

# ENCODE and modENCODE Guidelines For Experiments Generating ChIP, DNase, FAIRE, and DNA Methylation Genome Wide Location Data

Version 2.0, July 20, 2011

Every data producer aims to generate high-quality data sets. To help achieve that goal, this document aims to provide standards and guidelines for experiments that map the genomic 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: Quality control measures for ChIP-seq experiments

## I) Standard Measurements for Common ENCODE Cell Types

The ENCODE Consortium has designated common cell types that will be used by all investigators. This will 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 date at which cells were put into culture and when they were harvested should be recorded. Investigators should use the original stock after growing a culture for two months. Passage number should be assessed and recorded for primary cells. Primary cells should not exceed 6 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  $2.0 \times 10^5$  cells/ml and  $1.0 \times 10^6$  cells/ml.
  - K562 cells should be grown to a maximal density of  $7.5 \times 10^5$  cells/ml.
  - HepG2 cells should be grown to a maximum of 75% confluence.
  - HeLa-S3 should be grown to a maximal density of  $5 \times 10^5$  cells/ml.
- **Cell cycle and gene expression state.** Gene expression experiments were performed by seven different laboratories using cells of the same lot and recommended growth conditions. Strong concordance was observed. Thus, 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.
- **Presence of mycoplasma.** Cell cultures should be tested periodically for the presence of mycoplasma. This is particularly critical if the growth of cells is altered. The

mycoplasma testing protocol used by Bionique, which does mycoplasma testing for ATCC, is recommended.

- **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**

Despite their widespread use, there is considerable diversity in the way global ChIP experiments are designed, executed, scored and reported. There are substantial differences in the number of sites detected for each factor, ranging from hundreds to tens of thousands. This is partly a function of the underlying biology of different factors and modifications, but it is also a function of differences in the quality, scoring and reporting of experiments. To address the needs for reproducible high quality data and to facilitate analysis and dissemination of results, the ENCODE and modENCODE consortia (Feingold et al., 2003; Celniker et al., 2006; ENCODE Consortium, 2011) have worked to develop standards and best practices for ChIP-chip and ChIP-seq experiments. Because the methods for performing genome-wide ChIP continue to evolve, our standards and practices have also evolved during the course of these projects, informed by results from over one thousand experiments performed in different organisms (e.g. *D. melanogaster*, *C. elegans*, mouse and human), and multiple independent data production pipelines. The resulting guidelines include recommendations on study design, quality control, evaluation of results, reporting and archiving. Below are our current standards with the expectation that they will be revised as protocols and technologies change over time.

### **Ila. Antibody Characterization and Epitope Tagging**

ChIP experiments have been performed successfully using both polyclonal and monoclonal antibodies. However, the success of these experiments is heavily dependent upon the quality of the antibodies, which can vary considerably in terms of specificity and performance. Consequently, we propose a set of standards for antibody characterization, which differ for antibodies generated against transcription factors relative to those used to study histone modifications. The thresholds used in these standards, while somewhat arbitrary, provide a useful guide for helping to ensure that high quality data are generated.

#### **Characterization of antibodies directed against transcription factors**

The workflow for characterization of a new antibody that binds a transcription factor is summarized in Figure 1; both a primary and secondary characterization are required for each new antibody and for a new lot of a previously characterized antibody (Table 1).

##### **Primary mode of characterization**

Antibodies are characterized by **one of two** primary methods:

1. **Immunoblot analyses:** Immunoblot analysis should be performed on protein lysates from either whole cell extracts, nuclear extract, chromatin preparations, or immunoprecipitated material. The primary reactive band should contain at least 50% of the signal observed on the blot. Ideally, this band should correspond to the size expected for the protein of interest. However, the mobility of many factors deviates significantly from the expected size due to modifications, isoform differences or intrinsic properties of the factor. Therefore, main bands that differ from the expected size by more than 20% or multiple bands are acceptable under certain circumstances, such as if the unexpected mobility has been properly documented in published studies using the same antibody lot, if the signal in the band is reduced by siRNA knockdown or mutation, or if the factor can be identified in this band by mass spectrometry. Before proceeding to

ChIP assays, it is helpful to demonstrate that the protein of interest can be efficiently immunoprecipitated from a nuclear extract.

2. Immunofluorescence: Some antibodies that work well for ChIP-seq or ChIP-chip do not work well in immunoblots. If immunoblot analysis is not successful, immunofluorescence may be used as an alternative method. Staining should be of the expected pattern (e.g. nuclear and only in cell types or under specific growth conditions that express the factor). Because immunofluorescence does not provide evidence that the antibody detects only one protein, this validation method must be combined with a method which reduces the level of the protein, such as siRNA- or shRNA-mediated knockdown or use of a knockout cell line or organism (see below).

