Dynamic *in vivo* mapping of the methylproteome using a chemoenzymatic approach

Jonathan Farhi[1,2,3], Robert B. Jones[1], Ashish K. Verma[3], Kiran K. Parwani[1,2], Kuei-Chien Tang[3], Benjamin Emenike[3], Pritha Bagchi[4], Monika Raj[3], Jennifer M. Spangle[1, #]

[1] Department of Radiation Oncology, Winship Cancer Institute of Emory University School of Medicine, Atlanta, GA

[2] Cancer Biology Program, Emory University, Atlanta, GA

[3] Department of Chemistry, Emory University, Atlanta, GA

[4] Emory Integrated Proteomics Core, Emory University, Atlanta, GA

[#] Corresponding Author, Jennifer.spangle@emory.edu

Running Title: *In vivo* methylproteome mapping using a chemoenzymatic approach

Abstract:

Dynamic protein post-translation methylation is essential for cellular function, highlighted by the essential role of methylation in transcriptional regulation and its aberrant dysregulation in diseases including cancer. This underscores the importance of cataloging the cellular methylproteome. However, comprehensive analysis of the methylproteome remains elusive due to limitations in current enrichment and analysis pipelines. Here, we employ an L-Methionine analogue, ProSeMet, that is chemoenzymatically converted to the SAM analogue ProSeAM in cells and mice to tag proteins with a biorthogonal alkyne that can be functionalized for global detection, selective enrichment, and LC-MS/MS identification. We identify 486 proteins known to be methylated and 221 proteins with novel methylation sites encompassing diverse cellular functions. Systemic ProSeMet delivery in mice pseudomethylates proteins across organ systems with blood-brain barrier penetrance. Leveraging this pipeline to define the cellular methylproteome may have broad applications for understanding the methylproteome in the context of disease.

Introduction

Protein methylation is a widely occurring post-translational modification (PTM) that plays an important role in cell signaling, development, and disease[1]. Methylation most commonly occurs on the basic amino acids lysine (Lys), arginine (Arg), and histidine (His): lysine residues can be mono-, di-, or tri-methylated, arginine residues can be monomethylated or dimethylated symmetrically or asymmetrically, and histidine is monomethylated at both the 1 and 3 positions of the imidazole ring[1,2]. On proteins, these methylation events are catalyzed by methyltransferase enzymes using the metabolite *S*-adenosyl-L-methionine (SAM)[3]. SAM-dependent methyltransferases fall into three structurally distinct classes: Class I enzymes contain a conserved seven-stranded β-sheet and include all canonical arginine methyltransferases (PRMTs) as well as the lysine methyltransferase DOT1L; Class II enzymes contain a SET domain and encompass all other lysine methyltransferases (KMTs), and Class III enzymes are membrane-associated methyltransferases[4].

Histone protein methylation affects epigenetic regulation of gene expression by mediating the recruitment of chromatin modifying enzymes to chromatin and transcriptional regulators to gene regulatory regions including promoters and enhancers[2,5]. Non-histone protein methylation is likewise an essential PTM affecting cellular biology and disease pathogenesis[6,7]. For example, methylation of the PI3K effector AKT by SETDB1 increases the duration of AKT activation, promoting tumor growth in murine models[8,9]. Similarly, methylation of CRAF at R100 by PRMT6 interferes with RAS/RAF binding, suggesting a mechanism by which PRMT6 loss increases tumor initiation and metastasis in hepatocellular carcinoma[10]. Moreover, the oncogenic role of Enhancer of Zester Homolog 2 (EZH2) is hypothesized to extend beyond methylation of histone H3 on lysine 27 (H2K37) to include cytosolic targets: the transcription factors STAT3[11], GATA4 (K299)[12]; and FOXA1 (K295)[13], among others. The important role these methyltransferase enzymes play in normal cellular function and disease has prompted the clinical development of several inhibitors, including the recent first-in-class EZH2 inhibitor tazemetostat for the treatment of epithelioid sarcoma and follicular lymphoma[14] and several PRMT1/5 inhibitors[15,16].

Current efforts to identify histone and non-histone protein methylation often employ mass spectrometry (MS) or antibody-based approaches to identify methylated residues and changes to their methylation status[6]. However, unlike acetylation or phosphorylation, methylation does not change the charge or physiochemical properties of the amino acid residue, rendering MS-based characterization challenging without isotopic labeling, chemical modification, or biased enrichment to reduce the complexity of analysis[1,17]. The development of pan-methyl, pan-methyl-lysine, and pan-methyl-arginine specific antibodies has underscored the breadth of the methylproteome[18,19] but has been limited by cross-reactivity (e.g., between methylated Lys and Arg residues)[20], batch-to-batch irreproducibility, and an inability to identify novel methylation sites[21,22]. Furthermore, MS-based methods aimed at detecting methylation must overcome high false discovery rates because a high number of

amino acid combinations produce sequences that are isobaric to methylated peptides of a different sequence (e.g., the mass difference between Asp and Glu is equivalent to a single methylation event)[23].

The challenges presented by current MS- or antibody-based approaches have prompted the development of chemical strategies to selectively modify methylated residues[24] or hijack cellular machinery for a tag-and-modify approach[25]. In particular, SAM analogues have been employed to "pseudomethylate" (tag) proteins using engineered methyltransferases in live cells or on cell lysate[26-30]. Recently, treatment of cell lysates with the SAM analogue propargyl-*Se*–adenosyl-L-selenomethionine (ProSeAM) enabled profiling of the methylproteome and of methylhistidine targets of the histidine methyltransferase METTL9[31,32]. While these *in situ* methodologies demonstrate the ability of selected methyltransferases to methylate a substrate based on a consensus sequence, *in situ* studies cannot capture changes to the methylproteome in response to biological stimuli or reflect methyltransferase activation and localization into unique cellular compartments. Likewise, strategies involving engineered enzymes are unable to profile the broad spectrum of the methylproteome and are therefore restricted to specific applications.

These challenges have prompted the development of novel strategies to chemically label methyltransferase targets in live cells. It has been reported that the L-Met analogue propargyl-L-selenohomocysteine (ProSeMet) is chemoenzymatically converted by cellular methionine adenosyltransferase (MAT) enzymes to the SAM analogue ProSeAM *in vitro.* In turn, ProSeAM is used by native RNA methyltransferases to pseudomethylate RNA with a biorthogonal alkyne tag allowing for downstream analysis[33-35]. Previous reports have not, however, explored the ability of ProSeAM to pseudomethylate proteins using endogenous methyltransferases. As such, we envisioned that dynamic profiling of the methylproteome could be accomplished using ProSeMet through the chemoenzymatic approach (**Fig. 1**). In this work, we describe the application of ProSeMet for methylproteome profiling *in vitro* and *in vivo*. Using ProSeMet, we performed an unbiased analysis of the methylproteome in the EZH2-activated cell line G401[36] and identified 486 previously characterized methylated proteins. Our study further led to the discovery of 221 novel methylated proteins. Furthermore, we demonstrate that *in vivo* ProSeMet administration to BALB c/J mice leads to labeling of several key organ systems, including the heart, lungs, and brain. Our work demonstrates the development of an improved chemoenzymatic platform that has the potential to study dynamic changes to the methylproteome in response to biological stimulus and disease pathogenesis both in cell and murine models.

