Published in final edited form as:

Methods Mol Biol. 2009 ; 567: 87–98. doi:10.1007/978-1-60327-414-2_6.

Epitope tagging of endogenous proteins for genome wide Chromatin immunoprecipitation analysis

Zhenghe Wang

Department of Genetics and Case Comprehensive Cancer Center, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106

Zhenghe Wang: zwx22@case.edu

Abstract

The development of chromatin immunoprecipitation methods coupled with DNA microarray (ChIP-chip) technology has enabled genome-wide identification of cis-DNA regulatory elements to which transcription factors bind. Nonetheless, the ChIP-chip technology requires antibodies with extremely high affinity and specificity for the target transcription factors. Unfortunately, such antibodies are not available for most human transcription factors. In principle, this problem can be circumvented by utilizing ectopically expressed epitope-tagged proteins recognizable by well-characterized antibodies. However, such expression is no longer endogenous. To surmount this problem, we have successfully developed a facile method to knock-in a 3xFlag epitope into the endogenous gene loci of transcription factors. The knock-in approach provides a general solution for the study of proteins for which antibodies are substandard or not available.

Keywords

Epitope tag; ChIP-chip; recombinant adeno-associated virus; knock-in; colorectal cancer

1. Introduction

The human genome encodes approximately 25,000 proteins. Characterizing all 25,000 depends on availability of high quality antibodies that can be used for multiple applications including Western blot, immunofluorescence (IF), and immunoprecipitation (IP). For analysis of transcription factors and other DNA-binding proteins, “ChIP grade” antibodies capable of immunoprecipitating the protein of interest within the context of chromatin are most often desired [1]. Notwithstanding, ChIP-grade antibodies exist for only a small fraction of chromatin-associated proteins. This is particularly problematic for ChIP-chip or ChIP-sequencing studies, where the use of more than one antibody is highly recommended. The antibody problem can be circumvented by generating cell lines that stably express epitope-tagged proteins recognizable by available antibodies, but this approach is far from ideal given that expression is no longer endogenous, which may complicate interpretation of results. Moreover, the construction of recombinant plasmids containing both full length cDNA and epitope sequences can be cumbersome, particularly for proteins encoded by large transcripts.

Epitope-tagging by homologous recombination-mediated knockin (KI) is an effective means for biochemical and cellular studies of proteins in recombination prone organisms, such as yeast [2]. Applying this approach to somatic mammalian cells is not feasible due to low frequency of homologous recombination between exogenous plasmid and specific genomic loci. Recent studies have shown that this problem can be circumvented by delivering constructs with recombinant adeno-associated virus (rAAV), which can increase the

frequency of homologous recombination to as much as 2% [3]. We have successfully developed a method whereby rAAV is used to “knock in” epitope tag sequences into targeted loci in human somatic cells [4]. The tagged proteins, which harbor three Flag epitopes in tandem (3xFlag), can be exploited for Western blot, IP, IF, and ChIP-chip analyses [4]. Here, I describe step-by-step protocol of the 3xFlag knockin approach.

2. Materials

2.1. Targeting vector construction

1. pTK-Neo-USER-3xFlag targeting vector.
2. Restriction enzymes: Xba I, Nt.BbvC I (New England Biolabs, Ipswich, MA).
3. Hi-fidelity platinum Taq polymerase (Invitrogen, Carlsbad, CA).
4. USER Enzyme (New England Biolabs, Ipswich, MA).
5. Subcloning EfficiencyTM DH5 α TM Competent Cells (Invitrogen, Carlsbad, CA).
6. LB agar plates with 100 μ g/ml ampicillin.

2.2. rAAV targeting virus generation

1. Dulbecco's Modified Eagle's Medium (DMEM) (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum and 1% Pen/Strep (FBS, HyClone, Ogden, UT).
2. HEK 293 T cells (American Type Culture Collection (ATCC), Manassas, VA).
3. Phosphate Buffered Saline 1xsolution (PBS, HyClone, Ogden, UT).
4. Opti-MEM[®] I Reduced Serum Media (Invitrogen, Carlsbad, CA).
5. LipofectamineTM Transfection Reagent (Invitrogen, Carlsbad, CA).
6. pAAV-RC Plasmid and pHelper Plasmid (Stratagene, La Jolla, CA).
7. Cell scraper (Fisher, Pittsburgh, PA).

