Cell type specific labeling with amino acid precursors

by

Nicholas P. Gauthier

A Dissertation

Presented to the Faculty of the Louis V. Gerstner, Jr.

Graduate School of Biomedical Sciences,

Memorial Sloan-Kettering Cancer Center

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

New York, NY

December, 2012

Sander

Chris Sander, Ph.D.
Dissertation Mentor

December 3, 2012

Date

Acknowledgments

First and foremost, I would like to express my deepest appreciation to my mentor Dr. Chris Sander. Without his guidance, inspiration, and support, this work would never have been possible. Aside from his absolute dedication to science, on a personal level Chris has consistently worked to alleviate and understand the human issues that we face outside of the laboratory. I am absolutely grateful for the opportunities he has provided me and will forever be thankful for his supervision.

Dr. Martin Miller has been a best friend and scientific equal during this journey. He has contributed an untold amount to this project, taught me a great deal about all aspects of science, and his dedication and collaboration to this work has been invaluable.

I would also like to thank my Thesis Committee members Dr. Johanna Joyce and Dr. Andrew Koff. Both of whom have challenged and inspired not only this work, but also my understanding of life and science in general.

Dr. Boumediene Soufi at the Proteome Center of Tübingen has been a valuable experimental contributor, patient teacher (explaining the details of proteomics and mass spectrometry), and a personal friend.

A very special thanks to the those who collaborated and help produce the data that made this work possible. In particular, Dr. Konstantinos Mavrakis, Dr. Virginia Pedicord, and Bill Walkowicz at MSKCC have provided an enormous amount of experimental support. Our collaboration with Dr. Boris Maceks group at the Proteome Center of Tübingen has also been very fruitful and I am grateful for all of their help.

Abstract

Tissue development, homeostasis, and pathogenesis involve complex signaling between many cell types through both secreted factors and direct contact. There are currently no high-throughput methodologies to determine the cell-of-origin of proteins in multicellular environments. To address this limitation, we have developed a technique that selectively and continuously labels the proteome of individual cell types in co-culture. By expressing exogenous amino acid biosynthesis enzymes, we enable vertebrate cells to produce and metabolically incorporate canonical amino acids from labeled precursors. We have named this method **C**ell **T**ype specific labeling with **A**mino acid **P**recursors (CTAP). Using this new method, we demonstrate the ability to differentially label the proteome of distinct populations in co-culture and determine the relative expression of proteins from each cell type by quantitative mass spectrometry. Previously not accessible using other methods, we identify the cell-of-origin for secreted factors in a multicellular environment, further emphasizing the potential applicability of CTAP to a wide variety of scientific applications including biomarker discovery.

Contents

List of Figures

List of Tables

List of Abbreviations

2D-PAGE	Two-dimensional Polyacrylamide Gel Electrophoresis
BMDC	Bone marrow derived cells
CA125	Cancer Antigen 125
CID	Collision-Induced Dissociation
CRTL1	Cartilage-link protein
CTAP	Cell Type specific labeling with Amino acid Precursors
CTSZ	Cathepsin X
CyTOF	Mass cytometry
Da	Dalton
DAP	Meso-2,6-diaminopimelate
DDC	Diaminopimelate Decarboxylase
DMEM	Dulbecco's Modified Eagle's Medium
DTT	dithiothreitol
EGF	Epidermal growth factor
ESI	Electrospray Ionization
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
FGF	Fibroblast Growth Factor
Flow	Flow cytometry
GEF	guanine exchange factor
GPCR	G Protein-Coupled Receptor
GFP	Green Fluorescent Protein
H	Heavy isotopically labeled
H/L Ratio	Ratio of heavy to light isotopes
HGF	Hepatocyte Growth Factor
ICAT	Isotope-Coded Affinity Tag
IF	Immunofluorescence
IGF	Insulin-like growth factor
IHC	Immunohistochemistry
iTRAQ	Isobaric Tags for Relative and Absolute Quantitation
K0	Light L-lysine
K4	Medium $[^2H_4]$L-lysine

K8	Heavy [$^{13}C_6$,$^{15}N_2$]L-lysine
L	Light unlabeled
LC-MS/MS	Liquid Chromatography Tandem Mass Spectrometry
LGALS3BP	Galectin-3BP
lyr	Lysine racemase
m/z	Mass-to-charge
MALDI	Matrix-Assisted Laser Desorption/Ionization
MDSC	myeloid-derived suppressor cells
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MSC	Mesenchymal Stem Cells
MTS	Mitochondrial Targeting Sequence
N2A	N2-Acetyl-L-lysine
N2AAS	N2-Acetyl-L-aminoadipate semialdehyde
N2DCL	N2-(D-1-carboxyethyl)-L-lysine
N6A	N6-Acetyl-L-Lysine
NLS	Nuclear Localization Sequence
PAM	Protein antibody microarry
pcc	Pearson Correlation Coefficients
PSA	Prostate Specific Antigen
RTK	receptor tyrosine kinase
RNAi	RNA Interference
RPPA	Reverse Phase Protein Array
rsq	R-squared values
SERPINA3	Serpin A3
SILAC	Stable Isotope Labeling of Amino Acids in Cell Culture
SPARC	Osteonectin
TAIR	The Arabidopsis Information Resource
TEM	TIE2-expressing monocytes
VEGF	Vascular endothelial growth factor
WB	Western Blot
Y2H	Yeast two-hybrid

Chapter 1

Introduction

The complexity of multicellular environments is dependent on extensive cell-cell communication. Such communication is fundamental to the development of all organisms and when perturbed can lead to a range of diseases. Studies to elucidate the molecules involved in cell-cell communication would benefit from the development of new methods to comprehensively profile the secretome and intracellular proteome of distinct cell types cultured in multicellular environments.

In order to study the signaling molecules involved in tumor-stroma interactions, researchers have traditionally relied on low-throughput antibody-based staining together with microscopy and flow cytometry to identify distinct cell types in multicellular environments and quantify the proteins they contain. A major drawback of this approach is that researchers must choose the proteins they believe to be involved in a particular phenotype. Another weakness is that staining methods, although both informative and well established, require preselection of antibodies and one can only expect to identify a handful of the many proteins in any given cell type. Genetic screens have established some molecules involved in multicellular communication. However, these screens are labor intensive and

whether any individual identified molecule is acting in a cell-specific manner is largely speculation. Consequently, although progress has been made in identifying a handful of regulators of heterotypic cell-cell communication, one must wonder it there is a better way to identify others.

For cancer, it has long been appreciated that solid tumors consist of cancer cells and a variety of non-malignant "normal" cells (**Figure 1.1**). This tumor microenvironment plays a significant role in all aspects of tumor development and maintenance [1–3]. Recruited and resident stromal cells have also been shown to attenuate the effect of chemotherapeutic and targeted therapies [4–7]. Consequently, novel therapeutic approaches might one day be able to target signaling between cancer cells and their microenvironment. At this time, however, the molecules involved are largely undefined. The ability to comprehensively characterize intra- and intercellular communication between tumor and stroma would be helpful in elucidating the mechanisms that underlie stromal influence on tumorigenesis. In contrast to genetic and antibody-based methods, modern mass spectrometry (MS) can identify thousands of proteins in a single sample, but lacks the ability to determine from which cell type any given protein originated. The research described here is motivated by the desire to perform more comprehensive mapping of proteins in multicellular systems with MS, while retaining the ability to discriminate the cell-of-origin of identified factors.

This chapter is composed of several sections. First, an introduction to cell-cell communication is provided. Second, various genetic and proteomics methods used to identify and characterize inter- and intracellular molecular machinery are described. Third, I present other techniques that have aimed to comprehensively identify and determine the origin of proteins in multicellular environments. Finally, a short introduction is provided to the method we developed, titled **C**ell **T**ype specific labeling with **A**mino acid **P**recursors
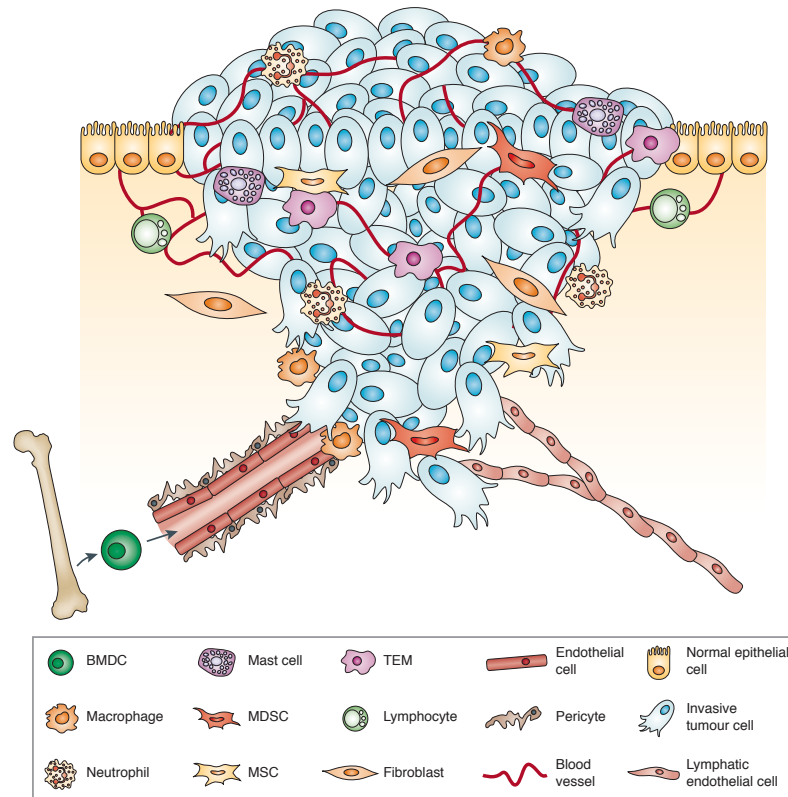
Figure 1.1: The tumor microenvironment is composed of a variety of cell types. Tumors are not only composed of tumor cells, but also contain a variety of non-neoplastic normal cells. These "normal" cells are known to play a variety of roles in tumor development, maintenance, and response to therapy. BMDC, bone marrow derived cells; MDSC, myeloid-derived suppressor cells; MSC, mesenchymal stem cells; TEM, TIE2-expressing monocytes. Figure adapted from Joyce and Pollard [8].

(CTAP), which allows for quantitative cell-of-origin assignment of proteins identified by MS in multicellular culture.

1.1   Cell-cell communication

Cells do not exist in a vacuum. In natural environments organisms are constantly in contact or close proximity to one another. Even in laboratory controlled systems, a single cell is rarely kept in complete isolation. The ability to sense, respond, and adapt to environmental changes is essential to any living organism. Sharing information and coordinating

distinct tasks offer many evolutionary advantages and it is therefore not surprising that cells have evolved the ability to detect and communicate with their neighbors . Unicellular organisms communicate with one another in order to coordinate a variety of tasks such as mating or reacting to changes in the environment [9, 10]. Multicellular organisms rely heavily on tightly-controlled communication in order to differentiate cells into functional subsets, coordinate the development and control of complex tissues, and mount responses to invading pathogens or rogue self [2, 11–14]. These gross phenotypic outcomes emerge from the assembly of a large set of complex molecular interactions and exactly how this occurs is an area of intense study.

There are a large number of molecules that cells use to "talk" to one another. Detection of these molecules generally leads to the propagation of the "signal" through a set of intermediates and ultimately culminates in a cellular decision. Cells are constantly inundated with multiple signals and, depending on their specific cellular context, the integration of these signals dictates the systematic cellular reaction. Our understanding of the molecular details of these processes has been worked out and validated over years of detailed study in very specific contexts on a molecule-by-molecule basis. The purpose of this section is to give a broad overview of what we know about how cell-cell communication works and how it leads to important phenotypic outcomes.

1.1.1 How do cells talk to each other?

The cells within multicellular organisms have developed a diverse toolkit for communicating with one another. There are a variety of small molecules that are able to instigate reactions in receiver cells and these molecules can be transmitted over both long and short distances. These signals can be almost anything that naturally exists within the organ-

ism and include molecules like ions, hormones, peptides, and proteins. Through the large repertoire of transmitting molecules, spatial and temporal controls, and by restricting the cells that can receive signals, organisms are able to fine-tune messages to distinct cells in very specific locations.

The classification of signaling is usually based on the distance over which a signaling molecule acts. Juxtacrine signaling is signaling that occurs through direct contact. Some examples of juxtacrine signaling include Ephrin and Notch signaling, both of which involve transmembrane protein interactions between adjacent cells [15, 16]. In addition, some adjacent cells also form junctions between their membranes that allow the passage of electrical current or small molecules in the cytoplasm. For example, cardiac cells exchange current through gap junctions, which coordinates the contraction of the heart [11]. Paracrine signaling is signaling over short distances to nearby neighboring cells that are not necessarily in direct contact. These signals do not diffuse far and many are deposited in the extracellular matrix. Neural cell communication is an example of paracrine signaling in which nearby cells are able recognize the release of neurotransmitters. Many growth factors, such as fibroblast growth factor (FGF) and Hedgehog family proteins, are able to signal to nearby neighbors in a paracrine fashion [17, 18]. Endocrine signaling represents the transmission of signals over long distances. Hormones (e.g., insulin, androgen, and estrogen) are often secreted from glands to cells in very distant organs. In addition, some growth factors such as vascular endothelial growth factor (VEGF) can also travel great distances through the organism before reaching target cells. Although this distance classification scheme is sometimes grey and each signaling type is not mutually exclusive, this grouping scheme highlights the diverse toolkit available to cells for transmitting messages to other cells.

The transmitted signal (also known as a ligand) then must be recognized by a target cell or the message will not be understood. For many of these ligands, this recognition takes the form of a protein receptor expressed within or on the surface of the target cell. By fine-tuning expression of these receptors to distinct cell types, organisms can further control exactly which cells respond to given signals. Binding of a target molecule to a receptor usually induces a conformational change in the receptor and kicks off a series of molecular events that propagate to the point of a phenotypic response. The next section contains an overview of just how these signals can be propagated and how this process eventually leads to a gross cellular response.

1.1.2 How does communication lead to phenotypic response?

The decoding of cell-cell communication messages is known as signal transduction. This abstract term is meant to encompass the molecular activities of proteins and other small molecules between initial detection of an signal and final response. Signaling pathways are the workhorses of signal transduction. The term signaling pathway generally refers to a collection of molecules that respond to stimuli by interacting and modifying one another resulting in a phenotypic response decision. Many different pathways exist and often these are able to detect multiple signals and can induce different responses depending on the context of any individual cell. Furthermore, pathways have been found to interact with one another and the combination of these interactions likely play a large role in determining the final cellular response [15, 19]. The evolutionary reward for this complexity, although daunting, allows organisms respond or adapt to almost an unlimited number of environmental perturbations.

The binding of a receptor to a ligand is the first step of pathway activation, inducing a

conformational change in the receptor. This structural change is ultimately the molecular mechanism that kicks off a cascade of signals. A variety of mechanisms underlie signal propagation including enzymatic addition or removal of post translational modifications (e.g., phosphorylation), release of secondary ions (e.g., Ca^{2+}), hydrolysis or exchange of protein-bound small molecules (e.g., GTP), complex formation or dissociation, and protein cleavage by protease activity. Individual pathways apply different combinations of these mechanisms when activated.

At the risk of oversimplification, there are several canonical pathways that warrant further discussion as they serve as good exemplars for pathway diversity. The largest class of pathways utilize G Protein-Coupled Receptors (GPCRs), which reside in the cell membrane and bind hormones and neurotransmitters [20]. After conformational changes induced by binding to ligand, GPCRs are able to associate with a trimeric complex consisting of the subunits G_α, G_β, and G_γ. In canonical GPCR signaling, the G_α subunit becomes activated when bound to GTP, dissociates from $G_{\beta\gamma}$, and binds to an effector protein. The diversity of effector types is large and includes adenylyl cyclases that produce cAMP, ion channels that alter ion concentrations, and phospholipase C that generates 1,2-diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3). Products of activated effectors, termed secondary messengers, can induce a variety of responses including activation of metabolic enzymes, kinases, and transcription factors. Another well characterized class of signaling pathways begins with the activation of receptor tyrosine kinases (RTKs). These receptors bind to specific growth factors and cytokines and are highly evolved for signaling related to development, differentiation, and intercellular communication in multicellular systems [21]. Binding of ligand to its RTK generally leads to dimerization and auto-phosphorylation. This form of the receptor is then able to bind proteins whose activities lead to activation

of RAS and a cascade of phosphorylations from RAF to MEK to MAPK, ultimately leading to changes in cytoskeletal components and gene transcription [22]. Another pathway that plays many roles in both development and cancer begins with binding of the Notch receptor to its ligand Delta. Both Notch and Delta are transmembrane proteins whose interaction across adjacent cells leads to two cleavage events that result in release of the Notch intracellular domain [13, 23]. This polypeptide then translocates into the nucleus where it alters transcription [13, 23]. Although only examples, these three pathways do illustrate some of the many mechanisms by which pathways sense and respond to signals.

1.1.3 Why is studying cell-cell communication important?

Although communication between cells is important for all living creatures, it is especially important for multicellular organisms, enabling the coordination of both complex development and physiological response. These processes, which normally play a role in promoting and maintaining homeostasis, can be co-opted and exploited in diseases like cancer. Rather than broadly describing how studies of cell-cell communication have increased our understanding of a variety of biology, this section will focus on several examples of the importance of multicellular communication to tumorigenesis.

Tumor-stroma interactions have been shown to play a variety of roles in tumor development, maintenance, and response to therapy (**Figure 1.2**). For example, the association between angiogenesis and tumor development has been known for over forty years [26]. It is now known that tumor cells and surrounding stromal cells are able to facilitate expansion of local vasculature and recruitment of new blood vessels through a variety of mechanisms including production of growth factors and secretion of angiogenic factors like VEGF [27]. Although largely unsuccessful in the clinic as of now, targeting of angiogenic processes
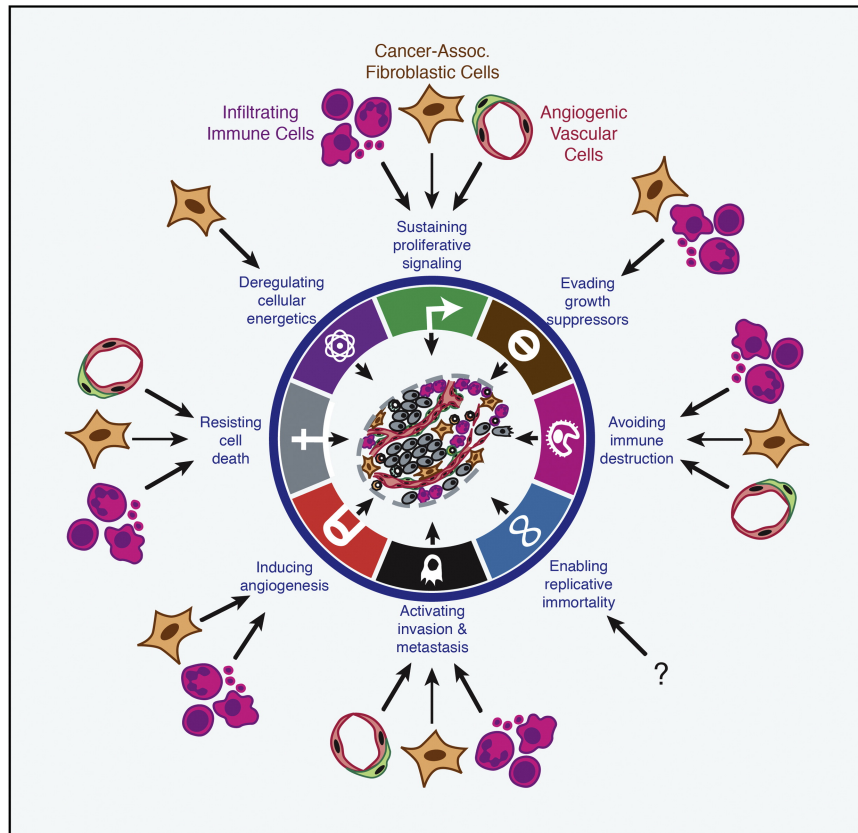
Figure 1.2: Stromal cell contributions to tumorigenesis. Of the eight hallmarks of cancer proposed by Hanahan and Weinberg in in [24] and [25], stromal cells have been shown to influence at least seven of these processes [2]. The extent of this contribution varies widely depending on the tumor type and includes both positive and negative contributions to tumorigenesis. Figure from Hanahan and Coussens [2].

has garnered much interest and is an area of intense study. Stromal cells are also able to mediate tumor-cell resistance to therapy. For example, macrophages have been shown to protect breast cancer cell response to taxol in a process that requires cathepsin proteases [28]. Similarly, inhibition of CD11b$^+$ myeloid cell recruitment to squamous cell carcinoma xenografts resulted in higher response to radiation [29]. These observations highlight the need to understand how stromal cells influence therapeutic response in order to make more informed and effective treatment decisions in the future. Stromal cells have also been shown to support aberrant proliferation. For example, fibroblasts have been shown to promote tumorigenesis through paracrine release factors such as hepatocyte growth fac-

tor (HGF), epidermal growth factor (EGF), insulin-like growth factor (IGF), and FGF [2]. Infiltrating immune cells also likely contribute to mitogenic bioavailability through variety of proteolytic enzymes that can degrade the extracellular matrix [2]. The evidence that tumor-stroma interactions play many roles in tumorigenesis is well established, while the mechanisms that underlie these roles are often elusive. A deeper understanding of how genotypically-normal cells are able to help tumors grow and survive will likely play an important role in how we view and treat this disease.

1.2 Methods to identify and profile proteins in biological samples

Researchers have employed a variety of techniques to identify, quantify, and profile the proteins associated with specific biological phenotypes. Genetic screens are often used to identify molecules important for interesting phenotypes. Another approach to narrow the search space of potential molecular players is the use of antibody or MS-based profiling to identify proteins altered between samples. Additional validation of significantly altered proteins must then be applied to establish causative roles.

1.2.1 Genetic approaches to identify and determine protein function

Genetic tools provide a powerful means to identify and study the genes and proteins involved in specific cellular processes. Optimized protocols now allow researchers to not only delete, alter, increase much of the molecular machinery within cells, but also determine whether proteins are able to interact with one another. The data that are produced from these techniques provides a means to identify genes important for specific processes and can also be used to validate predictions generated by profiling methods (detailed in the next section).

Several genetic methods exist to investigate genes that play a role in cell-cell communication. Random mutagenesis screens, using several techniques such as chemical mutagens or UV, followed by selection of interesting phenotypes, is an effective way to identify genes involved in specific processes. For example, the Notch receptor was found through a mutant fly that contained 'notches' in its wings and other members of the Notch pathway were elucidated through characterization of other mutants that contained similar wing patterns [13, 30]. Unfortunately, identification of the mutated region of DNA is labor intensive, multiple mutations can confound identification and interpretation, and it is difficult to study the functions of essential genes [31]. In another approach, knockdown using RNA interference (RNAi) is becoming the standard to identify and characterize loss-of-function of specific genes. RNAi provides control over the the identity of the disrupted gene through sequence complementarity to the target gene, and optimized reagents now make screening a relatively large number of genes feasible. Oricchio *et al.* utilized a short hairpin RNA screen to identify genes important in follicular lymphoma, identifying ephrin receptor A7 as a tumor suppressor [32]. The use of RNAi does have some limitations including difficult design of effective sequences, non-specific and off-target effects, and incomplete knockdown [33].

Although extremely powerful and useful, genetic techniques do have some limitations when studying cell-cell communication. By definition, genetic methods are external perturbations. The observed effects may therefore not be caused by direct involvement of the gene being studied, but may be the indirect result of disrupting of a different process altogether. Extensive validation of these screens is therefore required. When studying multicellular communication in particular, followup is needed to establish whether a protein is acting in a cell-autonomous or non-autonomous manner. In addition, most

of these techniques are designed to identify and characterize individual molecules, while processes that involve redundant control mechanisms are difficult to investigate due to the intractable nature of manipulating combinations of genes. System-wide understanding of cell-cell communication, a process that involves the coordination of many molecules, may benefit from more comprehensive profiling techniques such as the simultaneous detection and quantitation of many proteins.

1.2.2 Comprehensive identification and quantification of proteins

Another approach to understanding the molecular mechanisms by which cells communicate is to measure and compare the molecular "state" of cells in different experimental conditions. New technology and methods now allow researchers to profile thousands of genes and proteins in a single sample. It is hoped that simultaneous identification and quantitation of so many molecules will help us better describe complex multivariate processes such as cell-cell communication. This section contains a brief overview of various high-throughput methodologies for identifying and quantifying proteins in biological samples. A description of several proteomic methods will be given with an emphasis placed on quantitative MS (**Table 1.1**).

Historically, the most popular technique for studying entire proteomes has been separa-

	2D-PAGE	PAM	RPPA	MS
Dynamic Range	low	high	high	low
Specificity	high	medium	low	very high
Sample Throughput	low	high	very high	low
# Genes Measured	medium	medium	medium	very high
De novo Discovery	yes	no	no	yes

Table 1.1: Advantages and disadvantages of different proteomic technologies. 2D-PAGE = two-dimensional polyacrylamide gel electrophoresis, PAM = Protein antibody microarray MS = mass spectrometry, RPPA = reverse phase protein array, MS = mass spectrometry. Details of each technology are discussed in the text.

tion of proteins by isoelectric weight and charge, in an assay termed two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). Stained proteins appear as spots in the 2D gel, and those that differ in staining intensity between two samples (e.g., tumor vs. normal) are excised from in-gel digested protein, and sent for identification by MS. The major drawbacks of 2D-PAGE are low sensitivity and low throughput [34, 35]. Another tool for studying large numbers of proteins is the protein antibody microarray, which contains many different antibody-coated spots, each directed at a protein of interest. Intensity of secondary labeled antibodies is then used to quantitate the amount bound. Protein antibody microarrays have several advantages over 2D-PAGE including both higher sensitivity and higher throughput. The use of antibodies requires *a priori* decisions of the proteins to identify, making *de novo* discovery of unsuspected proteins impossible. In addition, the specificity of each identification varies depending on the antibody. Another method for proteomic research is the reverse phase protein array (RPPA). This technology prints hundreds to thousands of cell lysate samples on a single nitrocellulose-coated glass slide. Quantification of different proteins or post translational modifications for each sample is achieved by antibody staining [36, 37]. The use of RPPAs requires only minute amounts of sample and, because so many samples are spotted on a single slide, sample throughput is extremely high. RPPA is limited to only well-characterized antibodies and as with all antibody-based methods has limited discovery potential due to preselection of protein readout. All of these high-throughput antibody-based methods require cell lysis, abrogating the ability to differentiate the proteomes of multicellular systems.

