HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2019 May 01.

Published in final edited form as:

Nat Methods. 2018 May ; 15(5): 330–338. doi:10.1038/nmeth.4632.

A toolbox of immunoprecipitation-grade monoclonal antibodies to human transcription factors

A full list of authors and affiliations appears at the end of the article.

Abstract

A key component of efforts to address the reproducibility crisis in biomedical research is the development of rigorously validated and renewable protein-affinity reagents. As part of the US National Institutes of Health (NIH) Protein Capture Reagents Program (PCRP), we have generated a collection of 1,406 highly validated immunoprecipitation- and/or immunoblotting-grade mouse monoclonal antibodies (mAbs) to 737 human transcription factors, using an integrated production and validation pipeline. We used HuProt human protein microarrays as a primary validation tool to identify mAbs with high specificity for their cognate targets. We further validated PCRP mAbs by means of multiple experimental applications, including immunoprecipitation, immunoblotting, chromatin immunoprecipitation followed by sequencing (ChIP-seq), and immunohistochemistry. We also conducted a meta-analysis that identified critical variables that contribute to the generation of high-quality mAbs. All validation data, protocols, and links to PCRP mAb suppliers are available at http://proteincapture.org.

There are serious problems with the quality, consistency, and availability of research-grade antibodies. Globally, it is estimated that over $800 million is wasted annually because of the

Correspondence should be addressed to I.P. (ignacio.pino@cdi-lab.com), D.E. (dan.eichinger@cdi-lab.com), H.Z. (hzhu4@jhmi.edu), or S.B. (sblack@jhmi.edu).

[19]Present addresses: Department of Immunology, University of Texas Southwestern Medical Center, Dallas, Texas, USA (Z.K.); BioReliance, Sigma-Aldrich Corp., Rockville, Maryland, USA (D.B.K.).

Anand Venkataraman http://orcid.org/0000-0001-6303-4820
Mark Mackiewicz http://orcid.org/0000-0002-7088-1902
Paolo Mita http://orcid.org/0000-0002-2093-4906
Zheng Kuang http://orcid.org/0000-0001-5855-8371
Ignacio Pino http://orcid.org/0000-0002-6513-9498
Daniel J Eichinger http://orcid.org/0000-0002-7764-7668
Heng Zhu http://orcid.org/0000-0002-8426-2889
Seth Blackshaw http://orcid.org/0000-0002-1338-8476

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

A.V., M.M., P.M., Z.K., L.X., Y.L., D.G., S.L., P.R., S.H., D.B.K., H. Zhang, F.P.-B., G.S., E.A., L.A., L.R., L.L., G.M., J.R., K.R., R.A., L.N., K.M., I.V., Z.A.R.-P., C.R., M.V., J.M., B.S.C., S.Y., S.G.K., J.d.M., M.S., L.J., B.J., A.T. and E.C. performed experimental work. R.S. and S.C. performed independent validation of PCRP mAbs. K.Y., J.I. and S.K. designed algorithms and implemented software. W.Y.Y., S.A., G.T.M., R.M.M., J.D.B., D.F., G.W., D.J.E., J.S.B., I.P., H. Zhu and S.B. contributed expertise and supervision. All authors contributed to manuscript preparation.

COMPETING INTERESTS

S.B., H. Zhu, I.P., D.J.E., and J.D.B. are cofounders and shareholders of CDI Labs Inc. J.I., P.R., D.B.K., E.A., L.A., L.R., L.L., G.M., J.R., K.R., R.A., L.N., K.M., I.V., Z.A.R.-P., C.R., M.V., and W.Y.Y. are employees of CDI Labs Inc. A.V. and J.D.B. are consultants to CDI Labs Inc. J.D.B. serves on the Board of Directors of CDI Labs, and J.D.B.'s relationship with CDI Labs is managed by NYU Langone Health's committee on conflicts of interest. B.J. is an employee of NeoBiotechnologies, Inc. A.T. is the founder and sole owner of NeoBiotechnologies, Inc. G.T.M. is founder and shareholder of Nexomics Biosciences, Inc.

use of poor-quality antibodies[1]. Factors that aggravate this problem include the absence of standardized antibody-validation criteria in the research community, a lack of transparency from commercial antibody suppliers about their products, the use of polyclonal reagents that show extensive batch variation, and technical difficulties in comprehensive assessments of antibody cross-reactivity[2–6].