Results

ProSeMet competes with L-Met to pseudomethylate protein in the cytoplasm and nucleus

Previous literature has reported the use of ProSeMet as a viable substrate of the unaltered RNA methyltransferase METTL3-METTL14 in MOLM13, HEK293T, and HeLa cells[33,35]. We therefore hypothesized

that ProSeMet could also be used to deposit a biorthogonal alkyne tag (pseudomethylate) proteins in living cells using native methyltransferases (**Fig. 2A)**. Protein extracted from T47D cells incubated with ProSeMet subjected to copper-catalyzed azide-alkyne cycloaddition (CuAAC) to attach a fluorescent picolyl azide and separated via SDS-PAGE showed protein labeling that was detectable with as little as 5 µg input protein (**Fig. 2B**). Labeling efficacy scaled with increasing concentrations of ProSeMet until 100 µM concentrations (**Fig. 2C**). Since previously reported L-Met analogues have been shown to incorporate into proteins during translation[37], we collected lysates from cells treated with L-Met, ProSeMet, or the L-Met analogue azidohomoalanine (AHA) in the presence of the protein synthesis inhibitor cycloheximide (CHX). CHX treatment abrogated AHA labeling but did not affect ProSeMet signal, indicating that ProSeMet is not incorporated into newly synthesized proteins but rather converted to ProSeAM and used by native methyltransferases (**Fig. 2D**). Furthermore, competition with the unlabeled natural metabolite L-Met revealed a dose-dependent reduction of ProSeMet labeling with almost complete depletion of ProSeMet signal at 10 µM L-Met concentrations (**Fig. 2E**). Since protein methylation occurs in both the cytoplasm and nucleus, we next evaluated the cellular compartmentalization of ProSeMet-directed labeling in live cells. Cell fractionation of the cytosolic and nuclear compartments followed by SDS-PAGE fluorescent analysis revealed no fluorescent labeling of the L-Met control but robust ProSeMet labeling of protein across molecular weights in both cytosolic and nuclear fractions (**Fig. 2F**). We further performed immunofluorescence (IF) microscopy analysis of T47D, LNCaP, and MCF10A cells incubated with ProSeMet or L-Met. Cells treated with 100 µM ProSeMet or L-Met for 16 h were washed to remove unreacted ProSeMet, fixed, permeabilized, and ligated to a fluorophore-conjugated azide. Cells treated with ProSeMet exhibited labeling in both the cytosolic and nuclear compartments, while cells treated with L-Met remained unlabeled (**Fig. 2G, Fig. S1**). Evaluation of condensed chromatin in mitotic cells suggest ProSeMet-directed labeling of DNA is not achieved using this approach (**Fig. S2**). To ensure that ProSeMet would not cause cellular death during the experimental timeframe, we performed Annexin V/PI staining to assess the viability of cells treated with ProSeMet over a time course. ProSeMet treatment did not increase cytotoxicity compared to DMSO treatment at low (50 µM) or high (200 µM) concentrations (**Fig. S3**).

Enrichment of pseudomethylated proteins used to determine breadth of methylproteome

To determine the identity of pseudomethylated proteins in live cells following incubation with ProSeMet, the SMARB1-deficient G401 cell line was treated with ProSeMet or L-Met and cellular lysates were generated.

Whole cell lysates were subjected to CuAAC to attach a biotin handle, pulled down with streptavidin beads, and analyzed via label-free LC-MS/MS. Using this approach, we identified 707 proteins statistically enriched at *P* values at or below 0.05 and $\log_2$-fold change greater than 2 (**Fig. 3A**). Of these, 486 proteins have known methylation sites and 221 have uncharacterized methylation sites (**Fig. S4**)[38,39]. Furthermore, 47 proteins were identified at *P* values at or below 0.01 and $\log_2$-fold change greater than 5 and includes an enrichment of RNA processing proteins (SNRPD3, SNRPD1, SNRPB, SNRPE, QKI), enzymes critical to cell cycle regulation (CDK1), and several methyltransferases (CARM1, KMT2B) (**Fig. 3B**, top 15 genes).

Network analysis of ProSeMet enriched proteins using the gene ontology (GO) molecular signatures collection revealed a top list of GO terms involved in biological processes and molecular functions (**Fig. 3C**). As expected, proteins involved in methylation were highly enriched, suggesting pulldown of proteins involved in methyl transfer as well as methyltransferase targets. We further observed positive enrichment of genes associated with p53-mediated signal transduction and DNA damage response (DDR) (**Fig. 3D**). P53 is an important DNA sequence-specific transcription factor known to arrest growth by inducing apoptosis[40]. P53 activity is controlled in part by methylation at multiple key Lys and Arg residues: PRMT5-mediated p53 methylation at R333, R335, and R337 mediate p53 oligomerization and directly affects in DNA-induced apoptosis[41] whereas competing p53 methylation at K370/K372 by SMYD2/SETD7 can repress or enhance transcription of DDR target genes [7,42,43]. In addition to p53-mediated signaling, we also observed enrichment of proteins involved in RNA binding, processing, and stability, including previously identified PRMT5 targets SNRPD3, SNRPD1, and EEF1A1[44-46]. Our data supports previous studies suggesting heavy methylation of RNA-processing proteins.

Further pathway analysis using the REACTOME gene set identified gene clusters in actin dynamics, cell cycle control, WNT signaling, and transcriptional regulation (**Fig. S5**). Within these gene clusters, we found a series of proteins involved in adhesion dynamics, including the Rho GTPases RhoA, RhoG, and CDC42 and their downstream effectors, such as PAK2. Of these, only RhoA and PAK2 are known to be methylated[38,47]. We further identified key components of the WNT signaling pathway, including G3BP2, a DVL3-associated protein and positive regulator of WNT signaling that is methylated on R432, R438, R452, and R468 as well as WNT5A, which has no previously characterized methylation sites[38,39,48]. Collectively, these experiments identified dozens of known PRMT5, CARM1, EZH2, and other methyltransferase substrates in known pathways as well as novel targets, highlighting the prevalence of methylation in key signaling pathways and biological processes, and providing a chemical proteomics scaffold to survey the methylproteome in an unbiased manner.

ProSeMet-directed pseudomethylation is detectable *in vivo*

The ability of ProSeMet to produce pseudomethylated proteins *in vivo* was explored by administering adult male and female BALB c/J mice ProSeMet. We aimed to assess the impact of existing dietary L-Met on ProSeMet-directed labeling efficiency by dividing mice into the following cohorts: (1) dietary restriction for 12 h prior to L-Met administration; (2) no dietary restriction prior to ProSeMet administration, “fed;” or (3) dietary restriction for 12 h prior to ProSeMet administration, “starved.” In all cohorts, the dietary source of L-Met was restored immediately following L-Met/ProSeMet administration. Mice were administered 15 mg ProSeMet or the equimolar equivalent of L-Met in normal, sterile saline via intraperitoneal (IP) injection. This ProSeMet dose is equivalent to the daily dietary intake of L-Met for an adult mouse (**Fig. 4A**)[49]. To compare the labeling efficiency between groups, tissue from a diverse array of organs were collected from mice 12 h after ProSeMet or L-Met delivery. The tissues harvested from mice in all conditions were subjected to CuAAC to attach a fluorophore-

conjugated azide, then analyzed via fluorescent IHC or lysis followed by SDS-PAGE. Fluorescent labeling across organ systems demonstrates that ProSeMet can diffuse across tissues and penetrate the blood-brain barrier (BBB) (**Fig. 4B-D, Fig. S6**). Labeling was detected in both fed mice and mice experiencing dietary restriction to reduce intracellular L-Met abundance prior to ProSeMet administration (**Fig. 4B, S7**). This combined data suggests that mice starved prior to ProSeMet injection had increased ProSeMet labeling in the heart, whereas mice fed prior to ProSeMet administration had increased labeling in the brain and lungs. Additional SDS-PAGE experiments suggest ProSeMet labeling of the methylproteome occurs as early as 3 h (data not shown). Collectively, these data demonstrate ProSeMet can be utilized in *in vivo* animal models to pseudomethylate proteins from diverse organ systems.

