Biorthogonal masked acylating agents for proximity-dependent RNA labeling

Shubhashree Pani*, Tian Qiu*, Kaitlin Kentala, Saara-Anne Azizi, Bryan C. Dickinson
Department of Chemistry, University of Chicago, Chicago, IL, 60637

Abstract

RNA subcellular localization is highly regulated, with local enrichment driving geography-dependent cell physiology. While proximity-based labeling technologies that use highly reactive radicals or carbenes provide a powerful method for unbiased mapping of protein organization within a cell, methods for unbiased RNA mapping are scarce and comparably less robust. In this work, we develop novel α-alkoxy thioenol and chlorenol esters, which function as potent acylating agents upon controlled ester unmasking. We pair these probes with subcellular-localized expression of a bioorthogonal esterase to establish a novel method for spatial analysis of RNA: Bioorthogonal Acylating agents for Proximity labeling and sequencing (BAP-seq). We demonstrate that by selectively unmasking the enol probe in a locale of interest, we can map RNA distribution in membrane-bound and membrane-less organelles. The controlled-release acylating agent chemistry and corresponding BAP-seq method expands the scope of proximity labeling technologies and provides a powerful approach to interrogate the cellular organization of RNAs.

The distribution of biomolecules in mammalian systems is tightly regulated, with compartmentalization in membrane-bound organelles as the most traditionally understood mechanism to dictate biomolecular interactions and localized regulation. Noncovalent interactions between molecules, such as protein-protein and protein-RNA interactions[1, 2], can also regulate the organization of molecules in a cell, in some cases through "membrane-less compartments" such as stress granules and the nucleolus. This is especially true for RNA; localization and local translation enable spatiotemporal regulation critical for processes such as embryonic patterning and cell motility and polarization. Regulatory non-coding RNAs (ncRNAs), which often serve as scaffolds, mediate interactions between other RNAs and proteins within sub-organellar compartments, such as the nucleolus within the nucleus[3, 4]. Thus, the location-based regulation of RNA function is critical to protein homeostasis and cellular physiology[5, 6]. Despite the criticality of RNA localization for function, there are a few existing methods for transcriptome-wide, unbiased mapping of RNAs in cells. While in some cases, organelles can be isolated and their RNA content characterized, such methods are time-consuming, prone to contamination, and do not exist for RNAs that are localized to membrane-free or sub-organellar compartments[7].

Proximity labeling technologies employ the localized expression of an engineered protein or the delivery of a small molecule catalyst to generate reactive molecules that covalently modify biomolecules within a local vicinity, enabling subsequent isolation and analysis[8, 9]. For example, APEX2 is an engineered peroxidase that produces biotin phenol radicals from biotin tyramide in the presence of hydrogen peroxide[10, 11], primarily labeling electron-rich residues such as tyrosine on nearby proteins. The APEX system has revolutionized spatial proteomics by providing maps of the mitochondrial[12], macromolecular complexes[13], the endoplasmic reticulum (ER[14]), and the synaptic cleft[15] proteomes, among others. Various other peroxidase-based proximity biotinylation methods have likewise been developed for proximity labeling[16, 17]. More recently, the µMap platform uses small molecule iridium photocatalysis to generate carbenes from diazirines for proximity-dependent protein labeling[18, 19]. Finally, acylation-based proximity labeling technologies for protein such as BioID[20, 21], and TurboID[22] generate biotinoyl-5′-AMP without the use of potentially confounding hydrogen peroxide or light.

In contrast to the robust and diverse technologies for mapping proteins, proximity labeling technologies for RNA remain nascent. APEX-RIP combines peroxidase-catalyzed protein biotinylation with RNA-protein chemical crosslinking to isolate RNA polymers[23]. The discovery that APEX-derived biotin phenol radicals could directly modify RNAs for direct proximity-based labeling resulted in the development of APEX-seq[24, 25] and the first high-resolution maps of RNA distribution in live cells. Since then, APEX-derived biotin phenol radicals have been tuned for

Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling   1

higher reactivity toward RNA[26]. Light-activated RNA-targeting proximity labeling technologies have also been developed, including CAP-seq[27] and Halo-seq[28], which use light-activated enzyme (miniSOG) or ligand (dibromofluorescein) activation to achieve reactive oxygen species (ROS)-mediated RNA proximity-based labeling. While these methods show promise, all rely on radical or ROS-reactivity – species that primordial RNA-like polymers likely evolved to minimize reactivity. Thus, current RNA labeling methods require increased reagent concentration and/or labeling time to overcome poor RNA reactivity. This raises the possibility of incomplete labeling, biased by substrate abundance or structure and reflecting stress, not homeostatic, subcellular RNA localizations. Interestingly, RNA is a well-known nucleophile through the reaction of its 2′ hydroxyl group with acylating agents and other electrophilic species[29, 30, 31, 32, 33, 34, 35]. However, the non-radical, acylation-based proximity reagents available – such as the anhydride-based TurboID biotin-5′-AMP molecules - are not known to be reactive enough to label RNA. Therefore, we set out to develop a proximity labeling platform that generates highly reactive and tunable acylating agents to efficiently label RNA in live cells.

In this study, we develop Bioorthogonal Acylating agents for Proximity labeling and sequencing (BAP-seq), a new proximity labeling method that combines a subcellularly-targeted esterase from *Bacillus subtilis* (BS2) with a new class of masked enol-ester probes to generate highly reactive acylating agents that label nearby RNAs in live cells. Using previously unreported chemistry, we synthesize both a panel of masked aryl thioester probes with a spectrum of electrophilicity and a panel of masked acid chloride probes with diverse functional handles. Following *in vitro* and *in cellulo* analyses, AC-2 emerged as the leading probe. AC-2 generates a highly reactive acid chloride acylating agent upon reaction with BS2, labeling nearby nucleophiles with an alkyne for subsequent imaging or enrichment via click chemistry. Confocal imaging reveals that the AC-2 probe labels subcellular locales in a BS2 proximity-dependent manner, both in membrane-bound (e.g., the nucleus) and membrane-less (e.g., the nucleolus) compartments. We then demonstrate RNA labeling by the BS2/AC-2 system *in vitro* and *in cellulo*. Finally, we adapt BS2/AC-2 labeling and enrichment with RNAseq to successfully map mitochondrial, nuclear, and nucleolar RNAs. These results establish BAP-seq as a valuable addition to the repertoire of proximity labeling approaches.

Results

Design and synthesis of biorthogonal enol-ester masked acylating reagents

In designing novel acylation-based proximity labeling agents, we sought to develop a strategy in which an exogenously supplied enzyme could convert an inert probe to a reactive acylating agent. We envisioned that a bioorthogonal ester-masked enol with an α-leaving group ("LG") would be unreactive toward biomolecules, but selective ester hydrolysis would generate an enol that would rapidly tautomerize to the reactive acylating agent (**Fig. 1**). For the ester mask, we used the methyl-cyclopropyl ("mCP") ester, which is inefficiently cleaved by endogenous esterases in human and murine cells[36], but can be deprotected quickly by exogenously expressed esterases, such as PLE[36] and BS2[37]. We therefore selected the BS2 esterase/mCP ester pair as the bioorthogonal, genetically encoded release agent in our proposed proximity-labeling platform.

We first synthesized thioester-based ("TE") probes, for which the resultant acylating agent's electrophilicity can be predicted based on the pKa of the thiolate leaving group[38]. This necessitated development of methods to synthesize α -alkoxy thioenol esters, as this functional group has never been observed in natural or synthetic molecules. Attempts to use various silver-based catalysts previously reported for trans-addition of carboxylic acids to ynol ethers[38] on thioynol ether substrates yielded more than 90% of the 1,2-addition product, with less than 5% of the desired 1,1-addition product (α-alkoxy thioenol esters), likely due to the reversed polarity of the triple bond caused by electronegativity difference between sulfur and oxygen atom[39, 40] (**Fig. S1**). After systematically screening catalysts, ligands, and solvent conditions on a model compound **1** (**Fig. S1-S2**), we discovered [Ir(COD)OMe]$_2$ could catalyze the addition of carboxylic acids to thioynol ethers to generate α -alkoxy thioenol esters with good selectivity (**Fig. 2a**). Preliminary investigation on the substrate scope reveals the necessity of an α-position heteroatom, possibly to serve as a σ-donor to coordinate with iridium. Attempts to synthesize a minimalist probe failed due to the unsuccessful conversion of alcohol **3** to the azide, probably due to propargyl-allenyl isomerization (**Fig. S3a-b**). We reasoned that an extra PEG linker would avoid the isomerization reaction and successfully installed the click handle via Mitsunobu azidation reaction[41] (**Fig. 2a**). Having established a robust synthetic route, we synthesized four thioester-based probes with a range of acylation strengths of unmasked product by tuning the leaving group thiolate: phenyl thiol (TE-1), difluorophenyl thiol (TE-2), tetrafluoro phenyl thiol (TE-3), and 4-nitrophenyl thiol (TE-4), with pK$_a$s of the thiolate leaving groups ranging from 3.8 to 6.6 (**Fig. S3c**) [38].

Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling   3

To confirm that the TE probes are substrates of BS2 and can generate the proposed reactive thioester, we assessed the least reactive probe (TE-1) *in vitro*, which has a stable enough thioester product to detect before hydrolysis. While we observed minimal degradation of TE-1 (1 mM) after 3 hours in aqueous solution (**Fig. S4a**), the addition of 385 nM BS2 esterase to 700 μM of TE-1 resulted in the complete consumption of TE-1 in under 5 min and the formation of a mixture of the predicted S-phenyl thioester (**6**), the hydrolyzed thiophenol, and the corresponding disulfide (**Figs. 2b, S3b-d**). This data corroborated the hypothesis that mCP-caged TE probes are substrates of BS2 and that their BS2-mediated deprotection generates S-aryl thioesters, known acylating agents[42].