Several efforts have generated collections of high-quality antibodies to human proteins. Such efforts have produced useful tools, but they also have limitations. Projects that have generated renewable reagents—mAbs or recombinant antibodies—targeted only a limited subset of proteins[7–10]. More comprehensive efforts generated nonrenewable polyclonal antibodies[11]. Linear antigens are typically used for immunization, and yield reagents that are useful in immunoblotting and immunostaining but often cannot recognize native antigens in applications such as immunoprecipitation[11]. Finally, proof that these antibodies exclusively recognize their intended targets is usually lacking.

To extend these efforts, the NIH Common Fund initiated the PCRP[12]. Here we describe an integrated production and validation pipeline for generating mouse mAbs of high specificity and affinity that is also suitable for analyzing the specificity of existing mAbs. We used this pipeline to generate 1,406 mAbs, targeting 737 unique human transcription factors (TFs) and TF-associated proteins, that work for one or more research applications, including immunoprecipitation, immunoblotting, ChIP-seq, and immunohistochemistry (IHC). However, we note that researchers using these reagents should carry out their own validation experiments to assess their ability to detect endogenous protein. These PCRP mAbs are available to the research community through both the Developmental Studies Hybridoma Bank (DSHB) and commercial suppliers.

RESULTS

Design of the production pipeline

The workflow for production, validation, and distribution of mAbs, which is an adaptation of a previously published protocol[13], is shown in Figure 1 and described further below and in the Online Methods. Protocols and standard operating procedures are available at https://proteincapture.org/protocols. We produced and purified recombinant domains and full-length proteins from *Escherichia coli* and *Saccharomyces cerevisiae*, respectively, and used them for mouse immunization. After identifying IgG-positive hybridomas by ELISA, we screened supernatants against a minichip comprising the antigens used for immunization and up to 80 other proteins. We then analyzed all antibodies that recognized their cognate antigen as the top target on the minichip with the HuProt human protein microarray (hereinafter referred to as HuProt), which contains > 19,500 recombinant human proteins affinity-purified from yeast[14]. We then tested passing mAbs identified as monospecific in the HuProt analysis by carrying out immunoblotting and immunoprecipitation experiments using their full-length targets expressed at a range of concentrations in human cells, with a subset also tested by ChIP-seq in ENCODE cell lines and/or by immunohistochemical analysis. Validation data for all mAbs are available through the PCRP web portal (http://proteincapture.org). mAbs that work in one or more research-grade applications are available through multiple distributors, including the DSHB, via the PCRP web portal.

HuProt arrays contain a majority of the annotated, full-length proteome in native conformation

The HuProt array contains >75% of proteins in each major functional category of the proteome, based on Gene Ontology annotation (Supplementary Fig. 1a). The use of HuProt as a screening platform to identify protein interaction partners, enzyme substrates, and autoantibodies has already been reported[15–19]. In the current study we tested proteins on HuProt to determine whether they were in native conformation, as this is necessary for screening of immunoprecipitation-grade mAbs. First, we probed both native and denatured HuProt arrays with mAbs that selectively recognize either linear or folded epitopes of their cognate antigen. mAbs that recognize folded ACO2 (Supplementary Fig. 1b) or glutathione *S*-transferase (GST) (Supplementary Fig. 1c,d) recognized their targets only on native HuProt, whereas a mAb that specifically recognizes a linear epitope on SMAD4 recognized its target only on denatured HuProt (Supplementary Fig. 1b). Second, as a minimal test of whether proteins on HuProt retain function, we screened for proteins that bound the long noncoding RNA *Xist*. We found that 247 proteins interacted with *Xist* RNA with *z*-scores > 2 (Supplementary Table 1), 5 of which were previously identified as *Xist* interactors[20, 21]. We also identified many RNA-binding proteins as potentially novel *Xist* interactors (Supplementary Fig. 1e, Supplementary Table 1). Specific binding to *Xist* was lost after protein denaturation (Supplementary Fig. 1e,f, Supplementary Table 1).