Discussion

It is increasingly recognized that non-histone protein methylation is an important component of cell signaling and that comprehensive characterization of the methylproteome will require development of new experimental and analytical pipelines. Herein, we describe the application of the chemical probe ProSeMet, which can be used to pseudomethylate proteins in the natural cellular environment. In cells, ProSeMet is enzymatically converted to ProSeAM by cellular MAT enzymes, which can subsequently be used by diverse methyltransferases in both the cytoplasm and nucleus to deposit a biorthogonal alkyne on target proteins (**Fig. 1**). Our approach extends previous *in situ* methodologies[27-35] to allow consideration for the biological impetus behind methylation events; this approach allows dynamic, global mapping of the methylproteome.

Using ProSeMet, we developed a reproducible platform for analysis of protein methylation in live cells. For the first time, we demonstrated that ProSeMet-mediated pseudomethylation can be detected using SDS-PAGE and confocal microscopy in cancerous (T47D, LNCaP) and untransformed (MCF10A) cell lines. Our results indicate that ProSeMet can successfully penetrate the plasma and nuclear membranes to yield labelled proteins across molecular weights (**Fig. 2**). In addition to our *in vitro* results, we report, to our knowledge, the first application of L-Met/SAM analogues in a murine model. ProSeMet administered to Balb c/J mice via IP injection pseudomethylated proteins in the heart, lungs, and brain **(Fig. 4**). These data suggest that ProSeMet can be used to study the role of methylation in the pathophysiology of diseases affecting these and potentially other organ systems. We further hypothesize that ProSeMet can be widely leveraged in other *in vitro* and *in vivo* models.

Likewise, we developed a proteomics platform to enrich and identify pseudomethylated proteins. Using the SMARCB1-deficient cell line G401, we enriched non-histone proteins known to be methylated by PRMT1 (NONO)[50], PRMT5 (SNRPD3, SNRPD1)[45], CARM1 (SNRPB, SF3B4)[7], and EZH2 (STAT3, PCNA)[11,51] (**Fig. 3**). We also identified the catalytic subunits (PPP2CA, PPP4C, PPP6C) of several protein phosphatases known to

be methylated by the leucine carboxyl methyltransferase LCMT1[52]. Furthermore, we identified proteins suggested to be methylated by the promiscuous histidine methyltransferase METTL9 (SF3B4) as well as homologous members of the zinc transporter (SLC) and NADH:Ubiquinone Oxidoreductase (NDUF) families reported to contain histidine methylation[32]. In addition to these known methyltransferase targets, we identified 221 proteins with uncharacterized methylation sites with unknown biological function (**Fig. S4**). Our approach categorically demonstrates that ProSeMet treatment supports pseudomethylation of Arg and Lys residues while suggesting pseudomethylation of histidine and other non-canonical amino acid targets.

Our proteomics data also indicated enrichment of the SAM synthetase enzymes MAT2A/B, which are responsible for the coupling of ATP and L-Met to produce SAM[53]. While we cannot discount the possibility that these enzymes were identified because of ProSeAM present in their binding pockets, MAT2A is purportedly methylated on a critical Arg residue in the binding pocket (R264) and this methylation event may serve as a biological switch to control cellular SAM levels[54]. Similarly, our MS data indicates enrichment of several methyltransferases, including the Lys methyltransferases KMT2A/B, the Arg methyltransferases PRMT1, 3, 5 and CARM1, and the RNA methyltransferase NSUN2, all of which are known to be methylated[38]. To expand upon the approach taken here, future chemoproteomics analyses of ProSeMet targets could focus on confirmation of pseudomethylated peptides at high resolution, identification of the precise amino acid(s) that harbor methylation, and elucidation of the molecular mechanisms by which methylation of methyltransferases controls their function.

The ProSeMet pipeline described herein allows profiling of methyltransferase targets *in vitro* and *in vivo* while allowing a comprehensive survey of canonical and non-canonical methylated residues. Our approach has several advantages over previous methodologies: ProSeMet readily enters live cells without toxicity; ProSeMet-mediated pseudomethylation considers the biological impetus for a methylation event; ProSeMet and the corresponding SAM analogue ProSeAM can be used by native methyltransferases in cell lines and *in vivo*; and the CuAAC reaction is allows enrichment of pseudomethylated proteins and identification via LC-MS/MS. This platform provides a new avenue to dissect the methylation-controlled crosstalk and signal transduction responsible for cellular function.

Methods

Synthesis of ProSeMet

Propargyl-Selenium L-Methionine (ProSeMet) was synthesized follow a previously described protocol[33] with modified purification method. L-selenohomocystine (200 mg, 0.55 mmol) was dissolved in dry ethanol (25 mL), $NaBH_4$ (126 mg, 3.30 mmol) was added under nitrogen, and the reaction mixture was stirred at room temperature for 15 min. $NaHCO_3$ (152 mg, 3.3 mmol) was added to the reaction mixture in one portion, followed by propargyl bromide (250 μL, 3.3 mmol), and the resultant reaction mixture was stirred at room temperature for 12 h. The solvent was removed under reduced pressure, and the crude compound was dissolved in distilled water (6 mL) and 1% TFA (60 μL). The resulting solution was directly loaded to neutral alumina, and then 50% EtOAc and hexane (50 mL) were used to wash the nonpolar impurities, followed by 50% methanol and DCM (50 mL) to remove polar impurities. ProSeMet was eluted by NH_4OH solution (50 mL) from neutral alumina column, which was concentrated and lyophilized to afford the pure product (92 mg, 76%).

^{1}H NMR: ^{1}H NMR (600 MHz, D_2O) δ 3.84 – 3.81 (m, 1H), 3.30 (s, 2H), 2.84 (t, *J* = 8.0 Hz, 2H), 2.64 (s, 1H), 2.30 – 2.20 (m, 2H) (**Fig. S8**).

ESI-MS: Expected mass for $C_7H_{11}NO_2SeK^+$: 259.24 $[M + K]^+$; found: 259.815. (**Fig. S9**).

Cell Culture

Cells were purchased from ATCC and maintained at 37 °C and 5% CO_2. T47D, LNCaP, and G401 cells were cultured in RPMI supplemented with 10% (V/V) fetal bovine serum (FBS) and 1% (V/V) penicillin/streptomycin (pen/strep) (100 µg/mL). MCF10A cells were cultured in DMEM F-12 supplemented with 3% (V/V) FBS, cholera toxin (100 ng/mL), EGF (20 ng/mL), hydrocortisone (0.5 mg/mL), insulin (10 µg/mL), and 1% (V/V) pen/strep. For L-Met depletion experiments, cells were cultured in cysteine- and methionine-free DMEM (Gibco 21013024) supplemented with pen/strep, L-Cystine 2HCl (63 mg/mL), and dialyzed serum (Gibco A3382001). Cycloheximide (CHX) was purchased from Sigma (01810).

Cell Lysis and Immunoblotting: Whole cell lysate was generated by lysing cells on ice in RIPA buffer (50 mM TrisHCl [pH 8], 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with protease and phosphatase inhibitors. Lysates were centrifuged 6,500 *x* g, 10 m at 4°C, and soluble lysate collected. Proteins were separated using 12 or 16% SDS-PAGE. When applicable, proteins were fluorophore labeled via CuAAC. Proteins were then transferred to nitrocellulose membranes and visualized directly via fluorophore or visualized after primary antibody incubation (Odyssey, Li-Cor). Primary antibodies used were as follows: β-actin (Millipore-Sigma MAB1501) and HDAC1 (Upstate 05-614).

