Fully automated high-throughput chromatin immunoprecipitation for ChIP-seq: Identifying ChIP-quality p300 monoclonal antibodies

William C. Gasper†, Georgi K. Marinov¹, Florencia Pauli-Behn², Max T. Scott¹, Kimberly Newberry², Gilberto DeSalvo¹, Susan Ou¹, Richard M. Myers², Jost Vielmetter* & Barbara J. Wold¹*

¹Division of Biology and Bioengineering, California Institute of Technology, Pasadena, CA 91125, USA, ²HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA.

Chromatin immunoprecipitation coupled with DNA sequencing (ChIP-seq) is the major contemporary method for mapping in vivo protein-DNA interactions in the genome. It identifies sites of transcription factor, cofactor and RNA polymerase occupancy, as well as the distribution of histone marks. Consortia such as the ENCylopedia Of DNA Elements (ENCODE) have produced large datasets using manual protocols. However, future measurements of hundreds of additional factors in many cell types and physiological states call for higher throughput and consistency afforded by automation. Such automation advances, when provided by multiuser facilities, could also improve the quality and efficiency of individual small-scale projects. The immunoprecipitation process has become rate-limiting, and is a source of substantial variability when performed manually. Here we report a fully automated robotic ChIP (R-ChIP) pipeline that allows up to 96 reactions. A second bottleneck is the dearth of renewable ChIP-validated immune reagents, which do not yet exist for most mammalian transcription factors. We used R-ChIP to screen new mouse monoclonal antibodies raised against p300, a histone acetylase, well-known as a marker of active enhancers, for which ChIP-competent monoclonal reagents have been lacking. We identified, validated for ChIP-seq, and made publicly available a monoclonal reagent called ENCITp300-1.

Contemporary studies of gene regulation are often based, at least in part, on learning the patterns of chromatin mark distribution and the locations of specific transcription factor occupancy in the genome. The chromatin immunoprecipitation (ChIP) assay, in several variations, provides this information1-3. ChIP protocols typically begin by cross-linking proteins to DNA (usually with formaldehyde); then selectively retrieving DNA fragments associated with a protein of interest by immunoprecipitation; and finally analyzing the enriched DNA. Originally, ChIP-enrichment was analyzed using qPCR at predefined genomic regions⁴. Later, it was coupled with microarray readouts (ChIP-chip/ChIP-on-chip) which allowed many selected regions to be assayed in parallel (e.g. all promoters) or even whole genomes, especially in organisms with small genomes⁵-⁹. Eventually, high-throughput sequencing enabled truly genome-wide mapping of protein-DNA interactions, with high resolution, in the form of ChIP-seq¹⁰-¹⁴.

ChIP-seq has become the workhorse for mapping the whole-genome occupancy and genomic distribution of hundreds of transcription factors and numerous histone modifications in a wide variety of human, mouse, and worm cell lines and tissues by the ENCODE¹⁵-¹⁸, mouse ENCODE¹⁹ and modENCODE consortia²⁰,²¹, and the NIH Roadmap Epigenomics Mapping Consortium²². Despite the large number of datasets generated thus far, they are a small fraction of the expected future experiments from individual laboratories as well as consortia. Initially, DNA sequencing capacity and cost were major barriers to large scale ChIP-seq, but sequencing capacity has increased by several orders of magnitude and costs per ChIP have dropped significantly. The immunoprecipitation step has now emerged as rate-limiting. It is tedious, and in practice it is often variable from one practitioner to another, from experiment to experiment, and even among replicates in a single experiment. This suggested that a robust robotic ChIP protocol could stabilize and improve data quality, reproducibility, manpower use, and overall costs and efficiency per experiment. An automated system would offer these benefits.
to individual laboratories doing small numbers of experiments, through core facilities, in addition to enabling large-scale projects and consortia.