Recent technological advances in MS allow for the identification of thousands of proteins in a single sample. Due to extremely high-accuracy, modern MS is capable of also measuring many different chemical and structural characteristics of proteins (e.g., phospho-

rylations). A standard MS experiment is analogous to a very accurate scale, measuring the mass-to-charge (m/z) ratio and intensity for many molecules in a sample. Tandem mass spectrometry (MS/MS) instruments take the technology one step further, fragmenting molecules and measuring the fragmentation pattern (m/z ratios and intensity). When sample proteins are digested into peptides (usually via trypsin), this fragmentation pattern can be used to determine the sequence of each observed peptide. In order to propel the peptides through the mass spectrometer, it is necessary to vaporize them by ionization. The two most common ionization methods are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). Use of ESI allows coupling to a liquid chromotographer in order to further separate complex samples, whereas MALDI is generally used on less complex samples [38]. Following ionization, peptides are propelled forward through one of the many varieties of mass spectrometers (for an excellent review of different types of MS machines see [39]). A general overview of an LTQ Orbitrap MS machine, the type of instrument used for these studies, is provided in Figure 1.3. Acquisition of mass-to-charge (m/z) ratios along with intensity values are recorded over time for both the precursor and fragment ions. These values are subsequently used as input to complex computational algorithms [40–42], which aim to identify and map each peptide to its respective protein. Modern MS machinery, advances in soft ionization methods, and development of computational frameworks have made MS-based proteomics a powerful tool for studying complex biological systems.

Several methods have been developed that allow relative and absolute quantification of peptides identified in the MS. The correlation between the amount of peptide and the intensity measured by the mass spectrometer is complex and not well understood. This relationship depends on many variables such as ionization efficiency and detector response
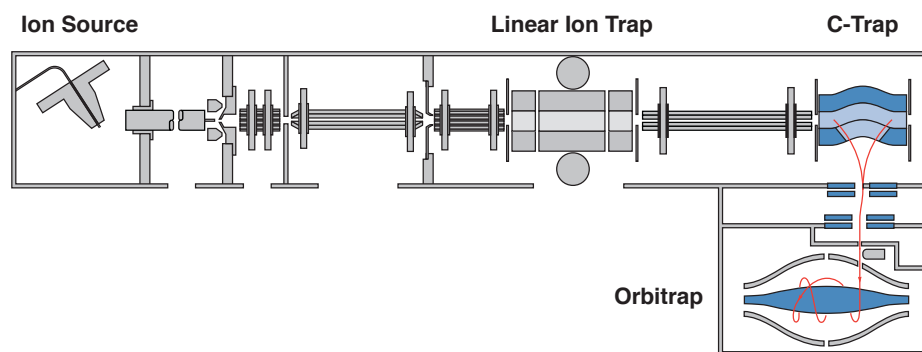
Figure 1.3: Overview of a hybrid linear ion trap / orbitrap mass spectrometer. Peptides are ionized by electrospray ionization as they elute from a liquid chromatographer into the machine. Electric fields and ion optics guide these ions into the linear ion trap where they are transferred into the C-trap. The ion population is subsequently injected into the orbitrap and begins to circle the electrode. Based on these oscillations, the mass-to-charge value can be calculated for each ion species. From full spectra analysis, the peptide with the highest intensity can be fragmented by collision with an inert gas (termed collision induced dissociation or CID) and sequence information can be determined by analyzing the fragment ions. Figure adapted from Yates *et al.* [43].

[44]. Several researchers have come up with clever ways to circumvent these issues by calculating relative peptide abundance. In general, these methods introduce stable isotopes or tags into sample peptides and by combining labeled and unlabeled samples early in a MS protocol, peptides that only differ by a small mass difference co-elute and ionize at the same time into the mass spectrometer, eliminating many cross-run biases. There are two general approaches to introducing labels into peptides. Chemical labeling approaches such as Isotope-Coded Affinity Tag (ICAT) and Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) introduce tags following protein or peptide isolation [45, 46]. Metabolic labeling using Stable Isotope Labeling of Amino Acids in Cell Culture (SILAC) introduces stable isotopes metabolically by supplying cultured cells or larger organisms with isotopically-labeled forms of essential amino acids (e.g., L-arginine and L-lysine, **Figure 1.4**) [44]. The sensitivity of modern mass spectrometers can differentiate minute mass shifts, which allows accurate quantification by comparing labeled to unlabeled peptide intensities. In addition to relative quantification, by 'spiking in' isotopically labeled peptides
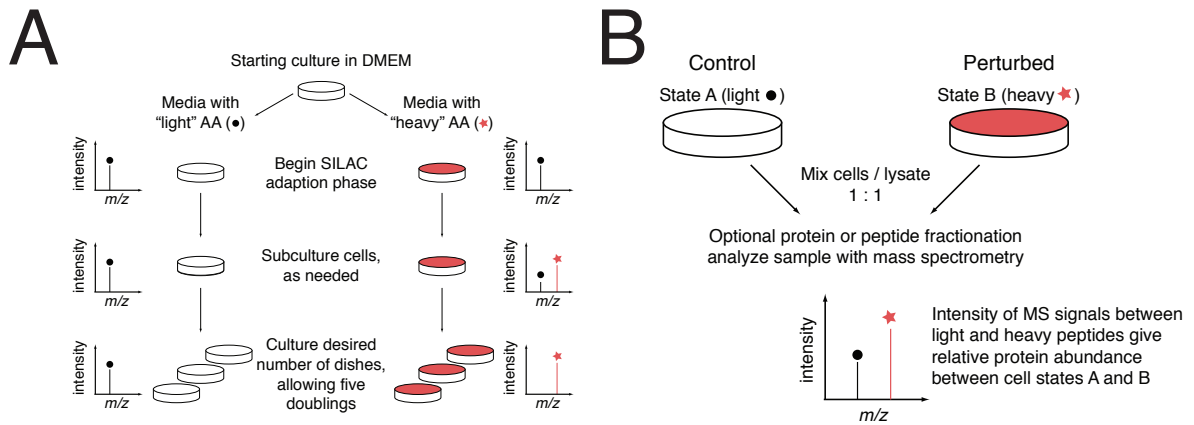
Figure 1.4: Overview of Stable Isotope Labeling of Amino Acids in Cell Culture (SILAC). (**A**) Cells are initially labeled in isolation using both media that contain light amino acids (black circle, typically L-lysine and L-arginine) and media that contains heavy-isotopically-labeled forms of the same amino acids (red star). After five doublings, these amino acids are completely incorporated into all proteins in these cells. (**B**) Each light and heavy labeled population of cells is treated differently, inducing changes in the proteome. Samples are mixed at the cell or lysate level, proteins fractionated, and digested peptides are identified by MS. The relative intensity differences between the heavy and light forms of identified proteins represent abundance differences between each cell population. Figure adapted from Ong and Mann [49].

or proteins of a known quantity, several groups have demonstrated the ability to determine absolute protein amounts [47, 48]. For those peptides that are identified, isotopic labeling and intensity comparison allow for both relative and absolute protein quantification. Unfortunately, these quantitative methods all require labeling in isolation. This limitation disallows cell-of-origin determination in multicellular culture, making it difficult to study cell-cell communication with these MS techniques.

1.3 Identifying and discriminating protein origin in multicellular systems

There are several different high-throughput protein identification methodologies. These techniques vary substantially in the number of proteins they can identify and in their abil-
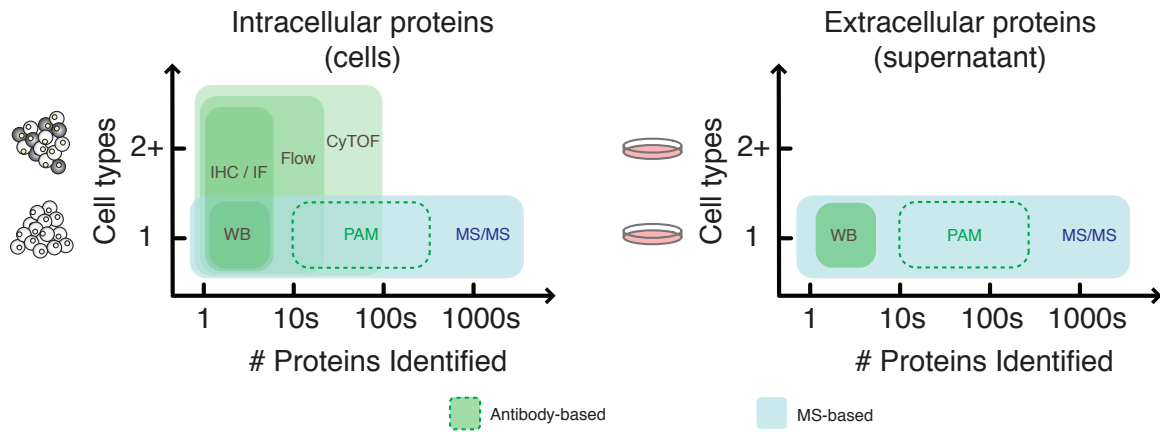
Figure 1.5: Methods for studying proteins in multicellular environments. The number of proteins each method is able to reasonably identify in a single sample is plotted against the whether the method can differentiate the cell-of-origin for identified proteins. The ability of each method to work with intra- or intercellular derived proteins is shown on the left and right, respectively. Antibody-based methods are shown in green and MS is shown in blue. WB = Western Blot, IHC/IF = Immunohistochemisty / Immunofluorescence, Flow = Flow Cytometry, CyTOF = mass cytometry, PAM = protein antibody microarray.

ity to discriminate the cell-of-origin for identified factors (**Figure 1.5**). The situation is particularly dismal for analyzing secreted proteins where, to the best of my knowledge, no standard technique is able to identify the cell-of-origin of identified proteins. This section contains a description of some of the high-throughput techniques that several groups have published for simultaneous identification and discrimination of protein origin.

1.3.1 Mass cytometry

One exciting new method that combines antibody-based MS identification with flow cytometry is mass cytometry [50, 51]. In this innovative approach, single cells are stained with a set of antibodies that have been tagged with distinct element isotopes (each with a different mass). These cells are then sent through a flow cytometer, vaporized, and analyzed by MS. Each identified isotope corresponds to a specific antibody and the intensity detected represents the abundance of its target. In contrast to flow cytometry, which detects fluorophores coupled to antibodies and has issues with overlapping signal when

using multiple fluorophores, the use of MS-detected isotopes allows for the simultaneous measurement of many target proteins.

The ability to simultaneously measure 34 protein targets in single cells has been demonstrated and up to 100 simultaneous protein measurements are possible [50–52]. Measuring this many parameters in single cells is not possible with other methods. Unfortunately, the reliance on antibodies requires well-validated reagents and limits the discovery potential due to preselection of measured proteins. This technology is still under development and will likely become a widely adopted tool for single-cell protein measurements.

1.3.2 Introduction of non-canonical amino acids into proteins via mutant tRNA-synthases

One very promising method to discriminate the proteome of cells in multicellular environments incorporates amino acid analogs into proteins [53–55]. These analogs are not normally recognized by the endogenous protein synthesis machinery and are therefore not integrated into proteins. However, transgenic expression of mutated tRNA-synthases enables the coupling (esterification) of these analogs to tRNAs and subsequent incorporation into proteins of specific cell types. The use of amino acid analogs may cause structural alterations to proteins and likely prevents complete labeling. However, the ability to perform affinity enrichment on some of these analogs (e.g., using azide chemistry) could be particularly useful for serum biomarker discovery in mouse models of disease. The use of this technology to discriminate the cell-of-origin of proteins in multicellular environments is relatively recent [56, 57] and further development and validation will be useful in determining the general applicability of the approach.

1.3.3 Differentiating the proteome by species-specific amino acid sequence differences

Several groups have utilized mixed-species cultures and xenografts to study stromal signaling in cancer and to identify disease-originated secreted factors [58, 59]. Protein assignment to specific cell types can be achieved for peptide sequences identified with MS that are unique to each species. Unfortunately, this approach has several limitations. First, signaling between cells of different species may not be physiologically relevant. Second, depending on evolutionary conservation between each organism, only a subset of identified peptides can be traced to a distinct cell type. For example, when comparing the *in silico* trypsin-digested human and mouse proteomes, most peptides exist in both species and only ~15% are unique and can be used to discriminate each organism (data not shown). Third, false positives or negatives are possible due to incomplete database annotations or mutations in sequenced peptides. Finally, the reliance on mixed-species models means that many established co-culture methods cannot be studied with this system.

1.3.4 Short term labeling with stable isotope-labeled canonical amino acids

A relatively straightforward approach to discriminating the proteome of distinct cell types in co-culture involves differentially labeling each cell type in isolation using standard SILAC procedures. The labeled cells are then combined together in co-culture and MS is used to quantitatively determine the cell-of-origin of identified proteins. This procedure has been successfully demonstrated in a model of ephrin signaling and as a way to identify proteins transferred between cell types [60, 61]. During culture, the label status will become diluted as cells synthesize new proteins and divide (approximately 50% dilution per cell division). Rapid loss of isotopic enrichment makes this method primarily

useful for studying quick signaling events. Although the differentially labeled cells could be plated in media lacking amino acids that dilute the signal, such media may perturb the phenotype of interest. Isolation-based labeling followed by co-culture is a creative, albeit a limited approach to high-throughput cell-of-origin determination in co-culture.

1.4   CTAP: Cell Type specific labeling with Amino acid Precursors

A technique that selectively introduces stable-isotope-labeled amino acids into the proteome of distinct cell types in multicellular environments would facilitate the use of high throughput MS to differentiate the cell-of-origin for identified proteins. Such a system would alleviate the requirement for preselection and reagent availability with antibodies, incomplete and possible functional alterations induced by non-canonical amino acid labeling, and short co-culture time requirements when prelabeling cells in isolation. Because stable isotopes are structurally and biochemically indistinguishable, introducing labeled forms into distinct cells in the same culture is challenging.

We have developed a system that is able to restrict distinct isotope-labeled amino acid use to specific cells in co-culture. In brief, our system replaces one or more essential amino acids, which are normally supplemented in the media, with isotopically labeled amino acid precursors. Ectopic expression of enzymes that catalyze precursor-to-amino acid reactions can be used to selectively and continuously label distinct cell-types of interest in co-culture. We have named this methodology CTAP for **C**ell **T**ype specific labeling with **A**mino acid **P**recursors. The development of the CTAP methodology is the primary focus of this body of work and the next chapters provide a detailed description of the hypothesis, its validation, and future development and use of this new technique.

Chapter 2

Cell type specific proteome labeling with precursors of essential amino acids

Nicholas P Gauthier[1,2], William E Walkowicz[2,3], Boumediene Soufi[4], Virginia A Pedicord[5], Konstantinos J Mavrakis[6], Boris Macek[4], David Y Gin[3,7], Chris Sander[1] & Martin L Miller[1]

[1] Computational Biology Center, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

[2] Louis V. Gerstner Jr. Graduate School of Biomedical Sciences, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

[3] Molecular Pharmacology and Chemistry Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

[4] Proteome Center Tuebingen, Auf der Morgenstelle 15, 72076 Tubingen, Germany.

[5] Howard Hughes Medical Institute, Department of Immunology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

[6] Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

[7] Deceased.

Correspondence should be addressed to N.P.G., C.S., or M.L.M. (ctap@cbio.mskcc.org)

2.1   Abstract

Tissue development, homeostasis, and pathogenesis involve complex signaling between many cell types through both secreted factors and direct contact. There are currently no high-throughput methodologies to determine the cell-of-origin of proteins in multicellular environments. To address this limitation, we have developed a technique that selectively and continuously labels the proteome of individual cell types in co-culture. By expressing

exogenous amino acid biosynthesis enzymes, we enable vertebrate cells to produce and metabolically incorporate canonical amino acids from labeled precursors. We have named this method **C**ell **T**ype specific labeling with **A**mino acid **P**recursors (CTAP). Using this new method, we demonstrate the ability to differentially label the proteome of distinct populations in co-culture and determine the relative expression of proteins from each cell type by quantitative mass spectrometry. Previously not accessible using other methods, we identify the cell-of-origin for secreted factors in a multicellular environment, further emphasizing the potential applicability of CTAP to a wide variety of scientific applications including biomarker discovery.

2.2 Introduction

Investigating protein signal transduction induced by secreted factors and cell-cell interactions is limited by current research methods. A notable example of these limitations is the inability of any current method to identify the cell-of-origin of growth factors, cytokines, and other secreted proteins. Antibodies are widely used for identification and differentiation of proteins specific to different cell types in tissue or co-culture (e.g., immunostaining or fluorescence-activated cell sorting, FACS), however antibody-based methods are relatively low throughput, vary in specificity, and are biased by preselection of protein readout and availability of reagents. High-throughput and unbiased methods, such as quantitative mass spectrometry (MS) based proteomics [44–46], might overcome some of these limitations. However, as MS is unable to differentiate from which cell-type proteins originate in complex cell mixtures, it has limited utility for intercellular communication studies. Research in cell-cell communication would greatly benefit from methods that overcome the

complimentary limitations with current antibody assays and MS-based proteomics.

Several recent efforts have been made to differentiate the proteome of individual cell populations in co-culture. In one such approach each distinct cell type is labeled in isolation (e.g., using heavy stable isotope-labeled L-Lysine or L-Arginine), and the fully labeled cells are subsequently mixed. Peptides identified in liquid chromatography tandem mass spectrometry (LC-MS/MS) can then be assigned a source cell-type from the isotopic label status. Two recent reports demonstrate the feasibility of such an approach for identifying early ephrin signaling responses [60] and determining proteins transferred between cell types [61]. Unfortunately, these labels become rapidly diluted as cells grow and divide in co-culture, making this experimental setup primarily useful for investigating early signaling events. In a different approach, protein sequence differences between species are used to determine cell-of-origin in cross-species co-cultures and xenografts [58, 59]. Although this approach has the ability to distinguish between proteins from different cell types, the major drawbacks are that only a subset of peptides can be differentiated, established same-species co-culture models are not applicable, and the findings from mixed-species models may not be physiologically relevant. Yet another technique utilizes tRNA-synthetases that specifically recognize and incorporate amino acid analogs into proteins [56, 57]. Using certain tRNA-synthetase / amino-acid-analog pairs, this method provides for both proteomic incorporation that is specific to transgenic cells as well as the ability to perform affinity enrichment on chemical moieties (e.g., azides). However, structural differences between the analogs and canonical amino acids might cause unpredictable functional alterations in mature proteins [62]. Given the caveats of each of these methods, there is a strong need for a technique that enables continuous cell-specific labeling with canonical amino acids.

In this study, we demonstrate a method for cell-selective proteomic labeling that over-

comes the problems of throughput and specificity of antibody-based cell staining, possible functional perturbations induced by amino acid analogs, physiological relevance of cross-species models, and the requirement of short co-culture time frames for cells labeled in isolation. This technique allows the proteome of distinct cell-types growing together in co-culture to be differentially labeled by canonical amino acids, avoiding the use of amino acid analogs that may perturb protein structure. Our method utilizes the inability of vertebrate cells to synthesize certain amino acids required for growth and homeostasis. These "essential" amino acids are produced in some plants, bacteria, and lower eukaryotes, and must be supplemented to the media of cultured vertebrate cells or obtained in the diet of animals [63]. Using transgenic expression of enzymes that synthesize essential amino acids, vertebrate cells are able to overcome auxotrophy by producing their own amino acids from supplemented precursors. These precursors can be isotopically-labeled, allowing cell-of-origin of proteins to be determined by label status identified by LC-MS/MS. For these studies we focus on L-Lysine, as the biosynthesis of this essential amino acid is well studied and it is commonly used in quantitative proteomic methods such as stable isotope labeling by amino acids in cell culture (SILAC) [44]. In this work, we test the validity and feasibility of the CTAP method and demonstrate its viability for continuous and differential metabolic labeling of cells in co-culture. Using this novel method, we are able to determine relative protein expression levels between two cell types in co-culture and identify cell-of-origin of secreted proteins.

2.3 Results

2.3.1 Engineering vertebrate cells to grow on L-Lysine precursors

Several enzymes have been found in bacteria, fungi, and plants that catalyze reactions leading to the production of L-Lysine from precursor compounds. We hypothesized that by engineering vertebrate cells to produce their own supply of L-Lysine from labeled precursors, we can achieve differential proteomic labeling of specific cell types in co-culture (**Figure 2.1**). We began by identifying a set of precursor-enzyme pairs in which the precursor was readily available and the enzyme had no described orthologs in verte-brate genomes (**Supplementary Figure 2.S1**). In a different context, one of the candidate precursor-enzyme pairs had successfully been used to rescue L-Lysine auxotrophy when creating a positive selection system for vector incorporation [64, 65]. To investigate the candidate precursors and eliminate those that autonomously rescue L-Lysine auxotrophy, we examined growth rates in SILAC media supplemented with L-Lysine, various precur-sors, or in L-Lysine-free conditions. With the exception of N2-acetyl-L-Lysine, the tested precursors alone had little or no effect on growth in wild-type cells (**Supplementary Figure 2.S2**).

We next investigated whether transgenic expression of enzymes involved in L-Lysine biosyn-thesis would allow cells to acquire the ability to grow on precursors. The genes encoding the enzymes diaminopimelate decarboxylase (DDC) from *Arabidopsis thaliana* and Lysine racemase (lyr) from *Proteus mirabilis* were stably expressed in several cell lines (**Table 2.1**). DDC-expressing mouse 3T3 and HEK293T cells, along with lyr-expressing human MDA-MB-231 cells, exhibited growth rates in media supplemented with the precursors meso-2,6-diaminopimelate (DAP) and D-Lysine, respectively, comparable to those in media con-
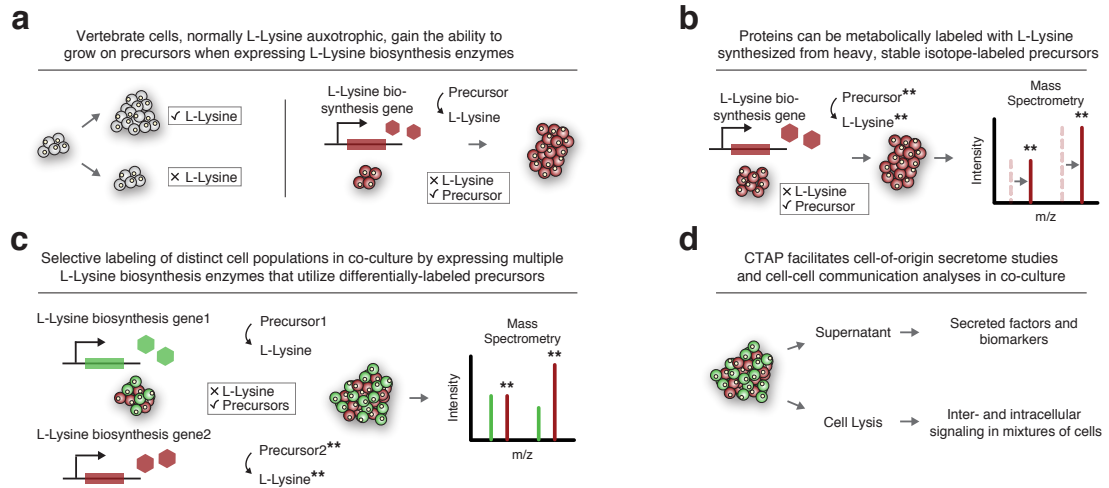
Figure 2.1: Overview of Cell Type specific labeling with Amino acid Precursors (CTAP). (**a**) The CTAP methodology takes advantage of vertebrate cells' inability to produce essential amino acids, resulting in the requirement that these molecules be supplemented in culture media or diet for cell growth. We focus on one of these amino acids, L-Lysine, and the enzymes used to produce it from precursor molecules. By expressing exogenous L-Lysine biosynthesis enzymes, transgenic cells produce their own supply of L-Lysine and (**b**) can be labeled selectively by supplementing the media with heavy isotope-labeled forms of the precursors. (**c**) Expressing distinct L-Lysine biosynthesis enzymes in different cell types enables continuous cell-selective proteome labeling with differentially-labeled precursors when grown in media lacking L-Lysine. (**d**) CTAP can be used to investigate direct contact or secreted factor mediated cell-cell communication, relevant for a range of biological phenomena.

taining L-Lysine (**Figure 2.2a**, **2.2b**, and **Supplementary Figure 2.S3**). Furthermore, the enzyme-precursor pairs were specific, as no growth was observed in the cross enzyme-precursor setup or in empty-vector controls (**Figure 2.2a** and **2.2b**). The time to reach normal growth rates varied between cell-types from immediate to a short passaging/selection period, indicating that certain cell-types may be more readily applicable to this method. These monoculture growth rescue results show that transgene enzyme expression together with supplementation of specific precursors is responsible for the growth rescue observed in L-Lysine free conditions.

Table 2.1: Transgenic cell lines and the precursor-enzyme pairs used in this study.

Cell type (origin)	Enzyme (origin)	Precursor of L-Lysine	Mass difference
MDA-MB-231 (human breast adenocarcinoma)	lyr (*P. mirabilis*)	D-Lysine D-Lysine*	0 Da 8 Da
HEK293T (human embryonic kidney)	DDC (*A. thaliana*)	DAP	0 Da
3T3 (mouse embryonic fibroblast)	DDC (*A. thaliana*)	DAP	0 Da

Da, Dalton; DAP, meso-2,6-diaminopimelate; DDC, Diaminopimelate decarboxylase; lyr, Lysine racemase; * heavy form (deuterated).

2.3.2 Cell-selective incorporation of L-Lysine produced from precursors

Although the phenotypic data served as a proxy for L-Lysine availability, they did not directly show molecular precursor-based incorporation. To investigate whether L-Lysine is directly produced by enzymatic-turnover of the supplemented precursors, we applied the SILAC principle of exchanging the isotopic label status of amino acids from one form to another (e.g., light L-Lysine to heavy L-Lysine) [44]. At the beginning of the experiments, DDC-expressing 3T3 cells were labeled with heavy [$^{13}C_6$,$^{15}N_2$]L-Lysine (H) and lyr-expressing MDA-MB-231 cells were labeled with light L-Lysine (L). These cells were then grown in monoculture for 13 days (3 passages) in L-Lysine-free media that contained unlabeled DAP (L), heavy-labeled [$^2H_8$]D-Lysine (H), or both precursors. Protein from cell lysate was trypsin-digested, submitted to high resolution LC-MS/MS, and H/L ratio for each peptide was determined by the MaxQuant software package [40].