Copper-catalyzed Azide-Alkyne Cycloaddition (CuAAC) reaction: CuAAC on whole cell lysate, fixed cells, or tissue sections was performed using 1 mM CuSO4, 1.5 mM THPTA, 3 mM sodium ascorbate (NaAsc) and indicated concentration of azide (**Table S1**). All azides and alkynes were purchased from ClickChemistryTools.

For whole cell lysate, CuAAC reagents were added directly to lysate, incubated for 30 or 60 m at RT, then prepared for downstream analysis. For confocal microscopy, cells were fixed in 3.7% PFA and permeabilized with 0.5% Triton-X. CuAAC reagents were added directly to cells, then incubated, rocking, for 45 m. For IHC, tissue sections were permeabilized as described below and CuAAC reagents supplemented with 5% DMSO and 0.2% TX-100 were added directly to tissue section for 1 h prior to proceeding with the confocal microscopy pipeline described below.

Biotin-Streptavidin Pulldown: G401 cells were incubated in L-Met free media for 30 m, treated with 100 μM ProSeMet or L-Met for 16 h, then lysed with RIPA buffer. Protein lysate (1 mg/IP) was diluted to 1 mL final volume in PBS, then subjected to CuAAC for 1 h at RT to attach an azide conjugated biotin moiety (ClickChemistryTools). After CuAAC, protein was precipitated in acetone. Briefly, 4 volumes of ice-cold acetone were added to the post-CuAAC lysate, followed by incubation at -20 °C for 1 h. Precipitated protein were centrifuged at 9,000 *x* g for 10 m and resuspended in PBS with 5% SDS. After resuspension, SDS solution was diluted to 0.5% in PBS, then magnetic streptavidin beads (Pierce) were added directly to the resuspended protein. Protein and beads were incubated for 2 h at RT, rocking, and then beads were washed 3X with ice-cold RIPA buffer and 3X with ice-cold PBS. Beads with attached proteins were stored at -80 °C until LC-MS/MS analysis.

Annexin V/PI Staining: T47D cells were incubated with 5 μM or 20 μM of ProSeMet for 24 h, detached with trypsin (Biolegend), and stained using Annexin V/PI following manufacturer's protocol (Biolegend). Cells were analyzed via flow cytometer (BDFACSymphony A3) within 1 h.

Confocal microscopy: Live cells were plated and after 16 h, cells were washed to remove excess L-Met and incubated in methionine-free supplemented DMEM (Gibco) for 30 m to remove endogenous methionine. Subsequently, 100 μM or 200 μM concentrations of ProSeMet was added to the cells. After 16 h, cells were fixed with 4% PFA and permeabilized using 0.5% Triton X-100. The CuAAC reaction was then performed directly on the fixed and permeabilized cells as described to attach a 488 or 568 nm picolyl azide-conjugated fluorophore (ClickChemistryTools) and nuclei were counterstained using Hoechst. Cells were imaged on Leica SP8 confocal microscope and images were processed and analyzed using ImageJ and Python.

Mass spectrometry

Sample Preparation:

For on-bead digestion, a published protocol was followed[55]. To the bead, digestion buffer (50 mM NH_4HCO_3) was added, and the mixture was then treated with 1 mM dithiothreitol (DTT) at RT for 30 m, followed by 5 mM iodoacetimide (IAA) at RT for 30 m in the dark. Proteins were digested with 2 μg of lysyl endopeptidase (Wako) at RT for overnight and were further digested overnight with 2 μg trypsin (Promega) at RT. Resulting peptides were desalted with HLB column (Waters) and were dried under vacuum.

LC-MS/MS:

The data acquisition by LC-MS/MS was adapted from a published procedure[56]. Derived peptides were resuspended in the loading buffer (0.1% trifluoroacetic acid, TFA) and were separated on a Water's Charged Surface Hybrid (CSH) column (150 µm internal diameter (ID) x 15 cm; particle size: 1.7 µm). The samples were run on an EVOSEP liquid chromatography system using the 15 samples per day preset gradient (88 min) and were monitored on a Q-Exactive Plus Hybrid Quadrupole-Orbitrap Mass Spectrometer (ThermoFisher Scientific). The mass spectrometer cycle was programmed to collect one full MS scan followed by 20 data dependent MS/MS scans. The MS scans (400-1600 m/z range, 3 x 10^6 AGC target, 100 ms maximum ion time) were collected at a resolution of 70,000 at m/z 200 in profile mode. The HCD MS/MS spectra (1.6 m/z isolation width, 28% collision energy, 1 x 10^5 AGC target, 100 ms maximum ion time) were acquired at a resolution of 17,500 at m/z 200. Dynamic exclusion was set to exclude previously sequenced precursor ions for 30 seconds. Precursor ions with +1, and +7, +8 or higher charge states were excluded from sequencing.

MaxQuant:

Relative intensity values were calculated from raw data using MaxQuant. Label-free quantification analysis was adapted from a published procedure[56]. Spectra were searched using the search engine Andromeda, integrated into MaxQuant, against Human Uniprot/Swiss-Prot database (20,379 target sequences). Methionine oxidation (+15.9949 Da), asparagine and glutamine deamidation (+0.9840 Da), and protein N-terminal acetylation (+42.0106 Da) were variable modifications (up to 5 allowed per peptide); cysteine was assigned as a fixed carbamidomethyl modification (+57.0215 Da). Only fully tryptic peptides were considered with up to 2 missed cleavages in the database search. A precursor mass tolerance of ±20 ppm was applied prior to mass accuracy calibration and ±4.5 ppm after internal MaxQuant calibration. Other search settings included a maximum peptide mass of 6,000 Da, a minimum peptide length of 6 residues, 0.05 Da tolerance for orbitrap and 0.6 Da tolerance for ion trap MS/MS scans. The false discovery rate (FDR) for peptide spectral matches, proteins, and site decoy fraction were all set to 1 percent. Quantification settings were as follows: re-quantify with a second peak finding attempt after protein identification has completed; match MS1 peaks between runs; a 0.7 min retention time match window was used after an alignment function was found with a 20-minute RT search space. Quantitation of proteins was performed using summed peptide intensities given by MaxQuant. The quantitation method only considered razor plus unique peptides for protein level quantitation.

Data analysis

Following MaxQuant, all data was log2-transformed before further analysis. Contaminant proteins were filtered and missing values were imputed using a normal distribution in Perseus. Differential protein expression was performed on normalized intensity values using the DEqMS R package (https://www.bioconductor.org/packages/release/bioc/html/DEqMS.html)[57]. Samples with a high level of variance

calculated using PCA were removed (**Fig. S6**). GSEA was performed on enriched proteins (LFC > 1) using the fgsea R package (https://bioconductor.org/packages/release/bioc/html/fgsea.html). Datasets with log_2-transformed fold change were analyzed using C5 (GO) and C2 (REACTOME) gene set collections in the Molecular Signatures Database (MSigDB v 7.5.1). Identification of proteins with known or unknown methylation sites was performed by cross referencing the PhosphoSitePlus and ActiveDriver DB databases[38,39].

In vivo animal studies

Male and female BALB c/J mice (Jackson labs) were maintained on a standard chow diet. 12 h prior to L-Met or ProSeMet delivery, mice were either starved or maintained on the standard chow diet, after which they were administered 15 mg ProSeMet or equimolar L-Met, resuspended in sterile 0.9N saline, via intraperitoneal (IP) injection. At the time of ProSeMet or L-Met administration, standard chow was returned. Mice were monitored hourly post administration and euthanized after 12 h. All tissue collected was isolated and prepared within 1 h following euthanasia. All mouse experiments were conducted in accordance with protocols approved by the Institutional Animal Care and Use Committee (IACUC) of Emory University School of Medicine (EUCM).