After validating the proposed mCP-masked enol mechanism, we next aimed to generate probes that produce extremely reactive acylating agents with short half-lives due to competitive hydrolysis, resulting in molecules that can label RNA in a proximity-dependent manner. Since acid chlorides are among the most reactive acylating agents known, with half-lives in water not reported due to their high reactivity, we next targeted to synthesize masked acid chloride (AC) probes. We reasoned that the functional warhead α-chloroenol ester could be synthesized by a chromium (II) chloride assisted transformation from trichloromethylcarbinol (**Figs. 2c, S5**; [43]). We then used this synthetic strategy to introduce a range of different handles, including alkynes with various linkers (AC-1, AC-2, and AC-3), an azide (AC-4), or a fluorophore for direct imaging (AC-5) (**Fig. 2d**). Importantly, the (Z)-α-chloroenol ester is also stable in aqueous solution; we found no detectable decomposition of AC-5 in PBS after one hour incubation (**Fig. S6a-b**). However, the treatment of AC-5 with BS2 esterase in PBS resulted in rapid (<2 min) formation of the corresponding carboxylic acid, the expected hydrolysis product of acid chloride (**Fig. S6c-d**). To confirm formation of the reactive acid chloride, we incubated AC-5 with BS2 in the presence of an amine nucleophile (Fmoc-Lysine ester) to "trap" the acid chloride product and observed the formation of the predicted amide adduct (**Figs. 2e-f, S5e-f**). Given that both the TE and AC series of probes are stable in aqueous conditions and can rapidly generate acylating reagents by BS2 esterase *in vitro*, we next sought to assess the entire panel of probes for both bioorthogonality (low signal in the absence of BS2) and robust BS2-dependent proximity labeling of biomolecules in live mammalian cells.

Assessing proximity labeling capabilities of AC and TE probes in live cells
With the nine TE and AC probes in hand, we next sought to assess their biocompatibility and ability to label mammalian cells in a BS2-dependent manner. To do this, we first evaluated *in cellulo* stability and BS2-dependent cellular labeling of each probe using nuclear-localized BS2. After confirmation of nuclear BS2 expression in transiently transfected HEK293T cells via

Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling   4

immunostaining, we treated the cells with 25 µM of each probe for 10 min. The cells were then fixed, derivatized with Alexa488-azide/alkyne via click chemistry (except AC-5) and visualized by epi-fluorescent microscopy. TE-1 (**Fig. S7**) and TE-2 (**Fig. S8**), which generate the acylating agents with lowest reactivity, showed no BS2-dependent labeling, aligned with the low reactivity of their thioester products. In contrast, TE-3, TE-4, and all the AC probes showed clear BS2-dependent, nuclear-localized signal (**Figs. S9-15**). We next compared the relative intensity of nuclear signal of each probe to determine which probe generates the highest signal-to-noise (fluorescence signal in presence versus absence of BS2). We found that the AC probes label more efficiently than TE-probes (**Fig. 3a**), with AC-2 generating a 50-fold increase of signal in nuclear labeling with respect to DMSO (**Fig. 3b**) and an 8-fold signal-to-noise ratio in the presence of BS2 (**Fig. S12)**. Examining the signal localization from AC-2 revealed evident nuclear staining (**Fig. 3c**). Therefore, we selected AC-2 as our lead probe for further characterization.

To further validate the proximity-dependent labeling of the BS2/AC-2 pair, we used confocal microscopy to visualize the co-localization of BS2 expression with AC-2 signal across a series of cellular compartments. Cells expressing BS2 localized to the nucleus (via a nuclear localization signal, NLS; **Fig. S16**), the cytosol (via a nuclear export signal, NES; **Fig. S17**), endoplasmic reticulum (ER; via a cytosol-facing outer ER membrane localization tag; **Fig. S18**), mitochondria (via a mitochondrial matrix localizing tag; **Fig. S19**), nucleolus (via NIK3x fusion, **Fig. S20**), or nuclear pore (via SENP2 fusion, **Fig. S21**) were treated with 25 µM AC-2 for 10 min, fixed, derivatized with Alexa488-azide via click chemistry, and visualized by confocal microscopy. We detected robust BS2-dependent AC-2 labeling with high colocalization with BS2 expression across all compartments (**Fig. 4**), even in non-membrane-bound organelles such as the nucleolus and nuclear pore. This data confirmed the suitability of the BS2/AC-2 system for proximity labeling. Given the high BS2-dependent labeling observed in several compartments, we next sought to assess whether RNA acylation by BS2-unmasked AC-2 occurs *in vitro* and *in cellulo*.

RNA labeling by BS2/AC-2

To assess whether the BS2/AC-2 probe system can label RNA, we performed an *in vitro* labeling reaction and observed BS2-dependent labeling of an RNA oligo by AC-5, the fluorophore-containing AC probe (**Fig. 5a**). To further test RNA labeling *in cellulo*, we incubated cells expressing BS2 in the cytosol, nucleus, or mitochondria with our lead probe, AC-2 (25 µM), for 10 min. We then lysed the cells, extracted, and purified total cellular RNA, and performed click chemistry with biotin-azide to install biotin on modified RNAs. We assessed the degree of RNA labeling using a dot blot and observed robust BS2-dependent RNA labeling across all organelles (**Fig 5b**). RNase treatment eliminated this signal, while proteinase K treatment did not, confirming

Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling   5

the signals were indeed from RNA. Importantly, we detected no labeling in cells without BS2, indicating the low background and further confirming the bioorthogonality of AC-2. This data confirms that BS2/AC-2 pair labels RNAs in cells, and as such, we sought to analyze which RNAs are labeled across each compartment.

BAP-seq

We next performed proximity-dependent labeling of RNA by BS2/AC-2 in multiple compartments, including the cytosol, nucleus, mitochondrial matrix, and nucleolus, and analyzed enriched RNAs by high-throughput sequencing, which we named BAP-seq (**Fig. 6a**). In brief, we transiently expressed BS2 in each compartment of interest in HEK293T cells and then treated the cells with 50 μM AC-2 for 3 min. We then lysed the cells and isolated total RNA – our initial trial enrichment was performed with 25 μg of RNA per sample. Labeled RNA was isolated by clicking on a biotin tag and subsequent enrichment with streptavidin-magnetic beads. We detected a significant amount of enriched RNA that is both BS2 expression and AC-2 dependent (**Fig. S22a**), in line with the dot blots. A subsequent enrichment with 50 μg of input RNA followed by analysis by high-resolution automated electrophoresis revealed a strong de-enrichment of ribosomal RNAs (rRNAs) in the mitochondria and nucleus, and an enrichment of rRNAs in the nucleolus, as expected (**Fig. S23a-d**). Given these results. we next subjected the enriched RNAs to polyA capture and analysis by RNAseq.

Principal component analysis (PCA) of the RNA-seq experiments showed distinct populations for mitochondria, nucleus, and nucleolus, suggesting that different subsets of RNA were enriched for each (**Fig. S22b**). For these analyses, we compared RNA sequencing results between pairs of AC-2 treated biological samples expressing BS2 in separate compartments. Enrichment values were obtained from differentially expressed gene (DEG) analysis using DESeq2 and cutoffs of log_2fold change of 0.75 and p-value of 0.05, based on similar pipelines[24, 25]. Our analysis pipeline, which compared relative enrichment between samples expressing BS2 in different compartments, identifies transcripts that are differentially localized in one locale versus another. In a simple assessment, we compared enriched transcripts when BS2 was expressed in the mitochondrial matrix to when BS2 expressed in the cytoplasm. As expected, we observed several mitochondrial encoded RNAs – all 13 mt-mRNAs, both mt-rRNAs (MT-RNR1, MT-RNR2) and some mt-tRNAs (MT-TL1, MT-TE, MT-TA) – were enriched in these datasets (**Fig. 6b**). Using the common cut-off of $p < 0.05$ [24, 28], a total of 190 RNAs were enriched in the mitochondria. A more stringent cut-off of $p < 0.01$, identified 70 RNAs enriched in mitochondria, including all the mt-mRNAs and mt-rRNAs. We then compared the enrichment of mitochondrial-encoded transcripts when BS2 was expressed in the mitochondrial matrix versus other compartments and found that samples

Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling   6

expressing mitochondrial BS2 have more normalized counts for mt-mRNAs than any other samples. These data demonstrate that mt-RNAs are always enriched in mitochondria samples, regardless of which compartment is used for comparison (**Fig. S22c**), as expected, and is an indication of the robustness of BAP-seq. In addition to the known mitochondrial transcripts, we found lncRNAs like MALAT1 and NEAT1 enriched in our mitochondrial datasets relative to the cytosol (**Fig. S24**), which have recently been discovered to localize to mitochondria[44, 45]. Interestingly, we also found a ncRNA host gene, MIR34AHG, whose microRNA product, miR34a, and microprotein product, miPEP133, are known residents of mitochondria, but the ncRNA itself was not previously known to localize to mitochondria[46, 47] (**Fig. 6b)**.

