CRISPR/Cas12a-assisted PCR tagging of mammalian genes

Julia Fueller[1,*], Matthias Meurer[1,*], Konrad Herbst[1,**], Krisztina Gubicza[1,**], Bahtiyar Kurtulmus[2,**], Julia D. Knopf[1], Daniel Kirrmaier[1,3], Benjamin Buchmuller[1,4], Gislene Pereira[2], Marius K. Lemberg[1], Michael Knop[1,3,#]

[1] Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), DKFZ-ZMBH Alliance, Heidelberg, Germany
[2] Center for Organismal Studies (COS), University of Heidelberg and German Cancer Research Centre (DKFZ), DKFZ-ZMBH Alliance, Heidelberg, Germany
[3] Cell Morphogenesis and Signal Transduction, DKFZ-ZMBH Alliance and German Cancer Research Center (DKFZ), Heidelberg, Germany
[4] present address: Technische Universität Dortmund, Chemische Biologie, Otto-Hahn-Straße 4a, 44227 Dortmund

* These authors contributed equally to this work

** These authors contributed equally to this work

[#] Corresponding author: m.knop@zmbh.uni-heidelberg.de

Abstract

Here we describe a simple strategy for tagging of genes in mammalian cells. The method enables efficient creation of endogenously expressed protein fusions. Only PCR for the generation of a DNA fragment is required. This avoids the handling of RNAs, recombinant proteins or cloning of plasmids. The fragment, termed 'PCR cassette', is then transfected into cells along with a CRISPR/Cas12a helper plasmid and integrates into the target locus specified by sequences provided by the oligonucleotides used for PCR. The method is robust and works in all cell lines tested with tagging efficiency of up to 20% without selection, and up to 60% when selection markers are used.

Introduction

In mammalian cells, chromosomal 'knock-ins' for applications such as gene tagging are typically done using an endonuclease-based strategy to promote integration of the desired fragment via homology directed repair (HDR) or non-homologous end joining (NHEJ). For this, suitable reagents are required, often including recombinant proteins, RNAs, single-stranded DNA (ssDNA) or the cloning of tailored and gene-specific plasmids, to provide all the necessary components for integration (Yamamoto and Gerbi, 2018). This makes tagging rather cumbersome, time consuming and costly. In yeast genomic tagging is done using a strategy based on PCR (Baudin et al., 1993; Wach et al., 1994), now commonly referred to as 'PCR tagging'. It requires two gene-specific DNA oligonucleotides and a generic 'template plasmid' that provides the tag and a selection marker to generate a 'PCR cassette'. Directed by the oligonucleotide sequence, this PCR cassette integrates into the genome by HDR, owing to the efficient homologous recombination machinery in this species. Despite of various improvements through CRISPR/Cas9 applications, this procedure still constitutes the *de facto* standard in yeast research for rapid functional analysis using genomic modifications of genes. A similar procedure for gene tagging in mammalian cells would be highly desirable.

Here, inspired by some improvements of the method in yeast (Buchmuller et al., 2018, bioRxiv) we now describe 'mammalian PCR tagging' where we engineered the PCR tagging process making it

compatible for gene tagging applications in mammalian cells. To enhance site specific integration of the PCR cassette we incorporated a CRISPR/Cas12a (Cpf1) (Zetsche et al., 2015) based strategy, while retaining the simplicity and effectiveness of the yeast procedure. Similar to yeast, mammalian PCR tagging involves the direct transfection of a PCR cassette into cells (Fig. 1a). Generation of the PCR cassette is simple – it requires two gene specific oligonucleotides (termed M1 and M2 tagging oligos) to conduct a PCR using a template plasmid (Fig. 1b). The M1 and M2 tagging oligos provide homology arms that specify the integration site. To promote integration by homologous recombination via a double strand break near the integration site, we incorporated into the M2 oligo a CRISPR RNA (crRNA) for Cas12a. PCR with a 'template plasmid', which provides a U6 polymerase III (Pol III) promoter, generates a functional gene expressing a crRNA from the PCR cassette. Therefore, in addition to the yeast method, a Cas12a helper plasmid needs to be co-transfected to provide a source for the endonuclease (Fig. 1a).

M1 and M2 tagging oligos can be rapidly designed using an online tool (www.pcr-tagging.com). Template plasmids are generic since they can be used with any M1/M2 tagging oligo (Fig. 1b). A template plasmid also provides the desired tag (e.g. GFP) along with a terminator and can contain additional features such as a selection marker. Upon PCR, a PCR cassette is generated that contains three essential functional elements (Supplementary Fig. S1): a Cas12a crRNA gene to direct the endonuclease to the target locus, flanking homologous sequences matching the target gene, and the tag itself. For integration, the cassette is co-transfected with a plasmid encoding Cas12a. Inside the cell, the crRNA and Cas12a are expressed and assemble into a functional complex that cleaves the target gene (Fig. 1c). The resulting double stand break (DSB) then stimulates DNA damage repair. DNA repair can occur via different pathways. One option is that the DSB is repaired using the transfected PCR cassette that contains homology arms that match the region adjacent to the cleaved site. Only this yields the desired integrands expressing the appropriately tagged proteins from the target locus.

Results

Implementation of mammalian PCR tagging and optimization of procedures

To test if our approach could be used for efficient gene tagging in mammalian cells we designed tagging plasmids containing the bright green fluorescent protein mNeonGreen (Shaner et al., 2013). For tagging we selected a list of 15 genes encoding proteins with a diverse range of cellular localizations (Supplementary Table S1) with sufficiently high endogenous expression levels (Geiger et al., 2012; Schaab et al., 2012) for easy detection by fluorescence microscopy of the corresponding mNeonGreen-tagged fusion proteins by fluorescence microscopy. We co-transfected the PCR cassettes together with a Cas12a encoding plasmid into HEK293T cells and inspected the sample three days later for the presence of fluorescence. For all genes, we observed between 0.2% and 13% of fluorescent cells with the expected protein-specific localization pattern (Fig. 1d), e.g. Endoplasmic Reticulum for CANX, mitochondrial staining for TOMM20, or a diffuse and a dotted nuclear staining for HNRNPA1 and PCNA, respectively (Fig. 1e). The formation of cells with correctly localized fluorescence signal depended on the presence of Cas12a and matching combinations of homology arms and crRNA, irrespective of whether they are on the same, or different PCR products (Fig. 1f). In the presence of a crRNA for a locus different from the one targeted by the homology arms, we found very rarely cells where the cassette became integrated into the foreign locus, indicating that in addition to HDR also other integration pathways such as NHEJ are used (Fig. 1f and Supplementary Fig. S2). Together, these results establish that the crRNA is transcribed from the transfected PCR cassette and that it directs Cas12a for cleavage of the target locus. Furthermore, we conclude that the Cpf1-mediated double strand break is repaired frequently using HDR.

In addition to cells with the expected localization of the green fluorescence we observed in several transfections also cells with diffuse cytoplasmic fluorescence of variable brightness (Fig. 1d-e, see examples labeled with arrows in 1e). This fluorescence was independent on Cas12a or matching combinations of crRNA and homology arms (Fig. 1f). This indicates that the non-specific cytoplasmic signal resulted from the transfected PCR cassettes alone.

Non-specific cytoplasmic fluorescence is caused by unstable extra-chromosomal fragments

The nature of the diffuse cytoplasmic fluorescence observed in a fraction of the cells was unclear. The cytoplasmic fluorescence could originate from extra-chromosomal fragments in the nucleo-cytoplasm or fragments that have chromosomally integrated at off-target loci. To investigate the fate of the transfected fragments in the cells three days after transfection we used Anchor-Seq (Meurer et al., 2018). This method amplifies all junctions between the tagging cassettes and their local DNA neighborhoods for analysis by next generation sequencing (Fig. 2a). With this analysis we detected junctions that resulted from correctly inserted PCR cassettes (Fig. 2b), consistent with the observation of correctly localized fluorescence signal. We inspected the sequences also for signatures of off-target integrations, but could not detect any. However, the detection sensitivity was limited because of a large number of reads that did not extend beyond the sequence of the M1 or M2 tagging oligos (Fig. 2b). This suggests that they result from PCR cassettes of the transfection that are still present in the cultured cells. In addition, we also detected a substantial fraction of reads that originate from ligated ends of transfected cassettes, consistent with the idea of their recognition and 'repair' by NHEJ. Different types of fusions were detected (Fig. 2b), and the frequency of the different fusion types was not distributed as one would expect from random joining. In fact, the most frequently observed fusion type comprised a fusion of the right and the left arm of the PCR fragment (LR homo fusion, Fig. 2b). This can best be explained by an intramolecular fusion of the ends of the same PCR cassette, fragments, leading to their circularization.

To validate the occurrence of cassette fusions, we transfected into the same cells a mixture of PCR cassettes for >15 different genes. This detected hybrid-fusions between PCR cassettes for different genes, validating the idea that after transfection the cassettes are ligated together, e.g. via NHEJ mediated DNA damage repair (Fig. 2b). Nevertheless, the LR homo fusion remained the most abundant event also in the transfection of the mixture. This can best be explained by a preference for intra-molecular ligation. Together, these data support the idea that small mini-circles are the most frequent outcome of DNA repair processes upon transfection of the PCR cassettes.

In such mini-circles the crRNA gene is fused to the 3' end of the mNeonGreen sequence with the homologies of the M1 and M2 tagging oligonucleotides in between. This could yield an mNeonGreen expressing DNA element driven by the U6 Pol III promoter of the crRNA gene. The used U6 Pol III promoter has previously been shown to also mediate Pol II driven expression (Rumi et al., 2006), in which case a translation competent capped mRNA could be produced. To assess whether the nonspecific cytoplasmic fluorescence involves the translation initiation codon of mNeonGreen, we next transfected a PCR cassette where the ATG codons at position 1 and 10 of the mNeonGreen open reading frame (ORF) have been deleted and substituted with a codon for valine, respectively. This largely, but not completely, suppressed the population of cells with unspecific cytoplasmic signal, while the fraction of cells with specific localization indicative for correct gene tagging was similar to unaltered mNeonGreen (Fig. 2c). This indicates that the necessary ATG is often provided by mNeonGreen itself. Additionally, the crRNA or homology sequences may provide an ATG in frame with the mNeonGreen ORF.

'Mini-circles' are unlikely to be stable over consecutive cell divisions. We tested this hypothesis by growing transfected cells for several days. Over the course of the experiment we observed a gradual

loss of the fraction of cells with unspecific cytoplasmic fluorescence, while the fraction of cells with correctly localized fluorescence signal remained constant (Fig. 2d). Our results confirmed the transient nature of the cytoplasmic signal and argue for a general applicability of mammalian PCR tagging for targeted 'knock-in' of PCR cassettes.

Parameters influencing tagging efficiency

To explore mammalian PCR tagging methodology further, we determined tagging efficiency as a function of various parameters.

DNA delivery - We first explored basic parameters such as amount of DNA and transfection method. We found that equal amounts of Cas12a plasmid DNA and PCR cassette DNA are optimal (Supplementary Fig. S3a), whereas the transfection method did not seem to influence the outcome (Supplementary Fig. S3b). We furthermore noticed that PCR cassette purification using standard DNA clean up columns (that do not remove long oligos) can be used. However, we observed that inefficient PCR amplification resulting in the presence of significant contamination of the final product with M1 and M2 tagging oligos can potentially lower the yield of integration at the correct loci (data not shown).

Length of homology arms – From yeast it is known that approx. 28 to 36 nucleotides (nts) of continuous sequence homology are minimally required for homologous recombination of transfected DNA with the genome (Rothstein, 1991). For PCR tagging in yeast, homology arms between 45 and 55 nts in length are routinely used. To obtain some insights into the requirement in mammalian cells, we tested the integration efficiency as a function of the length of the homology arms. This revealed that already short homology arms of 30 nts on both sides allow efficient integration of the cassette (Fig. 3a), but increasing the length results in more efficient integration.

Dependence on homology arms – Our control experiment (Fig. 1f) suggested that PCR tagging depends on the presence of homology arms. However, it could still be that a fraction of the productive events is not mediated by HDR, but by alternative DNA repair pathways. To test this directly we generated a series of PCR cassettes with different types of ends. In particular, we also generated a PCR cassette with compatible overhangs for direct ligation, by using a Type II restriction enzyme (*Hga*I). This enzyme generates ends that contain 3' overhangs of 5 nts on both sides, which were designed such that they are compatible with the ends produced by Cas12a in the corresponding genomic locus (Fig. 3b, D). We observed in-frame integration of the Hga1 cut fragment, but with lower frequency when compared to the integration in the presence of homology arms (Fig. 3b). This demonstrates the requirement of homology arms for efficient integration. Insertion of the PCR cassettes via NHEJ can be observed, but it is rather inefficient.

Modified oligonucleotides – Multimerization of transfected dsDNA inside cells can be hindered when bulky modifications such as Biotin are introduced at the 5'-end of the DNA fragment. This has been reported to enhance targeting efficiency ~2-fold in *Medaka* (Gutierrez-Triana et al., 2018) and the Biotin-modification could contribute to enhance targeting efficiency in mouse embryos (Gu et al., 2018), leading to the insertion of preferentially one copy of the donor DNA. We tested M1/M2 tagging oligos with multiple phosphorothioate bonds (to prevent exonuclease degradation) with and without Biotin at the 5'-end. Synthetic oligonucleotide synthesis occurs in the 3' to 5' direction, and oligo-preparations without size selection are contaminated by shorter species without the 5'-modifications. Therefore, we additionally included size selected (PAGE purified) oligos. Overall, we obtained mixed results (Fig. 3c). For TOMM20 we observed that with increasing number of modifications the tagging efficiency increased to a maximum of 2-fold. It was irrelevant, whether the oligos were size-selected or not. However, for HNRNPA1 and also CANX the modifications did not appear to change tagging efficiency, whereas for CLTC and DDX21 again a 2- to 3-fold improvement was observed.