In the presence of light-labeled DAP alone, peptides identified in DDC-expressing 3T3 cells switched from being predominantly labeled heavy (95%, median peptide) to light (97%) (**Figure 2.2c**). Similarly, the peptides identified in lyr-expressing MDA-MB-231 cells changed from 96% light to 95% heavy in the presence of heavy-labeled D-Lysine (**Figure 2.2d**). This amount of labeling can be considered complete as it is similar to the initial H/L

label status and levels typically reported in SILAC experiments [66, 67]. To test the amount of unspecific labeling (*i.e.*, cross contamination), cultures were also grown in the presence of both precursors. Supplementing the DDC precursor DAP had no effect on the label switch in lyr-expressing MDA-MB-231 cells, while the presence of D-Lysine (H) marginally increased the heavy label status (from 3% to 7%) in DDC-expressing 3T3 cells (**insets**, **Figure 2.2c** and **2.2d**). This difference was possibly due to heavy L-Lysine contamination in heavy D-Lysine and might be reduced with higher purity. Accordingly, the enantiomeric purity of heavy D-Lysine was ≥95% while the light form was ≥99.5% (C/D/N Isotopes and Sigma, respectively). Taken together, these data indicate that lyr and DDC-expressing cells are able to specifically incorporate and grow on L-Lysine synthesized directly from their respective precursors.

2.3.3 Limited perturbation to cells growing on precursors

We next investigated whether cells behave similarly when grown on precursors compared to L-Lysine. Cells were cultured for 3 days in media containing L-Lysine, precursor, or neither (starved, positive control for perturbed state) and mRNA expression levels were profiled using microarrays (**Figure 2.3, Supplementary Figure 2.S4**). Relative to the basal L-Lysine condition, no genes changed significantly when DDC-expressing 3T3 cells were grown on DAP, while 217 genes changed in the starved conditions ($FDR < 0.05$ and expression ratio greater than two, **Figure 2.3a**). The same pattern was also seen in lyr-expressing MDA-MB-231 cells when grown on L-Lysine or D-Lysine relative to starved cells (**Figure 2.3b**). Furthermore, several assays were performed to probe the effects of precursor-based growth, including measurement of protein changes by LC-MS/MS as well as growth and molecular response to drug perturbations (**Supplementary Figure 2.S5-**
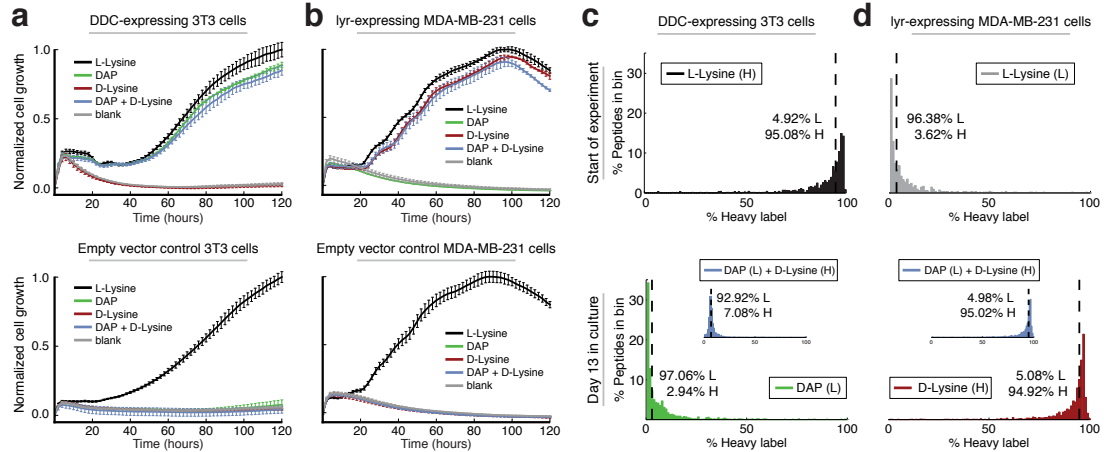
Figure 2.2: Vertebrate cell lines expressing L-Lysine biosynthesis enzymes grow and incorporate L-Lysine produced from their precursors. (**a**) Mouse fibroblast 3T3 cells that stably express DDC and (**b**) human breast carcinoma MDA-MB-231 cells that stably express lyr were plated in L-Lysine-free media supplemented with 10 mM DAP, 4 mM D-Lysine, both precursors, or 0.798 mM L-Lysine. Control (empty-vector) cells are shown in the lower panels. Cell growth, assessed with impedance (a correlate of the number of cells) using the xCELLigence system, was normalized to maximum growth. Error bars represent the standard deviation of three biological replicates. (**c**, **d**) Molecular incorporation assessed by LC-MS/MS. At the start of the experiment, cell lysates were collected from mono-cultured (**c**) DDC-expressing 3T3 cells labeled heavy (H) and (**d**) lyr-expressing MDA-MB-231 cells labeled light (L) (top panels). Cells were further passaged and samples harvested after 13 days in L-Lysine-free media containing the indicated precursors (bottom panels). Label status of Lysine-containing peptides was assessed by quantitative LC-MS/MS and percent incorporation of heavy label was determined using H/L ratios from MaxQuant analysis. Dashed black line indicates median peptide (percentages indicated).

2.S7). Although minor differences exist, overall these data demonstrate that growing cells on their precursors has little effect on gene expression, protein expression, or behavior.

2.3.4   Continuous and differential proteome labeling in co-culture

After demonstrating the principle of precursor-based L-Lysine production and incorporation in mono-culture, we next tested whether the same cells could be differentially labeled in co-culture with each population utilizing a distinct enzyme-precursor pair. To assess the specificity of labeling, we took advantage of species-specific sequence differences to compare label status between the enzyme-expressing mouse 3T3 and human MDA-MB-231 cell lines. Labeling each cell type in isolation, the 3T3 cells were initially cultured in heavy
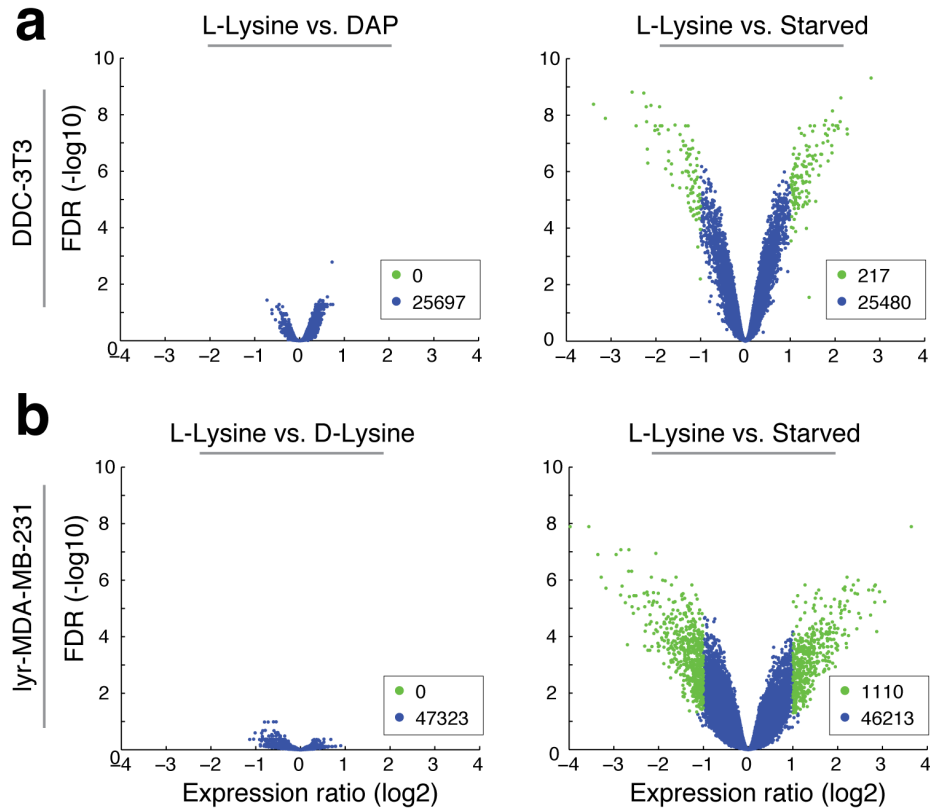
Figure 2.3: Limited gene expression changes observed when growing cells in precursor versus L-Lysine. (**a**) DDC-expressing 3T3 cells were plated in SILAC media supplemented with DAP, L-Lysine, or neither (starved). After 72 hours, mRNA was harvested and profiled for gene expression levels using the Illumina microarray platform. Expression differences of DAP versus L-Lysine (left panel) and starved vs L-Lysine (right panel) are plotted as a function of statistical significance (moderated t-statistics adjusted for multiple testing by the Benjamini and Hochberg method). Highlighted genes (green) are more than 2-fold differentially regulated at the level of $FDR < 0.05$. (**b**) As in (**a**) except MDA-MB-231 cells expressing lyr were plated on L-Lysine, D-Lysine, or in starved conditions. All experiments were performed in triplicate.

L-Lysine (H) and the MDA-MB-231 cells in light L-Lysine (L). A sample was harvested and combined 1:1 to verify the ability to differentiate label status based on species-specific peptide classification. As expected, labels of mouse-specific and human-specific peptides at the start of the experiment were confirmed to be primarily heavy and light, respectively (**Figure 2.4a, top panel**).

With the expectation that each cell type would exchange label status, the pre-labeled cells

Figure 2.4: Using two distinct enzyme-precursor pairs, co-cultured cells exhibit precursor-based differential proteome labeling. (**a**) DDC-expressing 3T3 cells (mouse) were labeled with heavy L-Lysine (H) and lyr-expressing MDA-MB-231 cells (human) with light L-Lysine (L) and mixed prior to sample analysis by LC-MS/MS (upper panel). Similarly labeled cells were co-cultured and analyzed after 3 passages on DAP (L) 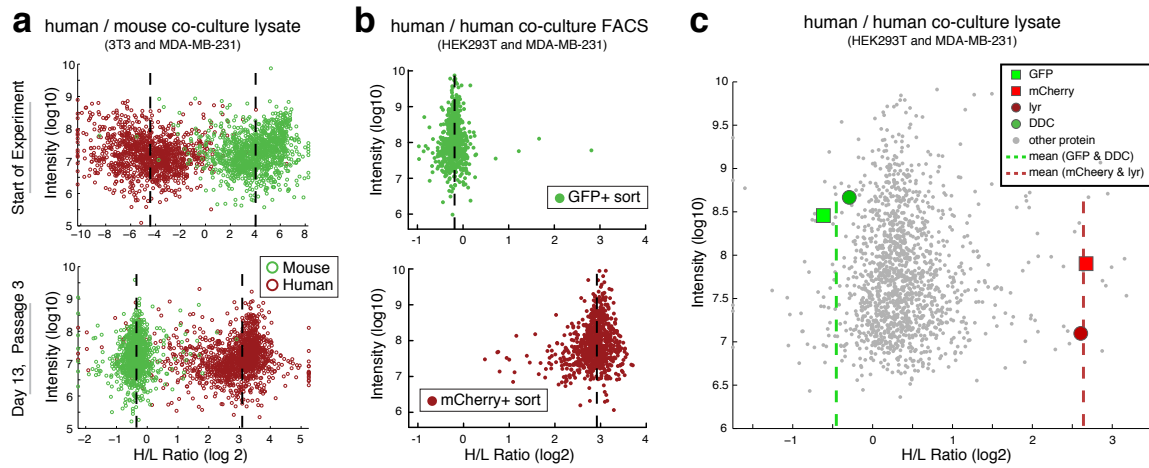and D-Lysine (H) (lower panel). Peptides unique to the mouse or human proteome are green and red, respectively. Median indicated by dashed black line. (**b**) GFP+ HEK293T expressing DDC were co-cultured with mCherry+ MDA-MB-231 cells expressing lyr in media containing DAP (L) and D-Lysine (H) for approximately 4-5 cellular doublings. Sorted GFP+ (upper panel) and mCherry+ (lower panel) cells were lysed, separately subjected to LC-MS/MS, and identified proteins are shown. (**c**) Proteins derived from unsorted co-culture of cells as in (**b**). Highlighted are proteins unique to each transgenic cell line (GFP and DDC in HEK293T, mCherry and lyr in MDA-MB-231 cells). Mean of transgenes for each HEK293T (DDC/GFP) and MDA-MB-231 (lyr/mCherry) are indicated with green and red lines, respectively.

were then combined in co-culture into media containing both DAP (L) and D-Lysine (H).

After three passages, with near equal growth rates of each cell population (**Supplementary Figure 2.S8**), the two cell types switched labels (**Figure 2.4a, bottom panel**). While the human MDA-MB-231 peptides became predominantly labeled from heavy precursor (90% or 3.1 log2 H/L), the mouse 3T3 peptides became approximately 57% (-0.4 log2 H/L) labeled from light precursor. For the 3T3 cells, the level of labeling was lower than expected from the results observed in mono-culture (see **Discussion** and **Supplementary Figure 2.S9**). Even with this lower labeling efficiency, the mouse and human peptides exhibit a similar number of overlapping H/L ratios as the SILAC labeled monocultures (top panel contains 3.2% peptides with H/L ratios not separable by cell type versus 4.7% in bottom panel,

Figure 2.4a). These distinct H/L ratios in species-specific sequences therefore demonstrate the ability to differentially label the proteome across cell types in co-culture.

Having validated continuous and differential labeling of human and mouse cells in co-culture, we next investigated whether the CTAP method could differentiate the proteome of a same-species co-culture system. DDC-expressing GFP+ HEK293T cells were plated together with lyr-expressing mCherry+ MDA-MB-231 cells. After five days of growth in DAP (L) and D-Lysine (H), a co-culture sample was sorted for mCherry+ and GFP+ cells by FACS (**Supplementary Figure 2.S10**) and each of the sorted populations was separately subjected to LC-MS/MS. Analysis of protein from the GFP+ and mCherry+ cells of this human-human co-culture showed similar labeling efficiency to that seen in the human and mouse co-culture, with each cell population exhibiting distinct H/L ratios (**Figure 2.4b**).

Another set of samples was collected directly from the non-sorted human-human co-cultures, subjected to LC-MS/MS, and 1362 proteins were identified in the union of three replicates. Focusing on the transgenic proteins exclusive to each cell population (GFP and DDC for the HEK293T as well as mCherry and lyr in MDA-MB-231 cells), we observed the expected H/L ratios corresponding to those determined by FACS (**Figure 2.4c**). This concordance confirms differential labeling in human-human co-culture lysates. When analyzing all identified proteins, the H/L ratios exhibited a near-normal distribution with the transgenes lying in the tails. Although these tails contain relatively few members, they likely represent cell type specific proteins (**Figure 2.4c** and **Supplementary Figure 2.S11**). This result is consistent with a recent report that found most proteins are ubiquitously expressed across different cell types but at different relative abundance levels [68]. In summary, these results demonstrate the ability to label the proteome in a cell-specific manner and show that label status (H/L ratio) is directly related to the relative protein abundance

level between the two cell types.

2.3.5   Determining cell-of-origin of secreted proteins in co-culture

To test the unique potential of the CTAP method to discriminate the cell-of-origin of secreted factors, supernatant was collected from the same human and mouse co-culture setup as the previous section. Prior to harvesting, the cells were grown for 24 hours in serum-free media to avoid overloading the sample with serum proteins. Secreted proteins were concentrated by ultra-centrifugation, precipitated by methanol-chloroform, and subjected to LC-MS/MS. Focusing on proteins identified only by species-specific peptides, nearly all could be completely distinguished by label alone (**Figure 2.5a**). These results demonstrate the ability of the method to determine cell-of-origin for secreted proteins in co-culture.

Applying a similar approach for analyzing secreted factors in a same-species co-culture, supernatant was collected and subjected to LC-MS/MS from the same co-cultured DDC-expressing HEK293T and lyr-expressing MDA-MB-231 cells as used previously. The H/L ratios of 245 identified proteins spanned a similar range as those detected intracellularly (**Figure 2.4c** and **Figure 2.5b**). Having shown that the H/L ratios are distinct for species-specific proteins in the human and mouse co-culture secretome, the tails of this human-human distribution likely represent cell type specific proteins. However, to gain more confidence that the H/L ratios reflect the relative protein abundance between each cell-type, we investigated whether extracellularly protein levels correlate with those found intracellularly. Quantitated protein ratios of mixed mono-culture lysates were therefore related to their secreted counterparts. Focusing on the subset of proteins that were common to both samples, good agreement was observed between the H/L ratios from the combined
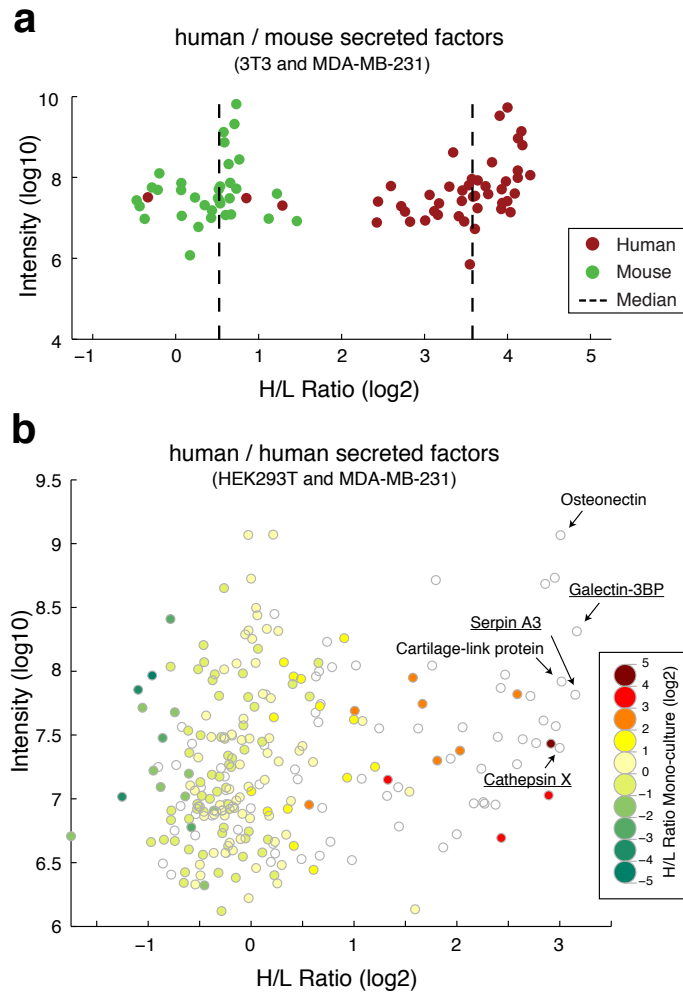
Figure 2.5: Application of CTAP for determining cell-of-origin for secreted factors. (**a**) DDC-expressing 3T3 cells (mouse) and lyr-expressing MDA-MB-231 cells (human) were co-cultured in DAP (L) and D-Lysine (H). Prior to sample collection, cells were grown for 24 hours in serum-free medium and the supernatant (medium) was collected. After concentrating proteins by ultra-centrifugation and methanol-chloroform extraction, the sample was analyzed by LC-MS/MS. Only proteins in which all identified peptides are unique to mouse (green) and human (red) are displayed. (**b**) Similar to (**a**) except the co-culture consisted of two human cell lines: HEK293T expressing DDC and MDA-MB-231 cells expressing lyr. Colors depict relative protein abundance as determined by SILAC quantitation of mixed, separately labeled mono-culture lysates. Uncolored points represent proteins that were not identified in the mono-culture sample. Annotated are the five proteins with highest H/L ratios: Galectin-3BP (LGALS3BP, Q08380); Serpin A3 (SERPINA3, P01011); Cartilage-link protein (CRTL1, P10915); Osteonectin (SPARC, P09486); Cathepsin X (CTSZ, Q9UBR2). Underlined protein names were identified as secreted from MDA-MB-231 cells in a recent study [69].

intracellular mono-cultures and secreted co-culture samples (R^2 = 0.66, pearson correlation = 0.81, **Figure 2.5b** and **Supplementary Figure 2.S12**). Considering the differences in culture conditions and localization of the harvested proteins, this correlation was surprisingly high. Further, this agreement indicates that the secreted proteins with the lowest and highest H/L ratios are cell type specific as they are most abundant in the HEK293T and MDA-MB-231 cells, respectively. As the secretome of MDA-MB-231 cells has been previously investigated and the high ratio proteins were readily separable from the majority of the identified proteins, we focused on the proteins the highest ratios. Indeed, of the top five proteins, three have recently been reported to be secreted by MDA-MB-231 cells (**Figure 2.5b**) [69]. Interestingly, a relatively large proportion of these putative MDA-MB-231-secreted proteins were not identified intracellularly (31%), highlighting the need for secretome profiling. Taken together with the species-verified secretome analysis, these results establish that the CTAP method can be applied to determine the cell-of-origin of secreted factors in co-culture.

2.4   Discussion

In this work, we demonstrate the validity and feasibility of the CTAP method for cell-selective proteome labeling in multicellular systems. Using precursors of the essential amino acid L-Lysine and enzymes that catalyze its synthesis, this work shows that the proteome of specific cell types in co-culture can be isotopically labeled by canonical amino acids produced in transgenic cells. Cell types of both mouse and human origin successfully overcome L-Lysine auxotrophy in the presence of specific enzyme-precursor pairs involved in the production of L-Lysine. Our studies demonstrate that there are limited

molecular and phenotypic consequences of culturing enzyme-expressing cells on their pre-cursors. Mass spectrometry analysis of enzyme-expressing cells in monoculture shows complete molecular labeling by L-Lysine derived from precursor. Differential-labeling of individual cell types in co-culture can be achieved using a dual-enzyme-precursor pair setup in the absence of L-Lysine, allowing relative expression levels for all identified pro-teins to be determined in each cell type. Supporting these results, we also found that CTAP is applicable for labeling a specific cell-type of interest in a mixed cell culture system us-ing only one enzyme-precursor pair, although titrating down the amount on L-Lysine in the media is necessary (for further details see **Supplementary Figure 2.S13**). In addition to DDC and lyr, we also tested and found specific but suboptimal growth rescue with the enzyme CBZcleaver and substrate Z-Lysine, supporting the adaptability of the CTAP method (**Supplementary Figure 2.S14**). Finally, analyzing the supernatant of cells in co-culture, cell-of-origin of secreted proteins can be readily established.

There are several features of the CTAP system that collectively distinguish it from other cell-selective protein labeling approaches. First, the products of enzymatic catalysis are canonical amino acids, allowing mature proteins to maintain their normal structure and avoiding functional alterations that may occur with methods based on amino acid analogs. Second, CTAP allows individual cell populations to be continuously labeled as they are grown and passaged over extended periods of time. Third, the genetic requirement of en-zyme activity to overcome essential amino acid auxotrophy makes labeling controllable by limiting transgenic expression. Fourth, utilizing multiple enzyme-precursor pairs per-mits differential labeling of multiple distinct cell types during co-culture. Fifth, CTAP can distinguish proteins from different cell types of the same organism rather than relying on artificial inter-species experimental setups. Finally, CTAP makes use of the same previ-

ously developed data-analysis workflows as the widely used SILAC method. To the best of our knowledge, CTAP is the only method in which the proteome of specific cell populations can be labeled continuously and differentially by canonical amino acids in a complex mixture of cells.

Although the results of this initial study demonstrate the feasibility and functionality of CTAP, there are several avenues for further method development. Optimally, CTAP should be quickly adaptable across many cell types without phenotypic or molecular disturbance. Focusing on the DDC/DAP and lyr/D-Lysine enzyme-precursor pairs, our results indicate that cells behave similarly when cultured on their specific precursor relative to L-Lysine. These similarities were measured after a period of growth that varied in length depending on the cell type tested. Two potential optimization steps could involve improving enzyme efficacy or increasing precursor uptake. For example, we observed that *A. thaliana* DDC is more effective at rescuing growth than *E. coli* DDC (data not shown), suggesting that screening additional organisms or mutagenesis approaches may lead to more effective enzymes. Additionally, Saqib *et al.* suggest that import of DAP is the primary limiting factor for production of L-Lysine [64], and therefore future studies aimed at optimizing precursor import could further increase the efficiency of L-Lysine production. While our attempts to rescue L-Lysine auxotrophy with DDC/DAP have been successful in all cells tested, another group has reported variable rescue efficiency of *E. coli* DDC across cell types [65] and it remains to be seen whether the enzyme-precursor pairs presented here can be applied to all vertebrate cells.

The principle of proteome labeling by amino acids produced from stable isotope-labeled precursors was demonstrated in mono-culture. Although this labeling was complete for both precursor-enzyme pairs (approximately 95%, **Figure 2.2c** and **2.2d**), when cells were

combined into co-culture we observed suboptimal labeling in one of the populations (approximately 50%, **Figure 2.4**). There are several possible explanations for this discrepancy. First, cells in co-culture might share amino acids or transfer proteins that are further metabolized. Second, phagocytosis might lead to amino-acid transfer. Third, transgenic enzymes may have extracellular activity. Fourth, minor label contamination may arise due to heavy L-Lysine impurities in the heavy D-Lysine stock. A combination of these possibilities or other unknown mechanisms may lead to the observed background labeling and will be addressed in future studies. Although desired, complete labeling was not necessary to determine relative protein expression levels between each cell type.

We anticipate that CTAP will be an important tool for gaining insight into intercellular signaling in a range of fundamental biological processes. For example, in various cancers the interaction between malignant cells and the surrounding stromal tissue has been shown to be important for disease progression, maintenance, and altered drug efficacy [3, 5, 8]. How stromal cells affect these processes is unclear, partly due to inadequate techniques for assaying their roles. The use of CTAP may address these limitations and offer an opportunity to understand the molecular mechanisms by which surrounding stroma alter tumor growth and response to treatment. Once precursor delivery, tolerance, and enzyme expression are optimized, another possible application of CTAP is identification of disease biomarkers *in vivo*. Current approaches for biomarker identification are limited by their inability to classify whether a potential marker originates from the diseased tissue itself or from normal tissue. Using the described technique we can circumvent these limitations, as proteins from specific cell types of interest can in principle be labeled continuously *in vivo*. Any labeled protein identified in the serum or proximal fluids will originate from the cell type of interest.

In these studies we show how utilization of exogenous amino acid biosynthesis compo-

nents allows for continuous cell-selective metabolic labeling of proteins. The principle be-

hind CTAP is generic and can theoretically be extended to more than two cell types and/or

applied to essential amino acids other than L-Lysine. CTAP may be an important step

forward in the field of proteomics, allowing the use unbiased and high-throughput LC-

MS/MS to differentiate peptides derived from distinct cells in complex cellular environ-

ments. The continued development and optimization of this method will allow researchers

to probe a variety of questions regarding cell-cell communication and cell-specific origin

of biomarkers not easily accessible with current methodologies.

2.5 Methods

2.5.1 Oligonucleotide Acquisition

The L-Lysine producing enzymes used in this study were DDC, lyr, and CBZcleaver. DDC

was directly amplified by PCR from *Arabidopsis thaliana* cDNA (TAIR id = AT3G14390,

primer sequences available in supplemental material). The lyr and CBZcleaver constructs

were synthesized by GeneArt with the amino acid sequences specified by either Kuan *et

al.* [70] (lyr) or Nanduri *et al.* [71] (CBZcleaver), and nucleotide sequences were optimized

for expression in mouse. All sequences are available in supplemental material. Sequences

were verified for all plasmids by the Sanger method of sequencing.