Given BAP-seq's ability to faithfully enrich known mitochondrial RNAs, we next analyzed the samples expressing nuclear BS2. We expected, and found, an enrichment of intronic sequences in the nucleus datasets when compared across all compartments (**Fig. S22d**), despite performing polyA enrichment, which may not capture some intronic RNA species. Next, we compared RNAs enriched in our nuclear dataset to the cytosol, where we found 2,375 selectively enriched RNAs in the nucleus, including over 1,000 lncRNAs, with the conventional *p*-value cut-off of 0.05 (**Fig. 6c**). With the more stringent *p*-value cut-off of 0.01, we found 1,473 RNAs enriched in our dataset, including several known nuclear lncRNAs. While the majority of the identified lncRNAs have not been characterized, our dataset identified several well-studied nuclear lncRNAs like MALAT1, XIST, and NEAT1, and also identified other lesser-studied lncRNAs that are known to reside in the nucleus, such as INE1, MIAT, TSIX, GAS5, PURPL, RAD51-AS1, MIR503HG and LINC00839 (**Fig. S26a**).

Finally, we compared transcripts enriched when BS2 is expressed in the nucleolus to BS2 expressed in the nucleus overall - a very stringent comparison as both BS2 proteins are expressed in the same membrane-bound compartment (the nucleus), but in different sub-compartments (nucleoplasm vs. nucleolus). Given the highly differentiated staining obtained following nucleolar BS2 expression compared to nuclear expression (**Fig. 3**), we reasoned that this comparison was feasible. Comparing the transcripts enriched in the nucleolus versus the nucleus, we found enrichment of several snoRNAs known to reside in the nucleolus, such as SNORA22, SNORD3B-2, SNORA14B[48], and corresponding de-enrichment of known nucleoplasm-residing lncRNAs, such as XIST, NEAT1 and MIAT (**Fig. 6d**; [49, 50]). We also found enrichment of the validated nucleolar-residing mRNA, IGF2[51], in our dataset. Taken together, these data confirm that BAP-seq can identify transcripts both selectively enriched between compartments (i.e., mitochondria vs. cytoplasm vs. nucleus), as well as between subcompartments with a single membrane-bound region of the cell (nucleolus vs. nucleus).

DISCUSSION

In this work, we developed "masked" enol-based probes, which release highly reactive acylating agents upon selective bioorthogonal ester unmasking by a delivered BS2 esterase, to create a new proximity labeling platform. After developing methods for the synthesis of a panel of masked aryl thioester and acid chloride-based probes we identified a lead probe, AC-2, which generates a highly active acid chloride upon BS2 unmasking and features a versatile alkyne for subsequent click chemistry with imaging or affinity reagents. We found the BS2/AC-2 system labels biomolecules in both membrane-bound and membrane-less subcellular compartments with high resolution and labels RNAs in a proximity-dependent manner. Finally, we used the BS2/AC-2 system to develop BAP-seq for RNA proximity analysis, which we used to map RNAs in the mitochondria, nucleus, and nucleolus, representing a highly differentiated new method to the repertoire of existing proximity-labeling platforms.

The BS2/AC-2 system and corresponding BAP-seq method offers several unique advantages and opportunities compared to existing methods. While we utilized a suite of thioesters and acid chloride-based acylating agents in this study, the concept presented here should, in principle, be amenable to a wide range of acylating reagents, opening opportunities to tune chemistries for acylation-based proximity labeling. The BS2/AC-2 pair and resultant BAP-seq does not require light or oxidants, which provides a potentially less perturbative method compared to existing RNA proximity labeling approaches and significantly simplifies the experimental workflow. The high temporal resolution of our method (3 min) is also likely less perturbative than other methods (30 min high concentration biotin-tyramide incubation + 1 min H_2O_2 for APEX-seq; 15 and 20 min for Halo and Cap-seq respectively) and would potentially allow for detection of short-lived RNAs[61]. Finally, our method could be adapted to uncage natural, bioactive electrophiles, a critical class of spatially regulated signaling molecules[52].

Intriguingly, based on our imaging data, we saw excellent µm-scale staining membrane-bound compartments, such as the nucleus, and large membrane-free compartments, such as the nucleolus (100's of nm) and nuclear pores (~100 nm). These data indicate a tight labeling radius around BS2 expression, which is expected based on the high reactivity of an acid chloride in aqueous biological conditions. While BAP-seq offers good subcellular resolution, it is limited by the localization specificity of the expressed BS2. As overexpression of proteins risks mislocalization, BS2 expression and proper subcellular localization requires empirical tuning and validation for any given experiment. However, the fact that BAP-seq works even with transient transfection adds more flexibility to future applications. Notably, the high activity level of BS2, as

Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling 8

a consequence of fast enzyme kinetics[37], bolsters the efficiency of the BS2/AC-2 system. Indeed, even in cells expressing very little BS2 – a common consequence due to transfection heterogeneity – we still observed robust AC-2 signal. This indicates that probe concentration, rather than enzyme (BS2) concentration, limits labeling, which is an advantage in such a system.

Aside from the technological advances in this work, we also generated new datasets identifying RNA species that are present in various locales using BAP-seq. The datasets yielded plenty of known targets, validating the method, as well as many new ones that warrant further investigation. When comparing our nuclear-enriched dataset with that of Apex-seq, we found that out of our top 25 enriched RNAs with the lowest p-value, 23 overlap with the Apex-dataset, and out of the top 50, 37 are present in the Apex-seq enriched nuclear datasets (**Fig. S25**). We listed the top 100 lncRNAs enriched in the nuclear dataset over cytosol, which consists of known nuclear lncRNAs like MALAT1, XIST, NEAT1, and several understudied lncRNAs with no validation of subcellular localization in HEK293T cells **(Fig. S26b)**. In the nucleolus, even when using a very stringent cutoff of $\log_2$fold change greater than 1.5 and adjusted p-value of 0.01, we identified 459 RNAs enriched in the nucleolus versus the nucleus, which includes IGF2. Among those, we also identified transcripts not previously known to localize to the nucleolus, such as HELLPAR, a lncRNA associated with HELLP syndrome[53], which warrants further investigation. Collectively, these observations indicate that BAP-seq provides a robust hypothesis generation pipeline across multiple cellular locales.

Since its debut in 2004, the biotin ligase-generated biotin-AMP anhydride system has been developed into the most widely used non-radical proximity-dependent protein labeling platform. However, despite significant advances in radical and carbene systems, non-radical based labeling chemistries have been considerably less developed. The synthetic strategies shown here introduce a new class of tunable, enzyme-activable, acylating agents that can be applied for proximity labeling and further applications. As acylation chemistry is prevalent in cellular processes involving all biomolecules, the approaches outlined here to selectively unmask acylating agents open new opportunities to study and control these biomolecules.
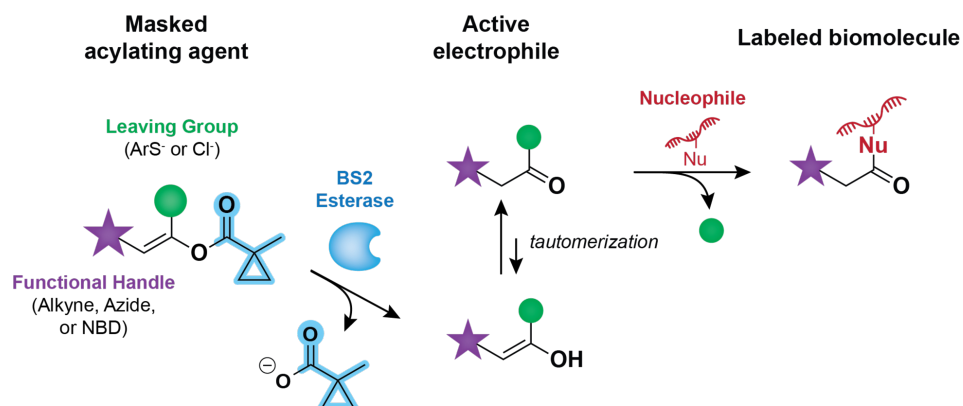
FIGURES

Fig 1: Conceptual framework for a biomolecular labeling strategy using masked acylation reagents.