Importantly, however, in all cases we observed a 2- to 3-fold reduced frequency of cells with diffuse cytoplasmic fluorescence. This is consistent with the idea that this fluorescence results from ligated PCR cassettes, and that the modifications are effective in suppressing such ligations, at least partially.

Taken together, these experiments demonstrate the robustness of the procedure and dependency on homology arms for efficient recombination with the target locus, leading to the tagged gene.

Selection of clones using antibiotics resistance markers and multi-loci tagging

Next, we generated template plasmids that additionally incorporated selection markers for different antibiotics and used them to generated PCR cassettes for some of the genes shown in the previous figures, but also including five genes that we have not tagged before (Supplementary Table S1). After amplification of the PCR cassettes we used *Dpn*I or *Fsp*EI digestion to selectively destroy the DAM methylated template plasmid DNA (which also contains the selection marker). Using Zeocin or Puromycin resistance as selection markers yielded cell populations highly enriched in cells exhibiting the correct localization of a fluorescent fusion protein (Fig. 4a). The selected populations still contained cells with the non-specific cytoplasmic fluorescence, but the fraction remained constant or even decreased sometimes, indicating the labile nature of the source of this signal.

After enrichment of positive cells by Zeocin selection, single cell clones for CANX tagging were obtained by limited dilution of cells transfected with the CANX-specific PCR cassette and analyzed in detail. PCR identified in all clones correct insertion junctions on the side of the fluorescent protein tag, and in 4 out of 5 also on the rear side of the PCR cassette. Antibodies detected the corresponding mNeonGreen fusion protein (Supplementary Fig. S4a). HEK293T cells are aneuploid and appear to have up to 5 copies of the CANX gene (Lin et al., 2014). We also detected the wt copy of CANX in all clones, indicating that not all copies were tagged (Supplementary Fig. S4a).

We aimed to amplify the inserted construct using primers that bind outside of the inserted fragment. Only in one out of five clones could the inserted construct be amplified by external primers, indicating more extensive genome alterations in the other clones. Indeed, using primers that bind at both ends of the PCR cassette and that are outwards oriented we detected bands that can best be explained from concatenated fragments (Supplementary Fig. S4b). This indicates that frequently not only single cassettes, but two or more ligated cassettes are inserted into the genome. However, since a STOP codon and a transcriptional terminator accompany the inserted tag. Therefore, these additional copies should not interfere with the function of the tagged gene. Using modified M1/M2 tagging oligos (Fig. 3c) it might be possible to reduce the frequency of multimeric insertions.

Using Anchor-Seq we next investigated off target integrations. We found and validated by PCR two off target integrations of the PCR cassette in the 5 positive clones (Supplementary Fig. S4c), one just downstream of the *CANX* gene, which resides on chromosome 5, and one site on chromosome 1, which contained an insertion in two clones. This indicates the occurrence of multiple insertions in the same clone, maybe caused by off-target activity of the crRNA.

To gain insight into the frequency of multiple tagging events, we generated for CANX and HNRNPA1 two PCR cassettes each, one for tagging with the red fluorescent protein mScarlet-i and one with mNeonGreen, respectively. The resulting four cassettes were then co-transfected into HEK293T cells in mixtures of pairs of two, using all four possible red-green and gene-gene combinations. This detected three types of cells, with green, red, or green and red fluorescence in the nucleus or the Endoplasmic Reticulum (ER) respectively, as shown for the example of the *HNRNPA1-mScarlet-i/HNRNPA1-mNeonGreen* transfection (Supplementary Fig. S5). The frequency of each of the three types of cells was roughly equal, no matter whether the same or two different genes were tagged

(Fig. 4b). This indicates high double tagging efficiency of different loci, and demonstrates that often more than one allele is tagged. This suggests applications of PCR tagging for the analysis of protein-protein interactions using epitope tagging, or protein co-localization using different fluorescent proteins. We validated this in different double tagging experiments (Fig. 4c), which demonstrated simultaneous detection of various cellular structures within one transfection.

Together, this analysis demonstrates that all positive clones contain insertions by homologous recombination that yield the correct fusion protein. Insertions are not necessarily single copy, but concatenated segments of ligated tagging cassettes. Since the PCR cassette provides STOP codon and a transcriptional terminator along with the tag, the generated transcript is properly defined.

Applications of PCR tagging: different cell lines

So far, we have provided a robust workflow for chromosomal tagging in HEK293T cells. To challenge the general applicability of PCR tagging, we tested additional human but also murine cell lines to tag genes already tagged successfully in our initial experiments. In each cell line we identified for most genes cells that showed correctly localized green fluorescence, with a frequency of 0.2 to 5% (Fig. 5a-d). Examples of tagged murine myoblast (C2C12) cells are shown in Supplementary Fig. S6a. For HeLa cells, we additionally subjected the cells to selection, and found up to 40% of cells exhibiting the correct localization (Supplementary Fig. S6b). In conclusion, these results demonstrate that PCR tagging works for different cell lines and species, including differentiated and stem cells.

crRNA design, PAM site selection and genomic coverage

Next, we asked how well Cas12a-targeted PCR tagging covers the human genome. Our tagging approach relies on relatively short homology arms of the PCR cassette. This constrains the target sequence space, since cleavage of the target locus must be inside the area of the homology arms, leaving enough sequence for recombination. Second, insertion of the cassette needs to destroy the crRNA cleavage site, in order to prevent re-cleavage of the locus. For C-terminal protein tagging these criteria confine potentially useful protospacer-associated motif (PAM) sites to a region of 17 nts on both sides of the STOP codon including the STOP codon, with the PAM site or protospacer sequence overlapping the STOP codon (Fig. 6a). So far, we have used Cas12a from *Lachnospiraceae bacterium* ND2006 (LbCpf1) (Zetsche et al., 2015), but PAM sites that are recognized by this Cas12a (TTTV) (Gao et al., 2017) and that are located in this area of a gene are relatively infrequent and would allow C-terminal tagging of about one third of all human genes (Fig. 6b). To increase this number we first tested different Cas12a variants with altered PAM specificities (Gao et al., 2017). The results demonstrated that other variants and PAM sites are also functional and can be used for PCR tagging (Fig. 6c). Considering these Cas12a variants renders approx. 72% of all human genes accessible for C-terminal PCR tagging (Fig. 6b). To increase this number further we extended the search space for suitable PAM sites into the 3'-UTR (typically 50 nts) (Fig. 6a) and adjusted the design of the M2 tagging oligo such that a small deletion occurs that removes the binding site of the crRNA. Since tagging introduces a generic terminator for proper termination of the tagged gene, this small deletion is unlikely to have an impact on the tagged gene. Considering the extended search space and the currently available palette of Cas12a variants (Fig. 6b) we calculated that potentially 98% of all human ORFs are amenable for C-terminal PCR tagging.

PCR tagging toolkit for mammalian cells

Our results outline PCR tagging as a rapid, efficient and cost-effective procedure facilitating chromosomal knock-ins of large DNA fragments in mammalian cells, e.g. for C-terminal tagging or gene disruption. To facilitate application of the method for various purposes we set up a webpage for oligo design (Fig. 1b). The online tool (www.pcr-tagging.com) requires as input the genomic DNA sequence around the desired insertion site, i.e. the STOP codon of the gene of interest for C-terminal tagging. The software then generates the sequence of the M1 oligo, which specifies the junction between the gene and the tag. Next, the software identifies all PAM sites for the available Cas12a variants and uses these to generate crRNA sequences and to assemble corresponding M2 tagging oligos. M2 tagging oligos are designed such that the integration of the PCR cassette does lead to a disruption of the crRNA binding site or PAM site in order to prevent re-cleavage of the locus. M2 tagging oligos are then ranked based on the quality of the PAM site and the presence of motifs that might interfere with crRNA synthesis or function. M1/M2 tagging oligos can be used with template plasmids based on different backbones: either without a marker, with the Zeocin or with the Puromycin resistance gene (Fig 7a). We generated a series of template plasmids containing different state of the art reporter genes (Table 1, examples shown in Fig. 7b).

Ongoing efforts continue to improve optimal crRNA prediction and to eliminate crRNAs with potential off-target binding activity. The current version of the server already allows to flexibly add novel Cas12a variants, by adjusting PAM site specificity and the sequence of the corresponding constant region of the crRNA.

In conclusion, Cas12a-mediated PCR tagging of mammalian genes using short homology arms is a rapid, robust and versatile method enabling endogenous gene tagging. The versatility of the method suggests many types of applications for functional or analytical gene and protein studies in mammalian cells.

Discussion

In this paper we demonstrate efficient targeted integration of DNA fragments of several kb in size into the genome of mammalian cells, guided by short homology arms. Integration is assisted by CRISPR/Cas12a and a crRNA that is expressed from the DNA fragment itself. This enables a PCR-only strategy for the production of the gene specific reagents for tagging of cell lines, thus allowing quick and low-cost experimentation. We developed a software tool for oligo design and established streamlined procedures for application in several cell lines.

PCR tagging is potentially useful to disrupt gene function. We tested this by generating a PCR cassette to disrupt genes by inserting STOP codons and a terminator and found that this did work as well (data not shown), making the method also applicable for gene KO studies.

Beyond mammalian cells, there may be other species where this strategy could boost tagging methodology, i.e. many fungal species that require a DNA double strand break for targeted integration of a foreign DNA fragment.

PCR tagging can be easily up-scaled and parallelized – since it needs only two oligonucleotides per gene. In yeast where PCR tagging is very efficient even in the absence of an endonuclease the ease of up scaling permitted the creation of many types of genome wide resources where all genes were modified in the same manner, i.e. by gene deletion or by tagging with a fluorescent protein or affinity tag (Gavin et al., 2002; Ghaemmaghami et al., 2003; Huh et al., 2003; Meurer et al., 2018; Winzeler et al., 1999).

The use of tagged genes always raises the question about the functionality of the tag-fusion. Here, two questions matter: How does tagging affect gene regulation, and how does it affect protein

function? Many aspects of protein tagging have been discussed in literature, i.e. from functional or structural points of view. But ultimately, one has to be aware of the fact that a cell expressing a tagged gene is a mutant, and that the tag does not necessarily report correctly about the behavior of the untagged protein. As part of good laboratory practice this demands for some sort of phenotypic analyses to investigate the functionality of the tagged gene/protein and/or orthogonal experiments to obtain independent validation of the conclusions that were derived with the tagged clone(s).

In yeast. genome wide C-terminal epitope tagging of haploid yeast cells revealed that >95% of the ~1000 essential yeast genes, when endogenously tagged with a large tag such as a fluorescence protein reporter, retain enough functionality to not cause an obvious growth phenotype under standard growth conditions (Khmelinskii et al., 2014)

When using PCR tagging, it needs furthermore to be considered that C-terminally inserted tag is accompanied by a generic transcription termination site that replaces the endogenous 3' untranslated region (UTR).

Various methods for gene tagging with long DNA fragments in mammalian cells have been developed (Agudelo et al., 2017; He et al., 2016; Lackner et al., 2015; Merkle et al., 2015; Suzuki et al., 2016; Zhang et al., 2017; Zhu et al., 2015). Besides methods that are tailored for particular DNA repair pathways such as NJEJ, also classical approach to use long homology arms to direct insertion via the repair of a double strand brake has been used in combination with CRISPR/Cas9 or other endonucleases in various implementations, using circular or linear repair templates which can be generated *ex vivo*, or *in vivo* upon endonuclease excision of the repair template. Because of low efficiency or the use of alternative repair pathways, often substantial number of clones need to be screened in order to obtain a few correct ones (Koch et al., 2018). In non-germline cells, the insertion precision seems to be not always satisfactory and errors such as small in-dels are observed frequently near one or the other side of the inserted fragment, thus compromising the sequence of the tagged gene. Since PCR tagging relies on a heterologous terminator to terminate transcription of the tagged gene, the insertion precision of the down stream end of the PCR cassette is rather unimportant, and, if erroneous, will only affect 3'UTR regions of the gene, which is not used for the tagged allele. Obviously, this constitutes a compromise, and bears the possibility that important gene regulatory sequences are omitted from the tagged gene. While for mammalian cells no global data set about the regulatory impact of the 3'UTR on gene expression is available, data from yeast, where seamless tagging was compared with tagging using a generic 3'UTR, demonstrated that only about 11% of the genes were impacted in their expression more than 2 fold (Meurer et al., 2018).

If it is essential to retain the endogenous 3' UTR, the PCR tagging strategy can be modified so that the crRNA gene and the tag are on different PCR fragments (see Fig. 1f). In this case the PCR product of the tag can be tailored to integrate seamless into the target site, without any additional sequence.

