Screening of PCRP transcription factor antibodies in biochemical assays

William K. M. Lai[1,2], Luca Mariani[3], Gerson Rothschild[5], Edwin R. Smith[6], Bryan J. Venters[7], Thomas R. Blanda[1], Prashant K. Kuntala[1], Kylie Bocklund[1], Joshua Mairose[1], Sarah N Dweikat[1], Katelyn Mistretta[1], Matthew J. Rossi[1], Daniela James[1], James T. Anderson[3], Sabrina K. Phanor[3], Wanwei Zhang[5], Avani P. Shaw[6], Katherine Novitzky[7], Eileen McAnarney[7], Michael-C. Keogh[7]*, Ali Shilatifard[6]*, Uttiya Basu[5]*, Martha L. Bulyk[3,4]*, B. Franklin Pugh[1,2]*

[1]Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA
[2]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853
[3]Division of Genetics, Department of Medicine; Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115.
[4]Department of Pathology; Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115.
[5]Department of Microbiology and Immunology, Vagelos College of Physicians and Surgeons, Columbia University, New York.
[6]Simpson Querrey Center for Epigenetics and the Department of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, 303 East Superior Street, Chicago, IL 60611, USA
[7]EpiCypher Inc., Durham NC 27709

*Co-corresponding authors

Contact Information:

William KM Lai <wkl29@cornell.edu>
Luca Mariani <lucamariani77@gmail.com>
Gerson Rothschild <ger2@cumc.columbia.edu>
Edwin R. Smith <edwinsmith@northwestern.edu>
Bryan J. Venters <bventers@epicypher.com>
Thomas R Blanda <trb5344@psu.edu>
Prashant K Kuntala <pxk284@psu.edu>
Kylie Bocklund <kxb357@psu.edu>
Joshua Mairose <josh.mairose@gmail.com>
Sarah N Dweikat <szd41@psu.edu>
Katelyn Mistretta <katelynsmistretta@gmail.com>
Matthew J. Rossi <mjr26@psu.edu>
Daniela James <daniela.james@psu.edu>
James T. Anderson <jamesanderson12358@gmail.com>
Sabrina K. Phanor <sphanor@bwh.harvard.edu>
Wanwei Zhang <wz2370@cumc.columbia.edu>
Avani P. Shaw <avani.shah1@northwestern.edu>
Katherine Novitzky <knovitzky@epicypher.com>
Eileen McAnarney <emcanarney@epicypher.com>

Michael-C. Keogh <mkeogh@epicypher.com>
Ali Shilatifard <ash@northwestern.edu>
Uttiya Basu <ub2121@cumc.columbia.edu>
Martha L. Bulyk <mlbulyk@genetics.med.harvard.edu>
B. Franklin Pugh <fp265@cornell.edu>

Abstract

Antibodies offer a powerful means to interrogate specific proteins in a complex milieu, and where epitope tagging is impractical. However, antibody availability and reliability are problematic. The Protein Capture Reagents Program (PCRP) generated over a thousand renewable monoclonal antibodies against human-presumptive chromatin proteins in an effort to improve reliability. However, these reagents have not been widely field-tested. We therefore screened their ability in a variety of assays. 887 unique antibodies against 681 unique chromatin proteins, of which 605 are putative sequence-specific transcription factors (TFs), were assayed by ChIP-exo. Subsets were further tested in ChIP-seq, CUT&RUN, STORM super-resolution microscopy, immunoblots, and protein binding microarray (PBM) experiments. At least 6% of the tested antibodies were validated for use in ChIP-based assays by the most stringent of our criteria. An additional 34% produced data suggesting they warranted further testing for clearer validation. We demonstrate and discuss the metrics and limitations to antibody validation in chromatin-based assays.

Introduction

Antibodies are a critical component of a wide variety of biochemical assays. They serve as protein-specific affinity-capture and detection reagents, useful *in vivo* and *in vitro*. Example assays include chromatin immunoprecipitation (ChIP) of protein-DNA interactions, immunofluorescence, immunoblotting, ELISA, purification of cells and proteins, protein binding microarray (PBM) experiments, and targeted in vivo delivery of effector molecules[1-6]. One advantage of antibodies is their ability to recognize proteins without their targets having an engineered affinity tag. The human proteome contains tens of thousands of distinct proteins, each requiring a different antibody for specific detection. The usage of a large variety of antibodies to diverse targets has been a critical component of large NIH-funded consortium projects such as the ENCODE and Roadmap Epigenomics Mapping[7, 8]. However, broad profiling of the genomic targets of human transcription factors (TFs) has been limited by the availability of 'ChIP-grade' antibodies.

Overall, there has been an acute lack of antibodies that effectively distinguish the many thousands of different chromatin proteins, and an inconsistency inherent in how many immunoreagents are produced and perform[9-11]. Polyclonal antibodies, being a mixed product of many antibody genes, have the advantage of recognizing multiple epitopes on a protein, thereby producing robust target detection[12]. However, their production is finite and variable from one animal source to another[13], which hampers reproducibility.

The NIH Protein Capture Reagent Program (PCRP) was initiated through the NIH Common Fund with the stated goal of testing the feasibility of producing low-cost, renewable, and reliable protein affinity reagents in a manner that can be scaled ultimately to the entire human proteome (PA-16-287) (https://proteincapture.org/)[14]. With an initial focus on putative TFs, this endeavor reported the production of 1,406 mouse monoclonal antibodies (mAb) against 737 chromatin protein targets[15]. This included two parallel production approaches: mouse hybridomas that release mAb into growth medium supernatant ("supernate"), and recombinant antibodies produced in *E. coli*. The advantages of these two approaches over polyclonal antibodies are, in principle, a renewable and consistent supply. Since mAb are produced from a single set of genes in an immortalized hybridoma or bacterial cells, they recognize only a single epitope[16-18]. To accommodate the potential shortcoming of a hybridoma recognizing a single

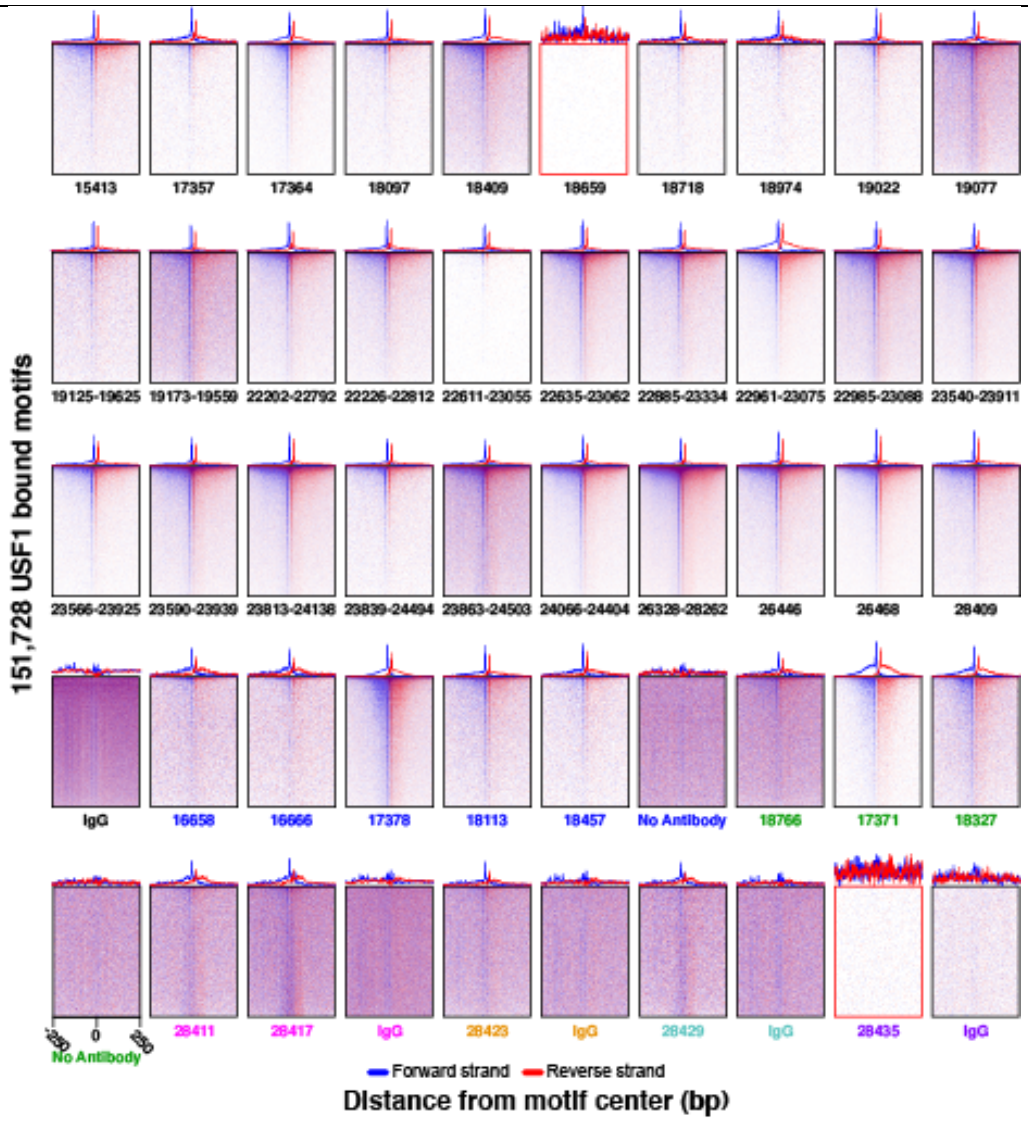
96 non-viable epitope, an effort by the NIH PCRP was made to generate at least two independent
97 clones for each target.
98 Antibody validation is critical to their utility. Validation exists at many levels ranging
99 from whether an antibody recognizes its intended target to whether it consistently performs
100 successfully in a particular assay. Each publicly available PCRP-generated antibody was
101 previously validated for its target recognition by *in vitro* human protein (HuProt) microarray
102 screening[15]. Whether the antibodies work in particular biochemical assays has been described for
103 a limited set in a limited number of approaches, including immunoblotting, immunoprecipitation,
104 immunohistochemistry, and ChIP-seq[15]. Previous work has reported that 46 of 305 mAb against
105 36 of 176 targets passed ENCODE ChIP-seq standards[15]. "Browser shots" of selected loci were
106 available for ~40 datasets against 31 targets. However, antibody validation is not credible when
107 using single-locus examples ("browser shots") of enrichment over background as the sole
108 criteria. Chromatin fragmentation and extraction may generate localized variation in yield that
109 varies from prep to prep (active promoters, enhancers, etc.) Numerous replicates (target and
110 control) are often needed to ensure against false-positives at a few individual loci due to
111 sampling variation. Furthermore, the datasets from Venkataraman et al. are not in the public
112 domain, and so we were unable to independently analyze them. As broader community use of the
113 PCRP-generated se antibodies will likely benefit from a wider survey, we conducted additional
114 tests of these reagents.
115 Since evaluating each of ~1,400 antibodies in a wide variety of assays is not practical, we
116 instead opted for broad coverage by ChIP-exo as a broad-screening assay, combined with
117 expanded capability testing of a limited number in one or more distinct experimental
118 approaches. Overall, we tested 950 antibodies, of which 642 targeted putative TFs, which
119 allowed computational validation through the enrichment of their cognate motifs. The antibodies
120 and assays were chosen to cover a wide range of end-user applications, with specific TFs chosen
121 in part based on the scientific interests of the investigators and a set of objective criteria. With a
122 deep dive on a single assay (ChIP-exo), we explored end-user practical issues related to antibody
123 sourcing, reproducibility and validation metrics, and specificity for cell types and states. More
124 limited evaluations in a subset of assays (ChIP-Seq, CUT&RUN, super-resolution cellular
125 microscopy, Western blots, and protein binding microarray experiments) resulted in 174
126 antibodies assayed by at least two independent methods.
127

128 ## Results

129 ### Screening PCRP mAb by ChIP-exo

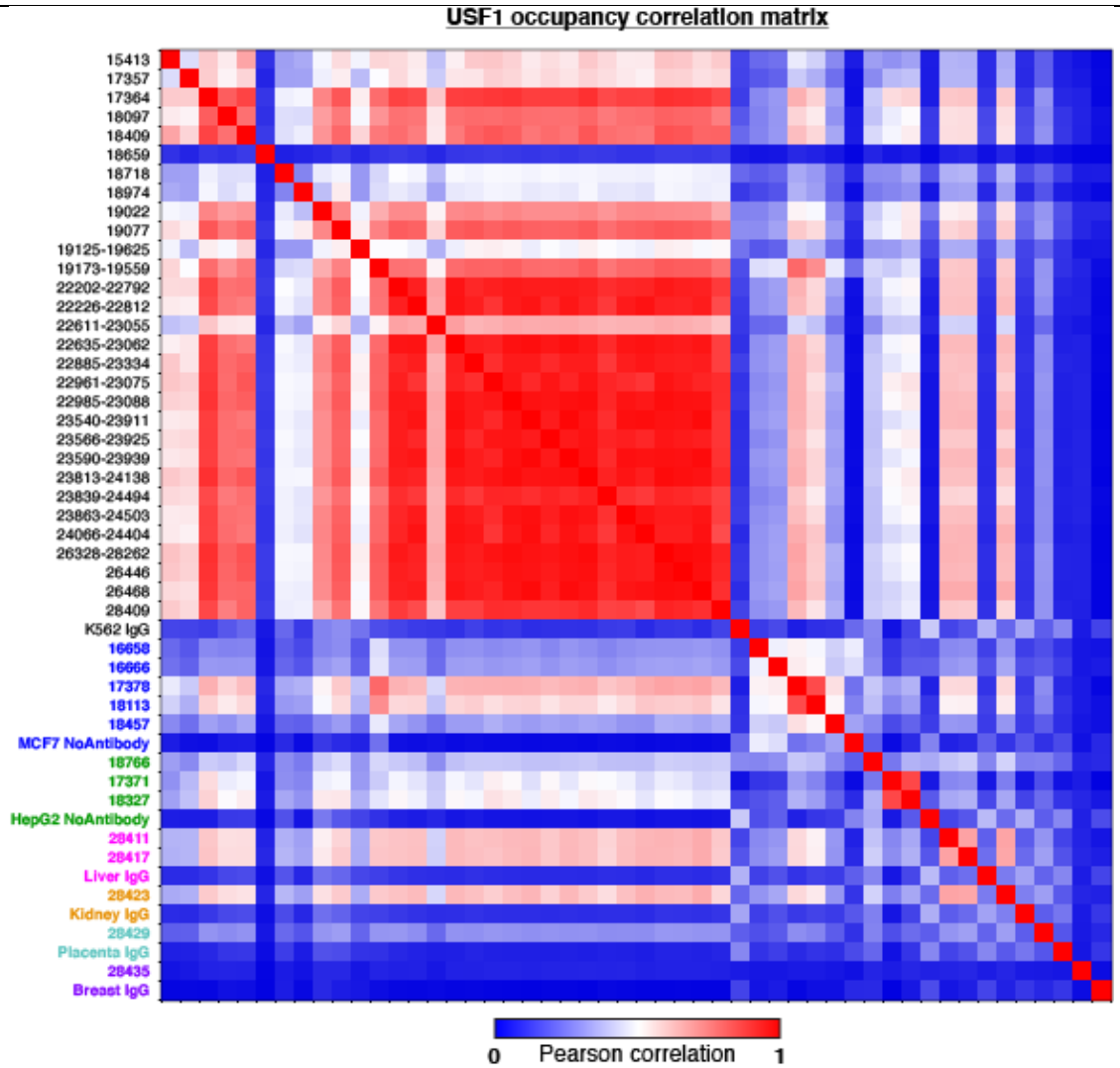
130 We used the massively parallel ChIP-exo version of ChIP-seq to screen PCRP Abs in 96-
131 well plate format for their ability to recognize their putative protein targets in a chromatinized,
132 cellular context[19, 20]. Briefly, proteins were crosslinked to DNA and each other within cells using
133 formaldehyde. Chromatin was then isolated, fragmented, and immunoprecipitated. While on the
134 beads, the fragmented DNA was trimmed with a strand-specific 5'-3' exonuclease up to the point
135 of crosslinking (i.e. protection), which was then mapped by DNA sequencing. For many proteins
136 this provides single-bp resolution in genome-wide detection[21].
137 Technical reproducibility of ChIP-exo with PCRP antibodies was evaluated with 43
138 independent replicates performed on the sequence-specific TF USF1, which also served as a
139 positive control in each cohort of 46 mAb assayed in K562, MCF7, HepG2, and human tissue.
140 From this and IgG negative controls, we defined a set of ~152,000 associated E-box motif
141 instances associated with a USF1 peak in at least one replicate ChIP-exo experiment. This

142 reflects an overly relaxed criterion, with a high level of false positives, for the purposes of
143 evaluating the gradient from true binding through nonspecific background at genomic E-boxes
144 (expected where a peak location occurred in only a small fraction of datasets). Greater stringency
145 on the number of replicates in which the same peak was found, resulted in peaks with higher
146 occupancy (not shown). 41 of the 43 USF1 datasets (>95%) produced a USF1-specific aligned
147 read pattern (*Supplementary Fig. 1*, vertical blue and red stripes in the heat maps and single-bp
148 peaks in the composite plots). These, but not IgG controls, were Pearson-pairwise correlated
149 (*Supplementary Fig. 2*). This level of reproducibility indicated that ChIP-exo was suitable for
150 screening and evaluating PCRP antibodies in ChIP.



Supplementary Fig. 1 Technical robustness of the USF1 PCRP mAb in the ChIP-exo assay. Heatmap and row-averaged composite plots for 43 USF1 ChIP-exo experiments with the indicated project sample ID were performed in K562 (black), MCF7 (blue), HepG2 (green), human liver (pink), kidney (orange), placenta (cyan), and breast (purple). The 5' end of aligned sequence reads for each replicate were plotted against their distance from cognate USF1 motifs (E-boxes). These motifs were present in the union of called peaks across all 43 USF1 datasets for a total of 151,728 peaks that intersected with an E-box motif. Reads are strand-separated (blue = motif strand, red = opposite strand). Rows are linked across samples and sorted based on their combined rank-order average in a 500 bp bin around each motif midpoint. Matching IgG or No-Antibody control experiments for each cell type are shown. Samples 18659 and 28435 are outlined in red and represent experiments which failed to show enrichment at USF1 peaks.

151



Supplementary Fig. 2 Correlation matrix of USF1 technical replicates. Pearson correlation was calculated between technical replicates and negative controls using the sum of tags in a 500 bp window centered around the motif midpoint for all 151,728 potential USF1 binding events. Samples are labelled and colored as defined in *Supplementary Fig. 1*.

We initiated our ChIP antibody survey by first considering the practicality of producing a large number of antibody preparations from recombinant *E. coli* or mouse hybridomas. Purification from *E. coli* included transformation of expressing plasmids, cell growth, followed by protein induction, and purification. After multiple attempts we determined that high throughput parallelized rAb production was not practical within the scope of this project, due to a need to optimize the protocol in our hands, particularly in comparison to the hybridoma approach. We therefore opted to pursue hybridoma-based mAb.

