

In Vivo Proximity Labeling for the Detection of Protein–Protein and Protein–RNA Interactions

David B. Beck,^{†,⊥} Varun Narendra,^{†,⊥} William J. Drury, III,^{†,#} Ryan Casey,[†] Pascal W. T. C. Jansen,^{‡,⊙} Zuo-Fei Yuan,[§] Benjamin A. Garcia,[§] Michiel Vermeulen,^{‡,⊙} and Roberto Bonasio^{*,||}

[†]Howard Hughes Medical Institute and Department of Biochemistry, New York University School of Medicine, 522 First Avenue, New York, New York 10016, United States

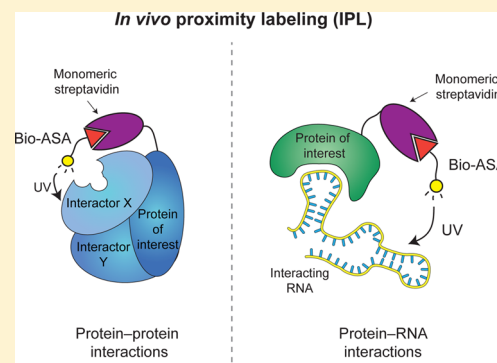
[‡]Department of Molecular Cancer Research, University Medical Center Utrecht, Universiteitsweg 100, 3584CG Utrecht, The Netherlands

[§]Department of Biochemistry and Biophysics, Epigenetics Program, ^{||}Department of Cell and Developmental Biology, Epigenetics Program, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19103, United States

S Supporting Information

ABSTRACT: Accurate and sensitive detection of protein–protein and protein–RNA interactions is key to understanding their biological functions. Traditional methods to identify these interactions require cell lysis and biochemical manipulations that exclude cellular compartments that cannot be solubilized under mild conditions. Here, we introduce an in vivo proximity labeling (IPL) technology that employs an affinity tag combined with a photoactivatable probe to label polypeptides and RNAs in the vicinity of a protein of interest in vivo. Using quantitative mass spectrometry and deep sequencing, we show that IPL correctly identifies known protein–protein and protein–RNA interactions in the nucleus of mammalian cells. Thus, IPL provides additional temporal and spatial information for the characterization of biological interactions in vivo.

KEYWORDS: Proximity labeling, protein–protein interactions, protein–RNA interactions, biotinylation, covalent tag, RNA-seq



INTRODUCTION

Cell survival depends on the network of molecular interactions among protein, RNA, and DNA. These interactions cover a vast range of affinities, from subnanomolar, as observed in stable protein complexes, to micromolar, as seen in transient and dynamic binding events that are nonetheless essential to cellular processes. In fact, physical interactions between biological macromolecules are so central to their function that they are often employed to infer the biological role of the interacting partners. On many occasions, when a new protein is studied, the first step in its characterization is to perform affinity purification followed by mass spectrometry (MS) with the goal of isolating and identifying its binding partners, thus revealing the biochemical pathways in which the protein is involved. Similarly, the accurate detection of protein–DNA and protein–RNA interactions is integral to the understanding of the complex processes that regulate genome output within the nucleus. The importance of these interactions has become increasingly evident with the recent deluge of studies based on genome-wide chromatin immunoprecipitation (ChIP-seq)¹ and RNA immunoprecipitation (RIP-seq).^{2,3}

The current gold standard for the detection of protein–protein interactions is the isolation of stable complexes and identification of associated polypeptides by MS.⁴ Although this approach has been invaluable in dissecting the interactome of

cells in a variety of conditions, its scope is limited to protein–protein interactions that persist through the purification steps and resist the physicochemical perturbations required to extract proteins from live cells and to maintain them in solution. Moreover, a considerable number of proteins resist extraction under the relatively mild conditions required to preserve weak protein–protein interactions and are thus inaccessible to investigation by conventional biochemical methods. These limitations are particularly evident in the study of nuclear proteins, which often remain associated with chromatin even at high salt concentrations.⁵

The problem of accurately detecting protein–RNA interactions in vivo is even more challenging because the solubilization of cellular structures during lysis facilitates spurious protein–RNA interactions⁶ and because RNA is notoriously unstable in cellular extracts. The reliable identification of protein–RNA interactions has become a key bottleneck in the understanding of noncoding RNAs in chromatin function and gene regulation.^{7,8} In fact, conventional RIP experiments have repeatedly identified very large numbers of mouse and human RNAs as associated with various

Received: March 5, 2014

Published: October 14, 2014

chromatin proteins,^{2,9,10} raising the question of whether some of these interactions may form *in vitro*, after lysis.

Here, we describe a technique that exploits physical proximity to covalently label interacting proteins and RNAs in live cells. Interacting molecules labeled *in vivo* can be recovered using harsh chemical conditions and identified *in vitro*. Our technology employs a genetic tag fused to a protein of interest (POI) that recruits a photoactivatable heterobifunctional probe that covalently labels proteins and RNAs in the proximity of the POI. We call this technology *in vivo* proximity labeling (IPL), and, as a proof of concept, we demonstrate that it successfully identifies known protein–protein and protein–RNA interactions.

MATERIALS AND METHODS

Plasmids and Stable Cell Lines

293T-REx cells were purchased from Invitrogen (CA). Cells were grown using DMEM containing 10% fetal bovine serum, 100 U/mL penicillin, 100 µg/mL streptomycin, and 2 mM L-glutamine. Codon-optimized monomeric streptavidin (mSA) corresponding to mutant M4 as described in Wu et al.¹¹ was synthesized *de novo* (GenScript, NJ) and cloned into the pINTO backbone, a modified pCDNA4-TO vector containing chicken insulator sequences, a generous gift from Dr. Gary Felsenfeld (NIH/NIDDK). The coding sequences, from ATG to the stop codon, of EZH2, SNRNP70, and the DNA binding domain of *Saccharomyces cerevisiae* Gal4p (GAL4) were cloned into pINTO-mSA and pINTO-Flag-HA¹² (pINTO-FH) using BamHI and XhoI. 293T-REx stable cell lines were generated for all pINTO constructs and maintained in 100 µg/mL zeocin (Invitrogen). A control 293T-REx line with an integrated transgene for the inducible expression of GAL4–EZH2 was previously described.¹³