A second independent challenge for contemporary ChIP-seq experiments is that the supply of high-quality sustainable immune reagents that have been experimentally validated for ChIP remains very limited. Many antibodies, including some marketed as “ChIP-grade” have failed in the ENCODE pipeline, and many that have succeeded are polyclonal, which means that different lots can vary radically in how well they perform in ChIP. At present, monoclonal antibodies are the most reliable renewable ChIP reagents, although they do not account for the majority of characterized reagents, and there are no ChIP-competent reagents for the majority of human and mouse transcription factors. The field therefore faces the twin challenges of generating large quantities of ChIP-seq data in reliable high-throughput manner for factors with extant affinity reagents, and having to screen and characterize new sustainable immune reagents.

In this work we develop a fully automated robotic pipeline for the chromatin immunoprecipitation reaction (R-ChIP). High-throughput 96-well plate methods for performing ChIP have been described before. However, those methods require substantial hands-on time and are subject to variability inherent in experiments done by humans. A conceptually similar robotic approach was recently developed independently, though it differs from the one presented here in requiring manual intervention at several steps. The R-ChIP protocol reported here is fully automated and employs a widely used, multipurpose programmable liquid handling robotic platform (Tecan Freedom EVO 200), which can be used for a multitude of other purposes, such as robotic plasmid cloning or automated ELISA screenings when it is not being used for ChIP. We test our protocol on factors that have previously been characterized in multiple ENCODE cell lines and show that it performs comparably to high-quality manual ChIP-seq in enrichment and in producing ChIP-seq libraries that are consistent within and between experiments. We then applied R-ChIP to screen candidate monoclonal antibodies directed against the transcriptional co-activator p300, a protein for which monoclonal ChIP-competent reagents have until now not been available, and for which polyclonal reagent lots have been highly variable.

Results

Automated ChIP protocol adaptations. The primary goal of this work was to fully automate ChIP without compromising yield and quality. Our design approach was to develop automation that mimics as closely as possible the established manual process, using the ENCODE ChIP protocol as the starting point (the current manual ENCODE ChIP protocol is provided in the supplementary Appendix). Where substantial changes were made to accommodate automation, we benchmarked the new process against results from the established protocol.

We configured a Tecan Freedom EVO 200 robot as detailed (Figure 1 and Supplementary Figure 2), and programmed it for running automated chromatin immunoprecipitation reactions in a 96-well format (the program itself is supplied as a supplementary file). Major considerations for automating ChIP revolved around magnetic bead-handling to achieve successful incubation, washing, and recovery of immunoprecipitated material, while effectively eliminating unbound chromatin. In the manual version of ChIP, bead agitation is achieved by tumbling the reaction mix in standard Eppendorf 1.5-mL micro tubes on a tumbler wheel. The agitation device available on the robot is an orbital shaker with a 2-mm shake radius and adjustable speeds ranging from 100 rpm to 1600 rpm. An alternate method for automated bead agitation mixes by repeated pipetting (trituration). We reasoned that pipetting would lead to inevitable bead losses to the pipette surface, especially as multiple tip changes would be required. We therefore focused on the orbital shaker. The second automation constraint comes from the 96-well plate format compared with individual microtubes in the manual protocol. This change requires effective robotic washing without cross-contamination between wells or sample spillage.

Finally, the 96-well format requires a plate magnet strong enough to efficiently pull down all beads. Several vendors offer plate magnets compatible with the robot platform, but most are designed for standard low profile micro plates. In contrast, our automated ChIP protocol requires deep-well plates for effective bead washing. A magnet designed specifically for deep-well plates (SPRIPlate Super Magnet Plate from Agencourt, Beckman Coulter) proved effective. Its success in our hands was optimal with a round-well-deep well plate with U-bottom wells (catalog # 278752, Nunc). A summary of major differences for the robotic protocol is below and both protocols are given in detail in the Supplementary Methods:

1. Bead agitation was changed from a tumbling motion in the manual protocol to rapid orbital shaking. The shake speed was optimized to keep beads fully suspended without spillage (1400 rpm).
2. The sample volume was reduced from 1000 μL to 500 μL to prevent spillage.
3. Wash steps after antibody and chromatin binding were increased in number from 3 to 4 to compensate for the smaller wash volume.
4. Bead recovery time on the magnet was extended to 7 min on the robot, a condition determined empirically using the criterion that no detectable beads were left behind in the supernatant upon microscopic inspection.
5. The step that dissociates bound chromatin from antibody-magnetic beads is done in the robot’s PCR module, thus eliminating constant bead agitation in the manual protocol during the 65 °C 1 hour incubation. The beads are now suspended by the orbital shaker after the 1 hour incubation.