PCR tagging is highly efficient, as it is easy to obtain enriched populations containing the correct gene fusion. Nevertheless, the presence of high NHEJ activity complicates the situation, as shown in our detailed analysis using *CANX* tagged clones where we detected inserted tandem fusions (Supplementary Fig. S4). Given the fact that enriched populations are composed from many different clones, it is possible to use such populations for a rapid first assessment of an experimental questions, for example the localization of an endogenously expressed protein in a specific condition, environment or cell line, by simply scoring multiple cells. Since they are derived from different clones, clone-specific effects can be spotted rapidly and considered in the analysis. This avoids the need of perfectly characterized cell lines with exactly the intended genomic modification. Depending on

experimental requirements, individual lines can then be isolated for detailed characterization prior to further experimentation.

The toolset available for PCR tagging can easily be expanded by constructing new template plasmids. Maintaining a certain level of standardization such as the preservation of the primer annealing sites for the M1 and M2 tagging oligos in new template cassettes, makes it is possible to re-use already purchased M1/M2 tagging oligos of the same gene for many different tagging experiments.

Further improvements of the tagging efficiency might be possible. While we found that inhibitors of NHEJ did not exhibit a positive impact on the integration efficiency (data not shown), it might be possible to target the repair template to the CRISPR endonuclease cut site (Gu et al., 2018; Roy et al., 2018), or further enhance Cas12a expression, to improve tagging efficiency further.

In conclusion, PCR mediated gene tagging has the potential to impact how research is done in an entire field, not only because of the simplicity of the method, but also because the required reagents are easy to handle, cost effective, and freely exchangeable. Moreover, PCR tagging is quicker than the construction of a plasmid for transient transfection, while simultaneously preventing the danger of studying overexpression artifacts.

With PCR tagging at hand, many different and exciting experimental avenues are becoming possible, from the rapid assessment of protein localizations to high throughput localization studies of many proteins.

Competing financial interest statement

The authors declare no competing financial interests.

Author contribution

M.K. and M.M designed the project and together with M.K.L and J.F. designed the experiments. J.F., K.H., M.M., B.K., J.D.K. and D.K. performed the experiments. K.H. analyzed the NGS data, K.G. wrote the web-tool for primer design. All authors analyzed the data and discussed the results. M.K. wrote the manuscript with input from all authors.

Acknowledgements

The authors wish to thank Cyril Mongis for help with IT infrastructure, Anne Schlaitz and Frauke Melchior for critical reading of the manuscript. We acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, grant KN498/12-1), the Collaborative Research Center SFB1036, the state of Baden-Württemberg through bwHPC for high-performance computing and SDS@hd for data storage (grant INST 35/1314-1 FUGG). K.H. is supported by a HBIGS graduate school fellowship. G.P. and B.K are supported by the Collaborative Research Center SFB873 and the Heisenberg Program of the DFG (granted to G.P.).

Materials and Methods

Plasmids and oligos

Plasmids are listed in Table 1 and Supplementary Tables S2 and S3. Sequences are provided for download. Plasmids can soon be obtained from www.addgene.org. All used oligos for cloning, Anchor-Seq and gene tagging are listed in Supplementary Table S4.

Construction of template cassettes

All clonings were performed by standard restriction enzyme digests or oligo annealing and ligations using enzymes from NEB. Most of the elements inside the template cassettes (M1-mNeonGreen-SV40polyA-ZeocinR-BGHpolyA-hU6promoter) were custom synthesized (gBlock, IDT) and cloned via *Bsi*WI and *Xba*I into a *Bsi*WI and *Spe*I cut pFA6a backbone. The SV40 promoter was cloned separately into the cassette via *Sal*I and *Eco*RI, since it contains repeats and could not be synthesized together with the other elements. In addition to the ZeocinR marker we have also introduced a PuromycinR marker. Because the standard DNA sequence for this marker is very GC-rich and difficult to amplify by PCR, we synthesized a new version with lower GC-content and cloned it via *Eco*RI and *Pst*I into the cassette. For a cassette without a marker the SV40promoter-ZeocinR-BGHpolyA sequence was removed by digest with *Sal*I and *Xho*I and subsequent religation of the backbone. This resulted in 3 different plasmids based upon the backbone pFA6 (see Fig. 7a).

The mNeonGreen ORF of these template plasmids is flanked by unique restriction sites and is therefore easily exchangeable. For introduction of new tags *Bam*HI and *Spe*I sites can be used. For a high flexibility in cloning, the sticky ends of both restriction sites are compatible to sticky ends produced by other enzymes (*Bcl*I/*Bgl*II and *Avr*II/*Nhe*I/*Xba*I, respectively).

All tags listed in Table 1 and Supplementary Table S2 are cloned either by amplification from template plasmids with oligos containing restriction sites or by annealing of two oligos and are ligated into *Bam*HI/*Spe*I cut backbones of pMaM523/526/541 (for detailed information see Supplementary Table S2) to retrieve template cassettes called pMaCTag (**p**lasmid for **Ma**mmalian **C**-terminal **Tag**ging) with the following naming scheme:

pMaCTag-xy: Tag xy, no marker, pMaM526 backbone

pMaCTag-**Z**xy: Tag xy, **Zeocin** marker, pMaM523 backbone

pMaCTag-**P**xy: Tag xy, **Puromycin** marker, pMaM541 backbone

M1 and M2 tagging oligo design

The online oligo design tool (www.pcr-tagging.com) was implemented using Shiny. The interactive web application was developed in R v3.4.4 (R Core Team, 2014) with the R packages shiny v1.1.0 (Chang et al., 2018) and shinyjs v1.0 (Attali, 2017). The R package Biostrings v2.46.0 (Pagès et al., 2017) is used for searching PAM sites. The latest code is available from our GitHub repository (www.github.com/knoplab). Oligo design principles are as follows:

M1 oligo: The design of the M1 oligo is straight forward as it contains only two functional elements: the primer annealing site for PCR, which is constant in all template cassettes (TCAGGTGGAGGAGGTAGTG), and the sequence of the homology arm, which is derived from the target locus.

***Example*: M1 tagging oligo (for TOMM70):**

Description of elements:
5'-homology (90 bases before the insertion site, direct orientation)---primer annealing site for PCR

Sequence:
ATGGAGATGGCCCATCTGTATTCACTTTGCGATGCCGCCCATGCCCAGACAGAAGTTGCAAA
GAAATACGGATTAAAACCACCAACATTATCAGGTGGAGGAGGTAGTG

M2 tagging oligo: The design of the M2 tagging oligo is more complex. It contains the annealing site for PCR (GCTAGCTGCATCGGTACC), the direct repeat sequence of the crRNA, which is

Cas12a-variant specific, and the protospacer sequence of the crRNA, which depends on available PAM sites at the target locus, a terminator for the Pol III RNA polymerase and the homology arm, as outlined below.

***Example*: M2 tagging oligo (for TOMM70):**

Description of elements:

3'-homology (55 bases after the insertion site, reverse orientation)---Pol III terminator---crRNA protospacer sequence---crRNA direct repeat sequence---primer annealing site for PCR

Sequence:

CAGTTGAAGAGGGGGTAAACTTTTAAAAAGAGGGTCAGTCTGCTTTCCCCCTGTTAAAAAAAG
TCTGCTTTCCCCCTGTTTATCTACAAGAGTAGAAATTAGCTAGCTGCATCGGTACC

Criteria used for ranking crRNAs currently implemented in www.pcr-tagging.com, listed according to priority:

1. Location of the crRNA binding site in the genome in a region where it becomes destroyed upon cassette integration in order to prevent re-cleavage. This can be on either side in close proximity of the insertion site (17 nts up and downstream of the insertion site). If no suitable crRNA binding site is found in this confined search space, the software offers the option to select PAM sites in the 3'-region of the insertion site (extended search space). In this case the design of the homology arm of the M2 tagging oligo is adjusted in such a manner that the target site of the crRNA is deleted. This results in a small deletion in the 3' UTR of the gene after the insertion site of the cassette. Since the PCR cassette contains a transcriptional terminator, we deem this to be non-critical. With these criteria, it is possible to design suitable crRNAs for C-terminal tagging of the vast majority of mammalian genes (Fig. 6b).
2. The protospacer sequence should preferably not contain 4 or more 'T'-s in a row, since this might lead to premature termination of the Pol III transcription of the crRNA (Arimbasseri et al., 2013). In practice, we observed that crRNAs with 'TTTT' are frequently functional.
3. PAM sites are ranked according to literature (Gao et al., 2017; Zetsche et al., 2015). In addition, unconventional PAM sites were considered (MCCC for AsCas12a and RATR for LbCas12a), based on depositor comments on the Addgene webpage. For ranking crRNAs, conventional PAM sites are preferred.
4. If multiple crRNAs are fulfilling these criteria, they are ranked according to position of the STOP codon, with a preference for closer distance after the STOP.

Synthesis of M1 and M2 tagging oligo

All M1 and M2 tagging oligos were obtained from Sigma-Aldrich using a 0.05 µmole synthesis scale and are RP1 cartridge purified, unless otherwise stated (as in Fig. 3c).

PCR of template cassettes using M1 and M2 tagging oligos

PCR using long oligos is not always easy and requires optimized protocols. We routinely use a self-purified DNA polymerase for PCR (Pfu-Sso7d (Wang et al., 2004)). Alternatively, for M1 and M2 PCR also commercial high-fidelity polymerases can be used. We have tested Phusion (ThermoFisher) and Velocity polymerase (Bioline). We note that the Phusion polymerase using the buffer provided by the manufacturer does not work for PCR cassette amplification with M1 and M2 tagging oligos, whereas good amounts of the product are obtained for Velocity polymerase using the buffer provided by the manufacturer.

We found that all polymerases work well using the buffer conditions and amplification scheme shown below, yielding similar amounts of PCR cassette.

PCR mixture:

- 5.0 µl of 10x HiFi-buffer (200 mM Tris-HCl, pH 8.8; 100 mM $(NH_4)_2SO_4$, 500 mM KCl, 1% (v/v) Triton X-100, 1 mg/ml BSA, 20 mM $MgCl_2$)
- 5.0 µl of dNTPs (10 mM stock, Bioline, BIO-39026)
- 1.0 µl of $MgCl_2$ (50 mM stock)
- 5.0 µl of betaine (5 M stock, Sigma-Aldrich, 61962)
- 0.3 µl of template DNA (200 ng/µl stock)
- 2.5 µl of M1 tagging oligo (10 µM stock)
- 2.5 µl of M2 tagging oligo (10 µM stock)
- x µl of H_2O up to 50 µl
- 1 µl self-purified DNA polymerase (1 U/µl), 0.5 µl Phusion or 0.25 µl Velocity polymerase

PCR was mixed on ice and was carried out in a Biometra TRIO (Analytik Jena) using the following program:

- 3 min at 95 °C
- 30 cycles of:
    - 20 s at 95 °C
    - 30 s at 64 °C
    - XX s at 72 °C (45 s per kb) (see Supplementary Table S2)
- 5 min at 72 °C
- 4 °C
- After PCR, 0.4 µl *Dpn*I or *FspEI* (and 1.67 µl Enzyme activator) was added to the reaction mixture and incubated at 37 °C for 1 h to digest the template that contains a selection marker that would contaminate the transfection.
- PCR products were analyzed by agarose gel electrophoresis and purified using column purification (Macherey-Nagel).

Note: Sometimes a particular pair of oligos does not yield a product upon PCR. In this case it is worth testing whether adding 2 min. on top of the calculated elongation time does solve the problem. If not, it might be that synthesis of the primer went wrong. To determine the faulty primer, pair-wise PCR with established M1 M2 primers can be used to identify the faulty primer. Usually, ordering the same primer again solves the problem. Providers may wave the cost of re-ordering.

Preparation of genomic DNA

Genomic DNA was isolated from HEK293T cells according to a protocol adapted from Sambrook et al. (Greene and Sambrook, 2012). After washing with PBS, a confluently grown 6-well was lysed in 600 µl SNET buffer (20 mM Tris pH 8.0, 400 mM NaCl, 5 mM EDTA pH 8.0, 1% SDS) and 2 µl of RNase A (10 mg/ml RNAse A, 10 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$) was added for 30 minutes at room temperature. Afterwards Proteinase K (20 mg/ml Proteinase K, 50 mM Tris-HCl pH 8.0, 1.5 mM $CaCl_2$, 50% glycerol) was added for another 30 minutes at room temperature. Proteins were precipitated using 200 µl 3 M K-Acetate solution, followed by precipitation of the DNA with Isopropanol and washing with 70% Ethanol. DNA was dried and dissolved in TE (10 mM Tris, 1mM EDTA) buffer.

Next-generation sequencing of genomic DNA with Anchor-Seq

Sequencing libraries for cassette integration sites were prepared based on our previously published Anchor-Seq protocol (Meurer et al., 2018) with some modifications to the adapter design to include

unique molecular identifiers (UMIs) (Supplementary Table S4; Buchmuller et al., 2018, bioRxiv). Quantified libraries were sequenced on a NextSeq 550 sequencing system (Illumina) with a spike-in of 20% phiX gDNA. Raw reads were trimmed from technical i.e. adapter and cassette sequences using custom scripts (Julia v.0.6.0 and BioSequences v0.8.0). The trimmed reads were aligned to the human reference genome (Genome Reference Consortium Human Build 38 for alignment pipelines, ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/) using bowtie2 (v2.3.3.1 (Langmead and Salzberg, 2012)). Template cassette sequences were included in the reference genome as decoy. Aligned reads were grouped with UMI-tools (Smith et al., 2017) based on unique molecular identifiers included in the Anchor-Seq adapters. Enriched integration sites were further evaluated and counted using IGV (v2.4.10 (Robinson et al., 2011)).

