

ENCODE and modENCODE

Guidelines For Experiments Generating Chromatin and Transcription Factor Data

Version 1.0, April 09, 2010

Every data producer aims to generate high-quality data sets. To help achieve that goal, this document aims to provide uniform standards and guidelines for experiments that map the location of DNA-associated proteins, chromatin modifications, chromatin organization, and DNA modifications.

Table of Contents

- I. Standard Measurements for Common ENCODE Cell Types
- II. ENCODE and modENCODE Standards for CHIP-chip and CHIP-seq Experiments
- III. Requirements for a DNase-seq and FAIRE-seq experiments
- IV. Standards for determining DNA methylation status
- V. Recommended Standards for Reporting mod/ENCODE Data

Appendix I. Recommended Standards for CHIP-seq

Appendix II: Validation of Commercial Histone Modification Antibodies to be Used for CHIP

I) Standard Measurements for Common ENCODE Cell Types

The ENCODE Consortium has designated common cell types that will be used by all investigators to aid in the integration and comparison of data produced using different technologies and platforms. To ensure consistency in cell cultures prepared in different laboratories, investigators should take the measurements below.

Required Measurements and Procedures

- **Growth time/passage number.** 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).
- **Cell density.** Cell density should be assessed for each cell culture, recorded, and submitted along with any data generated from that culture (see metadata standards).
 - The density of GM12878 cells should be maintained between 2.0×10^5 cells/ml and 1.0×10^6 cells/ml.
 - K562 cells should be grown to a maximal density of 7.5×10^5 cells/ml.
 - HepG2 cells should be grown to a maximum of 75% confluence.
 - HeLa-S3 should be grown to a maximal density of 5×10^5 cells/ml.

- **Cell cycle and gene expression state.** For all cell lines, genome-wide gene expression data taken under the growth conditions used for experiments must be collected in duplicate by each group. This needs to be performed once by a given group for each cell type, and not for each cell growth. For new groups entering the consortium, the gene expression data should be compared to established profiles to ensure concordance. If the guidelines for cell number and cell density are followed for Tier 1 and Tier 2 lines, analysis by FACS to determine cell cycle state is not required.
 - For cell lines new to the consortium (Tier 3), and for which no standard gene expression profile is available, it is recommended that cell cycle state should be measured by FACS analysis within 2 weeks of the time of harvest. Samples can be prepared at the time of harvest for later analysis. Staining should be done immediately before FACS analysis. A protocol for performing FACS analysis is available in the Cell Culture guidelines.
- **Presence of mycoplasma.** Cell cultures should be tested bi-monthly for the presence of mycoplasma. The mycoplasma testing protocol used by Bionique, which does mycoplasma testing for ATCC, is available in the Cell Culture guidelines.
- **Freezing Cell Aliquots.** Each ENCODE group should freeze a viable aliquot of each cell type used for any experiment for potential future phenotyping. The cells should be stored in the laboratory in which they are frozen.

II) ENCODE and modENCODE Standards for ChIP-chip and ChIP-seq Experiments

A number of researchers are mapping transcription factor binding sites, chromatin proteins, and chromatin modifications using chromatin immunoprecipitation followed by probing of genomic DNA microarrays (ChIP-chip), or by high throughput sequencing (ChIP-seq). To ensure the quality of ChIP data and the uniform reporting of results, a series of standards is proposed. Below is a plan for standards by which to monitor the quality, reproducibility, specificity, sensitivity and reporting of ChIP 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 comprehensive list of functional elements, and a high-confidence list of functional elements.

IIa. Antibody Characterization

Currently, there are a limited number of well-characterized antibodies against transcription factors and chromatin proteins. To ensure specificity of the antibodies used, the following tests must be employed. The data generated to characterize the antibody should be made publically available alongside the ENCODE dataset (see metadata standards).

Primary pathway to characterization

All antibodies used for ChIP experiments should be characterized using immunoblotting OR immunofluorescence.

