Cas9-mediated tagging of endogenous loci using HITAG

Joonwon Kim[1,2], Alexander F. Kratz[1], Jenny Sheng[1], Liudeng Zhang[1,2,3], Brijesh Kumar Singh[1,4], **Alejandro Chavez**[1,2,*]

[1] Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, NY, 10032, USA

[2] Department of Pediatrics, University of California San Diego, La Jolla, CA, 92093, USA

[3] College of Life Sciences, Zhejiang University, Hangzhou, 310058, China

[4] Current address: Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA

[*] Corresponding author (chavez2@health.ucsd.edu)

Abstract

To facilitate the interrogation of proteins at scale, we have developed High-throughput Insertion of Tags Across the Genome (HITAG). HITAG enables users to produce libraries of cells, each with a different protein of interest C-terminally tagged, to rapidly characterize protein function. To demonstrate the utility of HITAG, we fused mCherry to a set of 167 stress granule-associated proteins and characterized the factors which drive proteins to strongly accumulate within stress granules.

Introduction

To obtain an accurate working model of the cell, we must first understand the dynamic behavior and interaction partners for each of the thousands of proteins within it. A powerful approach to interrogate protein function is through the use of protein tags which are often appended to the C-terminus of a protein of interest, so as to minimize their influence on protein folding and localization[1,2]. When fused to a protein, these tags enable a myriad of studies such as the *in vivo* examination of protein localization (e.g. fluorescent protein tag), the affinity purification of a protein and its interaction partners from cells (e.g. FLAG epitope tag), or the rapid destruction of a target protein (e.g. FKBP-DD small molecule regulated degron tag)[3–6].

As CRISPR/Cas9 has simplified the modification of mammalian genomes, there has been growing interest in tagging all human proteins at their endogenous loci to facilitate the comprehensive mapping of protein behavior[7]. To achieve this goal, groups have relied on homologous recombination to insert tags at the C-terminus of target genes or used non-homologous end joining (NHEJ) to insert synthetic exons containing protein tags into the introns of target genes[7,8]. While powerful, these approaches still involve significant amounts of labor for each line generated or can perturb protein function due to the tag being inserted into the middle of the protein, respectively.

To fulfill the need for a high-throughput method of gene tagging that would minimally perturb protein function, we developed High-throughput Insertion of Tags Across the Genome (HITAG). HITAG uses Cas9 in combination with non-homologous end joining to insert protein tags into the C-terminus of target genes. Key to the HITAG approach is that

45 the process occurs within a mixed pool of cells wherein, at the end of the procedure, each
46 cell harbors a distinct protein that is C-terminally tagged. In analyzing the insertion events
47 mediated by HITAG, over 70% were found to be "perfect" fusions between the tag and the
48 target gene without the insertion or deletion of additional bases. To enable HITAG, we
49 performed iterative rounds of optimization including the development of a modified
50 selection marker to improve the enrichment of properly tagged cells. Using HITAG, a set
51 of 167 stress granule (SG)-associated proteins was fused to mCherry and the features
52 driving some of these proteins to strongly accumulate within SGs were explored. Our
53 results reveal a clear role for protein localization, the ability to bind to RNA, and the
54 propensity to phase separate in influencing efficient protein accumulation within SGs.
55

56 ## Results

57 ### Generating pools of tagged cells using Cas9+NHEJ

58 To develop HITAG, we began with a previously reported method of NHEJ-based
59 endogenous gene tagging termed CRISPaint[9]. In CRISPaint, a donor plasmid containing
60 the tag to be inserted into the genome is transfected into cells. Along with the donor
61 plasmid, 3 other plasmids containing Cas9, a gRNA against the C-terminus of the gene to
62 be tagged (target-gRNA), and a gRNA against the donor plasmid (donor-gRNA) are also
63 delivered (**Supplementary Figure 1**). Once all plasmids are inside the cell, Cas9 complexes
64 with the target-gRNA and the donor-gRNA to cut the target gene and the donor plasmid,
65 respectively. The cleaved donor plasmid can then become ligated into the endogenous
66 locus via NHEJ. If the tag gets knocked in-frame to the gene of interest, it will also lead to
67 the expression of a downstream drug resistant marker, enabling the facile enrichment of
68 properly tagged cells by applying drug selection.

69

70 While powerful, CRISPaint is limited in that it can only be applied towards one target gene
71 at a time, making it quite laborious to study a large number of genes. To adapt this system
72 into one suitable for rapidly generating libraries of tagged lines, the requirement for
73 performing independent transfections for each target gene needed to be removed. To
74 address this bottleneck, a mixture of target-gRNAs was packaged into lentiviruses and
75 delivered to cells at a low multiplicity of infection (MOI ~0.1) to ensure that on average,
76 each cell would integrate a single target-gRNA into its genome. This pool of cells was then
77 transfected *en masse* with the remaining components required for tagging (Cas9, donor
78 plasmid, donor-gRNA), thereby enabling each cell to uniquely tag the gene to which its
79 integrated target-gRNA is directed (**Figure 1a**). Finally, to enrich properly tagged cells, a
80 round of drug selection was applied.

81

82 In optimizing this high-throughput approach, we sought to avoid competition between
83 the target-gRNA and donor-gRNA for complexing with Cas9, since if competition did occur,
84 it would interfere with either cleavage of the endogenous locus or the donor plasmid and
85 in doing so decrease the rate of knock-in[10,11]. To achieve this goal, two orthogonal Cas9
86 proteins, SpCas9 and SaCas9, were employed, each of which interacts with a unique gRNA
87 scaffold. In this design, SpCas9 is used to cleave the endogenous target gene and SaCas9
88 is directed to linearize the donor plasmid (**Supplementary Figure 2**). Another benefit of
89 using two orthogonal Cas9 proteins is that properly tagged genes will not be recut, which
90 can occur when only using a single Cas9 protein should the target gene and the donor
91 plasmid have a similar sequence proximal to their PAM sites (**Supplementary Figure 3**).
92 Finally, to reduce off-target insertions which can occur if the donor-gRNA were to

93 inappropriately cleave the genome, several different SaCas9 donor-gRNAs were examined.
94 From this analysis, a donor-gRNA that showed high tagging efficiency and specificity was
95 identified and used throughout the remainder of the study (**Supplementary Figure 4**). As
96 a "perfect" fusion, without the loss or addition of bases, is the most common repair
97 outcome from the tagging, we developed three donor plasmids that upon fusion to the
98 target gene will place the tag in the appropriate translational reading frame
99 (**Supplementary Figure 5**)[9]. The design enables users to achieve precise tagging by
100 selecting an optimal donor plasmid depending upon the location within the C-terminal
101 codon that the target gene is cut by Cas9.
102
103 To demonstrate the utility of HITAG for studying protein behavior at scale, we used it to
104 investigate stress granules (SGs). SGs are transient liquid-liquid phase separated (LLPS)
105 RNA-protein complexes that form in response to environmental stress[12,13]. To date, the
106 factors which drive certain proteins to accumulate strongly within SGs remain unclear. We
107 hypothesized that by tagging a large number of SG-associated proteins with the
108 fluorescent protein, mCherry, we could gain insight into the properties that drive strong
109 versus weak accumulation in SGs. Having established our initial approach to high-
110 throughput gene tagging, a 193-member target-gRNA library against SG-associated
111 proteins was designed and delivered to HEK293T cells (**Figure 1a**). Into the resulting
112 mixture of cells, the remaining plasmids required for tagging were transfected in and
113 puromycin-resistant cells were then selected. To determine the bias in tagging in the
114 resulting library, the abundance of each gRNA within the population of cells before and
115 after the procedure was determined (**Figure 1b**). As expected, the distribution of gRNAs
116 in the initial population of cells before mCherry knock-in was even, with no gRNA

117 dominating the pool. In contrast, after mCherry insertion and puromycin selection, a
118 marked skew in the population was observed with 5 gRNAs occupying more than 60% of
119 the pool, suggesting that only a handful of genes were efficiently tagged during our initial
120 attempt, and that further optimization was required.
121
122 **Optimizing insertion rates and drug selection steps**
123 As numerous genes were poorly targeted in our initial library, we first sought to improve
124 the efficiency of the process by optimizing the ratio of the transfected plasmids. Using a
125 HEK293T cell line that stably expressed a single target-gRNA against the histone gene,
126 HIST1H4C, 10ng or 100ng of each of the plasmids required for tagging (Cas9, donor
127 plasmid, donor-gRNA) were delivered (**Figure 1c**). These studies revealed that a higher
128 amount of the donor plasmid (100ng) and a lower amount of the donor-gRNA (10ng) led
129 to a significant increase in the number of mCherry positive cells, before the application of
130 drug selection. Further optimization of the transfection conditions demonstrated an upper
131 limit to the benefit of delivering more donor plasmid (500ng), improvements by adding
132 additional Cas9 plasmid (100ng), and a decrease in insertion rates when the amount of
133 donor-gRNA is further reduced relative to the amount of donor plasmid (**Figure 1c**). These
134 results indicate that the amount of cleaved donor plasmid is a major factor determining
135 the tagging efficiency and suggest that cutting the donor plasmid too early, via increased
136 amounts of donor-gRNA, before the endogenous locus is cleaved, may be detrimental.
137
138 Given that the puromycin resistance marker is under the regulation of the endogenous
139 gene that is tagged, we next hypothesized that weakly transcribed genes may not produce
140 enough of the resistance protein to survive drug selection. To increase the level of

141 puromycin resistance conferred upon tagging, a modified donor plasmid (P3) was
142 constructed containing three tandem copies of the puromycin resistance gene downstream
143 of the mCherry tag (**Figure 1d**). As a control, a donor plasmid with three copies of the
144 puromycin resistance gene downstream of mCherry, but containing a stop codon after the
145 first PuroR copy was also constructed (P3S). These modified donor plasmids along with
146 the original donor plasmid with one copy of the PuroR gene (P1) were then transfected
147 along with Cas9 and the donor-gRNA into HEK293T cells containing a single target-gRNA
148 against either HIST1H4C, the chaperonin complex member, CCT7, or the nuclear pore
149 protein, NUP93. Minimal difference in the number of mCherry positive cells before
150 puromycin selection was observed among the different donors (**Supplementary Figure 6**).
151 In contrast, the number of viable cells dramatically increased after puromycin selection
152 using the P3 donor plasmid which expresses the three copies of the puromycin resistance
153 gene (**Figure 1d**). These results suggest that in our initial approach, many properly tagged
154 cells were likely lost due to insufficient expression of the drug selection marker.
155
156 Upon applying the improved transfection conditions and optimized P3 donor plasmid
157 design to the initial 193-member SG target-gRNA library, a marked improvement in the
158 number of tagged genes was observed. Of the genes that were tagged in the pool, 29
159 genes showed an abundance >1% following puromycin selection as compared to only 8
160 genes meeting this threshold in the initial approach (**Figure 1b**). In addition, a strong
161 correlation (ρ=0.78) among independent biological replicates was also found, suggesting
162 that the results were reproducible (**Supplementary Figure 7**). We designated this
163 optimized pipeline for generating pools of tagged cells HITAG.
164

165 **Application of HITAG to stress granule factors**

166 To ensure optimal tagging, our initial library of 193 SG targets had been selected such
167 that they all shared the same reading frame upon being cut by Cas9. To tag the remaining
168 proteins previously associated with SGs, two additional target-gRNA libraries of 190 and
169 205 members were created and used to generate additional pools of mCherry-tagged cells
170 via the HITAG approach.

171

172 To begin to characterize the fidelity of HITAG, the junctions between the various target
173 genes and the inserted mCherry tag (junction reads) were selectively enriched via a nested
174 PCR approach. From these analyses, 244 of the 588 genes that were targeted across the
175 three reading frame libraries were tagged and survived drug selection (**Supplementary**
176 **Figure 8**). The number of junction reads associated with each tagged gene was then
177 compared to its respective target-gRNA abundance within the same pool of cells. As
178 anticipated, a good correlation between these two metrics was observed (ρ = 0.84),
179 suggesting that the target-gRNA within each cell is driving the tagging events and that
180 target-gRNA abundance is a reasonable surrogate for which genes are tagged within the
181 library (**Supplementary Figure 9**).

182

183 While HITAG was able to target a large proportion of the genes within the library, many
184 failed to be tagged. To investigate the factors influencing these results, the characteristics
185 of tagged and untagged genes were interrogated. Given our previous findings regarding
186 the expression of the puromycin resistance marker limiting the efficiency of selection, we
187 first examined if gene expression influenced the results. On average tagged genes showed
188 a higher level of expression as compared to untagged genes (**Figure 1e**). Despite the

189 overall importance of gene expression, several genes showed high expression but remained
190 poorly tagged. In order to reduce the number of amino acids lost from the C-terminus of
191 the various target genes being examined, gRNAs were selected based on their proximity
192 to the stop codon. We hypothesized that a portion of the variation in tagging may have
193 been driven by inefficient target-gRNAs that poorly cut the endogenous target site. When
194 analyzing the on-target efficacy score of various target-gRNAs, tagged genes were
195 associated with target-gRNAs that had higher predicted efficiency scores as determined
196 by CRISPick[14] (**Figure 1f**).
197
198 A concern when using NHEJ to insert DNA sequences into endogenous loci is that it can
199 lead to error-prone repair with additional bases being lost from both the target gene and
200 donor construct. When analyzing the sequenced junctions from the pool of mCherry
201 tagged cells, 73.6% of all junctions were precise fusions between the endogenous locus
202 and the donor plasmid, in line with previous studies[9,15] (**Figure 1g**). Of the remaining
203 junctions, 21.7% retained the proper frame between the gene of interest and the mCherry
204 tag but showed a loss or gain of bases, with a strong bias towards smaller 1-2 amino acid
205 deletions and insertions within the observed repair products (**Figure 1g** and
206 **Supplementary Figure 10**). The remaining 4.7% of the repaired junctions were out-of-
207 frame products, presumed to have arisen from cells that have both a properly tagged allele
208 conferring drug resistance and an out-of-frame allele.
209
210 To investigate whether the HITAG approach can be applied to other cell lines, a set of 205
211 stress granule associated genes were tagged with mCherry using the human colorectal
212 carcinoma cell line, HCT116 (**Supplementary Figure 11**). The features of tagging such as

213 the efficiency, specificity, overall repair profile, and biases in tagging were similar between
214 HCT116 and HEK293T cells, suggesting that HITAG represents a general method for
215 tagging genes in high-throughput.
216
217 **Characterization of tagged proteins and their association with stress granules**
218 To examine the behavior of the tagged proteins within our mixed pool, single cells were
219 isolated, grown clonally, and the gRNA inside each was sequenced (for details see materials
220 and methods). Within the 806 clonal lines obtained, 167 unique proteins were mCherry
221 tagged, with each protein being represented by a median of 3 clonal isolates
222 (**Supplementary Figure 12**). To quantify the rates of improper tagging, proteins that were
223 tagged within at least 10 independent isolates and showed distinct subcellular localization
224 (e.g. nuclear localization) were examined (**Supplementary Figure 13**). On average ~95%
225 of the clones for a given protein showed the expected pattern of mCherry localization.
226 These results were further confirmed for two targets, BCALF1 and HNRNPA2B1, by using
227 PCR to amplify the junction between the target gene and the mCherry tag.
228
229 To probe if the mCherry tag alters protein localization, the subcellular distribution of the
230 tagged proteins (**Supplementary Figure 14**) was compared with the annotations listed
231 within the Human Protein Atlas[16]. Of the 155 proteins with localization data in the Human
232 Protein Atlas, 141 agreed with our findings. For the 14 proteins with conflicting results, 6
233 had annotations in other databases[7,17], with 5 of these agreeing with our findings[7]. As a
234 whole these data suggest that HITAG generates a high percentage of cells where the label
235 is inserted as designed and that the C-terminal fusion of mCherry to proteins of interest
236 rarely perturbs their localization.

