the unlabeled serine and glycine locates had the most negative rho values. Interestingly, when I obtained the full rank for lactate/glucose ratio, lactate secretion correlated strongly with it, but cell growth rate does not. It is interesting because glycolysis and lactate production are thought to be used for cell proliferation. In this way, we might be able to identify unexpected relationships.

The Spearman rank correlation of all features reveals surprising biological relationships

Tumor cells exhibit high levels of glycolysis despite the presence of ample oxygen, a phenomenon termed aerobic glycolysis, or Warburg effect. A universal property of primary and metastatic cancers is upregulation of glycolysis, resulting in increased

glucose consumption compared to normal tissues (Gambhir, 2002). It is widely accepted that in proliferating cells, the high glycolytic rate provides advantages such as energy generation and biosynthesis. However, there is still much to learn about how proliferating cell metabolism is regulated (DeBerardinis et al., 2008a). For example, does the nutrient uptake rate directly stimulate cancer cell growth? How do cells take glucose and glutamine? Will the higher nutrient consumption generate higher fluxes of glycolysis and glutaminolysis? This metabolic profiling data set will provide answers to these unclear basic questions.

Fig 4.5 shows the Spearman correlation coefficients in a heatmap grid version. This figure provides insight into how glucose and glutamine metabolism are related to each other, and with other metabolic phenotypes. Red color indicates a positive correlation while blue represents a negative correlation. There are many strong correlations between fumarate, malate and citrate isotopomer data, but only a few strong associations between nutrient utilization, nutrient dependence, and growth rate. These lack-of-association observations answer some important questions which will be addressed in following scatterplot figures.

There are some interesting correlations emerged when analyzed nutrient utilization. Fig 4.6a shows a positive association between glucose and glutamine utilization ($r=0.43$, $p=6.9e^{-05}$). Almost all cell lines use both nutrients, emphasizing the importance of glutamine. Moreover, the positive r value indicates that cells that consume more glucose usually consume more glutamine; they synchronize the utilization of the two

nutrients, rather than consuming cheat the expanse of the other. When we compared nutrient uptake with cell proliferation, surprisingly neither glucose nor glutamine utilization had significant association with growth rate. Fig 4.6b shows the p value of 0.8 and 0.87, indicating that most NSCLC cells spend the nutrition not only on growth.

When we compared nutrient uptake with ^{13}C isotope labeling, we found that more glucose utilization does not indicate higher carbon enrichment from ^{13}C glucose in citrate. Fig 4.6c shows a lack of correlation ($p>0.5$) between glucose utilization and citrate m+0 and m+2. Compared to glucose uptake, glutamine utilization significantly correlated to carbon enrichment from ^{13}C glutamine m+0 and m+4 in citrate (Fig 4.6d, $p < 0.005$). These observations are interesting because the high glycolysis rate were thought to be correlated with more nutrient consumption and faster cell growth in cancer cells.

Pyruvate carboxylase activity can be predicted by malate enrichment feature

Besides uncovering novel biological relationships, we wanted to use this data set to identify therapy-response, predictive biomarkers. The next paragraph is an example of how we predicted an enzyme activity and cell vulnerability by a metabolic feature.

Pyruvate carboxylase (PC) is a ligase which catalyzes the carboxylation of pyruvate to form oxaloacetate (OAA) which can then be funneled into the TCA cycle. In the Glucose-dependent anaplerosis pathway (Fig. 4.7a), cells use glucose as the major carbon source

to refill and form the majority TCA cycle metabolites. This allows cells to use glucose-derived pyruvate rather than glutamine for anaplerosis (Cheng et al., 2011). The usual way to measure PC activity is labeling cells with [3,4-$^{13}C_2$] glucose and check the m+1 enrichment on citrate. However, it would be useful to find a biomarker with universal-^{13}C glucose labeling condition so that we can apply our dataset. Malate carbon enrichment may be a good candidate because it derives from OAA as shown on the pathway diagram.

With the help of my lab mate Jiyeon Kim, we used around 20 NSCLC lines and compared their malate carbon enrichment (Mal m+3 from U-glucose labeling) with the PC activity (Cit m+1 from 3,4-^{13}C glucose labeling). We found Mal m+3 to be significantly correlated with Cit m+1 (Fig 4.7b) ($r=0.416$, $p=0.002$) which demonstrates that it can be used as a PC activity marker. Importantly, PC activity is associated with cell vulnerability. We picked 6 cell lines with high (Hcc515, H1792, H1648) and low PC activity (H920, PC9, H2444), knocked down PC and test their growth by colony formation assay. Cells with high PC activity displayed significantly lower colony formation when PC was silenced whereas cells with low PC activity tolerated PC silencing with no effect on colony formation (Fig 4.7c).

Discussion

In this chapter I described how we analyzed the correlation between cell-autonomous metabolic features, and uncovered novel biological relationships. This chapter also demonstrates how the phenotyping data set can be used to explain, demonstrate or predict many metabolic activities. Moreover, the consensus cluster result indicates many strong correlations which support and reinforce the reliability of this data set.

The Spearman rank correlation of all features also reveals surprising biological relationships. Some critical correlations increase our knowledge of how cells use glucose and glutamine. First, cell lines that consume more glucose usually consume more glutamine. Second, neither glucose nor glutamine utilization has significant association with cell growth rate. Third, higher glutamine utilization rates indicate higher carbon enrichment in citrate from ^{13}C glutamine, but not in glucose. The most possible explanation of these lack-of-association observations may simply because the cells use the nutrients in other ways, not only for proliferation.

Considering the highly positive correlations between glucose consumption and lactate secretion, a large part of glucose goes to lactate instead of entering TCA cycle, causing the low correlation between glucose uptake and ^{13}C glucose enrichment on citrate. Glutamate, as the major byproduct of glutamine, is an important amino acid for biosynthesis, gluconeogenesis and deamination. It is also a neurotransmitter that plays principal role in neural activation. Cells may consume glutamine to generate glutamate

and serve on different functions, not only cell growth. There is still much to learn about how proliferating cell metabolism is regulated. Despite a long and rich history of research, the complex connection between metabolism and proliferation remains an exciting area of investigation (Vander Heiden et al., 2009).

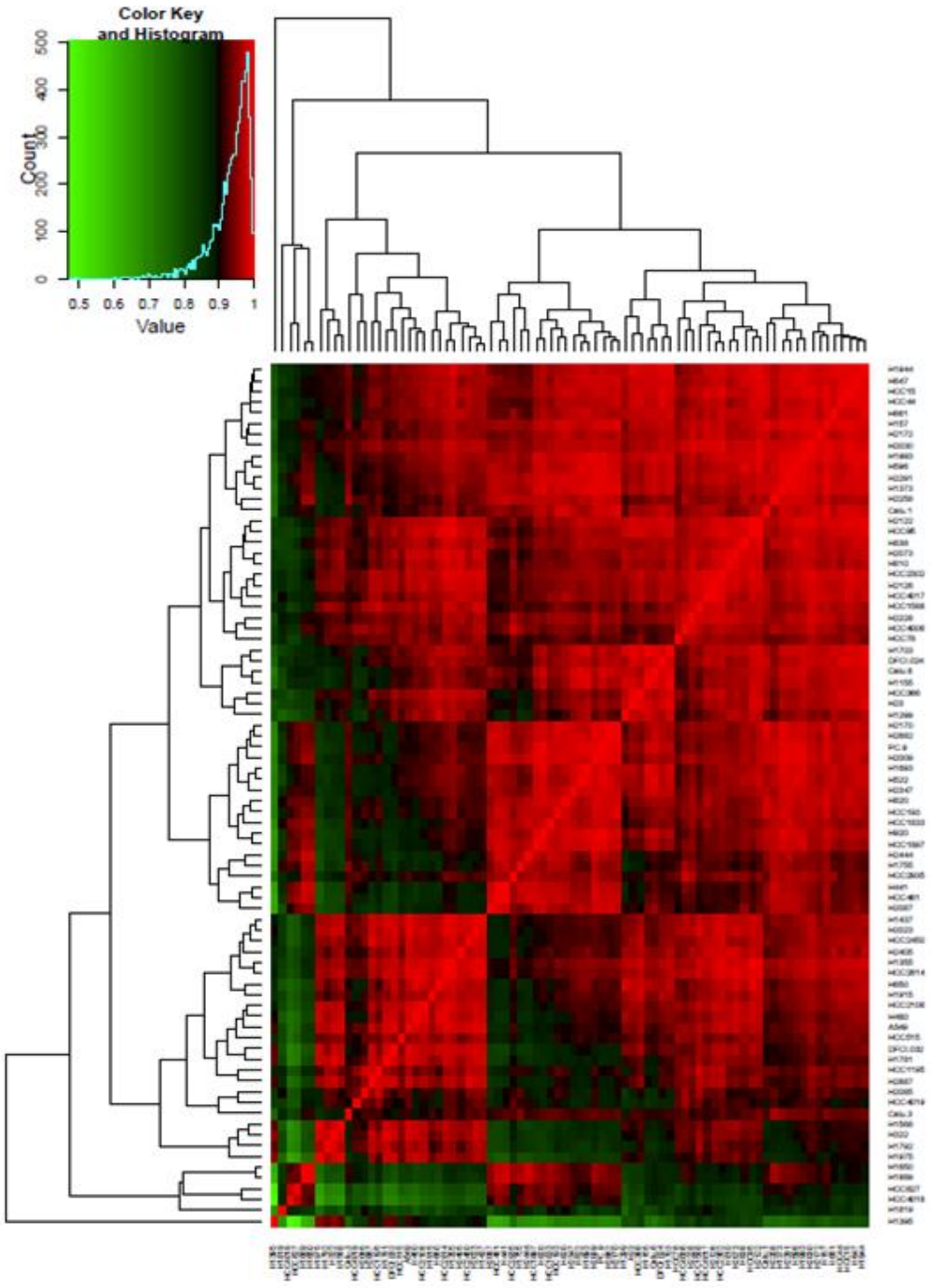


Fig 4.1 Consensus clusters of NSCLC cell lines

Consensus cluster of the 81 cell lines based on their 6 hours or 24 hours carbon enrichment features.

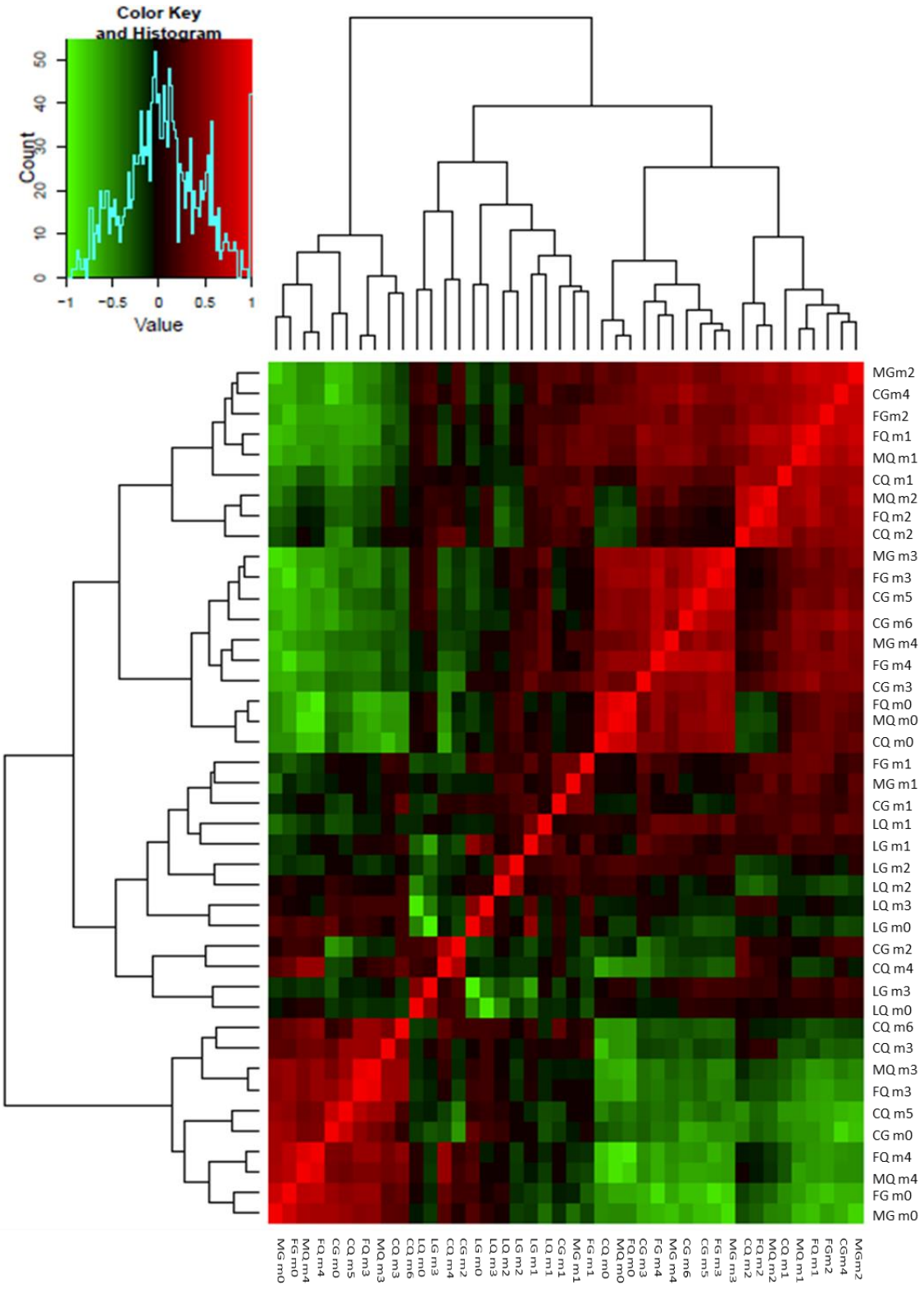

Color Key
and Histogram
Count
Value
MGm2
CGm4
FGm2
FQ m1
MQ m1
CQ m1
MQ m2
FQ m2
CQ m2
MG m3
FG m3
CG m5
CG m6
MG m4
FG m4
CG m3
FQ m0
MQ m0
CQ m0
FG m1
MG m1
CG m1
LQ m1
LG m1
LG m2
LQ m2
LQ m3
LG m0
CG m2
CQ m4
LG m3
LQ m0
CQ m6
CQ m3
MQ m3
FQ m3
CQ m5
CG m0
FQ m4
MQ m4
FG m0
MG m0

Fig 4.2 Consensus cluster of carbon enrichment features

Consensus cluster of mass isotopomer distribution features based on 6 hours stable isotopes labeling data. These features include the carbon enrichment fractions from ^{13}C-glucose (G) or ^{13}C-glutamine (Q) labeled citrate (C), malate (M), fumarate (F) and lactate (L).

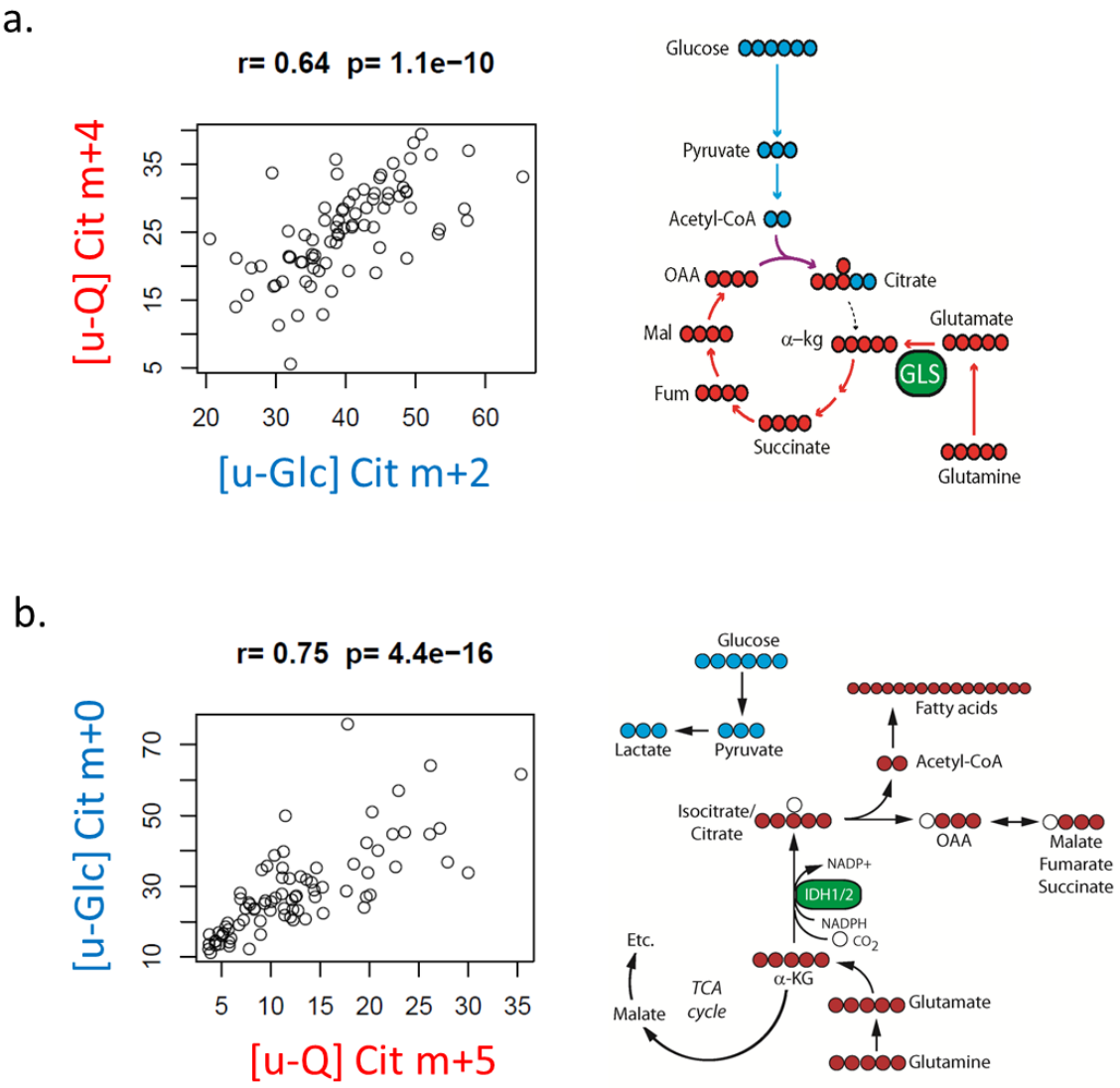


Fig 4.3 Different metabolic pathways revealed by MID analysis

Tracing metabolites and analysis their specific MID patterns may reflect the different pathways. (a) Glutamine-dependent anaplerosis pathway can be revealed by the positive correlation between citrate m+2 from ^{13}C-glucose and m+4 from ^{13}C-glutamine. Each dot presents individual cell line. (b) Glutamine-dependent reductive carboxylation pathway can be revealed by the positive correlation between citrate m+0 from ^{13}C-glucose and m+5 from ^{13}C-glutamine.