- Immunoblot Analysis. Successful antibodies should show essentially a single band of the predicted size by immunoblot analysis. The minimum standard to be met is that the predicted reactive band compose no less than one half of the total signal in the lane, as assessed by quantitative immunoblot analysis of immunoprecipitation using either nuclear extracts or whole cells (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).
- **In addition**, antibodies must be further characterized by the use of one of the four following methods:
 - a. Knock-down of the target protein by genetic mutation , RNAi, or siRNA.
 - i. siRNA or RNAi knockdowns should be conducted in duplicate.
 - ii. For immunoblots, the band of the expected size along with additional immunoreactive bands should be reduced to 30% or less of the signal of unaffected cells.
 - iii. For IF, any nuclear staining should be completely eliminated. Any cytoplasmic background remaining after genetic mutation, RNAi, or siRNA should be noted.
 - b. Immunoprecipitation followed by mass-spec verification;
 - i. Mass spectrometry of immunoprecipitated proteins extracted from major bands (or all material) separated through gel electrophoresis should identify peptides corresponding to the predicted target protein of the antibody. All immunoreactive bands identified by immunoblot analysis should be identified and shown to be that of the expected protein and/or not a known chromosomal associated protein. All proteins identified by MS should be reported.
 - c. Immunoprecipitation with multiple antibodies against different parts of the target protein.
 - i. 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 (see Appendix I). For antibodies generated against protein modifications, two independently raised antibodies are recommended, and antibodies should be thoroughly characterized to ensure that they recognize only the desired modification and not other related modifications using the minimum standards applied in **Appendix II**.
 - ii. Each antibody may be used for ChIP experiments, and a statistically significant overlap of targets will be constitute characterization. Any reasonable method of correlation can be used (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 (**Appendix I**). An alternative approach is to perform ChIP-chip or ChIP-seq for one antibody and to use a second antibody to test 24 targets selected from a range of positions using qPCR. At least 80% overlap of validated targets is expected. The full set of targets identified by each antibody, in addition to the intersection, should be reported.

- d. Immunoprecipitation with an epitope-tagged version of the protein.
 - i. An epitope-tagged version of the target protein may be used, preferably driven by the endogenous gene promoter. Experiments should be conducted and analyzed as described above for the use of multiple antibodies.

Alternate pathway to characterization

In recognition of the fact that there will be a subset of antibodies that work well for ChIP but not for IF or western, an alternate pathway to characterization is provided. For every sample, IF and Western must be attempted before this pathway is invoked.

- o Option #1. Prior knowledge may be a suitable method of antibody characterization, based on the ChIP pattern itself. In this case, absolute criteria are not practical or useful. The experimenter must express a rationale for why they think the pattern reported is valid. For example, comparison to proteins in the same complex can be useful. This rationale must be reviewed by the Data Analysis Working Group, and approved prior to data release on a case-by-case basis. The rationale should be released with the data track.
- o Option #2. If a robust signal is obtained by ChIP-chip or ChIP-seq, and a ChIP median signal by RNAi is reduced to <50% the original signal in duplicate, the experiment will be considered validated.

Additional Controls

The following additional controls may be performed, but will not substitute for any of the above methods.

- One may incubate the antibody with a blocking protein or peptide blocking protein prior to the chromatin immunoprecipitation, and report the results.

Other considerations

- For antibodies directed against members of a multi-gene family, antibodies will be prepared to protein regions that are unique to individual family members. Any potential cross-reaction should be noted when reporting data collected using that antibody.
- Antibodies that have been previously characterized in one organism or cell type can be used directly in that organism or cell type by other groups without additional characterization. For whole-organism characterization, testing in three growth stages is sufficient.

- Antibodies to be used on a new cell type or organism will be considered characterized if they are successfully tested by **any** of the single criteria above.
- Different Lots of antibody from the same manufacturer do not need to be characterized if they recapitulate previous ChIP-chip or ChIP-seq results to the standards established for biological replicates.