237

238 While there are hundreds of proteins that have been found in SGs, what drives their
239 accumulation and why some proteins are more efficiently recruited to SGs is under active
240 investigation. To probe these questions, all 167 clonal lines were treated with 0.5 mM
241 sodium arsenite for 1 hour to induce SG formation. Stress granules were then visualized
242 by staining for the canonical SG marker G3BP1 (**Figure 2a**). Of the 167 tagged proteins,
243 only 23 showed the formation of distinct foci that overlapped with the G3BP1 marker
244 (strong accumulation), as determined by fluorescence microscopy (**Figure 2b**). To
245 determine if the mCherry label might be affecting protein dynamics, 6 proteins that showed
246 either strong or weak accumulation (no G3BP1 overlapping foci) within SGs were stained
247 with antibodies against the target protein in wild-type cells and the results compared to
248 our findings (**Supplementary Figure 15**). Comparisons between the staining patterns
249 showed that the mCherry tag had a minimal effect on altering a given protein's interaction
250 with SGs.

251

252 In examining the 23 proteins with strong accumulation, several distinctive features
253 immediately became apparent. Among the hits, all showed predominantly cytoplasmic
254 localization and had associated RNA binding activity. In addition, the majority could be
255 clustered by protein-protein interactions into two groups centered around either EIF4G1
256 or G3BP1 (**Figure 2c**).

257

258 To further characterize the nature of proteins that showed a strong association with SGs,
259 the strong and weak accumulators were compared on a variety of features such as protein
260 length, size of intrinsically disordered regions, abundance, charge, and liquid-liquid phase

261 separation (LLPS) propensity (**Figure 2d** and **Supplementary Figure 16**). Somewhat
262 surprisingly, features such as protein length or number of protein interaction partners
263 showed no correlation with the strength of accumulation within SGs. Of the metrics
264 examined, a propensity to LLPS showed the greatest difference between strong and weak
265 accumulators, suggesting a critical role for multivalent intra and intermolecular interactions
266 in driving these proteins behavior within SGs. These findings are in line with the fact that
267 during times of translation inhibition, SGs form by stalled preinitiation complexes,
268 containing EIF4G1, and their associated RNAs which undergo condensation upon
269 interaction with key nucleating RBPs such as G3BP1[12]. They also highlight the known role
270 of protein-RNA and protein-protein interactions in driving SG formation[12,13].
271
272 As a whole, these results demonstrate that while many proteins may associate with SG,
273 only a relative few have the required features to enable their robust accumulation. In
274 agreement with these observation, several nuclear RNA-binding proteins were found to
275 have high LLPS scores but poor accumulation within SG within our dataset. Yet, previous
276 reports have shown that during times of stress, some of these proteins can become
277 localized to the cytoplasm, fulfilling the critical requirement we highlight, which then
278 enables them to readily accumulate within SGs[18,19].
279
280

281 ## Discussion

282 Here we report the development of High-throughput Insertion of Tags Across the Genome
283 (HITAG), a fast and accessible method for generating libraries of endogenously tagged
284 proteins. Although numerous methods exist for performing targeted knock-ins into the

285 mammalian genome, most allow only a single protein at a time to be targeted, are
286 inefficient, or place tags in between coding exons which can disrupt protein folding and
287 function[7,8,20–23]. By increasing the throughput of performing C-terminal tagging, HITAG will
288 help facilitate the scalable interrogation of protein function and dynamics. In future
289 developments the coupling of HITAG with single-cell approaches such as pooled optical
290 screens will remove the need to isolate clonal populations post-tagging and provide a
291 further boost to discovery throughput[24].
292
293 Despite the fact that all the proteins that were fused to mCherry in this study have been
294 previously annotated as stress granule-associated, only 23 showed fluorescent puncta that
295 overlapped with the canonical SG marker G3BP1 upon subjecting the cells to sodium
296 arsenite treatment[25,26]. While several factors were found to be associated with strong
297 accumulation within G3BP1 foci, cytosolic localization, the ability to interact with RNA, and
298 a propensity to phase separate were found to be among the most critical. In future studies,
299 the generated lines may be used in combination with live cell and single molecule imaging
300 modalities to better understand the kinetics and distribution of each tagged protein within
301 SGs to gain further insight into granule assembly, disassembly, and heterogeneity.
302
303 While HITAG is able to generate a diverse library of endogenously tagged proteins, there
304 are several areas that can be further explored to improve the quality of the generated
305 libraries. First, the correlation between gene expression and tagging efficiency, in
306 combination with the observation that knocking in multiple puromycin resistance genes
307 reduces our targeting bias, suggests that lowly expressed genes are being lost due to
308 insufficient levels of drug marker expression. To overcome this limitation, additional copies

309 of the puromycin resistance marker, beyond the three currently being used, should be
310 considered. In addition, efforts to identify more efficient drug markers via the mining of
311 metagenomes or performing laboratory evolution on existing markers are additional
312 options that can be considered[27,28]. Second, the finding that predicted gRNA activity
313 influences the rate of gene tagging suggests that Cas9 variants with reduced PAM
314 requirements may increase the efficiency of targeting by allowing greater flexibility in
315 selecting which gRNA to use while still enabling insertions to occur near the C-terminus
316 of the target genes[29,30]. In addition, although a perfect junction without any loss or insertion
317 of amino acids is the most common repair outcome observed during the use of HITAG,
318 there are still ~22% of junctions that show loss or gain in nucleotides. This raises the
319 possibility that one may be able to direct Cas9 to cut downstream of the stop codon and
320 rely on error-prone repair to process away the stop codon and place the tag in-frame of
321 the protein of interest. If this alternative approach was found to be efficient, it would
322 increase the number of gRNA available for use and may also reduce the number of amino
323 acids that are lost during tagging in cases where the most C-terminal PAM is several amino
324 acids away from the stop codon.
325
326 **Acknowledgements**
327 This work was supported by a Career Awards for Medical Scientists from the Burroughs
328 Wellcome Fund and NIH grant 1R21HG011855-01 to A.C. Flow cytometry experiments were
329 performed in CCTI supported by core-supporting NIH grant S10RR027050. Cell sorting
330 experiments were performed in CSCI Flow Cytometry core facility at Columbia University
331 Irving Medical Center under the leadership of Michael Kissner.
332

333 **Contributions**

334 J.K. and A.C. designed experiments, assisted with data interpretation, and directed the
335 research program. B.K.S. performed the initial feasibility studies. J.K. optimized the HITAG
336 methodology including the development of the triple puromycin marker. J.K. and J.S.
337 conducted the NGS sample preparation and isolated the clonal populations. J.K., J.S., and
338 A.F.K. analyzed the resulting NGS data. J.K. and L.Z. performed the image analysis and
339 interpreted the results of the localization data. J.K and A.C. wrote the manuscript with input
340 from all authors.

341

342 **Conflicts of interest**

343 A.C. has several issued and provisional patents managed by Harvard University and
344 Columbia University on CRISPR based tools and methods. In addition, AC, JK, and BKS have
345 filed patents related to HITAG with Columbia University.

346

347 **Data availability**

348 All sequencing data were uploaded to SRA under BioProject accession number:
349 PRJNA895413. The script for the junction read analysis is available on
350 https://github.com/afkratz/HITAG-analysis-code. Additional scripts can be obtained
351 upon written request.

352

353 # Materials and Methods

354 **Plasmid construction**

355 To construct the gRNA expression plasmids, pSB700-blasto (Addgene #167904) was used
356 for SpCas9-specific gRNA expression and a modified pSB700-vector containing a SaCas9
357 compatible gRNA scaffold with a zeocin-resistance gene was used for SaCas9-specific
358 gRNA expression. Vectors containing gRNAs were cloned by Golden Gate using Esp3I as
359 described previously[31]. pCAS plasmid was constructed from a previous dual-Cas9 plasmid
360 (Addgene #107320) by replacing the 3xHA sequence with a P2A sequence using Gibson
361 assembly. pDNR was constructed from pCRISPaint-TagGFP2-PuroR (Addgene #80970)
362 using Gibson assembly. To construct the modified P3 donor with additional copies of the
363 puromycin resistance gene two puromycin resistance genes were PCR amplified with
364 primers designed to add a T2A sequence on each end. These fragments were then
365 assembled into a version of pDNR that was digested with ZraI using Gibson assembly. All
366 plasmids were validated by Sanger sequencing and will be made available via Addgene.
367 Sequence of the P3 donor, pCAS, the target-gRNA, and donor-gRNA vectors are provided
368 in **Supplementary Table 2**.

369

370 **Target-gRNA design**

371 To design target-gRNAs the CRISPick tool from the Broad was used with settings Human
372 GRCh38, CRISPRko, and SpyoCas9[14]. Guide RNAs with Esp3I restriction sites inside of them,
373 a polyT stretch longer than 4 base pairs, or more than one exact match in the human
374 genome were excluded from use even if they were the gRNA closest to the stop codon.
375 Frame number was categorized as the number of bases required to complete the cut
376 codon after cleavage.

377

378 **Construction of stress granule library**

379 A list of stress granule associated genes was collected from the existing literature[25,26]. These
380 methods identified SG proteins based on biotinylation of proteins in close spatial proximity
381 to core SG components or affinity purification of SG followed by mass spectrometry. Using
382 this list a set of target-gRNAs against each gene was generated with any target-gRNAs
383 that would result in a loss of more than 6 amino acids from the C-terminus of a protein
384 being removed. The resultant list of target-gRNAs was then sorted into three libraries
385 based on the frame (**Supplementary Table 1**).

386

387 Three gRNA libraries with different frames were synthesized as oligo pools from Agilent.
388 Each gRNA library was then PCR amplified from the initial oligo pool and cloned into the
389 modified pSB700 vector using Golden Gate assembly as previously described[31].

390

391 **Construction of HEK293T cells containing a library of target-gRNAs**

392 HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) + 10% FBS +
393 1% penicillin/streptomycin and incubated at 37°C and 5.0% CO_2. On day 1, HEK293T cells
394 were seeded at ~3.5×10^6 cells per well in 6-well plates for lentivirus production. On day 2,
395 a mixture of 600ng psPAX2, 150ng pMD2.G, and 450ng of the target-gRNA plasmid was
396 transfected into each well using lipofectamine 2000 according to the manufacturer
397 instructions. On day 3, the media was changed. Lentivirus was harvested by collecting the
398 supernatant after centrifugation (500g for 5 minutes) of the media on day 4 and day 5.
399 The lentivirus stocks were stored at -80°C in 1ml aliquots. To generate HEK293T cells with

400 on average a single target-gRNA integrated into their genome viral stocks were tested for
401 infectivity and cells were infected at an MOI of ~0.1.
402
403 **Quantifying the efficiency of tagging**
404 When optimizing the tagging approach, cells were plated at a density of 5-7x10^4 cells per
405 well in a 24-well plate with a media volume of 600µl. After 24 hours transfections were
406 performed according to the manufacturer instructions with 100ng pCAS, 100ng pDNR-
407 gRNA, and 200ng pDNR being delivered. In cases where the amount of each vector being
408 delivered was changed an inert "stuffer" plasmid, pBR322, was delivered to ensure that the
409 same total amount of plasmid DNA was delivered across all conditions. Media was changed
410 on the next day of transfection.
411
412 On day 6 after transfection, the transfected population was split to poly-D-lysine (PDL)-
413 treated 96 well plates at a density of 1-2x10^4 cells per well. The following day cells were
414 fixed with 4% paraformaldehyde for 5 minutes. The cells were then washed with PBS,
415 stained with DAPI for 5 minutes, and washed twice with PBS. The cells were imaged using
416 the ImageXpress® Pico Automated Cell Imaging Systems. Tagging efficiency was
417 determined by the number of mCherry-positive cells showing the proper localization over
418 the total number of cells as determined by DAPI staining of nuclei.
419
420 **Creating HITAG libraries**
421 To initiate the HITAG procedure target-gRNA containing cell libraries were seeded into 6-
422 well plates at 2.5-5.0x10^5 cells per 6-well. One 6-well plate was used per library. On the
423 next day, HITAG plasmids were transfected with 5-times more plasmid and transfection