Target selection, antigen production, and immunization

We first sought to identify a core set of human proteins involved in transcriptional regulation by integrating previously annotated lists of high-confidence TFs[22], other proteins known to regulate transcription identified from literature sources, and proteins known to bind TFs (Supplementary Table 2). This resulted in a final target list of 2,229 proteins, 73% of which were annotated as high-confidence TFs (Fig. 2a).

We then used this list to guide the selection of proteins for expression, purification, and immunization. We tested 1,290 constructs expressing recombinant His- or Halo-tagged domains of individual target proteins in *E. coli*, and observed that 949 of them yielded >0.1 mg of purified protein[7] (Supplementary Table 3). Additionally, we purified 1,477 constructs encoding GST-tagged full-length target proteins from yeast, of which 768 yielded >30 μg of protein (Supplementary Table 4). We used a total of 1,453 unique proteins (~65%) from our target list for mouse immunization. We observed substantial variation in success rates for protein expression and immunization among different target classes (Fig. 2a). Immunizations were administered by either systemic intraperitoneal or local footpad routes.

Analysis of mAb specificity by HuProt binding

For our initial specificity screen for IgG^+ mAbs, we used a mini-chip comprising the intended target and 20–80 other antigens. We observed that 81.1% of domains and 80.1% of full-length proteins generated at least one mAb that passed minichip screening (minichip-positive), and 8.9% of all mAbs tested preferentially recognized their cognate target (Fig. 2b). We then analyzed the specificity of minichip-positive mAbs by examining HuProt binding. On the basis of previously established criteria[13], a mAb passed HuProt analysis ($HuProt^+$) if the *z*-score of the cognate target ranked highest among those of all ~19,500

antigens on the array, with an *S*-score (or the difference between the first-and second-ranked *z*-scores) > 2.5. We observed that 42.6% of minichip-positive mAbs passed HuProt analysis, corresponding to 73.3% and 70.7% of targets generated with domains and full-length antigens, respectively (Fig. 2b). A mAb could have a high *z*-score for the intended target and still fail to meet our validation criteria if the *z*-score of any off-target ranked higher. However, this seems unlikely, as HuProt$^+$ mAbs showed higher affinity for their targets, as measured by *z*-score, than did HuProt-negative (HuProt$^-$) mAbs, even when we considered only cases in which the intended target appeared in the top 50 proteins recognized (Fig. 2c). We observed strong correlation between *z*- and *S*-scores for passing mAbs (Fig. 2d), and both scores showed broad distribution (Fig. 2e). Some mAbs showed exceptionally high *S*-scores (Fig. 2f), with virtually undetectable binding to any proteins other than the cognate target.

To determine whether antibody specificity as measured by *S*-score can identify cross-reactivity, we profiled the specificity of six commercially available immunoprecipitation-grade mAbs that recognized the same target antigens as passing HuProt$^+$ mAbs (Supplementary Table 5). Using our standard quality criteria (*S*-score > 2.5), we found that five of six commercial antibodies showed cross-reactivity on the HuProt array with one or more proteins (Supplementary Table 6). We also carried out a competitive immunoprecipitation analysis (Supplementary Fig. 2, Supplementary Table 6). In the example shown in Figure 2g,h, TPM2 was identified as the top target for a commercially available anti-STAT4 mAb. Competitive immunoprecipitation analysis showed that the HuProt$^+$ PCRP mAb selectively recovered STAT4, whereas the HuProt$^-$ commercial antibody recovered TPM2 but not STAT4.

Highly specific antibodies typically bind their targets with high affinity[23, 24]. To test whether our HuProt$^+$ mAbs were also high affinity, we measured binding kinetics for 122 HuProt$^+$ mAbs and 106 minichip-positive, HuProt$^-$ mAbs using label-free Octet analysis[25]. Using K_d < 50 nM as a cutoff, we observed that 78 of 122 HuProt$^+$ mAbs and only 44 of 106 HuProt$^-$ mAbs bound their cognate targets with high affinity (χ^2 test, $P < 0.001$), indicating that HuProt$^+$ mAbs showed higher overall affinity than HuProt-mAbs. This difference in affinity affected only the k_{on} scores of HuProt$^+$ mAbs compared with those of HuProt$^-$ mAbs (Fig. 2i). However, the *z*- and *S*-scores of mAbs tested by HuProt did not correlate with any of the measured affinity parameters, which indicated that mAb specificity as measured by HuProt does not serve as a proxy for affinity.