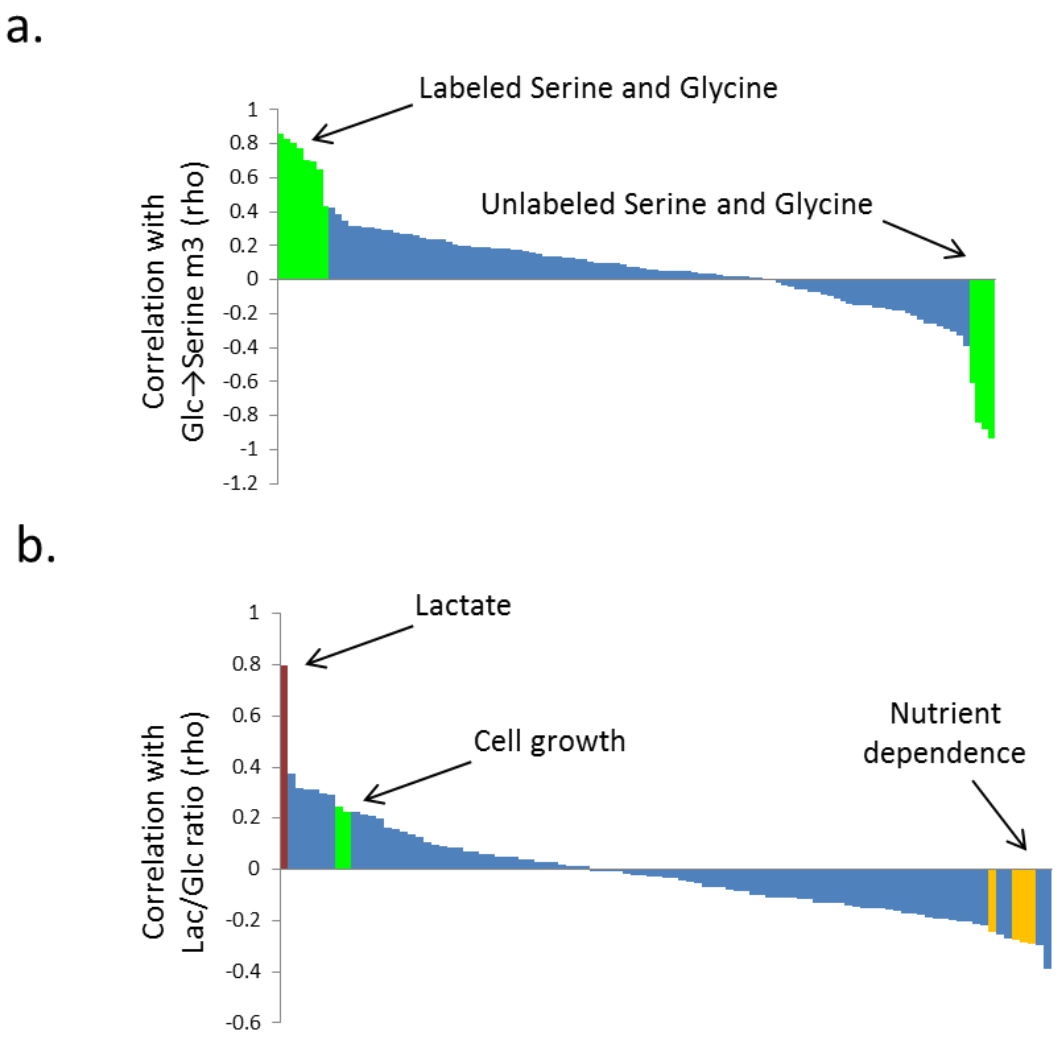


Fig 4.4 The full rank of Spearman correlation to target features

(a) Spearman correlation among all metabolic features and the ^{13}C-glucose labeled serine m+3 carbon enrichment. (b) Spearman correlation among all metabolic features and the lactate secretion/glucose utilization ratio.

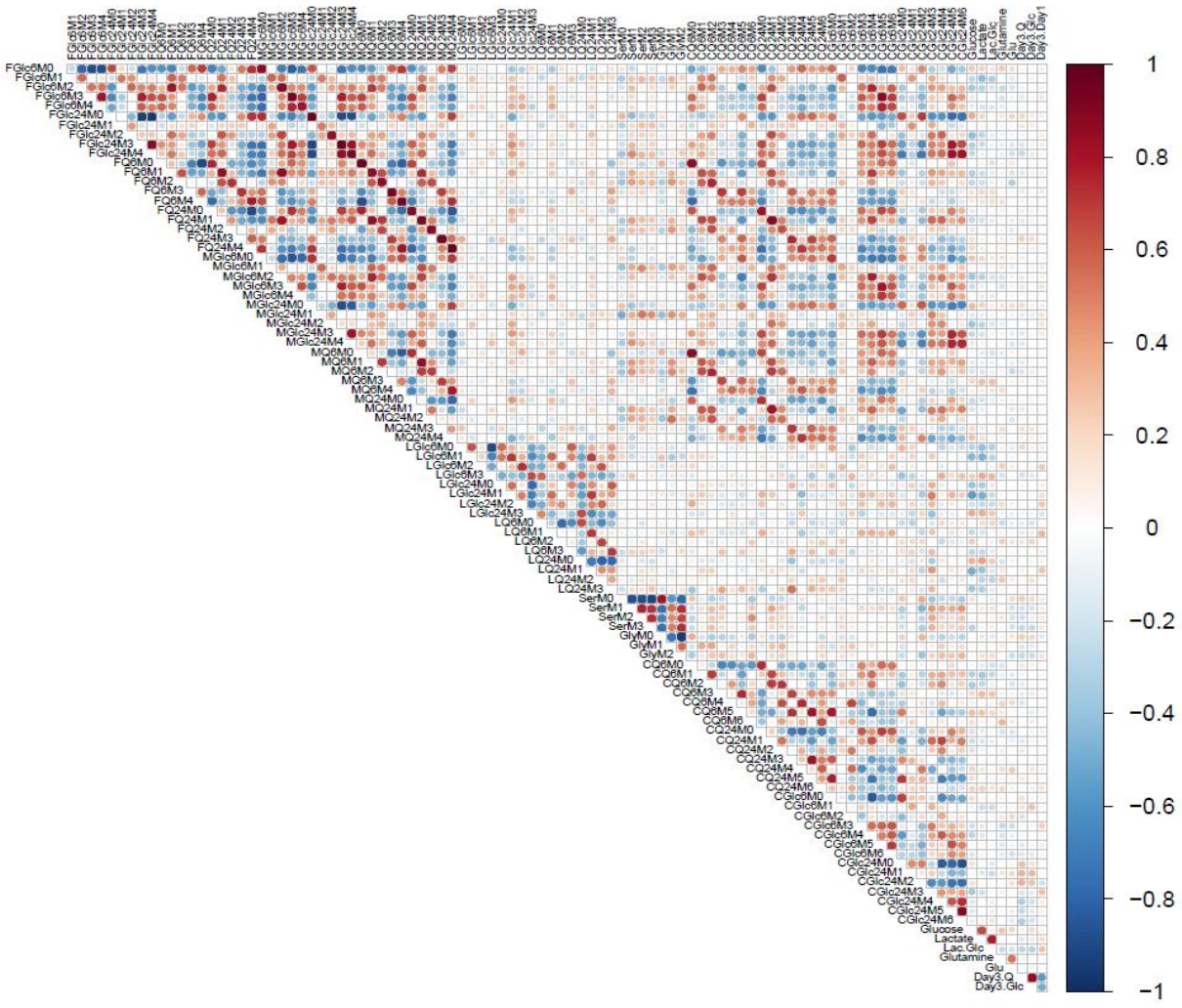
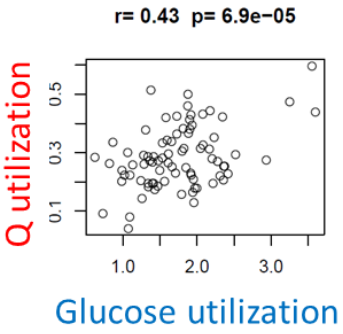


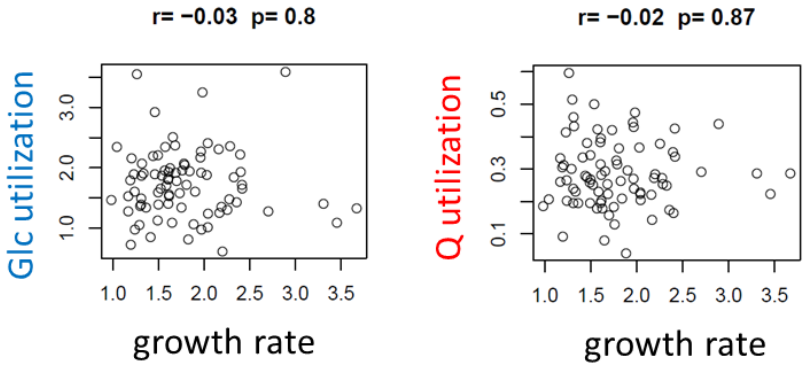
Fig 4.5 Spearman's rank correlation coefficients plot of all metabolic features

Spearman correlation coefficients were calculated for all pairwise correlations among 100 metabolic features. These correlation coefficients were plotted into a heatmap. Red dots stands for positive and blue dots stands for negative correlation. Rho values bigger than 0.2 are significant. Highly significant positive and negative correlations are seen in the correlation among mass isotopomer fractions from the same metabolite or from metabolite proximal to each other in metabolic pathways.

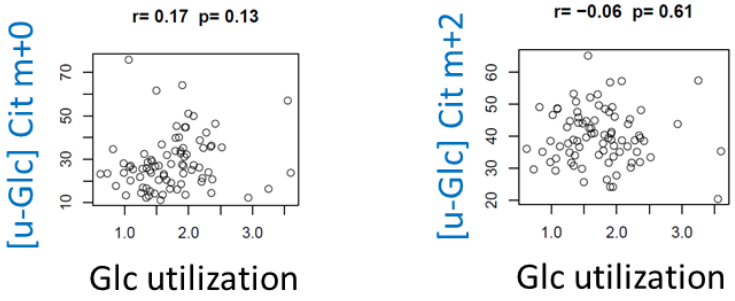
a.


r= 0.43 p= 6.9e−05
Q utilization
Glucose utilization

b.


r= −0.03 p= 0.8
Glc utilization
growth rate
r= −0.02 p= 0.87
Q utilization
growth rate

c.


r= 0.17 p= 0.13
[u-Glc] Cit m+0
Glc utilization
r= −0.06 p= 0.61
[u-Glc] Cit m+2
Glc utilization

d.

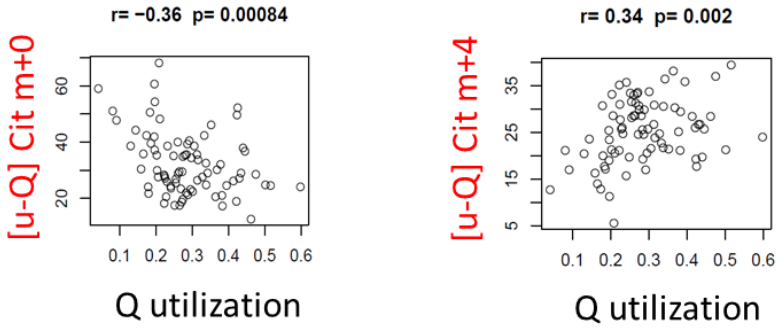

r= −0.36 p= 0.00084
[u-Q] Cit m+0
Q utilization
r= 0.34 p= 0.002
[u-Q] Cit m+4
Q utilization

Fig 4.6 The correlation among features reveal novel biological relationships

Pearson correlation coefficient and P-value was shown for the pairwise correlation between two metabolic parameters. Each dot presents individual cell line. (a) Positive correlation between glucose and glutamine utilization. (b) Insignificant correlation among nutrient utilization rates and cell growth rate. (c) Correlation between carbon enrichment from glucose and glucose utilization. (c) Correlation between carbon enrichment from glutamine and glutamine utilization.

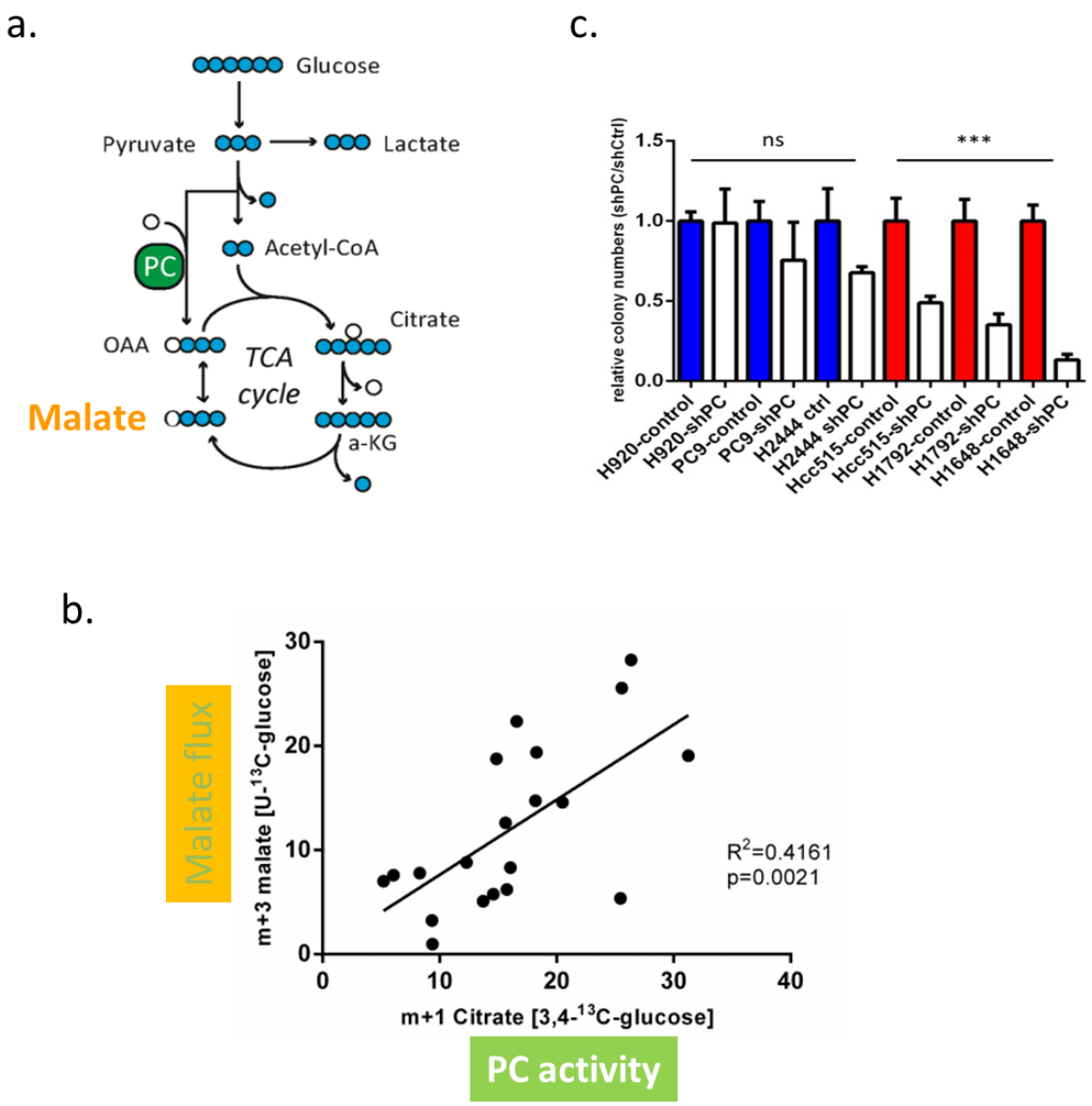


Fig 4.7 Pyruvate carboxylase dependency can be predicted by malate enrichment feature

(a) Pyruvate carboxylase pathway diagram (b) correlation between malate m+3 from [u-^{13}C] glucose labeling and Citrate m+1 from [3,4-$^{13}C_2$] glucose labeling. (c) Colony formation in soft agar of PC knockdown cells. Blue bars present cell lines with low PC activity while red ones are high PC lines. ***$p < 0.005$.

CHAPTER FIVE

Correlations between metabolic features and orthogonal datasets

Introduction

There are many orthogonal datasets available for the lung cancer lines I used for metabolic profiling, including clinical data, high-through put screening data, drug sensitivity data and various molecular data sets such as gene expression and targeted genotyping. Through the use of bioinformatics combined with biostatistics analysis, we sought to identify correlations between the datasets to identify novel determinations of metabolic preferences. In this chapter I will explain how we identified correlations between hypoxia, Myc and the glucose derived metabolic features by comparing the metabolic profiling data set with the Gene Set Enrichment Analysis data set. Additionally, we will introduce the association between KRAS-LKB1 co-mutation with metabolic features. The analysis of orthogonal datasets provides clues to molecular basis underlying metabolic phenotypes.

Gene Set Enrichment Analysis (GSEA) is one of the important tools we use. While the genomewide RNA expression analysis has become a routine tool in biomedical research, the challenge no longer lies in obtaining gene expression profiles, but rather in interpreting the results to gain insights into biological mechanisms (Schena et al., 1995).

The GSEA is a powerful analytical method for interpreting gene expression data. The method derives its power by focusing on gene sets, that is, groups of genes that share common biological function, chromosomal location, or regulation. Notably, where single-gene analysis finds little similarity between two independent studies, GSEA reveals many biological pathways in common. The GSEA method is embodied in a freely available software package, together with an initial database of 1,325 biologically defined gene sets (Chabner and Roberts, 2005; Mootha et al., 2003).

Background: Hypoxia and Myc regulates glycolysis

Glycolysis is the metabolic pathway that converts glucose into pyruvate. Pyruvate is generated in the final step of glycolysis, and then can either be converted into acetyl-CoA by pyruvate dehydrogenase (PDH) for further metabolism in the TCA cycle, or be converted into lactate by Lactate dehydrogenase (LDH). Two key transcription factors that play major roles in this metabolic reprogramming are hypoxia inducible factor-1 (HIF1) and MYC (Zwaans and Lombard, 2014).

Under hypoxic conditions, HIF-1 activates the transcription of the PDK1 gene encoding PDH kinase, which PDK1 phosphorylates and inactivates PDH complex, thereby inhibiting the conversion of pyruvate to acetyl-CoA (Kim et al., 2006 and Papandreou et al., 2006). HIF-1 also activates transcription of the LDHA gene encoding lactate dehydrogenase A, which catalyzes the conversion of pyruvate to lactate (Semenza et al.,

1996). Thus, the combined effect of PDK1 and LDHA is to decrease flux from pyruvate to acetyl-CoA and increase flux from pyruvate to lactate (Zeng et al., 2015). The breakdown of glucose into lactate provides cancer cells many advantages like ATP generation, ROS maintenance, and the production of intermediates that serve as building blocks to support rapid proliferation.

