ENCODE TF Antibody Characterization Guidelines

Approved on May 20, 2016

Background

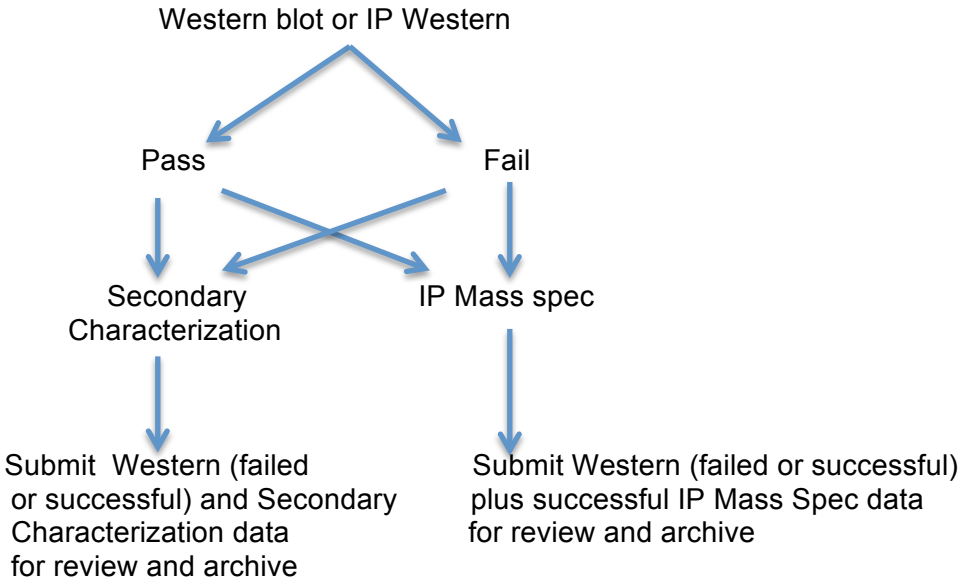
An integral part of the ENCODE Project is to characterize the antibodies used in the ChIP-seq experiments. This document describes the guidelines used by the ENCODE Consortium for characterizing antibodies used in **Transcription Factor (TF) ChIP-seq** experiments. We expect that these antibody characterization data will be pertinent to users of ENCODE ChIP-seq data as well as to future users of these antibody reagents.

As detailed below, the antibodies that pass our working thresholds for Primary and, as appropriate, a supporting Secondary Characterization method, are believed by ENCODE data producers to recognize the target factor and to be active in chromatin immunoprecipitation in the context of the cell type(s) tested.

The antibody characterization data themselves provide users with richer and more nuanced information than can be captured by pass/fail threshold parameters alone. Therefore, all antibody characterization data are available at the ENCODE data portal. As part of the ENCODE ChIP-seq data finalization process, the antibody characterization data are reviewed by a panel of ENCODE investigators and analysts, and successfully validated antibodies are designated as such. In some instances the characterization data fall short of current ENCODE thresholds or assay specifications, yet the antibody has generated ChIP-seq data that we judge potentially useful to the community. We flag such antibodies and corresponding ChIP-seq data as having a quality concern, and the nature of the characterization deficiency is reported. Because antibody reagents can vary considerably from one specific lot number to another, each lot is characterized independently.

It is important to recognize that our necessarily arbitrary thresholds might ultimately prove to have been either too strict or too relaxed for particular antibody/factor pairs. There may also be confounding differences between antibody-epitope recognition in fixed (ChIP experiment) versus the denatured (Western blot) or native (mass spec) states that are required by the Primary Characterization methods. *We therefore recommend that users review our primary antibody characterization data in light of their own expert knowledge, any additional data, or their own specific requirements.*

I-A. Characterization flowchart:



Primary characterization by Western blot or Immunoprecipitation/blot

For every TF ChIP-seq antibody, ENCODE data producers first perform an immunoblot characterization. This can be either a standard Western blot, or an immunoprecipitation followed by a Western blot ("IP Western"). If the blot results do not meet the parameters and thresholds given below, then Primary Characterization Method 2 (IP mass spec) is performed. In the latter case, the failed (or partially failed) immunoblot that preceded IP/mass spec is included in the report so that researchers and data users can independently evaluate the data for each antibody.