424 reagents used for the 24-well plate condition. Cells were split 1:4 in media with 0.5μg/ml
425 puromycin 7 days after the transfection. The selection stopped until the next day when the
426 negative control cells died out. Selected tagged cells were expanded by changing media.
427
428 A stress granule library of HCT116 cells was generated by going through the same
429 procedure as outlined for HEK293T cells except the concentration of drug (blastidicine,
430 5.0μg/ml; puromycin, 2.5μg/ml), transfection reagent (FugeneHD), and the scale at which
431 transfections were performed (4 T75 flasks) was varied.
432
433 **gRNA analysis**
434 1-2x10^6 cells were harvested with trypsin and washed once with PBS. The cell pellet was
435 resuspended in 500μl Lucigen DNA QuickExtract reagent and incubated at 65°C with
436 shaking at 750rpm for 15 minutes. After brief centrifugation, the sample was incubated at
437 95°C with shaking at 750rpm for 5 minutes. gRNA regions were PCR amplified using the
438 following conditions: 98°C 45s, [98°C 15s; 56°C 15s; 72°C 20s] x n , 72°C 2min, 4°C hold,
439 where n is the cycle number before the PCR reaction saturated (usually between 20-25
440 cycles). The second round PCR was performed to add the Illumina indices: 98°C 45s, [98°C
441 15s; 56°C 15s; 72°C 20s]x8, 72°C 2 min, hold at 4°C. The PCR products were then run on a
442 2% agarose gel and the band of interest was purified. Samples were then sequenced on
443 an Illumina NextSeq 500. The resulting reads were then analyzed by either aligning them
444 using Bowtie2 or the MAGeCK pipeline[32].
445
446 **High-throughput isolation and analysis of target-mCherry junctions**
447

448 Genomic DNA was extracted from 1-2x10^6 cells using the Qiagen DNA extraction kit
449 (#69504). Enzymatic DNA fragmentation was performed using the NEBNext® Ultra™ II FS
450 DNA Library Prep Kit (E7805S). 2µg of genomic DNA (500ng per reaction) was treated with
451 5 minutes of enzymatic fragmentation. All subsequent steps were performed as instructed
452 by the manufacturer. The DNA fragments containing the mCherry sequence was amplified
453 through a nested PCR approach. For the first round PCR, 500ng of fragmented and
454 adapter-ligated DNA were amplified (50ng per reaction) using a primer binding on the
455 reverse strand of the mCherry sequence (see **Supplementary Table 3** for primer sequences)
456 and another primer binding to the p7 adaptor using the following PCR condition: 98°C
457 45s, [98°C 15s, 65°C 15s, 72°C 90s] x20, 72°C 5min, hold 4°C. After the resulting sample
458 was purified using SPRI beads aiming to capture all products greater than 100 bp in size.
459 The second-round PCR reaction was then performed using 50 ng of the first-round PCR
460 product (10ng per reaction) using PJ3 (**Supplementary Table 3**) and index primers under
461 the following PCR condition: 98°C 45s, [98°C 15s, 65°C 15s, 72°C 90s] x8, 72°C 5min, hold
462 4°C. The final PCR products were then isolated using SPRI beads aiming to remove all
463 products smaller than 200 base pairs.
464
465 To characterize the target gene-mCherry tag junctions, we began by constructing a
466 database of all genomic regions adjacent to the target cut sites. NGS reads were then
467 blasted locally twice, once against the database of target genomic regions, and once
468 against the three linker sequences, which differ by a few bases in order to maintain the
469 appropriate reading frame. Reads which featured at least 20 bases of alignment to genomic
470 targets and at least 20 bases of alignment to linker sequences were analyzed by comparing
471 the locations of the alignments to the expected location based on the gRNA cut site.

472 Insertions were identified as sections of a read which did not align to either the genomic
473 target or the linker sequence, while deletions appeared as a difference between the
474 expected cut site and the observed cut site.
475
476 **Generation of clonal cell lines and identification of the gRNA in each well.**
477 The library of cells with mCherry tagged SG-associated proteins was detached from a T75
478 flask and washed once with prechilled PBS. The cells were resuspended in Ca/Mg-free PBS
479 + 1% FBS, filtered with a 50um mesh filter and kept in ice. Before single-cell sorting, SYTOX
480 Blue (1:1000, Thermo S34857) was added as a cell viability indicator. In preparation for
481 single cell sorting, 96-well plates were filled with 150µl of media and prewarmed to room
482 temperature. Viable cells were sorted on the 96-well plates using Sony MA900 in the
483 Single-Cell Mode. Cell were placed in the incubator and the media from all plates was
484 changed as needed. Each well was confirmed visually to have one colony per well after 10
485 days.
486
487 PCR-ready genomic DNA was prepared by mixing ~$2x10^4$ cells in each well with 30µl
488 Lucigen DNA QuickExtract. After incubating the plates for 15 minutes at 65°C followed by
489 10 minutes at 95°C, 1µl of the DNA extract was used for PCR to amplify the gRNA sequence.
490 The same gRNA PCR condition were used as described above except 35 cycles were used
491 during the first round of PCR. Read counts of gRNAs for each well were then analyzed. For
492 a gRNA to be identified in a given well it needed to be present at an abundance at least
493 3 times greater than the next most abundant gRNA in that same well.
494
495 **Immunofluorescence staining for stress granule formation**

496 Sodium arsenite stress was applied by incubating cells with media containing 0.5mM
497 NaAs2O3 for an hour. Cells were then washed once with PBS and immediately fixed with
498 4% paraformaldehyde for 5 minutes, washed twice with PBS + 0.1% TritonX-100 incubating
499 for 10 minutes between each wash step, and blocked with Superblock™ Blocking Buffer
500 (Thermofisher #37581) with 0.1% TritonX-100 for 2 hours at room temperature. For primary
501 antibody staining, cells were covered with 100µl Superblock with 0.1% TritonX-100 and
502 primary antibodies (G3BP1, proteintech 13057-2-AP, 1:1000 dilution; mCherry, proteintech
503 68088-1, 1:1000 dilution) overnight at 4°C. After washing the cells with PBS + 0.1% TritonX-
504 100 (15 minutes) and Superblock with 0.1% TritonX-100 (15 minutes), cells were incubated
505 with secondary antibodies (Goat anti-mouse, Invitrogen A32727; Goat anti-rabbit,
506 Invitrogen A32731, 1:1000 dilution) for an hour at room temperature. After one wash with
507 PBS + 0.1% TritonX-100, cells were stained with 0.1 µg/ml DAPI in PBS for 5 minutes
508 followed by two PBS washes. At this point plates were either immediately imaged or
509 covered with aluminum seals and stored at 4°C.
510
511 **Collection of protein features**
512 Three published algorithms including PSPredictor, CatGranule and Plaac were used to
513 predict various scores related to phase separation[33–35]. The number of intrinsically
514 disordered regions, number and fraction of charged residues, hydropathy, and net charge
515 were calculated using CIDER[36]. All the scores were predicted with default parameters using
516 the dominant isoform.
517
518 **Protein-protein interaction network analysis**

The protein-protein interaction network was extracted from the STRING database, with the network type set as physical network and a minimum required interaction score of 0.400[37]. All of the text mining, experimental data, and databases were accepted as active interaction sources. Orphan genes, without interacting partners, are not included in the final network. K-means was used for clustering, and the cluster number was set to 2. Visualization is made by Gephi 0.9.2, with different colors indicating different gene groups and node size indicating node degree.

528 **Figure Legends**:

529

530 **Figure 1**. Development of HITAG and properties of the generated lines (a) Schematic
531 summary of HITAG. (b) Distribution of gRNAs within the population of cells before tagging
532 and after tagging and drug selection with the initial versus optimized HITAG approach.
533 Each colored bar represents the abundance of one gRNA within the population. (c) Tagging
534 efficiency before drug selection as a function of different ng amounts of pCAS, pDNR-
535 gRNA, and pDNR plasmid. Data shown are from three biological replicates (independent
536 transfections), the error bars indicate ± the standard deviation. The optimal condition with
537 an asterisk over it showed a significant difference (two sample t-test, * = $p < 0.05$) to all
538 other conditions tested. (d) Relative number of cells surviving puromycin selection when
539 different donor plasmids with one or three copies of the selection marker were used, data
540 are normalized to the P1 condition. Data shown are from three biological replicates
541 (independent transfections), the error bars indicate ± the standard deviation, the donor
542 with an asterisk over it showed a significant difference (two sample t-test, * = $p < 0.05$) to
543 all other donors tested across targets. Comparison of how (e) RNA expression level,
544 represented as log (FPKM +1) or the (f) gRNA activity score of each target-gRNAs
545 influences if a gene is tagged or not within the pool of tagged HEK293T cells (Mann-
546 Whitney-Wilcoxon test, **** = $p < 10^{-4}$) . (g) Distribution of repair outcomes summed
547 across all targets upon performing HITAG in HEK293T cells. Perfect junction indicates
548 precise annealing of the endogenous locus to the donor plasmid without a loss or addition
549 of bases after Cas9 cutting.

550

551 **Figure 2**. Using HITAG to understand the properties of proteins that accumulate within
552 SGs. (a) Staining of proteins which show robust accumulation within SGs after treatment
553 with 0.5 mM of $NaAs_2O_3$ for 1 hour. (Blue:DAPI, Green: anti-G3BP1, Red: anti-mCherry). (b)
554 List of proteins found to strongly accumulate within SG as determined by their overlap
555 with the canonical SG marker, G3BP1. (c) Network depicting the interactions among
556 proteins which show robust accumulation within SGs. (d) Predicted ability to LLPS as
557 determined by LLPSDB. As all proteins that showed strong accumulation were
558 predominantly localized to the cytoplasm for comparison cytoplasmic proteins were
559 divided into those that strongly or weakly accumulate into SGs. As an additional group,
560 we included all other proteins that showed weak accumulation but did not primarily localize
561 to the cytoplasm (non-cytosolic) (Mann-Whitney-Wilcoxon test, ns = not significant; ****
562 = $p < 10^{-4}$).
563
564 **Supplementary Figure 1**. Schematic summary of CRISPaint approach for NHEJ-based gene
565 tagging. Four different components - Cas9, target-gRNA, donor-gRNA, and donor plasmid
566 - are transfected into cells and required for the CRISPaint approach. In CRISPaint, SpCas9
567 cleaves the target gene at its C-terminus and the donor plasmid (pDNR). Linearized pDNR
568 can then become inserted into the genome through NHEJ. The cells which are properly
569 tagged in-frame will express mCherry fused to the protein of interest along with the
570 puromycin resistance gene (PuroR), which enables properly tagged cells to grow in the
571 presence of puromycin.
572
573 **Supplementary Figure 2**. Schematic summary of HITAG approach for NHEJ-based high-
574 throughput gene tagging. In the HITAG approach the target-gRNA is designed to interact

575 with SpCas9, while the donor-gRNA is designed to interact with SaCas9. The library of
576 target-gRNAs against the various genes of interest is first integrated into the pool of cells
577 at low infectivity such that each cell gets on average a single target-gRNA or none. The
578 remaining CRISPR components are then transfected in. This then enables each cell in the
579 population to tag a unique gene to which their target-gRNA is against. When tagging
580 occurs in-frame the protein of interest gains the C-terminal tag and a drug resistance
581 marker (e.g. puromycin resistance gene) is also expressed, enabling properly tagged cells
582 to be readily isolated via a round of drug selection.
583
584 **Supplementary Figure 3**. Schematic depicting how properly tagged genes can be recut
585 when using only a single Cas9 protein. (a) When using only a single Cas9 protein, as is
586 done in the original CRISPaint approach, the properly tagged product can be recut if the
587 3 PAM-proximal base pairs of the target gene are identical or similar to those of the donor
588 plasmid. This occurs because the spacer sequence of the target-sgRNA with an accessible
589 SpCas9-PAM site is regenerated in the final knock-in product. This recutting is then
590 expected to inhibit the emergence of the perfectly tagged products. (b) Recutting can be
591 avoided when using two orthogonal Cas9 proteins, as is done in the HITAG approach. This
592 occurs because even if the final product that forms has homology to the spacer sequence
593 of the target-gRNA it will have a SaCas9-PAM site proximal to it, and thus not be an
594 appropriate substrate for SpCas9 which has different PAM requirements.
595
596 **Supplementary Figure 4**. Identifying efficient and specific gRNAs against the donor
597 plasmid. (a) The efficiency of three different donor plasmid targeting gRNAs (gRNA1-3)
598 were assessed for the ability to add a C-terminal mCherry tag to either CCND1, HIST1H4C,

599 or PCNA. To measure efficiency, the percentage of cells with the appropriate mCherry
600 localization as determined by fluorescence microscopy was quantified before the
601 application of drug selection. Frame refers to the reading frame of the donor plasmid when
602 cut. As there are three possible reading frames (i.e 0, 1, and 2) that a given target gene
603 could be cut in, 3 donor plasmids corresponding to each frame were prepared for each
604 donor-gRNA being tested. This enables the system to then cut the target gene and donor
605 plasmid in such a way that should they undergo a perfect fusion (without the addition or
606 removal of bases) the tag and drug marker will be in the appropriate reading frame. (b)
607 The same donor-gRNAs as in panel A were also examined for specificity by transfecting
608 them in combination with SaCas9 and a mCherry-containing donor plasmid. As no target-
609 gRNA is provided in this experiment, all drug resistant clones that arise are due to either
610 low levels of stochastic double strand breaks or the donor-gRNA cutting an endogenous
611 site in the genome to enable the donor plasmid to then insert itself. Based on these studies
612 gRNA2 was determined to be the optimal donor-gRNA and for all subsequent studies was
613 used. Three biological replicates (independent transfections) were performed for all
614 conditions and error bars represent ± standard deviation.
615
616 **Supplementary Figure 5**. Schematic summary of the results of tagging depending on the
617 "frame" in which the target gene is cleaved by SpCas9. Here, the frame number is defined
618 as the number of nucleotides that must be added by the donor so that the target gene
619 and the tag of interest are in-frame. In the original CRISPaint approach three unique donor-
620 gRNA are used against a single donor plasmid to generate each of the needed reading
621 frames, with each donor-gRNA potentially having its own unique set of off-target within
622 the genome. In contrast, HITAG employs three different donor plasmids that all work with

