bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Selective Labeling and Identification of the Tumor Cell Proteome of Pancreatic Cancer *In Vivo*

Nancy G. Azizian[1, 2], Litong Nie[3], Delaney K. Sullivan[4], Sammy Pardo[5], Dana Molleur[5], Junjie Chen[3], Susan T. Weintraub[5], Yulin Li[1, 2*]

[1]Center for Immunotherapy Research, Houston Methodist Research Institute, Houston, TX 77030, USA; [2]Department of Medicine, Weill Cornell Medical College, New York, NY 10065, USA; [3]Department of Experimental Radiation Oncology, Division of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; [4]UCLA-Caltech Medical Scientist Training Program, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [5]Department of Biochemistry and Structural Biology, The University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA.

***Corresponding author**: Yulin Li, Houston Methodist Research Institute. 6670 Bertner Avenue. Houston, TX 77030, USA. Phone: 713-441-7350; Fax: 713-441-7438; Email: yli@houstonmethodist.org

Key Words

Pancreatic ductal adenocarcinoma (PDAC), bioorthogonal non-canonical amino acid tagging (BONCAT), data-independent acquisition mass spectrometry (DIA-MS), methionyl-tRNA synthetase (MetRS), azidonorleucine (ANL), patient-derived xenografts (PDX).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is among the deadliest cancers. Dissecting the tumor cell proteome, from that of the non-tumor cells in the PDAC tumor bulk, is critical for tumorigenesis studies, biomarker discovery, and development of therapeutics. However, investigating the tumor cell proteome has proven evasive due to the tumor's extremely complex cellular composition. To circumvent this technical barrier, we have combined bioorthogonal non-canonical amino acid tagging (BONCAT) and data-independent acquisition mass spectrometry (DIA-MS) in an orthotopic PDAC model to specifically identify the tumor cell proteome *in vivo*. Utilizing the tumor cell-specific expression of a mutant tRNA synthetase transgene, this approach provides tumor cells with the exclusive ability to incorporate an azide-bearing methionine analog into newly synthesized proteins. The azide-tagged tumor cell proteome is subsequently enriched and purified via a bioorthogonal reaction, and then identified and quantified using DIA-MS. Applying this workflow to the orthotopic PDAC model, we have identified thousands of proteins expressed by the tumor cells. Furthermore, by comparing the tumor cell and tumor bulk proteomes, we showed that the approach can distinctly differentiate proteins produced by tumor-cells from non-tumor cells within the tumor microenvironment. Our study, for the first time, reveals the tumor cell proteome of pancreatic cancer under physiological conditions, providing broad applications for tumorigenesis, therapeutics, and biomarker studies in various human cancers.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Cancer is not only a mass of proliferating tumors cells, but also consists of a variety of non-tumor cells, secreted factors, and the extracellular matrix, which are collectively known as the tumor microenvironment (TME). The interaction between tumor cells and the surrounding TME has profound impacts on all stages of tumor development. Human pancreatic ductal adenocarcinoma (PDAC), in particular, has a highly complex TME, imparted by a dense desmoplastic stroma and a host of stromal fibroblasts, endothelial, inflammatory, and immune cells. The stromal components of human PDAC may account for up to 80% of the total tumor volume with tumor cells constituting a minor population.[1-3] Dissecting proteins produced by the PDAC tumor cells from those of non-tumor cells in TME is critical for tumorigenesis and therapeutic studies. However, the heterogeneous and complex cellular composition of the PDAC tumor mass, has thus far, precluded precise isolation and identification of the PDAC tumor cell proteome *in vivo*.

Direct investigation of the PDAC tumor cell proteome requires selective purification of proteins from the tumor cells and not the non-tumor cells within the tumor bulk. Several recent studies have focused on cell-selective metabolic labeling of the proteomes[4]. These approaches include cell-type-specific labeling using amino acid precursors (CTAP),[5] bioorthogonal non-canonical amino acid tagging (BONCAT),[6] and stochastic orthogonal recording of translation (SORT)[7]. BONCAT has been shown to label cell-selective proteomes in the fruit fly,[8] as well as mouse brain and muscle.[9-11] BONCAT works through bioorthogonal chemical reactions that do not exist in nature, and thus will not cross-react with any physiological processes in the cells.[12,13] This technique relies on bioorthogonal incorporation of azide-bearing methionine analogs, such

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

as azidonorleucine (ANL) and azidohomoalanine (AHA), into newly synthesized polypeptides. Due to the small size of the azide moiety, ANL or AHA incorporation has no apparent effect on protein function.[9-11] During protein translation, ANL is preferentially recognized and charged onto $tRNA^{Met}$ by a mutant methionyl-tRNA synthetase ($MetRS^{L274G}$), and is subsequently incorporated in the elongating polypeptide chains[14] (**Fig 1A-B**). ANL-tagged proteins may be selectively conjugated and enriched through azide-alkyne cycloaddition.[13] Further identification of the proteome is achieved through mass spectrometric (MS) analysis of the ANL-tagged proteins (**Fig 1C**). Incorporation of the azide-bearing amino acids is unbiased, non-toxic, biocompatible, and does not affect protein stability.[6]

The PDAC bulk tumor is composed of tumor cells and many types of non-tumor cells. The ectopic expression of $MetRS^{L274G}$ transgene in tumor cells but not the non-tumor cells in the tumor bulk, enables the exclusive tagging of the tumor cell proteome using ANL. Notably, the absence of the $MetRS^{L274G}$ transgene in various non-tumor cells in the TME precludes ANL incorporation into their proteomes. Following ANL labeling, the tumor cell proteome is enriched and purified for MS analysis. Thus, applying BONCAT to the animal model of PDAC facilitates the identification of tumor cell proteome in a physiological context.

Data-independent acquisition mass spectrometry (DIA-MS) is a highly reproducible state-of-the-art approach for quantitative proteomic analysis.[15-19] Traditionally, data-dependent acquisition mass spectrometry (DDA-MS) has been used in a variety of label-free and label-based methods to measure quantitative changes in global protein levels in biological samples.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

However, the stochastic nature of DDA bears a bias toward higher abundance peptides. Undersampling of medium and low abundance peptides causes inconsistencies in detection of peptides and hampers reproducibility among replicates. In the DIA-MS approach, all precursors are fragmented to yield tandem-MS data, providing sequence information from virtually all peptides in a sample with minimal loss of information. Due to its high accuracy and reproducibility, DIA-MS is a powerful method for comprehensive proteomic studies of complex samples, including tumor specimens.[20-22]

Here we have combined BONCAT bioorthogonal chemistry and DIA-MS proteomics to specifically investigate the tumor cell proteome in an orthotopic transplantation model of PDAC. We have identified approximately 3,000 proteins expressed in PDAC tumor cells, many of which are predominantly, if not exclusively, expressed in the tumor cells. Thus, we have established a robust technical platform for *in vivo* identification of the proteome of the tumor cells embedded within the bulk tumor, with broad applications in the studies of tumorigenesis, cancer therapeutics, and cancer detection.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Construction and validation of PDAC-BONCAT cells

The mutant murine methionyl-tRNA synthetase, $MetRS^{L274G}$, preferentially charges noncanonical amino acid azidonorleucine (ANL) to the elongator $tRNA^{Met}$, which is further incorporated into newly synthesized peptides. ANL is utilized by $MetRS^{L274G}$ and not by the wild-type translational machinery[23]. In cells expressing $MetRS^{L274G}$, the polypeptide incorporation of ANL containing the reactive azide moiety enables selective conjugation to dyes and functionalized beads for visualization and enrichment.

