Published in final edited form as:

Nat Methods. 2008 February ; 5(2): 163–165.

Epitope tagging of endogenous proteins for genome-wide ChIP-chip studies

Xiaodong Zhang^{1,7}, Chunguang Guo^{2,7}, Yueting Chen^{1,7}, Hennady P Shulha³, Michael P Schnetz¹, Thomas LaFramboise¹, Cynthia F Bartels¹, Sanford Markowitz⁴, Zhiping Weng^{3,5}, Peter C Scacheri¹, and Zhenghe Wang^{1,6}

¹Department of Genetics and Case Comprehensive Cancer Center, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA.

²Department of Molecular Biology and Microbiology, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA.

³Biomedical Engineering Department, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA.

⁴Department of Medicine and Ireland Cancer Center, Case Western Reserve University and University Hospitals of Cleveland, and Howard Hughes Medical Institute, Cleveland, Ohio 44106, USA.

⁵Bioinformatics Program, Boston University, 24 Cummington Street, Boston, Massachusetts 02215, USA.

⁶Genomic Medicine Institute, Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA.

Abstract

We developed a strategy to introduce epitope tag–encoding DNA into endogenous loci by homologous recombination–mediated ‘knock-in’. The tagging method is straightforward, can be applied to many loci and several human somatic cell lines, and can facilitate many functional analyses including western blot, immunoprecipitation, immunofluorescence and chromatin immunoprecipitation–microarray (ChIP-chip). The knock-in approach provides a general solution for the study of proteins to which antibodies are substandard or not available.

For ease in constructing directed knock-in of Flag-epitope tags, we developed a universal knock-in vector. The vector contains two multiple cloning sites, sequences that encode a triple Flag epitope tag (3×Flag), a neomycin gene flanked by *loxP* sites and two inverted terminal repeats (Fig. 1). For targeted knock-in of 3×Flag, we inserted sequences homologous to 5' and 3' regions flanking the target locus into the two respective cloning sites, and packaged the resulting vector into recombinant adeno-associated virus (rAAV). Then we infected cells with targeting virus and selected neomycin-resistant clones. We identified correctly targeted clones by genomic PCR and then excised the neomycin gene by infection with adenovirus expressing Cre recombinase. As detailed in Supplementary Methods online, we successfully used this strategy to Flag-tag the C terminus of the following five proteins in three different colorectal cancer cell (CRC) lines: STAT3 (signal transducer and activator of transcription 3; Fig. 1c), PTPN14 (protein tyrosine phosphatase non-receptor 14), MRE11 (meiotic recombination 11),

Correspondence should be addressed to P.C.S. (peter.scacheri@case.edu) or Z. Wang (zhenghe.wang@case.edu).

⁷These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS Z. Wang conceived and designed the experiments. X.Z. developed the tagging strategy and modified the targeting vector for uracil-specific excision reagent (USER) compatible cloning. C.G. performed experiments on the N protein. Y.C. tagged the MRE11 and PTPN14 proteins. H.P.S. and Z. Weng performed the motif identification analyses. P.C.S., M.P.S. and C.F.B. performed all ChIP analyses. T.L. analyzed ChIP data. P.C.S., Z. Wang and S.M. wrote the paper.

CHD7 (chromodomain helicase DNA-binding protein 7) and N, encoding a new protein (Supplementary Fig. 1 online). Knock-in of the 3×Flag epitope was highly efficient, with targeting frequency of 1–4% across these five different knock-in experiments (correctly targeted clones/total neomycin-resistant clones; Supplementary Table 1 online), and the efficiency of Cre-mediated excision of the neomycin cassette was 42–83% (Supplementary Table 1). The resultant epitope tagged proteins were readily detectable by western blot, immunoprecipitation and immunofluorescence, using commercially available anti-Flag (Fig. 2 and Supplementary Fig. 1). Moreover, Flag-STAT3 retained the ability to activate the expression of a target gene¹, and Flag-MRE11 remained associated with known interacting proteins², suggesting that the presence of the 3×Flag tag is not likely to interfere with targeted protein function (Supplementary Fig. 2 online). These data indicate that the rAAV-mediated targeting method is feasible for multiple loci across several different CRC cell lines and that 3×Flag can serve as a universal epitope for several antibody-based applications.