623 a single donor-gRNA. By designing each donor plasmid to have either 0, 1, or 2 bases
624 added it enables the same donor-gRNA, which is tested to ensure it has minimal off-target
625 activity, to be used across all studies. For example, pDNR (Frame 0) is used when tagging
626 a target gene that does not require additional nucleotides added for the tag to be in-
627 frame with the cut gene.
628
629 **Supplementary Figure 6**. Comparing tagging efficiency when using donor plasmids with
630 varying numbers of the puromycin resistance gene. P1: a single copy of puroR, P3: three
631 copies of puroR, P3S: three copies of puroR but a stop codon is placed after the first puroR
632 copy. Three biological replicates (independent transfections) were performed. Error bars
633 represent ± standard deviation. Results shown are the percentage of properly tagged cells
634 before drug selection as determined by fluorescence microscopy.
635
636 **Supplementary Figure 7**. Comparing gRNA abundance between two independent rounds of
637 HITAG on the same target-gRNA library. The abundance of each target-gRNA within the pool
638 of cells was internally normalized between 0 to 1 for each replicate. ρ represents the Spearman
639 correlation coefficient.
640
641
642 **Supplementary Figure 8**. Percentage of tagged gene members within each reading frame
643 and as a whole across the generated HITAG libraries. Red and grey boxes indicate the
644 tagged group and the untagged group of genes, respectively.
645

646 **Supplementary Figure 9**. Correlation between the read counts from each gRNA within the
647 pool of HITAG modified cells compared with the target-mCherry tag junction reads derived
648 from the same pool of cells. The abundance of each target-gRNA and junction reads within
649 the pool of cells was internally normalized to be between 0 to 1 before performing the
650 comparison. ρ represents the Spearman correlation coefficient between the two sets of data.
651

652 **Supplementary Figure 10**. Characterization of the junction between the target gene and
653 mCherry tag. Frequency of the number of amino acids lost from (a) the target genes or (b)
654 the linker, which connects the tag to the endogenous target gene, respectively. (c)
655 Frequency of the number of amino acids insertion. Insertion refers to the number of amino
656 acids that are inserted between the target gene and the tag. Red dots indicate the junctions
657 which show no deletion or insertion of additional amino acids.
658

659 **Supplementary Figure 11**. Result from applying HITAG to HCT116 cells. The SG target-
660 gRNA library 3 (frame 2) was integrated into HCT116 cells and the subsequent pool of
661 cells was taken through the remainder of the HITAG procedure to isolate a mixed
662 population of mCherry tagged cells. (a) Correlation between the normalized read counts
663 from each gRNA within the pool of HITAG modified HCT116 cells compared with the
664 normalized target-mCherry tag junction reads derived from the same pool of tagged
665 HCT116 cells. ρ represents the Spearman correlation coefficient between the two sets of
666 data. (b) Distribution of repair outcomes summed across all targets upon performing HITAG
667 in HCT116 cells. Perfect junction indicates precise annealing of the endogenous locus to
668 the donor plasmid without a loss or addition of bases after Cas9 cutting. Comparison of
669 how the (c) RNA expression level represented as log (FPKM +1) or the (d) On-target efficacy

670 score of the target-gRNAs influence if a gene is tagged or not within the pool of tagged
671 HCT116 cells (Mann-Whitney-Wilcoxon test; *** = $p < 10^{-3}$).
672
673 **Supplementary Figure 12**. The number of times a given gene was tagged within the set
674 of 806 clonal lines that were isolated after performing HITAG. The yellow bar represents
675 the median number of clones obtained for a given targeted gene.
676
677 **Supplementary Figure 13**. Examining the rates of off-target tagging by quantifying the
678 consistency of results when multiple clones of the same tagged gene were obtained. (a)
679 Proteins with either nuclear or ER localization which had 10 or more clones to examine
680 from within the 806 clonal isolates were studied to determine the number of clones that
681 showed the appropriate localization. (b) PCR validation using primers to show that
682 microscopy-based results were concordant with targeted PCR directed at the junction
683 between either HNRNPA2B1 or BCLAF1 and the mCherry tag. Yellow rectangle marks the
684 location of the expected PCR band. C: control DNA from untargeted HEK293T cells. L:
685 Ladder.
686
687 **Supplementary Figure 14**. Decision tree used to compare localization information between
688 this study and the Human Protein Atlas. Protein subcellular localization information was
689 available for 155 out of 167 genes in Human Protein Atlas database. 141 out of 155 genes
690 showed a similar subcellular localization within our data as described in Human Protein
691 Atlas. Of the 14 genes that disagreed with the data from Human Protein Atlas, 6 genes
692 were found in Opencell or the RBP Image Database. 5 out of 6 of these genes agreed with
693 our subcellular localization findings.

694

695 **Supplementary Figure 15**. Comparison of protein localization and behavior between
696 mCherry-tagged version of a protein and the endogenous untagged counterpart. Images
697 of stained cells under both homeostatic (normal) and stressed (arsenite) conditions are
698 shown.

699

700 **Supplementary Figure 16**. Analysis of the features which drive strong accumulation within
701 SGs. Because all proteins that strongly gather in SG were primarily cytosolic analyses are
702 done comparing cytosolic proteins that show strong vs weak accumulation. As an
703 additional category, proteins that show non-cytosolic localization are also examined. (a)
704 propensity to localize to a foci as determined by CatGranule, (b) prion-like propensity score
705 as determined by Plaac, (c) Protein length (number of amino acids), (d) Number of
706 intrinsically disorder regions, (e) RNA expression level represented as log (FPKM+1), (f)
707 protein abundance extracted from mass spectrometry data, (g-i) fraction of charged
708 residues, (j) net charge per residues at pH 7.4, (k) abosolute mean net charge at pH 7.4,
709 and (l) Uversky hydropathy were compared between the groups. (Mann-Whitney-Wilcoxon
710 test, ns: not significant; * = $p < 0.05$; ** = $p < 10^{-2}$; *** = $p < 10^{-3}$; **** = $p < 10^{-4}$)

711

712

713 **Reference**

714 1. Bukhari, H. & Müller, T. Endogenous Fluorescence Tagging by CRISPR. *Trends Cell*
715 *Biol.* **29**, 912–928 (2019).
716 2. Palmer, E. & Freeman, T. Investigation Into the use of C- and N-terminal GFP Fusion
717 Proteins for Subcellular Localization Studies Using Reverse Transfection Microarrays.
718 *Comp. Funct. Genomics* **5**, 742068 (1900).
719 3. Chudakov, D. M., Lukyanov, S. & Lukyanov, K. A. Fluorescent proteins as a toolkit for
720 in vivo imaging. *Trends Biotechnol.* **23**, 605–613 (2005).
721 4. Zhang, X. *et al.* Epitope tagging of endogenous proteins for genome-wide ChIP-chip
722 studies. *Nat. Methods* **5**, 163–165 (2008).
723 5. Nabet, B. *et al.* The dTAG system for immediate and target-specific protein
724 degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018).
725 6. Stadler, C. *et al.* Immunofluorescence and fluorescent-protein tagging show high
726 correlation for protein localization in mammalian cells. *Nat. Methods* **10**, 315–323
727 (2013).
728 7. Cho, N. H. *et al.* OpenCell: Endogenous tagging for the cartography of human cellular
729 organization. *Science* **375**, eabi6983.
730 8. Reicher, A., Koren, A. & Kubicek, S. Pooled protein tagging, cellular imaging, and in
731 situ sequencing for monitoring drug action in real time. *Genome Res.* **30**, 1846–1855
732 (2020).
733 9. Schmid-Burgk, J. L., Höning, K., Ebert, T. S. & Hornung, V. CRISPaint allows modular
734 base-specific gene tagging using a ligase-4-dependent mechanism. *Nat. Commun.* **7**,
735 12338 (2016).

736 10. McCarty, N. S., Graham, A. E., Studená, L. & Ledesma-Amaro, R. Multiplexed CRISPR
737 technologies for gene editing and transcriptional regulation. *Nat. Commun.* **11**, 1281
738 (2020).
739 11. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability
740 with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci.* **112**, 3570–3575
741 (2015).
742 12. Hofmann, S., Kedersha, N., Anderson, P. & Ivanov, P. Molecular mechanisms of stress
743 granule assembly and disassembly. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1868**,
744 118876 (2021).
745 13. Wolozin, B. & Ivanov, P. Stress granules and neurodegeneration. *Nat. Rev. Neurosci.*
746 **20**, 649–666 (2019).
747 14. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-
748 target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
749 15. Song, B., Yang, S., Hwang, G.-H., Yu, J. & Bae, S. Analysis of NHEJ-Based DNA Repair
750 after CRISPR-Mediated DNA Cleavage. *Int. J. Mol. Sci.* **22**, (2021).
751 16. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, eaal3321
752 (2017).
753 17. Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-
754 binding proteins. *Nature* **583**, 711–719 (2020).
755 18. Kim, H. J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause
756 multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).
757 19. Wall, M. L., Bera, A., Wong, F. K. & Lewis, S. M. Cellular stress orchestrates the
758 localization of hnRNP H to stress granules. *Exp. Cell Res.* **394**, 112111 (2020).

759 20. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-
760 independent targeted integration. *Nature* **540**, 144–149 (2016).
761 21. Anzalone, A. V. *et al.* Programmable deletion, replacement, integration and inversion
762 of large DNA sequences with twin prime editing. *Nat. Biotechnol.* **40**, 731–740 (2022).
763 22. Pallarès-Masmitjà, M. *et al.* Find and cut-and-transfer (FiCAT) mammalian genome
764 engineering. *Nat. Commun.* **12**, 7071 (2021).
765 23. Ioannidi, E. *et al.* Drag-and-drop genome insertion without DNA cleavage with
766 CRISPR-directed integrases. (2021) doi:10.1101/2021.11.01.466786.
767 24. Feldman, D. *et al.* Optical Pooled Screens in Human Cells. *Cell* **179**, 787-799.e17
768 (2019).
769 25. Markmiller, S. *et al.* Context-Dependent and Disease-Specific Diversity in Protein
770 Interactions within Stress Granules. *Cell* **172**, 590-604.e13 (2018).
771 26. Youn, J.-Y. *et al.* High-Density Proximity Mapping Reveals the Subcellular
772 Organization of mRNA-Associated Granules and Bodies. *Mol. Cell* **69**, 517-532.e11
773 (2018).
774 27. Molina, R. S. *et al.* In vivo hypermutation and continuous evolution. *Nat. Rev.*
775 *Methods Primer* **2**, 36 (2022).
776 28. Coughlan, L., Cotter, P., Hill, C. & Alvarez-Ordóñez, A. Biotechnological applications of
777 functional metagenomics in the food and pharmaceutical industries. *Front. Microbiol.*
778 **6**, (2015).
779 29. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained
780 genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**,
781 290–296 (2020).

782 30. Christie, K. A. & Kleinstiver, B. P. Making the cut with PAMless CRISPR-Cas enzymes.
783 *Trends Genet.* **37**, 1053–1055 (2021).
784 31. AU - Nageshwaran, S. *et al.* CRISPR Guide RNA Cloning for Mammalian Systems. *J.*
785 *Vis. Exp.* e57998 (2018) doi:10.3791/57998.
786 32. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-
787 scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
788 33. Bolognesi, B. *et al.* A Concentration-Dependent Liquid Phase Separation Can Cause
789 Toxicity upon Increased Protein Expression. *Cell Rep.* **16**, 222–231 (2016).
790 34. Chu, X. *et al.* Prediction of liquid–liquid phase separating proteins using machine
791 learning. *BMC Bioinformatics* **23**, 72 (2022).
792 35. Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: a web and
793 command-line application to identify proteins with prion-like amino acid
794 composition. *Bioinformatics* **30**, 2501–2502 (2014).
795 36. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G. & Pappu, R. V. CIDER:
796 Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered
797 Proteins. *Biophys. J.* **112**, 16–21 (2017).
798 37. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein–protein
799 networks, and functional characterization of user-uploaded gene/measurement sets.
800 *Nucleic Acids Res.* **49**, D605–D612 (2021).
801
802

Figures

Figure 1.

