Title
Development of cell-specific RNA metabolic labeling applications in cell and animal models

Author
Singha, Monika Kaur

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Development of Cell-Specific RNA Metabolic Labeling Applications in Cell and Animal Models

DISSERTATION

submitted in partial satisfaction of the requirements for
the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Monika Kaur Singha

Dissertation Committee:
Professor Robert C. Spitale, Chair
Professor Scott X. Atwood
Assistant Professor Christopher J. Halbrook

2022

DEDICATION

To

Damini, who constantly reminded me to laugh, imagine, love and create throughout everything.

My parents, Daljit and Gurbax, who relentlessly supported my lifelong ambitions.

Justin, who taught me to believe in myself and illuminated great possibilities.

All of my friends and mentors for helping grow and develop my strengths.

More diverse groups of people bring more creativity to solving problems

Joan Steitz
Yale School of Medicine

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Chapter 1 of this dissertation is a reprint of the material as it appears in Singha, Monika et. al. *WIREs RNA* **2021**, used with permission from John Wiley & Sons, Inc. The co-authors listed in this publication are Spitalny, Leslie; Nguyen, Kim; Vandewalle, Abigail; Spitale, Robert C. with the title "Chemical Methods for Measuring RNA Expression with Metabolic Labeling".

Chapter 2 of this dissertation is a reprint of the material as it appears in Nguyen, Kim et. al. *ACS Chem. Biol.* **2020**, used with permission from American Chemical Society. The co-authors listed in this publication are Kubota, Miles; Arco, Jon del; Feng, Chao; Singha, Monika; Beasley, Samantha; Sakr, Jasmine; Gandhi, Sunil P.; Blurton-Jones, Matthew; Fernández Lucas, Jesus; Spitale, Robert C. with the title "A Bump-Hole Strategy for Increased Stringency of Cell-Specific Metabolic Labeling of RNA."

Chapter 4 of this dissertation is a reprint of the material as it appears in Gupta, Mrityunjay et. al. *Org. Lett.* **2021**, used with permission from American Chemical Society. The co-authors listed in this publication are Singha, Monika; Rasale, Dnyaneshwar B.; Zhou, Zehao; Bhandari, Srijana; Beasley, Samantha; Sakr, Jasmine; Parker, Shane M.; Spitale, Robert. C. with the title "Mutually Orthogonal Bioconjugation of Vinyl Nucleosides for RNA Metabolic Labeling."

I thank Professor Mary Nolan-Riegle who inspired me towards this path. Her incessant enthusiasm towards plant biochemistry sparked my interest in biology and research.

I thank Professor Suzanne Sandmeyer and graduate student mentor James Yu who both encouraged me during my very first laboratory experiments.

I thank Professor Olga Razorenova who gave me the opportunity to work with her amazing group of Ph.D. students during my undergraduate research experience. Through this, I became motivated to pursue a Ph.D. degree.

I am thankful for the extensive support from my current mentor and advisor, Professor Rob Spitale, who guided me in the surf of many failures to become relentless. There was not a question I could not ask, or a project I could not pursue as he fostered an environment for creativity and collaboration for all in our group. I am grateful for his mentorship during all hours of the week helping me to define my career goals by providing many opportunities.

I am very appreciative to be able to work with Professors Scott Atwood and Chris Halbrook for they taught me how to develop these projects in mice, thanks to their generous guidance. I feel privileged to work within both of these lab groups and have them as part of my committee.

I thank Professor Brian Paegel for supporting me as an all-around person devoted to science.

I thank to all current and former Spitale Lab members. I've learned so much from this diverse group. Everyone taught me something new. I couldn't ask for a better lab group to develop my scientific expertise within the realms of RNA chemical biology.

I express my appreciation to my first friend in the Ph.D. program, Barbara Freda. Thank you for reminding me to touch grass and feel the warmth of the sun while also talking about science.

To my child, Damini Sangha, whose patience while I worked long and strange hours to spend more time with them was appreciated beyond anything I could ever expect. I appreciate you the most.

I am grateful to my parents, Daljit and Gurbax Singha, who supported us while I pursued my passion for science.

VITA

Monika Kaur Singha

EDUCATION

Doctor of Philosophy in Biological Sciences – June 2017 to August 2022

University of California, Irvine

Bachelor of Science in Biochemistry and Molecular Biology – June 2014 to June 2017

University of California, Irvine

RESEARCH EXPERIENCE

Spitale Lab- University of California, Irvine

June 2017 – August 2022

Razorenova Lab- University of California, Irvine

June 2016- June 2017

PUBLICATIONS

Exploiting Endogenous Enzymes for Cancer-Cell Selective Metabolic Labeling of RNA in Vivo. Samantha Beasley, Abigail Vandewalle, Monika Singha, Kim Nguyen, Whitney England, Eric Tarapore, Nan Dai, Ivan R. Corrêa, Scott X. Atwood, Robert C. Spitale. *Journal of the American Chemical Society.* 2022; 144 (16), 7085–7088.

Mutually Orthogonal Bioconjugation of Vinyl Nucleosides for RNA Metabolic Labeling. Mrityunjay Gupta, Monika Singha, Dnyaneshwar B. Rasale, Zehao Zhou, Srijana Bhandari, Samantha Beasley, Jasmine Sakr, Shane M. Parker, and Robert C. Spitale. *Organic Letters.* 2021; *23* (18), 7183-7187.

Chemical methods for measuring RNA expression with metabolic labeling. Monika Singha, Leslie Spitalny, Kim Nguyen, Abigail Vandewalle, Robert C. Spitale. *WIREs RNA.* 2021; 12:e1650.

A Bump-Hole Strategy for Increased Stringency of Cell-Specific Metabolic Labeling of RNA. Kim Nguyen, Miles Kubota, Jon del Arco, Chao Feng, Monika Singha, Samantha Beasley, Jasmine Sakr, Sunil P. Gandhi, Matthew Blurton-Jones, Jesus Fernández Lucas, and Robert C. Spitale. *ACS Chemical Biology.* 2020 ;*15* (12), 3099-3105.

Targeting the Mevalonate Pathway Suppresses VHL-Deficient CC-RCC through an HIF-Dependent Mechanism. Jordan M. Thompson, Alejandro Alvarez, Monika Singha, Matthew W. Pavesic, Quy H. Nguyen, Luke J. Nelson, David A. Fruman, Olga V. Razorenova; *Molecular Cancer Therapy.* 2018; 17 (8): 1781–1792.

ABSTRACT OF THE DISSERTATION

Development of Cell-Specific RNA Metabolic Labeling Applications in Cell and Animal Models

by

Monika Kaur Singha

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2022

Professor Robert C. Spitale, Chair

The development of new tools for RNA metabolic labeling is critical for analyzing nascent transcription in cells and animal models to elucidate specific regulatory mechanisms of RNA. These approaches have recently been demonstrated with commercially available methods, such as thiol (SH)-linked alkylation for the metabolic sequencing of RNA (SLAM-seq), yet there remain many opportunities to refine RNA metabolic labeling for cell-specific approaches. In the following dissertation, I will highlight the important work that has already been developed towards these efforts. I will also introduce a novel cell-specific RNA metabolic labeling method using a new enzyme-analog pair. For this approach, I will demonstrate how it can be used *in vivo* to capture nascent RNA within a metastatic breast cancer mouse model. Lastly, I will discuss a new chemo-selective reagent pair for vinyl nucleosides that can be used to analyze two RNA populations simultaneously in cells. Throughout the following document, I will emphasize the chemical and biochemical challenges and advantages of these labeling methods and discuss opportunities for future development. These tools are dedicated to improving biological insight to study transcription with greater physiological specificity.

INTRODUCTION

RNA are important regulatory molecules in cells that perform a variety of functions for maintaining cell homeostasis. In addition to being an integral component to protein translation, RNA molecules can act as a molecular signals, epigenetic regulators, intermolecular scaffolds and protein-binding modulators within cells[1]. These diverse functions of RNA have been rapidly identified over the last two decades with high-throughput RNA-sequencing applications to quantify genome-wide transcription[2]. As advancements in sequencing technologies continue to develop, the growing list of RNA functions is expected to expand rapidly with the implementation of multi-omics approaches. For instance, applying RNA immunoprecipitation sequencing (RIP-seq) and long read RNA-sequencing can each contribute unique biological insights into transcriptional pathways by measuring RNA-protein binding sites and splice variants. Combining new tools with RNA-sequencing to identify RNA expressed during critical cell processes has extensive potential to introduce unique strategies for developing RNA-targeting therapeutics to manage or prevent disease progression by targeting these multi-faceted and ubiquitous, functional molecules[3-5].

Many RNA molecules with regulatory importance are known to be expressed transiently or at low expression levels in specific cell types leading to challenges in detection with conventional RNA-sequencing methods that tend to measure the bulk, steady-state levels of RNA expressed at the time of cell lysis[6]. Through measuring RNA kinetics, RNA metabolic labeling experiments have shown that regulatory RNA have rapid turnover rates which may not be effectively represented in RNA-sequencing data sets. To address this challenge, RNA metabolic labeling is capable of tracking newly synthesized

RNA by treating cells with non-canonical nucleoside derivatives that contain chemical handles **(Figure 1a).** RNA metabolic labeling is used to determine RNA synthesis and decay rates through pulse-chase experiments applied to cells over incremental time points[7-8]. During the pulse, cells are treated with nucleoside analogs that are catalyzed through cell metabolism via nucleotide salvage enzymes and intracellular kinases converting these into nucleoside triphosphates for incorporation into newly synthesized RNA by RNA polymerase. The pulse is followed by a chase using canonical, unmodified nucleosides to "washout" the label-containing RNA within a designated time range (**Figure 1b)**. By calculating the ratio of newly transcribed, labeled RNA to the total RNA expressed, the half-life of specific RNA molecules are determined over the course of a cell process such as cellular differentiation, metabolic reprogramming or drug-response experiments in cells, as a few key examples[9]. The changes in RNA synthesis and decay rates are then used to determine the mechanisms of RNA in the context of these dynamic cell physiological processes[11].

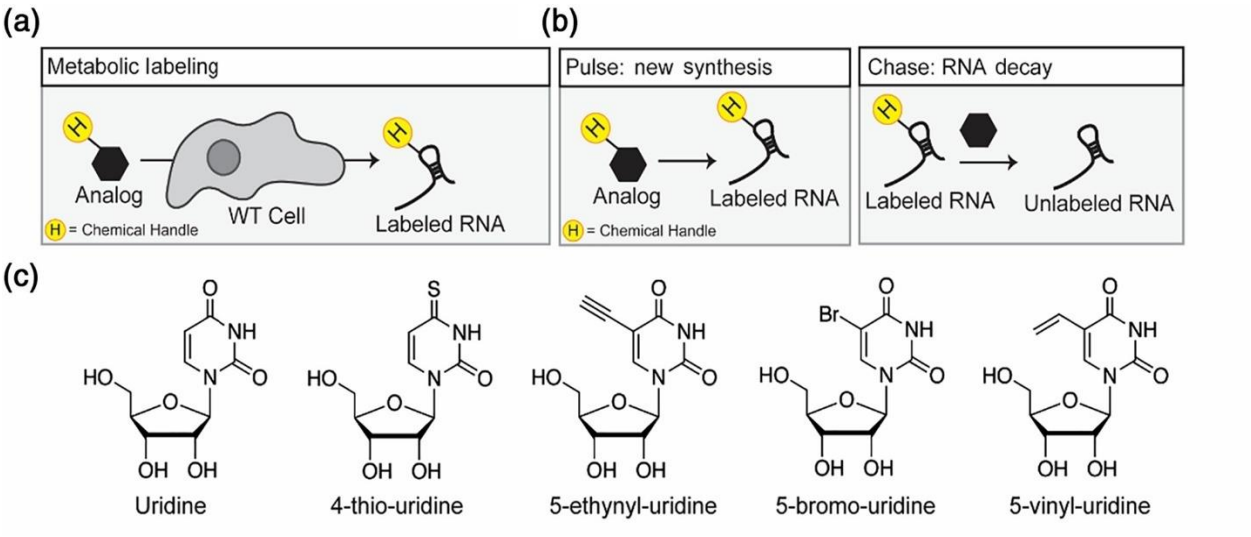


Figure 1: Chemical probes for metabolic labeling of RNA. **(a)** Schematic of metabolic labeling experiments. **(b)** Schematic of pulse and chase experiments: two classes of experiments are traditionally performed. First a pulse experiment where RNA is labeled and collected at desired times and enriched for newly transcribed RNA. A second type of experiment is one in which RNA is pulse labeled for longer time point and the metabolic label is taken away and RNA isolated to characterize RNA decay. **(c)** Chemical structures and derivatives of uridine[10].

RNA metabolic labeling is most often performed by treating cells with the two nucleoside analogs 4-thiouridine (4sU)[9] or bromo-uridine (BrU)[12] resulting in thiol-labeled or bromo-labeled transcripts, respectively, during the pulse experiment **(Figure 1c)**. Isolation of bromo-labeled RNA requires the use of anti-BrU antibodies which is more technically challenging to apply relative to isolation of thiol-labeled RNA using bioorthogonal reagents conjugated with biotin for streptavidin affinity-based biochemical enrichment. Alternatively, reacting the thiol-handle with other selective chemistries can generate cytidine mutations at the 4sU sites during reverse transcription that can then be

detected bioinformatically with RNA-sequencing [SLAM-seq[13], TimeLaspse-seq[14], thiouridine-to-cytidine conversion sequencing (TUC-seq[15])]. These nucleoside recoding strategies are of growing interest for using RNA metabolic labeling in cells and model organisms to analyze the nascent transcriptome, of which the first is commercially available to researchers[13].

In addition to 4-thiouridine, other uridine derivatives containing ethynyl groups have been used for RNA metabolic labeling applications, such as 5-ethynyluridine. Ethynyl-labeled RNA reacts selectively through copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) with biotin or fluorophore-containing azides for biochemical enrichment or fluorescent microscopy imaging of nascent RNA in cells and mouse tissues[16,17]. One clear advantage of using CuAAC bioorthogonal reactions for alkyne-labeled RNA is the relative stability of the covalent handle compared to reversible disulfide chemistry used with conjugating thiol-containing RNA[18]. However, CuAAC generates free radicals in solution thereby affecting RNA integrity in these reactions introducing advantages for, albeit less reactive, copper-free chemistries[19]. Finally, vinyl-nucleosides, such as 5-vinyluridine, are also applied for RNA metabolic labeling and can be reacted to contain biotin or fluorophores through inverse electron-demand Diels-Alder reactions with tetrazine conjugates. **(Figure 1c).** Contrary to 4-thiouridine[20] and 5-bromouridine[21] which are known to inhibit rRNA and mRNA processing, vinyl nucleosides demonstrate less toxicity in cells in comparison with thiol or ethynyl containing nucleoside derivatives[22]. These findings support the use of 5-vinyluridine for RNA metabolic labeling with minimal perturbance to nascent transcriptional profiling and for imaging nascent RNA in cells and animal tissues[23].

RNA metabolic labeling approaches have also been modified for cell-specific applications to profile nascent RNA within complex multicellular environments. These rely on the presence of an enzyme in a cell-type of interest that can selectively catalyze nucleoside or nucleobase derivatives with chemical handles into metabolic intermediates readily taken up by intercellular kinases and RNA polymerase into labeled RNA within that cell-type **(Figure 2).** The most widely applied method utilizes uracil phosphoribosyltransferase enzymes from the microorganism, *Toxoplasma gondii* (*Tg*UPRT) to selectively catalyze 4-thiouracil into 4-thiouridine monophosphate for downstream thiol-labeled RNA polymerization[24]. This method is known as "TU-tagging" and has been widely applied for *in vitro* and *in vivo* studies to profile nascent transcription

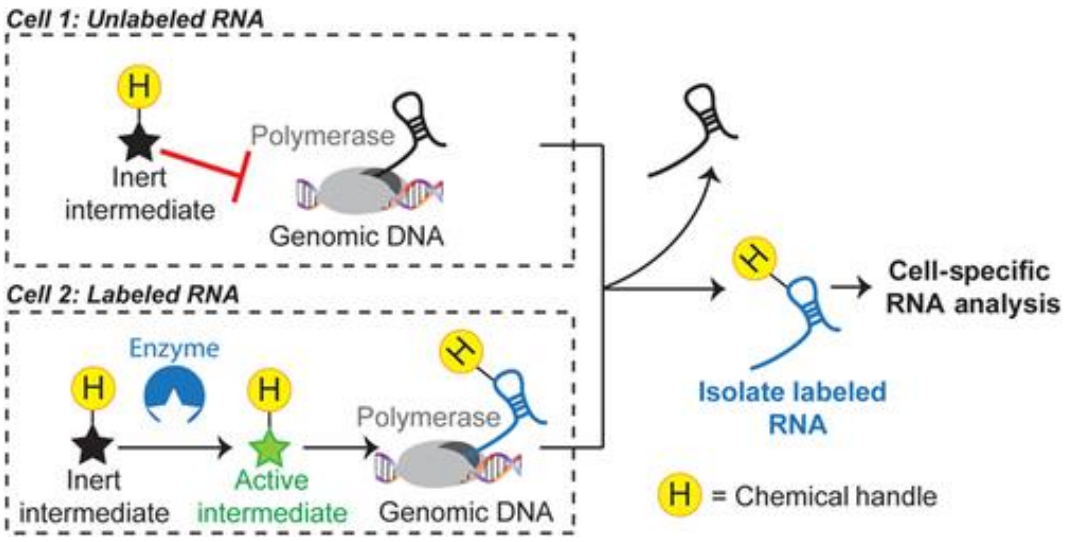


Figure 2 Schematic of cell-specific metabolic labeling of RNA. Cell 1 depicts a wild-type cell exposed to an inert intermediate that will not produce labeled RNA. Cell 2 is exposed to an inert intermediate that is activated by an exogenously expressed enzyme for subsequent metabolic incorporation by RNA polymerase producing labeled RNA. The chemical handle functionalizes RNA for isolation and analysis[10].

in cell types within their native environments enabling study of complex pathways in organisms[25-26].

Using this cell-specific approach, other metabolic labeling RNA reporters have been used in similarly complex cell-specific contexts to capture newly synthesized RNA in cell types that are rare or change their transcriptional profile after tissue dissociation with enzymatic, mechanical or flow cytometry-based techniques. To study cancer cells within their native environment, caged nucleosides were used for RNA labeling in cancer cell types expressing both endogenous enzymes, histone deacetylase and cathepsin L-protease, thereby facilitating the rapid de-caging of these reporters resulting in RNA metabolic labeling with 2'-azidoadenosine with high preference in mouse tumor tissues[27]. In a more generalizable strategy for mammalian cell types, exogenous expression of enzymes that are not active or rarely present in mammalian tissues can be exploiting for a wide range of applications in cells and whole organisms treated with metabolic reporter RNA substrates. For instance, EC-tagging relies on a cytidine deaminase and *Tg*UPRT (CD-UPRT), both which are not present or enzymatically active in mammalian salvage pathways, and proves useful for the two-step conversion of 5-ethynylcytosine into 5-ethynyluridine monophosphate for labeling RNA. Using CD-UPRT allowed for improved cell-specificity and stringency of RNA labeling rare cell-types expressed in *Drosophila melanogaster.*[28] Cell-specific labeling with *Tg*UPRT is advantageous due to enzymatic inactivity of high-order mammalian UPRT isoforms[29]. Due to the known toxicity and disulfide bond instability of 4-thiouracil, cell-specific RNA metabolic labeling with *Tg*UPRT expression was demonstrated with 5-ethynyluracil treatment demonstrating how these

applications are continuously improving by identification of new enzyme-analog pairs with defined bioorthogonal chemistries.

One approach to developing novel cell-specific RNA metabolic labeling methods is to engineer enzyme active sites to accommodate new metabolic labeling substrates. For instance, the Kleiner group developed mutant uridine-cytidine kinase 2 (UCK2) to accommodate 5-azidomethyl uridine substrates[31]. Through rapid screening of enzyme mutants with an array of metabolic labeling substrates, new cell-specific RNA metabolic labeling approaches can be introduced to the field for testing across different *in vitro* and *in vivo* applications. Recent findings in the Spitale lab support these developments by showing that 5-ethynyluracil acts as a substrate with uridine monophosphate synthase, a component of the *de novo* pyrimidine synthesis pathways that is active in mouse and human cell lines thereby limiting cell-specificity of RNA labeling[32]. These findings suggest the cause of background labeling in 4-thiouracil may be caused by this pathway, which explains the technical limitations of TU-tagging by time and concentration in these studies. Altogether, developing new enzyme-analog pairs for cell-specific RNA labeling will require continuous optimization however these current approaches illustrate important groundwork for developing new strategies to overcome these challenges.

As more metabolic reporters are discovered, synthesized, and developed for RNA metabolic labeling, more sophisticated options will become available for multiplexed RNA metabolic labeling in cells. The Akimitsu group combined bromo-uridine and 4-thiouridine to measure RNA kinetics simultaneously; they conclude that RNA turnover using both RNA metabolic labeling approaches provides useful information on the biological functions of regulatory RNA after applying genome-wide association analysis from both sets of

enriched transcripts. These results anticipate new developments for applying two (or more) nucleoside analogs with selective enrichment chemistries into cells for analysis of distinct RNA populations. Use of this strategy can infer causal relationships of RNA turnover during cellular perturbances such as inducing expression of genes or with drug treatment. As new orthogonal chemistries are developed for specific RNA derivative handles, there will be many new opportunities to analyze RNA mechanisms across synthetically induced and developmental changes in cells, eventually carrying out these methods into whole organism applications can be used for gaining deeper biological insight.

Summary

Different approaches for RNA metabolic labeling are developing rapidly as new chemistries, metabolic RNA reporters and applications are being applied to analyze nascent transcription in cells and animal models. As challenges to overcome technical limitations such as enrichment bias, non-cell-specific background labeling, and improved cell tolerance are achieved there will be growing opportunities to study RNA pathways within native, multicellular contexts. Pairing RNA metabolic labeling methods with RNA-sequencing and other types of high-throughput methods enables unique insights into these dynamic gene regulatory molecules. Future opportunities include developing transgenic mice with cell-specific RNA metabolic capabilities in tandem with variations of RIP-seq to analyze specific RNA subpopulations[34].

As advancements in genomics and model organisms expand for recapitulating human physiological and patho-physiological processes, applying these tools to tag nascent RNA will become increasingly important to define RNA regulatory pathways with

these contexts. For example, cancer metastasis continues to be a great challenge in treating patients although metastasis has been well-characterized by several transcription factors known to facilitate changes in cancer cell states during the course of migration to metastatic sites[35]. Insight into the order of events of how these transcription factors are regulated by interrogating each with metabolic RNA labeling reporters in these metastatically-destined cells could provide opportunities to understand these events in greater detail. By using cell-specific approaches in the metastatic cells or using multiple RNA analogs to label RNA for subsequent enrichment during specific stages of disease progression, new findings can potentially be used to identify novel therapeutic targets with greater specificity. As more biological targets become classified as "undruggable," there is an urgent need to find alternative methods to target molecules that can inhibit or prevent disease manifestation and limit the loss of life.

In conclusion, the purpose of this dissertation is to discuss the previous work established in the field of RNA metabolic labeling and to introduce new approaches. A comprehensive review is detailed in Chapter 2 which expands upon a few examples included in this brief introduction. The contents of Chapters 3-4 demonstrate new methods for cell-specific RNA metabolic labeling in cells and metastatic breast cancer in a mouse xenograft model. Finally, Chapter 5 will introduce a novel opportunity for chemo-selective reagents paired with vinyl-nucleosides for RNA metabolic labeling in cells with a dual-analog approach. These work detail the chemical, biological and computational efforts that are required for the continued development of RNA metabolic labeling tools for the study of nascent transcription dynamics.

References

1 Wang, K. C.; Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell* **2011**, *43* (6), 904–914.

2 Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A Revolutionary Tool for Transcriptomics. *Nat Rev Genet* **2009**, *10* (1), 57–63.

3 Turner, A. W.; Wong, D.; Khan, M. D.; Dreisbach, C. N.; Palmore, M.; Miller, C. L. Multi-Omics Approaches to Study Long Non-Coding RNA Function in Atherosclerosis. *Front. Cardiovasc. Med.* **2019**, *6*, 9.

4 Ergin, S.; Kherad, N.; Alagoz, M. RNA Sequencing and Its Applications in Cancer and Rare Diseases. *Mol Biol Rep* **2022**, *49* (3), 2325–2333.

5 Zhu, Y.; Zhu, L.; Wang, X.; Jin, H. RNA-Based Therapeutics: An Overview and Prospectus. *Cell Death Dis* **2022**, *13* (7), 644.

6 Tani, H.; Mizutani, R.; Salam, K. A.; Tano, K.; Ijiri, K.; Wakamatsu, A.; Isogai, T.; Suzuki, Y.; Akimitsu, N. Genome-Wide Determination of RNA Stability Reveals Hundreds of Short-Lived Noncoding Transcripts in Mammals. *Genome Res.* **2012**, *22* (5), 947–956.

7 Louloupi, A.; Ørom, U. A. V. Metabolic Pulse-Chase RNA Labeling for Pri-MiRNA Processing Dynamics. In *miRNA Biogenesis*; Ørom, U. A. V., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2018; Vol. 1823, pp 33–41.

8 Agarwal, S.; Vierbuchen, T.; Ghosh, S.; Chan, J.; Jiang, Z.; Kandasamy, R. K.; Ricci, E.; Fitzgerald, K. A. The Long Non-Coding RNA LUCAT1 Is a Negative Feedback Regulator of Interferon Responses in Humans. *Nat Commun* **2020**, *11* (1), 6348.

9 Biasini, A.; Marques, A. C. A Protocol for Transcriptome-Wide Inference of RNA Metabolic Rates in Mouse Embryonic Stem Cells. *Front. Cell Dev. Biol.* **2020**, *8*, 97.

10 Singha, M.; Spitalny, L.; Nguyen, K.; Vandewalle, A.; Spitale, R. C. Chemical Methods for Measuring RNA Expression with Metabolic Labeling. *WIREs RNA* **2021**, *12* (5).

11 Matsushita, K.; Takeuchi, O.; Standley, D. M.; Kumagai, Y.; Kawagoe, T.; Miyake, T.; Satoh, T.; Kato, H.; Tsujimura, T.; Nakamura, H.; Akira, S. Zc3h12a Is an RNase Essential for Controlling Immune Responses by Regulating MRNA Decay. *Nature* **2009**, *458* (7242), 1185–1190.

12 Yamada, T.; Imamachi, N.; Onoguchi-Mizutani, R.; Imamura, K.; Suzuki, Y.; Akimitsu, N. 5′-Bromouridine IP Chase (BRIC)-Seq to Determine RNA Half-Lives. In *mRNA Decay*; Lamandé, S. R., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2018; Vol. 1720, pp 1–13.

13 Herzog, V. A.; Reichholf, B.; Neumann, T.; Rescheneder, P.; Bhat, P.; Burkard, T. R.; Wlotzka, W.; von Haeseler, A.; Zuber, J.; Ameres, S. L. Thiol-Linked Alkylation of RNA to Assess Expression Dynamics. *Nat Methods* **2017**, *14* (12), 1198–1204..

14 Schofield, J. A.; Duffy, E. E.; Kiefer, L.; Sullivan, M. C.; Simon, M. D. TimeLapse-Seq: Adding a Temporal Dimension to RNA Sequencing through Nucleoside Recoding. *Nat Methods* **2018**, *15* (3), 221–225.

15 Lusser, A.; Gasser, C.; Trixl, L.; Piatti, P.; Delazer, I.; Rieder, D.; Bashin, J.; Riml, C.; Amort, T.; Micura, R. Thiouridine-to-Cytidine Conversion Sequencing (TUC-Seq) to Measure MRNA Transcription and Degradation Rates. In *The Eukaryotic RNA Exosome*; LaCava, J., Vaňáčová, Š., Eds.; Methods in Molecular Biology; Springer New York: New York, NY, 2020; Vol. 2062, pp 191–211.

16 Es-Haghi, M.; Godakumara, K.; Häling, A.; Lättekivi, F.; Lavrits, A.; Viil, J.; Andronowska, A.; Nafee, T.; James, V.; Jaakma, Ü.; Salumets, A.; Fazeli, A. Specific Trophoblast Transcripts Transferred by Extracellular Vesicles Affect Gene Expression in Endometrial Epithelial Cells and May Have a Role in Embryo-Maternal Crosstalk. *Cell Commun Signal* **2019**, *17* (1), 146.

17 Jao, C. Y.; Salic, A. Exploring RNA Transcription and Turnover *in Vivo* by Using Click Chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (41), 15779–15784.

18 Duffy, E. E.; Rutenberg-Schoenberg, M.; Stark, C. D.; Kitchen, R. R.; Gerstein, M. B.; Simon, M. D. Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. *Molecular Cell* **2015**, *59* (5), 858–866.

19 Paredes, E.; Das, S. R. Click Chemistry for Rapid Labeling and Ligation of RNA. *ChemBioChem* **2011**, *12* (1), 125–131.

20 Burger, K.; Mühl, B.; Kellner, M.; Rohrmoser, M.; Gruber-Eber, A.; Windhager, L.; Friedel, C. C.; Dölken, L.; Eick, D. 4-Thiouridine Inhibits RRNA Synthesis and Causes a Nucleolar Stress Response. *RNA Biology* **2013**, *10* (10), 1623–1630.

21 Sierakowska, H.; Shukla, R. R.; Dominski, Z.; Kole, R. Inhibition of Pre-MRNA Splicing by 5-Fluoro-, 5-Chloro-, and 5-Bromouridine. *Journal of Biological Chemistry* **1989**, *264* (32), 19185–19191.

22 Kubota, M.; Nainar, S.; Parker, S. M.; England, W.; Furche, F.; Spitale, R. C. Expanding the Scope of RNA Metabolic Labeling with Vinyl Nucleosides and Inverse Electron-Demand Diels–Alder Chemistry. *ACS Chem. Biol.* **2019**, *14* (8), 1698–1707.

23 Liu, H.; Ishizuka, T.; Kawaguchi, M.; Nishii, R.; Kataoka, H.; Xu, Y. A Nucleoside Derivative 5-Vinyluridine (VrU) for Imaging RNA in Cells and Animals. *Bioconjugate Chem.* **2019**, *30* (11), 2958–2966.

24 Miller, M. R.; Robinson, K. J.; Cleary, M. D.; Doe, C. Q. TU-Tagging: Cell Type–Specific RNA Isolation from Intact Complex Tissues. *Nat Methods* **2009**, *6* (6), 439–441.

25 Erickson, T.; Nicolson, T. Identification of Sensory Hair-Cell Transcripts by Thiouracil-Tagging in Zebrafish. *BMC Genomics* **2015**, *16* (1), 842.

26 Gay, L.; Miller, M. R.; Ventura, P. B.; Devasthali, V.; Vue, Z.; Thompson, H. L.; Temple, S.; Zong, H.; Cleary, M. D.; Stankunas, K.; Doe, C. Q. Mouse TU Tagging: A Chemical/Genetic Intersectional Method for Purifying Cell Type-Specific Nascent RNA. *Genes Dev.* **2013**, *27* (1), 98–115.

27 Beasley, S.; Vandewalle, A.; Singha, M.; Nguyen, K.; England, W.; Tarapore, E.; Dai, N.; Corrêa, I. R.; Atwood, S. X.; Spitale, R. C. Exploiting Endogenous Enzymes for Cancer-Cell Selective Metabolic Labeling of RNA in Vivo. *J. Am. Chem. Soc.* **2022**, *144* (16), 7085–7088.

28 Hida, N.; Aboukilila, M. Y.; Burow, D. A.; Paul, R.; Greenberg, M. M.; Fazio, M.; Beasley, S.; Spitale, R. C.; Cleary, M. D. EC-Tagging Allows Cell Type-Specific RNA Analysis. *Nucleic Acids Research* **2017**, *45* (15), e138–e138

29 Li, J.; Huang, S.; Chen, J.; Yang, Z.; Fei, X.; Zheng, M.; Ji, C.; Xie, Y.; Mao, Y. Identification and Characterization of Human Uracil Phosphoribosyltransferase (UPRTase). *J Hum Genet* **2007**, *52* (5), 415–422.

30 Nguyen, K.; Fazio, M.; Kubota, M.; Nainar, S.; Feng, C.; Li, X.; Atwood, S. X.; Bredy, T. W.; Spitale, R. C. Cell-Selective Bioorthogonal Metabolic Labeling of RNA. *J. Am. Chem. Soc.* **2017**, *139* (6), 2148–2151.

31 Zhang, Y.; Kleiner, R. E. A Metabolic Engineering Approach to Incorporate Modified Pyrimidine Nucleosides into Cellular RNA. *J. Am. Chem. Soc.* **2019**, *141* (8), 3347–3351.

32 Nainar, S.; Cuthbert, B. J.; Lim, N. M.; England, W. E.; Ke, K.; Sophal, K.; Quechol, R.; Mobley, D. L.; Goulding, C. W.; Spitale, R. C. An Optimized Chemical-Genetic Method for Cell-Specific Metabolic Labeling of RNA. *Nat Methods* **2020**, *17* (3), 311–318.

33 Kawata, K.; Wakida, H.; Yamada, T.; Taniue, K.; Han, H.; Seki, M.; Suzuki, Y.; Akimitsu, N. Metabolic Labeling of RNA Using Multiple Ribonucleoside Analogs Enables the Simultaneous Evaluation of RNA Synthesis and Degradation Rates. *Genome Res.* **2020**, *30* (10), 1481–1491.

34 Kaczmarczyk, L.; Bansal, V.; Rajput, A.; Rahman, R.; Krzyżak, W.; Degen, J.; Poll, S.; Fuhrmann, M.; Bonn, S.; Jackson, W. S. Tagger—A Swiss Army Knife for Multiomics to Dissect Cell Type–Specific Mechanisms of Gene Expression in Mice. *PLoS Biol* **2019**, *17* (8), e3000374.

35 Georgakopoulos-Soares, I.; Chartoumpekis, D. V.; Kyriazopoulou, V.; Zaravinos, A. EMT Factors and Metabolic Pathways in Cancer. *Front. Oncol.* **2020**, *10*, 499.

Chapter 1: Chemical Methods for Measuring RNA Expression with Metabolic Labeling

Introduction

As the list of RNA functions expands from its first proposed role in the central dogma, the need to study RNA in depth continues to necessitate the development of novel methods designed to identify RNA expression profiles from specific cell types to better understand how RNA molecules contribute to cellular identity[1-2]. Specific profiles of RNA expression in different cell types, both coding and non-coding, is a clear contributor to cellular function, thus, further characterizing RNA expression is critical to understanding what makes cells unique and how they respond to stimuli in order to control their function[3-4]. Cell sorting is a technique where cells can be separated by the expression of an endogenous fluorescent protein or through the affinity of fluorescently conjugated antibodies towards cell-specific surface markers. Although useful in isolating specific cell types, sorting can be extremely problematic as it often leads to the isolation of undesired cell types, relies upon conditions that can cause degradation of RNA[5], or puts undue stress on isolated cells resulting in transcriptional changes that do not represent the intact state of a cell *in vivo*[6]. Finally, cell sorting is incapable of ascertaining the dynamics of RNA expression, and the mixture of RNAs characterized by RNA-sequencing or reverse transcription-quantitative polymerase chain reaction consists of newly transcribed RNA and those RNAs that are at a steady state.

More sophisticated methods for isolation of cell-specific RNA have been developed. In one example, ribosomal proteins can be tagged in the form of fusion proteins expressed in a cell type of interest[7-8]. Although useful in identifying mRNAs that may be undergoing

translation, these techniques are not designed to capture non-coding RNAs. In addition, isolation of ribosome-bound RNAs occurs from cell lysate mixtures, the conditions of which have been demonstrated to cause re-association of RNA complexes that are not present within intact cells or tissues[9-10]. In another scenario, the poly(A)-binding protein (PAB), which binds to mRNA poly(A) tails, can be tagged and expressed in a cell-specific manner[11]. Again, this approach is useful in isolating mRNAs and non-coding RNAs which contain poly(A) tails; however, many of the biologically important non-coding RNAs are not included after purification. Similarly, PAB purification of associated RNAs is also done from a complex lysate mixture and is therefore susceptible to post-lysis re-association in RNA purification leading to a nonrepresentative interpretation of the in-cell state[9-10]. Lastly, just as in cell sorting, these methods do not report newly transcribed RNA or RNA expression dynamics. Overall, the associated challenges of RNA isolation have necessitated more sophisticated and perhaps more stringent methods to characterize RNA expression from cell-specific populations.

To address such gap, biochemical tools have been developed to be used in RNA metabolic labeling approach enabling analysis of nascent RNA expression in a cell-specific manner[12]. In RNA metabolic labeling, a chemically modified and orthogonal nucleoside (or nucleobase) metabolic precursor is added to cells and processed by endogenous machinery to form a triphosphate which can be incorporated into nascent RNA through transcription[13-19]. To extend these approaches to be cell-specific, researchers have identified chemically modified RNA analogs which are not processed by the endogenous cellular machinery ("inert metabolic intermediates")[13, 20-22]. These discoveries have opened the possibility to engineer cells with modified metabolic enzymes to transform

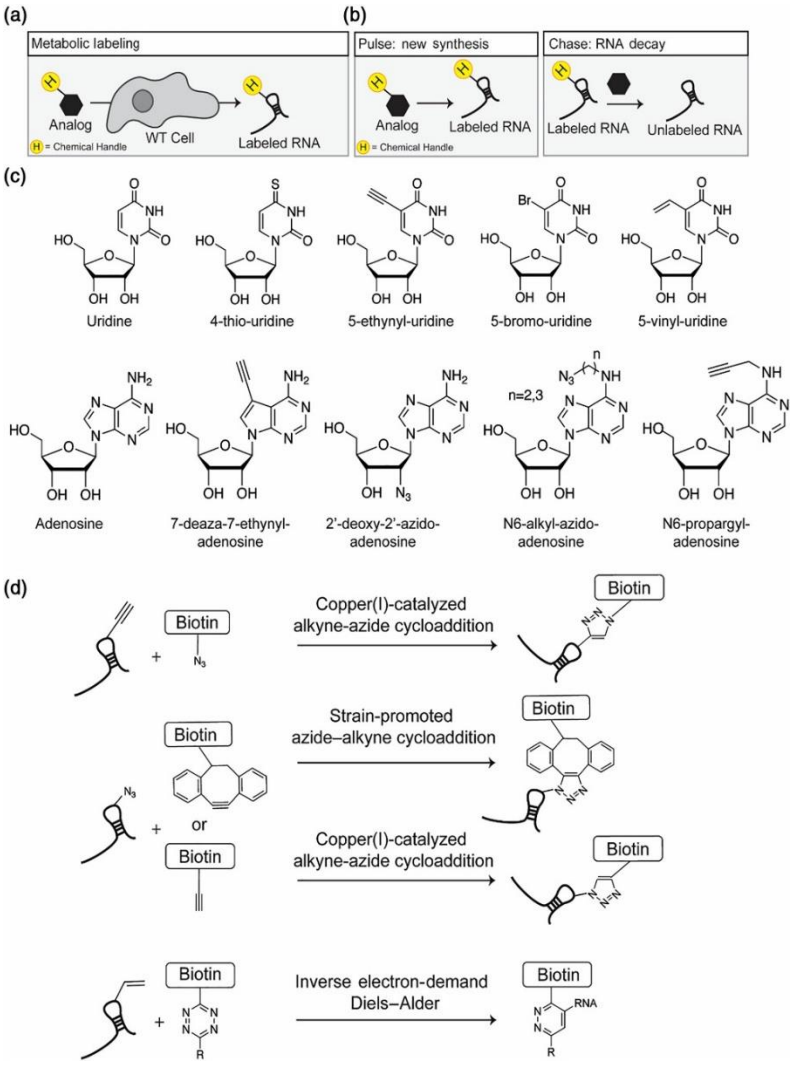


Figure 1-1 Chemical probes for metabolic labeling of RNA. (a) Schematic of metabolic labeling experiments. (b) Schematic of pulse and chase experiments: two classes of experiments are traditionally performed. First a pulse experiment where RNA is labeled and collected at desired times and enriched for newly transcribed RNA. A second type of experiment is one in which RNA is pulse labeled for longer time point and the metabolic label is taken away and RNA isolated to characterize RNA decay. (c) Chemical structures and derivatives of adenosine and uridine. (d) Bioorthogonal chemical approaches to label metabolically labeled RNA and append biotin groups for

"inert metabolic intermediates into active metabolic intermediates which can then be readily incorporated into RNA by polymerases. Here, I will highlight the efforts in RNA metabolic labeling studies.

Discussion of Approaches for RNA Metabolic Labeling

RNA metabolic labeling experiments are essential tools for understanding how cells respond to stimuli, stress, and signaling[23]. Pulse-chase experiments rely on introducing cells to an active metabolic intermediate that has a chemical handle for enrichment **(Figure 1-1a)**. To track transcription and decay rates of RNA, there are two main experiments employed: pulse and chase. During the pulse, a nucleoside analog is added to the cellular medium and the RNA is isolated at specific time intervals **(Figure 1-1b)**. Labeled RNA can be enriched from unlabeled RNA which are characterized as being synthesized within the specified time interval. In a chase experiment, a nucleoside analog is added to the medium for a longer pulsing time to ensure near-saturated incorporation. Then, the media is exchanged for media lacking the modified nucleoside analog (the chase) **(Figure 1-1b)**. Cells are then harvested at given times for RNA isolation and enrichment. The RNA that is enriched at the chase time points is the RNA which still retains the metabolic label indicating these are more stable RNA molecules with longer half-lives. Conversely, RNA that are only enriched at early chase times are less stable with shorter half-lives. Together, these approaches, although routine, are powerful and practical experimental procedures for characterizing RNA dynamics with metabolic labeling.

Cell-specific RNA metabolic labeling

The capability to profile cell-specific, genome-wide RNA expression can be achieved through an array of chemo-enzymatic strategies compatible with RNA-sequencing. These

techniques rely on cell-specific overexpression of metabolic enzymes capable of producing activated metabolic RNA intermediates within subpopulations of cells exposed to inert metabolic intermediates **(Figure 1-2a)**. As a result, labeled transcripts within can be isolated, purified, and identified with RNA-sequencing. To control the expression of metabolic enzymes, genetically encoded *cis*-acting elements such as cell-specific or condition-specific promoters and enhancers are used to determine the timing and extent of RNA labeling in cells. The selectivity of labeling depends on cell-specific enzyme expression plus the stability and availability of their corresponding RNA analog substrates. To reduce the potential of nonspecific cell labeling, researchers have synthesized novel nucleoside and nucleobase analogs as well as engineered metabolic enzymes guided by structural analysis to devise new approaches for cell-specific RNA metabolic labeling.