We tested NRF1, USF1, and YY1 PCRP mAb from two sources: DSHB and CDI. The former as hybridoma culture supernates (10-80 ug/ml antibody), and the latter as concentrates (**Figure 1a**). In general, we found that while antibodies from both sources specifically detect NRF1 and USF1, DSHB hybridoma supernates detected more binding events at cognate motifs compared to CDI concentrates, when assayed at the same reported antibody amounts (3 ug). The

165 improved performance, simplicity and lower cost of using hybridoma supernates led us to source
166 from DSHB for the remainder of this study.
167

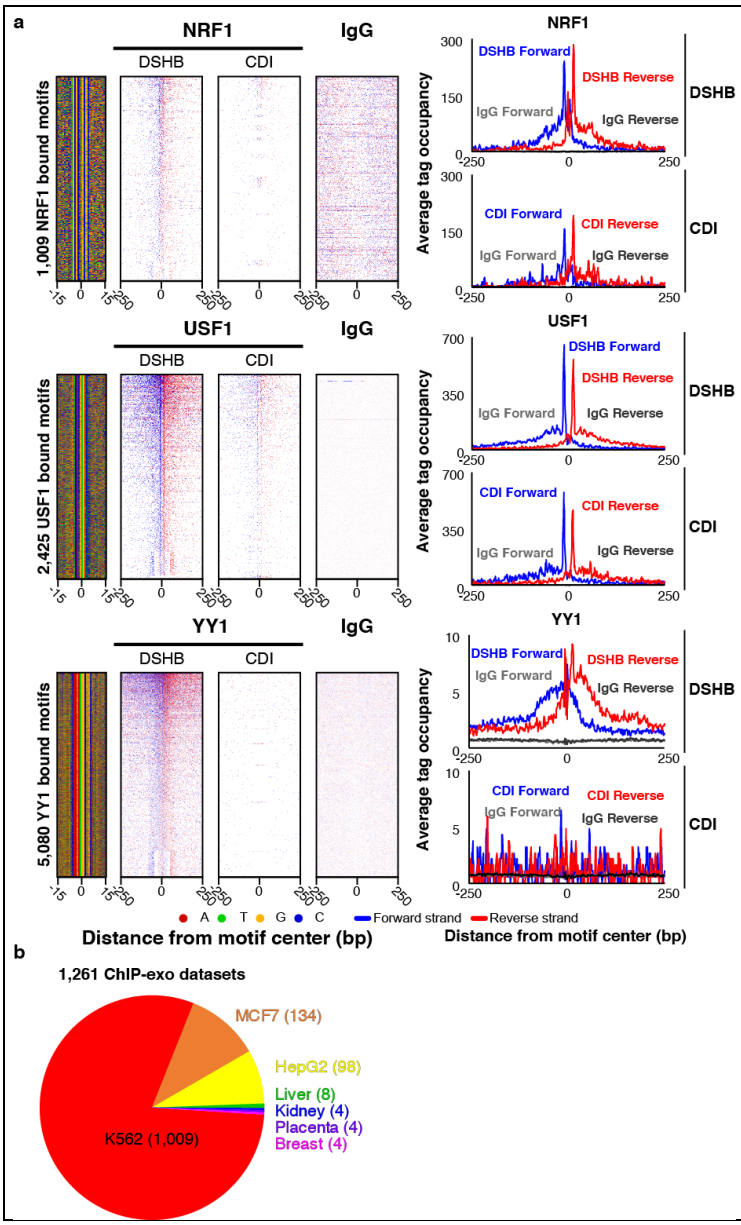
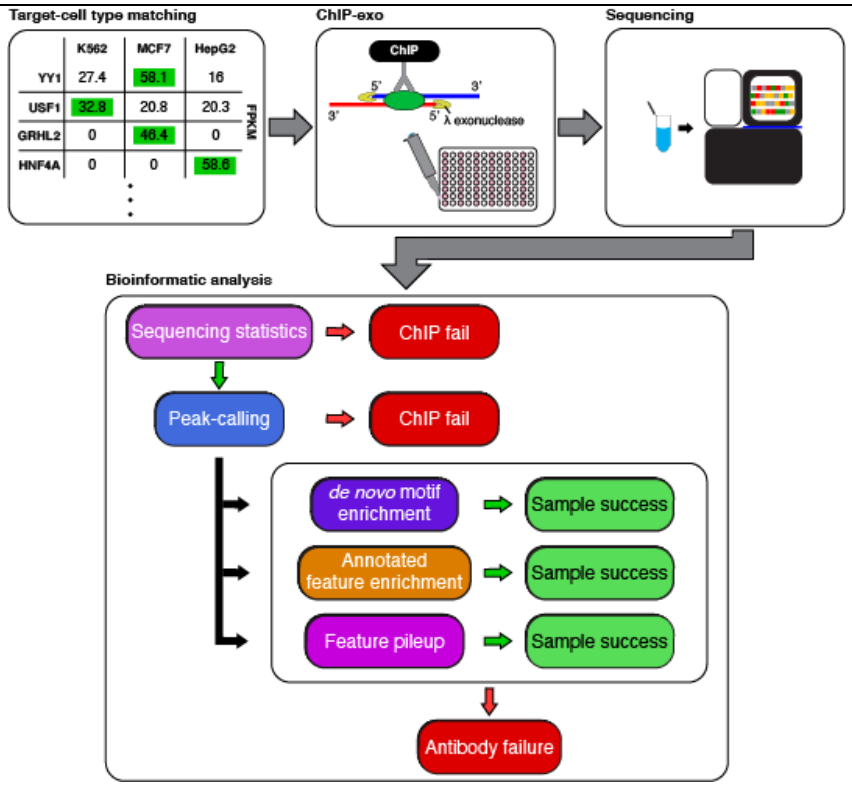


Figure 1 Assessment of antibody source (DSHB and CDI). (**a**) DNA-sequence 4-color plots (left), heatmaps (middle), and composites (right) were generated for the indicated targets, number of bound motifs, and antibody source, tested in K562 cells. The 5' end of aligned sequence reads for each set of experiments were plotted against distance from cognate motif, present in the union of all called peaks between the datasets for each indicated target. Reads are strand-separated (blue = motif strand, red = reverse strand). Rows are linked across samples and sorted based on their combined average in a 100 bp bin around each motif midpoint. (**b**) Pie chart shows the cell/tissue type of biological material used for 1,261 ChIP-exo datasets.

168
169 We assayed all 887 available hybridoma supernates, containing mAbs to 681 non-
170 redundant targets. Testing was initially performed in K562 (1,009 samples), MCF7 (134
171 samples), and HepG2 (96 samples) cells, based on reported target mRNA expression levels.
172 Testing defaulted to K562 if there was no substantial difference or the TF was not expressed at
173 appreciable levels in any of these three lines due to practical considerations including the ability
174 to grow K562 at scale (liquid culture) (*Supplementary Fig. 3*). Some targets had multiple
175 independent hybridoma clones, which were individually tested. In total, 887 mAb against 681
176 unique targets were assayed by ChIP-exo, some as replicates in the same or different cell types,
177 resulting in 1,261 datasets (**Figure 1b**). It was not feasible from a cost perspective to perform

178 replicates, deep sequencing, and antibody titrations for all mAbs, and so replicates were largely
179 limited to those that passed certain validation criteria (see below).
180



Supplementary Fig. 3 Workflow schematic of bulk PCRP antibody testing in the ChIP-exo assay. If possible, targets were assigned a primary cell line based on positive expression from RNA-seq[7]. Samples were processed in cohorts of 46 plus a USF1 positive control and an IgG or "No Antibody" negative control. After high-throughput sequencing, samples were automatically processed through a bioinformatics quality control pipeline. Sample were examined for sequencing depth, library complexity (% PCR duplication), and the ability to generate significant peaks. Peaks and raw tags were then examined to identify enriched sequence motifs, localization to annotated chromatin and sequence regions, and specific enrichment at genomic features such as transcription start sites.

181
182 As exemplified by NRF1, USF1, YY1, and an IgG control (**Figure 2**), the finding that
183 ChIP-exo peaks were enriched at a very precise distance from cognate but not noncognate motifs
184 provided support for detection of specifically bound target TF. We also compared independently
185 constructed hybridomas, exemplified by heat shock factor 1 (HSF1), as an alternative
186 independent validation. Both gave nearly identical ChIP-exo aligned read patterns around the
187 same set of features (heat shock elements, *Supplementary Fig. 4a*), thereby demonstrating
188 reproducibility of ChIP-exo profiles across independent PCRP antibodies. This is particularly
189 important where validation by motif enrichment is not applicable. However, two other HSF1
190 mAb clones failed (*Supplementary Fig. 4b*), indicating that independent PCRP Ab clones can
191 have different capabilities. Therefore, if one mAb clone fails, it may be productive to check
192 additional clones.
193 Protein targets that interact with each other or with the same sites may also provide cross-
194 validation for determining enrichment specificity. For example, in the case of USF1 and USF2
195 interaction partners and homologs[22], the two cross-validated (*Supplementary Fig. 4c*). However,

196 in this particular case we cannot exclude cross-reactivity of the mAb with the two homologs
197 (always a potential concern with target-specific antibodies). Additional validation may occur
198 with corresponding ChIP datasets in the public domain that use different antibodies (*e.g.*,
199 ENCODE, *Supplementary Fig. 5*).
200

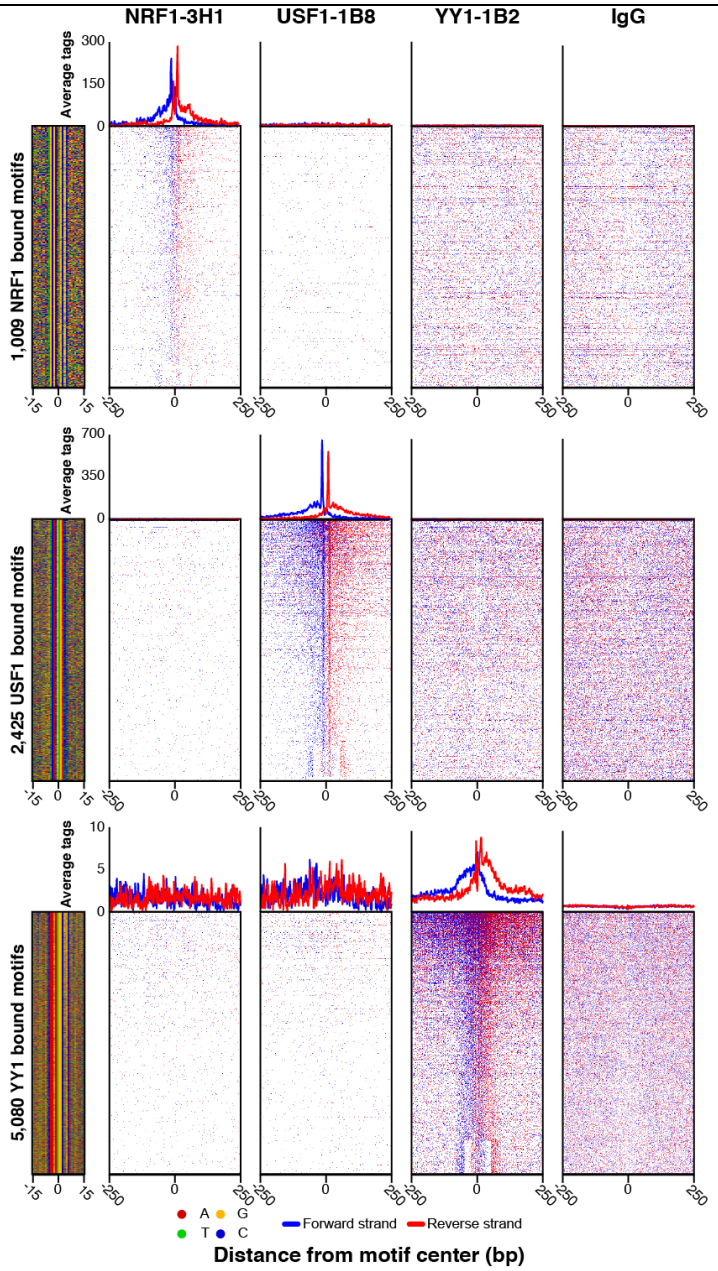
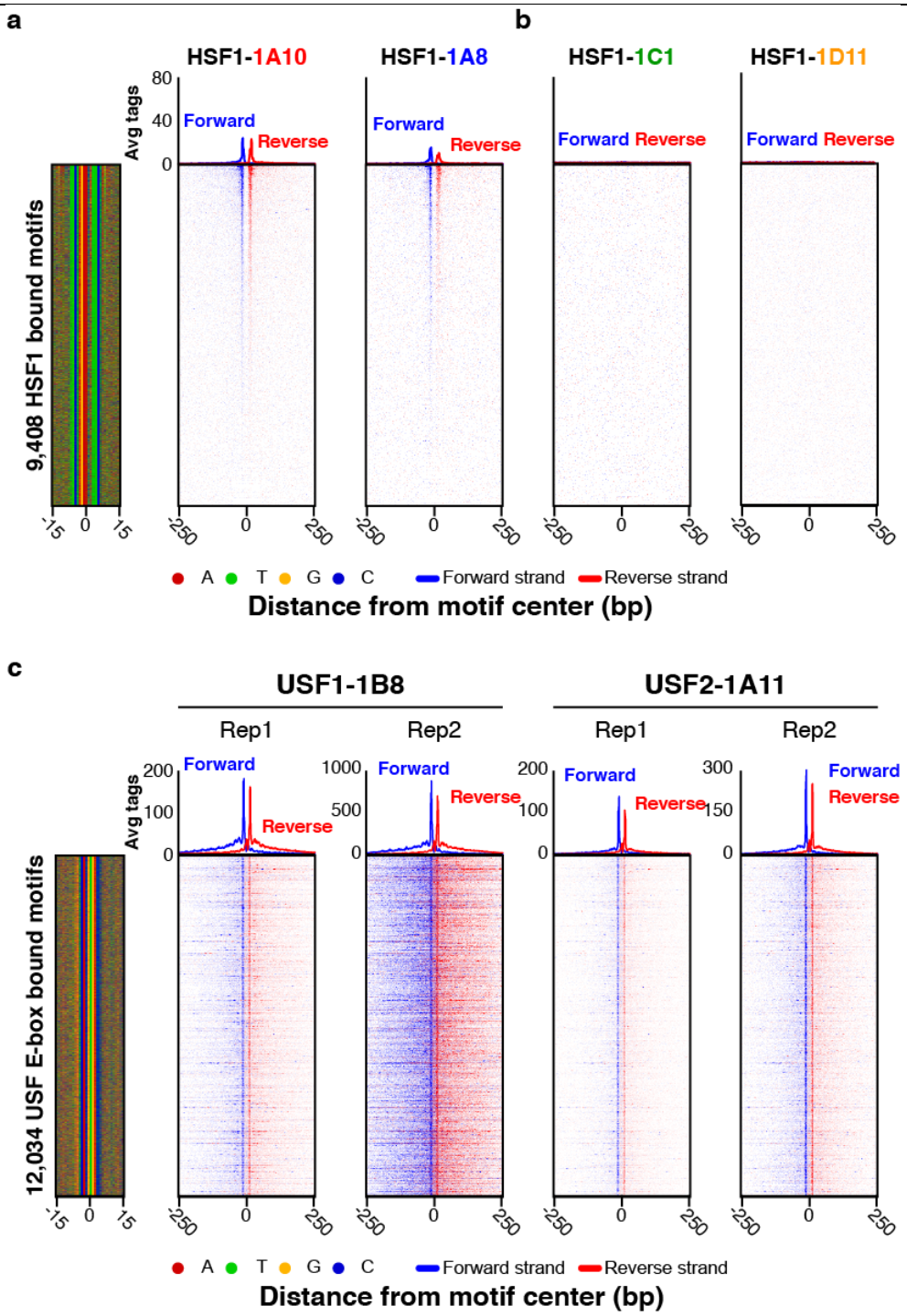
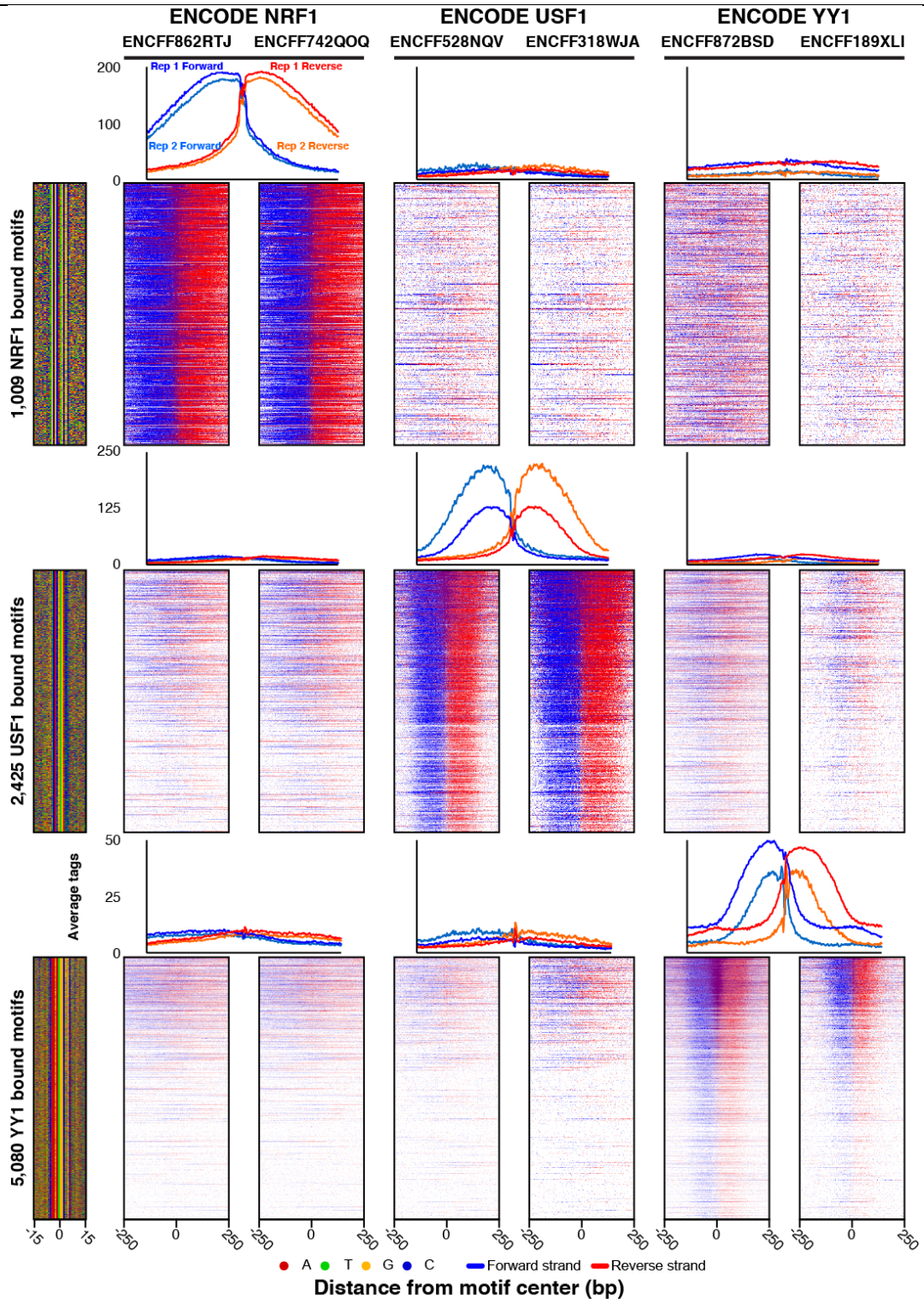


Figure 2. Comparison of ChIP-exo data at cognate vs non-cognate motifs. ChIP-exo heatmap, composite, and DNA-sequence 4-color plots were generated for NRF1, USF1, YY1, and IgG ChIP-exo datasets against the complete matrix of bound motifs from **Figure 1**. The 5' end of aligned sequence reads for each set of experiments were plotted relative to distance from cognate motif for each indicated target. Reads are strand-separated (blue = motif strand, red = opposite strand). Rows are linked across samples and sorted based on their combined average rank-order in a 100 bp bin around each motif midpoint. High levels of background result in a more uniform distribution of reads across the genome as in the IgG control.