In Vivo Proximity Labeling

Stable 293T-REx cell lines were induced to express transgenes by treatment with 1 µg/mL doxycycline for 24 h. For each experiment, an adequate number of 10 cm tissue culture dishes with cells at 60–80% confluence (~10⁷ cells/dish) was used. EZ-Link Biotin-LC-ASA (Pierce Biotechnologies, IL) was added to the cells in 10 mL of freshly prepared complete medium at a concentration of 10 µM (empirically determined as giving the best signal-to-noise ratio, see Figures S3A and 4B) followed by incubation for 2 h at 37 °C. After incubation, cells were washed once in ice-cold PBS. Photolabeling was induced with 500 mJ/cm² UVA (365 nm) in a Stratelinker irradiator (Stratagene, CA), with 8 mL of PBS covering the cell monolayer while the cells were kept on ice. After irradiation, cells were lysed in 10 mM Tris, 500 mM KCl, 0.1 mM EDTA, 10% glycerol, 1% IGEPAL-630, 0.5% sarkosyl and briefly sonicated to ensure DNA shearing. Lysates were incubated with streptavidin beads (Millipore, MA) for 1–2 h at room temperature. Because monomeric streptavidin has much lower affinity ($K_d \sim 10^{-7}$) for biotin,¹¹ it can be efficiently competed by wild-type streptavidin beads. After incubation beads were washed three times using lysis buffer followed by a final wash without detergent. In one replicate (see Figure 3A and Table S4), we performed an additional, high-stringency wash with lysis buffer plus 1% SDS. Elutions were performed in Laemmli loading buffer supplemented with 1 mM biotin and analyzed by western blotting or MS.

Conventional Mass Spectrometry

Samples were run as a gel plug using a Novex gel Bis-Tris 10% gel. The entire band was excised, and proteins in the gel were reduced, carboxymethylated, and digested with trypsin using standard protocols. Peptides were extracted, solubilized in 0.1% trifluoroacetic acid, and analyzed by nano LC–MS/MS using a RSLC system (ThermoFisher, MA) interfaced with a Velos-LTQ-Orbitrap (ThermoFisher). Samples were loaded onto a self-packed 100 µm × 2 cm trap packed with Magic C18AQ, 5 µm 200 Å (Michrom Bioresources Inc., Auburn, CA) and washed with buffer A (0.2% formic acid) for 5 min with flow-rate of 10 µL/min. The trap was brought in-line with the homemade analytical column (Magic C18AQ, 3 µm 200 Å, 75 µm × 50 cm), and peptides were fractionated at 300 nL/min with a multistep gradient: 4–15% buffer B (0.16% formic acid 80% acetonitrile) in 15 min, 15–25% B in 45 min, and 25–55% B in 30 min.

MS data was acquired using a data-dependent acquisition procedure with a cyclic series of a full scan acquired in Orbitrap with resolution of 60 000 followed by MS/MS scans of the 20 (for CID; replicates 1 and 3) or 10 (for HCD; replicate 2) most intense ions, with a repeat count of two and dynamic exclusion duration of 60 s. For CID, selected ions were fragmented and scanned in linear orbitrap, and centroid data were recorded. For HCD, selected ions were fragmented in the HCD cell using 40% of the collision energy and scanned in Orbitrap with resolution of 15 000 and recorded as centroid data.

The LC–MS/MS data was processed by pParse to calibrate the precursor isotopic mass to the monoisotopic mass and export coeluted precursors, to maximize the identification rate.¹⁴ The UniProt human database was searched (date 5/3/2011, number of sequences 39 703). Parameters for the database search engine pFind 2.8^{15,16} were set as follows: precursor mass tolerance ±10 ppm; fragment mass tolerance ±0.02 Th for HCD and ±0.4 Th for CID; trypsin cleaving after lysine and arginine with 2 miscleavages tolerated; cysteine carbamidomethylation as fixed modification; and methionine oxidation, protein N-terminal acetylation, and N-terminal Gln to pyro-Glu as variable modifications. The target-decoy approach was used to filter the search results,¹⁷ with an FDR < 1% at the spectral level. Only proteins identified by at least 2 unique peptides in each experiment were considered for further analysis.

SILAC

To achieve isotopic labeling, 293T-REx lines with integrated transgenes encoding GAL4–EZH2 or mSA–EZH2 were grown in heavy (H) medium, consisting of DMEM lacking arginine and leucine, supplemented with dialyzed FBS, 0.46 mM heavy Arg10 (¹³C6,¹⁵N4 L-arginine), 0.47 mM heavy Lys8 (¹³C6,¹⁵N2 L-lysine), and 2 mM light proline (all from Pierce/Thermo Scientific, IL), to minimize the conversion of heavy Arg into heavy Pro. Labels were incorporated by culturing in SILAC media for at least 14 days (4–5 passages) prior to IPL. After induction of the transgenes, IPL was performed as above. Before streptavidin pull-down, equal amounts of heavy and light lysates were mixed. After pull-down, peptides were acidified with TFA following trypsin digestion and desalted using Stagetips (Thermo Scientific, MA) prior to mass spec analyses. After elution from the Stagetips, peptides were applied to online nano LC–MS/MS using a 30 cm long fused silica emitter column with a 75 µm inner diameter (New Objective, MA) custom packed with 3 µm C-18 beads (Dr. Maisch,