Consistency of robotic ChIP results. We first tested the robustness and reproducibility of our robotic ChIP protocol by carrying out multiple manual and R-ChIP experiments for the NRSF/REST transcription factor. NRSF/REST is a negative transcriptional regulator of neuronal genes in non-neuronal cell types. It was the first transcription factor to which ChIP-seq was applied, its binding has been extensively mapped in multiple cell lines, and its recognition site (and its binding variants) is well studied. The monoclonal antibody used for NRSF ChIP has been well characterized in the ChIP-seq format. It is thus a good system to characterize the method.

We performed ChIP-seq experiments in two cell lines, GM12878 and Jurkat, producing at least three libraries from four separate plates for GM12878 and from four separate plates for Jurkat. We compared the resulting data to existing manually generated NRSF ChIP-seq datasets for GM18278 cells and to four additional manual ChIP-seq datasets generated in parallel with the R-ChIP ones. These data are summarized in Figure 2.

To assess ChIP quality, we used library and ChIP QC metrics that were developed previously by us and others as part of the ENCODE Consortium. The first question regarding ChIP quality is how well the immunoprecipitation step has enriched for DNA fragments attached to the antigen of interest. This can be assessed by calculating the fraction of reads falling within called peaks (FRIP) or by using cross-correlation. Both measures have limitations in some special cases, but when both are applied and concur, confidence in the results is high. Figure 2A shows the number of called peaks and Figures 2B and 2C show the RSC (Relative Strand Correlation) and FRIP values for manual and robotic NRSF ChIP-seq datasets. R-ChIP data consistently exhibited good RSC values (RSC ≈ 1) and FRIP and peak number values comparable to those of manually
generated libraries, with the exception of three Jurkat libraries (the first ChIP on plates 2, 3, and 4, Figure 2A, 2B and 2C) that scored as less successful. We do not presently know the cause of these lower-quality libraries, but their frequency is well within the range of variability of manually generated libraries we have observed over several years, during which sporadic unsuccessful experiments for factor/antibody pairs that are otherwise routinely successful have occurred. Finally, we asked how similar the final sets of called peaks are for the robotic protocol and how they compare with reference manual datasets for the same factor and cell type, by evaluating peaks called after sequencing. Figures 2D and 2E show the similarity of peak call sets for all libraries measured by calculating the size of the overlap between each pair of libraries. Overall, we observed consistently high overlap scores and thus high reproducibility between libraries. These observations applied both within and between plates, underscoring the consistency and robustness of the R-ChIP protocol.
Figure 2 | Reproducibility of R-ChIP experiments. Multiple ChIP-seq experiments on multiple plates were generated for the NRSF/REST repressor in GM12878 lymphoblastoid cells (n = 4 plates) and Jurkat T-cells (n = 4 plates) cell lines. The numbers (1 through 5) refer to the number of the plate a library came from, “M” refers to manually generated datasets. The first two manual GM12878 datasets were previously published as part of the ENCODE project, the next four were generated in parallel with the R-ChIP ones. (A) Number of called regions for each dataset (using ERANGE 4.0.11) (B) Assessment of ChIP enrichment using RSC (Relative Strand Correlation) cross-correlation scores; (C) Assessment of ChIP enrichment using FRiP (Fraction of Reads in Peaks) scores (Landt et al. 2012); (D) Overlap between called peaks in robotic and manual ChIP libraries in GM12878 cells; (E) Overlap between called peaks in robotic and manual ChIP libraries in Jurkat cells. The overlap score (OXY) shown in each box indicates the fraction of peaks in the dataset on the y-axis that are also found in the dataset on the x-axis, i.e. $O_{XY} = \frac{|X \cap Y|}{|Y|}$. 
To further characterize the consistency between the results from R-ChIP and manual ChIP experiments, we generated paired manual and robotic ChIP-seq datasets using matched chromatin samples for several additional targets (Figure 3). These included the H3K27ac histone modification (Figure 3A, 3B and 3C), the GABP transcription factor\(^{32-34}\) (Figure 3D, 3E and 3F), the ZBTB33/Kaiso zing-finger protein known for its preferential binding to methylated DNA\(^{35}\) (Figure 3G, 3H and 3I), and the important regulator of hematopoiesis SPI1/PU.1\(^{36,37}\) (Figure 3J, 3K and 3L). We observed comparable results between the manual and robotic datasets, further confirming the applicability of R-ChIP to large-scale ChIP-seq production.

