

# AI-assisted protein design to rapidly convert antibody sequences to intrabodies targeting diverse peptides and histone modifications

Gabriel Galindo<sup>1\*</sup>, Daiki Maejima<sup>2\*</sup>, Jacob DeRoo<sup>3</sup>, Scott R. Burlingham<sup>1</sup>, Gretchen Fixen<sup>1</sup>, Tatsuya Morisaki<sup>1</sup>, Hallie P. Febvre<sup>1</sup>, Ryan Hasbrook<sup>1</sup>, Ning Zhao<sup>4</sup>, Soham Ghosh<sup>3,6</sup>, E. Handly Mayton<sup>7</sup>, Christopher D. Snow<sup>1,3,5</sup>, Brian J. Geiss<sup>3,7</sup>, Yasuyuki Ohkawa<sup>8</sup>, Yuko Sato<sup>8,9</sup>, Hiroshi Kimura<sup>2,9,10#</sup>, and Timothy J. Stasevich<sup>1,3,5,10#</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO, USA

<sup>2</sup>School of Life Science and Technology, Institute of Science Tokyo, Yokohama, Japan

<sup>3</sup>School of Biomedical Engineering, Colorado State University, Fort Collins, CO, USA

<sup>4</sup>Department of Biochemistry and Molecular Genetics, University of Colorado-Anschutz Medical Campus, Aurora, CO, USA

<sup>5</sup>Department of Chemical & Biological Engineering, Colorado State University, Fort Collins, CO, USA

<sup>6</sup>Department of Mechanical Engineering, Colorado State University, Fort Collins, CO, USA

<sup>7</sup>Department of Microbiology, Immunology, & Pathology, Colorado State University, Fort Collins, CO, USA

<sup>8</sup>Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan

<sup>9</sup>Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan

<sup>10</sup>Cell Biology Center, Institute of Integrated Research, Institute of Science Tokyo, Yokohama, Japan

\*Contributed Equally

#Corresponding Authors

## ABSTRACT

Intrabodies are engineered antibodies that function inside living cells, enabling therapeutic, diagnostic, and imaging applications. While powerful, their development has been hindered by challenges associated with their folding, solubility, and stability in the reduced intracellular environment. Here, we present an AI-driven pipeline integrating AlphaFold2, ProteinMPNN, and live-cell screening to optimize antibody framework regions while preserving epitope-binding complementarity-determining regions. Using this approach, we successfully converted 19 out of 26 antibody sequences into functional single-chain variable fragment (scFv) intrabodies, including a panel targeting diverse histone modifications for real-time imaging of chromatin dynamics and gene regulation. Notably, 18 of these 19 sequences had failed to convert using the standard approach, demonstrating the unique effectiveness of our method. As antibody sequence databases expand, our method will accelerate intrabody design, making their development easier, more cost-effective, and broadly accessible for biological research.

## INTRODUCTION

Intrabodies are a special class of antibodies that are engineered to fold and function within the intracellular environment<sup>1,2</sup>. This property gives intrabodies great therapeutic and diagnostic potential because they can bind target biomolecules deep within living cells and tissue. Furthermore, when intrabodies are tagged with fluorescent proteins, they become powerful imaging reagents that can bind and light up challenging targets that cannot be imaged with standard genetic fusion tags like GFP. For example, fluorescent intrabodies have been used to image conformational<sup>3–6</sup> and linear epitopes<sup>7,8</sup>, macromolecular complexes<sup>9,10</sup>, and chemical modifications<sup>11–13</sup>.

Despite the broad applicability of intrabodies, engineering them has proven to be a significant challenge. The antibody scaffolds from which intrabodies are derived have evolved to function in extracellular environments like the bloodstream, where conditions support the formation of disulfide bonds that are critical for maintaining structure. In contrast, the reduced intracellular environment disrupts these bonds, often leading to issues such as misfolding, aggregation, and degradation<sup>14–16</sup>. Additionally, the intracellular environment features numerous differences from the extracellular environment in terms of off-target binding and aggregation. For these reasons, the straightforward placement of antibody domains into an intrabody format – whether it be a single-chain variable fragment<sup>17</sup> (scFv) or a nanobody<sup>2,18–20</sup> – often results in a non-functional protein. Thus, despite the wide availability of antibodies commercially and in protein databases, there are relatively few functional intrabodies<sup>21,22</sup>.

One specific area of research that could benefit from additional intrabodies is the study of histone modifications<sup>11</sup>. These reversible chemical marks – which include acetylation, methylation, and phosphorylation, among others – decorate the unstructured tails of nucleosomes across the genome. The modifications are believed to work both individually and in combination to help shape chromatin architecture and regulate gene expression<sup>23,24</sup>. As such, they are strongly correlated with human health and disease, making them promising therapeutic targets<sup>25,26</sup>. However, because the marks are reversible, they are difficult to image in living cells using standard genetic tags<sup>27,28</sup>. Instead, specialized imaging probes like intrabodies are needed that can dynamically bind and light up the modifications as they come and go. Although there has been some progress developing such probes<sup>27,28</sup>, the pace has been slow due to the inherent engineering challenges. For this reason, the dynamics of histone modifications have remained elusive, as has their precise roles in gene regulatory networks.

Recent advances in artificial intelligence, including tools like AlphaFold<sup>29</sup>, RFdiffusion<sup>30</sup>, and protein large language models<sup>31</sup>, offer new opportunities to overcome these challenges. These technologies not only enable the *de novo* creation of proteins with

novel functions<sup>32–34</sup>, but also provide a powerful new platform to improve existing proteins like antibodies. While numerous efforts have focused on optimizing the complementarity determining regions of antibodies to improve target binding affinity<sup>35–37</sup> or the optimization of therapeutic antibody sequences to minimize self association and off-target binding under standard formulation conditions<sup>38,39</sup>, it remains unclear if these technologies can similarly enhance the framework regions of antibodies to improve their functionality when expressed as intrabodies in living cells.

To address this issue, we here develop an AI-driven approach to accelerate the development of functional intrabodies. Using a novel pipeline that leverages antibody domain annotation<sup>40</sup>, AlphaFold2<sup>29,41</sup>, ProteinMPNN<sup>42</sup>, and live-cell screening, we demonstrate how existing antibody sequences can be rapidly converted into functional intrabodies. As a proof-of-principle, we first develop two peptide-binding intrabodies, one targeting the synthetic FLAG tag<sup>43</sup> and the other targeting a linear epitope of the SARS-CoV-2 nucleocapsid protein<sup>44</sup>. We then turn our attention to endogenous histone modifications. To highlight the speed and practicality of our approach, we create a panel of intrabodies that bind diverse residue-specific chemical modifications on both histones H3 and H4, including H3T3ph, H3K4me1, H3K4me2, H3K4me3, H3K9(un), H3S10ph, H3K14ac, H3K27me1, H3K27me2, H3K27ac, H3K36me3, H4K5ac, H4K8ac, H4K12ac, and H4K16ac, tripling the number of intrabodies available for this purpose<sup>27</sup>. We demonstrate the functionality of these intrabodies in living cells and provide their sequences so they can be further refined or modified to image or manipulate histone modifications *in vivo*. Furthermore, we provide open-source code implementing our pipeline, and provide useful metrics and tables to help predict which candidate designs will retain functionality inside cells. As the number of published antibody sequences continues to rapidly grow, we believe our approach will significantly increase the number, utility, and accessibility of intrabodies to scientists in the life sciences.

## RESULTS

### Developing a pipeline for transforming existing antibody sequences into enhanced scFvs for intracellular applications

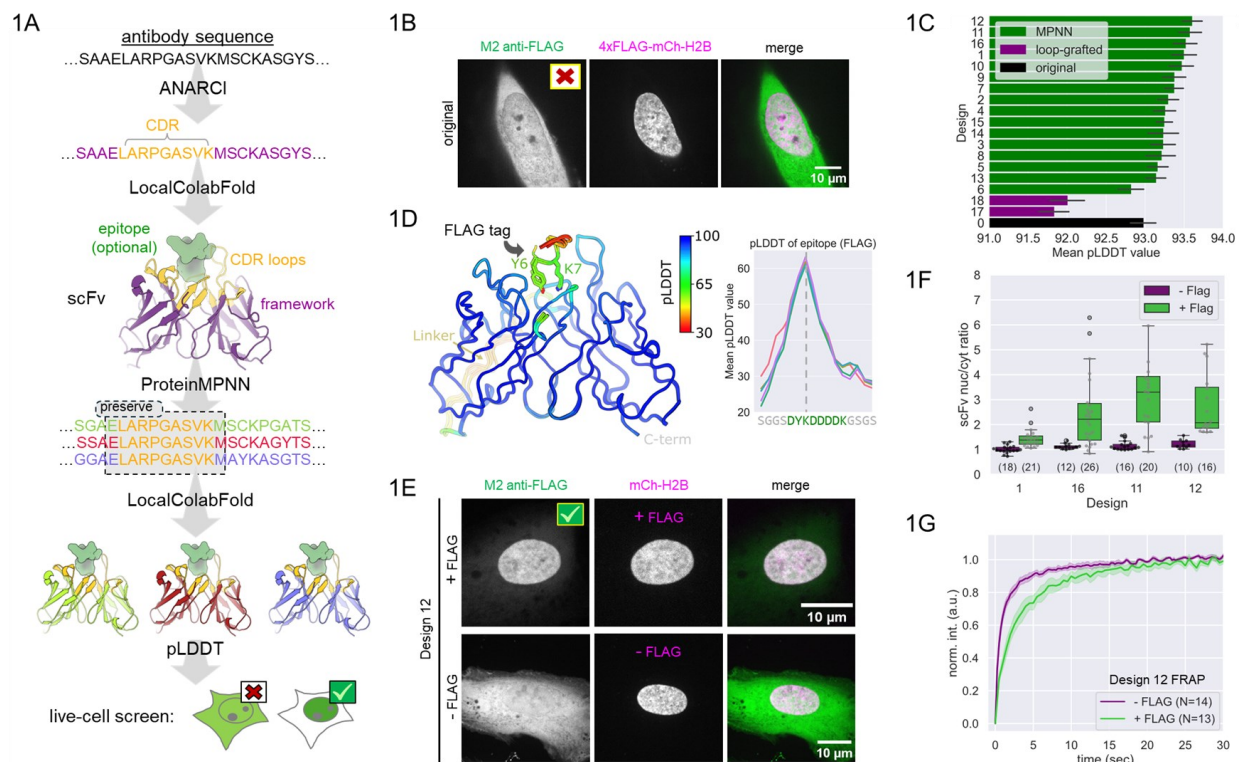
We developed a pipeline to explore the potential of AI-based protein design tools to enhance scFv engineering (Fig. 1A). We were driven by failures to produce a functional scFv based on the commercial M2 anti-FLAG antibody, whose peptide sequence had recently been published<sup>45</sup>. Initially, we reformatted the M2 sequence into a standard scFv with a flexible glycine-serine-rich linker. However, when fused to GFP and co-expressed with FLAG-tagged histone H2B, the scFv failed to colocalize with its target, instead showing uniform cellular distribution – a clear indication of it remaining unbound and freely