To label the proteome of PDAC tumor cells, $MetRS^{L274G}$ mutant transgene was cloned into a lentiviral vector and delivered to a murine pancreatic cancer cell line via lentiviral infection. Single cell clones were derived and the expression of FLAG-tagged $MetRS^{L274G}$ was confirmed by Western blot analysis (**Fig 2A**). ANL incorporation by $MetRS^{L274G}$ into the tumor cell proteome was visualized using the azide-reactive red-fluorescent tetramethyl rhodamine dibenzocyclooctyne (TAMRA-DBCO) alkyne probe. First, metabolic labeling was achieved by growing the tumor cells expressing $MetRS^{L274G}$ in media containing ANL (2μM) or control media for five hours. Cell lysates were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Next, to perform in-gel fluorescence, TAMRA was reacted to the ANL azide moiety in the proteins via a copper-free click reaction. Direct in-gel fluorescence detected the newly synthesized proteins labeled by ANL incorporation (**Fig 2B**). The same SDS-PAGE gel was stained with Coomassie blue to visualize the total protein load and size distribution. Highly specific signals were consistently detected in cells labeled with ANL, but not methionine (Met) (**Fig 2B-C**). Notably, ANL

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

incorporation was evenly distributed across the proteome as judged by the similarity of the band patterns between TAMRA and Coomassie blue staining of the ANL-labeled samples (**Fig 2B-C**). Thus, with the cell-specific expression of the $MetRS^{L274G}$ transgene, the PDAC-BONCAT system allows for effective and unbiased incorporation of ANL into the tumor cell proteome, facilitating subsequent enrichment and identification via mass-spectrometry.

In vivo validation of PDAC-BONCAT system

To examine whether the PDAC-BONCAT system allows for *in vivo* tumor cell-specific proteome labeling, we set up an orthotropic transplantation model. PDAC tumor cells expressing $MetRS^{L274G}$ transgene were surgically implanted in the pancreata of immunodeficient *NOD-scid IL2R*γ^{null} (NSG) mice. Tumor-bearing mice were metabolically labeled with either ANL (0.2mmol/kg,) or treated with normal saline (Met) via daily intraperitoneal injection for 10 days. Labeled tumor samples were collected for hematoxylin and eosin (H&E), and α-smooth muscle actin (α-SMA) immunohistochemistry (IHC) staining (**Fig 2D**). Additionally, *in situ* detection of ANL-incorporated proteins was performed using copper-catalyzed azide-alkyne cycloaddition (CuAAC).

H&E staining of the pancreata revealed a highly heterogeneous tumor invading the adjacent acinar tissues (**Fig 2E**). IHC analysis with α-SMA antibody identified abundant stromal fibroblasts in the tumor bulk (**Supplementary Fig 1**). Collectively, these features confirm the establishment of a murine PDAC model, capable of recapitulating the heterogeneous cellular composition and histological features of human pancreatic cancer. Further *in situ* detection of ANL incorporation via CuAAC click reactions showed highly specific signals in tumor cells

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

but not in the adjacent normal cells and tissues (**Fig 2F**), confirming that the tumor cells but not the non-tumor cells in the tumor bulk can incorporate ANL into their proteome. Contrary to the ANL labeling, no signal was detected in pancreatic tumors isolated from the control animals treated with Met, where ANL labeling had not taken place (**Fig 2F**). These *in vivo* observations demonstrate that BONCAT effectively tags the PDAC tumor cell proteome within its physiological milieu, and that the proteome labeling is highly specific to the tumor cells, distinguishing them from the various non-tumor cells in the TME.

Defining the in vivo tumor proteome through coupling BONCAT and DIA-MS

For *in vivo* identification of tumor cell-specific proteins, PDAC-MetRS^{L274G} cells were surgically implanted in the pancreata of a large cohort of NSG mice. Following the establishment of tumor growth in the engrafted animals, metabolic labeling of the proteome was carried out via intraperitoneal injection of ANL. Throughout the experiment, animals were provided with regular diet with no methionine depletion. Labeled tumors were collected, lysed, and subjected to BONCAT purification and DIA-MS proteomic analysis.

TAMRA-alkyne cycloaddition reaction detected ANL incorporation in the tumor bulk lysates collected from the ANL, and not the Met group (**Fig 3A-B**). Of note, the intensities of the bands in the Coomassie blue and TAMRA staining of the ANL-labeled samples did not match (lane 5-11), pointing to a minimal correlation between protein abundance and size distribution in the tumor bulk (represented by Coomassie blue staining) and the ANL-labeled tumor cells (represented by TAMRA staining). These data suggest that the proteome of bulk tumors,

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

typically identified in preclinical and/or clinical analysis of tumor samples, can not accurately represent the tumor cell proteome.

Four tumors from each group were randomly chosen for BONCAT enrichment and downstream DIA-MS proteomic analysis. Purification and enrichment of ANL-incorporated peptides in tumor lysates were achieved using a DBCO click chemistry reaction. Enriched proteins were subjected to DIA-MS analysis. In addition to the BONCAT enriched samples (BONCAT-ANL and BONCAT-Met), DIA-MS analysis was performed on the tumor bulk input lysates prior to BONCAT enrichment (Bulk-ANL and Bulk-Met). The proteomes of the four BONCAT-ANL samples were highly correlated among each other (r=0.94-0.97), while the BONCAT-ANL and BONCAT-Met samples showed only modest correlation (r=0.74-0.83) (**Supplementary Fig 2A**). Principal component analysis (PCA) categorized the samples into two distinct groups; BONCAT-ANL and BONCAT-Met samples, pointing to the specificity and efficacy of ANL labelling and BONCAT enrichment (**Fig 3C**). Interestingly, proteomic data among all bulk tumor samples (4 Bulk-ANL and 4 Bulk-Met) were highly correlated (r=0.98-0.99) (**Supplementary Fig 2B**), and PCA did not differentiate Bulk-ANL from Bulk-Met samples (**Fig 3D**), suggesting that the ANL labeling has no detectable influence on the bulk tumor proteome.

