Antibody and Genomic Characterization Guidelines for Epitope-tagged Transcription Factor ChIP-seq Experiments

Version 2, Approved on August 21, 2019

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. For guidelines on the validation of ChIP-seq experiments that involved the use of TF-specific antibodies (previous Phases of ENCODE), we refer the reader to the Antibody Characterization Standards for Transcription Factors document https://www.encodeproject.org/documents/c7cb0632-7e5f-455e-9119-46a54f160711/@@download/attachment/ENCODE_Approved_May_2016_TF_Antibody%20Characterization_Guidelines.pdf. In current epitope-tagged TF experiments, each engineered cell population for 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 pairs, and therefore, concerning epitope tagged datasets, these guidelines supersede the Antibody Characterization Standards of May 2016. 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-tag; FLAG-tag, HA-tag, etc.) is fused, usually by a linker peptide, with the protein–coding sequence of the transcription factor or other chromatin-associated protein of interest and produces a 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. In ENCODE, the most commonly used way to to achieve this is to modify the endogenous gene in a cell line to contain the tag in its normal location in the genome; this is typically done by a version of CRISPR/Cas gene editing. Alternatively, the tag segment can be added to the gene in a BAC clone and then used to transfer the modified gene into cultured cells. 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 uses a variety of methods to characterize tagged TFs in ChIP-seq experiments, and these methods are categorized as being either genomic characterizations (to ensure the correct locus of interest was tagged properly), or immunological characterizations (to ensure the antibody recognizes the epitope-tagged protein) Typically, one form of the experiments listed under part A (Genomic DNA Characterization) and one form of the experiments listed under part B (Immunocharacterization) 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.

What is reported for Genomic Characterizations:

A gel image of the PCR reaction products with a DNA sizing ladder. A negative control sample (amplification from wild-type DNA) should be included if available. The expected size should be indicated, along with the PCR primer sequences and thermocycling conditions used to generate the products. For sequencing data, an electropherogram (Sanger sequence trace) or genome browser screenshot with an indication of the integration region within the wild-type genomic DNA.

Genomic Characterization requirements to be considered fully validated for ENCODE data:

Ideally for PCR and sequencing data, results from both replicates should be represented. If, however, only one replicate is present or passes genomic validation, then a passing grade can be assigned if both replicates passed IDR from ChIP-seq.

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. 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 blot)

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 blot. 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 background control for immunocharacterization is the “wild type” cell line without a tag integration event. This control experiment is performed at least once for each parental cell line that is used.

What is reported for Immunoblot or Immunoprecipitation blot (IP-Western blot):

An image of the blot/gel showing affinity of the antibody for the epitope-tagged protein from either cell lysates (immunoblot) or immunoprecipitated proteins from cell lysates (immunoprecipitation blot). A protein sizing ladder should be included as well as a description of the blotting method and conditions for immunostaining. For immunoprecipitation blots, the antibodies used for both immunoprecipitation and visualization should be indicated. The

expected size of the tagged target protein should be indicated as well as other bands that might correspond to either lower size degradation products or putative post-translational modifications.

Immunoblot or Immunoprecipitation blot requirements to be considered fully validated for ENCODE data:

The protein band of interest must be within 20% of the size predicted by the coding region. If the Western blot or IP-Western blot result meets this criteria, we consider the engineered cell line to meet expectations. If however, protein sizes do not match expected sizes which include the tag, then Western blots with native antibodies from commercial vendors can be used for compliance if the sizes are equivalent. Protein modifications and degradation products are known to complicate the sizing and intensity of bands, therefore, all instances must be thoroughly explained in the corresponding captions so that users of the data are made aware.

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. The IP product 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:

An IP-Western blot gel image with an outline of gel slices that were submitted for mass spec should be reported. If, however, the entire IP was used for the mass spec analysis, a Western blot or IP-Western blot image is not required. A list of all peptides (with peptide counts) from all immunoreactive bands should be presented in tabular format. Fold enrichment of all the 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) should also be determined.

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

The target protein should be enriched within the top 20 ranked proteins 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. Ideally, the target TF would represent the highest ranking TF within this enrichment. If it is not however, then the production lab should indicate potential complexes or interacting partners (if known) that have co-immunoprecipitated with their target TF or provide an appropriate audit if the ChIP-seq data is deemed of high quality. In situations involving mock IPs for which the target protein has 0 peptides in the mock IP, a ranking by enrichment cannot be performed. In this case, the following criteria are considered for validation: the target TF is the top TF present as ranked by the number of detected peptides or, the target TF is not the top TF ranked by peptide counts but

is documented to be in a complex or have interactions with the other TFs having more counts. For situations where the target TF is not the top TF **and** there are no documented instances of interactions with other TFs having more counts, then an audit is assigned with the explanation that the enrichment could not be determined due to the lack of detected peptides in the IgG control and that no published data exists linking the target to the non-target TFs.

EXCEPTIONS

We realize that, in some 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. Often there is data from other sources that support a ChIP-seq dataset that has not passed both A and B standards. Examples include the same epitope tagging reagents having passed in another cell type, or a high overlap of peaks to an antibody based dataset in the same cell type, or a highly similar motif found to one previously published for that factor. Therefore, exceptions to these characterization standards are considered for special cases. The antibody characterization review committee of the ENCODE Consortium will consider each special request. If an exception is granted, the datasets using these “exempt” antibodies will be flagged in the ENCODE datasets.