SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
HIGH-THROUGHPUT SCREENING
FUNCTIONAL GENOMICS

Received
26 March 2014

Accepted
13 May 2014

Published
12 June 2014

Correspondence and requests for materials should be addressed to B.J.W. (woldb@caltech.edu)

† Current address: Department of Biology, Boston University, Boston, Massachusetts 02215.

* These authors contributed equally to this work.

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

William C. Gasper[1]†, Georgi K. Marinov[1], Florencia Pauli-Behn[2], Max T. Scott[1], Kimberly Newberry[2], Gilberto DeSalvo[1], Susan Ou[1], Richard M. Myers[2], Jost Vielmetter[1]* & Barbara J. Wold[1]*

[1]Division of Biology and Bioengineering, California Institute of Technology, Pasadena, CA 91125, USA, [2]HudsonAlpha 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 ENCyclopedia 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 information[1–3]. 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 regions[4]. 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 genomes[5–9]. Eventually, high-throughput sequencing enabled truly genome-wide mapping of protein-DNA interactions, with high resolution, in the form of ChIP-seq[10–14].

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 ENCODE[15–18], mouse ENCODE[19] and modENCODE consortia[20,21], and the NIH Roadmap Epigenomics Mapping Consortium[22]. 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[23]. 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[24,25]. 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[26], 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, the 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[27,28] 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[11], 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[16] 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[29,30]. 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[29]) or by using cross-correlation[29,31]. Both measures have limitations in some special cases[30], 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[29]) and FRiP values for manual and robotic NRSF ChIP-seq datasets. R-ChIP data consistently exhibited good RSC values (RSC $\geq$ 1) and FRiP and peak number values comparable to those of manually

