CRISPR/Cas9-mediated generic protein tagging in mammalian cells

Fabian M.B. Thöne[1,§], Nina S. Kurrle[1,2 §], Harald von Melchner[1,*], and Frank Schnütgen[1,2]

[1] Department of Molecular Hematology and LOEWE Center for Cell and Gene Therapy, University Hospital Frankfurt, Goethe University, 60590 Frankfurt am Main, Germany

[2] Department of Medicine 2, Hematology/Oncology and Frankfurt Cancer Institute, University Hospital Frankfurt, Goethe University, 60590 Frankfurt am Main, Germany

[§]equal contribution

*Correspondence to melchner@em.uni-frankfurt.de

This is an Accepted Manuscript version of the following article, accepted for publication in Methods: Thöne, F.M.B.; Kurrle, N.S.; von Melchner, H.; Schnütgen, H. (2019): CRISPR/Cas9-mediated generic protein tagging in mammalian cells. Methods, 164-165, p. 59-66, https://doi.org/10.1016/j.ymeth.2019.02.018

Abstract

Systematic protein localization and protein-protein interaction studies to characterize specific protein functions are most effectively performed using tag-based assays. Ideally, protein tags are introduced into a gene of interest by homologous recombination to ensure expression from endogenous control elements. However, inefficient homologous recombination makes this approach difficult in mammalian cells. Although gene targeting efficiency by homologous recombination increased dramatically with the development of designer endonuclease systems such as CRISPR/Cas9 capable of inducing DNA double-strand breaks with unprecedented accuracy, the strategies still require synthesis or cloning of homology templates for every single gene. Recently, Lackner et al reported a protein tagging method that relies on a homology-independent mechanism. By combining the CRISPR/Cas9 system with a generic tag-donor plasmid, proteins were successfully tagged in mammalian cells. We improved this technology by developing a tool kit comprising a CRISPR/Cas9 expression vector with several EGFP encoding plasmids that should enable tagging of almost every protein expressed in mammalian cells. By performing protein-protein interaction and subcellular localization studies of mTORC1 signal transduction pathway-related proteins expressed in HEK293T cells, we show that tagged proteins faithfully reflect the behavior of their native counterparts under physiological conditions.

Keywords: CRISPR/Cas, protein tagging, EGFP, proteomics, mTORC1

Highlights:

- Development of a generic *in situ* protein tagging kit
- Efficient tagging of proteins within the mTORC1 pathway
- Tagged proteins maintain protein binding affinities of their native counterparts
- Tagged proteins faithfully reproduce the subcellular localization patterns of their native counterparts
- The *in situ* protein tagging kit is directly applicable to almost every protein expressed in mammalian cells

1. Introduction

Understanding the role of genes in disease development requires analysis of signal transduction pathways involving multiple proteins because in most cases altered pathways rather than single proteins are responsible for disease. This is most effectively achieved using tag-based proteomics enabling systematic protein localization and protein-protein interaction studies.

Protein tags are preferentially introduced into a gene of interest by homologous recombination to ensure expression from endogenous control elements. While classic gene targeting is still quite effective in mouse embryonic stem cells (ESCs), it has been mostly ineffective in somatic and/or non-dividing cells. However, with the development of designer endonucleases, of which the CRISPR/Cas9 system has received widest attention because of its unmatched simplicity and flexibility, gene targeting efficiencies improved significantly in all cell types including differentiated- and non-dividing cells.

Cas9, a bacterial endonuclease from *Streptococcus pyogenes,* in combination with a small guide RNA (gRNA), can trigger site-specific DNA double-strand breaks (DSBs) at any genomic locus proximal to a protospacer-adjacent motif (PAM). These DSBs are either repaired by non-homologous end joining (NHEJ), which is often imprecise causing small insertions and/or deletions (indels), or by homology-directed repair (HDR) in presence of a suitable homology template [1]. However, even with CRISPR/Cas9, protein tag insertion by homologous recombination is dependent on unique homology templates for every single gene, for which synthesis and/or cloning are time consuming and laborious. Moreover, foreign DNA insertions (knock-ins) relying on HDR are still relatively inefficient in certain types of cells such as human ESCs [2], induced pluripotent stem cells (iPSCs) [3], hematopoietic stem cells [4] and in particular in non-dividing cells because HDR is restricted to the late S and G2 phases of the cell cycle [5].

A growing number of reports show that foreign DNA can be integrated into the genome at Cas9 induced DSB sites by homology-independent pathways [6]. Most effective among these is a recently published knock-in strategy for generic protein tagging using a tag-donor plasmid designed to release the tag upon transfection [7]. Unlike previously employed knock-in cassettes linearized *ex vivo*, knock-ins of *in vivo* liberated cassettes are highly efficient [8].

Based on this strategy, we developed a generic tool kit suitable for *in situ* N-terminal tagging of proteins expressed in mammalian cells. It consists of a CRISPR/Cas9 expression vector and a set of donor plasmids designed for tagging almost any protein expressed in mammalian cells with an EGFP encoding N-terminal localization and purification (nLAP) tag [9]. Proteomic validation of this tool kit in HEK293T cells expressing nLAP tagged proteins of the mTORC1 pathway revealed that the tagged proteins faithfully replicated the protein-protein interactions and subcellular localizations

of their native counterparts. Here, we describe the individual components of the tool kit and give guidance for its use and adaptation to specific proteins.

2. Materials

2.1. Growth medium

- 450 ml DMEM (Invitrogen)
- 50 ml FCS (Invitrogen)
- 100 U/ml Penicillin
- 100 U/ml Streptomycin

2.2. CoIP buffer

- 10 mM Tris-HCl pH8
- 150 mM NaCl
- 5 mM EDTA pH8
- 0.5 % Triton X-100 (Sigma-Aldrich)
- 60 mM Octyl-β-D-glucopyranoside (Sigma-Aldrich)

2.3. Molecular biology reagents

- Restriction enzyme (BstBI) (New England Biolabs)
- Polyethylenimine (Sigma-Aldrich)
- T4 DNA ligase (New England Biolabs)
- Hygromycin B (Invitrogen)
- GFP-Trap^{R}_A beads (ChromoTeck)
- Jumpstart^{TM} REDTaq$^{®}$ DNA Polymerase (Sigma-Aldrich)
- Paraformaldehyde (PFA) (Sigma-Aldrich)

2.4. Antibodies

For detection in Western blot all antibodies were diluted 1:1000 except for the GAPDH antibody which was applied in a 1:10000 dilution.

The following primary antibodies were used:

- mTOR (Cell Signaling #2972, dilution for confocal microscopy 1:200)
- MIOS (Cell Signaling #13557)
- RAPTOR (Cell Signaling #2280, dilution for confocal microscopy 1:50)
- RagA (Cell Signaling #4357)
- GAPDH (Abcam #ab8245)
- WDR59 (Abcam #ab179895)

- Beta Actin (Sigma #A5441)
- NPRL3 (Sigma #HPA011741, dilution in confocal microscopy 1:100)
- Nucleolin (Santa Cruz #sc-13057)
- GFP (Roche #11814460001)
- NPRL2 (Biorbyt #orb178648)
- WDR24 (Proteintech #20778-1-AP, dilution in confocal microscopy 1:25)

The following secondary antibodies were used:

- Anti-mouse HRP-coupled (Santa Cruz #sc-2005)
- Anti-rabbit Cy5-coupled (Jackson ImmunoResearch #111-175-144, dilution in confocal microscopy 1:300)

2.5. Oligonucleotide sequences for gRNAs

All oligonucleotides were obtained from Sigma-Aldrich.