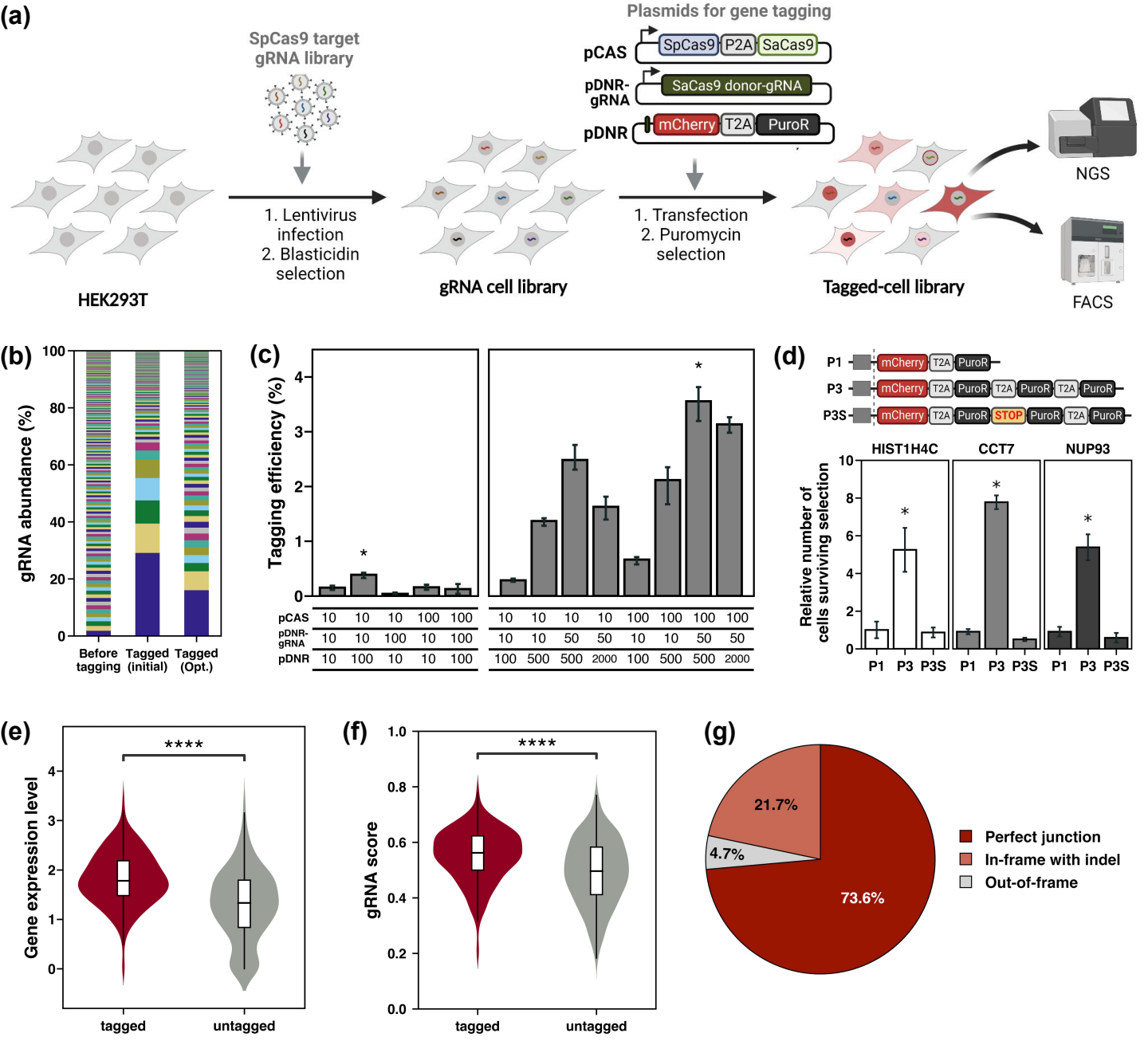
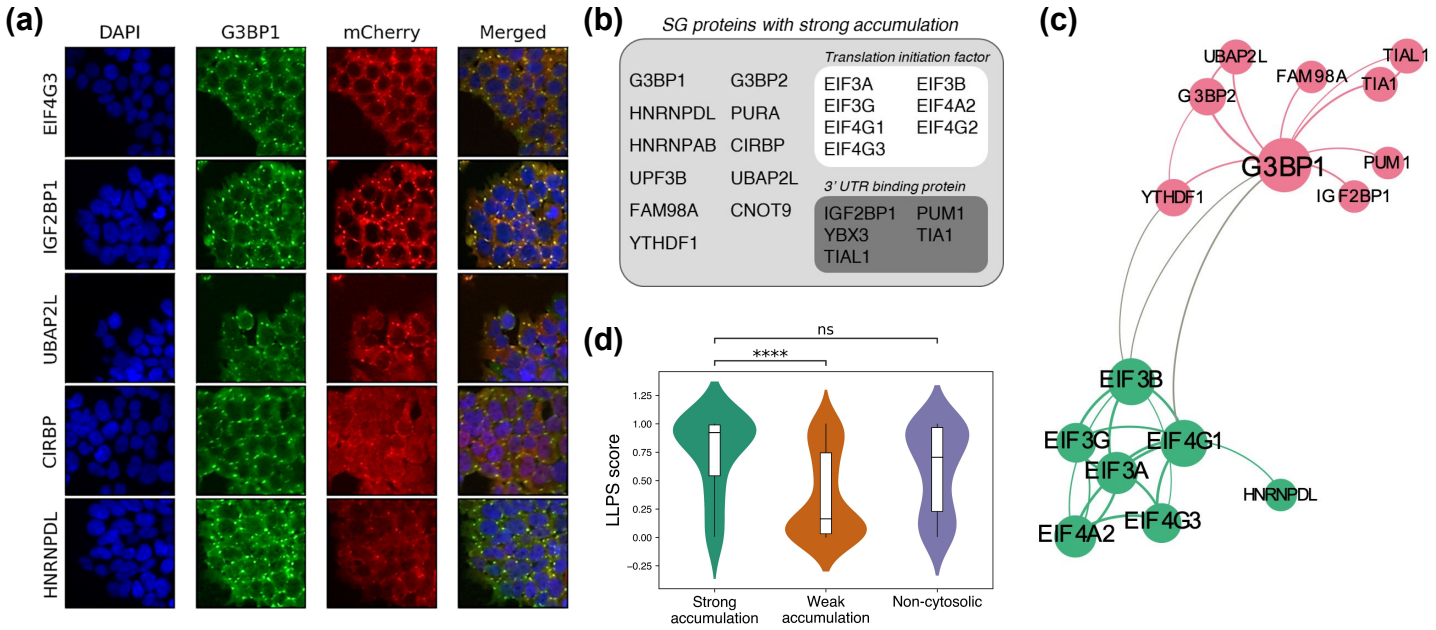
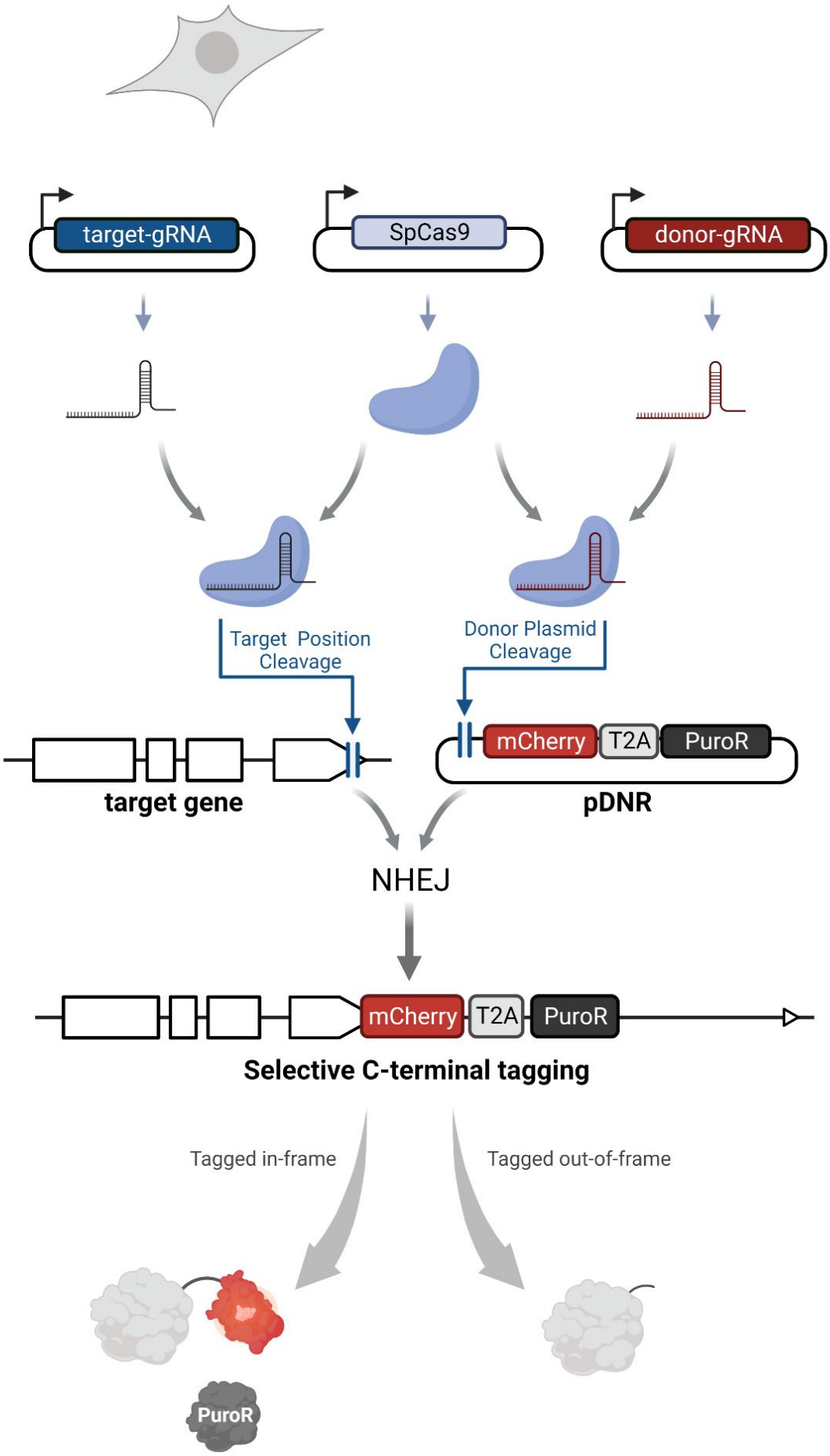

(a)
SpCas9 target gRNA library
Plasmids for gene tagging
pCAS
SpCas9
P2A
SaCas9
pDNR-gRNA
SaCas9 donor-gRNA
pDNR
mCherry
T2A
PuroR
HEK293T
1. Lentivirus infection
2. Blasticidin selection
gRNA cell library
1. Transfection
2. Puromycin selection
Tagged-cell library
NGS
FACS
(b)
gRNA abundance (%)
Before tagging
Tagged (initial)
Tagged (Opt.)
(c)
Tagging efficiency (%)
pCAS
pDNR-gRNA
pDNR
(d)
P1
P3
P3S
STOP
HIST1H4C
CCT7
NUP93
Relative number of cells surviving selection
(e)
Gene expression level
****
tagged
untagged
(f)
gRNA score
****
tagged
untagged
(g)
21.7%
4.7%
73.6%
Perfect junction
In-frame with indel
Out-of-frame

Figure 2.

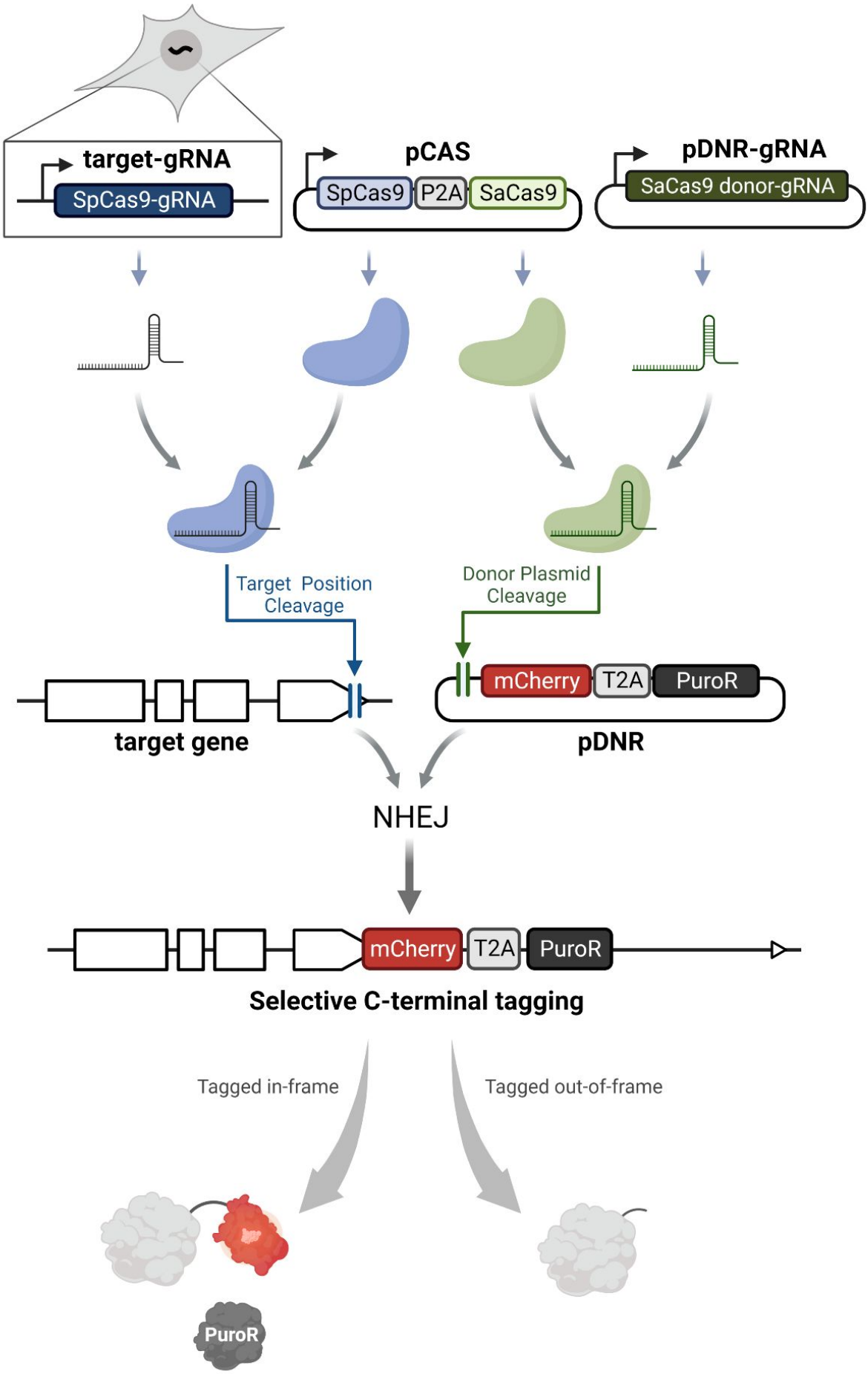

(a)
DAPI
G3BP1
mCherry
Merged
EIF4G3
IGF2BP1
UBAP2L
CIRBP
HNRNPDL
(b)
SG proteins with strong accumulation
G3BP1
G3BP2
HNRNPDL
PURA
HNRNPAB
CIRBP
UPF3B
UBAP2L
FAM98A
CNOT9
YTHDF1
Translation initiation factor
EIF3A
EIF3B
EIF3G
EIF4A2
EIF4G1
EIF4G2
EIF4G3
3' UTR binding protein
IGF2BP1
PUM1
YBX3
TIA1
TIAL1
(c)
UBAP2L
TIAL1
FAM98A
TIA1
G3BP2
G3BP1
PUM1
YTHDF1
IGF2BP1
EIF3B
EIF3G
EIF4G1
EIF3A
HNRNPDL
EIF4A2
EIF4G3
(d)
ns
****
LLPS score
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
Strong accumulation
Weak accumulation
Non-cytosolic

Supplementary Figure 1.



