ENCODE Antibody Characterization Guidelines for Histone Variants, Histone Modifications and Chromatin-associated Proteins

Approved on October 17, 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 that recognize histone variants, histone modifications and chromatin-associated proteins. 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 histone variant, histone modification or factor and to be active 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.*

Characterization of Antibodies to Histone Modifications and Chromatin-associated Proteins

For histone variants, modified histones, and associated chromatin proteins, each specific lot number of each antibody is characterized and reviewed before use. A primary and secondary characterization is submitted for review and deemed compliant to the current ENCODE standards, in order for the antibody lot to be considered characterized for use in new binding assays. Prior lot-specific review of this kind is recommended to any future users of these reagents, recognizing that significant lot-to-lot differences in performance have been observed.

1. Primary characterization. In all cases, a primary characterization of each lot number of an antibody for a modified histone, a histone variant, or a chromatin modifier is performed. For modified histones, the primary characterization of a histone antibody should include an immunoblot using nuclear or whole cell extracts from the species in which the antibody will be used for ChIP-seq. The specific histone band detected in the extracts should constitute at least 50% of the protein signal and show at least 10-fold enrichment relative to any other single band. In addition, this signal should be at least 10-fold enriched relative to that detected using the unmodified recombinant histone (if this is included on the western blot). Although it is preferable that a recombinant unmodified histone be included on the western blot, if unmodified relevant peptides are included in the dot blots, peptide arrays or peptide competitions that are used in the secondary characterization step, then it is not required that recombinant histones be included in the primary characterization step. Each immunoblot should label the cells and include size markers. For histone variants, detection by western blot under criteria described above constitutes primary validation. For chromatin-associated proteins, the same primary characterization methods as used for TFs (western or IP-western) should be used (see TF Antibody Characterization guidelines).

2. Use of the same lot number of a previously characterized antibody in a new cell type

Once a specific lot number for an antibody to a modified histone or a histone variant has passed both the primary and secondary characterizations, it is eligible for use in all cell types of that same species. However, for chromatin-associated proteins, immunoblot or immunoprecipitation is used to validate the specificity of the antibody in each new cell type in which it will be used.

3. Use of a different lot number of a previously characterized antibody

If this is the first time that a specific lot number has been used for a previously characterized antibody to a modified histone, a histone variant, or a chromatin-associated protein (same catalog number but different lot number), both a primary and a secondary characterization must still be performed for this lot number.

Secondary Characterization Methods for histone variants, modified histones, and chromatin-associated proteins:

1. For modified histones, a secondary characterization using dot blots, peptide arrays or peptide competitions can be used. Commercial arrays are acceptable and the vendor and catalog number should be provided. If arrays that include small numbers of peptides are used, it is critical that the array contain the most relevant peptides for the antibody being tested. If peptide competition is used to validate a histone antibody, it is also important that the most relevant peptides for the antibody being tested are used in the competition series. A 10-fold enriched binding signal for the modification of interest relative to other modifications is ideal. An annotated map of the peptides present on the blot or array should be provided.

2. For modified histones, histone variants, and chromatin-associated proteins, one can use the similarity of ChIP-seq data from a different antibody that recognizes the same modified histone, histone variant, or chromatin-associated protein as secondary validation; see Notes for details. For chromatin-associated proteins, one can also use ChIP-seq data for another factor in the same protein complex, or for a histone modification that the factor catalyzes or with which it has an established physical interaction (PMID numbers should be provided for publications supporting the interactions or activities). Correspondence between the datasets is criteria for secondary validation. When submitting secondary antibody validation documents, the specific antibody lot and ChIP-seq data used for the comparison, and the evidence for correspondence

of datasets are specified with the ENCODE antibody and experiment accessions. Note that the antibody to which the new antibody is being compared need not have already been deemed characterized to the standards current at the time of its review, since the high correlation between two antibodies raised against different proteins from the same multi-protein complex constitutes secondary validation of both antibodies**.**

3. For histone variants, demonstration by western blot that the antibody is specific to the variant form of the histone, and does not recognize recombinant forms of the corresponding canonical histone protein, can be used as a secondary validation.

4. For **chromatin-associated proteins**, the same secondary characterization methods as used for TFs (western or IP-western) can also be used (see TF Antibody Characterization guidelines).