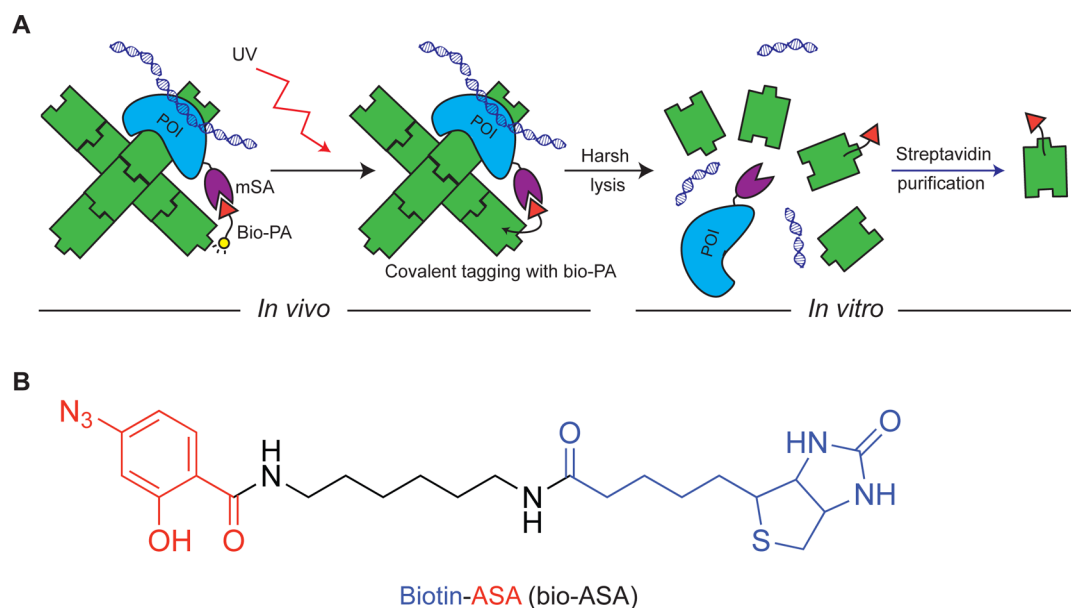


Figure 1. In vivo proximity labeling (IPL). (A) Schematic depiction of the IPL strategy. A protein of interest (POI) is fused with monomeric streptavidin (mSA), which recruits a probe (bio-PA) constituted of biotin (red triangle) linked to a photoactivatable group (yellow circle). After UV irradiation the photoactivatable group reacts with proteins and other macromolecules that are in close proximity to the POI *in vivo*. Because bio-PA is now covalently bound to putative POI interactors, the cells can be lysed under harsh conditions and the identity of the POI interactors can be revealed by streptavidin purification followed by mass spectrometry. (B) Chemical structure of the bio-ASA heterobifunctional probe used for IPL.

Germany) on a Proxeon EASY-nLC (Thermo Scientific, MA) and fractionated with a 120 min gradient from 7 until 32% acetonitrile followed by stepwise increases up to 95% acetonitrile. Mass spectra were recorded on an LTQ-Orbitrap-Velos (Thermo Scientific) with a resolution of 30 000 followed by MS/MS using CID of the 15 most intense precursor ions of every full scan with a dynamic exclusion duration of 50 s, repeat count of 1, repeat duration of 30 s, exclusion list size of 500, and exclusion mass width (low and high) of 10 ppm.

Raw data were analyzed by MaxQuant¹⁸ (version 1.1.1.25) using the Andromeda search engine.¹⁹ The MS/MS data was searched against the human International Protein Index database, version 3.38 (70 856 sequences). Parameters were set as follows: trypsin was set as enzyme and 2 missed cleavages were allowed; as variable modifications, oxidation on methionine and acetylation of the protein N-terminus were selected; for fixed modification, carbamidomethyl on cysteine was selected; in labels, “doublets” was chosen, and for the heavy labels, Arg10 and Lys8 were selected. We used standard settings for MS/MS and identification and quantification. The experimental design was uploaded to specify the samples and labeling combinations in the forward and reverse experiments.

We considered only proteins identified by at least two unique peptides in both the forward and the reverse pull-downs. *P*-values for enrichment were calculated by taking into account ratios and intensity values (significance B as described in Cox et al.¹⁸). Significant hits (+) were identified using a Benjamini–Hochberg FDR cutoff of 5%.

Antibodies and Immunoprecipitations

Antibodies against EZH2, EED, and SCML2 were kindly provided by the Reinberg laboratory.^{20,21} Antibodies against biotin (Bethyl laboratories, TX), SNRNP70 (Santa Cruz Biotech, TX), and SUZ12 (Cell Signaling Technology, MA) were obtained from their respective vendors. Immunoprecipitations were performed overnight in 10 mM Tris, 200 mM KCl,

0.1 mM EDTA, 0.05% IGEPAL-630, and washed three times prior to elution with Laemmli SDS-PAGE loading buffer and western blotting.

IPL for RNA and RNA-seq

Stable 293T-REx cell lines were induced to express mSA–SNRNP70 or FH–SNRNP70 by treatment with 1 μ g/mL doxycycline for 24 h and then incubated with increasing concentrations of bio-ASA for 2 h at 37 °C. Cells were washed once with ice-cold PBS. Photolabeling was induced with 500 mJ/cm² UVA (365 nm) in a Stratelinker irradiator (Stratagene, CA), with 8 mL of PBS covering the cell monolayer while the cells were kept on ice. RNA was isolated with TRIzol (Invitrogen), and biotinylated species were precipitated with MyOne Streptavidin C1 Dynabeads (Invitrogen) in RIP buffer (10 mM Tris, 200 mM KCl, 10 mM EDTA, 0.05% IGEPAL-630) with 3 washes in RIP-W buffer (10 mM Tris, 200 mM KCl, 0.1 mM EDTA, 0.05% IGEPAL-630, 1 mM MgCl₂). The abundance of different species of precipitated RNA was determined by reverse transcription quantitative PCR (RT-qPCR) using QuantiTect SYBR green PCR kits (Qiagen, MD) and the following primers: 5S sense, TCGTCTATCTCGG-AAGCTAAG; 5S antisense, GCCTACAGCACCC-GGTATTC; U1 sense, GGGAGATACCATGATCACGAAG; U1 antisense, CAAATTATGCAGTCGAGTTTCC. For deep sequencing, RNA was prepared as above and then converted to strand-specific Illumina libraries using a published protocol.²² Briefly, RNA was chemically fragmented, converted to cDNA, ligated to custom-designed barcoded adapters, size-selected, and amplified by ligation-mediated PCR. Input and streptavidin pull-down libraries from 2 biological replicates were sequenced on a HiSeq2000 and a NextSeq 500 (Illumina) at a depth of at least >20⁶ reads each. Reads were mapped and analyzed with a custom bioinformatic pipeline based on BOWTIE,²³ SAM-TOOLS,²⁴ and the R package DEGseq.²⁵ Only genes with 10 or more mapped reads were included in downstream analyses.