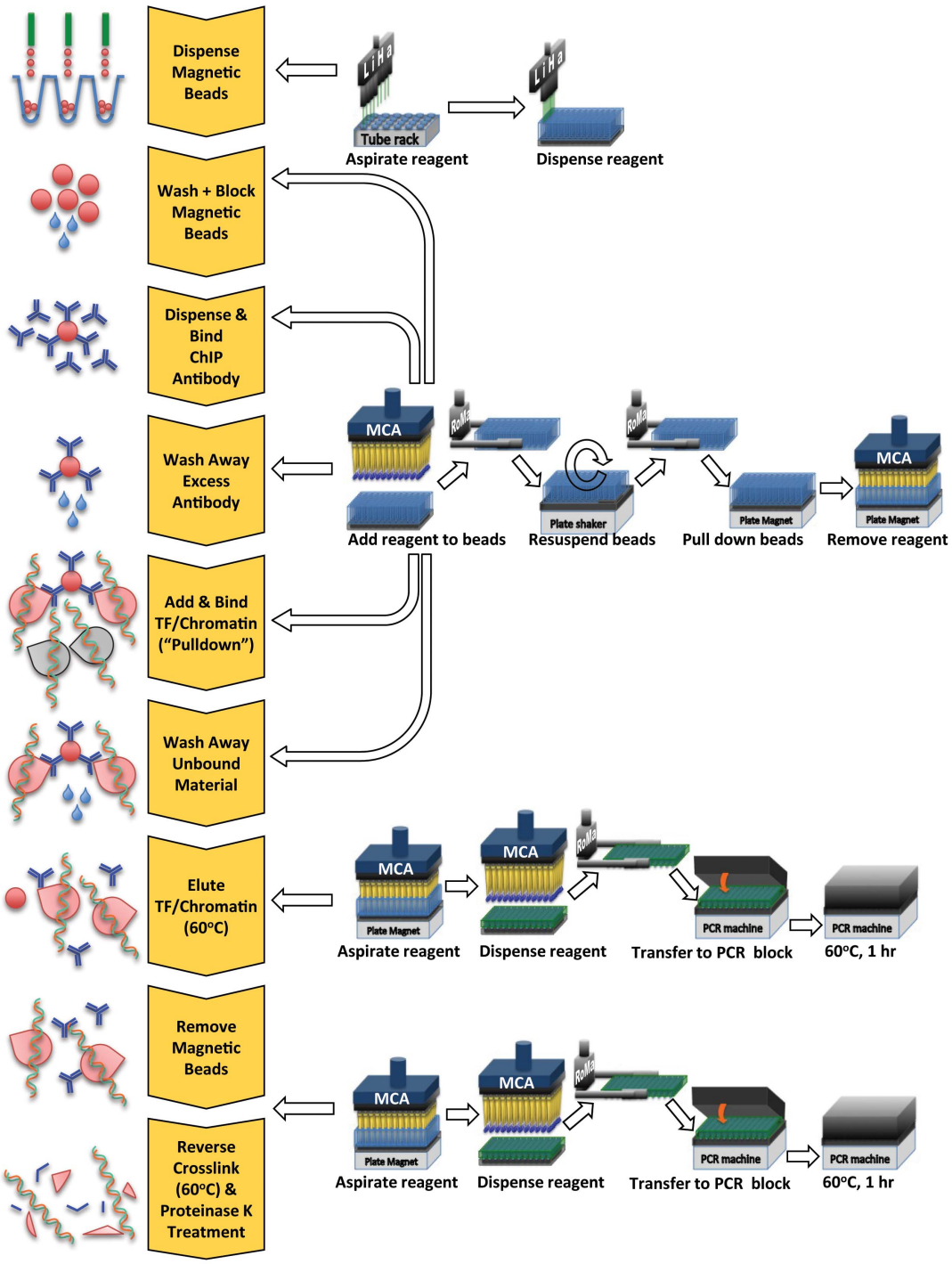

Figure 1 | Illustration of individual automated ChIP protocol steps. A Tecan Freedom EVO 200 robot equipped with a Liquid Handling arm (LiHa), a Multi Channel Arm (MCA) and Robotic Manipulator arm (RoMa) is used for all steps. Additional devices integrated into the robot are standard-size plate carriers, magnet plate, orbital plate shaker and PCR machine. The cartoons in the left column illustrate each protocol step, described in the flow diagram in the second column. The cartoon sequences on the right illustrate the robotic process step sequences used for each protocol step. The white arrows pointing to the protocol steps indicate which robot sequences apply to each protocol step.

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.