Expression of specific enzymes in a cell-type can be introduced by recombinant viruses or generation of transgenic model organisms. The first widely used example of this approach was to utilize the *Toxoplasma gondii* nucleotide salvage enzyme, uracil phosphoribosyltransferase (*Tg*UPRT). 4-thio-uracil (4TU) was identified as a suitable substrate for *Tg*UPRT to produce 4-thio-uridine (4SU) with 1% thiouridine substitution rates in nascent RNA. The capability to label newly synthesized RNA in *Tg*UPRT-expressing cells is possible because higher eukaryotic UPRT enzymes were not known to react with 4TU[14]. Using 4TU to tag (TU-tagging) RNAs has several limitations that reduce the stringency of cell specificity. First, there is detectable background RNA labeling with 4TU by UPRT-independent pathways in eukaryotic cells which could be minimized to some extent by experimental design, such as including lower concentrations of the probe and duration of the treatment. In *Drosophila melanogaster*, an active UPRT homolog was

identified thereby limiting the application of TU-tagging in this important model system[24]. Second, RNA enrichment depends on the purification of thiol-containing transcripts which is limited by the stability of disulfide bonds and reagents for enrichment chemistries, whereas other orthogonal handles demonstrate higher chemical stability and more efficient conjugation[25-26]. Finally, the inability to visualize 4TU or 4SU labeled transcripts in cell imaging and live animals is hindered by the widespread abundance of endogenous thiol-containing molecules limiting TU-tagging for RNA localization studies.

Other uracil analogs, such as 5-ethynyl-uracil (5EU) and 5-vinyl-uracil (5VU), are modified with different bioorthogonal handles that enable cellular imaging and more stable biochemical purification which is important for accurate and reliable identification of cell-specific nascent RNAs for analysis of transcriptional dynamics within living cells[19,22,27].

In order to develop new cell-specific metabolic RNA chemical reporters, it is essential to identify a variety of orthogonal and specific enzyme-analog pairs to label eukaryotic transcripts. By investigating the co-crystal structure of *Tg*UPRT binding to uracil, one study identified that 5-ethynyl-uracil (5EU) can be a substrate **(Figure 1-2b)**[22]. The 5-EU/UPRT system provided a useful technique to observe stimulus-induced transcriptome changes in neurons. The reduced labeling time window and elimination of constitutively expressed steady state transcripts improve the identification of transcriptional changes in response to neural stimulation. Using a neuronal

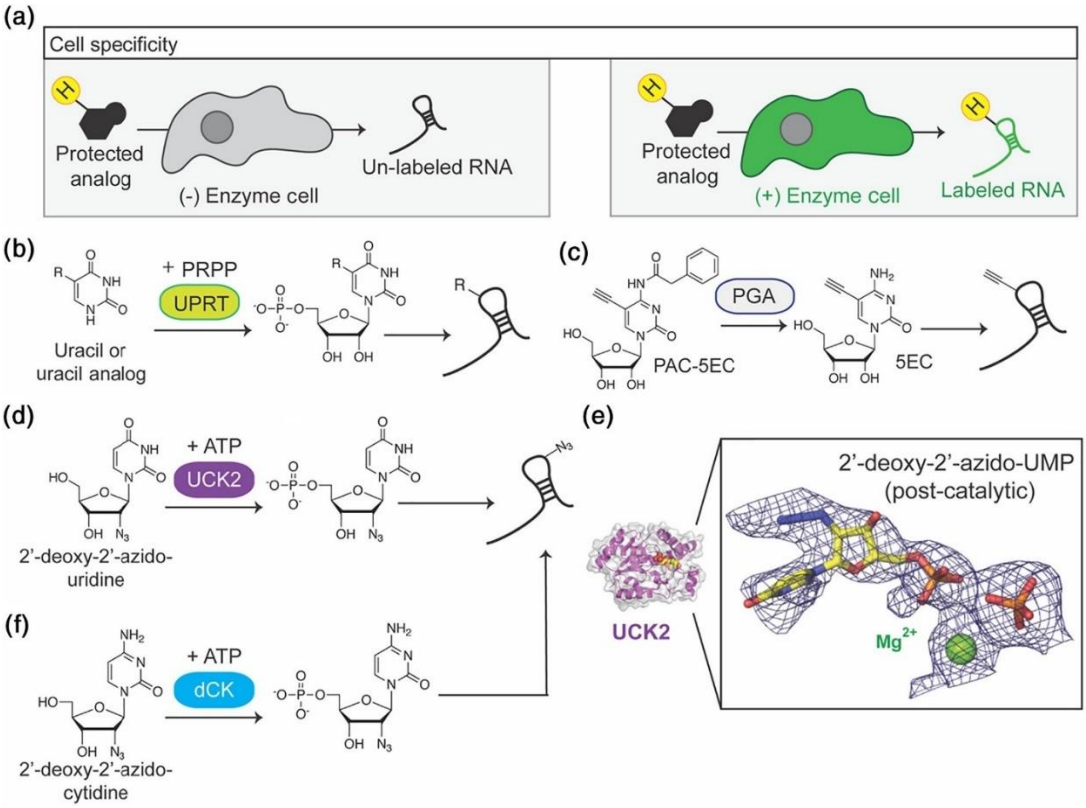


Figure 1-2 Chemical probes for cell-specific metabolic labeling of RNA. (a) Schematic of cell-specific metabolic labeling experiments. (b) Example reaction between uracil or uracil analogs reacting with uracil phosphoribosyltransferase (UPRT). (c) "Caged" cytidine analogs can be "uncaged" by penicillin G amidase to liberate 5-ethynyl-cytidine (5EC) for subsequent incorporation into RNA. (d) 2′-azido-uridine can be phosphorylated by uridine/cytidine kinase 2 (UCK2) for RNA metabolic labeling. (e) Co-crystal structure of UCK2 bound to 5′-phosphorylated 2′-azido-uridine (the kinase product. PDB ID for this structure is 6N53). (f) Chemical reaction of 2′-deoxy-2′azido-cytidine with deoxynucleotidyl kinase (dCK) for cell-specific metabolic incorporation into RNA.

specific promoter to express *Tg*UPRT, the Bredy group identified nascent transcripts in KCl-stimulated neurons treated with 5-EU for 3 h, reducing the time window to label transcripts and improving the temporal specificity for these fast-acting cell processes that regulate synaptic plasticity, memory and learning[27]. While useful, it was demonstrated that mammalian[28] and *D. melanogaster*[21,24] cells exhibit non-cell specific background incorporation of uracil analogs requiring higher stringency of RNA labeling for these approaches.

5-Ethynyl-cytosine tagging (EC-tagging) was developed to improve the specificity of cell-specific RNA metabolic labeling by expression of a dual-enzyme labeling strategy. The EC-tagging method requires expression of both cytidine deaminase (CD), expressed only in yeast and bacteria, and *Tg*UPRT to convert 5-EC into 5-ethynyl-uridine (5-EU) for incorporation into nascent RNA transcripts in cells expressing this gene fusion. Although there is some background labeling with the 5-ethynyl-uracil intermediate through metabolic pathways unrelated to UPRT, the co-expression of CD-UPRT drastically enhances the labeling signal[21]. These results suggest that low levels of background RNA labeling can be accomplished through developing novel enzymatic approaches that selectively control the metabolic flux of RNA labeling intermediates.

Previous work established the incorporation of 2′-AzA directly by RNA polymerase for preferential labeling of mRNAs and other polyadenylated transcripts which are enriched in adenosine nucleotides[19]. Addition of a bulky group at the N6 position inhibits 2′-AzA catalysis with RNA polymerase leading to the development of N6-phenylacetyladenosine-2′-AzA (N-PAC-2′-AzA), a “caged” nucleoside. The bulky group can

be removed by the bacteria-specific enzyme penicillin G amidase resulting in the "de-caged" 2′-AzA that can readily be incorporated into nascent RNA through RNA polymerase **(Figure 1-2c)**[22]. While useful, azido-containing caged small molecules are prone to spurious hydrolysis potentially introducing background issues with amide-caged nucleosides that are on electron-withdrawing nucleobases. Further exploration of caged analogs *in vivo* is needed to improve upon the cell-specificity of RNA labeling in complex and harsher environments that extend beyond cell culture experiments.

One of the major obstacles for cell-specific metabolic RNA labeling in complex environments, such as within higher organisms, is the potential of alternative pathways capable of metabolizing analogs; however, work to identify the biochemical mechanisms of background metabolic labeling is seldomly done. Background RNA metabolic labeling is a challenge presented when utilizing uracil analogs due to endogenous metabolic pathways, and this off-target RNA labeling can lead to false discoveries of nascent transcripts in the cell-type of interest. Investigating the source of background labeling was attributed to the de novo pyrimidine biosynthetic enzyme uridine 5′-monophosphate synthase (UMPS), in which background RNA labeling is increased with UMPS overexpression in cells treated with 5-EU and significantly decreased with siRNA targeting UMPS in human immortalized cell lines[28]. To some extent, human UPRT contributed to background labeling when overexpressed as well, which contradicts the findings that human UPRT enzymes are inactive[24]. As an alternative to using uracil analog labeling systems, novel pyrimidine analogs were developed that can label the transcriptome without bias if introduced into transcripts by a highly selective metabolic enzyme, such as those in the pyrimidine triphosphate pathways. The development of 2′-azido-uridine was found to be highly

specific for uridine/cytidine kinase-2 (UCK2), a rarely expressed isoform of ubiquitously expressed uridine/cytidine kinase-1 (UCK1). At low micromolar concentrations, cells with overexpression of UCK2 will selectively incorporate the nontoxic 2′-azido-uridine analogs at a ~1% substitution rate per uracil residue **(Figure 1-2d)**. The identification of this novel labeling technique was confirmed with extensive structural, functional, and cell-based studies that support high specificity between the enzyme-analog pair with comparable labeling efficiencies to uracil analogs in the *Tg*UPRT system, yet with undetectable background levels[28]. Another strategy reliant on UCK2 structure-based mutational engineering improved the enzyme's substrate tolerance towards 5-methylazidouridine and other pyrimidines with bulky 5-position groups[29] **(Figures 1-2d, 1-2e)**. This system extends the toolset for cell-specific RNA labeling in pulse-chase experiments using fluorescent microscopy to visualize RNA localization dynamics under stressed cell conditions.

Researchers have begun to extend beyond RNA synthesis enzymes for the incorporation of azido-containing RNA analogs. A recent manuscript from the Kleiner group demonstrates the utility of pairing 2′-azido-cytidine with deoxynucleotidyl kinase (dCK) as a robust strategy for cell-specific and polymerase-specific RNA labeling at high percentages (0.3% of C)[30] **(Figure 1-2f)**. However, 2′-azido-cytidine is also incorporated into cells with no overexpression of dCK[28,30]. These results mirror our observations that 2′-azido-cytidine, and not 2′-azido-uridine, is incorporated into RNA. Importantly, this study consistently demonstrates that 2′-modified nucleoside analogs have dramatically lower toxicity than the commonly used 4-thio-uridine. Collectively, these results present exciting

opportunities for continued cellular and enzymatic engineering to refine these methods for controlling cell-specific metabolic labeling of RNA.

Chemical methods to track analog incorporation into nascent RNA

Examining RNA expression dynamics through these versatile metabolic RNA labeling strategies relies on introduction of chemical handles into RNA. These chemical handles are then used to append biotin with orthogonal chemical reactions for subsequent enrichment with streptavidin magnetic beads. RNA is then eluted and can be converted into cDNA for preparing RNA-sequencing libraries. Although useful, it has been demonstrated that these approaches can introduce technical biases, due to varying enrichment conditions and challenges with elution of biotinylated RNAs[25,28]. To overcome these obstacles, researchers have been exploring novel chemical approaches that enable tracking of labeled RNA without the need for enrichment.

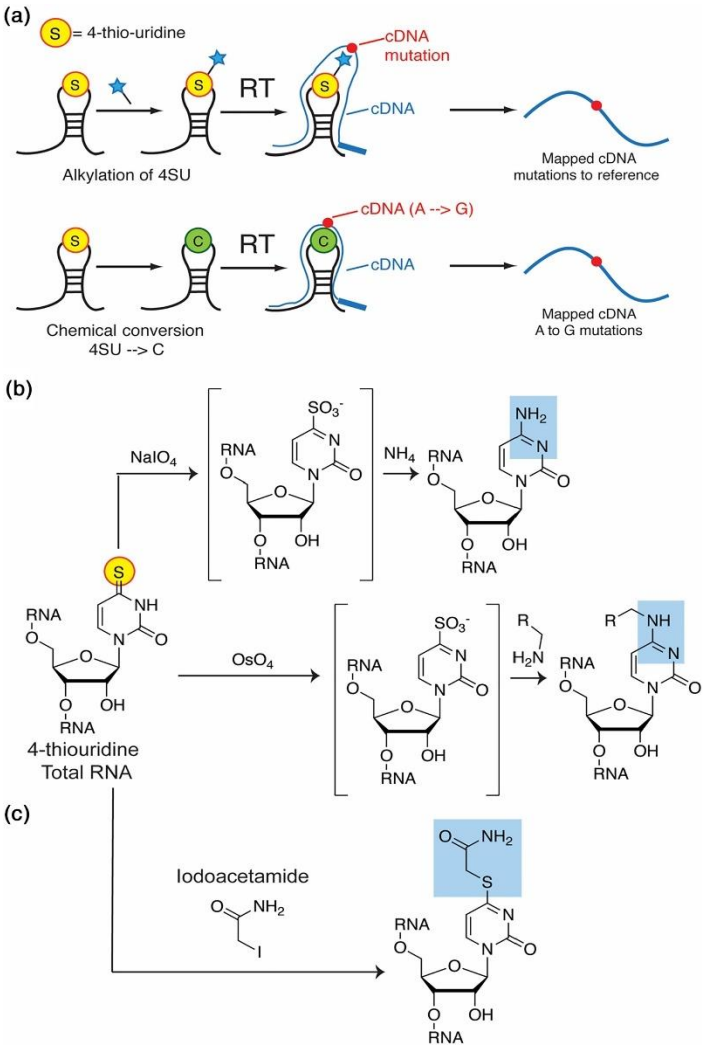


Figure 1-3 Nucleoside recoding to track nascent RNA synthesis. (a) Schematic of step-wide approach of nucleoside recoding for tracking nascent RNA synthesis. (b) Different methods for conversion of 4SU to cytidine or alkyl cytidine catalyzed by $NaIO_4$ or OsO_4, respectively. (c) Iodoacetamide mediated conversion of 4SU to 4-thiol (SH)-linked alkyl uridine. This modification is identified as cytidine after reverse transcription.

Exploiting the chemical reactivity of the thiol-modified nucleoside 4SU, new approaches are beginning to emerge to track nascent RNA expression. With conventional RNA-sequencing protocols, reverse transcription (RT) of enriched RNA is used to create cDNA libraries to map the expression and primary sequence RNAs to a reference transcriptome. As 4SU-containing transcripts are enriched and eluted, the 4SU is read during RT as a regular U residue and as such an adenosine is incorporated through normal Watson–Crick (W–C) pairing. Specific chemical modification protocols enable the transformation of 4SU into other nucleotide identities that can be identified by mapping U-mismatches. These mismatches can come from either chemical modification of 4SU to create cytidine mutations in the RNA or through the use of chemical adducts that cause RT enzymes to re-code the paired nucleoside during RT **(Figure 1-3a)**. Nucleotide re-coding reduces the additional technical steps of biochemical separation and enrichment, and instead utilizes chemical conversion of uridine to cytidine to subsequently identify the resulting mutations through RNA-sequencing methods and analysis.

The first approach to employ nucleoside recoding was to exploit the known oxidation potential of the 4SU thione, which had been shown previously to be converted from a thione to an amine following oxidation in the presence of ammonia. TUC-Seq **(Figure 1-3b)** developed by Riml et al. utilizes osmium-mediated 4SU conversion to cytidine in aqueous ammonia in 4 h with 98% efficiency[31,32]. The conversion of U to C in the sequencing reads enables tracking of the analog incorporation of 4SU and as such, labeled nascent transcripts. This approach is the most straightforward and effective chemical method for direct conversion and tracking of 4SU through nucleoside recoding. A similar approach was developed with TimeLapse-seq[33] where oxidative-nucleophilic-

aromatic substitution with sodium periodate ($NaIO_4$) and 2,2,2-trifluoroethylamine was employed to recode 4SU into an alkyl-cytidine analog with 80% conversion in 1 h **(Figure 1-3b)**. TimeLapse-seq mediated alkyl amine substitutions still preserve the W–C face of cytidine, and therefore also tracks cytidine for sequencing analysis.

As an alternative, employing the inherent nucleophilicity of 4SU thione is being utilized to place a "blemish" at the thione sulfur, which serves to disrupt W–C pairing. Through development of SLAM-seq, Herzog et al. reported the first application of this nucleoside recoding by alkylating 4SU with iodoacetamide to produce a cytidine analog in their thiol (SH)-linked alkylation for the metabolic sequencing of RNA **(Figure 1-3c)** with 98% conversion in 15 min[34]. These approaches were the first to detail the nucleoside recoding approach for monitoring RNA transcription and thereby enabling more precise measurements to evaluate RNA lifetimes which were previously undescribed. In summary, nucleotide re-coding methods that build on the techniques of 4SU metabolic labeling provide an efficient and simple means of monitoring RNA dynamics for a wide array of biochemical and sequencing applications both *in vitro* and *in vivo*.

Biological Insight

The examples discussed in this review primarily focused on developing the tools necessary for metabolic RNA labeling studies, including the chemical and biochemical methods for tracking and RNA isolation. These approaches have been merged to understand and characterize RNA expression *in vitro* and even more recently *in vivo*.

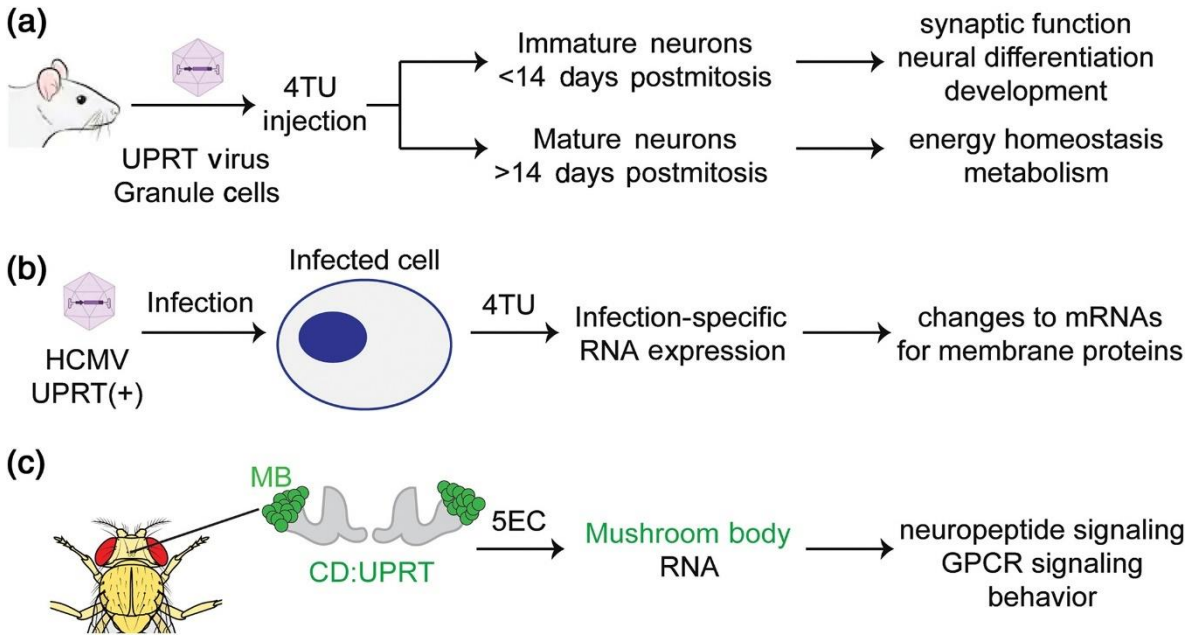


Figure 1-4 Biological applications for cell-specific metabolic labeling of RNA. (a,b) Application of 4TU-UPRT (TU-tagging) system for profiling differential RNA expression in neural development and human cytomegalovirus latency. (c) Utilization of 5EC-incorporating CD-UPRT fusion system to study signaling behavior and characterize transcriptomes of rare cell populations in *Drosophila melanogaster*.

Cell-specific metabolic RNA labeling has been employed to analyze important aspects of disease progression and development in different biological contexts which are difficult to interrogate in live model organisms. TU-tagging was initially applied in mice to

characterize RNA expression in brain endothelial cells[35], but has also been utilized to identify expression programs of newborn granule neurons in the dentate gyrus of the adult mouse hippocampus, which are important for learning and memory[36]. Mature and immature granule cell populations were selectively infected with lentiviral and retroviral constructs expressing *Tg*UPRT. TU-tagging in the immature neurons followed by RNA-sequencing of enriched transcripts revealed fold expression changes of nascent transcripts involved in neurogenesis, canonical and noncanonical Wnt signaling, and long-term potentiation required for the proper synaptic function in mature neurons. The nascent transcriptome of these mature neurons was enriched in metabolic genes, which is expected for cells that require extensive energy input to relay synaptic transmission through complex dendritic processes **(Figure 1-4a)**. As viral injection techniques are an invasive surgery and not suitable for all ages of mice, another group explored TU-tagging in the post-natal mouse visual cortex using the transgenic mouse method (from the original mouse TU tagging paper[35]; crossed with Cre-specific lines for two separate layers of the mouse visual cortex[37]. Upon subcutaneous injection of 4TU in the transgenic mouse lines, the workflow was optimized to increase RNA yields sufficient for RNA-sequencing and analysis to identify enrichment of developmentally significant transcripts specific to each layer of the mouse visual cortex by direct comparison of sequencing reads.

Metabolic labeling has also been employed to understand infection. TU-tagging was employed to label nascent RNA in human cytomegalovirus (HCMV) infected Kasumi-3 cells and CD34+ hematopoietic progenitors to monitor and compare host transcriptional changes in latent stages of infection[38]. In this model, a recombinant HCMV was generated with a bacterial artificial chromosome containing a fluorescent marker driven by a lytic

gene promoter and a latent gene promoter for UPRT expression. As a result, lytic infected cells could be separated from uninfected and latently infected cells by fluorescent-activated cell sorting. Latent infected cells exposed to 4TU labeled nascent RNA and transcriptional changes in uninfected cells are observed only by treatment with 4TU **(Figure 1-4b)**. An interesting observation from these experiments was the identification of differentially expressed mRNAs encoding for membrane-bound proteins, which may be a signal for changes in membrane protein content due to infection. These proof-of-concept experiments to characterize latently infected host cell transcriptomes overcome the barrier to study small populations of infected cells in a natural environment where it is difficult to separate infected cells by stage of infection or from uninfected cells in the bone marrow.

EC-tagging was applied in transgenic UAS-CD-UPRT flies exposed to 5-ethynyl-cytidine at the time of treatment enabling characterization of rare cell populations in the *D. melanogaster* nervous system[21]. EC-tagging employed cell type-specific gene expression data from intact *D. melanogaster* larvae, including transcriptome measurements from a small population of central brain neurons. EC-tagging successfully enriched for rare mushroom body mRNAs (from ~0.07% of total larvae cells) from a mixture of all larval mRNAs **(Figure 1-4c)**, whereas TU-tagging failed to identify mushroom body mRNAs in parallel experiments. The lack of specificity in our mushroom body TU-tagging experiments was likely due to widespread RNA-tagging via endogenous *D. melanogaster* UPRT activity. While data sets from 12B08 EC-tagging and mushroom body EC-tagging enriched for neural transcripts, comparing the data from each of these Gal4 lines allowed us to distinguish broadly expressed neural genes from mushroom body-specific neural genes.

These data should prove useful for identifying novel mushroom body properties, as demonstrated by our discovery of *pickpocket*-expressing mushroom body neurons. The discovery of these neurons reveals previously unknown cellular heterogeneity in the larval mushroom body and suggests that the *pickpocket*-expressing neurons respond to modalities that are distinct from those previously described in the mushroom body.

Conclusions

The examples highlighted in this review describe the flexibility of developing metabolic labeling methods and how their evolution is starting to bear fruit for novel biological insight into RNA expression. However, there is still a lot of work to be done in designing novel reagents and also protocol development. For example, although nucleoside recoding is very powerful in tracking RNA expression and reporter introduction into nascent transcripts, it is still limited in complex pools because the RNAs are not enriched. As such, employing this approach *in vivo* on rare cell populations will continue to be difficult. Expanding chemical approaches that enable enrichment and nucleoside recoding will be very powerful as they will enable enrichment for rarer-cell RNAs, but at the same time preserve the power of nucleoside recoding.

Another part of cell-specific metabolic RNA labeling methods that is ripe for improvement is designing more sophisticated enzymes and genetically-encoded controls for enzyme expression that can be utilized for tighter control over incorporation of metabolic probes. In EC-tagging it was hinted that split enzyme approaches can be explored through the demonstration that a split CD was reconstituted for eventual 5-ethynyl-uracil RNA incorporation through UPRT[21]. Split enzyme methods have been utilized in many

other biological applications and have proven to be extremely powerful for highly specific analysis of cell state and function[39]. This opens the door for designing signaling-gated enzyme reconstitution for controlled RNA labeling. Lastly, the types of sequencing that can be employed with metabolic labeling will continue to expand, but merging the kinetics of metabolic labeling with direct RNA-sequencing (Oxford Nanopore Technologies) has already been demonstrated for both 4SU and 5-EU metabolic probes[40-41]. Also, as long-read sequencing techniques such as Pac-Bio become more sophisticated, the list of metabolic labeling approaches which can be interrogated with long-read sequencing will undoubtedly expand. These efforts will shed additional light on the kinetics of RNA processing and alternative splicing.

The methods to develop cell-specific metabolic labeling have matured significantly with the help of identifying specific enzyme-nucleobase/nucleoside pairs to control incorporation into RNA. Although useful, these methods rely on expression of specific enzymes, which can present hurdles for efficiency and even cell-specific expression in rare cells. To overcome these challenges, alternative approaches employing other classes of small molecules for cell-specific labeling can be composed of functional groups with chemical compositions activated in in specific cell compartments, such as those with high or low pH and ion concentrations[42-43]. In addition, many endogenous cellular enzymes have been targeted to "uncage" a variety of functional groups for efficient liberation of other biomolecules[42]. Future efforts in this area should be focused on designing and synthesizing more sophisticated and specific groups to control cell-specific RNA labeling *in vivo*.

In this chapter, I have organized an analysis of the chemical approaches to achieving analysis of RNA expression by metabolic labeling. The chemical handles used have increased dramatically, while at the same time methods for cellular imaging and enrichment of profiled RNAs have matured significantly. Applications of cell-specific RNA expression methods have started to mature and have revealed their utility for the analysis of gene expression *in vivo*, in complex tissues, and in exciting biologically relevant settings. Future efforts will continue to expand the chemical methods for tagging RNAs, while at the same time continuing to exploit chemical approaches for enabling simpler applications to complex cell types.

References

1 Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol* 2013, 20:300-307.

2 Mattick JS. The genetic signatures of noncoding RNAs. *PLoS Genet* 2009, 5:e1000459.

3 Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 2011, 25:1915-1927.

4 Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol* 2016, 17:13.

5 Nishimoto KP, Newkirk D, Hou S, Fruehauf J, Nelson EL. Fluorescence activated cell sorting (FACS) using RNAlater to minimize RNA degradation and perturbation of mRNA expression from cells involved in initial host microbe interactions. *J Microbiol Methods* 2007, 70:205-208.

6 Clark NM, Fisher AP, Sozzani R. Identifying Differentially Expressed Genes Using Fluorescence-Activated Cell Sorting (FACS) and RNA Sequencing from Low Input Samples. *Methods Mol Biol* 2018, 1819:139-151.

7 Sanz E, Bean JC, Carey DP, Quintana A, McKnight GS. RiboTag: Ribosomal Tagging Strategy to Analyze Cell-Type-Specific mRNA Expression In Vivo. *Curr Protoc Neurosci* 2019, 88:e77.

8 Lesiak AJ, Brodsky M, Neumaier JF. RiboTag is a flexible tool for measuring the translational state of targeted cells in heterogeneous cell cultures. *Biotechniques* 2015, 58:308-317.

9 Riley KJ, Yario TA, Steitz JA. Association of Argonaute proteins and microRNAs can occur after cell lysis. *RNA* 2012, 18:1581-1585.

10 Mili S, Steitz JA. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* 2004, 10:1692-1694.

11 Yang Z, Edenberg HJ, Davis RL. Isolation of mRNA from specific tissues of Drosophila by mRNA tagging. *Nucleic Acids Res* 2005, 33:e148.

12 Wissink, E. M.; Vihervaara, A.; Tippens, N. D.; Lis, J. T. Nascent RNA Analyses: Tracking Transcription and Its Regulation. *Nat Rev Genet* 2019, *20* (12), 705–723.

13 Beasley S, Nguyen K, Fazio M, Spitale RC. Protected pyrimidine nucleosides for cell-specific metabolic labeling of RNA. *Tetrahedron Lett* 2018, 59:3912-3915.

14 Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotechnol* 2005, 23:232-237.

15 Curanovic D, Cohen M, Singh I, Slagle CE, Leslie CS, Jaffrey SR. Global profiling of stimulus-induced polyadenylation in cells using a poly(A) trap. *Nat Chem Biol* 2013, 9:671-673.

16 Grammel M, Hang H, Conrad NK. Chemical reporters for monitoring RNA synthesis and poly(A) tail dynamics. *Chembiochem* 2012, 13:1112-1115.

17 Jao CY, Salic A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc Natl Acad Sci U S A* 2008, 105:15779-15784.

18 Melvin WT, Milne HB, Slater AA, Allen HJ, Keir HM. Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. *Eur J Biochem* 1978, 92:373-379.

19 Nainar S, Beasley S, Fazio M, Kubota M, Dai N, Correa IR, Jr., Spitale RC. Metabolic Incorporation of Azide Functionality into Cellular RNA. *Chembiochem* 2016, 17:2149-2152.

20 Feng C, Li Y, Spitale RC. Photo-controlled cell-specific metabolic labeling of RNA. *Org Biomol Chem* 2017, 15:5117-5120.

21 Hida N, Aboukilila MY, Burow DA, Paul R, Greenberg MM, Fazio M, Beasley S, Spitale RC, Cleary MD. EC-tagging allows cell type-specific RNA analysis. *Nucleic Acids Res* 2017, 45:e138

22 Nguyen K, Fazio M, Kubota M, Nainar S, Feng C, Li X, Atwood SX, Bredy TW, Spitale RC. Cell-Selective Bioorthogonal Metabolic Labeling of RNA. *J Am Chem Soc* 2017, 139:2148-2151.

23 Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N, et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* 2011, 29:436-442.

24 Ghosh AC, Shimell M, Leof ER, Haley MJ, O'Connor MB. UPRT, a suicide-gene therapy candidate in higher eukaryotes, is required for Drosophila larval growth and normal adult lifespan. *Sci Rep* 2015, 5:13176.

25 Duffy EE, Rutenberg-Schoenberg M, Stark CD, Kitchen RR, Gerstein MB, Simon MD. Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. *Mol Cell* 2015, 59:858-866.

26 Duffy EE, Simon MD. Enriching s(4) U-RNA Using Methane Thiosulfonate (MTS) Chemistry. *Curr Protoc Chem Biol* 2016, 8:234-250.

27 Zajaczkowski EL, Zhao QY, Zhang ZH, Li X, Wei W, Marshall PR, Leighton LJ, Nainar S, Feng C, Spitale RC, et al. Bioorthogonal Metabolic Labeling of Nascent RNA in Neurons Improves the Sensitivity of Transcriptome-Wide Profiling. *ACS Chem Neurosci* 2018, 9:1858-1865.

28 Nainar S, Cuthbert BJ, Lim NM, England WE, Ke K, Sophal K, Quechol R, Mobley DL, Goulding CW, Spitale RC. An optimized chemical-genetic method for cell-specific metabolic labeling of RNA. *Nat Methods* 2020, 17:311-318.

29 Zhang Y, Kleiner RE. A Metabolic Engineering Approach to Incorporate Modified Pyrimidine Nucleosides into Cellular RNA. *J Am Chem Soc* 2019, 141:3347-3351.

30 Wang D, Zhang Y, Kleiner RE. Cell- and Polymerase-Selective Metabolic Labeling of Cellular RNA with 2'-Azidocytidine. *J Am Chem Soc* 2020, 142:14417-14421.

31 Riml C, Amort T, Rieder D, Gasser C, Lusser A, Micura R. Osmium-Mediated Transformation of 4-Thiouridine to Cytidine as Key To Study RNA Dynamics by Sequencing. *Angew Chem Int Ed Engl* 2017, 56:13479-13483.

32 Lusser A, Gasser C, Trixl L, Piatti P, Delazer I, Rieder D, Bashin J, Riml C, Amort T, Micura R. Thiouridine-to-Cytidine Conversion Sequencing (TUC-Seq) to Measure mRNA Transcription and Degradation Rates. *Methods Mol Biol* 2020, 2062:191-211.

33 Schofield JA, Duffy EE, Kiefer L, Sullivan MC, Simon MD. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat Methods* 2018, 15:221-225.

34 Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, Wlotzka W, von Haeseler A, Zuber J, Ameres SL. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* 2017, 14:1198-1204.

35 Gay L, Miller MR, Ventura PB, Devasthali V, Vue Z, Thompson HL, Temple S, Zong H, Cleary MD, Stankunas K, et al. Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. *Genes Dev* 2013, 27:98-115.

36 Chatzi C, Zhang Y, Shen R, Westbrook GL, Goodman RH. Transcriptional Profiling of Newly Generated Dentate Granule Cells Using TU Tagging Reveals Pattern Shifts in Gene Expression during Circuit Integration. *eNeuro* 2016, 3.

37 Tomorsky J, DeBlander L, Kentros CG, Doe CQ, Niell CM. TU-Tagging: A Method for Identifying Layer-Enriched Neuronal Genes in Developing Mouse Visual Cortex. *eNeuro* 2017, 4.

38 Roche KL, Nukui M, Krishna BA, O'Connor CM, Murphy EA. Selective 4-Thiouracil Labeling of RNA Transcripts within Latently Infected Cells after Infection with Human Cytomegalovirus Expressing Functional Uracil Phosphoribosyltransferase. *J Virol* 2018, 92.

39 Shekhawat SS, Ghosh I. Split-protein systems: beyond binary protein-protein interactions. *Curr Opin Chem Biol* 2011, 15:789-797.

40 Drexler HL, Choquet K, Churchman LS. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Mol Cell* 2020, 77:985-998 e988.

41 Maier KC, Gressel S, Cramer P, Schwalb B. Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *Genome Res* 2020, 30:1332-1344.

42 Fernandez A, Vendrell M. Smart fluorescent probes for imaging macrophage activity. *Chem Soc Rev* 2016, 45:1182-1196.

43 Wu L, Sedgwick AC, Sun X, Bull SD, He XP, James TD. Reaction-Based Fluorescent Probes for the Detection and Imaging of Reactive Oxygen, Nitrogen, and Sulfur Species. *Acc Chem Res* 2019, 52:2582-2597.

Chapter 2: A Bump-Hole Strategy for Increased Stringency of Cell-Specific Metabolic Labeling of RNA

Introduction

Profiling biomolecules in a cell-specific manner is a grand challenge in biochemical research. Chemical approaches toward this problem have expanded researchers' ability to understand protein expression and dynamics within specific cell types[1-5], better characterize cell cycle dynamics with modified DNA analogs[6-7], and characterize metabolic flux with chemical reporters of metabolites and glycans[8-10]. These approaches have matured because of efforts focused on pairing chemically modified metabolic intermediates with enzymes to control their flux and incorporation into endogenous cellular biomolecules.

Few researchers have developed chemically modified nucleoside analogs to track RNA expression, but these reagents are not cell specific[11-13]. We have recently expanded the chemical methods for cell-specific metabolic labeling of RNA, by chemically diversifying nucleobases to make them become either activated by enzymes or de- "caged" such that liberated nucleobases can eventually be incorporated into cellular RNA[14-16]. However, each of these approaches has their own limitations.

"Caged" nucleobases are often protected by carbonyl groups which may be susceptible to hydrolysis and may not be stable enough to work *in vivo*[17]. An alternative, and more actively pursued, approach is to utilize chemically modified nucleobases/nucleosides as metabolic intermediates that cannot be processed in normal cells. Pairing these with an enzyme that can convert inert intermediates into a form that

can be processed for eventual RNA incorporation is being currently pursued. As an example, we and others recently worked to identify modified nucleosides (2′-azidouridine) to match with the enzyme UCK2[18-20]. Despite the exciting observations, it is widely appreciated that sugar-functionalized nucleoside analogs have limitations of deep tissue

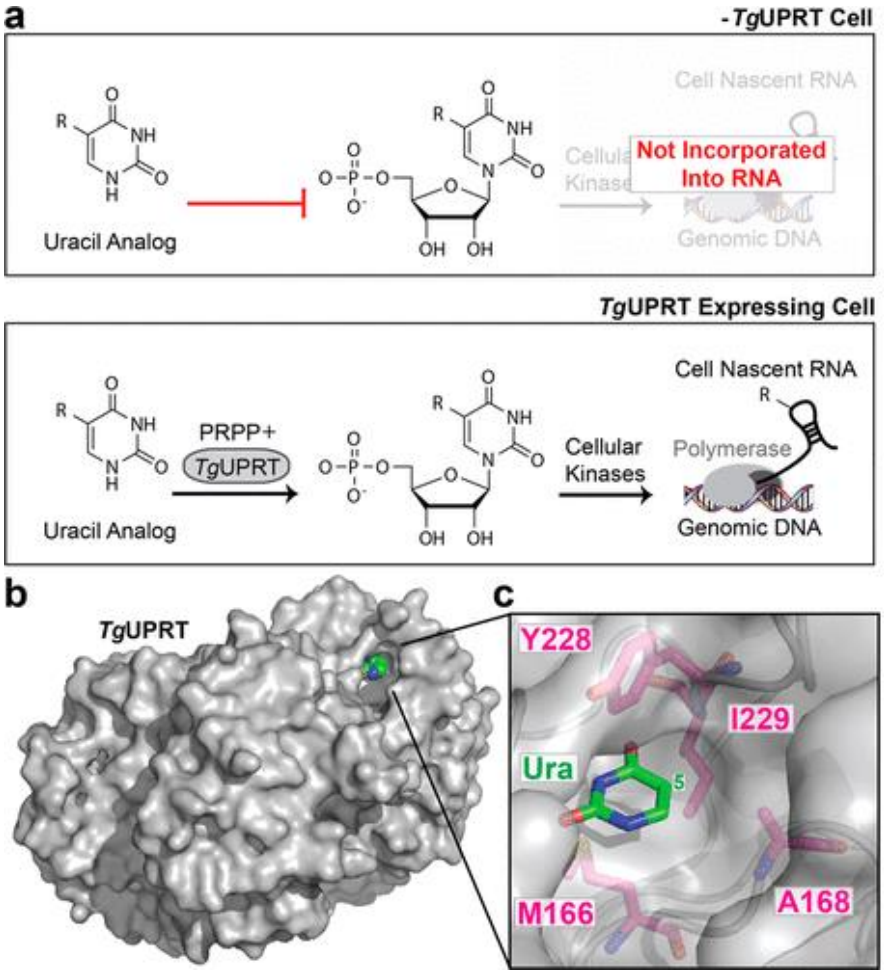


Figure 2-1 UPRT-dependent metabolic labeling of RNA. (a) Schematic of (−)*Tg*UPRT versus *Tg*UPRT expressing cells that enable cell-specific metabolic labeling of RNA. (b) Crystal structure of the *Tg*UPRT enzyme (PDB 1bd4). (c) Close-up view of the *Tg*UPRT active site. Positions chosen for mutagenesis are labeled.

penetrance, crossing the plasma membrane, and the lack of permeability across the blood–brain barrier due to their hydrophilic nature[21-23].

The potential problems with nucleoside analogs are in many cases not shared with nucleobases, which are much smaller in molecular weight and less hydrophilic[24-25]. In parallel, our laboratory has worked to expand the substrate capacity of *Toxoplasma gondii* uracil phosphoribosyltransferase (*Tg*UPRT) with modified 5-uracil analogs to produce modified 5′-phosphorylated uridines **(Figure 2-1a)**. However, mammalian cells are able to salvage uracil analogs for eventual incorporation into cellular RNA, without expression of *Tg*UPRT[26-27]. Together, these studies suggest that tailoring cell-specific metabolic labeling efforts to decrease the background associated nucleobase analogs leaves critical components to be optimized but presents a unique opportunity to expand the chemical repertoire of analogs to achieve high-stringency and cell-specific RNA labeling.

As mentioned, previous work in our laboratory investigated the ability of 5-modified uracil analogs to be eventually incorporated into cellular RNA in a *Tg*UPRT-dependent manner. We observed that 5-ethynyluracil **(Figure 2-S1)** was capable of such transformations [and later demonstrated in (−)*Tg*UPRT cells]. Consistent with our previous results, the most widely adopted uracil analog, 4-thiouracil, also has a low level of background incorporation in (−)*Tg*UPRT cells[18] **(Figure 2-S2)**. However, uracil analogs with bulkier functional groups, such as -vinyl and -methylazido were not, presumably due to steric clashes with the *Tg*UPRT active site[27]. Inspection of the *Tg*UPRT crystal structure in complex with uracil supports this notion as the uracil nucleobase is tightly surrounded

by many amino acid residues which likely clash with bulkier functional groups at the 5-

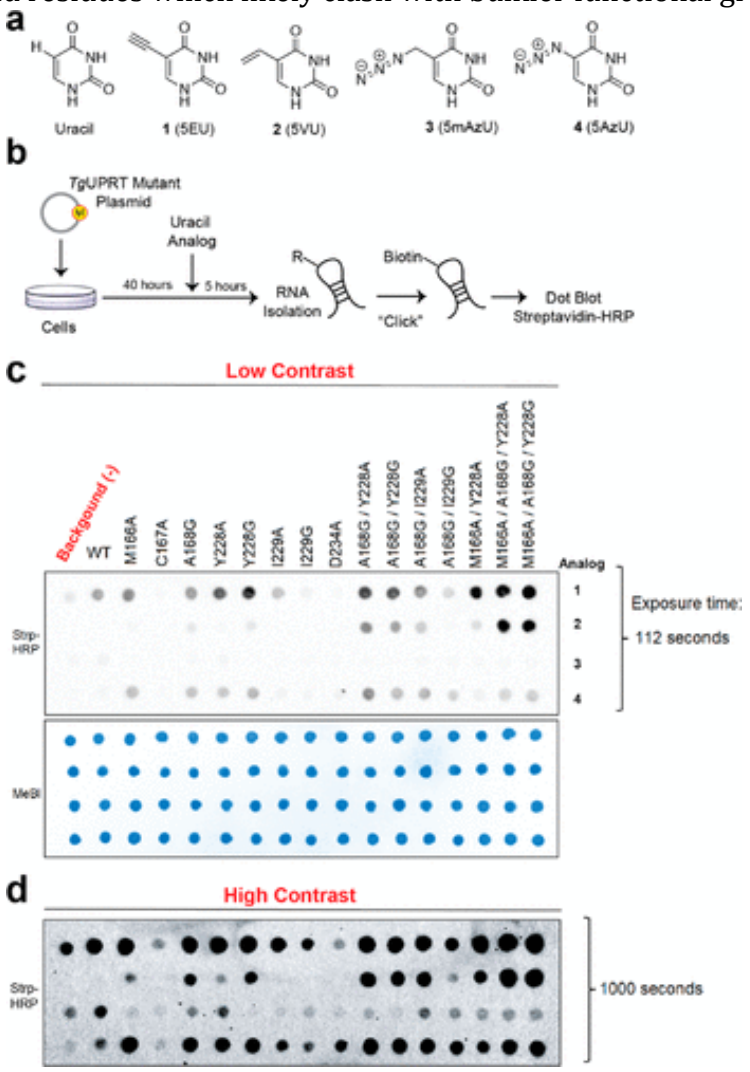


Figure 2-2 In-cell screening of *Tg*UPRT mutants matched with bioorthogonal analogs. (a) Chemical structures of uracil and uracil analogs used herein. (b) Schematic of in-cell screening experiments. HEK293T cells were transfected with *Tg*UPRT plasmids containing various mutations. Uracil analogs were added at 200 μM and incubated for 5 h. Following RNA isolation, biotinylation was performed and incorporation of the analog was determined by streptavidin dot blot. (c) Dot blot screening for RNA incorporation of four different uracil analogs by 15 *Tg*UPRT mutants. (d) Longer exposure of the dot blot shown in panel c. MeBl = methylene blue staining served as a loading control.

position on uracil **(Figure 2-2, b-c)**.