Comparison of DIA-MS results from the BONCAT-ANL and BONCAT-Met samples confirmed that the majority of the proteins are detected in the BONCAT-ANL samples, with some non-specific backgrounds present in the BONCAT-Met samples (**Fig 4A**). Among the highly enriched candidates, many proteins critical for pancreatic tumorigenesis, including

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

KRAS, YAP1, HMGB1, HMGB2, and LEG3 (Galectin-3) were identified (**Fig 4A**). There were 4360 proteins identified in the BONCAT-ANL samples (**Supplementary Table 1**). Only proteins enriched by at least four-fold in BONCAT-ANL compared to BONCAT-Met samples at a Student's t test significance level of $p<0.05$, were considered as true proteins expressed in tumor cells. Subtracting the non-specific background identified in the BONCAT-Met samples, a total of 3382 BONCAT-ANL specific proteins were identified (**Fig 4C**). These proteins together represent the tumor cell proteome of the murine pancreatic cancer. Pathway analysis of the identified tumor cell proteome revealed general as well as pancreatic cancer-specific pathways, such as RAS, Hippo, and MAPK signaling (**Supplementary Table 2**).

DIA-MS proteomic analysis of the bulk tumor, comprising tumor cells, non-tumor cell types, and the extracellular matrix components, identified >5800 proteins from ANL and Met groups with negligible differences between the two (**Fig 4B, Supplementary Table 3**), confirming that the ANL labeling process does not interfere with the general protein sysnthesis machinery in PDAC tumors. Notably, the protein levels of the candidates critical for pancreatic tumorigenesis, such as KRAS, YAP1, HMGB1, HMGB2, and LEG3 (Galectin-3), were not different between Bulk-ANL and Bulk-Met samples, further suggesting that the tumors from ANL and Met groups are biologically identical (**Fig 4B**). Thus, coupling BONCAT and DIA-MS allows for *in vivo* dissection of the PDAC tumor cell proteome in a physiological context.

Delineating proteins expressed in tumor cells from non-tumor cells within the PDAC tumor microenvironment

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

BONCAT-enriched proteins represent the tumor cell proteome, while the proteome of the pre-enrichment tumor bulk (Bulk-ANL and Bulk-Met) encompasses the entire tumor and non-tumor proteins within the TME. Notably, many proteins expressed by the tumor cells, such as housekeeping proteins, may also be abundantly produced by other cell types within the tumor bulk. Identifying the proteins preferentially expressed in either tumor cells or non-tumor cells within the TME is critical for the study of tumor cell intrinsic carcinogenesis, dynamic interaction between tumor cells and their environment, and the discovery of novel therapeutic targets and biomarkers.

To identify tumor cell-specific and TME-specific proteins, we compared the BONCAT-enriched tumor cell proteome to the pre-enrichment tumor bulk proteome within the ANL labeled group (BONCAT-ANL *vs* Bulk-ANL). For each protein, the ratio of abundance in BONCAT-enriched to tumor bulk indicates the preferential distribution in tumor or non-tumor cells. A high BONCAT-ANL/Bulk-ANL ratio points to tumor cell-specific expression, while a low ratio implies preferential expression in various non-tumor cells within the TME. Notably, the PDAC driver oncogene KRAS (RASK) was detected as a tumor cell-specific candidate with an average enrichment ratio of 10.9. Another tumor cell-specific protein, KI67, was also highly enriched with a ratio of 21.26. These data provide evidence to support that our analysis indeed identifies tumor cell specific protein expression (**Fig 4D**). To further validate our analysis, we examined signature proteins expressed exclusively in non-tumor cells within the TME. The PDAC TME of the NSG host mice contains various cell types, such as stromal fibroblasts, monocytes/macrophages, dendritic cells, neutrophils, and endothelial cells. We, therefore, searched for the signature proteins of these non-tumor cell types in our BONCAT-ANL and

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bulk-ANL data sets.[24,25] Among 34 signature proteins present in the bulk tumor lysates, the majority of them were either totally absent or highly depleted in the BONCAT-ANL tumor cell proteome. In contrast, housekeeping proteins, such as GAPDH and beta-tubulin, were not different between the BONCAT-enriched tumor cell proteome and the bulk proteome (**Supplementary Table 4**). These observations further argues that our BONCAT-DIA-MS approach delineates proteins expressed in the tumor cells from non-tumor cells in the TME, compartmentalizing tumor and non-tumor proteins within the tumor bulk.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Coupling BONCAT bioorthogonal chemistry with DIA-MS proteomics in an orthotopic pancreatic cancer model, we have developed an innovative technical framework that can specifically label, enrich, and identify the tumor cell proteome *in vivo*. The sensitivity and efficiency of this approach is validated through the identification of thousands of proteins expressed in pancreatic tumor cells within the tumor bulk. Comparative analysis of the BONCAT-enriched and pre-enrichment bulk tumor lysates facilitates the differentiation of proteins preferentially expressed in tumor cells from those of non-tumor cells within the TME.

Our approach has broad applications in the studies of tumorigenesis, cancer therapeutics, and biomarker discovery. Our platform may be applied to primary tumors isolated from human patients to systematically define their tumor cell proteomes. Patient-derived xenograft (PDX) models are increasingly utilized to investigate novel therapeutics and guide clinical cancer treatment.[26-28] Following well-established protocols, primary PDX tumors can express the $MetRS^{L274G}$ enzyme via lentiviral infection to enable tumor cell-specific proteomic labeling and characterization.[29,30] Our approach, therefore, allows *in vivo* tumor cell-specific proteomic characterization in PDX models, providing an unprecedented ability for systemic interrogation of therapeutic responses at the level of individual proteins. Additionally, this technical framework may be implemented to reveal the tumor cell-specific secretome. ANL labeling of tumor cell proteome in PDX models, allows selective purification and enrichment of various proteins secreted by tumor cells into the systemic circulation via BONCAT. Subsequent identification of the tumor cell-specific secretome using DIA-MS will open new opportunities for the development of novel biomarkers for early cancer detection, a particularly persistent challenge in pancreatic cancers.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Materials and Methods

Cell lines, constructs, and chemical reagents

4292 murine PDAC cell line was a generous gift from Dr. Marina Pasca di Magliano[31]. KRAS expression in this cell line is controlled by the Tet-ON system. $MetRS^{L274G}$ coding sequence was PCR amplified using pMarsL274G construct (Addgene 63177) as a template and inserted into BamHI and MluI sites within the pLV-EF1a-IRES-puro vector (Addgene 85132) to produce the Lentiviral-$MetRS^{L274G}$ vector. The FLAG M2 antibody used for the detection of $MetRS^{L274G}$ protein was purchase from Sigma. ANL, H-L-Lys(N3)-OH*HCL (HAA1625), was purchased from Peptide Solutions (Tuscon, AZ). DBCO agarose beads (1034), and DBCO-TAMRA (A131), were purchased from Click Chemistry Tools. Tris [(1-benzyl-1H-1, 2, 3-triazol-4-yl)methyl] amine was purchased from Fisher Scientific.