IIb. ChIP-chip and ChIP Seq Data Production

As outlined in the mod/ENCODE Data Verification Standards, ChIP-chip and ChIP-seq data should be verified to assess the reproducibility of the experiment by performing at least 2 biological replicates for ChIP chip or ChIP seq. Three biological replicates are recommended for ChIP-chip and two for ChIP 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 other list (See **Appendix I**). Once data have been verified, they should be released into public databases as consistent with the ENCODE/modENCODE Data Release Policy. See **Section V** for ChIP-chip and ChIP Seq Data Analysis and Reporting.

ChIP Validation by use of an alternate detection platform

Early standards called for validation of biochemical using an alternative detection platform. Concordant observation of peak signals using alternative detection methods raise confidence that the detected events are real. However alternate methods of detection do not test the biological validity of the results, or the accuracy of the underlying biological assay (for example, the accuracy of ChIP itself). Moreover disagreement between detection methods may tell you what differences exist, but not which answer is correct.

Because of the high cost and effort involved in validation using alternative platforms the decision to validate each experiment has been reconsidered. Moreover, practical considerations of cost along with validation experience thus far indicate that ChIP-Seq will be the current method of choice for mapping transcription factor binding sites, at least for mammalian experiments. The group therefore decided that detailed analysis of the quantitative accuracy of high-throughput DNA sequencing platforms using alternative detection methods (qPCR and nanostring) be performed. These experiments should determine what a ChIP-Seq signal really means and potentially reveal the optimal methods for quantifying signals relative to background. The tiling microarray platform has already been thoroughly and quantitatively analyzed (PMID: 18258921).

Consortium members wishing to validate ChIP experiments may use either Quantitative PCR (qPCR), high-throughput DNA sequencing, or ChIP-chip approach to validate ChIP experiments. This is highly recommended for investigators lacking ChIP experience.

qPCR: The quality of the primers used for qPCR should be tested using gel electrophoresis of the PCR products. Forty-eight ChIP-chip or ChIP Seq targets selected across a range of signal should be validated by qPCR to determine the number of false positive and false negative hits. Of the 48 targets selected for validation, 6 should be negatives. qPCR experiments should be

performed in triplicate using one biological replicate when testing unamplified samples and two biological replicates when testing amplified samples (but see below). The ChIP samples used in the ChIP-chip and ChIP-seq experiments and in the qPCR validation can be from different biological replicates. Positive and negative qPCR results should be assessed using a t-test with a 0.05 P-value. The P value can be determined by several means. One method is by comparing the values of multiple replicas of the targets with a pool of many negatives. A 0.05 P-value will normally provide a signal threshold of 2-4 fold. As an alternative to performing validation in triplicate, tests can be performed in duplicate; those samples exhibiting discrepancies of greater than 15% must be tested by a third test. Testing unamplified samples is always preferred.

Sequencing (to validate microarray):

Experiments are to be performed to a sequencing depth as specified for ChIP-seq experiments. The data should match to the degree specified for within-platform biological reproducibility.

DNA microarray (to validate sequencing):

Experiments are to be performed on tiling microarrays representing at least 1% of the genome. The data should match the sequencing data to the degree specified for within-platform biological reproducibility.

Specificity and Sensitivity

Specificity and sensitivity are difficult to accurately estimate for ChIP experiments since the number of true positives is difficult to obtain, the number of true negatives is not known, and methods for identification of *in vivo* binding independent of ChIP is difficult to ascertain. Nonetheless, the positive predictive value (defined as $TP/(TP+FP)$ where TP is true positives and FP is false positive) for ChIP experiments may be estimated using data from validation as defined above. In addition, for qPCR, sensitivity of the positives that can be detected by the particular experiment and dynamic range can be crudely estimated by extrapolating from the validation experiments. Sensitivity and specificity estimates should be reported at 3 values (25%, 10% and 5%) based on the validated data set or a justifiable parameter. It is possible that more than 48 samples will be needed in some cases to provide that estimate.

Specificity and sensitivity can also be estimated by comparing ChIP-chip and ChIP 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 promoters and/or motifs) and tested to attempt to ascertain a (potentially biased) false positive rate.