Results

A commonly used strategy for fitting bulky substrates into the active sites of enzymes is to create a corresponding "hole" to the bulky "bump" of large functional groups ("bump-and-hole")[16]. This approach has been employed to pair nucleotide kinase enzymes with nucleobases, but these modified nucleosides can be toxic to cells[20] and have limited tissue penetrance *in vivo*. This strategy has been used successfully for many classes of enzymes; therefore, we applied this method to screen for *Tg*UPRT mutants with different bulky modified uracil analogs.

Herein, we report an analysis of *Tg*UPRT mutants **(Table 2-S1)** and corresponding uracil analogs that provide a binary stringency for cell-specific metabolic RNA labeling. We began by inspecting the active site of *Tg*UPRT and identified several active site residues that could be amenable to mutation (**Figure 2-1, b-c**; **Figure 2-2)**. Single, double, and triple mutations were cloned and transiently transfected into HEK293T cells. Forty hours post-transfection, each uracil analog was added at a 200 μM final concentration for 5 h. RNA was subsequently isolated and appended with biotin using either copper-catalyzed azide–alkyne cycloaddition (CuAAC), with biotin-conjugated alkyne or azide (for azido– or alkynyl–uracil analogs, respectively). For analog **2**, we utilized inverse electron-demand Diels–Alder with a biotin-conjugated tetrazine (IEDDA), a well-established reaction used with RNA, DNA, and other biomolecules for conjugation[28-30]. Biotinylation was assayed using streptavidin-HRP dot blots **(Figure 2-2b)**. We observed background incorporation of **1** at this moderate concentration, as we had previously[15] **(Figure 2-2, c-d)**. Notably, at

higher concentrations (1 mM), we did not observe evidence of incorporation of **2 (Figure 2-S3)** but very robust incorporation of **1.**

From the comparison of uracil analog incorporation in the RNA of cells containing the wild-type (WT) or mutant *Tg*UPRT to (−)*Tg*UPRT (untransfected) cells, we identify several mutants that seemed to be compatible with most C-5 modified uracil analogs, and the triple mutants (3xMT), M166A/A168G/Y228A and M166A/A168G/Y228G, enabled robust incorporation of **2** (5-vinyluracil, 5VU). Cell viability measurement also demonstrated that this pair (M166A/A168G/Y228A and **2**) exhibits no significant differences from untreated cells **(Figure 2-S4)**, consistent with our recent evaluations with 5-vinyluridine analogs[28].

In previous work, we demonstrated that background labeling of uracil and uracil-like compounds (e.g., **1**) was due to the expression of uridine monophosphate synthase (UMPS)[18]. However, UMPS overexpression did not result in **2** incorporation into RNA **(Figure 2-S5)**. These results further suggest that **2** is not a viable substrate in endogenous enzyme pathways for eventual incorporation into RNA. Following these exciting observations, we aimed to obtain a more quantitative understanding of the differences between enzyme mutants and their enzyme kinetics with the analogs **(Figure 2-S6)**. To do so, we expressed and purified five representative mutants [single (M166A, A168G, Y228A), double (A168G/Y228A), and triple (M166A/A168G/Y228A)] and the WT *Tg*UPRT **(Figure 2-3; Tables 2-S2, 2-S3).** Compound **3** showed no detectable activity with any mutants. Compound **4** was reactive with only three *Tg*UPRT mutants (M166A, A168G/Y228A, and M166A/A168G/Y228A) but has no activity with the WT, A168G, and Y228A mutants). We

also observed reactivity of uracil for all five mutants but at different levels of specific activity, which suggests these enzymes are not orthogonal to host cell metabolism but are selective for the chemically modified uracil analogs described here.

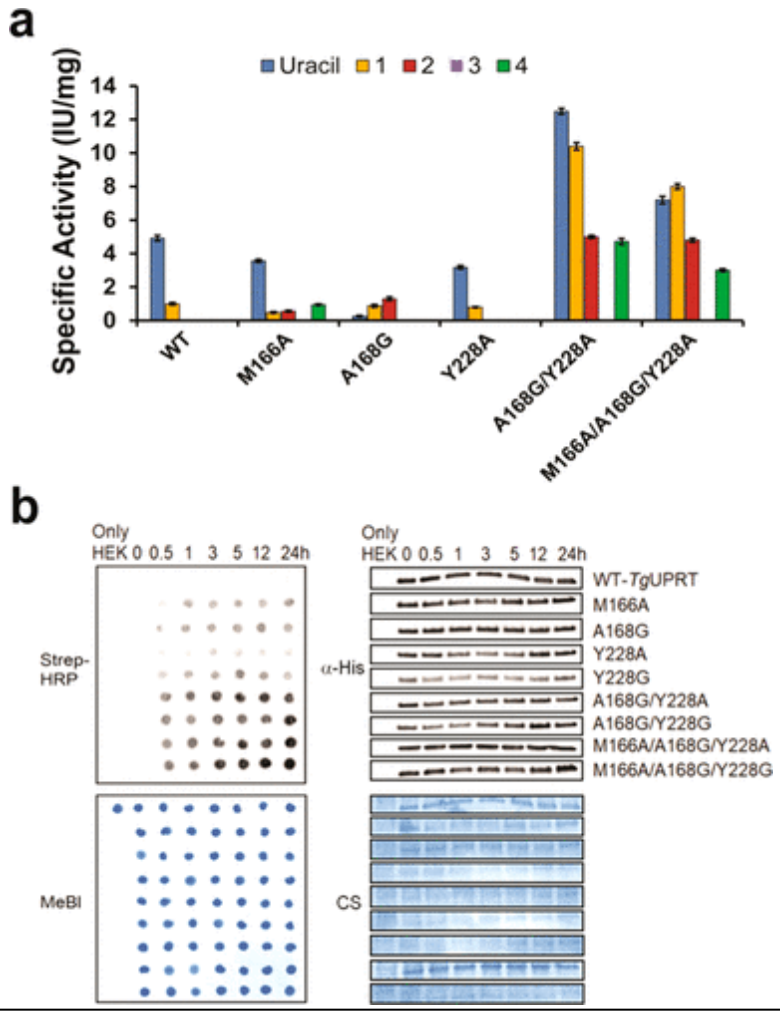


Figure 2-3 *In vitro* analysis of phosphoribosyltransferase activity of *Tg*UPRT variants with different uracil analogs. (a) Specific activity of *Tg*UPRT variants with different uracil analogs (**1** to **4** as shown in Figure 2A). *n* = 3 technical replicates. Error bars are standard deviation of the mean (b). Time course analysis of **2** (5VU) analog incorporation into RNA by streptavidin dot blot (left panels) along with immunoblot analysis (right panels) of corresponding 6xHis-*Tg*UPRT protein levels. α-His = anti-His antibody for immunoblot. Strep-HRP = streptavidin conjugated horseradish peroxidase was used for the assessment of biotin levels resulting from clicked RNA. MeBl = methylene blue stain and CS = Coomassie staining served as a loading control.

Consistent with our in-cell screening, compound **2** has undetectable activity in WT but increased dramatically with all mutants except Y228A, whereas compound **1** was reactive with WT enzymes, while also exhibiting increased reactivity with the mutants (10 IU/mg). Using dot blot analysis to evaluate the level of 5-vinyluracil incorporation into RNA, we also observed the analog incorporation in as little as 0.5 h with *Tg*UPRT mutants, which is faster and more robust than we previously reported **(Figure 2-3b)**. Overall, these results further support the specificity of our designed mutants and also demonstrate the increased metabolic incorporation and efficiency of the triple mutant 3xMT-*Tg*UPRT/**2** pair.

The above experiments demonstrate the specificity of the 3xMT-*Tg*UPRT/**2** pair; however, we aimed to better understand if these observations would enable cell-specific imaging of RNA incorporation of **2** and enrichment of RNA molecules. We began with an imaging experiment where **2** or DMSO was incubated with untransfected or transfected 3xMT-*Tg*UPRT (M166A/A168G/Y228A) HEK293T cells for 24 h at a 1 mM concentration **(Figures 2-S4, 2-S7)**, we observed a fluorescent signal only in cells with the presence of both **2** and the 3xMT-*Tg*UPRT mutant. However, cells in the experiments treated with either **2** or transfected with mutant *Tg*UPRT alone show no signal **(Figure 2-4a)**.

We next aimed to understand the stringency of RNA enrichment with the two analogs and the expression of the 3xMT-*Tg*UPRT enzyme. Compound **1** or **2** was incubated with untransfected or transfected 3xMT-*Tg*UPRT (M166A/A168G/Y228A) cells for 5 h at a 1 mM concentration. RNA was isolated from cells, and CuAAC- or IEDDA-biotinylated transcripts were subsequently enriched. The GAPDH mRNA copy number was quantified

using RT-qPCR of enriched RNAs. As shown in (**Figure 2-S8)**, we were able to enrich GAPDH mRNA from untransfected cells treated with **1**, but with negligible copies of GAPDH mRNA enriched in cells treated with **2** alone. GAPDH mRNA was enriched above 1 × 10^9 copies when **1** or **2** was added to cells transfected with 3xMT-*Tg*UPRT. These results

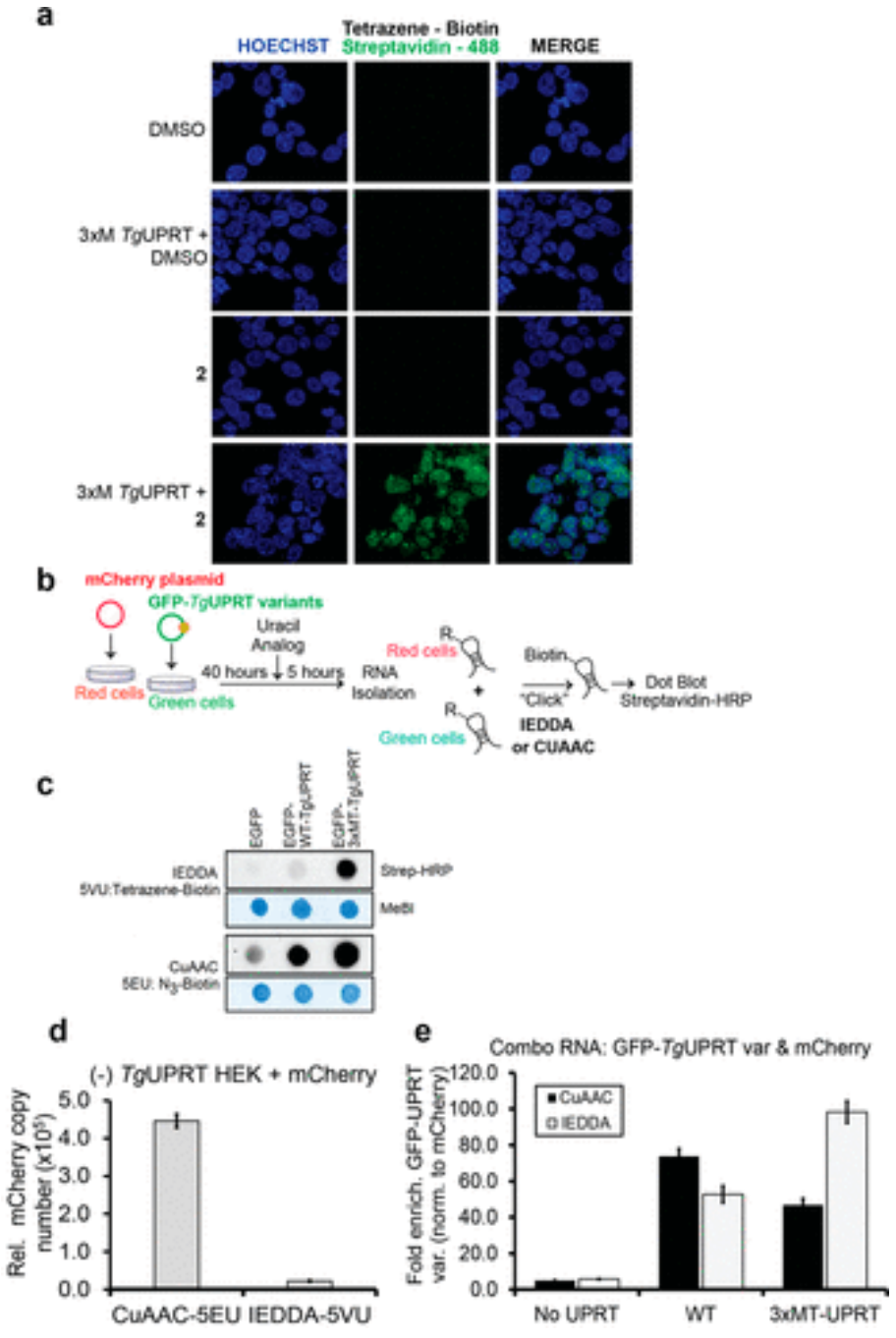


Figure 2-4 Characterizing the stringency of the 2 (5VU) mutant *Tg*UPRT pair. (a) Microscopy analysis RNA incorporation of **2** in a mutant *Tg*UPRT-dependent manner. Cells were transfected with *Tg*UPRT variants and incubated with **2** at a 1 mM final concentration for 24 h. **2** incorporation was imaged using two-step labeling: IEDDA using tetrazine-biotin then followed by Alexa488-streptavidin. (b) Schematic of experiment to assess specificity of **2** in cells transfected with mCherry (no UPRT) or GFP, GFP-*Tg*UPRT variants treated with both 200 μM **1** and 400 μM **2**. (c) Dot blot analysis of RNA isolated from **1** and **2** treated cells underwent either a CuAAC or IEDDA click reaction. Streptavidin-HRP was used for the assessment of biotin levels resulting from clicked RNA, and methylene blue (MeBl) staining served as a loading control. (d) QPCR analysis of enriched CuAAC- or IEDDA-biotinylated mCherry cDNA from (−)*Tg*UPRT HEK293T cells. (e) QPCR analysis of enriched CuAAC or IEDDA-biotinylated GFP cDNA from (−)*Tg*UPRT (no UPRT), WT-, and 3xMT-*Tg*UPRT (M166A/A168G/Y228A) transfected HEK293T cells.

suggest that **2** and 3xMT-*Tg*UPRT are a suitable pair for enrichment of mRNAs with negligible background due to spurious uracil analog incorporation into RNA.

To further demonstrate the specificity of RNA tagging, we performed coculture experiments. In these experiments, cells expressing transfected mCherry (off-target cells) were cocultured with cells expressing fusion of GFP-*Tg*UPRT variants (target cells) or GFP without *Tg*UPRT (negative control target cells) and treated with 200 μM of **1** and 400 μM of **2**. Total RNA from the coculture was subjected to biotinlyation via (a) CuAAC click reaction or (b) IEDDA ligation for the assessment of **1** or **2** incorporation in RNA,

respectively. Dot blot analysis of the coculture demonstrated again that **1** was incorporated into the coculture RNA (and presumably both cells) in the absence of UPRT as well as both variants (GFP-WT-*Tg*UPRT and GFP-3xMT-*Tg*UPRT). In contrast, **2** was only incorporated to high levels in the presence of 3xMT-*Tg*UPRT **(Figure 2-4c)**.

Finally, to test enrichment specificity of 3xMT-*Tg*UPRT and **2**, we repeated the above experiment and isolated total RNA. Enrichment and profiling of mCherry mRNA enrichment from the negative control cells ((−)*Tg*UPRT) demonstrated that the copy number of enriched CuAAC-biotinylated **1**-mCherry is almost 19-fold higher than the enriched IEDDA-biotinylated **2**-mCherry transcripts, indicating that the incorporation of **2** in RNA is much less than **1** in (−)*Tg*UPRT cells **(Figure 2-4d)**. We finally compared the enrichment of GFP transcripts in the combo RNA derived from green and red cells as described above, and compared the ratio of copy numbers of GFP or GFP-*Tg*UPRT variants (signal) to mCherry (noise), GFP/mCherry **(Figure 2-4d)**. A higher copy number of enriched mCherry (noise) would result in lower GFP/mCherry ratio and *vice versa*. In GFP or mCherry only expressing cells (No UPRT), which is a condition for testing and establishing the background baseline of this metabolic labeling approach, the CuAAC bar is slightly lower than IEDDA, indicating more **1**-tagged mCherry was enriched than **2**-tagged mCherry. When comparing enriched biotinylated GFP mediated by WT-*Tg*UPRT, we observed **2** incorporation into RNA at 50-fold higher than noise. However, such a ratio of GFP/mCherry does not improve over the ratio of signal-to-noise resulting from CuAAC-biotinylated **1** and the WT-*Tg*UPRT bioorthogonal pair. Consistently, the enrichment of IEDDA-biotinylated **2**-tagged GFP transcripts mediated by 3xMT-A-*Tg*UPRT is 100-fold higher than the background and has 50% improvement over the signal-to-noise ratio of

CuAAC-biotinylated **1**-labeled GFP. These results clearly demonstrate the background RNA incorporation of **1**, which can result in enrichment of transcripts in off-target cells. In stark contrast, our new 3xMT-*Tg*UPRT/**2** pair is highly stringent and can be used to enrich RNAs specifically from target cells.

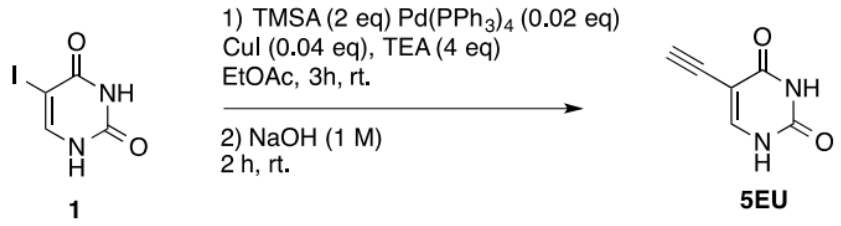
Conclusion

Herein, we have demonstrated that the nucleobase analog 5-vinyluracil (**2**) is likely not an amenable substrate for endogenous enzymes and incorporation into cellular RNA. We also demonstrate that in contrast to 5-ethynyluracil (**1**), **2** is not a substrate for the main enzyme responsible for the uracil analog background in cellular RNA, UMPS. Finally, through in-cell enzyme mutant screening, we identified that mutant UPRT enzymes with more open active sites are able to produce 5′-phosphorylated 5-vinyluracil (**2**) *in vitro*, and this enzymatic activity translates in cells to enable the incorporation of **2** into cellular RNA. Imaging and enrichment RT-qPCR experiments further support the stringency of utilizing a bump-hole approach for the control of cell-specific metabolic labeling. Overall, our results put forth a novel nucleobase-enzyme pair for highly stringent and cell-specific metabolic labeling of RNA. We anticipate these results will expand the scope of such experiments, which are appreciated to have reduced background issues due to endogenous enzymatic and metabolic activities. Future goals ongoing in the lab are to extend these findings to living animal settings. These experiments are currently being pursued in our lab and will be reported in due course.

Experimental Methods

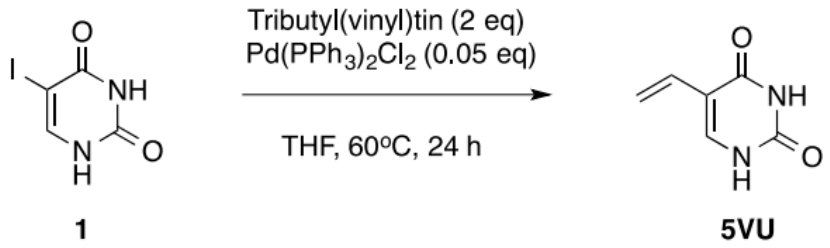
Synthesis and spectra:

5-ethynylpyrimidine-2,4(1H,3H)-dione (5EU)



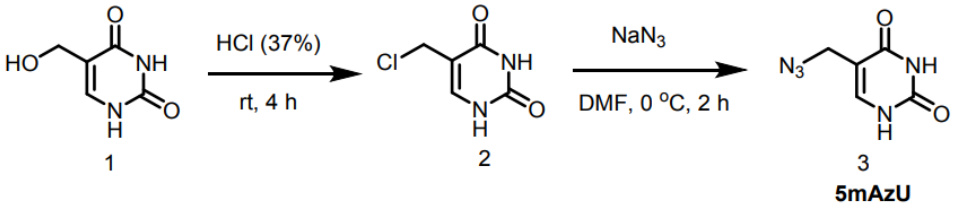
The product was prepared according to a modified procedure by Cooke, et al.1 5-Iodouracil (2000 mg, 8.4 mmol, 1 eq), TMS-acetylene (2.4 mL, 16.8 mmol, 2 eq), Et3N (4.7 mL, 33.6 mmol, 4 eq), Pd(PPh3)4 (196mg, 0.17 mmol, 0.02 eq), and CuI (65 mg, 0.34 mmol, 0.04 eq) were dissolved in 25 mL of degassed EtOAc. The suspension was stirred at r.t. for 3 hrs under Ar. The suspension was then filtered and washed with EtOAc. The extract was collected and dissolved in 10 mL of 1 M NaOH and stirred at r.t. for 2 hrs. The solution was then diluted with 10 mL of H2O and concentrated in vacuo. The residue was then redissolved in 10 mL of H2O and AcOH was added until a pH of 5 was reached. The suspension was then set on ice for 30 mins and filtered. The extract was washed with H2O, acetone, and Et2O. The extract was then dried in vacuo to give 5- ethynylpyrimidine-2,4(1H,3H)-dione (823 mg, 72%) as an off white solid. HRMS Calcd for C6H4N2O2 [M+] 136.11, found 135.02 [M-H-]; 1H NMR (400 MHz, DMSO) δ 11.29 (s, 2H), 7.78 (s, 1H), 3.99 (s, 1H). 13C NMR (126 MHz, DMSO) δ 163.22, 150.96, 147.02, 96.72, 83.71, 77.08.

5-Vinylpyrimidine-2,4(1H,3H)-dione (5VU)

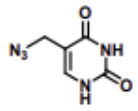


To a solution of 5-iodopyrimidine-2,4(1H,3H)-dione (1) (1.0 g, 4.20 mmol, 1 eq) and Pd(PPh3)2Cl2 (147.4 mg, 0.210 mmol, 0.05 eq) in dry THF (15 mL) was added tributyl(vinyl) stannane (2.45 mL, 8.40 mmol, 2 eq) dropwise at room temperature under argon. The resulting reaction mixture was warmed up to 60C and stirred at 60C for 24 h. The reaction mixture was then cooled down to room temperature and filtered through a pad of celite, washed with MeOH, concentrated to afford a yellowish slurry, which was recrystallized with Hexane/MeOH to afford 2 as a pale yellow solid (205.0 mg, 35% yield). 1H NMR (600 MHz, d6-DMSO) δ 5.06 (d, J = 11.4 Hz,13 H), 5.95 (d, J = 18 Hz, 1 H), 6.35 (dd, J = 18, 11.4 Hz, 1 H), 7.58 (s, 1 H), 11.04 (s, 1H), 11.13 (s, 1 H); 13C NMR (150 MHz, d6-DMSO) δ 109.6, 113.4, 129.0, 139.9, 150.6, 163.1; HRMS (ESI) calcd for C6H6N2O2Na [(M+Na)+] 161.0327, found 161.0333.

5-(Azidomethyl)pyrimidine-2,4(1H,3H)-dione

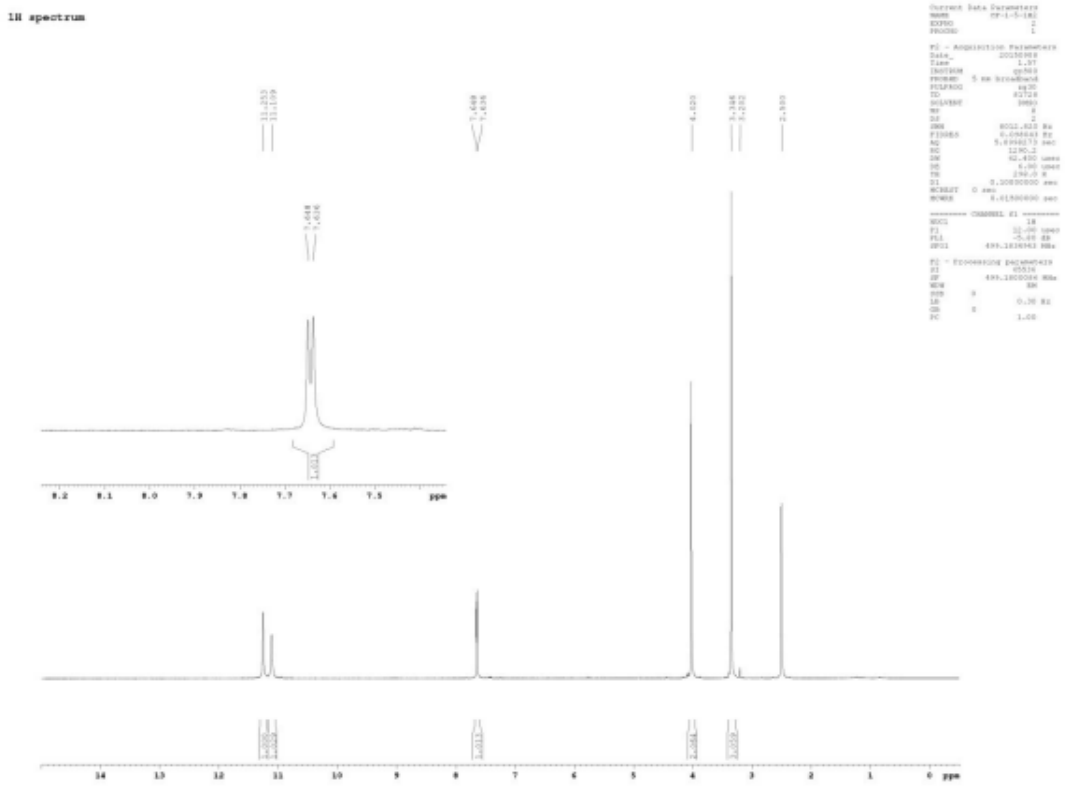


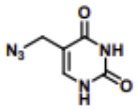
(3). Concentrated HCl (37% aq.) (15.0 mL) was added to 5-(hydroxymethyl)pyrimidine-2,4(1H,3H)-dione (1) (1.90 g, 13.2 mmol) at room temperature. The reaction mixture was stirred vigorously at room temperature for 4 h, then the precipitates were collected by filtration and washing with DI water (4 mLX2). After dried in high vacuum overnight, a white solid (1.17 g) was generated, which was used for next step without further purification. To an ice-cooled solution of sodium azide (521 mg, 8.02 mmol) in dry DMF (5 mL) was added a solution of the above white solid in dry DMF (16 mL) dropwise at room temperature. The solution was stirred at 0 oC for 2 h, and concentrated HCl (37% aq.) (1 mL) was added dropwise carefully into the solution at 0 oC to quench any remaining sodium azide. Then DMF was removed by reduced pressure to form a white solid, which was washed by ice-cooled water (4 mLX2). After dried under high vacuum overnight, a white solid (3) (1.22 g, 53% yield from 1) was formed. 1H NMR (500 MHz, d6-DMSO) δ 4.02 (s, 2 H), 7.64 (d, J = 6 Hz, 1 H), 11.11 (s, 1 H), 11.25 (s, 1 H) ppm; 13C NMR (125 MHz, d6-DMSO) δ 46.5, 106.6, 142.1, 151.2, 164.0 ppm; HRMS (ESI) calcd for C5H4N5O2 [(M-H)–] 166.0365, found 166.0369



5mAzU

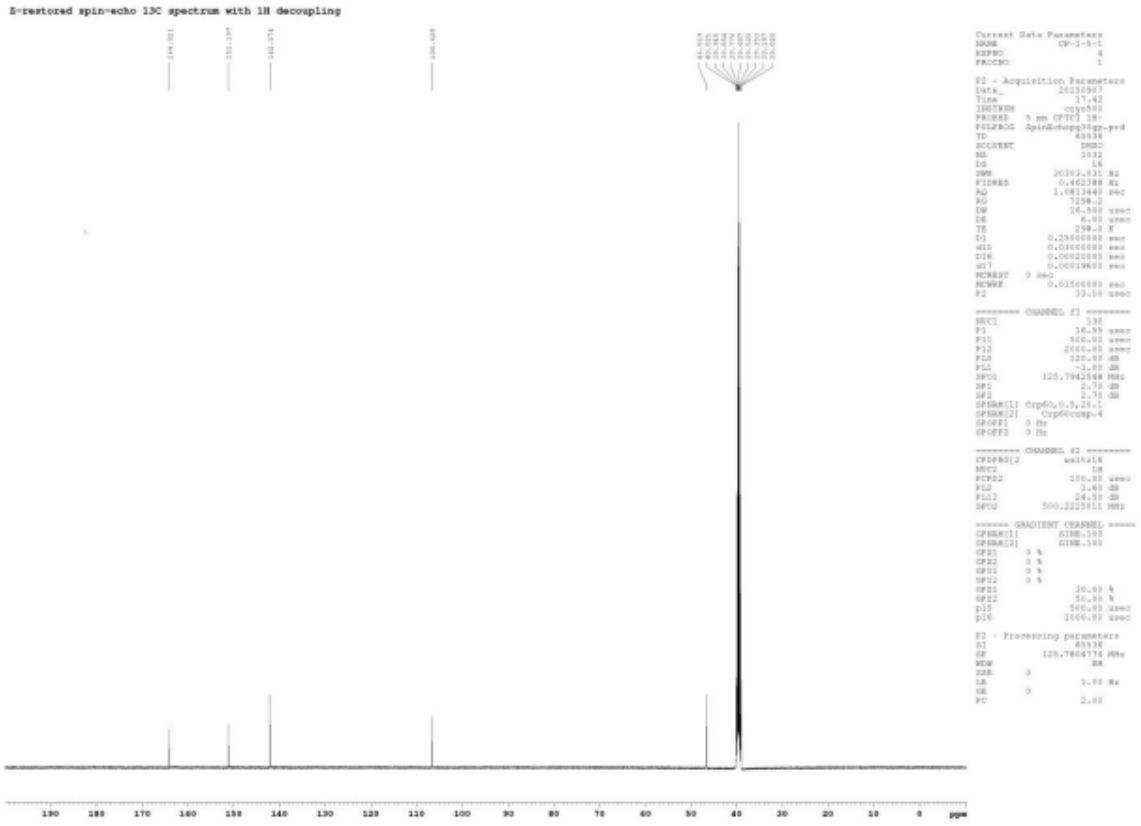
^{1}H NMR (500 MHz, d6-DMSO)



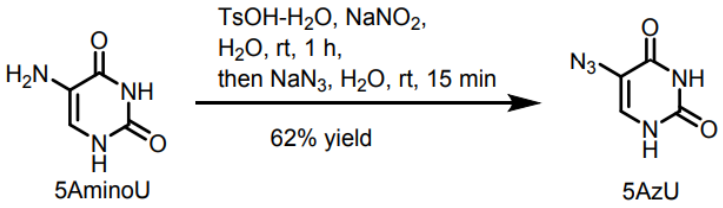


5mAzU

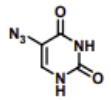
^{13}C NMR (125 MHz, *d*6-DMSO)



5-Azidopyrimidine-2,4(1H,3H)-dione (5AzU):

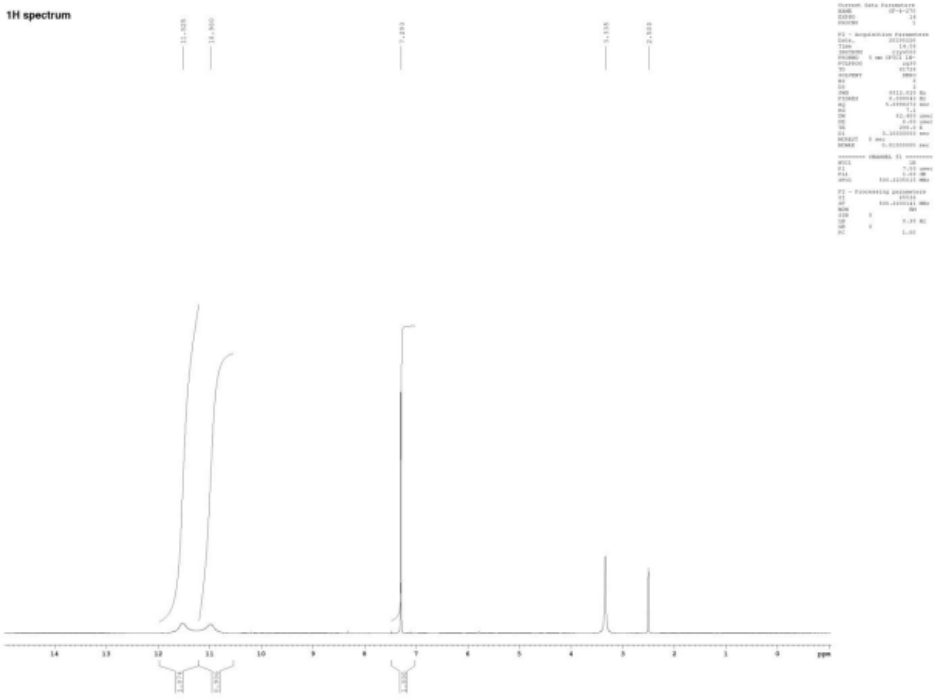


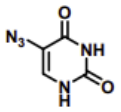
5AzU was synthesized according to a reported procedure.1 To a solution of 5-aminopyrimidine-2,4(1H,3H)-dione (5AminoU) (2.25 g, 17.7 mmol) in water (150 mL) at room temperature was added p-toluenesulfonic acid monohydrate (30.3 g, 160 mmol) and sodium nitric (11.0 g, 160 mmol) slowly and the reaction mixture was stirred vigorously at room temperature for 1 h. Sodium azide (1.84 g, 28.4 mmol) was added slowly to the reaction mixture and the reaction was stirred for additional 15 min until the bubbling was gone. The resulting reaction mixture was cooled down on ice bath to precipitate out the product 5AzU, which was obtained by filtration. After washing by water twice and diethyl ether twice, 5AzU was obtained as light beige solid (1.67 g, 62% yield). 1H NMR (500 MHz, d6-DMSO) δ 7.29 (s, 1 H), 11.0 (brs, 1 H), 11.5 (brs, 1 H) ppm; 13C NMR (125 MHz, d6-DMSO) δ 112.2, 130.1, 150.1, 160.9 ppm. The 1H and 13C NMR spectra were identical to those reported.1 1. M. hornum, A. Djukina, A.-K. Sassnau, P. Nielsen, Org. Biomol. Chem., 2016, 14, 4436.



5AzU

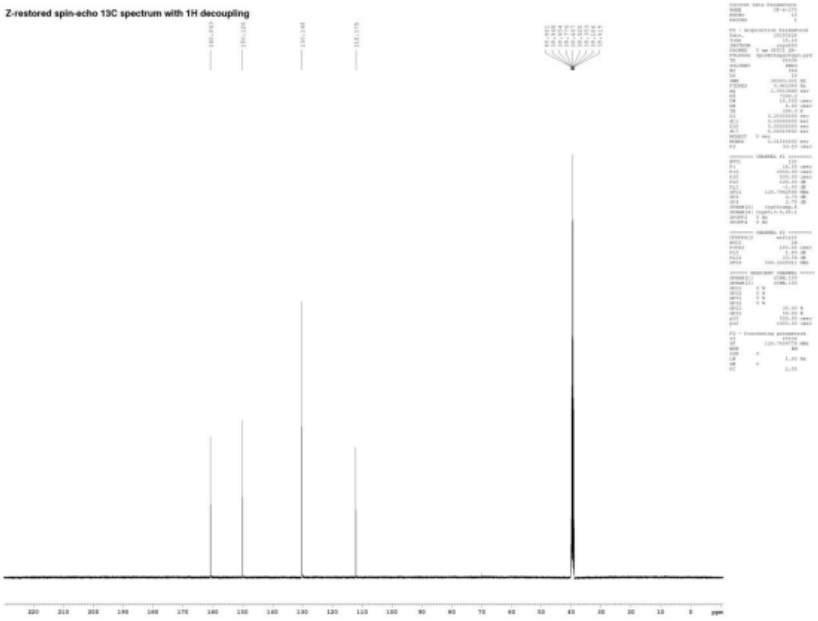
^{1}H NMR (500 MHz, d_6-DMSO)





5AzU

^{13}C NMR (125 MHz, d_6-DMSO)



Cloning and plasmids: The UPRT gene, which encodes a protein annotated as uracil phosphoribosyltransferase from Toxoplasma gondii (European Nucleotide Archive code: AAB60213.1; UniProtKB Q26998) was amplified as NdeI-BamHI fragment from pKN342 containing wild-type (WT) of *Tg*UPRT with 1xHA-tag at N-terminal (pET28a(+)-HA-*Tg*UPRT-WT) to remove HA-tag existing in the template. The PCR product was then sub-cloned into the same expression vector pET28a(+) backbone, leading to the recombinant vector pKN-T7-*Tg*UPRT-WT. The different UPRT mutants were generated by site-directed mutagenesis PCR from *Tg*UPRT WT. The resultant recombinant vectors are

pKN-T7-*Tg*UPRT-Mut1, pKN-T7-*Tg*UPRT-Mut3, pKNT7-*Tg*UPRT-Mut4, pKN-T7-*Tg*UPRT-Mut12 and pKN-T7-*Tg*UPRT-Mut16 (See Table 1) provided the recombinant N-terminal 6xHis-tagged fusion proteins with a thrombin cleavage site between the tag and the enzyme to be used *in vitro* analysis. PCR amplicons of AsiSI and MluI fragments from *Tg*UPRT-WT and variants were sub-cloned into mammalian expression pCMV6 vector resulting recombinant vectors pKN-CMV-*Tg*UPRT-WT and pKN-CMV-TgUPRT-Mutants (Mut) (See Table 1). Fragments of XbaI and XhoI mCherry were amplified from pRS35 template and sub-cloned into pCDNA3.3 backbone, resulting pKN-CMV-mCherry recombinant plasmid.

Cell lines, bacterial and mammalian culture conditions: HEK293 cells (un-transfected or *Tg*UPRT-transiently transfected) were cultured in DMEM (Corning, Cat#: 10-017-CM) supplemented with 10% FBS, 1% (1 mg/mL) penicillin and streptomycin and grown at 37ºC, 5% CO2. *Tg*UPRT and *Tg*UPRT variants were expressed in E. coli BL21(DE3) grown in LB medium at 37 °C with kanamycin 50 μg/mL. E. coli BL21(DE3) cell culture medium reagents were from Difco (St. Louis, United States). Trimethyl ammonium acetate buffer was purchased from Sigma-Aldrich (Madrid, Spain). All other reagents and organic solvents used *in vitro* studies were purchased from Scharlab (Barcelona, Spain) and Symta (Madrid, Spain). Nucleosides and nucleobases used in this work were provided by Carbosynth Ltd. (Compton, United Kingdom).

Transfection, labeling of cellular RNA and RNA isolation: After 48h seeding, HEK293T (HEK) cells were transfected with 2.5 μg of *Tg*UPRT WT or mutant plasmids (pKN-CMV-*Tg*UPRT-WT or -Mut1 to Mut17) per 5cm plate using JetPrime Transfection

reagent (Polyplus Transfection, France). At 40h post-transfection, cells were incubated with a final concentration of 200 μM at <1% DMSO from 400 mM stock of uracil analogs (5EU, 5VU, 5AU), or 200 mM 5AMU stock for 5h. For time course study, HEK cells were treated with 200 μM uracil analogs for 0h, 0,5h, 1h, 3h, 5h, 12h, and 24hours. For titrating uracil analog concentration experiments, HEK were treated with 0 (DMSO), 50, 100, 200, 500 μM or 1 mM uracil analog for 5h. Treated HEK cells were subjected to total RNA extraction using 1mL Trizol reagent (Invitrogen) following the manufacturer's instructions.

Supplementary Data

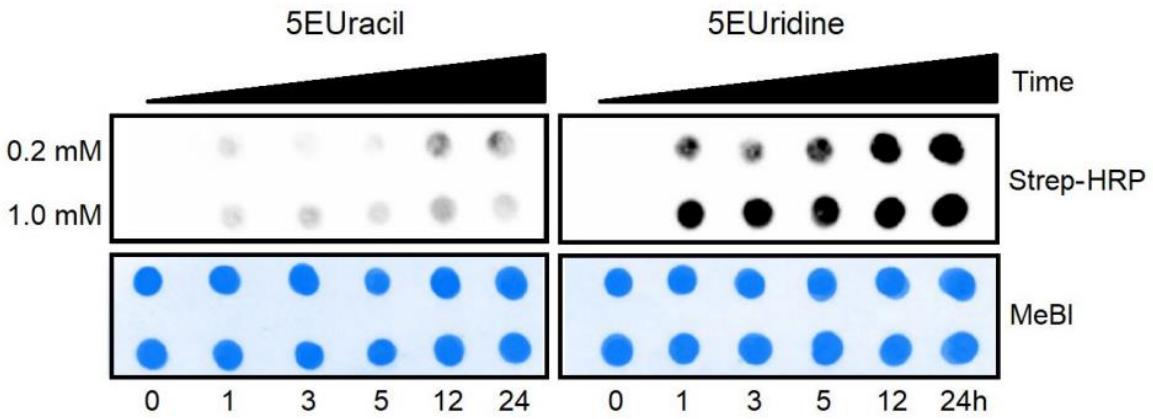


Figure 2-S1 Assessment of background level of 5EU incorporation into RNA in the absence of TgUPRT. HEK293T cells were treated with 5EU or 5EUridine (served as positive control) at two different concentrations (0.2 mM and 1 mM) in a time course. Strep-HRP = Streptavidin conjugated Horseradish peroxidase. MeBl = methylene blue staining serving as loading control.