Supplementary Fig. 4. Independent hybridoma clones and target interaction partners as antibody validation criteria. ChIP-exo heatmap, composite, and DNA-sequence 4-color plots are shown for the indicated number and type of bound motifs for (**a,b**) the indicated antibody hybridoma clones or (**c**) interaction partners, tested in K562 cells. The 5' end of aligned sequence reads for each set of experiments were plotted against distance from cognate motif, present in the union of all called peaks between the datasets for each indicated target. Reads are strand-separated (blue = motif strand, red = opposite strand). Rows are linked across samples and sorted based on their combined average rank-order in a 100 bp bin around each motif midpoint.



Supplementary Fig. 5. Overlay of ENCODE data at motifs defined by PCRP mAbs ChIP-exo. ChIP-seq heatmap, composite, and DNA-sequence 4-color plots at the bound motifs defined in **Figure 1** for the indicated targets in K562. The 5' end of aligned sequence reads for each set of experiments were plotted against distance from cognate motif of target. Reads are strand-separated (blue = motif strand, red = opposite strand). Rows are linked across samples and sorted as in **Figure 1**.

202 **Assessment by ChIP-seq**

203 Since ChIP-seq is a widely used assay (and related to ChIP-exo), we performed the assay
204 in HCT116 cells with 137 PCRP hybridomas corresponding to 70 targets associated with
205 chromatin binding / modification, enhancer function, or transcriptional elongation. We found 19
206 (14%) produced peaks (see Methods) (*Supplementary Table 1*). To provide highly stringent
207 validation for NRF1, NRF1 expression was knocked down in HCT116 cells by RNAi, and
208 detected NRF1 peaks were concomitantly diminished (**Figure 3a**), thereby demonstrating
209 specificity of the PCRP mAbs 3D4 and 3H1for this target.

210 Three hybridomas -- JAZF1-1C2, FOS-1A7, SMAD2-1B10 -- gave unexpected ChIP-seq
211 patterns for their intended targets; namely enrichment at transcription start sites (TSS in **Figure**
212 **3b**). Since JAZF1 is not expressed in HCT116 cells[23], JAZF1-1C2 mAb may have had off-target
213 interactions in ChIP-seq. This was not observed with a second mAb, JAZF1-1C1 (ChIP-seq and
214 ChIP-exo). SMAD2[24] and FOS are expected to be at enhancers. However, both ChIP-seq and
215 ChIP-exo showed enrichment at TSS, rather than at enhancers (see also online data). Deeper
216 biochemical validation (such as RNAi) will be needed to verify their specificity.
217

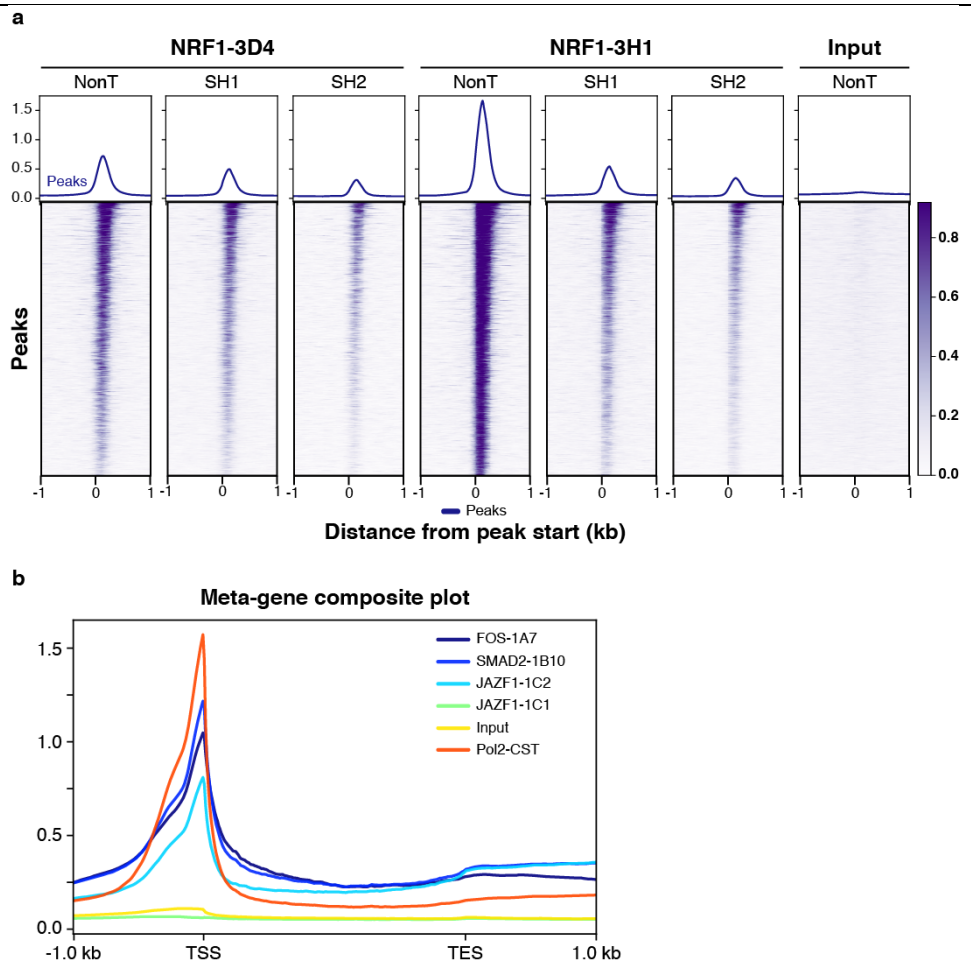


Figure 3. (**a**) Heatmaps and composite plots displaying the global loss of NRF1-3D4 and NRF1-3H1 ChIP-seq signal after NRF1 RNAi. ChIP-seq in HCT116 cells treated with non-targeting (sh Control) or two different NRF1-directed shRNAs (shRNA 1 and shRNA 2). (**b**) Composite plots of FOS-1A7, SMAD2-1B10, and JAZF1-1C1. Read counts are plotted along a linear x-axis, except between the TSS and TES (transcript end site) of gene bodies (N=7,309 genes).

218 **Validation through feature enrichment**
219 Thus far we have established the utility of three independent validation strategies inherent
220 to ChIP-analysis: 1) enrichment and patterning at a cognate motif; 2) correlation with an
221 independent mAb clone; and 3) co-localization with an interacting partner. In our large-scale
222 evaluation of mAbs, validation was often not attainable by all three approaches. We therefore
223 looked for additional criteria that might be useful where the preferred validation approaches were
224 inconclusive. The ChExMix algorithm was used to identify significant modes of protein binding
225 and *de novo* motif detection through a combination of DNA sequence enrichment and alternative
226 ChIP-exo patterns[25, 26]. Discovered motifs were identified using TOMTOM and the JASPAR
227 database[27, 28]. Relative enrichment in annotated genomic regions (e.g. promoters, enhancers, and
228 insulators) may be suggestive of function. These included chromHMM and Segway genome
229 segmentations (*Supplementary Fig. 3*)[29]. We also considered a low stringency test that did not
230 require peak-oriented statistical enrichment. Composite plots were generated around well-
231 defined general features like transcription start sites and CTCF binding sites and the average tag
232 enrichment was examined relative to a negative control (IgG) background.
233 These additional validation strategies generally offered less confidence compared to *a*
234 *priori* cognate motif validation because many of the defined chromatin states are relatively
235 abundant in the genome. Nevertheless, they provide a useful screening step for pursuing
236 additional validation and optimization. As examples, Rfx3 was enriched at HepG2 promoters and
237 enhancers, despite lacking *de novo* discovery of a motif; CTBP1 was primarily enriched in K562
238 heterochromatin; ZNF438 was enriched around K562 transcriptional start sites (TSS). Validation
239 metrics and analyses for each tested antibody can be found at www.PCRPvalidation.org and
240 *Supplementary Table 2***.**
241
242 **Validation through motif analysis.**
243 To further evaluate the ChIP-exo and ChIP-seq data for evidence of TF genomic
244 occupancy (direct or indirect), we analyzed 259 ChIP-exo and 21 ChIP-seq peak files for: (a)
245 motif enrichment by an AUROC metric in which ChIP 'bound' regions are compared to a
246 background set of unbound sequences, and (b) motif distance to the peak summit, where
247 centering indicates the motif of a directly bound putative TF (supporting primarily direct binding
248 to DNA by the profiled TF) or that of another putative TF (suggesting indirect binding to DNA
249 via a tethering TF)[30-33] (**Figure 4a,b**). We used a collection of 100 nonredundant position weight
250 matrices (PWMs) representative of the known repertoire of human TF binding specificity[32].
251 These approaches identified 23 PCRP antibodies, corresponding to 19 sequence-specific putative
252 TFs, for which their cognate DNA motif was both enriched and centered within the ChIP peaks
253 ("Direct Binding" in *Supplementary Table 3*).

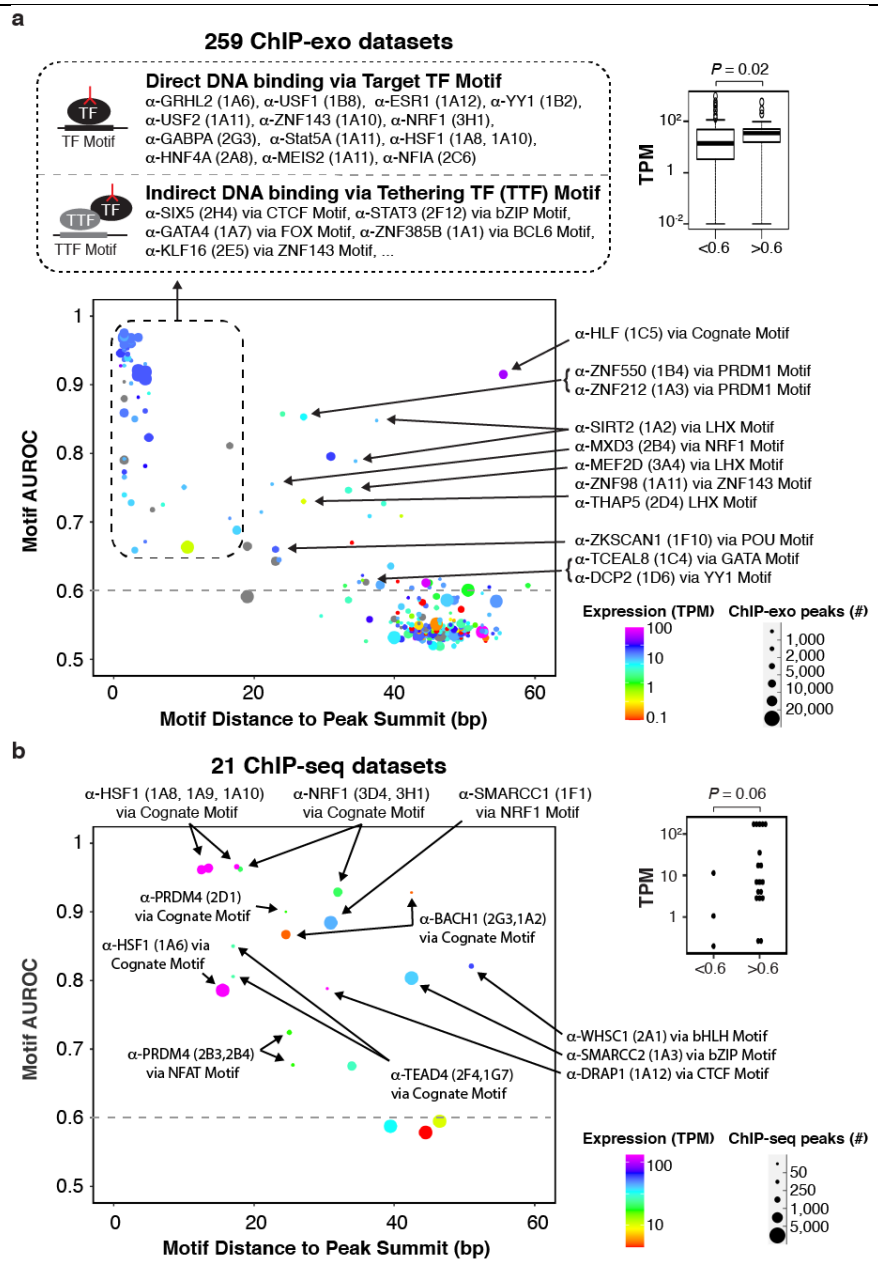
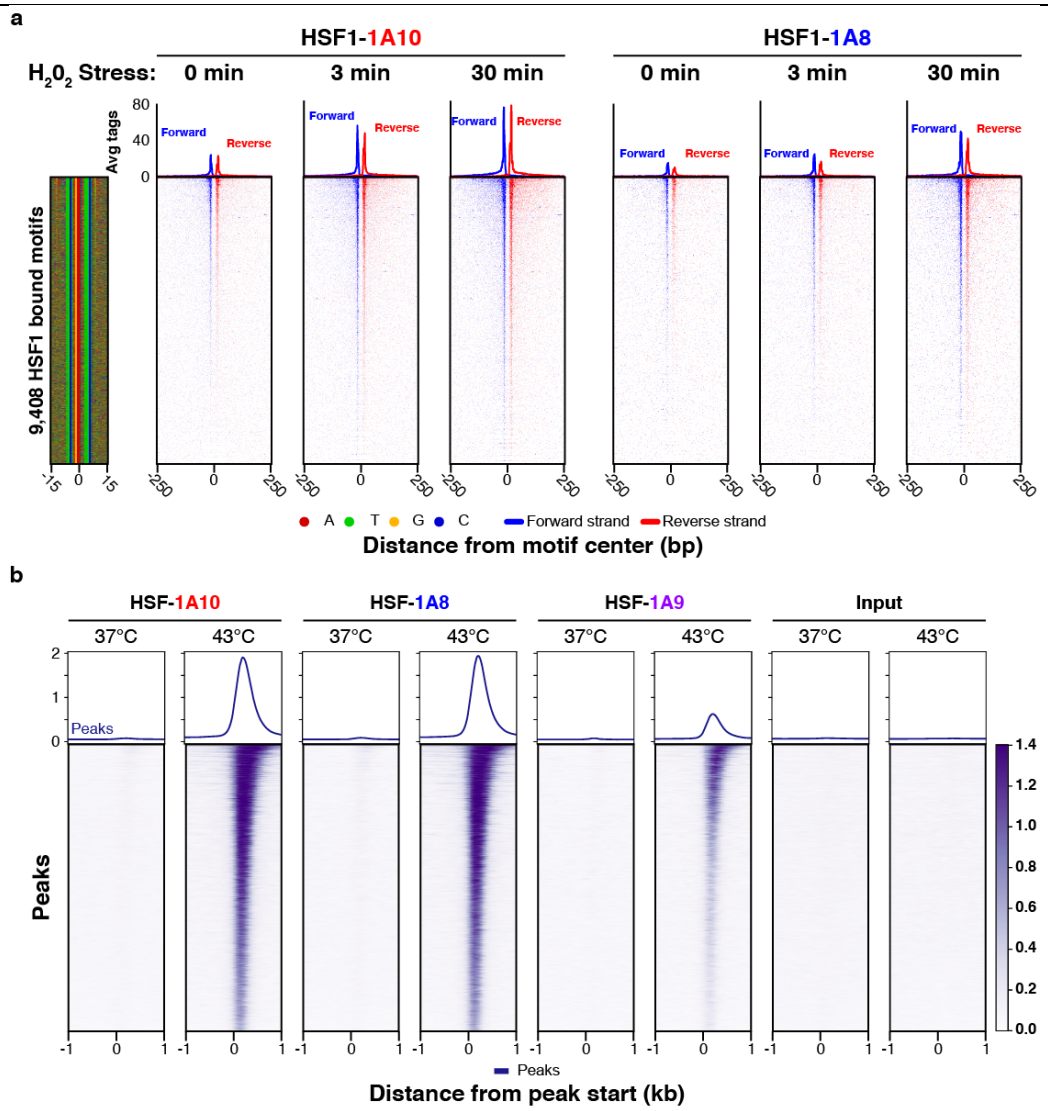


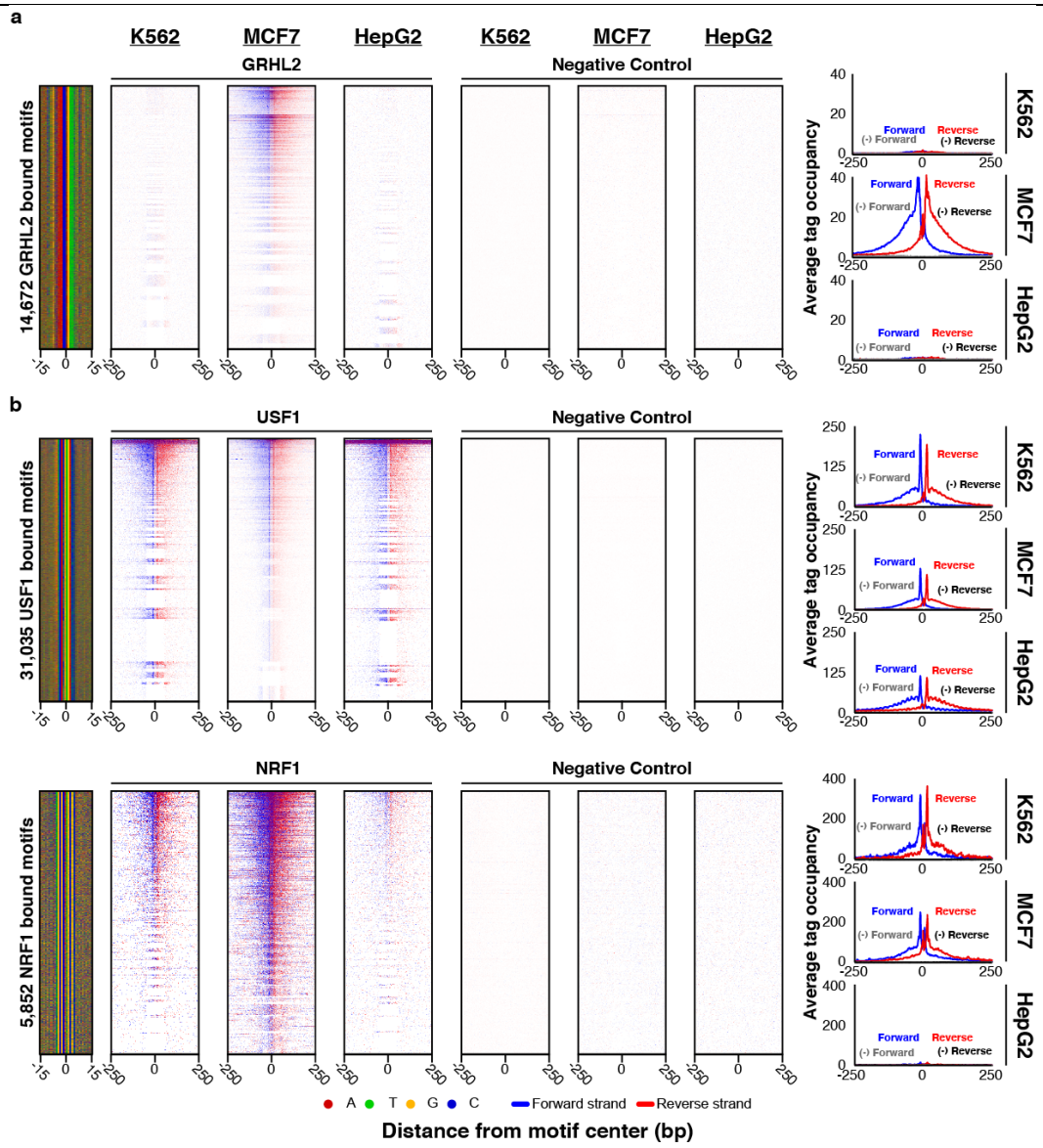
Figure 4. Motif enrichment analysis of ChIP-exo and ChIP-Seq datasets. **(a)** Cartoons depict models for direct binding via the cognate motif of the target TF or indirect binding via a sequence-specific tethering TF ("TTF"). Box plots of TPM expression values of target TFs associated to antibodies stratified by AUROC value. Results from analysis of 99 putative TF binding motifs within each ChIP-exo dataset with >500 peaks (259 datasets in total). We assigned to each ChIP-exo dataset the PWM with the highest AUROC ("Top Motif") and quantified its centering as the mean distance of the PWM match from the peak's summit. In the scatter plot, each point represents the enrichment/centering of the top motif in one of the 259 putative TF ChIP-exo datasets. Colors indicate the expression level (RNA-seq TPM value[7]; unavailable values are shown in gray) of the gene specific for the antibody used in the ChIP assay. Point sizes indicate the number of ChIP-exo peaks in the dataset. Top motifs with AUROC >0.6 (dashed line) and TPM values from duplicate RNA-seq experiments are indicated. **(b)** Results from enrichment analysis of 99 TF binding motifs within each of 21 differential ChIP-seq datasets. Points are formatted as in (**a**). Top right inset: dotplots were generated as in (**a**).