Data that do not meet the criteria

If after repeated attempts, data does not meet the release criteria, it may be released with a prominent note indicating that the criteria have not been met and explaining why the data is being released without meeting criteria. The displayed track on the browser should be prominently flagged, and explain that the data should be used with caution.

III) Requirements for a DNase-seq and FAIRE-seq experiments

Following an analysis of deeply sequenced DNase-seq and FAIRE-seq datasets, we suggest the following requirements.

1. Limited number of sequenced reference samples

Deeply sequenced reference samples such as Input DNA exhibit uneven coverage. In some cases this includes enrichment in the promoter regions, while other regions of the genome are under-represented in the reference sequence (alpha satellite regions, ribosomal subunits, etc.). We believe that the input promoter peaks are a result of either endogenous nuclease activity, or FAIRE-like regions that were not cross-linked during the sample preparation. These promoter peaks likely represent real open chromatin and therefore should not be excluded from our analysis. Three factors limit the impact of uneven background coverage. First, advances in computational methods to correct for such features are being incorporated into the analysis. Second, the DNase/FAIRE-seq data itself can be used to identify regions that exhibit copy number variations in our samples or that are under-represented in the sequenced reference genome. Third, the true signals from FAIRE and DNase exhibit a unique structure that differs greatly from the type of signal produced by uneven input coverage. Therefore, while it is always preferable to have deeply sequenced matched input for each sample, for DNase and FAIRE experiments, input sequencing from every cell type is not required at this time.

2. Depth of Sequencing

Since DNase and FAIRE data represent a continuum of chromatin openness, achieving true saturation may not be practical, or even definable. However, a decision must be made regarding adequate level of coverage. We propose that the optimal depth of sequencing be guided by our ability to identify regions that were also identified other methods such as by tiled arrays (ENCODE pilot arrays or equivalent), qPCR, or Southern blots. For DNase and FAIRE this is typically 20-50 million reads.

3. Number and Reproducibility of Biological Replicas

By definition, at least two biological replicates are necessary to ensure that the experiment is reproducible. Experiments completed to date indicate that there will not be a significant gain in information beyond two biological replicates, when they are in reasonable agreement. We propose to require the following agreement between biological replicas:

- a) The number of mapped reads from replicas should be within a factor of two of each other.
- b) The length of target lists should be within a factor of two of each other.
- c) Either of the following options
 - a. Intersect top fraction (40%) of target list from one replica against the entire other target list and require a threshold amount of overlap (80%). Repeat for the reciprocal.

- b. Target lists scored using all available reads from both replicates must share more than 75% of targets in common with each of the replica experiments.

These parameters may have to be revisited as more data sets are available, and methods to compare replicates of different quality or enrichment levels are developed.

IV) DNA Methylation Standards

The ENCODE group at the HudsonAlpha Institute has used three methods to determine DNA methylation throughout the human genome in ENCODE cell lines. One of these (Methyl-seq) was used early in the project, but has been subsequently replaced with Reduced Representation Bisulfite Sequencing (RRBS), which has and will continue to be used to determine the methylomes in all ENCODE lines. A third method, which uses Illumina's Methyl27 arrays to measure methylation, has been used since the beginning of the project and will also be used for all ENCODE cell lines. Each method is performed on biological replicates to determine appropriate consistency measures.

1. Reduced Representation Bisulfite Sequencing (RRBS)

In RRBS, genomic DNA is digested with a methylation-insensitive restriction enzyme and small restriction fragments covering the size range between 40 to 120 bp are purified. This method generates a specific, reduced representation of the genome of DNA fragments enriched for CpG dinucleotides. The selected fragments are treated with bisulfite to convert unmethylated C's to U's, and these fragments are then subjected to high-throughput, short-read DNA sequencing. Sequence reads are processed in a stringent manner to eliminate ambiguous alignments. Finally, the degree of methylation of each fragment, estimated from the number of converted reads compared to the unconverted reads in each CpG, is calculated. An RRBS library is considered complete only when more than 500,000 CpGs are assayed with at least 10 sequencing reads each. It has been determined that at least 10X coverage of a CpG is required for accurate measurement of percent methylation. RRBS libraries of replicates are required to exhibit 90% concordance in methylation levels at CpGs assayed with greater than 10X coverage in both replicates.