Using R-ChIP to characterize new monoclonal p300 antibodies. Having established the R-ChIP protocol, we next applied it to

Figure 3 | Comparison between manual ChIP and R-ChIP results for additional targets. Manual and robotic ChIP-seq datasets were generated in parallel using the same chromatin samples and the result compared. (A,B,C) ChIP-seq against H3K27ac in GM12878 cells. (A) FRIP score, (B) number of peaks called, (C) overlap between the sets of peaks; (D,E,F) ChIP-seq against GABP in GM12878 cells. (D) FRIP score, (E) number of peaks called, (F) overlap between the sets of peaks; (G,H,I) ChIP-seq against ZBTB33 in GM12878 cells. (G) FRIP score, (H) number of peaks called, (I) overlap between the sets of peaks. The overlap score \(O_{XY}\) shown in each box indicates the fraction of peaks in the dataset on the y-axis that are also found in the dataset on the x-axis, i.e. \(O_{XY} = |X \cap Y|/|Y|\).
characterize a set of monoclonal antibodies raised against the p300 transcriptional co-activator in the Beckman Institute Protein Expression Center and Monoclonal Antibody Facility at Caltech. The p300 protein is a histone acetyltransferase, best known for its role in the acetylation of histones\(^{36-41}\). It is used as a marker of active transcriptional enhancers in mammalian genomes\(^{42-44}\). Commercially available antibodies used to generate published p300 data are from a series of polyclonal nonrenewable reagents that have a reputation for lot-to-lot variability.

We generated 11 \(\geq\) p300 mouse monoclonal antibodies which were initially screened, cloned and then rescreened using a plate-based ELISA assay. We tested hundreds of individual hybridoma B-cells isolated after fusion with cells from spleens of mice injected with a GST-tagged p300 protein fragment (aa 152–213) or a synthetic KLH-coupled peptide (aa 1526–1545). The GST-tagged preparations were subjected to formaldehyde fixation (1% FA for 10 min) before immunization with the goal of increasing the likelihood of reactivity in ChIP. The resulting 11 p300 monoclonal antibodies were tested for ChIP together with two lots of rabbit polyclonal p300 antibodies (Santa Cruz sc-585, lot numbers F2711 and E3113) on chromatin from GM12878 cells. The resulting datasets were compared to each other and to publicly available ENCODE p300 data from the same cells (using two commercially available rabbit polyclonal antibodies, Santa Cruz sc-585 and sc-584) (Figure 4). Three of the mouse monoclonal antibodies raised against the synthetic KLH-coupled p300 peptide scored positive by ChIP-seq, identifying between 1,524 and 4,870 peaks (Figure 4A and 4B).

We sequenced multiple additional replicates for the best-scoring one, 1F4-E10P and identified an even higher number of peaks in some of the datasets, up to 8,430, with the typical number being \(\sim\)6,000. The peaks called in the monoclonal antibody dataset are a subset of those found in the polyclonal data (Figure 4C) confirming the specificity of the antibodies towards p300. While the monoclonal numbers are lower than the two most successful polyclonal datasets, they are within the range of what was previously observed in ENCODE data, and also within the range of published p300 datasets. We have renamed p300 1F4-E10P as ENCITp300-1 and submitted it under this name to the Developmental Studies Hybridoma Bank (http://dshb.biology.uiowa.edu/).