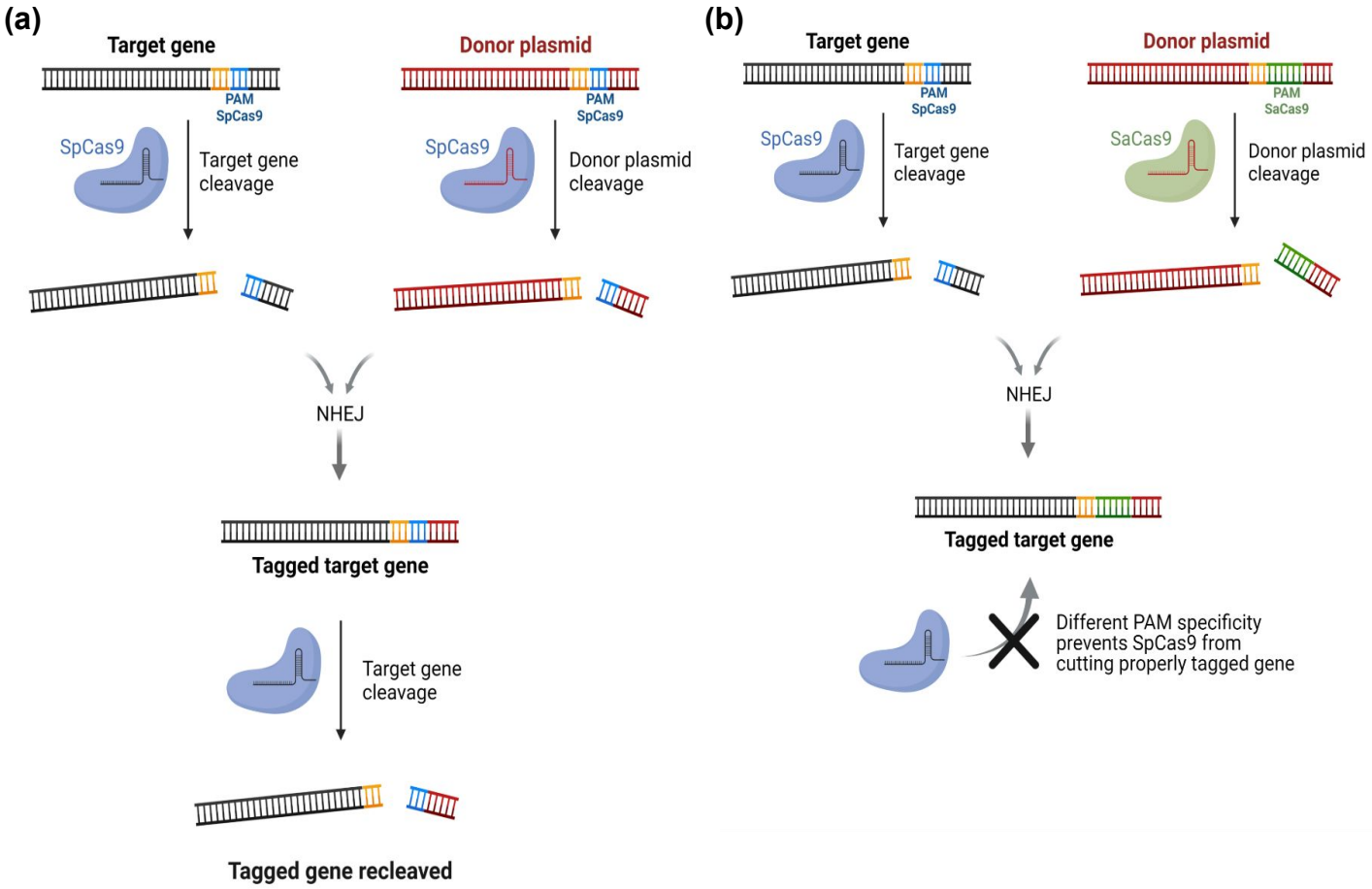
- Protein tagged with mCherry
- Puromycin resistance

Supplementary Figure 2.

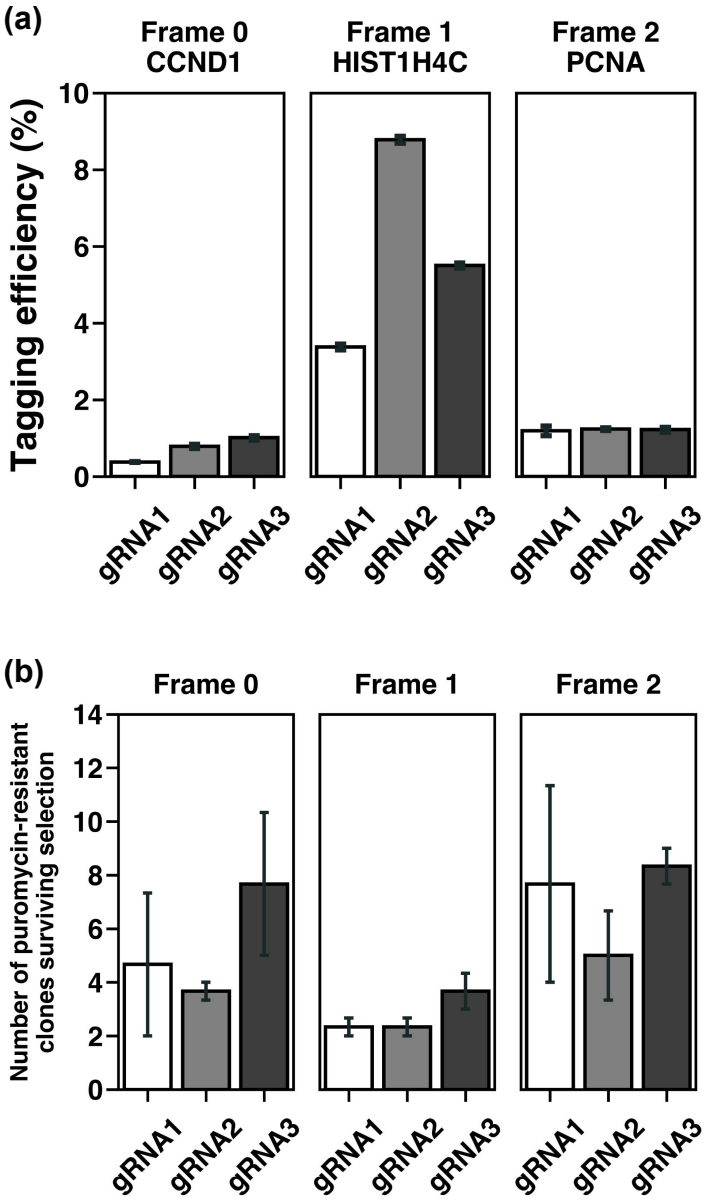


- Protein tagged with mCherry
- Puromycin resistance

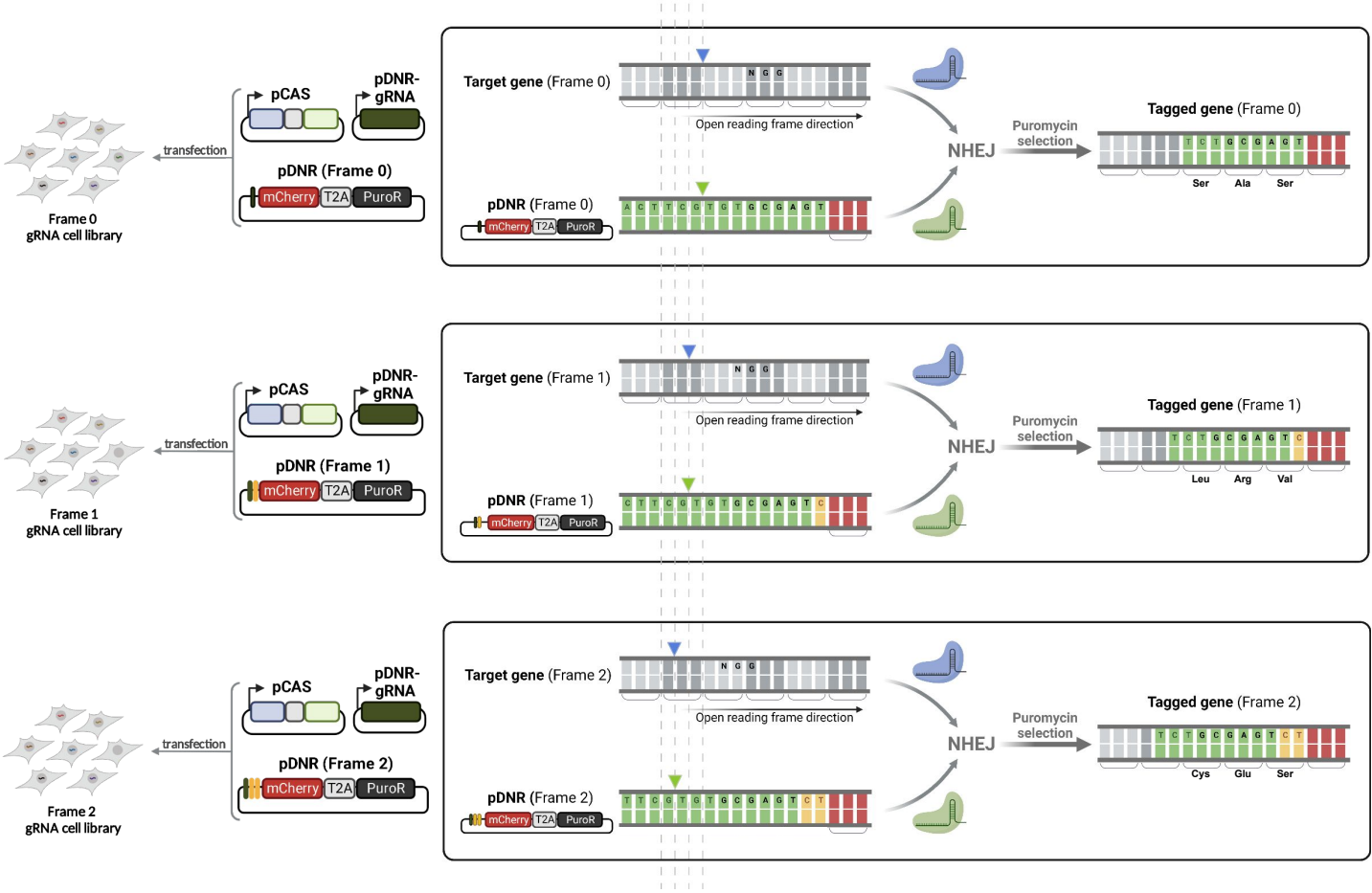
Supplementary Figure 3.



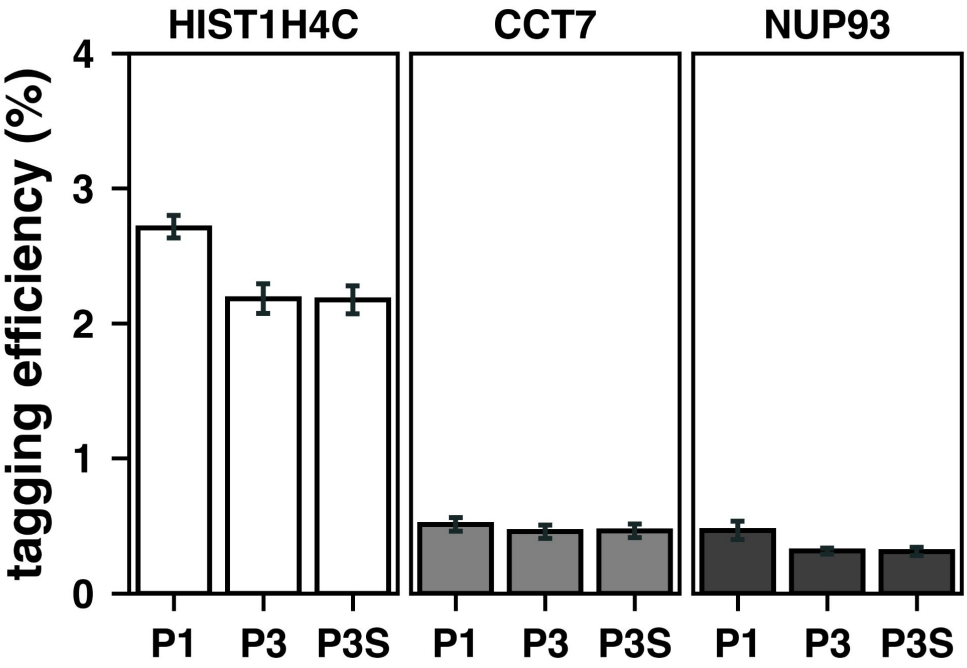
Supplementary Figure 4.



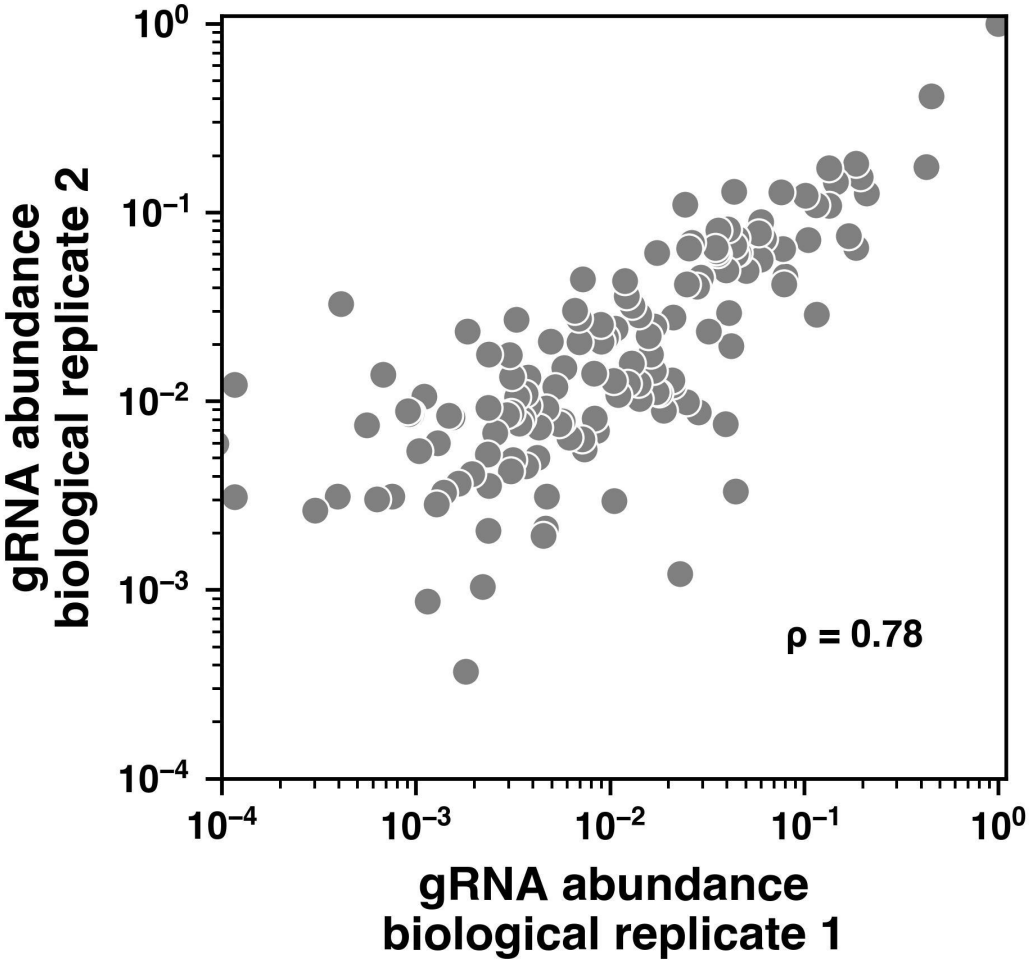
Supplementary Figure 5.