Immunoblot parameters:

a. ENCODE developed a set of working parameters and thresholds to identify antibodies with a high likelihood of being specific for the target factor. The parameters allow for modest variation in gel migration characteristics and in band number to accommodate known behaviors of typical nuclear factors. 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. For IP-westerns, a control IgG precipitation is also performed and analyzed on the same gel.

b. If the antibody fails to pass the immunoblot tests because the bands observed are too numerous, or too far from the predicted migration behavior, it can be “rescued” by a secondary characterization that supports the conclusion that the band(s) detected correspond to the correct protein (e.g. all bands are reduced by treatment with a specific siRNA to that protein; see secondary characterizations).

c. If the antibody passes the immunoblot tests, a further characterization is required to support the successful immunoblot. This can be Primary Characterization Method 2 (IP mass spec) or any one of the Secondary Characterization methods in IB.

Primary Characterization Method 2: Immunoprecipitation followed by mass spectrometry ("IP mass spec")

If the immunoblot characterization data was not successful (ranging from no bands to patterns that do not meet the thresholds given above), then Mass spec of an immunoprecipitation can be performed. The failed or ambiguous immunoblot is, however, shown as part of the antibody characterization dataset. Because the IP/mass spec assay provides explicit evidence about the identity of the TF detected, it can also be used in lieu of Secondary methods after a successful Immunoblot (see flowchart above).

For TF mass spec, a cell or nuclear extract is immunoprecipitated with the same antibody used to perform ChIP-seq. 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.
- 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 can not 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.

- If an antibody doesn't meet these characteristics, the antibody characterization document can be submitted for consideration as a special request (see Note 3).

Additional situations for Primary Characterization

a. Guidelines for using the same lot number of a previously characterized antibody in a new cell type

If a specific lot number for an antibody has previously passed characterization in another cell type, and if the banding pattern on the immunoblot or immunoprecipitation is the same in the new cell type as in the characterized cell type, then no further characterization is needed for the antibody in that new cell type. If the banding pattern is different in the new cell type, a secondary characterization is performed in the new cell type. Exceptions to this guideline will be considered for studies of human tissues (due to the fact that it is often hard to obtain sufficient tissue for the antibody characterization and a ChIP-seq experiment). If an antibody has passed characterization criteria in 2 different human cell lines and/or tissues, it does not have to be characterized in each tissue type.

b. Guidelines for using a different lot number of a previously characterized antibody

For the first time that a new lot number is used for a previously-characterized antibody, a Primary Characterization method (immunoblot or IP mass spec) is performed with one of the same cell types used to characterize the previous lot number plus the cell type for which ChIP-seq data will be deposited for the new lot number; the ENCODE antibody accession number of the specific previously characterized lot that should be used for comparison is indicated. If the patterns for the new lot number are the same in the previously characterized cell type and in the cell type for which ChIP-seq data will be deposited as shown in the characterization of the original lot number of that antibody, then no further characterization is required. If the banding patterns are different, a secondary characterization is performed. Exceptions to this guideline will be considered for analysis of tissues with antibodies that have been well-characterized and used extensively by the field (e.g a monoclonal antibody to RNAPII). In this case, if a previous lot number of an antibody has passed characterization criteria in 2 different human cell lines and/or tissues, the new lot number does not have to be characterized in each tissue type.

Other primary characterization methods. If other methods not specified above are used for primary characterization of an antibody, the antibody characterization document is submitted as a special request and is so annotated and flagged.

IB. Secondary Characterization Methods. These methods are used to support and clarify the Immunoblot data. In particular, they aim to verify that a band or bands observed on the prior blot correspond to the intended TF. At least one successful Secondary Characterization (or alternatively IP/Mass spec as shown in figure 1 above) is required to support a successful Western or IP/Western.

Secondary Characterization Method 1: siRNA or shRNA against the mRNA of the target protein

For siRNA or shRNA knockdown characterization, the band(s) detected by the antibody on a western blot should be reduced by at least 50% of the control signal. These methods are especially intended to address instances where the Western or IP Western data give multiple bands and unpredicted migration patterns. The sequence or vendor and catalog number of the oligonucleotide(s) reagent should be provided. A control knockdown should also be performed. Cell types will be labeled and size markers should be included on the immunoblot. A brief description of the transfection protocol will also be provided.

Secondary Characterization Method 2: ChIP-seq data from a previously characterized antibody

If ChIP-seq data for a different lot number of a previously characterized antibody or a previously characterized, but different, antibody for a given transcriptional regulator is available, this ChIP-seq data can be used to evaluate a new antibody or new lot number. The ChIP-seq data from the new antibody or new lot number are compared to the previous ChIP-seq using IDR. If the two datasets pass the ENCODE IDR cuts-offs for narrow peak ChIP-seq reproducibility (see below for current IDR standards), then the secondary characterization of the new antibody/lot number is scored as successful. For data submission, the specific antibody lot and ChIP-seq data used for the comparison are identified by their ENCODE antibody and experiment accessions, respectively. In a similar way, ChIP-seq data obtained using an endogenous epitope-tagged version of the target protein can be used for comparison.