Sense gRNA targeting mTOR	5′-CACCGTTCCAAGCATCTTGCCCTG-3′
Anti-sense gRNA targeting mTOR	5′-AAACCAGGGCAAGATGCTTGGAAC-3′
Sense gRNA targeting RAPTOR	5′-CACCGATTTCGGACTCCATCAGTG-3′
Anti-sense gRNA targeting RAPTOR	5′-AAACCACTGATGGAGTCCGAAATC-3′
Sense gRNA targeting NPRL3	5′-CACCGGTTGTCCCGCATCCCGCCG-3′
Anti-sense gRNA targeting NPRL3	5′-AAACCGGCGGGATGCGGGACAACC-3′
Sense gRNA targeting WDR24	5′-CACCGCTGAACTGATGACCCACCC-3′
Anti-sense gRNA targeting WDR24	5′-AAACGGGTGGGTCATCAGTTCAGC-3′
Sense gRNA targeting tia1l	5′-CACCGGTATGTCGGGAACCTCTCC-3′
Anti-sense gRNA targeting tia1l	5′-AAACGGAGAGGTTCCCGACATACC-3′

2.6. PCR primer sequences

All primers were obtained from Sigma-Aldrich.

Generic tag forward primer	5′-ATAGGTCAGGCTCTCGCTGA-3′
Generic tag reverse primer	5′-CAGCTGCTGCTAAATTCGAG-3′
Endogenous forward primer mTOR	5′-GCCCACAATTTAACCTCCCTA-3′
Endogenous reverse primer mTOR	5′-GTGCAACTGCTGCAAAAGAA-3′
Endogenous forward primer RAPTOR	5′-GGTTTTCGATTTCCCGTTTT-3′
Endogenous reverse primer RAPTOR	5′-TTTCCCAAGGACACTTTTCG-3′
Endogenous forward primer NPRL3	5′-CACGGTCCAGGGAGTACCAC-3′
Endogenous reverse primer NPRL3	5′-TTAGCTTTTGCTGGGCATTC-3′

Endogenous forward primer WDR24	5´-GGCCTAAGGGAAGCTAGGAA-3´
Endogenous reverse primer WDR24	5´-CCTTCGCTTAACCTGAGCTG-3´

3. Description of methods

3.1. The generic protein tagging method

Originally published by Lackner et al [7], the generic tagging method comprises a donor plasmid, Cas9 and a target locus specific gRNA. The donor plasmid contains a tag of interest flanked by two Cas9 cleavage sites corresponding to the *tia1l* genomic locus in zebrafish which is absent in mammalian cells. Furthermore, it contains a U6 promoter driving the expression of a *tia1l* gRNA. When the donor plasmid is delivered together with Cas9 and a gRNA specific for the chosen endogenous target site to mammalian cells, the tag is liberated from the donor and incorporated into the target site.

We used a tag encoding a hygromycin resistance gene linked to an EGFP expressing N-terminal protein localization and affinity purification (nLAP) tag by a P2A polyprotein cleavage sequence (Fig. 1A) [9]. Unlike the originally published method requiring FACS sorting of tagged protein expressing cells, the hygro-nLAP donor enables direct selection in hygromycin.

3.2. Hygro-nLAP donor plasmids for protein tagging

We further modified the Lackner et al [7] strategy by generating two donor constructs instead of one for each reading frame. By flanking the hygro-nLAP tags with *tia1l* gRNA targets in either sense or antisense orientation we aimed to prevent post-insertional gRNA target reconstitutions in instances where endogenous- and *tia1l* targets reside on the same DNA strand (Fig. 1B).

3.3. Cloning of locus specific targeting vectors

We designed all gRNAs using the Benchling software package. As targeting construct, we used the commercially available pLentiCRISPRv2 plasmid [11] (Addgene #52961), in which the puromycin resistance gene was replaced by a TagBFP encoding cDNA [12]. Cloning of target locus specific gRNAs into pLentiCRISPRv2 was performed using the Golden Gate protocol [13]. For this:

1. Anneal oligonucleotides complementary to preselected targets, preferably within the 5´UTR by adding 5 µl of each (100µM) to 40 µl ddH_2O. Incubate at 98°C for 5 min and cool down for 20 min at room temperature to achieve annealing.
2. Prepare a Golden Gate reaction mix containing 150 ng pLentiCRISPRv2 by adding 5 µl annealed oligonucleotides, 2 µl 10X ligase buffer, 1 µl T4 DNA ligase and 1 µl of *BstBI* restriction enzyme. Add ddH_2O to a total volume of 20 µl.

3. Perform Golden Gate reaction in a thermocycler as follows: initial digestion at 37°C for 5 min followed by 10 cycles at 16°C for 10 min (ligation), 37°C for 15 min (digestion) followed by denaturation at 80°C for 5 min.
4. Use the Golden Gate reaction product directly for transfection.

3.4. Transfection of HEK293T cells with targeting- and donor-constructs

Choose the donor plasmid that will preserve the target's reading frame assuming indel-less integration. Additionally, choose a donor with the *tia1I* sites on the opposite strand relative to the target. If the endogenous target is on the sense stand, the *tia1I* site should be on the anti-sense stand and vice versa. Using equimolar amounts of donor- and targeting-vectors gave us the best results. Proceed with transfection as follows:

1. Seed 3 x 10^6 HEK293T cells into a 10 cm cell culture dish. Make sure the cells are evenly distributed over the whole surface of the dish. Incubate at 37°C overnight.
2. On the next day, prepare the transfection mix as follows: (a) add 5 µg of each donor and pLentiCRISPRv2 vector to 700 µl plain DMEM, (b) add 50 µl 1 x polyethylenimine (PEI) to 650 µl plain DMEM. Vortex both mixes briefly and incubate for 5 min at room temperature.
3. Combine (a) and (b), vortex briefly and incubate for 20 min at room temperature.
4. Wash cell plates, which should be about 70% confluent, carefully with 5 ml PBS and add 6 ml growth medium (DMEM supplemented with 10% FCS and 100 U/ml Penicillin/Streptomycin) to each plate.
5. Add transfection mix dropwise to the cells and try to spread it evenly across the dish.
6. Incubate for 1 hour at 37°C.
7. Shake plates gently before re-Incubating overnight.