2.5.2 Plasmid Construction, Virus Production, and Cell Line Generation

Two MSCV based retroviral vector backbones, one expressing GFP (pMIG) and the other

mCherry (pMIC), were used to infect mouse cells. For insert into pMIG, the PCR product of

DDC was cloned into the EcoRI site of the vector. CBZcleaver was directly subcloned from the GeneArt supplied vector pMA-RQ into pMIC using EcoRI and XhoI restriction sites. Viral supernatants for pMIG and pMIC were produced by transfecting Phoenix cells with each plasmid and the supernatant was used to infect 3T3 cells 48 hours later as previously described [72, 73].

The lentiviral backbone pLM was used to infect human cells in this study. Overlapping PCR was performed to generate eGFP-DDC and mCherry-lyr constructs that were linked by a P2A peptide preceeded by a Gly-Ser-Gly linker [74]. The pLM-P2A-enzyme virus was packaged by calcium phosphate transfection of the HEK293T packaging cell line using 10 μg of transfer vector, 6.5 μg of CMVδR8.74, and 3.5 μg of the VSV.G plasmid. MDA-MB-231 and HEK293T cells were then infected with lentiviral supernatant produced from the pLM construct 48 hours post-transfection of the packaging line.

2.5.3 Cellular Growth Assays

Cell lines were grown in Dulbecco's modified Eagle's medium (DMEM) without L-Lysine and L-Arginine (SILAC-DMEM, Thermo Fisher Scientific) supplemented with 10% dialyzed FBS (Sigma, F0392), antibiotics, and L-glutamine. For mono-culture growth assays, 1 mM L-Arginine was added to the media and cells were seeded in 200 μL in 96-well plates with 4000 or 5000 cells per well in different concentrations of L-Lysine, meso-2,6-diaminopimelate (DAP, Sigma, 33240), D-Lysine HCL (Sigma, L5876), N-α-Cbz-L-Lysine (Z-Lysine, BaChem, C-2200), or N2-acetyl-L-Lysine (N2A, Sigma, A2010). Cell viability was measured using either the metabolic-activity based Resazurin (Sigma) reagent or the impedance-based xCELLigence system (Roche). For Resazurin experiments, 25 μL of the Resazurin reagent was added to each well and cellular growth was estimated after two

to three hours of incubation at $37\,^{\circ}$C as described by the manufacturer. For xCELLigence experiments, cells were plated in either 16 or 96-well E-plates, allowed to settle for 30 minutes at room temperature, and then placed in the RTCA DP or RTCA MP analyzer where impedance was measured every 15 minutes for 96-120 hours. At least three replicates were performed for each condition.

Measuring the percentage of mCherry+ and GFP+ cells in co-culture was performed by either flow cytometry (BD LSR II) or Tali image-based cytometry (Invitrogen). For flow cytometric assays, 25,000 cells from each cell line were seeded together in 6-well plates in 3-4 mL media supplemented with different concentrations of L-Lysine and/or L-Lysine precursors. After 72 hours, cells were trypsinized, washed, and resuspended in 200 μL PBS containing 2% dialysed FBS and 0.1%NaN$_3$. 20 μL was used for estimating total cell numbers using the ViaCount assay (Guava Technologies/Millipore) as described by the manufacturer. The remaining 180 μL was mixed with an equal volume of 2% paraformaldehyde. The percentage of GFP+ and mCherry+ cells in each sample was analyzed by flow cytometry. At least two replicates were performed for each condition. For Tali assays, cells were trypsinized, resuspended in media, 25 μL of co-culture cell suspension was used to determine the percentage of GFP+ and RFP+ cells in biological triplicate.

2.5.4 Stable Isotope Labeling and Cell Passaging

For exchange-of-label experiments (all monocultures, all human and mouse co-cultures, and **Supplementary Figure 2.S14**), cells were first metabolically labeled by growth for at least 10 cellular doublings in SILAC DMEM containing 798 μM light L-Lysine (L), medium [^{2}H$_4$]L-Lysine (M, +4 Daltons), or heavy [^{13}C$_6$,^{15}N$_2$]L-Lysine (H, +8 Daltons) (Cambridge Isotopes). Cells were then seeded in mono- or co-culture with 10 mM light DAP (L, Sigma),

2.5 mM or 4 mM heavy [^{2}H$_8$]D-Lysine (H, +8 Daltons, C/D/N Isotopes, 3,3,4,4,5,5,6,6-d8), 2.5 mM heavy labeled [^{13}C$_6$,^{15}N$_2$]Z-Lysine (H, +8 Daltons, **Supplementary Figure 2.S14**), or both DAP (L) and D-Lysine (H). For experiments that maintained label (all human-human co-cultures), cells were initially grown for at least 10 cellular doublings in their respective precursors: DDC-expressing in DAP (L) and lyr-expressing in D-Lysine (H). Populations were then combined in 10 mM DAP (L) and 3 mM D-Lysine (H) and grown together for 5 days in co-culture (approximately 4 cellular doublings). All cell lines were passaged 1:10-1:15 at 95% confluence.

2.5.5   mRNA Microarray Expression Profiling

Cells were seeded at equal densities into SILAC media containing 798 μM L-Lysine, 798 μM L-Lysine (M), 4 mM D-Lysine HCl, or 10 mM DAP. After 72 hours, cells were washed, trypsinized, pelleted, and frozen at $-80\,^\circ$C. RNA was extracted using the RNeasy mini kit (Qiagen), labeled, and hybridized to Illumina Mouseref-8 or Human HT-12 microarrays. After median centering the probe intensities for each array, moderated t-statistics and false discovery rate calculations for multiple hypothesis correction were performed using the eBayes method provided in LIMMA [75, 76].

2.5.6   Mass Spectrometry Sample Preparation

For harvesting of cell lysate, cells were trypsinized, resuspended in SILAC DMEM, washed three times in ice cold PBS, and cell pellets frozen at $-80\,^\circ$C. For FACS samples, co-cultures of GFP+ and mCherry+ cells were trypsinized, washed, and resuspended in PBS with 20% media (2% FBS) to a concentration of approximately $2\mathrm{x}10^7$ cells/mL. Cells were then sorted into single GFP+ and mCherry+ populations on a MoFlo cell sorter (Dako), washed twice

with ice cold PBS, and cell pellets were stored at $-80\,^{\circ}\text{C}$ for further analysis. A small aliquot of each sorted population was immediately reanalyzed to determine purity. For cultured media samples, cells were washed three times with PBS and supplied with serum-free SILAC DMEM 24 hours prior to supernatant sample collection. Media was collected, filtered with a 0.22 μm filter, and proteins were concentrated to around 1 mg/mL using a 3 KDa Amicon Ultra Centrifuge filter (Millipore) as described by the manufacturer.

2.5.7 Protein Extraction / Digestion

Cell pellets were resuspended with Denaturation buffer (6 M Urea/ 2 M thio Urea in 10 mM Tris), 1 μL of benzonase was added, followed by incubation for 10 minutes at room temperature. Cellular debris was removed by centrifugation at 4000 g for 30 min. For the supernatant samples, the secreted proteins were precipitated by chloroform/methanol extraction. Protein concentration was assessed by the Bradford assay (Bio-Rad). Crude protein extracts were subjected to either GelC or in-solution digest. For the GeLC-MS analysis, protein extracts were cleaned on a 10 cm, 4-12% gradient SDS-PAGE gel (Novex). The resulting lane was cut from the gel and subjected to in-gel digestion with trypsin as described previously [77]. Upon gel extraction, peptides were cleaned using Stage-tips and analyzed by nano-LC-MS. For in-solution digestion, proteins from the crude extract were reduced with 1 mM dithiothreitol (DTT), alkylated with 5 mM iodoacetamide, predigested with the endoproteinase Lys-C (Wako) for 3 h, and further digested with trypsin overnight [78]. The resulting peptide mixture was cleaned using Stage-tips [79] and subjected to nano-LC-MS without prior peptide separation.

2.5.8 LC-MS/MS Analysis

All samples were analyzed by online nanoflow liquid chromatography tandem mass spectrometry (LC-MS/MS) as previously described [80] with a few modifications. Briefly, nanoLC-MS/MS-experiments were performed on an EASY-nLC system (Proxeon Biosystems, Odense, Denmark) connected to an LTQ-Orbitrap XL or LTQ-Orbitrap Elite (Thermo Scientific, Bremen, Germany) through a nanoelectrospray ion source. Peptides were auto-sampled directly onto the 15 cm long 75 mm-inner diameter analytical column packed with reversed-phase C18 Reprosil AQUA-Pur 3 mm particles at a flow rate of 500 nl/min. The flow rate was reduced to 250 nl/min after loading, and the peptides were separated with a segmented linear gradient of acetonitrile from 5-50% in 0.5% acetic acid for either 100, 150, or 240 minutes. Eluted peptides from the column were directly electrosprayed into the mass spectrometer. For the LTQ-Orbitrap XL analyses, the instrument was operated in positive ion mode, with the following acquisition cycle: a full scan recorded in the orbitrap analyzer at resolution R 60,000 was followed by MS/MS (CID) of the top 10 most intense peptide ions in the LTQ analyzer. The total acquisition time was either 150 or 240 minutes. For LTQ-Orbitrap Elite data acquisition the instrument was operated in the positive ion mode, with the following acquisition cycle: a full scan recorded in the orbitrap analyzer at resolution R 120,000 was followed by MS/MS (CID Rapid Scan Rate) of the 20 most intense peptide ions in the LTQ analyzer. The total acquisition time was either 100 or 240 minutes depending on the method of sample preparation. Mono-enzyme co-culture samples were measured on the LTQ-Orbitrap XL with slight modifications: a full scan recorded in the orbitrap analyzer at resolution R 60,000 was followed by MS/MS (CID) of the top 5 most intense peptide ions, with a total acquisition time of 95 minutes.

2.5.9 Processing of MS Data

The MaxQuant software package (version 1.2.2.9) with the Andromeda search engine was used to identify and quantify proteins in cellular lysates and media [40, 81]. Mouse and human IPI protein databases (both version 3.84, `http://www.ebi.ac.uk/IPI/`) plus common contaminants and CTAP transgenes were used. With the exception of "second peptides", which was deselected, default parameters were selected. Detection and quantitation of L-Lysine containing peptides was specified as light L-Lysine, medium L-Lysine (Lys4), and heavy L-Lysine (Lys8). For L-Lysine derived from precursors DAP, Z-Lysine (H), and D-Lysine (H), variable labels were specified as light L-Lysine, heavy L-Lysine (Lys8), and a custom modification (8 deuterium atoms for L-Lysine), respectively.

Peptide and protein statistics (e.g., sequences, H/L ratios, intensities) were extracted from MaxQuant exported peptides.txt and proteingroups.txt, respectively. Entries that MaxQuant classified as contaminants were removed. Unless otherwise stated, no other filters or normalizations were applied to the H/L ratios. Peptides were determined to be species-specific if they only appeared in either one of the human or mouse IPI protein databases. For the species-specific sequence determination an exact peptide sequence match to the protein database was required, except Isoleucine (I) and Leucine (L) were used interchangeably. Percent heavy label was calculated from the H/L ratio ($HtoL$) as $= 100 * \frac{HtoL}{HtoL+1}$. In order to determine the overlap of H/L ratios between the human and mouse sequence-specific peptides, the median H/L ratio of each species was first determined. Next, the average of these two median values was used as a separator for each cell type and the miscategorizations were determined by the percentage of misclassified peptides on either side of this separator.

2.5.10 Data

The raw LC-MS/MS data are pending upload to the Peptide Atlas (`http://www.peptideatlas.org/`). Illumina microarray expreiments are pending upload to GEO (`http://www.ncbi.nlm.nih.gov/geo/`).

2.5.11 Accession Codes

The GeneArt optimized oligonucleotide sequences for lyr and CBZcleaver are pending upload to GenBank and are available in the Supplement.

2.5.12 Acknowledgments

2.5.13 Author Contributions

N.P.G. and M.L.M. designed, performed, and analyzed the experiments. W.E.W. generated reagents. B.S., K.J.M., and V.A.P. contributed with experiments. N.P.G. and M.L.M. wrote the manuscript. B.S., W.E.W., B.M., K.J.M., V.A.P., D.Y.G., and C.S. contributed to discussions and editing the manuscript. N.P.G. conceived the hypothesis. N.P.G., C.S., and M.L.M. developed the concept and managed the project.

2.6 Supplementary Information for:

Cell type specific proteome labeling with precursors of essential amino acids

Nicholas P Gauthier[1,2], William E Walkowicz[2,3], Boumediene Soufi[4], Virginia A Pedicord[5], Konstantinos J Mavrakis[6], Boris Macek[4], David Y Gin[3,7], Chris Sander[1] & Martin L Miller[1]

[1] Computational Biology Center, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

[2] Louis V. Gerstner Jr. Graduate School of Biomedical Sciences, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

[3] Molecular Pharmacology and Chemistry Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

[4] Proteome Center Tuebingen, Auf der Morgenstelle 15, 72076 Tubingen, Germany.

[5] Howard Hughes Medical Institute, Department of Immunology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

[6] Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

[7] Deceased.

Correspondence should be addressed to N.P.G., C.S., or M.L.M. (ctap@cbio.mskcc.org)

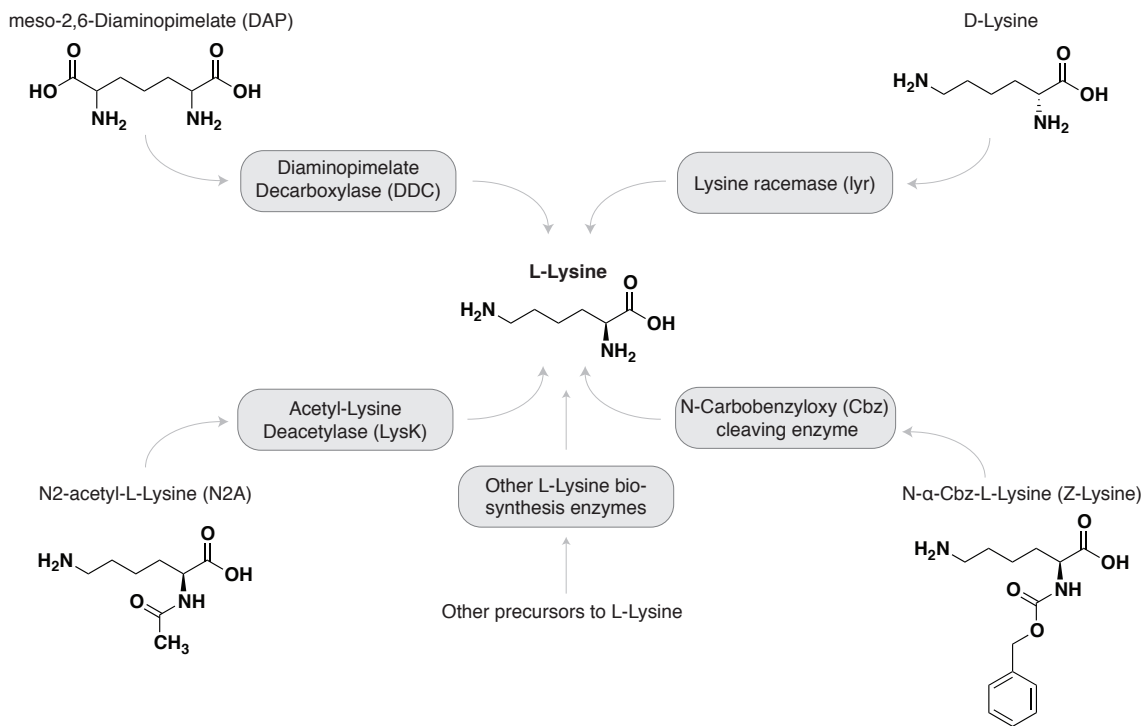## 2.6.1 Supplemental Figures

Figure 2.S1: Examples of L-Lysine producing enzymes and their substrates. Several enzymes have been found in bacteria, fungi, and plants that catalyze reactions leading to the production of L-Lysine from precursor compounds. Four examples of these enzymes and their respective precursors are indicated.
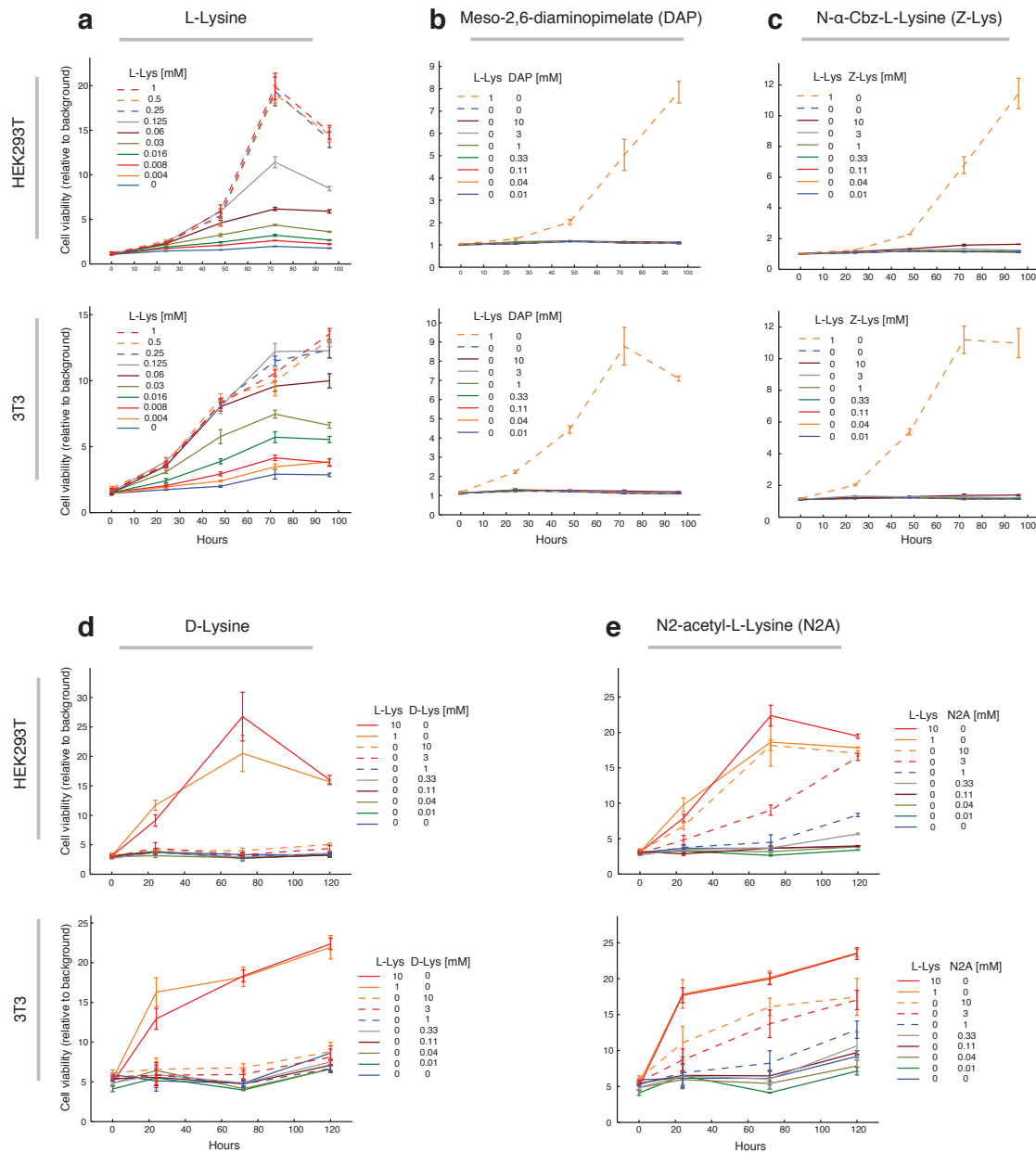
Figure 2.S2: Growth of human HEK293T and mouse 3T3 cell lines on L-Lysine and different precursors of L-Lysine. (**a**) Cells were seeded in 96-well format using at least 4 replicates per condition and cell proliferation was measured with the Resazurin (Alamar-Blue) assay at the time indicated. Note that both cell lines stop growing when no L-Lysine is present, confirming that mammalian cells are L-Lysine auxotrophic. Cells show no or limited growth response when the medium is supplemented with high (mM-range) concentrations of the L-Lysine precursors meso-2,6-diaminopimelate (DAP, **b**), N_α-cbz-L-Lys (Z-Lys, **c**), and D-Lysine (D-Lys, **d**). In contrast, both cell lines grow when the medium is supplemented with high concentrations of N2-acetyl-L-Lys (N2A, **e**).
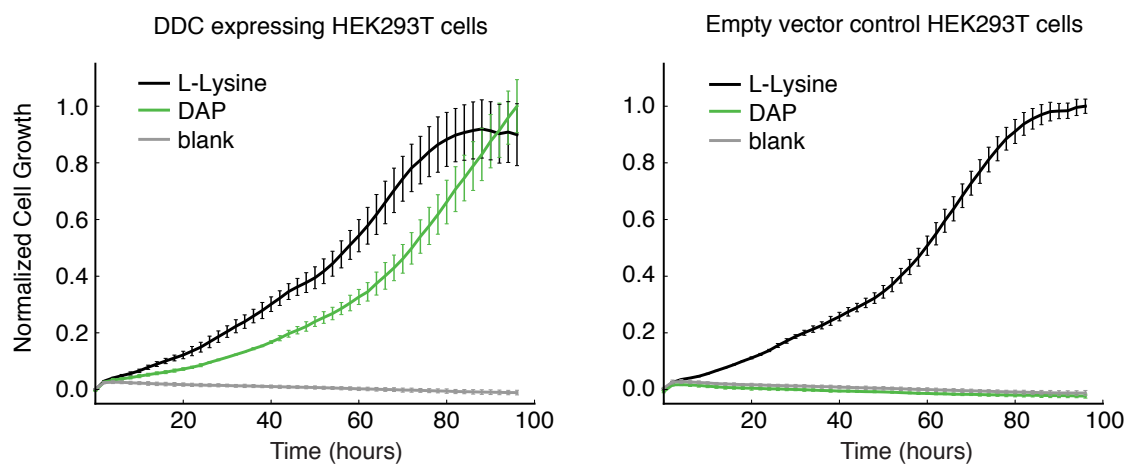
Figure 2.S3: HEK293T cells expressing the L-Lysine biosynthesis enzyme diaminopimelate decarboxylase (DDC) specifically grow on meso-2,6-diaminopimelate (DAP). HEK293T cells stably transfected with DDC (left panel) or empty control vector (right panel) were cultured in 0.798 mM L-Lysine, 10 mM DAP, or neither (blank). Cell growth was estimated by the impedance-based xCELLigence assay and data was normalized to the maximum value for each cell-type. Note that only HEK293T cells that express DDC grow on DAP. Error bars represent the standard deviation of three biological replicates.
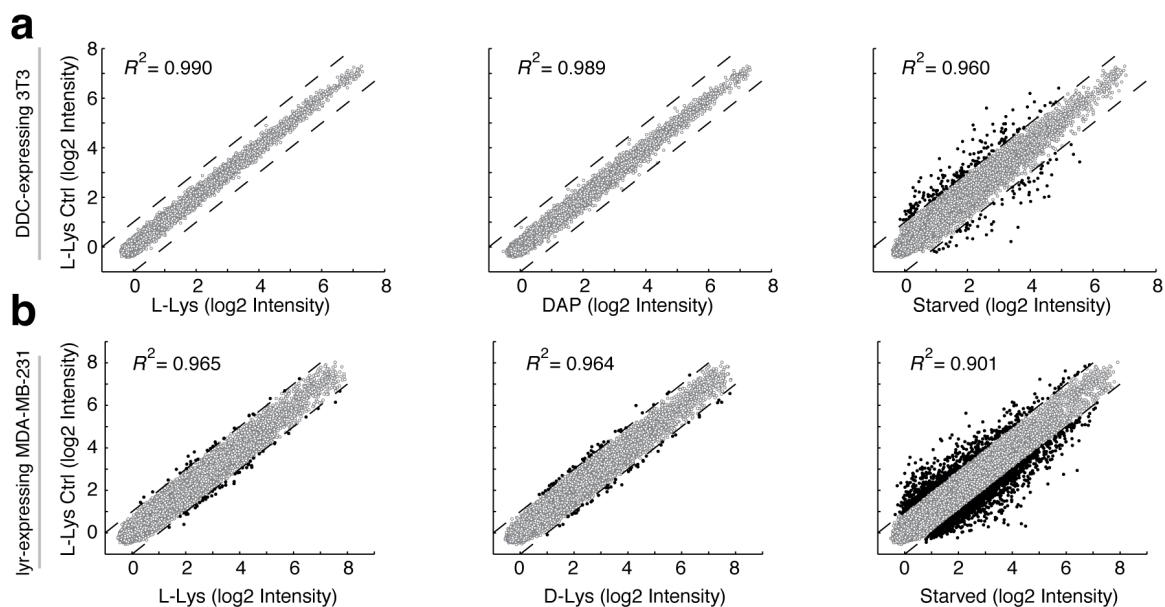
Figure 2.S4: Limited mRNA expression differences observed on growth of precursor vs L-Lysine (**a**) 3T3 cells expressing DDC were plated on L-Lysine, DAP, or in DAP/L-Lysine free (starved) conditions. After 72 hours, mRNA was harvested and run on the Illumina microarray platform. Representative arrays of three biological replicates are shown. Black dots represent genes that change more than two-fold between conditions. Dashed lines depict boundaries for 2-fold expression ratios between samples. (**b**) Similar to (**a**) except MDA-MB-231 cells expressing lyr were plated on L-Lysine, D-Lysine, or in starved conditions.