The MYC oncogene, which contributes to the genesis of many human cancers, is considered a master regulator of metabolism for cell growth and proliferation (Eilers and Eisenman, 2008). MYC positively regulates ribosomal biogenesis, glucose metabolism, and mitochondrial respiration in most cell types (Dang et al., 2009). Myc regulates expression of genes either directly, such as LDHA, a glycolytic gene, or indirectly, such as glutaminase (GLS) protein expression through repression of microRNAs miR-23a/b (Dang et al., 2009) . A study for a cell type-independent Myc core target gene signature revealed a 50-gene signature that is highly enriched for genes involved in RNA processing and ribosomal biogenesis (Ji et al., 2011). Also, myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis (Boon et al., 2001).

Lactate/glucose ratio correlates with gene sets relative to hypoxia, Myc and ribosome biogenesis

Using the metabolic profiling data and GSEA gene expression data set, we scored previously defined gene sets based on how the members in the gene set correlated with our metabolic features. We found that several gene sets that highly correlate with the Lactate/Glucose (Lac/Glc) ratio feature are involved in hypoxia and Myc activation. The Lac/Glc ratio feature was determined by measuring the rates of glucose utilization and lactate secretion separately in the NSCLC cells. Since lactate is generated through glycolysis from glucose, this result is reassuring as it is well known that hypoxia and Myc can enhance glycolysis.

Fig 5.1a shows the full range of Lac/Glc ratios among NSCLC cell lines. Theoretically a glucose molecule with six carbons can generate two three-carbon molecules of lactate. In the Lac/Glc ratios figure most of the ratios varied from one to two. Cell lines with a smaller Lac/Glc ratio may use more glucose for other pathways or use more lactate intracellularly instead of secreting out. On the other hand, cell lines with higher Lac/Glc ratios may have increased generation of lactate, like pyruvate from malate cycle.

Fig 5.1b shows the GSEA hypoxia gene set analysis. This WINTER_HYPOXIA metagene set contains 242 genes regulated by hypoxia. This dataset was obtained by analysis of genes whose in vivo expression clustered with the expression of ten well-known hypoxia-regulated genes (Winter et al., 2007). In order to validate this result, I choose nine cell lines with high (Hcc827, DHCI-024, H2347, H358 and Hcc515) and low (H2172, H1755, Hcc4018 and H661) Lac/Glc ratio, and check their expression of key hypoxia target genes by qPCR. Most lines with high Lac/Glc ratio have higher expression of many

top target genes such as ADORA2B and SLC2A1 (Fig 5.1c). LDH is not a top target gene in this gene set and it shows no difference between the high/low ratio lines.

Interestingly, the ribosome gene set also correlated with the Lac/Glc ratio feature. In Fig 5.2a, this KEGG_RIBOSOME metagene set contains 88 genes from KEGG human ribosomal RNA and protein component. For validation, I checked expression of the ribosomal protein S6 (RPS6) in the high and low Lac/Glc ratio cell lines by western blot since RPS6 expression directly correlates with glucose homeostasis (Ruvinsky et al., 2005). Fig 5.2b show that most lines with high Lac/Glc ratios expressed higher level of phospho RPS6 while most lines with low Lac/Glc ratios express much lower p-RPS6.

Correlations between Lac/Glc ratio with hypoxia, Myc and ribosome biogenesis gene expressions reinforce the utility of the metabolic profiling data set. HIF is widely reported to drive glycolysis, lactate production, and to regulate genes involved in glucose metabolism through Myc (Dang et al., 2008). Also, c-Myc regulates genes involved in the biogenesis of ribosomes and mitochondria. Therefore it is not surprising to see the correlations among Lac/Glc ratio with hypoxia, Myc and ribosome biogenesis by comparing GSEA and metabolic profiling and raises the hope that additional novel association will be identified using this method. These examples demonstrate how the orthogonal datasets provide clues to molecular basis underlying metabolic phenotypes.

Background: KRAS and LKB1 regulates tumor metabolism

Human Ras proteins are small GTPases. The three human RAS genes (HRAS, NRAS, and KRAS) encode four highly homologous small GTPases. Ras proteins are identified in 25% of human cancers (Bos, 1989), and KRAS is very frequently mutated in pancreatic, colon, and lung cancer (Karnoub and Weinberg, 2008). KRAS mutations predominantly include one of three point mutations at residues G12, G13, or Q61. Oncogenic substitutions of G12 or G13 leads to constitutive activation of Ras by creating steric hindrance that prevents the formation of van der Waals interactions between Ras and RasGAPS (Scheffzek et al., 1997), whereas substitutions of Q61 interfere with the coordination of a water molecule necessary for GTP hydrolysis (Scheidig et al., 1999). Oncogenic KRAS alleles are thought to reprogram metabolism, including increased glucose and glutamine consumption, lactic acid accumulation, altered expression of mitochondrial genes, increased reactive oxygen species (ROS) production, and reduced mitochondrial activity (Bos, 1989; Vizan et al, 2005; Chiaradonna et al, 2006a; Yun et al, 2009; Weinberg et al, 2010). However, the precise metabolic effects downstream of oncogenic Ras signaling in cancer cells have not been completely elucidated (Gaglio et al., 2011).

LKB1 (also known as serine-threonine kinase 11, STK11) is a tumor suppressive serine/threonine kinase that is mutated in 20-30% of NSCLC patients. LKB1 possesses multiple cellular functions in the regulation of cell metabolism, cell cycle arrest, embryo development, cell polarity, and apoptosis (Zhao et al., 2014). LKB1 is a key upstream

activator of the AMP-activated protein kinase (AMPK) and 12 kinases closely related to AMPK. AMPK is a central metabolic switch found in all eukaryotes that governs glucose and lipid metabolism in response to alterations in nutrients and intracellular energy levels (Shaw et al., 2004; Woods et al., 2003).

Recently, researchers performed an integrative analysis of genomic, transcriptomic and proteomic data from early-stage and chemo-refractory lung adenocarcinoma and identified three robust subsets of KRAS-mutant dominated by co-occurring genetic events in LKB1, TP53 and CDKN2A/B inactivation (Skoulidis et al., 2015). Furthermore, LKB1 deficiency in combination with KRAS mutation leads to an aggressive tumor phenotype at high prevalence in mouse models (Chen et al., 2012).

Combined KRAS/LKB1 mutation affects glucose-dependent TCA cycle

It is widely documented in the literature that KRAS regulates glucose and glutamine metabolism, and we did see the phenotype by comparing the metabolic profiling data set with the cell genetic background. Fig 5.3 shows increased glucose and glutamine ^{13}C labeling on KRAS-mutant cell lines. According to the pathway diagram of Fig 5.3a, ^{13}C labels glucose joining the TCA cycle as a two carbon unit Acetyl-coA, and generates citrate with a major m+2 MID pattern. Elizabeth McMillan analyzed the KRAS mutation status of all the cell line I used and compared it with the citrate m+2 from glucose labeling. All the red dots presented as mutant did shift to the higher fraction, indicating

that KRAS did affect the glucose labeling feature (p=0.0036). Similarly, Figure 5.3b shows the major citrate m+4 pattern from glutamine also shifted in the KRAS mutant cell lines (p=0.018). This non-supervised analysis demonstrated the influences of a specific type of mutation on cellular metabolic within the context of all other mutations in a molecularly heterogenous set of cells.

With the same method, there are many other examples by which the RAS family regulates metabolic features. The one that interested us most was the correlation between KRAS/LKB1 and glucose labeled citrate (Fig 5.4a with p=0.0019). We clustered the ^{13}C glucose labeled citrate carbon enrichment data and displayed it as heatmap (Fig5.4b). Cells were divided in two groups depending on their m+0, m+4 and m+5 groups, the top one shows lower m+0 and higher m+4 and m+5 compared to the bottom group. We checked the genetic background distribution of this heatmap, and found that most (12 out of 13) KRAS/LKB1 co-mutant cell lines belonged to the top group (P=0.0004). Neither KRAS nor LKB1 mutations alone were sufficient to cluster well by this analysis.

In order to further study this effect, I overexpressed LKB1 in three KRAS/LKB1 co-mutant cell lines (A549, H460, Hcc44) to generate three isogenic paired cell lines (Fig 5.4c). I first performed ^{13}C labeling on the three paired lines and found that, by over expressing LKB1, cells show higher citrate m+0 and moved to the bottom group on the heatmap. I also checked the full range of citrate m+0 fractions as shown in Fig 5.4d. Originally all the co-mutant cell lines clustered together on the low citrate m+0 side

(p=0.0004), but the three cell lines all moved to the far higher m+0 side with LKB1 overexpression.

Interestingly, this impact of LKB1 overexpression is specific for glucose metabolism. As shown in Fig 5.5, the nutrient utilization rates stayed close between original the KRAS/LKB1 co-mutant cell lines and their LKB1 overexpressed counterparts. Furthermore, I performed the ^{13}C-glutamine labeling on the three pair lines and showed no differences on citrate MID patterns.

The citrate m+0 increased in LKB1 overexpression lines may cause by lower TCA turnover rate

From previous results we found that the combined KRAS/LKB1 mutation cells cluster together in the ^{13}C-glucose heat map. Overexpression of LKB1 in these cells results in increased citrate m+0 from ^{13}C-glucose (which means more unlabeled carbons in citrate) but unaltered ^{13}C-glutamine labeling pattern and nutrient uptake rates. There are two possible explanations. First, LKB1 overexpression cells may decrease the turnover rate of the TCA cycle and delay the incorporation rate of carbons from ^{13}C-glucose. The second hypothesis is that LKB1 may activate other unlabeled carbon sources to join the TCA cycle and generate citrate. In the following paragraph, I discuss results in support of the first hypothesis.

In conventional form of the TCA cycle, cells pick up carbon molecules from multiple sources (usually glucose or glutamine) and use it for generating intermediate such as citrate. For example, glucose forms pyruvate through glycolysis and joins the TCA cycle as a two carbon molecule Acetyl-coA. If we provide ^{13}C-glucose as a stable tracer and use citrate carbon enrichment pattern as a readout, we will see mostly m+0 in the beginning because most of the carbons in citrate are still unlabeled ^{12}C. Then cells start to use the ^{13}C-labeled Acetyl-coA carbon from each TCA cycle, making the pattern move toward lower m+0 and higher m+2. In this way, I can simply focus the citrate m+0 carbon enrichment pattern on a period of time and assess the effect of LKB1 expression. I labeled the three paired lines (LKB1 overexpressed in LKB1/KRAS co-mutant cell lines) with ^{13}C-glucose and traced the citrate m+0 pattern from 6 to 48 hour of labeling time. Fig 5.6a indicated that although LKB1 overexpression induced a higher m+0 patter at the early time point, this effect was lost by 30 hours. This suggests that the LKB1 overexpression decreases the TCA cycle turnover rate and delays carbon incorporation from ^{13}C-glucose.

LKB1 is a tumor suppressor that is broadly inactivated in human lung cancers. In LKB1-deficient lung cancer cells, AMPK activity is suppressed and refractory to its pharmacological activators, leading to increased mTORC1 signaling, whereas the ability of AMPK to inhibit cell growth is restored when wild-type LKB1 is expressed (Carretero et al., 2007; Xiang et al., 2004). Overexpression of LKB1 activated AMPK signaling and induced G1 cell cycle arrest (Liang et al., 2014). Moreover, LKB1 was reported to

correlate with a substantial reduction of c-myc expression. The expression of c-Myc protein decreased by LKB1 in transfected cells may be a contributory factor in the process of cell proliferation (Liang et al., 2009). Combined with the fact that MYC positively regulates glucose metabolism, it may be the reason why cells overexpressing LKB1 delayed carbon incorporation from glucose.

LKB1 overexpression cell lines shows similar beta-oxidation rate with deficient lines

Another reason of why citrate m+0 is increased in lines that overexpress LKB1 is that LKB1 may activate other unlabeled carbon sources to join the TCA cycle and generate citrate with unlabeled ^{12}C, which would result in a higher citrate m+0 pattern. A possible unlabeled carbon source is Acetyl-coA generated from fatty acid metabolism.

AMPK is a major downstream target of LKB1 signaling, and it is known for regulating fatty acid oxidation (O'Neill et al., 2013). Fatty acid β-oxidation is the catabolic process by which fatty acid molecules are broken down in mitochondria to generate acetyl-CoA and energy. The beta-oxidation pathway includes four reactions that occur in repeating cycles with each fatty acid molecule. In each cycle, a fatty acid is progressively shortened by two carbons as it is oxidized and its energy captured by the reduced energy carriers NADH and FADH2. At the end of each cycle of four reactions, one acetyl-CoA two-carbon unit is released from the end of the fatty acid, which then goes through another round of beta-oxidation, continuing to oxidize and shorten even-chain fatty

acids until they are entirely converted to acetyl-CoA. This acetyl-CoA then enters the TCA cycle. The NADH and FADH2 are used by the electron transport chain to produce ATP (Su and Abumrad, 2009).

In order to test the β-oxidation rate of the pair cell lines A549, H460 and Hcc44, I performed the ^{14}C-Palmitate CO_2 assay on the six lines with LKB1 mutation or overexpression. Radiolabeled fatty acids are oxidized ex vivo and measured by capturing CO_2. Moreover, this ex vivo fatty acid oxidation is sensitive to inhibitors such as Etomoxir, and I used it to treat the cells as negative controls. Fig 5.6b shows that the CO_2 capturing rates in cells that overexpress LKB1 were not significantly increased compare to the original LKB1 mutant lines, but with the Etomoxir treatment the CO_2 dropped to near zero in every cell line. That means the inhibitor and the experiment system worked properly, but β-oxidation may not be the reason of increasing citrate m+0 in my cell pairs. To further support this result, I treated the paired cells with Etomoxir directly and determined the cell survival by the DNA content assay. If the fatty acid oxidation pathway plays a role in supporting the Acetyl-coA formation in the cell lines that overexpress LKB1, inhibiting the pathway by Etomoxir should impair cell growth and survival. However, I did not a find significant difference in cell survival between the paired lines (Fig 5.6c), which is consistent with the ^{14}C-Palmitate CO_2 assay results.

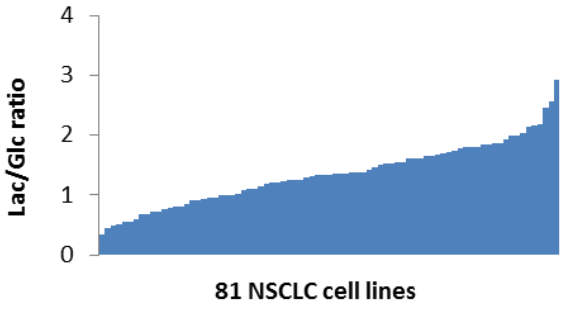
Discussion

In this chapter, we examined the cross associations between the metabolic features and other orthogonal datasets, and demonstrated that the analysis of orthogonal datasets provide clues to molecular basis underlying metabolic phenotypes.

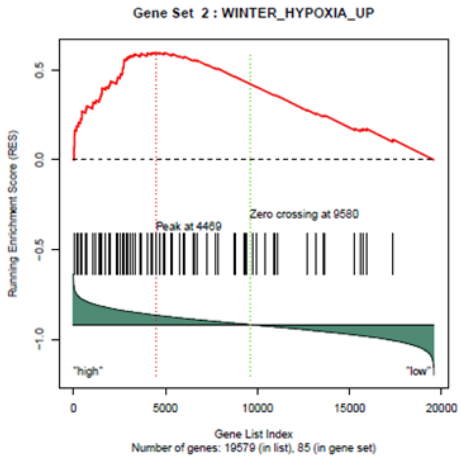
The first example revealed the proof of principle correlations between Lactate/glucose ratio and hypoxia, Myc and ribosome biogenesis gene expressions. HIF1 and Myc are known to play a role on glucose metabolism, and the ribosome biogenesis could be linked to glucose homeostasis through Myc regulation or other genes such as SFP1 or SIRT6 (Cipollina et al., 2008; Sebastian et al., 2012).

The second example, we found that the combined KRAS/LKB1 mutation affects the glucose-dependent TCA cycle. I overexpressed LKB1 in NSCLC cell lines and found the different TCA turnover rate may be the reason of the increased citrate m+0 these cells. Recent reports found that KRAS/LKB1 co-mutant cells display a more progressive disease and had increased sensitivity to MAPK and mTOR signaling inhibition (Chen et al., 2012; Mahoney et al., 2009). The deoxythymidylate kinase was reported as a target, and the lysosome maturation may be the metabolic bottleneck for KRAS/LKB1 tumors (Kim et al., 2013; Liu et al., 2013a). Recently, our lab also found that the KRAS/LKB1 co-mutation induces altered mitochondrial carbon and nitrogen metabolism (manuscript in preparation).

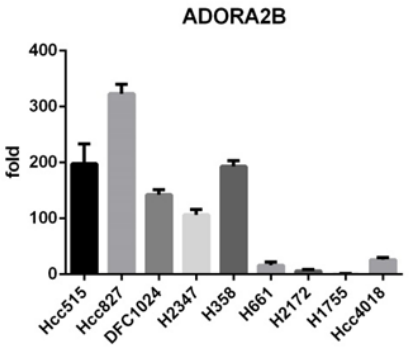
a.

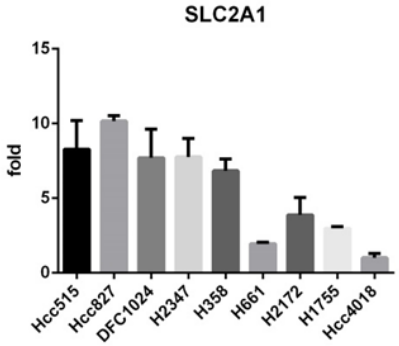

Lac/Glc ratio
4
3
2
1
0
81 NSCLC cell lines

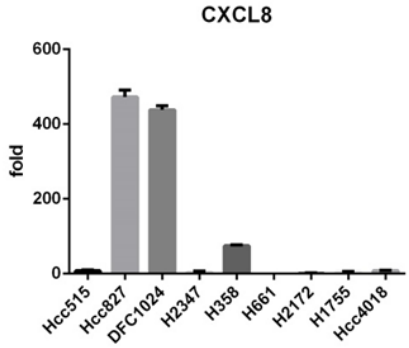
b.


Gene Set 2 : WINTER_HYPOXIA_UP
Running Enrichment Score (RES)
0.5
0.0
−0.5
−1.0
Peak at 4469
Zero crossing at 9580
"high"
"low"
0
5000
10000
15000
20000
Gene List Index
Number of genes: 19579 (in list), 85 (in gene set)

c.