diffusing (Fig. 1B). We attempted to fix this problem by loop grafting the M2 complementary determining regions (CDRs) onto a stable scFv framework region that we had previously used to generate anti-HA<sup>7</sup> and anti-FLAG<sup>46</sup> ‘frankenbodies’. We also tested a new scFv framework region derived from an anti-Pol II scFv<sup>47</sup>, selected due to its high sequence similarity to the M2 frameworks. Unfortunately, none of these variants achieved specific colocalization with FLAG-tagged H2B (Fig. S1A).

Given the widespread use of the M2 antibody, we were convinced our scFv failed in living cells due to misfolding or insolubility, rather than a failure in antigen binding and specificity. We hypothesized that if we could design more soluble scFv framework regions based on the original sequence, then this would enhance their compatibility with the original CDRs, thereby increasing the likelihood of successful loop grafting. To test this hypothesis, we created a computational pipeline that integrates several AI-based tools. The pipeline begins by annotating the antibody sequence to extract the CDRs and framework regions using ANARCI<sup>40</sup>. Second, we predict the 3D structure of the scFv:target complex using LocalColabFold<sup>41</sup>, which combines AlphaFold2<sup>29</sup> with the fast multiple sequence generation method MMseqs2<sup>48,49</sup>. Third, we input the predicted structure into ProteinMPNN<sup>42</sup>, holding regions near the annotated CDRs fixed. For the remaining residues, ProteinMPNN attempts to predict favorable new amino acid selections. Crucially, the resulting ProteinMPNN framework sequence differs from the original, but tends to encode a protein that is more well behaved because the underlying neural network has been trained on well-behaved, soluble proteins with solved crystal structures<sup>42,50</sup>. We hypothesized that these new framework sequences would be preferable with respect to combating off-target aggregation of the designed proteins, while maintaining the critical 3D structure needed for compatibility with the original CDRs.

Running the M2 anti-FLAG scFv through the pipeline resulted in 15 designs. Compared to the original M2 anti-FLAG scFv and the two unsuccessful loop-grafted versions, the new designs exhibited considerably improved pLDDT (predicted local distance difference) scores (Fig. 1C). AlphaFold2 predicted 3D structures of these designs closely matched the original model (Fig. 1D, left). While the per-residue confidence (pLDDT) was significantly lower for the FLAG epitope than the scFv (Fig. 1D) (excluding the linker), we noted that the confidence in the authentic epitope was nonetheless higher than the confidence for flanking residues, hinting at latent detection of binding specificity by AlphaFold2. Notably, the pLDDT scores for the FLAG epitope were significantly higher than those for adjacent residues (Fig. 1D, right), even when the FLAG epitope was challenged with a nearby HA epitope (Fig. S1B), indicating binding specificity. Furthermore, the predicted orientation of the FLAG epitope matched that of a recently determined structure based on the full-length parental M2 antibody (8RMO)<sup>51</sup>, despite being predicted prior to the deposition of the structure in the Protein Data Bank (Fig. S1C).

We next screened the top four designs in living cells, selected based on pLDDT scores (designs 12, 11, 16, and 1 in Fig. 1C). Remarkably, all four candidates successfully colocalized with FLAG-H2B in the nucleus, as quantified by the ratio of nuclear to cytoplasmic fluorescence (Fig. 1E,F). The specificity of the best scFv (design 12) was further validated in cells without FLAG tags, where the scFv displayed homogenous localization, indicating minimal off-target binding (Fig. 1E, bottom; Fig. 1F, purple). Consistently, FRAP on design 12 was significantly slower in the presence of FLAG-H2B (Fig. 1G), with a recovery half-time of ~2 seconds. The timescale of this recovery is similar to FRAPs we performed on monovalent antibodies (Fab) generated from full-length antibodies with affinities in the 10 nM range<sup>52</sup>. Consistent with this comparison, the affinity of the full-length M2 anti-FLAG antibody was recently estimated to be ~2 nM<sup>51</sup>. These data therefore suggest the CDRs of our M2 scFvs remain intact despite the large number of changes we made to their framework regions. This underscores the success of our pipeline in generating functionally improved scFvs that maintain specific antigen recognition.



**Fig 1. Pipeline to convert antibody sequences into functional intrabodies.** **A.** Schematic showing the pipeline, from antibody sequence to live-cell screening. **B.** The original M2 anti-FLAG scFv (green) does not colocalize with 4xFLAG-mCh-H2B (magenta) in the nucleus. **C.** The ranked pLDDT scores of all M2 anti-FLAG scFv designs (includes scFv and FLAG epitope), including the original (black bar), two loop-grafted variants (purple), and 16 from the pipeline (green). Error bars show the standard deviation from LocalColabFold predicted models (N=5 each). **D.** The top four predicted structures (models 1, 16, 11, and 12) overlaid with the original M2 anti-FLAG structure (model 0), colored by pLDDT. Right, the corresponding



overlaid pLDDTs of the tested binder, with the FLAG tag (DYKDDDDK) indicated in green text. **E.** Sample cell from the live-cell screen of design 12. This scFv (green) colocalizes with the FLAG-tagged H2B (magenta). When there is no FLAG-tagged protein co-expressed, the scFv shows uniform localization (bottom). **F.** Live-cell measurement of the nuclear to cytoplasmic ratio of the indicated scFv designs (number of cells (N) indicated; box is 25-75% interquartile range; whiskers mark points within 1.5× the interquartile range). **G.** FRAP experiments in the nucleus of cells expressing design 12 scFv in the presence and absence of FLAG-H2B. The number of cells (N) is indicated; the shaded region shows the 95% CI.

## Using the pipeline to create an intrabody against the SARS-CoV-2 nucleocapsid protein

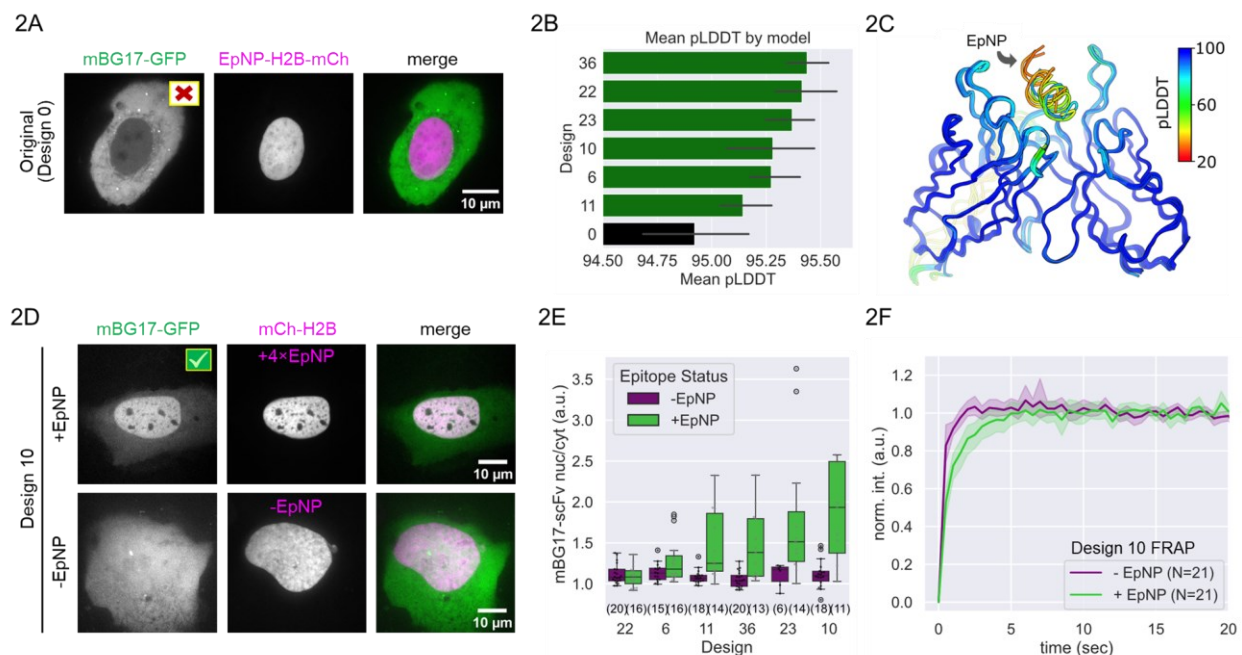
To determine if our technique could be generalized, we next targeted a medically relevant antigen: the nucleocapsid protein (NP) of SARS-CoV-2. The NP is a promising diagnostic target due to its early and abundant expression during viral infection<sup>53</sup>. We had previously developed and sequenced an anti-nucleocapsid mouse monoclonal antibody (mBG17), identifying its epitope as a short 10-amino acid peptide (EpNP: DDFSKQLQQS) that forms an alpha-helical structure<sup>44,54</sup>. Similar to our experiences with the M2 anti-FLAG scFv, when we attempted to convert this sequence into an intracellularly functional scFv or frankenbody, we were unsuccessful (Fig. S2A). These previous failures made the anti-EpNP scFv an ideal candidate to further test our computational pipeline. Furthermore, unlike the FLAG epitope and M2 anti-FLAG antibody, which are well-characterized and supported by crystal structures available in the Protein Data Bank, the mBG17 antibody lacks structural information, offering an unbiased test case for evaluating the predictions of our pipeline.