Secondary validation of mAbs

All HuProt$^+$ mAbs, along with a limited number of HuProt-high-affinity (K_d < 50 nM) mAbs, were considered candidates for secondary validation. This involved testing whether the mAbs could recognize their targets in immunoblotting (IB) and immunoprecipitation (IP) experiments. Because most TFs are expressed at low or undetectable levels in readily cultured cell lines, it was impractical to conduct large-scale IP and IB validation of these mAbs against endogenous proteins, so we instead used an inducible overexpression system (Fig. 3a). We cloned full-length target genes into a doxycycline-inducible expression vector, transiently transfected that vector into Tet-ON HeLa[26] or HEK293$_{LD}$ suspension cells, and

induced expression with both low and high doses of doxycycline for 24–48 h. Using anti-target mAb, anti-Flag M2 (positive control), or mouse pan-IgG sera (negative control), we carried out IB and IP. We used LiCOR two-color infrared fluorescence imaging for simultaneous evaluation of the IP and IB performance of each mAb.

To evaluate IP efficiency, we used input samples of lysates obtained by induction with or without doxycycline (Fig. 3b). A given mAb was scored as IP^+ when anti-Flag IB showed a positive signal at the predicted molecular weight after IP of doxycycline-induced samples, but not after IP with mouse IgG. Similarly, a given mAb was scored as IB^+ when anti-target mAb showed signal at the predicted molecular weight in any lane except the uninduced or pan-IgG IP lanes. The performance of each mAb in IP and IB was evaluated by two independent observers using standardized criteria. In this manner, we identified 1,406 of 2,017 mAbs (69.7%), corresponding to 737 out of 848 targets that passed our quality thresholds in IP and/or IB experiments (Fig. 3c). The vast majority (96.1%) of passing mAbs were IP-grade (808 IP^+, IB^- and 544 IP^+, IB^+ mAbs), whereas only 54 mAbs were IP^-IB^+.

We analyzed a subset of IP^+ and/or IB^+ mAbs by ChIP-seq against endogenous protein in human cell lines. In our ChIP-seq experiments, 46/305 mAbs (15%) against 36/176 (20.5%) targets passed as per ENCODE standards (Supplementary Table 7). For example, 61.2% of ChIP-seq peaks identified with an anti-NRF1 mAb (clone R157.1.3H1) included the known NRF1 consensus site (Fig. 3d, Supplementary Table 7).

We also evaluated passing IP^+ and/or IB^+ mAbs in IHC assays. In a human tissue microarray, 12/129 mAbs tested (9.3%) corresponding to 9/73 targets showed positive IHC signals in at least one sample (Fig. 3e), and the observed signals corresponded well with those generated by mAbs widely used by pathologists (Supplementary Fig. 3). We also carried out IHC experiments on a subset of passing mAbs using mouse brain and retinal sections, and found that 58/192 mAbs for 39/93 targets scored as positive (Figs. 3e). We further validated mAb specificity by IHC using retinal tissues from *Rax-CreER^{T2};Lhx2*$^{lox/lox}$ mice[27]. IHC signal detected with anti-LHX2 mAb (clone R911.1.2G5) was lost in mice with conditional knockout of *Lhx2* (Fig. 3e).

The PCRP consortium commissioned an independent evaluation of a subset of the mAbs that passed secondary validation by IP coupled with mass spectrometry (IP-MS), surface plasmon resonance (SPR), and ELISA (Supplementary Table 8). We tested 43 mAbs to 39 targets using standard US National Cancer Institute criteria, and found that 39, 34, and 22 mAbs passed in ELISA, SPR, and IP-MS experiments, respectively; examples of passing mAbs are shown in Figure 3f–h.