### **Secondary mode of characterization**

In addition to the primary mode of characterization, a secondary method must be performed using one of five assays:

1. Knockdown or knockout of the target protein. Immunoblots or IPs must be performed using extracts from siRNA or shRNA knockdowns or from knockout mutants. The primary immunoblot signal, along with additional immunoreactive bands, should be reduced to no more than 30% of the original signal. For immunofluorescence assays, nuclear staining should be to no more than 30% of the original signal, and any signal remaining after genetic mutation, RNAi, or siRNA should be noted. If the primary immunoblot or immunofluorescence characterization methods fail, knockdown can also be measured with ChIP experiments: statistically significant reduction of ChIP-chip or ChIP-seq signals in 30% of peaks (or in >30% in the size of enriched regions for broad marks) relative to knockdown controls is deemed acceptable as long as suitable control knockdown (nucleotides of similar sequence but not complementary to the target protein knockdown, e.g., scrambled siRNA sequences) is performed and analyzed at the same time.
2. Immunoprecipitation followed by mass spectrometry. All immunoreactive bands identified by immunoblot analysis should be analyzed. A protein identified in such bands should correspond to the protein of interest; if additional chromosomal proteins are identified in an immunoreactive band they should be present at lower prevalence than the desired protein (as measured by peptide counts or other methods). All proteins identified by mass spectrometry and their number of peptide counts should be reported
3. Immunoprecipitation with multiple antibodies against different parts of the target protein or members of the same complex. Different antibodies against different parts of the same protein or other members of a known protein complex may be used in ChIP experiments for antibody validation. For ChIP-seq experiments statistically significant overlap of enriched loci by one of three methods is required: a) the target is for 80% of the top 40% of the enriched loci (i.e. peaks) identified using one antibody should overlap the loci identified using the other antibody (and vice versa; see below) b) an  $R^2$  greater than 0.5 of signals from peaks on the two target list should be observed c) Reproducible peaks that pass a fixed IDR threshold of 1% by comparing replicate experiments of each antibody should be similar in number and overlap (factor of 2) to those obtained by comparing experiments across different antibodies (see below). Alternatively, a second antibody can be used to test 24 targets identified from the first antibody using qPCR. These targets should be selected from a range of enrichments. At least 80% of the tested loci should give signals within 25% of one another or the positive targets should show an overall enrichment of at least 2-fold. Other strong correlations can be considered. The full set of enriched loci identified by each antibody, in addition to the intersection, should be reported. Note that for different proteins that are members of a complex, there may be some functions that are independent of one another. Thus the targets lists for two different proteins may not entirely overlap.

4. Immunoprecipitation with an epitope-tagged version of the protein. An epitope-tagged version of the target protein may be used, preferably expressed from the endogenous gene promoter. Experiments should be conducted and analyzed as described above for the use of multiple antibodies.
5. Motif enrichment. For validation, an enrichment of at least 4-fold using a known or justifiable de novo motif of at least six nucleotides long will be required. In addition, the motif should be present in at least 10% of analyzed peaks.

#### *Other considerations*

1. Additional controls are welcomed but cannot replace the criteria indicated above. These include competition of immunoblot, immunofluorescence and/or ChIP signals using peptide and/or protein for the transcription factor of interest. Signals are expected to be diminished. Correlation of the pattern of protein expression at different developmental timepoints or different conditions using immunoblot analysis or immunofluorescence with that expected for the protein based on RNA expression is also a useful criteria.
2. For antibodies directed against members of a multi-gene family, antibodies should 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.
3. For antibodies that have been previously characterized for one cell type, only one validation method is required when used for ChIP-chip or ChIP-seq with a new cell type or organism. If an antibody has been validated in at least 3 different cell types, no further validation is needed for use with additional cell types. For whole organisms, if the antibody has been characterized in three growth stages, no further characterization is required. Alternatively, when an antibody has been previously validated in one cell line/tissue/developmental stage, motif enrichment (or annotation enrichment for histone modification) can be used for validation in another cell line/tissue/developmental stage. In this case the motif enrichment on the new datasets should be no less than 15% below the enrichment percentage observed in the original dataset.
4. Characterized antibodies may be used by different groups without further characterization, assuming they are derived from the same lot. Antibodies from different lots must be characterized as if they were new antibodies.

#### ***Epitope-tagged proteins***

The creation of cell lines or organisms expressing an epitope-tagged transcription factor for ChIP is a useful alternative and is particularly valuable in cases where a ChIP quality antibody is not available against the endogenous protein. A concern with these experiments is that if the tagged factor is expressed at significantly higher levels than the endogenous protein, some fraction of the identified binding sites may be occupied only because the protein is present at artificially high levels. Therefore it is recommended that tagged factors be expressed from their natural promoters in low copy. If the tagged protein is expressed from a heterologous promoter, data comparing expression levels of the tagged and endogenous proteins (i.e. immunoblots to measure protein levels or qPCR to measure RNA levels) under conditions in which the ChIP experiments are performed should be presented. In cases where ChIP signals cannot be obtained when the tagged protein is expressed at endogenous levels or when the endogenous protein cannot be measured, these observations must be noted when the data is reported. The recommended control for experiments with epitope-tagged factors is an immunoprecipitation using the same antibody against the epitope tag in otherwise identical cells which do not contain the expression construct or which do not express the tagged factor (if the promoter is inducible). Scoring can be compared against these control or input DNAs.