ADORA2B
fold
400
300
200
100
0
Hcc515
Hcc827
DFC1024
H2347
H358
H661
H2172
H1755
Hcc4018


SLC2A1
fold
15
10
5
0
Hcc515
Hcc827
DFC1024
H2347
H358
H661
H2172
H1755
Hcc4018


CXCL8
fold
600
400
200
0
Hcc515
Hcc827
DFC1024
H2347
H358
H661
H2172
H1755
Hcc4018

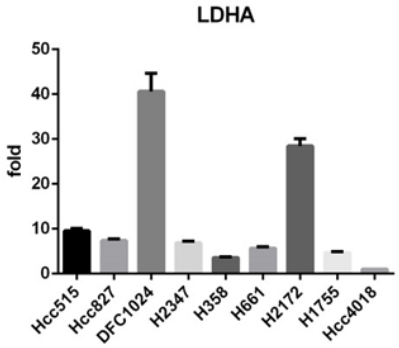

LDHA
fold
50
40
30
20
10
0
Hcc515
Hcc827
DFC1024
H2347
H358
H661
H2172
H1755
Hcc4018

Fig 5.1 Lactate/Glucose ratio correlates with gene sets relative to hypoxia and Myc

(a) The lactate secretion/glucose utilization ratio of NSCLC cell lines, ranked low to high. (b) Gene set enrichment analysis (GSEA) was run for genes correlated with lactate/glucose ratio. Spearman correlation coefficient was used as the ranking metrics. The enrichment plot provides a graphical view of the enrichment score for the hypoxia gene set, the majority of genes in the gene set are positively correlated with the metabolic feature. (c) qPCR of genes on the top hits (ADORA2B, SLC2A1, CXCL8) and bottom hit (LDHA) of the GSEA result. High ratio cell lines include Hcc515, Hcc827, DFCI024, H2347 and H358. Low lines include H661, H2172, H1755 and Hcc4018. Error bars represent SD.

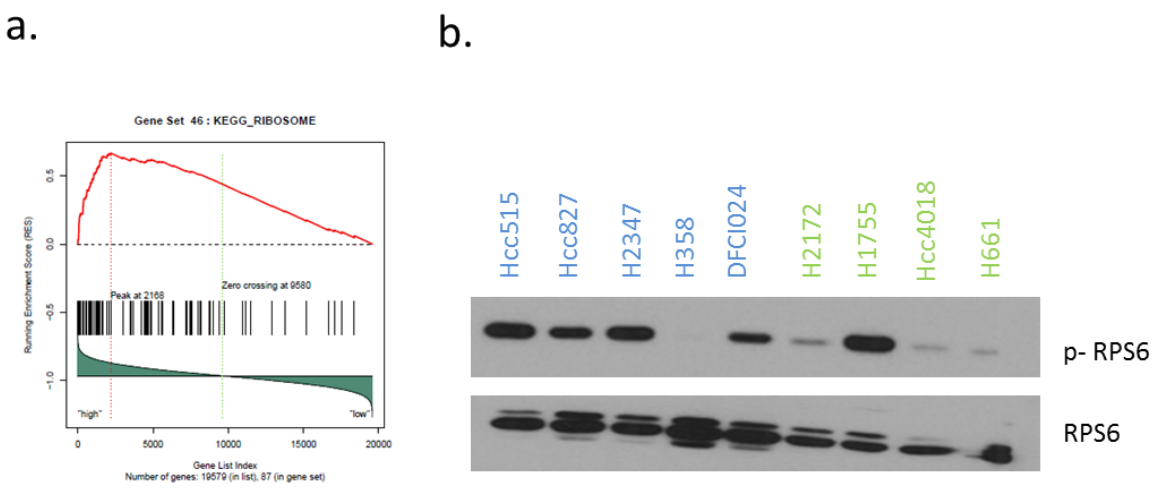


Fig 5.2 Lactate/Glucose ratio correlates with gene sets relative to ribosome biogenesis

(a) The GSEA enrichment plot provides a graphical view of the enrichment score for the ribosome biogenesis gene set, the majority of genes in the gene set are positively correlated with the lactate/glucose ratio metabolic feature. (b) Ribosomal protein S6 (RPS6) protein expressions in high Lac/Glc ratio cell lines were labeled by blue, and green color for low ratio lines.

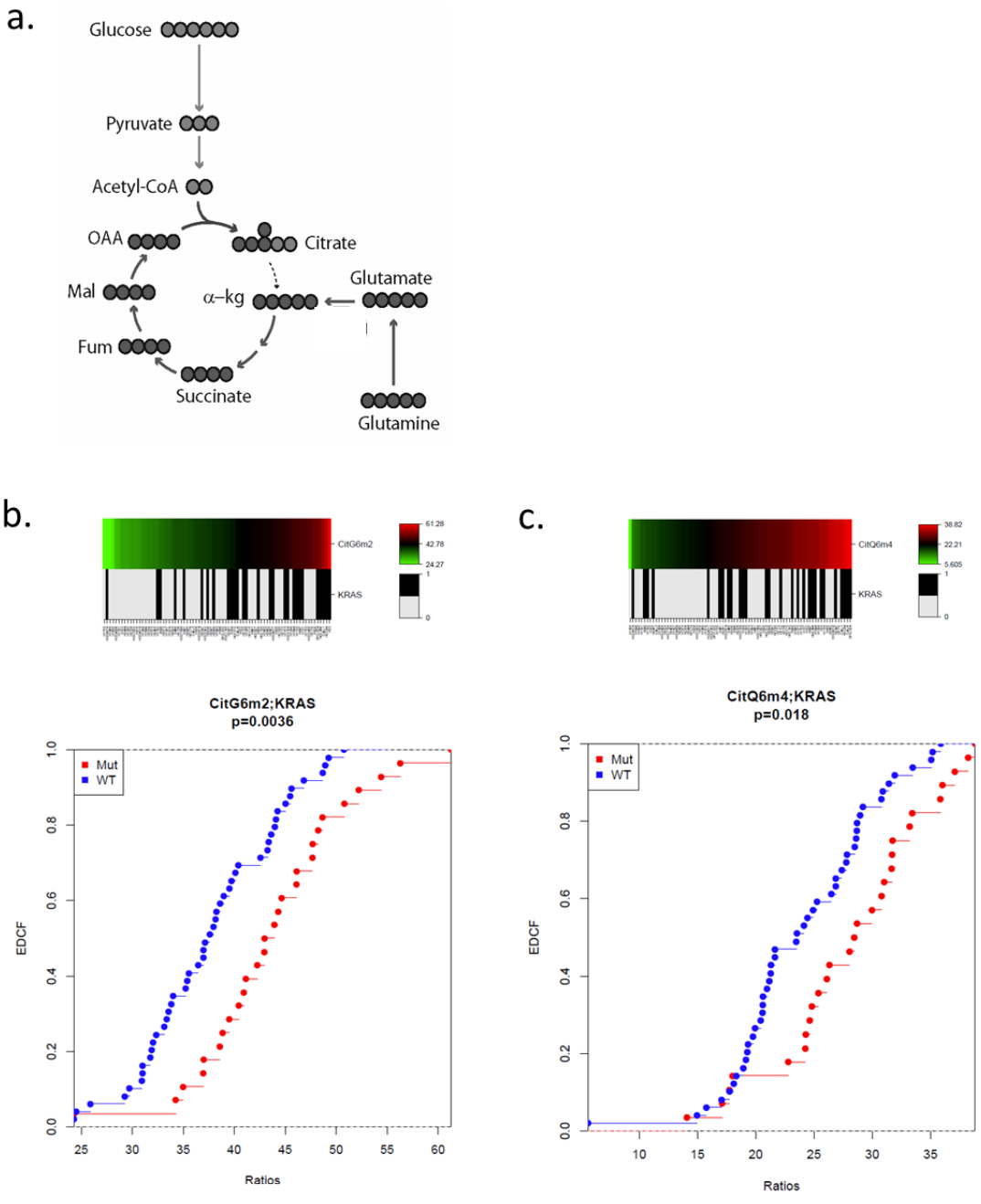


Fig 5.3 KRAS mutations impact how cells metabolize glucose and glutamine

(a) Glucose and glutamine fuel the TCA cycle. Increased glucose and glutamine ^{13}C labeling on citrate were show on KRAS-mutant cell lines, by comparing the glucose m+2 labeling fraction (b) and glutamine m+4 labeling fraction (c) with the KRAS mutation background.

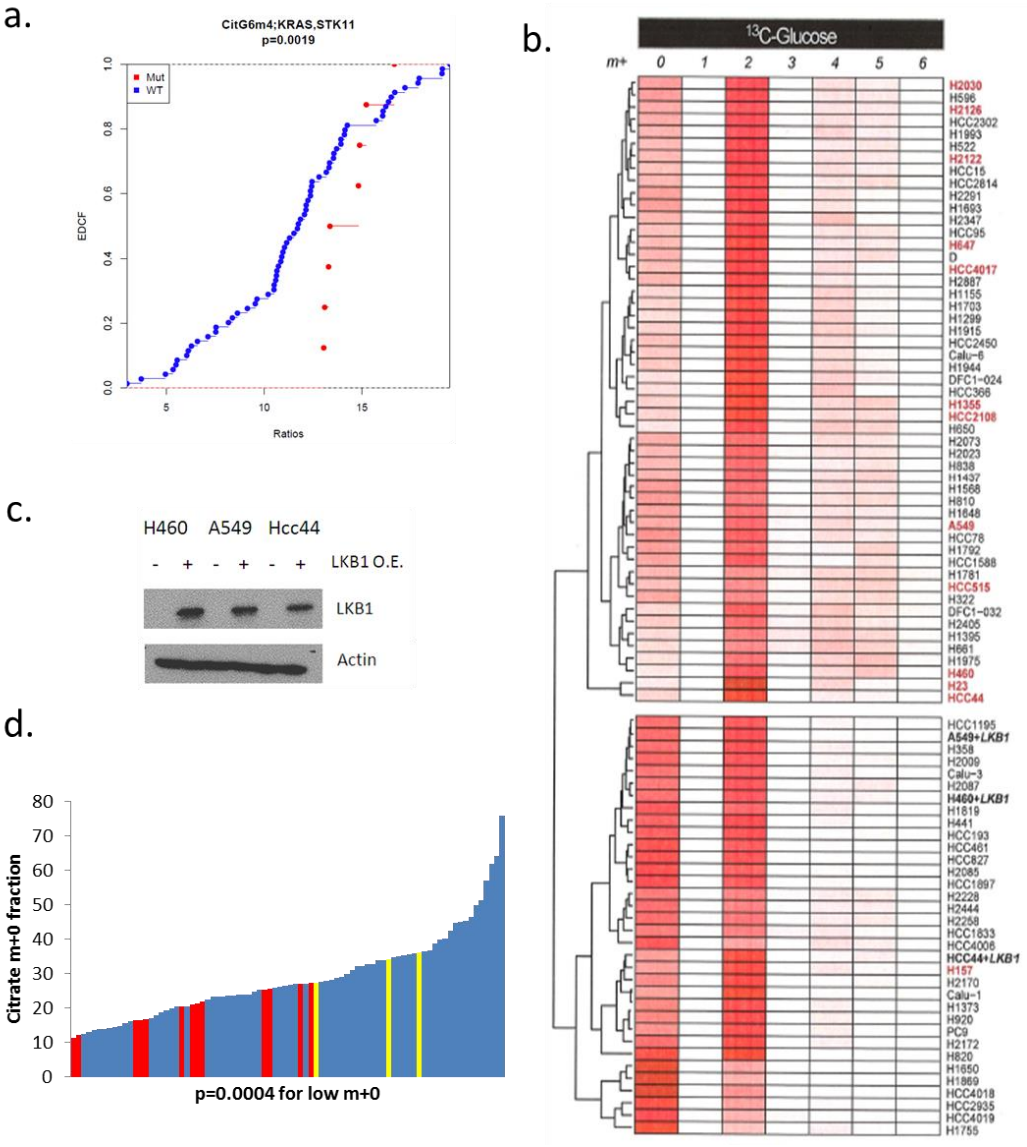


Fig 5.4 KRAS/LKB1 co-mutants require glucose-dependent TCA cycle

(a) KRAS/LKB1 co-mutants significantly associated with citrate m+2 fractions from ^{13}C-glucose. (b) Heatmap based on citrate MID from glucose labeling. Cell lines with combined KRAS/LKB1 mutation (label as red) cluster together in the heat map (p=0.01). Over-expressing LKB1 in co-mutants is sufficient to drive cell lines out of the group (bold labeled). (c) LKB1 protein expression in the CD8+ control and LKB1 overexpression cell lines. (d) Full rank of citrate m+0 fractions. Red bars present KRAS/LKB1 co-mutants; yellow bars present the three LKB1 overexpression cell lines; all other cell lines are blue bars.

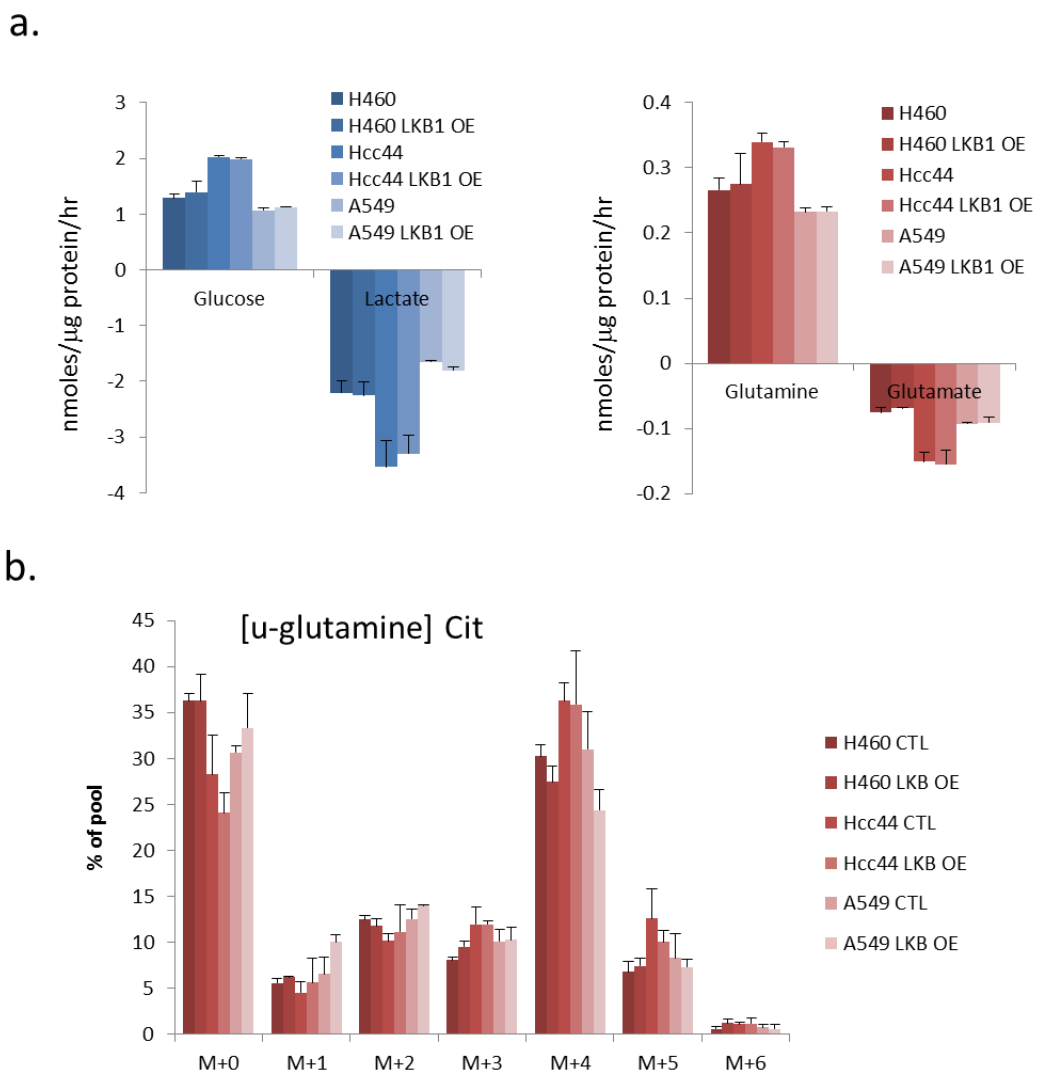


Fig 5.5 The impact of LKB1 overexpression is specific for glucose metabolism

(a) LKB1 overexpression cell pairs were cultured in the presence of 10 mM glucose and 2 mM glutamine for 6 hours. Media was analyzed for metabolite concentrations. Values are an average of biological triplicates with positive values representing consumption and negative values representing secretion. Error bars represent SD. (b) Cells were labeled by [u-^{13}C] glutamine for 6 hours and analyzed for the citrate MID. Error bars represent SD.