Table S1. *Tg*UPRT plasmids used in this study

Plasmid ID	Description
Under CMV promoter	
pKN-CMV-*Tg*UPRT-WT	Wild type *Toxoplasma gondii* Uracil PhosphoRiboxylTransferase (*Tg*UPRT-WT)
pKN-CMV-*Tg*UPRT-Mut1	*Tg*UPRT-M166A
pKN-CMV-*Tg*UPRT-Mut2	*Tg*UPRT-C167A
pKN-CMV-*Tg*UPRT-Mut3	*Tg*UPRT-A168G
pKN-CMV-*Tg*UPRT-Mut4	*Tg*UPRT-Y228A
pKN-CMV-*Tg*UPRT-Mut5	*Tg*UPRT-Y228G
pKN-CMV-*Tg*UPRT-Mut6	*Tg*UPRT-I229A
pKN-CMV-*Tg*UPRT-Mut7	*Tg*UPRT-I229G
pKN-CMV-*Tg*UPRT-Mut8	*Tg*UPRT-D235A
pKN-CMV-*Tg*UPRT-Mut9	*Tg*UPRT-M166A-A168G
pKN-CMV-*Tg*UPRT-Mut10	*Tg*UPRT-M166A-Y228A
pKN-CMV-*Tg*UPRT-Mut11	*Tg*UPRT-M166A-Y228G
pKN-CMV-*Tg*UPRT-Mut12	*Tg*UPRT-A168G-Y228A
pKN-CMV-*Tg*UPRT-Mut13	*Tg*UPRT-A168G-Y228G
pKN-CMV-*Tg*UPRT-Mut14	*Tg*UPRT-A168G-I229A
pKN-CMV-*Tg*UPRT-Mut15	*Tg*UPRT-A168G-I229G
pKN-CMV-*Tg*UPRT-Mut16	*Tg*UPRT-M166A-A168G-Y228A
pKN-CMV-*Tg*UPRT-Mut17	*Tg*UPRT-M166A-A168G-Y228G
pKN-CMV-EGFP-*Tg*UPRT-WT	EGFP-*Tg*UPRT-WT
pKN-CMV-EGFP-*Tg*UPRT-Mut16	EGFP-*Tg*UPRT-M166A-A168G-Y228A
pKN-CMV-EGFP-*Tg*UPRT-Mut17	EGFP-*Tg*UPRT-M166A-A168G-Y228G
pKN-CMV-EGFP	pCDNA3.3-EGFP (Addgene, Cat#)
pKN-CMV-mCherry	mCherry
Under T7 promoter	
pKN-T7-TgUPRT-WT	*Tg*UPRT-WT
pKN-T7-TgUPRT-Mut1	*Tg*UPRT-M166A
pKN-T7-TgUPRT-Mut3	*Tg*UPRT-A168G
pKN-T7-TgUPRT-Mut4	*Tg*UPRT-Y228A
pKN-T7-TgUPRT-Mut12	*Tg*UPRT-A168G-Y228A
pKN-T7-TgUPRT-Mut16	*Tg*UPRT-M166A-A168G-Y228A

Table 3-S1 *Tg*UPRT plasmids used in this study.

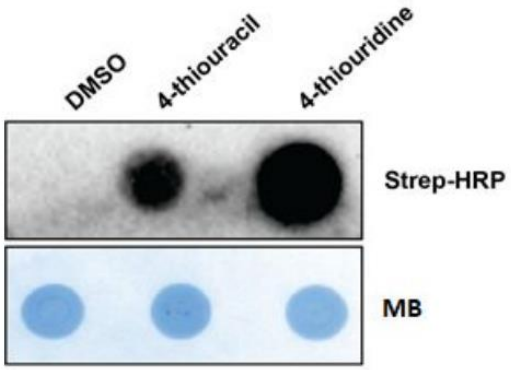


Figure 2-S2 Dot blot demonstrating background incorporation of 4-thiouracil in comparison to 4-thiouridine in HEK293T, (-)TgUPRT cells. Both analogs were added in HEK293T culture at 1 mM concentration for 24 hours. RNA was extracted and biotinylated using MTSEA-Biotin, followed by dot-blot analysis. Strep-HRP = Streptavidin conjugated Horseradish peroxidase. MeBl = methylene blue staining serving as loading control.

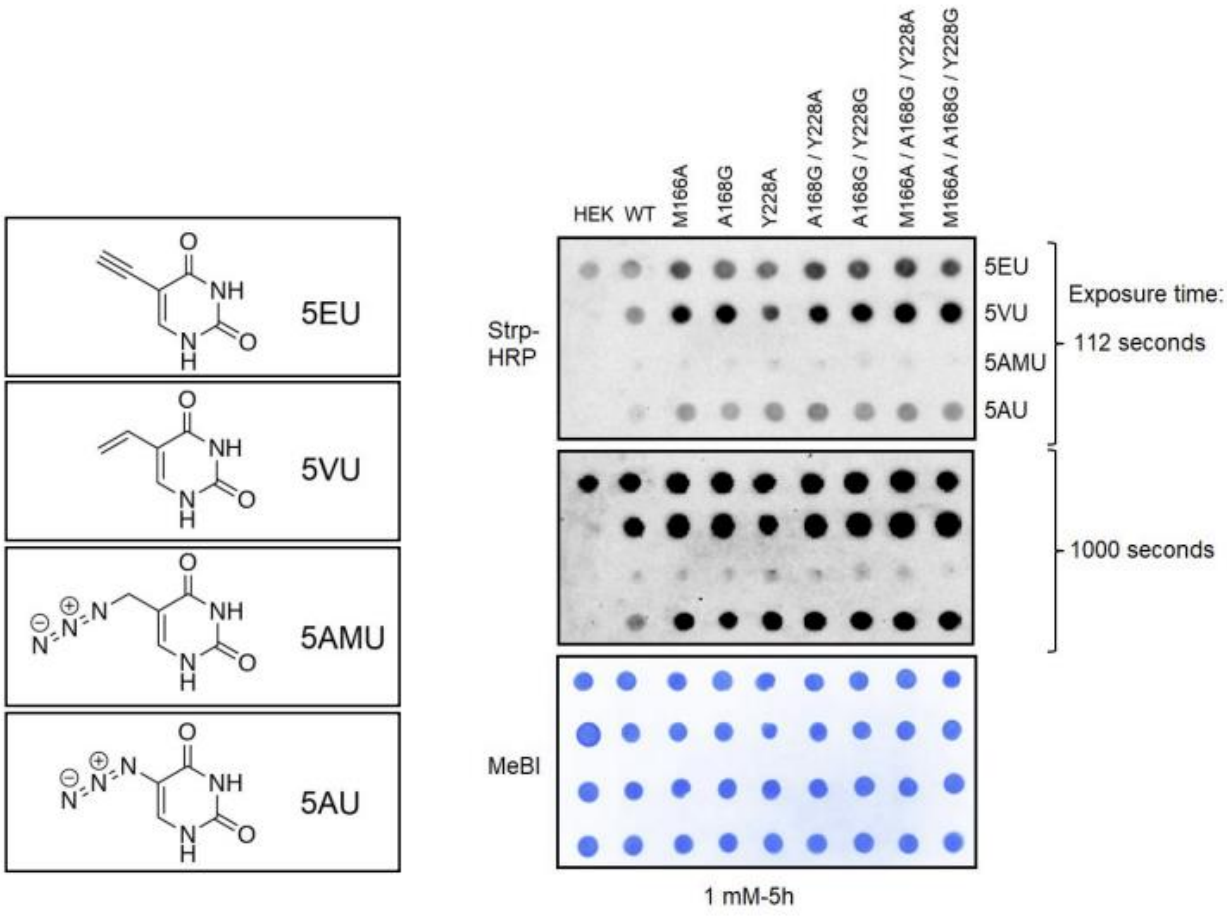


Figure 2-S3 Assessment of different uracil analogs incorporated into RNA without or with *Tg*UPRT-WT or mutants by dot blot analysis. HEK293T transfected with transfect reagent carrier only, with WT or with *Tg*UPRT mutants. Forty hours post-transfection, each of 1-4 uracil analogs was added to a final concentration of 1mM and incubated for 5h. Two µg biotinylated RNA was loaded per spot. (Left) Structures of different uracil analogs. (Right) Images of the same blots acquired at different exposure times: (top) when signal is still in linear range, (middle) when signal reached saturation, (bottom) loading control. Strep-HRP = Streptavidin conjugated Horseradish peroxidase. MeBl = methylene blue staining serving as loading control.

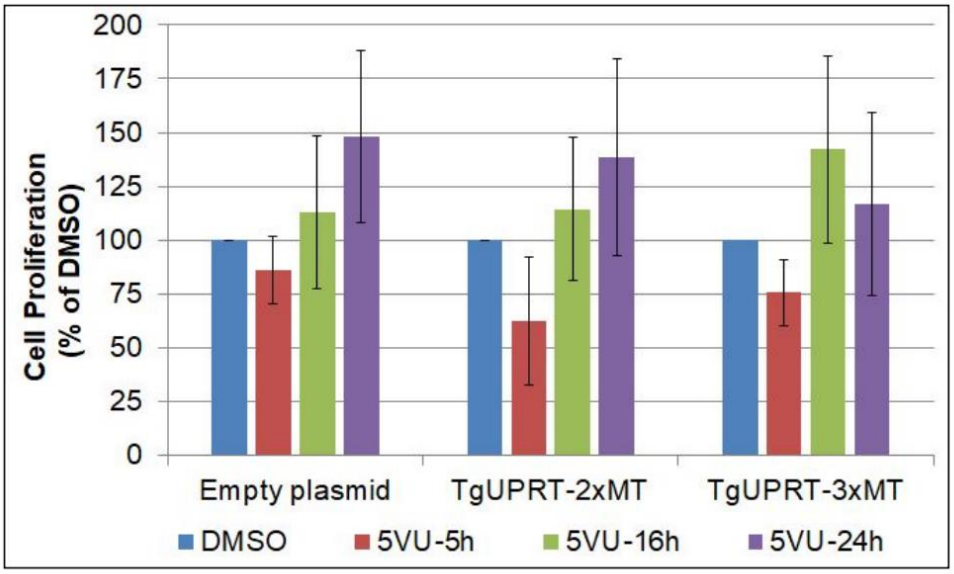


Figure 2-S4 Assessment of cell proliferation under treatment of 5-vinyl uracil time course in the presence or absence of mutant *Tg*UPRT (double = 2xMT or triple = 3xMT mutations). Equal amount HEK293T cells ($2x10^5$) were seeded in 6-well plate with 2mL media. Twenty-four hours post seeding, cells were transfected with 1µg empty plasmid (No GFP nor *Tg*UPRT), double (2xMT: M166A/A168G) or triple (3xMT: M166A/A168G/Y228A) *Tg*UPRT mutants. 5VU was added to transfected cells at 200 µM final concentration for 5, 16 and 24h treatment. Treated cells were assayed with Trypan blue at 48h post transfection. DMSO (carrier) treatment (negative control) was done for 24h. Cells were stained with 0.04% Trypan blue for 5 minutes and subjected to cell counting using Countess II FL (Invitrogen, USA) to count both live and dead cells. Per condition, counting was done with n = 2x biological and 3x technical replicates. The number of live cells was used for calculated cell proliferation percentage.

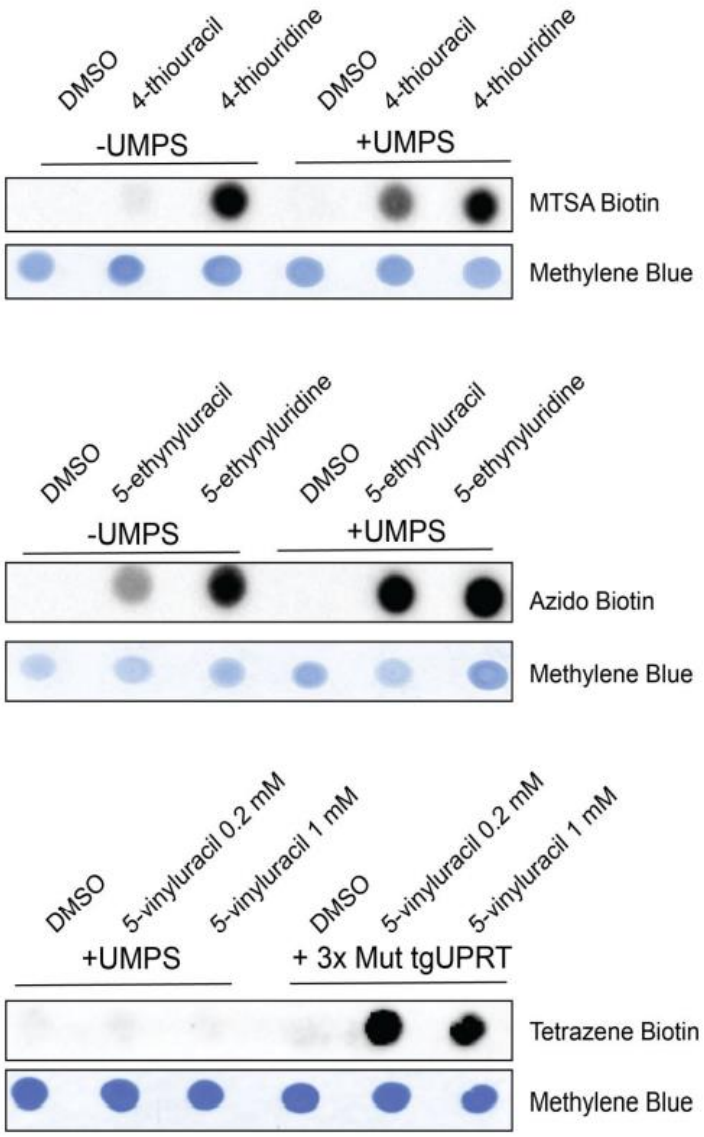


Figure 2-S5 Dot blot demonstrating incorporation of 1 and 4-thiouracil, but not 2 into cellular RNA with the overexpression of UMPS. Analogs were incubated for 24 hours at 1 mM concentration in HEK293T cells. RNA was extracted and biotinylated using MTSEA-Biotin, followed by dot-blot analysis. Strep-HRP = Streptavidin conjugated Horseradish peroxidase. MeBl = methylene blue staining serving as loading control.

Specific activity analysis of *Tg*UPRT variants with different uracil analog substrates in vitro

	Specific activity (IU/mg)					
	*Tg*UPRT	*Tg*UPRT$_{A168G}$	*Tg*UPRT$_{M166A}$	*Tg*UPRT$_{Y228A}$	*Tg*UPRT$_{A168G/Y228A}$	*Tg*UPRT$_{M166A/A168G/Y228A}$
Uracil[a]	4.92±0.17	0.26±0.05	3.57±0.09	3.18±0.11	12.49±0.18	7.20±0.23
5-ethynyl uracil (5-EtUra)[b]	1.00±0.08	0.88±0.09	0.48±0.05	0.80±0.03	10.40±0.22	8.00±0.17
5-vinyl uracil (5-VUra)[c]	n.d.	1.30±0.12	0.56±0.06	n.d.	5.0±0.09	4.80±0.12
5-azido uracil (5-AU)[b]	n.d.	n.d.	0.95±0.05	n.d.	4.72±0.19	3.01±0.09
5-azido methyl uracil (5AMU)[b]	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
5-bromovinyl uracil (5-BrVUra)[a]	0.10±0.01	0.12±0.03	0.10±0.02	0.10±0.01	2.80±0.25	2.34±0.15

Table 2-S2 Enzymatic synthesis of uridine-5'-monophosphate analogues catalyzed by TgUPRT variants.

Table S3	NMP (% peak area) (average)	Base (% peak area) (average)	Conversion (%)	µmol NMP	µmol MP/min (IU)	Spec.activity (IU/mg$_{enz}$)
Uracil [a]						
*Tg*UPRTWT	6.174	93.826	6.174	0.0049392	0.0024696	4.90±0.17
*Tg*UPRTA$_{168G}$	0.3276	99.6724	0.3276	0.00026208	0.00013104	0.26±0.05
*Tg*UPRT$_{M166A}$	4.4982	95.5018	4.4982	0.00359856	0.00179928	3.57±0.09
*Tg*UPRTY$_{228A}$	4.0068	95.9932	4.0068	0.00320544	0.00160272	3.18±0.11
*Tg*UPRT$_{A168G/Y228A}$	15.7374	84.2626	15.7374	0.01258992	0.00629496	12.49±0.18
*Tg*UPRT$_{M166A/A168G/Y228A}$	9.072	90.928	9.072	0.0072576	0.0036288	7.20±0.23
5-EtUra[b]						
*Tg*UPRTWT	1.26	98.74	1.26	0.001008	0.000504	1.00±0.08
*Tg*UPRTA$_{168G}$	1.1088	98.8912	1.1088	0.00088704	0.00044352	0.88±0.09
*Tg*UPRT$_{M166A}$	0.6048	99.3952	0.6048	0.00048384	0.00024192	0.48±0.05
*Tg*UPRTY$_{228A}$	1.008	98.992	1.008	0.0008064	0.0004032	0.80±0.03
*Tg*UPRT$_{A168G/Y228A}$	13.104	86.896	13.104	0.0104832	0.0052416	10.40±0.22
*Tg*UPRT$_{M166A/A168G/Y228A}$	10.08	89.92	10.08	0.008064	0.004032	8.00±0.17
5-VUra[c]						
*Tg*UPRTWT	n.d.	100	n.d.	n.d	n.d.	n.d.
*Tg*UPRTA$_{168G}$	1.638	98.362	1.638	0.0013104	0.0006552	1.30±0.12
*Tg*UPRT$_{M166A}$	0.7056	99.2944	0.7056	0.00056448	0.00028224	0.56±0.06
*Tg*UPRT$_{Y228A}$	n.d.	100	n.d.	n.d.	n.d.	n.d.
*Tg*UPRT$_{A168G/Y228A}$	6.3	93.7	6.3	0.00504	0.00252	5.00±0.09
*Tg*UPRT$_{M166A/A168G/Y228A}$	6.048	93.952	6.048	0.0048384	0.0024192	4.80±0.12
5-AU[b]						
*Tg*UPRTWT	n.d.	100	n.d.	n.d.	n.d.	n.d.
*Tg*UPRTA$_{168G}$	n.d.	100	n.d.	n.d.	n.d.	n.d.
*Tg*UPRT$_{M166A}$	1.197	98.803	1.197	0.0009576	0.0004788	0.95±0.05
*Tg*UPRTY$_{228A}$	n.d.	100	n.d.	n.d.	n.d.	n.d.
*Tg*UPRT$_{A168G/Y228A}$	5.9472	94.0528	5.9472	0.00475776	0.00237888	4.72±0.19
*Tg*UPRT$_{M166A/A168G/Y228A}$	3.7926	96.2074	3.7926	0.00303408	0.00151704	3.01±0.09
5-AMU[b]						
*Tg*UPRTWT	n.d.	100	n.d.	n.d	n.d.	n.d.
*Tg*UPRTA$_{168G}$	n.d.	100	n.d.	n.d	n.d.	n.d.
*Tg*UPRT$_{M166A}$	n.d.	100	n.d.	n.d.	n.d.	n.d.
*Tg*UPRTY$_{228A}$	n.d.	100	n.d.	n.d	n.d.	n.d.
*Tg*UPRT$_{A168G/Y228A}$	n.d.	100	n.d.	n.d.	n.d.	n.d.
*Tg*UPRT$_{M166A/A168G/Y228A}$	n.d.	100	n.d.	n.d.	n.d.	n.d.
5-BrVUra[a]						
*Tg*UPRTWT	0.126	99.874	0.126	0.0001008	0.0000504	0.10±0.01
*Tg*UPRTA$_{168G}$	0.1512	99.8488	0.1512	0.00012096	0.00006048	0.12±0.03
*Tg*UPRT$_{M166A}$	0.126	99.874	0.126	0.0001008	0.0000504	0.10±0.02
*Tg*UPRTY$_{228A}$	0.126	99.874	0.126	0.0001008	0.0000504	0.10±0.01
*Tg*UPRT$_{A168G/Y228A}$	3.528	96.472	3.528	0.0028224	0.0014112	2.80±0.25
*Tg*UPRT$_{M166A/A168G/Y228A}$	2.9484	97.0516	2.9484	0.00235872	0.00117936	2.34±0.20

Table 2-S3 Enzymatic synthesis of uridine-5'-monophosphate analogues catalyzed by *Tg*UPRT. Raw data for all data analyses for enzyme activity.

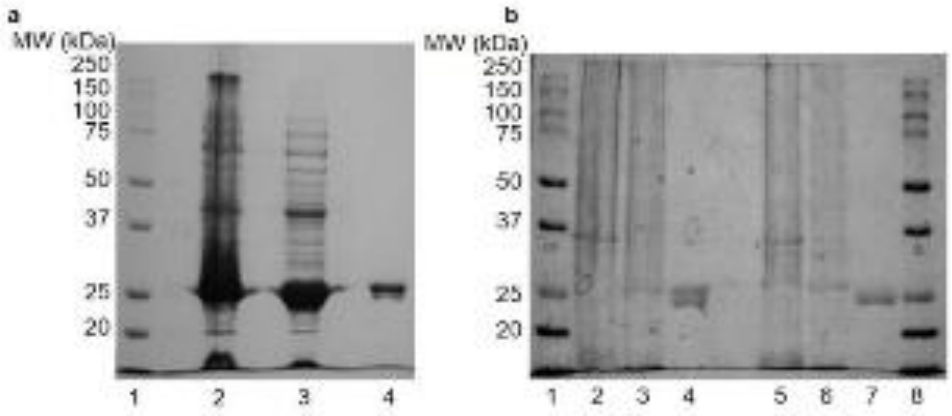
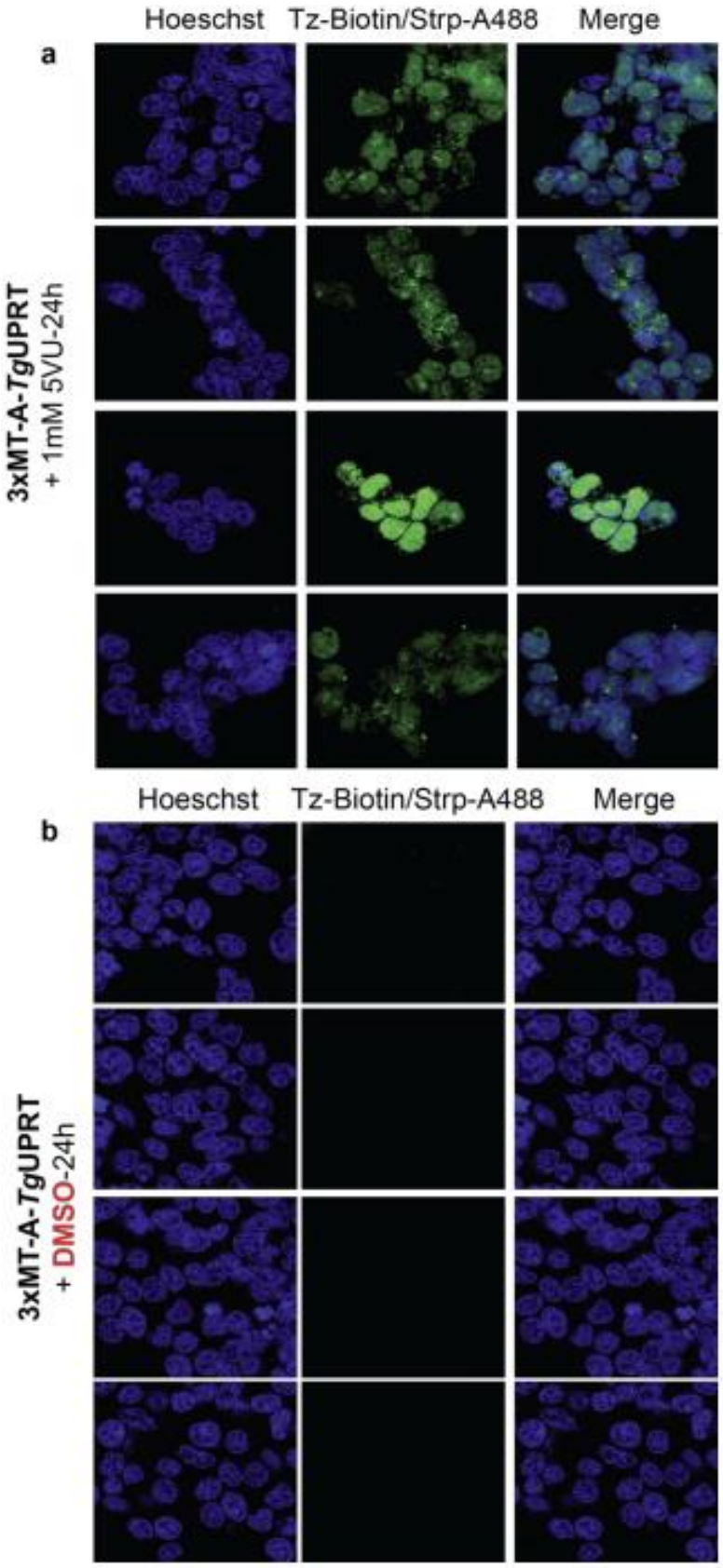
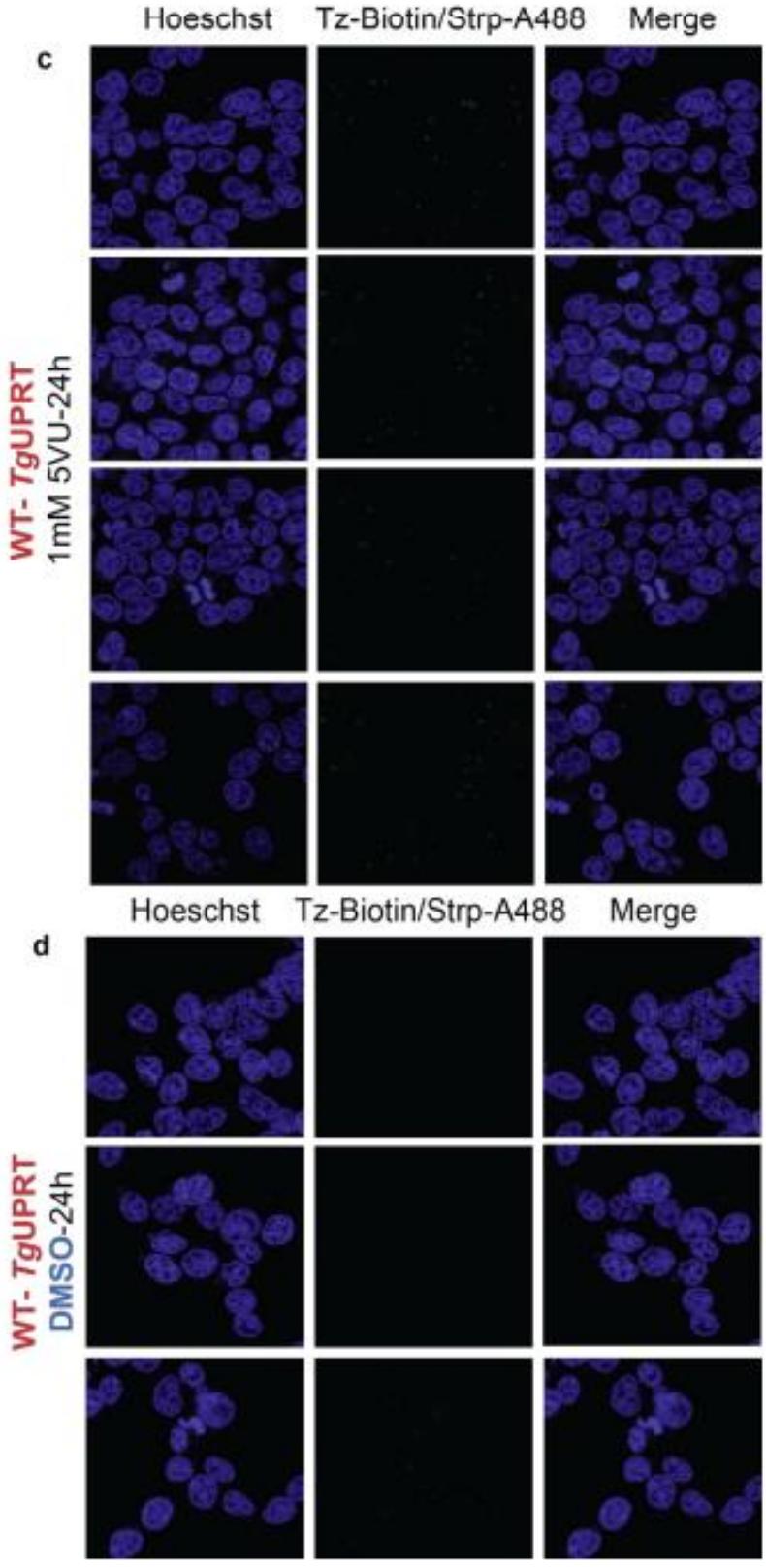


Figure 2-S6 SDS-PAGE analysis of the purification process of *Tg*UPRT enzymes. a. *Tg*UPRT wt. Lane 1. Precision Plus Protein™ prestained standard from BioRad used as molecular weight marker. Lane 2. Protein fraction obtained after cell lysis. Lane 3. Protein fraction obtained after centrifugation of the lysed sample. Lane 4. *Tg*UPRT wt after protein purification by affinity chromatography and size exclusion chromatography. b. SDS-PAGE analysis of the purification process of *Tg*UPRTA168G/Y228A and *Tg*UPRTM166A/A168G/Y228A. Lane 1. Precision Plus Protein™ prestained standard from BioRad used as molecular weight marker. Lane 2. Protein fraction containing TgUPRTA168G/Y228A obtained after cell lysis. Lane 3. Protein fraction obtained after centrifugation of the lysed sample containing *Tg*UPRTA168G/Y228A. Lane 4. *Tg*UPRTA168G/Y228A after protein purification by affinity chromatography and size exclusion chromatography. Lane 5. Protein fraction containing *Tg*UPRTM166A/A168G/Y228A obtained after cell lysis. Lane 6. Protein fraction obtained after centrifugation of the lysed sample containing *Tg*UPRTM166A/A168G/Y228A. Lane 7. *Tg*UPRTM166A/A168G/Y228A after protein purification by affinity chromatography and size exclusion chromatography. Lane 8. Precision Plus Protein™ prestained standard from BioRad used as molecular weight marker.


Hoeschst
Tz-Biotin/Strp-A488
Merge
a
3xMT-A-TgUPRT
+ 1mM 5VU-24h
Hoeschst
Tz-Biotin/Strp-A488
Merge
b
3xMT-A-TgUPRT
+ DMSO-24h


Hoeschst
Tz-Biotin/Strp-A488
Merge
c
WT- TgUPRT
1mM 5VU-24h
Hoeschst
Tz-Biotin/Strp-A488
Merge
d
WT- TgUPRT
DMSO-24h

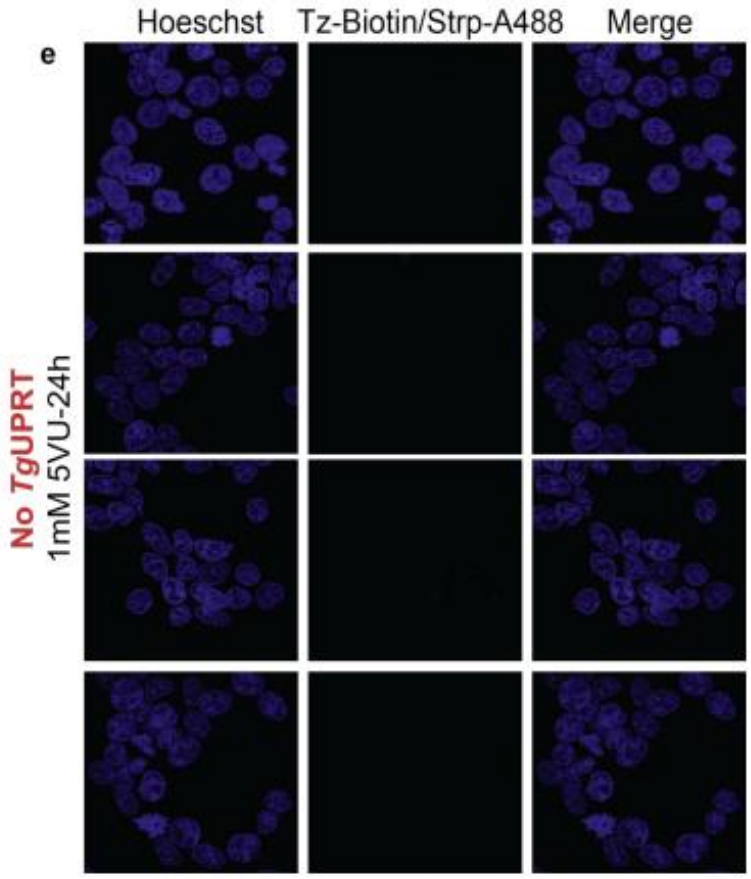


Figure 2-S7 Microscopy analysis RNA labeling using 5VU and HEK293T containing *Tg*UPRT variants. (a) Triple mutant 3xMT-A-*Tg*UPRT (M166A/A168G/Y228A) + 1 mM 5VU-24h, (b) triple mutant 3xMT-A-*Tg*UPRT + DMSO-24h, (c) WT-*Tg*UPRT + 1 mM 5VU-24h, (d) WT-*Tg*UPRT + DMSO-24h, (e) (-)*Tg*UPRT HEK293T + 1 mM 5VU-24h. Imaging was done via fluorescence confocal microscopy using a 63x oil immersion objective on a Leica 700 Carl Zeiss microscope. Hoeschst for staining nuclei, Tz-Biotin = tetrazine-biotin, Strp-A488 = Streptavidin-Alexa488. Merge = superimposed Hoeschst staining and Strep-A488 images.

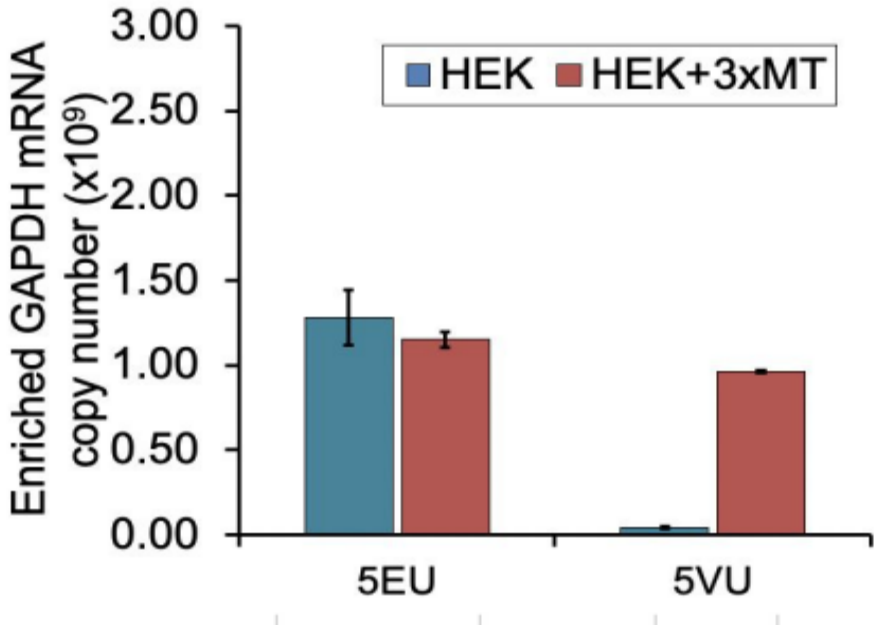


Figure 2-S8 RT-qPCR analysis of enriched CuAAC- or IEDDA-biotinylated GAPDH cDNA from (+/-) 3x-A-*Tg*UPRT HEK293T cells. Cells were treated with 1mM analog for 5 hours. RNA was isolated and biotinylated for enrichment, followed by cDNA synthesis and qPCR.

References

1 Landgraf, P., Antileo, E. R., Schuman, E. M., and Dieterich, D. C. (2015) BONCAT: metabolic labeling, click chemistry, and affinity purification of newly synthesized proteomes. *Methods Mol. Biol. 1266*, 199– 215.

2 Krogager, T. P., Ernst, R. J., Elliott, T. sS., Calo, L., Beranek, V., Ciabatti, E., Spillantini, M. G., Tripodi, M., Hastings, M. H., and Chin, J. W. (2018) Labeling and identifying cell-specific proteomes in the mouse brain. *Nat. Biotechnol. 36* (2), 156– 159.

3 Ernst, R. J., Krogager, T. P., Maywood, E. S., Zanchi, R., Beranek, V., Elliott, T. S., Barry, N. P., Hastings, M. H., and Chin, J. W. (2016) Genetic code expansion in the mouse brain. *Nat. Chem. Biol. 12* (10), 776– 778.

4 Barrett, R. M., Liu, H. W., Jin, H., Goodman, R. H., and Cohen, M. S. (2016) Cell-specific Profiling of Nascent Proteomes Using Orthogonal Enzyme-mediated Puromycin Incorporation. *ACS Chem. Biol. 11* (6), 1532– 6.

5 Li, Z., Zhu, Y., Sun, Y., Qin, K., Liu, W., Zhou, W., and Chen, X. (2016) Nitrilase-Activatable Noncanonical Amino Acid Precursors for Cell-Selective Metabolic Labeling of Proteomes. *ACS Chem. Biol. 11* (12), 3273– 3277.

6 Triemer, T., Messikommer, A., Glasauer, S. M. K., Alzeer, J., Paulisch, M. H., and Luedtke, N. W. (2018) Superresolution imaging of individual replication forks reveals unexpected prodrug resistance mechanism. *Proc. Natl. Acad. Sci. U. S. A. 115* (7), E1366– E1373.

7 Neef, A. B., Pernot, L., Schreier, V. N., Scapozza, L., and Luedtke, N. W. (2015) A Bioorthogonal Chemical Reporter of Viral Infection. *Angew. Chem. 127* (27), 8022– 8025.

8 Hubbard, S. C., Boyce, M., McVaugh, C. T., Peehl, D. M., and Bertozzi, C. R. (2011) Cell surface glycoproteomic analysis of prostate cancer-derived PC-3 cells. *Bioorg. Med. Chem. Lett. 21* (17), 4945– 50.

9 Rabuka, D., Forstner, M. B., Groves, J. T., and Bertozzi, C. R. (2008) Noncovalent cell surface engineering: incorporation of bioactive synthetic glycopolymers into cellular membranes. *J. Am. Chem. Soc. 130* (18), 5947– 53.

10 Chang, P. V., Prescher, J. A., Hangauer, M. J., and Bertozzi, C. R. (2007) Imaging cell surface glycans with bioorthogonal chemical reporters. *J. Am. Chem. Soc. 129* (27), 8400– 1.

11 Jao, C. Y. and Salic, A. (2008) Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci. U. S. A. 105* (41), 15779– 84.

12 Zheng, Y. and Beal, P. A. (2016) Synthesis and evaluation of an alkyne-modified ATP analog for enzymatic incorporation into RNA. *Bioorg. Med. Chem. Lett. 26* (7), 1799– 802.

13 Nainar, S., Beasley, S., Fazio, M., Kubota, M., Dai, N., Correa, I. R., Jr., and Spitale, R. C. (2016) Metabolic Incorporation of Azide Functionality into Cellular RNA. *ChemBioChem 17* (22), 2149– 2152.

14 Hida, N., Aboukilila, M. Y., Burow, D. A., Paul, R., Greenberg, M. M., Fazio, M., Beasley, S., Spitale, R. C., and Cleary, M. D. (2017) EC-tagging allows cell type-specific RNA analysis. *Nucleic Acids Res. 45* (15).

15 Abud, E. M., Ramirez, R. N., Martinez, E. S., Healy, L. M., Nguyen, C. H. H., Newman, S. A., Yeromin, A. V., Scarfone, V. M., Marsh, S. E., Fimbres, C., Caraway, C. A., Fote, G. M., Madany, A. M., Agrawal, A., Kayed, R., Gylys, K. H., Cahalan, M. D., Cummings, B. J., Antel, J. P., Mortazavi, A., Carson, M. J., Poon, W. W., and Blurton-Jones, M. (2017) iPSC-Derived Human Microglia-like Cells to Study Neurological Diseases. *Neuron 94* (2), 278– 293.

16 Islam, K. (2018) The Bump-and-Hole Tactic: Expanding the Scope of Chemical Genetics. *Cell Chem. Biol. 25* (10), 1171– 1184.

17 Yu, H., Li, J., Wu, D., Qiu, Z., and Zhang, Y. (2010) Chemistry and biological applications of photo-labile organic molecules. *Chem. Soc. Rev. 39* (2), 464– 73.

18 Nainar, S., Cuthbert, B. J., Lim, N. M., England, W. E., Ke, K., Sophal, K., Quechol, R., Mobley, D. L., Goulding, C. W., and Spitale, R. C. (2020) An optimized chemical-genetic method for cell-specific metabolic labeling of RNA. *Nat. Methods 17* (3), 311– 318.

19 Wang, D., Zhang, Y., and Kleiner, R. E. (2020) Cell- and Polymerase-Selective Metabolic Labeling of Cellular RNA with 2'-Azidocytidine. *J. Am. Chem. Soc. 142* (34), 14417– 14421.

20 Zhang, Y. and Kleiner, R. E. (2019) A Metabolic Engineering Approach to Incorporate Modified Pyrimidine Nucleosides into Cellular RNA. *J. Am. Chem. Soc. 141* (8), 3347– 3351.

21 Xie, R., Dong, L., Du, Y., Zhu, Y., Hua, R., Zhang, C., and Chen, X. (2016) In vivo metabolic labeling of sialoglycans in the mouse brain by using a liposome-assisted bioorthogonal reporter strategy. *Proc. Natl. Acad. Sci. U. S. A. 113* (19), 5173– 8.

22 Vinogradov, S. V. (2007) Polymeric nanogel formulations of nucleoside analogs. *Expert Opin. Drug Delivery 4* (1), 5– 17.

23 Balimane, P. V. and Sinko, P. J. (1999) Involvement of multiple transporters in the oral absorption of nucleoside analogues. *Adv. Drug Delivery Rev. 39* (1–3), 183– 209.

24 Tomorsky, J., DeBlander, L., Kentros, C. G., Doe, C. Q., and Niell, C. M. (2017) TU-Tagging: A Method for Identifying Layer-Enriched Neuronal Genes in Developing Mouse Visual Cortex. *eNeuro 4* (5), ENEURO.0181-17.2017.

25 Gay, L., Miller, M. R., Ventura, P. B., Devasthali, V., Vue, Z., Thompson, H. L., Temple, S., Zong, H., Cleary, M. D., Stankunas, K., and Doe, C. Q. (2013) Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. *Genes Dev. 27* (1), 98– 115.

26 Basnet, H., Tian, L., Ganesh, K., Huang, Y. H., Macalinao, D. G., Brogi, E., Finley, L. W., and Massague, J. (2019) Flura-seq identifies organ-specific metabolic adaptations during early metastatic colonization. *eLife 8*, e43627.

27 Nguyen, K., Fazio, M., Kubota, M., Nainar, S., Feng, C., Li, X., Atwood, S. X., Bredy, T. W., and Spitale, R. C. (2017) Cell-Selective Bioorthogonal Metabolic Labeling of RNA. *J. Am. Chem. Soc. 139* (6), 2148– 2151.

28 Kubota, M., Nainar, S., Parker, S. M., England, W., Furche, F., and Spitale, R. C. (2019) Expanding the Scope of RNA Metabolic Labeling with Vinyl Nucleosides and Inverse Electron-Demand Diels-Alder Chemistry. *ACS Chem. Biol. 14* (8), 1698– 1707.