3.5. Isolation of hygromycin resistant clones

1. Wash cells one day after transfection with 5 ml PBS and add 2 ml 1 X Trypsin for 3 min at room temperature. Apply 10 ml growth medium, resuspend cells carefully and transfer to 10 cm dishes containing 10 ml growth medium.
2. After incubating overnight, replace the medium with 10 ml of fresh growth medium containing hygromycin B. Use the minimum hygromycin B concentration required for killing your selected target cells to ensure capturing weakly expressed genes. In our case, as little as 400 µg/ml hygromycin were sufficient to eliminate all wild-type HEK293T cells within 10 days of selection.
3. After 10 days, replace the selection medium with normal growth medium and re-incubate at 37°C.
4. Examine plates daily until single colonies become visible by eye.
5. Carefully detach single colonies using a 200 µl pipette and place them individually into the wells of a 48-well dish containing 500 µl growth medium.

6. Incubate 48-well dishes for one day at 37°C, wash with 2 ml PBS and add 50 µl 1 X Trypsin per well.
7. After incubating for 3 min at room temperature add 500 µl growth medium, gently pipette the colonies up and down with a 200 µl pipette to obtain a single cell suspension and grow cells to 70% confluency. Note that incubation times may vary from clone to clone.
8. Divide cells of each clonal cell line into two wells of a 24-well plate and incubate again until 70% confluent. Use one well for DNA analysis and the other for proteomic assays.

3.6. Molecular analysis of integration sites

Correct integrations are identified by PCR using primers directing the amplification of tag-endogenous gene junctions (Fig.1A), followed by Sanger sequencing. Note that unlike the gene specific primers, the primers binding to the tag are generic and can be used for any target site. Correct tag integrations generate PCR products from both 5´ and 3´junctions whereby the 3´junction sequence should be in frame with the endogenous open reading frame (ORF). This is not required for the 5´junction because the translation initiating ATG is located downstream of the junction (Fig. 1A). Therefore, sequencing the 5' junctions is not necessarily required.

Set up PCR reaction as follows:

1. Extract and purify genomic DNA using any standard protocol.
2. Set up PCR mix containing 5 µl 10 x Red Taq Buffer, 2 µl dNTPs (10 µM each), 2.5 µl of each primer (10 µM each), 2 µl genomic DNA (150 ng/µl), 2 µl Jumpstart™ REDTaq® DNA Polymerase and 34 µl ddH_2O.
3. Run the following PCR program: Initial denaturation at 94°C for 2 min followed by 34 cycles at 94°C for 30 s (denaturation), 60°C for 30 s (annealing), 72°C for 1 min (elongation) followed by a final elongation at 72°C for 10 min. Note that annealing temperatures are primer dependent and should be adjusted accordingly.

3.7. Analysis of tagged protein expression

Confirm nLAP-tagged protein expression by Western blot and flow cytometry.

1. Use whole cell lysates for standard Western blotting and anti-EGFP antibodies to visualize the nLAP tagged proteins. In addition, where available you can probe the blots with high affinity antibodies against the native protein to identify both tagged and native proteins.
2. For flow cytometry grow cells to 70% confluency and trypsinize by covering the cells 1 X Trypsin solution for 3 min at room temperature. Resuspend cells carefully and directly transfer a small aliquot to a flow cytometry tube. Add 4 ml PBS and centrifuge for 5 min at 158 x g. Resuspend

cells in PBS and subject to flow cytometry using a LSRFortessa FACS machine or equivalent (excitation at 488 nm, detection at 530 nm).

3.8. Analysis of protein-protein interactions

To study protein-protein interactions use EGFP specific antibodies coupled to agarose beads (see Section 2.4) in co-immunoprecipitation experiments as follows:

1. Seed 4 x 10^6 cells on a 10 cm tissue culture dish and grow cells for 48h until confuent
2. Wash cells with 10 ml ice cold PBS. Prepare cell lysates with 400 µl CoIP buffer and measure the protein concentration (Bradford assay).
3. Incubate between 1000 - 3000 µg of protein overnight with freshly vortexed 15 µl GFP-Trap^R_A beads at 4°C on a rotating wheel.
4. Centrifuge for 2 min, 3000 x g at 4°C and wash the pelleted beads 5 times in CoIP buffer.
5. Run SDS-PAGE and transfer to blotting membrane.

3.9. Preparation of cells for confocal microscopy

1. Grow nLAP tagged HEK293T cells on round coverslips (15 mm diameter) until 50% confluent. Wash cells carefully with PBS and fix in 4% paraformaldehyde (PFA) in PBS for 10 min at room temperature in the dark.
2. Remove PFA and wash cells 2 X with PBS.
3. Apply PBS supplemented with 1% Bovine Serum Albumine (BSA) for blocking and 1 mg/ml Digitonin for permeabilzation. Incubate in the dark for 15 min at room temperature.
4. For co-staining, flip cover slips upside down onto parafilm carrying a droplet of 35 µl PBS supplemented with 1% BSA, 1 mg/ml Digitonin and the desired primary antibody (see Section 2.4 for dilutions). Incubate for 1 hour at room temperature in the dark. Lift coverslips by flushing with PBS to avoid cell sheering.
5. Wash in PBS and cover with PBS containing 1% BSA, 1 mg/ml Digitonin and the desired secondary antibody (see Section 2.4 for dilutions). Incubate in the dark for 1 hour at room temperature.
6. Mount coverslips with 25 µl Fluoromount containing 50 mg/ml 1,4-diazabicyclo[2.2.2]octane (DABCO) and examine the coverslips under a Zeiss LSM 710 confocal microscope or equivalent.

4. Results

4.1. Tagging mTORC1 proteins in HEK293T cells

To demonstrate allele-specific tagging, we chose 4 proteins related to mTORC1 signaling, - a major topic of the laboratory (Fig. 2) [14]. Accordingly, we co-transfected HEK293T cells with donor- and

pLentiCRISPRv2-TagBFP vectors targeting the following proteins: NPRL3, WDR24, RAPTOR and mTOR (Fig. 2).

Sixty-seven clones were isolated and subjected to molecular analysis by genomic PCR and sequencing. Twenty-nine of these exhibited correct tag integrations predicting fusion protein expression (Table 1). The rest either failed to generate a PCR product for at least one junction or incorporated frameshifting indels into the 3´junctions, suggesting that these clones expressed nLAP tags from off target integrations. Of the 19 nLAP-mTOR clones with 3´ ORFs, 17 preserved the wild-type sequence whereas two revealed 3 bp and 9 bp deletions resulting in loss of 1 and 3 amino acids, respectively. Of the eight nLAP-NPRL3 clones with 3´ ORFs, only one showed a 12 bp deletion. Furthermore, two clones with in-frame 3´junctions failed to yield PCR products from the 5´junctions, resulting in six clones with correct nLAP tag incorporation (Table 1). Of ten nLAP-WDR24 clones only one showed nLAP tag integration with a preserved ORF at the 3´junction. Finally, only three of 12 nLAP-RAPTOR clones displayed correct 3´ ORF preserving integrations. Thus, tagging efficiencies were variable and target locus dependent.

Table 1: Summary of protein tagging results within the mTORC1 pathway.