## ***Histone Modifications***

There are numerous issues that need to be addressed with respect to validating commercial histone modification antibodies to be used for ChIP-chip and ChIP-seq analyses (Egelhofer et al., 2011). 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 trimethylation 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. All commercial histone antibodies must be validated by at least two independent methods, as described below. New lots of antibody must be analyzed independently. The tests may be performed by the laboratory performing ChIP or the antibody supplier, but only if the supplier provides data for the specific lots of antibody. The tests need only be performed once for each antibody lot.

### ***Primary Test***

All antibodies must be checked for reactivity with non-histone proteins and with unmodified histone by performing immunoblot analysis on 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 those of 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.

### ***Secondary test***

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

1. 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 binding strength 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 provides confidence in the antibody specificity.

2. Mass spectrometry. 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 spectrometry analysis of material immunoprecipitated from histone preparations. The target modification should constitute at least 80% of the immunoprecipitated histone signal, and contaminating bands should not contribute more than 20%. It is recognized that this test may often not be successful since IP for one modification can simultaneously isolate coassociated histones with other modifications. Thus, only a positive result (i.e. a specific modification) is interpretable.

3. Mutants defective in modifying histones. Strains or cell lines harboring knockouts or catalytically inactive mutants of enzymes responsible for particular histone modifications offer the opportunity to test antibody specificity. Such mutants exist for *S. cerevisiae*, *S. pombe*,

*Drosophila*, *C. elegans*, and can, in cases where the modifying enzymes are non-redundant, be created for mammalian cells. [Antibody signal should be reduced to below 10% of wild-type signal in mutant samples, compared to wild type. RNAi or siRNA depletion of histone modifying activity may be substituted for mutants. Mutant or RNAi or siRNA reduction of signal can be assayed by immunoblot analysis or by immunofluorescence staining. Mutant/RNAi/siRNA 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. Replicates of this test are encouraged but not required. However, positive controls showing that the antibody works on wild-type samples processed in parallel, and positive controls showing that the mutant extract is amenable to the assay employed, are required.

4. Mutant histones. Mutant histones (e.g. histone H3 with Lys4 mutated to Arg or Ala) expressed in yeast provide another avenue to test specificity by immunoblot analysis or even by ChIP. When analyzing a strain containing a mutated histone that cannot be modified, we expect at least a 10-fold reduction in immunoblot or IP signal relative to wild-type histone preparations. Mutant histone tests cannot distinguish whether antibodies discriminate between mono, di, and trimethylation.

5. Annotation enrichment. Enrichment at annotated features (e.g. transcription start sites) can be used as a validation criterion for chromatin associated modifications and proteins. If a well characterized modification (e.g. Histone H3 lysine 4 trimethylation) is analyzed, the observed localization to annotations should be similar to that of known overlap standards derived from the literature or existing ChIP-seq datasets (for point source peaks, 80% of the top 40% of peaks should overlap known annotations).

#### *Flow of tests*

For antibodies to histone modifications, we envision the flow of tests should be:

- 1) Immunoblot 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;
- 2B) For problematic modifications, mass spectrometry analysis to make sure the antibody does not IP unmodified histone or histone bearing other modifications.

#### *Use of 2 different antibodies*

Even if antibodies pass the specificity tests described above, observing very similar ChIP results with 2 independent antibodies would provide 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. Similarly, new antibodies can be used to validate the results of previously used antibodies and “grandfather such datasets”.

## **IIb. ChIP-chip and ChIP-seq Data Production Standards**

In order to ensure that experiments are reproducible and high quality, standards have been established for performance of ChIP-seq and ChIP-chip experiments.

### ***ChIP-Seq***

#### *Sequencing Depth*

For ChIP-seq experiments, the number of targets identified varies substantially depending on the factor, antibody, and the algorithm used for peak calling (Wilbanks and Facciotti, 2010). It also depends on the depth to which the sample is sequenced (Appendix I). This latter point has been addressed in several studies with different factor/antibody combinations (RNA Pol II, Stat1, CTCF) and the number of reads necessary to approach saturation in peak identification varied substantially, from 14M to 40M uniquely mapped reads, to cases where no evidence of saturation was observed (Rozowsky et al, 2009; ENCODE consortium, 2011). For practical purposes and cost considerations for mammalian cells, beginning in June 1, 2010, a minimum depth of 10 million uniquely mapped reads per replicate was established as a standard for transcription factors and chromatin modifications that yields sharp peaks (note the previous requirement was 3 million reads); for flies and worms this figure is 2 million uniquely mapped reads. (Most of the reads (at least 80%) should be distinct; see quality control section below). For chromatin modifications that are broadly distributed such as H3K27me3 and H3K36me3, sequencing depth figures have not yet been established but deeper coverage is necessary (many groups currently use 15-20 million reads per replicate for mammalian cells. Recent analysis suggests that at least 5-10 million reads are needed for mixed or broad marks in flies (Jung et al, in preparation). It is recommended that the number of reads obtained for each replicate be similar. Control DNA needs to be sequenced to a similar or greater depth.