29 Rieder, U. and Luedtke, N. W. (2014) Alkene-tetrazine ligation for imaging cellular DNA. *Angew. Chem., Int. Ed. 53* (35), 9168– 72.

30 Knall, A. C. and Slugovc, C. (2013) Inverse electron demand Diels-Alder (iEDDA)-initiated conjugation: a (high) potential click chemistry scheme. *Chem. Soc. Rev. 42* (12), 5131– 42.

Chapter 3: *In Vivo* RNA Metabolic Labeling with Increased Cell-Specificity

Introduction

Advancements in transcriptomics have revealed a myriad of functional RNA involved in gene regulation. Deconvolution of RNA pathways is increasingly complex when analyzing RNA expression in multicellular contexts. To achieve time-resolution of transcription, RNA metabolic labeling methods are used to tag nascent RNA for downstream affinity-based enrichment and analysis. Herein, we describe RNA metabolic labeling with additional cell-specific resolution of RNA-tagging by the significant reduction of background, non-specific labeling when applied into model organisms. Through stable expression of an engineered metabolic enzyme (3xUPRT), we demonstrate MDA-MB-231 LM2 cells selectively catalyze 5-vinyluracil resulting in cell-specific RNA metabolic labeling. We report using this new approach in a metastatic human breast cancer mouse model to demonstrate the capabilities for profiling cell-specific nascent RNA in vivo for further investigation of RNA mechanisms.

During the last decade, RNA has been redefined as a macromolecule with diverse biological functions that drive many physiological processes[1]. Identifying regulatory RNA in a specific cell-type remains particularly challenging within *in vivo* experimental models due to the presence of many cell types. To capture RNA from cell-types of interest, extensive dissociation and isolation steps are required prior to capturing RNA for analysis[2]. These approaches rely on mechanical separation, enzymatic digestion or flow cytometry which are known to introduce transcriptional artifacts, as the cellular transcriptome is transformed after removal from the original whole-organism environmental niche[3].

Recapitulation of RNA expression pathways from intact, living cells within model organisms is difficult to analyze with current methods due to these technical challenges.

RNA metabolic labeling methods tag nascent RNA within living cells which provide opportunities for capturing transcriptional information present during key biological events. These methods expose cells to non-canonical nucleobase or nucleosides derivatized with chemical handles for subsequent metabolic incorporation into polymerizing RNA[4]. Isolation of total RNA includes a portion of metabolically-tagged nascent transcripts that are reactive with biotin-conjugated, orthogonal chemical reagents for subsequent streptavidin affinity-based enrichment and analysis. Attenuating RNA metabolic labeling into specific cell-types can be achieved by expressing non-mammalian exogenous metabolic enzymes that will selectively catalyze nucleobase analogs.

Our lab has demonstrated cell-specific RNA metabolic labeling with uracil phosphoribosyl transferase (UPRT) expression (from *Toxoplasma gondii), Tg*UPRT, paired with 5-ethynyluracil (5-eu) consisting of a more stable, bioorthogonal chemical handle for imaging experiments with fluorophores and biochemical affinity-based separation of nascent transcripts[5-6]. RNA labeled with ethynyl groups can react with biotin-azide through Cu(I)-catalyzed azide-alkyne addition (CuAAC) which can deteriorate RNA integrity due to the production of free metal-ion radicals[7]. Recent findings identified endogenous metabolic enzymes in the *de novo* pyrimidine biosynthetic pathway will catalyze 5-eu in mammalian cell types without UPRT expression limiting the applications for *in vivo* cell-specific RNA metabolic labeling[8]. These results necessitate the development of a more stringent enzyme-analog pair fitting for cell-specific RNA metabolic labeling that can eliminate the

background, non-specific labeling demonstrated by previous nucleobase analogs paired with *Tg*UPRT.

Results

To overcome challenges with 5-eu background RNA metabolic labeling in other mammalian cell types, our lab engineered the active site of *Tg*UPRT to generate triple-mutant *Tg*UPRT (3xUPRT) capable of catalyzing 5-vinyuracil (5-vu) into nascent, vinyl-labeled RNA[9]. Advantages of using 5-vu include increased cell-specific stringency of labeling, as 5-vu is not recognized as a metabolic substrate by endogenous mammalian enzymes[9]. Additionally, the vinyl-handle reacts selectively with tetrazines through inverse electron-demand Diels-Alder which does not affect RNA integrity[9]. Finally, 5-vinyluridine (5-vuD) treatment is cells is less detrimental to proliferation at longer incubation times compared with 5-ethynyluridine for RNA metabolic labeling applications in cells treated with each analog for >12 hours[10]. These findings support use of the 3xUPRT enzyme paired with 5-vu for improved cell-type specific RNA labeling within a broad range of applications in animal models and for maintaining the level of RNA integrity necessary for downstream sequencing analysis.

We predicted the 3xUPRT/5-vu cell-specific RNA labeling approach would be advantageous for profiling RNA in animal models where it is difficult to capture nascent RNA expression within specific cell types without inducing transcriptional artifacts. Additionally, these technical improvements improve the flexibility of these protocols to include longer incubations beyond a few minutes and at high concentrations which may be a necessity in the larger context of a whole organism to label a specific cell population. To

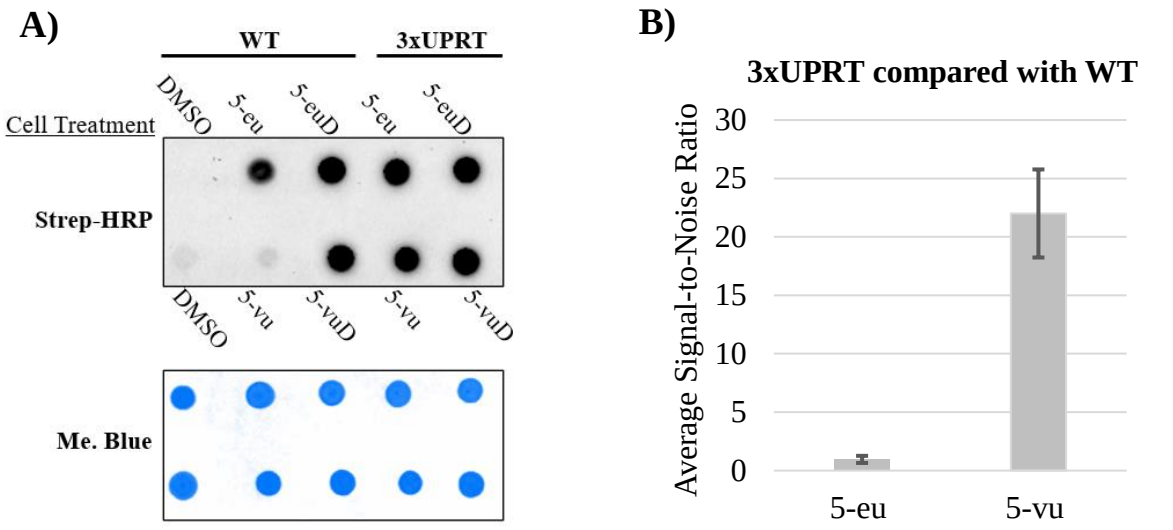


Figure 3-1 Characterizing Cell-Specific RNA metabolic labeling in MDA-MB-231 cells. 3h cell treatment with 1 mM of DMSO (negative control), uracil analog, or uridine analog (positive control) in WT or (+)-3xUPRT cell lines. A) RNA dot blot is performed after reacting with biotin-azide (for ethynyl RNA) or biotin-tetrazine (for vinyl RNA). RNA was crosslinked to membrane and incubated with Streptavidin-horseradish peroxidase (Strep-HRP) followed by staining with methylene blue (Me. Blue) as a loading control. 5-eu= 5-ethynyluracil, 5-euD= 5-ethynyluridine, 5-vu= 5-vinyluracil, 5-vuD=5-vinyluridine. B) ImageJ quantification of chemiluminescence was used to calculate signal/noise ratios.

test this, 3xUPRT was stably and constitutively expressed in highly metastatic MDA-MB-231 LM2 (LM2) cells using CRISPR/Cas9 genome-editing[11]. LM2 human breast cancer cells are capable of preferentially metastasizing to the lung in mouse xenograft models without overburdening mice with the primary tumor[12]. Before applying this approach to mouse models, *in vitro* RNA metabolic labeling was characterized with dot blot analysis. LM2 wild-

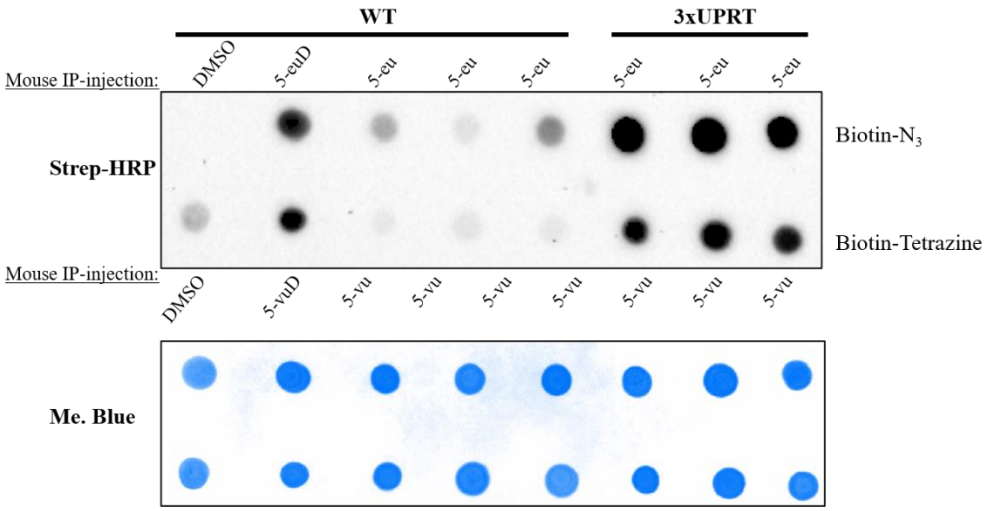
type (WT) and (+)-3xUPRT cells were incubated with 5-eu or 5-vu at 1 mM for 3 hours. As a positive control for RNA metabolic labeling, 5-ethynyluridine (5-euD) and 5-vinyluridine (5-vuD) were included in these experiments which label RNA in all cells regardless of 3xUPRT expression. Finally, DMSO was used as a negative control in cell treatments as the solvent of choice for these water-insoluble molecules. The chemiluminescent signal from these dot blots was quantified with ImageJ to compare the differences in signal-to-noise between 5-eu and 5-vu treatments in (+)-3xUPRT LM2 cells compared with WT cells. To verify 3xUPRT expression in cells, mCherry was co-expressed with 3xUPRT, whereas GFP is expressed in all the LM2 cell types in this study[12] **(Figure 3-1)**.

Prior to establishing xenograft mouse models, optimization for *in vivo* RNA metabolic labeling protocols were tested in (WT) C57Bl/6J mice. Rigorous testing of the solubility of these compounds in DMSO was shown to be maximized at 500 mM resulting in using this concentration for intraperitoneal (IP) injections. Initially, 5-vinyluridine (5-vuD) was tested in two sets of WT mice which demonstrated strong reproducibility of labeling across different tissue types depicted with RNA dot blot analysis **(Figure 3-S1)**. Next, RNA labeling was performed in mice treated with uracil analogs, 5-eu and 5-vu, and uridine analogs, 5euD and 5-vuD, as positive controls. These results clearly indicate stronger background labeling with 5-eu treatment compared with 5-vu over a 24-hour treatment period indicating 5-vu is advantageous for improved cell-specificity of RNA labeling in mice (**Figure 3-S2)**. In order to determine the range of background labeling with 5-vu, two biological replicates of WT mice were tested with increasing time points and number of injections after which organs were surveyed with RNA dot blot analysis. Although pancreas-derived RNA was extensively degraded due to high concentrations of

endonucleases[13], the majority of the RNA was compared effectively with this study to conclude that single-injection treatments at 1-hr have significantly reduced background labeling compared to >6-hour time points with >2 injections **(Figure 3-S3)**. As a result, the 3-hour time point, 500 mM injection was chosen as the optimal method [5-vu dose (150 mg/kg) and 5-vuD (300 mg/kg)] for testing cell-specificity of RNA labeling in mouse tumor xenografts.

Using the stable LM2 WT (negative control) and (+)-3xUPRT cells, mouse xenografts were generated through mammary-fat pad implantation into NSG female mice. After 3-4 weeks, tumors were visible and mice were IP-injected for 3 hours with 5-eu or 5-vu in triplicate. 5-euD and 5-vuD were similarly IP-injected as positive controls into one replicate mouse with WT LM2 tumors. After sacrificing mice, tumor and organ RNA were isolated and reacted with biotin-tetrazine, purified and analyzed with RNA dot blot to determine the extent of labeling **(Figure 3-2a and 3-S4)**. Signal-to-noise ratios in mouse xenografts were lower with more variability after 5-eu treatments compared with 5-vu treatments **(Figure 3-2b)**. To determine if 5-vu background could be reduced to any extent, 2-fold reduced titrations were included in (+)-3xUPRT tumor containing mice **(Figure 3-S5a)**. *Ex vivo* mCherry fluorescence shows similar size and indicates consistent expression of 3xUPRT across these tumors **(Figure 3-S5b)**. Flow cytometry was used to quantify GFP fluorescence in all LM2 cells used within mouse xenografts and mCherry to quantify (+)-3xUPRT positive cell populations prior to xenotransplantation experiments **(Figure 3-S6)**.

A)



B)

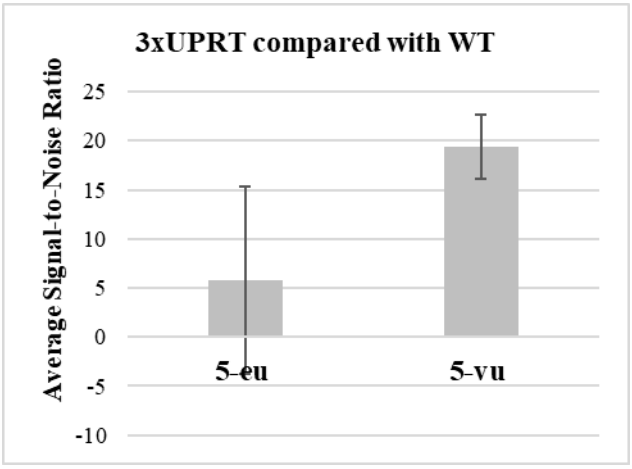


Figure 3-2 Cell-specific RNA metabolic labeling in mouse xenograft WT and (+)-3xUPRT LM2 primary tumors. 3-week xenografted mice were IP-injected with 500 mM of each indicated substance for 3 hours in biological triplicates for uracil analog treatments. A) RNA was extracted from tumors, reacted with Biotin-N_3 with ethynyl-RNA and Biotin-tetrazine with vinyl-RNA prior to dot blot analysis. B) ImageJ quantification of chemiluminescence was used to calculate signal/noise (n=3).

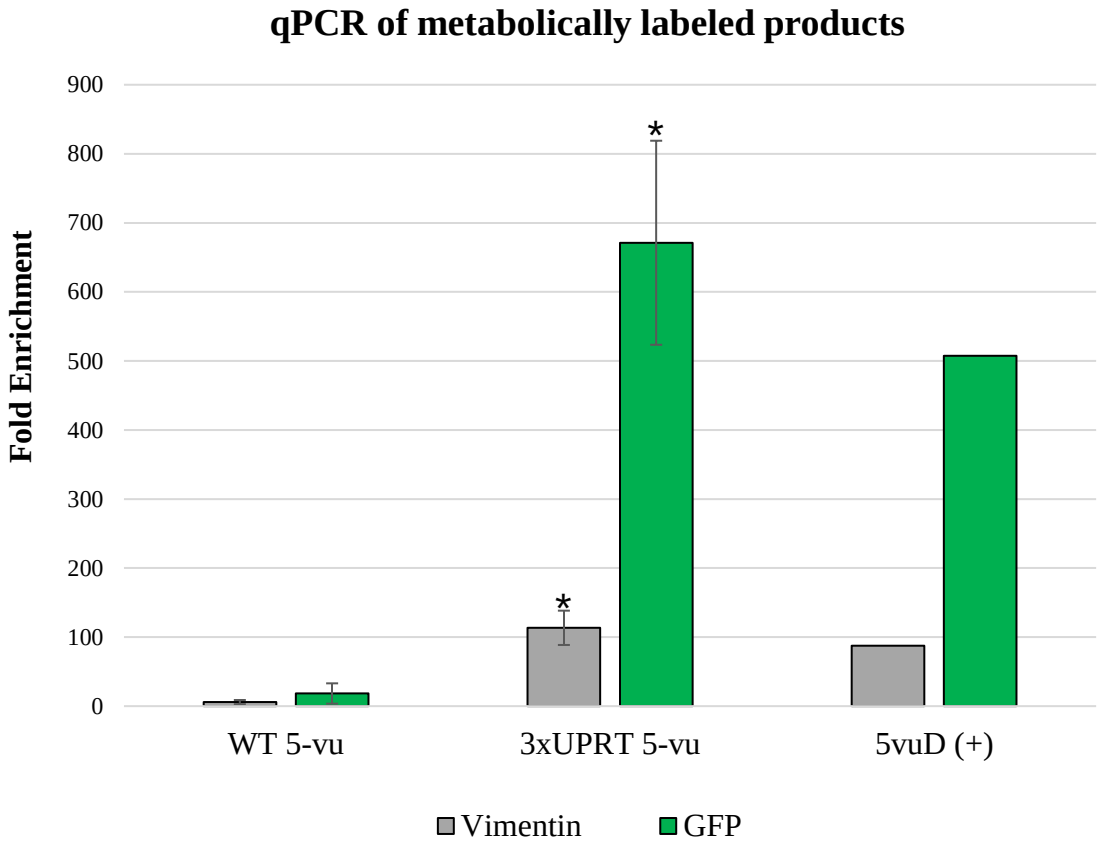


Figure 3-3 Streptavidin bead enrichment of biotin-RNA:cDNA for qPCR analysis. Biotinylated RNA was reverse transcribed to make in-tact RNA-cDNA which were subsequently enriched with streptavidin beads, eluted by RNase hydrolysis and quantified with qPCR. Fold enrichment was determined through 2^{-dCT} standardized to the untreated mouse for vimentin and GFP labeled RNA. Statistical significance relative to enrichment from untreated mouse was determined using a one-tailed Student's *t* test indicated as follows: P<0.1; *. WT=wild-type, UPRT=3xUPRT, 5-vu=5-vinyluracil, 5vuD=5-vinyluridine.

To quantify the biochemical enrichment of metabolically labeled transcripts from WT and (+)-3xUPRT tumors, RNA was subjected to reverse transcription after reaction with tetrazine-biotin. Bypassing the RNA-H step, RNA:cDNA hybrids were incubated with MyOne C1 streptavidin magnetic beads for separation and elution of cDNA from bead with RNAse and heat treatment; eluted cDNA was quantified with qPCR. Endogenous vimentin, known to be highly expressed in MDA-MB-231 cells[14], and GFP, which is exogenously expressed within all MDA-MB-231 LM2 cells[12] in these experiments were both amplified with qPCR. Untreated mouse RNA was included as a negative control and 5-vinyluridine treated WT tumors as a positive control for enrichment. The fold enrichment was calculated by 2^{-dCT} with normalization to the untreated mouse enrichment levels for each

detected gene[8]. A minimum of 10-fold enrichment for both gene of interest was

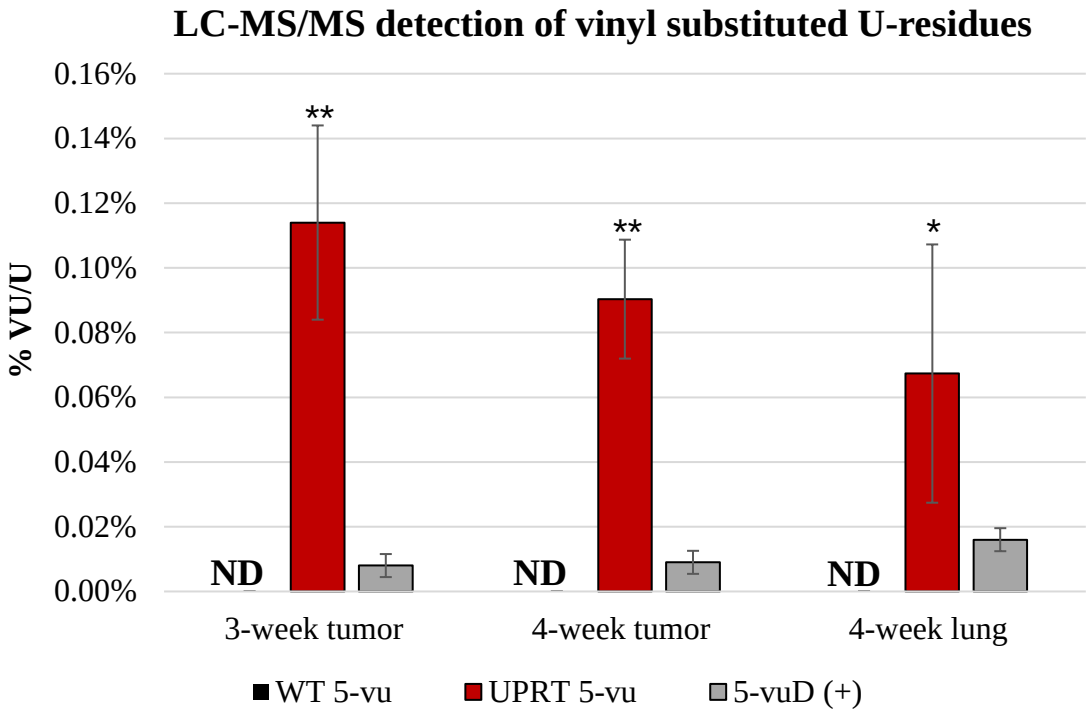


Figure 3-4 LC-MS/MS for analysis of vinyl-U modification from tumors and metastatic lungs. Purified RNA was analyzed for % of vinyl-U substation of total U normalized to untreated mouse tissue. 3-week xenograft mouse lungs did not produce metastases and were not included in the analysis. Biological triplicates were used in this data set. Statistical significance relative to wild-type tumors and metastases was determined using a one-tailed Student's *t* test indicated as follows: P<0.05; **, P<0.5 *. WT=wild-type, UPRT=3xUPRT, 5-vu=5-vinyluracil, 5vuD=5-vinyluridine, ND=not detected.

reproducibly detected in (+)-3xUPRT/5-vu treated tumor samples **(Figure 3-3).**

To quantify vinyl-substituted RNA, LC-MS experiments were used to detect the level of vinyl-U relative to total-U in tumor and lung samples. Both WT and (+)-3xUPRT tumor

RNA from 3-week and 4-week xenografts were quantified using this method to determine % vinyl substitution in uridine residues with metabolic labeling. Due to appreciable levels of lung metastases detected in 4-week xenografts, only 4-week lung RNA were included for this analysis. The ratio of vinyl-U/total-U was used to determine the % vinyl substitution rates across biological triplicates. WT tumor and lung-derived RNA from 5-vuD treated mice was used as a positive control. These samples were normalized to RNA from mouse tissues without any analog treatment **(Figure 3-4).** The percent incorporation of VU/U is an average of 0.1% which is remarkable considering *in vitro* vinyl-incorporation is 0.8% for 5-vud[10] RNA labeling in cells that are subjected to higher local analog concentrations and incubation times compared with those used in our mouse xenograft studies.

In sum, the results from LM2 cell-specific RNA metabolic labeling in mouse xenografts support using the 3xUPRT/5-vu pair in living animals through similar approaches. Our results show this method reduces the non-specific RNA metabolic labeling in other cell-types within tumor and lung tissues. To add, vinyl-modified analogs are less detrimental to the model system, due to reduced toxicity which allows for longer labeling with 5-vu incubation times. Furthermore, the orthogonal tetrazine chemistry is shown to be less detrimental to RNA integrity compared to CuAAC extending these applications for reliable sequencing analysis. The selectivity of metabolic labeling in LM2 cells is highlighted by significant levels of enrichment of cell specific genes through qPCR analysis.

Finally, cell-specific metabolic RNA labeling was tested with tetrazine-Cy5 to conjugate nascent, vinyl-labeled RNA in mouse tumor and lung tissue sections. Tumor sections show uniformity of mCherry expression indicating cells with 3xUPRT expression.

4-week lung sections demonstrate dispersed metastatic (+)-3xUPRT LM2 cells within lung tissue and co-localization of Cy5 signal demonstrating the level of cell-specificity achieved when captured through confocal microscopy. No background signal was detected in WT tumor sections reacted with tetrazine-Cy5 and the surrounding WT cell-types in the lung tissue. Although overall fluorescent dye signal was reduced in lung tissue compared to tumor tissue, this can be attributed to the lungs being further from the IP-injection site which may have reduced bioavailability of 5-vu. These results clearly show that only (+)-3xUPRT cells in mouse tissues label nascent RNA with 5-vu, which can be detected with tetrazine-Cy5. These findings support the utility for 3xUPRT/5-vu for *in vivo* imaging of cell-specific nascent RNA **(Figure 3-5 and Figure 3-S7)**.

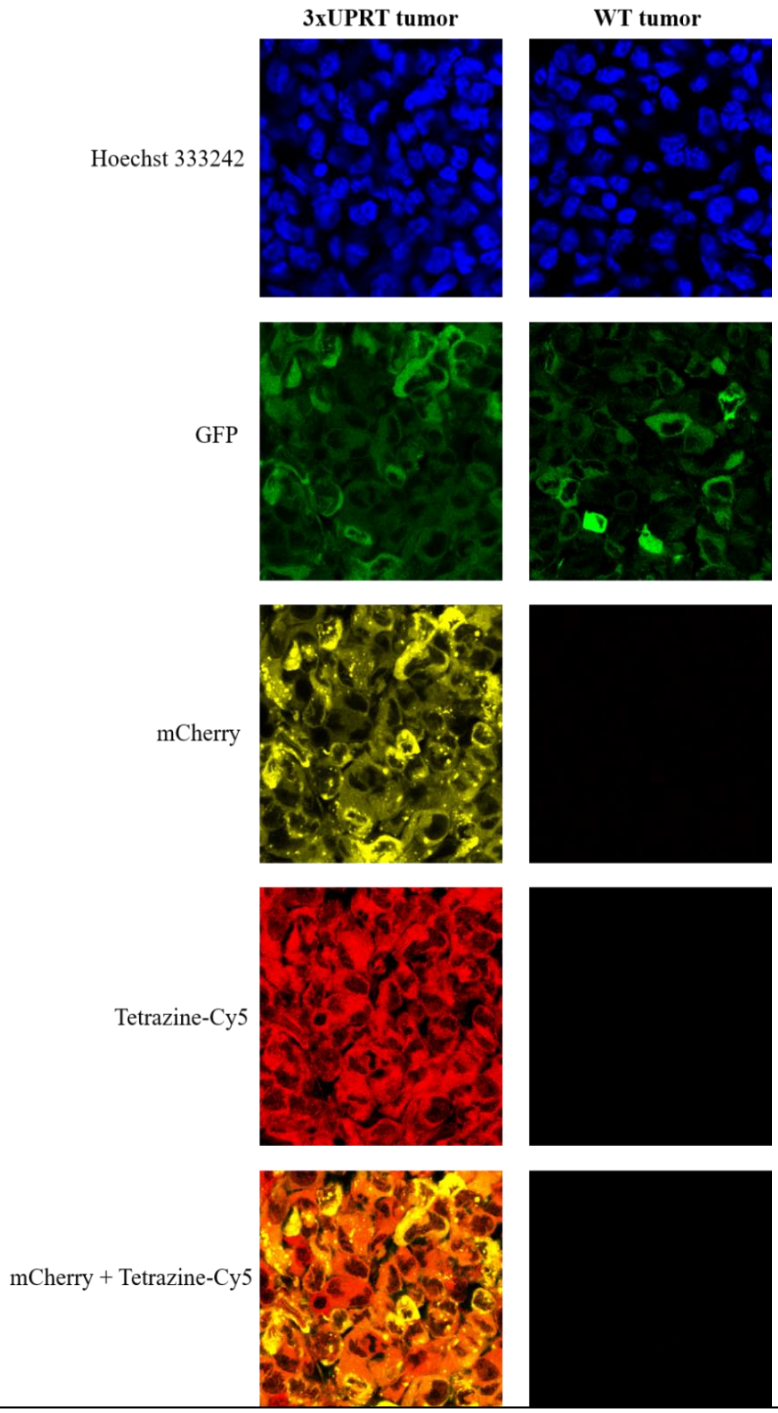


Figure 3-5 Metabolic labeling in (+)-3xUPRT-mCherry MDA-MB-231 LM2 cells in tumors tissues. Vinyl-containing RNA are fluorophore-conjugated with tetrazine-Cy5. GFP signal is present in WT MDA-MB-231 LM2 cells (parental line). mCherry indicated (+)-3xUPRT MDA-MB-231 cells.

Conclusions

The cell-specific RNA metabolic labeling method described herein has the potential to transform the ability to analyze nascent RNA *in vivo* using relatively common laboratory techniques. The most technically challenging aspect involved is to generate the stable expression of 3xUPRT into cell-types of interest which can be performed with CRISPR/Cas9 genome-editing strategies. Secondly, 5-vu, biotin-tetrazine, Cy5-tetrazine are commercially available for researchers to independently access if they lack resources to synthesize these reagents. Lastly, through the expression downstream of gene-specific promoters, 3xUPRT expression can label RNA when a gene-of-interest is actively expressed thereby extending the utility of these tools to capture and visualize nascent RNA in a gene-specific context.

As a proof-of-principle, we generated MDA-MB-231 LM2 cells that stably express 3xUPRT driven by the endogenous vimentin promoter. We chose to generate vimentin-3xUPRT for the purpose of labeling RNA important for the maintenance of mesenchymal cell states which occurs during the epithelial-to-mesenchymal transition in early metastatic progression [14] **(Figure 3-S8)**. The extent of labeling in vimentin-3xUPRT LM2 cells is overall comparable in chemiluminescent signal to the constitutively-expressed (+)-3xUPRT LM2 cells used throughout this study and to the 5-vuD positive control signal. These exciting results support using this experimental approach for analyzing actively transcribed RNA pertinent to the epithelial-to-mesenchymal transition in greater detail by co-expressing 3xUPRT with other relevant genes and transcription factors of interest with the expectation of reduced non-specific, background RNA labeling shown throughout this work. These results have encouraging potential for researchers to design unique *in vivo*

cell-specific RNA metabolic labeling experiments for a model organism of interest in order to elucidate complex transcriptional pathways in important cell-types. As the comprehension of RNA mechanisms deepens, more possibilities will become present for designing highly cell-specific RNA-targeting therapeutics that interfere in these pathways to prevent the progression of aggressive diseases, such as human metastatic breast cancer.

Experimental Methods

Cell lines used in the study: MDA-MB-231 LM2 4175 cells[11] were cultured in DMEM containing high glucose (4.5 g/L) and L-glutamine (Corning Ref# 10-017-CM) supplemented with 10% FBS, 1% penicillin-streptomycin (complete DMEM). These were purchased from Memorial Sloan Kettering Cancer Center (MSKCC) and developed in the lab of Joan Massagué. Cells were grown at 37 °C, 5% CO_2. Stable 3xUPRT-mCherry cell lines were selected in 1-10 μg/ml puromycin for selection in complete DMEM.

IP-injection experiments in mice. NSG or C57/Bl mice were treated with nucleoside or nucleobase analogs intraperitoneally by injection into the lower left quadrant of the abdomen (peritoneum), lateral to the animal's midline at a <45-degree angle. Injections did not exceed 70 μl. For uridine analogs, 300 mg/kg were used per injection. For uracil analogs, 150 mg/kg were used per injection. DMSO was used as a solvent to avoid precipitation of analogs after injection. Optionally, researchers may use 95% corn oil and 5% DMSO only for 5-vu injections. WT mice were treated for 1-24 hours for optimizing these studies. Xenograft mice were treated for 3 hours before sacrificing with CO_2 and cervical dislocation. Animal use protocols were reviewed and approved by the Institutional Animal Care and Use Committee of University of California, Irvine.

Mouse orthotopic xenografts for metastatic breast cancer and organ harvesting (followed by RNA preservation): NSG female mice were shaved at the injection site and sterilized with ethanol wipes. Prior to injection, isoflurane levels of 1.5-2% were was used for anesthesia prior to cell transplantation. Mammary fat pad L4 was injected with $1x10^6$ MDA-MB-231 LM2 cells in (WT or with 3xUPRT expression) with 100 µl PBS into the at 14-weeks of age. These mice grew tumors for 3-4 weeks to develop until 1000 mm^3 volumes maximum. Orthotopic injections into the fat pad were performed without incision by direct injection into the site of the mammary fat pad, below the nipple and into the fatty layer[17]. Post-operative care included heat pad for anesthesia recovery and acetaminophen (1.6 ml oral suspension into 150 ml drinking water). Xenograft mice were IP-injected for 3 hours with analogs (as previously indicated) before sacrificing with CO_2 and cervical dislocation for harvesting organs and tumors. After dissection, organs were immediately places into RNAlater (Thermofisher AM7021) solution and cut into 0.5 cm segments, submerged in solution for 24-48 hours at 4°C. Following this, tissue was minced avoiding RNAse contamination with RNAse Zap (Thermofisher AM9780) and directly homogenized with 0.9-2mm RNAse-free stainless-steel beads (Fisher NC0586331) in 1 mL TriZol (Thermofisher #15596026) for 1-minute increments. Between 1-minute increments, tubes were placed on ice, and homogenized for 2-3 more times. Homogenization tubes were centrifuged for 5 minutes prior to lysate collection and processed following manufacturer's instruction for RNA extraction. RNA was reconstituted in nuclease-free water and the concentration analyzed with Nanodrop 2000 spectrophotometer.

RNA biotinylation using biotin-tetrazine click reaction: 5-15 µg of total cellular, tumor or organ derived DNAse-treated RNA was reacted in solutions with 1 mM tetrazine-biotin (Sigma #793329) or 1 mM DMSO (negative controls), all of which were 10% DMSO final reactions in nuclease-free water and incubated in dark conditions. Tetrazine-biotin stocks were all dissolved and prepared in DMSO as 10 mM solution stocks stored in -20°C using amber tubes. RNA was purified after biotinylation using RNA Clean and Concentrator Kit #5 (Zymo Research #R1013) and eluted in 20 µl of nuclease-free water prior to dot blot analysis.

RNA dot blot analysis with chemiluminescence detection[9]: To detected metabolically labeled RNA, biotinylated RNA was crosslinked to a Hybond-N+ (GE Healthcare) membrane with 254 nm light using a UV stratalinker (Stratagene). The membrane was first equilibrated in 2x SSC buffer before spotting RNA and fixing with UV-light. After UV-crosslinking the RNA, membranes were blocked in blocking buffer[15] (0.12 M NaCl, 0.016 M Na2HPO4, 0.008 M NaH2PO4, 0.17 M SDS) for 30 minutes and followed by incubation with high sensitivity streptavidin-HRP (Fisher Scientific, Cat#: PI21130) at 1:5000 dilution in blocking buffer for 5 minutes. The membrane was washed twice in a Wash A buffer (1:10 dilution of blocking buffer in D.I. water) for 30 minutes and twice in Wash B buffer (0.1 M Tris-base, 0.1 M NaCl, 0.02M MgCl2, pH 9.5) for 5-10 minutes. Membranes were then incubated for 1-5 minutes in ECL Chemiluminescent Substrate (Fisher Scientific, Cat#: PI32106) and imaged on a ChemiDoc MP imaging system (Bio-Rad) to detect chemiluminescence on the membrane. After imaging is completed, methylene blue (0.04% methylene blue and 0.3 M sodium acetate) was used to stain onto membranes overnight, rinsed briefly with methanol, then D.I. water to remove excess stain, followed by

colorimetric imaging with the ChemiDoc MP imaging systems to detect consistent loading of RNA onto the membranes.

Enrichment of metabolically labeled RNA:cDNA for qPCR analysis[8]: RNA was purified after biotinylation as described before dot blot analysis and treated with 2U DNase TURBO (Invitrogen, RNAse free) for 30 min at 37°C followed by column purification using Zymo iic columns and eluted in nuclease-free water. Purified RNA was quantified by Nanodrop 2000 and 500 ng or 1ug of RNA included in for reverse transcription using SuperScript IV (ThermoFisher), following the manufacturer's protocol except for the RNAse-H step. The resulting RNA:cDNA hybrids were gently mixed and separated into equal volumes. One half was used as input and the other half for streptavidin magnetic bead separation. 10 µl of MyOne C1 streptavidin beads (ThermoFisher) were aliquoted per enrichment and washed 4x with binding buffer (1 M NaCl, 100 mM Tris-HCl, pH 7.0, 10 mM EDTA and 0.2% (vol/vol) Tween-20), then blocked in 10 µl of RNAse-free glycogen (Fisher FERR0551) for 1 hour at room temperature with adequate mixing (end-over-end). After blocking, beads were washed 2x in binding buffer, resuspended in 10 µl with equal volume of RNA:cDNA sample. To ensure improved mixing, samples were pipetted every 3-5 minutes over a 20-minute period at room temperature. Then, beads were immobilized by a Dynabeads 1.5 mL tube magnetic rack. Supernatant was removed and beads were washed 4x in high salt wash buffer ((4 M NaCl, 100 mM Tris-HCl, pH 7.0, 10 mM EDTA and 0.2% (vol/vol) Tween-20) followed by 2x washed in nuclease-free water. Beads were centrifuged and resuspended in 50 elution buffer (10 U RNase H, 1× RNase H buffer (New England Biolabs), 4 µg RNase A/T1 mix (Thermo Fisher Scientific) and 12.5 mM D-biotin) and incubated at 37 °C for 30 min at 750 rpm. Then 1 µl of DMSO was added and beads

were heated at 95 °C for 8 minutes. The eluted cDNA and the input cDNA were equally purified at the same time using Zymo DNA Clean and Concentrator-5 kit, with avoidance of carryover and guanidine contamination through extra ethanol washes, and eluted in 40 µl of nuclease-free water. The resulting cDNA was quantified with using Luna qPCR master mix (New England Bioland) with primers specific for human vimentin and GFP using a CFX Connect Real-Time PCR Detection System (Bio-Rad). Raw Cq were quantified for fold enrichment using 2^{-dcT} normalized to the negative control (untreated mouse).

LC-MS analysis of vinyl-U modification: RNA samples were digested to nucleosides using Nucleoside Digestion Mix (New England Biolabs). LC-MS/MS analysis was performed by injecting digested RNAs on an Agilent 1290 Infinity II UHPLC equipped with a G7117A diode array detector and a 6495C triple quadrupole mass detector operating in the positive electrospray ionization mode (+ESI). UHPLC was carried out on a Waters XSelect HSS T3 XP column (2.1 × 100 mm, 2.5 µm) with a gradient mobile phase consisting of methanol and 10 mM aqueous ammonium acetate (pH 4.5). MS data acquisition was performed in the dynamic multiple reaction monitoring (DMRM) mode. Each nucleoside was identified in the extracted chromatogram associated with its specific MS/MS transition: rC $[M+H]^+$ at m/z 244.1→112.1, rU $[M+H]^+$ at m/z 245.1→113, rG $[M+H]^+$ at m/z 284.1→152.1, rA $[M+H]^+$ at m/z 268.1→136.1, Y $[M+H]^+$ at m/z 245.1→209, and 5-VU $[M+H]^+$ at m/z 271.1→139.1. External calibration curves with known amounts of the nucleosides were used to calculate their ratios within the samples analyzed.

Tumor and lung cryopreservation for tissue sectioning[16]: After mice were sacrificed, and the tumors and organs were collected and placed in a solution of 4%

paraformaldehyde. The tumors and organs were fixed for 1 h at 4 °C with gentle shaking. After 1 hour, the PFA was removed and the samples were washed in PBS buffer for 10 minutes at 4 °C with gentle shaking. The wash was repeated twice. After the third wash, the PBS was removed and the samples were place in a 30% sucrose solution and left at 4 °C with gentle shaking overnight. The samples were then removed from the sucrose solution and placed in Tissue Tek-OCT (VWR 25608-930) and frozen. The frozen blocks were sectioned on a ThermoScientific Cryostar NX50 Cryostat. Tumor and lung sections were cut at 10 μM thickness and placed onto VWR Superfrost Plus Micro slides (VWR 48311-703). Slides were placed at - 80 °C until further needed.

Fluorescent imaging of vinyl-RNA in mouse tissues: Cryo-preserved tumor and lung slices were washed in 2x PBS gently to remove all of the visible OCT off samples. A PAP pen (Immedge, Vector Labs) was used to create a hydrophobic barrier around the samples. 1 mg/ml BSA was used to block the tissue at room temperature without shaking, ensuring complete contact with tissue. Slides were rinsed 2x in PBS. 2 mM tetrazine-Cy5 (Lumiprobe) was dissolved in DMSO and kept in amber tubes and 1:1000 was dissolved in nuclease-free water, added to the slide, and incubated for 3h at 37°C in the dark. Slides were washed with 0.1% Triton-X100 (Sigma-Aldrich) in PBS 5x to ensure excess dye is removed and then washed 3x in PBS to remove detergent. PBS was removed and 1:1000 Hoechst (Cat#H3570, Invitrogen/LifeTech) was added to the tissue for a 10-minute incubation to stain nuclei, then washed 3x in PBS. Excess solution was removed from the slide and Vectasheild antifade mounting medium used to add a coverslip to the slide. A

Leica Stellaris was used to capture nuclear signal, GFP, mCherry and Cy5 signals at 40x magnification with a water-immersion lens.

Supplementary Figures

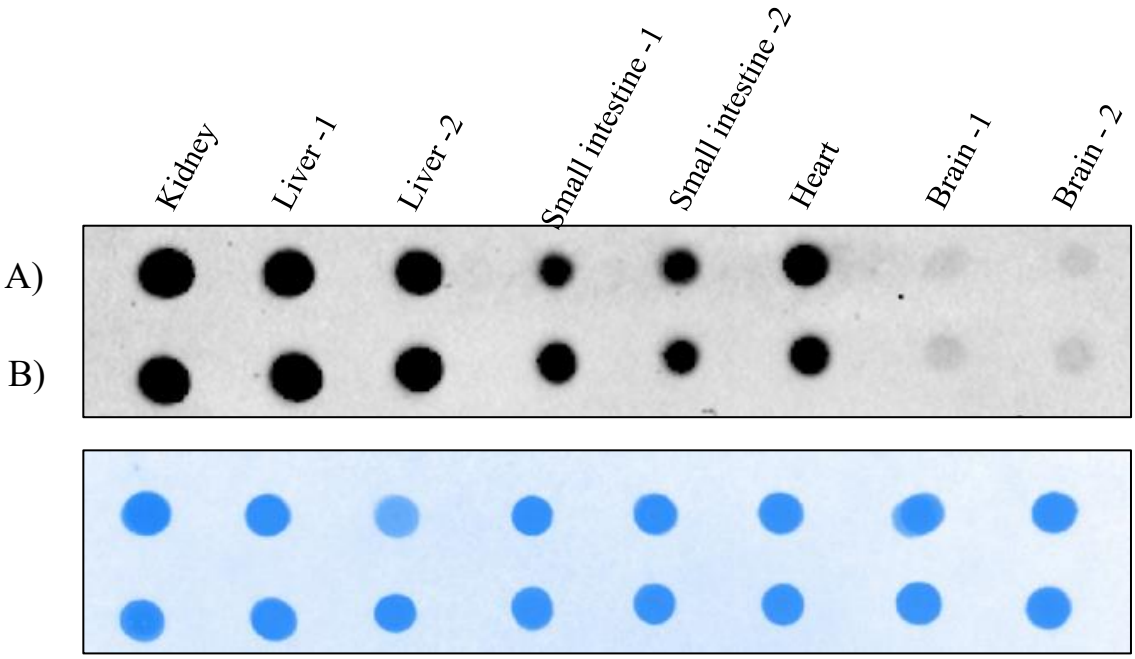


Figure 3-S1 IP-injection into WT mice. Wild-type (WT) C57Bl mice were treated intraperitoneally with 500 mM IP injections (300 mg/kg) of 5-vinyluridine over 5 hours prior to isolation of organs. RNA was reacted with tetrazine-biotin for dot blot analysis. A) and B) indicate biological replicates. 1 and 2 indicate technical replicates from organs.