2. Methyl27

To perform Methyl27, 1 µg genomic DNA is treated with bisulfite and then hybridized to a microarray. DNA methylation levels are then determined by using Illumina's Methyl27 technology. Appropriate quality control is done on each array, including the requirement of an initial call of at least 90% of features. Biological replicates of each cell line are required to show 90% concordance in methylation levels across all CpGs that pass QC.

3. Comparison of RRBS and Methyl27

When both RRBS and Methyl27 are performed on the same samples, a comparison is made at the CpGs that are assayed by both experiment types. Methyl27 and RRBS are

considered to be in agreement when overlapping CpGs show 70% concordance in methylation levels.

4. Methyl-seq

Methyl-seq is a sequence census method that measures the DNA methylation status at about 250,000 CpGs in the human genome. Genomic DNA is digested separately with MspI, a methyl-insensitive restriction enzyme, and HpaII, a methyl-sensitive restriction enzyme. Size-selected fragments from each digest are sequenced and read counts are compared to determine locations of DNA methylation. Methyl-seq was discontinued as a standard ENCODE assay after RRBS was implemented, as RRBS measures at least twice as many CpG's as Methyl-seq, requires only one sequencing lane instead of two, and is more quantitative in determining methylation frequencies.

Despite its discontinued use, we mention data standards because there are several useful Methyl-seq datasets in the ENCODE data releases. A Methyl-seq experiment is required to contain 7 million or more aligned reads for each biological replicate and the replicates are required to exhibit 80% concordance in methylation in regions assayed by both replicates. Additionally, we require 99% of the regions with HpaII reads to have MspI reads.

V) Recommended Standards for Reporting mod/ENCODE Data

Storing High Throughput Sequencing Data

- Image files from sequencing experiments do not need to be stored for the long term.

Submitting mod/ENCODE Data

For ChIP-chip, DNase-chip/array, and FAIRE-chip:

- Raw data should be submitted to GEO simultaneous to DCC submission (note 1).
 - Data should be flagged as being part of the mod/ENCODE project upon submission to NCBI
 - Currently (5-30-2010) The ENCODE DCC is submitting previously submitted ENCODE sequences to the NCBI on behalf of the production labs. If this goes smoothly, the ENCODE DCC will provide this service for future submissions.
- Processed data should be submitted to the relevant DCC as:
 - Ratio tracks
 - Called peaks (see below)
 - Metadata, including peak caller version used (see below)

For ChIP-seq, DNase-seq and FAIRE-seq:

- Raw data should be submitted to GEO (Small Read Archive; SRA) simultaneous to DCC submission
 - Data should be flagged as being part of the mod/ENCODE project through the use of the appropriate genome project ID

- Each replicate should be submitted independently
- NCBI intends for sequence data to be submitted directly to GEO. GEO will pass primary sequence data to the SRA. If more convenient, data producers can make dual submissions to the SRA and GEO. If this is done, sequences should be submitted to the SRA and metadata, processed data, and a link to the SRA accession should be submitted to GEO.
- Processed data should be submitted to the relevant DCC as:
 - Input signal or alignments (but not for DNase/FAIRE Tier 3 lines)
 - CHIP, DNase, and FAIRE signal or alignments
 - The ENCODE DCC will accept sequence alignments and/or signal graphs for input and CHIP data. Sequence alignments will be posted for download, and will be loaded and displayed as a 'Counts graph' (count of tags overlapping each base) if no signal graph is submitted.
 - As of now (5-30-2010) All ENCODE labs are generating their own signal tracks, so the ENCODE DCC does not plan to continue this service.
 - Interpreted data signal
 - Called peaks (see below)
 - FASTQ files
 - The ENCODE DCC will host FASTQ format sequence data and, where feasible, provide a co-located cluster computing for analysis since the GEO/SRA pipeline and access tools are not mature enough at this stage to adequately support ENCODE analyses.
 - Note added 5/30/2010. This may now be obsolete (access may now be sufficient).
 - Metadata, including peak caller version used (see below)