It was not our purpose in this study to characterize new polyclonal reagent lots, but the p300 reagents used previously by ENCODE were no longer available. We therefore used two additional rabbit polyclonal antibodies in R-ChIP (Santa Cruz sc-585, lot numbers F2711 and E3113), and they identified up to \(\sim\)30,000 peaks. This number greatly exceeds previously published p300 datasets, including currently available ENCODE data for the same GM12878 B-cell line (for which between 2,610 and 12,924 peaks were called previously) (Figure 4A and 4B). This increase has two likely causes, and they are not mutually exclusive. The first well-appreciated variable is different performance by polyclonal antibody lots. In principle, individual lots can vary in the number and identity of epitopes recognized, in effective antibody concentration, and in non-specific reactivity. A second difference from the prior ENCODE data is the fixation condition. For p300, we fixed cells at 37°C for 30 minutes, versus room temperature for 10 minutes for the historic ENCODE data. This condition was suggested to us by Dr. Bing Ren, UC San Diego, and is based on the idea that a longer time and elevated temperature might increase p300 cross-linking via indirect links to DNA-bound transcription factors or histones. This condition significantly improves p300 ChIP in our hands (for example, we generated datasets using the 1F4-E10P antibody on chromatin fixed under standard conditions and they were all unsuccessful; Supplementary Figure 4). Of the 30,000 p300 peaks called using the 37°C fixation condition, the majority (between 76% and 88%) overlap with one or more chromatin marks associated with enhancers and promoters in ENCODE data (H3K27ac; H3K4me1) or with regions of DNase hypersensitivity (Supplementary Figure 3). This is consistent with these regions being active enhancers and promoters. For multiple cell types, the numbers of DNase hypersensitive regions\(^{46-47}\), H3K27ac and H3K4me1 positive regions, reported previously are typically in the tens of thousands (ENCODF Project Consortium 2012), and the number of expressed genes per cell type is between 5 and \(10^5\). Thus the expected number of enhancers (and potentially p300-positive regions) is larger than the single-digit thousands of p300 peaks called in most previously available data. While reagent-specific background, including possible polyclonal cross-reactivity, could explain the modest number of p300 peaks that lack additional enhancer or promoter marks, the most parsimonious explanation for the overall very large number of newly revealed p300 peaks is that prior ChIP measurements have underestimated p300 occupancy. The monoclonal peaks also correspond preferentially to the strongest sites called for the polyclonals under the 37°C fixation condition (Supplementary Figure 8). Our best-performing monoclonal antibody did not produce comparably high peak numbers using the same chromatin substrate, but 99% of its peaks overlap those called in the polyclonal datasets. We note that the polyclonal antibody lots used here are also no longer in stock from the manufacturer.

We then tested additional factors with the 37°C fixation condition. Results were very similar to those with the standard condition for NRSF, H3K27ac and GABP (Supplementary Figure 5, Supplementary Figure 6A–F), suggesting that the more aggressive fixation condition does not result in general nonspecific background. However, for PU.1 and ZBTB33, the 37°C fixation led to fewer peaks and overall lower quality ChIP data (Supplementary Figure 6G–I). As discussed below, our data suggest that the impact of fixation condition on ChIP outcome might be more complex than is generally acknowledged, and that a wider survey of factor-antibody pairs relative to fixation conditions will be needed.

**Discussion**

The robotic ChIP (R-ChIP) reported here was developed on a widely used commercial liquid-handling platform whose configuration and running program for ChIP are provided. It was developed to increase ChIP-seq throughput, uniformity, and quality, while reducing investigator tedium and error. It should be equally applicable for large-scale projects or for smaller, individual-lab projects carried out in multiuser core facilities.