Cell counting and Fluorescence microscopy

For Fig. 5b-d - Cells were grown on coverslips (No. 1.5, Thermo Fischer Scientific), washed once with PBS and fixed with 3% PFA for 10 minutes at 37 °C. After fixation, coverslips were washed 3 times with PBS, incubated in PBS containing 0.1 µg/ml 4 ', 6-Diamidin-2-phenylindole (DAPI) for 10 minutes and embedded in Mowiol. Coverslips were coated with 0.1% gelatin Type B (Sigma Aldrich) for culturing C2C12 and 0.2% gelatin Type A (Sigma Aldrich) for C2C12 and mES cells, respectively. Images of RPE-1 and C2C12 cells were acquired as Z stacks using Zeiss Axio Observer Z1 equipped with 40x NA 1.3 PlanNeo oil immersion objective, and AxioCam MRm CCD camera using ZEN software. Images of mESC colonies were acquired as Z stacks using Nikon A1R confocal microscope equipped with Nikon Plan Apo λ 20x NA 0.75 objective, using NIS elements software. Maximum intensity projections of the Z stacks were prepared using Fiji (Schindelin et al., 2012; Schneider et al., 2012).

For cell counting, random fields of view were inspected in the HOECHST/DAPI channel and all nuclei present in the entire field of view were counted. Cells containing transfected fluorescent protein expressing cassettes were then counted subsequently in the same fields of view using the appropriate illumination wavelengths. In some experiments counting was done using images recorded in the same manner.

For Fig. 4c images were taken with Zeiss LSM 780 confocal microscope using a Plan-APOCHROMAT 63x, 1.40 NA Oil Objective (panels i-iii) or a Leica Spinning DMi8 Spinning Disk microscope with HC PL APO 63x, 1.40 NA Oil Objective (panel iv).

For all other Figs. For live cell imaging, cells were splitted 24h after transfection into 8 well µ-slides (Ibidi). Analyses of transfected cells were performed 3 days after transfection or as described in the figure legends. Cells were stained with Hoechst 33342 (4 µg/ml in PBS, Thermo Fisher Scientific) for 5 minutes and then the medium was changed to FluoroBrite (Thermo Fisher Scientific) supplemented with 10% FBS (Gibco) and 20 mM HEPES-KOH, pH 7.4 (Thermo Fisher Scientific).

For counting and imaging different microscopes were used: Nikon Ti-E widefield epifluorescence microscope or a DeltaVison with each 60x oil immersion objectives (1.49 NA, Nikon, 1.40 NA, DeltaVision). Z-stacks of 11 planes with 0.5 µm spacing were recorded with 100 ms exposure time. Single plane images and maximum intensity z-projections are shown. Subcellular localizations were identified and scored visually.

Western Blotting

Cells were solubilized in SDS sample buffer (50 mM Tris-Cl pH 6.8; 10 mM EDTA, 5% glycerol, 2% SDS, 0.01% bromphenol blue) containing 5% β-mercaptoethanol. All samples were incubated for 15 min at 65 °C. Denatured and fully-reduced proteins were resolved on Tris-glycine SDS-PAGE

followed by western blot analysis using the following antibodies: rat monoclonal anti-HA (11867423001; Roche), mouse monoclonal anti-V5 (V8012; Sigma), anti-S-tag mouse monoclonal antibody (MA1-981; Thermo Fisher), rabbit polyclonal anti mNeonGreen Tag (53061S, Cell Signaling), rabbit anti Calnexin (ab22595; abcam).

Tissue culture

h-TERT-immortalized Retinal Pigment Epithelial (RPE-1, ATCC, CRL-4000, USA) cells were grown in DMEM/F12 (Sigma Aldrich) supplemented with 10% fetal bovine serum (FBS, Biochrom), 2 mM L-glutamine (Thermo Fisher Scientific) and 0.348% sodium bicarbonate (Sigma Aldrich). Mouse myoblast C2C12 cells (gift from Edgar R. Gomis, iMM, Portugal) were grown in DMEM High Glucose (Sigma Aldrich) supplemented with 20% fetal bovine serum (FBS, Biochrom). Mouse embryonic stem cell line E14 (gift from Frank van der Hoeven, DKFZ, Germany) were grown in Knockout DMEM (Thermo Fisher Scientific) supplemented with 10% ESC qualified FBS (Thermo Fisher Scientific), 2 mM GlutaMax (Thermo Fisher Scientific), 0.1 mM β-mercaptoethanol, 10^3 units of murine leukemia inhibitory factor (LIF from ESGRO, Millipore). mES cells were grown under feeder-free conditions on 0.2% gelatin Type B coated dishes (Sigma Aldrich).

HEK293T, HeLa and U2OS cells were grown in DMEM High Glucose (Life technologies) supplemented with 10% (vol/vol) fetal bovine serum (Gibco).

All cell lines were grown at 37 °C with 5% CO_2, and regularly screened for mycoplasma contamination.

Selection was performed using 1 µg/ml Puromycin (Sigma Aldrich) or 500 µg/ml Zeocin (Invitrogen) for HEK293T cells. For HeLa cells 300 µg/ml Zeocin was used.

Transfection

Chemical transfection - Transfection of HEK293T, HeLa and U2OS cells was performed using Lipofectamine 2000 (Invitrogen) according to protocol of the manufacturer and using a 24-well format. If not otherwise described, 500 µg Cas12a Plasmid and 500 µg of the PCR cassette were used for transfection of one well in a 24-well plate.

Electroporation - Plasmids containing Cas12a variants and PCR cassettes were electroporated into RPE-1, C2C12, and mESCs using 2 mm gap cuvettes and NEPA-21 electroporator (Nepa Gene, Japan) according to manufacturer's instructions. OPTI-MEM (Thermo Fisher Scientific) was used as electroporation buffer.

For electroporation of HEK293T cells the Neon Transfection System (Thermo scientific) was used according to the protocol of the manufacturer using 2 pulses of 20 ms and 1150 V.

Colony picking and generation of clonal lines

After Zeocin selection cells were trypsinized from a confluent plate and counted in a Neubauer chamber. Three cells per well were calculated and seeded in a 96-well plate. After 5 d wells were checked for single clones. After another 7-10 days cells were checked for fluorescence and positive clones were transferred to a 24-well plate.

References

Agudelo, D., Duringer, A., Bozoyan, L., Huard, C.C., Carter, S., Loehr, J., Synodinou, D., Drouin, M., Salsman, J., Dellaire, G., et al. (2017). Marker-free coselection for CRISPR-driven genome editing in human cells. Nat Methods *14*, 615–620.

Arimbasseri, A.G., Rijal, K., and Maraia, R.J. (2013). Transcription termination by the eukaryotic RNA polymerase III. Biochim. Biophys. Acta *1829*, 318–330.

Attali, D. (2017). Shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds. R package version 1.0.

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C. (1993). A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. Nucleic Acids Res. *21*, 3329–3330.

Chang, W., Cheng, J., Allaire, J.J., and Xie, Y. (2018). shiny: Web Application Framework for R Package Version 1.1.0; https://CRAN.

Gao, L., Cox, D.B.T., Yan, W.X., Manteiga, J.C., Schneider, M.W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N., and Zhang, F. (2017). Engineered Cpf1 variants with altered PAM specificities. Nat. Biotechnol. *35*, 789–792.

Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141–147.

Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012). Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Mol. Cell Proteomics *11*, M111.014050.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

Greene, M.R., and Sambrook, J. (2012). Molecular Cloning: a laboratory manual (Cold Spring Harbor).

Gu, B., Posfai, E., and Rossant, J. (2018). Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. Nat. Biotechnol. *36*, 632–637.

Gutierrez-Triana, J.A., Tavhelidse, T., Thumberger, T., Thomas, I., Wittbrodt, B., Kellner, T., Anlas, K., Tsingos, E., and Wittbrodt, J. (2018). Efficient single-copy HDR by 5' modified long dsDNA donors. eLife *7*.

He, X., Tan, C., Wang, F., Wang, Y., Zhou, R., Cui, D., You, W., Zhao, H., Ren, J., and Feng, B. (2016). Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. Nucleic Acids Res. *44*, e85–e85.

Huh, W.-K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Khmelinskii, A., Blaszczak, E., Pantazopoulou, M., Fischer, B., Omnus, D.J., Le Dez, G., Brossard, A., Gunnarsson, A., Barry, J.D., Meurer, M., et al. (2014). Protein quality control at the inner nuclear membrane. Nature *516*, 410–413.

Koch, B., Nijmeijer, B., Kueblbeck, M., Cai, Y., Walther, N., and Ellenberg, J. (2018). Generation and validation of homozygous fluorescent knock-in cells using CRISPR-Cas9 genome editing. Nat

Protoc *13*, 1465–1487.

Lackner, D.H., Carré, A., Guzzardo, P.M., Banning, C., Mangena, R., Henley, T., Oberndorfer, S., Gapp, B.V., Nijman, S.M.B., Brummelkamp, T.R., et al. (2015). A generic strategy for CRISPR-Cas9-mediated gene tagging. Nat Commun *6*, 10237.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359.

Lin, Y.-C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., et al. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. Nat Commun *5*, 4767.

Merkle, F.T., Neuhausser, W.M., Santos, D., Valen, E., Gagnon, J.A., Maas, K., Sandoe, J., Schier, A.F., and Eggan, K. (2015). Efficient CRISPR-Cas9-mediated generation of knockin human pluripotent stem cells lacking undesired mutations at the targeted locus. Cell Rep *11*, 875–883.

Meurer, M., Duan, Y., Sass, E., Kats, I., Herbst, K., Buchmuller, B.C., Dederer, V., Huber, F., Kirrmaier, D., Stefl, M., et al. (2018). Genome-wide C-SWAT library for high-throughput yeast genome tagging. Nat Methods *425*, 1.

Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2017). Biostrings: Efficient manipulation of biological strings.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Rothstein, R. (1991). Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Meth. Enzymol. *194*, 281–301.

Roy, K.R., Smith, J.D., Vonesch, S.C., Lin, G., Tu, C.S., Lederer, A.R., Chu, A., Suresh, S., Nguyen, M., Horecka, J., et al. (2018). Multiplexed precision genome editing with trackable genomic barcodes in yeast. Nat. Biotechnol. *36*, 512–520.

Rumi, M., Ishihara, S., Aziz, M., Kazumori, H., Ishimura, N., Yuki, T., Kadota, C., Kadowaki, Y., and Kinoshita, Y. (2006). RNA polymerase II mediated transcription from the polymerase III promoters in short hairpin RNA expression vector. Biochem. Biophys. Res. Commun. *339*, 540–547.

Schaab, C., Geiger, T., Stoehr, G., Cox, J., and Mann, M. (2012). Analysis of high accuracy, quantitative proteomics data in the MaxQB database. Mol. Cell Proteomics *11*, M111.014068.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676–682.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Methods *9*, 671–675.

Shaner, N.C., Lambert, G.G., Chammas, A., Ni, Y., Cranfill, P.J., Baird, M.A., Sell, B.R., Allen, J.R., Day, R.N., Israelsson, M., et al. (2013). A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nat Methods *10*, 407–409.

Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. *27*, 491–499.

Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z., et al. (2016). In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature *540*, 144–149.

Wach, A., Brachat, A., Pöhlmann, R., and Philippsen, P. (1994). New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast *10*, 1793–1808.

Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M., and Vander Horn, P.B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res. *32*, 1197–1207.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science *285*, 901–906.

Yamamoto, Y., and Gerbi, S.A. (2018). Making ends meet: targeted integration of DNA fragments by genome editing. Chromosoma *16*, 87–16.