254 Of the analyzed datasets, an additional 28 PCRP mAbs enriched for motifs that did not
255 match a cognate motif of the targeted TFs ("Indirect Binding" in *Supplementary Table 3*). For 17
256 of these TFs (ATAD2, DCP2, DRAP1, HMGB1, MXD3, NMI, PPP1R10, SIRT2, THAP5,
257 TCEAL8, WHSC1, ZNF550, ZNF98, ZNF212, ZNF385B, ZNF550, ZZZ3) there are (to our
258 knowledge) no reported DNA binding motifs[34]; our results provide support for their genomic
259 occupancy being mediated at least in part via a putative tethering TF. For example, the
260 enrichment of the GATA motif within ChIP-exo peaks for TCEAL8 in K562 cells agrees with
261 the known role of GATA1/2 TFs as lineage specifiers that recruit additional factors to their
262 bound regions[31]. Similarly, the enrichment of CTCF (in DRAP1 and HMGB1 peaks), RUNX (in
263 NMI and DDIT peaks), YY1 (in DCP2 peaks), NRF1 (in SMARCC1 peaks), bHLH (in WHSC1
264 peaks), bZIP (in SMARCC2 peaks) and ETS (in PPP1R10 peaks) motifs are consistent with an
265 indirect recruitment mechanism. Each are associated with well-expressed TFs in the underlying
266 cell lines (see color-coding for RNA-seq TPM values in **Figure 4a,b**). Indirect binding could
267 also explain the observed noncognate motif enrichment for some of the 13 mAbs raised against
268 sequence-specific TFs. For example, the enrichment and the centering of the FOX motif within
269 GATA4 peaks in HepG2 cells agrees with the known role of FOXA1/2 as lineage specifiers.
270 Similarly, the enrichment and centering of the bZIP motif (*e.g.*, AP1) in STAT3 peaks, CTCF
271 motif in SIX5 peaks, and ZNF143 motif in KLF16 peaks is consistent with tethering mechanisms
272 for binding by these factors in K562 cells, which would be a novel finding for SIX5 and KLF16.
273
274 **ChIP assessment in multiple cell states and types**
275 Any number of targets may be sequestered in a state that prevents their interaction with
276 chromatin (and thus detection by ChIP), unless activated by specific signaling: i.e. a change in
277 cell state. We addressed this with HSF1, which is rapidly induced to bind in the nucleus upon
278 heat shock to activate heat shock response genes[35]. HSF1 was bound to cognate motifs at
279 relatively low levels under non-stressed conditions, but increased in binding upon treatment of
280 cells with hydrogen peroxide (0.3 mM) for 3 min and 30 min. (*Supplementary Fig. 6a*), or upon
281 heat shock (shift from 37˚C to 42˚C for 1 hr) (*Supplementary Fig. 6b*).
282



Supplementary Fig. 6. Validation of HSF1 mAb in a cell state change (stress). (**a**) ChIP-exo heatmap, composite, and DNA-sequence 4-color plots are shown for the indicated number of bound motifs for the indicated mAb in response to hydrogen peroxide treatment (0.3mM); where binding increases with treatment time) in K562 cells. The 5' end of aligned sequence reads for each set of experiments were plotted against distance from cognate motif, present in the union of all called peaks between the datasets for each indicated target. Reads are strand-separated (blue = motif strand, red = opposite strand). Rows are linked across samples and sorted based on their combined average in a 100 bp bin around each motif midpoint. (**b**) ChIP-seq heatmap and composite plot are shown for the indicated number of bound loci for the indicated antibody hybridoma clone and input in HCT116 cells in response to 1 hr. of heat shock (42˚C) or mock (37˚C).

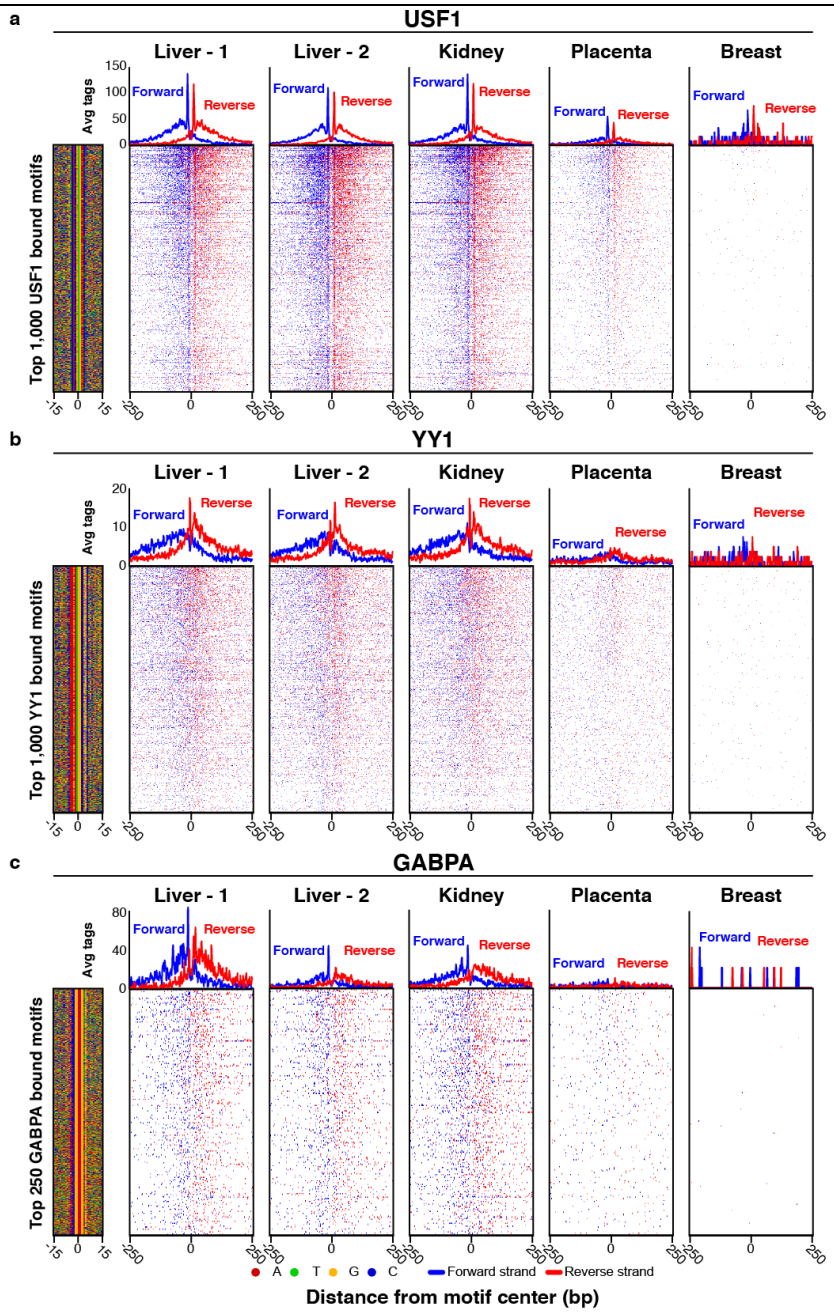
While ChIP-exo antibody validation was primarily performed in K562 cells (see above) any number of the targets may only interact with chromatin in other cell types. An example of cell-type specific expression was observed with the breast cancer factor GRHL2, where binding was detected in MCF7 cells but not in other cell types (*Supplementary Fig. 7a*). Other targets like USF1 and NRF1 did not appear to possess obvious cell type-specific binding, although we do not exclude selective cell-type enrichment at subsets of sites (*Supplementary Fig. 7b*).



Supplementary Fig. 7. Cell line comparison of antibody performance. (**a**,**b**) ChIP-exo heatmap, composite, and DNA-sequence 4-color plots are shown for the indicated number of bound motifs for the indicated targets, in the indicated cell types. The 5' end of aligned sequence reads for each set of experiments were plotted against distance from cognate motif, present in the union of all called peaks between the datasets for each indicated target. Reads are strand-separated (blue = motif strand, red = opposite strand). Rows are linked across samples and sorted based on their combined average in a 100 bp bin around each motif midpoint.

We next examined the capability of some mAb for clinical specimens (donated organs). ChIP-exo was performed using PCRP mAbs against USF1, YY1, and GABPA in chromatin from human liver (2 different specimens), kidney, placenta, and breast tissue (*Supplementary Fig. 8*). Largely consistent with the cell line ChIP's, enrichment and aligned read patterning was observed with all three mAb's for the liver and kidney, that was diminished in placenta and not detectable in breast. While we suspect technical differences in chromatin yields from different

tissue types drove these differences, we cannot rule out effects of tissue specificity for these at this point. Nonetheless, these findings demonstrate the utility of at least some PCRP mAb in epigenomic profiling of human clinical specimens.



Supplementary Fig. 8. Application of ChIP-exo using renewable mAbs in human tissue. ChIP-exo heatmap, composite, and DNA-sequence 4-color plots are shown for the indicated number and type of bound motifs for the indicated targets, in the indicated organ types (liver includes two donors). The 5' end of aligned sequence reads for each set of experiments were plotted against distance from cognate motif, present in the union of all called peaks between the datasets for each indicated target. Reads are strand-separated (blue = motif strand, red = opposite strand). Rows are linked across samples and sorted based on their combined average in a 100 bp bin around each motif midpoint.

301 **Validation using CUT&RUN.**
302 CUT&RUN uses a tethering approach: antibodies are added to permeabilized cells or
303 nuclei, where they bind their targets and recruit antibody-binding pAG-MNase to cleave DNA in
304 the locality. The result is a selective release of chromatin from the otherwise insoluble nucleus,
305 where genomic enrichment can then be identified by sequencing. We tested ~40 PCRP
306 antibodies in K562 by native CUT&RUN of which only 25 had some level of enrichment
307 previously detected by ChIP-exo (see Methods). Upon examination, USF2-1A11 worked the best
308 out of the tested PCRP mAbs, as shown with its tag enrichment at previously defined ChIP-exo
309 bound motifs. (**Figure 5a**). In the same way, we also examined NRF1, USF1, and YY1
310 antibodies by CUT&RUN (**Figure 5b**). While we identified a robust USF1 CUT&RUN
311 footprint, we did not observe enrichment of NRF1 or YY1 relative to the IgG control. As all
312 three antibodies had performed well in alternate assays, this likely reflects a need for additional
313 assay optimization. Of note a DNA-accessibility footprint was observed with the IgG negative
314 control in these CUT&RUN experiments, indicating background cleavage by pAG-MNase at
315 open (i.e. TF-bound) locations. Thus, specificity in most cases could not be established. All
316 CUT&RUN samples were processed through the ChIP-exo bioinformatic pipeline
317 (*Supplementary Fig. 3*) and individual sample results are available in *Supplementary Table 4*.

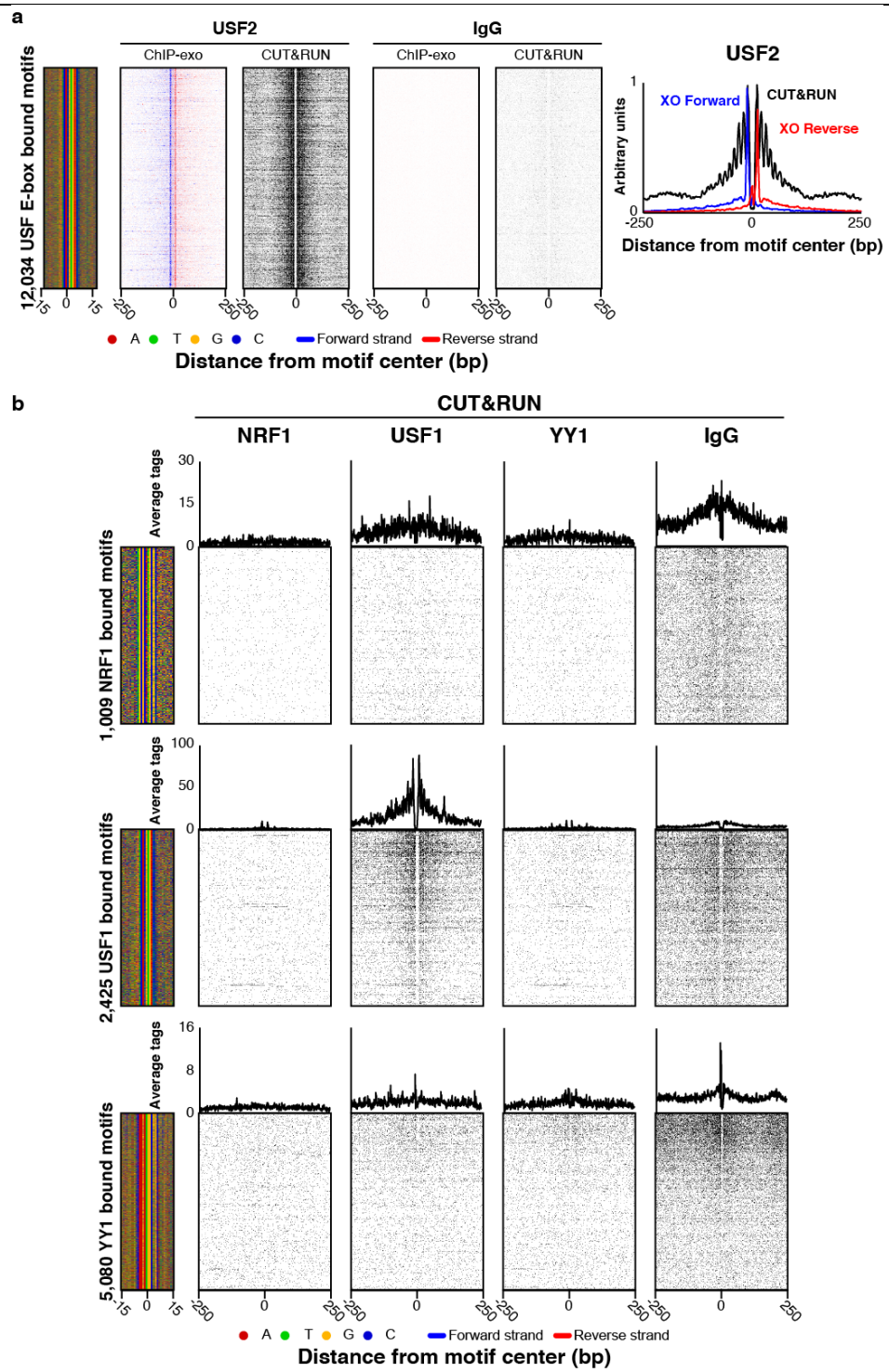
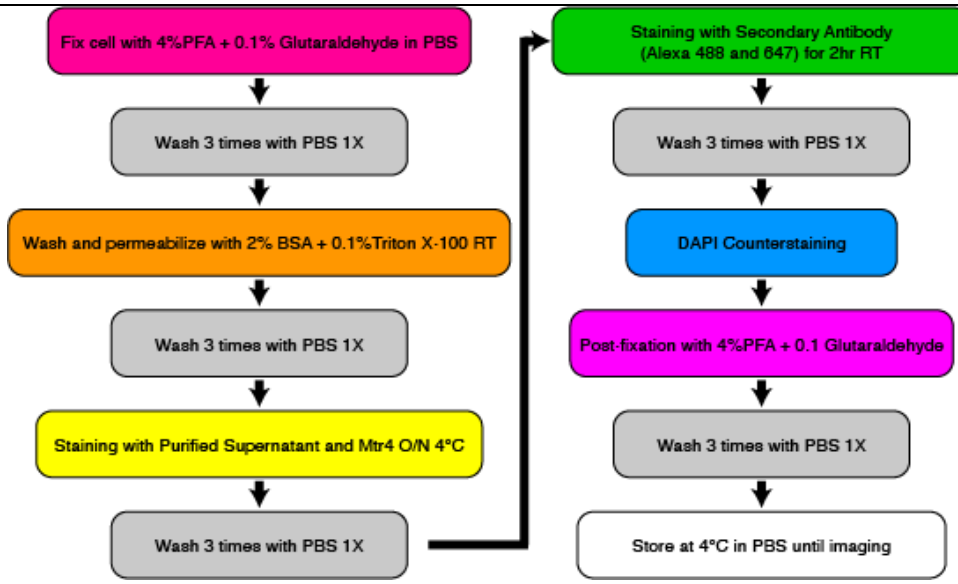


Figure 5. Enrichment validation of ChIP-exo using CUT&RUN. (**a**) CUT&RUN heatmap, composite, and DNA-sequence 4-color plots are shown relative to the peaks defined and sorted in **Figure 2**. The 5' end of aligned sequence reads for each set of experiments were plotted against distance from cognate motif. Reads are strand-separated (blue = motif strand, red = opposite strand) for ChIP-exo and combined (black) for CUT&RUN. (**b**) CUT&RUN heatmap, composite, and DNA-sequence 4-color plots were generated relative to the peaks defined in **Figure 2** for NRF1, USF1, YY1, and IgG datasets. Reads are aligned as above.