Our R-ChIP results were comparable in quality to those from a large-scale manual pipeline by multiple metrics. However the platform’s performance is not perfect. We expect that we and others will continue to make improvements. Specifically, we have observed sporadic single reaction failures among replicate samples within an individual plate, possibly associated with edge and corner well positions (Supplementary Figure 7), but sporadic failures are also observed with manual experiments. It is our standard practice to include on each R-ChIP plate a minimum of triplicate control samples deployed across the plate geometry. For these controls, we use a monoclonal reagent and a large batch of control chromatin that permits comparisons between plates through time. This allows us to evaluate each plate run and to compare it with other runs. This evaluation can be done as a QC step before committing to building and sequencing all remaining libraries. This gating step has clear economic implications.

ChIP troubleshooting is aided by R-ChIP. If a group of failed samples are embedded in a large R-ChIP run where the controls and other samples are successful, it becomes unlikely that the ChIP process is the source of failure, and a user can turn attention to the input sample and immune reagent (or any post-ChIP variation) as a more likely problem. Of course, the overall success of a ChIP-seq measurement includes the local DNA sequencing process, with spe-
Specific sequencing protocols being differentially sensitive to the mass, fragment size, and other characteristics of a ChIP sample.

We used R-ChIP to screen for monoclonal ChIP-competent antibodies for the histone acetyltransferase, p300. Many antibodies made against transcription factors fail in ChIP reactions, even though they work well in one or more conventional uses (e.g. standard immunoprecipitation, western blots or immunocytochemical stains). Moreover, polyclonal reagents that are ChIP-seq compatible typically vary, sometimes greatly, in their specificity and performance from lot to lot\(^23\). The upshot has been that the only way to identify a ChIP-quality antibody is to test it directly for ChIP, and the most general way to ensure reliability and unlimited supply is with a monoclonal. Whether the final readout for ChIP competence is DNA sequencing or qPCR (the latter requiring known targets for the factor), the capacity to test many ChIP reactions is critical for screening. The ChIP peaks identified by our top monoclonal, ENCITp300-1 (originally called 1F4-E10P during screening) overlap highly with prior measurements from ENCODE for the same cell line and with concurrent polyclonal determinations, confirming its specificity for p300 and the utility of R-ChIP for screening new immune reagents for ChIP. This reagent is being made publicly available at the Developmental Studies Hybridoma Bank. We note, however, that the data obtained with it are not as inclusive as the best datasets produced using polyclonal rabbit reagents.

Figure 4 | Characterization of novel monoclonal p300 antibodies using robotic ChIP. ChIP-seq against p300 was carried out in GM12878 cells and prior ENCODE data for it in that cell line (from the “SYDH” production group) was used as a reference. ENCODE data was generated using two different rabbit polyclonal antibodies from Santa Cruz (sc-584 and sc-585). We carried out robotic ChIP testing of two different lots of the sc-585 antibody and 11 different monoclonals we raised. The 1F4-E10P clone scored best and additional replicates were generated in subsequent experiments. (A) Number of called regions; (B) ChIP enrichment as measured by FRIP scores (Landt et al. 2012); (C) Overlap between called peaks with different antibodies. The overlap score \(O_{XY}\) shown in each box indicates the fraction of peaks in the dataset on the \(y\)-axis that are also found in the dataset on the \(x\)-axis, i.e. \(O_{XY} = |X \cap Y|/|Y|\); (D) Representative browser snapshot of p300 ChIP enrichment in polyclonal and monoclonal datasets around the IL13 and IL4 locus.
DNA cleanup. Samples from the R-ChIP ChIP experiments presented here were then cleaned up manually using the protocol described by Qiagen in their Quickamp PCR purification kit with the adjustment that the UB buffer is heated to 55°C prior to elution and eluted in a 50 μL volume using DNA Lo-bind 1.5 μL tubes from Eppendorf. We anticipate automating this step.