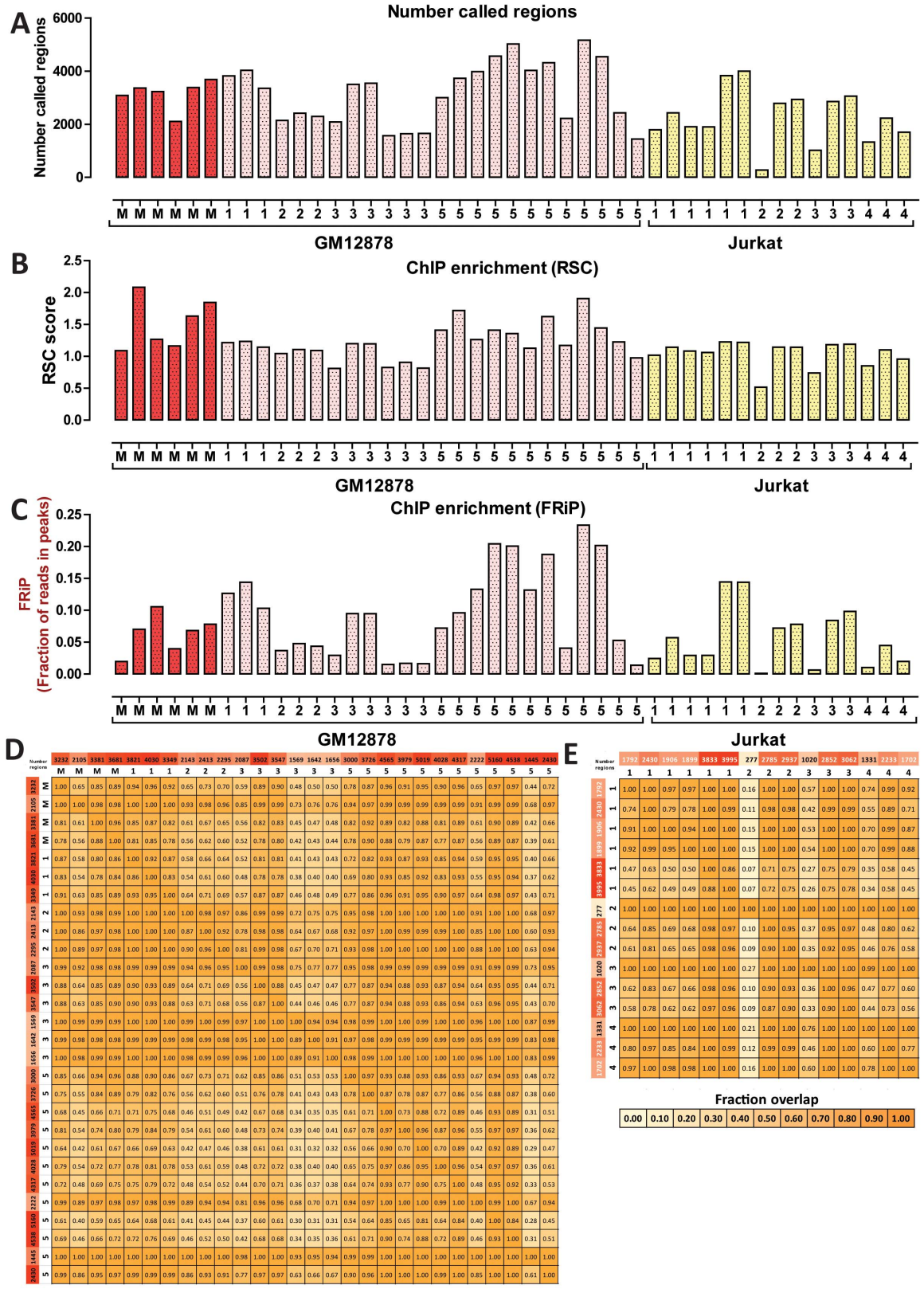

Figure 2 | Reproducibility of R-ChIP experiments. Multiple ChIP-seq experiments on multiple plates were generated for the NRSF/REST repressor in GM12878 lymphoblastoid cells ($n = 4$ plates) and Jurkat T-cells ($n = 4$ plates) cell lines. The numbers (1 through 5) refer to the number of the plate a library 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[29]; (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 (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. $O_{XY} = |X \cap Y|/|Y|$.