Target locus	PCR positive 3´junction	in frame 3´junction	3`WT sequence preserved	PCR positive 5´junction	correct integration
mTOR	19/21	19/19	17/19	19/19	19/21
NPRL3	8/24	8/8	7/8	6/8	6/24
WDR24	3/10	1/3	1/1	1/1	1/10
RAPTOR	4/12	3/4	3/3	3/3	3/12
Sum	34/67	31/34	28/31	29/31	29/67

4.2. Tagged proteins are expressed in HEK293T cells

To verify tagged protein expression, we subjected several clones with correct nLAP tag incorporations to Western blot analysis using anti-EGFP and anti-native protein antibodies. Figure 3 (left panel) shows that all modified clones expressed the respective nLAP-tagged protein although expression levels varied, presumably reflecting clonal variation and/or allele-specific variegation. Moreover, all clones expressed the respective wild-type protein when probed with antibodies against the native protein except for nLAP-WDR24, suggesting that in the latter either all alleles were tagged successfully or untagged alleles incorporated large indels resulting in gene inactivation. When analyzed by flow cytometry using EGFP autofluorescence, fusion proteins were detected in all clones

whereby fluorescence intensity mostly correlated with protein expression levels revealed by the Western blots (Fig. 3, right panel and data not shown).

4.3. Tagged proteins bind to their native interaction partners

To demonstrate whether the nLAP tagged proteins reproduce the native protein-protein interactions reported for the mTORC1/GATOR complex (Fig. 2), we selected nLAP-mTOR clone 4, nLAP-NPRL3 clone 37, nLAP-WDR24 clone 36 and nLAP-RAPTOR clone 3 for co-immunoprecipitation experiments. As shown in Figure 4, nLAP-mTOR pulled down RAPTOR, nLAP-RAPTOR mTOR and RagA, nLAP-NPRL3 NPRL2 and nLAP-WDR24 WDR59 and MIOS. Overall, the expected protein-protein interaction patterns could be largely reproduced with the tagged proteins confirming previous studies showing that the nLAP tag (35,55 kDa) does not significantly alter native protein binding affinities [10].

4.4. Tagged proteins mirror the native protein localization patterns

To test whether nLAP tagged proteins interfere with canonical subcellular localization patterns of the mTORC1 pathway proteins, we used the nLAP tag's EGFP autofluorescence and immunofluorescence staining to simultaneously visualize the tagged proteins and their specific interactors. Confocal microscopy revealed partial perinuclear co-localization of mTOR with RAPTOR, WDR24 and NPRL3 (Fig. 5A). The punctiform localization pattern observed on all slides likely reflects lysosomal co-localization consistent with the lysosomes' role as mTORC1 signaling hubs [15]. To verify that the observed green fluorescence corresponded to the tagged proteins, we counterstained nLAP-WDR24 with a WDR24 specific antibody. As expected, nLAP-WDR24 and WDR24 showed identical localization patterns (Fig. 5B). Taken together, these results confirm previous observations showing that the N-terminal nLAP tag does not interfere with subcellular protein localization [16].

5. Discussion

Due to an overall shortage of high affinity antibodies against cellular proteins, protein-protein interactions studies frequently rely on protein tags that are easily detectable by dedicated antibodies. Ideally, the tagged proteins faithfully reproduce the interactomes of their corresponding native proteins. However, reproduction is frequently compromised by tagged protein overexpression from cDNA expression vectors which are still widely used in the analysis of signal transduction pathways. To achieve physiological expression of tagged proteins, tags need to be introduced into the endogenous coding genes to ensure endogenous control of gene expression. This is preferably accomplished by knocking-in the tag into a chosen target gene by homologous recombination. Until recently conventional gene targeting was reasonably successful only in mouse ESCs, but the development of designer endonucleases and in particular of the CRISPR/Cas9 system capable of

introducing DSBs at prespecified genomic sites, gene targeting also became available for a large variety of somatic cells. However, targeted knock-ins of foreign DNA are still cumbersome even with CRISPR/Cas9 technology because correct insertions are dependent on homology templates individually designed for every single gene. To circumvent this problem, attempts have been made to target foreign DNA into prespecified loci in a homology-independent manner. As has been recently shown, homology-independent integration of foreign DNA into CRISPR/Cas9 triggered DSBs was successfully achieved in zebrafish and human ESCs and even in non-dividing somatic cells *in vivo* [17□,18□,19]□. The efficiency of this homology independent knock-in strategy has been enhanced significantly by Lackner et al [7] and used successfully for *in situ* protein tagging in mammalian cells.

In this study, we simplified the Lackner et al strategy by using a generic protein tagging (donor) plasmid enabling drug selection for correct knock-in events. Unlike the Lackner et al method requiring limiting dilution and FACS sorting of numerous clones to identify correctly tagged recombinants, the method employed here using a hygro-nLAP tag enabled recovery of tagged protein expressing cells from less than a dozen hygromycin resistant clones.

To enable tagging of every protein expressed in a particular target cell, we developed a tool kit consisting of a pLentiCRISPRv2-TagBFP targeting vector expressing Cas9 and a target locus specific gRNA plus a set of six hygro-nLAP donor vectors covering all three reading frames, each equipped with either sense or anti-sense *tia1l* sequences to prevent post-insertional gRNA target site reconstitution.

Use of the tool kit for tagging proteins of the mTORC1 pathway revealed variable tagging efficiencies, presumably due to different targeting efficiencies of individual gRNAs or positional variegation, as was shown previously in high-throughput gene targeting performed in mouse ESCs [20]. However, the tagging efficiencies ranged from 10 to 90%, suggesting that in practice, screening as few as 10 hygromycin resistant clones per target would yield at least one clone expressing a correctly nLAP tagged protein. Finally, as demonstrated in protein-protein interaction- and subcellular-localization studies, the tagging strategy largely preserves the properties of the native proteins, confirming previous observations showing that LAP-tags are compatible with native protein function [10□,21□,22]□.

Acknowledgements

We thank Jenny Bleeck for expert technical assistance. This work was supported by grants from the Bundesministerium für Bildung und Forschung (BMBF) to HvM (NGFNplus-DiGtoP/01GS0858), the Deutsche Forschungsgemeinschaft to HvM (ME 820/5-1, ME 820/6-1) and FS (SCHN1166/4-1), the SFB815 and by individual grants from the Loewe Center for Cell and Gene Therapy to NK & FS.

References

[1] Ahmad HI, Ahmad MJ, Asif AR, Adnan M, Iqbal MK, Mehmood K, et al. A review of crispr-based genome editing: Survival, evolution and challenges, Curr Issues Mol Biol. (2018);28:47–68.

[2] Zhang JP, Li XL, Li GH, Chen W, Arakaki C, Botimer GD, et al. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage, Genome Biol. Genome Biology; (2017);18(1):1–18.

[3] Byrne SM, Ortiz L, Mali P, Aach J, Church GM. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells, Nucleic Acids Res. (2015);43(3):e21.

[4] Wang J, Exline CM, DeClercq JJ, Llewellyn GN, Hayward SB, Li PW-L, et al. Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors, Nat Biotechnol (2015) Nov 9;33:1256.

[5] Branzei D, Foiani M. Regulation of DNA repair throughout the cell cycle, Nat Rev Mol Cell Bio (2008) Feb 20;9:297.