### *Controls*

For both ChIP-seq and ChIP-chip, control experiments must be performed. Breakage during sonication can occur preferentially in regions of open chromatin resulting in non-uniform background signal (Auerbach et al., 2009). In addition, many cell lines are aneuploid and have many large regions of genomic duplications, which can heavily influence peak sizes and rankings. Control DNAs include "Input" DNA, in which DNA is isolated from cells that have been crosslinked and sonicated under conditions similar to the experimental sample or "IgG" in which control immunoprecipitations are performed using an antibody fraction (Immunoglobulin G from an unimmunized animal) that does not recognize DNA or chromatin associated proteins. In cases where factor binding to DNA is environmentally induced, (e.g. after hormone induction of the glucocorticoid receptor or when analyzing a protein expressed from an inducible promoter), a potential appropriate control is a parallel ChIP experiment done on cells in the uninduced conditions. Similarly, for epitope tagged constructs suitable controls are performed with cells lacking the epitope tag. If amplification is used to prepare the experimental sample, then the control DNA must be prepared using the same amplification procedure; note that biases in amplification can increase the chances of overrepresentation and underrepresentation of sequences. Control experiments need to be performed for each cell line, developmental stage and different condition/treatment since the open chromatin regions are likely to change under the different cell types/stages/conditions.

### *Number of Replicates*

In order to ensure that experiments are reproducible at least two replicates must be performed. For ChIP-seq experiments with RNA Pol II, the use of more than two replicates did not significantly increase the number of targets identified (Rozowsky et al., 2009). Therefore, it is recommended that ChIP-seq experiments be performed in duplicate using two independent samples (i.e. biological replicates) and that the number of mapped reads and identified targets be within two fold of one another.

80% of the top 40% of the targets identified from one replicate using an acceptable scoring method (discussed below) should overlap the list of targets from the other replicate. This standard was chosen based on the experiences of the ENCODE production groups to allow an achievable threshold of reproducibility while generally producing high quality target lists.

Alternatively, target lists scored using all available reads from both replicates must share more than 75% of targets in common with each of the replicate experiments. Reads from replicate experiments that meet the above criteria are usually pooled. The recently described Irreproducible discovery rate (IDR) analysis methodology provides another alternative approach to assessing replicate agreement and is discussed below in Appendix II.

Broad chromosomal patterns are particularly challenging to quantify for reproducibility as the size and number of such regions can vary considerably. Metrics are being developed for assessing reproducibility for experiments that map broad marks and will be reported at a future date.

### *Scoring*

A variety of scoring methods exist to process ChIP data and to evaluate how successful an individual experiment has been. In general, the top targets measured by the absolute signal (read number) and significance of enrichment (e.g. p-values and false discovery rates) are found by nearly all methods. For ChIP-seq, the various methods include scoring peaks that are enriched relative to control experiments (e.g. PeakSeq, Rozowsky et al., 2009; SPP, Kharchenko et al., 2008; MACS, Zhang et al., 2008). Any reasonable scoring method is acceptable, but since background can be nonuniform, those that account for control signals are superior. Thresholds should be no higher than a 5% false discovery rate; 1% is commonly used for many algorithms. IDR tests in which similarity between replicates is assessed, comparison to known targets, and validation tests (see below) can be used to help set thresholds

### *Quality control*

A number of additional metrics provide useful indications of the quality of a ChIP-Seq experiment, but are not currently required to submit data to ENCODE databases (UCSC). These are summarized in Appendix II and include 1) visualization of tracks to ensure that replicates appear similar with ChIP peaks not present in input/IgG controls. 2) analysis of library complexity. The number of *distinct* (i.e. non-redundant) reads should be analyzed to ensure peaks are not due to PCR amplification artifacts 3) quantification of the fraction of reads present in the called regions (depending upon the scoring method and threshold, this figure is typically greater than 2% for factors with large number of peaks (>10,000). This method is most valuable when analyzing a factor/modification that is similarly expressed in multiple lines; similar levels of enrichment are often observed). 4) cross correlation analysis and 5) Irreproducible discovery rate (IDR) analysis of replicate agreement. The latter two methods were developed by the ENCODE Consortium, and assist in both quality control and the uniform thresholding of data, and are described briefly below and in detail in the appendix. Many of the methods (visualization of tracks, library complexity and cross correlation are performed on single experiments; IDR is performed on replicates). It is recommended, but not required, that as many of these different QC steps be applied and the results reported.

### *Cross correlation analysis*

A high-quality ChIP-seq experiment shows a significant degree of IP DNA fragment clustering around genomic locations that are stably bound by the protein of interest. This results in peaks of sequenced tag density on both the forward and reverse strand around the bound locations, separated on average by a characteristic distance which depends on the distribution of fragment sizes. A control experiment such as sequenced input DNA, will typically show a weak pattern of shifted stranded tag densities. One can characterize the degree of fragment clustering (IP enrichment) by computing a strand cross-correlation profile of a dataset (Kharchenko et al., 2008; Kundaje et al., 2011). We compute the correlation between tag densities on the Watson and Crick strand at different offsets with respect to each other i.e. uniformly shifting all reads on



the Watson strand to the left or right by  $k$  base pairs keeping read positions fixed on the Crick strand [Fig 1].