Mirroring our work with the M2 anti-FLAG antibody sequence, we generated 45 new scFv sequences targeting the NP epitope. All had higher pLDDT scores than the original, with some designs better than others (Fig. S2B). To assess how predictive these pLDDT scores were for live-cell functionality, we focused on six designs across the pLDDT score range (Fig. 2B). The predicted 3D structures for these six designs showed consistent placement of the epitope within the CDR binding pocket (Fig. 2C), with the central portion having especially high pLDDT values (Fig. 2C). This aligned well with our earlier work<sup>44</sup> that identified the key residues within the NP epitope that are critical for binding to the antibody, bolstering our confidence in the predicted structures.

To screen individual designs, we co-transfected each in cells together with a plasmid encoding a 4×EpNP-tagged histone H2B. Three of the six tested anti-EpNP scFv designs successfully colocalized with the EpNP-H2B in the nucleus (Figs. 2D,E). Interestingly, the successful candidates were not those with the highest pLDDT scores, suggesting that high scores alone might not always predict solubility and functional success. To verify specific binding, we expressed the successful scFv in cells lacking H2B-EpNP. In these cells, we observed a uniform distribution, indicating that the scFv was freely diffusing and

unbound, with no detectable off-target binding (Fig. 2D, lower and Fig. 2E, purple). Fluorescence recovery after photobleaching (FRAP) further confirmed this specificity, showing significantly slower recovery rates in cells containing H2B-EpNP compared to those without it (Fig. 2F).

In our analysis of the sequence alignment of the screened scFv variants, we identified a feature that correlated with success: residue 80 was K in more effective designs versus S or N in less effective designs (Fig. S2C,D). This residue is situated in a beta hairpin turn outside the CDR loops, and could play a role in stabilizing CDR conformations (Fig. S2D). Intriguingly, our previous studies that utilized random mutagenesis on different scFvs also identified residue 80 as crucial for scFv solubility<sup>55</sup>. However, the presence of K80 alone is not sufficient for enhanced performance, as the original sequence also contained K80. This suggests that the improved solubility and functionality likely result from modifications to multiple residues. Nonetheless, these observations highlight the effectiveness of our strategy in rapidly generating functional scFv sequences and also demonstrates how a comparison of different designs can be used to quickly generate hypotheses about key residues.



**Fig. 2. Engineering a functional intrabody against the nucleocapsid protein from SARS-CoV-2. A.** An scFv (mBG17-GFP; green) based on the sequence of a wildtype antibody against the SARS-CoV-2 nucleocapsid protein does not colocalize with its target epitope (EpNP-H2B-mCh, magenta) in living cells. **B.** The ranked pLDDT scores of a selection of mBG17 pipeline designs (includes both scFv and EpNP; green bars) along with the original (black bar). Error bars show the standard deviation from LocalColabFold predicted models (N=5 each). **C.** The predicted structures (models 11, 6, 10, 23, 22, and 36) overlaid, colored by pLDDT. **D.** Design 10 (green) colocalizes with its target epitope (top; +EpNP = EpNP-mCh-H2B; magenta) and displays uniform fluorescence throughout cells in the absence of its target epitope (bottom; -EpNP = mCh-HB; magenta). **E.** Live-cell measurement of the nuclear to cytoplasmic ratio of the indicated

scFv designs (number of cells (N) indicated; box is 25-75% interquartile range; whiskers mark points within 1.5X the interquartile range). **F.** FRAP experiments in the nucleus of cells expressing design 10 mBG17 scFv in the presence (+EpNP = EpNP-mCh-H2B) and absence (-EpNP = mCh-H2B) of EpNP. The number of cells (N) is indicated; the shaded region shows the 95% CI.

## Developing intrabodies targeting post-translational histone H3 modifications

Encouraged by these successes, we next set out to develop intrabodies that target endogenous post-translational histone modifications. We refer to such modification-specific intracellular antibodies as 'mintbodies'<sup>13,27</sup>. Mintbodies dynamically bind their target modifications, enabling real-time imaging and tracking in living cells without affecting cell divisions and animal development<sup>27,56</sup>.

Since 2013, five mintbodies against histone and RNA polymerase II modifications have been developed<sup>13,47,55,57,58</sup>. The first two were selected from several hybridoma clones<sup>13,57</sup> and the other three involved random and/or site-directed mutagenesis with extensive screening<sup>47,55,58</sup>. In the latter cases, the original scFv were only a few mutations away from being functional in the intracellular environment, suggesting they may have been fortuitously close to working from the start. Based on our experience with anti-EpNP, we suspect other scFvs may require many mutations before they can properly fold and function inside cells, in which case random mutagenesis becomes impractical due to the immense combinatorial possibilities. We therefore hoped our computational pipeline could significantly reduce the search space and expedite the production of useful mintbodies.

To test this, we decided to focus on a set of mouse monoclonal antibodies targeting three distinct modifications to histone H3: Lysine 27 acetylation (anti-H3K27ac, clone 9E2H10/CMA309)<sup>59</sup>, Serine 10 phosphorylation (anti-H3S10ph; clone 7G1G7/CMA311)<sup>60</sup>, and Lysine 9 trimethylation (anti-H3K9me3; clone 2F3/CMA318)<sup>61</sup>. These antibodies are excellent candidates for our pipeline because – despite being sequenced and well characterized – each fails to bind its target modification in living cells when converted into an scFv format (Fig. 3A). Furthermore, the structures of these antibodies have not been experimentally determined, so our pipeline predictions should be free from any bias that could arise from AlphaFold having prior access to the data during its training.

For each antibody sequence, we generated 15 candidate scFvs with our pipeline and chose the top four or five by pLDDT to screen in living cells. To see if modeling the target epitope was critical, we tested a couple of strategies. For H3K27ac, we inputted the scFv along with the target peptide (the H3 N-terminal tail with glutamine at position 27 to mimic acetylation), while for H3S10ph and H3K9me3, we just input the scFv sequence without



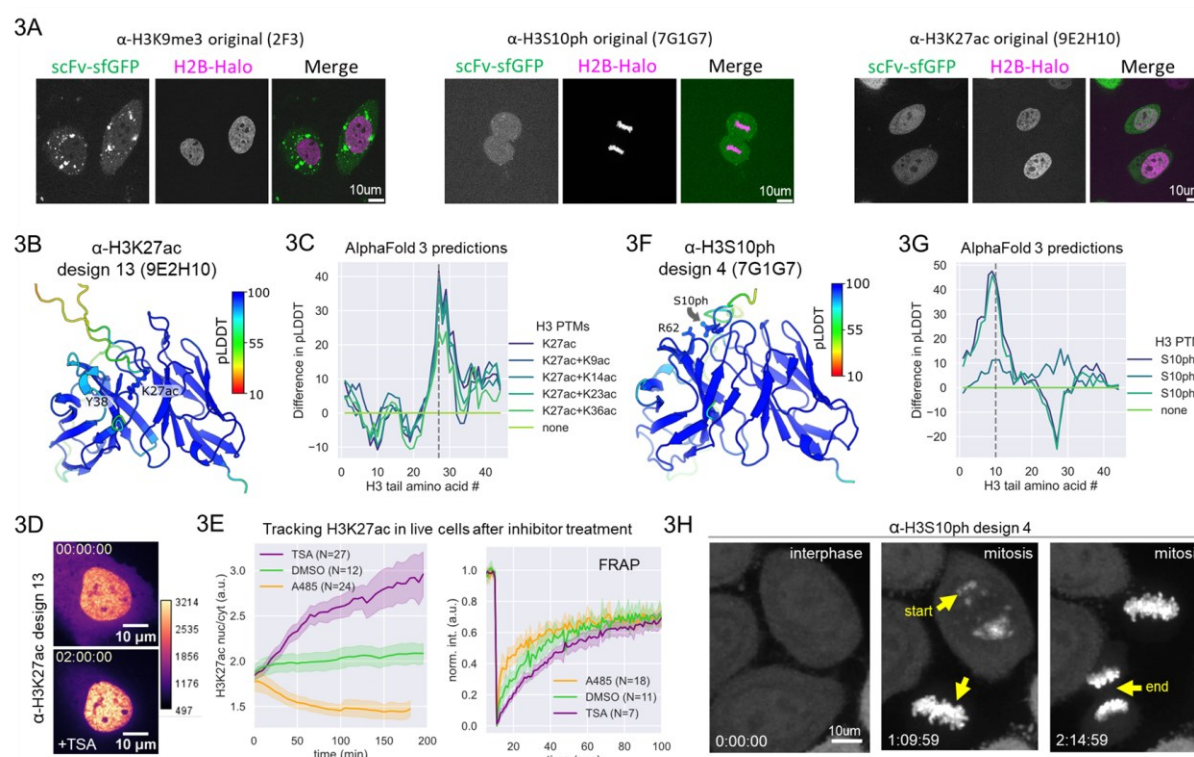
the target modified peptide. Although we could have used AlphaFold3<sup>62</sup> to more properly model post-translational modifications, we chose AlphaFold2 because at the time it was more open and could be used without restriction.