Target Region and Peak Calling for CHIP, DNase and FAIRE Experiments

Point Source Peaks

For point source peaks (e.g., DNase, FAIRE, or signals from CHIP experiments with antibodies to sequence-specific transcription factors), common features that should be reported to the DCC are:

- Peak, defined as a single base pair
- Start and end, defined as specific base pairs
- Significance statistics using a three slot model (the inclusion of slots 2 and 3 is optional for data submitters):
 - Slot 1: Signal value (e.g., fold enrichment) using an algorithm chosen by the submitter
 - Slot 2: P -value determined using a method chosen by the submitter
 - Slot 3: Q-value (false discovery rate correction) determined using a method chosen by the submitter

- Metadata, including peak caller approach used (see below) and methods for determining signal values, P-values, and Q-values, as applicable

Broad Regions

- Start and end, defined as specific base pairs
- Significance statistics using a three slot model (the inclusion of slots 2 and 3 is optional for data submitters):
 - Slot 1: Average signal value across region (e.g., fold enrichment) using an algorithm chosen by the submitter
 - Slot 2: P -value determined using a method chosen by the submitter
 - Slot 3: Q-value (false discovery rate correction) determined using a method chosen by the submitter
- Metadata, including peak caller approach used (see below) and methods for determining signal values, P-values, and Q-values, as applicable
- Point-source peaks can be called in addition to broad regions (i.e. one can have "peaks" and potentially "valleys" within "regions").
 - This situation has never been formally defined, so the ENCODE DCC does not currently support this feature 5/30/2010. Perhaps additional discussion on this topic is needed if this is an important feature.

It is up to the investigator to determine whether their data best fits the broad region/point source peak data or both.

Metadata for Peak Caller

As an overall aim, mod/ENCODE should strive to provide public access to peak calling software so outside data users can replicate the findings of the mod/ENCODE data producers. Currently, peak calling software can be downloaded from the websites of the individual data production groups and initially the consortium and should build on this arrangement. In the longer term, it may be worthwhile to attempt to standardize peak and region calling approaches so as to achieve optimal integration of data.

The metadata for each experiment should include a free form field where data producers are required to include information about the peak caller version that was used to produce the hit list along with any information on parameters used for a particular experiment.

The modENCODE metadata field and ENCODE track documentation will be linked to a DCC web page that lists the peak callers (and versions) that have been used for mod/ENCODE data, which will in turn link to the websites maintained by the individual groups that allow for the downloading of peak caller software by outside data users. Data producers are expected to update information about peak calling software (including versions) on their websites as soon as new or updated software is implemented.

Metadata for Antibodies and experimental procedures

The DCC should implement a rubric for submission that requires submission of certain basic experimental data, including minimally the following:

1. Confirm that the experiment has been performed in duplicate (sign off).
2. Confirm that the experiment meets data quality standards (sign off).
3. Attach a cell growth protocol for this experiment, or reference a previously submitted protocol.
4. Provide the catalog and lot number for any antibody used. If not a commercial antibody, indicate the precise source of the antibody. What criteria were used to assure that the antibody correctly recognizes the target?
5. Provide any images of Western or IF experiments that characterize the antibody according to the recommendations above.
6. Summarize the number of reads and number of targets for each replicate and for the merged dataset. Provide the overlap results (using the top 40% rule). What criteria were used to validate the resultant ChIP-seq data?
7. Is this experiment linked to an INPUT or other control track? If so, specify the track name.
8. Does the experiment fail to meet any of the data or experimental standards? If so, explain why and FLAG the track on the browser.