Fluorescent immunohistochemistry (IHC) staining

Organs extracted from mice treated with ProSeMet or L-Met were immediately placed in 4% formaldehyde solution and left at RT for 24 h. Subsequently, organs were paraffin embedded and sectioned. Tissue slides were rehydrated by washing 3 times with xylenes (mixed isomers), twice with 100% EtOH, once with 95% EtOH, once with 80% EtOH, once with 70% EtOH, and once with PBS. All washes were for 5 m. After the final wash, tissue sections were permeabilized using 0.5% TX-100 in PBS for 1 h. After permeabilization, CuAAC was performed directly on slides using 1 mM CuSO4, 1.5 mM THPTA, 3 mM NaAsc, 25 μM picolyl azide conjugated to a 568 nm fluorophore (ClickChemistryTools), 5% DMSO, and 0.2% TX-100, for 1 h at RT. After reaction completed, slides were washed for 30 m in PBS supplemented with 5% DMSO and 0.2% TX-100, then for 30 m in PBS. Slides were then stained with DAPI (300 nM) for 5 m at RT. Prior to mounting, slides were dehydrated by washing once with 70% EtOH, once with 80% EtOH, twice with 100% EtOH, and twice with xylenes (mixed isomers). All washes were for 5 m. After dehydration, slides were mounted and imaged using a confocal microscope (Leica SP8).

Acknowledgements and funding sources

The authors thank Dr. David Lynn, Dr. Andrew Hong, and the Spangle, Raj, and Hong labs for helpful project discussions. This study was supported by funds from Winship Cancer Institute of Emory University (to JMS). M.R. acknowledges funding from NIH (1R35GM133719-01). This study was supported in part by the Cancer Tissue Pathology Core and Emory Integrated Proteomics Core, both of which are shared resources of Winship Cancer Institute of Emory University and the NIH/NCI (P30CA138292)

Author Contributions:

J.F., M.R., and J.M.S. designed the research. J.F., R.B.J., K.K.P., B.E., K.T., A.K.V., and P.B. performed experiments and analyzed the results. J.M.S. supervised the studies. J.F. and J.M.S. wrote the manuscript.

Data availability statement:

The data generated in this study are publicly available in synapse (in progress).

Correspondence and requests for materials should be addressed to:
Dr. Jennifer M. Spangle, Ph.D., Jennifer.spangle@emory.edu

References

1 Murn, J. & Shi, Y. The winding path of protein methylation research: milestones and new frontiers. *Nature Reviews Molecular Cell Biology* **18**, 517--527 (2017).

2 Michalak, E. M., Burr, M. L., Bannister, A. J. & Dawson, M. A. The roles of DNA, RNA and histone methylation in ageing and cancer. *Nature reviews Molecular cell biology* **20**, 573--589 (2019).

3 Fontecave, M., Atta, M. & Mulliez, E. S-adenosylmethionine: nothing goes to waste. *Trends in biochemical sciences* **29**, 243--249 (2004).

4 Bedford, M. T. Arginine methylation at a glance. *Journal of cell science* **120**, 4243-4246 (2007).

5 Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell research* **21**, 381--395 (2011).

6 Wu, Z., Connolly, J. & Biggar, K. K. Beyond histones--The expanding roles of protein lysine methylation. *The FEBS journal* **284**, 2732--2744 (2017).

7 Biggar, K. K. & Li, S. S.-C. Non-histone protein methylation as a regulator of cellular signalling and function. *Nature reviews Molecular cell biology* **16**, 5-17 (2015).

8 Guo, J. *et al.* AKT methylation by SETDB1 promotes AKT kinase activity and oncogenic functions. *Nature Cell Biology* **21**, 226--237 , pmid = 30692625, doi:10.1038/s41556-018-0261-6 , file = :Users/yonatan/Library/Application Support/Mendeley Desktop/Downloaded/Guo et al. - 2019 - AKT methylation by SETDB1 promotes AKT kinase activity and oncogenic functions.pdf:pdf (2019).

9 Wang, G. *et al.* SETDB1-mediated methylation of Akt promotes its K63-linked ubiquitination and activation leading to tumorigenesis. *Nature cell biology* **21**, 214--225 (2019).

10 Chan, L. H. *et al.* PRMT6 regulates RAS/RAF binding and MEK/ERK-mediated cancer stemness activities in hepatocellular carcinoma through CRAF methylation. *Cell reports* **25**, 690--701 (2018).

11 Kim, E. *et al.* Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells. *Cancer cell* **23**, 839--852 (2013).

12 He, A. *et al.* PRC2 directly methylates GATA4 and represses its transcriptional activity. *Genes \& development* **26**, 37--42 (2012).

13 Park, S. H. *et al.* Posttranslational regulation of FOXA1 by Polycomb and BUB3/USP7 deubiquitin complex in prostate cancer. *Science advances* **7**, eabe2261 (2021).

14 Hoy, S. M. Tazemetostat: first approval. *Drugs* **80**, 513-521 (2020).

15 Siu, L. L. *et al.* METEOR-1: A phase I study of GSK3326595, a first-in-class protein arginine methyltransferase 5 (PRMT5) inhibitor, in advanced solid tumours. *Annals of Oncology* **30**, v159 (2019).

16 Watts, J. M. *et al.* A phase I/II study to investigate the safety and clinical activity of the protein arginine methyltransferase 5 inhibitor GSK3326595 in subjects with myelodysplastic syndrome and acute myeloid leukemia. *Blood* **134**, 2656 (2019).

17 Ong, S.-E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nature methods* **1**, 119--126 (2004).

18 Cao, X.-J., Arnaudo, A. M. & Garcia, B. A. Large-scale global identification of protein lysine methylation in vivo. *Epigenetics* **8**, 477--485 (2013).

19 Guo, A. *et al.* Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. *Molecular \& Cellular Proteomics* **13**, 372--387 (2014).

20 Levy, D. *et al.* A proteomic approach for the identification of novel lysine methyltransferase substrates. *Epigenetics \& chromatin* **4**, 1--12 (2011).

21 Bremang, M. *et al.* Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. *Molecular bioSystems* **9**, 2231--2247 (2013).

22 Moore, K. E. & Gozani, O. An unexpected journey: lysine methylation across the proteome. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1839**, 1395--1403 (2014).

23 Hart-Smith, G., Yagoub, D., Tay, A. P., Pickford, R. & Wilkins, M. R. Large scale mass spectrometry-based identifications of enzyme-mediated protein methylation are subject to high false discovery rates. *Molecular & Cellular Proteomics* **15**, 989-1006 (2016).

24 Raj, M., Nwajiobi, O., Mahesh, S. & Streety, X. Triazenation Unable Selective Reaction of Secondary Amines (TrUSS) for Tagging Monomethyl Lysine PTMs. *Angewandte Chemie (International ed. in English)* (2020).

25 Chalker, J. M., Bernardes, G. J. L. & Davis, B. G. A “tag-and-modify” approach to site-selective protein modification. *Accounts of chemical research* **44**, 730--741 (2011).

26 Islam, K. The bump-and-hole tactic: expanding the scope of chemical genetics. *Cell chemical biology* **25**, 1171--1184 (2018).

27 Wang, R. *et al.* Profiling genome-wide chromatin methylation with engineered posttranslation apparatus within living cells. *Journal of the American Chemical Society* **135**, 1048--1056 (2013).