Of the four designed anti-H3K27ac mintbodies we screened, one localized in the nucleus of U2OS cells, where histone acetylation is expected to be (Figs. S3A). Reassuringly, when this sequence was later input into AlphaFold3 together with its target H3 tail peptide (Fig. 3B), the model confidence (pLDDT) of the H3 peptide peaked at K27 when that residue was acetylated, even when competing acetylation marks were nearby (Fig. 3C). To more directly test specificity in an experimental setting, we tracked the cellular localization of the mintbody in response to the addition of the histone deacetylase inhibitor trichostatin A (TSA; to increase H3K27ac) and the p300 histone acetyltransferase inhibitor A485 (to decrease H3K27ac). As expected of a functional anti-H3K27ac intrabody, the ratio of nuclear to cytoplasmic signal increased in the presence of TSA (Fig. 3D and Fig. 3E, purple; Movie S1) and decreased in the presence of A485 (Fig. 3E, orange and Movie S2). As well, FRAP recoveries were faster in TSA than they were in control cells, and faster yet in A485-treated cells, indicating progressively fewer target binding sites (Fig. 3E, right). The FRAP recovery ( $t_{1/2} \sim 15$  sec) was also on par with similar measurements using purified fragmented antibodies (Fab) physically loaded into cells ( $t_{1/2} \sim 21$  sec)<sup>63</sup>, indicating the CDRs are largely intact in this mintbody.

Regarding anti-H3S10ph, four out of five scFv designs localized with mitotic chromatin, as would be expected (Fig. S3B). One design (design 4) had the best signal-to-noise. When we input this sequence into AlphaFold3 together with its target H3 tail peptide (Fig. 3F), the model confidence (pLDDT) of the H3 peptide peaked at S10 when that residue was phosphorylated, even when competing phosphorylation marks were nearby (Fig. 3G). This success suggested the epitope may not be necessary in the original model used to generate the mintbody designs. Time-lapse imaging of design 4 in living cells undergoing mitosis revealed the rapid dynamics of H3S10 phosphorylation and dephosphorylation (Fig. 3H). The anti-H3S10ph mintbody marked phosphorylation in patchy spots across the nucleus nearly five minutes before mitosis could be detected based on histone H2B localization (Movie S3), a result also seen using purified Fab physically loaded into cells<sup>60</sup>. Loss of phosphorylation was also rapid as cells approached telophase, with the condensed mitotic chromatin closest to the metaphase plate harboring phosphorylation longest.

Unfortunately, all designs we tested for the anti-H3K9me3 mintbody failed to properly localize within the nucleus (Fig. S3C), despite having high pLDDT scores according to AlphaFold2 and 3. Some designs were worse than others, however. In particular, design 1 could not be detected in cells at all, and design 3 showed characteristic aggregates that

are most likely indicative of misfolding. It remained unclear why this particular antibody suffered, although a clue may come from the parental antibody (2F3/CMA318), from which it was difficult to generate Fab from due to the presence of protease susceptible sites in the heavy chain<sup>52</sup>. It is possible these sites may also hinder scFv folding, stability, and overall performance.



**Fig. 3. Intrabodies targeting post-translational histone H3 modifications.** **A.** Three original scFvs (green) against histone H3 Lysine 9 trimethylation (H3K9me3; 2F3), Serine 10 phosphorylation (H3S10ph; 7G1G7), and Lysine 27 acetylation (H3K27ac; 9E2H10), respectively, do not colocalize with their target epitopes in the nucleus (H2B-Halo; magenta). **B.** Predicted AlphaFold3 structure of design 13 of the anti-H3K27ac scFv (9E2H10) bound to the H3 tail with Lysine 27 acetylated (color coded by pLDDT). **C.** The difference in predicted pLDDT scores between AlphaFold3 models with an H3 peptide harboring acetylation (as in B) versus those without. The first 44 amino-acids of the H3 tail are shown along the x-axis, with a dashed line marking K27ac. **D.** Image of a cell expressing design 13 of the anti-H3K27ac scFv pre- (top) and post-TSA (bottom). **E.** Left, the nuclear to cytoplasmic ratio of anti-H3K27ac scFv (design 13) in living cells post treatment with the deacetylation inhibitor TSA (purple), the vehicle control DMSO (green), or the p300 inhibitor A485 (orange). Nuclear FRAP recovery curves from cells, as in E (3 hrs post inhibitor treatment). **F.** Predicted AlphaFold3 structure of design 4 of the anti-H3S10ph scFv (7G1G7) bound to the H3 tail with Serine 10 phosphorylated (color coded by pLDDT). **G.** The difference in predicted pLDDT scores between AlphaFold3 models with an H3 peptide harboring phosphorylation (as in F) versus those without. The first 44 amino-acids of the H3 tail are shown along the x-axis, with a vertical dashed line marking S10ph. **H.** Snapshots from a movie of cells expressing the anti-H3S10ph scFv (design 4) during mitosis. Scale bars, 10  $\mu$ m.

## Developing a panel of intrabodies against diverse histone modifications

As a final demonstration of the practicality and speed of our approach, we applied our scFv pipeline to a full panel of histone modification-specific mouse monoclonal antibodies that were previously developed, sequenced, and validated for binding specificity and strength<sup>52,59,61,64–68</sup>. We were especially curious if there are specific histone modifications that are inherently difficult to target. Among the 19 antibody sequences we ran through the pipeline, 14 resulted in successful scFv, having at least one design that exhibited enrichment in nuclei or mitotic chromosomes (Figs. 4A and S4). For example, the original anti-H4K16ac scFv failed to localize to the nucleus, but 4 out of 5 scFv designs exhibited nuclear enrichment (Fig. 4B). Although the nuclear to cytoplasmic ratios were high for designs 2 and 5 (Fig. 4C), their expression levels – both the intensity and the number of expressing cells – were lower than design 3, as revealed by Western blotting (Fig. 4D). Other acetylation-specific antibodies (H3K14ac, H4K5ac, H4K8ac, and H4K12ac) were also vastly improved, exhibiting strong nuclear enrichment that was completely lacking in their original counterparts (Fig. S4A).

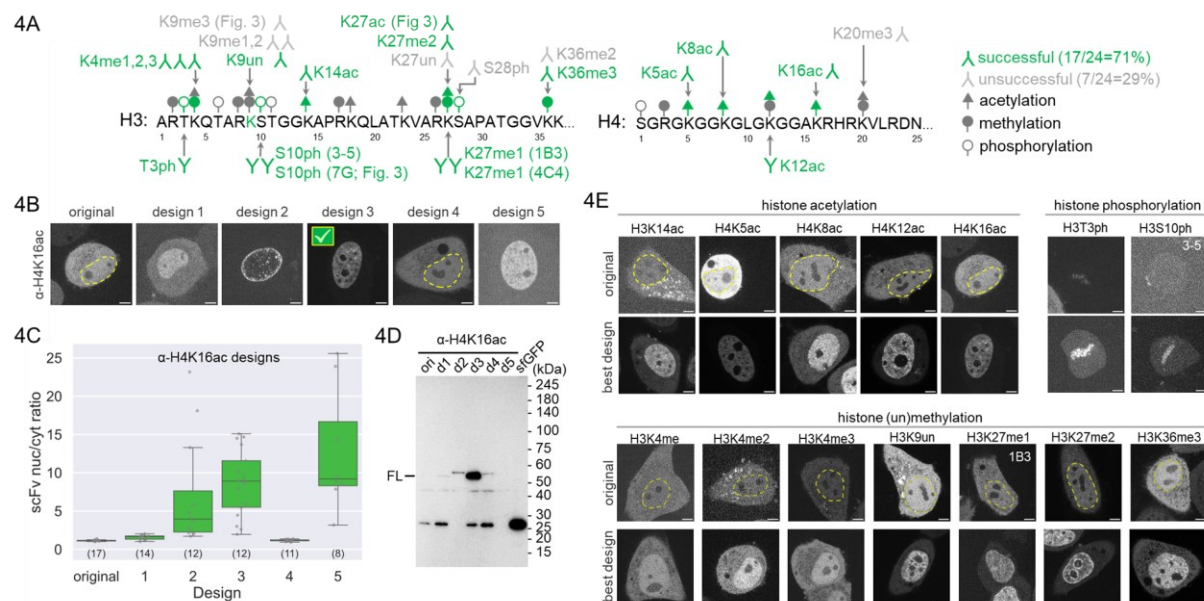
Among the 12 methylation-specific antibodies we tested, 8 localized to the nucleus, including three targeting monomethylation (H3K4me1 clone 19A5/CMA302 and H3K27me1 clones 4C4 and 1B3/CMA321), two targeting dimethylation (H3K4me2, clone 27A6/CMA303 and H3K27me2 clone 5A12/CMA322), and two targeting trimethylation (H3K4me3 clone 16H10/CMA304 and H3K36me3 clone 2C3/CMA333) (Fig. S4C). This suggests our inability to target H3K9me3 (Fig. S3E) was most likely due to chance rather than being indicative of an inherent problem targeting methylation more generally.