a

DDC-expressing 3T3 cells

pcc: 0.88
rsq: 0.78

y-axis: L-Lysine (H) (log10 intensity)
x-axis: DAP (L) (log10 intensity)

pcc: 0.83
rsq: 0.69

y-axis: L-Lysine (H) (log10 intensity)
x-axis: L-Lysine (M) (log10 intensity)

b

lyr-expressing MDA-MB-231 cells

pcc: 0.98
rsq: 0.96

y-axis: L-Lysine (L) (log10 intensity)
x-axis: D-Lysine (L) (log10 intensity)

pcc: 0.95
rsq: 0.91

y-axis: L-Lysine (L) (log10 intensity)
x-axis: L-Lysine (M) (log10 intensity)
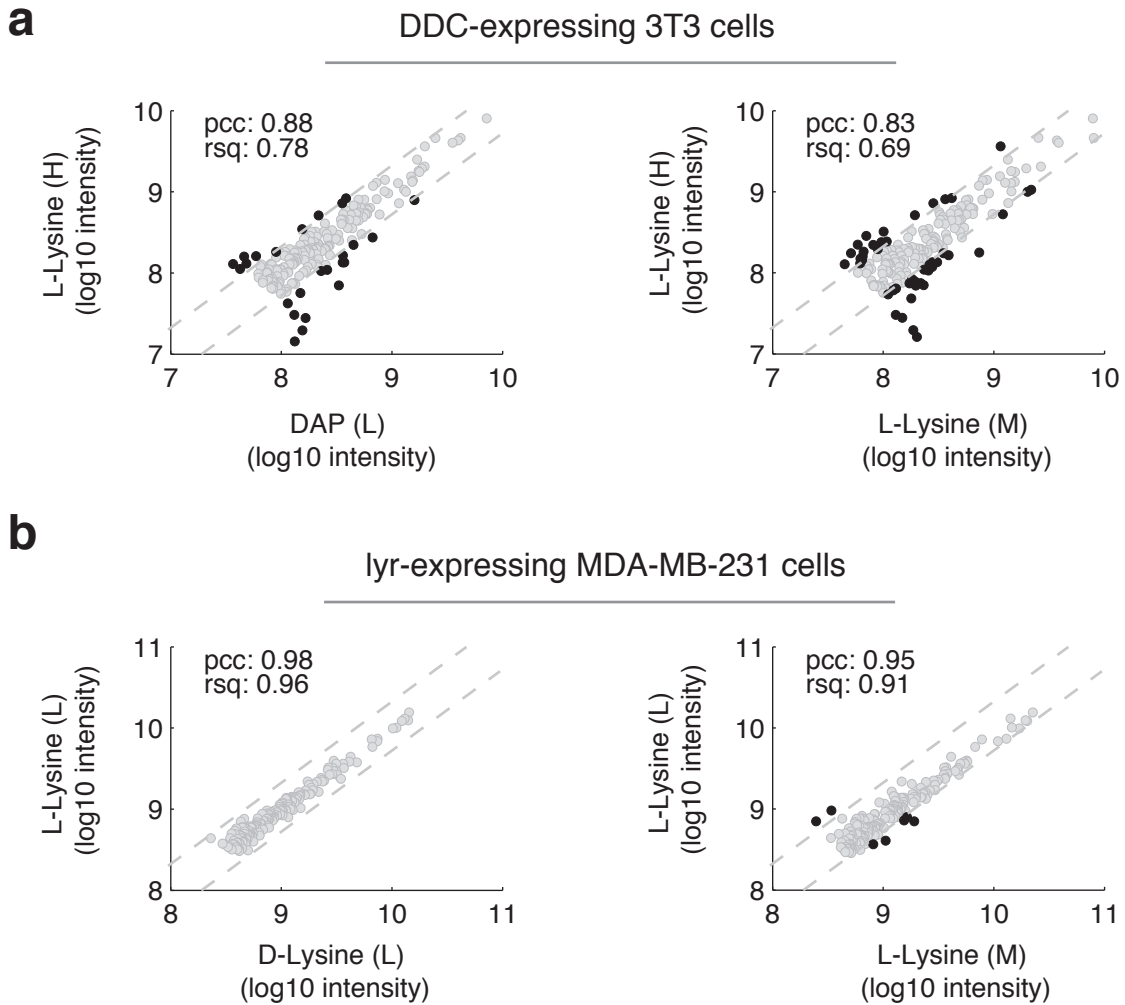
Figure 2.S5: Cells grown on precursors exhibit few or no protein abundance changes relative to those grown on L-Lysine. (**a**) DDC-expressing 3T3 cells were grown on either 10 mM DAP, 0.798 mM medium L-Lysine (M), or 0.798 mM heavy L-Lysine (H), and were analyzed by LC-MS/MS. Using label-free quantitation by the MaxQuant software, the intensities of the top 200 most intense proteins (minimum two peptides quantified) were compared between the conditions. Pearson correlation coefficients (pcc) and r-squared values (rsq) are provided. Intensity ratios greater than 2 are indicated (black dots). (**b**) Similar to (**a**) except lyr-expressing MDA-MB-231 were grown on 4 mM light D-Lysine (L), light 0.798 mM L-Lysine (L), or 0.798 mM medium L-Lysine (M). Note that the correlation between cells grown on precursor versus L-Lysine (left panels) is similar to that of cells grown on two different stable isotopes of L-Lysine (SILAC-labeled biological replicate, right panels).
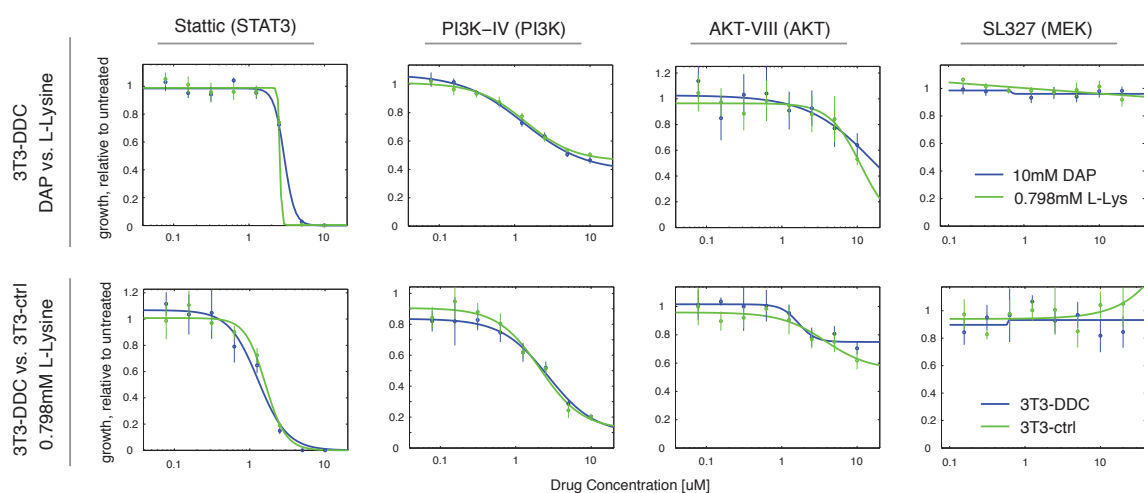
Figure 2.S6: Drug perturbation induces comparable effects to cell viability for both cells on DAP versus L-Lysine and enzyme-expressing versus empty-vector control cells. In the upper panel, DDC-expressing 3T3 cells were grown in the presence of either 10 mM DAP (green) or 0.798 mM L-Lysine (blue) in various concentrations of drugs as indicated (target of drug is indicated in parenthesis). Cell viability was measured after 48 hours of drug exposure with the Resazurin assay and normalized to untreated control cells. The lower panel compares DDC-expressing 3T3 cells (green) to empty vector control cells (blue) in the presence of 0.798 mM L-Lysine.
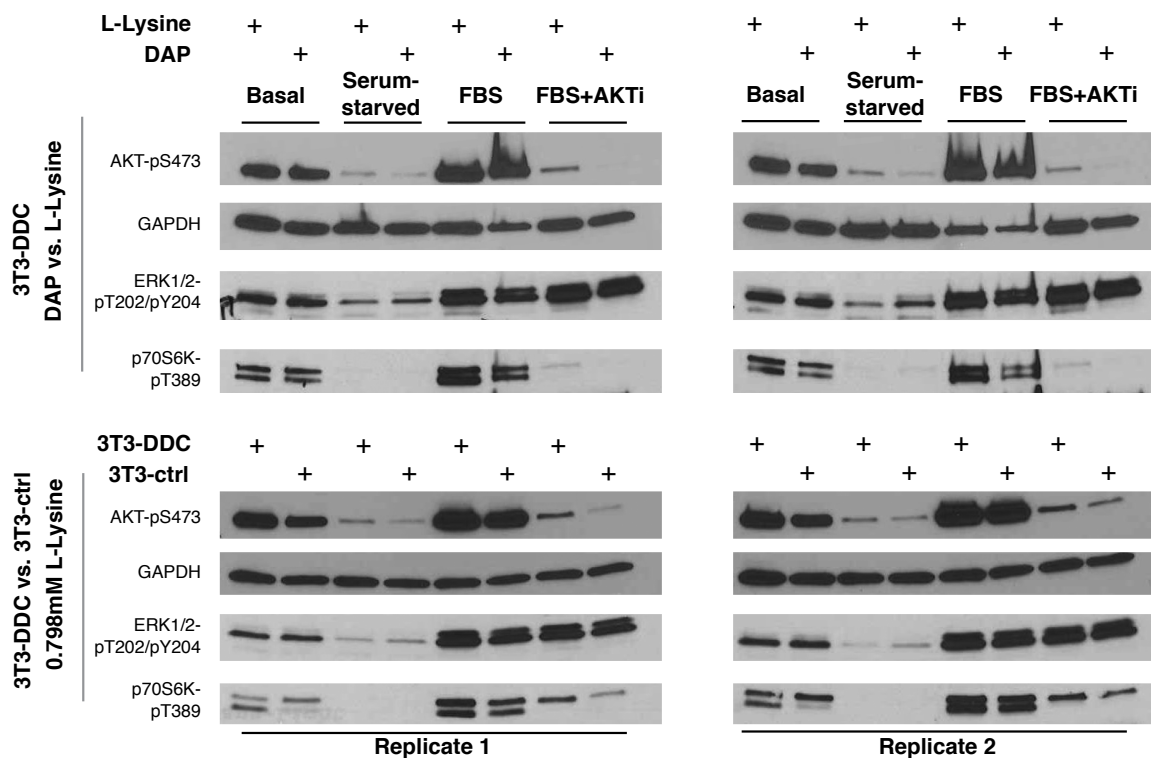
Figure 2.S7: Molecular response to starvation, FBS stimulation, and drug perturbation are largely similar for both cells on DAP versus L-Lysine as well as enzyme-expressing versus empty-vector control cells. In the upper panel, DDC-expressing 3T3 cells were grown in the presence of either 10 mM DAP or 0.798 mM L-Lysine in media with 10% FBS (basal), without FBS (serum-starved), starved for 24h and stimulated with 10% FBS for 1h (FBS), or stimulated with FBS and perturbed with 5 μM AKT Inhibitor VIII (EMD Chemicals) for 1h (FBS+AKTi). In the lower panel, DDC-expressing 3T3 cells and empty vector control cells were grown in the presence of 0.798 mM L-Lysine and exposed to similar conditions. For both experiments, cells were lysed and the response of several phosphoproteins was assessed by western blotting. Loading is indicated with GAPDH. Two biological replicates are shown. For western blot protocol, see **Supplementary Methods**.
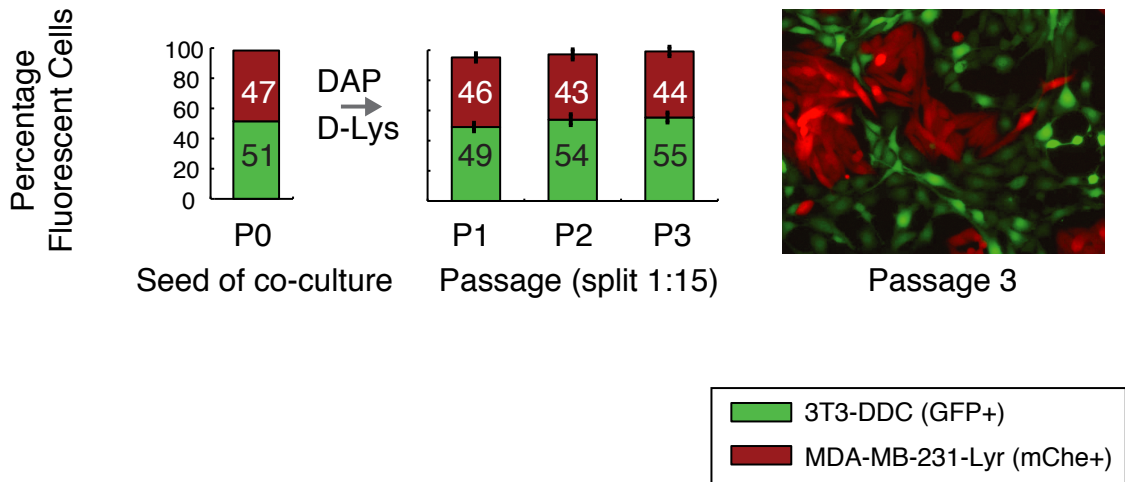
Figure 2.S8: Using two distinct enzyme-precursor pairs, co-cultured cells grow on precursors in L-Lysine free conditions and maintain similar proportion of each cell type over several passages. DDC-expressing 3T3 GFP+ cells were plated with lyr-expressing MDA-MB-231 mCherry+ cells and the media was supplemented with 10 mM DAP and 4mM D-Lysine in L-Lysine-free conditions. The co-cultures were split 3 times (1:15) and the ratio of GFP+ and mCherry+ was determined at each passage using image-based flow cytometer (Tali, Invitrogen). A representative fluorescent microscopic image at passage 3 is depicted.
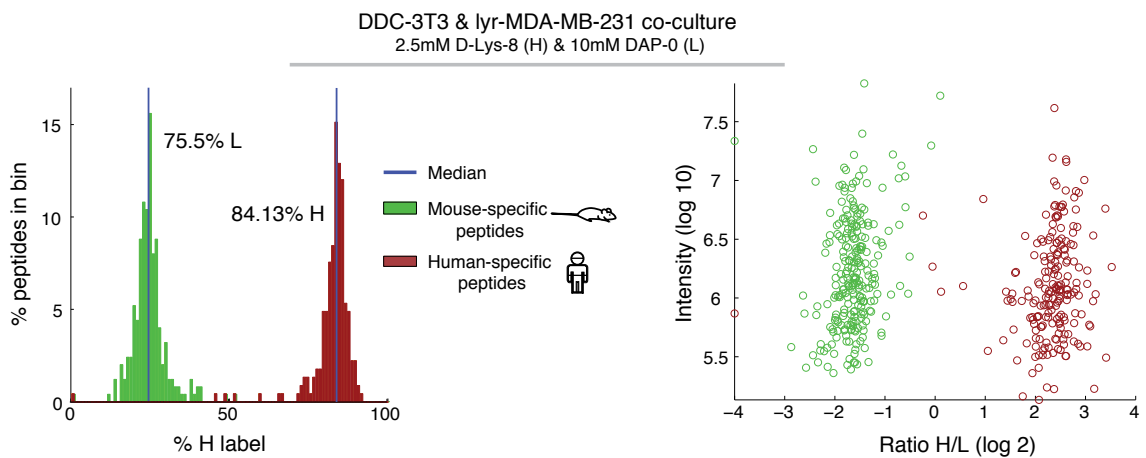
Figure 2.S9: Lowering the concentration of D-Lysine decreases the amount of nonspecific labeling in DDC-expressing 3T3 cells. DDC-expressing 3T3 cells were plated with lyr-expressing MDA-MB-231 cells and the media was supplemented with 10 mM DAP (L) and 2.5 mM D-Lysine (H). A lysate sample was collected after 3 passages (13 days in culture) and analyzed by LC-MS/MS for label status of L-Lysine containing peptides (left). Peptide intensities of the same sample is plotted against the H/L ratio (right). Only peptides that are unique to the mouse (green) or human (red) proteome by sequence are analyzed. Note that lowering the concentration of D-Lysine to 2.5 mM from previous used levels (4mM, **Figure 2.5**) decreases the amount of nonspecific labeling in DDC-expressing 3T3 cells.
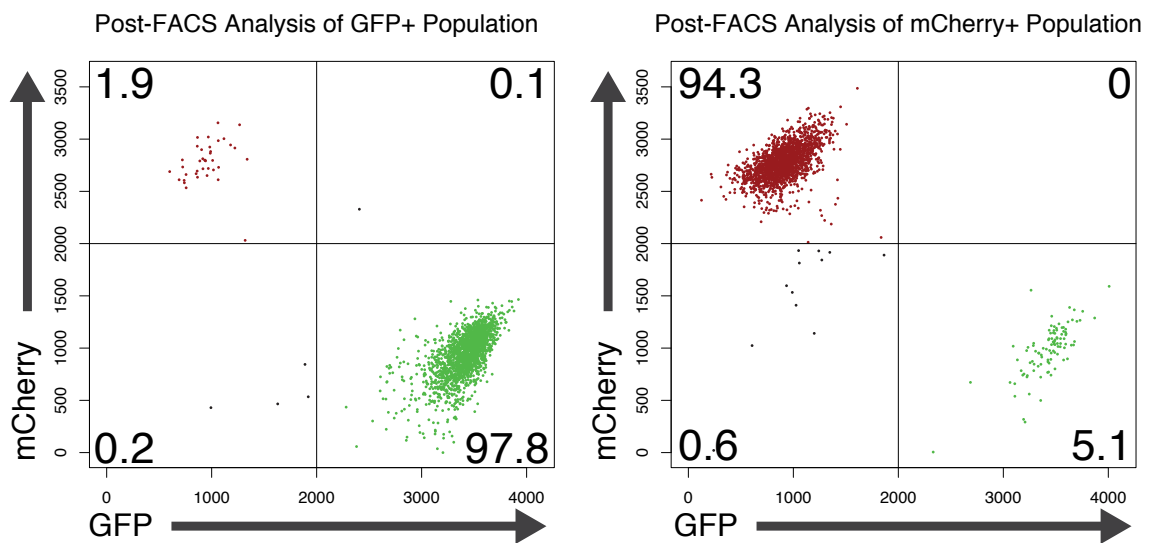
Figure 2.S10: Post sort FACS analysis of co-cultured human HEK293T and MDA-MB-231 cells. GFP+ HEK293T expressing DDC were co-cultured with mCherry+ MDA-MB-231 cells expressing lyr and sorted for GFP+ and mCherry+ cells by FACS. Depicted is a post-sort analysis showing the purity of each of the sorted populations as assessed by flow cytometry. Percentages are indicated. Although a post-sort analysis of the sorted populations showed a high enrichment for the expected fluorophores, there was 2-5% cross-contamination.
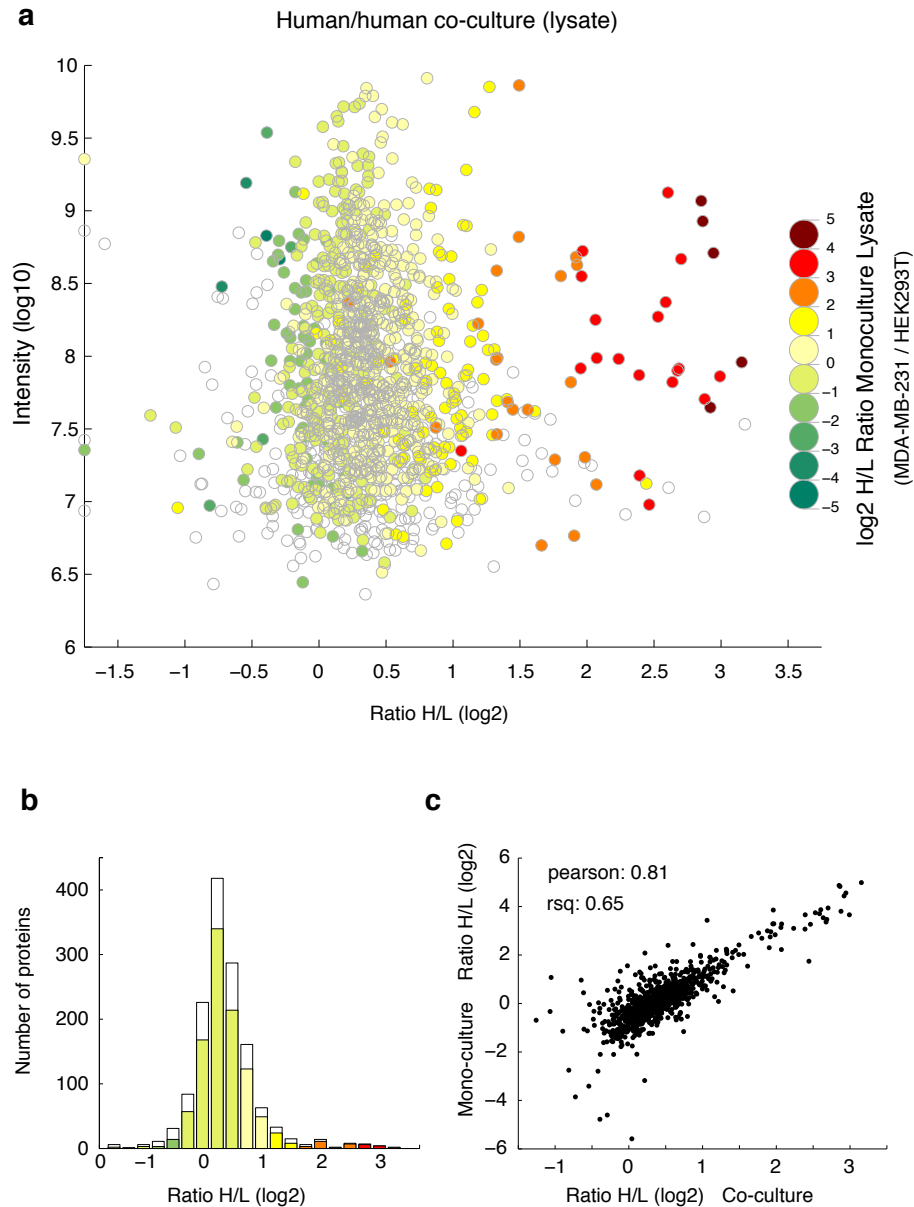
Figure 2.S11: Label status of differentially labeled co-culture cells shows good agreement with SILAC-labeled mono-cultures. (**a**) HEK293T expressing DDC cells were co-cultured with MDA-MB-231 cells expressing lyr in 10 mM DAP (L) and 4mM D-Lysine (H). Cell lysate was collected, proteins were digested, and the sample was subjected to LC-MS/MS. Colors depict relative protein abundance as determined by quantitation (median-centered H/L ratios) of mixed mono-cultures that were separately labeled using standard SILAC labeling. Uncolored points represent proteins that were not identified in the mono-culture sample. (**b**) Co-culture H/L ratios were binned and the average mono-culture H/L ratio in each bin was determined and depicted using a similar color scheme as in (**a**). (**c**) Correlation between mono-culture and co-culture H/L ratios.
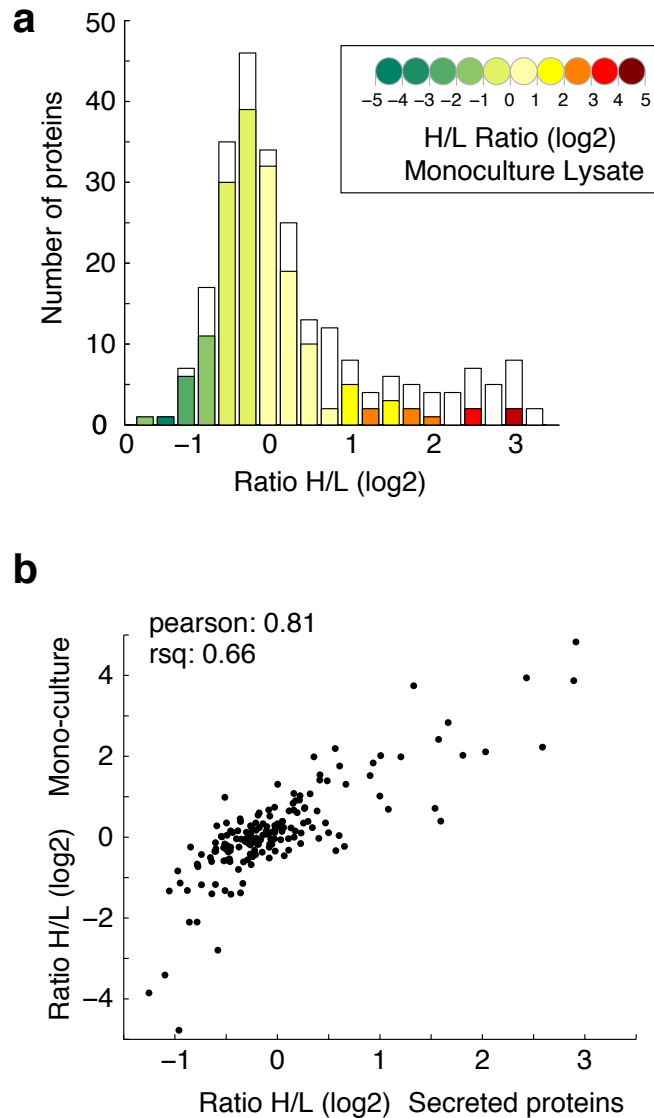
Figure 2.S12: Label status of secreted proteins from differentially labeled co-culture cells shows good agreement with SILAC-labeled mono-culture lysate. (**a**) HEK293T expressing DDC cells were co-cultured with MDA-MB-231 cells expressing lyr in 10 mM DAP (L) and 4mM D-Lysine (H). Prior to harvest of supernatant (24 h), cells were grown in serum-free medium. Proteins were concentrated by ultra-centrifugation, precipitated by methanol-chloroform, digested, and subjected to LC-MS/MS. To investigate if the H/L ratios reflect the relative protein abundance between each cell-type, we analyzed whether monoculture intracellular protein levels correlate with those found extracellularly. To test this, the quantified H/L ratios of the secreted proteins were compared to median-centered H/L ratios from mixed mono-cultures that were separately labeled using standard SILAC procedures. The histogram depicts binned H/L ratios from the secreted proteins, with the color set to the average mono-culture H/L ratio for that bin. Note that a relatively high proportion of the proteins identified with high H/L ratios could not be identified intra-cellularly (uncolored portion of bars). (**b**) Correlation between mono-culture lysate and co-culture secreted protein H/L ratios.

2.6.2 Differential proteome labeling of co-cultures using one enzyme-precursor pair

In certain co-culture models, it may be desirable or necessary to use CTAP for labeling a single cell type of interest utilizing only one enzyme-precursor pair. In such a situation, supplementing L-Lysine is necessary to allow for growth of the wild-type cells, but creates competition between L-Lysine and precursor-based L-Lysine for the cell-type of interest. To investigate the balance between precursor- and L-Lysine-based growth, we plated GFP+ DDC-expressing 3T3 cells cells together with control 3T3 mCherry+ cells in the presence and absence of DAP. Various concentrations of L-Lysine were added to the media and the number of GFP+ and mCherry+ cells were measured by flow cytometry after 3 days in co-culture. In our assays, the presence of DAP allowed the DDC-expressing cells to outgrow control cells in low levels of L-Lysine. In the absence of DAP, both DDC-expressing and control cells exhibited similar growth rates at all L-Lysine concentrations tested (**Figure 2.S13a**). At approximately 40 μM L-Lysine allowed for growth of both cell types as well as precursor-based growth of the cell-type of interest.