[6] Suzuki K, Izpisua Belmonte JC. In vivo genome editing via the HITI method as a tool for gene therapy, J Hum Genet (2018);63(2):157–64.

[7] Lackner DH, Carré A, Guzzardo PM, Banning C, Mangena R, Henley T, et al. A generic strategy for CRISPR-Cas9-mediated gene tagging, Nat Commun (2015);6:10237.

[8] Cristea S, Freyvert Y, Santiago Y, Holmes MC, Urnov FD, Gregory PD, et al. In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration, Biotechnol Bioeng. (2013);110(3):871–80.

[9] Schnütgen F, Ehrmann F, Ruiz-Noppinger P, von Melchner H. High throughput gene trapping and postinsertional modifications of gene trap alleles, Methods. (2011);53(4):347–55.

[10] Poser I, Sarov M, Hutchins JRA, Hériché J-K, Toyoda Y, Pozniakovsky A, et al. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals, Nat Methods (2008) May;5(5):409–15.

[11] Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening, Nat Methods (2014) Aug;11(8):783–4.

[12] Sürün D, Schwäble J, Tomasovic A, Ehling R, Stein S, Kurrle N, et al. High Efficiency Gene Correction in Hematopoietic Cells by Donor-Template-Free CRISPR/Cas9 Genome Editing, Mol Ther - Nucleic Acids (2018);10(March):1–8.

[13] Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. PLoS One. (2008);3(11).

[14] Bar-Peled L, Sabatini DM. Regulation of mTORC1 by amino acids, Trends Cell Biol (2014) Jul;24(7):400–6.

[15] Betz C, and Hall MN, Where is mTOR and what is it doing there?, J Cell Biol (2013) 203(4), pp. 563–574.

[16] Schnütgen F, Ehrmann F, Poser I, Hubner NC, Hansen J, Floss T, et al. Resources for proteomics in mouse embryonic stem cells, Nat Methods (2011) ;8(2):103–4.

[17] Auer TO, Duroure K, De Cian A, Concordet JP, Del Bene F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair, Genome Res. (2014);24(1):142–53.

[18] He X, Tan C, Wang F, Wang Y, Zhou R, Cui D, et al. Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair, Nucleic Acids Res (2016);44(9):e85–e85.

[19] Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ, et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration, Nature (2016);540(7631):144–9.

[20] Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, et al. A conditional knockout resource for the genome-wide study of mouse gene function, Nature (2011) Jun 15;474:337.

[21] Hubner NC, Bird AW, Cox J, Splettstoesser B, Bandilla P, Poser I, et al. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions, J Cell Biol. (2010);189(4):739–54.

[22] Bird AW and Hyman AA, Building a spindle of the correct length in human cells requires the interaction between TPX2 and Aurora A, *Journal of Cell Biology*, (2008)182(2), pp. 289–300.

Figure legends

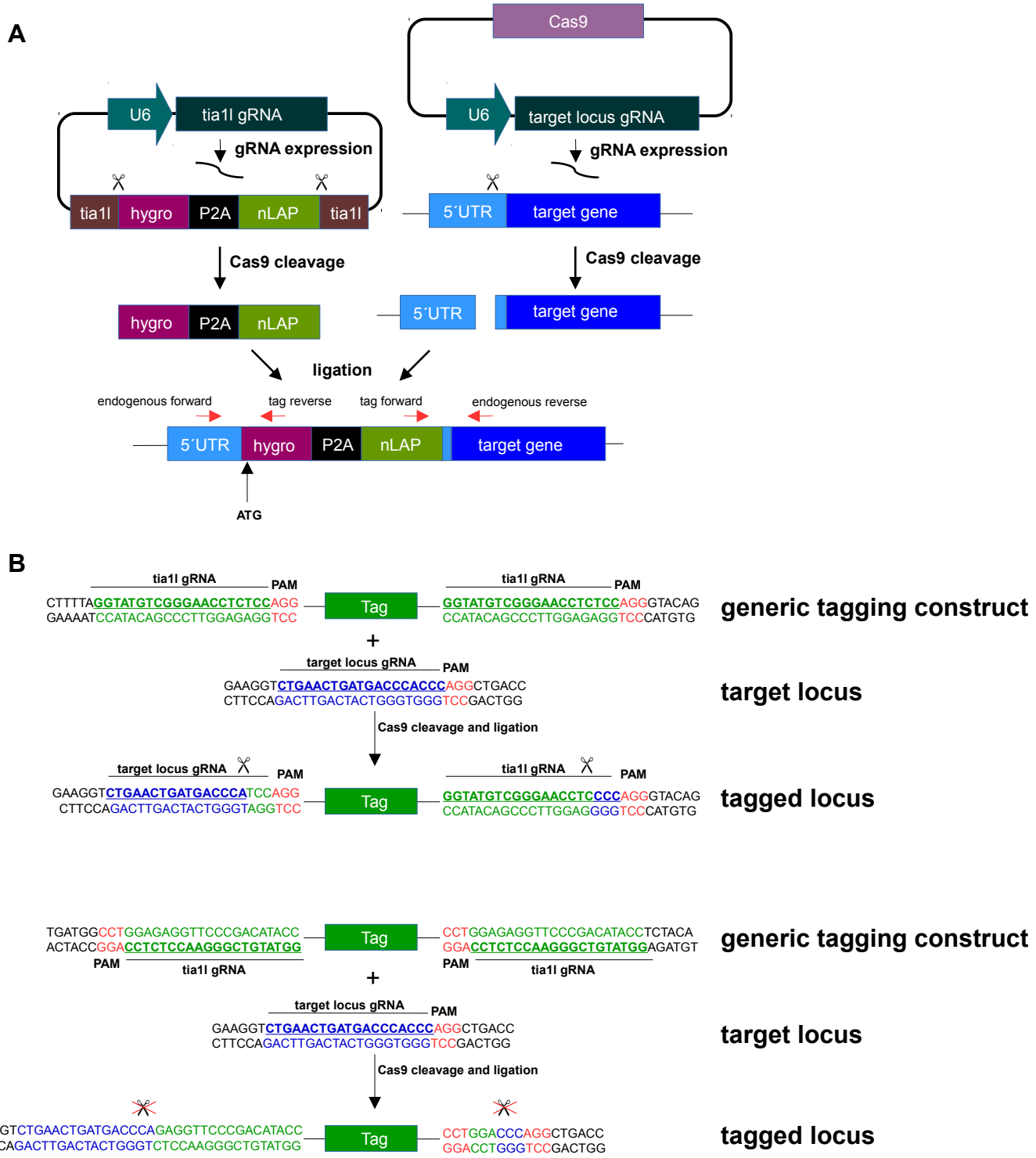
Figure 1. Homology independent tagging strategy. (**A**) Schematic representation of the tagging strategy. Cells are co-transfected with a targeting construct (here pLentiCRISPRv2) expressing Cas9 and a target locus specific gRNA and a generic donor construct comprising a hygro-P2A-nLAP cassette tag flanked by *tia1l* sites and a U6 promoter driving the expression of the gRNA targeting the *tia1l* sites. Expression of Cas9 releases the hygro-P2A-nLAP tag from the donor which is inserted into the target site, here inside the 5´UTR. Red arrows indicate the position of the primers used for verifying tag-endogenous gene junctions. (**B**) Mechanism of gRNA target recreation. *Upper panel:* Same strand position of *tia1l* and endogenous gRNA targets can lead to target sequences amenable to Cas9 cleavage in presence of the existing gRNAs. *Lower panel:* Target recreation is prevented by having *tia1l* and endogenous targets positioned on opposite strands. gRNA, guide RNA; hygro, hygromycin resistance gene; P2A, 2A virus polyprotein cleavage sequence; nLAP, N-terminal protein localization and protein purification tag nLAP tag [10], 5'UTR, 5' untranslated region.

