

**ENCODE and modENCODE**  
**Guidelines For Experiments Generating Data using RNA-Binding Proteins (RBPs):**  
**ENCODE and modENCODE Standards for RIP-Chip and RIP-Seq Experiments**  
**Version 2.0, 9 January 2012**

**Introduction:** In eukaryotic organisms, gene regulatory networks require an additional level of coordination that links transcriptional and post-transcriptional processes. Messenger RNAs have traditionally been viewed as passive molecules in the pathway from transcription to translation. However, it is now clear that RNA-binding proteins (RBPs) play a major role in regulating multiple mRNAs in order to facilitate gene expression patterns. Several ENCODE/modENCODE researchers are mapping RBP-binding sites using RBP immunoprecipitation (RIP) followed by extraction of associated RNA and characterization using genomic microarrays (RIP-Chip), ENCODE Tiled-array (RIP-T-Chip) or by high throughput sequencing (RIP-Seq).

In an effort to facilitate and unify high-quality RBP-based ENCODE/modENCODE data sets, this document aims to provide uniform standards and guidelines for experiments that map RBP-associated RNA networks and the location of RBP-associated binding sites and regulatory elements. These are proposed standards by which to monitor the quality, reproducibility, specificity, sensitivity and reporting of RIP experiments. We realize that there is no single path to validating and verifying an experiment, and that these standards may require future revision, as technologies change and our understanding of the data increases. We also recognize that there is an inherent trade-off between reporting a complete list of RNA-based functional elements, and a high-confidence list of RNA-based functional elements.

## **I. General Required Measurements and Procedures**

- Growth and maintenance of cells. The protocol for growing and maintaining cells should follow the general outline described in the *Guidelines For Experiments Generating Chromatin and Transcription Factor Data, Version 1.0*. Briefly, for each experiment, the growth time of cell lines should be determined by recording the date and time at which cells were put into culture, and when they were harvested. Investigators should go back to the original stock after growing a culture for one month. Passage number should be assessed and recorded for primary cells. Primary cells should not exceed 4 passages. Any experiment that does not follow the officially approved protocol for that cell line/type should be noted (see metadata standards).

## **II. Antibody Characterization**

- Presently, there are a limited number of well-characterized antibodies against RBPs. To ensure specificity of the antibodies used, the following tests are suggested. The data generated to characterize the antibody should be made publically available alongside the ENCODE dataset and must be registered with the DCC prior to submission of the data.
- It is expected that as the number of antibodies increase in the near future, the use of multiple antibodies against the same protein will become the standard for the field.
- Multiple antibodies, when available, should be used for RIP experiments, and a statistically significant overlap of targets will be used for characterization. Any reasonable method of correlation can be used (such as  $r^2$  greater or equal to 0.5 or 80% of the top 40% of the targets of one list should overlap that of the list from the second antibody).