High quality ChIP-seq datasets tend to show a peak in the cross-correlation profile at a shift equal to the mode of the fragment length distribution whereas input datasets tend to show a peak in the cross-correlation profile at a shift equal to the read length [Fig 1]. Hence, the normalized cross-correlation at the fragment length offset as well as its value relative to the cross-correlation at the read length offset provide an intuitive and useful measure of data quality (Kundaje et al., 2011). We call these measures the normalized strand-cross correlation (NSC) coefficient and the relative strand cross correlation (RSC) coefficient respectively. The NSC and RSC coefficients are an inherent property of the dataset and are not affected by peak calling methods and thresholds. High quality ChIP-seq datasets typically have NSC values  $> 1.1$  and RSC values  $> 1$  (See Appendix and [Kundaje et al. 2011] for more details).

### *IDR analysis*

Since at least two replicates of each ChIP-seq experiment are required, we can leverage reproducibility of signal to select appropriate thresholds for peak calling and to evaluate data quality (Li et al., 2011). Given a set of peak calls for a pair of replicate datasets, the peaks can be ranked based on a criterion of significance, such as the p-value, q-value or ChIP to input enrichment, or the read coverage for each peak. If two replicates measure the same underlying biology, the most significant peaks, which are likely to be genuine signals, are expected to have high self-consistency, whereas peaks with low significance, which are more likely to be noise, are expected to have low self-consistency. This manifests as reshuffling of the ranks of the scores at the noise level. Thus, if the consistency between a pair of rank lists that contains both significant and insignificant findings is plotted in order of decreasing significance, a change in consistency is expected. This change of consistency provides an internal indicator of the transition from signal to noise and indicates how many peaks have been reliably detected.

The IDR statistic formally quantifies this notion of reproducibility by modeling all pairs of peaks present in both replicates as belonging to one of two groups: a reproducible group and an irreproducible group. In general, the signals in the reproducible group are more consistent (i.e. have a larger correlation coefficient) and are ranked higher than the irreproducible group. The proportion of identifications that belong to the "noise" component and the correlation of the significant component are estimated adaptively from the data. The IDR provides a score for each peak, which reflects the posterior probability that the peak belongs to the irreproducible group.

A major advantage of this approach is that it provides a stable threshold for called peaks that is much more consistent across laboratories, antibodies, and analysis protocols (e.g. peak callers) than the FDR. The increased consistency comes from the fact that the IDR uses information from replicates, whereas the FDR is computed on each replicate independently.

### **ChIP-Chip Experimental Standards**

For ChIP-chip control, replicates, scoring and quality control are also performed.

Controls and scoring: For hybridization to arrays input DNA or genomic DNA is typically prepared and labeled with a different dye relative to the ChIP DNA (e.g. Cy3 vs Cy5). The sample is hybridized to an array. For two-color arrays (e.g., NimbleGen), control DNA (typically input DNA) is labeled with a different dye relative to the ChIP DNA (e.g. Cy3 vs Cy5) and hybridized on the same array. For one-color arrays (e.g., Affymetrix), ChIP and control DNA are

hybridized separately. The data are normalized and scored using one of many algorithms (ENCODE consortium, 2011).

Replicates: In general, the observed enrichments for ChIP-chip are lower than for ChIP-seq (Euskirchen et al. 2007). As such, more replicates may be needed for ChIP-chip than ChIP-seq for some factors that give weaker signals, especially for larger genomes in which there is higher noise due to cross-hybridization. At least two replicates should be done initially but a third is recommended if the first two are not highly concordant. The ChIP-chip experiments in modENCODE/ENCODE were done on multiple array platforms and an appropriate consistency criteria was developed for each platform. In general, the same criterion mentioned above for ChIP-seq (overlap of 80% for the top 40% of the targets) can be used here.

### **IIc. 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.

### **IIId. ChIP-seq and ChIP-chip Experiment Validation**

Methods orthogonal to ChIP-seq and ChIP-chip can be used to test the biological validity of results obtained by ChIP. Furthermore, concordant observation of ChIP peak signals using alternative detection methods raise confidence that the detected events are real. Validation using each class of method is recommend but not required.

#### ***Orthogonal Methods***

Validation is first performed by comparison with known targets already identified by other means. Orthogonal methods that are capable of assessing some targets identified by ChIP include: knockdown/knockout experiments followed by gene expression to search for altered expression, transient reporter assays in conjunction with knockout/knockdowns or mutation of binding sites, one hybrid experiments, DamID, and *in vitro* binding experiments. However, each of these methods does not directly assess binding *in vivo* and has limitations. Many targets identified by ChIP do not exhibit alterations in gene expression upon knockdown of the factor of interest, potentially because the presence of redundant factors or the lack of function at those sites. Thus only a positive result is interpretable.

#### ***Other Detection Methods***

One approach to validating the information from ChIP experiments is ChIP itself and use of alternative detection platforms. These platforms include quantitative PCR (qPCR), ChIP-chip (for ChIP-seq experiments, arrays representing 1% of the genome or more should be used) and ChIP-seq (for ChIP-chip experiments). It is generally recognized that ChIP-seq experiments are more sensitive and more specific than ChIP-chip (Robertson et al., 2007; Liu et al., in preparation). As such, more targets are identified with ChIP-seq and only the top ChIP-seq targets will generally validate using ChIP-chip.