Of the two phospho-specific antibodies we ran through our pipeline, one (anti-H3S10ph clone 3-5/CMA313) clearly marked mitotic chromosomes (Fig. S4B). To further test our pipeline, we also examined an scFv that was functional from the start (anti-H3T3ph clone 16B3). In this case, all five pipeline designs localized to mitotic chromosomes like the original (Fig. S5A). Although these designs did not have superior signal-to-noise compared to the original (Fig. S5B), the fact that all five retained functionality highlights the robustness of our pipeline.

We hypothesized that the success of our designs stemmed from improved solubility and stability. To test this directly in a controlled setting, we expressed anti-H4K16ac mintbodies (1B2/CMA416) using an *in vitro* translation system. After centrifugation, the original mintbody largely precipitated, whereas the redesigned version was more evenly distributed between the supernatant and precipitate (Fig. S5C). Furthermore, nearly all of the original mintbody in the supernatant precipitated after heating to 60°C and 70°C, whereas a substantial fraction of the redesigned version remained soluble under the

same conditions (Fig. S5C). These results suggest that scFvs redesigned using our pipeline have greater thermal stability than their original counterparts.

Finally, we generated stable polyclonal cell lines for each of our 17 newly designed mintbodies in human (HeLa or U2OS) and mouse (NIH3T3) cells. The successful establishment of these cell lines indicates that the new mintbodies can be expressed at an appropriate concentration without overtly disrupting cellular viability. Mintbodies bind their targets reversibly and in a monovalent manner, which limits the duration of target engagement compared to full-length antibodies<sup>27</sup>. However, at higher concentrations, they may still affect endogenous interactions, depending on the experimental conditions.



**Fig. 4. A panel of intrabodies targeting diverse histone modifications.** **A.** 24 antibodies against histone H3 and H4 post-translational modifications were passed through the pipeline to generate candidate mintbodies (modification-specific intrabodies). Top-ranked (by pLDDT) candidates (4-5 per antibody) were tested in living HeLa cells. 17 of 24 attempts were successful (highlighted in green versus gray). **B.** Top, sample images of cells expressing candidate mintbodies against H4K16ac to compare which localize better to the nucleus (which is marked by a dashed yellow line when needed). Design 3 showed the best localization pattern. **C.** Sample live-cell measurement of the nuclear to cytoplasmic ratio of the anti-H4K16ac mintbody designs (number of cells (N) indicated; box is 25-75% interquartile range; whiskers mark points within 1.5× the interquartile range). **D.** A Western blot showing that design 3 (d3) of the anti-H4K16ac mintbody has the highest full-length (FL) expression level in cells. **E.** Sample live-cell images of each successful mintbody comparing the localization of the original mintbody (top rows) compared to the best design (bottom rows). The best designs localize better to the nucleus (which is marked by a dashed yellow line when needed) or, in the case of phosphorylation, to mitotic chromatin. n.d., mitotic cells not detected. Scale bars, 5 μm.

## DISCUSSION



In this study, we used AI-based tools to accelerate the development of functional intrabodies by modeling and optimizing their framework regions. Much effort has been made to evolve and design CDRs to obtain antigen binders with high specificity and affinity. In contrast, although the framework regions play a critical role in securing antibody structure and stability, efforts to develop a general strategy to improve their intracellular function have been limited<sup>14,69</sup>. By focusing on refining these framework regions while leaving the CDRs intact, we achieved a high success rate in converting existing antibody sequences into functional intrabodies.

Out of the 26 parental antibody sequences we tested, 19 were successfully converted into functional scFv intrabodies, yielding an overall success rate of approximately 73%. This success rate is significantly higher than the 5-10% success rate we've experienced when directly converting endogenous antibody sequences. Importantly, we began with designs that all initially failed as scFv intrabodies (aside from anti-H3T3ph), many of which also failed after standard loop grafting onto stable scaffolds having high sequence similarity, usually an indicator of grafting success<sup>7</sup>. Despite these difficult starting points, out of a total of 127 designs the pipeline generated, 51 functioned effectively in living cells, so the average probability that any single design worked was roughly 41%.

These statistics highlight the effectiveness of our pipeline to optimize scFv framework regions for intracellular expression. As a first application, we created intrabodies against the popular synthetic FLAG tag and an epitope within the SARS-CoV-2 nucleocapsid protein. We then focused on histone modifications, tripling the number of intrabodies available for tracking the dynamics of these elusive marks *in vivo*, a traditionally challenging task<sup>11,27</sup>. With these intrabody sequences now stored on plasmids, it is easier to distribute them through platforms like Addgene and to modify them, for example, to create bispecific antibodies that could be used to explore the combinatorics of the histone code hypothesis<sup>24,70,71</sup>. These applications will complement existing tools like GFP-tagging, further expanding our ability to probe complex biological processes.

The potential impact of our approach is substantial. With over 2,000 solved antibody structures available in the Structural Antibody Database<sup>72</sup> (SAbDab) and more than 147,000 antibody sequences from the Observed Antibody Space<sup>73</sup> (OAS), each represents a candidate for conversion into a functional intrabody. Importantly, our pipeline requires only the antibody sequence, not the structure, and can even work without inputting the target epitope sequence. This is ideal because the number of publicly available antibody sequences is rapidly growing, as researchers prioritize preserving these sequences to address the antibody crisis caused by variability in commercial antibody batches<sup>74,75</sup>. Our method offers a timely approach to take advantage of this expanding resource.



Interestingly, our results suggest that while AI-based tools like AlphaFold offer valuable structural predictions, they alone are insufficient for ensuring successful intrabody design. For example, high pLDDT scores, which indicate predicted structural confidence, did not consistently correlate with functional success in our live-cell screens. This finding highlights the importance of experimental validation and suggests a need for new metrics that better predict intrabody functionality. Notably, the failures in our pipeline were just as informative as the successes. By comparing successful and unsuccessful designs in anti-EpNP scFv, we identified a key residue within the antibody framework that had previously been shown to play a crucial role in stabilizing scFv structure in living cells<sup>55</sup>. In contrast, we observed that some antibodies, such as anti-H3K9me3 or anti-H3S28ph, consistently failed for reasons that remain unclear, whether due to issues with the parental antibody or other hidden factors. Despite these occasional failures, we believe our approach will be instrumental in mapping the key residues that stabilize scFv structures, thereby improving future intrabody designs.

While our work was being prepared, new methods such as AbMPNN<sup>76</sup>, IgMPNN<sup>37</sup>, and AntiFold<sup>35</sup> have emerged, building on ProteinMPNN<sup>42</sup> and AlphaFold2<sup>29</sup> to improve antibody structure prediction and design by incorporating antibody-specific data. Although these tools focus on optimizing CDRs to improve binding affinity<sup>37,76</sup> and also prevent self-binding and non-specific targeting<sup>38</sup>, the extent to which these tools can also improve the framework regions of intrabodies for functionality inside living cells remains unclear. Developing tools to optimize these regions will be useful, and comparing their predictions with our experimentally validated designs will offer insights into what drives intrabody success in living cells.

To facilitate these sorts of comparisons, we provide a comprehensive list of all sequences we designed and tested, ranked by visualization and quantification in living cells (Table S1). As more designs are generated and analyzed, this resource will grow, making it easier to distinguish successful from unsuccessful intrabody designs. Ultimately, we hope to train improved algorithms with this data, with the goal of developing a next-generation pipeline specifically trained on functional intrabodies to further reduce the need for labor-intensive live-cell screening.

## MATERIALS AND METHODS

### Antibody to Intrabody Pipeline

Our pipeline to convert antibody sequences into designed-intrabody sequences consists of five steps. First, we aligned and annotated the antibody heavy and light chain sequences, typically using Clustal<sup>77</sup> for alignment and ANARCI<sup>40</sup> or Benchling's built-in

antibody property prediction tool for CDR and framework region annotation. We sometimes compared predicted CDRs and chose a conservative set of residues encompassing all predictions. Second, based on the annotations from step one, we designed the scFv sequence, starting with the variable region of the heavy chain at the N-terminus (containing four framework regions and three CDRs), followed by a GS-rich linker and ending in the variable region of the light chain (also containing four framework regions and three CDRs). Third, a structure was predicted for this sequence with AlphaFold2<sup>29</sup> using ColabFold<sup>41</sup>. Fourth, this structure was used as input into ProteinMPNN<sup>42</sup>. During the ProteinMPNN sequence design the fixed positions included the CDRs as well as all residues in the scFv's framework regions that were within 3Å of the CDRs, along with the GS-rich linker. We experimented with fewer angstroms when designing M2 anti-FLAG scFvs, but this seemed to lower the success rate (e.g. we tested 0 Å with the M2 anti-FLAG antibody and this produced 0 successful intrabodies). By contrast, a value of 5Å left important residues we had previously identified in the framework regions unchanged (e.g., M83, which we previously discovered<sup>47</sup> was an indicator of non-functional scFv). The remaining residues form the framework of the scFv, to be redesigned by ProteinMPNN. The settings were: Number of sequences = 15, Sampling temperature = 0.1, Model = vanilla-v\_48\_020 or soluble-v\_48\_020, Backbone noise = 0.02. Fifth, we predicted the structure of all ProteinMPNN guided sequence designs using ColabFold and ranked these by pLDDT.