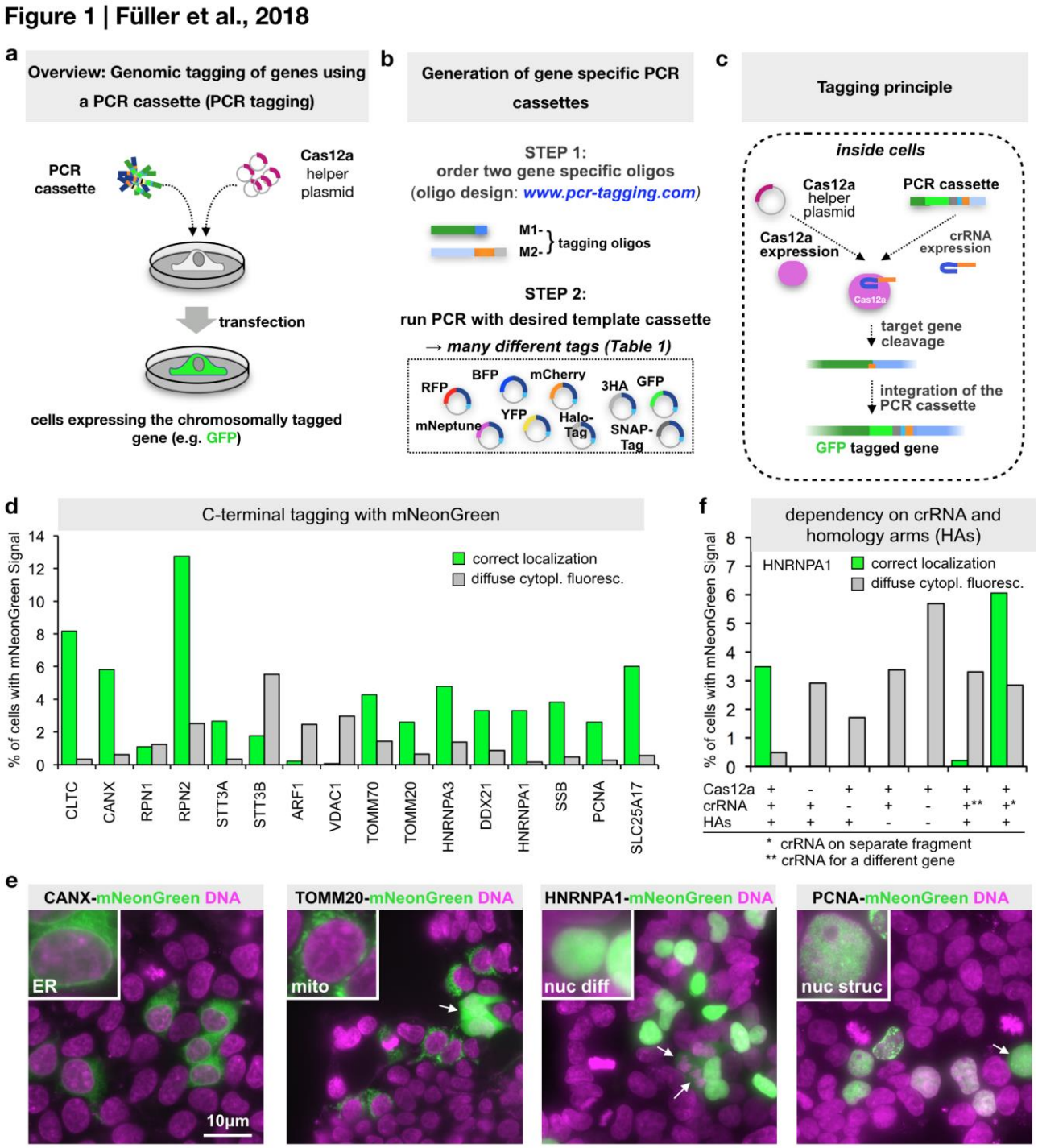
Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell *163*, 759–771.

Zhang, J.-P., Li, X.-L., Li, G.-H., Chen, W., Arakaki, C., Botimer, G.D., Baylink, D., Zhang, L., Wen, W., Fu, Y.-W., et al. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. Genome Biol *18*, 35.

Zhu, Z., Verma, N., González, F., Shi, Z.-D., and Huangfu, D. (2015). A CRISPR/Cas-Mediated Selection-free Knockin Strategy in Human Embryonic Stem Cells. Stem Cell Reports *4*, 1103–1111.

Figures and Legends

Figure 1: Endogenous C-terminal tagging in mammalian cells using PCR tagging.



(**a**) Overview: for gene tagging a PCR cassette is transfected into the target cell together with a helper plasmid containing a Cas12a endonuclease. This leads to insertion of the PCR cassette into the chromosome, which yields a fusion of the tag (e.g. GFP) with the target gene.

(**b**) The gene specific PCR cassette is generated using two gene specific tagging oligos (termed M1 and M2) and PCR with a generic template plasmid. Assisted oligo design is available from www.pcr-tagging.com, based on the principles described in Materials and Methods. The template plasmid provides the tag (e.g. a fluorescent protein) and a possible selection marker. Template plasmids with different tags can be used.

(**c**) Tagging Principle: The PCR cassette contains a crRNA sequence (orange) that is expressed inside the cell. The crRNA directs Cas12a (which is expressed from the helper plasmid) to the target locus close to the insertion site. Stimulated by the double strand break the linear PCR cassette is then inserted into the genome. The homology arm of the M1-oligo thereby directs in frame fusion of the tag with the target ORF, leading to the expression of a tagged protein from the target locus. Integration leads to destruction of the crRNA target site, thus preventing re-cleavage of the locus.

(**d**) Efficiency of C-terminal mNeonGreen-tagging for 15 organelle specific genes. For each gene, specific M1/M2 tagging oligos were used to amplify an mNeonGreen containing PCR cassette. The resulting PCR cassettes were transfected in HEK293T cells. HOECHST staining of live cells and analysis by fluorescence microscopy was performed three days after transfection. Fractions of cells exhibiting the expected localization or diffuse cytoplasmic green fluorescence are shown. For information on selected genes, see Supplementary Table S1. Data from one representative experiment is shown.

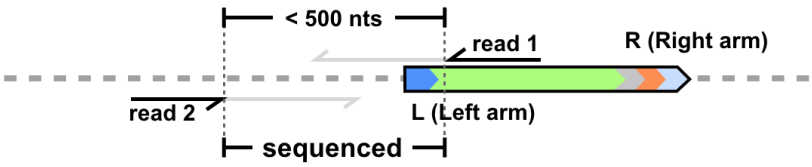
(**e**) Representative images from HEK293T cells 3 days after transfection. mNeonGreen fluorescence and HOECHST staining (DNA) is shown. In addition to the expected localization, cells showing non-specific cytoplasmic fluorescence (arrows) are detected.

(**f**) Tagging is specific for the crRNA and guided by the homology arms (HAs). Efficiency of control transfections (see Supplementary Fig. S2 for representative examples). * in this transfection indicates that a matching combination of crRNA and HAs was used, but the crRNA was expressed from a different PCR fragment. ** indicates that in this case a PCR cassette was used where the crRNA (for CANX) did cleave a different gene than the one specified by the HAs (HNRNPA1). A small fraction of cells (<0.02%, corresponding to 5 cells in the entire well) exhibiting an ER localization pattern typically seen for CANX was observed, indicating cassette integration at the CANX locus, e.g. via NHEJ.

Figure 2. Analysis of the fate of the transfected PCR cassette using target enrichment sequencing

Figure 2 | Füller et al., 2018

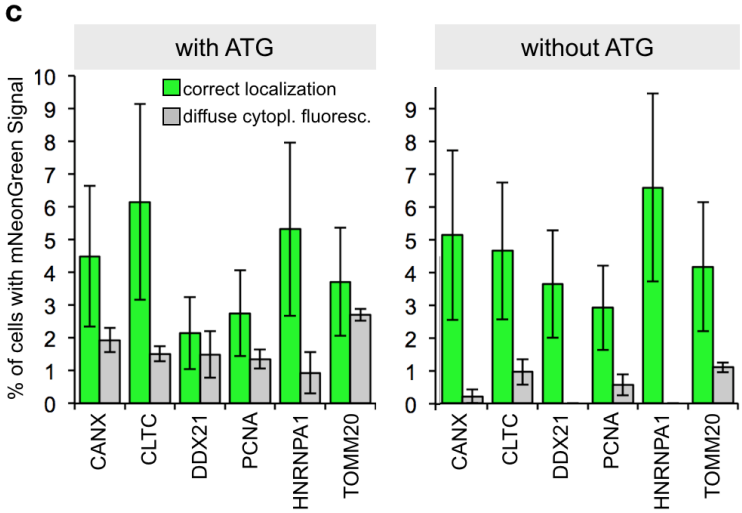
a



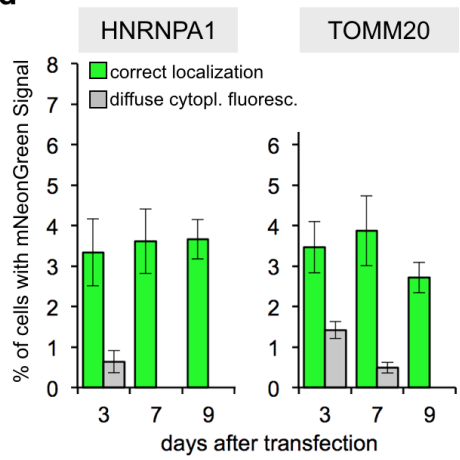
b

observed read pair types	RPN2	STT3A	TOMM70	TOMM20	Mixture -Cas12a	Mixture +Cas12a
correct integration	0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
PCR cassette	85.0	81.3	83.7	89.3	89.4	90.2
template plasmid	0.5	0.3	0.1	< 0.1	< 0.1	< 0.1
LL homo fusion	2.4	3.0	2.1	1.2	1.3	0.8
LL hetero fusion	-	-	-	-	1.4	1.0
LR homo fusion	12.1	15.4	14.1	9.5	6.5	6.8
LR hetero fusion	-	-	-	-	1.4	1.2

c



d



(**a**) Anchor-Seq (Meurer et al., 2018) is based on a target enrichment procedure that uses an oligo in the mNeonGreen gene to enrich adjacent sequences for analysis by next generation sequencing using a paired end sequencing protocol (reads 1 & 2).

(**b**) Anchor-Seq analysis of adjacent sequences of the PCR cassette from transfected HEK293T cells three days after transfection, for the 4 genes shown individually, and from cells transfected with a mixture of PCR cassettes for all the genes shown in Fig. 1d (labeled with 'Mixture'). Fraction of reads (in %) observed for the different categories, where R and L stand for Right and Left arm of the PCR cassette, respectively. Combinations of the letter denote the detected fusion, homo denotes fusion of two ends from the same PCR cassette type, hetero from PCR cassettes targeting different genes.

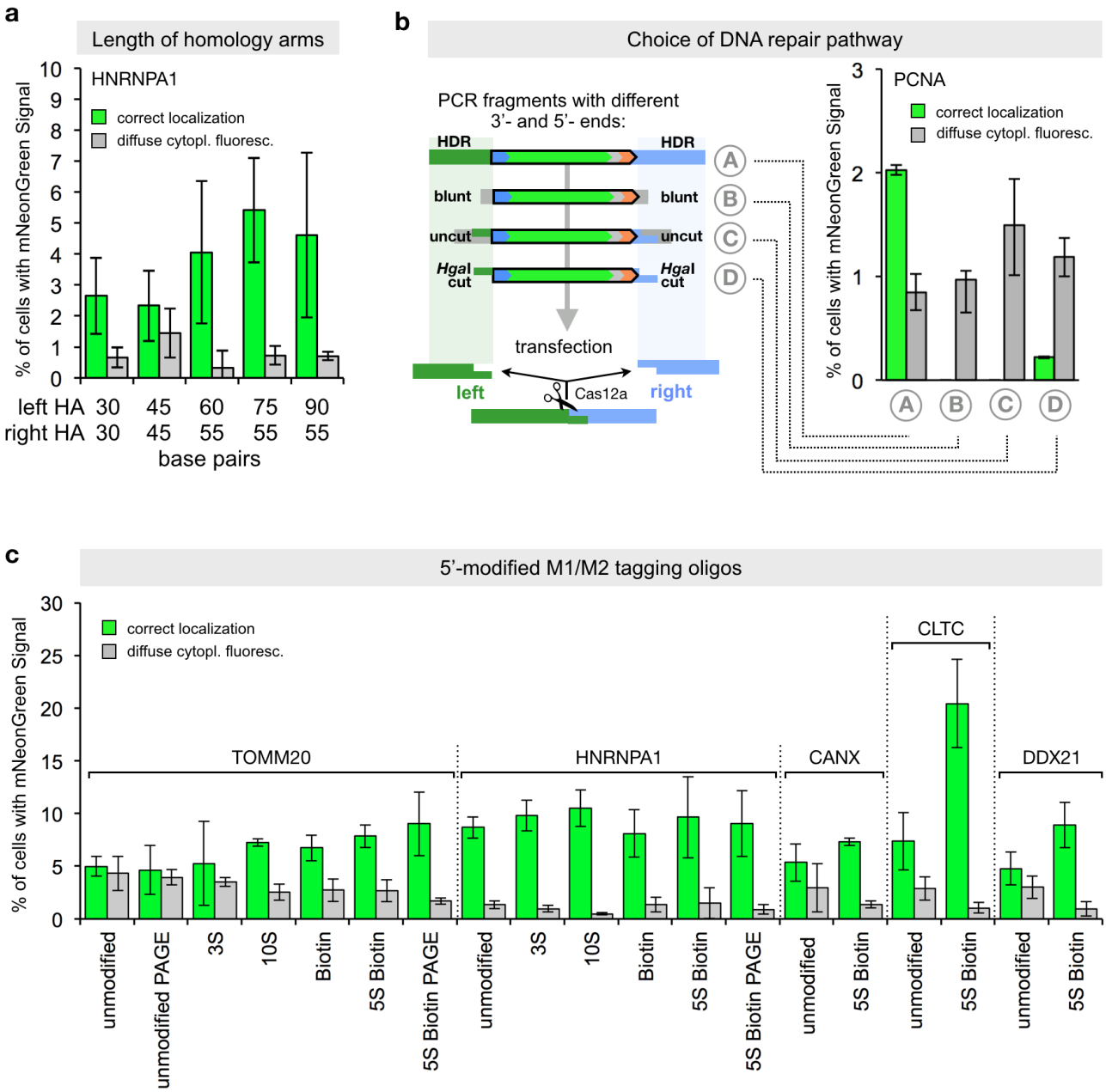
(**c**) HEK293T cells transfected with PCR cassettes as indicated using wild type mNeonGreen gene or lacking ATG translation initiation codons within the first 30 codons of the mNeonGreen ORF. Live cell fluorescence microscopy of HOECHST stained cells was used to determine the fraction of cells

(in %) with correct localization and diffuse cytoplasmic fluorescence. Data from three replicates is shown. Error bars indicate SD.