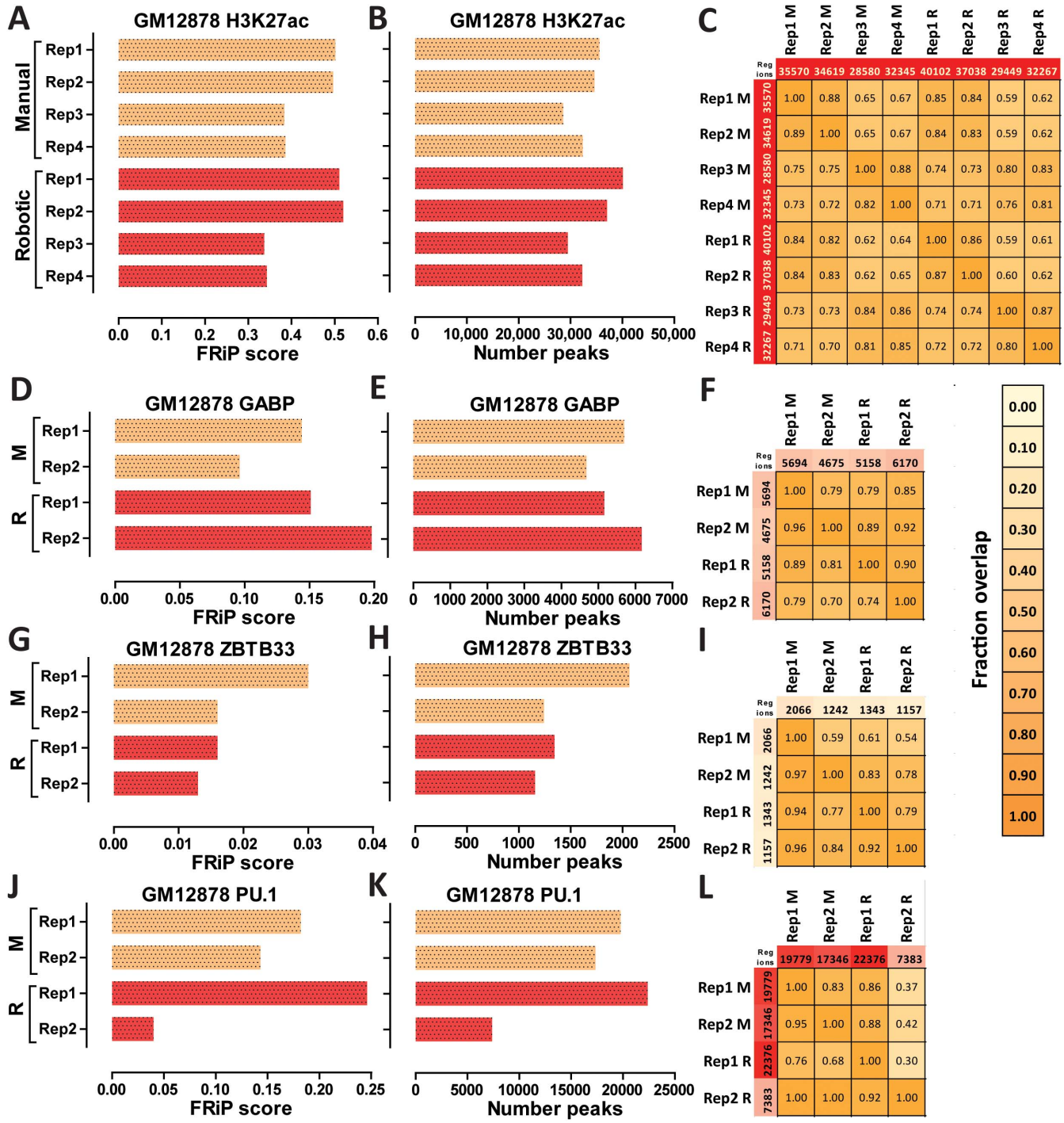

Figure 3 | Comparison between manual ChIP and R-ChIP results for additional targets. Manual and robotic ChIP-seq datasets were generated in parallel using the same chromatin samples and the result compared. (A,B,C) ChIP-seq against H3K27ac in GM12878 cells. (A) FRiP score, (B) number of peaks called, (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; (J), (K), (L) ChIP-seq against PU.1 in GM12878 cells. (J) FRiP score, (K) number of peaks called, (L) 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. $O_{XY} = |X \cap Y|/|Y|$.

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

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[38–41]. It is used as a marker of active transcriptional enhancers in mammalian genomes[42–45]. 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 α-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 ~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 E3113 and F2711), and they identified up to ~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 (ENCODE Project Consortium 2012), and the number of expressed genes per cell type is between 5 and 10×10^3. 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–L). 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-