We next used mass spectrometry analysis to test if the cell-type of interest could be selectively labeled in this co-culture setup. At the start of the experiment, heavy L-Lysine (H) labeled DDC-expressing mouse 3T3 cells were mixed with medium L-Lysine (M) labeled human MDA-MB-231 cells and plated in media supplemented with 40 μM heavy L-Lysine (H) and light 10 mM DAP (L). Using two sets of peptides unique to either human or mouse and focusing on the 200 most intense peptides from each set, the expected labels of human-specific and mouse-specific peptides at the start of the experiment were confirmed to be primarily medium and heavy, respectively (**Figure 2.S13b**). After passaging, the human cells fully exchanged their proteome and became heavy labeled, while the

mouse proteome was labeled by both light and heavy. The percentage of light label was significantly increased only in the mouse-specific peptides (from 5% to 23%, p < 5.6e-34, two-tailed students t-test), while the human-specific light label remained unchanged (from 7% to 8%, p < 0.31). As expected, full precursor-based labeling was not obtained, likely due to supplementation of L-Lysine to the co-culture media. The shift from isotopically-labeled heavy L-Lysine (H) to precursor-based light L-Lysine (L), which is observed only in enzyme-expressing mouse cells, demonstrates cell-selective labeling in co-culture using a single enzyme-precursor pair. These findings are potentially relevant for identifying biomarkers originating from a cell-type of interest in the context of its natural multicellular environment.
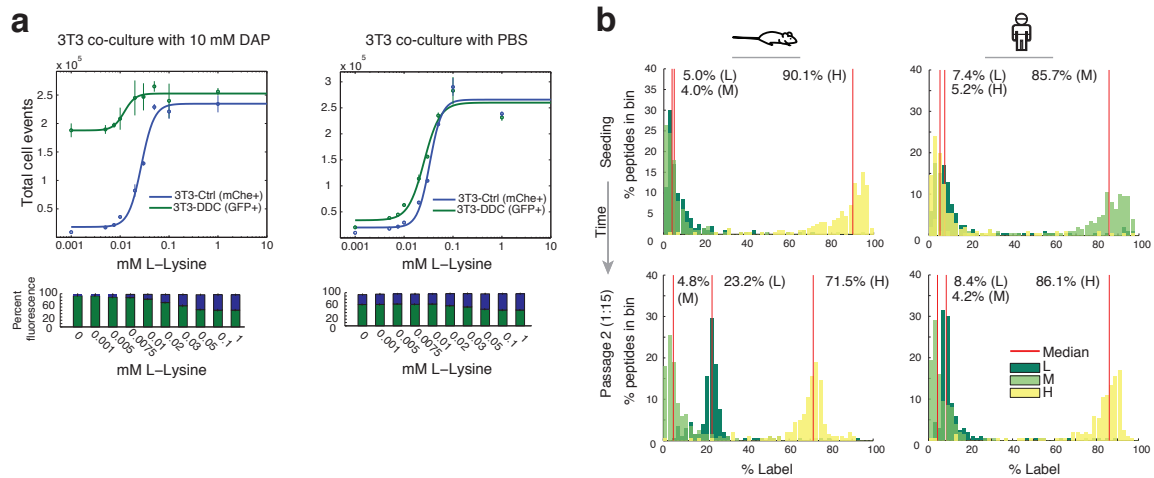
Figure 2.S13: Cell-selective labeling of co-cultures using one enzyme-precursor pair. (a) Co-culture of DDC expressing GFP+ 3T3 cells and empty vector control mCherry+ 3T3 cells with (left panel) or without (right panel) 10 mM DAP and various concentrations of L-Lysine. After 72 h in co-culture, flow cytometry was used to determine the number of GFP+ and mCherry+ cells. Data points represent at least two biological replicates. **(b)** Mouse 3T3 cells expressing DDC were labeled with heavy L-Lysine (H) and human wild-type MDA-MB-231 cells were labeled with medium L-Lysine (M). Cell lysates from these separately labeled cells were combined, analyzed by LC-MS/MS, and labeling status of peptides unique to the mouse (top panel, left) and human (top panel, right) proteome were determined. Co-cultures with similarly labeled cells were also grown for two passages in 40 μM L-Lysine (H) and 10 mM DAP (L) and were analyzed as above (bottom panels). Species-nonspecific peptides were ignored.
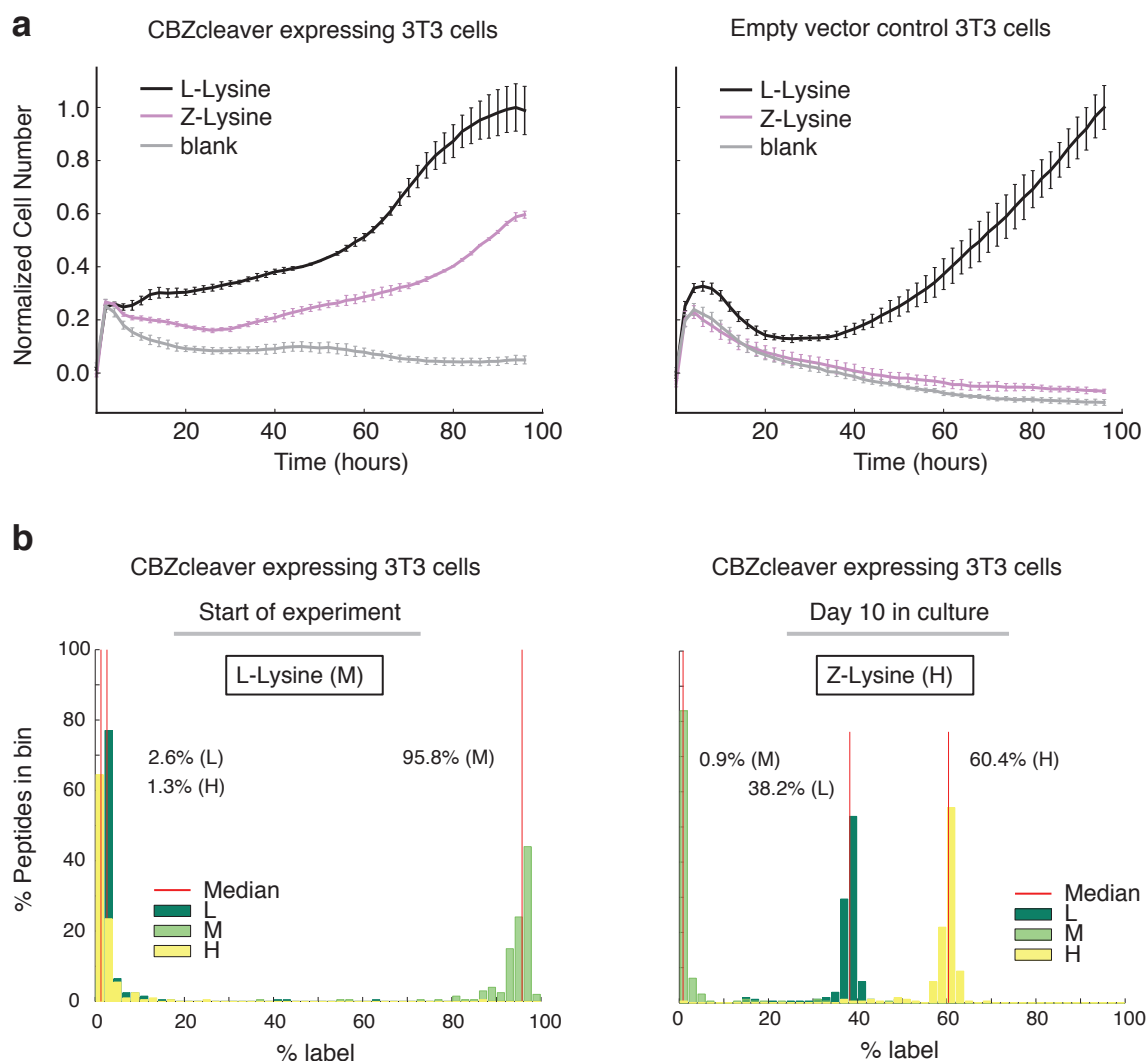
Figure 2.S14: 3T3 cells expressing the CBZcleaver enzyme grow suboptimally on Z-Lysine and partially incorporate L-Lysine produced from Z-Lysine (N_α-cbz-L-Lys). (a) 3T3 cells stably transfected with CBZcleaver (left panel) or empty control vector (right panel) were cultured in 0.798 mM L-Lysine, 2.5 mM Z-Lysine, or without either (blank). Cell growth was estimated by the impedance-based xCELLigence assay and data was normalized to maximum values for each cell-type. Error bars represent the standard deviation of three biological replicates. **(b)** Peptide histograms depicting the light L-Lysine (L), medium L-Lysine (M), and heavy L-Lysine (H) status of the 200 most intense peptides (that contain L-Lysine) in CBZcleaver-expressing 3T3 cells. The labeling status was assessed by quantitative LC-MS/MS at the beginning of the experiment where the cells were labeled with medium L-Lysine (left, M) and after 10 days in L-Lysine-free media with heavy labeled Z-Lysine (right, H). For synthesis of heavy Z-Lysine, see **Supplementary Methods**. The percent label incorporation for the median peptide is indicated (red bars). Concentration of L-Lysine (M) used was 0.798 mM, and Z-Lysine (H) was 2.5 mM. Although specific to CBZcleaver-expressing cells, both growth on Z-Lysine and L-Lysine incorporation based on Z-Lysine were incomplete, and therefore, we discontinued further experimentation with the CBZcleaver-Z-Lysine enzyme-precursor pair.

2.7 Supplementary Methods

2.7.1 Drug Perturbation Assay

Cells were seeded in 96-well plates (2000 cells/well) and grown to 40% confluence in SILAC media containing 0.798 mM K0 or 10 mM DAP DMEM with 10% dialyzed fetal bovine serum (FBS). Cells were then inhibited with eight different drug concentrations (2 fold dilution) in eight replicates. Drugs used were Stattic (STAT3 inhibitor), PI3K-IV (PI3K inhibitor), AKT-VIII (AKT inhibitor), and SL327 (MEK inhibitor). After 48 hours drug treatment cell viability was measured by Resazurin (Sigma) as described by manufacturer. Cell viability relative to untreated cells was calculated to obtain dose-response curves.

2.7.2 Western Blotting

Frozen cell pellets were thawed and lysed for 20 min with NP40 lysis buffer, which contained 1% Nonident P-40, 1 mM sodium orthovanadate, and Complete protease inhibitors (Roche Diagnostics) in PBS. Protein concentrations were determined by the Bradford assay (Bio-Rad) and adjusted to 1-1.5 mg/mL. Protein was then denatured in 2% SDS for 5 minutes at 95°C. Approximately 20 μg of each sample was then separated by SDS-PAGE, transferred to PVDF membrane, and immunoblotted using primary and secondary antibodies. All antibodies were from Cell Signaling. Chemoluminescence visualization was performed on Kodak or HyBlotCL films and films were scanned by a microTEK scanner at 600 d.p.i. in gray scale. The membranes were stripped and reprobed with anti-GAPDH (Cell Signaling) to test for protein loading.

2.7.3 Synthesis of Z-Lysine [N^α-Cbz-lysine(K8)]

To a solution of saturated aqueous NaHCO$_3$ (1.25 mL) and L-lysine·2HCl (250 mg, 1.11 mmol, 1.00 equiv) was added solid NaHCO$_3$ (105 mg, 1.13 equiv, 1.25 mmol) followed by aqueous CuSO$_4$ (1.5 mL, 0.50 M, 0.68 mmol 0.60 equiv), immediately forming a blue copper complex. After stirring for 10 min, di-*tert*-butyl dicarbonate (325 mg, 1.49 mmol, 1.35 equiv) was added in 1 mL acetone. After stirring for 16 h, additional di-*tert*-butyl dicarbonate solid (150 mg, 0.621 equiv, 0.690 mmol) was added. After 24 h, the reaction was quenched with methanol (1 mL) and stirred for an additional 16 h. Ethyl acetate (1 mL) and water (1 mL) were added and the heterogeneous suspension was filtered. The recovered blue solid was taken up in H$_2$O (3 mL), sonicated for 30 s, and filtered. After air drying, the N^ϵ-Boc-protected copper complex was collected as a fine periwinkle blue powder (235 mg, 0.423 mmol, 74.2% yield), which was used without further purification.

To a suspension of N^ϵ-Boc-protected copper complex (235 mg, 0.417 mmol, 1.00 equiv) in acetone (1.5 mL) was added 8-hydroxyquinoline (130 mg, 0.900 mmol, 2.13 equiv) and 10% Na$_2$CO$_3$ (1.8 mL). After 1 h, N-(Benzyloxycarbonyloxy)succinimide (205 mg, 0.821 mmol, 1.97 equiv) in 1 mL acetone was added dropwise over 10 min and stirred for 1 h. The reaction mixture was filtered, and the residue washed with water (3 x 1 mL). The pale green filtrate was acidified carefully with 1 N HCl to a pH of 2, and extracted with ethyl acetate (2 x 5 mL). The combined organics were washed with brine, dried over sodium sulfate, filtered, and concentrated by rotary evaporation to afford crude N^ϵ-Boc-N^α-Cbz-L-lysine(K8) (148 mg, 45.7% yield, 0.381 mmol), where K8 refers to [^{13}C$_6$,^{15}N$_2$]L-Lysine.

To a solution of crude N^ϵ-Boc-N^α-Cbz-L-lysine(K8) (148 mg, 0.381 mmol, 1.00 equiv) in acetone (1.7 mL) was added TsOH·H$_2$O (145 mg, 0.762 mmol, 2.00 equiv). After 16 h,

crystals were collected by vacuum filtration and washed sparingly with cold acetone, giving N^α-Cbz-lysine(K8)·TsOH (124 mg, 71.0% yield, 0.270 mmol).

Crude N^α-Cbz-lysine(K8)·TsOH was dissolved in 1.0 mL 5% acetonitrile (v/v in water), treated with triethylamine (37.5 μL, 0.269 μmol, 1.00 equiv), and purified on a 5.5 g C-18 ISCO RediSep Gold column (5→90% acetonitrile in H_2O). Lyophilization furnished N^α-Cbz-lysine(K8) as a fluffy white amorphous solid (77 mg, 0.27 mmol, 99% yield).

¹H NMR (D_2O, 600 MHz) δ 7.25–7.35 (m, 5H), 5.04 (d, J = 12.5 Hz, 1H), 4.97 (d, J = 12.5 Hz, 1H), 3.83 (dm, J_{CH} = 140.4 Hz, 1H), 2.84 (dm, J_{CH} = 142.8 Hz, 2H), 1.66 (dm, J_{CH} = 128.4 Hz, 1H), 1.53 (dm, J_{CH} = 131.4 Hz, 3H), 1.28 (dm, J_{CH} = 132.6 Hz, 2H); **¹³C-NMR** (D_2O, 151 MHz) δ 179.8 (d, J = 8.4 Hz), 179.5 (d, J = 8.4 Hz), 128.7 (s), 128.2 (s), 127.6 (s), 66.8 (s), 56.2 (ddd, J = 138.0, 46.2, 14.4 Hz), 55.8 (ddd, J = 139.2, 46.8, 15.0 Hz), 34.2 (dt, J = 161.0, 18.6 Hz), 31.1 (td, J = 138.6, 18.0 Hz), 23.2 (td, J = 138.6, 18.8 Hz), 22.0 (t, J = 137.4 Hz); $[\alpha]_D^{19}$:–12.50 ± 0.04° (c = 2.00, 0.2 N HCl); **FTIR** (solid, cm^{-1}) 3306, 3031, 2931, 1717, 1654, 1497, 1402, 1369, 1344, 1232; **ESI-HRMS** (*m/z*): calcd for $C_8{}^{13}C_6H_{21}{}^{15}N_2O_4$ (M+H)$^+$ 289.1643, found 289.1650.

2.7.4   Supplementary Tables

Reaction	Primer Name	Oligonucleotide Sequence (5′→3′)	Product Size
Clone DDC (TAIR id = AT3G14390) from *Arabidopsis thaliana* cDNA			
1	FWD-DDC-XhoI	*GCC***ctcgag**ATGGCGGCAGCTACTCAAT	1475nt
	REV-DDC-EcoRI	*CGC***gaattc**GTTCATAGACCTTCAAAGAAACGC	
Subclone DDC into MSCV-IRES-GFP (pMIG)			
1	FWD-DDC-EcoRI	*ATC***gaattc**ATGGCGGCAGCTACTCAAT	1475nt
	REV-DDC-EcoRI	*CGC***gaattc**GTTCATAGACCTTCAAAGAAACGC	
Subclone DDC into pLM-GFP			
1	FWD-FluorescentGene-AgeI	*CCGGTT***accggt**ATGGTGAGCAAGGGCGAGGAG	795nt
	REV-P2A-FluorescentGene	agggccgggattctcctccacgtcacctgcttgtt tgagtagtgagaagtttgttgctccagatccCTTG TACAGCTCGTCCATGCCG	
2	FWD-P2A-DDC	ggatctggagcaacaaacttctcactactcaaaca agcaggtgacgtggaggagaatcccggccctATGG CGGCAGCTACTCAAT	1533nt
	REV-DDC-SalI	*CCGGTT***gtcgac**TCATAGACCTTCAAAGAAACGCA	
3	FWD-FluorescentGene-AgeI	*CCGGTT***accggt**ATGGTGAGCAAGGGCGAGGAG	2262nt
	REV-DDC-SalI	*CCGGTT***gtcgac**TCATAGACCTTCAAAGAAACGCA	

Three PCR reactions were to generate pLM-GFP-P2A-DDC for insert into pLM using the AgeI and SalI restriction enzymes. In the first reaction, a GFP-P2A oligonucleotide fusion that began with an AgeI site was created. The second reaction generated a PCR fragment of P2A-DDC flanked by SalI. Finally, an overlapping PCR reaction created AgeI-GFP-P2A-DDC-SalI. This sequence was then ligated into the AgeI-SalI digested pLM vector.

Reaction	Primer Name	Oligonucleotide Sequence (5′→3′)	Product Size
Subclone lyr into pLM-mCherry			
1	FWD-FluorescentGene-AgeI	*CCGGTT***accggt**ATGGTGAGCAAGGGCGAGGAG	786nt
	REV-P2A-FluorescentGene	agggccgggattctcctccacgtcacctgcttgtt tgagtagtgagaagtttgttgctccagatccCTTG TACAGCTCGTCCATGCCG	
2	FWD-P2A-lyr	ggatctggagcaacaaacttctcactactcaaaca agcaggtgacgtggaggagaatcccggccctATGA GCCTGGGCATCAGATAC	1300nt
	REV-lyr-SalI	*TGTT***gtcgac**TCAATCCACCAGCACGCG	
3	FWD-FluorescentGene-AgeI	*CCGGTT***accggt**ATGGTGAGCAAGGGCGAGGAG	2020nt
	REV-lyr-SalI	*TGTT***gtcgac**TCAATCCACCAGCACGCG	

Three PCR reactions were to generate pLM-GFP-P2A-lyr for insert into pLM using the AgeI and SalI restriction enzymes. In the first reaction, a mCherry-P2A oligonucleotide fusion that began with an AgeI site was created. The second reaction generated a PCR fragment of P2A-lyr flanked by SalI. Finally, an overlapping PCR reaction created AgeI-mCherry-P2A-lyr-SalI. This sequence was then ligated into the AgeI-SalI digested pLM vector.

Table 2.S1: Primers used in this study. Note that the clamp / extra sequences are italicized and the restriction enzyme sites are highlighted in lowercase bold. Gene sequences are all uppercase and the P2A sequence is in lowercase.

2.7.5 Supplementary Sequences

Diaminopimelate decarboxylase (DDC) from *Arabidopsis thaliana*

Note: The DDC gene was cloned directly from *A. thaliana* cDNA using the primers above. For up-to-date information about DDC in *A. thaliana*, please see AT3G14390 at the Arabidopsis Information Resource (TAIR).

>AT3G14390 from *Arabidopsis thaliana*,Diaminopimelate decarboxylase (DDC)

```
atggcggcagctactcaatttctctcccaaccttcgtctctcaatccacaccaactgaag
aaccaaacctcacaacgctccagaagcatccctgtcttgtctcttaaatccacattgaag
ccacttaaacgcctctccgtgaaagccgccgtcgtttctcaaaactcgtccaaaaccgtg
acgaagttcgatcactgtttcaagaaatcatcagatgggtttctctattgtgaaggaact
aaagttgaggatatcatggagtcagtggagagaagaccctttacttatatagcaaacct
cagatcactagaaacctcgaggcttataaagaagcattggaaggagtgagctctgtgatt
ggttacgctatcaaagctaataacaatcttaaaattttggagcatttgagaagtttaggc
tgtggtgctgtgctcgttagtggaaatgagcttagacttgctcttcgtgctggtttcgat
cccacaaagtgcattttcaatggaaatggcaagtctttggaagatttagttctagctgct
caagaaggtgttttcgttaatgtcgatagtgagtttgacttgaataacattgtggaagct
tcaagaatttctggtaagcaggtcaatgtactgctgcgtatcaatcctgatgttgatcct
caggtgcatccatatgttgctactgggaacaagaactcaaagtttggtatcaggaacgag
aagcttcaatggtttctggatcaggtcaaggcacatcccaaagagctgaagcttgttgga
gctcattgccatctaggctctaccattactaaggtggatatattcagagatgcggcagtt
ctcatgatagaatacattgacgagatccggcgtcaaggttttgaagttagttacttgaac
attggtggtggtttagggattgattataccatgccggcgctgtccttcccacacccatg
gatctcatcaacactgtaagagagcttgttctttcacgagacctgaatctaataatcgag
ccagggagatctctgattgcaaacacttgctgtttcgtcaaccatgtaactggtgtgaag
acgaatggaactaagaacttcatagtcattgatggaagtatggctgagcttatccgtccc
agtctttatgatgcttatcagcacattgagttggtctctcctccaccggctgaagcagag
gttaccaaattcgacgtagtgggtcctgtctgtgaatctgctgatttcctgggcaaagac
agagagcttcccactcctccacagggagctggtctggtggttcatgacgctggtgcatac
tgtatgagcatggcttccacttacaatctcaagatgcgtcctccggaatactgggttgaa
gaagatgggtcgatcactaagataaggcatgctgagacattcgatgaccatttgcgtttc
tttgaaggtctatga
```

Lysine racemase (lyr) from *Proteus mirabilis* as synthesized by GeneArt

>lyr from *Proteus mirabilis*,Nucleotide Sequence Optimized for Mouse Expression by GeneArt

atgagcctgggcatcagatacctggccctgctgcccctgttcgtgatcaccgcttgtcag
cagcccgtgaactacaacccccctgccacacaggtggcccaggtgcagcctgccatcgtg
aacaacagctggatcgagatcagcagaagcgccctggacttcaacgtgaagaaggtgcag
agcctgctgggcaagcagagcagcctgtgtgctgtgctgaagggcgacgcctacggccac
gatctgtctctggtggcccccatcatgatcgagaacaatgtgaagtgcatcggcgtgacc
aacaaccaggaactgaaagaagtgcgggacctgggcttcaagggcagactgatgagagtg
cggaacgccaccgagcaggaaatggcccaggccaccaactacaacgtggaagaactgatc
ggcgacctggacatggccaagagactggacgctatcgccaagcagcagaacaaagtgatc
cccatccacctggctctgaacagcggcggcatgagcagaaacggcctggaagtggacaac
aagtctggcctggaaaaggccaagcagatctcccagctggccaacctgaaggtcgtgggc
atcatgagccactaccccgaagaggacgccaacaaagtgcgcgaggacctggcccggttt
aagcagcagtctcagcaggtgctggaagtgatgggcctggaacggaacaacgtgaccctg
cacatggctaacaccttcgccaccatcaccgtgcccgagagctggctggatatggtgcga
gtgggcggcatcttctacggcgacacaatcgccagcaccgactacaagagagtgatgacc
ttcaagagcaatatcgcctccatcaactactaccccaagggcaacaccgtgggctacgac
agaacctacaccctgaagagggacagcgtgctggctaacatccccgtgggatacgccgac
ggctacagaagagtgttcagcaacgccggccacgccctgatcgctggacagagggtgcca
gtgctgggaaagaccagcatgaacaccgtgatcgtggacatcaccagcctgaacaacatc
aagcccggcgacgaggtggtgttcttcggcaagcagggcaacagcgagatcaccgccgag
gaaatcgaggacatctctggcgccctgttcaccgagatgagcatcctgtggggcgccacc
aatcagcgcgtgctggtggattga

CBZcleaver from *Sphingomonas paucimobilis* as synthesized by GeneArt

>CBZcleaver from *Sphingomonas paucimobilis*,Nucleotide Sequence Optimized for Mouse Expression by GeneArt

```
atggtgcagcccacccctacccctcagagcgagctgcctggcctgatcgccagagacatg
gaaggcctgatgacactgtacagggacctgcacgccaaccccgagctgagcctgcaggaa
gtgaacaccgccgccaagctggccaagagactgaaggccatgaagttcgacgtgacagag
aaagtgggcggcaccggcgtggtggccgtgatgaagaacggcagcggacctgtgctgctg
atcagagccgatatggacggcctgcccgtggtggaacagaccggcctggacttcgccagc
aaagtgcggaccaagacccccgagggcgtggaaacaggcgtgatgcacgcttgcggccac
gacacccacatgaccgccttcatcgagacagccaagctgctgagcagccagaaagacaag
tggaagggcaccctggtgatgatcctgcagcccgccgaggaagtgggcaagggcgccagg
gacatgctggaagatggcctgtacaccagattccccagacccacccacgctatcgccttc
cacgacgccgccaacctccaggctggcgtggtgggatacacccctggctacgccctggcc
aacgtggacagcgtggacatcgtggtgaaaggcctgggaggacacggcgcctacccccag
accaccagagatcccatcgtgctgggcagcagaatcgtgaccagcctgcagaccctggtg
tccagagagcaggacccccaggaccctgccgtggtgacagtgggctctttccaggctggc
gccaagcacaacatcatccccgaccaggctctgctgctgctgaccgtgcggagctacagc
gacgagacaagagccaagctgatcaagggcatcgagagaatcgccagaggcgaggctatc
gccgctggcgtgcccgacgacaagatgcctgtggtgtccgtgaaggacgagttcacccccc
agcacctacaacccccccgagttcgccgagcagatgggcgctctgctgaagggacacttc
gccgagggcagagtggtgaaaacccctgccgtgatgggcggcgaggacttcggcagattc
tacagagccgacaagtctatcaacagcttcatcttctgggtgggaggcgtgccagccgat
aagatggccgctgcccaggccggccagatcacactgcctagcctgcacagcccctttctgg
gcccctgaggccgacaaagtgatcgccaccgccagcgaggccatgaccgtgctggccatg
gacatcctgaagaaggactga
```

Chapter 3

Increasing Isotopic Enrichment in CTAP Labeled Cells

3.1 Abstract

We recently introduced a method, termed **C**ell **T**ype specific labeling with **A**mino acid **P**recursors (CTAP), which extends mass spectrometry based proteomics to allow identified proteins from multicellular environments to be assigned a cell-of-origin. Transgenic expression of enzymes that produce essential amino acids and supplementation of stable isotope labeled precursors to these enzymes provides genetic control of labeling specificity. Although initial results were promising and show that complete differential labeling is possible, the system still exhibited suboptimal labeling in one of the cell types across several co-culture experiments. In this work, we investigate the cause and attempt to implement a solution to the observed incomplete labeling.