A masked acylation reagent is unreactive towards biomolecules. Selective unmasking by BS2 esterase through cleavage of a bioorthogonal methyl cyclopropyl ester cage releases an enol, which rapidly tautomerizes to a carbonyl, acylating reagent. Nucleophilic biomolecules react with the acylating agent, covalently modifying them, which installs a visualization or affinity handle onto the target biomolecule.
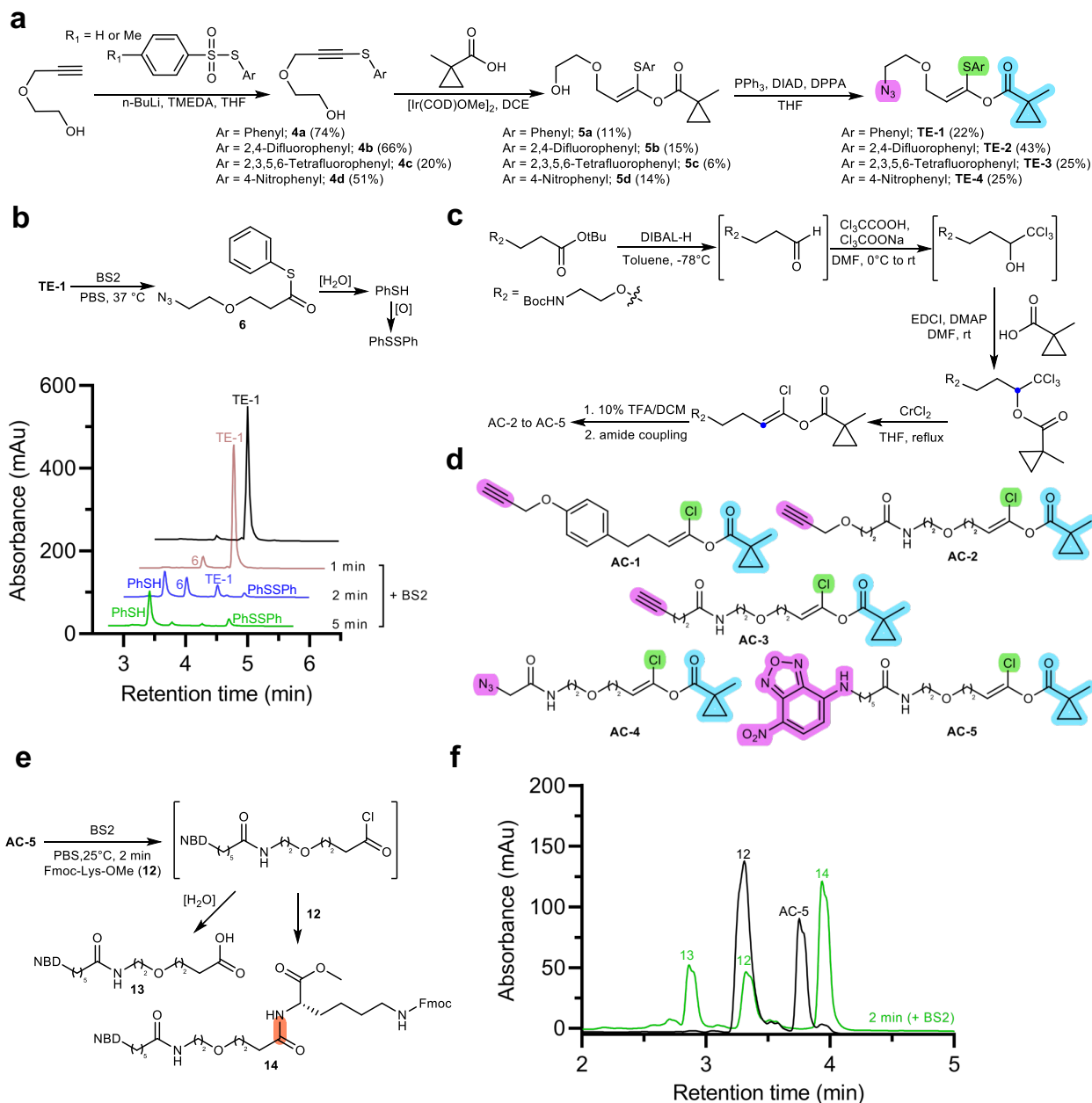
Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling   10

Fig 2: Synthetic design and validation of masked acylating probes.