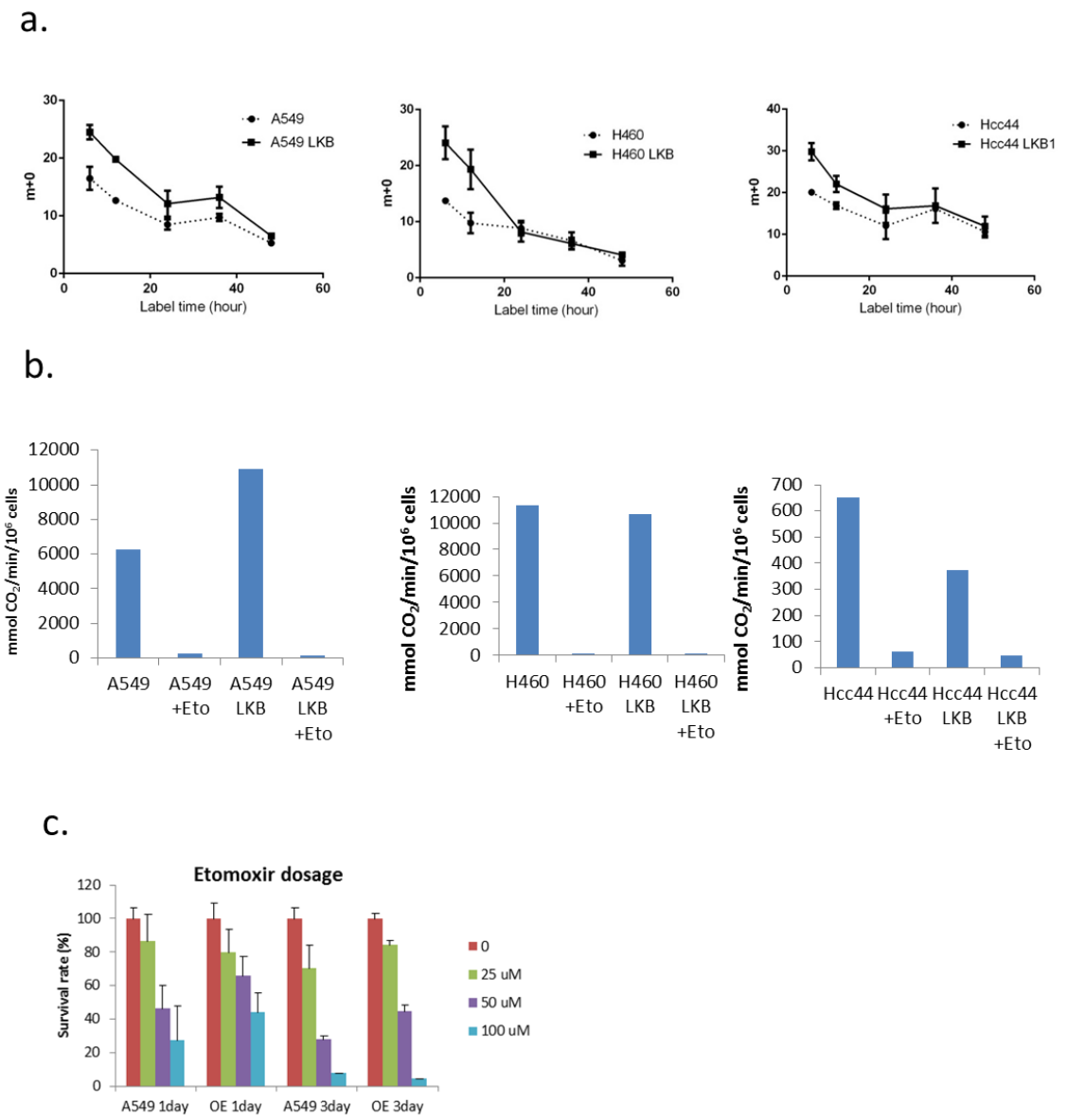


Fig 5.6 KRAS/LKB1 co-mutants show quick TCA turnover rate

(a) LKB1 overexpression cell pairs were labeled with [u-^{13}C] glucose from 6 to 48 hours. Fresh labeling medium were supplied every 24 hours. Citrate m+0 carbon fraction were shown. (b) The beta-oxidation rate of LKB1 overexpression cell lines were tested by ^{14}C-Palmitate oxidation assay. 200 uM Etomoxir were used as negative control. Each bar presents ^{14}CO2 production rate. (c) Cells were treated with different Etomoxir dosages for 1 and 3 days, and the survive rate were tested by DNA content assay.

CHAPTER SIX
Activity of the de novo serine biosynthetic pathway predicts pemetrexed sensitivity in NSCLC cells

Introduction

In the last decade, targeted chemotherapy of NSCLC has been focused on mutations of the epidermal growth factor receptor (EGFR) and on the abnormal fusion of the anaplastic lymphoma kinase (ALK) being inhibited successfully with EGFR tyrosine kinase inhibitors (gefitinib and erlotinib) and crizotinib respectively (Chan and Hughes, 2015). Unfortunately, the responses are not durable and resistance inevitably occurs after a median of 9 to 14 months (Matikas et al., 2015). In order to improve patient outcomes and the quality of life, new biomarkers are needed to predict sensitivity of targeted therapy and estimate prognosis (Thunnissen et al., 2014).

In this project, metabolic features derived from nutrient utilization, nutrient addiction, and isotope labeling patterns were cross-queried with orthogonal data sets including the sensitivity to over 40 chemotherapeutic agents. The results may identify robust, novel relationships connecting metabolic preferences to therapeutic responses. In this chapter I will describe one example: sensitivity to folate antagonists in cell lines displaying high flux through the de novo serine biosynthetic pathway. We first analyzed the glucose dependent serine biosynthesis and found the association between serine

carbon enrichment and the sensitivity of Pemetrexed in NSCLC. I further validated this correlation and identified a possible mechanism. This finding may provide a novel metabolic biomarker for Pemetrexed therapy.

Background: serine, glycine, and one-carbon network in metabolism

Serine and glycine are non-essential amino acid that fuel metabolic pathways including the one-carbon network. One-carbon metabolism involving the folate and methionine cycles integrates nutritional status from amino acids, glucose and vitamins, and generates diverse outputs such as the biosynthesis of lipids, nucleotides and proteins, the maintenance of redox status and the substrates for methylation reactions (Locasale, 2013). Imported serine and serine derived from a branch of glycolysis can be converted to glycine, which in turn provides methyl groups for one-carbon metabolism. These methyl groups are used in the folate cycle and distribute to diverse cellular processes including nucleotide biosynthesis. Inhibitor of folate metabolism and its downstream pathway nucleotide biosynthesis has been used in chemotherapy for decades, with as 5-fluorouracil (5-FU) and methotrexate being two effective agents (Chabner and Roberts, 2005).

Recent work has pointed for new roles of the serine, glycine, and one-carbon network in cancer pathogenesis (Chaneton et al., 2012; Maddocks et al., 2013; Zhang et al., 2012). Whereas a subset of cancer cells increase de novo serine biosynthesis (Locasale

et al., 2011; Possemato et al., 2011), other cancers benefit from an increased serine and glycine uptake rate, which allows them to metabolize these amino acids for their biosynthetic needs (Jain et al., 2012; Maddocks et al., 2013). A critical role for the mitochondrial folate pathway in rapidly proliferating cancer cells has also been recently elucidated. In one study, researchers compared messenger RNA profiles of 1,454 metabolic enzymes across 1,981 tumors spanning 19 cancer types to identify enzymes that are consistently differentially expressed, and the highest scoring pathway was mitochondrial one-carbon metabolism (Nilsson et al., 2014). Despite these advances, the general coordinated usages and different contexts in which serine and glycine flux contributes to different metabolic functions within and across cancer types remain largely unknown (Mehrmohamadi et al., 2014).

Metabolic flux through de novo serine synthesis correlates with serine MID

Besides providing energy, glycolysis is extensively used to sustain anabolism, which is necessary for tumor growth. In anabolic pathways, the serine biosynthetic pathway represents a crucial turning point in glucose conversion (Kalhan and Hanson, 2012). As indicated by Fig 6.1a, the glycolytic intermediate 3-phosphoglycerate (3-PG) is converted to serine following a three-step pathway. Cells use phosphoglycerate dehydrogenase (PHGDH) and NAD to oxidize the 3-PG into the serine precursor 3-phosphohydroxypyruvate (3-PP). Subsequent enzymes in the pathway convert 3-PP into

serine via transamination (PSAT1) and phosphate ester hydrolysis (PSPH) reactions. After serine synthesis, conversion of serine to glycine is catalyzed by serine hydroxylmethyltransferase (SHMT). This enzyme provides the majority of methyl groups to the one-carbon pools for biosynthesis and DNA methylation (DeBerardinis, 2011).

When labeling cells with [U-^{13}C] glucose, the six ^{13}C carbons from each glucose molecule generate two units of ^{13}C labeled 3-PG, and contribute to serine carbon enrichment through the de novo serine synthesis pathway. The major serine MID pattern that derives from [U-^{13}C] glucose is Ser m+3 because of complete labeling of 3-PG. Then glycine shows a major m+2 pattern because of the loss of one carbon through the SHMT reaction. Fig 6.1b shows the serine and glycine MID of 80 cell lines from ^{13}C-glucose labeling. Both Ser m+3 and Gly m+2 distribute within 40% enrichment. The low level or carbon enrichment results from the relatively small input from glycolysis compared to unlabeled serine and glycine imported from medium. Furthermore, the significant correlation between Ser m+3 and Gly m+2 labelling indicates that the glycine produced from glucose was derived from serine ($r=0.79$, $p=7e^{-14}$).

GSEA revealed that the KEGG Glycine Serine and Threonine Metabolism gene set was the top gene-set positively correlated with Ser m+3 from glucose. This dataset contains 31 genes closely associated with glycine, serine and threonine metabolism. Furthermore, the top genes that showed the most significant positive correlation with the Ser m+3 feature are all involved in serine metabolic pathways, including PHGDH, PSAT1 and PSPH (Fig 6.1c). Since PHGDH was the top hit, we analyzed its expression

independently against Ser m+3. Figure 6.1d shows significant positive correlation between Ser m+3 and PHGDH expression ($r=0.47$, $p=2.7e^{-5}$).

Cell lines with high Ser m+3 pattern are more sensitive to Pemetrexed treatment

One of the NSCLC orthogonal data sets from the Minna lab is their response to over 40 chemotherapeutic agents. Appendix online file shows some examples of the distribution of chemotherapeutic in NSCLC cell lines. The IC50 distributions of various chemotherapeutic agents were shown in density plots, indicating the diversity of responses of NSCLC cells to each agent. Ling Cai compared each metabolic feature with each IC50 and ranked the results by P-value from low to high. Table 6.1 shows that many of the top hits of Ser m+3 belong to a group of antifolates and drugs involved in nucleotide metabolism in DNA damage. Considering the correlation between serine synthesis and one-carbon metabolism, this result is encouraging. Pemetrexed, the only anti-folate agent in the data set, located on the top of the hits list.

Pemetrexed (brand name Alimta) is a chemotherapeutic agent manufactured by the Eli Lilly Company. This folate analog works by inhibiting three enzymes used in purine and pyrimidine synthesis - thymidylate synthase (TS), dihydrofolate reductase (DHFR), and glycinamide ribonucleotide formyltransferase (GARFT). By inhibiting the formation of precursor purine and pyrimidine nucleotides, pemetrexed prevents DNA and RNA synthesis in cancer cells. Pemetrexed (PEM) was approved by FDA to treat malignant

pleural mesothelioma on 2004, and is currently used in NSCLC patients as a first-line agent with cisplatin or solely as second-line treatment (Cohen et al., 2009). Observational and preclinical studies suggested an association between the expression of TS and clinical effects of PEM-based chemotherapy in NSCLC patients. However, the predictive value of TS for PEM-containing chemotherapy regimen remained controversial (Liu et al., 2013b). In order to improve patient outcomes, new biomarkers are urgently needed.

Fig 6.2a shows the correlation plot between PEM sensitivity and Ser m+3 enrichment. There are only 12 cell lines that present Ser m+3 enrichment higher than 15%, and 11 of them are PEM sensitive. In other words, although the PEM sensitive lines include both Ser m+3 low and high lines, almost all of the high lines are PEM sensitive. Our results indicate that metabolic phenotype may be a more powerful method of prediction compared to TS expression in NSCLC cell lines (Fig 6.2b).

In order to validate the sensitivity, I picked five cell lines with high Ser m+3 (PC9, H2170, H596, H460, H1299) and five cell lines with low Ser m+3 (Hcc44, H2882, H650, H1395, Hcc1195) and treated them with different doses of PEM. Fig 6.2c shows that all high Ser m+3 cell lines were sensitive to PEM doses above 0.1uM (dosage data not shown). The area under curve plot also indicated that the Ser m+3 high lines are more sensitive to PEM treatment compared to the Ser m+3 low lines. These results supported the correlation between Ser m+3 and sensitivity to PEM, but the underlying mechanisms remain to be determined.

The high serine m+3 enrichment comes from de novo serine biosynthesis pathway

In order to find the mechanism behind Ser flux and PEM sensitivity, I needed to first identify the different serine enrichment between high and low lines, and figure out the pathway that contributes to the high ser m+3. There are four possibilities that can affect serine flux and carbon enrichment patterns. First, different cell growth rate may increase the cell turnover rate and incorporate more labeled carbon from glycolysis. Second, every cell line has different intrinsic serine pool sizes; for instance the high lines may have small serine pool size and show higher labeling even with low flux. Third, the low lines may just incorporate more unlabeled serine from medium, not from labeled glucose. Fourth, the high serine m+3 may come from bone side flux through glycolysis and de novo serine biosynthesis. I performed experiments to test and exclude the first three possibilities, and provided evidence to support that differential de novo serine biosynthesis accounts for differential Ser m+3 content.

To examine the effect of cell growth rate, I checked the correlation between growth rate and Ser m+3 (Fig 6.3a) and found the two parameters to be insignificantly associated. To confirm this result, I performed a BrdU staining assay to check the S phase in high and low lines. S phase fraction between high and low lines was comparable (Fig. 6.3b). Then using an isotope dilution method, I found that there was no significant difference in the serine pool size between high and low lines (Fig 6.3c). For

the third possibility, I used NOVA and Amino Acid Analyzer to examine how much serine, glycine, glucose and glutamine was consumed by cells from the medium. There was similar nutrient utilization among the high and low lines (Fig 6.3d). These data exclude the first three possibilities mentioned on last paragraph.

By using qPCR, I found that the expression levels of most of the major genes involved in the serine biosynthesis pathway were increased in high Ser m+3 cell lines. Important enzymes like PHGDH and PSAT, and upstream genes like NRF2 and ATF4 were all upregulated as shown in Fig 6.4ab. I also checked the PHGDH protein expression by western blot and found that most high lines show higher level of PHGDH (Fig 6.4c). PHGDH is a rate-limiting step in the conversion of 3-PG to serine. It is known that the expression of PHGDH has been found upregulated in triple negative breast cancer and in melanoma, suggesting that elevated serine biosynthesis is advantage in these tumors (Locasale et al., 2011; Possemato et al., 2011). The PHGDH expression result not only supports the qPCR data, but also provides further evidence that the de novo serine biosynthesis pathway is upregulated in high ser m+3 lines, and contributes to the higher serine flux. By revealing the link between serine carbon enrichment and PEM sensitivity, ser m+3 may be a proper biomarker candidate for PEM treatment.

SHMT may be the target of Pemetrexed treatment in serine biosynthesis pathway

Though we showed that the ser M+3 may be a biomarker through de novo serine biosynthesis pathway, how PEM affects this pathway remains unclear. PEM is well

known to inhibit three enzymes in folate cycle (TS, DHFR and GARFT); however, SHMT1 has also been reported as a PEM target (Daidone et al., 2011), as well as inhibited by some other antifolate agents (Paiardini et al., 2015). Fig 6.5a shows the pathway diagram between serine and glycine through SHMT as a double-direction reaction. By labeling with ^{13}C-glucose, the fully labeled serine (Ser m+3) can be converted to fully labeled glycine (Gly m+2) by SHMT. Since this is a reversible reaction, the labeled glycine may be converted back to serine and appear as Ser m+2 as a unlabeled carbon is added. If PEM inhibits SHMT activity, I should detect the Gly m+2 decrease as a direct downstream effect. Indeed, the ^{13}C labeled serine and glycine enrichment change after PEM treatment (Fig6.5b). In all the cell lines, Gly m+2 dramatically decreased with PEM treatment as I hypothesized. Moreover, the Ser m+2 also dropped because of a lack of supply from reversed glycine though SHMT. Ser M+3 did not change much since serine is upstream of SHMT. Here I used only high serine lines due to the low levels of labeling in the Ser low lines. Besides the stable isotope tracing experiment, I also checked the serine pool size and found it to be increased in all cell lines after treating PEM (Fig 6.5c). This result further supported SHMT as a target of PEM.

To further support the result that SHMT is inhibited by PEM and decreases Gly m+2 enrichment, I knocked down SHMT in H460 cells and examined the change in labeling with ^{13}C-glucose. Fig 6.6a and b shows the labeling results in SHMT1 knockdown and SHMT2 knockout cells separately. Both SHMT knockdown cell lines have decreased Gly m+2 and Ser m+2 MID patterns, which is similar to the PEM treatment results.

Altogether these data provide strong evidence that PEM targets serine biosynthesis pathway through SHMT.

Cell lines show consistent sensitivity to PEM treatment in mouse tumor model.

NSCLC cell lines are a useful model system to identify new biomarkers (such as Serine flux) and study the underlying mechanisms. To test if the high lines and low lines keep the same sensitivity to PEM treatment *in vivo,* I used a xenograft mouse model. First, I subcutaneously implanted all the 10 high and low cell lines using nude mice to determine their tumorigenicity. Then I selected one high line (PC9) and one low line (H650) for PEM treatment test, which are strongly sensitive and resistant to PEM treatment, respectively (Fig 6.2c). Once tumors grew to a palpable size, I started treating the mice with daily saline control or 200 mg/kg PEM for ten days, and kept measuring tumor size. As expected, PC9 tumors showed a significant tumor regression in response to treatment while H650 tumors were highly resistant to PEM (Fig 6.7). These data provide primary evidence that this cell intrinsic effect is maintained in vivo, and further validates serine flux as a promising biomarker for chemosensitivity to PEM.

Conclusion

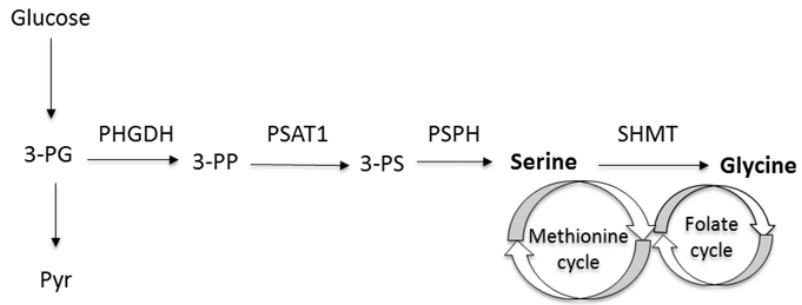
In this chapter I discovered and validated a novel biomarker for sensitivity to Pemetrexed, a folate antimetabolite drug used in NSCLC patients. By analyzing our metabolic profiling data with other orthogonal data sets, we found an interesting association linking Ser m+3 feature to serine biosynthesis and PEM drug sensitivity. I further validated this correlation and demonstrated that the underlying mechanism comes from a more active de novo serine biosynthesis pathway. Then I examined if PEM treatment affects the serine biosynthesis and found that the SHMT enzyme may be the target by analyzing the MID patterns from ^{13}C tracing assay. Moreover, SHMT silencing phenocopies PEM treatment. Finally I implanted the cell lines to mice and revealed a consistent sensitivity of cell lines to PEM treatment *in vivo*. This represents a promising new metabolic biomarker of Pemetrexed sensitivity via de novo serine biosynthesis pathway.