318 **Utility of PCRP mAb in STORM**

319 Super-resolution microscopy has permitted the visualization of cellular structures and
320 processes at nanometer resolution[36]. Such visualization involves the use of fluorescently
321 conjugated antibodies that bind to specific structures. In Stochastic Optical Reconstruction
322 Microscopy (STORM), the fluorophores conjugated to the antibodies are sequentially activated
323 and imaged in a time-resolved manner. Centroid positions of individually detected photons
324 provide 2-20 nanometer-scale resolution. This contrasts with conventional confocal microscopy,
325 which loses resolution below 200 nm. To evaluate PCRP mAbs in STORM we processed a
326 subset of 41 in K562 cells, having relatively high validation scores by ChIP-exo (*Supplementary*
327 *Fig. 9*). Briefly, cells were plated on glass dishes, fixed, and anti-MTR4 used to (stain the
328 nucleus and cytoplasm), providing a robust positive control[37]. The process was then repeated on
329 the same set of cells with individual PCRP mAbs. Samples were then visualized by STORM.
330



Supplementary Figure 9. Workflow of processing hybridoma supernates for imaging.

331

332 Of the 41 antibodies analyzed, only three produced images with substantial nuclear
333 localization, and undetectable in the cytoplasm, as would be expected for active TFs
334 (*Supplementary Table 5*). A key positive control is MED4, a component of the Mediator
335 transcription co-factor, which has been demonstrated to form nuclear puncta[38]. As in **Figure 6a**,
336 both Mtr4 and MED4 were detected in the nucleus of K562 cells. In contrast, PCRP antibodies
337 against other factors (e.g., Mdx4) showed cytoplasmic staining (**Figure 6b**). This may be due to
338 Mdx4 being cytoplasmic, or alternatively that the Mdx4 PCRP mAb is not specific in these
339 experiments. The low validation rate indicates that care should be taken in evaluating their
340 general usage for STORM.

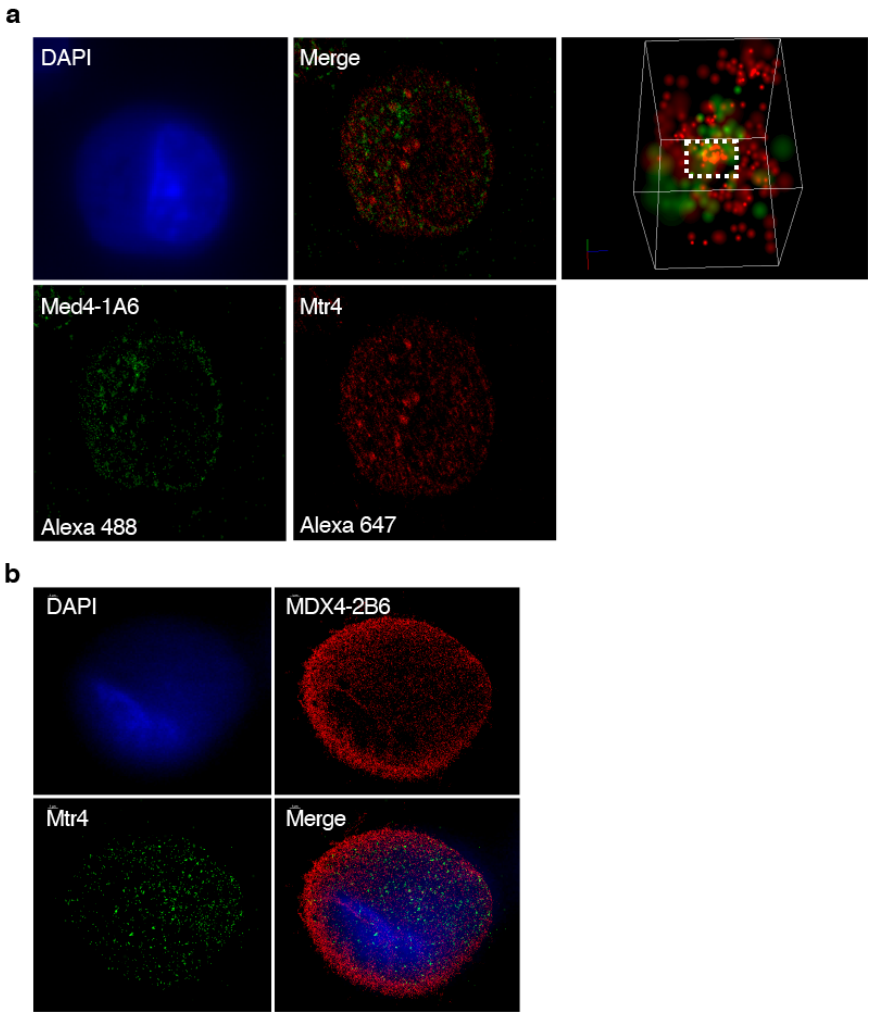
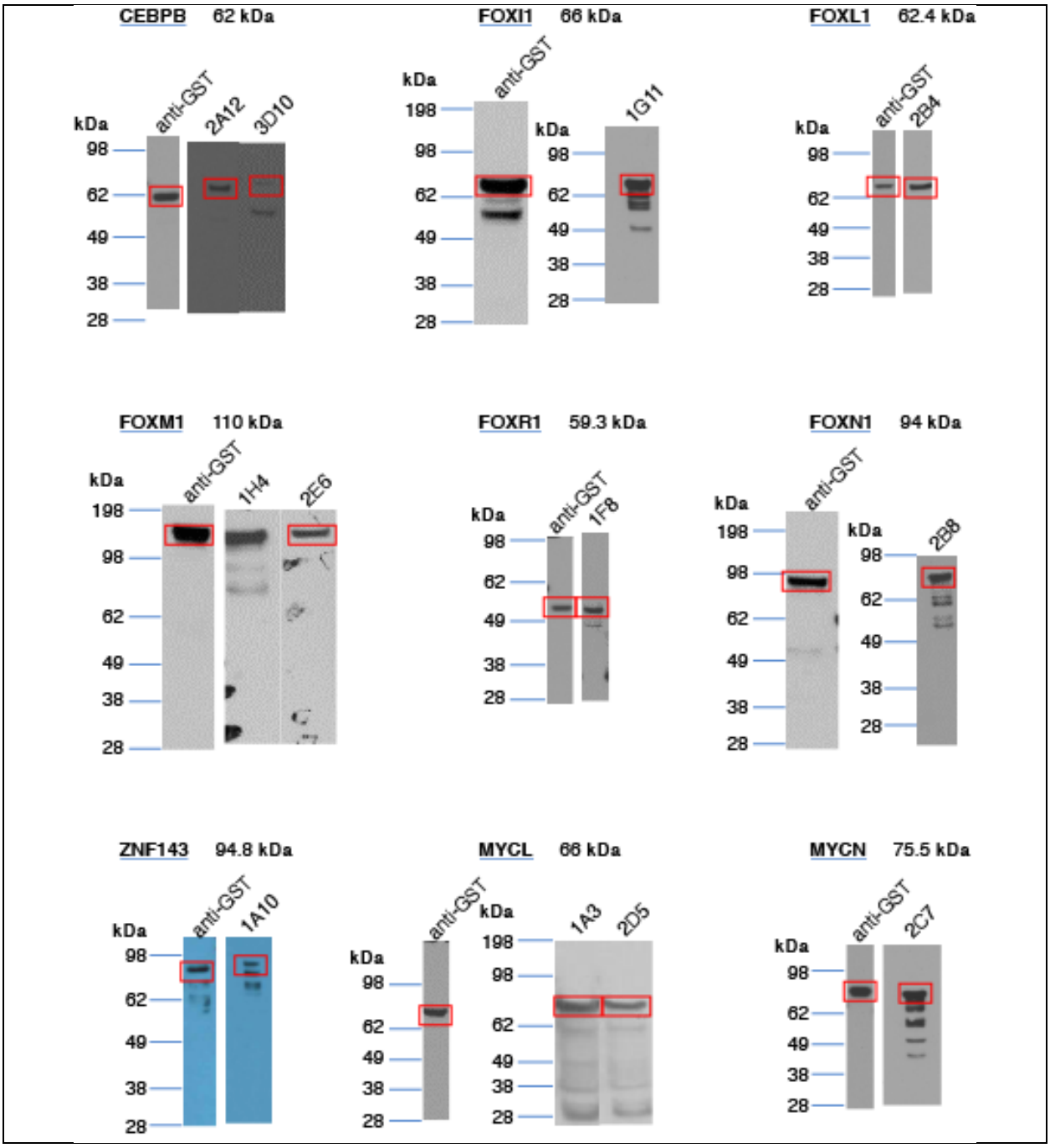


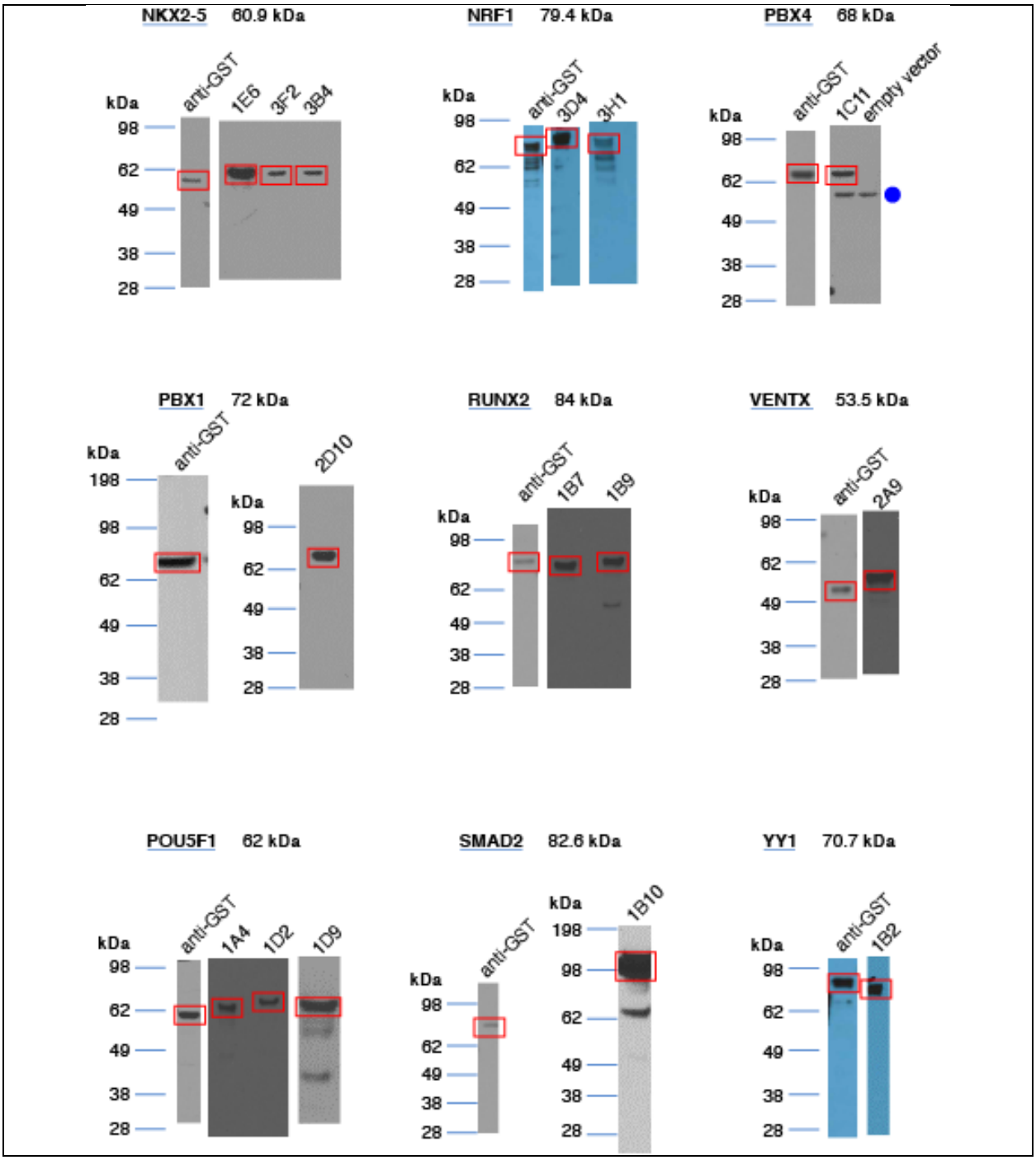
Figure 6. Example of STORM capabilities employing three different fluorescence channels examining one K562 cell for nuclear localization. (**a**) Upper left: DAPI (405nm) to localize the nucleus. Lower left: Staining of cell with concentrated PCRP supernatant (Med4, clone 1A6), with secondary antibody conjugated to Alexa-488 (green). As the mediator complex (Med4) interacts with DNA-binding gene-specific TFs to modulate transcription by RNA polymerase II it would be expected to be found in the nucleus. Lower right: Staining of cell with commercial anti-Mtr4 (a helicase; expected to be found in the nucleus and used as a standard STORM positive control[37]. Secondary anti-rabbit conjugated to Alexa-647 (red). Upper right: merged image. Far right: Zoom-in on merged image, demonstrating ability of STORM to spatially resolve the localization of individual molecules at sub-light microscopy resolution. (**b**) STORM with possibly non-specific antibody staining. Upper left: staining of K562 cell with DAPI for nuclear localization. Lower left. Stain with commercial anti-Mtr4 antibody Secondary antibody is coupled to Alexa 488 (green). Upper right: stain of cell with anti-Mdx4 (Max dimerization 4 protein). Secondary antibody is conjugated to Alexa 647 (red). Note that the antibody staining forms a ring at the periphery of the nucleus, possibly indicating non-specific binding. Lower right: merged image.

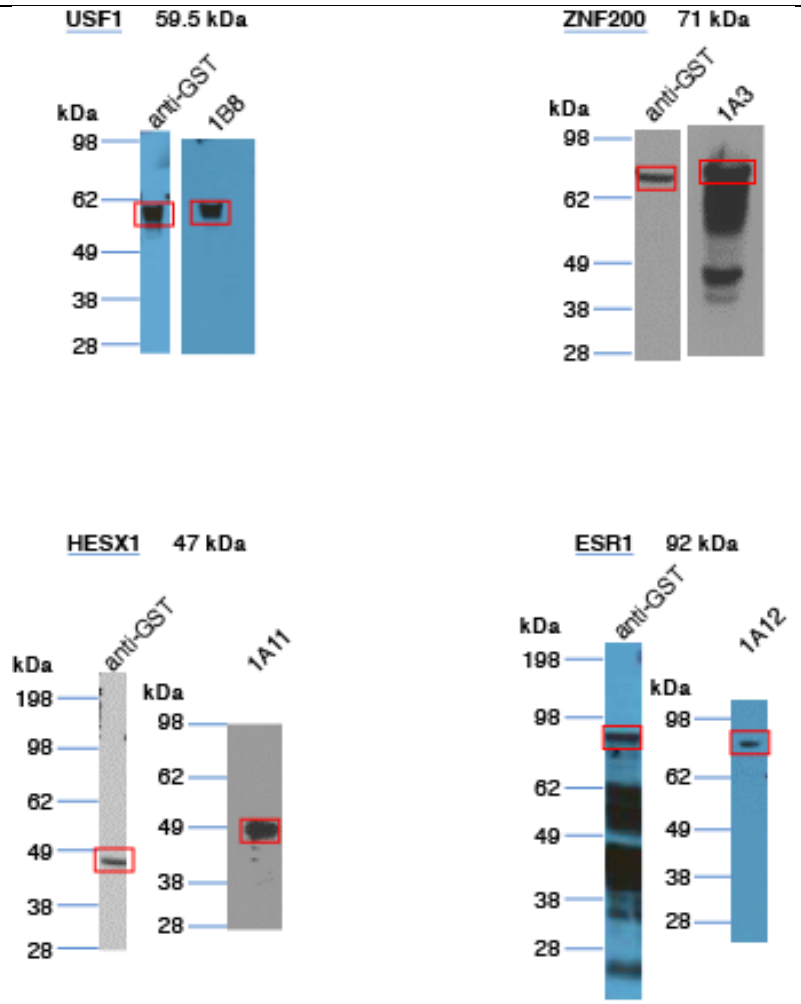
Utility of PCRP mAb in detecting recombinant protein in Western blots

Recombinant proteins are often expressed with an epitope tag for in-depth biochemical analysis. A common means of detecting specific proteins in a complex mixture is through size separation using denaturing gel electrophoresis (SDS-PAGE), followed by immunoblotting.

346 Here, we used coupled-in vitro transcription / translation (IVT) to produce 33 TFs as N-term
347 GST-fusion proteins (*Supplementary Table 6*), which we then used to assay a corresponding set
348 of 46 PCRP mAbs by immunoblotting. 31 of the 46 (67%) PCRP Abs detected a band of the
349 expected molecular weight (*Supplementary Fig. 10*). As a positive control, anti-GST antibody
350 detected all 33 GST-fusion TFs. Thus, despite limitations inherent in cell-based assays, the
351 majority of the PCRP mAbs are expected to be active in recognizing their target proteins as
352 GST-fusions in Western blots.


CEBPB 62 kDa
anti-GST
2A12
3D10
kDa
98
62
49
38
28
FOXI1 66 kDa
anti-GST
1G11
kDa
198
98
62
49
38
28
FOXL1 62.4 kDa
anti-GST
2B4
kDa
98
62
49
38
28
FOXM1 110 kDa
anti-GST
1H4
2E6
kDa
198
98
62
49
38
28
FOXR1 59.3 kDa
anti-GST
1F8
kDa
98
62
49
38
28
FOXN1 94 kDa
anti-GST
2B8
kDa
198
98
62
49
38
28
ZNF143 94.8 kDa
anti-GST
1A10
kDa
98
62
49
38
28
MYCL 66 kDa
anti-GST
1A3
2D5
kDa
198
98
62
49
38
28
MYCN 75.5 kDa
anti-GST
2C7
kDa
98
62
49
38
28


NKX2-5 60.9 kDa
anti-GST
1E6
3F2
3B4
kDa
98
62
49
38
28
NRF1 79.4 kDa
anti-GST
3D4
3H1
PBX4 68 kDa
anti-GST
1C11
empty vector
PBX1 72 kDa
anti-GST
2D10
198
RUNX2 84 kDa
anti-GST
1B7
1B9
VENTX 53.5 kDa
anti-GST
2A9
POU5F1 62 kDa
anti-GST
1A4
1D2
1D9
SMAD2 82.6 kDa
anti-GST
1B10
YY1 70.7 kDa
anti-GST
1B2



Supplementary Fig. 10. Immunoblots with PCRP mAb. The indicated full-length GST-tagged targets were synthesized by *in vitro* transcription and translation (IVT), resolved by SDS-PAGE, electroblotted to a membrane, and probed with hybridoma supernatant from the indicated PCRP hybridoma clone ID or with anti-GST (positive control). The red boxes indicate the correct bands corresponding to the sizes of the full-length TFs. The blue dot indicates cross-reactivity of the secondary antibody with the 1-Step Human Coupled IVT Kit. The expected molecular weight of each target is indicated above each immunoblot. Species located below the full-length protein may be protein degradation or incomplete synthesis.