2.3. Gene targeting of human cells

1. McCoy's 5A Medium (Invitrogen, Carlsbad, CA) supplemented with 10% FBS and 1% Pen/Strep.
2. DLD1 colorectal cancer cells (ATCC, Manassas, VA).
3. Trypsin-EDTA (Invitrogen, Carlsbad, CA).
4. 96-well tissue culture plates (Fisher).
5. Geneticin (Invitrogen, Carlsbad, CA).

2.4. Genomic DNA preparation

1. Lyse-N-go reagent (Pierce, Rockford, IL)
2. Trypsin-EDTA without phenol red (Invitrogen, Carlsbad, CA).

2.5. Targeted clone screening

1. 96-well PCR plates (Fisher).
2. Platinum taq polymerase (Invitrogen, Carlsbad, CA).

2.6. Excision of the Neomycin resistance gene

1. Adeno-Cre recombinase (Adeno-Cre).
2. 6-well plates (Fisher).
3. 24-well plates (Fisher).

3. Method

The 3xFlag tag sequences are inserted before the stop codon of target genes through rAAV-mediated homologous recombination (outlined in Figure 1). The entire procedure can be arbitrarily divided into 6 major steps: (1) Targeting vector construction; (2) rAAV targeting virus generation; (3) Gene targeting of human cells; (4) Genomic DNA preparation; (5) Targeted clone screening; and (6) Excision of the Neomycin resistance gene. It takes ~ 45 days to generate 3xFlag knock-in clones in DLD1 cells.

We also developed a one-step highly efficient targeting vector construction strategy (Figure 2). Recently, the New England Biolabs has developed the USER (uracil-specific excision reagent) cloning technique, which facilitates assembly of multiple DNA fragments in a single reaction by in vitro homologous recombination and single-strand annealing [5]. In this system, the vector contains a cassette with two inversely oriented nicking endonuclease sites separated by restriction endonuclease site(s). The vector is then digested and nicked with restriction endonucleases, yielding a linearized vector with 8-nucleotide single-stranded, non-complimentary overhangs. To generate target molecules for cloning into this vector, a single deoxyuridine (dU) residue is placed 8 nucleotides from the 5'-end of each PCR primer. In addition to the dU, the PCR primers contain sequence that is compatible with each unique overhang on the vector. After amplification, the dU is excised from the PCR products with a uracil DNA glycosylase and an endonuclease (the USER enzyme), generating PCR products flanked by 3-prime, 8 nucleotide single-stranded extensions that are complementary to the vector overhangs. When mixed together, the linearized vector and PCR products directionally assemble into a recombinant molecule through complementary single-stranded extensions. To make the rAAV-mediated targeting vector compatible with the USER cloning system, we inserted cassette A (Cst A) between L-ITR and 3xFlag sequences, and cassette B (Cst B) between the right lox P site and R-ITR of the AAV-3xFlag knockin vector to generate the AAV-USER-3xFlag-KI vector (Figure 2). These cassettes contain two inversely oriented nicking endonuclease sites (Nt. BbvCI) separated by restriction endonuclease sites (Xba I). After treatment with Nt.BbvC I and Xba I restriction enzymes, the AAV-USER-3xFlag-KI vector is digested into a 3xFlag-lox P-Neo-lox P fragment flanked by two 5' single-stranded overhangs (Figure 2) and a vector backbone flanked by two 5' overhangs (Figure 2). PCR is then used to amplify left and right homologous arms from genomic DNA. The sequence GGGAAAGdU is added to the 5' of the forward left arm primers, and GGAGACAdU is added to the reverse left arm primers. GGTCCCAAdU is added to the forward right arm primers and GGCATAGdU to the reverse right arm primers. The PCR products are then treated with the USER enzymes to generate single-stranded overhangs. Finally, the left and right arms are mixed with the two vector fragments followed by bacterial transformation (Figure 2).

3.1 Targeting vector construction

1. Design PCR Primers using Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) as follows:

Left arm:

Forward primer: add GGGAAAGdU to the 5' end of the designed PCR primer.

Reverse primer: add GGAGACAdUnn to the 5' end of the reverse sequences of the upstream of stop codon (the first n could be A, T, G, or C; the second n could be any nucleotides but A so that the 3xFlag is in frame fused with the targeted gene, and avoid to introduce a stop codon before the 3xFlag).

Right arm:

Forward primer: add GGTCCCAAdU to the downstream sequences of stop codon.

Reverse primer: add GGCATAGdU to the 5' end of the designed PCR primer.