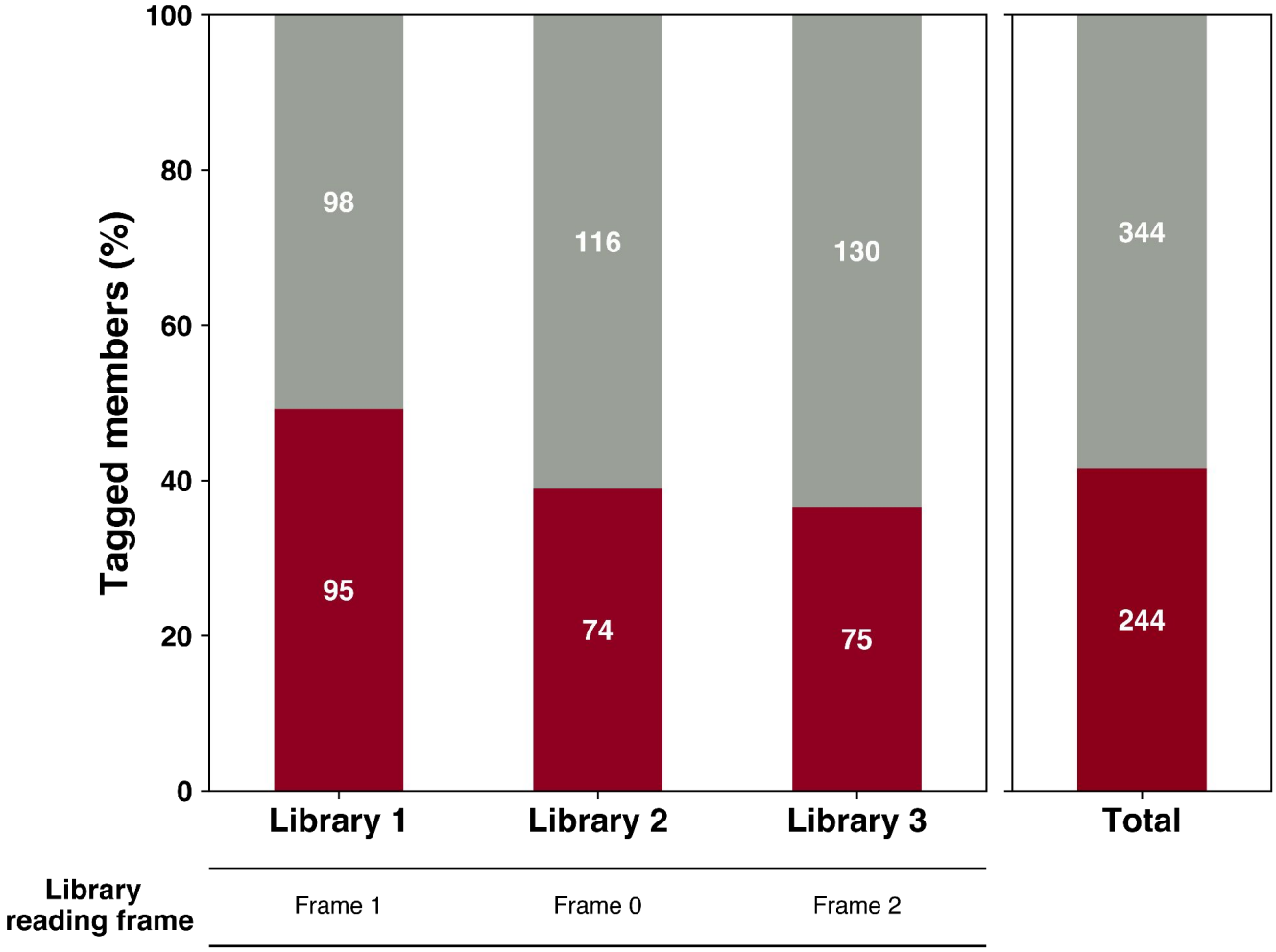
Supplementary Figure 6.



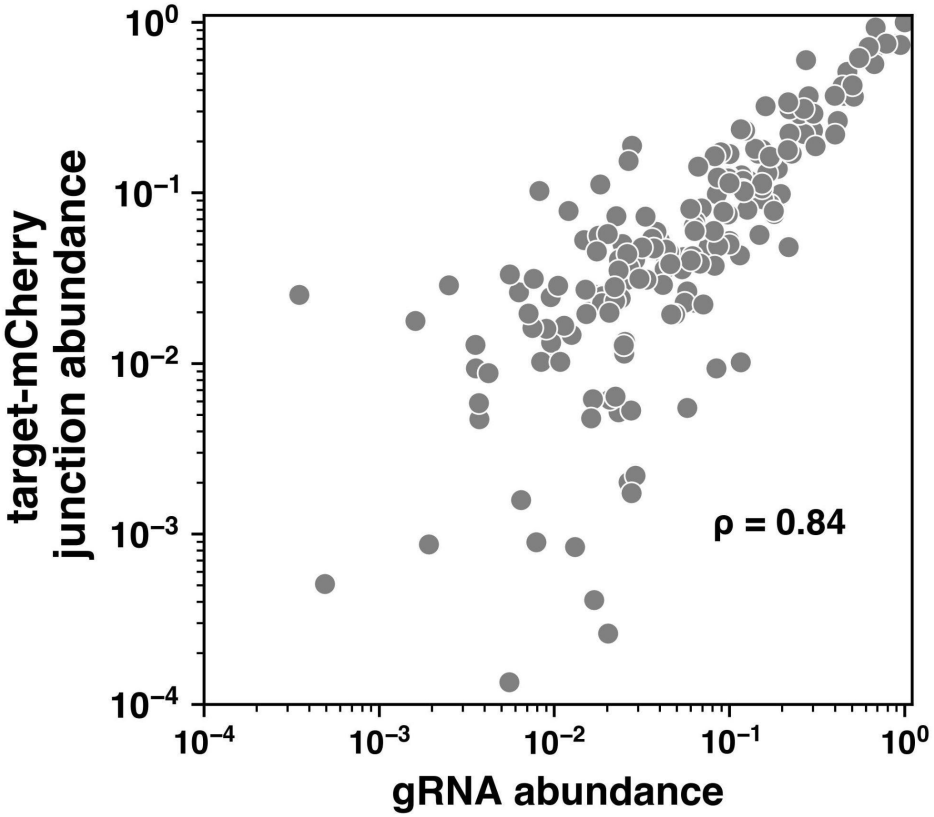
Supplementary Figure 7.



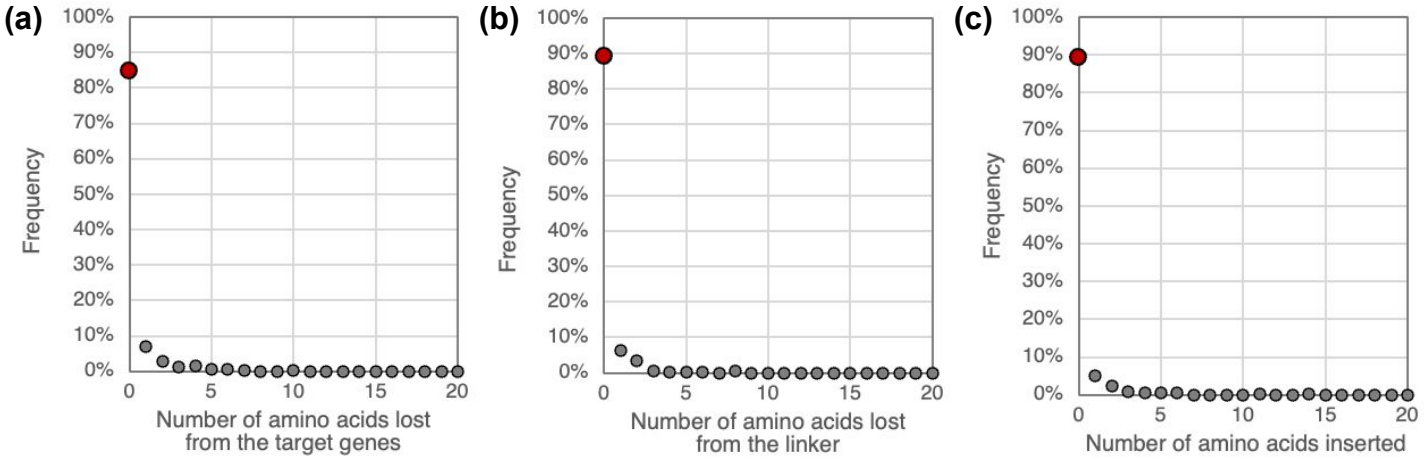
Supplementary Figure 8.



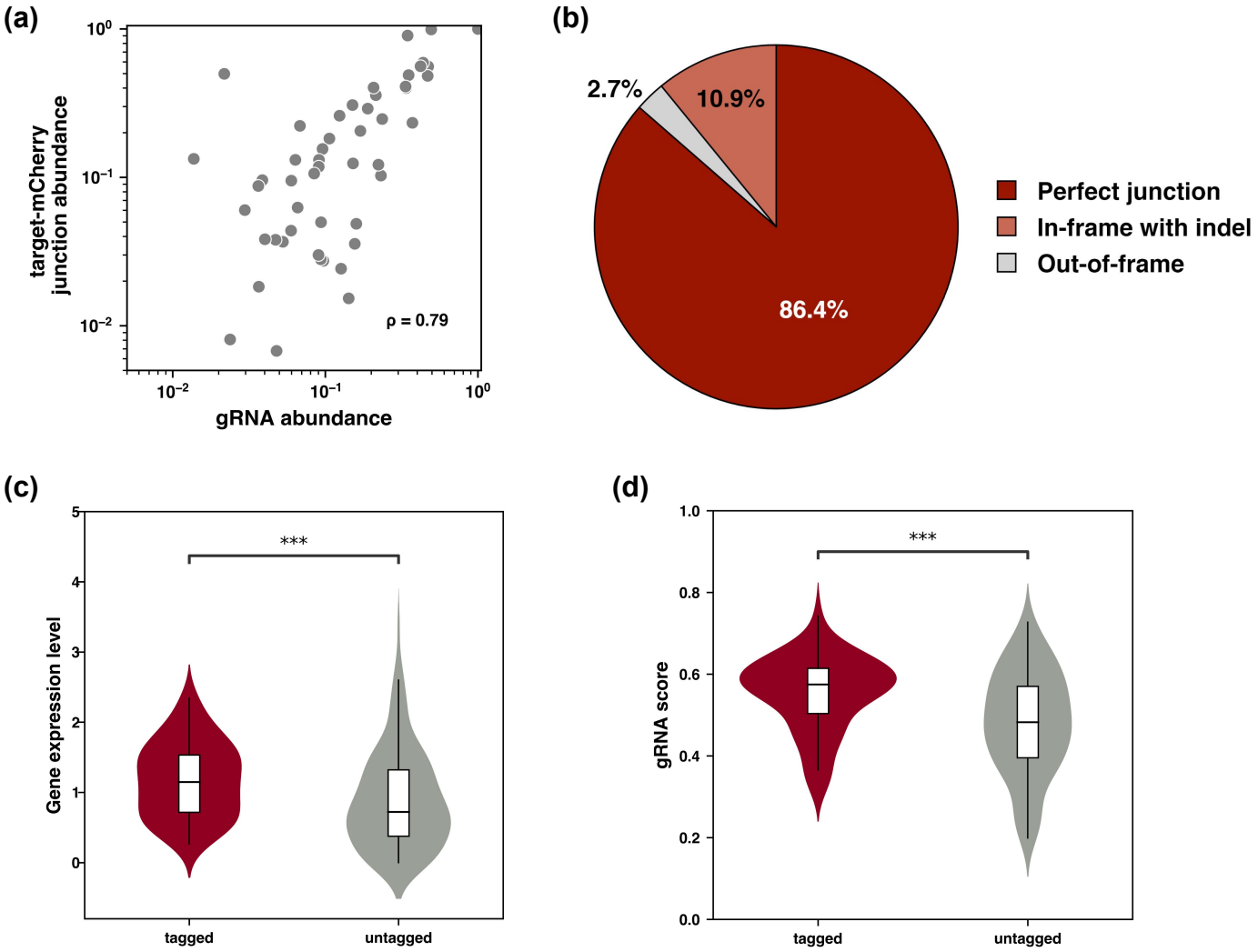
Supplementary Figure 9.