Click Chemistry Reactions

TAMRA reaction: To examine labeling efficiency 20 µg protein lysate was incubated with 5-30 µM DBCO-TAMRA (Absorbance/Emission of 548/562 nm)/PBS (pH 7.4) for 1 hour at room temperature. Samples were boiled in Laemmli buffer and run on SDS-PAGE gel. Electrophoresed samples were visualized using Biorad gel imaging system with Pro-Q Diamond filters. The gel was subsequently stained with Imperial stain (Coomassie dye R-250, ThermoFisher 24615) according to manufacturer's recommendation.

CuAAC: Copper-assisted click reaction was performed on paraffin-embedded slides. Slides were deparaffinized in 2 changes of xylene, 5 minutes per change, and rehydrated sequentially in 2 changes, 5 minutes each of 100% ethanol and 95% ethanol, and 5 minutes 70% ethanol, and changed into water. To quench endogenous peroxidase enzyme slides were immersed in 3%

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

H_2O_2 for 15 minutes at room temperature. Slides were washed 3 times with PBS/0.1% triton X-100. Copper-assisted reaction was essentially performed as described[32]. Briefly, orthogonal tagging reaction was assembled in dark. 5 µl of 200 mM TBTA, 5 µl of 500 mM TCEP, 5 µl of 2 mM biotin-alkyne-tag (Click Chemistry Tools, 1266), and 5 µl of 200 mM CuSO4, were added at the specified order to 5 ml of PBS (pH 7.8), and the mixture was vortexed for 10 seconds after each addition. The slides were reacted with the mixture overnight at room temperature. Slides were subsequently washed three times, 20 minutes each, in PBS (pH 7.8), 0.5 mM EDTA, 1% Tween 20, followed by two washes, 10 minutes each, of PBS (pH7.8), 0.1% Tween 20. Slides were finally washed twice with PBS (pH 7.4). For signal amplification and HRP conjugation samples were incubated with VECTASTAIN Elite ABC reagent for 30 minutes, washed for 15 minutes in PBS (pH 7.4), and changed into water. Signals were detected using ImmPACT DAB Peroxidase (HRP) Substrate (Vector laboratories, SK-4105).

IHC: Slides were incubated in three washes of xylene, 100% ethanol, and 95% ethanol for 5 minutes each. Sections were then washed in water twice, 5 minutes per wash. Antigens were unmasked by boiling the slides for 15 minutes in antigen unmasking citrate buffer (Cell Signaling Technology 14746), and cooled at room temperature. Staining was performed with VECTASTAIN ABC Elite kit according to the accompanying protocol and detected as mentioned. α-SMA antibody (19245) was purchased from Cell Signaling Technology.

Animal models, orthotopic transplantation, and ANL labeling

Animal studies and experimental protocols were approved by Institutional Animal Care and Use Committee at Houston Methodist Research Institute. All experimental methods were performed in accordance with the relevant national and institutional guidelines and regulations. 6-8 week

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

old *NOD-scid IL2Rγ*^{null} (NSG) mice underwent surgical orthotopic injection of the pancreatic cancer cells into the pancreas. Carprofen medicated gel (5 mg/kg/day) was used for analgesia prior to the surgery, and within 3 days following surgical procedure. Mice were anesthetized with isoflurane. The abdominal skin directly above the spleen was incised, pancreas was retracted laterally and positioned outside the body. Direct injection of 1 million cells was performed using a 28.5 G needle. The needle was inserted through the knot into the pancreas tail and passed into the pancreas head to deliver the cells. Following cell injection the spleen and pancreas were returned to the peritoneal cavity and the abdominal muscle and the skin layers were sequentially sutured. One day following surgery doxycycline was administered through the drinking water, at a concentration of 0.2g/L in a solution of 5% sucrose, and replaced every 3-4 days. On day 4 post-surgery, experimental and control mice were intraperitoneally injected with 0.1mg/g per day of the amino acid analog or PBS respectively, for one week[33].

BONCAT enrichment

Tumor nodules were harvested and snap frozen in liquid nitrogen until further use. Frozen tumor samples were homogenized for 20-60 seconds in PBS (pH 7.4), 1% SDS, 100 mM chloroacetamide, and protease inhibitors. The homogenate was left at room temperature for 20-30 minutes to allow protein solubilization. Lysates were boiled for 10 minutes, and centrifuged at room temperature at 16,000 g for 10 minutes. The supernatant was separated and aliquoted. Protein concentration was determined using BCA protein assay. The supernatants were used for the identification of tumor bulk (Bulk-ANL and Bulk Met) and tumor cell proteomes following BONCAT enrichment (BONCAT-ANL and BONCAT-Met). Approximately 1.5 mg protein was diluted with 8 M urea/0.15 M NaCl/PBS (pH 7.4) to a total volume of 1 ml. 50 µl of DBCO-

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

agarose bead slurry was washed 3 times with 0.8% SDS in PBS (pH 7.4). Diluted protein sample was added to the washed resin, and shaken at 12,000 rpm at room temperature for more than 12 hours. Unreacted DBCO was quenched by the addition of 2 mM ANL for 30 minutes. Resins were washed with 1 ml water and reduced with 1mM DTT for 15 minutes at 70°C on a shaking platform. Free thiols were subsequently blocked with 40 mM iodoacetamide in PBS/ 0.8 SDS for 30 minutes in the dark at 12,000 rpm. Resins were then subjected to the following washes: 40 ml 0.8% SDS in PBS, 40 ml 8 M urea, and 40 ml 20% acetonitrile. Beads were then washed with 10% acetonitrile in 50 mM ammonium bicarbonate. The beads were spun at 2,000 g to remove the liquid, and resuspended in 100 µl 10% acetonitrile in 50 mM ammonium bicarbonate and 100 ng trypsin (Thermo Scientific Pierce, 90057). Beads were digested at 37 °C on a shaking platform overnight, and subsequently removed using centrifuge columns, and digested peptides were dried at room temperature using speed vacuum. Digested peptides were subsequently subjected to DIA-MS analysis.

Lysis and digestion of tumor bulk

Tumor bulk cells were lysed in a buffer containing 5% SDS/50 mM triethylammonium bicarbonate (TEAB) in the presence of protease and phosphatase inhibitors (Halt; Thermo Scientific) and nuclease (Pierce™ Universal Nuclease for Cell Lysis; Thermo Scientific). Aliquots corresponding to 100 µg protein (EZQ™ Protein Quantitation Kit; Thermo Scientific) were reduced with tris (2-carboxyethyl) phosphine hydrochloride (TCEP), alkylated in the dark with iodoacetamide and applied to S-Traps (mini; Protifi) for tryptic digestion (sequencing grade; Promega) in 50 mM TEAB. Peptides were eluted from the S-Traps with 0.2% formic acid

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in 50% aqueous acetonitrile, quantified using Pierce™ Quantitative Fluorometric Peptide Assay (Thermo Scientific) and diluted as needed to achieve a concentration of 0.4 µl/µl.