For qPCR, it is recommended that 24-48 targets be selected from a range of p-values. At least six must be negative control regions where no significant enrichment is indicated by the ChIP-seq or ChIP-chip data. Immunoprecipitation relative to control IgG immunoprecipitation or reverse crosslinked "input" DNA should be performed and experiments should be performed in triplicate (duplicate experiments are acceptable if the results lie within 15% of one another). The false positive rate and accuracy can be assessed across the thresholds and used to help set the

scoring threshold. For ChIP-chip experiments, a q-value of less than 0.05 usually yields greater than two-fold enrichment by qPCR. Note that assessment of false negative cannot be readily be determined using qPCR as large numbers of negative regions would need to be assessed. False negatives can be deduced from existing known targets, if this information is available.

Currently, the ENCODE consortium is compiling a summary of efforts to quantitatively characterize ChIP-seq platforms using alternate detection strategies, primarily qPCR and nanostring. The tiling microarray platform has already been thoroughly and quantitatively analyzed (Euskirchen et al., 2007; Liu et al., 2011).

### **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.

#### *Controls*

Deeply sequenced reference samples such as input DNA exhibit uneven coverage. For example, peaks in promoters have been observed in some input samples, perhaps as the result of endogenous nuclease activity, or sonication and solubility biases (Auerbach et al., 2009, Giresi, et al., 2007). These promoter peaks likely represent real open chromatin and therefore should not be excluded from analysis. Other reasons for uneven input signals are copy number variation, and under-representation of repetitive DNA sequence in the reference genome. However, the impact of uneven coverage in input chromatin is limited. Advances in computational methods to correct for such features are being incorporated into the analysis. For example, using reads that are not in peaks, the DNase/FAIRE-seq data itself can be used to identify regions that exhibit copy number variations in samples. In addition, the true signals from FAIRE and DNase exhibit a unique structure that differs greatly from the type of signal produced by uneven input coverage. 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.

#### *Sequencing Depth.*

Since DNase and FAIRE data represent a continuum of the degree to which chromatin is “open”, 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 by other methods such as tiled arrays (Giresi 2009, Giresi et al., 2007, Sabo et al., 2006, Crawford et al., 2006), qPCR (Boyle et al., Cell 2008), or Southern blots (Sabo et al., 2006). For DNase and FAIRE this is typically 20-50 million reads. In general, it is best to sequence replicates to a similar depth. We have found that similar sequence depth matters most for replicates on the lower end of the recommended read depth.

#### *Number of Replicates.*

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. For DNase, we recommend that at least 80% of the top 50,000 peaks in one replicate are detected in the top 100,000 peaks in the second replicate, and vice-versa. For FAIRE, we recommend that at least 50% of the top 50,000 peaks in one replicate are detected in the top 200,000 peaks in the second replicate, and vice-versa.

#### *Scoring.*

Similar to ChIP-seq, a variety of peak calling methods can be used to score peak intensity, including Fseq (Boyle et al., Bioinformatics 2008), Hotspot, and others. The following suggestions can be used to identify a statistical significant cutoff by one of the following methods. 1) Fitting the data to a gamma distribution to calculate p-values, and contiguous regions where p-values were below a 0.05 threshold can be considered significant. 2) Irreproducible discovery rate (IDR) analysis described in section IIb above.

#### **IV) DNA Methylation Standards**

Three methods have been used 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 mapped 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 10-fold 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 10-fold coverage in both replicates.

##### **2. Methyl27**

To perform Methyl27, 1 µg genomic DNA is treated with bisulfite and then hybridized to a Illumina Methyl27 microarray. DNA methylation levels are then determined by the manufacturer's recommendation. Appropriate quality control is done on each array, including the requirement of an initial call at least 90% of array probes. 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, Methyl-seq is mentioned as there are several useful 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**

In order to facilitate data sharing among laboratories both within and outside the consortium, and to ensure that the results can be reproduced, requirements have been established for data sharing.

### Storing High Throughput Sequencing Data

- Image files from sequencing experiments do not need to be stored.

### Submitting mod/ENCODE Data

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

- Raw data should be submitted to both GEO and the DCC. For modENCODE, the DCC is responsible for submission of data to GEO.
- 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 both GEO/SRA and the DCC.
  - For modENCODE, the DCC is responsible for submission of data to GEO.
  - 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.
- Data should be flagged as part of the mod/ENCODE project through the use of the appropriate genome project ID
- Each replicate should be submitted independently
- 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).
  - Interpreted data signal

- Called peaks (see below)
- FASTQ files
- 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").

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

### *Metadata*

The DCC requires submission of basic experimental data, including minimally the following:

1. Investigator, organism or cell line, experimental protocol (or reference a known protocol).
2. Indication as to whether an experiment is a technical or biological replicate (at least two must be performed).
3. Provide the catalog and lot number for any antibody used. If not a commercial antibody, indicate the precise source of the antibody.
4. Provide information used to characterize the antibody, including summary of results (images of immunoblots, immunofluorescence, etc. )
5. Peak calling algorithm and parameters used, including threshold, Reference genome used to map peaks.