Notes:

1. At the present, the only approved IDR standards for ChIP-seq datasets are for narrow-peak datasets. The 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 (RR_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

2. Another method for comparing ChIP-seq datasets for broad marks of chromatin-associated proteins is to use a Pearson Correlation. In the work performed by the Bernstein lab at the Broad Institute under Encode3, metrics were derived as follows: (1) *Genomic windows for validation test:* We collated a set of ~15,000 3KB genomic windows whose ChIP-seq signals vary between histone marks and cell types, based on a manually curated set of ~1000 ChIP-seq experiments (data from the Encode2 and NIH Roadmap Epigenomics projects [https://www.encodeproject.org/matrix/?type=Experiment&lab.title=Bradley+Bernstein%2C+Broad&assay_title=ChIP-seq&award.rfa=Roadmap&award.rfa=ENCODE2]). Note that we excluded ~1000 windows with false-positive constitutive signals across cell types and/or marks. The genomic coordinates of these ~15,000 3KB windows are available here: https://docs.google.com/spreadsheets/d/1LaCo-KjnOP6dLoLLPmTYcR2V9Ht_N-Uu9MzPSjX6YqE/edit?usp=sharing. (2) *Antibody validation test:* For ChIP-seq experiment, numerical values corresponding to normalized read density for each window were calculated. Pearson correlations between any/all pairs of datasets across the 15,000 windows were then computed. For secondary validations of chromatin-associated proteins, genome-wide correlations between documented members of multi-protein complexes should be higher than 0.8; references documenting the presence of the proteins in the same complex must be provided. Cases with correlations that are significant but do not meet this threshold are flagged in the secondary validation documentation.

Examples of a secondary validation using comparison of a) two antibodies that recognize the same protein and b) antibodies that recognize two members of the same complex are provided below.

Broad Institute - Encode3 Secondary Antibody Validation

Millipore # 05-1317 Lot NRG1665726 (mouse monoclonal)

Target: SUZ12

Approved name: SUZ12 polycomb repressive complex 2 subunit

Function: plays a role in repressive Histone H3 K27 methylation

Member of complex: PRC2

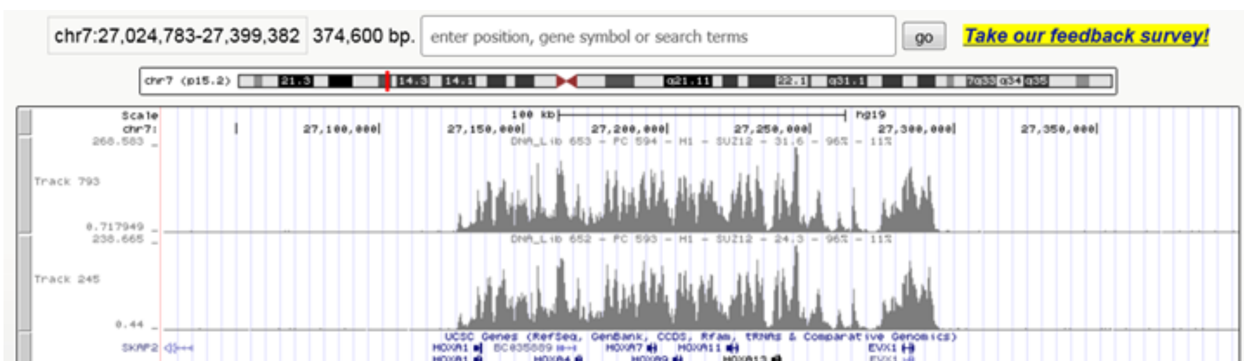
Other complex members: EZH2, EED PMID: 12435631, PMID: 12408864

Antibody being validated:

1. Millipore # 05-1317 Lot NRG1665726 (mouse monoclonal)
2. Broad Alias: PchAb 168
3. Immunogen: Recombinant protein corresponding to human SUZ12.
4. https://www.encodeproject.org/antibodies/ENCAB032FOO/

This validation relies on the use of **two different antibodies to the same protein**, and the demonstration that highly similar patterns of enrichment are obtained with each antibody. The first track shown used the antibody being validated (mouse monoclonal Millipore 05-1317, raised against full length SUZ12), and the second track shown used a different antibody to SUZ12 (Active Motif 39357), a rabbit polyclonal raised against a C-terminal peptide (https://www.encodeproject.org/antibodies/ENCAB823XVS/). The high correlation between two antibodies raised against different peptides from the same protein constitutes secondary validation of both antibodies.

Overall correlation score: 0.9956.



Broad Institute - Encode3 Secondary Antibody Validation
Active Motif # 39639 Lot 23809001
Target: EZH2

Approved name: enhancer of zeste 2 polycomb repressive complex 2 subunit
Function: plays a role in repressive Histone H3 K27 methylation. Catalytic subunit.
Member of complex: PRC2
Other complex members: SUZ12, EED PMID: 12435631, PMID: 12408864
Antibody being validated:
1. Active Motif # 39639 Lot 23809001 (rabbit polyclonal)
2. Broad Alias: PchAb 58-V.
3. Immunogen: a recombinant fusion protein corresponding to amino acids 1-370 of mouse EZH2
4. https://www.encodeproject.org/antibodies/ENCAB000AYA/

This validation relies on the use of antibodies to different members of a known complex, and the demonstration that highly similar patterns of enrichment are obtained with each antibody. The first track shown used the antibody being validated whereas the second track shown employed an antibody (https://www.encodeproject.org/antibodies/ENCAB823XVS/) raised against SUZ12, another member of the PRC2 complex. The high correlation between two antibodies raised against different proteins from the same multi-protein complex constitutes secondary validation of both antibodies.

Overall correlation score: 0.9833