Meta-analysis identifies factors for successful mAb generation

We next conducted a meta-analysis of collected data (Online Methods). We observed that the *z*- and *S*-scores of mAbs that passed secondary validation were higher than those of mAbs that failed secondary validation (Fig. 4a). However, none of the individual parameters significantly correlated with success rates in secondary validation. We therefore investigated whether multiple parameters in combination influenced success at different stages of the pipeline. We carried out three types of parametric comparisons, analyzing (i) individual

mAbs (Supplementary Table 8), (ii) targets (Supplementary Table 9), and (iii) target subclasses. Matrix comparison of parameters (Supplementary Table 10) generated a comprehensive list of factors that could predict success in the pipeline (Supplementary Tables 11–16).

We had hypothesized that immunization with multiple antigens might increase the yield of passing mAbs. However, parametric comparison clearly showed that mAbs generated by immunization with a single antigen had higher *z*- and *S*-scores (Fig. 4b; Supplementary Table 11, comparisons 2 and 17; $P = 6.5 \times 10^{-9}$ and 6.8×10^{-9}, respectively) and significantly improved success at the HuProt stage and in the overall pipeline (Fig. 4c; Supplementary Table 12, comparisons 12 and 34; $P = 2.5 \times 10^{-6}$ and $P = 7.18 \times 10^{-4}$, respectively) compared with those generated by immunization with multiple antigens.

We grouped mAbs on the basis of the antigen and immunization protocol used, and used abbreviations to identify each group as detailed below.

- **i.** Type of immunogen used: domain (D), full-length (F), or all (A)
- **ii.** Type of epitope recognized by the mAb: its immunized domain but not full-length counterpart on the HuProt (D), both its domain immunogen and its full-length counterpart on HuProt (F), or considering all types (A)
- **iii.** Route of immunization: intraperitoneal (i.p.), footpad (f.p.), or all immunizations (A)

In subsequent text, group designations include one abbreviation from each of these three categories, separated by hyphens, and presented in the order shown above.

We compared mAbs in the D-A-i.p. and D-A-f.p. groups generated with both single and multiple immunogens. I.p. immunization generated mAbs with higher *z*- and *S*-scores compared with those of mAbs generated by f.p. immunization (Fig. 4d; Supplementary Table 13, comparisons 2 and 10; $P = 3.1 \times 10^{-3}$ and $P = 6.0 \times 10^{-5}$, respectively). Although no differences in K_d were observed, mAbs generated via the f.p. route showed higher k_{on} scores than those of mAbs generated via i.p. immunization (Fig. 4e; Supplementary Table 13, comparisons 34 and 18; $P = 9.8 \times 10^{-1}$ and $P = 6.7 \times 10^{-3}$, respectively). Furthermore, i.p. immunization yielded a higher fraction of passing mAbs, at minichip and secondary-validation stages, than did f.p. immunization (Fig. 4f; Supplementary Table 14, comparisons 2 and 18; $P = 1.2 \times 10^{-6}$ and $P = 7.1 \times 10^{-3}$, respectively). This comparatively reduced efficiency of f.p. immunization may in practical terms be offset by lower costs of antigen production, as i.p. immunization uses on average two orders of magnitude more antigen than f.p. immunization.

Next, we compared parameters and success rates for the D-D-A and D-F-A groups. For this, we compared only mAbs generated with a single immunogen. Strikingly, D-D-A mAbs had lower *z*- and *S*-scores on average (Fig. 4g; Supplementary Table 15, comparisons 5 and 27; $P = 1.83 \times 10^{-13}$ and $P = 1.08 \times 10^{-8}$, respectively). These differences were also visible at the level of targets. D-F-A mAbs had higher success rates in secondary validation and

therefore in the overall pipeline (Fig. 4h; Supplementary Table 16, comparisons 43 and 62; $P = 7.11 \times 10^{-3}$ and 5.01×10^{-3}, respectively).

Although full-length protein production in yeast gave substantially lower yields than protein production in *E. coli*, immunization with full-length antigens by the f.p. route yielded mAbs with superior *z*- and *S*-scores (Supplementary Fig. 4a; Supplementary Table 15, comparisons 1 and 23; $P = 1.39 \times 10^{-6}$ and $P = 0.01$, respectively) and similar success rates compared with those of mAbs generated by immunization with protein domains by the same route (Supplementary Fig. 4b; Supplementary Table 16, comparisons 2, 21, 40, and 59; $P > 0.05$).