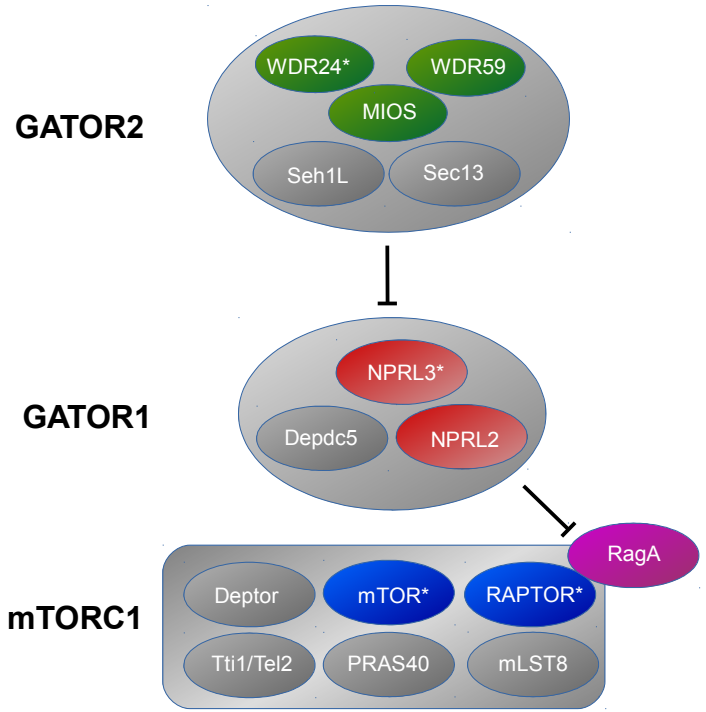
Figure 2. The mTORC1/GATOR interactome. mTORC1 signaling is activated by RagA binding to RAPTOR. Biding is regulated by the upstream GATOR1 and GATOR2 protein complexes [14]. nLAP-tagged proteins are indicated by asterisks.

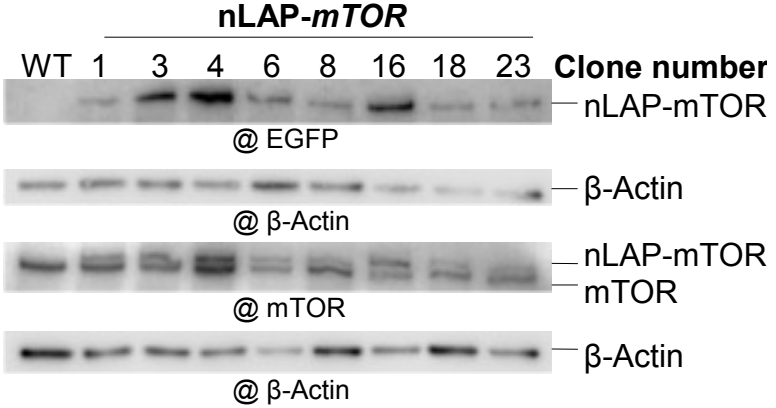
Figure 3. Expression of nLAP tagged proteins. *Left panels:* Protein expression from tagged and wild-type alleles visualized on Western blots probed with anti-EGFP and anti-native protein antibodies. *Right panels:* FACS profiles of nLAP-tagged proteins in selected clones.

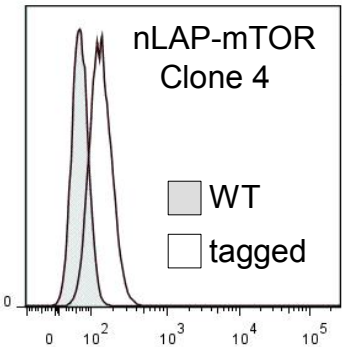
Figure 4. Binding of native mTORC1/GATOR proteins to nLAP-tagged proteins. nLAP-mTOR, nLAP-NPRL3, nLAP-WDR24 and nLAP-RAPTOR were immunoprecipitated from selected HEK293T clones using anti-EGFP antibodies. Interactors were identified on Western blots with high affinity anti-native protein antibodies.

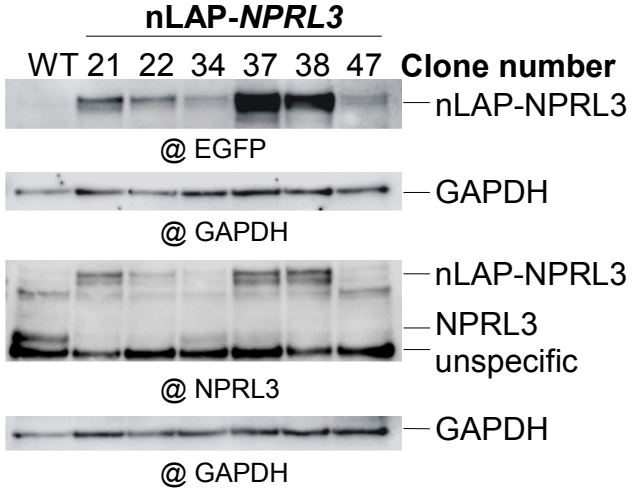
Figure 5. Lysosomal localization of mTORC1/GATOR proteins. (**A**) Subcellular co-localization of nLAP-RAPTOR, nLAP-WDR24 and nLAP-NPRL3 in selected HEK293T clones visualized by EGFP autofluorescence and indirect immunofluorescence staining. (**B)** Control staining of nLAP-WDR24 expressing cells showing identical EGFP autofluorescence and WDR24 immunostaining patterns patterns. Scale bar: 20 µm.