Drug	n	r	p	Target
Pemetrexed	61	-0.41	0.000951	anti-folate
Crenolanib	12	-0.75	0.004547	PDGFRA
Carboplatin	60	-0.33	0.010278	DNA intercalator
Doxorubicin	48	-0.35	0.014584	DNA intercalator
Irinotecan	40	-0.35	0.025469	topoisomerase 1
Pemetrexed+Cisplatin	60	-0.28	0.031548	
Gemcitabine	62	-0.25	0.047271	nucleoside analog
Paclitaxel+Carboplatin	62	-0.25	0.048309	

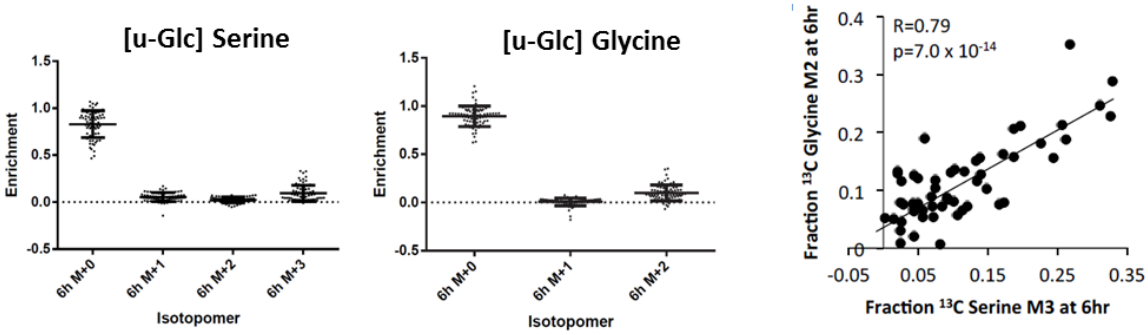
Table 6.1 Correlations between Ser m+3 fraction and drug sensitivity

Pearson correlation between Ser M3 and IC50 of various chemotherapeutic agents in log 10 scale were ranked by P-value from low to high, number of cell lines tested, pearson correlation coefficient, P-value and drug targets are shown in the table.

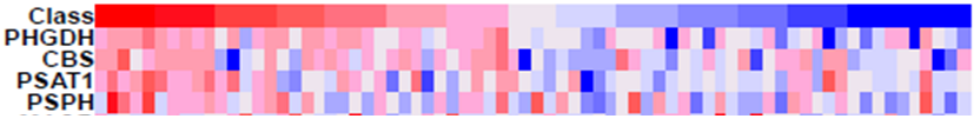
a.


Glucose
3-PG
PHGDH
3-PP
PSAT1
3-PS
PSPH
Serine
SHMT
Glycine
Pyr
Methionine cycle
Folate cycle

b.


[u-Glc] Serine
Enrichment
Isotopomer
6h M+0
6h M+1
6h M+2
6h M+3
[u-Glc] Glycine
R=0.79
p=7.0 x 10-14
Fraction 13C Glycine M2 at 6hr
Fraction 13C Serine M3 at 6hr

c.


Class
PHGDH
CBS
PSAT1
PSPH

d.

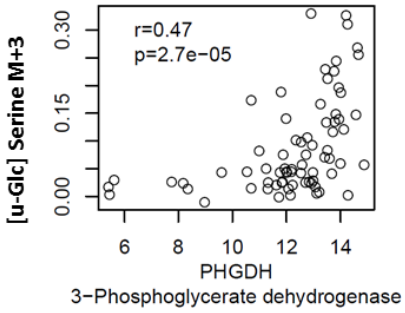

r=0.47
p=2.7e-05
[u-Glc] Serine M+3
PHGDH
3-Phosphoglycerate dehydrogenase

Fig 6.1 Metabolic flux through de novo serine synthesis correlates with serine MID

(a) De novo serine synthesis pathway diagram. Abbreviations: 3-PG, 3-Phosphoglycerate; 3-PP, 3-phosphohydroxypyruvate; Pyr, pyruvate. (b) Left, the serine and glycine MID from [u-^{13}C]-glucose labeling assay. Each dot presents a NSCLC cell line. Right, the correlation between serine m+3 and glycine m+2 fraction. (c) From GSEA, KEGG Glycine Serine and Threonine Metabolism returned as the top gene-set that positively correlated with Ser m+3. The top genes in this gene set that showed the most significant positive correlation with Ser m+3 are all involved in serine metabolic pathways. (d) Pearson correlation coefficient and P-value was shown for the pairwise correlation between Ser m+3 and PHGDH gene expression (log2 scale).

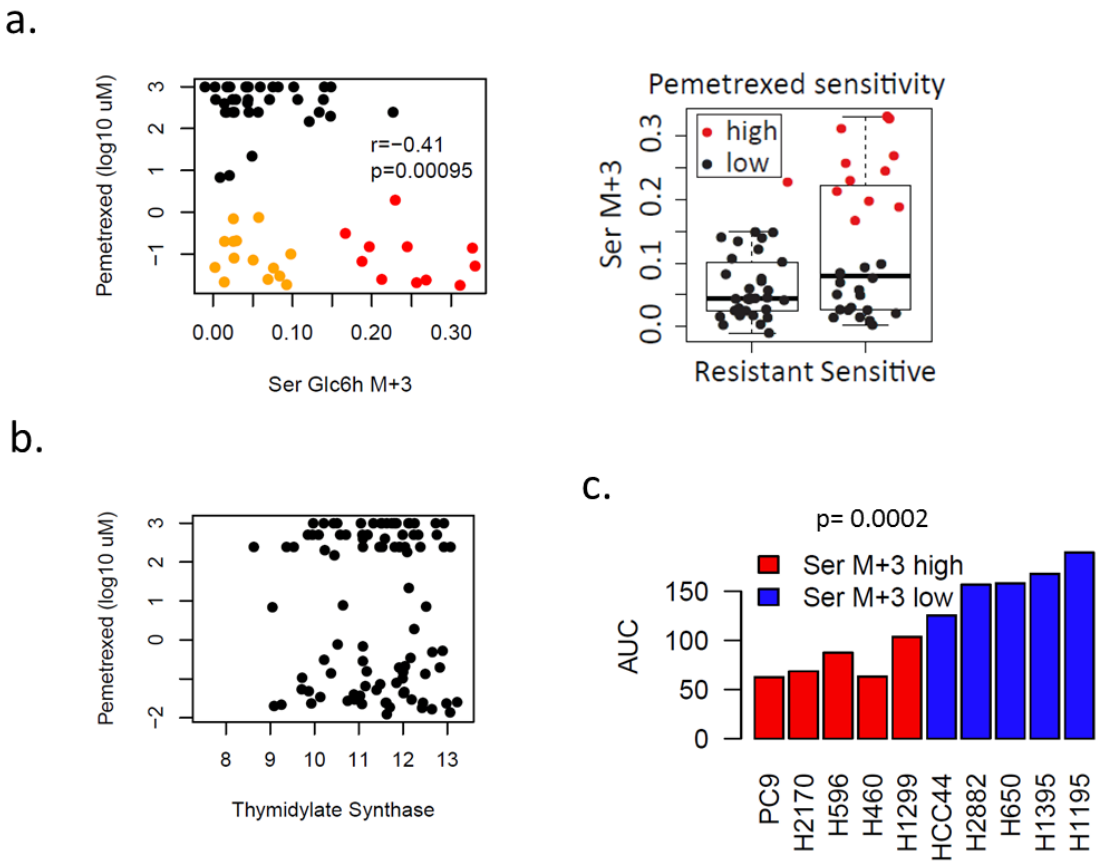


Fig 6.2 Cell lines with high Ser m+3 pattern are more sensitive to Pemetrexed treatment

(a) Left, Pearson correlation coefficient and P-value was shown for the pairwise correlation between Ser m+3 from [u-^{13}C] glucose labeling and Pemetrexed (PEM) sensitivity (log10 scale). Right, 11 of the 12 cell lines that with high Ser m+3 enrichment are PEM sensitive. (b) Insignificant correlation between TS gene expression and PEM sensitivity. (d) Cells were treated with different PEM dosages for 3 days, and the survive rate were analyzed to show as area under curve (AUC). The lower red bars present lower survival rate with PEM treatment (p= 0.0002).

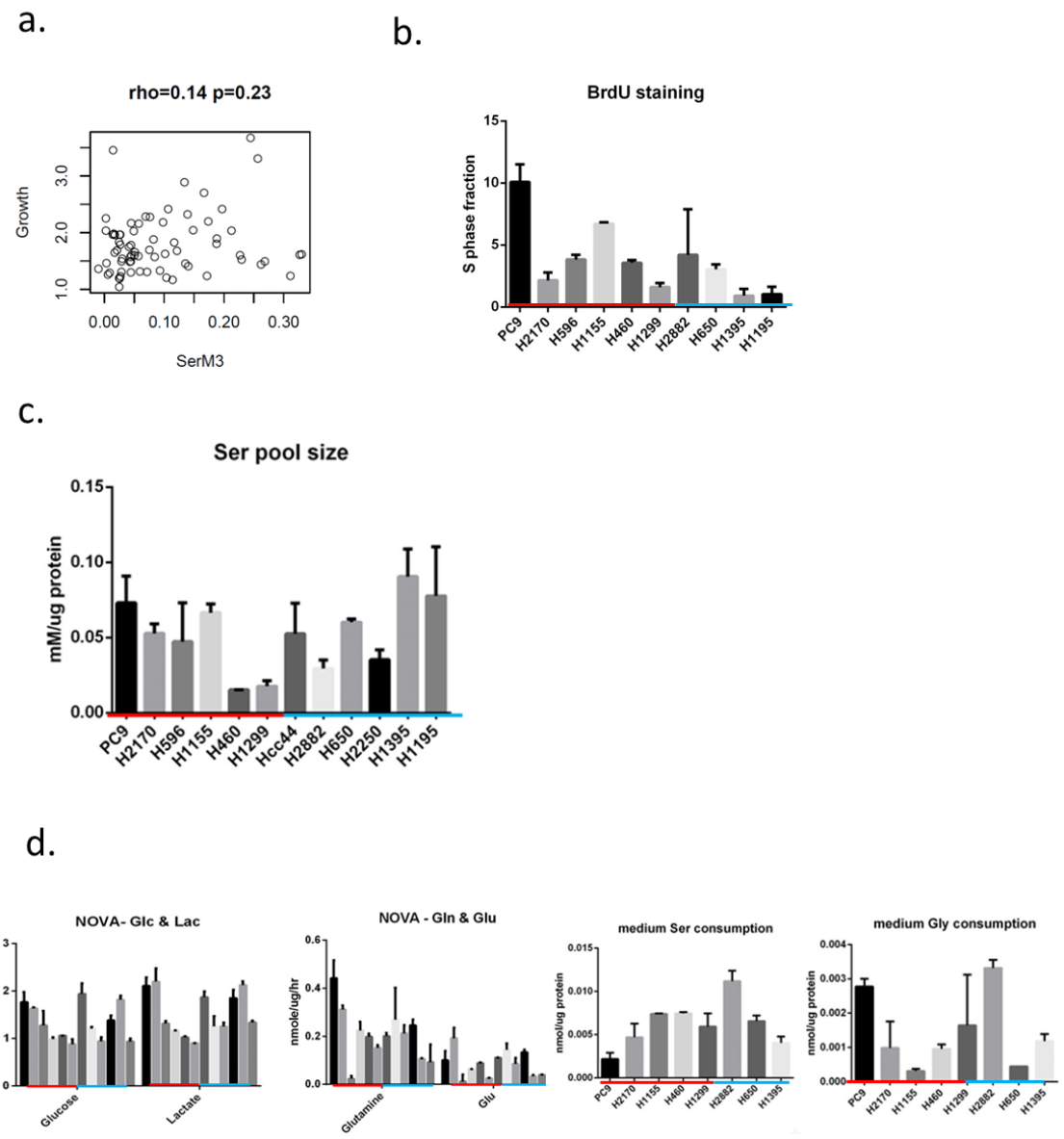
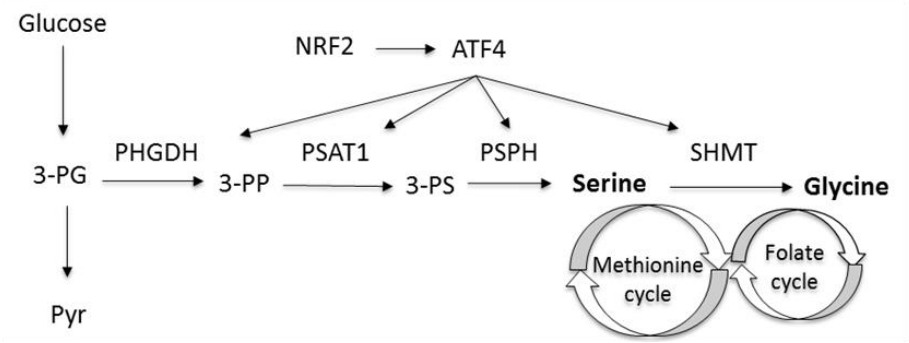


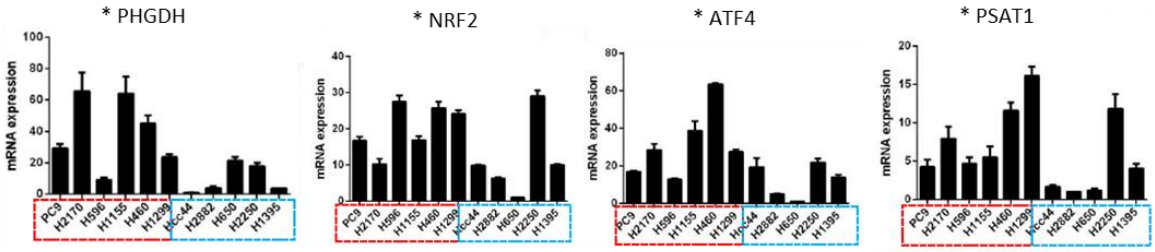
Fig 6.3 Identify the pathway that contributes to different serine enrichments

(a) Insignificant Pearson correlation coefficient and P-value was shown for the pairwise correlation between serine m+3 and cell growth rate. (b) S phase fractions of cell lines were examined by BrdU staining assay. (c) Intrinsic serine pool size was tested by isotope dilution method. Error bars represent SD. (d) Left, glucose/ glutamine utilization and lactate/ glutamate secretion were examined by NOVA. Right, serine and glycine utilization were examined amino acid analyzer. Cell lines with high Ser m+3 fractions were labeled by red while the cells with low Ser m+3 fractions were labeled by blue.

a.



b.



c.

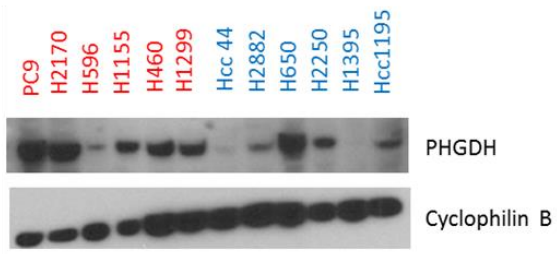
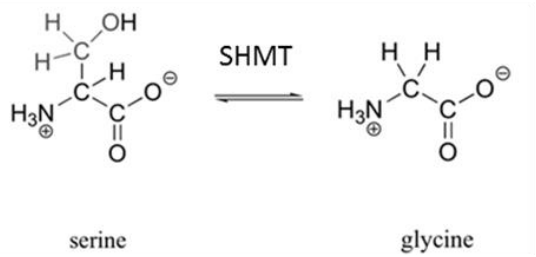


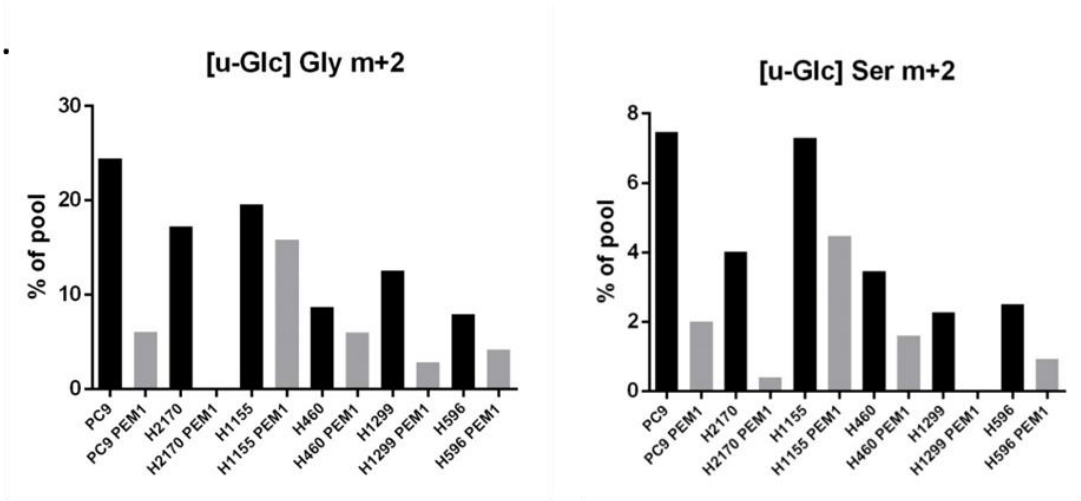
Fig 6.4 The high serine m+3 enrichment comes from de novo serine biosynthesis pathway

(a) De novo serine synthesis pathway with two upstream genes, NRF2 and ATF4. (b) mRNA expressions of genes on serine synthesis pathway. Cell lines with high Ser m+3 fractions were labeled by red while the cells with low Ser m+3 fractions were labeled by blue. *P <0.05 (c) PHGDH protein expression on Ser m+3 high lines (red) and low lines (blue).

a.


SHMT
serine
glycine

b.


[u-Glc] Gly m+2
% of pool
[u-Glc] Ser m+2
% of pool
PC9
PC9 PEM1
H2170
H2170 PEM1
H1155
H1155 PEM1
H460
H460 PEM1
H1299
H1299 PEM1
H596
H596 PEM1

c.

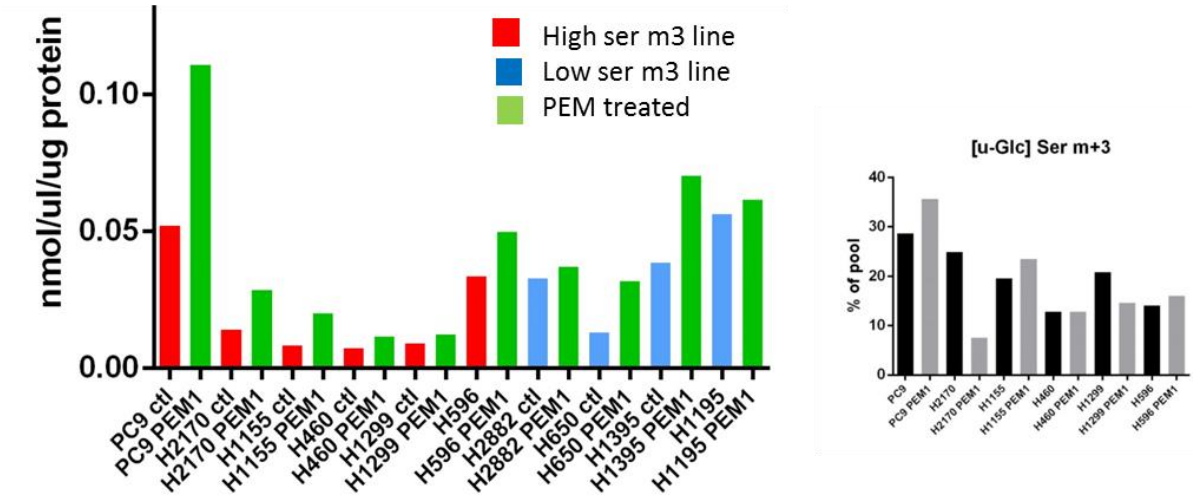

High ser m3 line
Low ser m3 line
PEM treated
nmol/ul/ug protein
PC9 ctl
PC9 PEM1
H2170 ctl
H2170 PEM1
H1155 ctl
H1155 PEM1
H460 ctl
H460 PEM1
H1299 ctl
H1299 PEM1
H596
H596 PEM1
H2882 ctl
H2882 PEM1
H650 ctl
H650 PEM1
H1395 ctl
H1395 PEM1
H1195
H1195 PEM1
[u-Glc] Ser m+3
% of pool

Fig 6.5 SHMT may be the target of Pemetrexed treatment

(a) SHMT is the enzyme convert between serine and glycine. (b) Cells were pretreated with or without 1 uM PEM for overnight, then labeled with [u-^{13}C]-glucose for 6 hour. Serine and glycine MID were analyzed. Black bars present no treatment control bars and grey bars present the PEM treatment result. (c) Cells were pretreated with or without 1 uM PEM for overnight, and the intrinsic serine pool size was tested by isotope dilution method. Green bars present the PEM treatment result.