DIA-MS proteomic analyses

Experimental samples were randomized for sample preparation and analysis. DIA-MS analyses were conducted on an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific). On-line HPLC separation was accomplished with an RSLC NANO HPLC system (Thermo Scientific/Dionex): column, PicoFrit™ (New Objective; 75 µm i.d.) packed to 15 cm with C18 adsorbent (Vydac; 218MS 5 µm, 300 Å); mobile phase A, 0.5% acetic acid (HAc)/0.005% trifluoroacetic acid (TFA) in water; mobile phase B, 90% acetonitrile/0.5% HAc/0.005% TFA/9.5% water; gradient 3 to 42% B in 120 min; flow rate, 0.4 µl/min. Separate pools were made of all of the samples in each experiment (equal volumes from the BONCAT-ANL/BONCAT-Met digests; equal quantities for the tumor bulk lysate digests). For the tumor bulk lysates, injections of 2µg peptides of the pooled samples were used for chromatogram library generation. For the BONCAT-ANL and BONCAT-Met samples, aliquots of the pool of equal volumes of the digests were injected. To create the DIA chromatogram library[34] for each sample type, the indicated peptide quantities were analyzed using gas-phase fractionation and 4-m/z windows (staggered; 30k resolution for precursor and product ion scans, all in the orbitrap) and the MS files processed in Scaffold DIA (v2.1.0; Proteome Software) and searched against a predicted spectral library generated from the UniProt_mouse (2019_01) protein database by Prosit.[35] Injections of 2 µg of peptides were employed for DIA-MS analysis of the individual bulk tumor lysate digests while injections corresponding to equal volumes were used for the BONCAT-ANL and BONCAT-Met samples. MS data for all individual digests were acquired in

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the orbitrap using 12-m/z windows (staggered; 30k resolution for precursor and product ion scans) and searched against the chromatogram library. Scaffold DIA (v2.1.0; Proteome Software) was used for processing the DIA data from the experimental samples. Only peptides that were exclusively assigned to a protein were used for relative quantification, with 2 minimum peptides required for each protein and a protein-level FDR of 1%.

Correlations among different BONCAT-enriched samples and total lysate samples were analyzed by Pearson correlation. Differentially abundant proteins were analyzed by a two-sided Student's t-test. For BONCAT-ANL and BONCAT-Met comparison, proteins with a missing value in any of the four BONCAT-ANL samples were excluded from the analysis. For the analysis of proteins from tumor cells versus non-tumor cells, BONCAT-ANL and Bulk-ANL data were first normalized by the total intensity in each sample and then compared by a two-sided t-test. Perseus (version 1.6.7.0)[36] was used to generate the volcano plots for the three paired comparisons displayed in figure 4 (FDR = 0.01, S0 = 0.1). Proteins with missing values in any of the samples were excluded from the plots. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were analyzed using DAVID[37].

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1. Construction of the PDAC-BONCAT cells

(A-B) Diagram showing the cell selective labeling of proteome by BONCAT.

(C) Diagram showing selective coupling of ANL labeled peptides using DBCO-alkyne beads.

Figure 2. *in vitro* and *in vivo* validation of the PDAC-BONCAT system

(A) Western blot analysis of single cell clones ectopically expressing the $MetRS^{L274G}$ transgene. The $MetRS^{L274G}$ protein is indicated by the arrow. n.s. indicates the non-specific bands recognized by the FLAG antibody.

(B-C) Detection of ANL and Met labeling by TAMRA-DBCO and SDS-PAGE *in vitro*. 10, 20, and 30, indicate micrograms of protein lysate loaded per lane.

(D) Overview of the *in vivo* ANL labeling.

(E) H&E staining showing the tumor infiltration in the adjacent normal acinar tissues. Tumor nodule is indicated by a red asterisk.

(F) IHC analysis of ANL incorporation by CuAAC. Met samples serve as the negative control. Tumor nodules are indicated by red asterisks.

Figure 3. Quality analysis of BONCAT-enriched samples for DIA-MS

(A-B) TAMRA-alkyne and SDS-PAGE analysis of eleven *in vivo* tumor samples labeled with ANL or Met.

(C) PCA of DIA-MS data for BONCAT-ANL and BONCAT-Met samples

(D) PCA of DIA-MS data for Bulk-ANL and Bulk-Met samples

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4. Tumor cell and tumor bulk proteomes

(A) Comparison of the BONCAT-ANL and BONCAT-Met samples (FDR = 0.01, S0 = 0.1). Some genes critical for pancreatic tumorigenesis were labeled.

(B) Comparison of the bulk proteomes from ANL and Met total tumor lysates.

(C) Venn diagram of the number of proteins identified by BONCAT enrichment in the ANL labeled samples.

(D) Comparison of BONCAT-ANL proteome and the corresponding Bulk-ANL proteome differentiates tumor cell specific proteins from those of non-tumor cells within the TME.

Supplementary Figure Legends

Supplementary Fig S1. α-SMA antibody identifies stromal fibroblasts in the tumor bulk.

Supplementary Fig S2. Pearson correlation of DIA-MS data from the BONCAT (S2A), and the Bulk samples (S2B). The correlation coefficients are labeled above the dot plots.

Supplementary Tables

Supplementary Table 1. List of tumor cell proteins identified in BONCAT samples by DIA-MS.

Supplementary Table 2. Pathway analysis of the PDAC tumor cell proteome.

Supplementary Table 3. List of proteins identified in the tumor bulk lysates by DIA-MS.

Supplementary Table 4. Protein levels of marker genes from tumor and non-tumor cells within the tumor microenvironment.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author Contributions

N.G.A., S.T.W. and Y.L. designed the study and developed the approach. N.G.A. carried out all cell culture, animal studies, and click chemistry experiments. S.P. and D.M. assisted with the DIA-MS analysis. L.N., N.G.A, D.K.S, S.T.W, and Y.L. analyzed the data. J.C. advised the study. N.G.A and Y.L. wrote and S.T.W. edited the manuscript. All authors reviewed and approved the manuscript.

Acknowledgments

We greatly appreciate Dr. Marina Pasca di Magliano, for generously providing us the murine pancreatic cancer cell line (4292). We would also like to thank members of David Tirrell lab and Erin Schuman lab for sharing protocols on BONCAT procedures. Mass spectrometry analyses were conducted in the Mass Spectrometry Laboratory at the University of Texas Health Science Center at San Antonio. This work was supported in part NIH K22CA207598 (Y.L.) and NIH GM008042 (D.K.S.; a grant from the UCLA-Caltech Medical Scientist Training Program). Support from the University of Texas System Proteomics Core Network for purchase of the Lumos mass spectrometer is gratefully acknowledged.

Competing Interests

The authors declare no financial interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reference

1 Ohlund, D. *et al.* Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med* **214**, 579-596, doi:10.1084/jem.20162024 (2017).

2 Lafaro, K. J. & Melstrom, L. G. The Paradoxical Web of Pancreatic Cancer Tumor Microenvironment. *Am J Pathol* **189**, 44-57, doi:10.1016/j.ajpath.2018.09.009 (2019).