- For antibodies directed against members of a multi-gene family, antibodies should be directed to protein regions that are unique to individual family members. Any potential cross-reaction should be noted when reporting data collected using that antibody.
- Different Lots of antibody from the same manufacturer do not need to be characterized if they recapitulate previous RIP results to the standards established for biological replicates.
- RIP using epitope-tagged proteins - As there are few antibodies available to RNA binding proteins, RIP experiments may also be conducted in cell lines expressing epitope-tagged RNA binding proteins. In these cases, the antibody that recognizes the epitope-tag should be characterized in the same manner as antibodies that directly recognize RNA binding proteins. Western blotting and/or immunofluorescence should be used to monitor the expression level of the epitope-tagged RNA binding protein. Additionally, Western blotting should monitor the efficiency and specificity of the RIP.
- Immunoblot Analysis. Successful antibodies should show essentially a single band of the predicted size by immunoblot analysis of RIP material. The minimum standard to be met is that the predicted reactive band composes at least half of the total signal in the lane, as assessed by quantitative immunoblot analysis of RIP-material from whole cells lysate (e.g. quantitative imaging of chemiluminescence or chemifluorescence).
- Immunofluorescence (IF) Analysis. If immunoblots are unsuccessful, immunofluorescence may be used as a characterization measure. The immunofluorescence pattern must conform to expectations (for example, nuclear staining for a chromatin protein or TF).
- To be considered “RIP” quality, an antibody RIP-IP Western blot activity should be reactive against preferably one prominent band of the appropriate molecular weight. Although multiple bands due to isoform, phosphorylation or glycosylation differences are to be expected. With the exception of the immunoglobulin heavy and light chain, no other activity should be present and the negative control Ab or beads alone IP should be negative. If other bands are present, the appropriate target band should be at least 2-times the intensity of any other band.
- As discussed in the March mod/ENCODE Consortia meeting, documents describing characterization of antibodies to RBPs, and epitope-tagged RBPs will be required for complete data submission. Effective March 1, 2011, 1) New antibodies submitted for registration to the DCC must include a complete antibody characterization PDF document (see below template) and 2) New data submissions that involve antibodies must include the registered antibody information or they will not be accepted. For more information regarding the antibody characterization policy, including a copy of the characterization form and a proposed workflow for registering future antibodies, please visit the ENCODE Antibodies wiki page: <http://genomewiki.ucsc.edu/EncodeDCC/index.php/Antibodies>

# ENCODE DCC Antibody Validation Document

Date of Submission:

Name:  Email:

Lab:

Antibody Name:  Target:

Company/  
Source:

Catalog Number, (database ID), laboratory:  Lot Number:

Antibody Description:

Target Description:

Species/Target:  Species Host:

Validation Method #1:  Validation Method #2:

Purification Method:  Polyclonal/  
Monoclonal:

Vendor URL:

Reference (if  
Publication  
Information):

Please complete the following for a ribbed or to list any modifications:  
If your peptide sequence is not listed in the shop database,  
please enter in the appropriate information.

Peptide Name:  AA modified:  AA Position:  Modification:

Validation #1  
Analysis

Validation #2  
Analysis

Submit by Email

### **RIP-chip and RIP-Seq Data Production**

- As outlined in the mod/ENCODE Data Verification Standards, RIP-Chip, RIP-T-Chip and RIP-Seq data should be verified to assess the reproducibility of the experiment by performing at least 2 biological replicates. Three biological replicates are recommended for RIP-Chip and RIP-T-Chip and two for RIP-Seq. The replicas should have a significant correlation coefficient when analyzing the top targets on the list. 80% of the top 40% of the targets of one list should overlap that of the second list. Once data have been verified, they should be released into public databases as consistent with the ENCODE/modENCODE Data Release Policy.

### **RIP Validation by use of an alternate detection platform**

- Similar observation using alternative detection methods increases confidence that detected events are reproducible. However alternate methods of detection do not test the biological validity of the results and may not assess the accuracy of the underlying biological assay. Small-scale validation can be performed by subjecting individual transcripts of interest to Northern blot following RIP, to demonstrate specificity with respect to other control transcripts not associated with the factor of interest. Northern blotting is also advantageous with respect to sequencing and microarray technologies, as it reports directly on the size(s) of the associated transcripts.

### **Specificity and Sensitivity**

- Specificity and sensitivity may be difficult to accurately estimate for RIP experiments since the number of true positives and negatives is frequently unknown and methods for identification of *in vivo* binding independent of RIP are difficult to ascertain. Nonetheless, the positive predictive value (defined as  $TP/(TP+FP)$  where TP is true positives and FP is false positive) for RIP experiments may be estimated using data comparing RIP-Chip or RIP-Seq results to data available from other sources (e.g. data obtained using other methods and knowledge from the literature) with the realization that results from the literature are not necessarily accurate. In addition, depending upon the results, comparisons to expectations can be made (e.g. binding to known RNA-binding sites and/or motifs) and tested to attempt to ascertain a (potentially biased) false positive rate.