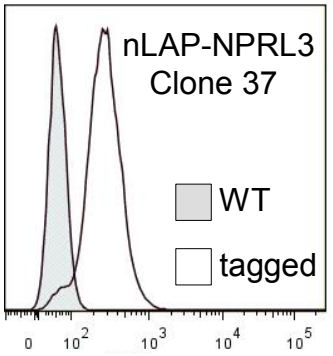

A
Cas9
U6
tia1l gRNA
gRNA expression
U6
target locus gRNA
gRNA expression
tia1l
hygro
P2A
nLAP
tia1l
5´UTR
target gene
Cas9 cleavage
Cas9 cleavage
hygro
P2A
nLAP
5´UTR
target gene
ligation
endogenous forward
tag reverse
tag forward
endogenous reverse
5´UTR
hygro
P2A
nLAP
target gene
ATG
B
tia1l gRNA
PAM
CTTTTAGGTATGTCGGGAACCTCTCCAGG
GAAAATCCATACAGCCCTTGGAGAGGTCC
Tag
tia1l gRNA
PAM
GGTATGTCGGGAACCTCTCCAGGGTACAG
CCATACAGCCCTTGGAGAGGTCCCATGTG
generic tagging construct
+
target locus gRNA
PAM
GAAGGTCTGAACTGATGACCCACCCAGGCTGACC
CTTCCAGACTTGACTACTGGGTGGGTCCGACTGG
target locus
Cas9 cleavage and ligation
target locus gRNA
PAM
GAAGGTCTGAACTGATGACCCATCCAGG
CTTCCAGACTTGACTACTGGGTAGGTCC
Tag
tia1l gRNA
PAM
GGTATGTCGGGAACCTCCCCAGGGTACAG
CCATACAGCCCTTGGAGGGGTCCCATGTG
tagged locus
TGATGGCCTGGAGAGGTTCCCGACATACC
ACTACCGGACCTCTCCAAGGGCTGTATGG
PAM
tia1l gRNA
Tag
CCTGGAGAGGTTCCCGACATACCTCTACA
GGACCTCTCCAAGGGCTGTATGGAGATGT
PAM
tia1l gRNA
generic tagging construct
+
target locus gRNA
PAM
GAAGGTCTGAACTGATGACCCACCCAGGCTGACC
CTTCCAGACTTGACTACTGGGTGGGTCCGACTGG
target locus
Cas9 cleavage and ligation
GAAGGTCTGAACTGATGACCCAGAGGTTCCCGACATACC
CTTCCAGACTTGACTACTGGGTCTCCAAGGGCTGTATGG
Tag
CCTGGACCCAGGCTGACC
GGACCTGGGTCCGACTGG
tagged locus

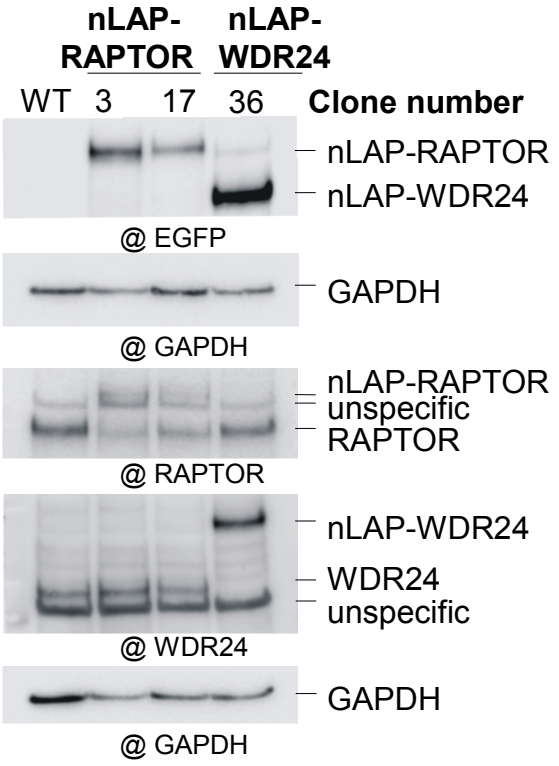

GATOR2
WDR24*
WDR59
MIOS
Seh1L
Sec13
GATOR1
NPRL3*
Depdc5
NPRL2
RagA
mTORC1
Deptor
mTOR*
RAPTOR*
Tti1/Tel2
PRAS40
mLST8

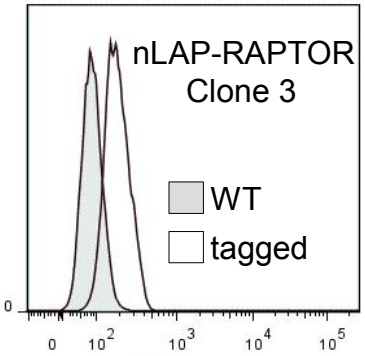

nLAP-*mTOR*
WT 1 3 4 6 8 16 18 23 Clone number
nLAP-mTOR
@ EGFP
β-Actin
@ β-Actin
nLAP-mTOR
mTOR
@ mTOR
β-Actin
@ β-Actin