We provide python code on Github that implements our complete pipeline in python ([https://github.com/jbderoo/scFv\\_Pmpnn\\_AF2](https://github.com/jbderoo/scFv_Pmpnn_AF2)), using ANARCI<sup>40</sup> to annotate CDRs (using the Martin numbering scheme by default), LocalColabFold<sup>41</sup> to predict intrabody structures, and ProteinMPNN<sup>42</sup> to generate new framework sequences for each input.

## Plasmid Construction

The nucleotide sequence of histone modification-specific antibodies was determined as previously described<sup>78</sup>. AI-assisted scFv library designs were reverse translated into human codon optimized DNA sequences. The library scFv DNA sequences were flanked with sequences homologous to the PB533\_15F11-sfGFP vector (Addgene Plasmid #167526) to be inserted at the NheI/AgeI cut sites with isothermal assembly. scFv sequences were synthesized by Twist Bioscience. In some cases, scFv sequences were synthesized by IDT (eBlocks) and cloned into the PB533 vector (System Biosciences) harboring an sfGFP sequence<sup>57</sup> using the In-Fusion Cloning system (Takara Bio). Plasmids for all successful scFvs will be made available at Addgene upon acceptance of the manuscript.

## Stable cell line generation

Stable cell lines were generated using the PiggyBac transposon system. All scFv vectors contain PiggyBac LTRs with neomycin resistance for selection. Cells were co-transfected with Super PiggyBac transposase (NovoPro or Systems Biosciences) at 1:5 transposase to transposon in a 35-mm dish at 50% confluency. Once cells reached confluency, cells were expanded into 10-cm dishes and selected for a week with G418 750 µg/mL (or 1 mg/mL for HeLa). Individual colonies were hand picked for monoclonal cell lines or pooled together for polyclonal lines.

## Cell Culture

Human U2OS osteosarcoma or HeLa cells were grown in DMEM (ThermoFisher Scientific or Nacalai Tesque) supplemented with 10% v/v FBS (Atlas or ThermoFisher Scientific), 1% v/v Pen/Strep (ThermoFisher Scientific or Sigma), and 1 mM L-glutamine (ThermoFisher Scientific or Sigma) at 37°C in 5% CO<sub>2</sub>. Mouse fibroblast NIH3T3 cells were grown in DMEM (ATCC 30–2002) supplemented with 10% v/v FBS (EqualFETAL, Atlas Biologicals) and 1% v/v Pen/Strep (Gibco 15140122) at 37°C in 5% CO<sub>2</sub>. Trypsinization for cell passaging was performed using TrypLE Express (Thermo Fisher 12605010).

## Transfection

The day prior to imaging, cells were seeded on 35-mm MatTek Chambers or 24-well glass-bottom plates (AGC Techno Glass) at ~70% confluency and transfected. Transfection was performed using Lipofectamine LTX or 2000 (Thermo Fisher Scientific) with 0.8-1 µg total plasmid DNA. Cells were washed 3 times with DMEM the next morning. In experiments using HaloTag, staining was performed using 200 nM of JF646-HaloLigand for 30 min then washed 3 times with DMEM. Cells were then kept in supplemented DMEM lacking phenol red or FluoroBrite (ThermoFisher Scientific) containing 10% FBS for imaging.

For colocalization imaging experiments (Figs. 1B, 1E, S1A, 2A, 2D, and S2A), the cells were seeded on 35-mm MatTek chambers with 70% confluency. The next day, 1.25 µg of the scFv-GFP construct and 1.25 µg of the epitope-tagged mCh-H2B construct were transfected into cells using Lipofectamine LTX. The medium was changed to complete DMEM medium at 3 hours post transfection. The cells were imaged at 20 h post transfection. Cell treatment for anti-H3K27ac design 13 imaging experiments were performed as above except 2.0 µg of scFv-GFP and 1.0 µg of mCh-H2B constructs were used (Figs. 3D, 3E).

## Western blotting

For Western blotting analysis in Fig. 4C, cells were plated on 12-well plates (Corning, 3513) at  $1.0 \times 10^5$  cells/well and transfected with 1.6  $\mu$ g plasmid DNA using Lipofectamine 2000 (Thermo Fisher Scientific). The next day after transfection, cells were washed 3 times with phosphate-buffered saline (PBS) and lysed with 1X SDS-gel loading buffer (2% SDS, 10% glycerol, 100 mM Tris-HCl [pH 6.8], 100 mM dithiothreitol, 0.01% bromophenol blue; 150  $\mu$ L/well), before incubating at 95°C for 5 min.

For Fig. S5B, mintbodies were produced by an in vitro transcription-translation coupled system using PURE flex 2.0 (Gene Frontier). Approximately 90 ng of each template DNA in a 20  $\mu$ L reaction mixture was incubated for 6 h at 37°C. After centrifugation at 16,000  $\times g$  for 30 min at 4°C, supernatant was transferred to a new tube. After incubating supernatant fractions for 30 min at 50, 60 and 70°C, precipitated materials were removed by centrifugation as above. All fractions were added with an equal volume of 2X SDS sample buffer and heated for 10 min at 95 °C.

Protein samples were separated in polyacrylamide gels (SuperSep Ace, 5-20 %, 17 well; Fujifilm Wako) and transferred onto a 90 mm x 90 mm polyvinylidene difluoride membranes (FluoroTrans W; Fujifilm Wako) using a semi-dry system (Trans-Blot Turbo; Bio-Rad; 1.3 A for 15 min) with EzFastBlot (Atto) as a transfer buffer. Membranes were washed with TBST (20 mM Tris-HCl [pH 8.0], 150 mM NaCl, and 0.05% Tween 20), blocked for 20 min in Blocking-One (Nacalai Tesque) with gentle shaking, and washed with TBST. Membranes were then incubated with mouse anti-GFP antibody (GF200, Nacalai Tesque; 1:2,000 dilution) in Can-Get-Signal Solution 1 (Toyobo) for 1 h at room temperature, washed three times with TBST for 5 min each, incubated with horseradish peroxidase-conjugated sheep anti-mouse Ig (Jackson ImmunoResearch; 1:10,000 dilution) in the Can-Get-Signal Solution 2 (Toyobo) for 1h at room temperature, and washed three times with TBST for 5 min each. Chemiluminescence signals were developed using ImmunoStar LD (Fujifilm Wako) and detected using a luminescent imager LuminoGraph II (Atto).

## Live-cell Imaging

### *Olympus Spinning Disk Confocal Microscope*

For Figs. 1B,E-G, S1A (top), 2A,D-F, S2A, 3D-E, S3A (right), and Movies S2, images were captured on an Olympus IX8 Spinning Disk Confocal Microscope (CSU-22 head) equipped with an iXon Ultra 888 EMCCD camera (Andor) operated by SlideBook (Intelligent Imaging Innovations) coupled with a Phasor photomanipulation unit (Intelligent Imaging Innovations), and a 60x (NA 1.42) oil immersion objective. Cells were enclosed

in a custom-built 37°C chamber with 5% CO<sub>2</sub> throughout imaging. A 488-nm laser was used to capture GFP tagged scFv's and a 561-nm laser was used to capture H2B-mCh.

Imaging for 1F and 2E were acquired at full cell volumes with 17 z-stacks at 0.5-μm step size with the 488-nm laser set to 50% power and 100 ms and the 561-nm laser at 30% power with 100 ms exposure. Time lapse imaging in 3E (left) was captured with 21 z-stacks at 0.5-μm step size and 5 min interval for 175-200 min. The 488-nm laser was set to 40% power at 200-ms exposure time, and the 561 nm laser was set to 20% power at 100 ms exposure time. FRAP experiments (1G, 2F, and 3E right) were performed on cells transiently transfected (as described above). Single plane imaging was performed about the center of the nucleus. Before photobleaching, 10 frames were acquired at 1s time interval. Between frames 10 and 11 the Phasor 488-nm laser was set to bleach a circle of ~1 μm in diameter for 100 ms at 90% power. Image captures continued every 1s for an additional 170-250 times points.

For drug treatment imaging and FRAP of U2OS cells expressing anti-H3K27ac mintbody (Fig. 3D and 3E), the position of cells expressing both the scFv and mCh-H2b were recorded and an initial image was taken. Cells were imaged in 2.5 ml of media. 0.5 ml of media was removed and mixed with 2 μl of drug in DMSO or DMSO only, and then added back to the imaging dish. Working:Final concentrations of TSA and A485 was 625 μM:500 nM and 12.5 mM:10 μM respectively. Starting 5 minutes after drug addition, the cells were imaged every 5 minutes for 2 to 3 hrs. Cells which were not imaged during the 3 hr time lapses were used for the FRAP experiments as described in the previous section.

### *Custom HILO microscope*

For S1A (bottom) and S3A imaging, a custom built high inclined thin illumination fluorescence microscope<sup>79</sup> equipped with a 60x (NA 1.49, Olympus) oil-immersion objective lens, three solid-state laser lines (488, 561, and 637 nm, Voltran), a T660lpxr ultra-flat imaging grade emission image splitter (Chroma), and two iXon Ultra 888 EMCCD cameras (Andor) was used. A 300 mm achromatic lens (Thorlabs) was used to focus images onto the camera instead of the standard 180 mm Olympus tube lens, resulting in an effective magnification of 100x. Cells were incubated in a 37°C chamber with 5% CO<sub>2</sub> (Okolab). Images were captured at 512 x 512 px<sup>2</sup> size, 53.64 msec exposure time. A single z plane was captured about the center of the nucleus.