Library building and sequencing. Library building for sequencing on the Illumina HiSeq platform was performed conventionally with barcoding to allow multiple ChIP libraries to be sequenced in a single cell line according to the ENCODE ChIP protocol (Supplementary Methods). Standard methods were used for end repair and A-d addition of DNA fragments recovered from ChIPs or chromatin controls. The fragments were then ligated to Illumina Paired-End adapter sequences and PCR-amplified to complete the adapter sequences and introduce a 7-bp DNA barcode in the i7 position. The fragments were then linearized using the restriction enzyme Mbol and the umbilical cord DNA was purified on an agarose gel and a Tecan 96-well plate. The barcode in the i7 position was sequenced on the Illumina HiSeq2000 or HiSeq2500 with 50 bp single-end reads following the manufacturer’s recommendations. Raw sequencing reads are available from GEO accession number GSE53366.

Data processing and analysis. Reads were aligned using Bowtie2 version 0.12.7, with the following settings: "-v 2 -t 2 -k 2 -m 1 --best --strata", which allow for two mismatches relative to the reference and only retain unique alignments, against the hg19 version of the human genome (assembly downloaded from the UCSC genome

For p300 R-ChIP, we used chromatin from cells fixed with a modified condition (37°C, 1% formaldehyde, 30 min; personal communication, Bing Ren, UC San Diego). This improved p300 ChIP significantly in our hands compared with our standard fixation condition (see Results and Methods), while it had no detectable effect on NRSF, GABP and H3K27ac ChIP. We also observed that the PU.1 and ZBTB33 ChIP-seq datasets we generated were not as good with 37°C chromatin as with the standard condition. Until more factors are tested directly under both fixation conditions, we cannot predict which antibody-factor pairs will be affected or how. Epitope destruction or occlusion, or elevated signals from lower affinity interactions, are among the plausible negative effects. The most positive impact is expected for proteins that interact indirectly with DNA, as p300 is understood to do.

We anticipate that R-ChIP, and variations on it, will advance a wide range of functional genomics studies by relieving the emerging data production bottleneck, increasing efficiency and improving overall data quality.

Methods

Cell growth and harvesting. Cells were grown and harvested following established ENCODE protocols (available at http://genome.ucsc.edu/ENCODE/cellTypes.html) with the exception of GM1287 p300 experiments for which chromatin was fixed at 37°C for 30 minutes.

Chromatin preparation. Chromatin was cross-linked by adding formaldehyde directly to the cell culture media at a final concentration of 1% and gently mixed for 10 minutes. The exception was when (where indicated) fixation at 37°C for 30 minutes, which was used for p300 experiments. In all cases, the formaldehyde reaction was quenched by adding glycine to a final concentration of 0.125 M for 10 minutes. Cells were then pelleted, rinsed once in cold phosphate-buffered saline (PBS) with 1 mM PMSF and once in cold MC lysis buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM MgCl2, 0.5% NP-40, and Roche Complete Protease Inhibitor Cocktail) to obtain nuclear pellets. Nuclei were sonicated in RIPA buffer (PBS, 1% NP-40 Substitute, 0.5% Sodium Deoxycholate, 0.1% SDS, and Roche Complete Protease Inhibitor Cocktail) at a concentration of at least 5 x 10^7 nuclei/mL using a probe sonic disseminator from Fischer Scientific (Model 550). To check for fragment size distribution after sonication, a small fraction of the sample was reverse cross-linked for two hours at 37°C, purified using DNA PCR purification from Qiagen, then loaded onto a 2% agarose EtBr E-Gel from Invitrogen (Supplementary Figure 9).

Antibodies used. The following antibodies were used: an α-NRSF mouse monoclonal (12C11) from the Anderson Lab at Caltech1, an α-p300 rabbit polyclonal (sc-585) from Santa Cruz Biotechnology, a mouse monoclonal α-GAB (sc-28312) from Santa Cruz Biotechnology, a mouse monoclonal α-Zvai/ZBTB33 (sc-23871) from Santa Cruz Biotechnology, a rabbit polyclonal α-SPI1/PU.1 (sc-22805) from Santa Cruz Biotechnology, and a mouse monoclonal α-H3K27ac (306–34849) from Wako. In