a, Key synthetic steps for the synthesis of thioester-based (TE) probes. b, Unmasking reaction of TE-1 by BS2 yields the predicted thioester product, confirmed by LC-MS. c, Synthetic scheme for acid chloride-based (AC) probes. d, Structures of all AC probes. Leaving groups are highlighted in green, the bioorthogonal BS2 recognition site is highlighted in blue, and the functional handle is highlighted in purple. e, Unmasking reaction of AC-5 by BS2 in the presence of a nucleophilic substrate (Fmoc-Lys-OMe) yields the predicted amide product. The newly formed bond is highlighted in red. f, LC-MS trace of the reaction shown in e.
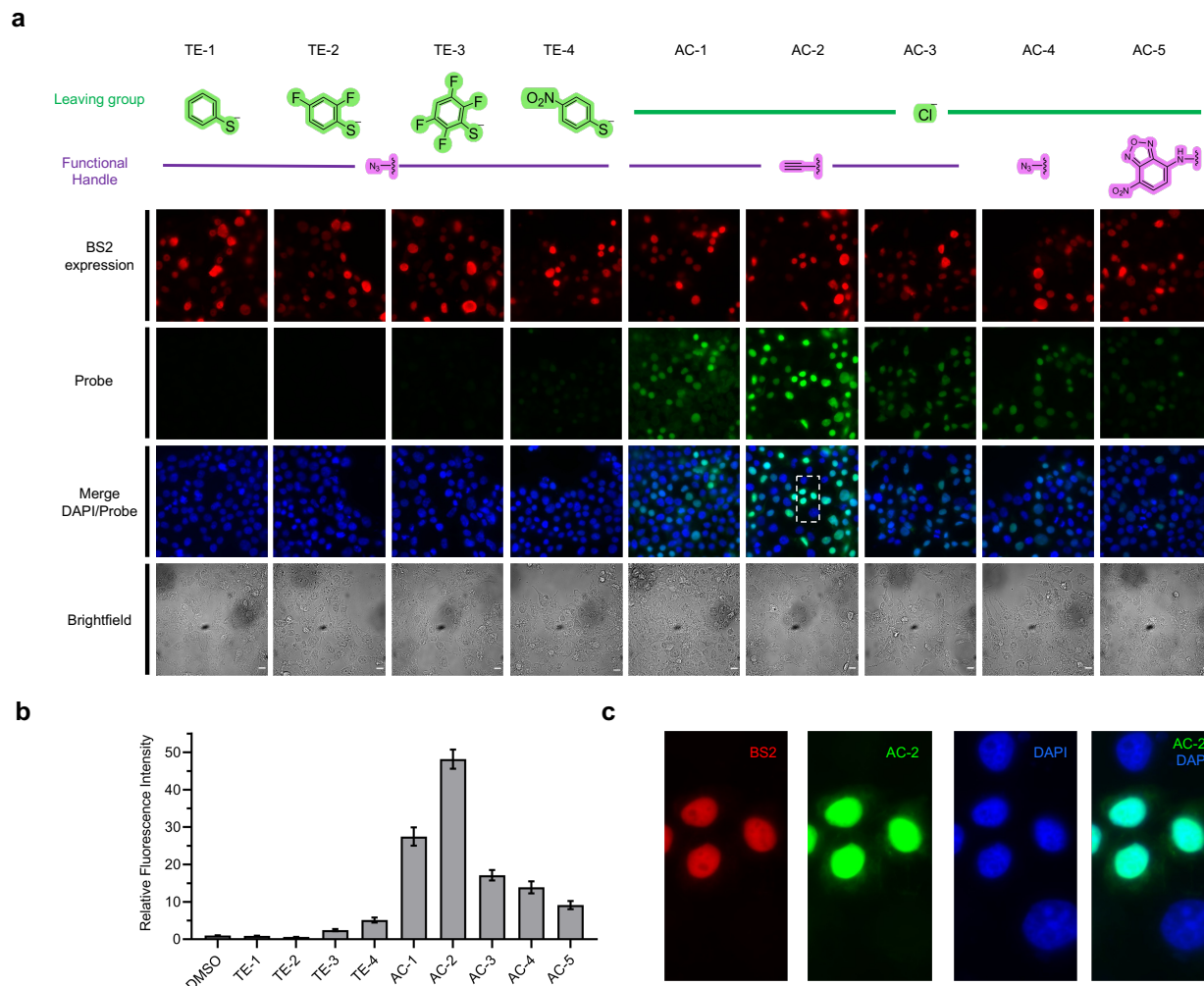
Fig 3: TE and AC probes label sites proximal to nuclear localized BS2. *Comparison of nuclear proximity labeling in live cells across all AC and TE probes.* **a,** *Representative fluorescence images of HEK293T cells transiently transfected with a vector expressing a NLS-BS2 esterase construct then treated with 25 μM of the indicated probes for 10 min. Probe localization, corresponding to labeled sites, is shown in green (via click-chemistry installation of AF-488), BS2 is shown in red (via IHC), and nuclear DAPI staining is shown in blue. Scale bar = 10 μm.* **b,** *Quantification of probe labeling from images in a. Data are shown as the mean ± SD from n=4 images per probe.* **c,** *Detail of inset outlined for AC-2-treated cells in a.*
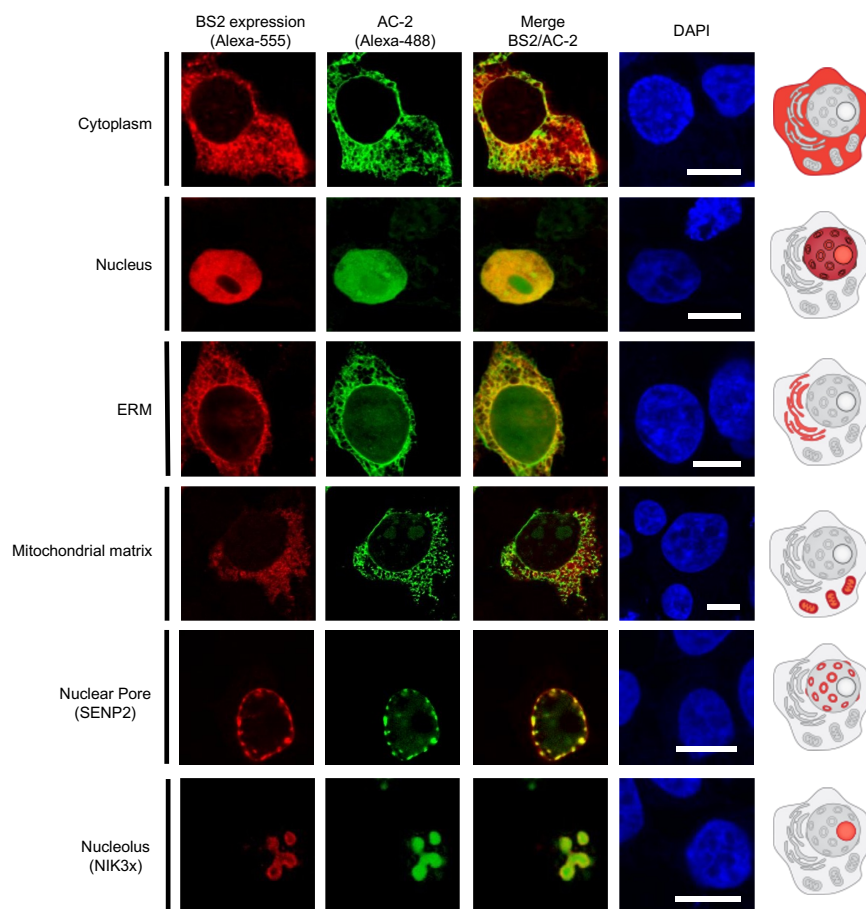
Fig 4: AC-2 labeling is restricted to the vicinity of BS2 expression across multiple compartments.
Representative confocal images of HEK293T cells expressing BS2 in the cytosol, nucleus, ER membrane (ERM), mitochondrial matrix, nuclear pore, or nucleolus. All cells were treated with 25 μM of AC-2 for 10 min. Probe localization, corresponding to labeled sites, is shown in green (via click-chemistry installation of AF-488), BS2 is shown in red (via IHC), and nuclear DAPI staining is shown in blue. Scale bar = 10 μm.
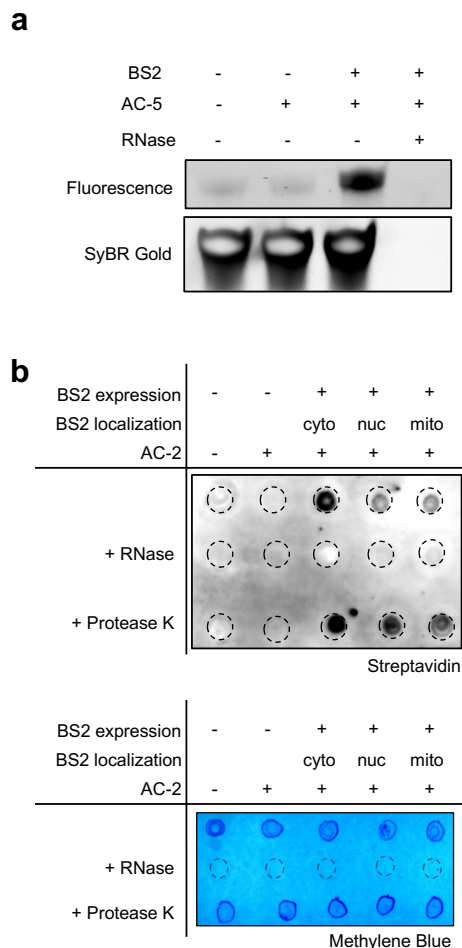
Fig 5: AC-2 labels RNA in vitro and in cellulo. a, *In vitro labeling of 2.5 µM of a 74-mer RNA by 500 µM AC-1 and 600 nM BS2 in 5 min.* **b,** *Dot blot showing BS2-dependent RNA labeling by AC-2 when BS2 is expressed in cytosol, nucleus and mitochondria. HEK293T cells were transfected with a control vector, or a vector expressing BS2 in the cytoplasm, nucleus, or mitochondria, then treated with 25 µM AC-2 for 10 min, followed by lysis, RNA isolation, and click chemistry to install biotin on labeled RNAs. RNA labeling was analyzed via dot blot with 500 ng of isolated RNA using a streptavidin antibody. Methylene Blue blot shows even loading of RNA across all samples. For RNAse- or protease-treated samples, RNA was treated with 85 µg/mL of RNase A or 85 µg/mL of Proteinase K for 15 min, purified, and then blotted onto the nylon membrane.*
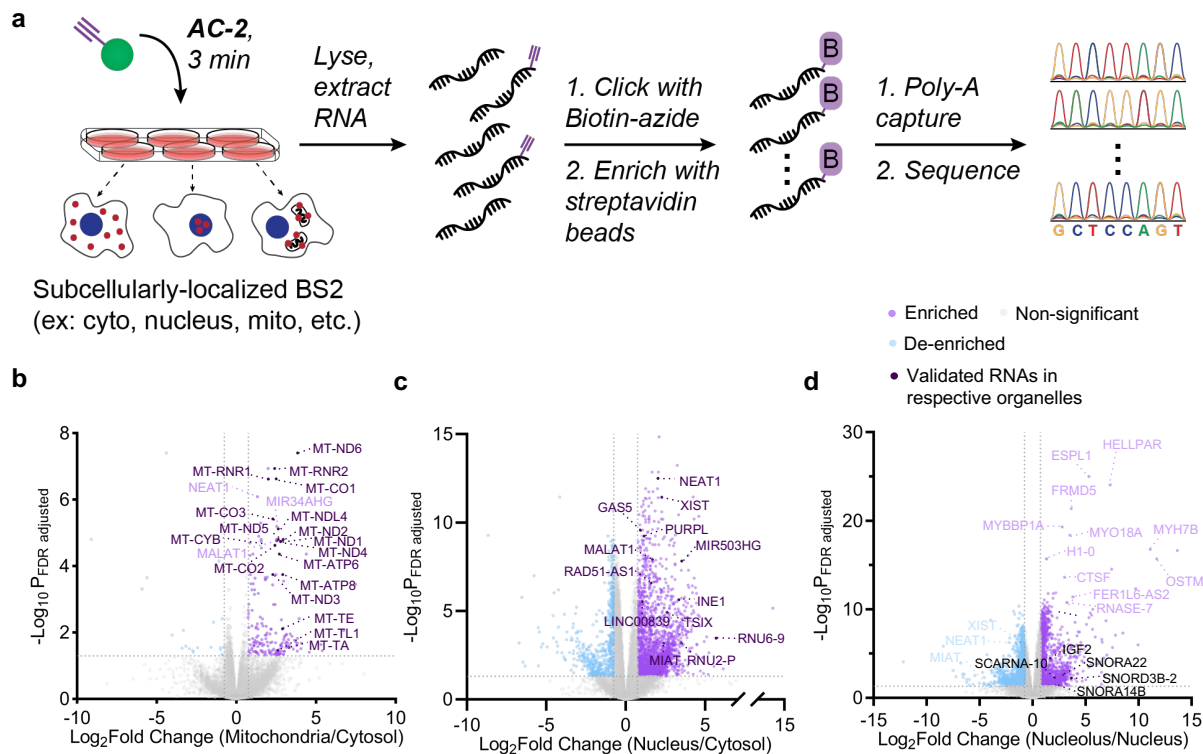
Fig 6: BS2-dependent proximity labeling of RNA using AC-2 paired with quantitative sequencing provides an unbiased spatial transcriptomic map. a, Schematic workflow of Bioorthogonal Acylating agents for Proximity labeling and sequencing (BAP-seq). **b,** Volcano plot depicting enrichment of mitochondrial transcripts via BAP-seq in cells expressing mitochondrial BS2, as compared with cells expressing BS2 in the cytoplasm with cut-offs of log_2(fold change) > 0.75 and adjusted p-value < 0.05. **c,** Volcano plot depicting enrichment of mitochondrial transcripts via BAP-seq in cells expressing nuclear BS2, as compared with cells expressing BS2 in the cytoplasm with cut-offs of log_2(fold change) > 0.75 and adjusted p-value < 0.05. **d,** Volcano plot depicting enrichment of mitochondrial transcripts via BAP-seq in cells expressing BS2 in the nucleolus, as compared with the nucleus with cut-offs of log_2(fold change) > 0.75 and adjusted p-value < 0.05.

Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling 1

METHODS

General Methods. For chemical synthesis, reagents and dry solvents were purchased from commercial sources (Sigma Aldrich, Combi-blocks, Thermo Fisher) and used without further purification. Silica gel P60 (SiliCycle, 40–63 µM, 230–400 mesh) was used for column chromatography. Analytical thin-layer chromatography was performed using SiliCycle 60 F254 silica gel (pre-coated sheets, 0.25 mm thick) with detection at 214 nm. Low-resolution-mass spectral analysis and liquid chromatography analysis were carried out on an Advion Expression-L mass spectrometer (Ithaca, NY) with electron spray ionization (ESI) in the positive mode coupled to an Agilent 1220 Infinity LC System with an Agilent Poroshell 120 column (SantaClara, CA). Automated flash column chromatography purification was carried out on a Biotage system Isolera One using SNAP Biotage columns. NMR spectra were recorded on the BRUKER Ascend 400 at the Department of Chemistry NMR Facility, University of Chicago, for ^{1}H-400 MHz and ^{13}C-101 MHz measurements. Chemical shifts are given in parts per million (δ) referenced to TMS (δ = 0.00 ppm ^{1}H-, ^{13}C-NMR). Coupling constants are given in Hertz. High-resolution mass spectra measurements were performed on an Agilent 6224 TOF. Using a combination of atmospheric pressure chemical ionization and electrospray ionization at the Department of Chemistry Mass Spectrometry Facility, University of Chicago.

Synthesis Methods. The detailed synthetic methods for all the probes are provided in the Supporting Information.

Cloning. All the plasmids were constructed by Gibson Assembly with PCR products generated using Q5 or Phusion DNA polymerases (NEB). The plasmids were sequenced by the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping facility. Full vector sequences are included in Fig. S25.

Mammalian cell culture and plasmid transfection. HEK293T (ATCC) cells were cultured in DMEM (GlutaMAX™, high glucose, sodium pyruvate, phenol red; Thermo Fisher) supplemented with 10% (vol/vol) fetal bovine serum (FBS, Gemini Benchmark) and 1% (vol/vol) penicillin/streptomycin (Gibco/Life Technologies). Cells were maintained in a water-saturated, 5% CO_2-containing, 37°C incubator. Cells used for experiments never exceeded passage number 25. Fresh HEK293T cells were obtained from ATCC and frozen at an early passage number (5) in individual aliquots. There was no testing for mycoplasma infection as a result. Transient

Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling 2

transfections were performed using Lipofectamine 3000 (Invitrogen; Thermo, L3000015) following the manufacturer's protocol.