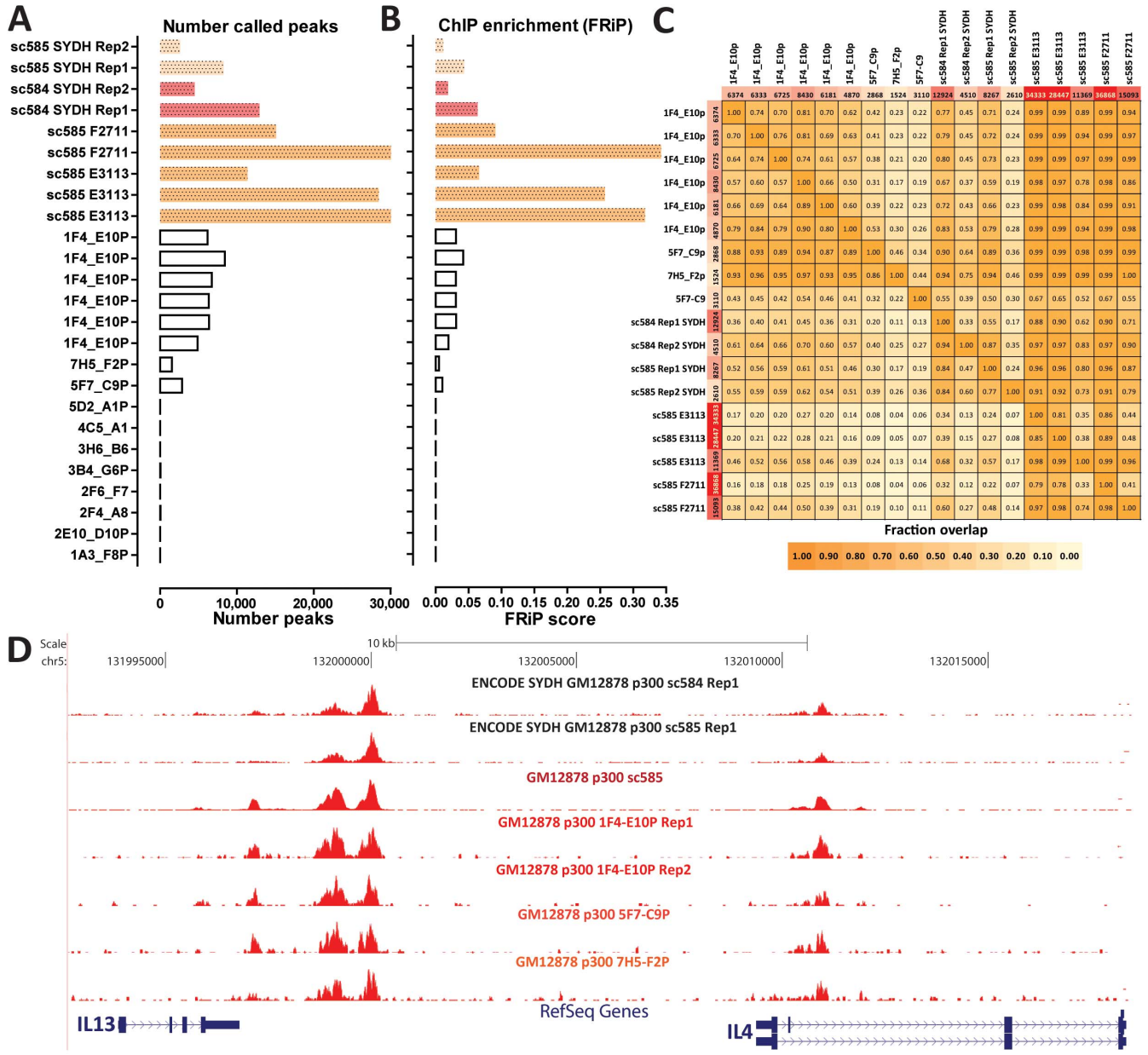

Figure 4 | Characterization of novel monoclonal p300 antibodies using robotic ChIP. ChIP-seq against p300 was carried out in GM12878 cells and prior ENCODE data for it in that cell line (from the "SYDH" production group) was used as a reference. ENCODE data was generated using two different rabbit polyclonal 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.

cific 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.

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 GM12878 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 (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 5×10^7 nuclei/mL using a probe sonic dismembrator from Fisher 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 65°C, purified using DNA purification columns 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 Caltech[11,48], an α-p300 rabbit polyclonal (sc-585) from Santa Cruz Biotechnology, a mouse monoclonal α-GABP (sc-28312) from Santa Cruz Biotechnology, a mouse monoclonal α-Kaiso/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 addition 11 α-p300 mouse monoclonals were generated in the Caltech Mouse Monoclonal Facility. Four of the α-p300 mouse monoclonals were raised against a bacterially expressed GST fusion protein containing N-terminal residues 152–213. The remaining seven antibodies were raised against a synthetic peptide from GenScript containing residues 1526–1545.