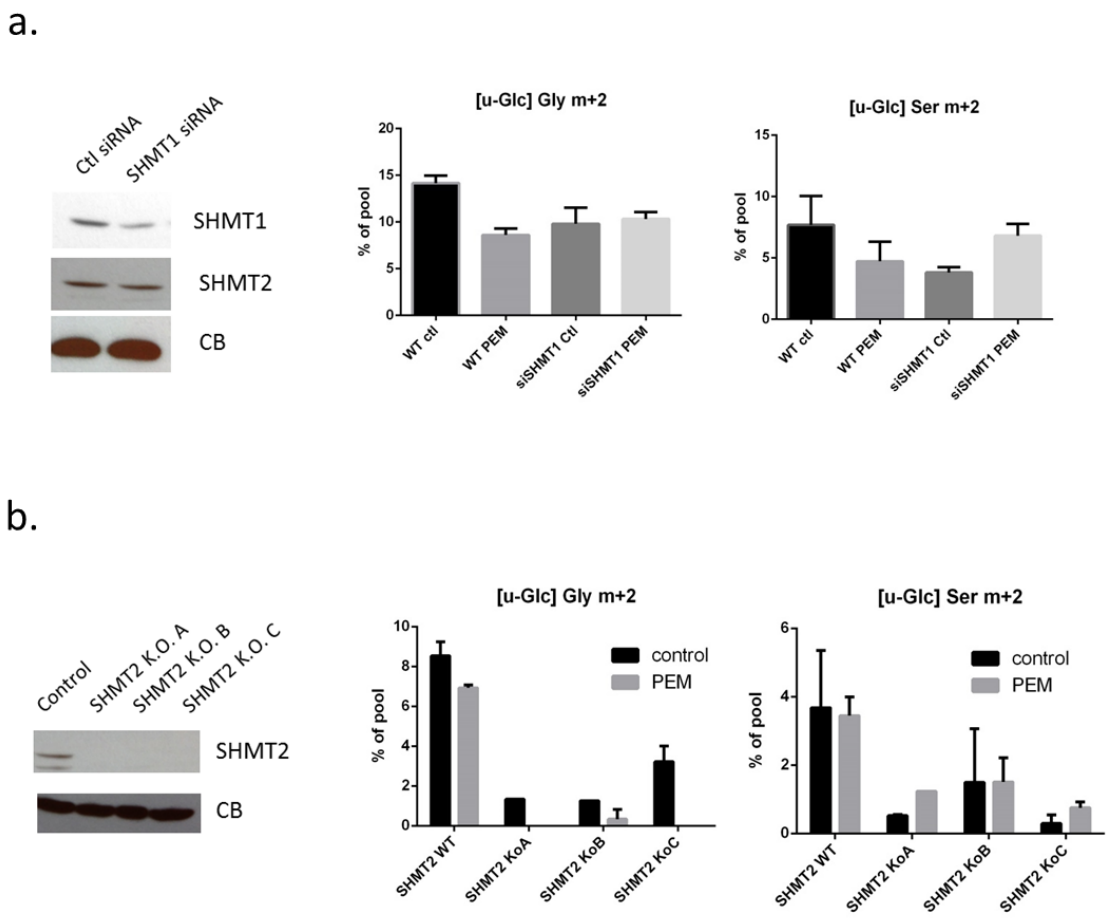


Fig 6.6 SHMT silencing phenocopies Pemetrexed treatment

(a) Left, SHMT1 knockdown was checked by western blot. Right, cells were pretreated with or without 1 uM PEM for overnight, then labeled with [u-^{13}C]-glucose for 6 hour. (b) Left, SHMT2 knockout was checked by western blot. KoA, B and C are CRISPR knockout colonies under same selection. Right, cells were pretreated with or without 1 uM PEM for overnight, then labeled with [u-^{13}C]-glucose for 6 hour.

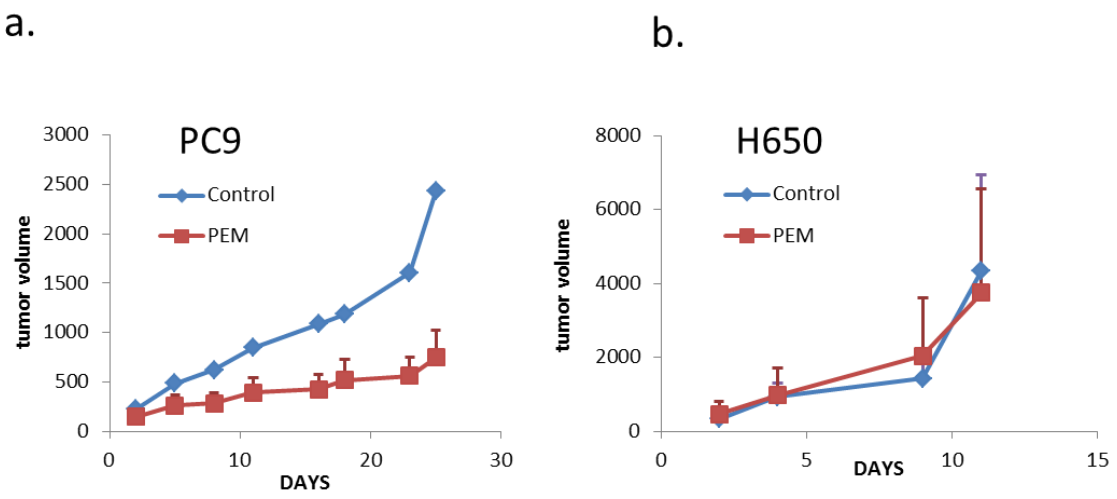


Fig 6.7 Cell lines show consistent sensitivity to PEM treatment *in vivo*

Xenograft mice were subcutaneous implanted with a high Ser m+3 cell line (a, PC9) and a low line (b, H650). The xenografts were treated with vehicle (saline) or 200 mg/kg Pemetrexed for 10 days and the tumor volume were measured. Data are the average and SEM.

CHAPTER SEVEN
Summary and future directions

Summary

In this project, I developed the most comprehensive database of cancer cell metabolism to date by using a panel NSCLC cell lines as a model. A major goal is to understand whether these metabolic phenotypes predict therapeutic vulnerabilities to novel metabolic inhibitors, targeted therapies, or conventional chemotherapeutic agents.

In order to obtain an integrated view of cancer metabolism, a highly annotated panel of more than 80 NSCLC cell lines was analyzed for about 100 metabolic parameters derived from nutrient utilization, nutrient addiction, and isotope labeling patterns. By building the metabolic profile, we characterized cell-autonomous metabolic heterogeneity in NSCLC, and compared it with orthogonal data sets including genome, proteome, and chemotherapeutic agent sensitivity. The project provided answers to three basic questions long presented in cancer metabolism field.

First, how much metabolic diversity exists among cancer cell lines grown under identical conditions?

To reveal the intrinsic diversity of NSCLC cell metabolism, I performed three metabolic assays that are the most systematic, informative and appropriate to apply on large scale analysis. I was able to extract 100 metabolic features to build an informative, quantitative and broad view of cancer metabolism profiling. Not only the stable isotope labeling assays but also nutrient utilization rates revealed astonishing diversity among NSCLC cell lines. This is the first systematic evidence to describe the highly variable metabolic phenotypes of cancer metabolism.

Second, how do metabolic features cross-correlate with each other?

Combined with bio statistical power, I analyzed the correlation between cell-autonomous metabolic features. The Spearman rank correlation of all features not only uncovered novel biological relationships, but also revealed how this data set can be used to explain or even predict many metabolic activities. We discover some critical correlations about how cells use glucose and glutamine. For example, neither glucose nor glutamine utilization has significant association with cell growth rate. We also found that the malate enrichment feature can be used as biomarker to predict pyruvate carboxylase activity and dependence.

Third, how well does the metabolic phenotype correlate with other datasets?

Analysis of orthogonal datasets provides clues to the molecular basis underlying metabolic phenotypes. In this thesis, I explained three examples. First, the

lactate/glucose ratio correlates with gene sets related to hypoxia, Myc and ribosome biogenesis. This can be considered a proof of principle since HIF1 and Myc are known to regulate glucose metabolism. Second, we found that the combined KRAS/LKB1 mutation correlated with glucose-dependent citrate MID. I then overexpressed LKB1 and found the different TCA turnover rate may explain the increased citrate m+0. For the third example, I presented the sensitivity to the folate antagonist Pemetrexed in cell lines displaying high flux through the de novo serine biosynthetic pathway. Briefly, I proved the correlations between serine flux feature and PEM drug sensitivity, found the mechanism, and revealed a novel metabolic biomarker of PEM.

Significance

Lung cancer is still the leading cause of cancer-related death worldwide, so improvements in clinical care are desperately needed. The study of NSCLC metabolism will yield biological and potentially translatable insights into the metabolic basis of transformation and tumor cell growth. This project differs from every published study in cancer metabolism in its scope in the use of targeted ^{13}C enrichment analysis as a screening tool, and in the unique opportunity to correlate the findings with genomic, gene expression, and high throughput screening data sets. This diversity produces an opportunity to derive novel functional families that cannot be recapitulated solely through analysis of the genome or transcriptome. Functional metabolic families are significant because they predict sensitivity to existing therapies and nominate new

therapeutic targets from a set of reprogrammed metabolic networks. Together, the information from this project may lead to new avenues for diagnostic imaging, therapy, and biomarker detection.

Future directions

Functional metabolic families provide novel tools to study metabolic phenotypes

Metabolic networks are highly complex and therefore there are many areas that remain undeveloped or uncovered. As we generated a large-scale and quantitative atlas of cell metabolism profiling using NSCLC, this methodology and dataset may be of broad utility for investigating cellular metabolism. For example, the 100 metabolic features can be a useful tool to further characterize mechanisms that determine cell resistance to therapy based on metabolic phenotype. Additionally, the full rank of Spearman correlations may identify more unexpected relationships between features. Moreover, it groups cell lines in novel ways. The affinity propagation clustering using metabolic features alone produced families that were largely distinct from clusters based solely on gene expression. By further examining the families, we may find novel clues to the molecular underpinnings responsible for such metabolic phenotypes. As we demonstrated in this thesis, analysis of the correlation between metabolic families and orthogonal datasets may determine new therapeutic targets or biomarkers.

Compare metabolic preferences in culture and in vivo to see if phenotypes are maintained when the cells are implanted into mice

Cell culture systems are convenient models to study regulatory mechanisms and metabolic dependencies. At the same time, it is also important to translation from cultured cancer cells to live tumors. However, to directly examine metabolic pathway utilization within intact human tumors has proven to be highly complex. The paucity of metabolic data from patient tumors has produced a disconnect between reductionist models of cell-autonomous metabolism in culture and the much more complex and clinically relevant conditions experienced by cancer cells in vivo (Hensley and DeBerardinis, 2015).

There are some papers that emphasize the prominence of mitochondrial metabolism in tumors, both in mice (Marin-Valencia et al., 2012; Yuneva et al., 2012) and humans (Fan et al., 2009; Sellers et al., 2015). For example, MYC-induced mouse liver tumors show significantly increased glucose and glutamine catabolism (Yuneva et al., 2012). By infusing ^{13}C-labeled nutrients into orthotopic glioblastomas in mice, tumors were found to utilize mitochondrial glucose oxidation during aggressive tumor growth, and many metabolic activities were conserved in cells cultured ex vivo from the tumors (Marin-Valencia et al., 2012). In NSCLC patients, using ^{13}C-glucose revealed that both glycolysis and oxidative metabolism of pyruvate in the TCA cycle were apparent in the tumor, and that the abundance of metabolites labeled by these pathways was generally higher in

the tumors compared to the lung (Fan et al., 2009). Recently, PC was reported to be highly expressed in tumors and contributed to the labeling differences between tumor and surrounding lung (Sellers et al., 2015).

In this project I determined a full range of cell-autonomous metabolic features, which urge its recapitulation in mouse pre-clinical models. For example, I implanted the high serine flux cell lines to mice and demonstrated a consistent sensitivity of cell lines to PEM treatment *in vitro* and *in vivo*. There are more interesting questions we can ask, such as choosing cell lines from the Lac/Glc list and see if the high lines show increased glycolysis in mice model, or be more sensitive to inhibitors target to HIF/Myc/ribosome biogenesis. We can also simply select top and bottom cell lines from a target feature (glutamine utilization, Cit Glc m+2, Cit Q m+5 ...) and test if the phenotypes are maintained when the cells are implanted into the mice.

The serine, glycine and one-carbon metabolism may be new target of chemotherapy

The first antimetabolites drugs were introduced in cancer therapy more than 50 years ago and are still the most widely used drugs in cancer chemotherapy. With the development of the cancer metabolism field, researchers re-focused their efforts in the development of novel antimetabolites drugs or in seeking new potential therapeutic targets (Antonov et al., 2014). Among them, the emerging role of serine, glycine and one-carbon metabolism in cancer biology opens the opportunity of alternative chemotherapeutic approaches (Amelio et al., 2014). Some preclinical studies are

underway for small molecules targeting the catalytic site of metabolic enzymes in this pathway, including PHGDH PSAT, PSPH, and GLDC (DeBerardinis, 2011; Zogg, 2014).

Being an interesting chemotherapeutic intervention target, SHMT occupies a critical position at the convergence of two key pathways, serine biosynthesis and folate cycle. SHMT converts serine to glycine and provides methyl groups for the one-carbon pool and DNA methylation. There are two SHMT genes in human genome, which are cytoplasmic form SHMT1 and mitochondrial form SHMT2 (Garrow et al., 1993). Both isoforms have been shown to be highly expressed and to be a crucial factor for the serine/glycine metabolism of several cancer cell types (Jain et al., 2012; Kim et al., 2015; Pandey et al., 2014; Paone et al., 2014). In this project, I revealed that SHMT may be a target of antifolate drug Pemetrexed, and discovered the serine flux may be a novel biomarker for the sensitivity of Pemetrexed. However, many open questions remain. For example, does PEM target to one or both SHMT isoform? Do other antifolate drugs also target SHMT or the serine biosynthesis pathway? Will the inhibition be rescued by adding back folate cycle intermediates? Will serine/glycine flux serves as biomarker for other antifolate drugs? Further studies are required to fully understand the phenotype.

BIBLIOGRAPHY

Abu-Jamous, B., Fa, R., Roberts, D.J., and Nandi, A.K. (2013). Paradigm of tunable clustering using Binarization of Consensus Partition Matrices (Bi-CoPaM) for gene discovery. PloS one *8*, e56432.

Amelio, I., Cutruzzola, F., Antonov, A., Agostini, M., and Melino, G. (2014). Serine and glycine metabolism in cancer. Trends in biochemical sciences *39*, 191-198.

Antonov, A., Agostini, M., Morello, M., Minieri, M., Melino, G., and Amelio, I. (2014). Bioinformatics analysis of the serine and glycine pathway in cancer cells. Oncotarget *5*, 11004-11013.

Boon, K., Caron, H.N., van Asperen, R., Valentijn, L., Hermus, M.C., van Sluis, P., Roobeek, I., Weis, I., Voute, P.A., Schwab, M., *et al.* (2001). N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. The EMBO journal *20*, 1383-1393.

Carretero, J., Medina, P.P., Blanco, R., Smit, L., Tang, M., Roncador, G., Maestre, L., Conde, E., Lopez-Rios, F., Clevers, H.C., *et al.* (2007). Dysfunctional AMPK activity, signalling through mTOR and survival in response to energetic stress in LKB1-deficient lung cancer. Oncogene *26*, 1616-1625.

Chabner, B.A., and Roberts, T.G., Jr. (2005). Timeline: Chemotherapy and the war on cancer. Nature reviews Cancer *5*, 65-72.

Chan, B.A., and Hughes, B.G. (2015). Targeted therapy for non-small cell lung cancer: current standards and the promise of the future. Translational lung cancer research *4*, 36-54.

Chaneton, B., Hillmann, P., Zheng, L., Martin, A.C., Maddocks, O.D., Chokkathukalam, A., Coyle, J.E., Jankevics, A., Holding, F.P., Vousden, K.H., *et al.* (2012). Serine is a natural ligand and allosteric activator of pyruvate kinase M2. Nature *491*, 458-462.

Chen, Z., Cheng, K., Walton, Z., Wang, Y., Ebi, H., Shimamura, T., Liu, Y., Tupper, T., Ouyang, J., Li, J., *et al.* (2012). A murine lung cancer co-clinical trial identifies genetic modifiers of therapeutic response. Nature *483*, 613-617.

Cheng, T., Sudderth, J., Yang, C., Mullen, A.R., Jin, E.S., Mates, J.M., and DeBerardinis, R.J. (2011). Pyruvate carboxylase is required for glutamine-independent growth of tumor cells. Proceedings of the National Academy of Sciences of the United States of America *108*, 8674-8679.

Cipollina, C., van den Brink, J., Daran-Lapujade, P., Pronk, J.T., Porro, D., and de Winde, J.H. (2008). Saccharomyces cerevisiae SFP1: at the crossroads of central metabolism and ribosome biogenesis. Microbiology *154*, 1686-1699.

Clarke, R., Ressom, H.W., Wang, A., Xuan, J., Liu, M.C., Gehan, E.A., and Wang, Y. (2008). The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. Nature reviews Cancer *8*, 37-49.

Cohen, M.H., Justice, R., and Pazdur, R. (2009). Approval summary: pemetrexed in the initial treatment of advanced/metastatic non-small cell lung cancer. The oncologist *14*, 930-935.