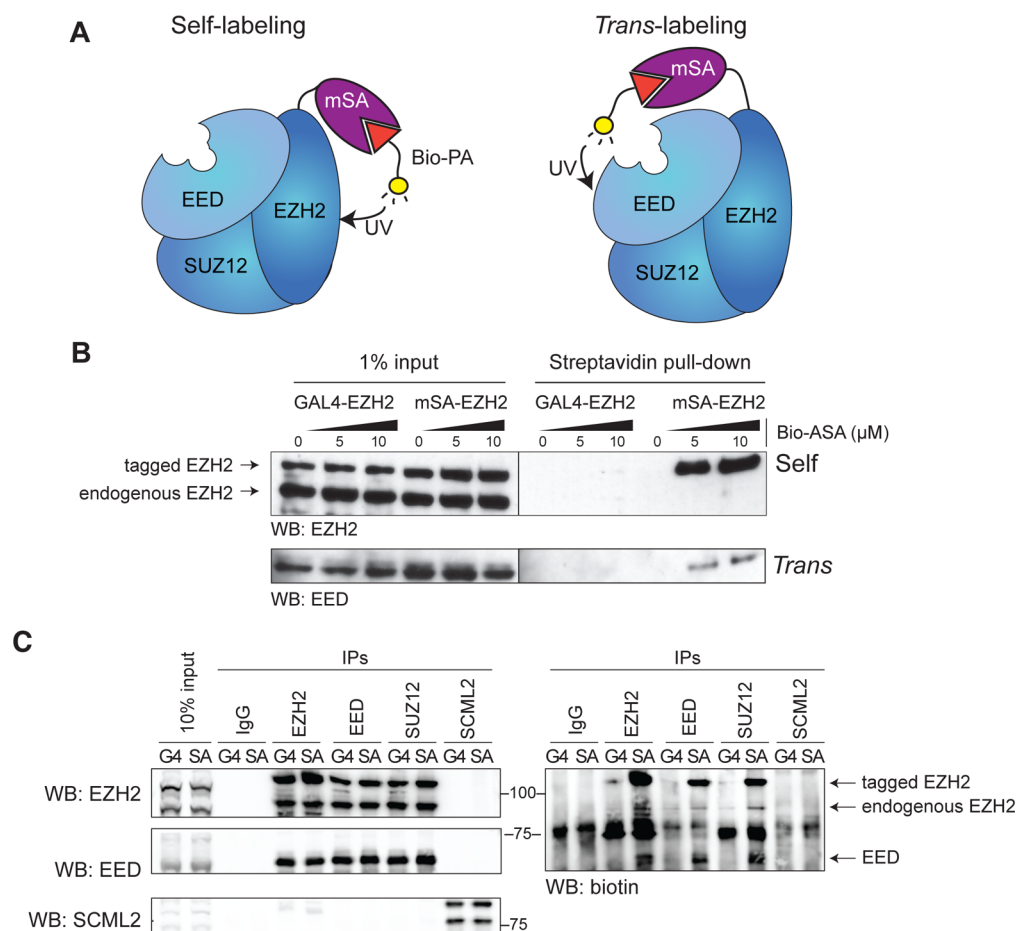


Figure 2. Self- and trans-labeling by IPL in the PRC2 complex. (A) Schematic depiction of self-labeling reactions (left) and trans-labeling reaction (right) in the context of the PRC2 complex used for this proof of concept. (B) 293T-REX cells expressing the N-terminal fusions mSA-EZH2 or GAL4-EZH2 (negative control) were subjected to IPL with the indicated concentrations of bio-ASA. Biotinylated proteins were recovered by streptavidin pull-down and revealed by western blots with EZH2 and EED antibodies. (C) IPL of GAL4-EZH2 (control) and mSA-EZH2 followed by IP for PRC2 components (EZH2, EED, and SUZ12) as well as a PRC1-associated factor (SCML2) as an additional control. Western blots for EZH2, EED, and SCML2 (left) and biotin (right) are shown.

Sequencing Data

All sequencing data has been deposited to the GEO as series GSE55370.

RESULTS

We conceived and developed IPL with the goal of overcoming some of the limitations found in traditional affinity purification schemes. Specifically, being interested in nuclear processes and nuclear organization, we were concerned that many chromatin-associated POIs or their interacting partners would not be easily extracted from nuclei without disrupting their biochemical milieu.

We were inspired by the DamID approach, in which an *Escherichia coli* adenine methyltransferase is fused to a POI so that chromosome regions that come in contact with it are methylated at adenine positions (which does not naturally occur in eukaryotic cells) and can be later detected by methyl-sensitive PCR.²⁶ In other words, DamID converts spatial information (the proximity of certain DNA sequences to the POI) into chemical information encoded within the DNA molecules in the form of methyl-adenines. We sought to develop an analogous methodology for protein–protein and protein–RNA interactions using a chemical biology approach.

IPL utilizes a monomeric mutant of streptavidin (mSA)¹¹ that, when fused to a POI, recruits to its vicinity a small chemical probe that contains a biotin and a photoactivatable moiety (bio-PA; Figure 1A, left). IPL probes can be directly added to the cell culture medium without adverse effects on cell viability and growth (Figure S1); they cross the plasma membrane and reach the POI inside the cells. Once equilibrium is achieved, the cells are irradiated with UV light of suitable wavelength for activation of the photochemical group, which reacts promptly and irreversibly with molecules in the vicinity. The result of this procedure is that proteins (and RNAs) inside the cell are biotinylated with increasing efficiency as a function of their proximity to the POI at the moment of irradiation. Photoactivated biotinylation gives rise to covalent bonds; therefore, cells subjected to IPL can be lysed under extremely harsh chemical conditions without losing information regarding the *in vivo* interactions, which remains encoded in the distribution of the covalent biotin tag. The relative biotinylation enrichment for each protein is obtained by performing stringent streptavidin precipitation followed by MS (Figure 1A, right) and comparison with appropriate negative controls. Thus, IPL offers the possibility of identifying candidate interactors for proteins beyond the reach of traditional biochemistry and does so by setting very stringent require-