To test whether the tagging method can be used for global chromatin immunoprecipitation (ChIP) analyses, which require antibodies with very high specificity, we performed ChIP-chip analyses on homozygously tagged *STAT3* and hemizygotously tagged *CHD7* loci. We performed ChIP using antibodies either to the Flag tag, or to native STAT3 or CHD7 proteins. Hybridizations were carried out on tiled microarrays that span the Encyclopedia of DNA Elements (ENCODE) regions³. A histogram displaying the intensity ratios of oligonucleotide probes after hybridization showed a skewing of the data in the positive direction, consistent with enrichment of DNA fragments captured by ChIP with Flag antibody (Fig. 3a and data not shown). We also plotted intensity ratios by their position along each chromosome. Representative examples of the STAT3 data are shown in Figure 3b and Supplementary Figure 3 online. We used a computer program incorporating a sliding window and threshold approach to identify genomic sites enriched for STAT3 binding at high confidence ($P < 1 \times 10^{-15}$)^{4,5}. Within the ENCODE regions, we identified 179 binding sites using Flag antibodies and 153 binding sites using STAT3 antibodies in cells expressing Flag-tagged STAT3, and 161 binding sites using STAT3 antibodies in wild-type DLD1 cells (Supplementary Table 2 online). Using conventional ChIP-PCR we verified 15 of 18 Flag-STAT3 binding sites (83%; Supplementary Fig. 4 online). Moreover, the relative amounts of enrichment for each site tested were similar between cells expressing Flag-tagged and wild-type STAT3 ($R^2 = 0.79$), indicating that the efficiency of ChIP was independent of the cell lines and antibodies used. Although we performed these studies on homozygous 3×Flag-tagged STAT3, ChIP-chip analysis of heterozygous 3×Flag-tagged CHD7 yielded similar amounts of enrichment, indicating that tagging only one copy of a given allele is sufficient for ChIP analyses (data not shown).

Binding profiles from Flag-STAT3 and wild-type STAT3 ChIP-chip experiments appear strikingly similar for the 0.5 Mb ENCODE regions shown in Figure 3b. To systematically determine the overlap of STAT3 binding sites for the remaining 29.5 Mb within the ENCODE regions, we selected all sites that were identified by ChIP-chip with antibodies to Flag or wild-type STAT3 ($n = 214$), and plotted the maximum mean signal intensity value for each site in a scatter plot (Fig. 3c) and heatmap (Fig. 3d). The plots revealed excellent correlation between sites identified using Flag antibodies and those identified using STAT3 antibodies, suggesting that the vast majority of binding sites identified between experiments overlap, and indicating the fidelity of ChIP results obtained using the epitope-tagging method. Some of the nonoverlapping sites could be due to differences in antibody sensitivity, subtle variations in growth conditions or experimental variability. However, we think that most of the variation is the result of threshold issues related to data processing and is not due to true false negatives. This is supported by both the heatmap (Fig. 3d) and by visual examination of the raw data (Supplementary Fig. 3). Regardless of the minor differences, the data suggest that the Flag antibodies are specific for STAT3, and by and large the presence of the tag does not alter the genomic distribution of STAT3.

Using the Clover algorithm⁶ and all motifs in the TRANSFAC database⁷, we tested ChIP-chip hits for enrichment of motifs that correspond to known transcription-factor binding sites. As expected, the STAT3 motif was significantly enriched in ChIP-chip hits from cells expressing either Flag-tagged or wild-type STAT3 ($P < 0.01$; data not shown). Notably, we also detected enrichment of AP-1 and HNF-3 motifs. Interactions between STAT3, c-Jun and HNF-3 have been reported previously^{8,9}, and the enrichment of these motifs may reflect cooperative binding to DNA.

The tagging approach offers several advantages over transgenic expression of recombinant proteins. First, as the epitope-tag sequences are knocked into endogenous loci by homologous recombination, transcriptional regulation by native promoters and enhancers is maintained. Second, the tagging method obviates the need for cloning tagged full-length cDNAs, which can be particularly challenging for large transcripts. Third, as the 3×Flag tag serves as a universal epitope for multiple applications, detection methods can be standardized. The targeted knock-in method is fast and inexpensive. Additionally, we have modified the vector so that construction of the targeting vector can be achieved in one step with very high efficiency (>80%) (Supplementary Fig. 5a–c online)¹⁰. Using the modified vector, an epitope-tagged protein can be generated in approximately one-half the time required for polyclonal antibody production, and this assumes that the resulting antibody would be suitable for ChIP (Supplementary Fig. 5d). In the future, as high-throughput gene targeting strategies evolve, it should be possible to target virtually every transcription factor within the same or different cell types. Combining this effort with genome-wide ChIP-chip or ChIP-Seq¹¹ analyses could facilitate the characterization of transcription factor–DNA interaction networks in mammalian cells, similar to studies that are now only feasible in yeast¹².