Recombinant protein expression and purification. The pET-BS2 plasmid (Fig. S25) was transformed into chemically-competent E. coli BL21 cells. 500 mL cultures were grown at 37°C in LB broth (40 µg/mL Kan) in a 2 L flask to mid-log-phase (OD600 of 0.65). The cultures were induced with 0.2 mM IPTG and incubated at 30°C overnight. Cells were harvested by 4°C centrifugation (4000xg for 15 min), the supernatants were discarded, and the pellets were resuspended in 30 mL of lysis buffer (50 mM Tris (pH 7.5), 1 M NaCl, 10 mM TCEP, 20% glycerol) and protease inhibitor tablet (Thermo, A32963). The mixtures were then sonicated (total on time: 10 min; 10s on/20s off; amplitude 30%) and spun down at 12,000 rpm for 40 min. BS2 was purified from the supernatants by nickel affinity chromatography. After discarding the cell debris, the supernatant was incubated with 1 mL Takara His60 Ni Superflow Resin with gentle rotation at 4 °C for 1 h. The His-tagged proteins were purified using a standard protocol of washing and eluting the resin with wash buffer-1 (20 mM imidazole, 50 mM Tris, 1 M NaCl, 20% glycerol, 10 mM TCEP, pH 7.5), wash buffer-2 (40 mM imidazole, 50 mM Tris, 1 M NaCl, 20% glycerol, 10 mM TCEP, pH 7.5) and elution buffer (300 mM imidazole, 50 mM Tris, 1 M NaCl, 20% glycerol, 10 mM TCEP, pH 7.5). The purified proteins were then desalted on GE Disposable PD-10 desalting columns and stored in the protein storage buffer (50 mM Tris, 100 mM NaCl, 2 mM TCEP, 50% glycerol, pH 7.5). Final BS2 concentrations were measured using a standard BCA assay.

LC/MS detection of TE-1 stability and BS2 activity in PBS. 0.7 mM of TE-1 was incubated in 500µL PBS containing 15% DMSO by volume at room temperature in the presence of a reference compound-1,3,5-trimethoxybenxzene (Sigma Aldrich, 621-23-8). At regular intervals (1 hour), 20 µL of the solution was removed, diluted with 20 µL of methanol, and assayed by LC/MS. The absorbance maxima, normalized to that of the reference peak, were plotted as a function of time. To test the activity of BS2 on TE-1, 0.85 mM TE-1 was dissolved in 350 µL PBS containing 15% DMSO by volume. 30 µL of the solution was taken out at this stage, diluted with an equal volume (30 µL) of methanol, and assayed by LC/MS. BS2 stock solution was added to the TE-1 solution in PBS to achieve 385 nM final BS2 concentration and incubated at 37 °C with shaking at 900 rpm. After 1, 2, and 5 minutes, 30 uL of the solution was removed, quenched with an equal volume of methanol, and assayed by LC/MS.

LC/MS detection of AC-5 stability, BS2 activity, and trapping of acid chloride with Fmoc-Lys in PBS. 0.5 mM AC-5 was dissolved in 500 µL PBS containing 30% DMSO by volume at room temperature. After 0, 30, and 60 min, 50 µL aliquots of the solution was removed and assayed by LC/MS. To test BS2 activity on AC-5, 0.5 mM AC-5 was dissolved in 100 µL of PBS containing 30% DMSO by volume. After recording LC/MS spectra for the no BS2 control, 1 µL of BS2 stock solution was added to the solution at room temperature to achieve 500 nM final BS2 concentration, which was then assayed by LC/MS immediately. To trap the active acylating agent, 0.5 mM of Fmoc-L-Lysine-methyl ester and 0.5 mM AC-5 was dissolved in PBS containing 30% DMSO by volume and assayed by LC/MS before and after adding BS2 at a final concentration of 500 nM.

Fluorescence microscopy. For epifluorescence microscopy, an inverted epifluorescence microscope (Lieca DMi8) equipped with a Hamamatsu Orca-Flash 4.0 camera, a ×63 oil objective (numerical aperture 1.4) and a 300 W Xenon light source (Sutter Lamda XL) was used. Leica LASX software was used to obtain images for Alexa 488 (490/20x, ET Quad-S, ET 525/36m), Hoechst 33342 (ET 402/15×, Quad-S, ET 455/50 m), Alexa 555 (ET 555/25x, Quad-S, ET 605/52 m), and bright field. Acquisition time ranged from 30 to 800 ms. The acquisition time and Fluorescence Intensity Manager (FIM) was kept constant for a given channel across all samples in an experiment that are compared. For confocal microscopy, the slides were imaged on a Leica Stellaris 8 Laser Scanning Confocal microscope (DMI8-CS) equipped with near UV and white light laser (WLL). DAPI (HyD S1 detector, 405 nm laser excitation, 430/500 emission), Alexa 488 (HyD X2, 499 nm laser excitation, 510/560 nm emission), Alexa 555 (HyD S3 detector, WLL 80mHz, 558 nm laser excitation, 590/650 nm emission), BrightField (Trans PMT, WLL 80mHz) images were taken on 63x/1.4 UV oil objective with 3x magnification and 8-line averaging.

HEK293T cells were plated without antibiotics on glass coverslips (Neuvitro Corporation) pre-treated with 0.1 mg/mL Poly-D-Lysine (for at least 2 h at rt) and contained within a 24-well dish. 18-24 hrs post-plating, the cells were transfected with 500 ng of the respective BS2 plasmids or no-BS2 containing dummy plasmid at 60% confluency. 24 h post-transfection, 25 µM of the corresponding probe or DMSO control in pre-warmed DMEM was added for 10 min at 37 °C. The cells were then incubated with 1 mM PMSF in DPBS for a minute to stop labeling by quenching BS2 followed by 2x wash with DPBS. The cells were then fixed with 4% paraformaldehyde (Electron Microscopy Science, 15710) solution for 15 min, washed twice with PBS, permeabilized with ice cold MeOH for 8 min at -20°C, and washed twice with PBS. The samples were then incubated in a CuAAC reaction solution in PBS with the following final concentrations 2 mM

Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling 4

BTTAA (Click Chemistry Tools, 1236-100), 1 mM CuSO$_4$, 10 µM of 488-PEG4-alkyne (Sigma, 761621) for azide-handle probes or AzDye 488 azide (Click Chemistry Tools, 1275-5) for alkyne handle probes, and 10 mM sodium ascorbate (prepared fresh). The BTTAA, CuSO$_4$, and 488 alkyne/azide were pre-mixed, and the sodium ascorbate was added at last, immediately prior to cell bathing. The samples were incubated in the CuAAC solution for 1 h at room temperature, followed by three washes with PBS for 5 min each. The samples were then blocked with 3% BSA in PBS for 1 hr at room temperature, followed by overnight incubation with 1:1000 diluted Anti-V5 antibody (Invitrogen, R960-25) in 3% BSA in PBS at 4°C. The following day, after three washes with PBS, 1:1000 diluted secondary antibody-anti mouse Alexa 555 (Invitrogen, A21427) in 3% BSA in PBS was added for 1 hr at room temperature. The samples were washed twice with PBS, and incubated with 1 µM of DAPI in PBS for 5 minutes, followed by 2x more washes with PBS. After removing the last PBS wash, the cell-containing coverslips were then removed from the 24-well plate, inverted, and transferred to fixative pre-dotted on a glass slide. The slides were dried for at least 5 h in the dark prior to imaging. The samples were imaged on the inverted epifluorescence microscope or Leica Stellaris 8 Laser Scanning Confocal microscope.

Image analysis. Measurements of four different images of the same slide were taken for each condition. For Fig 3b, signal intensity for the probe channel (AF488) was measured with thresholding from the BS2 channel (AF555) to count the signal from transfected cells only. All signals were normalized to that of the nucleus of non-transfected DMSO-treated cells by DAPI thresholding. To compare the signal-to-noise ratio for each probe in Fig. S6-S14, the signal intensity for the probe channel (AF488) was measured with thresholding from the DAPI channel to select the nuclear signal. This signal was normalized to the signal intensity of the probe/AF488 channel with DAPI thresholding of non-transfected samples treated with the respective probes.

In vitro **RNA labeling.** 3 µg of a 74-mer RNA oligo (a gift from Szostak lab) (ucugugugggguaggugggugcugggguaauggggacugcauugcauaucccuaggcaccuauugagauuucucug) in PBS was boiled at 92°C for a minute and immediately cooled to 4°C for 5 min. Purified BS2 and AC-5 were added to the folded RNA solution to get a final concentration of 600 nM BS2 and 500 µM AC-5 in a reaction volume of 50 µL in PBS. The reaction mixture was incubated at 37 °C at 900 rpm for 5 min. The labeled RNA was then purified using RNA Clean and Concentrator Kit (Zymo, RCC-5) and eluted in 10 µL of water. For RNase treatment, purified, labeled RNA was incubated with 85 µg/mL of RNase A at 37 °C for 15 min, re-purified using the RCC-5 kit, and eluted in 10 µL of water. The entire RNA product for all the treatment conditions was loaded onto

a 10% Urea-PAGE gel along with 2X RNA loading dye (NEB) and run at 80V for 2 hr. AC-5 labeling was observed by fluorescence imaging using the Amersham Typhoon Laser Scanner in the Cy2 channel. After fluorescence imaging, the gel was incubated in SyBr Gold for 15 min and imaged under UV for RNA loading control.