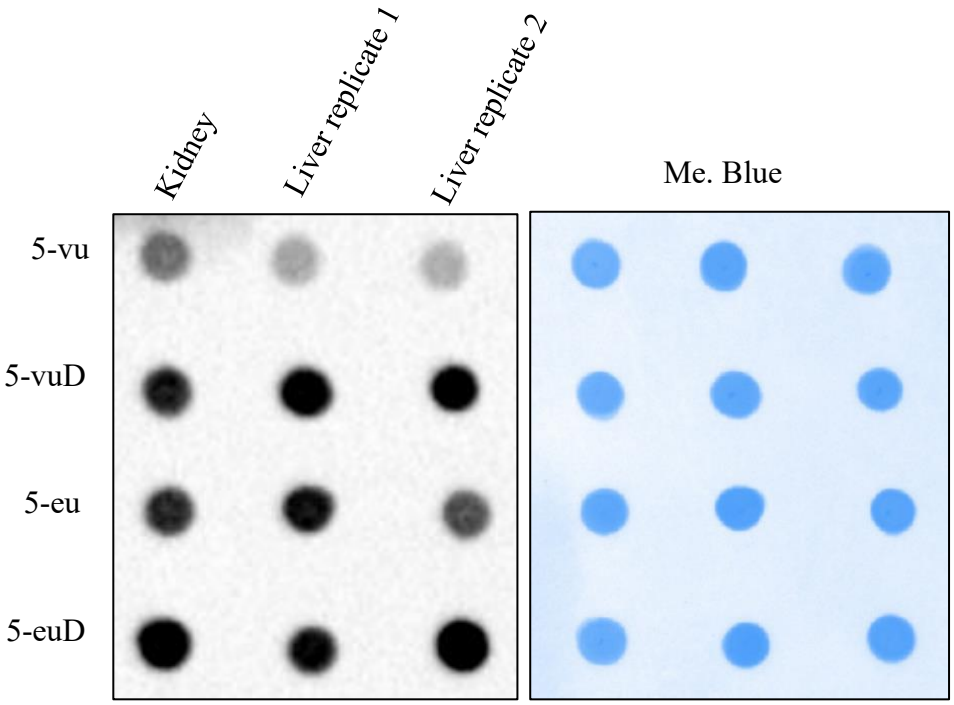


Figure 3-S2 IP-injection with different uracil and uridine analogs into WT mice. WT C57Bl mice were treated with 500 mM per injection over 24 hours. Organ RNA was isolated and reacted with biotin-tetrazine for dot blot analysis. 1 and 2 indicate technical replicates from a liver.

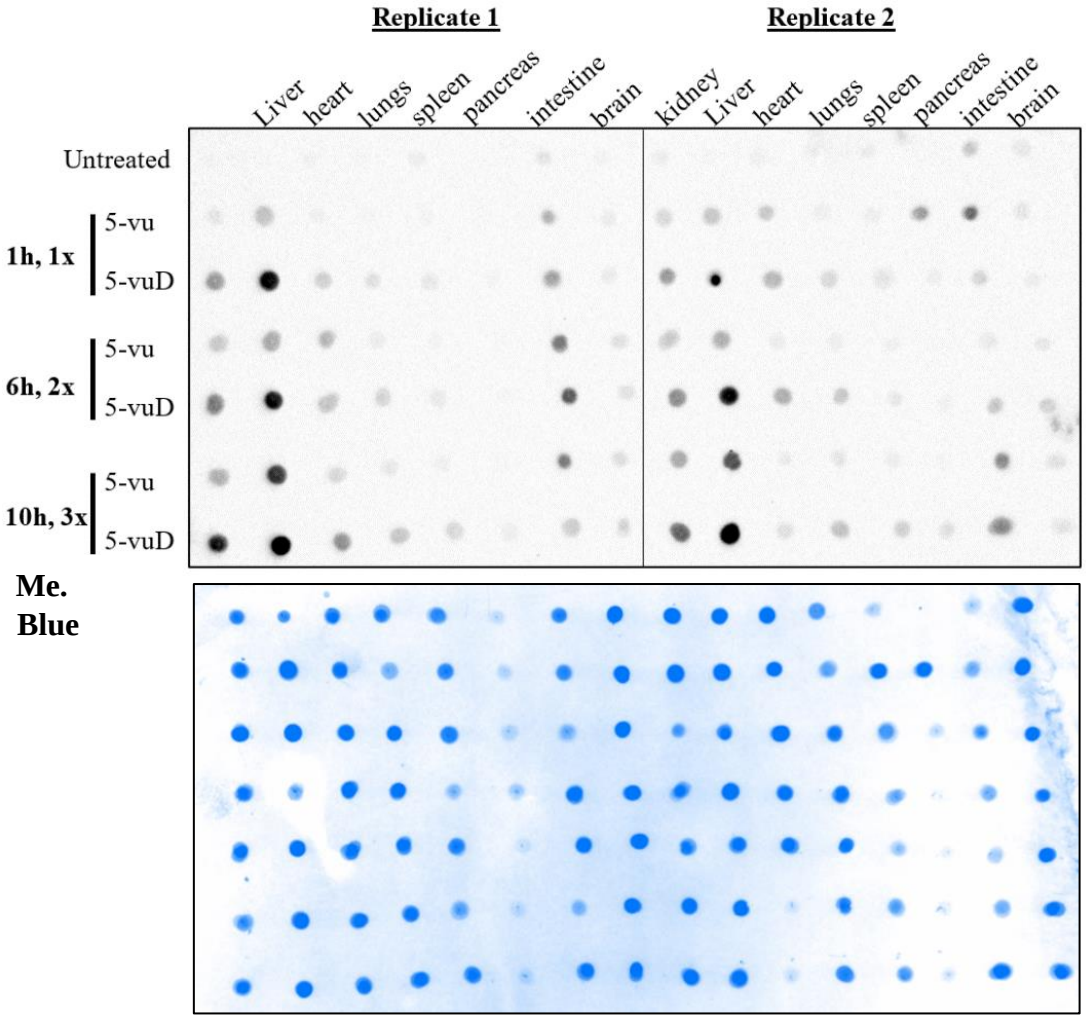


Figure 3-S3 Comparison of 5-vu background labeling with increased time and injections. Two biological replicates of mice were treated with 500 mM of 5-v or 5-vuD with increasing injections over time at 1h, 6h, and 10h. Doubly and triply injected mice were treated at 3-hour time increments between injections. Organs were isolated at the indicated time points. RNA was extracted and reacted with tetrazine-biotin for dot blot analysis. Untreated mice were used as negative controls.

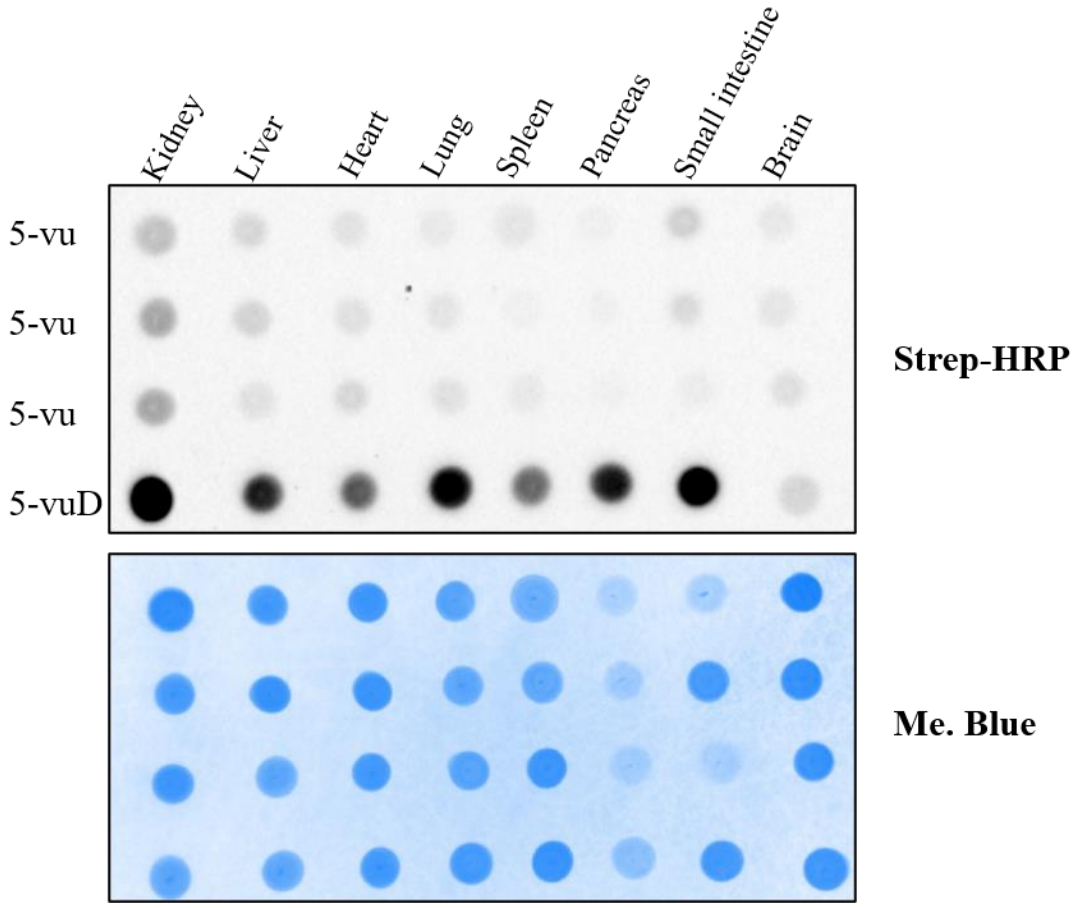


Figure 3-S4 Organs from (Figure 3-2) xenografts to background RNA labeling. Dot blot analysis shows the signal of RNA labeling from tissues that do not express 3xUPRT. No or low levels of 3xUPRT cells may be expressed in the lung below *ex vivo* fluorescence detection limit.

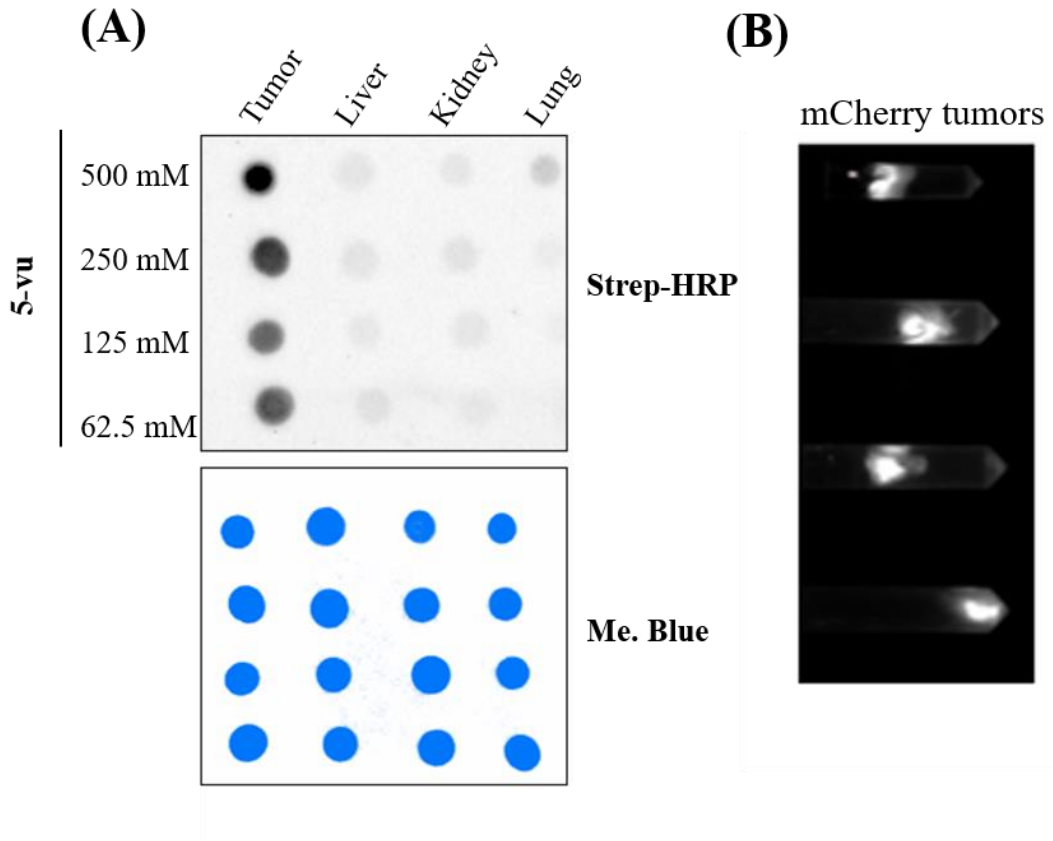


Figure 3-S5 IP-injection of standard dose and 2-fold reduced dose in (+)-3xUPRT xenograft mice. A) Dot blot analysis of 5-vu treatment titration in mice with 3x-UPRT expressing tumors. Liver and kidney tissue do not have 3xUPRT expression. No or low levels of 3xUPRT tumors were detected in matching lung samples. B) *Ex vivo* mCherry fluorescence depicts co-expressed 3xUPRT in tumors.

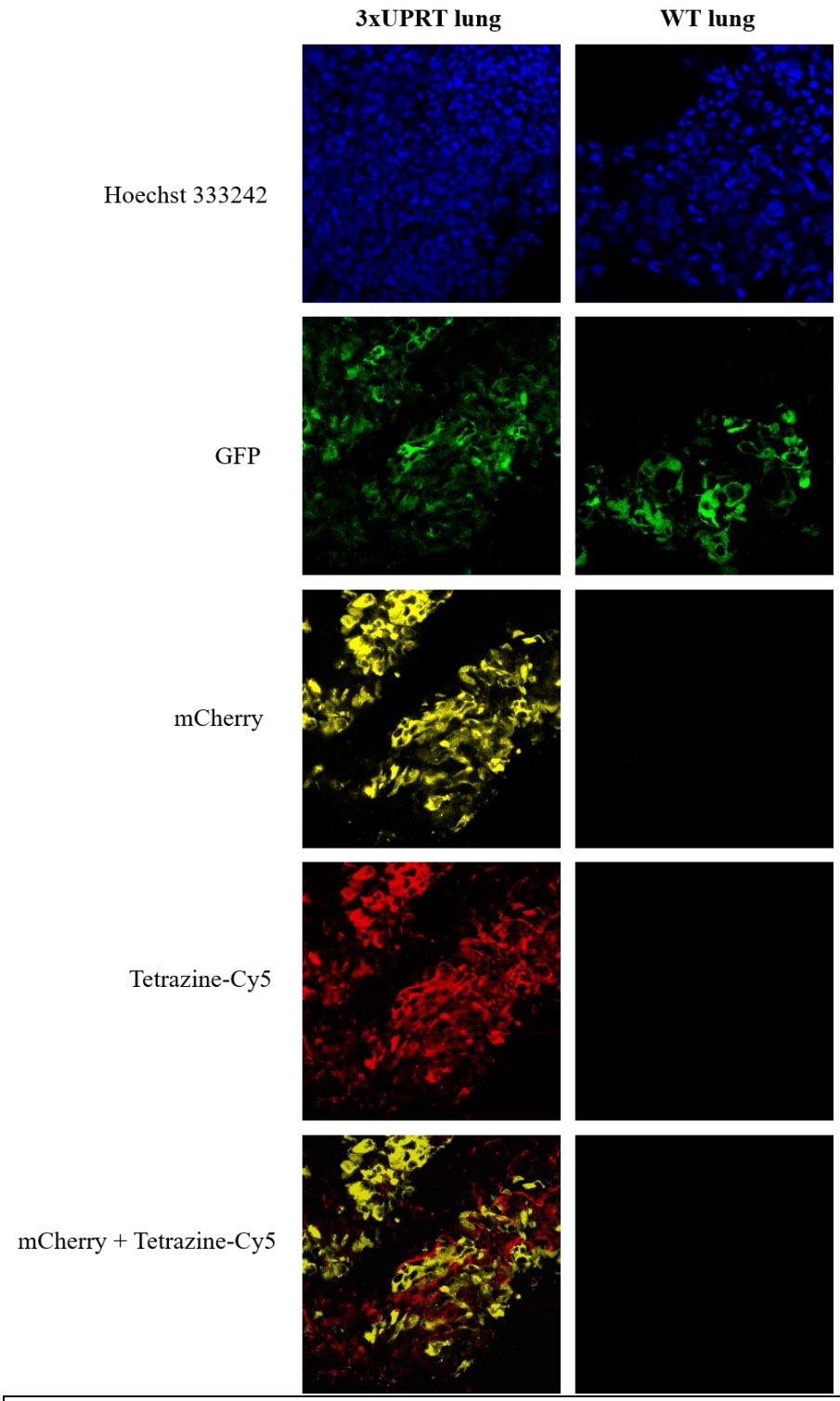


Figure 3-S7 Metabolic labeling in (+)-3xUPRT-mCherry MDA-MB-231 LM2 metastatic cells in lung tissues. Vinyl-containing RNA are fluorophore-conjugated with tetrazine-Cy5. GFP signal is present in WT MDA-MB-231 LM2 cells (parental

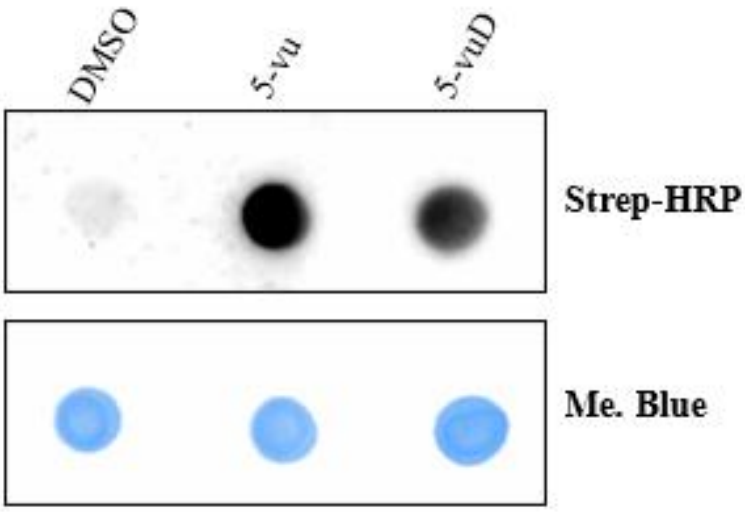


Figure 3-S8 RNA dot blot with vimentin-3xUPRT stable MDA-MB-231 cells. Cells were treated for 3h with 1 mM of DMSO, 5-vu or 5-vuD, biotinylated with tetrazine-biotin and metabolically labeled RNA were detected with streptavidin-HRP chemiluminescent signal.

References

1 Wei, J.-W.; Huang, K.; Yang, C.; Kang, C.-S. Non-Coding RNAs as Regulators in Epigenetics. *Oncology Reports* **2017**, *37* (1), 3–9.

2 Richardson, G. M.; Lannigan, J.; Macara, I. G. Does FACS Perturb Gene Expression? *Cytometry* **2015**, *87* (2), 166–175.

3 Pfister, G.; Toor, S. M.; Sasidharan Nair, V.; Elkord, E. An Evaluation of Sorter Induced Cell Stress (SICS) on Peripheral Blood Mononuclear Cells (PBMCs) after Different

Sort Conditions - Are Your Sorted Cells Getting SICS? *Journal of Immunological Methods* **2020**, *487*, 112902.

4 Singha, M.; Spitalny, L.; Nguyen, K.; Vandewalle, A.; Spitale, R. C. Chemical Methods for Measuring RNA Expression with Metabolic Labeling. *WIREs RNA* **2021**, *12* (5).

5 Nguyen, K.; Fazio, M.; Kubota, M.; Nainar, S.; Feng, C.; Li, X.; Atwood, S. X.; Bredy, T. W.; Spitale, R. C. Cell-Selective Bioorthogonal Metabolic Labeling of RNA. *J. Am. Chem. Soc.* **2017**, *139* (6), 2148–2151.

6 Zajaczkowski, E. L.; Zhao, Q.-Y.; Zhang, Z. H.; Li, X.; Wei, W.; Marshall, P. R.; Leighton, L. J.; Nainar, S.; Feng, C.; Spitale, R. C.; Bredy, T. W. Bioorthogonal Metabolic Labeling of Nascent RNA in Neurons Improves the Sensitivity of Transcriptome-Wide Profiling. *ACS Chem. Neurosci.* **2018**, *9* (7), 1858–1865.

7 Paredes, E.; Das, S. R. Click Chemistry for Rapid Labeling and Ligation of RNA. *ChemBioChem* **2011**, *12* (1), 125–131.

8 Nainar, S.; Cuthbert, B. J.; Lim, N. M.; England, W. E.; Ke, K.; Sophal, K.; Quechol, R.; Mobley, D. L.; Goulding, C. W.; Spitale, R. C. An Optimized Chemical-Genetic Method for Cell-Specific Metabolic Labeling of RNA. *Nat Methods* **2020**, *17* (3), 311–318.

9 Nguyen, K.; Kubota, M.; Arco, J. del; Feng, C.; Singha, M.; Beasley, S.; Sakr, J.; Gandhi, S. P.; Blurton-Jones, M.; Fernández Lucas, J.; Spitale, R. C. A Bump-Hole Strategy for Increased Stringency of Cell-Specific Metabolic Labeling of RNA. *ACS Chem. Biol.* **2020**, *15* (12), 3099–3105.

10 Kubota, M.; Nainar, S.; Parker, S. M.; England, W.; Furche, F.; Spitale, R. C. Expanding the Scope of RNA Metabolic Labeling with Vinyl Nucleosides and Inverse Electron-Demand Diels–Alder Chemistry. *ACS Chem. Biol.* **2019**, *14* (8), 1698–1707.

11 Suzuki, K.; Tsunekawa, Y.; Hernandez-Benitez, R.; Wu, J.; Zhu, J.; Kim, E. J.; Hatanaka, F.; Yamamoto, M.; Araoka, T.; Li, Z.; Kurita, M.; Hishida, T.; Li, M.; Aizawa, E.; Guo, S.; Chen, S.; Goebl, A.; Soligalla, R. D.; Qu, J.; Jiang, T.; Fu, X.; Jafari, M.; Esteban, C. R.; Berggren, W. T.; Lajara, J.; Nuñez-Delicado, E.; Guillen, P.; Campistol, J. M.; Matsuzaki, F.; Liu, G.-H.; Magistretti, P.; Zhang, K.; Callaway, E. M.; Zhang, K.; Belmonte, J. C. I. In Vivo Genome Editing via CRISPR/Cas9 Mediated Homology-Independent Targeted Integration. *Nature* **2016**, *540* (7631), 144–149.

12 Minn, A. J.; Gupta, G. P.; Siegel, P. M.; Bos, P. D.; Shu, W.; Giri, D. D.; Viale, A.; Olshen, A. B.; Gerald, W. L.; Massagué, J. Genes That Mediate Breast Cancer Metastasis to Lung. *Nature* **2005**, *436* (7050), 518–524.

13 Azevedo-Pouly, A. C. P.; Elgamal, O. A.; Schmittgen, T. D. RNA Isolation from Mouse Pancreas: A Ribonuclease-Rich Tissue. *JoVE* **2014**, No. 90, 51779.

14 Satelli, A.; Li, S. Vimentin in Cancer and Its Potential as a Molecular Target for Cancer Therapy. *Cell. Mol. Life Sci.* **2011**, *68* (18), 3033–3046.

15 Neef, A. B., Pernot, L., Schreier, V. N., Scapozza, L., and Luedtke, N. W. (2015) A Bioorthogonal Chemical Reporter of Viral Infection. *Angew. Chem. 127* (27), 8022– 8025.

16 Beasley, S.; Vandewalle, A.; Singha, M.; Nguyen, K.; England, W.; Tarapore, E.; Dai, N.; Corrêa, I. R.; Atwood, S. X.; Spitale, R. C. Exploiting Endogenous Enzymes for Cancer-Cell Selective Metabolic Labeling of RNA in Vivo. *J. Am. Chem. Soc.* **2022**, *144* (16), 7085–7088.

17 Zhang, G.-L.; Zhang, Y.; Cao, K.-X.; Wang, X.-M. Orthotopic Injection of Breast Cancer Cells into the Mice Mammary Fat Pad. *JoVE* **2019**, No. 143, 58604.

Chapter 4: Mutually Orthogonal Bioconjugation of Vinyl Nucleosides for RNA Metabolic Labeling

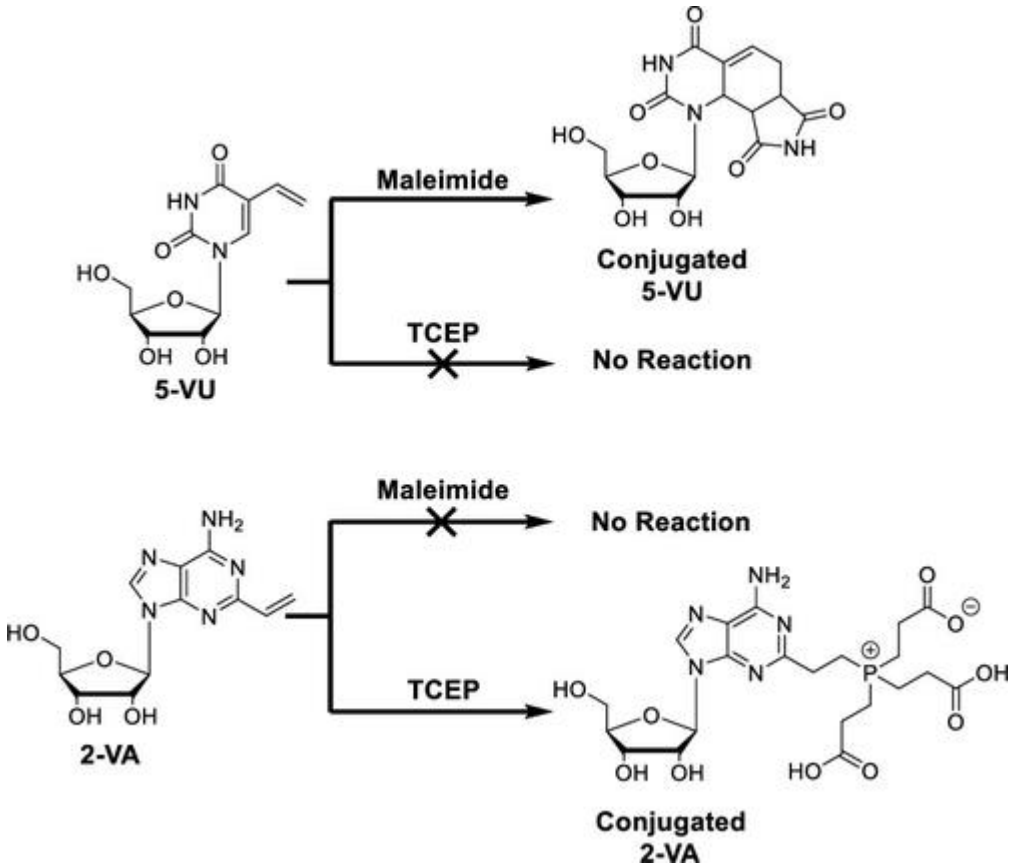
Introduction

Gene expression within biological systems produces active intermediates such as RNA and proteins that drive cellular processes. Using RNA-sequencing, genome-wide RNA expression analysis has shown that RNA transcripts are capable of regulating genes through a variety of interactions, modulating key mechanisms that control cell metabolism, identity, and signaling[1]. Monitoring RNA expression as a function of time provides information about the relative synthesis and decay rates of specific transcripts[2]. RNA metabolic labeling methods accomplish this by exposing cells to chemically modified, unnatural ribonucleoside analogues that are incorporated into elongating RNA strands[3]. Applying bioconjugation chemistries to the labeled RNA allows for the affinity-based separation of labeled transcripts from unlabeled, steady-state ones. It also permits localization studies of labeled RNA in cells with microscopy[4]. Temporal analysis of labeled transcripts illuminates RNA regulatory pathways providing novel insights into cellular mechanisms.

Recently, the use of multiple nucleoside analogues for RNA metabolic labeling has been demonstrated by Akimitsu et al[5]. They performed an analysis of both RNA transcription and degradation via the simultaneous use of the unnatural nucleosides 5-bromo-uridine (Br-U) and 4-thio-uridine (4-SU). Although this approach is promising for deconvoluting complex RNA pathways, their method is hindered by two factors: (i) 4-SU is reported to be toxic to cells[7] and (ii) the isolation of Br-U-labeled RNA relies on the

antibody specificity for Br-U. Chemical methods for selective modification of multiple nucleosides in RNA metabolic labeling can overcome these challenges. This has the potential to expand RNA metabolic labeling applications toward profiling differentially expressed RNA pools in cells. For instance, this strategy allows for probing the changes in RNA expression in response to experimental stimuli, enhancing our understanding of biological mechanisms that occur in tandem.

Here, we report a strategy for the chemo-selective modification of nontoxic vinyl modified RNA nucleosides that causes minimal perturbation to living cells. Previously, our lab and others have demonstrated the use of vinyl nucleosides metabolic labeling, and have



Scheme 4-1 5-VU Is Modified Selectively with Maleimide, and 2-VA Is Selectively Modified by TCEP.

observed they exhibit lower toxicity compared to currently employed unnatural nucleosides for both RNA and DNA labeling[6,8].

Results

We chose 5-vinyl uridine (5-VU) and 2-vinyl adenosine (2-VA) for developing an orthogonal labeling strategy, and we predicted that the terminal alkene in 5-VU could react as a diene in a formal [4 + 2] cycloaddition with a dienophile, because it has limited resonance with the adjacent alkene in uracil **(Scheme 4-1)**. Similarly, we reasoned that the terminal alkene in 2-VA, which is part of a conjugated imine, is more delocalized and therefore could act as a Michael acceptor by stabilizing the negative charge of the phosphorylated intermediate in a phospha-Michael (PM) reaction.

To support the hypothesis of differential reactivity *in silico,* we first evaluated the proposed reactions using density functional theory (DFT) calculations. We started by analyzing the selectivity of 5-VU over 2-VA with maleimide in a Diels–Alder (DA) reaction. Maleimide was chosen as the dienophile because it has been widely used for the conjugation of free cysteines[9] and recently was used as a dienophile for modifying diene modified nucleosides[10]. When the energies and shapes of the participating frontier molecular orbitals (MOs) of the dienes were analyzed, we found that the HOMO of 5-VU matches the phase of the lowest unoccupied MO (LUMO) of maleimide and that the associated energy gap is 2.11 eV **(Figure 4-1).** By contrast, for 2-VA, the phase of the HOMO does not match the LUMO of maleimide indicating poor overlap. Instead, the fourth highest occupied MO (HOMO-3) is the MO with the correct phase for favorable overlap with the LUMO of maleimide. However, the associated energy gap between the LUMO and

HOMO-3 is 2.86 eV, significantly larger than that for 5-VU. Thus, this predicts that the driving force for the reaction of 2-VA with maleimide will be significantly reduced relative

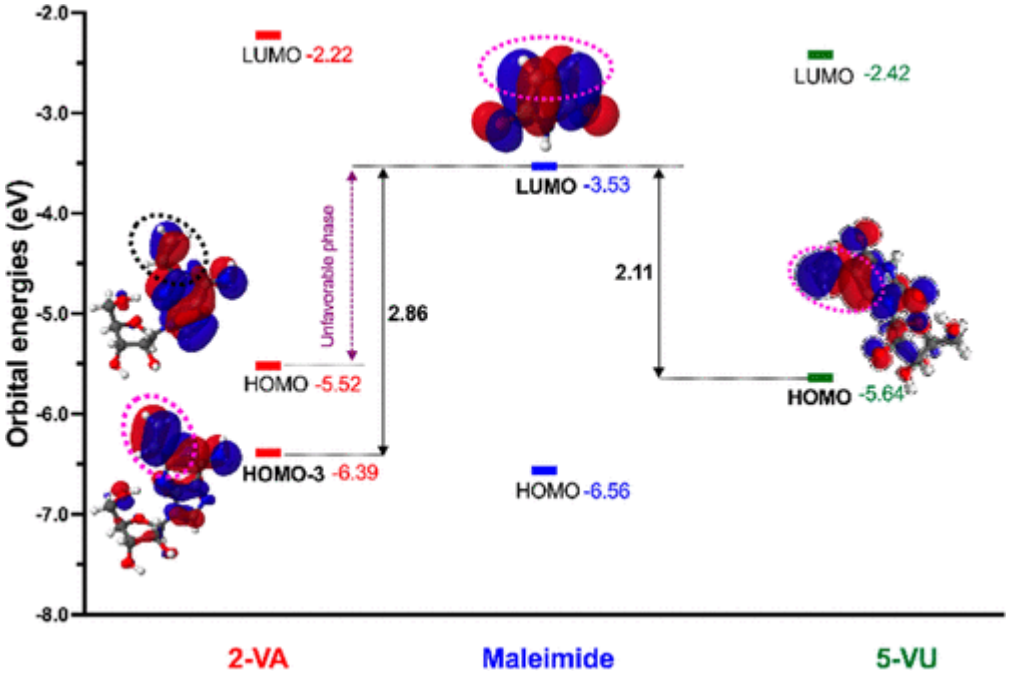


Figure 4-1 Energy level diagram to illustrate the feasibility of DA reaction between vinyl modified nucleosides and maleimide along with isosurface plots of respective MOs. The isovalue used was 0.02 ($Å^{-3}$).

to the reaction of 5-VU with maleimide.

Next, we evaluated the PM addition of the soft nucleophile tris(2-carboxyethyl)phosphine (TCEP) to the vinyl nucleosides. TCEP was chosen because it is a reactive phosphine reagent known to reduce disulfide bonds in proteins[11] and was recently demonstrated to detect crotonyl groups on proteins via PM reactions[12]. We began by manually searching for the intermediate containing the C–P bond formation, where we noticed that, regardless of the initial structure, the C–P bond in 5-VU was always broken by the geometry optimization, indicating that the C–P bond in 5-VU is unstable. On the other

hand, a stable intermediate of 2-VA complexed with C–P was obtained readily **(Figure 4-S1)**. To clarify the instability of C–P bond formation, we systematically scanned the potential energy curve as a function of C–P distance for the 2-VA/TCEP complex in comparison to the 5-VU/TCEP complex **(Figure 4-2a)**. Strikingly, the 5-VU/TCEP complex shows no minimum with respect to the C–P distance, whereas the 2-VA/TCEP complex has a shallow minimum at a C–P distance of 1.88 Å.

To test the hypothesis that extended conjugation in 2-VA could stabilize the negative charge of the phosphorylated intermediate more effectively compared to 5-VU, we consider the resonance stabilization by computing partial atomic charges using natural bonding orbital (NBO) analysis[13]. For the following analysis, we use the 5-VU/TCEP constrained to a C–P bond distance of 1.86 Å (i.e., lowest energy bond distance in the scan for 2-VA/TCEP) as an approximate intermediate structure, but we emphasize that this is not a local minimum. We focused on the charge accumulated during C–P bond formation, Δq **(Figure 4-2a)**. The largest changes in charge occur for the vinyl carbons and phosphorus and are similar for both 2-VA and 5-VU, with the sum of Δq for all three atoms being significantly positive in both cases (0.19 in 2-VA and 0.18 in 5-VU), indicating a net loss of electron density during C–P bond formation. We find that this electron density transfers to the rings and for 2-VA/TCEP the excess electron density is evenly distributed across the whole adenine ring, with a root-mean-square Δq (RMSΔq) of 0.02 and maximum $|\Delta q|$ = 0.08. By contrast, in 5-VU, almost the entirety of the excess electron density accumulates on the C6 of uracil (Δq = −0.22) and the RMSΔq is 0.05, twice as large as that for 2-VA. Thus, we conclude that the greater resonance stabilization of excess electron density in 2-VA leads to a stable intermediate, whereas the concentration of electron density on the C6 position in

5-VU prevents C–P bond formation. We computationally tested our conclusion by repeating the C–P bond scan using a C6 nitro modified uracil (i.e., a strong π-acceptor) and we found

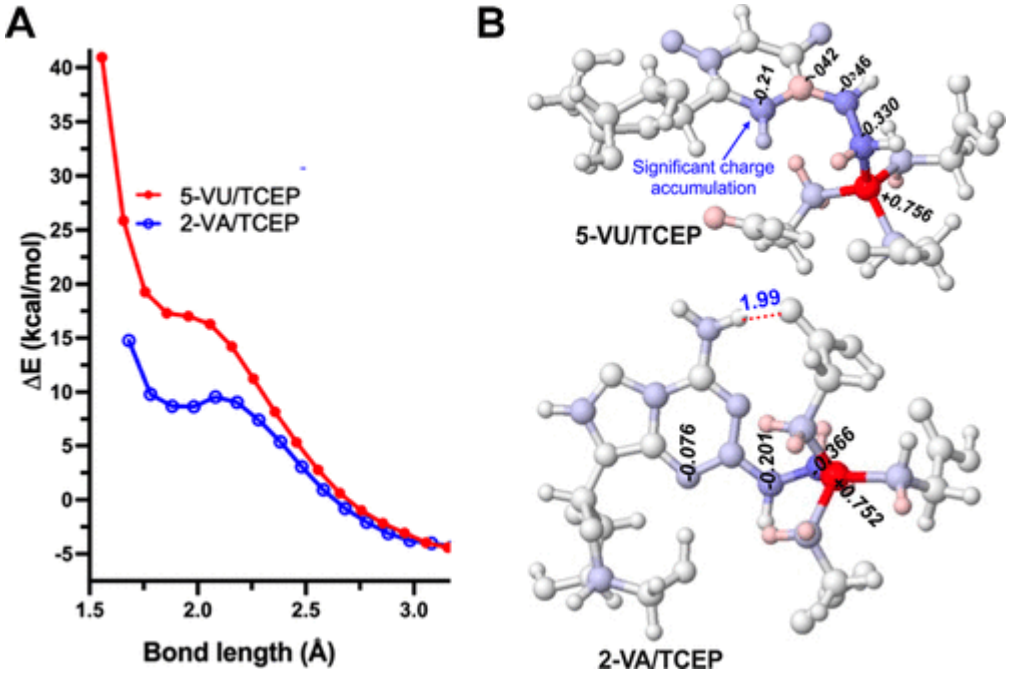


Figure 4-2 (A) Relaxed potential energy surface of nucleoside/TCEP as a function of C–P bond length. Energies (kcal/mol) are relative to the energies of separated reactants. (B) Nucleoside/TCEP complex structures showing differences between in partial atomic charges of separated reactants and complexes, Δq, for 5-VU/TCEP and 2-VA/TCEP. Atoms colored red indicate $\Delta q > 0$, while blue indicates $\Delta q < 0$. Atomic charges (black) are written in atomic units, and hydrogen bonding distance (blue) in ångstroms.

a strongly bound stable intermediate **(Figure 4-S2)**. Finally, we also investigated the possible role of H-bonding in the PM addition, but observed no effect on the selectivity of the reaction.

With the theoretical validation of our hypothesis, we began our investigation into the reactivity of vinyl nucleosides by NMR spectroscopy. First, we observed the reactions between maleimide and vinyl nucleosides, 5-VU and 2-VA, via ^{1}H NMR spectroscopy. Upon incubating maleimide with 5-VU in D_2O, we observed the formation of a new compound. The signal from vinyl protons on 5-VU's terminal alkene disappeared, and the 1′ H underwent a significant upfield shift with near quantitative conversion of 5-VU observed in 3 h **(Figure 4-3a)**. This new compound was then isolated by separate reaction in H_2O and identified as **1**, the desired product of the proposed formal [4 + 2] cycloaddition between 5-VU and maleimide. In contrast, upon incubation of excess maleimide with 2-VA in D_2O, we did not observe any product formation within 48 h and the 2-VA remained unreacted in solution **(Figure 4-S6)**. Next, we investigated the reactivity of TCEP with 2-VA and 5-VU. Here, we incubated the vinyl nucleosides with excess TCEP in D_2O and again monitored the reactions by ^{1}H NMR. Upon incubation of 2-VA with TCEP, we observed the complete consumption of the vinyl nucleoside within 20 min, based on the loss of the signal from the terminal alkene's protons **(Figure 4-3b)**. The reaction was also observed with ^{31}P NMR, where an upfield shift was observed for the newly formed phosphonium ion **(Figure 4-S8)**. The reaction of TCEP with 2-VA was repeated in H_2O and product **2** was isolated, which was confirmed to be product of proposed PM reaction upon characterization. In contrast,

after incubation of 5-VU with TCEP, no reaction was observed within 48 h and 5-VU remained unreacted in solution **(Figure 4-S9)**.