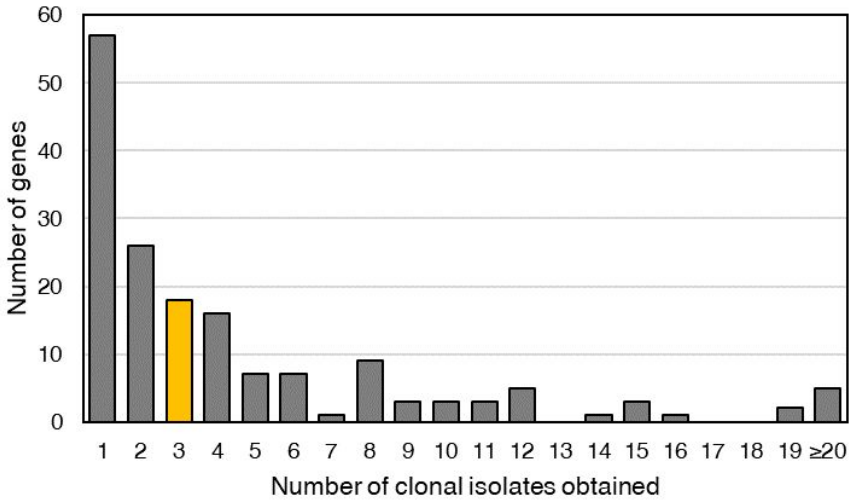
Supplementary Figure 10.



Supplementary Figure 11.



Supplementary Figure 12.

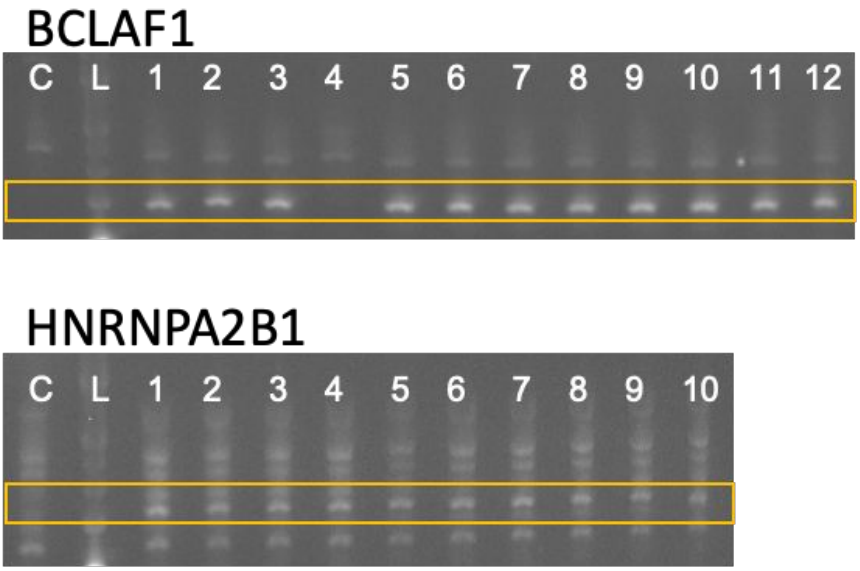


Supplementary Figure 13.

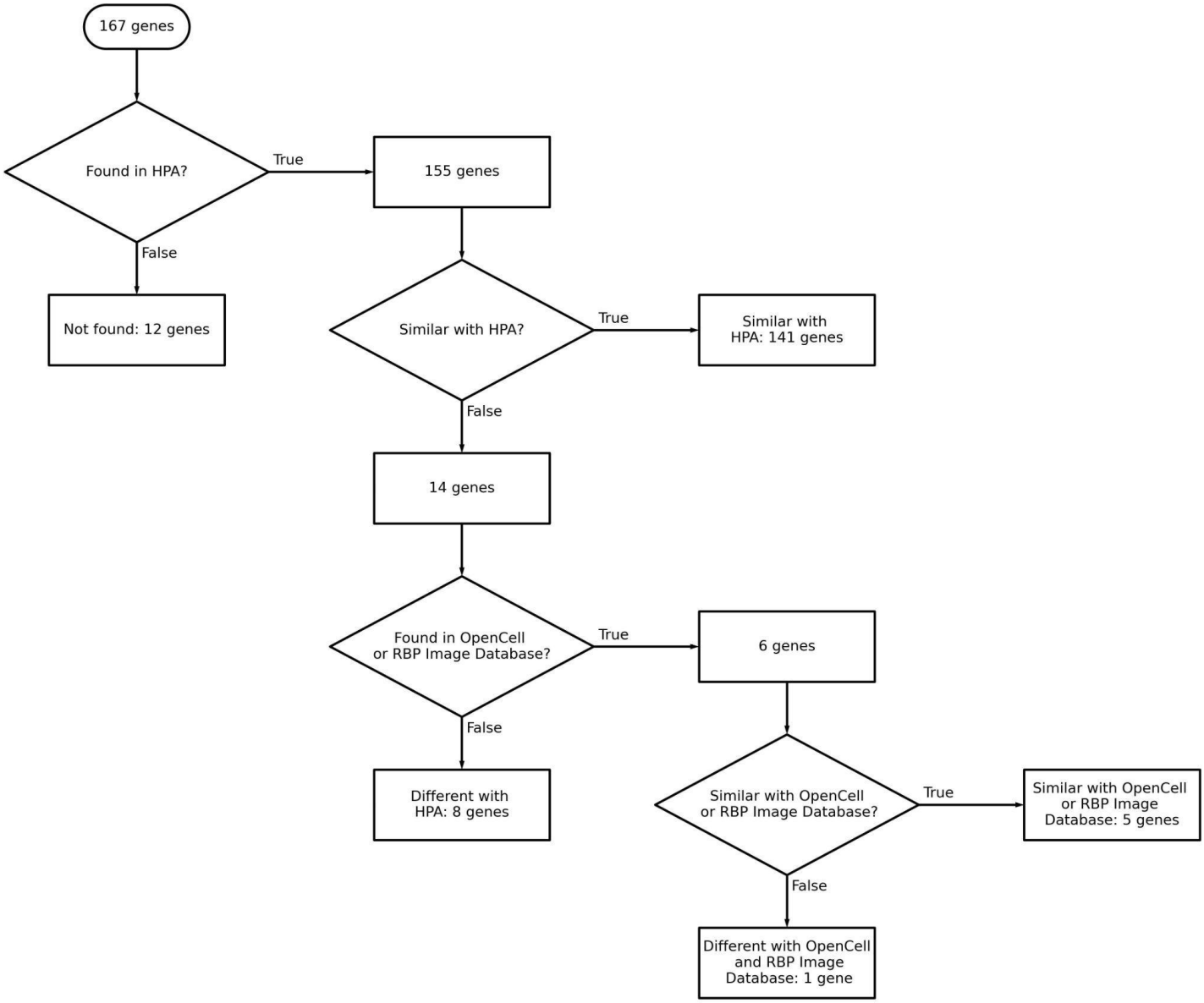
(a)

Gene	Localization	Correct % (correct/total)
AKAP8	Nucleus	**100%** (10/10)
BCLAF1	Nucleus	**91.7%** (11/12)
HNRNPA2B1	Nucleus	**100%** (10/10)
HNRNPAB	Nucleus	**85.0%** (34/40)
HNRNPC	Nucleus	**100%** (10/10)
HNRNPH3	Nucleus	**93.5%** (29/31)
ILF2	Nucleus	**100%** (13/13)
MATR3	Nucleus	**84.6%** (11/13)
MCM7	Nucleus	**90.0%** (9/10)
PPIG	Nucleus	**90.9%** (10/11)
SRSF5	Nucleus	**100%** (31/31)
CALR	ER	**100%** (16/16)
PDIA3	ER	**100%** (18/18)

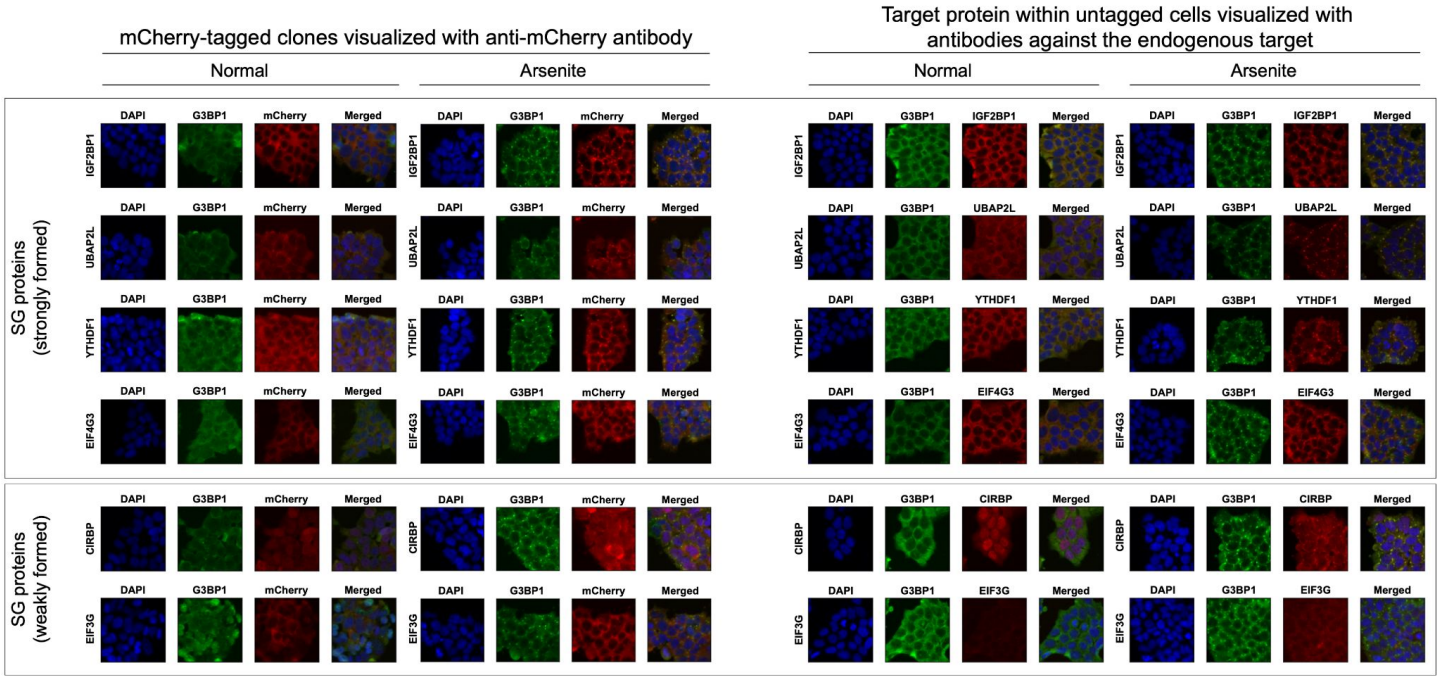
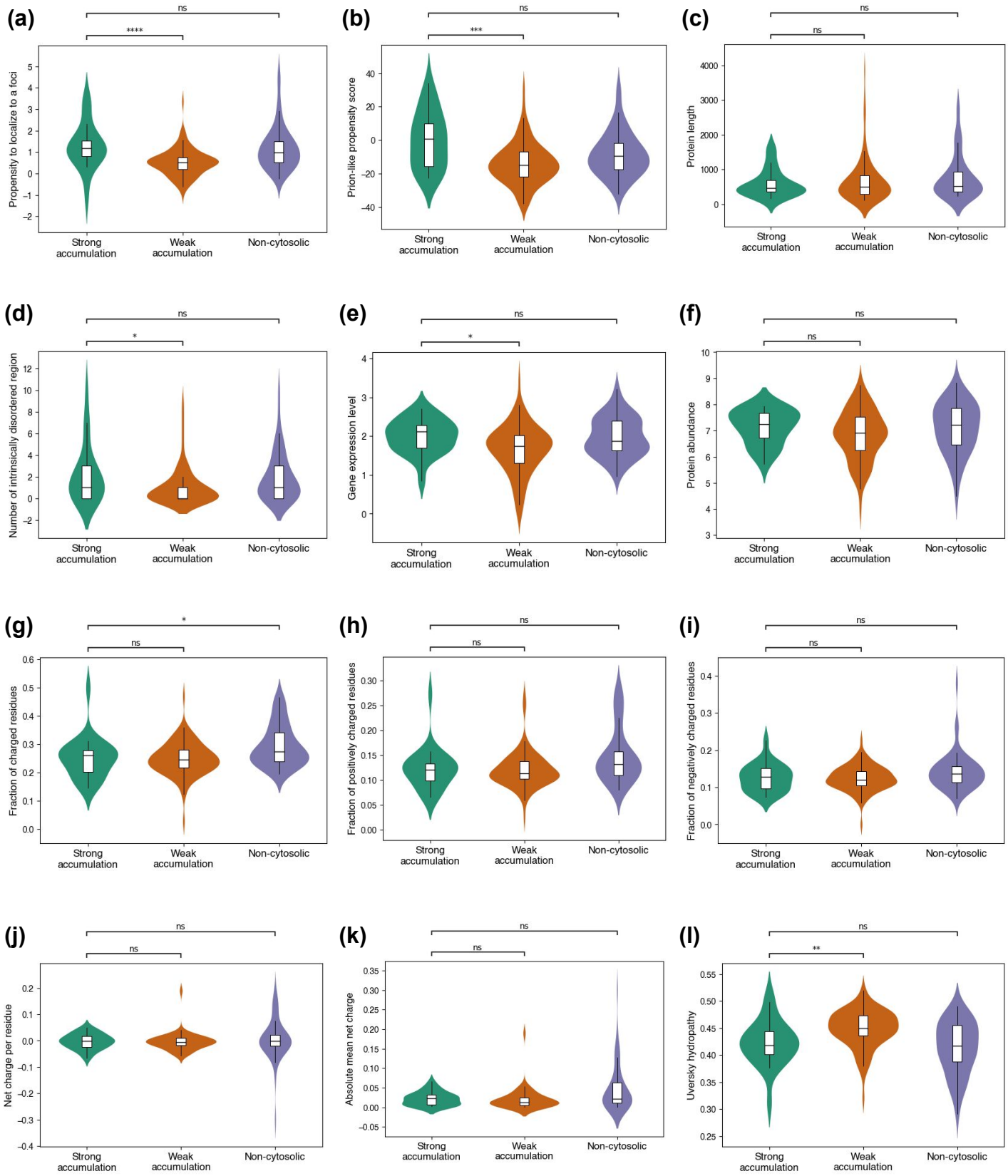
(b)



Supplementary Figure 14.



Supplementary Figure 15.



Supplementary Figure 16.