Appendix I. Recommended Standards for ChIP-seq

1. Matched sequenced reference sample

Deeply sequenced reference samples such as Input DNA sometimes exhibit uneven coverage, including “peaks” in the promoter regions of many known genes. To identify regions of transcription factor binding in a ChIP-Seq experiment it is necessary to have sequenced a relevant reference sample in the same cell-line (and same cellular condition) to show that the region is enriched in the ChIP sample compared to the control. For the case of ChIP-Seq experiments for different transcription factors, only one control sample need be produced for each cell-line. The control (i.e. Input DNA) should be sequenced to at least the same depth as the greatest depth of any matching individual transcription factor ChIP-Seq experiment.

2. Depth of Sequencing

The required depth of sequencing needed will vary depending on the nature of the binding of the transcription factor (number of binding sites, domain vs. point source binding, efficiency of the antibody). Some transcription factors/chromatin modifications might require exceptionally deep sequencing to achieve probable saturation of all biologically relevant sites. We recognize that for many factors, we will not have a large number of previously known biologically relevant sites at the time of the ChIP-Seq measurement. One of three criteria should therefore be used for depth of sequencing:

- a) When the number of targets begins to saturate (i.e. approach the asymptotic number of identifiable targets). The criteria should be $\geq 95\%$ of the extrapolated total number of targets (for HeLa-S3 Pol II 12 million mapped reads yields greater than 95% of the approximately 30,000 extrapolated total targets).
- b) If the total number of targets does not approach saturation, one should detect 99% of targets that show at least 2-fold enrichment over control with 90% of the data.
- c) If either a) or b) are not satisfied, then starting June 1, 2010 at least 10 million mapped reads should be sequenced per replicate for mammalian experiments, with a minimum of 4 million mapped reads per replicate for worm or fly. Note that older data sets did not require this depth of sequencing and were acceptable under standards at the time.

3. Number and Reproducibility of Biological Replicas

At least two biological replicates are necessary to ensure that the experiment is reproducible. It does not seem that there will be a significant gain in information beyond two biological replicates, when they are in reasonable agreement. The data from replicates can then analyzed and reported. We propose to require the following agreement between biological replicas (target lists identified using common number of mapped reads):

- a) The number of mapped reads from different replicas should be within a factor of two of each other.
- b) The length of target lists should be within a factor of two of each other.
- c) Plus, either of the following options:

- I. Intersect top fraction (40%) of target list from one replica against the entire other target list and require a threshold amount of overlap (80%) and repeat for the reciprocal. The longer list may have to be trimmed to the length of the shorter list.
 - II. Target lists scored using all available reads share more than 75% of targets in common.
- d) Any experiment not conforming to the guidelines above should be noted, and submitted with a paragraph documenting the reproducibility. This would include number of reads/replicate, number targets/replicate, the overlap information and so on, so that data users can have all the information that they need to assess the value of the experiment.

Appendix II: Validation of Commercial Histone Modification Antibodies to be Used for ChIP

Overview: There are numerous issues that need to be addressed with respect to validating commercial histone modification antibodies used extensively for ChIP-chip and ChIP-seq analyses. The main issues are 1) specificity of antibodies with respect to other nuclear/chromatin proteins, 2) specificity with respect to unmodified histones and other modified histone residues (e.g. H3K9me and H3K27me), 3) specificity with respect to mono-, di-, and tri- methylation at the same residue (e.g. H3K9me1, me2 and me3), and 4) lot-to-lot variation. Validation of commercial antibodies is necessary to produce the high quality data sets desired by the modENCODE and ENCODE consortia. We propose that all commercial histone antibodies be validated by at least 2 independent methods, as described below. New lots of antibody must be analyzed independently. The tests may be performed by ENCODE/modENCODE labs or by the companies that sell the antibodies, but only if the companies provide data for the specific lots of antibody. The tests need only be performed once for each antibody lot.

Test #1: All antibodies must be checked for reactivity with non-histone proteins and with unmodified histone by performing Western blot analysis of total nuclear extract and recombinant histone. To enable visual quantification of reactivity, a concentration series of both extract and recombinant histone should be analyzed, using recombinant histone levels that are comparable to the respective histone in nuclear extract. Since cross-reactivity may vary between species, this test should be performed using nuclear extracts from each species to be studied by ChIP. In nuclear extracts, the specific nuclear histone band should constitute at least 50% of the nuclear protein signal, show at least 10-fold enrichment relative to any other single band, and show at least 10-fold enriched signal relative to unmodified histone.