The results of these ^{1}H NMR studies demonstrate the desired orthogonality of vinyl nucleoside labeling, where maleimide reacts preferentially with 5-VU and TCEP reacts preferentially with 2-VA. To further evaluate the specificity of these reactions, we incubated natural ribonucleosides with maleimide and TCEP under the same reaction conditions and monitored the reactions by HPLC. We observed no reactivity between either reagents and any of the natural nucleosides **(Figures 4-S10, 4-S11)**, thus demonstrating specific reactivity of these labeling reagents with vinyl nucleosides. Finally, we evaluated the aqueous stability of the vinyl nucleosides. Upon incubation of 5-VU and 2-VA in D_2O at room temperature, we did not observe any signs of chemical degradation within 48 h **(Figures 4-S12, 4-S13)**,

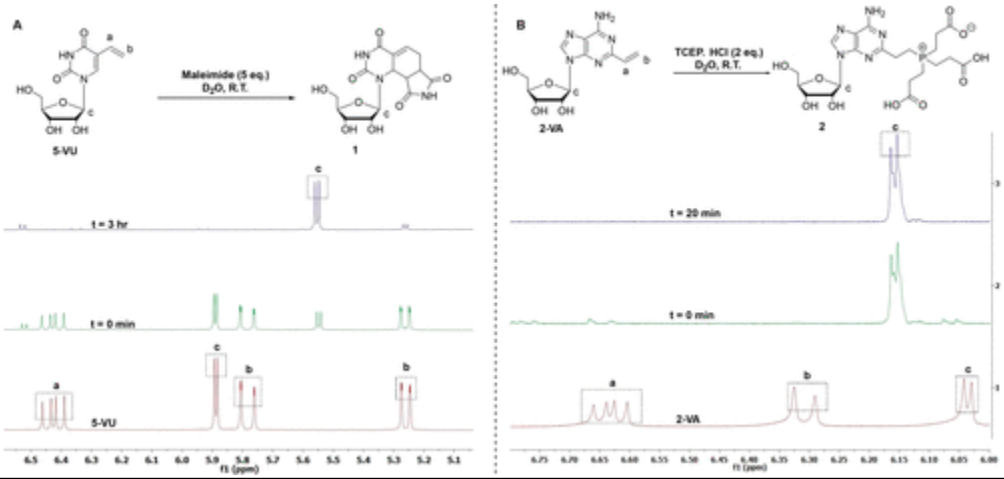
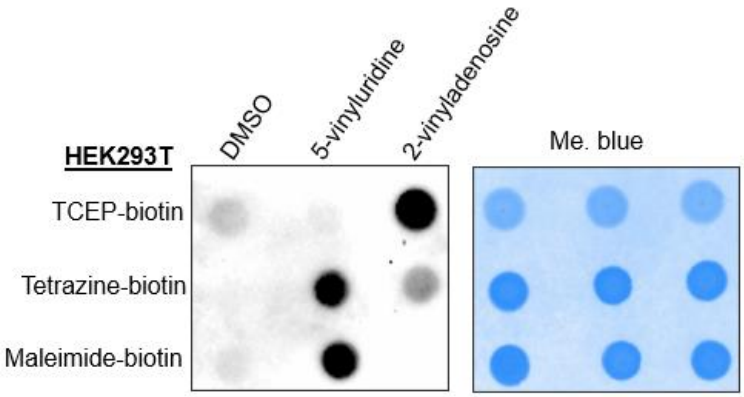


Figure 4-3 NMR spectra demonstrating reactivity of vinyl nucleosides in DA and PM reactions. (A) Reaction of 5-VU with maleimide. (B) Reaction of 2-VA with TCEP.

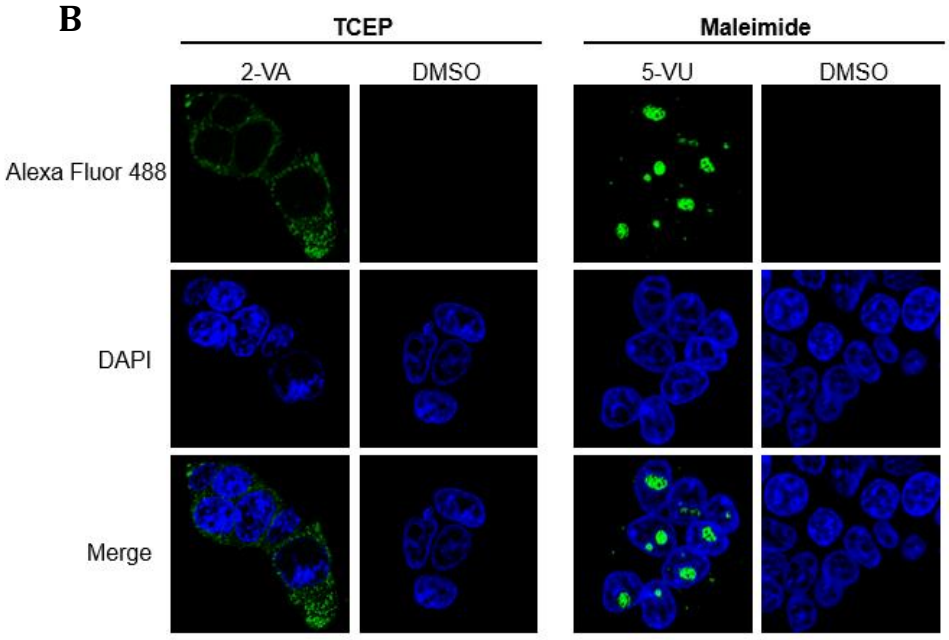
To evaluate the chemo selectivity of these reagents in a biological context, we tested the system by reacting RNA from HEK293T cells treated with 1 mM 5-VU or 2-VA for 5 h.

The total cellular RNA isolated from these cells was incubated with biotin-conjugated maleimide or biotin-conjugated TCEP for testing these chemical reactions via dot blot analysis **(Figure 4-4a)**. Tetrazine-biotin performed as a positive control since it has been shown to react with both 5-VU and 2-VA labeled RNA[14]. Consistent with previous observations, tetrazine-biotin reacted robustly with 5-VU labeled RNA and to a lesser extent with 2-VA labeled RNA. As observed with single nucleosides, maleimide-biotin reacted exclusively with total RNA extracted from 5-VU treated cells and TCEP-biotin also reacted specifically with total RNA from 2-VA treated cells. Additional time and concentration studies showed comparable reactivities between maleimide-biotin and tetrazine-biotin with 5-VU labeled RNA. Conversely, robust reactivity was observed with 2-VA labeled RNA and TCEP-biotin within shorter reactions times **(Figure 4-S14)**. Lastly, to demonstrate the utility of these chemistries for *in vitro* modification of vinyl RNA, we applied each bioconjugation reagent to formaldehyde-fixed HEK293T cells pretreated with 5-VU or 2-VA at 1 mM for 5 h. Using our time and concentration analysis to inform reaction conditions within fixed cells, bioconjugation of metabolically labeled transcripts was detected **(Figure 4-4b)**. 5-VU in cells predominantly showed a signal in the nucleolus, whereas cells with fluorophore-conjugated 2-VA showed a higher signal in the cytoplasm, which contains mature poly-A tailed mRNA consistent with earlier findings[14].

A


HEK293T
DMSO
5-vinyluridine
2-vinyladenosine
Me. blue
TCEP-biotin
Tetrazine-biotin
Maleimide-biotin

B


TCEP
Maleimide
2-VA
DMSO
5-VU
DMSO
Alexa Fluor 488
DAPI
Merge

C

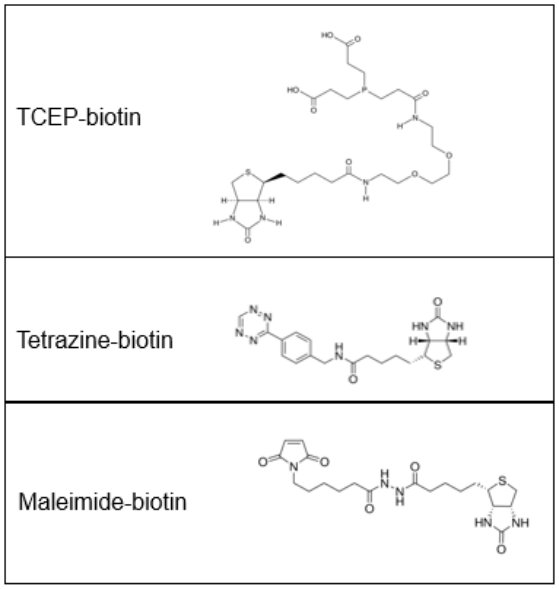


Figure 4-4 Chemical orthogonality applied for RNA metabolic labeling. (A) RNA extracted from HEK293T cells treated with 1 mM 5-VU and 2-VA for 5 h was reacted with 1 mM TCEP-biotin, tetrazine-biotin or maleimide-biotin. Signal from labeled RNA was detected with streptavidin-HRP chemiluminescence. Methylene blue = loading control. (B) Confocal microscopy of formaldehyde-fixed HEK293T cells treated with 1 mM 5-VU or 2-VA for 5 h and conjugated with AlexaFluor 488. DMSO = negative control. (C) Chemical structures of biotinylation reagents used for dot blot assay.

Conclusions

In summary, we demonstrate an orthogonal labeling strategy for the modification of vinyl nucleosides that expands the capabilities for RNA metabolic labeling techniques. The orthogonal reactions were first studied by DFT and later confirmed with single nucleoside reactions. Further, these chemistries were successfully applied for modification of vinyl

RNA isolated from metabolic labeling experiments and utilized for the selective visualization of metabolically labeled RNA in 5-VU or 2-VA treated fixed cells with confocal microscopy. We envision that these methods will be useful for analyzing RNA expression dynamics within cells subjected to experimental stimuli in order to investigate complex biological pathways. These approaches enable multiplexed RNA metabolic labeling at desired time points that can be subsequently distinguished through applying these orthogonal chemistries.

Experimental Methods

DFT Studies: Relaxed structures for vinyl-nucleosides (2-VA and 5-VU), maleimide and TCEP were obtained through a combination of conformational search and density functional theory (DFT) optimization. First, we generated conformers for each compound using the CREST with the GFN2-xTB semiempirical model1,2. Next, the ten lowest energy conformers according to GFN2-xTB were selected for further refinement with DFT. All density functional calculations were performed using Turbomole 7.53 . DFT optimizations were performed using the TPSS4,5 density functional including D36,7 dispersion corrections and def2-TZVP8 basis set (TPSS-D3/def2-TZVP). The resolution-of-the-identity (RI)9 approximation was used to accelerate Coulomb two-electron integrals, and a tight integration grid m5 and convergence criteria (scfconv 8, denconv 1d-8) were used. The lowest energy conformer after TPSS optimization was selected as the equilibrium structure for each compound. All DFT calculations were solvated using the conductor-like screening model (COSMO)10 ($\varepsilon = 80.1$). The propensity to form C-P bonds between the vinyl nucleosides and TCEP was evaluated by generating relaxed potential energy surfaces as a

function of the C-P bond length. Constrained optimizations were performed by fixing the distance between the beta [4] carbon of the vinyl nucleoside and the phosphorus in TCEP to values from 3.6 Å to 1.8 Å (Figure S1).

Hydrogen bonding DFT studies: We investigate the effect of hydrogen bonding in the stability of the intermediate complexes. We see evidence that the 2-VA/TCEP complex has intermolecular hydrogen bonding between an adenine C6NH to TCEP carbonyl with a distance of 1.99 Å (Fig. 2). By contrast, in the 5-VU/TCEP complex, we find no hydrogen bond within 4.0Å. That said, we expect our model to exaggerate the role of hydrogen bonding because the implicit solvation neglects hydrogen bonding with the solvent which will net destabilize the products (with artificially suppressed hydrogen bonding) relative to the complex structure. Therefore, we generated a comparison with a structure where TCEP is orientated in such a way that hydrogen bond formation is suppressed (Fig. S3). Once again, we found a stable 2-VA/TCEP intermediate but no stable 5-VU/TCEP intermediate. Finally, this behavior was verified with a different density functional (Fig. S4). Thus, we conclude that intermolecular hydrogen bonding contributes to the relative stability of the 2-VA/TCEP complex but does not explain the selectivity of the phospha-Michael reaction.

NMR studies for reactivity. Reactivity of vinyl nucleosides was observed by dissolving the nucleotides in D2O to achieve 100mM concentration. 5 equivalents of maleimide/TCEP.HCl was added to the solution and mixed thoroughly. 400µL of this mixture was transferred to a fresh NMR tube [5] and NMR spectroscopy was conducted at regular intervals, incubating the NMR tubes at room temperature between the intervals.

Background reactivity studies: 100 µL of 10 mM Adenosine, Uridine, Guanosine and Cytidine solutions in H2O each were added into an HPLC vial. The solution was diluted with 500 µL of H2O and 100 µL of 50 mM maleimide or TCEP solution were added. The solution was then analyzed just after addition and then after incubation at room temperature for 3 hours by analytical RP-HPLC (Thermo-Scientific Hypersil BDS C18 column, 150 x 4.6 mm, 3 micron) at 254 nm. Mobile phase A: H2O + 1% TFA, mobile phase B: ACN. Flow rate: 0.5 mL/min. Gradient: 0–10% B in 4 min, 10–80% B in 3 min, 80% B for 3 min, 80-50% B in 1 min and 50% B for 1 minute.

NMR aqueous stability data: Aqueous stability of 5-VU and 2-VA was assessed by dissolving the vinyl nucleotides in D2O to achieve 100mM final concentration. 400µL of this solution was transferred to a fresh NMR tube and NMR spectroscopy was performed at 24-hour intervals for two days, incubating the NMR tube at room temperature between the experiments.

RNA biotinylation with Maleimide-biotin and TCEP-biotin: HEK293T cells were plated at 1x105 cells per well in a 6-well plate and grown in Dulbecco's modified Eagle medium (DMEM, high glucose, Corning #10-017-CM) supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin (10,000 U/mL, ATCC #30-2300) for 24-36 hours in a 37°C, 5% carbon dioxide incubator. At the start of [6] treatment, HEK293T cells were incubated in a final concentration of 1 mM 5-vinyl-uridine (Jena Bioscience #CLK-049), 2-vinyl-adenosine or DMSO for 5 hours. The final concentration of DMSO in cell culture medium for all treatments was 0.25% DMSO for all treatments. After 5 hours, total cellular RNA was extracted with TriZol (Thermofisher #15596026) according to the

manufacturer's protocol. RNA pellets were reconstituted in nuclease-free water and quantified using a NanoDrop 2000. 15 µg of each RNA sample was reacted in solutions with 1 mM maleimide-biotin (Sigma # B1267), 1 mM TCEP-biotin (Cayman Chemical #25365), 1 mM tetrazine-biotin (Sigma #793329), or 1mM DMSO (negative controls), all of which were 10% DMSO final reactions in nuclease-free water. For time and concentration tests with each biotinylating reagent or DMSO (negative control), different reaction conditions were tested and compared which are specified in Fi S14. Reactions were performed for 3 hours or less at room temperature and with 500 RPM shaking and under dark conditions. All biotinylating reagents were all dissolved and prepared in DMSO as 10 mM solution stocks stored in -20°C. RNA was purified after biotinylation using RNA Clean and Concentrator Kit #5 (Zymo Research #R1013) and eluted in 20 ul of nuclease-free water prior to dot blot analysis.

RNA dot blot analysis with chemiluminescent imaging: To detected biotinylated RNA with dot blot analysis 2 ug of purified, biotinylated RNA was fixed to a Hybond-N+ (GE Healthcare) membrane with 254 nm light using a UV stratalinker (Stratagene). The membrane was first equilibrated in 2x SSC buffer before spotting RNA and fixing with UV-light. After fixing RNA, membranes were blocked in [7] blocking buffer (0.12 M NaCl, 0.016 M Na2HPO4, 0.008 M NaH2PO4, 0.17 M SDS) for 30 minutes and followed by incubation with high sensitivity streptavidin-HRP (Fisher Scientific, Cat#: PI21130) at 1:5000 dilution in blocking buffer for 5 minutes. The membrane was washed twice in a Wash A buffer (1:10 dilution of blocking buffer in D.I. water) for 30 minutes and twice in Wash B buffer (0.1 M Tris-base, 0.1 M NaCl, 0.02M MgCl2, pH 9.5) for 5-10 minutes. Membranes were then incubated for 1-5 minutes in ECL Chemiluminescent Substrate (Fisher Scientific, Cat#:

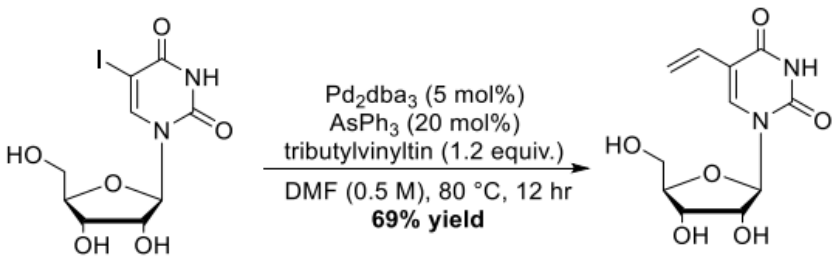
PI32106) and imaged on a ChemiDoc MP imaging system (Bio-Rad) to detect chemiluminesence on the membrane. After imaging is completed, methylene blue (0.04% methylene blue and 0.3 M sodium acetate) was used to stain onto membranes over night, rinsed in D.I water, followed by colorimetric imaging with the ChemiDoc MP imaging systems to detect consistent loading of RNA onto the membranes.

Cell imaging with fluorophore-conjugated reagents in HEK293T: HEK293T cells were grown in 6-well plates (growth conditions in section VII) onto polyD-lysine (10µg/ml) coated coverslips for 24-36 hours to ensure attachment to the coverslips. At the time of treatment, cells were incubated with 1 mM of 5-vinyl-uridine, 2- vinyl-uridine or DMSO (negative control) for 5 hours in 2 mL of complete DMEM per well in 6-well tissue culture plates. Cells were washed gently twice in DPBS (GenClone # 25- 508) to remove residual complete media and subsequently fixed in 3.7% methanol-free paraformaldehyde prepared with 0.15% Triton-X100 dissolved in DPBS for 30 minutes at room temperature to ensure complete permeabilization and fixation of cells. Cells were washed twice in DPBS for 10 minutes, with gentle shaking. The following protocols [8] described in the sub-sections below were used for biotinylation of fixed cellular RNA with maleimide-AlexaFluor488 or with TCEP-biotin conjugated using streptavidinAlexaFluor488. After the completion of each respective biotinylation reaction in fixed cells, DNA was stained by incubating coverslips with cells in 1:2000 Hoeschst 33342. Then, coverslips were mounted onto slides with VectaShield Anti-fade (Vector Labs) for fluorescent confocal microscopy with a Zeiss LSM 700 microscope. Maleimide-AlexaFluor488 to image 5-vinyl-uridine labeled RNA in fixed HEK293T Using DMSO as a negative control, cells treated with 1 mM 5-vinyl-uridine for 5 hours are fixed with paraformaldehyde and washed in DPBS, as

indicated in the procedure above. 1 mM of unconjugated TCEP (Sigma #75259) was prepared in DPBS to reduce disulfide bonds in which cells were incubated for 30 minutes in the dark. TCEP was removed and freshly prepared 500 mM iodoacetamide (Sigma # I6125) in DPBS was immediately applied to fixed cells and incubated overnight (18-20 hours) at 37°C in the dark carefully ensuring cells are completely immersed in solution for alkylation of endogenous thiols. These were washed with DPBS twice for 10 minutes followed by incubation in bovine serum albumin (BSA) (1mg/ ml in DPBS) for 25 minutes, then washed again twice for 10 minutes in DPBS. Maleimide-AlexaFluor488 diluted in DPBS to 2 uM final concertation was applied (Thermofisher # A10254) for 30-45 minutes in the dark at room temperature, without shaking. After aspirating the maleimide solution, 3 total 0.1% Triton-X100 in DPBS washes were done for 5 minutes each with gentle shaking. Hoechst 33342 (Thermofisher #H3570) diluted in DPBS 1:2000 was incubated at room temperature in the dark for 10- [9] 20 minutes and washes with DPBS 3 times for 5 minutes. Coverslips were mounted onto slides for imaging. TCEP-biotin conjugated with streptavidin-AlexaFluor488 to image 2-vinyl-adenosine labeled RNA in fixed HEK293T Using DMSO as a negative control, cells treated with 2-vinyl-adenosine are fixed with paraformaldehyde and washed in DPBS, identical to the protocols above. To block endogenous biotin-containing proteins, 2 mg/ml unconjugated streptavidin (Thermofisher # 21125) was dissolved in DPBS and incubated with fixed cells for 30 minutes. Without aspirating, BSA was added (1 mg/ml in DPBS) for 35 minutes and then washed again twice for 10 minutes in DPBS with gentle shaking. After aspirating DPBS, 100 μM of TCEP-biotin was prepared in nuclease-free water and incubated on cells for 30 minutes in dark conditions without shaking, at room temperature. After washing in DPBS for 10 minutes, twice, 2mg/ml of streptavidin-

AlexaFluor488 (Thermofisher #S11223) was dissolved 1:1000 in DPBS and applied to cells for 30 minutes in order to conjugate the fluorophore onto biotinylated, labeled RNA. After aspirating, 0.1% Triton-X100 in DPBS washes were done for 5 minutes. Hoechst 33342 (Thermofisher #H3570) diluted in DPBS 1:2000 was incubated at room temperature in the dark for 10-20 minutes and washes with DPBS 3 times for 5 minutes. Coverslips were mounted onto slides for imaging.

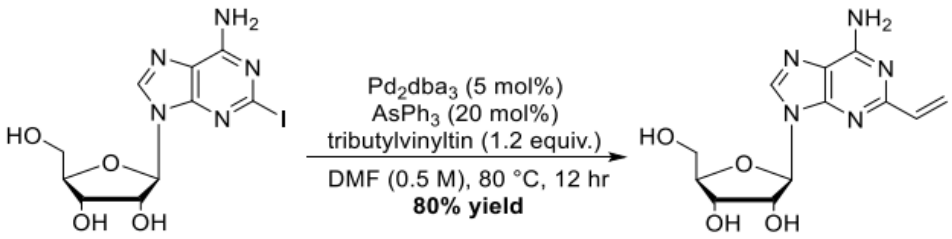
Synthesis of 5-vinyl-uridine (5-VU):



5-vinyl-uridine was prepared according to previously reported method.12 Briefly, 5- Iodouridine (0.50 g, 1.35 mmol, Pd2(dba)3 (67.5 mg, 67.6 µmol), AsPh3 (82.7 mg, 0.27 mmol) was dissolved in anhydrous DMF and purged with N2 for 15 minutes (3 mL). Tributylvinylstannane (0.47 mL, 1.62 mmol) was added to the mixture and stirred with heating in an oil bath for 10 h at 80 °C. Upon completion, the mixture was concentrated in vacuo and purified by column chromatography (EtOAc followed by 0-10% MeOH/DCM) to give 5-VU (0.25 mg, 69%) as a white solid. 1H NMR (400 MHz, DMSO) δ 11.43 (s, 1H), 8.20 (s, 1H), 6.37 (dd, J = 17.7, 11.6 Hz, 1H), 5.91 (dd, J = 17.7, 2.0 Hz, 1H), 5.78 (d, J = 4.7 Hz, 1H), 5.41 (d, J = 5.5 Hz, 1H), 5.23 (t, J = 4.9 Hz, 1H), 5.12 (dd, J = 11.5, 1.9 Hz, 1H), 5.06 (d, J = 5.4 Hz, 1H), 4.03 (ddd, J = 24.8, 10.0, 5.0 Hz, 2H), 3.86 (dd, J = 7.4, 2.8 Hz, 1H), 3.73 – 3.66 (m,

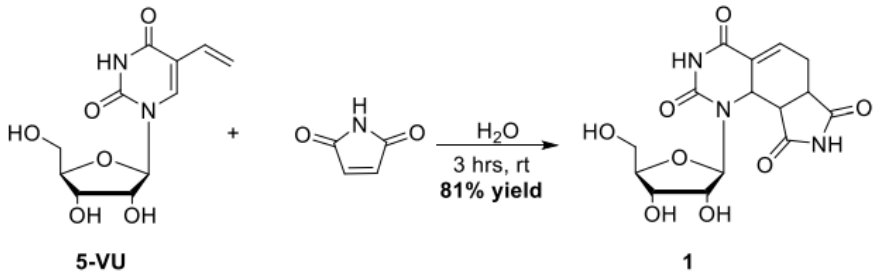
1H), 3.62 – 3.54 (m, 1H). 13C NMR (126 Hz, DMSO) δ 162.1, 149.8, 138.0, 128.6, 114.0, 110.8, 88.2, 84.6, 73.9, 69.4, 60.4.

Synthesis of 2-vinyl-adenosine (2-VA):



[11] 2-vinyl-adenosine was synthesized according to previously reported method.12 Briefly, 2- Iodoadenosine (500 mg, 1.27 mmol), Pd2dba3 (58.2 mg, 63.5 μmol), and AsPh3 (77.9 mg, 0.25 mmol) was dissolved in DMF (1.3 mL) and stirred with heating in an oil bath at 80 °C. Tributylvinylstannane (0.45 mL, 1.53 mmol) was added to the mixture and stirred for 10 h. The reaction was then concentrated in vacuo and purified by silica flash column chromatography (10% MeOH/DCM) to give 2-VA (335 mg, 90%) as a tan solid. 1H NMR (400 MHz, DMSO) δ 8.32 (s, 1H), 7.30 (s, 2H), 6.60 (dd, J = 17.3, 10.4 Hz, 1H), 6.36 (dd, J = 17.3, 2.3 Hz, 1H), 5.88 (d, J = 6.5 Hz, 1H), 5.55 (dd, J = 10.4, 2.3 Hz, 1H), 5.46 – 5.28 (m, 1H), 5.19 (d, J = 4.4 Hz, 1H), 4.65 (dd, J = 11.2, 6.0 Hz, 1H), 4.15 (dd, J = 7.2, 4.5 Hz, 1H), 3.97 (dd, J = 6.3, 3.4 Hz, 1H), 3.68 (dt, J = 12.0, 4.2 Hz, 1H), 3.61 – 3.49 (m, 1H).13C NMR (126 MHz, DMSO) δ 157.8, 155.8, 149.7, 140.4, 137.1, 121.2, 118.6, 87.7, 86.0, 73.3, 70.8, 61.8.

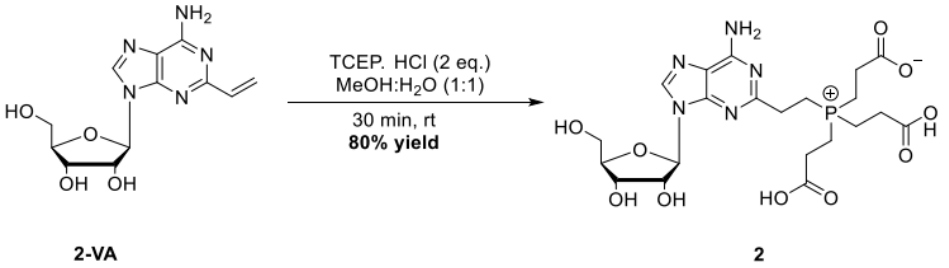
Synthesis of 1-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2- yl)-6,6a,9a,9b-tetrahydro-1H-pyrrolo[3,4-h]quinazoline-2,4,7,9(3H,8H)-tetraone (1):



5-VU 1

Maleimide (7.2 mg, 0.074 mmol) was added to a solution of 5-VU (10 mg, 0.037 mmol) in H2O (1 mL) in a 1.5 mL eppendorf tube at room temperature. The resulting solution was incubated in a shaker at room temperature. After 3 hours, [12] 2,3-dimethyl-1,3-butadiene (84 µL, 0.74 mmol) was added in the eppendorf tube and incubated in a shaker for 1 hour. The resulting solution was washed with chloroform (300 µL x 3) to remove the organic impurities and aqueous layer was dried in vacuo to obtain compound 1 (11 mg, 81%) as white solid. Extinction coefficient ε (220 nm, absorbance maxima) = 10670.2 ± 612.5 M-1 cm-1 , ε (260 nm) = 2081.3 ± 201.2 M-1 cm-1 . HRMS Calcd for C15H17N3O8Na [M+Na+] 390.0913, found 390.0928 [M+Na+]. 1H NMR (400 MHz, DMSO) δ 11.14 (s, 1H), 10.43 (s, 1H), 7.02 (s, 1H), 5.58 (d, J = 5.5 Hz, 1H), 5.06 (d, J = 5.5 Hz, 1H), 4.90 (d, J = 5.4 Hz, 1H), 4.82 (t, J = 5.1 Hz, 1H), 4.43 – 4.39 (m, 1H), 4.29 (q, J = 5.5 Hz, 1H), 3.98 (dd, J = 10.1, 5.2 Hz, 1H), 3.90 (dd, J = 8.6, 5.1 Hz, 1H), 3.72 (dd, J = 7.8, 4.0 Hz, 1H), 3.58 (ddd, J = 11.6, 5.0, 3.5 Hz, 1H), 3.53 – 3.48 (m, 1H), 3.14 (t, J = 8.4 Hz, 1H), 2.72 (dd, J = 15.0, 7.9 Hz, 1H), 2.33 (s, 1H). 13C NMR (101 MHz, DMSO) δ 180.1, 178.0, 160.0, 151.6, 136.2, 128.8, 90.7, 83.7, 70.5, 69.7, 61.4, 50.9, 45.5, 25.2.

Synthesis of 3-((2-(6-amino-9-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-9H-purin-2-yl)ethyl)bis(2-carboxyethyl)phosphonio)propanoate (2):



[13] TCEP.HCl (19.5 mg, 0.068 mmol) was added to a solution of 2-VA (10 mg, 0.034 mmol) in MeOH/H2O (1 mL, 1:1) at room temperature. The resulting solution was stirred at room temperature for 30 minutes. The reaction mixture was evaporated to dryness in vacuo and then purified by semi-preparative RP-HPLC (Thermo-Scientific Hypersil BDS C18 column, 150 x 4.6 mm, 3 micron) at 254 nm to obtain compound 2 (15 mg, 80%) as a white solid. Mobile phase A: H2O + 1% TFA, mobile phase B: ACN. Flow rate: 3 mL/min. Gradient: 0−10% B in 20 min and 10−100% B in 15 min. HRMS Calcd for C21H30N5O10PH [M+H+] 544.1808, found 544.1809 [M+H+]. Extinction coefficient ε (210 nm, absorbance maxima) = 12063.0 ± 457.5 M-1 cm-1 , ε (260 nm) = 8353.4 ± 278.6 M-1 cm-1 . 1H NMR (400 MHz, D2O) δ 8.50 (s, 1H), 6.13 (d, J = 5.2 Hz, 1H), 4.41 (t, J = 4.8 Hz, 1H), 4.22 (dd, J = 7.6, 4.2 Hz, 1H), 3.87 (dd, J = 12.7, 3.1 Hz, 1H), 3.79 (dd, J = 12.7, 4.4 Hz, 1H), 3.37 (dt, J = 10.8, 7.9 Hz, 2H), 2.99 (dt, J = 12.3, 4.5 Hz, 2H), 2.83 (dt, J = 7.3, 6.3 Hz, 8H), 2.71 (dt, J = 14.9, 7.5 Hz, 8H). 13C NMR (125Hz, D2O) 174.4, 174.3, 163.0, 162.7, 154.9, 154.7, 150.5,

148.5, 142.8, 120.6, 117.7, 117.6, 114.8, 88.3, 85.4, 74.2, 70.2, 61.0, 26.0, 25.7, 25.7, 15.8, 15.3, 14.5, 14.1, 14.0.

Supplementary Figures

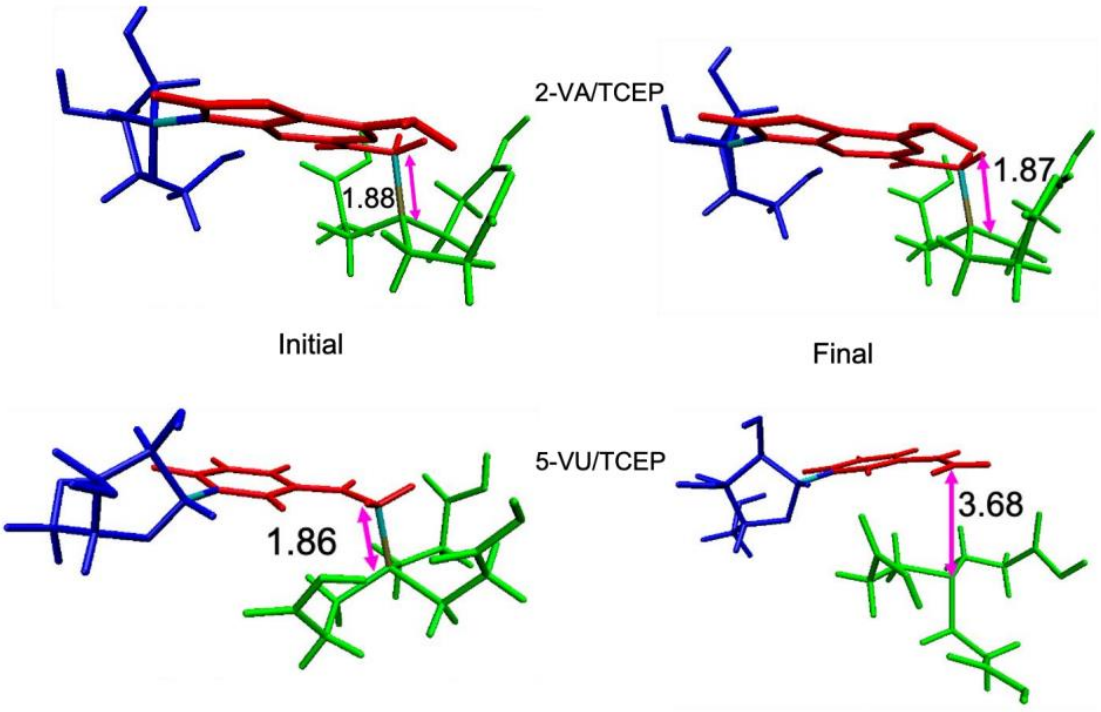


Figure 4-S1 Illustration of bond C-P bond breaking in 5-VU/TCEP intermediate complex during optimization indicated by a longer C-P bond length in lower right panel (3.68 Å).

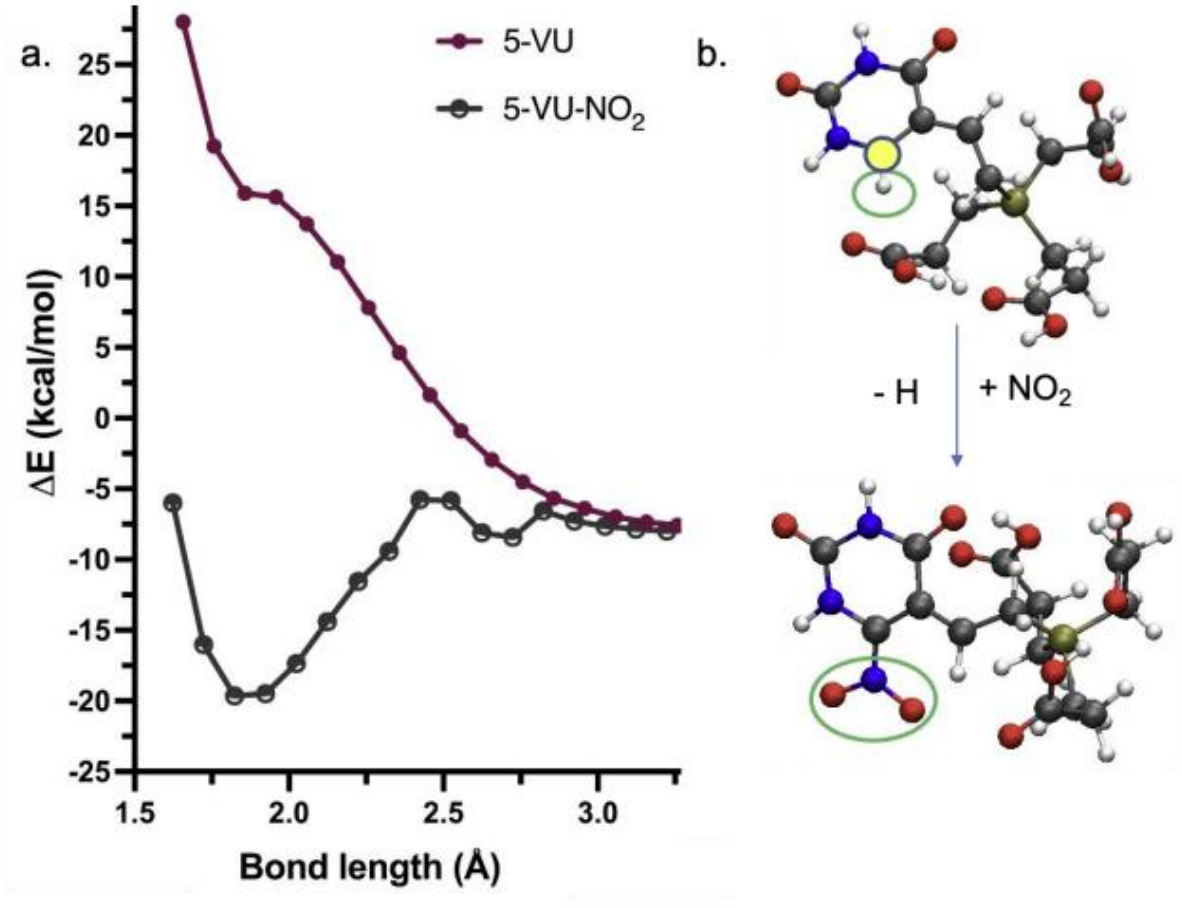


Figure 4-S2 Stabilization of the 5-VU intermediate complex indicated by shallower potential energy surface when H of C-6 of uracil (highlighted yellow) is substituted by an electron withdrawing -NO2 group. The potential energy surface of 5-VU is duplicated for easy comparison. b. Schematic of 5-VU and 5-VU-NO2 molecular structures.

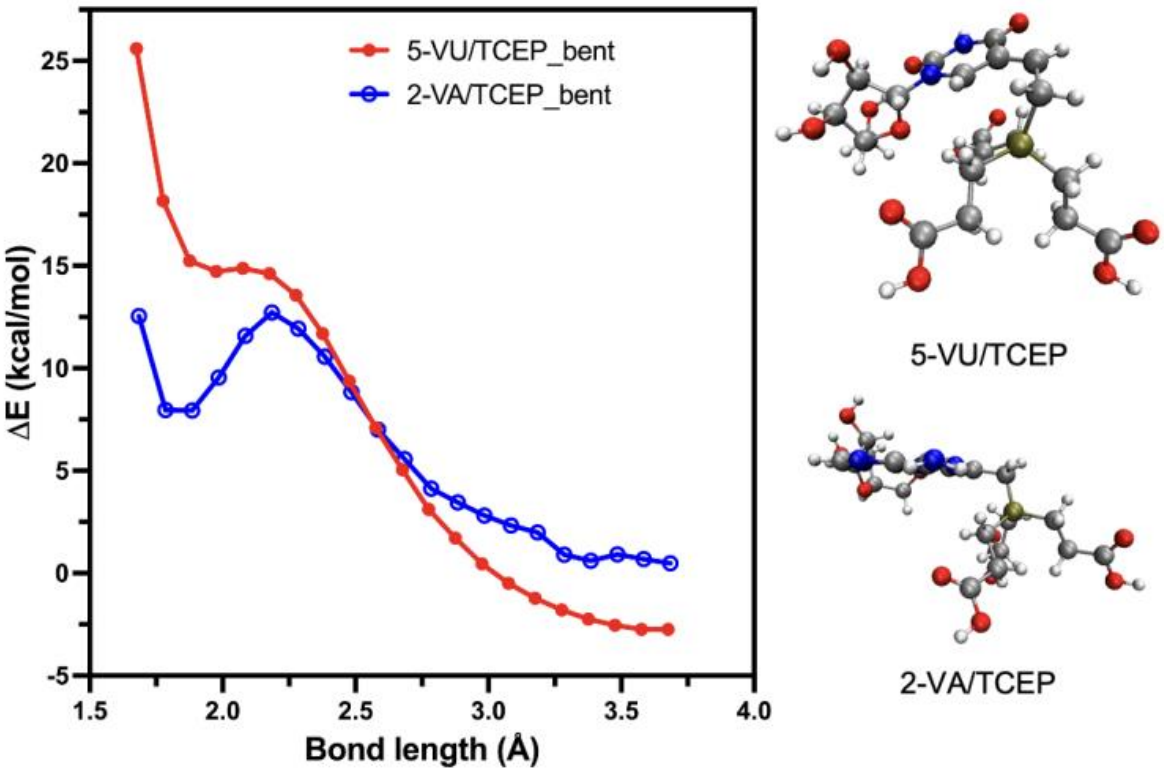


Figure 4-S3 Potential energy scan where TCEP is kept spatially away from the nucleoside.

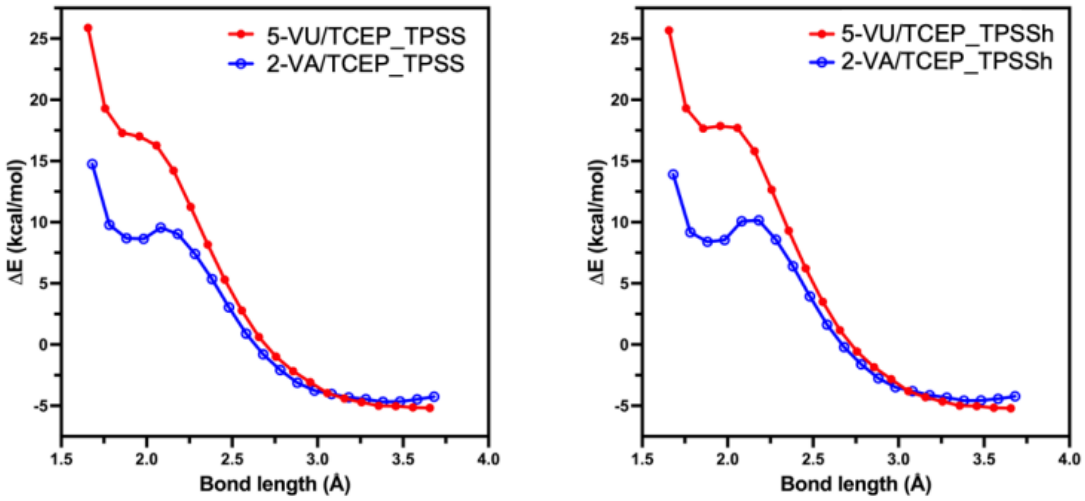


Figure 4-S4 Potential energy scan with TPSSh functional (right panel). The scan with TPSS (left panel) is duplicated for easy comparison.

