Epitope-tagged transcription factor ChIP-seq

Approved on Jan 2, 2017

Context and Overview

ChIP-seq can be performed by epitope tagging the transcription factor, co-factor or chromatin-associated protein of interest. Epitope-tag experiments are especially appropriate to discriminate multiple similar TF paralogs or isoforms, to map occupancy when no ChIP-reactive antibody is available for a factor, or when post-translational modifications disrupt recognition of the TF by existing antibodies. The following section describes how ENCODE has characterized epitope tag-specific reagents for ChIP-seq.

Each engineered cell population into which a tag has been introduced, is quality characterized for correct tagging and factor expression. Because of experimental design differences, the procedures and quality measures are not identical to those for a native TF/antibody pair.

For epitope-tagged ChIP, ENCODE characterization includes the following: i) evidence that the fusion protein is constructed and expressed in the engineered cell population as designed; ii) evidence that the antibody recognizes the tagged fusion protein; and iii) evidence that the antibody does not produce significant off-target background signals that would be confused with true occupancy signals.

As is the case for antibodies directed against individual TFs in classical ChIP-seq, the measurements used for quality assessment of a given epitope-tagged ChIP-seq experiment draws from a menu of several tests. We note that other measures and metrics might also be informative or could be used to substitute for the ones that ENCODE employs.

Experiment Definition

An ENCODE epitope-tagged TF experiment is a cell population engineered so that a specified epitope (e.g., GFP; FLAG-tag, HA-tag, etc.) is fused with the protein–coding sequence of the transcription factor or other chromatin-associated protein of interest and produces a functional chimeric protein containing the epitope. Ideally, the expression level of the tagged gene would be similar to that of the native factor. This can be achieved by using the gene's native promoter in the same genomic context as the rest of the gene and flanking regions. One way to achieve this is to add the tag segment to the gene in a BAC clone and then use that to transfer the modified gene into cultured cells. Alternatively, the endogenous gene in a cell line can be modified to contain the tag in its normal location in the genome; this is typically done by a version of CRISPR/Cas gene editing. In some cases, the tagged TF is put under the control of a different promoter complex, either inducible or constitutive, and the tagged gene is transfected into cultured cells. In these cases, the expression level of the transcription factor is not necessarily similar to that of the endogenous gene and this may affect the resulting ChIP-seq map.

Guidelines for ENCODE Epitope-tagged transcription factor ChIP-seq

ENCODE has used a variety of methods to characterize tagged TFs in ChIP-seq experiments. Typically, one form of the experiments listed under part A **and** one form of the experiments listed under part B is used for a given TF.

A. Genomic DNA characterization (A-1 or A-2 should be performed)

The experimental design relies on correct integration of the epitope tag sequence into genomic DNA of the recipient cell line. One of the following genomic characterizations is performed:

A-1. PCR analysis

PCR is used to verify the presence of the intended integrated sequence at the intended site of integration. PCR primers are designed such that the amplification product is generated only if the epitope tag is integrated correctly in the genomic DNA. In this design, one primer is selected to anneal outside the region used for the homology-directed repair (the mechanism used for integration), and one primer is located inside the tag sequence.

A-2. DNA sequencing of integrated tag segment

Genomic DNA is used to show epitope-tag integration at the designed target site. Sanger or next-generation DNA sequencing of genomic DNA showing correct integration of the tag sequence is performed for this determination.

B. Immunocharacterization (B-1 or B2 should be performed)

The epitope-tagged ChIP-seq experiment relies on a well-characterized antibody raised against the epitope tag. Immunological characterization of the antibody in each parental target cell population or type, prior to introduction of the tag, is performed. This characterization is used to detect any significant off-target ChIP signals due to cross-reactivity of the antibody with proteins other than the designed tagged protein.

The background control for immunocharacterization is the “wild type” cell line without a tag integration event. This control experiment is performed for each parental cell line.

Epitope tagged cell immunocharacterization is done by performing one of the methods below (B-1 or B-2).

B-1. Immunoblot (Western blot) or Immunoprecipitation blot (IP-Western)

It is preferred that the antibody used for the blots is the same one as used in the ChIP-seq experiment. However, it is recognized that antibodies differ in their ability to detect denatured and native proteins. Therefore, if necessary, another antibody raised against the epitope tag can be used for the western. A band (or bands) corresponding to predicted migration for the epitope tagged protein (or multiple forms, if they are predicted) should be visible when comparing the epitope-tagged cell line versus the “wild type” cell line. The parental “wild type” cell line, lacking the tag, is performed as a negative control in each western/IP-western experiment. Current acceptable parameters are that the major band is within 20% of the size predicted by the size of the coding region and corresponds to >50% of all bands on the gel (excluding the antibody bands in the case of an immunoprecipitation). If the western or IP-western results meet these criteria, we consider the antibody to meet expectations for the primary characterization. The immunoblot results (which must include appropriate size markers) are submitted as evidence for each cell type or tissue tested.