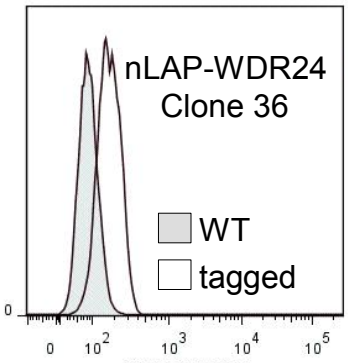

nLAP-mTOR
Clone 4
WT
tagged

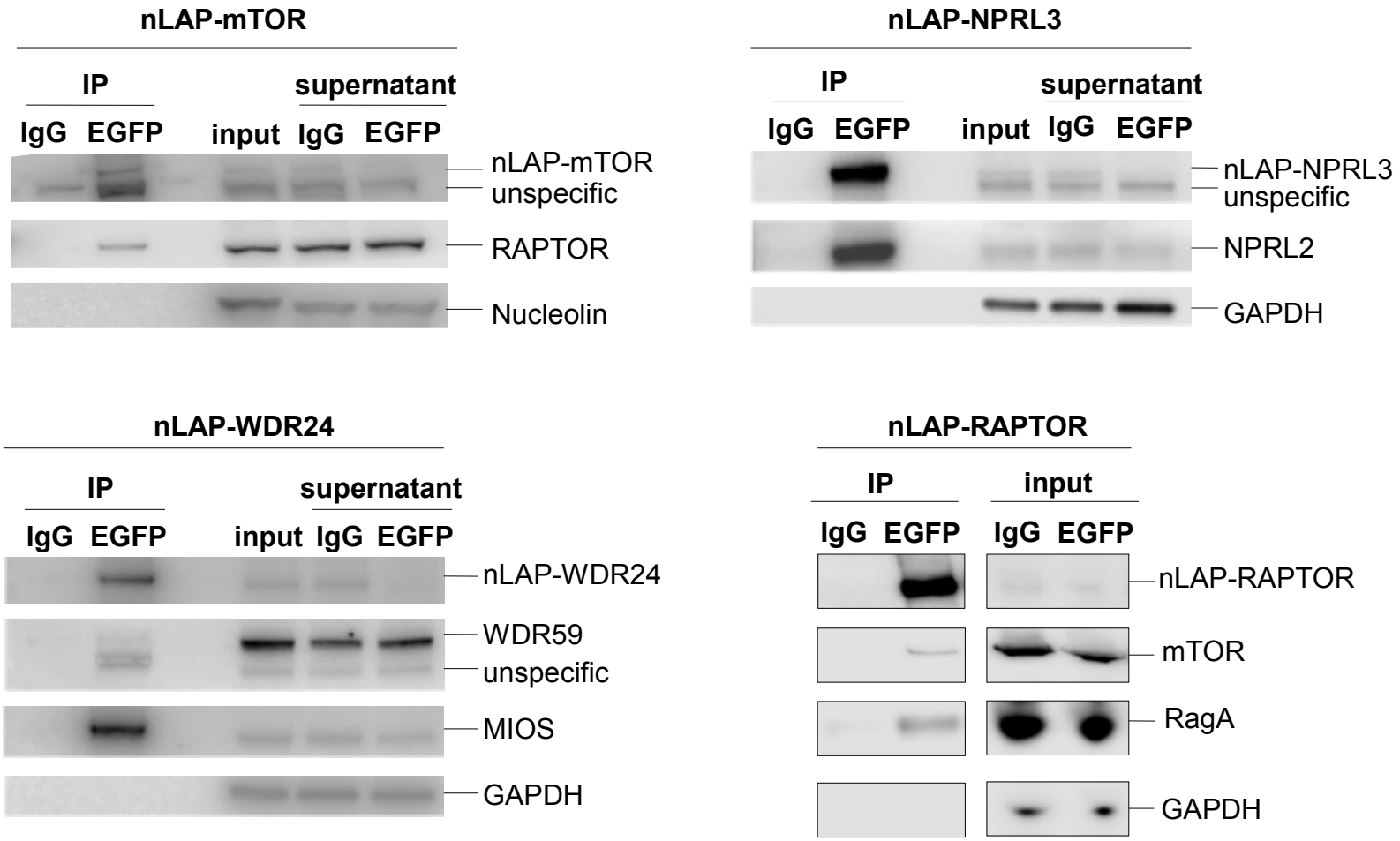

nLAP-*NPRL3*
WT 21 22 34 37 38 47 Clone number
nLAP-NPRL3
@ EGFP
GAPDH
@ GAPDH
nLAP-NPRL3
NPRL3
unspecific
@ NPRL3
GAPDH
@ GAPDH

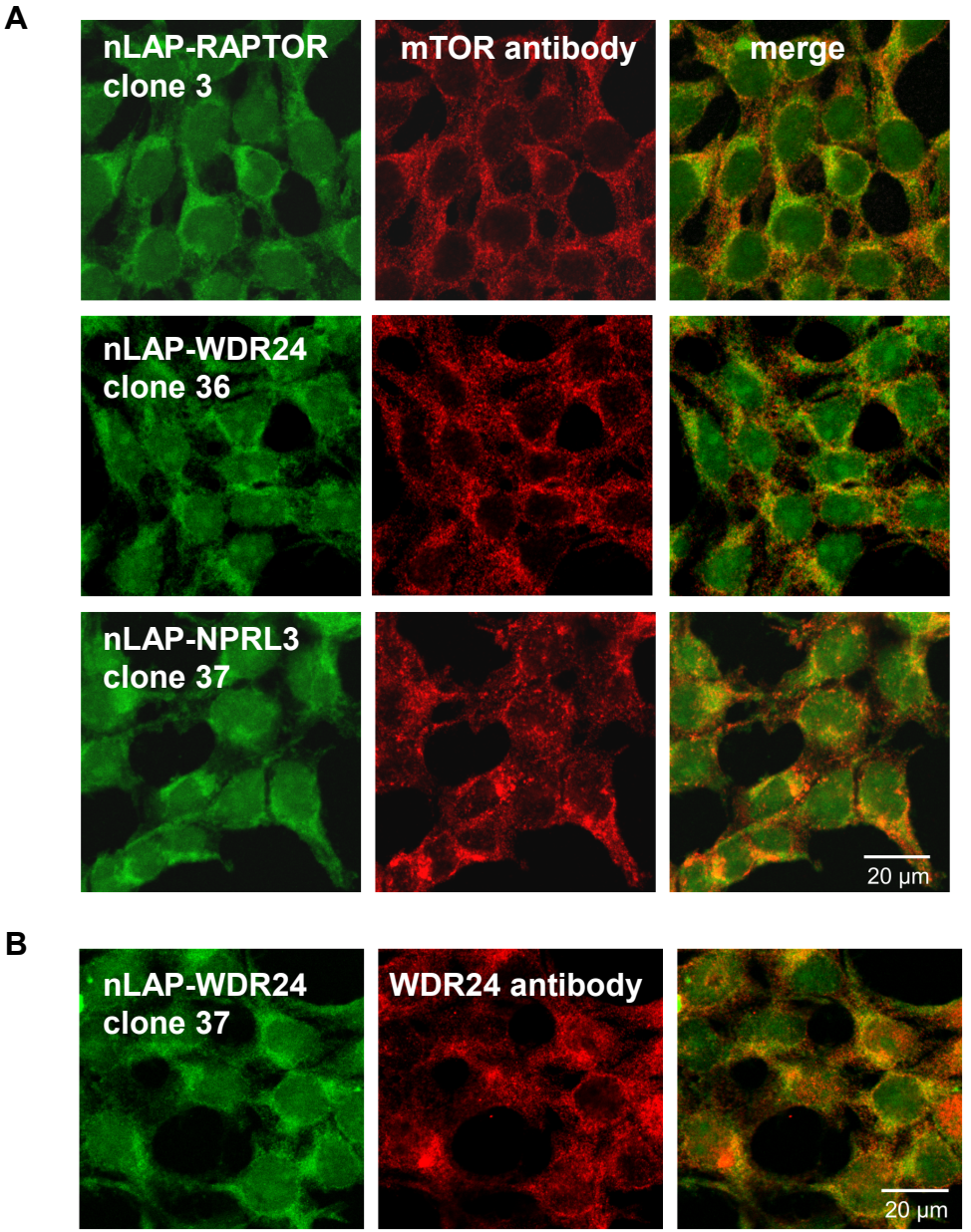

nLAP-NPRL3
Clone 37
WT
tagged


nLAP-
RAPTOR
nLAP-
WDR24
WT 3 17 36 Clone number
nLAP-RAPTOR
nLAP-WDR24
@ EGFP
GAPDH
@ GAPDH
nLAP-RAPTOR
unspecific
RAPTOR
@ RAPTOR
nLAP-WDR24
WDR24
unspecific
@ WDR24
GAPDH
@ GAPDH


nLAP-RAPTOR
Clone 3
WT
tagged


nLAP-WDR24
Clone 36
WT
tagged


nLAP-mTOR
IP
supernatant
IgG
EGFP
input
IgG
EGFP
nLAP-mTOR
unspecific
RAPTOR
Nucleolin
nLAP-NPRL3
IP
supernatant
IgG
EGFP
input
IgG
EGFP
nLAP-NPRL3
unspecific
NPRL2
GAPDH
nLAP-WDR24
IP
supernatant
IgG
EGFP
input
IgG
EGFP
nLAP-WDR24
WDR59
unspecific
MIOS
GAPDH
nLAP-RAPTOR
IP
input
IgG
EGFP
IgG
EGFP
nLAP-RAPTOR
mTOR
RagA
GAPDH


A
nLAP-RAPTOR clone 3
mTOR antibody
merge
nLAP-WDR24 clone 36
nLAP-NPRL3 clone 37
20 µm
B
nLAP-WDR24 clone 37
WDR24 antibody
20 µm