2. Amplification of left and right arms.

- a.** Use DLD1 genomic DNA (or genomic DNA from the cell that you intend to target) as the templates. The left and right arms are generated by PCR in two separate reactions (20 μ l each) according to the following receipt and cycling conditions:

10 μ l reaction	1	15
H ₂ O	5.58	83.7
10 $\times$ HiFi Buffer	1	15
10mM dNTPs	0.2	3
50mM MgSO ₄	0.4	6
DMSO	0.6	9
Primer 1 (50 μ M)	0.06	0.9
Primer 2 (50 μ M)	0.06	0.9
DNA	2	30
HiFiTaq (5 u/ μ l)	0.1	1.5

94 °C for 2 min; 1 cycle

94 °C for 10 sec, 64 °C for 30 sec, 68 °C for 1 min; 4 cycles

94 °C for 10 sec, 61 °C for 30 sec, 68 °C for 1 min; 4 cycles

94 °C for 10 sec, 58 °C for 30 sec, 68 °C for 1 min; 4 cycles

94 °C for 10 sec, 55 °C for 30 sec, 68 °C for 1 min; 20 cycles

68 °C for 5 min; 1 cycle

(Extension time should vary according to the length of the arm at 1 kb per min).

- b.** Run the PCR products on a 1% agarose gel and purify both fragments.

3. Vector digestion.

- a.** Digest 5 μ g of pTK-Neo-USER-3Flag vectors DNA with 40 units of Xba I overnight at 37°C in a total volume of 100 μ l.
- b.** Add 20 U of Xba I in the next morning together with 20 units of Nt.BbvCI to the digestion mixture, and incubate for 2 hours at 37°C.
- c.** Run the digestion mixture on 1% agarose gel and excise both fragments. The large fragment is named as B and the small fragment is named as S.
- d.** Extract both the B and S fragments with a gel extraction kit.

4. Insertion of PCR fragments into the USER vectors

- a.** Mix B (~30 ng), S, left arm and right arm together in a 1:10:10:10 molar ratio.

- b. Add 1 μ l of 10x TE buffer (pH8.0) and 1 μ l of USERTM enzyme mixture (1 unit/ μ l) to 8 μ l of the mixture prepared in a.
 - c. Incubate the reaction mixture for 20 min at 37°C, followed by 20 min at 25°C.
5. Transformation
 - a. Mix the entire USER-treated reaction mixture (10 μ l) with 50 μ l of chemically competent *E. coli* cells by heat shock (**Don't use electroporation method**).
 - b. Plate them on LB agar plates supplemented with ampicillin (100 μ g/ml).

3.2. rAAV targeting virus generation

1. Plate HEK293 T cells in a T75 flask one day prior to transfection to achieve a 40–80% confluence at the time of transfection.
2. Prepare two wells of a 24-well tissue culture plate and add 750 μ l of OptiMEM into each well.
3. In one well, add 3 μ g each of the targeting vector, pAAV-RC and pHelper plasmids and mix well. In the second well, add 54 μ l of lipofectamine transfection reagent.
4. Drop the DNA mixture into the lipofectamine mixture and let it sit for 10 – 30 minutes while preparing the HEK293 T cells to be transfected.
5. Rinse the cells once with sterile PBS and once with OptiMEM, then add 7.5 ml of OptiMEM and keep the cells in incubator.
6. Add the lipofectamine/DNA mixture into the HEK 293T cells, rock gently, and return the cells to the incubator.
7. After 3–4 hours, remove the OptiMEM medium and replace with complete medium (DMEM supplemented with 10% FBS and 1% Pen/Strep).
8. Grow the cells for 72 hours prior to harvesting virus.
9. Scrap the transfected cells and pool them with the culture medium in a 15 ml conical tube (The floating cells contain a lot of viruses).
10. Spin cells down at 1000 rpm for 3 min and aspirate medium.
11. Suspend the cells into 1 ml of sterile PBS.
12. Freeze and thaw the pellet three cycles. (Each cycle consists of 10 min freezing in a dry ice-ethanol bath, and 10 min thawing in a 37 °C water bath, vortex after each thawing.).
13. Spin the lysate at 12,000 rpm for 5 min in a micro-centrifuge to remove cell debris.
14. Divide the supernatant containing rAAV into 3 aliquots (~330 μ l each) and freeze them at –80°C. (In general, 1/3 of the virus generated from one T75 cm² flask is sufficient for infection of one 25 cm² flask containing the cells to be targeted).