3 Pothula, S. P. *et al.* Key role of pancreatic stellate cells in pancreatic cancer. *Cancer Lett* **381**, 194-200, doi:10.1016/j.canlet.2015.10.035 (2016).

4 Stone, S. E., Glenn, W. S., Hamblin, G. D. & Tirrell, D. A. Cell-selective proteomics for biological discovery. *Curr Opin Chem Biol* **36**, 50-57, doi:10.1016/j.cbpa.2016.12.026 (2017).

5 Gauthier, N. P. *et al.* Cell-selective labeling using amino acid precursors for proteomic studies of multicellular environments. *Nat Methods* **10**, 768-773, doi:10.1038/nmeth.2529 (2013).

6 Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci U S A* **103**, 9482-9487, doi:10.1073/pnas.0601637103 (2006).

7 Elliott, T. S., Bianco, A. & Chin, J. W. Genetic code expansion and bioorthogonal labelling enables cell specific proteomics in an animal. *Curr Opin Chem Biol* **21**, 154-160, doi:10.1016/j.cbpa.2014.07.001 (2014).

8 Erdmann, I. *et al.* Cell-selective labelling of proteomes in Drosophila melanogaster. *Nat Commun* **6**, 7521, doi:10.1038/ncomms8521 (2015).

9 Liu, Y. *et al.* Application of bio-orthogonal proteome labeling to cell transplantation and heterochronic parabiosis. *Nat Commun* **8**, 643, doi:10.1038/s41467-017-00698-y (2017).

10 Schanzenbacher, C. T., Langer, J. D. & Schuman, E. M. Time- and polarity-dependent proteomic changes associated with homeostatic scaling at central synapses. *Elife* **7**, doi:10.7554/eLife.33322 (2018).

11 Alvarez-Castelao, B. *et al.* Cell-type-specific metabolic labeling of nascent proteomes in vivo. *Nat Biotechnol* **35**, 1196-1201, doi:10.1038/nbt.4016 (2017).

12 Prescher, J. A., Dube, D. H. & Bertozzi, C. R. Chemical remodelling of cell surfaces in living animals. *Nature* **430**, 873-877, doi:10.1038/nature02791 (2004).

13 Baskin, J. M. *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proc Natl Acad Sci U S A* **104**, 16793-16797, doi:10.1073/pnas.0707090104 (2007).

14 Mahdavi, A. *et al.* Identification of secreted bacterial proteins by noncanonical amino acid tagging. *Proc Natl Acad Sci U S A* **111**, 433-438, doi:10.1073/pnas.1301740111 (2014).

15 Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods* **1**, 39-45, doi:10.1038/nmeth705 (2004).

16 Bruderer, R. *et al.* Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol Cell Proteomics* **16**, 2296-2309, doi:10.1074/mcp.RA117.000314 (2017).

17 Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **11**, O111 016717, doi:10.1074/mcp.O111.016717 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18 Michalski, A., Cox, J. & Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res* **10**, 1785-1793, doi:10.1021/pr101060v (2011).

19 Wilson, R. S. *et al.* Development of Targeted Mass Spectrometry-Based Approaches for Quantitation of Proteins Enriched in the Postsynaptic Density (PSD). *Proteomes* **7**, doi:10.3390/proteomes7020012 (2019).

20 Nguyen, E. V. *et al.* Identification of Novel Response and Predictive Biomarkers to Hsp90 Inhibitors Through Proteomic Profiling of Patient-derived Prostate Tumor Explants. *Mol Cell Proteomics* **17**, 1470-1486, doi:10.1074/mcp.RA118.000633 (2018).

21 Keam, S. P. *et al.* Exploring the oncoproteomic response of human prostate cancer to therapeutic radiation using data-independent acquisition (DIA) mass spectrometry. *Prostate* **78**, 563-575, doi:10.1002/pros.23500 (2018).

22 Kim, Y. J. *et al.* Data-Independent Acquisition Mass Spectrometry To Quantify Protein Levels in FFPE Tumor Biopsies for Molecular Diagnostics. *J Proteome Res* **18**, 426-435, doi:10.1021/acs.jproteome.8b00699 (2019).

23 Mahdavi, A. *et al.* Engineered Aminoacyl-tRNA Synthetase for Cell-Selective Analysis of Mammalian Protein Synthesis. *J Am Chem Soc* **138**, 4278-4281, doi:10.1021/jacs.5b08980 (2016).

24 Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-457, doi:10.1038/nmeth.3337 (2015).

25 Ren, B. *et al.* Tumor microenvironment participates in metastasis of pancreatic cancer. *Molecular cancer* **17**, 108, doi:10.1186/s12943-018-0858-1 (2018).

26 Crystal, A. S. *et al.* Patient-derived models of acquired resistance can identify effective drug combinations for cancer. *Science* **346**, 1480-1486, doi:10.1126/science.1254721 (2014).

27 Bertotti, A. *et al.* A molecularly annotated platform of patient-derived xenografts ("xenopatients") identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer discovery* **1**, 508-523, doi:10.1158/2159-8290.CD-11-0109 (2011).

28 Zhang, X. *et al.* A renewable tissue resource of phenotypically stable, biologically and ethnically diverse, patient-derived human breast cancer xenograft models. *Cancer research* **73**, 4885-4897, doi:10.1158/0008-5472.CAN-12-4081 (2013).

29 Hanna, C., Kwok, L., Finlay-Schultz, J., Sartorius, C. A. & Cittelly, D. M. Labeling of Breast Cancer Patient-derived Xenografts with Traceable Reporters for Tumor Growth and Metastasis Studies. *J Vis Exp*, doi:10.3791/54944 (2016).

30 Liu, J. F. *et al.* Establishment of Patient-Derived Tumor Xenograft Models of Epithelial Ovarian Cancer for Preclinical Evaluation of Novel Therapeutics. *Clinical cancer research : an official journal of the American Association for Cancer Research* **23**, 1263-1273, doi:10.1158/1078-0432.CCR-16-1237 (2017).

31 Collins, M. A. *et al.* Oncogenic Kras is required for both the initiation and maintenance of pancreatic cancer in mice. *The Journal of clinical investigation* **122**, 639-653, doi:10.1172/JCI59227 (2012).

32 Tom Dieck, S. *et al.* Metabolic labeling with noncanonical amino acids and visualization by chemoselective fluorescent tagging. *Curr Protoc Cell Biol* **Chapter 7**, Unit7 11, doi:10.1002/0471143030.cb0711s56 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

33 Calve, S., Witten, A. J., Ocken, A. R. & Kinzer-Ursem, T. L. Incorporation of non-canonical amino acids into the developing murine proteome. *Sci Rep* **6**, 32377, doi:10.1038/srep32377 (2016).

34 Searle, B. C. *et al.* Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nat Commun* **9**, 5128, doi:10.1038/s41467-018-07454-w (2018).