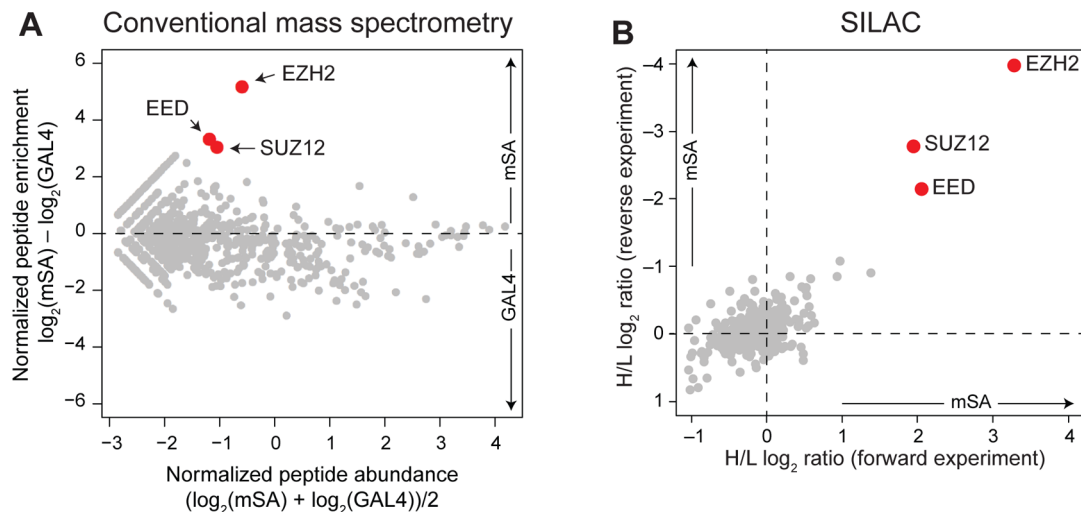


Figure 3. Unbiased IPL proteomics. (A) 293T-REx expressing mSA–EZH2 or GAL4–EZH2 were subjected to IPL. Biotinylated proteins were purified using streptavidin and identified by MS. Each identified protein is represented by a dot in the scatter plot. The x axis indicates the normalized and log-converted average of unique peptide abundance in mSA–EZH2 and GAL4–EZH2; the y axis indicates the specific enrichment in mSA–EZH2 samples (above the dotted line) compared to the GAL4–EZH2 control. Red dots indicate the position of the PRC2 core components. Data is averaged from 3 biological replicates. (B) Scatter plot with SILAC enrichment scores for polypeptides biotinylated by IPL in mSA–EZH2 cells. The two axes indicate the normalized ratios of spectral counts for each polypeptide in the heavy sample vs the light sample (H/L) in the forward SILAC (GAL4–EZH2 light, mSA–EZH2 heavy) and reverse SILAC (GAL4–EZH2 heavy, mSA–EZH2 light).

ments for the temporal and spatial parameters of protein–protein and protein–RNA interactions being reported.

To identify the biotin probe most suitable for *in vivo* labeling, we tested a number of commercially available compounds that comprise a biotin linked by inert spacers of comparable lengths to different photoreactive groups, such as psoralen, tetrafluorophenyl azide (TFPA), aryl azide (ASA), and benzophenone (BP) (Figure S2A). Despite some non-specific biotinylation of the GAL4–EZH2 control bait, the most robust specific labeling of mSA–EZH2 occurred in cells treated with biotin conjugated with ASA (bio-ASA) (Figure S2B), and we concluded that this compound was the most efficient in photolabeling proteins *in vivo*. We also reasoned that efficient photolabeling of the mSA–EZH2 bait would correspond to efficient photolabeling of potential interactors *in trans*, as the chemical reactions involved are the same, regardless of the identity of the target.

In this embodiment, IPL exploits the reversible interaction between mSA and bio-ASA to recruit the probe to the vicinity of the POI. Because the monomeric mutant of streptavidin has much lower affinity for biotin than that of its wild-type counterpart ($K_d \sim 10^{-7}$ compared to $\sim 10^{-15}$), the interaction with the mSA fusion tag is quickly reversed after labeling, and biotinylated interactors can be purified using wild-type streptavidin. In the bio-ASA probe, the photoreactive group is connected to the biotin by a ~ 14 Å linker (Figure S2A), which sets an upper limit to the distance from the mSA site where an interactor can be still detected with this probe. Upon UV irradiation at 365 nm, the aryl azide in bio-ASA is converted to a highly reactive aryl nitrene that may directly insert onto C–H bonds and N–H bonds or rearrange to a dehydroazepine and subsequently react with nucleophiles in the biological milieu (i.e., proteinaceous or nucleobase amines).²⁷ This reactivity profile maximizes its chances of reacting with biomolecules rather than with water.

IPL Detects the Composition of the PRC2 Protein Complex

We sought to validate our technology on a known protein complex and to determine whether well-established protein–protein interactions would be identified by IPL. For these experiments, we selected polycomb repressive complex-2 (PRC2), a chromatin-modifying protein complex that has been extensively characterized by conventional and affinity purifications.²⁸ The core PRC2 complex is composed of EZH2, the catalytic subunit responsible for its histone methyltransferase activity, SUZ12, EED, and the ubiquitous histone binding proteins RBBP4/7²⁸ (Figure 2A). We reasoned that, in addition to self-labeling of mSA–EZH2 (Figure 2A, left), photo-activation of bio-ASA should give rise to *trans*-biotinylation of other, untagged subunits of the PRC2 complex (Figure 2A, right), thus identifying them as *in vivo* interactors.

To determine the concentration of bio-ASA probe to use for IPL, we performed a titration experiment. We found that 10 μM bio-ASA in the culture medium maximized specific biotinylation of mSA–EZH2 compared to nonspecific biotinylation of untagged endogenous EZH2 (Figure S3A). To control for residual background biotinylation and distinguish specific from nonspecific labeling, in all subsequent experiments we used cells expressing the bait fused to an irrelevant tag (either GAL4 or FH) that cannot bind bio-ASA and considered only the specific enrichment in biotinylation caused by fusion of the bait to the mSA tag.

After IPL, only mSA–EZH2 bound to streptavidin-coupled beads, whereas GAL4–EZH2 did not (Figure 2B, top), demonstrating that the self-labeling reaction took place as expected and in a specific manner. We also recovered endogenous EED in the streptavidin precipitation in a specific manner, only when IPL was performed on mSA–EZH2-expressing cells (Figure 2B, bottom), suggesting that EED had been biotinylated *in trans*. Importantly, all streptavidin precipitations were performed in the presence of 0.5% sarkosyl, which disrupted interactions between EZH2 and EED within the PRC2 complex (Figure S3B).