3.3. Gene targeting of human cells

1. Grow DLD1 cells or cells of your interest to be targeted in a T25 flask at 60–80% confluence.
2. Wash cells once with PBS.

3. Add ~330 μ l of rAAV and then 1.5 ml of the appropriate growth media (McCoy's 5A for DLD1 cells) to the flask.
4. Incubate at 37 °C for 2–5 hours.
5. Add 5 ml of growth media the flask and grow for 48 hours.
6. Harvest cells by trypsinization and resuspend cells in 100 ml of medium containing 1 mg/ml Geneticin.
7. Distribute 50 ml of cell suspension into two 96-well plates (250 μ l/well).
8. Add 50 ml of Geneticin-containing medium to the remaining 50 ml of cell suspension.
9. Repeat step 7 and 8 until you have a stack of 10–20 96 well plates. (The purpose of this is to serially dilute cells so that you will get 1 Geneticin resistance clone/well).
10. Wrap the plates with Saran Wrap to minimize evaporation and incubate them at 37°C for 10–14 days prior to consolidating single clones.
11. Check the plates on day 10 and mark the single clones under the microscope.
12. Consolidate the single clones, once they grow to 1/3–1/2 of the wells
13. Dump the medium from the 96-well plates, add 50 μ l of trypsin into each of the marked wells and incubate the plates at 37°C for >20 min.
14. Prepare a set of 96-well plates with 200 μ l medium added into each well.
15. Transfer all of single clones into the new 96-well plates and grow cells to confluence (If you can't get enough single clones, you can also screen multiple clones).

3.4. Genomic DNA preparation

1. To a monolayer or a large colony in a 96-well tissue culture plate, add 25–30 μ l **trypsin/EDTA WITHOUT PHENOL RED**. This should be roughly 2000–5000 cells/ μ l. Incubate at 37°C for 10 minutes.
2. Using a multi-channel pipette, aliquot 5 μ l of Lyse-N-Go reagent to each well of a 96-well PCR plate.
3. Shake the tissue culture plate gently to dislodge cells. Pipette 2 μ l of cell suspension from each well to the PCR plate containing Lyse-N-Go reagent.
4. Add 200 μ l of fresh medium back to the plate with the trypsinized cells and keep growing them.
5. Cycle as per manufacturer's recommendations:
 - 65 °C, 30 sec
 - 8 °C, 30 sec
 - 65 °C, 1 min and 30 sec
 - 97 °C, 3 min
 - 8 °C, 1 min
 - 65 °C, 3 min
 - 97 °C, 1 min

65 °C, 1 min

80 °C, 5 min

6. Spin down the reactions to get it at the bottom of the tube.
7. Add 20 µl of ddH₂O (PCR grade) to each well, spin down and use 2 µl for the PCR.

3.5. Targeted clone screening

1. Design forward PCR primers upstream of the left arm (close to 5' end of left arm and avoid repetitive sequences). Those primers are designated as left arm screening primers.
2. Design reverse PCR primers downstream of the right arm (close to 3' end of left arm and avoid repetitive sequences). Those primers are designated as right arm screening primers.
3. Pair the left arm screening primers with NR (GTTGTGCCCAGTCATAGCCG) or pair the right arm screening primers with NF (TCTGGATTCATCGACTGTGG) to perform PCRs for screening targeted clones.
4. All PCR reactions are performed with Platinum Taq DNA Polymerase using the conditions specified by the manufacturer. The reaction volume is 10 µl in 96-well plates using the following receipt and cycling conditions:

10µl reaction	1	13	26	50	75	102
H ₂ O	5.68	73.84	147.68	284	426	579.36
10 × Buffer	1	13	26	50	75	102
10mM dNTPs	0.2	2.6	5.2	10	15	20.4
50mM MgCl ₂	0.3	3.9	7.8	15	22.5	30.6
DMSO	0.6	7.8	15.6	30	45	61.2
Primer 1 (50 µM)	0.06	0.78	1.56	3	4.5	6.12
Primer 2 (50 µM)	0.06	0.78	1.56	3	4.5	6.12
DNA	2					
Taq (5U/µl)	0.1	1.3	2.6	5	7.5	10.2

94 °C for 2 min; 1 cycle

94 °C for 10 sec, 64 °C for 30 sec, 68 °C for 1–3 min; 4 cycles

94 °C for 10 sec, 61 °C for 30 sec, 68 °C for 1–3 min; 4 cycles

94 °C for 10 sec, 58 °C for 30 sec, 68 °C for 1–3 min; 4 cycles

94 °C for 10 sec, 55 °C for 30 sec, 68 °C for 1–3 min; 35 cycles

(Extension time should vary according to the length of the arm at 1 kb per min).