### *Dragonfly and Yokogawa CSU-W1 Microscopes*

For Fig. 3A and S3-5, a spinning-disk confocal system (Andor DragonFly or Yokogawa CSU-W1 for Fig. S3B) with an inverted microscope (Ti-E; Nikon), operated by Fusion software ver. 2.2.0.50 (Andor) or NIS-elements software ver. 5.11.03 (Nikon; for Fig. 3H



and S3B), was used with a heated stage at 37°C supplemented with 5% CO<sub>2</sub> (Tokai Hit). Single optical sections were acquired with a 100x PlanApo VC (NA 1.4) oil-immersion objective lens, a 405/488/561/640 or a 405/470/555/640NIR (for Fig. S3B) dichroic mirror, a 521/38 emission filter or a 520/60 emission filter (for Fig. S3B), a laser unit (5-20% transmission of 488 nm; LC-ILE-400-M; Andor or 30% transmission of 470 nm with a 10% neutral density filter; LDI-7 Laser Diode Illuminator; Chroma Technologies Japan for Fig. S3B), and an EM-CCD (iXonUltra; Andor; gain multiplier 300; exposure time 0.3-0.5 s or iXon+; Andor; gain multiplier 300; exposure time 0.1-1 s for Fig. S3B). Transmission images were also acquired.

For time lapse imaging (Fig. 3H and Movie S3), HeLa cells stably expressing sfGFP-tagged anti-H3S10ph (7G1G7) scFv (design 4) and H2B-Halo<sup>55</sup> were plated on a 35-mm glass-bottom dish (AGC Technology Solutions). The next day, the cells were stained with 100 nM Janelia Fluor 646 HaloTag Ligand (Promega) for 1 h, before replacing the medium with FluoroBrite containing FBS (Thermo Fisher Scientific) and setting onto a heated stage at 37°C supplemented with 5% CO<sub>2</sub> (Tokai Hit) featured on a spinning-disk confocal system (Yokogawa CSU-W1) with an inverted microscope (Ti-E; Nikon), operated by NIS-elements software ver. 5.11.03 (Nikon). Images were acquired with a 100x PlanApo VC (NA 1.4) oil-immersion objective lens, a 405/470/555/640NIR dichroic mirror, 520/60 and 690/50 emission filters, a laser unit (20% transmission of 470 nm and 15% of 640 nm with a 10% neutral density filter; LDI-7 Laser Diode Illuminator; Chroma Technologies Japan), and an EM-CCD (iXon+; Andor; gain multiplier 300; exposure time 500 ms). Fifteen different focal planes were imaged at 1 µm intervals every 5 min. Max intensity projection images are shown in Fig. 3H and Movie S3.

## Image Quantification

### *Nuclear/Chromatin to cytoplasmic ratio*

The ratio of nuclear to cytoplasmic signal intensity of cells from movies (Figs. 1F, 2E, and 3E, left) was calculated using custom scripts (Nuc2Cyt scripts on <https://github.com/Colorado-State-University-Stasevich-Lab/StasLabImageAnalysis>).

The maximum intensity projection of image stacks of full cell volumes was used for quantification. Background fluorescence was removed by subtracting the 1st percentile of intensity values, and negative values were set to zero. Nuclei were identified using Otsu thresholding of H2B and a binary mask was created to measure the nuclear signal. To measure cytoplasmic signals, a ring mask was generated around each nucleus with a 6-pixel width. The average intensity of intrabodies for both nuclear and cytoplasmic regions was measured and the ratios were then calculated. For live cell tracking experiments (Fig 3D, 3E Left), each cell in a time lapse was given a unique identifier using

Trackpy which was used to link the nuclear and cytoplasmic signals of individual cells for the duration of the experiment.

For Fig. 4C, and S3-5, nuclear and cytoplasmic areas in mintbody-expressing cells were selected manually, assisted with transmission images of cells. After the background intensity in a cell-free area was subtracted from the nuclear and cytoplasmic intensities, the nuclear to cytoplasm intensity ratio was determined.

### *Fluorescence Recovery After Photobleaching (FRAP)*

FRAP curves (Figs. 1G and 2F) were calculated using a custom script written in Python ([https://github.com/Colorado-State-University-Stasevich-Lab/frap\\_analysis](https://github.com/Colorado-State-University-Stasevich-Lab/frap_analysis)). Briefly, the FRAP spot location for the measurement was determined by subtracting the first post-FRAP frame (frame 11) from the previous frame (frame 10). The resulting image was used to create a mask that marked the bleached area, and a Difference-of-Gaussians filter was applied to further enhance the precise location of the center of the FRAP spot. A circular region was then inscribed into the bleached area, and the average signal from this region was plotted for the duration of the experiment (normalized to the intensity from frames 1-10).

**ACKNOWLEDGEMENTS:** We thank Prakash Shyam Karuppiiah and Jacob Leavitt for assistance with preliminary experiments during their rotation in the Stasevich lab. We thank the Jennifer Deluca lab (Colorado State University) for helpful scientific discussions. We thank Harumi Ueno (Tokyo Institute of Technology) for technical assistance and the Center for Integrative Biosciences and the Biomaterials Analysis Division, Open Facility Center at Tokyo Institute of Technology for DNA sequencing. We note that ChatGPT was used to assist with text grammar and clarity.

**FUNDING:** TJS, GG, GF, TM, HPF, and RH were supported by the National Institutes of Health (NIH) R35GM119728. TJS, CDS, JD, CNS, and BG were supported by NIAID R01AI168459. HK was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (JP17H01417, JP21H04764, and JP24H02325), Japan Agency for Medical Research and Development (AMED) Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS; JP23ama121020) and ASPIRE (24jf0126008h0001), and Shimadzu Corporation. DM was supported by the Japan Science and Technology Agency (JST) Support for Promoting Research Initiated by the Next Generation (SPRING) (JPMJSP2106). YS was supported by JST CREST (JPMJCR20S6). YO was supported by JSPS KAKENHI (JP23H00372, JP24H02323, JP24K21949), AMED BINDS (JP22ama121017j0001) and ASPIRE (24jf0126008h0001), MEXT Promotion of Development of a Joint Usage / Research System Project: Pan-

Omics DDRIC, MRCI for High Depth Omics, CURE:JPMXP1323015486 for MIB and RIIT in Kyushu Univ. N.Z was supported by NIH R00GM141453 and Cystic Fibrosis Foundation grant 005749A123. SRB and SG was supported by National Science Foundation (2236710).

## BIBLIOGRAPHY

1. Cardella, D., Sanchez-Guzman, D. & Rabbitts, T. H. Intracellular antibodies and biodegraders: Beyond small molecules and back again. *Curr. Opin. Biomed. Eng.* **27**, 100455 (2023).
2. Wagner, T. R. & Rothbauer, U. Nanobodies Right in the Middle: Intrabodies as Toolbox to Visualize and Modulate Antigens in the Living Cell. *Biomolecules* **10**, 1701 (2020).
3. Nizak, C. *et al.* Recombinant Antibodies to the Small GTPase Rab6 as Conformation Sensors. *Science* **300**, 984–987 (2003).
4. Manglik, A., Kobilka, B. K. & Steyaert, J. Nanobodies to Study G Protein–Coupled Receptor Structure and Function. *Annu. Rev. Pharmacol. Toxicol.* **57**, 19–37 (2017).
5. Irannejad, R. *et al.* Conformational biosensors reveal GPCR signalling from endosomes. *Nature* **495**, 534–538 (2013).
6. Kirchhofer, A. *et al.* Modulation of protein properties in living cells using nanobodies. *Nat. Struct. Mol. Biol.* **17**, 133–138 (2010).
7. Zhao, N. *et al.* A genetically encoded probe for imaging nascent and mature HA-tagged proteins in vivo. *Nat. Commun.* **10**, 2947 (2019).
8. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
9. Pleiner, T., Bates, M. & Görlich, D. A toolbox of anti-mouse and anti-rabbit IgG secondary nanobodies. *J. Cell Biol.* **217**, 1143–1154 (2018).
10. Tanaka, T., Williams, R. L. & Rabbitts, T. H. Tumour prevention by a single antibody domain targeting the interaction of signal transduction proteins with RAS. *EMBO J.* **26**, 3250–3259 (2007).
11. Kimura, H., Hayashi-Takanaka, Y., Stasevich, T. J. & Sato, Y. Visualizing posttranslational and epigenetic modifications of endogenous proteins in vivo. *Histochem. Cell Biol.* **144**, 101–109 (2015).
12. DeLuca, K. F. *et al.* Generation and diversification of recombinant monoclonal antibodies. *eLife* **10**, e72093 (2021).
13. Sato, Y. *et al.* Genetically encoded system to track histone modification in vivo. *Sci. Rep.* **3**, 2436 (2013).
14. Wörn, A. & Plückthun, A. Stability engineering of antibody single-chain Fv fragments. *J. Mol. Biol.* **305**, 989–1010 (2001).
15. Cattaneo, A. & Biocca, S. The selection of intracellular antibodies. *Trends Biotechnol.* **17**, 115–121 (1999).
16. Lo, A. S.-Y., Zhu, Q. & Marasco, W. A. Intracellular Antibodies (Intrabodies) and Their Therapeutic Potential. in *Therapeutic Antibodies* (eds. Chernajovsky, Y. & Nissim, A.) 343–373 (Springer, Berlin, Heidelberg, 2008). doi:10.1007/978-3-540-73259-4\_15.
17. Bird, R. E. *et al.* Single-Chain Antigen-Binding Proteins. *Science* **242**, 423–426 (1988).
18. Hamers-Casterman, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1993).
19. Rothbauer, U. *et al.* Targeting and tracing antigens in live cells with fluorescent