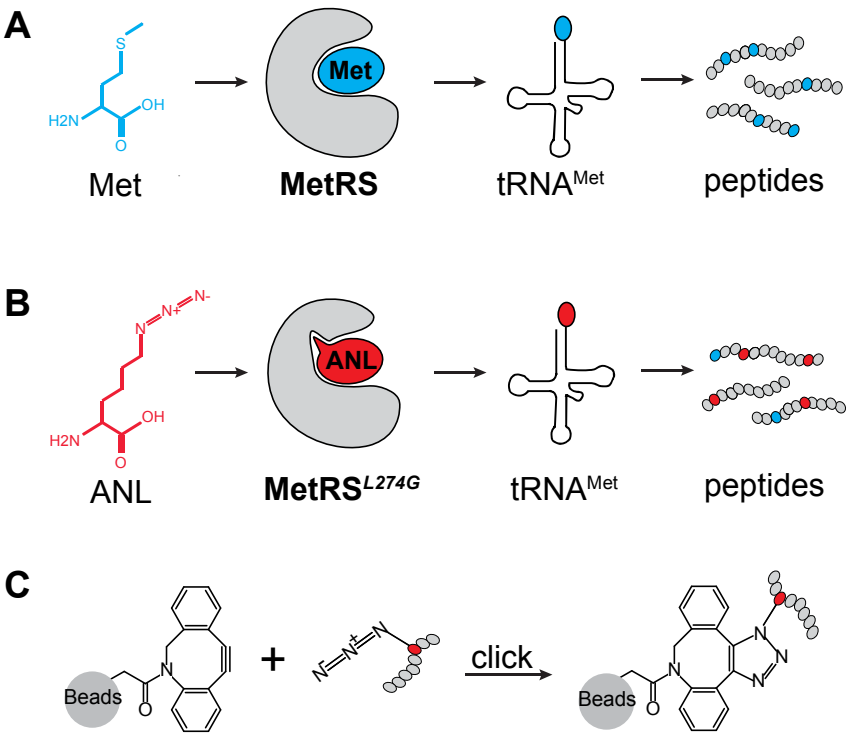
35 Gessulat, S. *et al.* Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat Methods* **16**, 509-518, doi:10.1038/s41592-019-0426-7 (2019).

36 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740, doi:10.1038/nmeth.3901 (2016).

37 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

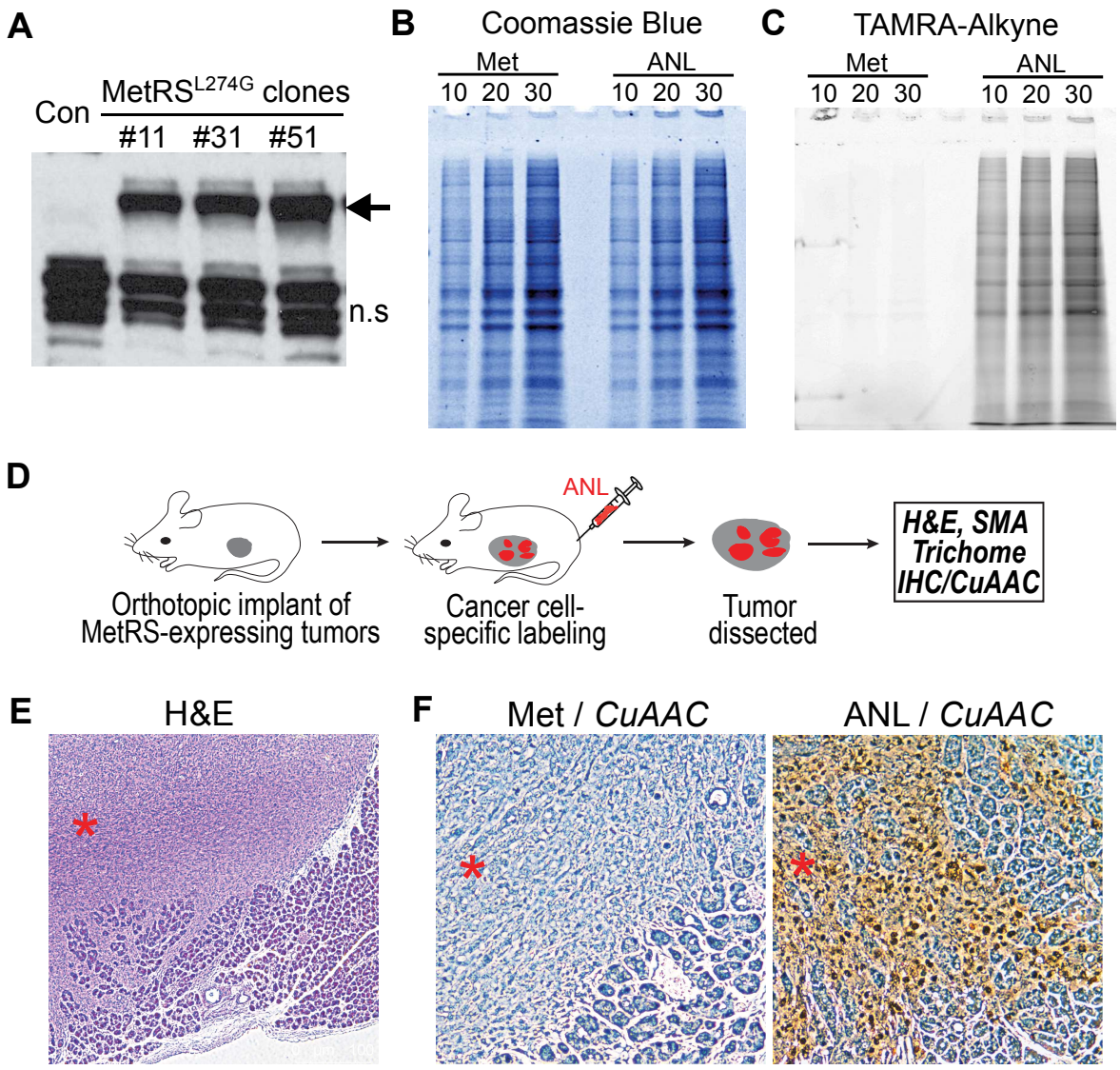
bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1