3.6 Excision of the Neomycin resistance gene

1. Design a pair of primers surrounding the stop codon to amplify a fragment ~200 bp (Cre screening primers).
2. Transfer the positive clones to 24-well plate to expand them (From now on, do not add Geneticin into medium). Pick at least two of the targeted clones for excision of the neomycin resistance gene.
3. Once confluence, split 2/3 of cells to a 6-well plate to grow as a stock, and transfer the remaining 1/3 cells to a new 24-well plate for adeno-Cre virus infection.

4. Add adeno-Cre virus to the 24-well and grow for 24 hours.
5. Dilute the cells and plate into 96-well plates so that you will have single clones. Incubate the plates for 2 weeks. On day 10, mark single clones.
6. Consolidate 24 clones for each of the Cre-ed clones. Prepare genomic DNA as describe in section 3.4.
7. Perform PCR with the Cre screening primers. The clones with neomycin resistance gene being excised should give two bands (as shown in Figure 3).

4. Notes

1. For left arm reverse and right arm forward primers, you don't have many choices. Just use the sequences around stop codon. Sometimes, it is hard to find good pairs of PCR primers. In this case, amplify a big fragment using left forward primer (P1) and the reverse cre screening primer P4 (Figure 1) first, and then perform nest-PCR to amplify the left arm. You can use the same strategy to amplify the right arm.
2. The USER cloning system is rapid and highly efficient (>80% cloning efficiency). If you have trouble with this system, we also have a targeting vector for the traditional restriction and ligation cloning method. We are happy to send it to you per request.
3. Lyse-N-Go is a reagent from Pierce that is useful for the rapid, inexpensive production of template DNA from cells. Such templates have been used successfully for a number of PCR reactions in which products of up to 5 kb have been amplified robustly. However, Qiagen genomic DNA prep kit is an expensive alternative to produce better quality DNA.
4. After getting the positive clones, make new genome DNA using QIAamp DNA Blood Mini Kit and confirm with two pairs of screening primers across both arms (i.g. Left arm screen primer + NR and Right arm screening primer + NF, Figure 1).
5. It is imperative to confirm expression of Flag tagged proteins by Western blot!
6. We have successfully targeted DLD1, RKO, LOVO and HCT116 colorectal cancer cells so far. Other cell lines should be targetable too.

Acknowledgments

The author would like to thank Dr. Chao Wang for proof reading. This work was supported by RO1 CA127590 and RHG004722A.

References

1. Kim TH, Ren B. Genome-Wide Analysis of Protein-DNA Interactions. *Annu Rev Genomics Hum Genet* 2006;7:81–102. [PubMed: 16722805]
2. Lee TI, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 2002;298(5594):799–804. [PubMed: 12399584]
3. Kohli M, et al. Facile methods for generating human somatic cell gene knockouts using recombinant adeno-associated viruses. *Nucleic Acids Res* 2004;32(1):e3. [PubMed: 14704360]
4. Zhang X, et al. Epitope tagging of endogenous proteins for genome-wide ChIP-chip studies. *Nat Methods* 2008;5(2):163–5. [PubMed: 18176569]
5. Bitinaite J, et al. USER friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Res* 2007;35(6):1992–2002. [PubMed: 17341463]

Figure 1. Schematic diagram of tagging endogenous protein with 3xFlag

rAAV targeting vectors contain a left and right arm homologous to sequences in the target gene, flanking a *NEO Lox P-3xFlag* cassette. Clones are then screened by genomic PCR with primers complementary to the neomycin resistance gene and upstream of the left (indicated as P1 and NR) or downstream of the right (indicated as NF and P2) homologous arms. The neomycin gene cassette is excised with Cre-recombinase and genomic PCR using primers P3 and P4 identifies clones with the correct excision.

Left arm left overhang sequence: GGGAAAGT
 Left arm right overhang sequence: GGAGACAT

Right arm left overhang sequence: GGTCCCAT
 Right arm right overhang sequence: GGCATAGT

Figure 2. One-step USER cloning vector

A. Sequences of cassette A (Cst A) and cassette B (Cst B). **B.** Diagram of targeting vector construction by USER cloning.

Figure 3. Genomic PCR 3xFlag knockin clones

Parental (P) and 3xFlag knockin cells (clone 1 and 2). Arrow indicates the targeted allele.