- nanobodies. *Nat. Methods* **3**, 887–889 (2006).
20. Muyldermans, S. Applications of Nanobodies. *Annu. Rev. Anim. Biosci.* **9**, 401–421 (2021).
21. Philibert, P. *et al.* A focused antibody library for selecting scFvs expressed at high levels in the cytoplasm. *BMC Biotechnol.* **7**, 81 (2007).
22. Auf der Maur, A., Tissot, K. & Barberis, A. Antigen-independent selection of intracellular stable antibody frameworks. *Methods* **34**, 215–224 (2004).
23. Talbert, P. B. & Henikoff, S. The Yin and Yang of Histone Marks in Transcription. *Annu. Rev. Genomics Hum. Genet.* **22**, 147–170 (2021).
24. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
25. Dai, W. *et al.* Epigenetics-targeted drugs: current paradigms and future challenges. *Signal Transduct. Target. Ther.* **9**, 332 (2024).
26. Feehley, T., O'Donnell, C. W., Mendlein, J., Karande, M. & McCauley, T. Drugging the epigenome in the age of precision medicine. *Clin. Epigenetics* **15**, 6 (2023).
27. Sato, Y., Nakao, M. & Kimura, H. Live-cell imaging probes to track chromatin modification dynamics. *Microsc. Oxf. Engl.* **70**, 415–422 (2021).
28. Stepanov, A. I., Besedovskaia, Z. V., Moshareva, M. A., Lukyanov, K. A. & Putlyaeva, L. V. Studying Chromatin Epigenetics with Fluorescence Microscopy. *Int. J. Mol. Sci.* **23**, 8988 (2022).
29. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
30. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
31. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
32. Anishchenko, I. *et al.* De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).
33. Torres, S. V. *et al.* De novo design of high-affinity binders of bioactive helical peptides. *Nature* (2023) doi:10.1038/s41586-023-06953-1.
34. Wicky, B. I. M. *et al.* Hallucinating symmetric protein assemblies. *Science* **378**, 56–61 (2022).
35. Høie, M. H. *et al.* AntiFold: Improved antibody structure-based design using inverse folding. Preprint at <http://arxiv.org/abs/2405.03370> (2024).
36. Bennett, N. R. *et al.* Atomically accurate de novo design of single-domain antibodies. Preprint at <https://doi.org/10.1101/2024.03.14.585103> (2024).
37. Shanehsazzadeh, A. *et al.* *In vitro* validated antibody design against multiple therapeutic antigens using generative inverse folding. Preprint at <https://doi.org/10.1101/2023.12.08.570889> (2023).
38. Makowski, E. K. *et al.* Optimization of therapeutic antibodies for reduced self-association and non-specific binding via interpretable machine learning. *Nat. Biomed. Eng.* **8**, 45–56 (2023).
39. Makowski, E. K. *et al.* Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space. *Nat. Commun.* **13**, 3788 (2022).
40. Dunbar, J. & Deane, C. M. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* **32**, 298–300 (2016).
41. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
42. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).



43. Einhauer, A. & Jungbauer, A. The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* **49**, 455–465 (2001).
44. DeRoo, J. *et al.* PAbFold: Linear Antibody Epitope Prediction using AlphaFold2. Preprint at <https://doi.org/10.7554/eLife.98369.1> (2024).
45. Peng, W., Pronker, M. F. & Snijder, J. Mass Spectrometry-Based *De Novo* Sequencing of Monoclonal Antibodies Using Multiple Proteases and a Dual Fragmentation Scheme. *J. Proteome Res.* **20**, 3559–3566 (2021).
46. Liu, Y. *et al.* Visualizing looping of two endogenous genomic loci using synthetic zinc-finger proteins with anti-FLAG and anti-HA frankenbodies in living cells. *Genes Cells* **26**, 905–926 (2021).
47. Ohishi, H. *et al.* STREAMING-tag system reveals spatiotemporal relationships between transcriptional regulatory factors and transcriptional activity. *Nat. Commun.* **13**, 7672 (2022).
48. Mirdita, M., Steinegger, M. & Söding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **35**, 2856–2858 (2019).
49. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
50. Goverde, C. A. *et al.* Computational design of soluble and functional membrane protein analogues. *Nature* **631**, 449–458 (2024).
51. Beugelink, J. W., Sweep, E., Janssen, B. J. C., Snijder, J. & Pronker, M. F. Structural Basis for Recognition of the FLAG-tag by Anti-FLAG M2. *J. Mol. Biol.* **436**, 168649 (2024).
52. Hayashi-Takanaka, Y. *et al.* Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. *Nucleic Acids Res.* **39**, 6475–6488 (2011).
53. Bar-On, Y. M., Flamholz, A., Phillips, R. & Milo, R. SARS-CoV-2 (COVID-19) by the numbers. *eLife* **9**, e57309 (2020).
54. Terry, J. S. *et al.* Development of a SARS-CoV-2 nucleocapsid specific monoclonal antibody. *Virology* **558**, 28–37 (2021).
55. Uchino, S. *et al.* Live imaging of transcription sites using an elongating RNA polymerase II-specific probe. *J. Cell Biol.* **221**, e202104134 (2022).
56. Sato, Y. *et al.* Visualizing histone H4K20me1 in knock-in mice expressing the mCherry-tagged modification-specific intracellular antibody. *Histochem. Cell Biol.* **162**, 41–52 (2024).
57. Tjalsma, S. J. D. *et al.* H4K20me1 and H3K27me3 are concurrently loaded onto the inactive X chromosome but dispensable for inducing gene silencing. *EMBO Rep.* **22**, e51989 (2021).
58. Sato, Y. *et al.* A Genetically Encoded Probe for Live-Cell Imaging of H4K20 Monomethylation. *J. Mol. Biol.* **428**, 3885–3902 (2016).
59. Kimura, H., Hayashi-Takanaka, Y., Goto, Y., Takizawa, N. & Nozaki, N. The organization of histone H3 modifications as revealed by a panel of specific monoclonal antibodies. *Cell Struct. Funct.* **33**, 61–73 (2008).
60. Hayashi-Takanaka, Y., Yamagata, K., Nozaki, N. & Kimura, H. Visualizing histone modifications in living cells: spatiotemporal dynamics of H3 phosphorylation during interphase. *J. Cell Biol.* **187**, 781–790 (2009).
61. Chandra, T. *et al.* Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. *Mol. Cell* **47**, 203–214 (2012).
62. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
63. Stasevich, T. J. *et al.* Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* **516**, 272–275 (2014).
64. Kelly, A. E. *et al.* Survivin reads phosphorylated histone H3 threonine 3 to activate the mitotic kinase Aurora B. *Science* **330**, 235–239 (2010).
65. Rechtsteiner, A. *et al.* The histone H3K36 methyltransferase MES-4 acts epigenetically



- to transmit the memory of germline gene expression to progeny. *PLoS Genet.* **6**, e1001091 (2010).
66. Hayashi-Takanaka, Y., Stasevich, T. J., Kurumizaka, H., Nozaki, N. & Kimura, H. Evaluation of chemical fluorescent dyes as a protein conjugation partner for live cell imaging. *PloS One* **9**, e106271 (2014).
67. Hayashi-Takanaka, Y. *et al.* Distribution of histone H4 modifications as revealed by a panel of specific monoclonal antibodies. *Chromosome Res.* **23**, 753–766 (2015).
68. Hayashi-Takanaka, Y. *et al.* Histone modification dynamics as revealed by multicolor immunofluorescence-based single-cell analysis. *J. Cell Sci.* **133**, jcs243444 (2020).
69. Kabayama, H. *et al.* An ultra-stable cytoplasmic antibody engineered for in vivo applications. *Nat. Commun.* **11**, 336 (2020).
70. Taylor, B. C. & Young, N. L. Combinations of histone post-translational modifications. *Biochem. J.* **478**, 511–532 (2021).
71. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).
72. Dunbar, J. *et al.* SABDab: the structural antibody database. *Nucleic Acids Res.* **42**, D1140–D1146 (2014).
73. Olsen, T. H., Boyles, F. & Deane, C. M. Observed Antibody Space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Sci.* **31**, 141–146 (2022).
74. Kahn, R. A. *et al.* Antibody characterization is critical to enhance reproducibility in biomedical research. *eLife* **13**, e100211 (2024).
75. Baker, M. Reproducibility crisis: Blame it on the antibodies. *Nature* **521**, 274–276 (2015).
76. Dreyer, F. A., Cutting, D., Schneider, C., Kenlay, H. & Deane, C. M. Inverse folding for antibody sequence design using deep learning. Preprint at <http://arxiv.org/abs/2310.19513> (2023).
77. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
78. Kuniyoshi, Y. *et al.* Identification of Immunoglobulin Gene Sequences from a Small Read Number of mRNA-Seq Using Hybridomas. *PloS One* **11**, e0165473 (2016).
79. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 159–161 (2008).