Secondary Characterization Method 3: Expression patterns of an epitope-tagged transcription factor

Especially useful for TFs that are resistant to knockdown using shRNA or siRNAs (e.g. very stable proteins) is a secondary characterization method that involves comparison to overexpressed or endogenously epitope-tagged TF proteins. In this case, the primary characterization of the TF antibody must first show the appropriate specificity on the western or IP-western. Then, two side-by-side immunoblots can be performed using control cells and cells expressing the tagged-factor. The first immunoblot employs the antibody to the tag to show the position of the exogenous factor band(s) and the second immunoblot employs the antibody to the endogenous factor to show that the band(s) in the control and ectopically expressing cases correspond.

Secondary Characterization Method 4: Motif analysis

Motif enrichment for antibody characterization requires pre-existing information about the DNA sequence to which the factor binds. Enrichment of a known motif for a target TF in a ChIP experiment is evidence that the antibody does in fact recognize the target TF.

Motif enrichment can be used as a validation method for antibodies that meet the following criteria:

i. The antibody under consideration binds a sequence-specific transcription factor

ii. The DNA motif sequence bound by the transcription factor has been previously well-characterized by either in vitro or in vivo experiments

iii. The antibody is raised to a unique region of the transcription factor (in relation to other TFs in the same family)

Motif analysis can be performed using high-quality peaks (0.01 IDR cut off) from the ChIP experiment. Proper use of motif enrichment analysis for antibody validation should include metrics indicative of:

i. Global Enrichment z-score: Enrichment of the motif sequence in the ChIP peak over shuffled randomized motifs of the same sequence composition

ii. Positional Bias z-score: A measure of the distance of the motif to the peak center

iii. Peak Rank Bias z-score: A measure of the distribution of the motif in peaks ranked by ChIP intensity

The mean of these three z-scores is used in computing the final enrichment rank among 282 motif groups, as well as the “accept probability”. The “accept probability” is a combined metric that measures confidence in the antibody under investigation being of high quality for ChIP experiments. An accept probability greater that 0.6 is the current criteria for accepting an antibody as passing secondary characterization by motif enrichment (see note 4)..

The Characterization report where Motif enrichment is used for antibody secondary validations includes:

i. The ENCODE DCC file identifiers for the peaks files (.bed files) used in the analysis

ii. A brief description of the analysis method and a reference to the standards documents

iii. The accept probability score from the motif analysis pipeline

iv. The identified motif (PWM) and its enrichment rank

v. The positional bias score as well as the peak rank score

Motif analysis cannot be used when:

i. The transcription factor does not bind in a sequence-specific manner

ii. There is no information for the DNA motif bound by the TF

iii. When it has been shown that the TF bind to DNA indirectly by interacting with other proteins that directly bind DNA

Because transcription factors are recruited by multiple mechanisms, failure of a data set to meet the motif enrichment criteria does not indicate poor antibody quality or poor data quality. Such antibodies can be validated using other Secondary Characterization methods.

Additional notes on methods for antibody characterization:

1. These methods refer to characterization of antibodies that recognize endogenously expressed proteins. The requirements for characterization of antibodies that recognize epitope-tagged proteins are described elsewhere.

2. Current IDR standards for a narrow-peak ChIP-seq dataset are: Rescue Ratio RR_new = | Np U Nt | / | Np ^ Nt | Self consistency ratio SR_new = | N1 U N2 | / | N1 ^ N2 | where ^ = intersection (common) of 2 peak sets U = union (merge) of 2 peaks sets

If (R_new > 2) AND (SR_new > 2) then the replicates are proclaimed to have low reproducibility (failed) and flagged with -1 quality score If (RR_new > 2) OR (SR_new > 2) but not both, then the replicates are proclaimed to have moderate reproducibility (passed) and flagged with a 0 quality score If (RR_new <= 2) AND (SR_new <= 2) then the replicates are proclaimed to have high reproducibility (passed) and flagged with +1 quality score.

3. Scientists within and outside ENCODE have learned over time that some antibodies that perform well in ChIP assays nevertheless fail to pass the conventional tests that comprise primary and secondary analyses. Therefore, exceptions to the basic characterization can be considered for such cases. The antibody characterization review committee together with the ENCODE Production PIs, will consider these on a case by case basis. Datasets using such reagents, referred to as “exempt” antibodies, will be flagged in the ENCODE data.