(**d**) HEK293T cells transfected with PCR cassettes for HNRNPA1 or TOMM20 were passaged for the indicated time periods. Analysis as in (c). Data from three replicates is shown. Error bars indicate SD.

Figure 3. Tagging efficiency as a function of different parameters

Figure 3 | Füller et al., 2018



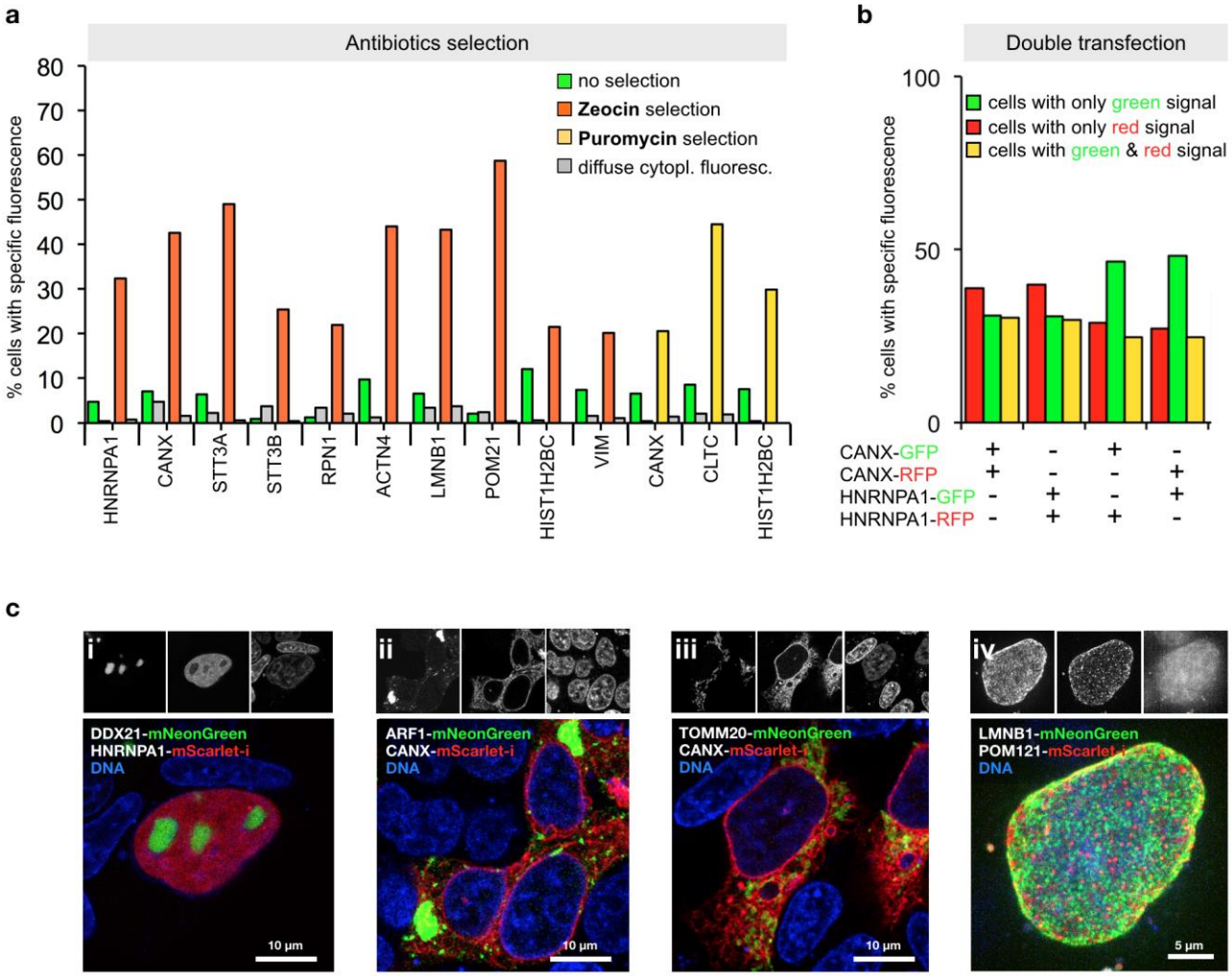
(**a**) Length of homology arms. M1 and M2 tagging oligos containing the indicated sequence lengths of homology (left HA and right HA, respectively) to the destination locus were used for PCR tagging of the HNRNPA1 locus using HEK293T cells. Tagging efficiency was estimated 3 days after transfection as described before. Data from three replicates is shown. Error bars indicate SD.

(**b**) PCR cassettes containing various types of ends to direct the choice of DNA repair pathway: Homology arms (90 bp and 55 bp homology, for HDR; *A*), blunt ended arms without homology to the target locus (blunt; *B*), *Hga*I cut (*D*) and uncut ends (*C*). Cutting with the type 2 restriction enzyme *Hga*I results in 5 nt 3'-overhangs that are complementary to the overhangs generated by the crRNA directed Cas12a-cleavage of the destination locus. Tagging efficiency was estimated 3 days later as described in (a) using HEK293T cells. Data from three replicates is shown. Error bars indicate SD.

(**c**) Use of modified oligonucleotides. M1/M2 oligonucleotides with the indicated number of phosphorothioate bonds and/or Biotin as indicated were used for generation of PCR cassettes. All oligos were 'cartridge' purified except for the ones denoted with 'PAGE', which were size selected using polyacrylamid gel electrophoresis. Tagging efficiency was estimated 3 days after transfection as described before using HEK293T cells. Data from three replicates is shown. Error bars indicate SD.

Figure 4. Antibiotic selection and simultaneous tagging of two loci

Figure 4 | Füller et al., 2018

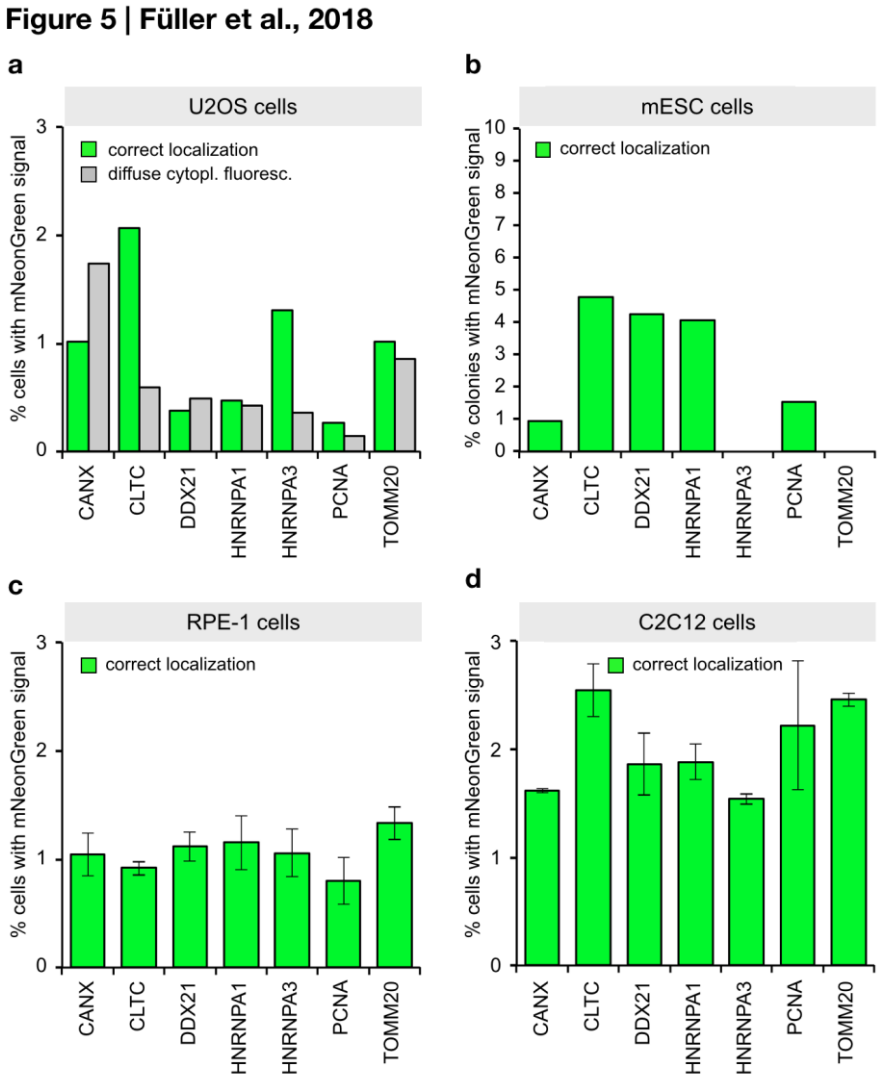


(**a**) Enrichment of HEK293T cells expressing correctly localized fusion proteins using Zeocin or Puromycin selection as indicated. Antibiotics selection was started 3 days after transfection. Fractions of cells exhibiting localized or diffuse cytoplasmic fluorescence are shown. Data from one representative experiment is shown.

(**b**) Double transfection of cells using PCR cassette reporters for the indicated genes and with the indicated fluorescent protein. For counting, only cells exhibiting correctly localized fluorescence signals were counted (ER localization for CANX tagging, nuclear localization for HNRNPA1 tagging, see Supplementary Fig. S5). Data from one representative experiment is shown.

(**c**) Double tagging of the genes indicated in the images. Representative cells are shown. (i to iii) single plane images, (iv) a maximum projection of multiple planes spanning the upper half of a cell nucleus is shown.

Figure 5. PCR tagging in different cell lines



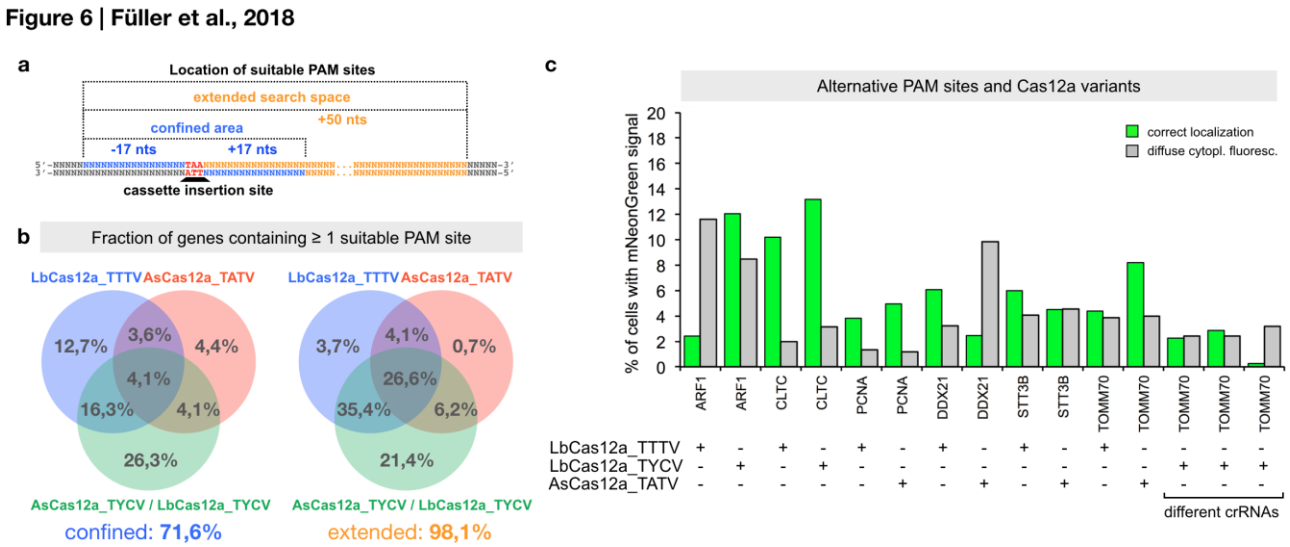
(**a**) Transfection of U2OS cells using Lipofectamine 2000. After three days the cells were analyzed using HOECHST staining and live cell imaging. Data from one representative experiment is shown.

(**b**) Electroporation of mESC cells with PCR cassettes for tagging the indicated genes. After three days the cells were fixed using paraformaldehyde and analyzed. We counted microcolonies that have at least 1 positive cell. Please note: For these cells we did not quantify cells with diffuse cytoplasmic fluorescence, since paraformaldehyde fixation prior to imaging leads to an increase in non-specific fluorescence. This prevented the detection of the weak diffuse mNeonGreen fluorescence. Data from one representative experiment is shown.

(**c**) Electroporation of RPE-1 cells. Cells were analyzed 2 days later. Experimental setup similar to (b). Data from three replicates is shown. Error bars indicate SD.

(**d**) Electroporation of C2C12 cells. Cells were analyzed 2 days later. Experimental setup similar to (b). Data from two replicates is shown. Error bars indicate SD.

Figure 6. PCR tagging enables C-terminal tagging of the majority of human genes.