28 Wang, R. & Luo, M. A journey toward Bioorthogonal Profiling of Protein Methylation inside living cells. *Current opinion in chemical biology* **17**, 729--737 (2013).

29 Peters, W. *et al.* Enzymatic site-specific functionalization of protein methyltransferase substrates with alkynes for click labeling. *Angewandte Chemie International Edition* **49**, 5170--5173 (2010).

30 Bothwell, I. R. *et al.* Se-adenosyl-L-selenomethionine cofactor analogue as a reporter of protein methylation. *Journal of the American Chemical Society* **134**, 14905--14912 (2012).

31 Sohtome, Y., Shimazu, T., Shinkai, Y. & Sodeoka, M. Propargylic Se-adenosyl-l-selenomethionine: A Chemical Tool for Methylome Analysis. *Accounts of Chemical Research* **54**, 3818--3827 (2021).

32 Davydova, E. *et al.* The methyltransferase METTL9 mediates pervasive 1-methylhistidine modification in mammalian proteomes. *Nature communications* **12**, 1--14 (2021).

33 Hartstock, K. *et al.* Enzymatic or in vivo installation of propargyl groups in combination with click chemistry for the enrichment and detection of methyltransferase target sites in RNA. *Angewandte Chemie International Edition* **57**, 6342--6346 (2018).

34 Muttach, F. & Rentmeister, A. A biocatalytic cascade for versatile one-pot modification of mRNA starting from methionine analogues. *Angewandte Chemie International Edition* **55**, 1917--1920 (2016).

35 Mikutis, S. *et al.* meCLICK-Seq, a substrate-hijacking and RNA degradation strategy for the study of RNA methylation. *ACS central science* **6**, 2196--2208 (2020).

36 Knutson, S. K. *et al.* Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proceedings of the National Academy of Sciences* **110**, 7922--7927 (2013).

37 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). *Proceedings of the National Academy of Sciences* **103**, 9482-9487 (2006).

38 Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic acids research* **43**, D512-D520 (2015).

39 Krassowski, M. *et al.* ActiveDriverDB: human disease mutations and genome variation in post-translational modification sites of proteins. *Nucleic acids research* **46**, D901-D910 (2018).

40 Menendez, D., Inga, A. & Resnick, M. A. The expanding universe of p53 targets. *Nature reviews Cancer* **9**, 724-737 (2009).

41 Jansson, M. *et al.* Arginine methylation regulates the p53 response. *Nature cell biology* **10**, 1431-1439 (2008).

42 Chuikov, S. *et al.* Regulation of p53 activity through lysine methylation. *Nature* **432**, 353-360 (2004).

43 Huang, J. *et al.* Repression of p53 activity by Smyd2-mediated methylation. *Nature* **444**, 629-632 (2006).

44 Fong, J. Y. *et al.* Therapeutic targeting of RNA splicing catalysis through inhibition of protein arginine methylation. *Cancer cell* **36**, 194-209. e199 (2019).

45 Meister, G. *et al.* Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Current Biology* **11**, 1990-1994 (2001).

46 Bezzi, M. *et al.* Regulation of constitutive and alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery. *Genes & development* **27**, 1903-1916 (2013).

47 Hodge, R. G. & Ridley, A. J. Regulating Rho GTPases and their regulators. *Nature reviews Molecular cell biology* **17**, 496-510 (2016).

48 Bikkavilli, R. K. & Malbon, C. C. Wnt3a-stimulated LRP6 phosphorylation is dependent upon arginine methylation of G3BP2. *Journal of cell science* **125**, 2446-2456 (2012).

49 Brown-Borg, H. M. Reduced growth hormone signaling and methionine restriction: interventions that improve metabolic health and extend life span. *Annals of the New York Academy of Sciences* **1363**, 40--49 (2016).

50 Yin, X.-K. *et al.* PRMT1 enhances oncogenic arginine methylation of NONO in colorectal cancer. *Oncogene* **40**, 1375-1389 (2021).

51 Xu, X. *et al.* EZH2 promotes DNA replication by stabilizing interaction of POLδ and PCNA via methylation-mediated PCNA trimerization. *Epigenetics & Chromatin* **11**, 1-14 (2018).

52 Hwang, J., Lee, J. A. & Pallas, D. C. Leucine carboxyl methyltransferase 1 (LCMT-1) methylates protein phosphatase 4 (PP4) and protein phosphatase 6 (PP6) and differentially regulates the stable formation of different PP4 holoenzymes. *Journal of Biological Chemistry* **291**, 21008-21019 (2016).

53 Singh, S. *et al.* Facile chemoenzymatic strategies for the synthesis and utilization of S-adenosyl-L-methionine analogues. *Angewandte Chemie* **126**, 4046-4050 (2014).

54 Massignani, E. *et al.* ProMetheusDB: an in-depth analysis of the high-quality human methyl-proteome. *bioRxiv* (2021).

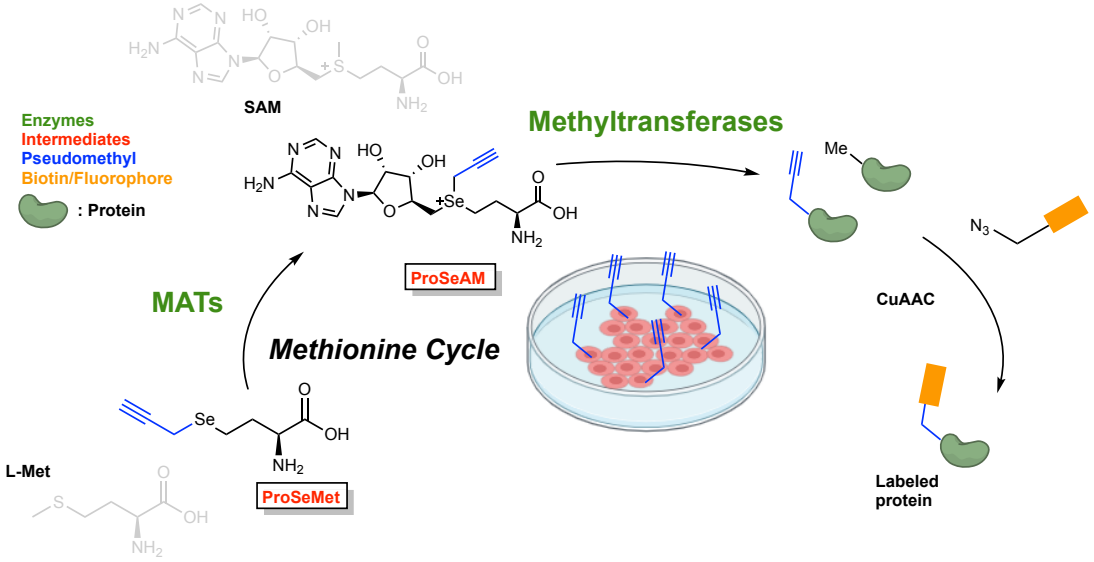
55 Soucek, S. *et al.* Evolutionarily conserved polyadenosine RNA binding protein Nab2 cooperates with splicing machinery to regulate the fate of pre-mRNA. *Molecular and cellular biology* **36**, 2697-2714 (2016).

56 Seyfried, N. T. *et al.* A multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer's disease. *Cell systems* **4**, 60-72. e64 (2017).

57 Zhu, Y. *et al.* DEqMS: a method for accurate variance estimation in differential protein expression analysis. *Molecular & Cellular Proteomics* **19**, 1047-1057 (2020).