6. Summarize the number of reads and number of targets for each replicate and for the merged dataset.
7. Criteria that were used to validate the quality of the resultant ChIP-seq data. Provide the overlap results (using the top 40% rule).
8. Downstream validation results.(e.g. qPCR)
9. Link to the control track that was used.
10. If the experiments fails to meet any of the standards an explanation is required.

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.

## Appendix I. Recommended Standards for 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 <sup>3</sup> 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 uniquely mapped reads should be sequenced per replicate for mammalian experiments, with a minimum of 2 million mapped reads per replicate for worm or fly. (80% should be distinct (i.e. nonidentical). Note that older data sets did not require this depth of sequencing and were acceptable under standards at the time.

## APPENDIX II. Quality control measures for ChIP-Seq experiments

A number of additional parameters can provide important insight into the quality of an individual ChIP-Seq replicate in the absence of a second replicate for comparison. Five of these are described in this appendix. In addition, the recently developed Irreproducible Discovery Rate (IDR) method provides a formal means of assessing replicate agreement and is also described here. None of these methods are currently required to be carried out and reported for data submitted to ENCODE, but they are practiced internally by some or all ENCODE ChIP-Seq production groups.

1) **Visual inspection.** In a typical ChIP experiment, mapped reads are uploaded into a "genome browser" and examine for enriched peaks relative to those for control reads. Biological replicate experiments should generally appear similar.

2) **Library complexity.** Most library preparations involve some level of DNA amplification. Overamplification of specific genome regions due to either biases or too little DNA in the initial preparation both do not provide enough information to identify all true binding sites and can artificially amplify some regions giving the impression of peaks. Therefore it is important to both visually inspect peaks as well as determine the library complexity to ensure that they do not result from overamplification artifacts.

3) **Peak enrichment.** Library enrichment refers to the number of reads that reside in the peak regions. In general, enrichment values <1% may indicate a poor ChIP experiment, especially where a factor is known to have many binding sites. However, there are multiple examples of factors that have few binding sites and thus generate low enrichment values even in successful ChIP experiments.

4) **Cross correlation analysis.** Several important issues with cross correlation analysis are (Kundaje et al., 2011):

- a. High quality ChIP-seq experiments show a predominant peak in the cross-correlation function at the characteristic shift of 'd' (fragment length peak) [Fig. 1]. However, sequenced



input DNA experiments typically show a predominant peak in the cross-correlation function at a shift equal to the sequenced tag length (phantom peak) [Fig. 1]. These peaks should be distinct from those observed in control experiments. For any ChIP-seq experiment, the normalized cross-correlation at the fragment length peak as well as its value relative to the height of the phantom peak in cross-correlation profile provides an intuitive and useful measure of data quality (Kundaje et al. 2011). We call these measures the normalized strand-cross correlation (NSC) coefficient and the relative strand cross correlation (RSC) coefficient respectively.

$$NSC = cc[fraglen] / min(cc)$$

$$RSC = ( cc[fraglen] - min(cc) ) / ( cc[readlen] - min(cc) )$$

where  $cc[.]$  is the cross-correlation value as a function of varying strand shifts,  $fraglen$  is the predominant fragment length,  $readlen$  is the read length.

b. High quality ChIP-seq datasets for punctate binding factors and histone modifications (such as H3k4me1/2/3, H3k27ac and H3k9ac) typically have NSC and RSC significantly greater than 1 whereas low quality datasets (including input experiments) have NSC values very close to 1 and RSC values smaller than 1 and close to 0.

c. A ChIP-seq experiment for a transcription factor that binds very few sites (< 500) in the genome and has a moderate non-specific binding component can also have a low NSC/RSC value. Hence, it is important to consider the biology of the factor in the relevant cellular context.

e. We also consider the library complexity alongside the phantom peak coefficient before interpreting the quality of a ChIP-seq dataset. A dataset with low library complexity and high PCR bottlenecking can show a high phantom peak coefficient due to the artificially high fragment clustering.

f. Cross-correlation profiles of ChIP-seq datasets for histone modifications that are characterized by broad domains (> 5kb) of enrichments (such as H3k27me3, H3k36me3 and H4k20me1) typically exhibit a weak and dispersed fragment length peak. NSC values tend to be < 1.1 and RSC values tend to be less close to 1. This is expected due to dispersed fragment clustering. Also, significantly deeper sequencing is required to achieve a signal to noise ratio equivalent to that of punctate binding factors.

g. The cross-correlation analysis has an added benefit that it can be used to detect problems with size selection and also be used to determine optimal sequencing depth for ChIP-seq datasets [Kundaje et al. 2011]

#### *IDR analysis.*

The IDR method (Li et al., 2011) includes three components:

1. A correspondence curve: It is a graphical tool for visualizing and inspecting the reproducibility of the replicates and the transition from reproducible to irreproducible signals. It provides a quick graphical view of the data without making any model assumptions. It is a convenient tool for diagnosis and quality control, but not adequate for selecting signals. It is independent of the other two components
2. An inference procedure: It quantitatively summarizes the consistency structure, including the proportion of reproducible and irreproducible signals, how strong the association between replicates is for the reproducible group, and how separate the reproducible group is from the irreproducible group. It also assigns each signal a probability to be from the irreproducible group which is the local IDR score.