In cellulo **AC-2 RNA labeling and detection by streptavidin dot-blot.** HEK293T cells were plated in a 12-well plate without antibiotics and transfected with 800 ng of cyto-BS2, NLS-BS2, mito-BS2, or no BS2-dummy plasmid at 60% confluency. 24 hr after transfection, all the wells were treated with 25 μM of AC-2 or DMSO in pre-warmed DMEM for 10 min at 37°C. The labeling reaction was quenched by treating the cells with 1mM PMSF in DPBS for 1 min, followed by two rounds of DPBS wash. The cells were lysed immediately, and RNA was extracted using the QIAGEN kit. Eluted RNA was then clicked with biotin at 25°C for 30 min by adding 5x Click mixture to get the final concentration of 0.4 mM $CuSO_4$, 2 mM Biotin picolyl azide (Sigma-Aldrich, 900912), 1 mM THPTA (Click Chemistry Tools, 760952-88-3) and 10mM sodium ascorbate along with 1 mM SuperAs.In RNase inhibitor (Thermo Fisher, AM2694) in PBS. The click reaction was quenched with 10 mM EDTA, and RNA was purified using the Zymo RCC-25 kit. For every sample, 1.5 μg of biotinylated RNA was then treated with each condition-DMSO, 85 μg/mL of RNase A, and 85 μg/mL of Proteinase K at 25°C and 900 rpm for 30 min followed by Zymo RCC-5 clean up. RNA was eluted in 12 μL of water, split equally, and blotted in two Hybond membranes (Cytiva, RPN1210B) presoaked in 2x SSC buffer and air-dried. The blotted RNA on the membrane was allowed to dry at room temperature for 15 min before UV cross-linking at 2400 mJ/cm^2 (254 nm wavelength, UV Stratalinker 2400). Both the membranes were washed twice TBST, and one of the blots was used for methylene blue staining for 5 min, washed several times with water, and imaged for RNA loading control. The other blot was blocked with 3% BSA solution in TBST at room temperature for 1 hr, followed by overnight incubation with 1:2000 Streptavidin-HRP antibody (Sigma Aldrich, S2438) in 3% BSA solution in TBST. The next day, the blot was washed thrice with TBST and imaged for chemiluminescence using SuperSignal West Pico PLUS Chemiluminescent Substrate (Fisher Scientific, 34577) in Azure c300 imager.

Sample preparation for BAP-seq. HEK293T cells transfected with cyto-BS2, nuclear-BS2, mitochondrial-BS2 NIK3x-BS2 (nucleolus targeting), or no BS2-dummy plasmid with 2-3 biological replicates at 60% confluency. 24hrs post transfection, cells were treated with 50 μM of AC-2 probe in pre-warmed DMEM for 3 min at 37°C. The labeling reaction was quenched by incubating with 1 mM PMSF in DPBS for 1 min, followed by a quick DPBS wash. The cells were

Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling   6

then lysed, and RNA was extracted using the Qiagen kit. For each replicate of a BS2 construct, 50 µg of RNA (2 wells of a 6-well plate) was used for the next steps. The labeled RNA was clicked with biotin using the click reaction cocktail of 0.4 mM CuSO₄, 2 mM Biotin picolyl azide, 1 mM THPTA, and 10mM sodium ascorbate along with 1mM SUPERase-in RNase inhibitor in PBS. The click reaction was quenched with 10mM EDTA, and biotinylated RNA were purified using the Zymo RNA Clean and Concentrator Kit-25. The biotinylated RNAs were enriched using APEX-seq enrichment method [24], with some modification. Briefly, 50 µL Pierce Streptavidin Magnetic Beads (Thermo, 88816) were utilized per 50 µg RNA sample. Beads were washed three times on/off the magnet with 500 µL binding & wash buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.1% Tween-20) per sample, followed by two 500 µL washes with solution A (100 mM NaOH, 50 mM NaCl) per sample, one 500 µL wash with solution B (100 mM NaCl), and a final resuspension of the beads in 250 µL solution B containing 1 µL SUPERase-in per sample. The resuspended beads were mixed thoroughly with the 250µL of eluted, biotin-labeled RNA from above and incubated at 4°C for 2 h on a rotator. After this incubation, samples were washed five times with binding & wash buffer as described above, and enriched RNAs were digested off with Trizol [45]. For elution of enriched RNAs off the beads, they were first resuspended in 100µL of PBS, followed by the addition of 300µL of trizol (Invitrogen, 16096040) and were incubated at 37°C for 10 minutes to dissociate the RNAs. The beads were placed on the magnet and dissociated RNAs in the supernatant were pipetted out. Finally, the enriched RNAs were recovered using the Zymo Direct-zol RNA Microprep kit.

PolyA enrichment and sequencing. The enriched RNAs for all the samples were subjected to polyA capture. Subsequently, the DNA library with dual-index adapters was generated and amplified using the mRNA Hyper Prep Kit (Kapa Biosystems, KK8581) following manufacturer's protocol. After amplification, the quality check of the library was performed with tapestation analysis. Libraries were sequenced on Next-seq 550 using a High Output kit at ~13 million single 75-cycle reads per library.

Analysis of BAP-seq data. The RNA-seq library generated was mapped to the Homo sapiens GRCh38 reference genome using STAR[54], version 2.7.9. aligned reads were quantified using Salmon (http://combine-lab.github.io/salmon/), version 1.4.0. After read mapping with Salmon, (https://bioconductor.org/packages/release/bioc/html/tximport.html) was used to read Salmon outputs into the R environment. Annotation data from Gencode V34 was used to summarize data from transcript-level to gene-level. Filtering was performed to remove particular gene biotypes or

Biorthogonal masked acylating agents for proximity-dependent biomolecule labeling 7

genes with low expression. We first selected gene biotypes of interest: protein coding, lncRNA, Mt_rRNA, Mt_tRNA, rRNA, snRNA, scRNA, and snoRNA. This reduced the number of genes from 60728 to 39823. Genes were then filtered for low expression, keeping only genes with greater than 1 count per million (CPM). This reduced the number of genes from 39823 to 24878. To identify differentially expressed genes, precision weights were applied to TMM-normalized gene counts based on within-group sample-level variance and gene-level mean-variance trends using VOOM[55]. The count data was fitted to a gene-wise linear model with group status as a coefficient, and contrasts were specified for comparisons of interest. The Limma [56] empirical Bayes method was used to estimate the posterior odds of differential expression ($|logFC| > 0$) after adjusting for gene-level posterior residual standard deviations. Significant differentially expressed genes were decided with $|log_2$fold change$| > 0.75$ and false-discovery rate (FDR) of 0.05. Results of differential expression testing are visualized by volcano plots comparing samples with BS2 expression in two different locales (cytoplasm versus nucleus, cytoplasm versus mitochondria, and nucleus versus nucleolus).

Data Availability

All BAP-seq data are available in the Sequence Read Archive through accession number GSE229451. All other data generated or analyzed in this study are available within the main text and Supplementary Information. Source data are provided in this paper. Links to vector maps are included in **Fig. S27,** and key physical vectors will be deposited to Addgene.

REFERENCES

1. Keskin O, Tuncbag N, Gursoy A. Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. *Chem Rev* 2016, **116**(8)**:** 4884-4909.

2. Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* 2018, **19**(5)**:** 327-341.

3. Engreitz JM, Ollikainen N, Guttman M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol* 2016, **17**(12)**:** 756-770.

4. Rinn J, Guttman M. RNA Function. RNA and dynamic nuclear organization. *Science* 2014, **345**(6202)**:** 1240-1241.

5. Tiruchinapalli DM, Oleynikov Y, Kelic S, Shenoy SM, Hartley A, Stanton PK*, et al.* Activity-dependent trafficking and dynamic localization of zipcode binding protein 1 and beta-actin mRNA in dendrites and spines of hippocampal neurons. *J Neurosci* 2003, **23**(8)**:** 3251-3261.

6. Huttelmaier S, Zenklusen D, Lederer M, Dictenberg J, Lorenz M, Meng X*, et al.* Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1. *Nature* 2005, **438**(7067)**:** 512-515.

7. Satori CP, Kostal V, Arriaga EA. Review on recent advances in the analysis of isolated organelles. *Anal Chim Acta* 2012, **753:** 8-18.

8. Qin W, Cho KF, Cavanagh PE, Ting AY. Deciphering molecular interactions by proximity labeling. *Nat Methods* 2021, **18**(2)**:** 133-143.

9. Seath CP, Trowbridge AD, Muir TW, MacMillan DWC. Reactive intermediates for interactome mapping. *Chem Soc Rev* 2021, **50**(5)**:** 2911-2926.

10. Rhee HW, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA*, et al.* Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* 2013, **339**(6125)**:** 1328-1331.

11. Lam SS, Martell JD, Kamer KJ, Deerinck TJ, Ellisman MH, Mootha VK*, et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* 2015, **12**(1)**:** 51-54.

12. Hung V, Zou P, Rhee HW, Udeshi ND, Cracan V, Svinkina T*, et al.* Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Mol Cell* 2014, **55**(2)**:** 332-341.

13. Han S, Udeshi ND, Deerinck TJ, Svinkina T, Ellisman MH, Carr SA*, et al.* Proximity Biotinylation as a Method for Mapping Proteins Associated with mtDNA in Living Cells. *Cell Chem Biol* 2017, **24**(3)**:** 404-414.

14. Hung V, Lam SS, Udeshi ND, Svinkina T, Guzman G, Mootha VK, *et al.* Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *Elife* 2017, **6**.