1 **Figures and Figure Legends:**

2 **Figure 1. Chemoenzymatic approach for metabolic methyltransferase labeling.**



3

4 **Figure 1. Chemoenzymatic approach for metabolic MTase labeling**. ProSeMet can be converted to
5 ProSeAM by MAT enzymes in live cells. ProSeAM can then be used by diverse methyltransferases to
6 pseudomethylate (tag) target protein.

7

Figure 2. ProSeMet pseudomethylates nuclear and cytosolic proteins in live cells.

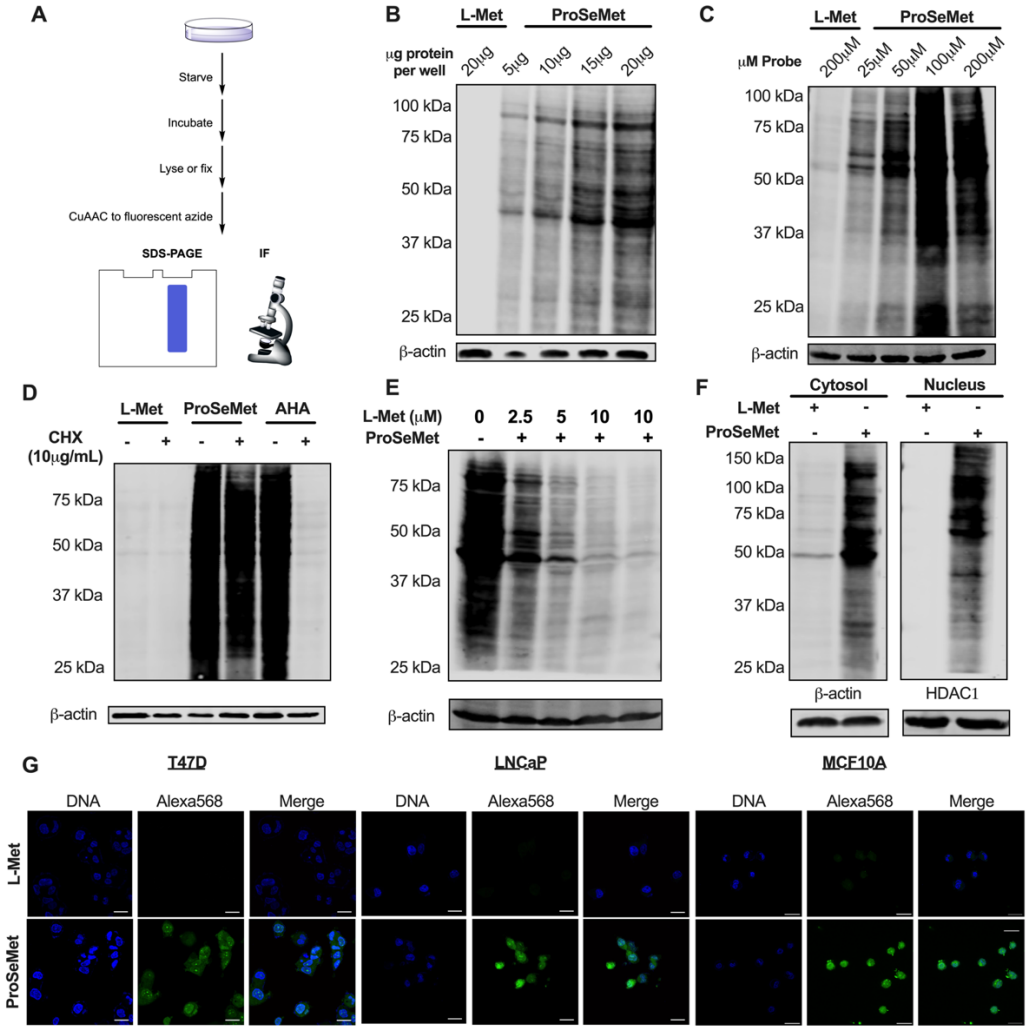
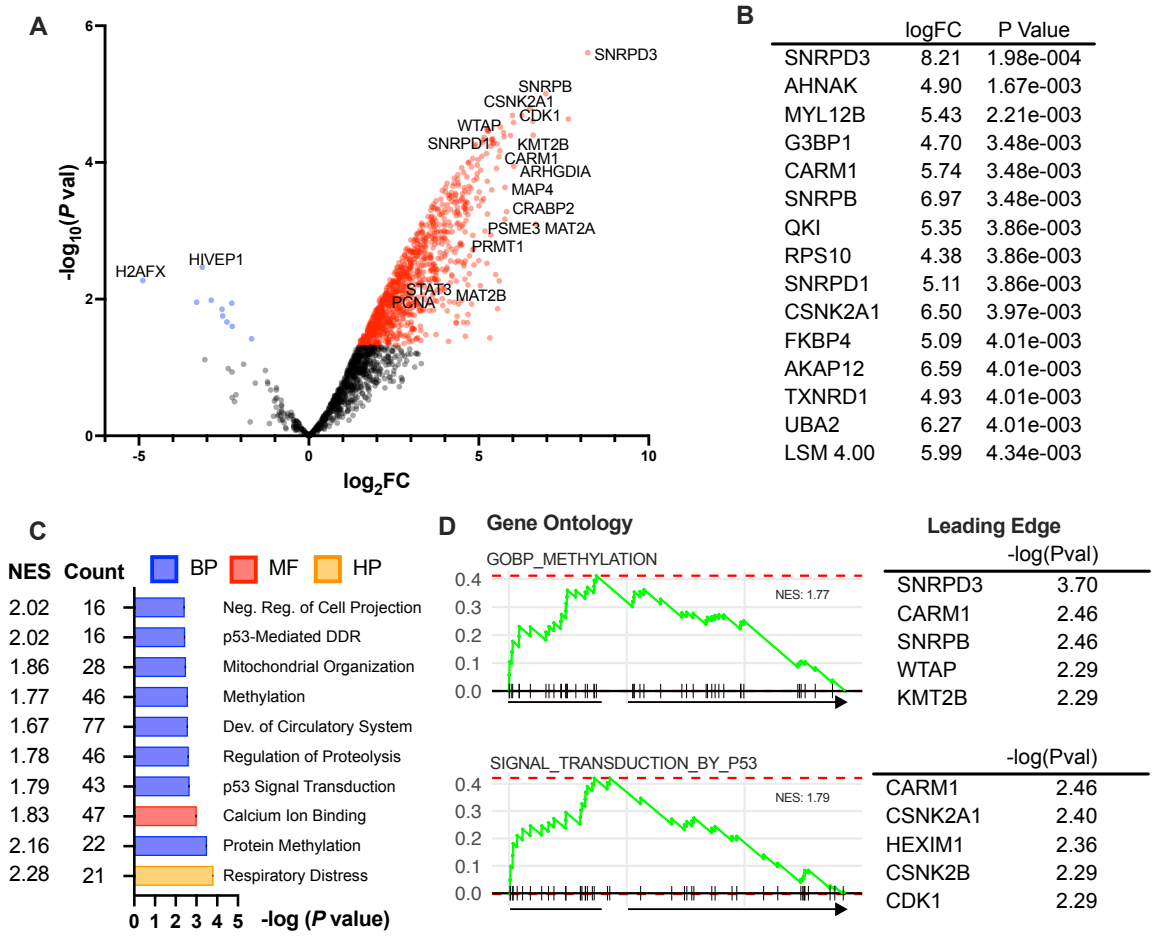