Protein binding microarrays (PBMs) are a high-throughput technology to assay the DNA binding specificities of proteins or protein complexes[39-41]. Proteins used in PBMs are typically expressed as epitope tag-fusions, which permits detection on the array with a fluorescent anti-tag antibody. Broader availability of TF-specific antibodies would increase the toolkit of reagents for protein detection in PBMs. Furthermore, a possible explanation for why some antibodies may fail to work in ChIP experiments is that their target epitope may become inaccessible when the TF is bound to an accessory protein / DNA and/or subjected to modification by formaldehyde. Therefore, for a set of 31 TFs that were of interest or performed well in ChIP, we used PBMs to test 44 PCRP mAb for their ability to recognize their DNA-bound target TF. Briefly, we expressed the relevant TFs by IVT and used each in universal PBM experiments, where all 10-bp sequences were represented within ~44,000 60-bp probes on double-stranded Agilent

365 oligonucleotide arrays[39]. For 20 of these 44 PCRP Abs (45%), the PBM experiments
366 successfully identified a DNA binding motif consistent with the known or anticipated element
367 (**Figure 7,** *Supplementary Fig. 11*). Of the 20 PCRP antibodies that successfully yielded the
368 expected motif in PBMs, 11 had at least some validation support by ChIP.
369

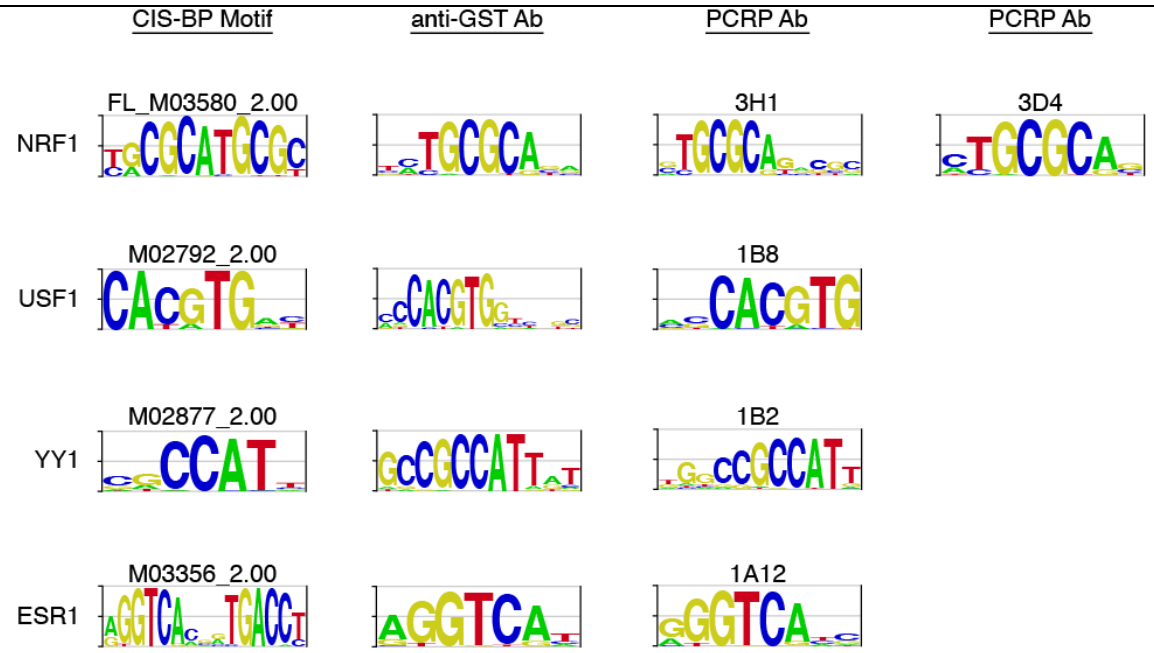
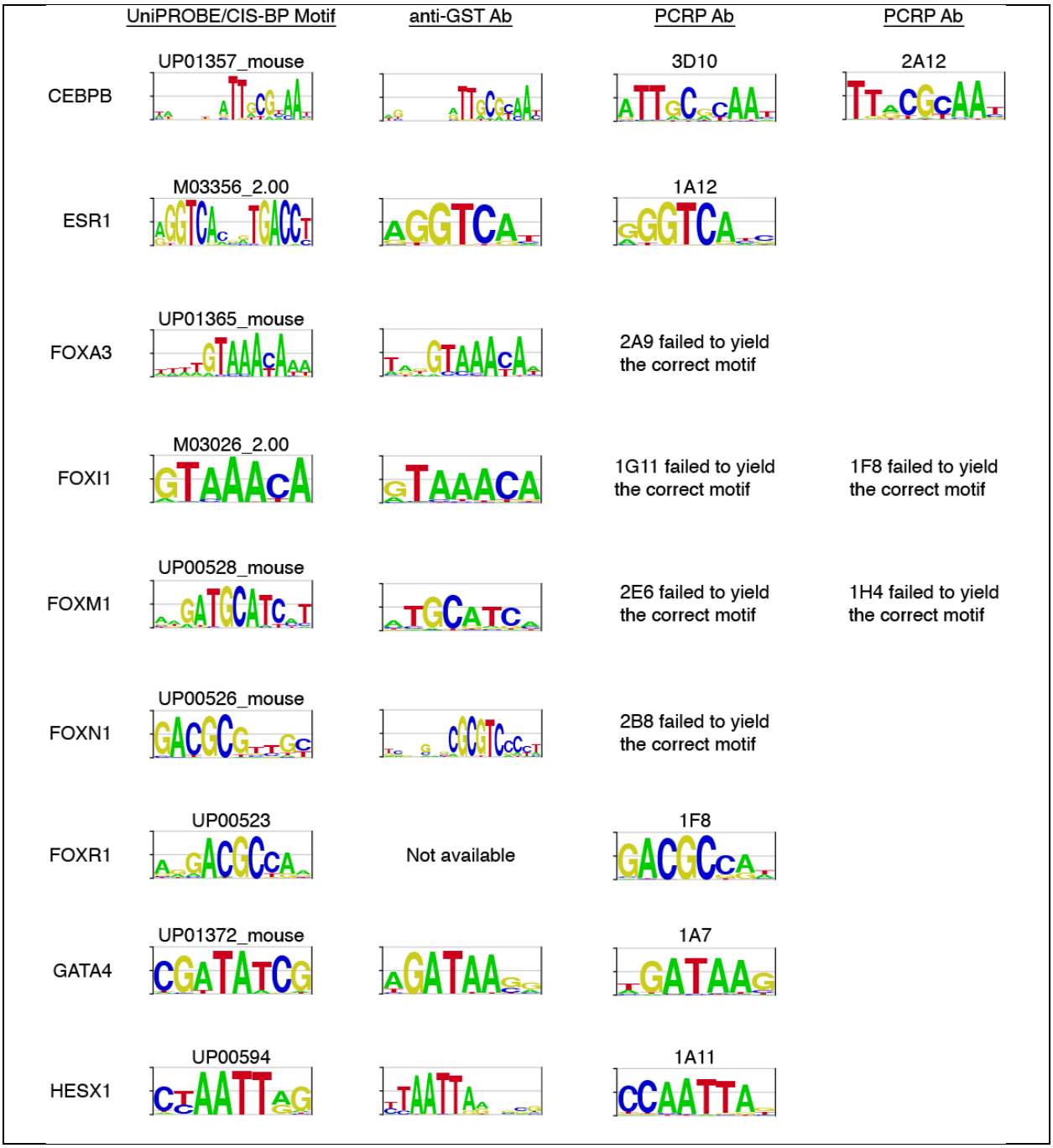
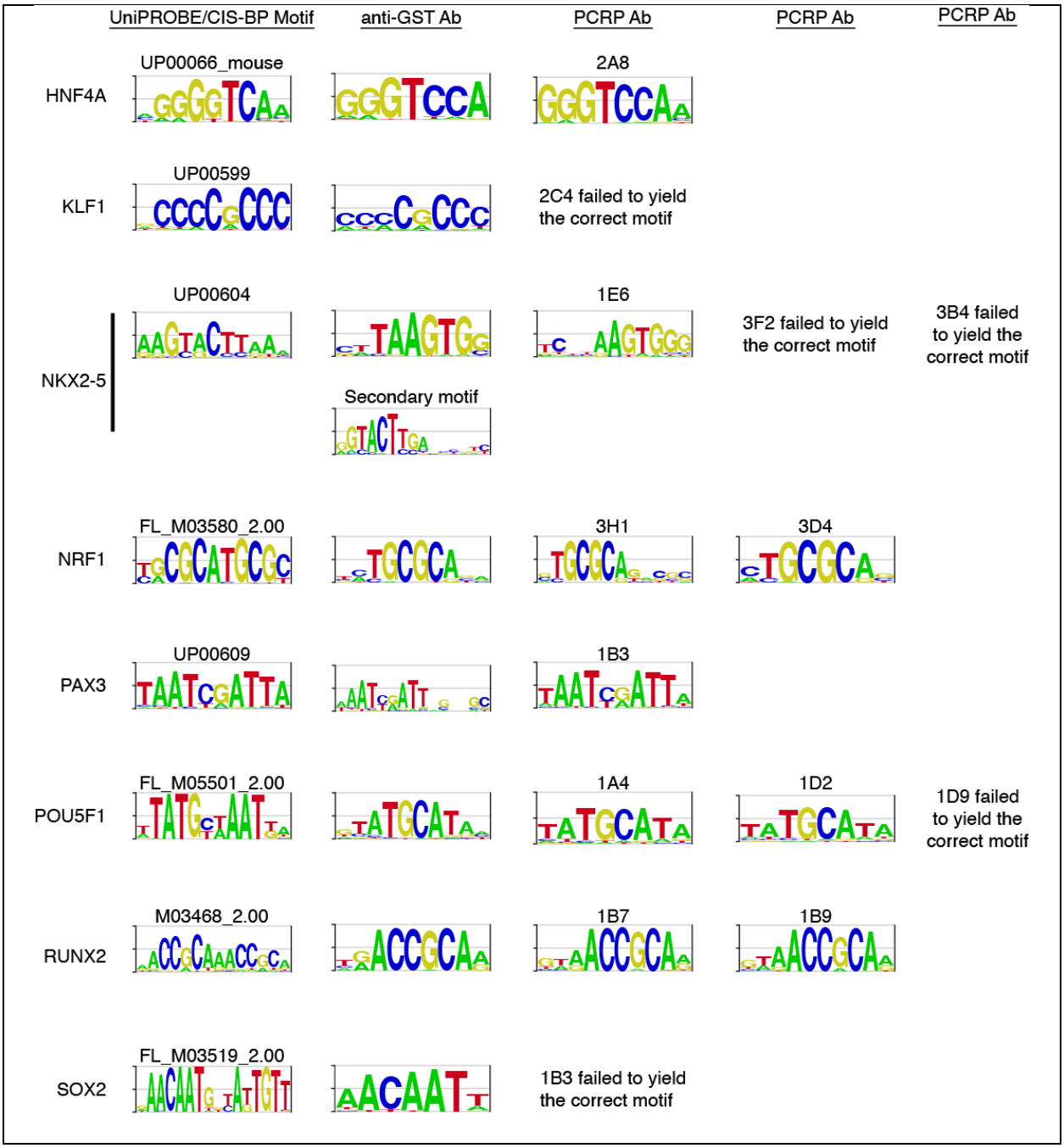
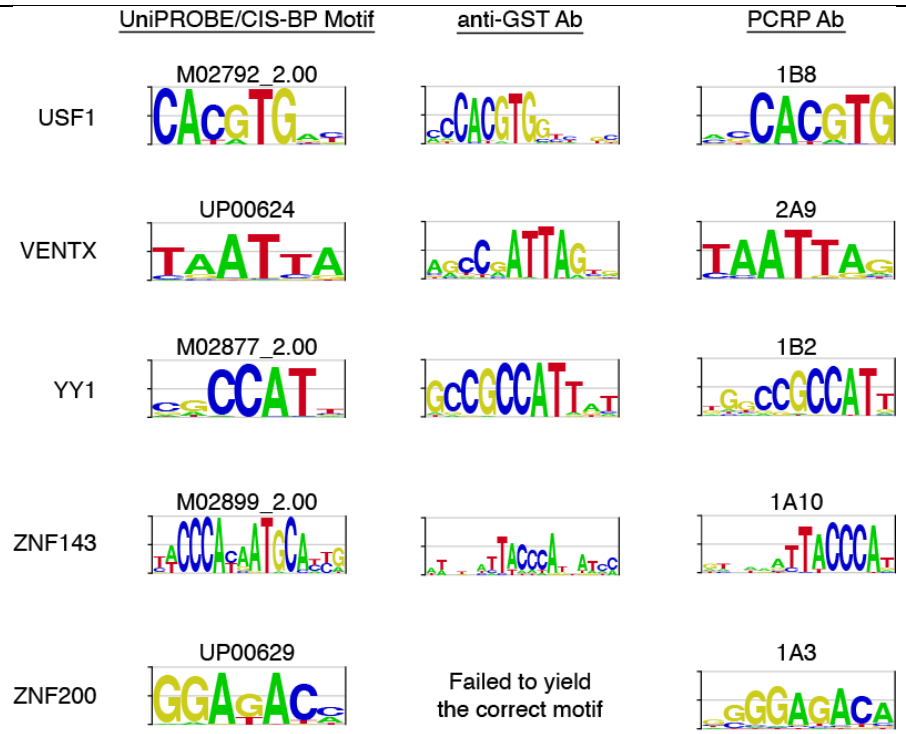


Figure 7. Examples of TF binding motifs derived from PBM experiments performed using anti-GST or PCRP antibodies. Transcription factors were assayed with their PCRP mAbs to compare their motif against their corresponding motif derived from an anti-GST PBM experiment and/or expected motif from the CIS-BP database[42]. NRF1, USF1, YY1, and ESR1 assayed with Alexa488-conjugated anti-GST antibody or, separately, with the indicated PCRP mAbs, yielded DNA binding motifs consistent with those from CIS-BP. For display of sequence motifs, probability matrices were trimmed from left and right until 2 consecutive positions with information content of 0.3 or greater were encountered, and logos were generated from the resulting trimmed matrices using enoLOGOS (DOI: 10.1093/nar/gki439)

370


UniPROBE/CIS-BP Motif
anti-GST Ab
PCRP Ab
PCRP Ab
CEBPB
UP01357_mouse
3D10
2A12
ESR1
M03356_2.00
1A12
FOXA3
UP01365_mouse
2A9 failed to yield the correct motif
FOXI1
M03026_2.00
1G11 failed to yield the correct motif
1F8 failed to yield the correct motif
FOXM1
UP00528_mouse
2E6 failed to yield the correct motif
1H4 failed to yield the correct motif
FOXN1
UP00526_mouse
2B8 failed to yield the correct motif
FOXR1
UP00523
Not available
1F8
GATA4
UP01372_mouse
1A7
HESX1
UP00594
1A11


UniPROBE/CIS-BP Motif
anti-GST Ab
PCRP Ab
PCRP Ab
PCRP Ab
HNF4A
UP00066_mouse
2A8
KLF1
UP00599
2C4 failed to yield the correct motif
NKX2-5
UP00604
1E6
3F2 failed to yield the correct motif
3B4 failed to yield the correct motif
Secondary motif
NRF1
FL_M03580_2.00
3H1
3D4
PAX3
UP00609
1B3
POU5F1
FL_M05501_2.00
1A4
1D2
1D9 failed to yield the correct motif
RUNX2
M03468_2.00
1B7
1B9
SOX2
FL_M03519_2.00
1B3 failed to yield the correct motif



Supplementary Fig. 11. TF binding motifs derived from PBM experiments performed using anti-GST or PCRP antibodies. Each full-length TF was assayed with its PCRP mAb(s) and compared against its corresponding motif derived from an anti-GST PBM experiment and anticipated motif from the UniPROBE or CIS-BP databases. The TFs from UniPROBE or CIS-BP were assayed as extended DNA binding domains. For display of sequence motifs, probability matrices were trimmed from left and right until 2 consecutive positions with information content of 0.3 or greater were encountered, and logos were generated from the resulting trimmed matrices using enoLOGOS (DOI: 10.1093/nar/gki439)

Discussion

The ability to interrogate the diverse human proteome is often dependent on the use and specificity of affinity capture reagents, of which antibodies are the most widely used. PCRP was intended to be a pilot project for the entire human proteome, but with a focus on nuclear proteins. To this end, nearly all PCRP mAb against ~700 putative chromatin targets or TFs were assayed genome-wide by ChIP-exo. A smaller subset was analyzed by other genome-wide location assays (ChIP-seq, CUT&RUN), by super-resolution microscopy (STORM), and in detection of recombinant target protein in Western blots and bound to DNA in PBMs. The purpose of this antibody screen is to present a technical assessment of these reagents and to provide a 'first-pass' assessment of those that demonstrate utility in one or more biochemical or cellular assays. To this end, we report those for which we obtained successful assay outcomes either "as is" as hybridoma supernates or after further processing (see Methods).