B-2. Immunoprecipitation followed by mass spectrometry

A cell or nuclear extract from cells expressing the tagged protein is immunoprecipitated with the same antibody used to perform ChIP-seq. These characterizations should be performed using the same lot number of antibody as used in the reported ChIP-seq experiments. That IP is then

fractionated on a denaturing polyacrylamide gel, and the fractions are prepared and analyzed by mass spec as described below.

What is reported for IP mass spec:

- IP-western blot of gel image with outline of gel slices submitted for mass spec. If the entire IP is used for the mass spec analysis, a western or IP-western image is not required, but would be informative additional information if available.
- All peptides (with peptide counts) from all immunoreactive bands.
- Fold enrichment of all peptides in the immunoreactive bands vs either mock IP or a set of proteins that have been immunoprecipitated from the same cell type using a collection of other antibodies from the same host species (the list of proteins used as the set of IP contaminants list must be provided).
- Indication as to which proteins above the target protein on the ranked list (ranked by fold enrichment) are TFs and which TFs are members of the same TF family as the target protein.

IP mass spec requirements to be considered fully validated for ENCODE data:

- The target protein should be enriched in the IP when compared to a mock IP or to a set of proteins that have been immunoprecipitated from the same cell type using a collection of other antibodies from the same host species.
- The target should be in the top 25 ranked proteins and the top most-enriched TF (by fold enrichment) in the immunoreactive band, unless the higher ranked TFs are known interacting partners of the target TF and/or a known interacting partner of one of the other higher ranked TFs that is a known partner of the target TF. Evidence for interaction can come from publications or refer to records in interaction databases such as BioGRID, or other sources.
- The target should be the top ranked member of that family of TFs (exceptions will be allowed if a publication is provided that demonstrates that a higher ranked family member is known to dimerize with the target protein).
- In situations for which the target protein has 0 peptides in the mock IP, a ranking by enrichment cannot be performed. In that case, the following criteria are used:
  - If the target TF is the top TF as ranked by number of detected peptides, then the antibody passes this characterization method.
  - If the target TF is not the top TF but the TFs having more counts have previously been documented to be in the same complex and/or interact directly with the target TF, then the antibody passes this characterization method.
  - If the target TF is not the top TF but the non-target TFs (having a greater number of detected peptides) were detected using mass spec analysis of two different antibodies to the target TF, then the antibody passes this characterization method (with the assumption that the other TFs are bona fide interacting TFs that have not yet been documented in the literature).
  - If the target TF is not the top TF and the TFs having more counts have never been linked to the target TF then this antibody is flagged, with the explanation that enrichment could not be determined due to the lack of detected peptides in the IgG and that no published data exists linking the target to the non-target TFs.

Final characterization report for tagged TF ChIP-seq

The following report is prepared for each epitope-tag ChIP-seq experiment:

1. The identity of the epitope tag used and the target TF, the general location of the tag on the target (e.g., amino terminal, carboxy terminal or internal in the protein), and the cell line in which the tag was inserted.

2. Method of epitope-tag integration. Indicate specifically whether the tagging experiment involved single clone isolation or a polyclonal pool.

3. The identity of the antibody used, including company, catalog number and lot number.

4. Genomic characterization: Gel image (or equivalent) of PCR reaction, showing expected size band, along with PCR primer sequences; or sequencing data with explanation of integration region of genomic DNA.

5. Immunocharacterization: Western or IP-western image, showing both the control ("wild type") cell line and the epitope-tagged cell line; or mass spec report showing fold enrichment in epitope-tagged cell line compared to "wild type" cell line.

6. Result of peak-calling using the specific antibody in this experiment as the ChIP antibody in the parental, non-tagged cell line. This experiment should be performed for every unique parental cell line/antibody/antibody lot combination.

EXCEPTIONS

We realize that, in rare cases, situations may arise in which antibodies or tagged factor lines do not pass the above standards, but the data producers feel that the datasets should be made available to users. Therefore, exceptions to these characterization standards are considered for special cases. The antibody characterization review committee of the DCC, along with the Production PIs, will consider each special request. If an exception is granted, the datasets using these "exempt" antibodies will be flagged in the ENCODE datasets.