Library building and sequencing. Library building for sequencing on the Illumina HiSeq platform was performed conventionally with barcoding to allow multiple ChIP libraries to be sequenced in a single cell line according to the ENCODE ChIP protocol (Supplementary Methods). Standard methods were used for end repair and A-d addition of DNA fragments recovered from ChIPs or chromatin controls. The fragments were then ligated to Illumina Paired-End adapter sequences and PCR-amplified to complete the adapter sequences and introduce a 7-bp DNA barcode in the i7 position. The fragments were then linearized using the restriction enzyme Mbol and the umbilical cord DNA was purified on an agarose gel and a Tecan 96-well plate. The barcode in the i7 position was sequenced on the Illumina HiSeq2000 or HiSeq2500 with 50 bp single-end reads following the manufacturer’s recommendations. Raw sequencing reads are available from GEO accession number GSE53366.

Data processing and analysis. Reads were aligned using Bowtie2 version 0.12.7, with the following settings: "-v 2 -t 2 -k 2 -m 1 --best --strata", which allow for two mismatches relative to the reference and only retain unique alignments, against the hg19 version of the human genome (assembly downloaded from the UCSC genome
31. Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq. Cell 149, 17–27 (2012).

32. Collins, P. J., Kobayashi, Y., Nguyen, L., Trinklein, N. D. & Myers, R. M. The es-focused transcription factor GARP binds directly transcriptional. PLoS Genet 3, e208 (2007).

33. Watanabe, H., Wada, T. & Handa, H. Transcription factor E47FT1 contains two subunits with different functions. EMBO J 9, 841–847 (1990).

34. Thompson, C. C., Brown, T. A. & McKnight, S. L. Convergence of Ets- and notch-related structural motifs in a heteromeric DNA binding complex. Science 253, 762–768 (1991).

35. Prokhorchoukh, A. et al. The pi20 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev 15, 1613–1618 (2001).

36. Burda, P., Laslo, P. & Stoppa, T. The role of P.U.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. Leukemia 24, 1249–1257 (2010).

37. Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C. & Maki, R. A. The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. Cell 61, 113–124 (1990).

38. Eckner, R. et al. Molecular cloning and functional analysis of the adenovirus EIA-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. Genes Dev 8, 869–884 (1994).

39. Arany, Z., Sellers, W. R., Livingston, D. M. & Eckner, R. EIA-associated p300 and CREB-associated CBP belong to a conserved family of coactivators. Cell 77, 799–800 (1994).

40. Lundblad, J. R., Kwok, R. P., Laurance, M. E., Marter, M. L. & Goodman, R. H. Adenoviral EIA-associated protein p300 as a functional homologue of the transcriptional co-activator CBP. Nature 374, 85–88 (1995).

41. Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell 87, 953–959 (1996).

42. How, M. J. et al. ChIP-Seq identification of weakly conserved heart enhancers. Nat Genet 42, 806–810 (2010).

43. Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854–858 (2009).

44. Visel, A. et al. A high-resolution enhancer atlas of the developing telencephalon. Cell 152, 895–908 (2013).

45. May, D. et al. Large-scale discovery of enhancers from human heart tissue. Nat Genet 44, 89–93 (2011).

46. Nep, S. et al. An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 489, 83–90 (2012).

47. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. Nature 489(7441), 75–82 (2012).

48. Mortazavi, A., Lee, P. C., Garcia, S. T., Myers, R. M. & Wold, B. Comparative genomics modeling of the NRSF/REST repressor network: from single conserved sites to genome-wide repertoire. Genome Res 16, 1208–1221 (2006).

49. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25 (2009).

Acknowledgments

The authors thank Henry Amrhein, Diane Troun and Sean Upchurch for technical and computational assistance, Dan Savic for discussions and contributions to our manual ChIP
copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/

Additional information

Supplementary information accompanies this paper at http://www.nature.com/

Copyright and Reprints

Published online: 13 May 2014

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. The images in this article are included in the
article's Creative Commons license, unless indicated otherwise in the image credit;
if the image is not included under the Creative Commons license, users will need to
obtain permission from the license holder in order to reproduce the image. To view a

copyright notice as well as the terms and conditions governing copying and
republication of the images, please visit:

www.nature.com/scientificreports