15. Loh KH, Stawski PS, Draycott AS, Udeshi ND, Lehrman EK, Wilton DK, *et al.* Proteomic Analysis of Unbounded Cellular Compartments: Synaptic Clefts. *Cell* 2016, **166**(5)**:** 1295-1307 e1221.

16. Rees JS, Li XW, Perrett S, Lilley KS, Jackson AP. Selective Proteomic Proximity Labeling Assay Using Tyramide (SPPLAT): A Quantitative Method for the Proteomic Analysis of Localized Membrane-Bound Protein Clusters. *Curr Protoc Protein Sci* 2015, **80:** 19 27 11-19 27 18.

17. Bar DZ, Atkatsh K, Tavarez U, Erdos MR, Gruenbaum Y, Collins FS. Biotinylation by antibody recognition-a method for proximity labeling. *Nat Methods* 2018, **15**(2)**:** 127-133.

18. Geri JB, Oakley JV, Reyes-Robles T, Wang T, McCarver SJ, White CH, *et al.* Microenvironment mapping via Dexter energy transfer on immune cells. *Science* 2020, **367**(6482)**:** 1091-1097.

19. Oakley JV, Buksh BF, Fernández DF, Oblinsky DG, Seath CP, Geri JB, *et al.* Radius measurement via super-resolution microscopy enables the development of a variable radii proximity labeling platform. *Proceedings of the National Academy of Sciences* 2022, **119**(32).

20. Choi-Rhee E, Schulman H, Cronan JE. Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. *Protein Sci* 2004, **13**(11)**:** 3043-3050.

21. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* 2012, **196**(6)**:** 801-810.

22. Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* 2018, **36**(9)**:** 880-887.

23. Kaewsapsak P, Shechner DM, Mallard W, Rinn JL, Ting AY. Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *Elife* 2017, **6**.

24. Fazal FM, Han S, Parker KR, Kaewsapsak P, Xu J, Boettiger AN, *et al.* Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* 2019, **178**(2)**:** 473-490 e426.

25. Padron A, Iwasaki S, Ingolia NT. Proximity RNA Labeling by APEX-Seq Reveals the Organization of Translation Initiation Complexes and Repressive RNA Granules. *Mol Cell* 2019, **75**(4)**:** 875-887 e875.

26. Zhou Y, Wang G, Wang P, Li Z, Yue T, Wang J, *et al.* Expanding APEX2 Substrates for Proximity-Dependent Labeling of Nucleic Acids and Proteins in Living Cells. *Angew Chem Int Ed Engl* 2019, **58**(34)**:** 11763-11767.

27. Wang P, Tang W, Li Z, Zou Z, Zhou Y, Li R, *et al.* Mapping spatial transcriptome with light-activated proximity-dependent RNA labeling. *Nat Chem Biol* 2019, **15**(11)**:** 1110-1119.

28. Engel KL, Lo HG, Goering R, Li Y, Spitale RC, Taliaferro JM. Analysis of subcellular transcriptomes by RNA proximity labeling with Halo-seq. *Nucleic Acids Res* 2022, **50**(4)**:** e24.

29. McGinnis JL, Dunkle JA, Cate JH, Weeks KM. The mechanisms of RNA SHAPE chemistry. *J Am Chem Soc* 2012, **134**(15)**:** 6617-6624.

30. Lee B, Flynn RA, Kadina A, Guo JK, Kool ET, Chang HY. Comparison of SHAPE reagents for mapping RNA structures inside living cells. *RNA* 2017, **23**(2)**:** 169-174.

31. Xiao L, Fang L, Chatterjee S, Kool ET. Diverse Reagent Scaffolds Provide Differential Selectivity of 2'-OH Acylation in RNA. *J Am Chem Soc* 2023, **145**(1)**:** 143-151.

32. Sun W, Wang N, Liu H, Yu B, Jin L, Ren X, *et al.* Genetically encoded chemical crosslinking of RNA in vivo. *Nat Chem* 2023, **15**(1)**:** 21-32.

33. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, *et al.* Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* 2015, **519**(7544)**:** 486-490.

34. Xiao L, Jun YW, Kool ET. DNA Tiling Enables Precise Acylation-Based Labeling and Control of mRNA. *Angew Chem Int Ed Engl* 2021, **60**(51)**:** 26798-26805.

35. Xiao L, Habibian M, Kool ET. Site-Selective RNA Functionalization via DNA-Induced Structure. *J Am Chem Soc* 2020, **142**(38)**:** 16357-16363.

36. Tian L, Yang Y, Wysocki LM, Arnold AC, Hu A, Ravichandran B, *et al.* Selective esterase-ester pair for targeting small molecules with cellular specificity. *Proc Natl Acad Sci U S A* 2012, **109**(13)**:** 4756-4761.

37. Jones KA, Kentala K, Beck MW, An W, Lippert AR, Lewis JC, *et al.* Development of a Split Esterase for Protein-Protein Interaction-Dependent Small-Molecule Activation. *ACS Cent Sci* 2019, **5**(11)**:** 1768-1776.

38. Yin J, Bai Y, Mao M, Zhu G. Silver-catalyzed regio- and stereoselective addition of carboxylic acids to ynol ethers. *J Org Chem* 2014, **79**(19)**:** 9179-9185.

39. Stang PJ. Ynol esters and alkynyl(phenyl)iodonium chemistry. *Russian Chemical Bulletin* 1993, **42**(1)**:** 12-23.

40. Gray VJ, Wilden JD. The chemistry of ynol and thioynol ethers. *Org Biomol Chem* 2016, **14**(41)**:** 9695-9711.

41. Lal B, Pramanik BN, Manhas MS, Bose AK. Diphenylphosphoryl azide a novel reagent for the stereospecific synthesis of azides from alcohols. *Tetrahedron Letters* 1977, **18**(23)**:** 1977-1980.

42. Dawson PE, Muir TW, Clark-Lewis I, Kent SB. Synthesis of proteins by native chemical ligation. *Science* 1994, **266**(5186)**:** 776-779.

43. Bejot R, Tisserand S, Reddy LM, Barma DK, Baati R, Falck JR, *et al.* Stereoselective transformations of trihalomethylcarbinols induced by chromous chloride. *Angew Chem Int Ed Engl* 2005, **44**(13)**:** 2008-2011.

44. Zhao Y, Zhou L, Li H, Sun T, Wen X, Li X, *et al.* Nuclear-Encoded lncRNA MALAT1 Epigenetically Controls Metabolic Reprogramming in HCC Cells through the Mitophagy Pathway. *Mol Ther Nucleic Acids* 2021, **23:** 264-276.

45. Mohammad G, Kowluru RA. Nuclear Genome-Encoded Long Noncoding RNAs and Mitochondrial Damage in Diabetic Retinopathy. *Cells* 2021, **10**(12).

46. Rippo MR, Olivieri F, Monsurro V, Prattichizzo F, Albertini MC, Procopio AD. MitomiRs in human inflamm-aging: a hypothesis involving miR-181a, miR-34a and miR-146a. *Exp Gerontol* 2014, **56:** 154-163.

47. Kang M, Tang B, Li J, Zhou Z, Liu K, Wang R, *et al.* Identification of miPEP133 as a novel tumor-suppressor microprotein encoded by miR-34a pri-miRNA. *Mol Cancer* 2020, **19**(1)**:** 143.

48. Wang H, Liang L, Dong Q, Huan L, He J, Li B, *et al.* Long noncoding RNA miR503HG, a prognostic indicator, inhibits tumor metastasis by regulating the HNRNPA2B1/NF-kappaB pathway in hepatocellular carcinoma. *Theranostics* 2018, **8**(10)**:** 2814-2829.

49. Sone M, Hayashi T, Tarui H, Agata K, Takeichi M, Nakagawa S. The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *J Cell Sci* 2007, **120**(Pt 15)**:** 2498-2506.

50. Yu B, Shan G. Functions of long noncoding RNAs in the nucleus. *Nucleus* 2016, **7**(2)**:** 155-166.

51. Reyes-Gutierrez P, Ritland Politz JC, Pederson T. A mRNA and cognate microRNAs localize in the nucleolus. *Nucleus* 2014, **5**(6)**:** 636-642.

52. Long MJC, Huang K-T, Aye Y. The not so identical twins: (dis)similarities between reactive electrophile and oxidant sensing and signaling. *Chemical Society Reviews* 2021, **50**(22)**:** 12269-12291.

53. van Dijk M, Visser A, Buabeng KM, Poutsma A, van der Schors RC, Oudejans CB. Mutations within the LINC-HELLP non-coding RNA differentially bind ribosomal and RNA splicing complexes and negatively affect trophoblast differentiation. *Hum Mol Genet* 2015, **24**(19)**:** 5475-5485.

54. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013, **29**(1)**:** 15-21.

55. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014, **15**(2)**:** R29.

56.   Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W*, et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015, **43**(7)**:** e47.

Acknowledgements

Author Information

These authors contributed equally: Shubhashree Pani and Tian Qiu.

Contributions

S.P., T.Q., K.K, and B.C.D. conceived of the project. S.P. and T.Q. synthesized all compounds and performed experiments. S-A.A. assisted with experiments and data analysis. S.P., T.Q., and B.C.D. wrote the paper with input from all authors.

Corresponding author

Correspondence to Bryan Dickinson, dickinson@uchiago.edu

Ethics Declaration

Competing interests: B.C.D. is a founder and holds equity in Tornado Bio, Inc.

Supporting Information

Supplementary Information (Supplementary Figs. 1–27, synthetic methods, BAP-seq Protocol).

Supplementary Table 1 (Filtered and TMM-normalized cpm reads for all the samples)

Supplementary Table 2 (Differential expression gene analysis data used for comparison of samples with BS2 expressed in different locales in the volcano plots)