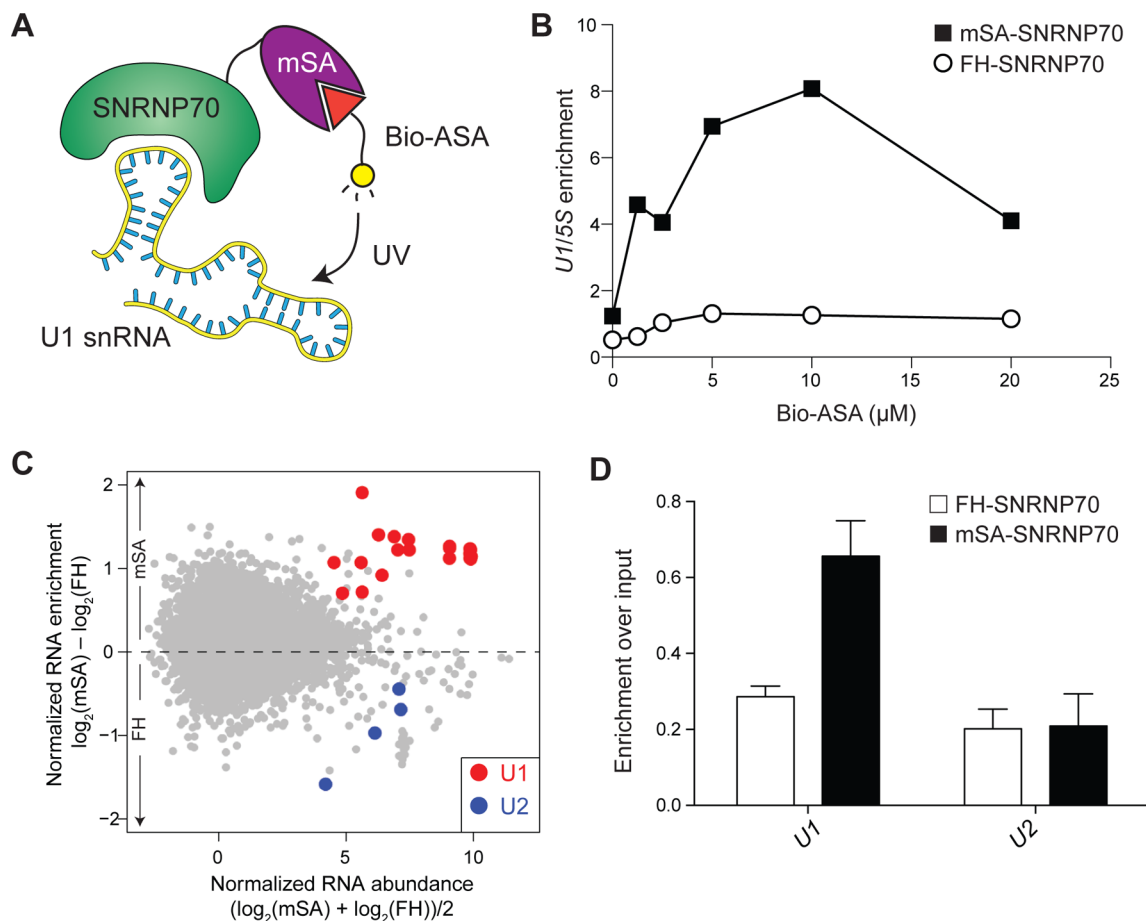


Figure 4. IPL of protein-interacting noncoding RNAs. (A) Schematic depiction of the labeling reaction. The mSA tag was fused to the N terminus of SNRNP70, which interacts with the U1 spliceosomal snRNA. After IPL, part of the label is deposited on the RNA. (B) IPL was performed with a different concentration of bio-ASA probe (x axis) as determined by 293T-REx expressing either mSA-SNRNP70 (black squares) or FH-SNRNP70 (white circles) as a control. RNA was extracted with TRIzol and precipitated with streptavidin-coupled magnetic beads. The y axis shows the enrichment of U1 RNA compared to 5S rRNA after precipitation, as determined by RT-qPCR. (C) Enrichment of U1 RNA in SNRNP70 IPL, as determined by deep sequencing and mapping to annotations in ENSEMBL 71. Mean abundance is plotted on the x axis, and input-corrected enrichment is plotted on the y axis. U1 and U2 genes are highlighted in red and blue, respectively. Data is from 2 biological replicates. (D) Quantification of U1 and U2 IPL enrichment according to deep sequencing. Reads per kilobase per million (RPKM) for each U1 and U2 locus were calculated in FH and mSA SNRNP70 IPLs and divided for the RPKM of the respective genes in the input RNA. Bars represent mean + SEM.

Covalent biotinylation in trans was also demonstrated by western blot of the biotin tag not only on the mSA-EZH2 bait but also on EED after immunoprecipitations (IPs) with antibodies against EZH2, EED, and SUZ12 (Figure 2C). No biotinylated protein of size similar to EED was observed in control IPs with an irrelevant IgG or with antibodies against SCML2, a component of the PRC1 complex (Figure 2C).

To determine whether IPL would identify EZH2-interacting proteins in an unbiased manner, we repeated the procedure but rather than testing the presence of EED by western blot, we subjected the entire streptavidin-bound fraction to MS. A comparison of number of peptides identified by MS in 3 biological replicates revealed EZH2, SUZ12, and EED as the most reproducibly enriched proteins in streptavidin pull-downs from mSA-EZH2-expressing cells compared to the GAL4-EZH2 controls (Figure 3A and Tables S1–S5). To confirm the direct biotinylation of EED and SUZ12, we included a high-stringency wash in 1% SDS for the third biological replicate, which resulted in the highest enrichment of EED and SUZ12 among the three replicates (Table S4), the opposite of what would be expected if their recovery depended on residual interactions with EZH2. Although it would have been desirable

to directly identify by MS the EED peptides labeled by bio-ASA, pilot experiments with an in vitro-biotinylated recombinant protein revealed that bio-ASA adducts could not be discerned by conventional MS/MS, suggesting that bio-ASA undergoes source decay or fragments in an unpredictable manner at the MS2 stage.

As conventional MS is only semiquantitative, we reproduced this result by utilizing stable isotope labeling with amino acids in cell culture (SILAC) followed by IPL and quantitative MS of streptavidin-purified material (Figure 3B and Tables S6–S7).

Together, these data demonstrate that IPL correctly identifies known interactors (EED, SUZ12) of a test protein (EZH2) in vivo, by both candidate-based (western blot) and unbiased (MS) approaches.

IPL Identifies Protein–RNA Interactions

To further explore the possibilities offered by IPL, we wished to determine whether it could be used to detect protein–RNA interactions in vivo. The importance of these interactions in nuclear processes, especially in the epigenetic regulation of gene expression, can hardly be overestimated.^{7,8,29} As the affinities of some protein–RNA interactions are relatively low

and binding appears rather promiscuous,^{9,30,31} several technologies have been developed to distinguish protein–RNA interactions that occur *in vivo* from spurious associations that take place after lysis.⁶ These technologies typically rely on cross-linking of RNA to proteins either with UV light^{32,33} or formaldehyde, but they present drawbacks: the former requires very tight contacts and specific molecular configurations, and the latter has the potential to create large networks of cross-linked proteins and nucleic acids that may confound the results.

We reasoned that the photoconversion chemistry of the aryl azide in bio-ASA should permit trans-labeling not only of proteins associated with the mSA-tagged POI but also of RNAs, such as, for example, the spliceosomal U1 small nuclear RNA (snRNA) that associates with the SNRNP70 protein (Figure 4A).³⁴ To test this hypothesis, we performed IPL on 293T-REx cells expressing mSA–SNRNP70, extracted total RNA, and purified biotinylated species by streptavidin precipitation. We quantified the amount of precipitated U1 by RT-qPCR and normalized it against the amount of 5S rRNA. The addition of bio-ASA to the cells and subsequent photoactivation resulted in the specific labeling of U1 RNA in mSA–SNRNP70-expressing cells but not in control cells expressing SNRNP70 fused to an irrelevant tag (FH–SNRNP70) (Figure 4B). At higher concentrations, the signal-to-noise ratio deteriorated, suggesting that the optimal concentration for efficient IPL of RNA was 10 μ M bio-ASA (Figure 4B). Next, we wished to determine whether the same approach could retrieve this protein–RNA interaction in an unbiased manner. We performed larger scale IPL on cells expressing mSA–SNRNP70 and FH–SNRNP70, confirmed the specific self-labeling of the mSA-tagged bait (Figure S4), and then purified the RNA biotinylated *in trans* with streptavidin-coupled beads. The RNA was eluted from the beads using TRIzol and constructed into libraries. Deep sequencing of two biological replicates revealed that U1 RNAs were specifically enriched in libraries from mSA–SNRNP70 cells compared to U2 RNAs as negative control (Figure 4C–D).