Robotic-ChIP (R-ChIP) workflow. ChIP experiments were adapted from methods previously described and optimized for performance in a 96-well plate format using a Tecan Freedom EVO 200 liquid handling robot. Reagents and labware are placed on deck of the robot (Supplementary Figure 1).

After setup the R-ChIP workflow is completely hands-off and consists of a series of modules with a run time of ~24 hours, including the 12-hour reverse cross-linking step. All aspects of the setup are checked thoroughly to ensure a smooth run.

1. **Blocking and washing of magnetic beads.** The Tecan begins by resuspending magnetic beads (Invitrogen M-280 Dynabeads) from the source tube with the liquid-handling arm (LiHa) and dispenses the magnetic beads into a Fisher 96-Well DeepWell™ Polypropylene known as the ChIP plate. 100 μL of beads is used for a monoclonal IP antibody and 200 μL for a polyclonal. The LiHa tips are evacuated and rinsed with ddH_2O between subsequent dispenses to prevent cross-contamination. 500 μL of PBS containing 5% bovine serum albumin (BSA) is then dispensed by the LiHa from a buffer reservoir (Te-Fill) to block and wash the magnetic beads. The plate containing the beads is transferred to an orbital mixer (Te-Shake) with the robotic manipulator arm (RoMa) and mixed several times for 20 seconds with a 20 second pause between each mix. The RoMa moves the bead plate to a magnetic plate for seven minutes where the beads are then pelleted in a ring allowing the multi-channel arm (MCA96) fitted with natural 200 μL tips from TipOne to aspirate liquid. These steps are repeated three more times and include an ethanol rinse of the MCA96 tips as needed to prevent cross-contamination.
2. **Binding of antibody to magnetic beads.** The LiHa adds 400 μL of PBS-BSA to the antibody plate bringing the final volume to 500 μL. The antibody is then added to the beads using the MCA 96 which transfers the diluted antibody from the a 2.0-mL 96-well PlateOne V-bottom plate to the ChIP plate. For monoclonal antibodies, 5 μg of antibody were diluted in 500 μL (10 μg for polyclonals). The beads and antibody are incubated together for one hour with mixing using the Te-Shake. Any unbound antibody is then aspirated with the MCA96 and deposited into a fresh 2.0-mL 96-well PlateOne V-bottom plate for further analysis if needed.
3. **Incubation of chromatin and antibody-bead complex.** The MCA96 transfers 500 μL of chromatin containing 2.5×10^7 cells from the Matrix tube rack to the ChIP plate. The chromatin and antibody bead complex are then incubated together for 2 hours during which the ChIP plate alternates between the Te-Shake and a 4°C cool plate using the RoMa. The chromatin is stored in 1.2 mL screw-top Matrix tubes that can be arrayed on the chromatin plate as needed. Any unbound chromatin is then aspirated with the MCA96 and deposited into a fresh 2.0-mL 96-well PlateOne V-bottom plate for further analysis if needed.
4. **Washing of IP complex.** The LiHa dispenses 500 μL of LiCl wash buffer (100 mM Tris pH 7.5, 500 mM Lithium Chloride, 1% NP-40, 1% sodium deoxycholate) from the Te-Fill onto the beads, which are then mixed for 20 seconds with 20 second pauses between each mix. The beads are then pelleted and the wash with LiCl buffer is repeated four more times. The LiHa then adds 500 μL of TE buffer (10 mM Tris pH 7.5 and 1 mM EDTA) and resuspends the beads with the Te-Shake for 20 seconds. Beads are then pelleted with the magnetic plate and any remaining buffer is aspirated and discarded by the MCA96.
5. **Elution from beads.** The LiHa dispenses 100 μL of IP elution buffer (1% SDS and 0.1 M $NaHCO_3$) from the Te-Fill and the beads are resuspended by mixing for 20 seconds with the Te-Shake. The MCA96 then aspirates the beads and transfers them from the ChIP plate to a Hard-Shell Semi-Skirted PCR Plate from Bio-Rad. The RoMa transfers a PCR lid from the storage hotel and places it on top of the PCR plate then transfers the lidded PCR plate to a DNA Engine Peltier Thermal Cycler with Remote Alpha Dock System from Bio-Rad. The top of the thermal cycler closes and places force on the PCR plate lid creating a seal. The beads are then heated for one hour at 65°C to disassociate the IP complex from the magnetic beads.
6. **Reversal of Cross-links.** The RoMa takes the PCR plate from the thermal cycler and transfers it to the Te-Shake to resuspend the beads. The PCR plate is then transferred to the magnetic plate for pelleting of the beads. The MCA96 mounts 50 μL Tecan Pure Disposable tips from Tecan, slowly aspirates the supernatant, and transfers it to a fresh PCR plate. 10 μL of proteinase K from Epicentre diluted 1 : 5 in proteinase K buffer (50% glycerol, 50 mM Tris-HCl pH 7.5, 0.1 M NaCl, 0.1 mM EDTA, 1 mM DTT, 10 mM $CaCl_2$, 0.1% Triton® X-100) is then added to the supernatants with the LiHa. The RoMa places a lid on the fresh PCR plate and transfers both back to the thermal cycler for a 12 hour incubation at 65°C to reverse the cross-links. Once the incubation is finished the plate is transferred to the deck with the RoMa and the R-ChIP is complete.