3. Irreproducible Discovery Rate (IDR): It is a reproducibility criterion derived from the inference procedure (#2) in a fashion similar to FDR, and can be used to control the level of irreproducibility rate when selecting signals.

A valuable advantage of the IDR method for thresholding datasets is that one can select a single IDR threshold (1%) over datasets of varying quality. False discovery rate (FDR) thresholds that are typically used for peak calling need significant fine tuning to extract the optimal amount of signal from datasets of varying quality. FDR thresholds can also be quite unstable i.e. small changes in the FDR threshold can result in large changes in the number of peaks selected. The IDR thresholds on the other hand tend to exhibit smoother behavior and so selection of an optimal threshold does not require much tuning. Also, a particular FDR threshold applied to a dataset using different peak callers can translate into vastly different number of selected peaks. However, this is typically because different peak callers often use different types of null models to compute measures of significance. The IDR method works internally with peak ranks and is hence very robust to different ranking measures. Multiple peak callers (such as SPP and MACS) select very similar number of overlapping peaks for a defined IDR threshold.

1. Auerbach RK, Euskirchen G, Rozowsky J, Lamarre-Vincent N, Moqtaderi Z, Lefrançois P, Struhl K, Gerstein M, Snyder M. Mapping accessible chromatin regions using Sono-Seq. *Proc Natl Acad Sci U S A*. 2009 Aug 18. [Epub ahead of print] PMID: 19706456
2. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. *Cell*. 2008 Jan 25;132(2):311-22. PMID: PMC2669738
3. Boyle AP, Guinney J, Crawford GE, Furey TS. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics*. 2008 Nov 1;24(21):2537-8. Epub 2008 Sep 10. PMID: PMC2732284
4. Crawford GE, Davis S, Scacheri PC, Renaud G, Halawi MJ, Erdos MR, Green R, Meltzer PS, Wolfsberg TG, Collins FS. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat Methods*. 2006 Jul;3(7):503-9. PMID: PMC2698431
5. Euskirchen GM, Rozowsky J, Wei CL, Lee WH, Zhang ZD, Hartman S, Emanuelsson O, Stolc V, Weissman S, Gerstein M, Ruan Y, Snyder M. Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. *Genome Res*. 2007;17: 898-909.
6. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res*. 2007 Jun;17(6):877-85. PMID: PMC1891346
7. Giresi PG, Lieb JD. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods*. 2009 Jul;48(3):233-9. PMID: PMC2710428
8. The ENCODE Project Consortium. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol*. 2011 Apr; 9(4)
9. Iyer VI,\* Horak CA,\* Scafe CS, Botstein D, Snyder M, Brown PO. Genomic binding distribution of the yeast cell-cycle transcription factors SBF and MBF. *Nature*. 2001;409: 533-538.
10. Johnson DS, Li W, Gordon DB, Bhattacharjee A, Curry B, Ghosh J, Brizuela L, Carroll JS, Brown M, Flicek P, Koch CM, Dunham I, Bieda M, Xu X, Farnham PJ, Kapranov P, Nix DA, Gingeras TR, Zhang X, Holster H, Jiang N, Green R, Song JS, McCuine SA, Anton E, Nguyen L, Trinklein ND, Ye Z, Ching K, Hawkins D, Ren B, Scacheri PC, Rozowsky J,

- Karpikov A, Euskirchen G, Weissman S, Gerstein M, Snyder M, Yang A, Moqtaderi Z, Hirsch H, Shulha HP, Fu Y, Weng Z, Struhl K, Myers RM, Lieb JD, Liu XS. Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets. *Genome Res.* 2008. Mar;18(3):393-403. Epub 2008 Feb 7. PMID: 18258921.
11. Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol.* 2008 Dec;26(12):1351-9. Epub 2008 Nov 16.
  12. Kundaje A, Jung YL, Kharchenko PV, Sidow A, Batzoglou S, Park PJ. (2011) Assessment of ChIP-seq data quality using strand cross-correlation analysis, Submitted
  13. Li Q, Brown JB, Huang H, Bickel PJ. (2011) Measuring reproducibility of high-throughput experiments, *Annals of Applied Statistics*, In press.
  14. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods.* 2007;4: 651-7.
  15. Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, Bjornson R, Carriero N, Snyder M, Gerstein MB. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol.* 2009 Jan;27(1):66-75. Epub 2009 Jan 4. PMID: 19122651.
  16. Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldy J, Haydock A, Weaver M, Shafer A, Lee K, Neri F, Humbert R, Singer MA, Richmond TA, Dorschner MO, McArthur M, Hawrylycz M, Green RD, Navas PA, Noble WS, Stamatoyannopoulos JA. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods.* 2006 Jul;3(7):511-8. PMID: 16791208

## Figures

Fig. 1 Crosscorrelation analysis for assessing the quality of a ChIP-seq experiment.

F). The absolute and relative height of the two peaks are useful determinants of the success of a ChIP-seq. A high-quality IP is characterized by a ChIP peak that is much higher and large bp than the low bp peak; in successful IP experiments often lvery small low bp peaks (left) or no peaks are detected. There is considerable continuity between these extremes.