Together, these results show that IPL can be used for detection not only of protein–protein but also protein–RNA interactions in an unbiased manner.

DISCUSSION

Using chemical biology, genetic tagging, and high-throughput proteomics and sequencing strategies, we have developed and validated IPL as a technology that identifies presumptive protein–protein and protein–RNA interactions by converting information about their *in vivo* proximity into irreversible chemical modifications.

The need for alternative methods to detect macromolecular interactions *in vivo* is particularly pressing in the study of chromatin regulation and other nuclear processes, as they often take place in cellular compartments that are impervious to conventional biochemical approaches. It is estimated that 10% of total nuclear protein remains insoluble even after harsh extraction with 2 M salt, 1% Triton X-100, and nucleases.³⁵ An even larger portion of the nuclear proteome remains insoluble under the mild conditions typically used for nuclear extraction.³⁶ Even those proteins that are extracted with these procedures are likely to lose weak and transient binding partners during the *in vitro* manipulations required for affinity or conventional purifications. Due to these limitations, fluorescence resonance energy transfer and yeast two-hybrid

have been utilized to detect novel interactions *in vivo*, but the former is suitable only for candidate-based approaches, when both binding partners are already known, and the latter tests for interactions in a non-physiological molecular context, as protein fragments are artificially tethered to the yeast transcription apparatus to test for interactions.³⁷

Even more pressing is the need to develop strategies for detection of weak and transient protein–RNA interactions *in vivo*. This need originates from the recent realization that various classes of noncoding RNAs play fundamental roles in a variety of biological processes and in particular in targeting and regulating chromatin-associated complexes that in turn modulate gene expression.^{8,38} However, many of these ncRNA–protein interactions in chromatin-associated complexes display considerable promiscuity when probed with affinity-based approaches.^{2,9,10} Although it is possible that this promiscuity reflects true *in vivo* diversity in protein–RNA interactions, we argue that some of it may be a consequence of the *in vitro* RIP procedure.

We have shown that IPL offers an alternative to affinity-based and cross-linking-based approaches, and, as a proof of concept, we identified *de novo* known protein–protein and protein–RNA interactions. We utilized biotin–streptavidin interactions to both target a commercial photoactivatable probe, bio-ASA, and to isolate the labeled macromolecules. As a protein tag, we chose a monomeric mutant of streptavidin to avoid artificial multimerization of the bait,¹¹ which would have interfered with protein localization and complex formation. The interaction of biotin with mSA has the additional benefit of lower affinity and faster reversibility compared to its interaction with wild-type streptavidin, which facilitates subsequent purification steps.

Our results with EZH2 confirm that the presence of the 15 kDa mSA tag does not impair the formation of the PRC2 complex *in vivo*. Similarly, mSA–SNRNP70 bound to its physiological RNA partner, U1, is unaffected by the presence of the tag.

The idea of utilizing spatial proximity to detect interacting partners was pioneered with the DamID approach for the detection of protein–DNA interactions²⁶ and was recently extended to protein–protein interactions using a promiscuous biotin ligase, in a technology called BioID,^{39–42} and fusions to a peroxidase enzyme, in a technology named APEX.⁴³ However, the chemical nature of our approach has features that differentiate it from other strategies.

First, the chemical flexibility of IPL probes allowed us to detect not only protein–protein interactions but also protein–RNA interactions, which are garnering considerable interest, especially in the field of chromatin function and gene regulation.^{7,8,29} Furthermore, although currently untested, IPL may also have the potential to identify protein–DNA interactions, further expanding its utility.

Second, by using a brief UV irradiation to activate the labeling chemistry, IPL allows for very fine time resolution, resulting in a snapshot of macromolecular interactions as they exist in the cell at a particular time point. In contrast, DamID deposits methyl groups on DNA continuously from the time the fusion protein is produced to when the cells are harvested.²⁶ BioID allows for some degree of time resolution by manipulating the intracellular levels of biotin available for the reaction, but labeling still occurs continuously during the 6–24 h required to achieve optimal biotin concentrations.⁴¹ By using UV irradiation, IPL uncouples the time required for the probe to reach equilibrium inside the cell (2 h at 37 °C) from the

labeling phase, which remains exceedingly short. Thus, only proteins and RNAs that are in the vicinity of the POI for the 5 min of UV irradiation are labeled and detected as interactors, limiting the number of false positives. The peroxidase-based technique, APEX, also offers a tight temporal resolution but not the same degree of spatial resolution as that of IPL, and it has not been shown to label RNA.⁴³

Third, the chemical nature of the IPL approach and the modularity of potential IPL probes will allow further optimization of both the photolabeling reaction to preferentially target certain macromolecular species and the targeting step by exploring other high-affinity and high-specificity interactions between small protein tags and small organic molecules. We have begun exploring these additional possibilities by developing second-generation probes that contain click chemistry handles, facilitating subsequent *in vitro* manipulations.

Fourth, the use of a chemically synthesized molecular probe allows for a high degree of spatial resolution, as the distance of labeled interactors from the POI can be easily controlled by changing the size of the linker that connects the photoactivatable group to the affinity tag. This fine tuning is not possible with BioID or APEX given that they rely on the free diffusion of activated biotin from the catalytic site of the enzyme tag.^{41,43}

To our knowledge, IPL is the first proximity-based trans-labeling approach that can identify both protein–protein and protein–RNA interactions. IPL-compatible probes are commercially available, and the approach can be implemented in any biological laboratory. Thus, IPL offers unmatched flexibility, temporal resolution, and spatial control, which make it useful to identify protein–protein and protein–RNA interactions that remain undetectable by conventional means.