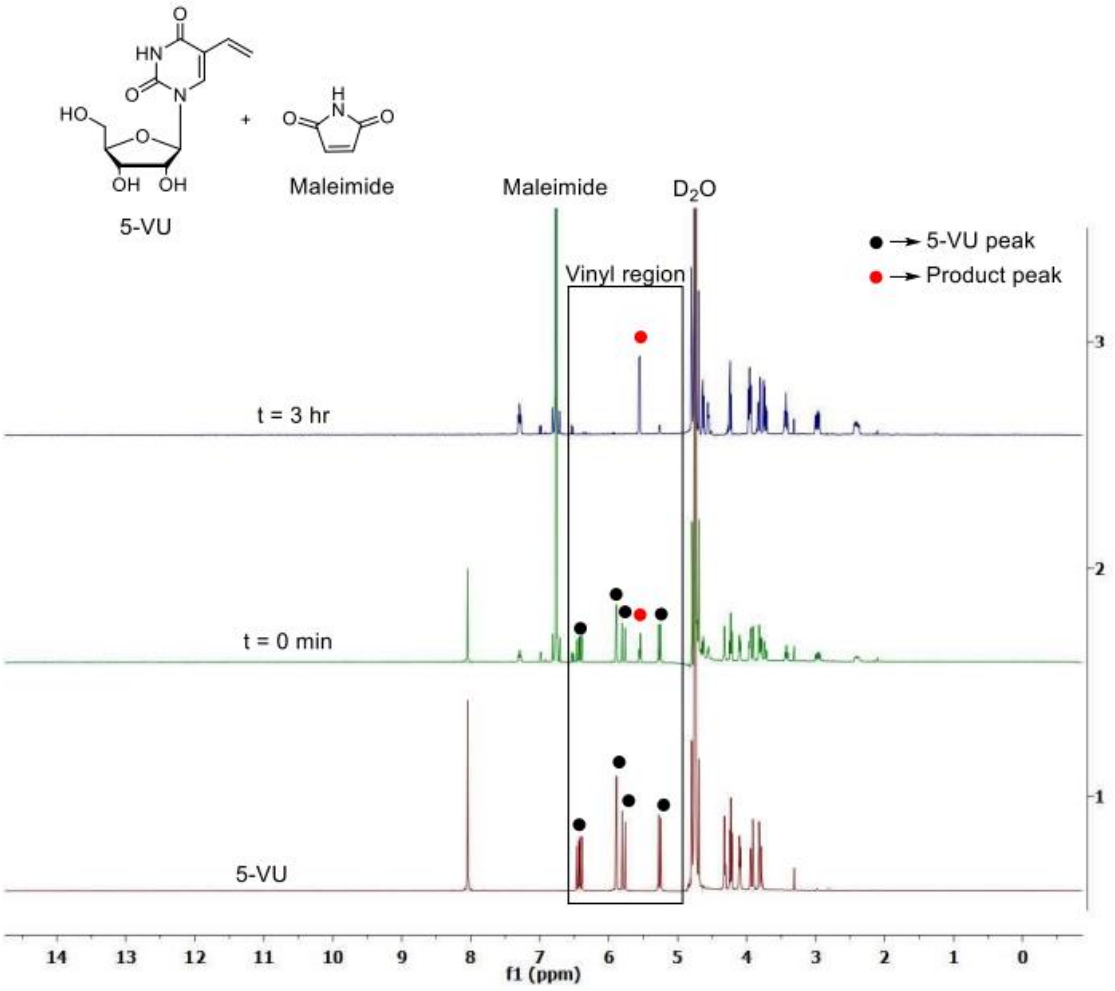


Figure 4-S5 Reactivity study of 5-VU with maleimide. Characteristic vinyl peaks of 5-VU were observed to disappear after 3-hour incubation at room temperature.

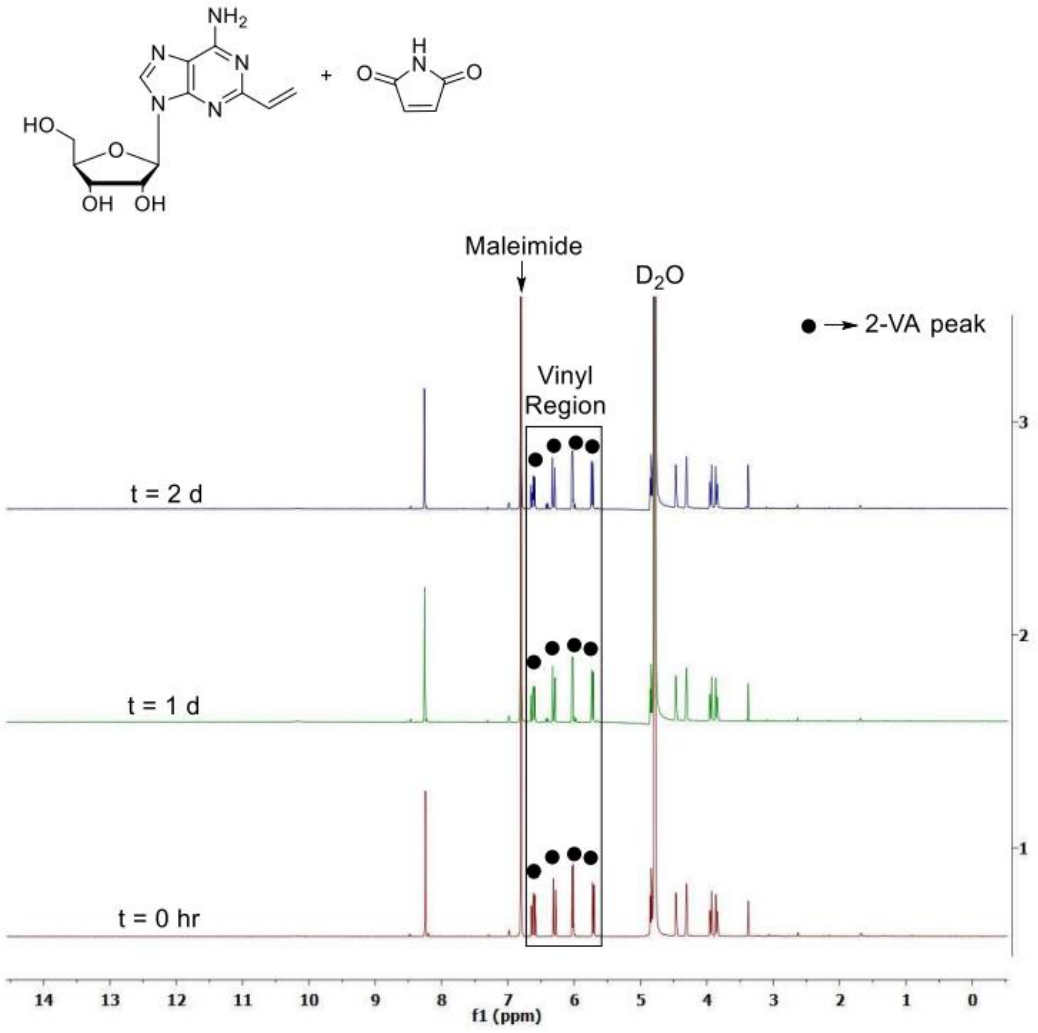


Figure 4-S6 Reactivity study of 2-VA with maleimide. Characteristic vinyl peaks of 2-VA and other peaks were observed to remain intact after 48-hour incubation at room temperature.

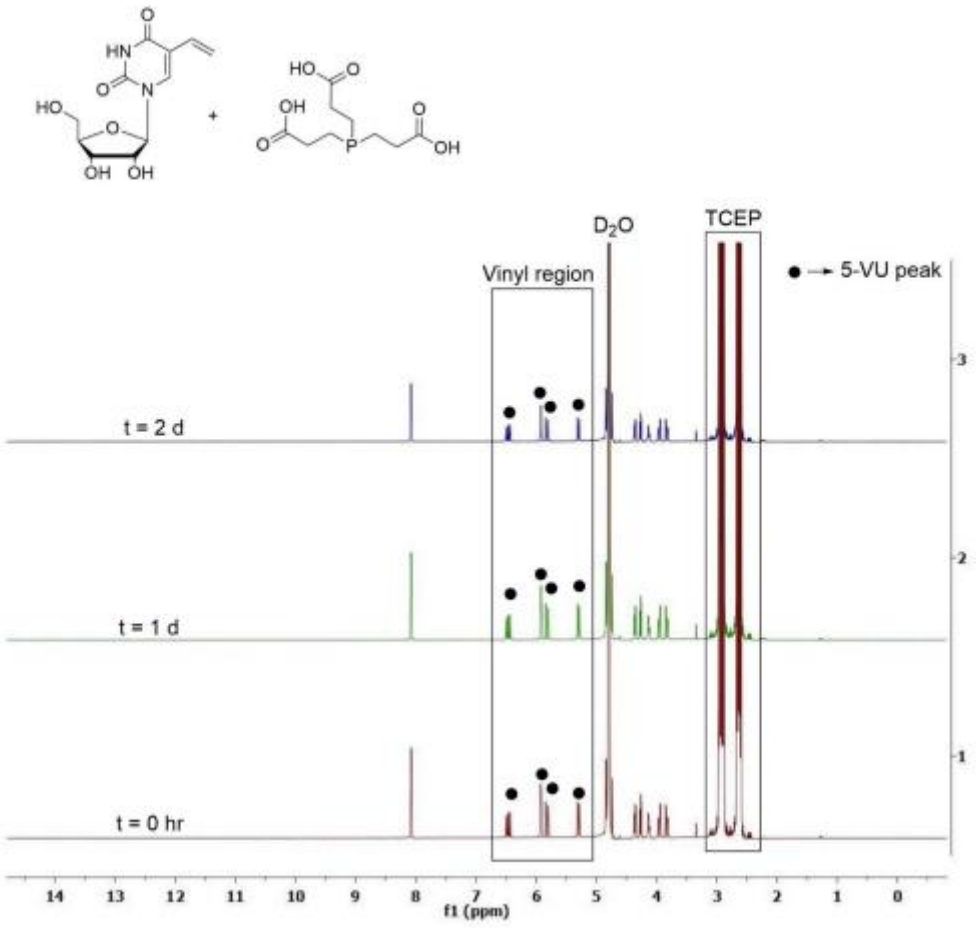


Figure 4-S7 Reactivity study of 5-VU with TCEP.HCl. Characteristic vinyl peaks of 5-VU and other peaks were observed to remain intact after 48-hour incubation at room temperature.

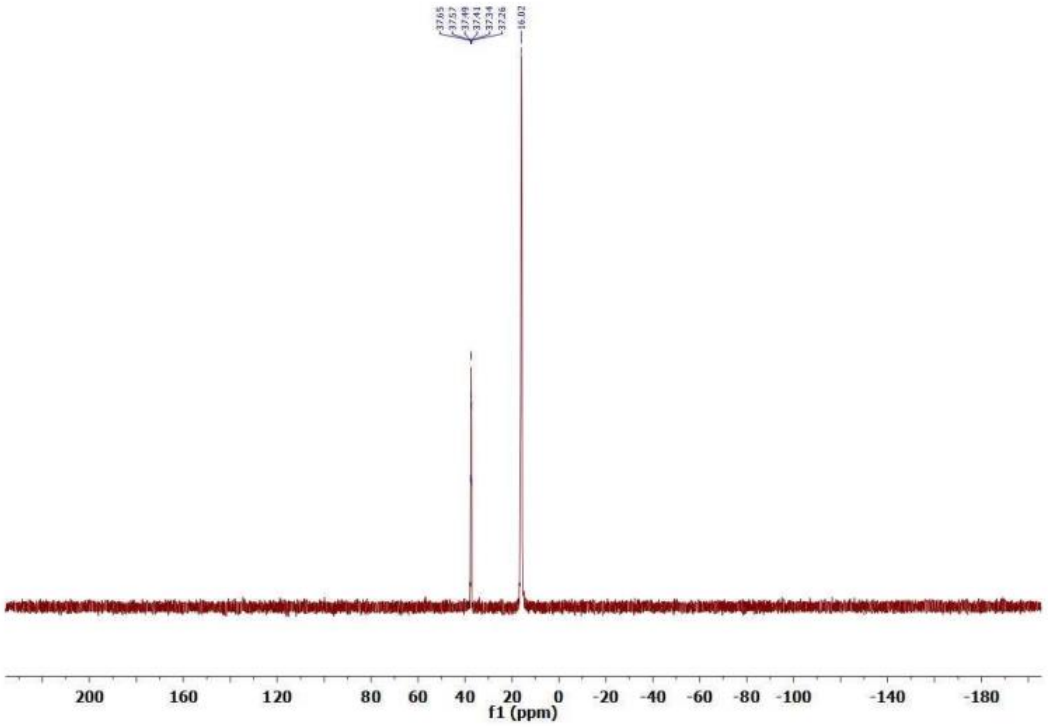


Figure 4-S8 Reactivity study of 2-VA with TCEP by 31P NMR. An upfield peak for product appears at 37.45 ppm. Chemical shift calibrated with residual TCEP at 16.02 ppm.

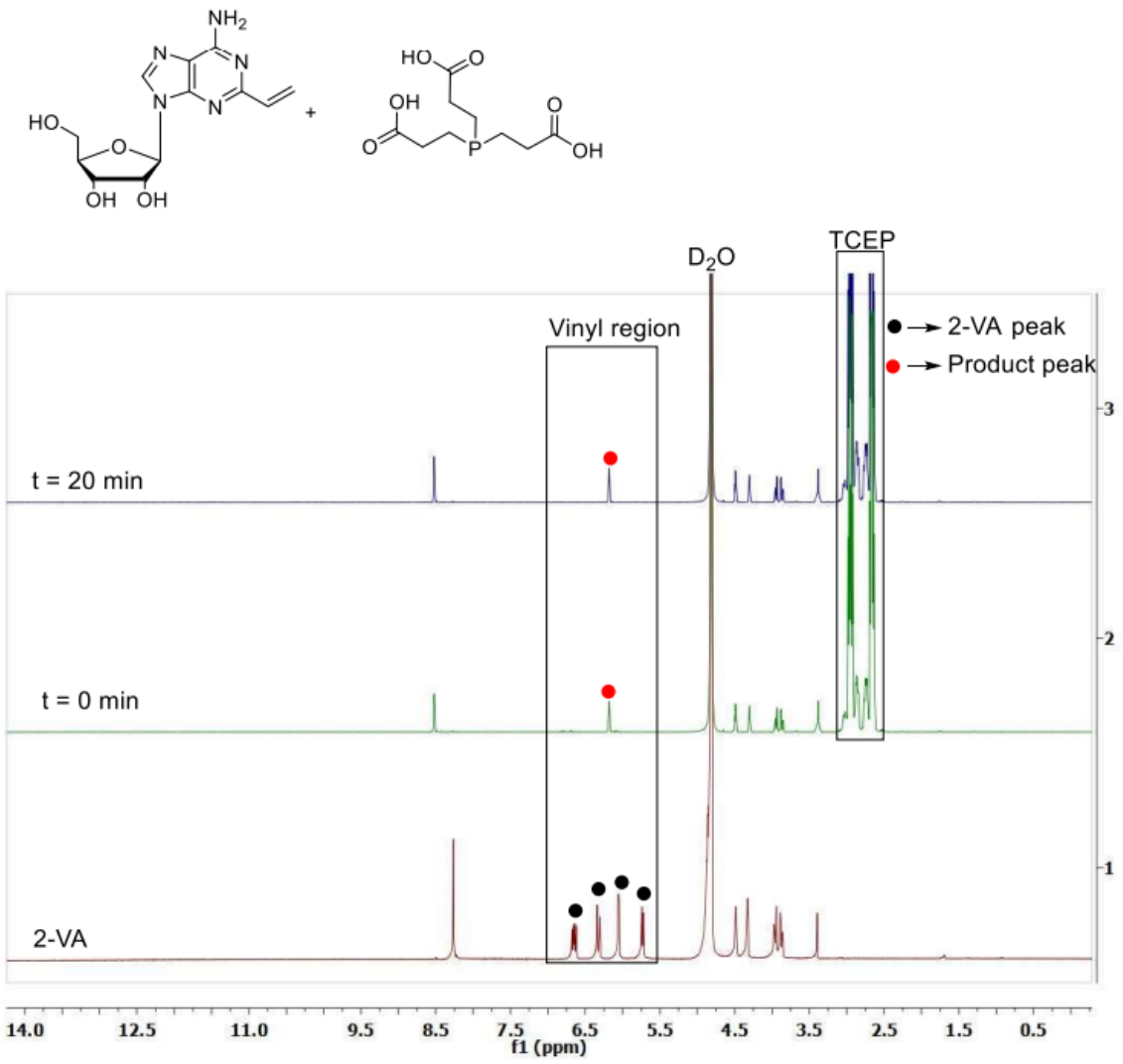


Figure 4-S9 Reactivity study of 2-VA with TCEP.HCl. Characteristic vinyl peaks of 2-VA were observed to completely disappear after 20 minutes incubation at room temperature.

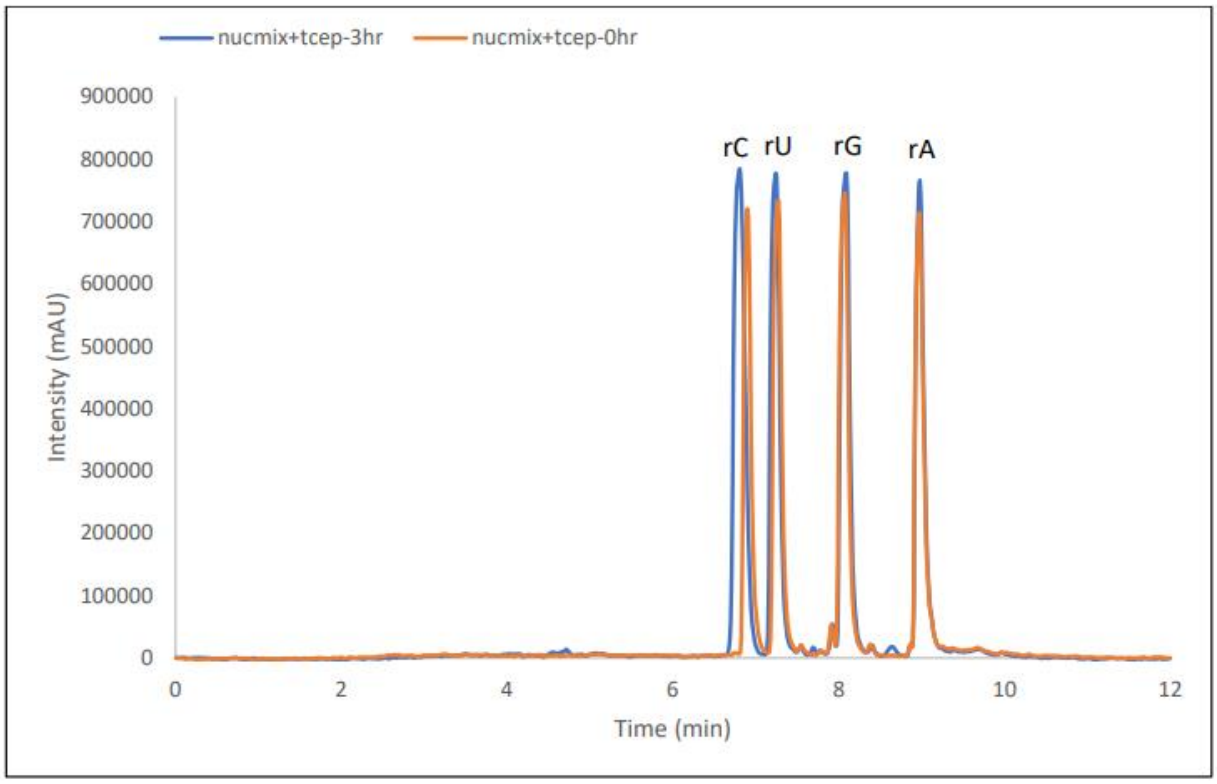


Figure 4-S10 Background reactivity study of natural ribonucleosides incubated with TCEP.

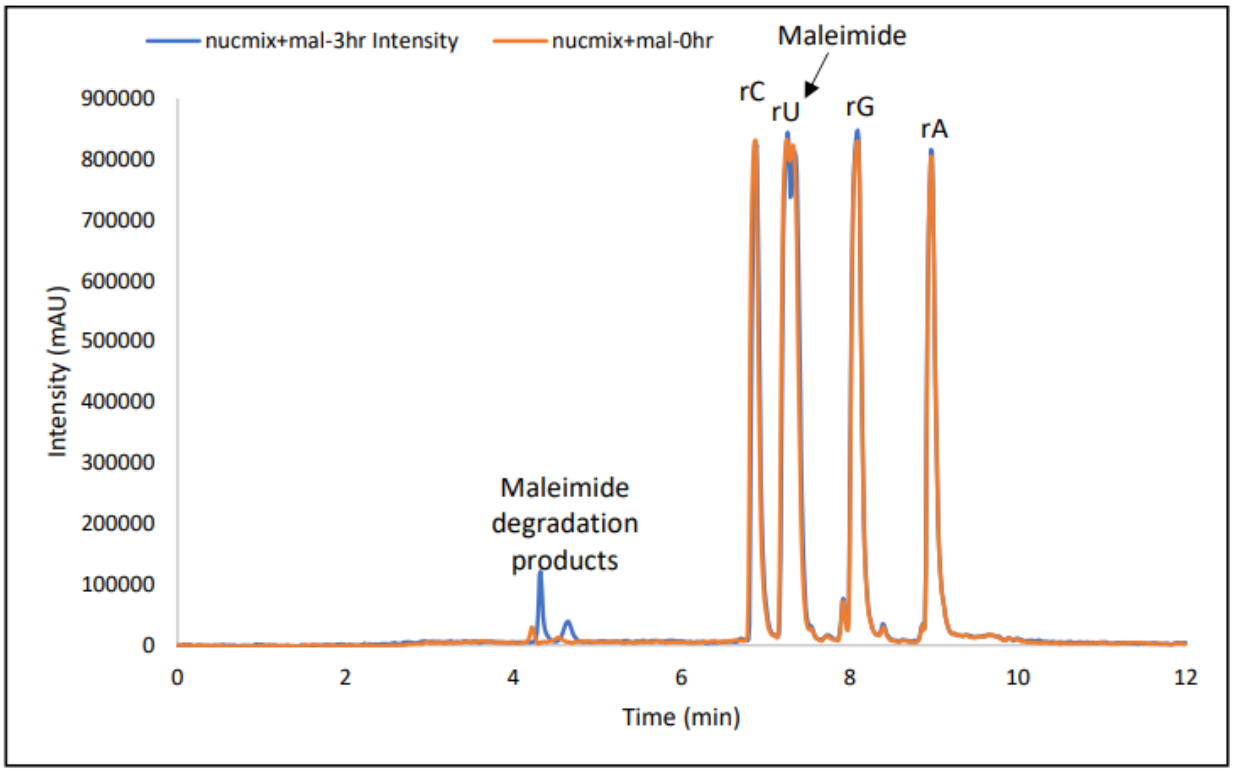


Figure 4-S11 Background reactivity study of natural ribonucleosides incubated with maleimide.

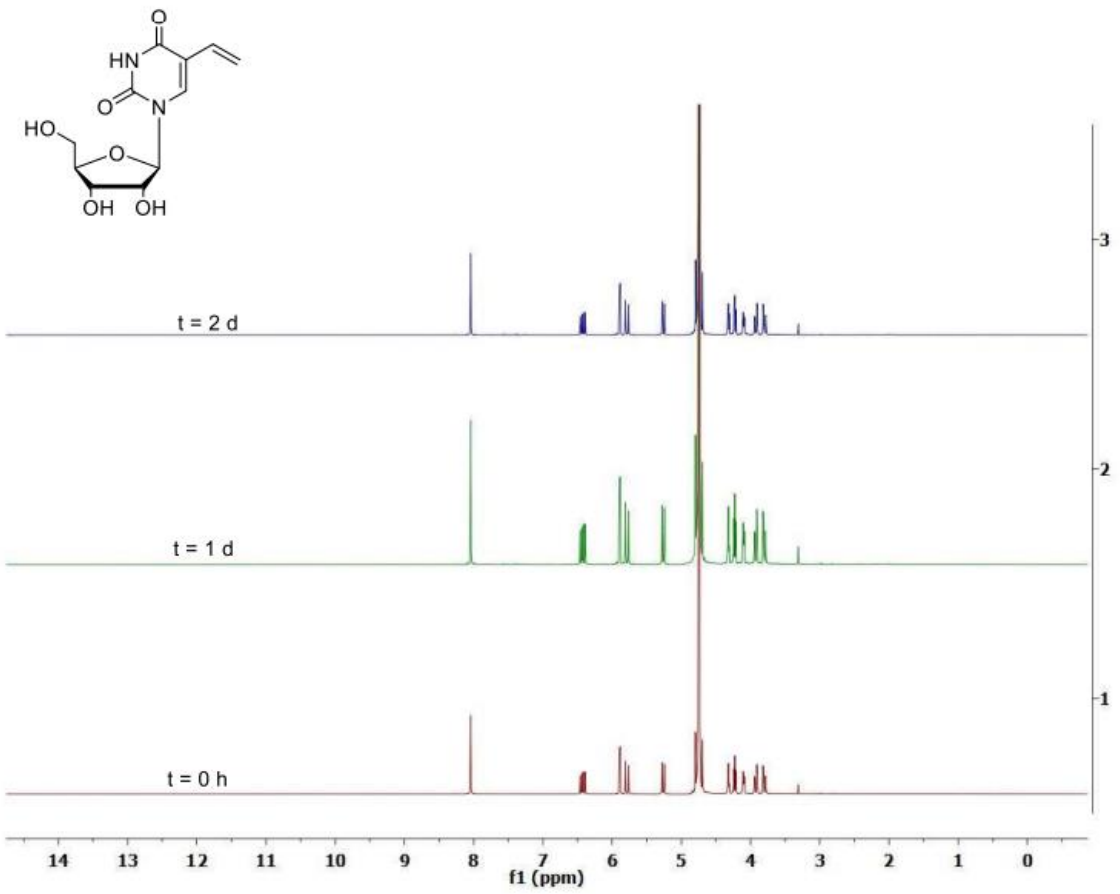


Figure 4-S12 Aqueous stability profile of 5-VU observed over 48 hours. No signs of chemical modification were observed.

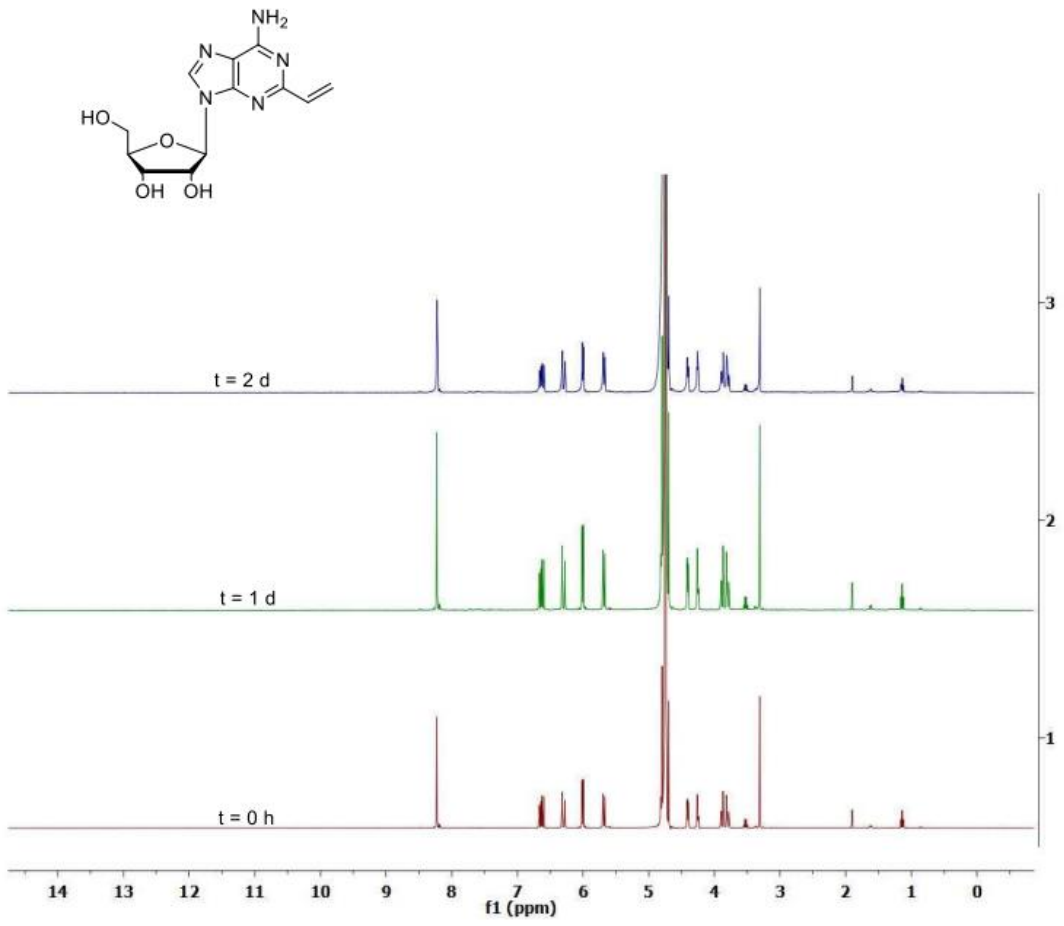


Figure 4-S13 Aqueous stability profile of 2-VA observed over 48 hours. No signs of chemical modification were observed.

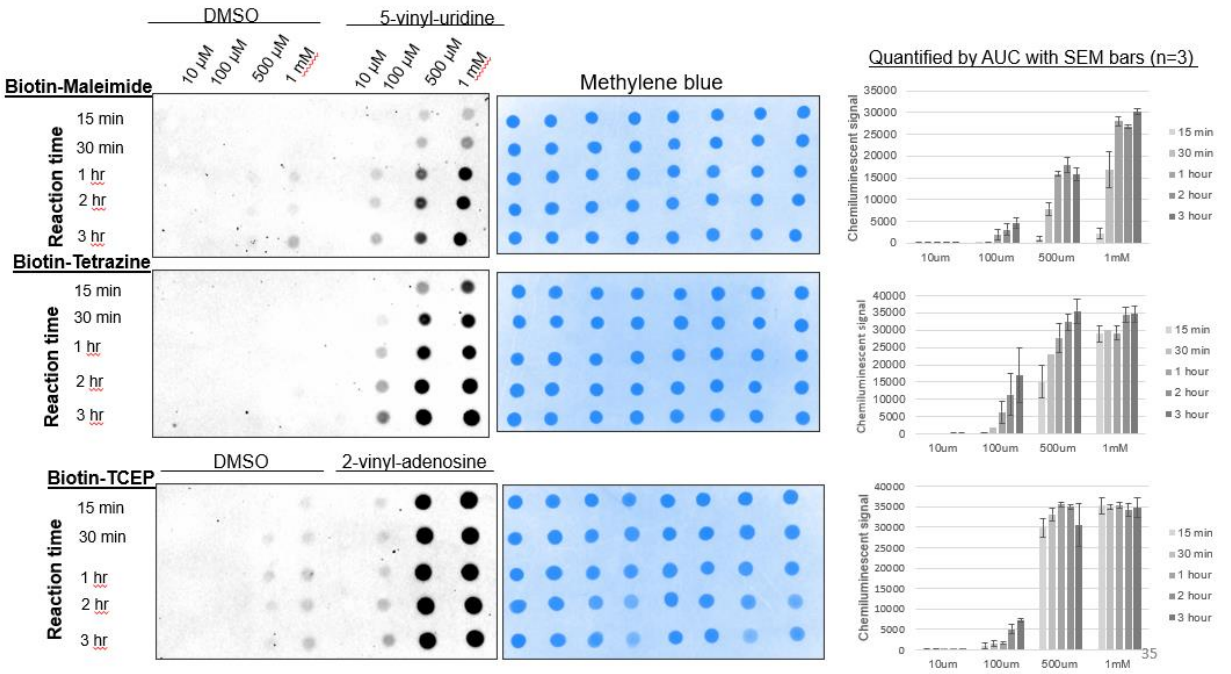


Figure 4-S14 Time and concentration study with biotin-maleimide, biotin-tetrazine and biotin-TCEP depicted through dot blot assay and signal quantification with ImageJ software.

References

1 Yao, R.-W.; Wang, Y.; Chen, L.-L. Cellular Functions of Long Noncoding RNAs. *Nat. Cell Biol.* **2019**, *21*, 542– 551.

2 Neymotin, B.; Athanasiadou, R.; Gresham, D. Determination of *in vivo* RNA kinetics using RATE-seq. *RNA* **2014**, *20* (10), 1645– 1652.

3 Singha, M.; Spitalny, L.; Nguyen, K.; Vandewalle, A.; Spitale, R. C. Chemical Methods for Measuring RNA Expression with Metabolic Labeling. *Wiley Interdiscip. Rev.: RNA* **2021**, *12*, e1650.

4 Muthmann, N.; Hartstock, K.; Rentmeister, A. Chemo-Enzymatic Treatment of RNA to Facilitate Analyses. *Wiley Interdiscip. Rev.: RNA* **2020**, *11* (1), e1561.

5 Kawata, K.; Wakida, H.; Yamada, T.; Taniue, K.; Han, H.; Seki, M.; Suzuki, Y.; Akimitsu, N. Metabolic Labeling of RNA Using Multiple Ribonucleoside Analogs Enables the Simultaneous Evaluation of RNA Synthesis and Degradation Rates. *Genome Res.* **2020**, *30*, 1481– 1491.

6 Rieder, U.; Luedtke, N. W. Alkene-tetrazine ligation for imaging cellular DNA. *Angew. Chem., Int. Ed.* **2014**, *53* (35), 9168– 72.

7 Burger, K.; Mühl, B.; Kellner, M.; Rohrmoser, M.; Gruber-Eber, A.; Windhager, L.; Friedel, C. C.; Dölken, L.; Eick, D. 4-Thiouridine Inhibits rRNA Synthesis and Causes a Nucleolar Stress Response. *RNA Biol.* **2013**, *10*, 1623– 1630.

8 Kubota, M.; Nainar, S.; Parker, S. M.; England, W.; Furche, F.; Spitale, R. C. Expanding the Scope of RNA Metabolic Labeling with Vinyl Nucleosides and Inverse Electron-Demand Diels–Alder Chemistry. *ACS Chem. Biol.* **2019**, *14*, 1698– 1707.

9 Ravasco, J. M. J. M.; Faustino, H.; Trindade, A.; Gois, P. M. P. Bioconjugation with Maleimides: A Useful Tool for Chemical Biology. *Chem. - Eur. J.* **2019**, *25*, 43– 59.

10 Borsenberger, V.; Howorka, S. Diene-Modified Nucleotides for the Diels–Alder-Mediated Functional Tagging of DNA. *Nucleic Acids Res.* **2009**, *37*, 1477– 1485.

11 Burns, J. A.; Butler, J. C.; Moran, J.; Whitesides, G. M. Selective Reduction of Disulfides by Tris(2-Carboxyethyl)Phosphine. *J. Org. Chem.* **1991**, *56*, 2648– 2650.

12 Bos, J.; Muir, T. W. A Chemical Probe for Protein Crotonylation. *J. Am. Chem. Soc.* **2018**, *140*, 4757– 4760.

13 Weinhold, F. Natural Bond Orbital Analysis: A Critical Overview of Relationships to Alternative Bonding Perspectives. *J. Comput. Chem.* **2012**, *33*, 2363– 2379.

14 Nguyen, K.; Fazio, M.; Kubota, M.; Nainar, S.; Feng, C.; Li, X.; Atwood, S. X.; Bredy, T. W.; Spitale, R. C. Cell-Selective Bioorthogonal Metabolic Labeling of RNA. *J. Am. Chem. Soc.* **2017**, *139*.

CONCLUSION

The ability to capture nascent transcription has far-reaching implications for the field of RNA biology. These rapidly expressed molecules have a wide range of activities that are important at every level of regulation, often with cell-specific functions[1]. To profile the dynamics of transcription in cells, RNA metabolic labeling techniques provide time-resolved detection of transcripts though labeling of nascent RNA. RNA metabolic labeling methods are developed with a variety of metabolic reporters consisting of non-canonical nucleobase or nucleoside analogs derivatized with chemical handles. Technical advancement with new metabolic reporters and refinement of existing chemical strategies will help to expand the tools available for transcriptome-wide nascent RNA profiling[2].

As many functional RNA are expressed in cell-type specific contexts, cell-specific RNA metabolic labeling becomes a useful laboratory tool. These efforts are continuously maturing as new enzyme-analog pairs with improved cell-selectivity are being developed[2]. Widely employed 4-tu and 5-eu have significant challenges for cell-specific RNA metabolic labeling when applied in wild-type and (+)-*Tg*UPRT cell mixtures due to prevalence of background RNA labeling in wild-type cells[8]. These analogs are also limited by the stability of thiol-conjugation[3], cell toxicity[4-5], short treatment times[6] and detriment to RNA integrity[7]. The 3xUPRT engineered and discovered by our work. demonstrated the increased cell-specific stringency with 5-vu labeling[8]. These findings demonstrate, unlike 4-tu and 5-eu, that 5-vu is not a substrate for endogenously expressed uridine monophosphate synthase therefore limiting non-specific background RNA labeling in wild-type cell populations[8-9].

Additionally, vinyl nucleosides have demonstrated no significant change in cell proliferation with longer treatments supporting overall cell viability[5]. Finally, the RNA quality after tetrazine reactions is maintained, extending the utility of this approach for sequencing workflows and imaging of nascent RNA localization in specific cell-types within tissues[8]. I have tested this approach to label nascent RNA in metastatic breast cancer cells proliferating within mouse models. Cell-specific nascent RNA was visually demonstrated in two tissue types, the mammary fat pad and lungs, with tetrazine-Cy5 signal detected through fluorescent microscopy. Similarly, biotin-streptavidin affinity pulldowns showed high levels of enrichment for metabolically labeled RNA in only the (+)-3xUPRT tumors compared with WT tumors.

Together, these results support future applications for transcriptome-wide sequencing analysis to identify new functional RNA expressed *in viv*o within metastatic breast cancer cells. In further context of this work, cell-specific RNA metabolic labeling was demonstrated with additional gene-specificity through expression of vimentin-3xUPRT in metastatic breast cancer cells. This approach can potentially label nascent RNA preferentially during mesenchymal cell states. Although the complexity of transcriptional states over the course of metastasis are difficult to mechanistically define, tagging the functional RNA critical to these processes[10] is an important goal for identifying new strategies to develop RNA-targeting therapies specific for the prevention of human breast cancer metastasis.

Although the tetrazine ligation reaction is useful for conjugating vinyl-containing RNA, maleimide and TCEP provide additional chemo-selectivity with 5-vinyluridine and 2-vinyladenosine, respectively, for RNA metabolic labeling in cells[11]. With these two vinyl-

nucleosides, nascent RNA can be profiled within two distinct RNA populations expressed at separate time points. For instance, applying each vinyl nucleoside to the same cell culture dish before and after treatment with a drug can potentially differentiate RNA turnover rates[12] before and after dosing. Finally, RNA can be localized to different cell compartments and sequestered based on cell cycle, differentiation states and during metabolic changes[13]. These vinyl-nucleosides can be distinctly detected with fluorescence to show two populations of nascent RNA for capturing changes in RNA localization within cells.

The tools I developed demonstrate a basis for the continued improvement of RNA metabolic labeling strategies in cells and whole organisms to gain a deeper mechanistic understanding of important regulatory RNA. Currently, there are already many unique tools to use for RNA metabolic labeling with each presenting unique challenges and advantages[2]. I have demonstrated overcoming some of these challenges by applying new cell-specific tools in mouse models and applying orthogonal chemo-selective reagent pairs for two vinyl-nucleosides in cells. As sequencing technologies continue to grow rapidly, applying these RNA metabolic labeling approaches in tandem will enhance information about nascent RNA activation providing greater biological insight into transcriptional regulation[14]. Finally, as more tools become readily available to generate complex animal models, adding RNA metabolic labeling approaches into multi-omics studies becomes increasingly promising for defining functional mechanisms of RNA in cell-types within complex tissues [15].

References

1 Wang, K. C.; Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell* **2011**, *43* (6), 904–914.

2 Singha, M.; Spitalny, L.; Nguyen, K.; Vandewalle, A.; Spitale, R. C. Chemical Methods for Measuring RNA Expression with Metabolic Labeling. *WIREs RNA* **2021**, *12* (5).

3 Duffy, E. E.; Rutenberg-Schoenberg, M.; Stark, C. D.; Kitchen, R. R.; Gerstein, M. B.; Simon, M. D. Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. *Molecular Cell* **2015**, *59* (5), 858–866.

4 Burger, K.; Mühl, B.; Kellner, M.; Rohrmoser, M.; Gruber-Eber, A.; Windhager, L.; Friedel, C. C.; Dölken, L.; Eick, D. 4-Thiouridine Inhibits RRNA Synthesis and Causes a Nucleolar Stress Response. *RNA Biology* **2013**, *10* (10), 1623–1630.

5 Kubota, M.; Nainar, S.; Parker, S. M.; England, W.; Furche, F.; Spitale, R. C. Expanding the Scope of RNA Metabolic Labeling with Vinyl Nucleosides and Inverse Electron-Demand Diels–Alder Chemistry. *ACS Chem. Biol.* **2019**, *14* (8), 1698–1707.

6 Biasini, A.; Marques, A. C. A Protocol for Transcriptome-Wide Inference of RNA Metabolic Rates in Mouse Embryonic Stem Cells. *Front. Cell Dev. Biol.* **2020**, *8*, 97.

7 Paredes, E.; Das, S. R. Click Chemistry for Rapid Labeling and Ligation of RNA. *ChemBioChem* **2011**, *12* (1), 125–131.

8 Nguyen, K.; Kubota, M.; Arco, J. del; Feng, C.; Singha, M.; Beasley, S.; Sakr, J.; Gandhi, S. P.; Blurton-Jones, M.; Fernández Lucas, J.; Spitale, R. C. A Bump-Hole Strategy for Increased Stringency of Cell-Specific Metabolic Labeling of RNA. *ACS Chem. Biol.* **2020**, *15* (12), 3099–3105.

9 Nainar, S.; Cuthbert, B. J.; Lim, N. M.; England, W. E.; Ke, K.; Sophal, K.; Quechol, R.; Mobley, D. L.; Goulding, C. W.; Spitale, R. C. An Optimized Chemical-Genetic Method for Cell-Specific Metabolic Labeling of RNA. *Nat Methods* **2020**, *17* (3), 311–318

10 Beasley, S.; Vandewalle, A.; Singha, M.; Nguyen, K.; England, W.; Tarapore, E.; Dai, N.; Corrêa, I. R.; Atwood, S. X.; Spitale, R. C. Exploiting Endogenous Enzymes for Cancer-Cell Selective Metabolic Labeling of RNA in Vivo. *J. Am. Chem. Soc.* **2022**, *144* (16), 7085–7088.

11 Gupta, M.; Singha, M.; Rasale, D. B.; Zhou, Z.; Bhandari, S.; Beasley, S.; Sakr, J.; Parker, S. M.; Spitale, R. C. Mutually Orthogonal Bioconjugation of Vinyl Nucleosides for RNA Metabolic Labeling. *Org. Lett.* **2021**, *23* (18), 7183–7187.

12 Kawata, K.; Wakida, H.; Yamada, T.; Taniue, K.; Han, H.; Seki, M.; Suzuki, Y.; Akimitsu, N. Metabolic Labeling of RNA Using Multiple Ribonucleoside Analogs Enables the Simultaneous Evaluation of RNA Synthesis and Degradation Rates. *Genome Res.* **2020**, *30* (10), 1481–1491.

13 Kloc, M.; Zearfoss, N. R.; Etkin, L. D. Mechanisms of Subcellular MRNA Localization. *Cell* **2002**, *108* (4), 533–544.

14 Battich, N.; Beumer, J.; de Barbanson, B.; Krenning, L.; Baron, C. S.; Tanenbaum, M. E.; Clevers, H.; van Oudenaarden, A. Sequencing Metabolically Labeled Transcripts in Single Cells Reveals mRNA Turnover Strategies. *Science* **2020**, *367* (6482), 1151–1156.

15 Kaczmarczyk, L.; Bansal, V.; Rajput, A.; Rahman, R.; Krzyżak, W.; Degen, J.; Poll, S.; Fuhrmann, M.; Bonn, S.; Jackson, W. S. Tagger—A Swiss Army Knife for Multiomics to Dissect Cell Type–Specific Mechanisms of Gene Expression in Mice. *PLoS Biol* **2019**, *17* (8), e3000374.