Test #2 options: In addition to test #1, antibody specificity must be verified by at least one additional test. The tests and the pros and cons of each are described below, followed by the likely flow of tests.

A. Peptide binding tests. Histone tail peptides with particular modifications can be purchased commercially. Peptide binding and peptide competition assays provide a fast method to initially evaluate the specificity and relative strength of binding of antibodies to histone tails with different modifications (e.g. H3K9 and H3K27 and me1, me2, and me3 levels of methylation). A potential drawback is that antibodies may differ in their binding specificity toward histone tail peptides in vitro versus toward full-length histones in the context of chromatin in IP experiments. Nevertheless, observing at least a 10-fold enriched binding signal for the modification of interest relative to other modifications would contribute to confidence in antibody specificity.

B. Mass spec. For antibodies generated against related and historically problematic modifications, the ability of the antibody to effectively distinguish between similar histone

marks (e.g. H3K9me and H3K27me) and between different levels of methylation (e.g. H3K9me1, me2 and me3) should be tested by mass spec analysis of material IP'd from histone preparations. The target modification should constitute at least 80% of the IP'd histone signal, and contaminating bands should not contribute more than 20%.

C. Mutants defective in modifying histones. Mutants (in *S. cerevisiae*, *S. pombe*, *Drosophila*, *C. elegans*, or mammalian cells) that lack the ability to catalyze particular histone modifications offer the opportunity to test antibody specificity. Antibody signal should be reduced to below 10% of wild-type signal in mutant samples, compared to wild type. RNAi depletion of histone modifying activity may be substituted for mutants. Mutant or RNAi reduction of signal can be assayed by Western blot analysis or by immunofluorescence staining. Mutant/RNAi tests usually don't allow testing antibodies for the ability to discriminate between mono, di, and trimethylation. In cases where more than one enzyme modifies the same residue (e.g. K9 methylation in *Drosophila*), double mutants or RNAi may be required.

D. Mutant histones. Mutant histones (e.g. histone H3 with Lys4 mutated to Arg or Ala) expressed in yeast provide another avenue to alter histones for testing specificity by western blot analysis or even by ChIP. Since the modified residue is mutated, we expect at least a 10-fold reduction in signal relative to wild-type histone preparations. Mutant histone tests cannot distinguish whether antibodies discriminate between mono, di, and trimethylation.

Flow of tests:

For antibodies to methylated histone residues, we envision the flow of tests will be:

- #1 western blot analysis to make sure the antibody does not show significant cross-reaction with unmodified histone or non-histone proteins;
- #2A peptide binding/competition tests to make sure the antibody does not interact with histone tail peptides lacking modifications or bearing other modifications; for similar and problematic modifications,
- #2B mass spec analysis to make sure the antibody does not IP unmodified histone or histone bearing other modifications.

For antibodies to histone modifications other than methylation, we envision the flow of tests will be:

- #1 western blot analysis to make sure the antibody does not show significant cross-reaction with unmodified histone or non-histone proteins;
- #2A peptide binding/competition tests to make sure the antibody does not interact with histone tail peptides lacking modifications or bearing other modifications OR one of the other #2 tests listed above.

Grandfathering old antibodies and testing new lots of antibodies: Once ChIP-chip or ChIP-seq results are obtained with antibodies validated as described above, then old or new antibodies can be validated by similar testing or by obtaining ChIP results using old/new antibodies that

are statistically consistent with CHIP results using validated antibodies (based on ENCODE and modENCODE standards for statistical agreement between replicates).

Use of 2 different antibodies: Even if antibodies pass the specificity tests described above, observing very similar CHIP results with 2 independent antibodies would give added confidence in the results. Therefore, we encourage using 2 independent antibodies whenever possible, providing statistical comparisons of the results, and presenting the intersection of the peak sets obtained with the 2 antibodies.