3.2 Introduction

Technological advances in mass spectrometry (MS) based proteomics enable the identification and quantification of thousands of proteins and post translational modifications in a single sample [68, 82, 83]. This high number of identifications combined with extremely high specificity for each identification make MS an exciting proteomics tool. Until recently, traditional MS lacked the ability to differentiate and quantitate proteins belonging to distinct cell types in multicellular environments. In order to study interactions and communication in such complex systems, researchers have primarily utilized traditional antibody staining together with microscopy to identify distinct cell types and quantify the proteins they contain. These antibody-based techniques are extremely useful but antibody staining, unlike MS, is unable to simultaneously measure many proteins and exhibits variable specificity depending on the antibody being used.

Our lab recently developed a method, titled **C**ell **T**ype specific labeling with **A**mino acid **P**recursors (CTAP), that allows proteins identified with MS to be quantitatively assigned a cell-of-origin. In short, cells transgenic for amino acid producing enzymes generate their own supply of labeled essential amino acids from labeled precursors. Genetic control of enzyme expression is used to restrict labeling to specific cells of interest. By using two distinct enzyme-precursor pairs in individual cell types, the proteome of each cell type becomes differentially labeled in media supplemented with precursors that contain unequal stable isotopes. The CTAP principle was validated using the enzymes diaminopimelate decarboxylase (DDC) and Lysine racemase (lyr) to produce L-lysine from meso-2,6-diaminopimelate (DAP) and D-lysine, respectively (**Chapter 2**).

Although distinct and differential labeling was obtained in co-culture experiments, only

lyr-expressing cells exhibited complete labeling from the expected precursor (D-lysine). In contrast, the DDC-expressing cells became approximately 55% labeled from the expected precursor DAP, but 45% contamination of L-lysine derived from D-lysine was observed. Similar results were obtained in two separate co-culture systems with different DDC-expressing cell types (HEK293T and 3T3 cells). Importantly, the label status of DDC-expressing cells in mono-culture containing both precursors resulted in complete labeling from only the expected precursor DAP. This result suggests that D-lysine itself causes minimal contamination. The simplest and most likely explanation for this discrepancy in mono- vs. co-culture labeling is that DDC-expressing cells obtained L-lysine produced from lyr catalysis of D-lysine.

In this work, we find that the most probable molecular mechanism for incomplete labeling is lyr secretion and activity within the media. The overall goal of this work is to create a CTAP system that results in complete labeling of the expected stable isotopes in both cell types in co-culture. To try to rectify lyr secretion, engineering a molecular solution is also explored.

3.3 Results

To investigate incomplete labeling, we utilized the phenotypic readout of DDC-expressing to lyr-expressing co-cultured cells in the presence of only one precursor. In this system, we expected that DDC-expressing cells would outcompete lyr-expressing cells in DAP-only conditions and the reverse in D-lysine-only conditions. Note that as this research is exploratory and many of the trial-and-error experiments are descriptive in nature and data are not shown.

Incomplete labeling could be the result of several possible mechanisms. First, lyr-expressing cells could be actively secreting L-lysine directly (or indirectly through metabolized protein), which is picked up by DDC-expressing cells. If this explanation is correct, we surmised that less sharing and therefore less DDC-expressing cell growth would be observed in co-culture at low relative to high D-lysine levels. This decrease in the DDC-expressing population was not observed (data not shown) and we therefore believe this mechanism has little impact on incomplete labeling. Second, phagocytosis of lyr-expressing cells may lead to amino acid exchange. Using a transwell plate, in which the cells are not in direct contact yet small molecules diffuse between the populations, we observed similar growth rates as with direct co-culture (data not shown). This result suggests that phagocytosis is not playing a major role in L-lysine exchange. The last mechanism is that the extracellular lyr leads to a shared pool of L-lysine in the culture media. Secretion is the mechanism we believe to be most responsible for the observed cross contamination.

3.3.1 The lyr enzyme secreted from eukaryotic cells

In two separate co-culture systems (3T3 with MDA cells and HEK with MDA cells) we observed similar amounts of contamination in DDC-expressing cells. In order to determine whether it was the lyr enzyme that was causing the contamination or the MDA cells that produce it, we generated both lyr- and DDC-expressing MDA cells. We co-cultured these two enzyme-expressing cell lines with empty vector control HEK293T cells and asked whether the enzyme-precursor pairs (DDC/DAP or lyr/D-lysine) performed similarly in a competition experiment. At the beginning of the experiment each cell type was seeded at similar levels (**Table 3.1**). After two passages in 10mM DAP, the DDC-expressing (GFP+) MDA cells had essentially outcompeted the HEK293T cells. In contrast, two passages in

Passage	lyr-expressing MDA (mCherry+)		DDC-expressing MDA (GFP+)	
	4mM D-Lysine		10mM DAP	
start	45% HEK	55% MDA	52% HEK	48% MDA
P1	29% HEK	68% MDA	8% HEK	92% MDA
P2	24% HEK	76% MDA	2% HEK	97% MDA

Table 3.1: Co-culture competition of enzyme-expressing MDA cells with control HEK293T cells. DDC- or lyr-expressing MDA cells were plated together with fluorophore-only vector control HEK293T cells in either D-lysine or DAP. At the start of the experiment and after two passages, cells were counted with the Tali image-based cytometer (Invitrogen). The percent of each cell type is given as calculated from the relative GFP+ and mCherry+ cells.

4mM D-lysine yielded only a marginal increase in lyr-expressing MDA cells relative to HEK293T cells. These data indicate that it is something innate about the lyr enzyme, rather than just MDA cells, which leads to the contamination.

In order to determine whether lyr is likely to be secreted from cells, we utilized the "classical signal peptide" predictor SignalP 4.0 [84]. Indeed, the first 19 amino acids were predicted to contain a signal peptide with a cleavage site between positions 19 and 20 (**Figure 3.1A**). The SignalP 3.0 algorithm has an accuracy of over 90% [85], strongly suggesting that lyr contains a signal peptide.

To gain experimental evidence of lyr secretion, we decided to search the mass spectra of the secreted media of DDC-expressing 3T3 cells co-cultured with lyr-expressing MDA cells. These data are summarized in the top of Table 3.2. More lyr peptides were observed in the media than intracellularly and the overall intensity score for lyr was in the top 1% of proteins in the media but only the top 17% intracellularly. Surprisingly, we also observed a number of mCherry and GFP peptides in the secreted media and cell lysate. Overall, the competition experiments, signal peptide prediction, and MS results strongly argue that much of the lyr enzyme is being exported into the media.
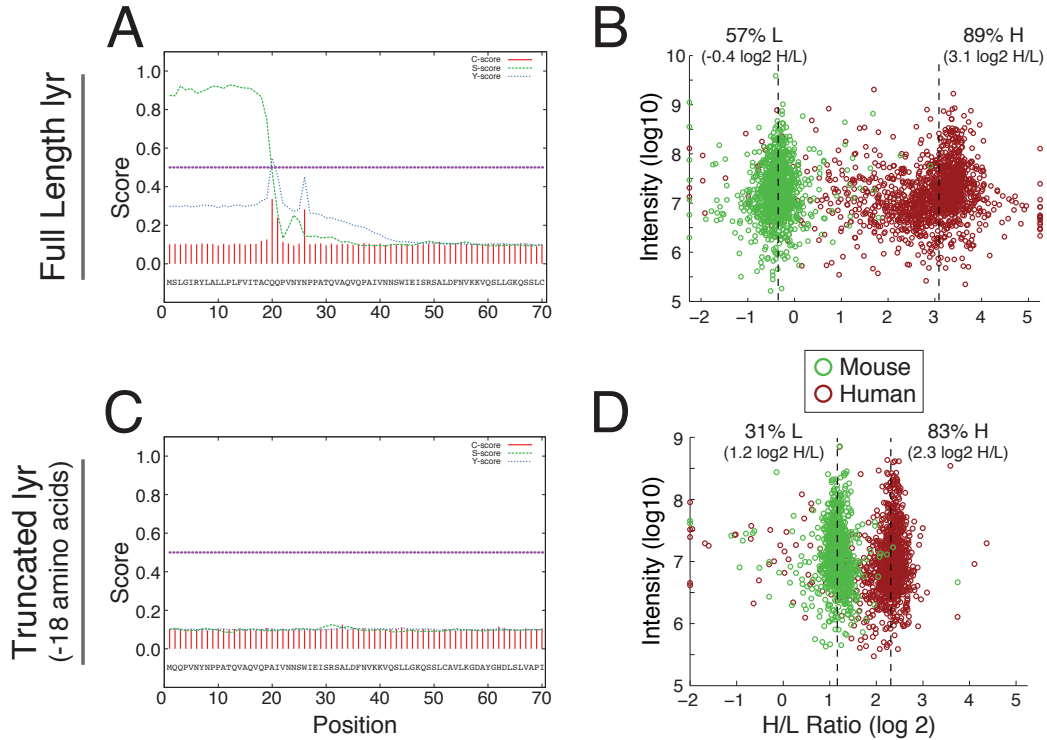
Figure 3.1: The lyr enzyme contains a predicted signal peptide, but removal worsens isotopic enrichment. (**A**) Predicted secretion of lyr from SignalP 4.0. (**B**) H/L peptide ratios observed in co-culture of MDA-MB-231 cells that express full-length lyr (with predicted signal peptide) and DDC-expressing 3T3 cells. Media contained 10mM Light DAP and 4mM Heavy D-lysine and cells were propagated for 13 days. (**C**) Removal of 18 amino acids from the 5 terminus of lyr removes the signal peptide predicted from SignalP 4.0. (**D**) H/L peptide ratios observed in co-culture of MDA-MB-231 cells that express truncated lyr (-18 amino acids) and DDC-expressing 3T3 cells. Cells were propagated for 5 days in 10mM Light DAP and 1mM Heavy D-lysine. For SignalP prediction, standard parameters were used with Organism group set to Eukaryotic.

3.3.2   Engineering a solution to lyr secretion

In order to label distinct cells in co-culture, each enzyme should act on its respective precursor inside of the cells. We therefore decided to rationally engineer a form of the protein that would would not be secreted.

Signal peptide removal

The most obvious way to keep lyr inside of cells is to remove the signal peptide. Several crystal structures of amino acid racemases with high sequence homology to lyr have been

	Protein	Intracellular # Peptides	Media # Peptides	Intracellular Rank of 100	Media Rank of 100
Full-length lyr	lyr	5	30	17	1
	mCherry	4	7	37	27
	DDC	5	0	52	NA
	GFP	7	8	8	4
Truncated lyr	lyr	11	13	2	2
	mCherry	7	7	6	12
	DDC	0	0	NA	NA
	GFP	6	8	8	2

Table 3.2: MS/MS analysis of transgenic proteins in the secreted media and cell lysate of a DDC-expressing 3T3 and lyr-expressing MDA co-culture. The **Rank of 100** indicates the percentage of proteins identified that have equal or higher intensities than the given exogenous protein. Maxquant was used to determine the identifications and calculate the intensities. NA, not identified.

published [86–90]. In order to maintain a functioning form of enzyme, we first explored the sequence-structure relationship by comparing the crystal structure of one of these alanine racemases [87] with the lyr sequence. Many of the active residues in the alanine racemase, including the PLP-binding and catalytic sites, are conserved in lyr. Importantly, the alanine racemase is not predicted to contain a signal peptide by SignalP (data not shown), suggesting that the enzyme may not require secretion to exhibit activity.

In order to remove the signal peptide, we *in silico* truncated the lyr sequence one residue at a time and asked whether each sequence contained a predicted signal peptide by SignalP. Although the removal of 12 or more amino acids abrogated the signal peptide prediction, we chose to truncate the first 18 amino acids to increase the chances of signal peptide removal (**Figure 3.1C**). The cloned construct that contained truncated lyr was used to produce lentivirus and infect MDA-MB-231 cells. In contrast to cells expressing the non-truncated form, these stably infected cells grew immediately on D-lysine without any observed selection period (qualitative, data not shown). This immediate rescue was also observed in other cell types including 3T3 cells, B16 cells, and HEK293T cells. Less en-

Passage	Condition	%GFP+ (DDC)	%mCherry+ (truncated-lyr)	% Neither	% Both
start		51%	46%	0%	0%
P1	L-Lys	61%	36%	0%	1%
P1	1mM D-Lys	42%	56%	0%	0%
P1	10mM DAP	97%	1%	0%	0%

Table 3.3: Competition of DDC-expressing MDA (GFP+) and truncated-lyr-expressing MDA (mCherry+) cells. Passage 0 (P0) shows cells were seeded in roughly equal numbers (51% DDC-expressing and 46% truncated-lyr-expressing). After 1 passage (P1), DDC-cells completely outcompete the lyr-cells in DAP (97% to 1%), but the lyr-cells are unable to outcompete the DDC-cells in D-lysine (56% to 42%). Measurements were made with the Tali image-based cytometer (Invitrogen).

couragingly, co-culture of truncated-lyr-expressing MDA cells with DDC-expressing MDA cells, the lyr-expressing cells were still unable to outcompete DDC-expressing cells (**Table 3.3**).

We next decided to determine whether the truncated form of lyr was able to increase the isotopic enrichment in a DDC-expressing 3T3 and lyr-expressing MDA-MB-231 mix-species co-culture. After 5 days in co-culture, cells were lysed, protein extracted and trypsin digested, and peptides were identified by MS/MS. To our surprise, the truncation actually led to worse isotopic enrichment, with the DDC-expressing 3T3 mouse cells going from 57% to only 31% light enrichment (compare **Figure 3.1B** to **Figure 3.1D**). Analysis of co-culture secreted media revealed that the truncated form of the enzyme was still ending up outside of the cells (**Table 3.2, bottom**). The lyr enzyme was still one of the most intense proteins in the media (top 2%), but there did appear to be more of the enzyme intracellularly based on the ratio of the number of inter- versus intracellular identified peptides ($\frac{11 \text{ intracellular}}{13 \text{ extracellular}}$ truncated-lyr versus $\frac{5 \text{ intracellular}}{30 \text{ extracellular}}$ for full-length-lyr). Taken together the MS/MS and co-culture competition results demonstrate strong evidence that truncated-lyr is still secreted and is active in the media.

Forcing intracellular lyr localization and changing the media to remove lyr

In order to force lyr to remain inside of transgenic cells, we decided introduce a nuclear lo-
calization sequence (NLS) and mitochondrial targeting sequence (MTS) into lyr. To make
validation of localization easier, we first created a mCherry-lyr fusion gene using the 18
amino acid truncated form of lyr (**Figure 3.2A**). The ability of mCherry-lyr to rescue growth
in D-lysine only conditions was verified. This form of lyr was also unable to rapidly out-
compete lyr-free cells in D-lysine-only conditions (data not shown).

Working on the hypothesis that incomplete labeling occurs due to extracellular activity
of lyr, one possible way to alleviate this activity is to remove the protein from the me-
dia. To test this, we prepared a 3T3 and MDA-MB-231 co-culture (as above) and changed
the media daily for 5 days. Mouse-specific peptides derived from the mCherry-lyr fusion
control sample (without media change) showed a higher enrichment of light L-lysine rel-
ative to the unfused construct (from 31% to 44%, compare **Figure 3.1D** with **Figure 3.2B**).
This enrichment was even further increased when the media was changed daily, leading
to a much higher enrichment of light isotopes in the mouse peptides (from 44% to 77%,
compare **Figure 3.2B** with **Figure 3.2C**). Unfortunately, the human peptides were not as
enriched with heavy lysine as previously measured, possibly due to the short culture time
of only 5 days in heavy D-lysine. Regardless of this caveat, these results are encouraging
and suggest that engineering a non-secreted form of the enzyme combined with removal
of enzyme from the media may be a possible solution to the observed contamination.

The addition of localization signals to lyr may force the protein to stay inside the cells,
disallowing lyr activity in the media. Using the mCherry-lyr fusion protein as a starting
point, we cloned a construct that contained a MTS and a construct that contained a NLS on
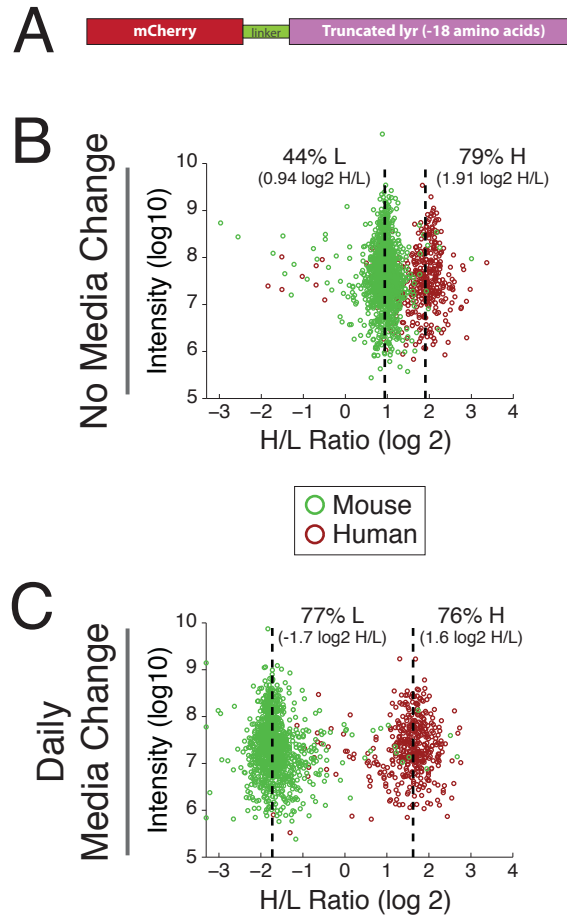
Figure 3.2: Generating a mCherry-lyr fusion and testing its effect on isotopic enrichment in co-culture. (**A**) Schematic of the mCherry-lyr fusion generated and transgenically expressed in MDA-MB-231 cells. (**B**) H/L ratio of all species-specific peptides identified by MS/MS from a co-culture containing MDA-MB-231 cells that express the mCherry-lyr fusion (above) 3T3 cells that express DDC. Cells were grown for 5 days in media containing 10mM Light DAP and 1mM Heavy D-lysine. Media was not changed during the course of the experiment. (**C**) As in (**B**), but media was changed daily throughout the course of the experiment.

the 5′ end of the gene (**Figure 3.3A** and **3.3C**). Using these plasmids, virus was created and used to infect MDA-MB-231 cells. The localization of both of the targeted proteins was verified (**Figure 3.3B** and **3.3D**) and preliminary (qualitative) experiments indicate that forcing localization allows lyr-expressing cells to better outcompete DDC-expressing cells in D-lysine only conditions (data not shown). These results are encouraging, but unfortunately also suggest that lyr is still able to provide L-lysine non-autonomously. We have yet
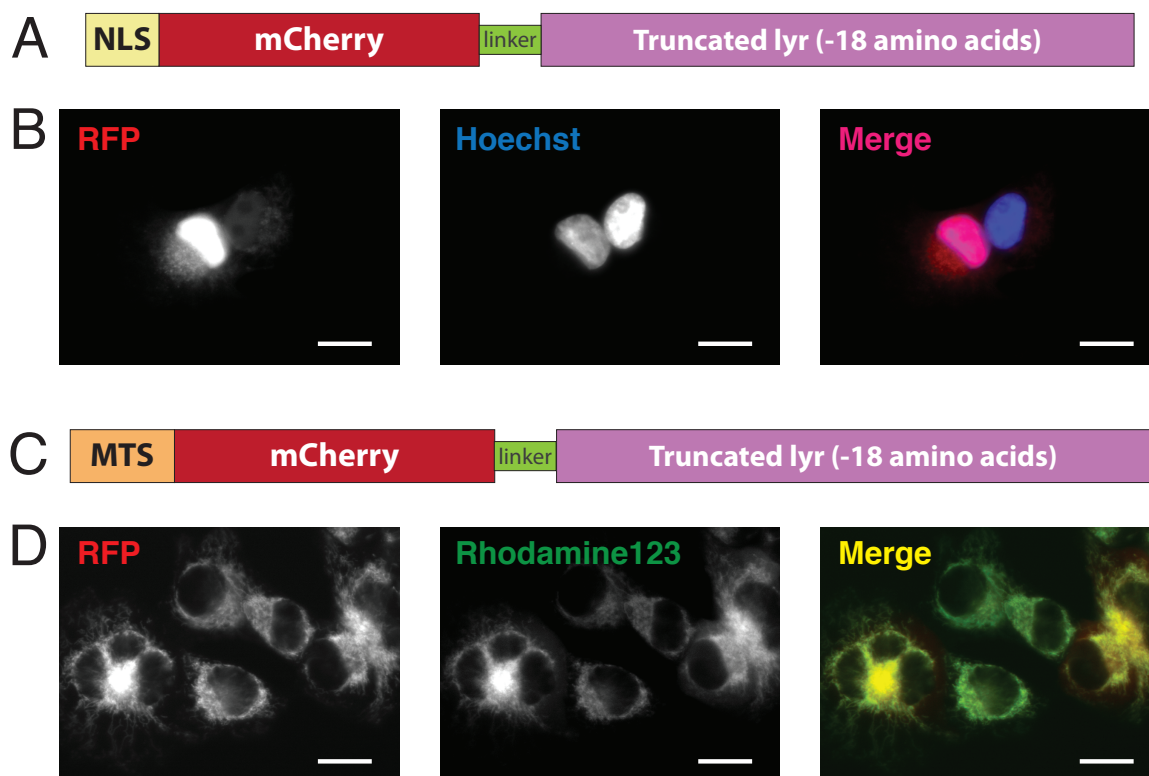
Figure 3.3: Localizing lyr to the nucleus and mitochondria. (**A**) Schematic representation of the lyr construct targeted to the nucleus with a nuclear localization sequence (NLS). (**B**) MDA-MB-231 cells that stably express the construct in **A**. The RFP filed (left) represents the localization of the NLS-mCherry-lyr construct and Hoechst (middle) stains the DNA (nuclear portion). (**C**) Schematic representation of the lyr construct targeted to the mitochondrial with a mitochondrial targeting sequence (MTS). (**D**) MDA-MB-231 cells that stably express the construct in **C**. The RFP filed (left) represents the localization of the MTS-mCherry-lyr construct and Rhodamine123 stains the mitochondria. White bar = 20 μm.

to determine the effect on L-lysine isotopic enrichment using MS/MS, but hopefully the combination of localization and daily media changes will lead to near-complete labeling in co-culture.

3.4 Discussion

In this study, we find that the most likely explanation for lyr-based contamination in co-culture is its secretion and activity in the media. We utilize several approaches to attempt to keep lyr from being secreted. First, truncation to remove the signal peptide paradoxi-

cally led to decreased isotopic enrichment. This result, however, may be explained by our observation that lyr was still present in the media even with this modification. Second, fusing mCherry to truncated-lyr combined with daily media changes led to a substantial decrease in heavy L-lysine containing peptides in DDC-expressing cells. Third, fusion constructs were tagged with MTS and NLS sequences in order to force its localization to the mitochondria and nucleus, respectively. Preliminary data indicate forced localization may decrease contamination and experiments are ongoing to determine the enrichment effect of these forms of lyr. As of yet, however, research undertaken in this study has not completely fixed the incomplete labeling problem.

Our experiences with the lack of contamination from the DDC enzyme make us confident that this issue is not innate to the CTAP hypothesis, but rather a technological issue with the lyr enzyme itself. We believe that several of the strategies discussed have the potential to solve these issues and a solution to this problem is actively under investigation.

Chapter 4

Discussion

The advances we have made in introducing stable isotopes into the proteins of specific cell types in multicellular environments have created a compelling experimental methodology for studying a wide variety of cell-cell communication. I plan to apply this methodology to several established *in vitro* tumor-stroma co-culture systems in the near term, asking questions about the role of macrophages in metastasis and the mechanisms by which the microenvironment is able to attenuate tumor-cell response to therapy. Algorithmic analysis of these data will be used to interpret the alterations induced in multicellular culture. In addition, future work will also include optimization and extension of the CTAP methodology, potentially allowing its future application *in vivo*. In the long term I envision the simultaneous use of CTAP with mouse models of cancer as a novel system for biomarker discovery.

4.1 Comprehensive mapping and modeling of tumor-stroma interactions *in vitro*

The lack of high-throughput proteomic methods that are able to discriminate the cell-of-origin of proteins identified in multicellular environments has hindered the ability of researchers to address important biological problems. In the context of cancer research, understanding the molecular mechanisms by which tumor-stroma interactions influence tumorigenesis would particularly benefit from more comprehensive proteomic methodologies. In this section, I will introduce several specific biological questions and describe how systems-wide proteomic profiling using CTAP can be used to begin to address these problems.

4.1.1 Understanding the role of tumor-associated macrophages in metastasis.

Macrophages have been shown to promote angiogenesis [91], protect cancer cells from chemotherapy [28], and be directly involved in metastasis [92]. What intracellular signaling events occur when macrophages interact with breast cancer cells that have metastasized to the brain, lung, or bone? Does the signature of metastatic-macrophage interaction resemble macrophage communication with the primary tumor? Can we predict the site of metastasis based on the interaction profile between tumor and macrophage? Mapping this reciprocal communication using CTAP will help elucidate the pleiotropic roles macrophages play in tumorigenesis.

We have implemented an experimental system that enables the generation of macrophages from PMA treated THP1 monocytes [93]. CTAP relevant derivatives of the THP1 line, which stably express L-lysine producing enzymes, have been generated and their ability

to differentiate into macrophages has been verified. By co-culturing these macrophages with a variety of primary and metastatic tumor cells, we plan to comprehensively profile the proteomic and phosphoproteomic response of these lines to one another. Computational analysis will focus on finding an interaction signature of primary- and metastatic-macrophage co-cultures. This research will help us understand how macrophages influence tumorigenesis in a variety of cancer types and may point to drug targets for inhibition or activation of these interactions.

4.1.2   Investigating stromal attenuation of drug response.

Building evidence indicates that the tumor microenvironment plays an important contributing role in drug resistance. Several molecular mechanisms for this attenuation have been described. For example, macrophage-derived cathepsins protect breast cancer cell death following Taxol treatment [28], stromal secretion of hepatocyte growth factor (HGF) has been shown to induce resistance to RAF inhibitors in melanoma [6, 94], and microenvironment mediated activation of the Wnt pathway in tumor cells can attenuate chemotherapeutic effects in prostate cancer models [7]. For many co-culture systems the details of these mechanisms are unknown [6]. I am interested in investigating the global signaling response of these understudied tumor-stroma co-cultures to drug perturbation. Can proteomic signals between these cell types describe the mechanisms of drug resistance? Is it possible to re-sensitize tumor cells by interfering with key stromal signals or targeting resistance-activated pathways?