Figure 2. ProSeMet pseudomethylates nuclear and cytosolic proteins in live cells. **a.** Workflow for gel- and IF-based profiling of ProSeMet target proteins. Cells were starved of L-Met by incubating for 30 m in L-Met free media. Cells were then lysed or fixed, subjected to CuAAC to attach a fluorophore, and separated via SDS-PAGE or imaged via confocal microscopy. **b**. T47D cells treated with 100 μM ProSeMet or 100 μM L-Met 16 h were lysed and subjected to click reaction to attach a fluorescent picolyl azide (680 nm). Increasing protein concentration (5-20 μg) was loaded into each well and separated by SDS-PAGE (*n*=3). **c**. T47D cells were treated with 200 μM L-Met or increasing concentration (25-200 μM) ProSeMet 16 h. Cellular lysates were subjected to CuAAC and separated via SDS-PAGE. Peak ProSeMet labeling occurs at 100 μM concentrations (*n*=3). **d.** T47D cells were treated with 100 μM L-Met, ProSeMet, or AHA in the presence or absence of 10 μg/mL of cycloheximide (CHX). After 16 h, lysates were collected and subjected to CuAAC to attach a fluorescent picolyl azide (L-Met, ProSeMet) or fluorescent alkyne (AHA) then separated by SDS-PAGE (*n*=3). **e**. Competition by increasing concentrations of L-Met during incubation of T47D cells with 100 μM ProSeMet reduces ProSeMet

labeling across molecular weights and in a dose-dependent manner (*n*=3). **f**. Cell fractionation of T47D cells treated with 100 μM ProSeMet or 100 μM L-Met demonstrates ProSeMet-mediated pseudomethylation occurs in both the cytoplasm and nucleus (*n*=3). **g**. T47D, LNCaP, and MCF10A cells were treated with 100 μM ProSeMet or 100 μM L-Met 16 h then fixed, permeabilized, subjected to CuAAC to attach a fluorescent picolyl azide (568 nm, pseudocolored green), and counterstained with Hoescht (blue). ProSeMet treated cells show robust labeling in cytosol and nucleus across conditions.

Figure 3. Pseudomethylated proteins converge into predicted gene ontologies while extending the methylproteome



	logFC	P Value
SNRPD3	8.21	1.98e-004
AHNAK	4.90	1.67e-003
MYL12B	5.43	2.21e-003
G3BP1	4.70	3.48e-003
CARM1	5.74	3.48e-003
SNRPB	6.97	3.48e-003
QKI	5.35	3.86e-003
RPS10	4.38	3.86e-003
SNRPD1	5.11	3.86e-003
CSNK2A1	6.50	3.97e-003
FKBP4	5.09	4.01e-003
AKAP12	6.59	4.01e-003
TXNRD1	4.93	4.01e-003
UBA2	6.27	4.01e-003
LSM 4.00	5.99	4.34e-003

Figure 3. Pseudomethylated proteins converge into predicted gene ontologies while extending the methylproteome. G401 cells treated with 100 μM ProSeMet or 100 μM L-Met 16 h were lysed, subjected to a click reaction to attach a biotin-conjugated picolyl azide, and enriched using magnetic streptavidin beads. Enriched proteins were digested with trypsin and analyzed via label-free LC-MS/MS. **a.** Volcano plot showing statistical significance vs $\log_2$-fold change (logFC) of proteins isolated from G401 cells treated with ProSeMet or L-Met. Data represents 3 biological replicates consisting of matched experiments. Statistical significance was calculated using a two-tailed unpaired *t*-test. **b**. Top 15 enriched proteins. Statistical significance represents scaled *P* values after differential expression analysis. **c**. Gene set enrichment analysis (GSEA) of ProSeMet enriched proteins using GO gene set. The top ten pathways are shown with corresponding number of proteins (count) and normalized enrichment score (NES). **d**. GSEA analysis of methylation (top) and p53-mediated signal transduction (bottom) gene sets. Top 5 enriched proteins by *P* value in the leading edge are shown to the right. For **c** and **d**, statistical significance was assessed using Fast Gene Set Enrichment Analysis (fgsea).

Figure 4. ProSeMet pseudomethylates proteins *in vivo.*

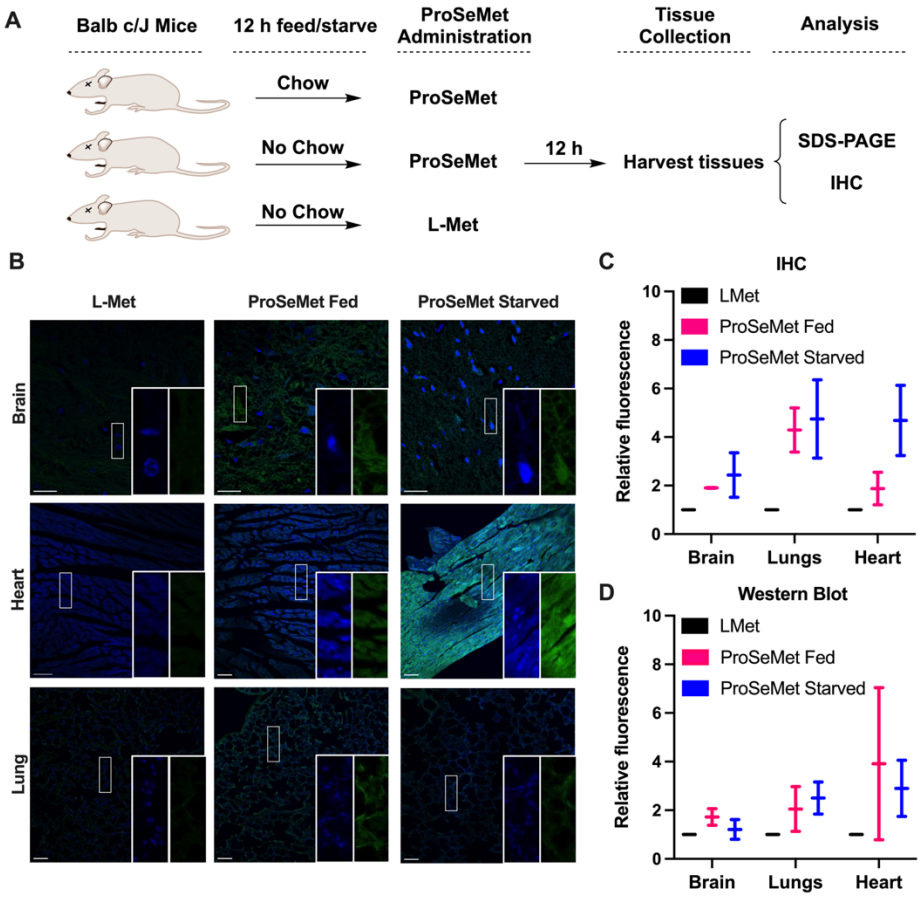


Figure 4. ProSeMet pseudomethylates proteins *in vivo*. a. Workflow for *in vivo* administration of ProSeMet or L-Met and subsequent analysis of organs for pseudomethylation. **b**. Organs extracted from mice treated with 15 mg ProSeMet or equimolar L-Met via IP injection while being fed or starved for 12 h were formaldehyde fixed and paraffin embedded. Tissue sections were subjected to CuAAC to attach a fluorescent picolyl azide (568 nm, pseudocolored green) and counterstained with DAPI (blue). Fluorescent IHC analysis of brain (top), heart (middle), and lung (bottom) pseudomethylation demonstrates successful blood-brain barrier penetrance of ProSeMet, as well as pan-organ labeling. Scalebar represents 25 μM (top) or 50 μM (middle, bottom) (*n*=2). **c**. Quantification of IHC fluorescence in brain, heart, and lungs (*n*=2). **d**. Western blot densitometry normalized to β-actin and relative to L-Met control (*n*=2). Brain, heart, and lung show increased protein labeling compared to L-met; no significant difference observed between mice that were fed or starved prior to ProSeMet administration.