ACKNOWLEDGMENTS

We thank D. Sedwick for helpful discussions, J. Yu for technical assistance, and P. Harte and G. Crawford for critically reading this manuscript. This work was supported by grants from US National Institutes of Health (CA127590, U54CA116867), Concern Foundation and V foundation to Z. Wang, and National Institutes of Health grants KCA103843A and RHD056369A to P.C.S.

References

1. Zhang X, et al. *Proc. Natl. Acad. Sci. USA* 2007;104:4060–4064. [PubMed: 17360477]
2. Cherry SM, et al. *Curr. Biol* 2007;17:373–378. [PubMed: 17291760]
3. The ENCODE Project Consortium. *Science* 2004;306:636–640. [PubMed: 15499007]
4. Scacheri PC, Crawford GE, Davis S. *Methods Enzymol* 2006;411:270–282. [PubMed: 16939795]
5. Scacheri PC, et al. *PLoS Genet* 2006;2:e51. [PubMed: 16604156]
6. Frith MC, et al. *Nucleic Acids Res* 2004;32:1372–1381. [PubMed: 14988425]
7. Wingender E, Dietze P, Karas H, Knuppel R. *Nucleic Acids Res* 1996;24:238–241. [PubMed: 8594589]
8. Ginsberg M, et al. *Mol. Cell. Biol* 2007;27:6300–6308. [PubMed: 17636030]
9. Waris G, Siddiqui A. J. *Virology* 2002;76:2721–2729. [PubMed: 11861839]
10. Bitinaite J, et al. *Nucleic Acids Res* 2007;35:1992–2002. [PubMed: 17341463]
11. Johnson DS, Mortazavi A, Myers RM, Wold B. *Science* 2007;316:1497–1502. [PubMed: 17540862]
12. Lee TI, et al. *Science* 2002;298:799–804. [PubMed: 12399584]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Figure 1.

Schematic diagram of tagging endogenous protein with 3xFlag. **(a)** Targeting (*NEO-loxP*-3xFlag) vector. L-ITR and R-ITR, left and right inverted terminal repeats, respectively; MCS, multiple cloning site; CMV, cytomegalovirus promoter; *NEO*, neomycin resistance gene. **(b)** Diagram of knock-in strategy. rAAV targeting vectors contain a left and right 'arm' homologous to sequences in the target gene, flanking a *NEO-loxP*-3xFlag cassette. Clones are then screened by genomic PCR with primers complementary to the neomycin resistance gene and upstream of the left (P1 and NR) or downstream of the right (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. **(c)** Genomic PCR of parental (P) and *STAT3* 3xFlag knock-in cells (clones 1 and 2). WT, wild-type *STAT3*; Flag, Flag-tagged *STAT3*. Arrow indicates the targeted allele.

Figure 2. 3×Flag tagged proteins are detectable by western blot, immunoprecipitation and immunofluorescence. **(a)** Western blots of wild-type and 3×Flag-tagged STAT3 in DLD1 cells with either anti-STAT3 or anti-Flag. **(b)** Cell lysates of STAT3 3×Flag knock-in cells immunoprecipitated with anti-Flag and western blotted with anti-STAT3. **(c)** Parental and STAT3 3×Flag tagged cells were treated with or without IL-6 for 30 min and fixed. Immunofluorescence staining was performed with a rabbit anti-STAT3 and a mouse monoclonal anti-Flag. Note that similar to wild-type STAT3, Flag-tagged STAT3 translocates to the nucleus upon IL-6 stimulation. Scale bar, 10 μ m.

Figure 3.

ChIP analysis of wild-type and Flag-tagged STAT3. **(a)** Histogram of mean signal ratios of Flag-STAT3 chromatin-immunoprecipitated DNA versus random-sheared total genomic DNA. The distinct tail at the right-hand end corresponds to DNA fragments enriched by Flag-STAT3 ChIP (see insert). Tiled oligos that displayed the top 0.25% ratios are located to the right of the red bar. **(b)** STAT3 binding profiles from a 500-kb region on chromosome 21. Normalized raw ratio data from the indicated ChIP-chip experiments are plotted. The top 0.25% is displayed as a dotted horizontal line. An expanded view of a positive signal from the left is shown on the right. **(c)** The maximum signal intensity ratios for each STAT3-occupied site are plotted on the x and y axes and correspond to the filled circles ($n = 214$). For comparison, 15 randomly selected regions are plotted as open circles. ($R = 0.73$, $P < 2.2 \times 10^{-16}$). **(d)** Signal intensities of STAT3 bound regions are plotted as a heatmap. Note the high degree of overlap of STAT3-bound sites found in all three experiments. For comparison, 15 randomly selected sites were included (bracket on the right).