Daidone, F., Florio, R., Rinaldo, S., Contestabile, R., di Salvo, M.L., Cutruzzola, F., Bossa, F., and Paiardini, A. (2011). In silico and in vitro validation of serine hydroxymethyltransferase as a chemotherapeutic target of the antifolate drug pemetrexed. European journal of medicinal chemistry *46*, 1616-1621.
Dang, C.V., Kim, J.W., Gao, P., and Yustein, J. (2008). The interplay between MYC and HIF in cancer. Nature reviews Cancer *8*, 51-56.
Dang, C.V., Le, A., and Gao, P. (2009). MYC-induced cancer cell energy metabolism and therapeutic opportunities. Clinical cancer research : an official journal of the American Association for Cancer Research *15*, 6479-6483.
DeBerardinis, R.J. (2011). Serine metabolism: some tumors take the road less traveled. Cell metabolism *14*, 285-286.
DeBerardinis, R.J., Lum, J.J., Hatzivassiliou, G., and Thompson, C.B. (2008a). The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. Cell metabolism *7*, 11-20.
DeBerardinis, R.J., Mancuso, A., Daikhin, E., Nissim, I., Yudkoff, M., Wehrli, S., and Thompson, C.B. (2007). Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. Proceedings of the National Academy of Sciences of the United States of America *104*, 19345-19350.
Deberardinis, R.J., Sayed, N., Ditsworth, D., and Thompson, C.B. (2008b). Brick by brick: metabolism and tumor cell growth. Current opinion in genetics & development *18*, 54-61.
Dong, Y., Bey, E.A., Li, L.S., Kabbani, W., Yan, J., Xie, X.J., Hsieh, J.T., Gao, J., and Boothman, D.A. (2010). Prostate cancer radiosensitization through poly(ADP-Ribose) polymerase-1 hyperactivation. Cancer research *70*, 8088-8096.
Eilers, M., and Eisenman, R.N. (2008). Myc's broad reach. Genes & development *22*, 2755-2766.
Fan, T.W., Lane, A.N., Higashi, R.M., Farag, M.A., Gao, H., Bousamra, M., and Miller, D.M. (2009). Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM). Molecular cancer *8*, 41.
Fediuc, S., Gaidhu, M.P., and Ceddia, R.B. (2006). Regulation of AMP-activated protein kinase and acetyl-CoA carboxylase phosphorylation by palmitate in skeletal muscle cells. Journal of lipid research *47*, 412-420.
Fernandez, C.A., Des Rosiers, C., Previs, S.F., David, F., and Brunengraber, H. (1996). Correction of 13C mass isotopomer distributions for natural stable isotope abundance. Journal of mass spectrometry : JMS *31*, 255-262.
Gaglio, D., Metallo, C.M., Gameiro, P.A., Hiller, K., Danna, L.S., Balestrieri, C., Alberghina, L., Stephanopoulos, G., and Chiaradonna, F. (2011). Oncogenic K-Ras decouples glucose and glutamine metabolism to support cancer cell growth. Molecular systems biology *7*, 523.
Gambhir, S.S. (2002). Molecular imaging of cancer with positron emission tomography. Nature reviews Cancer *2*, 683-693.
Garrow, T.A., Brenner, A.A., Whitehead, V.M., Chen, X.N., Duncan, R.G., Korenberg, J.R., and Shane, B. (1993). Cloning of human cDNAs encoding mitochondrial and

cytosolic serine hydroxymethyltransferases and chromosomal localization. The Journal of biological chemistry *268*, 11910-11916.
Gibbons, F.D., and Roth, F.P. (2002). Judging the quality of gene expression-based clustering methods using gene annotation. Genome research *12*, 1574-1581.
Gower, A., Wang, Y., and Giaccone, G. (2014). Oncogenic drivers, targeted therapies, and acquired resistance in non-small-cell lung cancer. Journal of molecular medicine *92*, 697-707.
Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.
Hensley, C.T., and DeBerardinis, R.J. (2015). In vivo analysis of lung cancer metabolism: nothing like the real thing. The Journal of clinical investigation *125*, 495-497.
Hrustanovic, G., Lee, B.J., and Bivona, T.G. (2013). Mechanisms of resistance to EGFR targeted therapies. Cancer biology & therapy *14*, 304-314.
Hsu, P.P., and Sabatini, D.M. (2008). Cancer cell metabolism: Warburg and beyond. Cell *134*, 703-707.
Jain, M., Nilsson, R., Sharma, S., Madhusudhan, N., Kitami, T., Souza, A.L., Kafri, R., Kirschner, M.W., Clish, C.B., and Mootha, V.K. (2012). Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. Science *336*, 1040-1044.
Jemal, A., Siegel, R., Xu, J., and Ward, E. (2010). Cancer statistics, 2010. CA: a cancer journal for clinicians *60*, 277-300.
Ji, H., Wu, G., Zhan, X., Nolan, A., Koh, C., De Marzo, A., Doan, H.M., Fan, J., Cheadle, C., Fallahi, M., *et al.* (2011). Cell-type independent MYC target genes reveal a primordial signature involved in biomass accumulation. PloS one *6*, e26057.
Kalhan, S.C., and Hanson, R.W. (2012). Resurgence of serine: an often neglected but indispensable amino Acid. The Journal of biological chemistry *287*, 19786-19791.
Karnoub, A.E., and Weinberg, R.A. (2008). Ras oncogenes: split personalities. Nature reviews Molecular cell biology *9*, 517-531.
Kim, D., Fiske, B.P., Birsoy, K., Freinkman, E., Kami, K., Possemato, R.L., Chudnovsky, Y., Pacold, M.E., Chen, W.W., Cantor, J.R., *et al.* (2015). SHMT2 drives glioma cell survival in ischaemia but imposes a dependence on glycine clearance. Nature *520*, 363-367.
Kim, H.S., Mendiratta, S., Kim, J., Pecot, C.V., Larsen, J.E., Zubovych, I., Seo, B.Y., Kim, J., Eskiocak, B., Chung, H., *et al.* (2013). Systematic identification of molecular subtype-selective vulnerabilities in non-small-cell lung cancer. Cell *155*, 552-566.
Li, F., Han, X., Li, F., Wang, R., Wang, H., Gao, Y., Wang, X., Fang, Z., Zhang, W., Yao, S., *et al.* (2015). LKB1 Inactivation Elicits a Redox Imbalance to Modulate Non-small Cell Lung Cancer Plasticity and Therapeutic Response. Cancer cell *27*, 698-711.
Liang, X., Nan, K.J., Li, Z.L., and Xu, Q.Z. (2009). Overexpression of the LKB1 gene inhibits lung carcinoma cell proliferation partly through degradation of c-myc protein. Oncology reports *21*, 925-931.
Liang, X., Wang, P., Gao, Q., and Tao, X. (2014). Exogenous activation of LKB1/AMPK signaling induces G(1) arrest in cells with endogenous LKB1 expression. Molecular medicine reports *9*, 1019-1024.

Lindeman, N.I., Cagle, P.T., Beasley, M.B., Chitale, D.A., Dacic, S., Giaccone, G., Jenkins, R.B., Kwiatkowski, D.J., Saldivar, J.S., Squire, J., *et al.* (2013). Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. The Journal of molecular diagnostics : JMD *15*, 415-453.
Liu, Y., Marks, K., Cowley, G.S., Carretero, J., Liu, Q., Nieland, T.J., Xu, C., Cohoon, T.J., Gao, P., Zhang, Y., *et al.* (2013a). Metabolic and functional genomic studies identify deoxythymidylate kinase as a target in LKB1-mutant lung cancer. Cancer discovery *3*, 870-879.
Liu, Y., Yin, T.J., Zhou, R., Zhou, S., Fan, L., and Zhang, R.G. (2013b). Expression of thymidylate synthase predicts clinical outcomes of pemetrexed-containing chemotherapy for non-small-cell lung cancer: a systemic review and meta-analysis. Cancer chemotherapy and pharmacology *72*, 1125-1132.
Locasale, J.W. (2013). Serine, glycine and one-carbon units: cancer metabolism in full circle. Nature reviews Cancer *13*, 572-583.
Locasale, J.W., Grassian, A.R., Melman, T., Lyssiotis, C.A., Mattaini, K.R., Bass, A.J., Heffron, G., Metallo, C.M., Muranen, T., Sharfi, H., *et al.* (2011). Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. Nature genetics *43*, 869-874.
Maddocks, O.D., Berkers, C.R., Mason, S.M., Zheng, L., Blyth, K., Gottlieb, E., and Vousden, K.H. (2013). Serine starvation induces stress and p53-dependent metabolic remodelling in cancer cells. Nature *493*, 542-546.
Mahoney, C.L., Choudhury, B., Davies, H., Edkins, S., Greenman, C., Haaften, G., Mironenko, T., Santarius, T., Stevens, C., Stratton, M.R., *et al.* (2009). LKB1/KRAS mutant lung cancers constitute a genetic subset of NSCLC with increased sensitivity to MAPK and mTOR signalling inhibition. British journal of cancer *100*, 370-375.
Marin-Valencia, I., Yang, C., Mashimo, T., Cho, S., Baek, H., Yang, X.L., Rajagopalan, K.N., Maddie, M., Vemireddy, V., Zhao, Z., *et al.* (2012). Analysis of tumor metabolism reveals mitochondrial glucose oxidation in genetically diverse human glioblastomas in the mouse brain in vivo. Cell metabolism *15*, 827-837.
Matikas, A., Mistriotis, D., Georgoulias, V., and Kotsakis, A. (2015). Current and Future Approaches in the Management of Non-Small-Cell Lung Cancer Patients With Resistance to EGFR TKIs. Clinical lung cancer *16*, 252-261.
Mazurek, S., and Eigenbrodt, E. (2003). The tumor metabolome. Anticancer research *23*, 1149-1154.
Mehrmohamadi, M., Liu, X., Shestov, A.A., and Locasale, J.W. (2014). Characterization of the usage of the serine metabolic network in human cancer. Cell reports *9*, 1507-1519.
Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., *et al.* (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature genetics *34*, 267-273.

Mullen, A.R., Wheaton, W.W., Jin, E.S., Chen, P.H., Sullivan, L.B., Cheng, T., Yang, Y., Linehan, W.M., Chandel, N.S., and DeBerardinis, R.J. (2012). Reductive carboxylation supports growth in tumour cells with defective mitochondria. Nature *481*, 385-388.
Niederst, M.J., and Engelman, J.A. (2013). Bypass mechanisms of resistance to receptor tyrosine kinase inhibition in lung cancer. Science signaling *6*, re6.
Nilsson, R., Jain, M., Madhusudhan, N., Sheppard, N.G., Strittmatter, L., Kampf, C., Huang, J., Asplund, A., and Mootha, V.K. (2014). Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. Nature communications *5*, 3128.
O'Neill, H.M., Holloway, G.P., and Steinberg, G.R. (2013). AMPK regulation of fatty acid metabolism and mitochondrial biogenesis: implications for obesity. Molecular and cellular endocrinology *366*, 135-151.
Paiardini, A., Fiascarelli, A., Rinaldo, S., Daidone, F., Giardina, G., Koes, D.R., Parroni, A., Montini, G., Marani, M., Paone, A., *et al.* (2015). Screening and in vitro testing of antifolate inhibitors of human cytosolic serine hydroxymethyltransferase. ChemMedChem *10*, 490-497.
Pandey, S., Garg, P., Lee, S., Choung, H.W., Choung, Y.H., Choung, P.H., and Chung, J.H. (2014). Nucleotide biosynthesis arrest by silencing SHMT1 function via vitamin B6-coupled vector and effects on tumor growth inhibition. Biomaterials *35*, 9332-9342.
Pao, W., and Chmielecki, J. (2010). Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. Nature reviews Cancer *10*, 760-774.
Paone, A., Marani, M., Fiascarelli, A., Rinaldo, S., Giardina, G., Contestabile, R., Paiardini, A., and Cutruzzola, F. (2014). SHMT1 knockdown induces apoptosis in lung cancer cells by causing uracil misincorporation. Cell death & disease *5*, e1525.
Parlo, R.A., and Coleman, P.S. (1984). Enhanced rate of citrate export from cholesterol-rich hepatoma mitochondria. The truncated Krebs cycle and other metabolic ramifications of mitochondrial membrane cholesterol. The Journal of biological chemistry *259*, 9997-10003.
Possemato, R., Marks, K.M., Shaul, Y.D., Pacold, M.E., Kim, D., Birsoy, K., Sethumadhavan, S., Woo, H.K., Jang, H.G., Jha, A.K., *et al.* (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature *476*, 346-350.
Rapp, E., Pater, J.L., Willan, A., Cormier, Y., Murray, N., Evans, W.K., Hodson, D.I., Clark, D.A., Feld, R., Arnold, A.M., *et al.* (1988). Chemotherapy can prolong survival in patients with advanced non-small-cell lung cancer--report of a Canadian multicenter randomized trial. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *6*, 633-641.
Ruvinsky, I., Sharon, N., Lerer, T., Cohen, H., Stolovich-Rain, M., Nir, T., Dor, Y., Zisman, P., and Meyuhas, O. (2005). Ribosomal protein S6 phosphorylation is a determinant of cell size and glucose homeostasis. Genes & development *19*, 2199-2211.
Sanchez-Cespedes, M., Parrella, P., Esteller, M., Nomoto, S., Trink, B., Engles, J.M., Westra, W.H., Herman, J.G., and Sidransky, D. (2002). Inactivation of LKB1/STK11 is a common event in adenocarcinomas of the lung. Cancer research *62*, 3659-3662.

Scheffzek, K., Ahmadian, M.R., Kabsch, W., Wiesmuller, L., Lautwein, A., Schmitz, F., and Wittinghofer, A. (1997). The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. Science *277*, 333-338.
Scheidig, A.J., Burmester, C., and Goody, R.S. (1999). The pre-hydrolysis state of p21(ras) in complex with GTP: new insights into the role of water molecules in the GTP hydrolysis reaction of ras-like proteins. Structure *7*, 1311-1324.
Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467-470.
Sebastian, C., Zwaans, B.M., Silberman, D.M., Gymrek, M., Goren, A., Zhong, L., Ram, O., Truelove, J., Guimaraes, A.R., Toiber, D., *et al.* (2012). The histone deacetylase SIRT6 is a tumor suppressor that controls cancer metabolism. Cell *151*, 1185-1199.
Sellers, K., Fox, M.P., Bousamra, M., 2nd, Slone, S.P., Higashi, R.M., Miller, D.M., Wang, Y., Yan, J., Yuneva, M.O., Deshpande, R., *et al.* (2015). Pyruvate carboxylase is critical for non-small-cell lung cancer proliferation. The Journal of clinical investigation *125*, 687-698.
Shajahan-Haq, A.N., Cheema, M.S., and Clarke, R. (2015). Application of metabolomics in drug resistant breast cancer research. Metabolites *5*, 100-118.
Shaw, R.J., Kosmatka, M., Bardeesy, N., Hurley, R.L., Witters, L.A., DePinho, R.A., and Cantley, L.C. (2004). The tumor suppressor LKB1 kinase directly activates AMP-activated kinase and regulates apoptosis in response to energy stress. Proceedings of the National Academy of Sciences of the United States of America *101*, 3329-3335.
Skoulidis, F., Byers, L.A., Diao, L., Papadimitrakopoulou, V.A., Tong, P., Izzo, J., Behrens, C., Kadara, H., Parra, E.R., Rodriguez-Canales, J., *et al.* (2015). Co-occurring genomic alterations define major subsets of KRAS - mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities. Cancer discovery.
Su, X., and Abumrad, N.A. (2009). Cellular fatty acid uptake: a pathway under construction. Trends in endocrinology and metabolism: TEM *20*, 72-77.
Thunnissen, E., van der Oord, K., and den Bakker, M. (2014). Prognostic and predictive biomarkers in lung cancer. A review. Virchows Archiv : an international journal of pathology *464*, 347-358.
Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science *324*, 1029-1033.
Whang, Y.M., Park, S.I., Trenary, I.A., Egnatchik, R.A., Fessel, J.P., Kaufman, J.M., Carbone, D.P., and Young, J.D. (2015). LKB1 deficiency enhances sensitivity to energetic stress induced by erlotinib treatment in non-small-cell lung cancer (NSCLC) cells. Oncogene.
Winter, S.C., Buffa, F.M., Silva, P., Miller, C., Valentine, H.R., Turley, H., Shah, K.A., Cox, G.J., Corbridge, R.J., Homer, J.J., *et al.* (2007). Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. Cancer research *67*, 3441-3449.
Wise, D.R., DeBerardinis, R.J., Mancuso, A., Sayed, N., Zhang, X.Y., Pfeiffer, H.K., Nissim, I., Daikhin, E., Yudkoff, M., McMahon, S.B., *et al.* (2008). Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to

glutamine addiction. Proceedings of the National Academy of Sciences of the United States of America *105*, 18782-18787.
Woods, A., Johnstone, S.R., Dickerson, K., Leiper, F.C., Fryer, L.G., Neumann, D., Schlattner, U., Wallimann, T., Carlson, M., and Carling, D. (2003). LKB1 is the upstream kinase in the AMP-activated protein kinase cascade. Current biology : CB *13*, 2004-2008.
Xiang, X., Saha, A.K., Wen, R., Ruderman, N.B., and Luo, Z. (2004). AMP-activated protein kinase activators can inhibit the growth of prostate cancer cells by multiple mechanisms. Biochemical and biophysical research communications *321*, 161-167.
Yang, C., Sudderth, J., Dang, T., Bachoo, R.M., McDonald, J.G., and DeBerardinis, R.J. (2009). Glioblastoma cells require glutamate dehydrogenase to survive impairments of glucose metabolism or Akt signaling. Cancer research *69*, 7986-7993.
Yuneva, M.O., Fan, T.W., Allen, T.D., Higashi, R.M., Ferraris, D.V., Tsukamoto, T., Mates, J.M., Alonso, F.J., Wang, C., Seo, Y., *et al.* (2012). The metabolic profile of tumors depends on both the responsible genetic lesion and tissue type. Cell metabolism *15*, 157-170.
Zeng, W., Liu, P., Pan, W., Singh, S.R., and Wei, Y. (2015). Hypoxia and hypoxia inducible factors in tumor metabolism. Cancer letters *356*, 263-267.
Zhang, W.C., Shyh-Chang, N., Yang, H., Rai, A., Umashankar, S., Ma, S., Soh, B.S., Sun, L.L., Tai, B.C., Nga, M.E., *et al.* (2012). Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. Cell *148*, 259-272.
Zhao, Y., Butler, E.B., and Tan, M. (2013). Targeting cellular metabolism to improve cancer therapeutics. Cell death & disease *4*, e532.
Zogg, C.K. (2014). Phosphoglycerate dehydrogenase: potential therapeutic target and putative metabolic oncogene. Journal of oncology *2014*, 524101.
Zwaans, B.M., and Lombard, D.B. (2014). Interplay between sirtuins, MYC and hypoxia-inducible factor in cancer-associated metabolic reprogramming. Disease models & mechanisms *7*, 1023-1032.