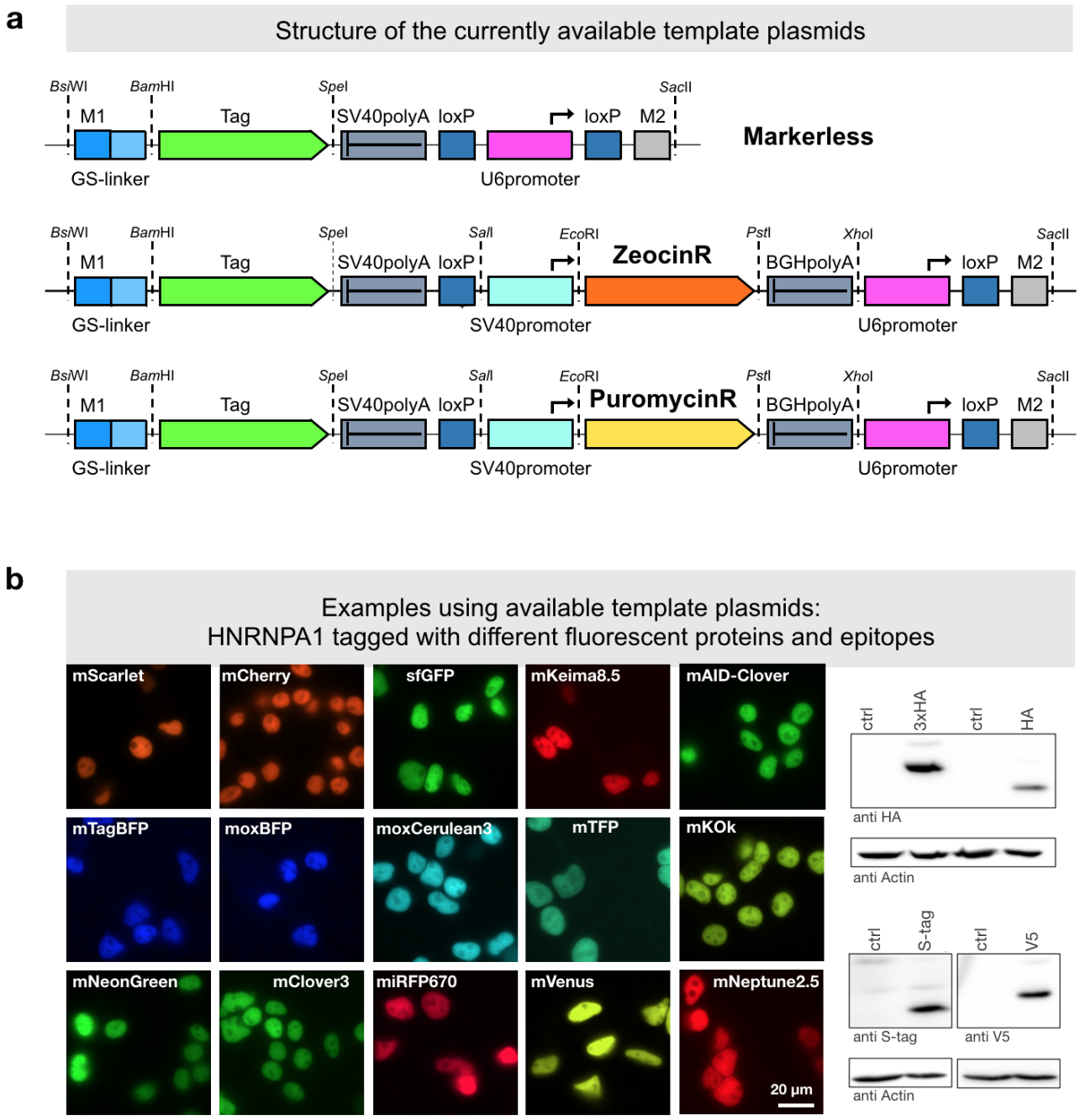
(**a**) Search space for Cas12a-PAM sites suitable for C-terminal protein tagging. PCR cassette insertion into the genome using PAM sites located in the confined search space (blue) lead to a disruption of the crRNA target sequence. This would not be the case for PAM sites in the extended search space (yellow). To prevent re-cleavage after insertion the homology arm of the PCR fragment (provided by the M2-oligo) is designed such that a small deletion in the region after the STOP codon does lead to the disruption of the crRNA target site.

(**b**) Fraction (in %) of human genes with suitable PAM sites near the STOP codon, as a function of the confined and extended search spaces (a) and different Cas12a variants, as indicated.

(**c**) Tagging of the indicated genes in HEK293T cells. Helper plasmids with different Cas12a genes, as indicated. PCR cassettes contained crRNA genes with matching PAM site specificity. For *TOMM70* three different Cas12a variants were tested, using three different crRNA sequences for AsCas12a, as indicated. Tagging efficiency was determined three days after transfection. Data from one representative experiment is shown.

Figure 7. PCR tagging Toolkit for mammalian cells.

Figure 7 | Füller et al., 2018



(**a**) Schematic outline of the template plasmids provided.

(**b**) Examples of HNRNPA1 tagging using different available cassettes. Complete list of features and sequence files are provided in Table 1 and Supplementary Table S2. Western blot analysis was performed 3 days after transfection with crude lysate of a cell pool. Fluorescence microscopy was performed using cells 3 days after transfection.

Table 1. Overview about available template plasmids for PCR tagging

Tag	Type	Oligos	Markerless	Zeocin	Puromycin	Application
moxBFP	fluorescent protein	M1/M2	pMaCTag-01	pMaCTag-Z01	pMaCTag-P01	Imaging
TagBFP	fluorescent protein	M1/M2	pMaCTag-02	pMaCTag-Z02	pMaCTag-P02	Imaging
moxCerulean3	fluorescent protein	M1/M2	pMaCTag-03	pMaCTag-Z03	pMaCTag-P03	Imaging
mTFP1	fluorescent protein	M1/M2	pMaCTag-04	pMaCTag-Z04	pMaCTag-P04	Imaging
eGFP	fluorescent protein	M1/M2	pMaCTag-05	pMaCTag-Z05	pMaCTag-P05	Imaging
sfGFP	fluorescent protein	M1/M2	pMaCTag-06	pMaCTag-Z06	pMaCTag-P06	Imaging
mNeonGreen	fluorescent protein	M1/M2	pMaCTag-07	pMaCTag-Z07	pMaCTag-P07	Imaging
mClover3	fluorescent protein	M1/M2	pMaCTag-08	pMaCTag-Z08	pMaCTag-P08	Imaging
mVenus	fluorescent protein	M1/M2	pMaCTag-09	pMaCTag-Z09	pMaCTag-P09	Imaging
mKOk	fluorescent protein	M1/M2	pMaCTag-10	pMaCTag-Z10	pMaCTag-P10	Imaging
mScarlet-i	fluorescent protein	M1/M2	pMaCTag-11	pMaCTag-Z11	pMaCTag-P11	Imaging
mScarlet	fluorescent protein	M1/M2	pMaCTag-12	pMaCTag-Z12	pMaCTag-P12	Imaging
mCherry	fluorescent protein	M1/M2	pMaCTag-13	pMaCTag-Z13	pMaCTag-P13	Imaging
mNeptune2.5	fluorescent protein	M1/M2	pMaCTag-14	pMaCTag-Z14	pMaCTag-P14	Imaging
miRFP670	fluorescent protein	M1/M2	pMaCTag-15	pMaCTag-Z15	pMaCTag-P15	Imaging
Dronpa3	fluorescent protein	M1/M2	pMaCTag-16	pMaCTag-Z16	pMaCTag-P16	Imaging
mEos4b	fluorescent protein	M1/M2	pMaCTag-17	pMaCTag-Z17	pMaCTag-P17	Imaging
mKeima8.5	fluorescent protein	M1/M2	pMaCTag-18	pMaCTag-Z18	pMaCTag-P18	Imaging
mCherry-mNeonGreen	fluorescent protein	M1/M2	pMaCTag-19	pMaCTag-Z19	pMaCTag-P19	Imaging
10xSuntag	detection tag	M1/M2	pMaCTag-20	pMaCTag-Z20	pMaCTag-P20	Imaging
24xSunTag	detection tag	M1/M2	pMaCTag-21	pMaCTag-Z21	pMaCTag-P21	Imaging
SNAP	chem. modification	M1/M2	pMaCTag-22	pMaCTag-Z22	pMaCTag-P22	Imaging
Halo	chem. modification	M1/M2	pMaCTag-23	pMaCTag-Z23	pMaCTag-P23	Imaging
TAP	Affinitiy tag	M1/M2	pMaCTag-24	pMaCTag-Z24	pMaCTag-P24	PPIs
YFP-LID	degron	M1/M2	pMaCTag-25	pMaCTag-Z25	pMaCTag-P25	function
mAID-Clover	degron	M1/M2	pMaCTag-26	pMaCTag-Z26	pMaCTag-P26	function
1xHA	antibody epitope	M1/M2	pMaCTag-27	pMaCTag-Z27	pMaCTag-P27	detection
3xHA	antibody epitope	M1/M2	pMaCTag-28	pMaCTag-Z28	pMaCTag-P28	detection
Strep-II	Affinity-tag	M1/M2	pMaCTag-29	pMaCTag-Z29	pMaCTag-P29	detection
VSV-G	antibody epitope	M1/M2	pMaCTag-30	pMaCTag-Z30	pMaCTag-P30	detection
V5	antibody epitope	M1/M2	pMaCTag-31	pMaCTag-Z31	pMaCTag-P31	detection
S-Tag	bimolecular compl.	M1/M2	pMaCTag-32	pMaCTag-Z32	pMaCTag-P32	detection
FLAG	antibody epitope	M1/M2	pMaCTag-33	pMaCTag-Z33	pMaCTag-P33	detection
3xFLAG	antibody epitope	M1/M2	pMaCTag-34	pMaCTag-Z34	pMaCTag-P34	detection

Supplementary Information

Content

Terminology

Supplementary Figures and Legends

Terminology

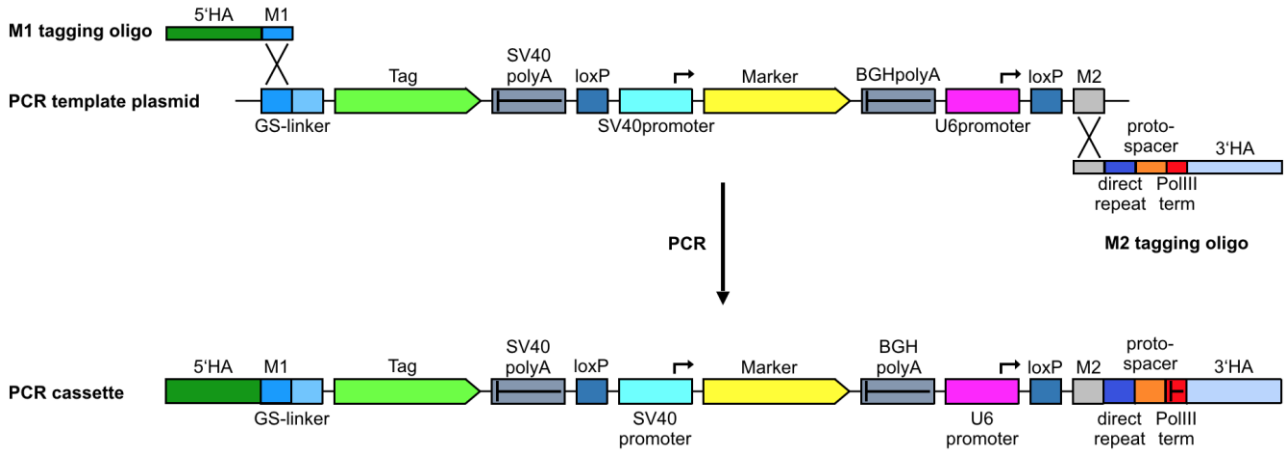
Throughout the manuscript we use the following terms in a consistent manner in order to denote the different components and processes:

Method:	Mammalian PCR tagging	
Oligos:	M1 und M2 tagging oligos (forward and reverse oligo), tagging oligos	
Templates:	(C-terminal) template plasmid with ... (e.g. GFP): Name of Cassettes:	
	pMaCTag-01/P01/Z01 (P..Puromycin/Z..Zeocin)	
Tagging PCR:	cassette PCR	
PCR product:	PCR cassette	
Tagged gene:	*before*:	target gene
	after:	tagged gene
Webpage:	domain name: www.pcr-tagging.com	

Supplementary Figures and Tables

Supplementary Fig. S1: PCR Strategy

Supplementary Figure S1 | Füller et al., 2018

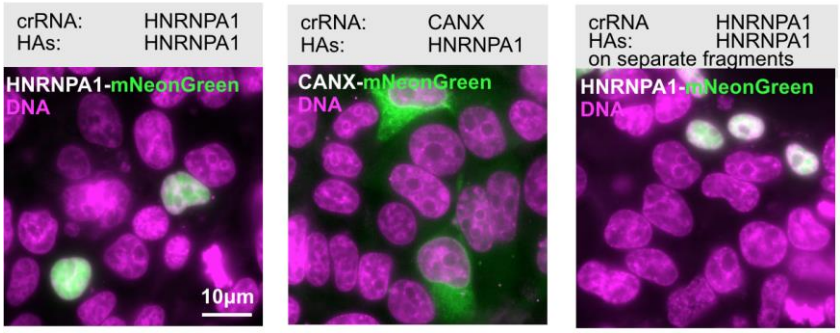


PCR using M1 and M2 tagging oligos and a template cassette with Tag. The M1 and M2 tagging oligos provide the homology arms (~55 to 90 nts in length) for targeted integration. The M2 tagging oligo additionally provides a protospacer sequence (orange) for a Cas12a endonuclease. The template cassette contains the desired tag and additional features, such as a selection marker. It also contains the U6 Pol III promoter for the crRNA. PCR yields a linear DNA fragment (PCR

cassette) that contains homology arms to the target locus and a functional crRNA gene to cleave the locus.