In order to study these questions, we plan to recapitulate several *in vitro* microenvironment models of tumor-stromal drug response. A recent publication by Straussman *et al.* [6], which explored approximately 1000 tumor-stroma co-cultures in the presence of anti-

cancer drugs, will serve as a reference to select interesting tumor-stroma-drug combinations. Using CTAP we plan to profile the proteomic, phosphoproteomic, and secretome responses induced by interactions between select cell types and drug perturbation. These data will be passed through a computational analysis pipeline to delineate how stromal co-culture allows the tumor cells to escape drug sensitivity. To find common and distinct mechanisms of resistance in the different co-cultures, altered pathways will be compared across all models. I anticipate that this research will provide clinically-translatable insight into how the tumor microenvironment influences response to therapy.

4.2   Future CTAP method development

Although the CTAP method in its current state allows for complete discrimination of the proteome of cells in co-culture, there are areas where the method can be improved. In this section, I will discuss some of the experimental optimizations and extensions to CTAP that we are exploring. There are two overall goals of these improvements. First, increase isotopic enrichment in both cell types in co-culture to levels observed in cells completely labeled by SILAC protocols. Second, optimize the system to allow for quick application in a variety of cell types with minimal phenotypic and molecular disturbance. This section is organized into individual projects aimed at contributing to these goals.

4.2.1   Further increasing cell-specific isotopic enrichment

Secretion of lyr and its subsequent extracellular activity is the most likely source of L-lysine cross contamination. This issue is the biggest hurdle we face in obtaining complete labeling in both cell types and we have begun to make inroads into solving this problem (**Chapter 3**). Finding constructs to keep lyr from being secreted are our top priority and we

are actively pursuing this line of research. In addition to these strategies, there are several other approaches that may restrict lyr activity to lyr-expressing cells.

Removal of lyr from the media via selective protease degradation

One way to decrease extracellular lyr activity is to selectively degrade the enzyme in the media. Selective proteolysis of lyr is possible through the introduction of a protease-specific sequence into the polypeptide. Either recombinant addition of the protease or ectopic expression of a secreted form of the protease by DDC-expressing cells would result in lyr degradation.

Several challenges exist in setting up a proteolytic solution to specifically degrade lyr. First, it will be important to select an appropriate protease that is highly specific and does not target any vertebrate antigens. There has been a significant amount of research into engineering novel proteases and this will provide background for selection [95]. Second, it will be important to choose a sequence location for the cleavage site that is both proteolytically accessible and in which breakage of the amino acid bond results in enzyme inactivation. Structural analysis of several published lyr orthologs will help guide a rational decision for this location [86–90]. Successful implementation of lyr degradation may effectively disallow L-lysine production outside of lyr-expressing cells.

Antagonizing lyr function in the media with a dominant negative mutant

Another possible strategy to keep lyr from producing L-lysine in the media is through dominant negative inactivation. Orthologous crystal structures of alanine racemases suggest that the lyr enzyme likely exists as a dimer [86–90] and multimers can be subject to inactivation when binding to inactivated forms of their binding partners through several

mechanisms [96]. Ectopic addition of a catalytically-inactive form of lyr could block the function of activated lyr in the media through one of these mechanisms. Further, successful inactivation of multimers using structure based design decision has been demonstrated [97] and we plan to use orthologous structures of lyr to rationally choose and introduce mutations that render the final dimer non-functional. These mutations may effectively prohibit L-lysine production in the media, decreasing lyr-based contamination.

Inhibiting L-lysine import

Inhibition of L-lysine import could keep lyr-produced L-lysine from ever entering in DDC-expressing cells. We have initiated shRNA knockdown of the primary transporter of L-lysine in mammalian cells (CAT-1, [98, 99]) in an attempt to prevent uptake of soluble L-lysine. There are several potential challenges in implementing this approach. For example, multiple L-lysine transporters exist and this redundancy may make it infeasible to prevent import. Further, the CAT-1 transporter also serves to import L-arginine and inhibition may cause starvation of this or other essential amino acids. Finally, it is possible the the precursors themselves utilize the CAT-1 transporter, effectively disallowing enzyme-precursor enabled growth. Future studies will need to be carefully designed to determine the feasibility of this strategy.

4.2.2 Increasing the enzyme-precursor toolkit

Additional enzyme-precursor pairs that produce L-lysine would allow for more complex culture systems (e.g., three cell-types) and would also provide options when engineering new cell types. Further, the use of another non-secreted enzyme is another possible solution to cross-contamination caused by lyr activity in the media. There are several re-

quirements for enzyme-precursor pairs to be suitable for the CTAP method. First, the enzyme and precursor must not exist in vertebrate systems. Second, any enzyme must be catalytically active upon transgenic expression in vertebrate cells. Third, precursors must be able to cross the membrane barrier and be available for catalysis inside cells. Fourth, the enzyme should not cross react with other CTAP precursors and would optimally not perturb endogenous metabolites. From literature review, we have curated a set of candidate precursors and these can be used as a starting point for experimental validation (**Table 4.1**).

Phenotypic comparison of cell growth in L-lysine-free media with and without each of the precursors can be used to rapidly test for the unlikely possibility that vertebrate cells have a mechanism to catalyze a reaction of each compound to L-lysine. For candidate precursors
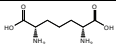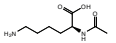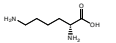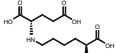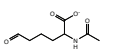
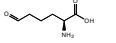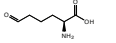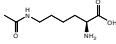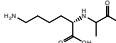|	Compund	Availability	Enzyme (EC#)	# Reactions to L-lysine	In Vertebrates ⋆
	DAP	Sigma (33240)	DDC (4.1.1.20)	1	No
	N2A	Sigma (A2010)	LysK	1	No
	D-Lysine	Sigma (L8021)	lyr (5.1.1.5)	1	No
	Saccharopine	Sigma (S1634)	LYS1 (1.5.1.7)	1	Yes
	N2AAS	custom	LysJ	2	No
	Allysine	custom	(2.6.1.36)	1	unknown
	Allysine	custom	(1.5.1.10+1.5.1.7)	2	unknown
	N6A	custom	(3.5.1.17)	1	enzyme in rats?
	N2DCL	custom	(1.5.1.16)	1	unknown

Table 4.1: Candidate precursors for the enzymatic production of L-lysine in vertebrate cells. Several requirements are essential for their use in our system. It is important that these metabolites are not found endogenously within vertebrate systems, that mammalian cells are unable to innately utilize them for lysine production, and that each is not toxic. Further, we desire easy import and easy synthesis. DAP = meso-2,6-Diaminopimelic acid, N2A = N2-Acetyl-L-lysine, DDC = diaminopimelate decarboxylase, N2AAS = N2-Acetyl-L-aminoadipate semialdehyde, N6A = N6-Acetyl-L-Lysine, N2DCL = N2-(D-1-carboxyethyl)-L-lysine. ⋆ = from literature searches and to the best of our knowledge.

that are unable to rescue growth, the enzymes which catalyze their conversion to L-lysine will be synthesized or cloned. Transgenic cell lines will be generated and validation of growth rescue will be used to determine if the enzyme is suitable for the CTAP method.

4.3   A unique approach to biomarker discovery

The past several decades have witnessed significant improvement in our ability to understand and treat cancer. In contrast to this revolution in treatment, clinical tools for finding markers indicative of disease onset or response to therapy have not witnessed such progress [100]. This lack of advancement is in part due to technical limitations. For example, the abundance of proteins in biological fluids (e.g., plasma, urine, breast ductal fluid, etc.) span a large dynamic range and proteins originating from diseased tissue likely exist in relative low abundance [101, 102]. Profiling sub-stoichiometric proteins is difficult, but slowly becoming possible with advances in MS technology and fractionation techniques [102–106]. Even as these technical issues are solved, the biological interpretation of altered protein abundance (i.e., in diseased individuals versus controls) will be difficult to interpret. Many of the clinically useful biomarkers, such as prostate specific antigen (PSA), cancer antigen 125 (CA125), and α-fetoprotein, are thought to originate from the organ that houses the diseased cells or from the diseased cells themselves [59, 100]. If researchers had a handle on the origin of the proteins identified in proximal fluids, would this lead to more biologically relevant markers of tumor emergence, response to therapy, and disease recurrence?

One unique aspect of CTAP is the ability to discriminate the cell-of-origin of secreted factors (**Chapter 2**). In theory, this methodology is applicable *in vivo*, providing a way to

link proteins identified in the blood of animal models to their cell-of-origin. Such a link would provide biological meaning, rather than just statistical significance, to diagnostic and prognostic biomarkers found to originate directly from diseased cells in mouse models of cancer. The identification of this link in mice may help narrow the search space of potential markers when looking for similar proteins in human patients.

There are several possible approaches to applying CTAP *in vivo* for biomarker discovery. Initial studies will likely involve transplanting transgenic cell lines generated *in vitro* into mice [107, 108], followed by supplementation of labeled precursors to the animals food supply. This approach will allow us to validate and troubleshoot the technology with the ultimate goal of creating transgenic animals that can be crossed with a variety of cancer models. A number of challenges will likely need to be overcome for successful application of CTAP *in vivo*, however with proper optimization this technology would provide a unique approach to biomarker discovery.

4.4   Concluding Remarks

This thesis has presented a new method—CTAP—for selective and continuous proteome labeling of distinct cell populations in multicellular environments. Utilizing quantitative mass spectrometry, this technology allows the cell-of-origin of secreted factors and intracellular signaling events to be determined in co-culture systems at a scale not possible with other methods. The hypothesis and feasibility of the method was demonstrated in co-culture (**Chapter 2**) and some initial steps we have taken to increase isotopic separation across the cell types was presented (**Chapter 3**). In the near future, we plan to use CTAP to interrogate several established co-culture systems, asking asking questions about how

tumor-stroma interactions affect cancer progression, maintenance, and altered drug efficacy. These studies will be performed in parallel with several projects aimed at enhancing isotopic enrichment and extending the method to *in vivo* models. In the long term, I envision the use of CTAP as a platform for diagnostic and prognostic biomarker discovery, linking labeled proteins directly to diseased cells in mouse models of cancer.

Bibliography

[1] Gocheva, V. *et al.* Il-4 induces cathepsin protease activity in tumor-associated macrophages to promote cancer growth and invasion. *Genes & Development* **24**(3), 241–255 Feb (2010).

[2] Hanahan, D. & Coussens, L. M. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* **21**(3), 309–22 Mar (2012).

[3] Olumi, A. F. *et al.* Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res* **59**(19), 5002–11 Oct (1999).

[4] Joyce, J. A. Therapeutic targeting of the tumor microenvironment. *Cancer Cell* **7**(6), 513–20 Jun (2005).

[5] Mcmillin, D. W. *et al.* Tumor cell-specific bioluminescence platform to identify stroma-induced changes to anticancer drug activity. *Nat Med* **16**(4), 483–9 Apr (2010).

[6] Straussman, R. *et al.* Tumour micro-environment elicits innate resistance to raf inhibitors through hgf secretion. *Nature* **487**(7408), 500–4 Jul (2012).

[7] Sun, Y. *et al.* Treatment-induced damage to the tumor microenvironment promotes prostate cancer therapy resistance through wnt16b. *Nat Med* Aug (2012).

[8] Joyce, J. A. & Pollard, J. W. Microenvironmental regulation of metastasis. *Nat Rev Cancer* **9**(4), 239–52 Apr (2009).

[9] Waters, C. M. & Bassler, B. L. Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **21**, 319–46 Jan (2005).

[10] Sprague, G. F., Blair, L. C. & Thorner, J. Cell interactions and regulation of cell type in the yeast saccharomyces cerevisiae. *Annu Rev Microbiol* **37**, 623–60 Jan (1983).

[11] Dhein, S. Pharmacology of gap junctions in the cardiovascular system. *Cardiovasc Res* **62**(2), 287–98 May (2004).

[12] Herbert, S. P. & Stainier, D. Y. R. Molecular control of endothelial cell behaviour during blood vessel morphogenesis. *Nat Rev Mol Cell Biol* **12**(9), 551–64 Sep (2011).

[13] Lai, E. C. Notch signaling: control of cell communication and cell fate. *Development* **131**(5), 965–73 Mar (2004).

[14] Springer, T. A. Adhesion receptors of the immune system. *Nature* **346**(6283), 425–34 Aug (1990).

[15] Arvanitis, D. & Davy, A. Eph/ephrin signaling: networks. *Genes & Development* **22**(4), 416–29 Feb (2008).

[16] Fortini, M. E. Notch signaling: the core pathway and its posttranslational regulation. *Developmental Cell* **16**(5), 633–47 May (2009).

[17] Itoh, N. & Ornitz, D. M. Fibroblast growth factors: from molecular evolution to roles in development, metabolism and disease. *Journal of Biochemistry* **149**(2), 121–30 Feb (2011).

[18] Theunissen, J.-W. & de Sauvage, F. J. Paracrine hedgehog signaling in cancer. *Cancer Res* **69**(15), 6007–10 Aug (2009).

[19] Fagan, D. H. & Yee, D. Crosstalk between igf1r and estrogen receptor signaling in breast cancer. *J Mammary Gland Biol Neoplasia* **13**(4), 423–9 Dec (2008).

[20] Audet, M. & Bouvier, M. Restructuring g-protein- coupled receptor activation. *Cell* **151**(1), 14–23 Sep (2012).

[21] Manning, G., Plowman, G. D., Hunter, T. & Sudarsanam, S. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* **27**(10), 514–20 Oct (2002).

[22] Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**(7), 1117–34 Jun (2010).

[23] Bray, S. J. Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Biol* **7**(9), 678–89 Sep (2006).

[24] Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**(1), 57–70 Jan (2000).

[25] Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**(5), 646–74 Mar (2011).

[26] Folkman, J. Tumor angiogenesis: therapeutic implications. *N Engl J Med* **285**(21), 1182–6 Nov (1971).

[27] Weis, S. M. & Cheresh, D. A. Tumor angiogenesis: molecular pathways and therapeutic targets. *Nat Med* **17**(11), 1359–70 Jan (2011).

[28] Shree, T. *et al.* Macrophages and cathepsin proteases blunt chemotherapeutic response in breast cancer. *Genes & Development* **25**(23), 2465–79 Dec (2011).

[29] Ahn, G.-O. *et al.* Inhibition of mac-1 (cd11b/cd18) enhances tumor response to radiation by reducing myeloid cell recruitment. *Proc Natl Acad Sci USA* **107**(18), 8363–8 May (2010).

[30] Morgan, T. The theory of the gene. *American Naturalist* , 513–544 (1917).

[31] Baum, B. & Craig, G. Rnai in a postmodern, postgenomic era. *Oncogene* **23**(51), 8336–9 Nov (2004).

[32] Oricchio, E. *et al.* The eph-receptor a7 is a soluble tumor suppressor for follicular lymphoma. *Cell* **147**(3), 554–564 (2011).

[33] Sledz, C. A. & Williams, B. R. G. Rna interference in biology and disease. *Blood* **106**(3), 787–94 Aug (2005).

[34] Chen, R., Pan, S., Aebersold, R. & Brentnall, T. Proteomics studies of pancreatic cancer. *Proteomics Clinical applications* **1**(12), 1582–1591 Jan (2007).

[35] Wulfkuhle, J. D., Liotta, L. A. & Petricoin, E. F. Proteomic applications for the early detection of cancer. *Nat Rev Cancer* **3**(4), 267–75 Apr (2003).

[36] Tibes, R. *et al.* Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Molecular Cancer Therapeutics* **5**(10), 2512–21 Oct (2006).

[37] Espina, V. *et al.* Protein microarrays: molecular profiling technologies for clinical specimens. *Proteomics* **3**(11), 2091–100 Nov (2003).

[38] Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**(6928), 198–207 Mar (2003).

[39] Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**(5771), 212–7 Apr (2006).

[40] Cox, J. & Mann, M. Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* Nov (2008).

[41] Mann, M. & Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* **66**(24), 4390–9 Dec (1994).

[42] Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**(18), 3551–67 Dec (1999).

[43] Yates, J. R., Cociorva, D., Liao, L. & Zabrouskov, V. Performance of a linear ion trap-orbitrap hybrid for peptide analysis. *Anal Chem* **78**(2), 493–500 Jan (2006).

[44] Ong, S.-E. *et al.* Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* **1**(5), 376–86 May (2002).

[45] Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**(10), 994–9 Oct (1999).

[46] Ross, P. L. *et al.* Multiplexed protein quantitation in saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics* **3**(12), 1154–69 Dec (2004).

[47] Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quan-

tification of proteins and phosphoproteins from cell lysates by tandem ms. *Proc Natl Acad Sci USA* **100**(12), 6940–5 Jun (2003).

[48] Hanke, S., Besir, H., Oesterhelt, D. & Mann, M. Absolute silac for accurate quantitation of proteins in complex mixtures down to the attomole level. *J Proteome Res* **7**(3), 1118–30 Mar (2008).

[49] Ong, S.-E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (silac). *Nat Protoc* **1**(6), 2650–60 Jan (2006).

[50] Bandura, D. R. *et al.* Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem* **81**(16), 6813–22 Aug (2009).

[51] Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**(6030), 687–96 May (2011).

[52] Ornatsky, O. *et al.* Highly multiparametric analysis by mass cytometry. *Journal of Immunological Methods* **361**(1-2), 1–20 Sep (2010).

[53] Connor, R., Piatkov, K., Varshavsky, A. & Tirrell, D. Enzymatic nterminal addition of noncanonical amino acids to peptides and proteins. *ChemBioChem* **9**(3), 366–369 (2008).

[54] Johnson, J., Lu, Y. & Deventer..., J. V. Residue-specific incorporation of noncanonical amino acids into proteins: recent developments and applications. *Current Opinion in ...* Jan (2010).

[55] Mukai, T. *et al.* Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-trna synthetases. *Biochem Biophys Res Commun* **371**(4), 818–22 Jul (2008).

[56] Ngo, J. T. *et al.* Cell-selective metabolic labeling of proteins. *Nat Chem Biol* **5**(10), 715–7 Oct (2009).

[57] Truong, F., Yoo, T. H., Lampo, T. J. & Tirrell, D. A. Two-strain, cell-selective protein labeling in mixed bacterial cultures. *Journal of the American Chemical Society* May (2012).

[58] Naba, A. *et al.* The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol Cell Proteomics* **11**(4), M111.014647 Apr (2012).

[59] van den Bemd, G.-J. C. M. *et al.* Mass spectrometric identification of human prostate cancer-derived proteins in serum of xenograft-bearing mice. *Mol Cell Proteomics* **5**(10), 1830–9 Oct (2006).

[60] Jørgensen, C. *et al.* Cell-specific information processing in segregating populations of eph receptor ephrin-expressing cells. *Science* **326**(5959), 1502–9 Dec (2009).

[61] Rechavi, O. *et al.* Trans-silac: sorting out the non-cell-autonomous proteome. *Nat Methods* **7**(11), 923–7 Nov (2010).

[62] Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annual review of biochemistry* **79**, 413–44 Jan (2010).

[63] Xu, H., Andi, B., Qian, J., West, A. H. & Cook, P. F. The alpha-aminoadipate pathway for lysine biosynthesis in fungi. *Cell Biochem Biophys* **46**(1), 43–64 Jan (2006).

[64] Saqib, K. M., Hay, S. M. & Rees, W. D. The expression of escherichia coli diaminopimelate decarboxylase in mouse 3t3 cells. *Biochim Biophys Acta* **1219**(2), 398–404 Oct (1994).

[65] Jouanneau, J., Stragier, P., Bouvier, J., Patte, J. C. & Yaniv, M. Expression in mammalian cells of the diaminopimelic acid decarboxylase of escherichia coli permits cell growth in lysine-free medium. *Eur J Biochem* **146**(1), 173–8 Jan (1985).

[66] Sury, M. D., Chen, J.-X. & Selbach, M. The silac fly allows for accurate protein quantification in vivo. *Mol Cell Proteomics* **9**(10), 2173–83 Oct (2010).

[67] Spellman, D. S., Deinhardt, K., Darie, C. C., Chao, M. V. & Neubert, T. A. Stable isotopic labeling by amino acids in cultured primary neurons: application to brain-derived neurotrophic factor-dependent phosphotyrosine-associated signaling. *Mol Cell Proteomics* **7**(6), 1067–76 Jun (2008).

[68] Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics* **11**(3), M111.014050 Mar (2012).

[69] Zhang, Y. *et al.* Lectin capture strategy for effective analysis of cell secretome. *Proteomics* **12**(1), 32–6 Jan (2012).

[70] Kuan, Y. *et al.* Biochemical characterization of a novel lysine racemase from proteus mirabilis bcrc10725. *Process Biochemistry* (2011).

[71] Nanduri, V., Goldberg, S., Johnston, R. & Patel, R. Cloning and expression of a novel enantioselective n-carbobenzyloxy-cleaving enzyme. *Enzyme and Microbial Technology* **34**(3-4), 304–312 (2004).

[72] Mavrakis, K. J. *et al.* Tumorigenic activity and therapeutic inhibition of rheb gtpase. *Genes & Development* **22**(16), 2178–88 Aug (2008).

[73] Swift, S., Lorens, J., Achacoso, P. & Nolan, G. P. Rapid production of retroviruses for efficient gene delivery to mammalian cells using 293t cell-based systems. *Curr Protoc Immunol* **Chapter 10**, Unit 10.17C May (2001).

[74] Szymczak, A. L. *et al.* Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2a peptide-based retroviral vector. *Nat Biotechnol* **22**(5), 589–94 May (2004).

[75] Smyth, G. Limma: linear models for microarray data. *Bioinformatics and computational biology solutions using R and Bioconductor* , 397–420 (2005).

[76] Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**(1), 289–300 (1995).

[77] Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**(6), 2856–60 Jan (2006).

[78] Macek, B. *et al.* Phosphorylation of the human full-length protein kinase ciota. *J Proteome Res* **7**(7), 2928–35 Jul (2008).

[79] Ishihama, Y., Rappsilber, J. & Mann, M. Modular stop and go extraction tips with stacked disks for parallel and multidimensional peptide fractionation in proteomics. *J Proteome Res* **5**(4), 988–94 Apr (2006).

[80] Olsen, J. V. *et al.* Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**(3), 635–48 Nov (2006).

[81] Cox, J. *et al.* Andromeda: a peptide search engine integrated into the maxquant environment. *J Proteome Res* **10**(4), 1794–805 Apr (2011).

[82] Olsen, J. V. *et al.* Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Science signaling* **3**(104), ra3 Jan (2010).

[83] Rikova, K. *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**(6), 1190–203 Dec (2007).

[84] Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. Signalp 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**(10), 785–6 Jan (2011).

[85] Klee, E. W. & Ellis, L. B. M. Evaluating eukaryotic secreted protein prediction. *BMC Bioinformatics* **6**, 256 Jan (2005).

[86] Fenn, T. D., Stamper, G. F., Morollo, A. A. & Ringe, D. A side reaction of alanine racemase: transamination of cycloserine. *Biochemistry* **42**(19), 5775–83 May (2003).

[87] Im, H., Sharpe, M. L., Strych, U., Davlieva, M. & Krause, K. L. The crystal structure of alanine racemase from streptococcus pneumoniae, a target for structure-based drug design. *BMC Microbiology* **11**, 116 Jan (2011).

[88] Morollo, A. A., Petsko, G. A. & Ringe, D. Structure of a michaelis complex analogue: propionate binds in the substrate carboxylate site of alanine racemase. *Biochemistry* **38**(11), 3293–301 Mar (1999).

[89] Noda, M., Matoba, Y., Kumagai, T. & Sugiyama, M. Structural evidence that alanine

racemase from a d-cycloserine-producing microorganism exhibits resistance to its own product. *J Biol Chem* **279**(44), 46153–61 Oct (2004).

[90] Shaw, J. P., Petsko, G. A. & Ringe, D. Determination of the structure of alanine racemase from bacillus stearothermophilus at 1.9-a resolution. *Biochemistry* **36**(6), 1329–42 Feb (1997).

[91] Lin, E. Y. *et al.* Vascular endothelial growth factor restores delayed tumor progression in tumors depleted of macrophages. *Mol Oncol* **1**(3), 288–302 Dec (2007).

[92] Condeelis, J. & Pollard, J. W. Macrophages: obligate partners for tumor cell migration, invasion, and metastasis. *Cell* **124**(2), 263–6 Jan (2006).

[93] Tsuchiya, S. *et al.* Induction of maturation in cultured human monocytic leukemia cells by a phorbol diester. *Cancer Res* **42**(4), 1530–6 Apr (1982).

[94] Wilson, T. R. *et al.* Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors. *Nature* **487**(7408), 505–9 Jul (2012).

[95] Pogson, M., Georgiou, G. & Iverson, B. L. Engineering next generation proteases. *Current Opinion in Biotechnology* **20**(4), 390–7 Aug (2009).

[96] Sheppard, D. Dominant negative mutants: tools for the study of protein function in vitro and in vivo. *Am J Respir Cell Mol Biol* **11**(1), 1–6 Jul (1994).

[97] Steed, P. M. *et al.* Inactivation of tnf signaling by rationally designed dominant-negative tnf variants. *Science* **301**(5641), 1895–8 Sep (2003).

[98] Closs, E. I., Boissel, J.-P., Habermeier, A. & Rotmann, A. Structure and function of cationic amino acid transporters (cats). *J Membr Biol* **213**(2), 67–77 Jan (2006).

[99] Bröer, S. Amino acid transport across mammalian intestinal and renal epithelia. *Physiol Rev* **88**(1), 249–86 Jan (2008).

[100] Sawyers, C. L. The cancer biomarker problem. *Nature* **452**(7187), 548–52 Apr (2008).

[101] Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**(11), 845–67 Nov (2002).

[102] Hanash, S. M., Pitteri, S. J. & Faca, V. M. Mining the plasma proteome for cancer biomarkers. *Nature* **452**(7187), 571–9 Apr (2008).

[103] Addona, T. A. *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* **27**(7), 633–41 Jul (2009).

[104] Keshishian, H., Addona, T., Burgess, M., Kuhn, E. & Carr, S. A. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics* **6**(12), 2212–29 Dec (2007).

[105] Tirumalai, R. S. *et al.* Characterization of the low molecular weight human serum proteome. *Mol Cell Proteomics* **2**(10), 1096–103 Oct (2003).

[106] Zhang, H., Li, X.-J., Martin, D. B. & Aebersold, R. Identification and quantification of n-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **21**(6), 660–6 Jun (2003).

[107] Wang, X. *et al.* A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature* **461**(7263), 495–500 Sep (2009).

[108] Zender, L. *et al.* An oncogenomics-based in vivo rnai screen identifies tumor suppressors in liver cancer. *Cell* **135**(5), 852–64 Nov (2008).