■ ASSOCIATED CONTENT

● Supporting Information

Figure S1: Effect of bio-ASA on 293T-REx proliferation. Figure S2: Comparison of photoactivatable groups for IPL. Figure S3: Optimization of probe and detergent concentration for PRC2 IPL. Figure S4: Self-labeling of mSA–SNRNP70. Table S1: Polypeptides at least 3-fold enriched by IPL-MS for EZH2. Table S2: List of proteins identified by conventional IPL-MS for EZH2 (first replicate). Table S3: List of proteins identified by conventional IPL-MS for EZH2 (second replicate). Table S4: List of proteins identified by conventional IPL-MS for EZH2 (third replicate). Table S5: Full results of protein identification by conventional MS. Table S6: Top 20 proteins with highest mSA–EZH2 vs GAL4–EZH2 ratio by IPL-SILAC. Table S7: Full results of protein identification by SILAC. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rbon@mail.med.upenn.edu. Phone: 215-573-2598.

Present Addresses

[#]Institute of Organic Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057, Zurich, Switzerland.

[○]Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, The Netherlands.

Author Contributions

[†]D.B.B. and V.N. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful to Danny Reinberg for his support and advice, Adrian Carcamo for help with generating constructs, Tim Christopher for technical assistance, the NYU Genome Technology Center for next-generation sequencing, and Emilio Lecona for critical reading of the manuscript. B.A.G. gratefully acknowledges support from an NIH Director's New Innovator award (DP2OD007447). Work performed in the Reinberg lab was supported by the National Institutes of Health (GM-64844 and R37-37120 to Danny Reinberg) and the Howard Hughes Medical Institute.

■ REFERENCES

- (1) Zhou, V. W.; Goren, A.; Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **2011**, *12*, 7–18.
- (2) Zhao, J.; et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **2010**, *40*, 939–953.
- (3) Bonasio, R.; et al. Interactions with RNA direct the Polycomb group protein SCML2 to chromatin where it represses target genes. *eLife* **2014**, *3*, e02637.
- (4) Gingras, A.-C.; Aebersold, R.; Raught, B. Advances in protein complex analysis using mass spectrometry. *J. Physiol.* **2005**, *S63*, 11–21.
- (5) Berezney, R.; Coffey, D. S. Nuclear matrix. Isolation and characterization of a framework structure from rat liver nuclei. *J. Cell Biol.* **1977**, *73*, 616–637.
- (6) Mili, S.; Steitz, J. A. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **2004**, *10*, 1692–1694.
- (7) Guttman, M.; Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **2012**, *482*, 339–346.
- (8) Wang, K. C.; Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **2011**, *43*, 904–914.
- (9) Khalil, A. M.; et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11667–11672.
- (10) Guttman, M.; et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **2011**, *477*, 295–300.
- (11) Wu, S.-C.; Wong, S.-L. Engineering soluble monomeric streptavidin with reversible biotin binding capability. *J. Biol. Chem.* **2005**, *280*, 23225–23231.
- (12) Gao, Z.; et al. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* **2012**, *45*, 344–356.
- (13) Sarma, K.; Margueron, R.; Ivanov, A.; Pirrotta, V.; Reinberg, D. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation *in vivo*. *Mol. Cell Biol.* **2008**, *28*, 2718–2731.
- (14) Yuan, Z. F.; et al. pParse: a method for accurate determination of monoisotopic peaks in high-resolution mass spectra. *Proteomics* **2012**, *12*, 226–235.
- (15) Fu, Y.; et al. Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry. *Bioinformatics* **2004**, *20*, 1948–1954.
- (16) Wang, L. H.; et al. pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2985–2991.
- (17) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4*, 207–214.

- (18) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372.
- (19) Cox, J.; et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **2011**, *10*, 1794–1805.
- (20) Margueron, R.; et al. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* **2008**, *32*, 503–518.
- (21) Lecona, E.; et al. Polycomb protein SCML2 regulates the cell cycle by binding and modulating CDK/CYCLIN/p21 complexes. *PLoS Biol.* **2013**, *11*, e1001737.
- (22) Parkhomchuk, D.; et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* **2009**, *37*, e123.
- (23) Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25.
- (24) Li, H.; et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079.
- (25) Wang, L.; Feng, Z.; Wang, X.; Zhang, X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **2010**, *26*, 136–138.
- (26) van Steensel, B.; Delrow, J.; Henikoff, S. Chromatin profiling using targeted DNA adenine methyltransferase. *Nat. Genet.* **2001**, *27*, 304–308.
- (27) James, V. S. Aryl azide photolabels in biochemistry. *Trends Biochem. Sci.* **1980**, *5*, 320–322.
- (28) Margueron, R.; Reinberg, D. The polycomb complex PRC2 and its mark in life. *Nature* **2011**, *469*, 343–349.
- (29) Bonasio, R.; Tu, S.; Reinberg, D. Molecular signals of epigenetic states. *Science* **2010**, *330*, 612–616.
- (30) Davidovich, C.; Zheng, L.; Goodrich, K. J.; Cech, T. R. Promiscuous RNA binding by polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1250–1257.
- (31) Kaneko, S.; Son, J.; Shen, S. S.; Reinberg, D.; Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1258–1264.
- (32) Jensen, K. B.; Darnell, R. B. CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. *Methods Mol. Biol.* **2008**, *488*, 85–98.
- (33) Hafner, M. PAR-CLIP—a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J. Vis. Exp.* **2010**, e2034.
- (34) Pomeranz Krummel, D. A.; Oubridge, C.; Leung, A. K.; Li, J.; Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **2009**, *458*, 475–480.
- (35) Pederson, T. Half a century of “the nuclear matrix”. *Mol. Biol. Cell* **2000**, *11*, 799–805.
- (36) Dignam, J. D.; Lebovitz, R. M.; Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **1983**, *11*, 1475–1489.
- (37) Lalonde, S.; et al. Molecular and cellular approaches for the detection of protein–protein interactions: latest techniques and current limitations. *Plant J.* **2008**, *53*, 610–635.
- (38) Koziol, M. J.; Rinn, J. L. RNA traffic control of chromatin complexes. *Curr. Opin. Genet. Dev.* **2010**, *20*, 142–148.
- (39) Van Itallie, C. M.; et al. Biotin ligase tagging identifies proteins proximal to E-cadherin, including lipoma preferred partner, a regulator of epithelial cell–cell and cell–substrate adhesion. *J. Cell Sci.* **2014**, *127*, 885–895.
- (40) Van Itallie, C. M.; et al. The N and C termini of ZO-1 are surrounded by distinct proteins and functional protein networks. *J. Biol. Chem.* **2013**, *288*, 13775–13788.
- (41) Roux, K. J.; Kim, D. L.; Raida, M.; Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **2012**, *196*, 801–810.
- (42) Morriswood, B.; et al. Novel bilobe components in *Trypanosoma brucei* identified using proximity-dependent biotinylation. *Eukaryotic Cell* **2013**, *12*, 356–367.
- (43) Rhee, H. W.; et al. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **2013**, *339*, 1328–1331.