DNA cleanup. Samples from the R-ChIP ChIP experiments presented here were then cleaned up manually using the protocol described by Qiagen in their Qiaquick PCR purification kit with the addition that the EB buffer is heated to 55°C prior to elution and eluted in a 50 μL volume using DNA lo-bind 1.5 mL 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 flow cell lane, according to the HudsonAlpha ENCODE ChIP protocol (Supplementary Methods). Standard methods were used for end repair and dA addition of DNA fragments recovered from ChIPs or chromatin controls. The fragments were then ligated to Illumina Paired-End adaptor sequences and PCR-amplified to complete the adaptor sequences and introduce a 7-base DNA barcode in the i7 position. The barcodes allowed mixing of multiple samples per flowcell lane. Control libraries were prepared from 500 ng of DNA from reverse crosslinked sonicated chromatin. ChIP library starting amounts varied, with a median of 7.5 ng. Fragment size selection was achieved at the lower threshold with solid phase reversible immobilization (SPRI) technology to recover dsDNA greater than 100 bp after adaptor ligation (thereby excluding unligated adaptors) and at the upper threshold with an extension time of 30 seconds during PCR amplification. This size selection method consistently produced final DNA library fragments that ranged from ~100 to 400 bp, as determined by BioAnalysis. Final library amounts varied by ChIP, with a median of 546 ng.

Detailed description of the protocol can be found in the Supplementary Methods.

Libraries were pooled in equimolar amounts and 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 Bowtie[49], version 0.12.7, with the following settings: "-v 2 -t -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

browser) with the Y chromosome retained or removed depending on the sex of the cell line. Peak calling was carried out using ERANGE[11], version 4.0, with the following settings: "--minimum 2 --ratio 3 --listPeak --shift learn --revbackground", against matching control samples. Library complexity was estimated as described previously[29]. Cross-correlation analysis was carried out using version 1.10.1 of SPP[31]) and the following parameters: "-s = 0 : 2 : 400".

All additional analysis was carried out using custom-written Python scripts.