A large number of antibodies did not validate by the assays and criteria we employed in this study. However, proteins lacking known or putative DNA binding specificities may not have been accommodated within our discovery framework. For example, a target protein may bind a class of DNA sequences that was not captured by current motif discovery algorithms, or alternatively may interact with a genomic target sites that contain a wide range of different DNA

389 sequence motifs, and are not restricted to a small number of motifs, as is typically observed for
390 sequence-specific DNA binding proteins. Another possible scenario is that a target protein might
391 only bind genomic regions that form an as-yet undiscovered chromatin class. We often found
392 motifs that were long, simple, semi-repetitive, and highly degenerate (See online). These are not
393 typical properties of sequence-specific DNA binding proteins and the ChIP-exo patterns at these
394 motifs were often quite distinct from well-validated targets. Historically, some of these locations
395 may have been set aside as within so-called "black-listed" regions. Whether these regions are
396 artifactual or have some unknown biology remains to be determined. While we accepted these
397 motifs as evidence of enrichment, we urge caution when interpreting such atypical binding
398 events.
399 We believe that some mAbs that produced data classed as marginal or with unconvincing
400 evidence of specificity might perform better with assay optimization, a different cell type, or
401 through a different metric or threshold of validity. Notably, many of the TFs whose
402 corresponding PCRP mAbs were screened by ChIP-exo in K562 cells, are not expressed at
403 appreciable levels, if at all, in this cell type. However as shown in *Supplementary Fig. 6*, even
404 where a target is abundantly expressed, it may not interact with chromatin (and thus escape
405 detection) unless activated to do so. Therefore, knowledge of the underlying biology of the target
406 can be critical in how ChIP quality is assessed.
407 It was not practical in our high-throughput ChIP-exo screen to profile each PCRP mAb
408 in a wide range of cell types. However, in future studies, such PCRP Abs will need to be assayed
409 in cells in which their associated TFs are expressed at higher levels, in order to make a more
410 definitive statement of their capability. Furthermore, some targets may simply not be crosslinked
411 to chromatin in the assayed cell type (or any cell type), making ChIP an inappropriate assay for
412 their antibody validation (or indeed use). Conversely, in the case of CUT&RUN, the addition of
413 cross-linking (versus the native approach used) may stabilize the association of TFs with their
414 DNA motifs through early sample processing steps. Other explanations for poor performance
415 may include epitope inaccessibility, either due to the presence of an interfering ligand (protein or
416 nucleic acid), or epitope modification (by enzymes or formaldehyde). Furthermore, unlike
417 engineered epitope tags, each target-specific antibody may have a substantially different affinity
418 for its cognate target. Therefore, we cannot rule out that at least some low- or non-performing
419 antibodies could perform better under different immunoprecipitation conditions. Still other
420 potential reasons for antibody non-performance may be due to lot expiration or mis-identification
421 along the supply chain.
422 In total, 950 unique hybridoma clones were tested in at least one of the assays. We
423 identified 57 clones (6%) that worked with high confidence in at least one assay. However, only
424 a very small portion of the validation spectrum has been explored. Using relaxed criteria, that
425 may reflect significant but off-target or unknown behavior, we find 40% (380/950) of the tested
426 PCRP mAb had at least some evidence in at least one assay. However, such marginal validation
427 requires deeper biochemical validation such as target depletion/deletion or negative control cell
428 lines for a more robust validation. Our analysis identifies suitable candidates. A detailed
429 summary of each assay's results along with all of the measured quality control metrics is
430 available in **Table 1,** along with an interactive searchable web-interface online at
431 www.PCRPvalidation.org.
432 We have described here a principled strategy in evaluating antibody performance across
433 various assays . This strategy may be useful in evaluating other antibodies in ChIP or other
434 experimental formats.
435

436 # Methods
437 **ChIP protocols**
438 *Antibodies*. 1,308 TF hybridomas were reported through the PCRP portal at the start of
439 this study (September, 2017). Hybridoma supernates were purchased from Developmental
440 Studies Hybridoma Bank (DSHB, U. Iowa, IA) as 1 ml aliquots. Monoclonal antibody (mAb)
441 concentration (average 36 ug/ml ELISA-quantified antibody) and preparation dates were
442 provided. Concentrated mAbs and their concentration were generously provided by CDI
443 laboratories (Mayaguez, PR).
444 *Cell material.* Cell stocks were obtained by the Pugh laboratory from ATCC. K562 were
445 grown in suspension using IMDM media and periodically checked for mycoplasma
446 contamination. HepG2 and MCF7 were grown as adherent cells in DMEM media. MCF7 cells
447 were additionally grown in phenol red-free DMEM and treated with Beta-estradiol 30 min prior
448 to cell harvest. Cells were pelleted, re-suspended in PBS, crosslinked with 1% formaldehyde for
449 10 min, and then quenched with a molar excess of glycine. Donated human organs were obtained
450 from NDRI (Philadelphia, PA) and then cryoground to fine powder using the SPEX cryomill
451 cyrogrinder. Frozen tissue powder was re-suspended in room-temperature PBS containing
452 formaldehyde to a final concentration of 1% and quenched with a molar excess of glycine. All
453 cells and tissue for ChIP-exo then proceeded through standard lysis and sonication protocol
454 described below after crosslink quenching. HCT116 cells (ATCC CCL-247) were grown in the
455 Shilatifard laboratory in DMEM supplemented with 10% FBS (Fisher Scientific, 35-015-CV).
456 70%–80% confluent HCT116 cells were heat shocked for 1 hr by adding pre-heated conditioned
457 media pre-heated to 42°C[37]. Heat shock and non-heat shock HCT116 cells were washed with
458 PBS before fixing with 1% formaldehyde (Sigma, 252549) in PBS for 15 minutes and processing
459 for ChIP-seq.
460 ChIP-exo testing was initially prioritized in K562, MCF7, and HepG2, using gene
461 expression values (FPKM in RNA-seq) generated from the ENCODE project as the basis for the
462 cell type used[7, 43]. Targets were assigned to the cell type most likely to express the protein of
463 interest. If no cell line had a clear high expression for a specific target (>25% FPKM relative to
464 all other considered cell lines), testing defaulted to K562. K562 was selected as the default due to
465 its status as a Tier 1 ENCODE cell line and the plethora of existing genomic data that could
466 orthogonally support any findings. Samples were processed in batches of 48 in 96 well plate
467 format. The PCRP-derived USF1 or NRF1 antibody served as a positive control for every
468 processed cohort as well as an IgG or "No Antibody" mock ChIP negative control. Crosslinked
469 sonicated chromatin from ~7 million cells was incubated with antibody-bound beads, then
470 subjected to the ChIP-exo 5.0 assay[20].
471 *ChIP-exo 5.0 assay*. Chromatin for ChIP-exo was prepared by resuspending crosslinked
472 and quenched chromatin in Farnham cell lysis buffer at a ratio of 25 million cells to 1 mL of
473 buffer for 20 min at 4°C. At the 10 min mark, cells were pushed through a 25G needle 5 times to
474 enhance cellular lysis. Nuclei were then isolated by pelleting at 2,500g for 5min. Nuclei were
475 resuspended in RIPA buffer (25 million cells to 1 mL of buffer) for an additional 20 min at 4°C
476 and then pelleted again at 2,500g for 5 min. Disrupted nuclei were then finally resuspended in
477 1X PBS (25 million cells to 1 mL of buffer) and sonicated for 10 cycles (30on/30off) in a
478 Diagenode Pico. Solubilized chromatin was then processed through ChIP-exo. Production-scale
479 ChIP-exo 5.0 was generally performed in batches of 48 in a 96-well plate, alternating every
480 column to reduce risk of cross-contamination[18]. Briefly, solubilized chromatin was incubated
481 with Protein A/G Dynabeads, preloaded with 3 ug of antibody, overnight then sequentially

482 processed through A-tailing, 1st adapter ligation, Phi29 Fill-in, Lambda exonuclease digestion,
483 cross-link reversal, 2nd adapter ligation, and PCR for final high-throughput sequencing. Equal
484 proportions of ChIP samples were barcoded, pooled, and sequenced. Illumina paired-end read
485 (40 bp Read 1 and 36 bp Read 2) sequencing was performed on a NextSeq 500 and 550. While,
486 on average, we sought ~10 million total paired-end reads per ChIP, we accepted less if there was
487 strong evidence of target enrichment. Otherwise, we performed an additional round of
488 sequencing. The 5' end of Read_1 corresponded to the exonuclease stop site, located ~6 bp
489 upstream of a protein-DNA crosslink. Read_2 served two indirect functions: to provide added
490 specificity to genome-wide mapping, and to remove PCR duplicates.
491 *ChIP-seq.* 1x10^8 cells were fixed in 1% formaldehyde (Sigma, 252549) in PBS for 15-
492 20 min at room temperature and quenched with 1/10th volume of 1.25 M glycine for 5 minutes at
493 room temperature. Cells were collected at 1,000xg for 5 minutes, washed in PBS, pelleted at
494 1,000xg for 5 min and pellets were flash frozen in liquid nitrogen and stored at -80°C until use.
495 Pellets were thawed on ice and resuspended in 10 ml lysis buffer 1 (50 mM HEPES, pH 7.5, 140
496 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% IGEPAL CA-630, 0.25% Triton X-100 with
497 5µl/ml Sigma 8340 protease inhibitor cocktail incubated on ice 10 minutes, pelleted 1,500xg and
498 subsequently washed in lysis buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA,
499 0.5 mM EGTA and 5ul /ml protease inhibitor) as with lysis buffer 1 before resuspending in 1 ml
500 lysis buffer 3 (10 mM tris-HCl pH 8.0, 1 mM EDTA, 0.1% SDS and 5 ul/ml protease inhibitors)
501 for sonication as previously described[44]. Chromatin was sheared in a 1 ml milliTUBE with AFA
502 fiber on a Covaris E220 using 10% duty factor for 2 min. The sheared chromatin concentration
503 was estimated with Nanodrop at OD 260 and diluted to 1 mg/ml in ChIP Dilution Buffer (10%
504 Triton X-100, 1M NaCl and 1% Sodium deoxycholate). 1 mg chromatin was combined with 4
505 µg hybridoma tissue culture supernatant and rotated overnight at 4°C. 40 µl protein A/G
506 Dynabeads was added and incubated for 2-4 hr rotating at 4°C. Samples were washed 5 times
507 with 1 ml RIPA buffer, 2 times with TE with 50 mM NaCl. Chromatin was eluted with 800 µl
508 elution buffer (50 mM Tris pH 8.0, 1 mM EDTA, 0.1% SDS) for 30 min at 65°C shaking at
509 1,500 rpm in a ThermoMixer (Eppendorf). Supernatants were collected, digested with 20 µl of
510 20 mg/ml Proteinase K and incubated overnight at 65°C. DNA was purified with phenol
511 chloroform extraction. 500 µl of the aqueous phase was precipitated with 20 µl 5M NaCl, 1.5 µg
512 glycogen and 1 ml EtOH on ice for 1 hr or at -20°C overnight.
513 Sequencing libraries were prepared with the KAPA HTP library prep kit (Roche) using 1-
514 10 ng DNA and libraries were size selected with AMPure XP beads (Beckman Coulter). Illumina
515 50 bp single-end read sequencing was performed on a NextSeq 500 or NovaSeq 6000. Reads
516 were aligned to hg19 with Bowtie 1.1.2[45], keeping only uniquely mapped reads, allowing up to
517 two mismatches.
518 Peaks were called with MACS2 2.1.0[46]. For coverage tracks, reads were extended to 150
519 bp and coverage was normalized to total mapped reads (reads per million). Heatmaps and
520 metaplots were made with deepTools[47]. For metagene plots, the Ensembl version 75 transcripts
521 with the highest total coverage from the annotated TSS to 200 nt downstream and were at least 2
522 kb long, and 1 kb away from the nearest gene.
523 *RNAi.* Lentiviruses were packaged in HEK293T cells transfected with 1 µg pME-VSVG,
524 2 µg PAX2 and 4 µg shRNA in pLKO.1 backbone using Lipofectamine 3000 (Thermo Fisher
525 Scientific, Waltham, MA) according to the manufacturer's instructions. The virus particles were
526 then harvested 24 to 48 hours later by passing through a 0.45 µm syringe filter (Thermo
527 Scientific). Viruses were mixed with an equal volume of fresh media supplemented with 10%

528 FBS and polybrene was added at a final concentration of 5 µg/mL to increase infection
529 efficiency. The medium was changed 6 hours after infection. Cells were selected with 2 µg/ml
530 puromycin for three days before Western blotting and ChIP-seq experiments. The following
531 shRNA sequences were used:

shNRF1-F2	CCGGCCTCATGTATTTGAGTCTAATCTCGAGATTAGACTCAAATACATGAGGTTTTTG
shNRF1-R2	AATTCAAAAACCTCATGTATTTGAGTCTAATCTCGAGATTAGACTCAAATACATGAGG
shNRF1-F3	CCGGCCGTTGCCCAAGTGAATTATTCTCGAGAATAATTCACTTGGGCAACGGTTTTTG
shNRF1-R3	AATTCAAAAACCGTTGCCCAAGTGAATTATTCTCGAGAATAATTCACTTGGGCAACGG

532

533 **Bioinformatic protocols**

534 *Technical performance.* A series of modular bioinformatic analyses were implemented to
535 evaluate technical success of ChIP-exo library construction and sequencing, independent of
536 whether the antibody found its target or not. 1) Sequencing depth (standard is 8-10M), which is
537 the total number of sequencing reads having a target-specific barcode. 2) Adapter dimers
538 (standard is <2%), which is the fraction of reads that contain the sequencing adapters but lack a
539 genomic insert. 3) Alignment (standard is 70-90%), which is the percent of reads that map to the
540 reference human genome after removing adapter dimers. 4) PCR duplicates (standard is <40%),
541 which are expected to have identically mapped Read_1 and Read_2 5' ends. Since PCR
542 duplicates are not a direct product of ChIP, they add no value to enrichment metrics. High PCR
543 duplicates, at normal sequencing depths, often means technical loss of material during library
544 construction prior to PCR.

545 *Peak calling.* We utilized two distinct algorithms for ChIP-exo peak-calling. The first
546 algorithm was GeneTrack which used a gaussian kernel to call strand-separated peaks at the 5'
547 ends of reads[25]. The reads were then paired across strands and the tag occupancy was normalized
548 using the NCIS approach[48]. Peak significance was called using either the binomial or poisson test
549 (taking whichever p-value was higher) with Benjamini-Hochberg correction and a q-value cutoff
550 of q<0.01. The other peak-caller used was the ChExMix algorithm, which is a high-resolution
551 peak-caller designed to simultaneously identify enriched sequence motifs and distinct sub-types
552 of binding using a combination of clustering and hierarchical mixture modeling[26]. ChExMix was
553 designed to take advantage of ChIP-exo's ability to identify protein co-factors through indirect
554 crosslinking events by modeling detected tag distributions in ChIP-exo data and using tree-
555 clustering based approaches to determine the significant peak subtypes that exist with ChIP-exo
556 data.

557 *Motif enrichment via ChExMix.* De novo motif discovery was performed by ChExMix.
558 Each motif was compared against the JASPAR database using TOMTOM with default
559 parameters to identify similarity to known motifs[28, 49]. Heatmaps and composite plots were
560 generated of sequence reads aligned relative to motif midpoints of all peaks containing an
561 enriched motif. For samples with high background, low complexity, and/or low sequencing
562 depth, it is possible the antibody is valid, but that standard de novo motif discovery may fail. We
563 developed an orthogonal method for motif detection. By first identifying all non-redundant motif
564 classes[50] in the genome, we then overlap low-threshold ChExMix peaks and determined which
565 motif class possesses overlapping peaks above background (>2 $\log_2$).

566 *Motif enrichment and centering analysis.* For ChIP-exo data generated from K562,
567 HepG2 and MCF-7 cells, narrowpeak data were called using ChExMix[26]; we restricted our motif
568 enrichment analysis to narrowpeak datasets that contained more than 500 peaks. For ChIP-seq
569 data generated from HCT116 and K562 cells, we did not require the presence of more than 500
570 peaks due to the smaller number of datasets available. Motif enrichment analysis of ChIP-exo

571 and ChIP-seq peaks was then performed as described previously[31]. Briefly, we trimmed each
572 peak to 200 bp centered around the peak summit. We generated background sequences for the
573 trimmed peaks using GENRE software with the default human setting, to ensure the same level
574 of promoter overlap, repeat overlap, GC content and CpG dinucleotide frequency between each
575 peak and its associated background sequence[31]. We manually curated a collection of 99 position
576 weight matrices (PWMs), primarily from biochemical TF DNA binding assays (*i.e.*, PB or HT-
577 SELEX), from the UniPROBE and CisBP databases, as a representative repertoire of human
578 sequence-specific TF binding motifs[34, 42, 51]. We scored each sequence for matches to each of the
579 motifs using the function "matchPWM" from the "Biostrings" R package. Motif enrichment was
580 quantified using an established AUROC metric that assesses the presence of a motif among the
581 500 highest confidence peaks (foreground set) as compared to the corresponding background set
582 of sequences using publicly available tools for analysis of TF ChIP-seq data[31]. We also assessed
583 each motif for its enrichment towards the centers of each ChIP-exo or ChIP-Seq peak set as
584 described previously[31]. Briefly, we first identified the PWM score threshold that maximized the
585 difference between foreground and corresponding background sets in the number of sequences
586 containing at least one PWM match (*Optimal PWM Match Score*). If a sequence had multiple
587 PWM matches, we considered only the highest score site. We then calculated the distance from
588 each of these sites to the corresponding peak summits, and used the mean of these distances in
589 the foreground or background set to quantify the motif enrichment towards the centers of ChIP
590 peaks. The *P*-values associated with motif enrichment (*i.e.*, AUROC value) and enrichment
591 towards the peak summits (*i.e.*, mean motif distance to peak summit) were both calculated by
592 using a Wilcoxon signed-rank test comparing their scores (PWM match score and PWM match
593 distance to distance to peak summit, respectively) for foreground and background sequences
594 when the PWM threshold was set to the optimal PWM match score. We then adjusted the *P*-
595 values across the PWM collection with a false discovery rate test for multiple hypothesis testing.
596 To test the significance of the difference in the TPM distributions between ChIP datasets with
597 enriched versus non-enriched motifs, we calculated the *P*-value by a Wilcoxon test using the
598 function wilcox.test in R.
599 *Genome annotation enrichment.* Only a small fraction of DNA-interacting factors binds
600 sequence-specific motifs. In the case of targets with either no expected motif or no known
601 function, determining peak enrichments at annotated regions of the genome can provide evidence
602 of ChIP success. The relative frequency of peaks occurring in different functional genomic
603 regions as defined by chromHMM[52] was calculated for each target, an IgG negative control, and
604 for random expectation. The log2 frequency enrichment of sample over IgG control was used to
605 identify regions of enrichment, as well as significant areas of de-enrichment (regions that
606 selectively avoid the target).
607 Significant peaks were intersected with chromHMM and Segway states to generate
608 frequency histograms for overlap with predicted chromatin states[52, 53]. Peaks derived from the
609 matched negative control dataset were also intersected with annotated states. The log2 ratio of
610 sample state frequency over control state frequency was then calculated in order to identify
611 general state enrichment of the sample throughout the genome.
612 *Positional enrichment at promoters and insulators.* In order to identify enrichment in
613 well-characterized promoter regions, sequence reads for the target, a matched "No Antibody"
614 control, and an IgG were aligned relative to annotated transcription start sites (TSS). Heatmaps
615 of all genes and composite plots of the top 1,000 TSS by gene expression (RNA-seq FPKM)
616 were generated from the data[7].