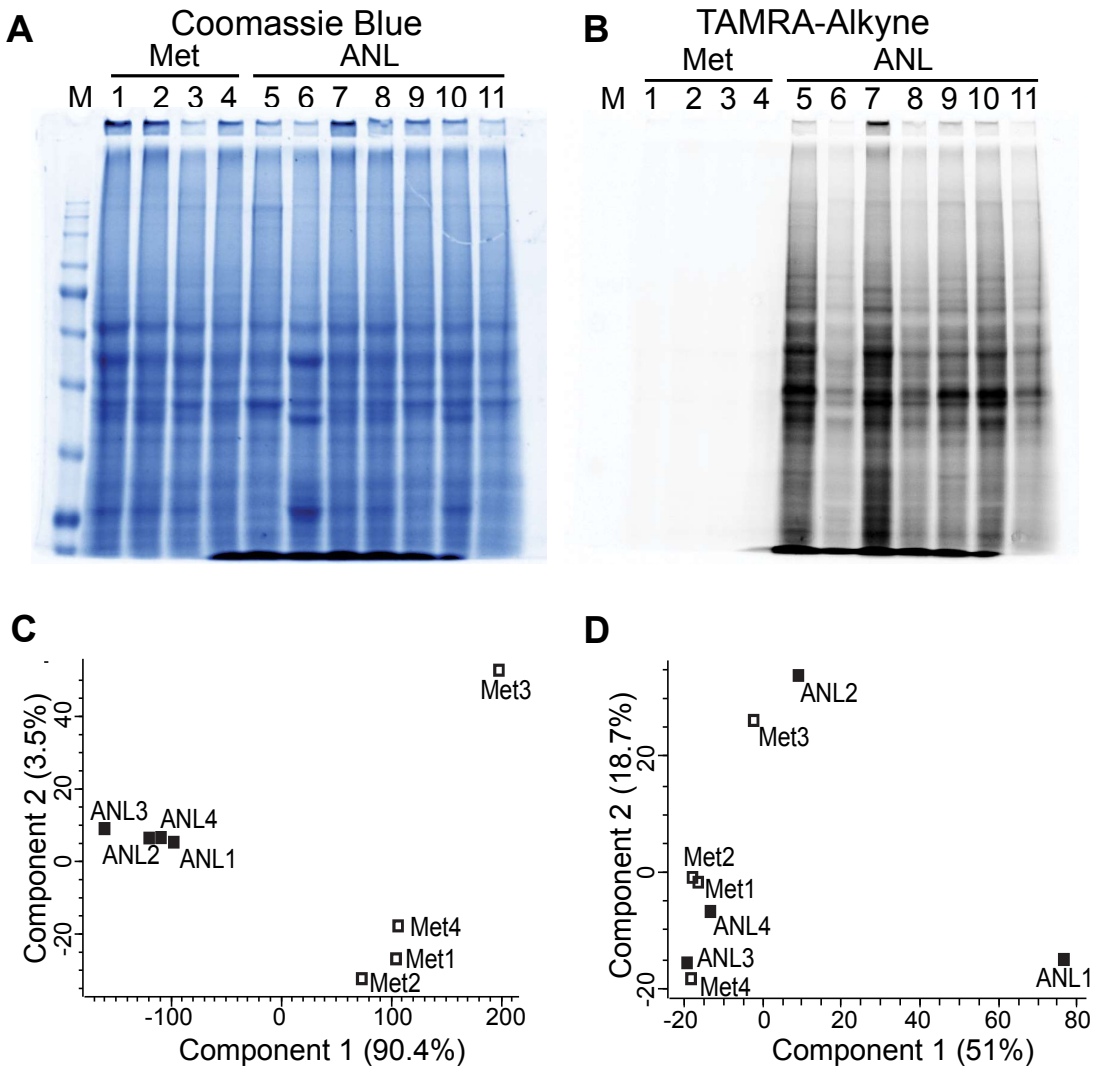
bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2



bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3



bioRxiv preprint doi: https://doi.org/10.1101/2020.05.25.113670; this version posted May 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

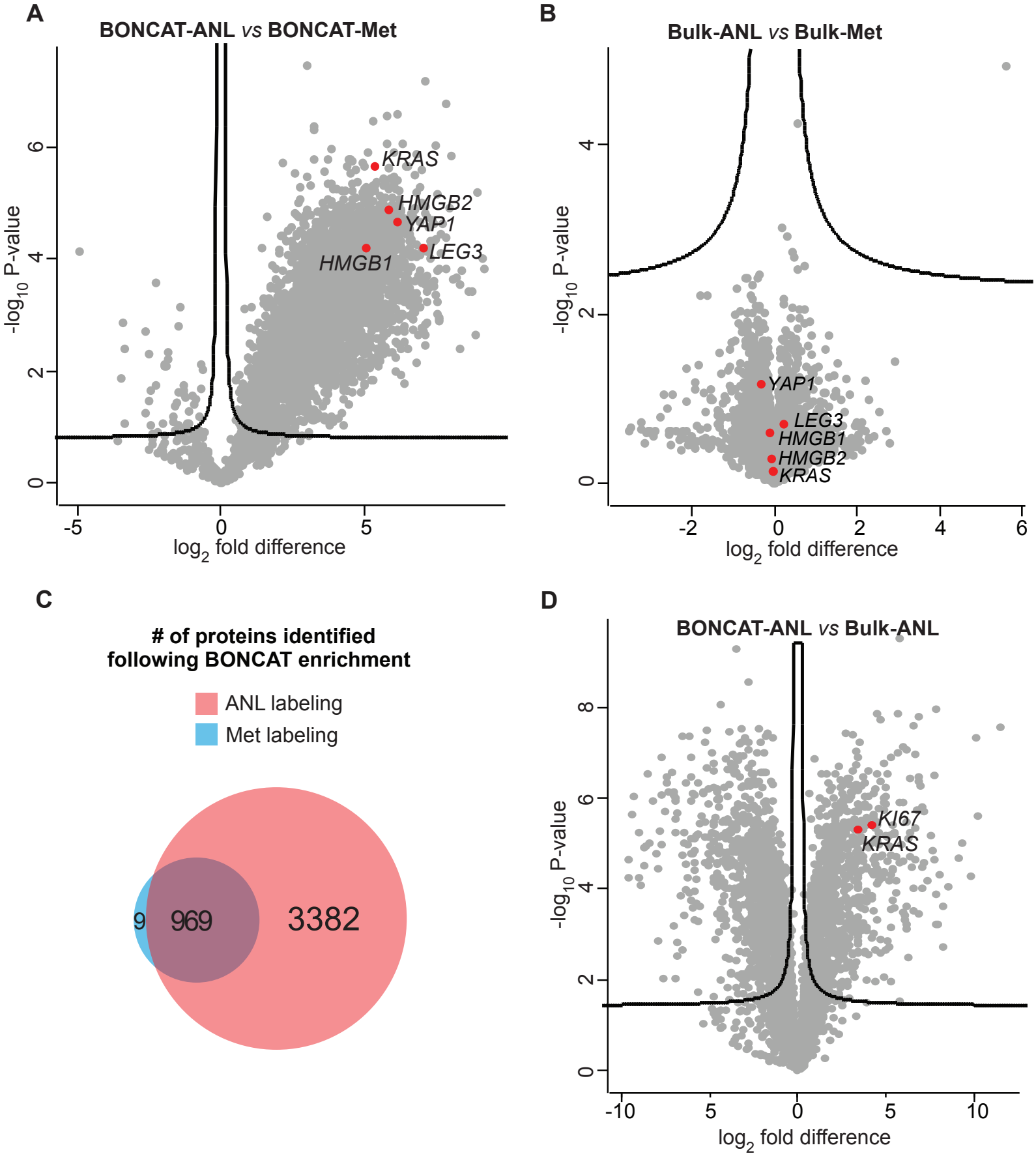



Figure 4