1. Gilmour, D. S. & Lis, J. T. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. *Proc Natl Acad Sci U S A* **81**, 4275–4279 (1984).
2. Gilmour, D. S. & Lis, J. T. In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. *Mol Cell Biol* **5**, 2009–2018 (1985).
3. Solomon, M. J., Larsen, P. L. & Varshavsky, A. Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* **53**, 937–947 (1988).
4. Hecht, A., Strahl-Bolsinger, S. & Grunstein, M. Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* **383**, 92–96 (1996).
5. Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
6. Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M. & Brown, P. O. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533538 (2001).
7. Horak, C. E. & Snyder, M. ChIP-chip: A genomic approach for identifying transcription factor binding sites. *Methods Enzymol* **350**, 469483 (2002).
8. Lieb, J. D., Liu, X., Botstein, D. & Brown, P. O. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat Genet* **28**, 327334 (2001).
9. Weinmann, A. S., Yan, P. S., Oberley, M. J., Huang, T. H. & Farnham, P. J. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev* **16**, 235244 (2002).
10. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823837 (2007).
11. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–502 (2007).
12. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
13. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **4**, 651–657 (2007).
14. Wold, W. & Myers, R. M. Sequence census methods for functional genomics. *Nat Methods* **5**, 19–21 (2008).
15. ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **9**, e1001046 (2011).
16. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
17. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
18. Wang, J. *et al.* Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* **22**, 1798–1812 (2012).
19. Mouse ENCODE Consortium. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol* **13**, 418 (2012).
20. Gerstein, M. B. *et al.* Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
21. modENCODE Consortium. Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330**, 1787–1797 (2010).
22. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045–1048 (2010).
23. Egelhofer, T. A. *et al.* n assessment of histone-modification antibody quality. *Nat Struct Mol Biol* **18**, 91–93 (2011).
24. Garber, M. *et al.* A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol Cell* **47**, 810–822 (2012).
25. Blecher-Gonen, R., Barnett-Itzhaki, Z., Jaitin, D., Amann-Zalcenstein, D., Lara-Astiaso, D. & Amit, I. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat Protoc* **8**, 539–554 (2013).
26. Aldridge, S. *et al.* AHT-ChIP-seq: a completely automated robotic protocol for high-throughput chromatin immunoprecipitation. *Genome Biol* **14**, R124 (2013).
27. Chong, J. A. *et al.* REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**, 949–957 (1995).
28. Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267**, 1360–1336 (1995).
29. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813–1831 (2012).
30. Marinov, G. K., Kundaje, A., Park, P. J. & Wold, B. J. Large-Scale Quality Analysis of Published ChIP-seq Data. *G3 (Bethesda)* **4**, 209–223 (2014).
31. Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**, 1351–1359 (2008).
32. Collins, P. J., Kobayashi, Y., Nguyen, L., Trinklein, N. D. & Myers, R. M. The ets-related transcription factor GABP directs bidirectional transcription. *PLoS Genet* **3**, e208 (2007).
33. Watanabe, H., Wada, T. & Handa, H. Transcription factor E4TF1 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. Prokhortchouk, A. *et al.* The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev* **15**, 1613–1618 (2001).
36. Burda, P., Laslo, P. & Stopka, T. The role of PU.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 E1A-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. E1A-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., Harter, M. L. & Goodman, R. H. Adenoviral E1A-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. Blow, 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. Neph, 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**(7414), 75–82 (2012).
48. Mortazavi, A., Leeper Thompson, E. 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 Trout and Sean Upchurch for technical and computational assistance, Dan Savic for discussions and contributions to our manual ChIP production and development efforts and the Caltech Protein Expression Center and Monoclonal Antibody Facility for production of monoclonal antibodies. This work was supported by the Beckman Institute Functional Genomics Center, the Donald Bren Endowment, and NIH grants U54 HG004576 and U54 HG006998.

Author contributions

B.J.W., J.V. and W.C.G. designed the R-ChIP protocol and experiments to test it; W.C.G., J.V. and M.T.S. established and refined the protocol and performed the R-ChIP experiments; F.P. and K.N. developed and carried out all ChIP library construction for sequencing and executed comparative manual ChIP, G.D. and J.V. designed and screened, and S.O. immunized and produced the p300 monoclonal antibodies; G.K.M. performed computational analyses; G.K.M., W.C.G., J.V., F.P., R.M.M. and B.J.W. wrote the paper.

Additional information

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

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Gasper, W.C. *et al.* Fully automated high-throughput chromatin immunoprecipitation for ChIP-seq: Identifying ChIP-quality p300 monoclonal antibodies. *Sci. Rep.* **4**, 5152; DOI:10.1038/srep05152 (2014).