Supplementary Fig. S2: Control transfections

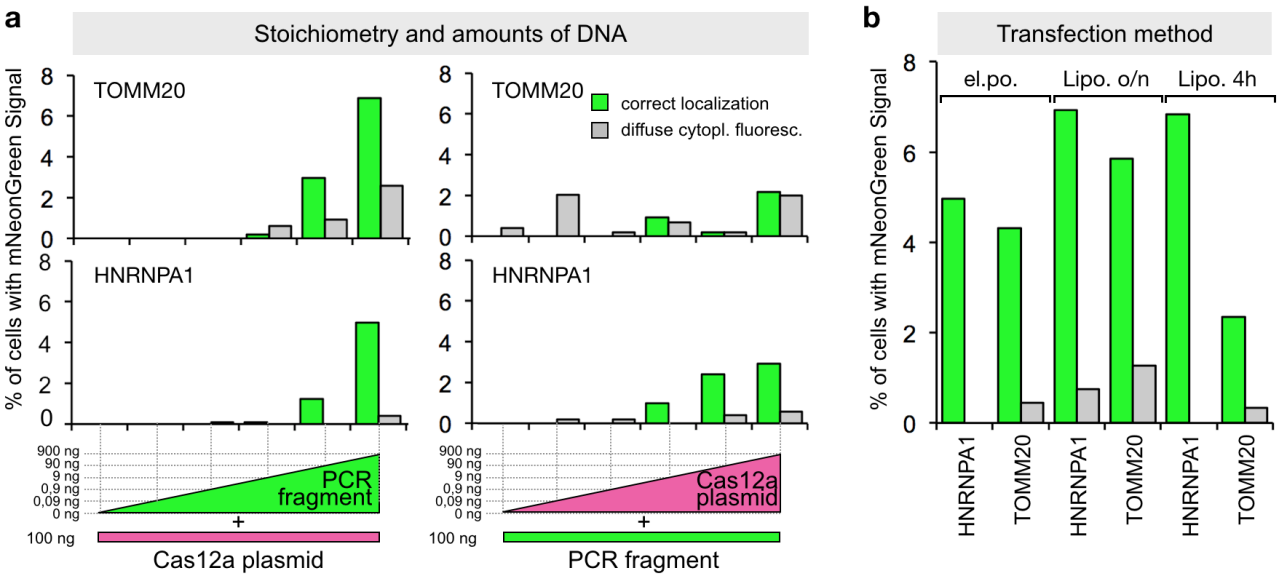
Supplementary Figure S2 | Füller et al., 2018



Control transfections to demonstrate the effect of the crRNA, the Cas12a plasmid and the presence of homology arms (HAs). Locus specificity of the HAs and the crRNA as indicated.

Supplementary Figure S3: Effect of transfection parameters on tagging efficiency

Supplementary Figure S3 | Füller et al., 2018

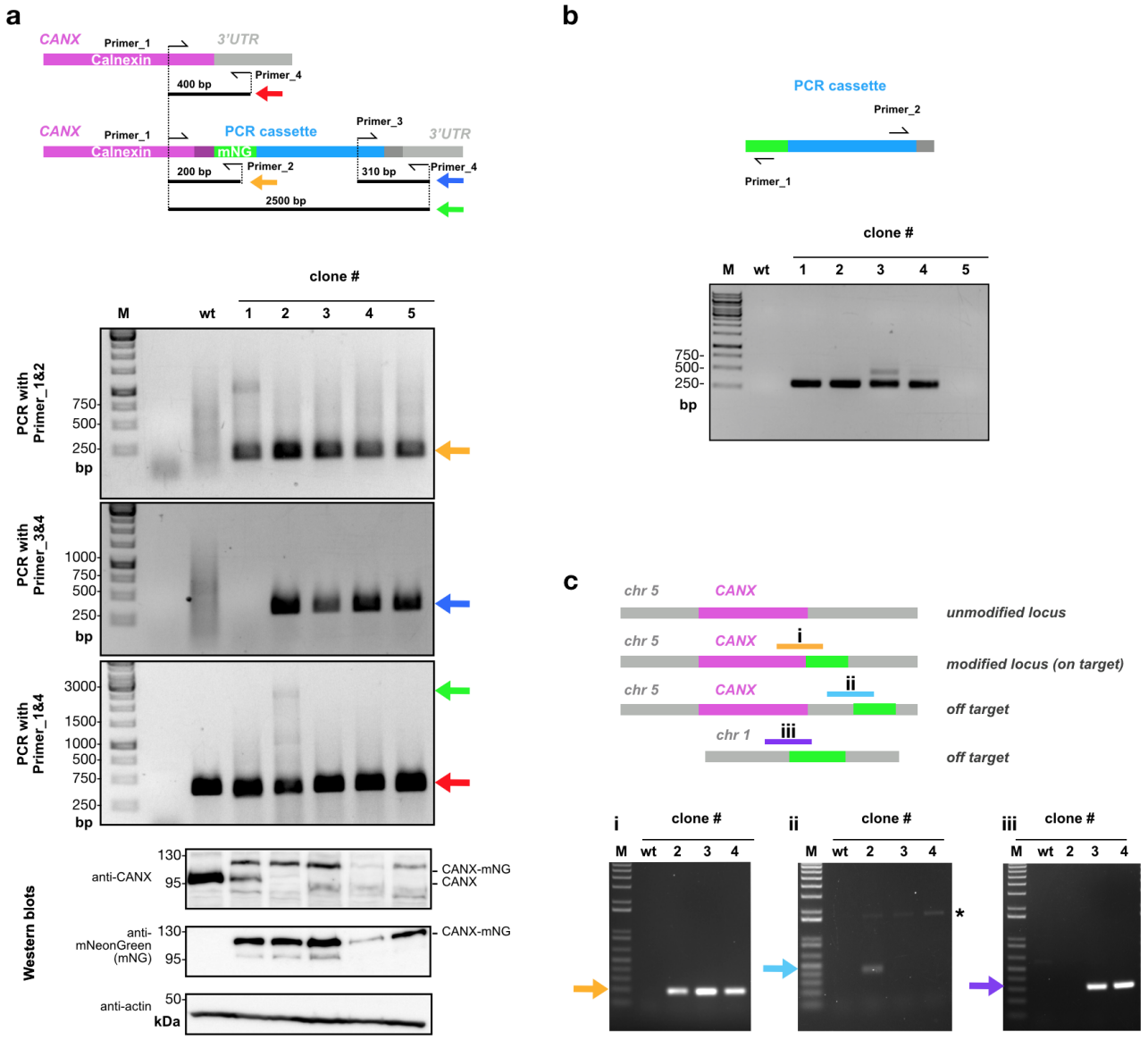


(**a**) Impact of transfected amounts of DNA on tagging efficiency using HEK293T cells. Transfected amounts of PCR cassette and Cas12a plasmid as indicated. Always 1 µg of DNA was transfected using lipofectamine. pUC18 was used as a neutral DNA. Tagging efficiency was determined 3 days later using HOECHST staining and live cell imaging. Data from one representative experiment is shown.

(**b**) HEK293T cells were transfected for 4 hours or overnight using Lipofectamine 2000 or transfected using electroporation, as indicated. Tagging efficiency was determined three days later as described in (a). Data from one representative experiment is shown.

Supplementary Figure S4. Analysis of clones from a CANX-mNeonGreen tagging experiment.

Supplementary Figure S4 | Füller et al., 2018

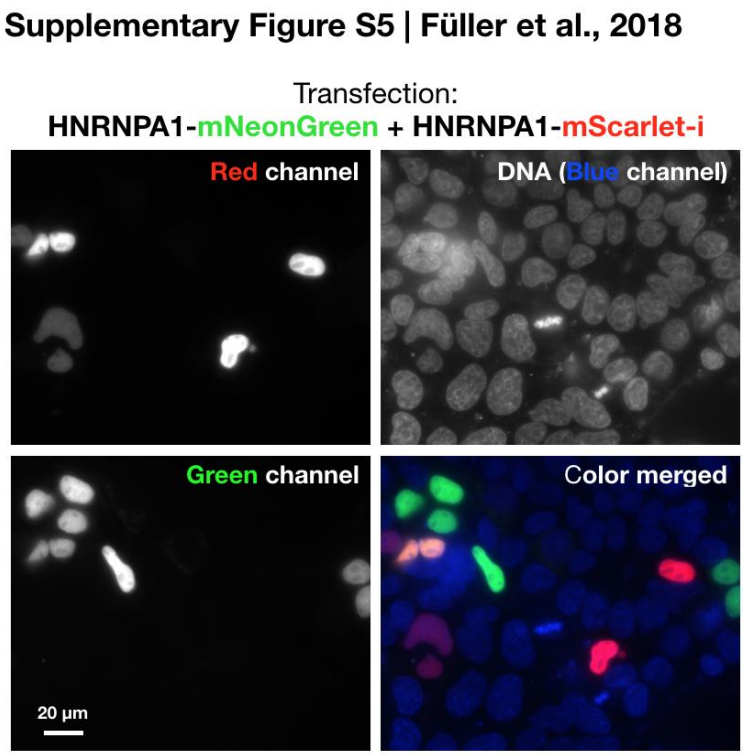


(**a**) PCR analysis of single clones using primer for PCR of characteristic fragments indicative for correctly inserted fragments. Primer that anneal to chromosomal DNA were chosen to reside outside of the sequences that are contained in the homology arms for recombination. For western blot analysis antibodies specific to mNeonGreen or to Calnexin were used.

(**b**) PCR analysis for the detection of concatenated PCR cassettes using primers 2 and 3 from (a).

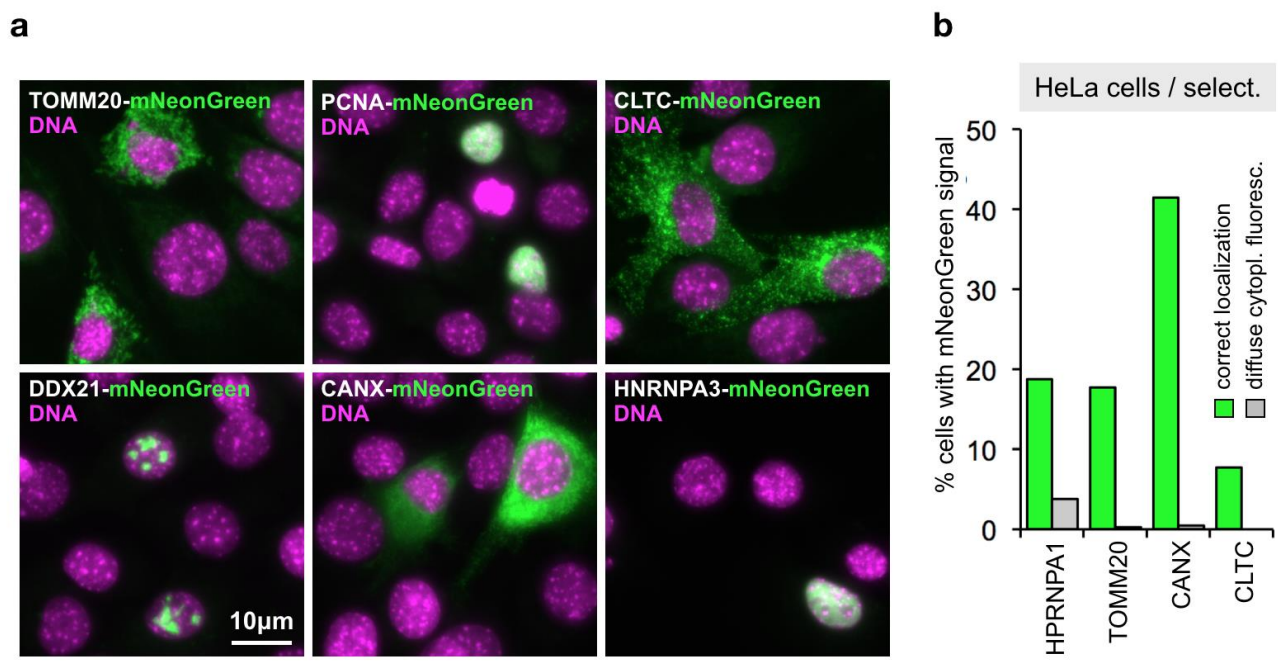
(**c**) PCR validation of the insertion of off-target inserted cassettes. For PCR, dedicated primers in the off-target genomic region were used in combination with a primer in the mNeonGreen sequence in order to detect the insertion junction. Colored arrows indicate specifc product of expected size and asterisk unspecifc product. Off-target products were Sanger sequenced for additional sequence confirmation. As a size marker (M) 1 Kb Plus DNA ladder (Invitrogen) was used.

Supplementary Figure S5. Multi-color integration



Double tagging using a mixture of HNRNPA1-mScarlet-i and HNRNPA1-mNeonGreen PCR cassettes. For analysis, dual color fluorescence images were acquired.

Supplementary Figure S6. Multi-color integration



(**a**) Sample images from C2C12 cells (Fig. 5d), 5 days after transfection.

(**b**) HeLa cells transfected using Lipofectamine 2000. Cells were grown for 3 days without, and 10 days in the presence of Zeocin using HOECHST staining and live cell imaging. Data from one representative experiment is shown.