617 *Heatmaps, composite plots, and 4color sequence plots.* All heatmaps, composite plots,
618 and 4 color sequence plots were generated using ScriptManager v0.12
619 (https://github.com/CEGRcode/scriptmanager). ScriptManager is a Java-based GUI tool that
620 contains a series of interactive wizards that guide the user through transforming aligned BAM
621 files into publication-ready figures.
622
623 **CUT&RUN protocols**
624 *Antibody sourcing and concentration.* Antibody hybridoma supernatants, name, clone ID,
625 and lot) were from DSHB. mAbs were concentrated using Amicon Ultra-4 Centrifugal Filter
626 Units with a 50 kDa cut-off (Millipore Sigma Cat # UFC805024) following manufacturer's
627 recommendations. All centrifugation steps (including 3x 4ml washes with 1X Tris buffered
628 Saline [TBS]) were performed at 4,000 x g for 15 minutes at room temperature. Final
629 concentrations for recovered mAbs (stored at 4ºC in [TBS, 0.1% BSA, 0.09% Sodium Azide])
630 were assumed based on initial concentrations / final recovery volumes and 1 µg used per
631 CUT&RUN experiment.
632 *CUT&RUN.* CUT&RUN was performed on 500k native nuclei extracted from K562 cells
633 using CUTANA® protocol v1.5.1 [http://www.epicypher.com] which is an optimized version of
634 that previously described[54]. For each sample, nuclei were extracted by incubating cells on ice for
635 10 min in Nuclei Extraction buffer (NE: 20 mM HEPES–KOH, pH 7.9; 10 mM KCl; 0.1%
636 Triton X-100; 20% Glycerol; 0.5mM spermidine; 1x complete protease inhibitor [*Roche* #
637 11836170001]), collecting by centrifugation (600 g, 3 min, 4ºC), discarding the supernatant, and
638 resuspending at [100 µl / 500K nuclei] sample in NE buffer. For each target 500K nuclei were
639 immobilized onto Concanavalin-A beads (*EpiCypher* #21-1401) and incubated overnight (4ºC
640 with gentle rocking) with 1 µg of antibody (40x PCRP as above ; IgG (*EpiCypher* 13-0042, lot
641 20036001-52) ; CTCF (*Millipore* 07-729, lot 3205452).
642 *Modified CUT&RUN library prep.* Illumina sequencing libraries were prepared from 1ng
643 to 10ng of purified CUT&RUN DNA using NEBNext Ultra II DNA Library Prep Kit (*New*
644 *England Biolabs* # E7645) as previously[55] with the following modifications to preserve and
645 enrich smaller DNA fragments (20-70 bp). Briefly, during end repair the cycling time was
646 decreased to 30 mins at 50ºC. After adapter ligation, to purify fragments >50bp, 1.75x volumes
647 of Agencourt AMPure XP beads (*Beckman Coulter* #A63881) were added for the first bead
648 clean-up before amplification following manufacturer's recommendations. PCR amplification
649 cycling parameters were as described[54]. Post-PCR, two rounds of DNA size selection were
650 performed. For the first selection, 0.8x volume of AMPure XP beads was added to the PCR
651 reaction to remove products >350bp. The supernatant, containing fragments <350bp, was moved
652 forward to a second round of size selection using 1.2x volumes of AMPure XP beads, to remove
653 products <150 bp. Libraries were quantified using Qubit Fluorometer (*Invitrogen*) and checked
654 for size distribution with a Bioanalyzer (*Agilent*).
655 *CUT&RUN library sequencing and data analysis.* Libraries were sequenced on the
656 *Illumina* NextSeq 550, obtaining ~5 million paired-end reads (75 x 75 nucleotides) on average.
657 Paired-end FASTQ files were aligned to the hg19 reference genome using the ChIP-exo pipeline.
658
659 **STORM protocols**
660 *Supernatant concentration.* 3 milliliters of PCRP supernatant were concentrated using the
661 Amicon® Pro Affinity Concentration Kit Protein G with 10kDa Amicon® Ultra-0.5 Device

662 following the manufacturer's recommendations. Supernatant was quantitated by
663 spectrophotometer for rough approximation of concentration.
664 *Cellular preparation and staining***.** K562 cells obtained from the ATCC (cat number
665 CCL243) were grown in a humidified 5% CO2 incubator. 3-5x10^5 cells were centrifuged at
666 1,500 rpm for 5 minutes, washed with PBS, and plated on MatTek-brand glass bottom dishes
667 (P35-G.1) prepared appropriately (washes with increasing concentrations of ethanol, followed by
668 coating with poly-L-Lysine (Sigma P4707) for 5 minutes and subsequent washes with water and
669 airdrying for 2h.) After plating, the cells were allowed to adhere for 2h and then washed with
670 PBS. For fixation, 1ml of 4% paraformaldehyde and 0.1% glutaraldehyde in PBS were added
671 for 10 minutes at room temperature with gentle rocking followed by blocking and
672 permeabilization in 2% normal goat serum/1% Triton X-100 in PBS for 1h at room temperature
673 with gentle shaking. Immunostaining was performed with anti-MTR4 antibody (Abcam 70551)
674 at a dilution of 1:250 in 0.1% normal goat serum, 0.05% Triton X-100 overnight at 4 degrees C.
675 The next morning cells were incubated with the secondary antibody (conjugated to Alexa-Fluor
676 647 (Life Technologies)) at a 1:1000 dilution in 0.1% normal goat serum in PBS and incubated
677 for 2h at room temperature followed by washes in PBS. This material was then subjected to
678 immunostaining with PCRP antibodies, by repeating the above procedure. PCRP mAb
679 hybridoma supernatant (3 ml) was first concentrated 30-100 fold since raw supernatants were
680 unsuccessful in both confocal microscopy and STORM (data not shown). The secondary
681 antibody was conjugated to Alexa-Fluor 488 (Life Technologies). At the end of the application
682 of the second secondary antibody, the cells were washed 3x with PBS and DAPI (Sigma D8542)
683 at a concentration of 1:500 in 0.1% normal goat serum in 1x PBS was applied for 10 minutes at
684 room temperature. The cells were washed 3x with PBS and finally 4% paraformaldehyde/0.1%
685 glutaraldehyde in PBS was applied for 10 minutes at RT before 3 washes in 1x PBS were
686 performed and dishes stored at 4 degrees until microscopy could be performed.
687 *Microscopy.* Cells were brought to the imaging facility and OXEA buffer was applied (50
688 mM Cysteimine, 3% v/v Oxyfluor, 20% v/v sodium DL Lactate, with pH adjusted to
689 approximately 8.5, as necessary.) The two colors (Alexa fluor 488 and Alexa fluor 647) were
690 imaged sequentially. Imaging buffer helped to keep dye molecules in a transient dark state.
691 Subsequently, individual dye molecules were excited stochastically with high laser power at their
692 excitation wavelength (488 nm for Alexa fluor 488 or 647 nm for Alexa fluor 647, respectively)
693 to induce blinking on millisecond timescales. STORM images and the correlated high-power
694 confocal stacks were acquired via a CFI Apo TIRF 100 × objective (1.49 NA) on a Nikon Ti-E
695 inverted microscope equipped with a Nikon N-STORM system, an Agilent laser launch system,
696 an Andor iXon Ultra 897 EMCCD (with a cylindrical lens for astigmatic 3D-STORM imaging)
697 camera, and an NSTORM Quad cube (Chroma). This setup was controlled by Nikon NIS-
698 Element AR software with N-STORM module. To obtain images, the field of view was selected
699 based on the live EMCCD image under 488-nm illumination. 3D STORM datasets of 50,000
700 frames were collected. Lateral drift between frames was corrected by tracking 488, 561, and 647
701 fluorescent beads (TetraSpeck, Life Technologies). STORM images were processed to acquire
702 coordinates of localization points using the N-STORM module in NIS-Elements AR software.
703 Identical settings were used for every image. Each localization is depicted in the STORM image
704 as a Gaussian peak, the width of which is determined by the number of photons detected[36]. All of
705 the 3D STORM imaging was performed on a minimum of two different K562 cells.
706
707

TF cloning, protein expression, western blots, and PBM protocols

Full-length TFs were either obtained from the hORFeome clone collection or synthesized as gBlocks (Integrated DNA Technologies) (**Supplementary Table 7**), full-length sequence-verified, and transferred by Gateway recombinational cloning into either the pDEST15 (ThermoFisher Scientific) or pT7CFE1-NHIS-GST (ThermoFisher) vectors for expression as N-terminal GST fusion proteins[56]. TFs were expressed by a coupled *in vitro* transcription and translation kit according to the manufacturer's protocols (**Supplementary Table 6**). Protein concentrations were approximated by an anti-GST western blot as described previously[39]. PCRP antibodies were used at a final concentration of 40 ng/mL in western blots. 8X60K GSE 'all 10-mer universal' oligonucleotide arrays (AMADID #030236; Agilent Technologies, Inc.) were double-stranded and used in PBM experiments essentially as described previously, with minor modifications as described below[39, 57, 58]. GST-tagged TFs assayed in PBMs were detected either with Alexa488-conjugated anti-GST antibody (Invitrogen A-11131), or with a TF-specific PCRP antibody, followed by washes and detection with Alexa488-conjugated goat anti-mouse IgG(H+L) Cross-Adsorbed Secondary Antibody (Invitrogen A-11001), essentially as described previously[41] (**Supplementary Table 6**). All PCRP Abs were used undiluted in PBM experiments; a subset of the PCRP Abs were also tested at a 1:5 or 1:20 dilution (**Supplementary Table 6**). All PBM experiments using PCRP antibodies were performed using fresh arrays or arrays that had been stripped once, as described previously[39, 57]. PBMs were scanned in a GenePix 4400A Microarray Scanner and raw data files were quantified and processed using the Universal PBM Analysis Suite[39, 57].

Code Availability

Our automated bioinformatic ChIP-exo analysis pipeline can be found at (https://github.com/CEGRcode/PCRPpipeline).

Data Availability

Raw ChIP-exo sequencing files are available at NCBI GEO archive (GSE151287). PBM data will be available via the UniPROBE database (accession ID: LAI20A) upon publication. CUT&RUN data has been deposited at GEO: Accession GSE151326 [(5.27.20) NCBI tracking system #20907832].

Acknowledgements

The Basu lab would like to thank Drs. Teresa Swayne and Emilia Laura Munteanu of the Herbert Irving Comprehensive Cancer Center Core Microscopy facility at Columbia University. The Pugh lab would like to thank Kyle Nilson for donating the peroxide stressed cells. The Shilatifard lab would like to thank Anna Whelan and Jordan Harris for technical assistance and Zibo Zhao and Marc Morgan for helpful discussions. The Bulyk lab thanks Steve S. Gisselbrecht for assistance with preparation of figures. This work was supported by an administrative supplement to NIAID grant 1R01AI099195 to U.B., an administrative supplement to NIH grant R21 HG009268 to M.L.B., an administrative supplement to NIH grant R01-ES013768 to B.F.P., an administrative supplement to NIH grant R01CA214035 to A.S., NIH grant R50CA211428 to E.R.S., and NIH grant R44 DE029633 to EpiCypher.

753 **Author Contributions**
754 In the Basu lab, G.R. performed STORM experiments; W.Z. performed required bioinformatics
755 analyses of STORM data; U.B. analyzed data, provided oversight and co-wrote the manuscript.
756 In the Bulyk lab, J.T.A. and S.K.P. performed cloning, protein expression, western blots, PBM
757 experiments, and PBM analysis; L.M. performed analysis of motif enrichment and centering;
758 S.K.P. and L.M. prepared figures and Supplementary Tables; M.L.B. supervised research; and
759 S.K.P., L.M., and M.L.B. co-wrote the manuscript. In the Pugh lab, TRB, KB, JM, SND, and
760 KM performed ChIP-exo assays. PK designed and implemented the web portal. MJR and DJ
761 performed ChIP-exo library quantitation and sequencing. WKML directed the ChIP-exo
762 experiments, processed and analyzed the ChIP-exo, and co-wrote the manuscript writing. BFP
763 provided oversight for ChIP-exo and co-wrote the manuscript. In the Shilatifard lab, A.P.S.
764 conducted experiments. E.R.S. analyzed ChIP-seq data. E.R.S. and A.S. provided oversight and
765 co-wrote the manuscript. From EpiCypher, B.J.V., K.N. and E.M. performed CUT&RUN
766 studies, and M-C.K. provided oversight and co-wrote the manuscript.
767
768 **Competing Interests**
769 M.L.B. is a co-inventor on U.S. patent #8,530,638 on universal PBM technology. BFP has a
770 financial interest in Peconic, LLC, which utilizes the ChIP-exo technology implemented in this
771 study and could potentially benefit from the outcomes of this research. EpiCypher is a
772 commercial developer of reagents to support CUTANA® CUT&RUN. The authors in the Basu
773 and Shilatifard labs declare no competing financial interests.

774 **REFERENCES**

775 1. Park, P.J. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*
776 **10**, 669-680 (2009).
777 2. Chames, P., Van Regenmortel, M., Weiss, E. & Baty, D. Therapeutic antibodies:
778 successes, limitations and hopes for the future. *Br J Pharmacol* **157**, 220-233 (2009).
779 3. Mahmood, T. & Yang, P.C. Western blot: technique, theory, and trouble shooting. *N Am*
780 *J Med Sci* **4**, 429-434 (2012).
781 4. Lin, J.R., Fallahi-Sichani, M. & Sorger, P.K. Highly multiplexed imaging of single cells
782 using a high-throughput cyclic immunofluorescence method. *Nat Commun* **6**, 8390
783 (2015).
784 5. Engelen, E. et al. Proteins that bind regulatory regions identified by histone modification
785 chromatin immunoprecipitations and mass spectrometry. *Nat Commun* **6**, 7155 (2015).
786 6. Siggers, T. et al. Principles of dimer-specific gene regulation revealed by a
787 comprehensive characterization of NF-kappaB family DNA binding. *Nat Immunol* **13**,
788 95-102 (2011).
789 7. Consortium, T.E.P. An integrated encyclopedia of DNA elements in the human genome.
790 *Nature* **489**, 57-74 (2012).
791 8. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human
792 epigenomes. *Nature* **518**, 317-330 (2015).
793 9. Egelhofer, T.A. et al. An assessment of histone-modification antibody quality. *Nat Struct*
794 *Mol Biol* **18**, 91-93 (2011).
795 10. Baker, M. Reproducibility crisis: Blame it on the antibodies. *Nature* **521**, 274-276 (2015).
796 11. Shah, R.N. et al. Examining the Roles of H3K4 Methylation States with Systematically
797 Characterized Antibodies. *Mol Cell* **72**, 162-177 e167 (2018).
798 12. Hanly, W.C., Artwohl, J.E. & Bennett, B.T. Review of Polyclonal Antibody Production
799 Procedures in Mammals and Poultry. *ILAR J* **37**, 93-118 (1995).
800 13. Reardon, S. in Nature (2016).
801 14. Blackshaw, S. et al. The NIH Protein Capture Reagents Program (PCRP): a standardized
802 protein affinity reagent toolbox. *Nat Methods* **13**, 805-806 (2016).
803 15. Venkataraman, A. et al. A toolbox of immunoprecipitation-grade monoclonal antibodies
804 to human transcription factors. *Nat Methods* (2018).
805 16. Kohler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of
806 predefined specificity. *Nature* **256**, 495-497 (1975).
807 17. Winter, G., Griffiths, A.D., Hawkins, R.E. & Hoogenboom, H.R. Making antibodies by
808 phage display technology. *Annu Rev Immunol* **12**, 433-455 (1994).
809 18. Liu, J.K. The history of monoclonal antibody development - Progress, remaining
810 challenges and future innovations. *Ann Med Surg (Lond)* **3**, 113-116 (2014).
811 19. Rhee, H.S. & Pugh, B.F. ChIP-exo method for identifying genomic location of DNA-
812 binding proteins with near-single-nucleotide accuracy. *Curr Protoc Mol Biol* **Chapter**
813 **21**, Unit 21 24 (2012).
814 20. Rossi, M.J., Lai, W.K.M. & Pugh, B.F. Simplified ChIP-exo assays. *Nat Commun* **9**,
815 2842 (2018).
816 21. Rhee, H.S. & Pugh, B.F. Comprehensive genome-wide protein-DNA interactions
817 detected at single-nucleotide resolution. *Cell* **147**, 1408-1419 (2011).

818 22. Rada-Iglesias, A. et al. Whole-genome maps of USF1 and USF2 binding and histone H3
819 acetylation reveal new aspects of promoter structure and candidate genes for common
820 human disorders. *Genome Res* **18**, 380-392 (2008).
821 23. Piunti, A. et al. CATACOMB: An endogenous inducible gene that antagonizes H3K27
822 methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism.
823 *Sci Adv* **5**, eaax2887 (2019).
824 24. Su, J. et al. TGF-beta orchestrates fibrogenic and developmental EMTs via the RAS
825 effector RREB1. *Nature* **577**, 566-571 (2020).
826 25. Albert, I., Wachi, S., Jiang, C. & Pugh, B.F. GeneTrack--a genomic data processing and
827 visualization framework. *Bioinformatics* **24**, 1305-1306 (2008).
828 26. Yamada, N., Lai, W.K.M., Farrell, N., Pugh, B.F. & Mahony, S. Characterizing protein-
829 DNA binding event subtypes in ChIP-exo data. *Bioinformatics* **35**, 903-913 (2019).
830 27. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. Quantifying similarity
831 between motifs. *Genome Biol* **8**, R24 (2007).
832 28. Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor
833 binding profiles. *Nucleic Acids Res* **48**, D87-D92 (2020).
834 29. Hoffman, M.M. et al. Integrative annotation of chromatin elements from ENCODE data.
835 *Nucleic Acids Res* **41**, 827-841 (2013).
836 30. Gordan, R., Hartemink, A.J. & Bulyk, M.L. Distinguishing direct versus indirect
837 transcription factor-DNA interactions. *Genome Res* **19**, 2090-2100 (2009).
838 31. Mariani, L., Weinand, K., Vedenko, A., Barrera, L.A. & Bulyk, M.L. Identification of
839 Human Lineage-Specific Transcriptional Coregulators Enabled by a Glossary of Binding
840 Modules and Tunable Genomic Backgrounds. *Cell Syst* **5**, 187-201 e187 (2017).
841 32. Bailey, T.L. & Machanick, P. Inferring direct DNA binding from ChIP-seq. *Nucleic*
842 *Acids Res* **40**, e128 (2012).
843 33. Wang, J. et al. Sequence features and chromatin structure around the genomic regions
844 bound by 119 human transcription factors. *Genome Res* **22**, 1798-1812 (2012).
845 34. Hume, M.A., Barrera, L.A., Gisselbrecht, S.S. & Bulyk, M.L. UniPROBE, update 2015:
846 new tools and content for the online database of protein-binding microarray data on
847 protein-DNA interactions. *Nucleic Acids Res* **43**, D117-122 (2015).
848 35. Baler, R., Dahl, G. & Voellmy, R. Activation of human heat shock genes is accompanied
849 by oligomerization, modification, and rapid translocation of heat shock transcription
850 factor HSF1. *Mol Cell Biol* **13**, 2486-2496 (1993).
851 36. Betzig, E. et al. Imaging intracellular fluorescent proteins at nanometer resolution.
852 *Science* **313**, 1642-1645 (2006).
853 37. Lim, J. et al. Nuclear Proximity of Mtr4 to RNA Exosome Restricts DNA Mutational
854 Asymmetry. *Cell* **169**, 523-537 e515 (2017).
855 38. Sabari, B.R. et al. Coactivator condensation at super-enhancers links phase separation
856 and gene control. *Science* **361** (2018).
857 39. Berger, M.F. et al. Compact, universal DNA microarrays to comprehensively determine
858 transcription-factor binding site specificities. *Nat Biotechnol* **24**, 1429-1435 (2006).
859 40. Mukherjee, S. et al. Rapid analysis of the DNA-binding specificities of transcription
860 factors with DNA microarrays. *Nat Genet* **36**, 1331-1339 (2004).
861 41. Siggers, T., Duyzend, M.H., Reddy, J., Khan, S. & Bulyk, M.L. Non-DNA-binding
862 cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Mol*
863 *Syst Biol* **7**, 555 (2011).

42. Weirauch, M.T. et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431-1443 (2014).
43. Uhlen, M. et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* **4**, 1920-1932 (2005).
44. Lee, T.I., Johnstone, S.E. & Young, R.A. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* **1**, 729-748 (2006).
45. 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).
46. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
47. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).
48. Liang, K. & Keles, S. Normalization of ChIP-seq data with control. *BMC Bioinformatics* **13**, 199 (2012).
49. Bailey, T.L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-208 (2009).
50. Castro-Mondragon, J.A., Jaeger, S., Thieffry, D., Thomas-Chollier, M. & van Helden, J. RSAT matrix-clustering: dynamic exploration and redundancy reduction of transcription factor binding motif collections. *Nucleic Acids Res* **45**, e119 (2017).
51. Mariani, L., Weinand, K., Gisselbrecht, S.S. & Bulyk, M.L. MEDEA: Analysis of Transcription Factor Binding Motifs in Accessible Chromatin. *Genome Res* **In Press** (2020).
52. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49 (2011).
53. Hoffman, M.M. et al. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* **9**, 473-476 (2012).
54. Skene, P.J., Henikoff, J.G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006-1019 (2018).
55. Liu, N. et al. Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. *Cell* **173**, 430-442 e417 (2018).
56. Collaboration, O.R. The ORFeome Collaboration: a genome-scale human ORF-clone resource. *Nat Methods* **13**, 191-192 (2016).
57. Berger, M.F. & Bulyk, M.L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat Protoc* **4**, 393-411 (2009).
58. Nakagawa, S., Gisselbrecht, S.S., Rogers, J.M., Hartl, D.L. & Bulyk, M.L. DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proc Natl Acad Sci U S A* **110**, 12349-12354 (2013).