Analysis of denatured HuProt identifies IB-grade mAbs

Proteins on the HuProt were primarily in native conformation (Supplementary Fig. 1). Furthermore, only 3.6% of passing mAbs were IP^- and IB^+. This suggested that analysis of a denatured HuProt array might selectively identify IB-grade mAbs. To test this, we determined *z*- and *S*-scores for both IB^+IP^+ (47 mAbs to 38 targets) and IB^-IP^+ (29 mAbs to 28 targets) mAbs using denatured HuProt arrays (Supplementary Table 17). We observed that IB^+ mAbs passed our denatured HuProt screen more often than IB^- mAbs (Fisher's exact test, $P = 0.002$; Fig. 4i).

Antibody specificity is influenced by immunogen source

A number of mAbs that met our validation criteria (IP^+ and/or IB^+ with *S*-score > 2.5) bound to at least four potential cross-reactors (rank 2–5, *z*-score > 3), although with lower specificity than for their intended targets. We analyzed these and found that ~71% of all off-targets were not in the same target class/subclass as the intended target (Fig. 4j). Cross-reactors of mAbs generated from domains were more likely to be in the same target class/subclass than were those generated by full-length antigens (Fig. 4k; Fisher's exact test, $P = 2.8 \times 10^{-35}$). This suggests that mAb specificity screens that test only homologs of the intended target will not identify the majority of cross-reactive proteins.

DISCUSSION

Figure 5 summarizes our pipeline that resulted in the generation of 1,406 IP-grade and/or IB-grade PCRP mAbs (Supplementary Table 8) to 737 target proteins (Supplementary Table 9). Supplementary Figure 5 and Supplementary Tables 18 and 19 provide information about mAbs by target class/subclass. Supplementary Table 20 presents success rates across the pipeline by antigen type and immunization route. The collection of mAbs generated and described here is currently distributed through both DSHB and commercial suppliers; further information and links are provided at http://proteincapture.org.

Antigen production is a major bottleneck in all antibody-production efforts[28], and was a challenge in early phases of this project. To a large extent, we were able to overcome this problem by implementing f.p. immunization, which requires substantially less antigen than i.p. injection. The great majority of mAbs obtained were IP-grade, with many also IB-grade. Proteins on the HuProt are predominantly in native conformation, and mAbs identified as specific by HuProt proved to be good candidates for use in IP. We also showed that mAbs that recognize their targets only when they are conformationally intact, as indicated by an IP

$^{+}IB^{-}$ result, concurrently did not recognize their cognate target on denatured HuProt. However, mAbs that also recognized linear epitopes, as indicated by an $IP^{+}IB^{+}$ result, were more likely to selectively recognize their cognate target on denatured HuProt arrays (Fig. 5i). This suggests that array-based screening can be adapted to identify antibodies according to the desired application, which could lead to substantial overall savings in both costs and time when integrated into an antibody-production pipeline.

A substantially smaller fraction of these mAbs were suitable for IHC or ChIP-seq. In fact, stringent *post hoc* evaluation led to the reclassification of some $HuProt^{+}$ mAbs as non-passers ($IP^{-}IB^{-}$); some of these same $IP^{-}IB^{-}$ mAbs, however, had been previously identified as passers in ChIP-seq (three mAbs to two targets) and IHC (eight mAbs to eight targets) experiments. This indicates that mAbs that do not pass HuProt analysis might still be useful in specific contexts, particularly if cross-reacting proteins are not expressed in the cell or tissue type under investigation.

It is important to note a number of limitations in our pipeline that might affect the usefulness of the mAbs described here. Not all full-length proteins in the human proteome are present on HuProt, and HuProt-based analysis will not query most alternative splice variants or post-translational modifications. Target proteins are typically present at high concentrations on HuProt, which may lead to incorrect classification of mAbs as cross-reactive. Furthermore, because overexpressed proteins were used for our secondary-validation analysis, the ability of each mAb to selectively detect endogenous protein will need to be validated by individual investigators. Finally, this resource will become truly stable and renewable only after these mAbs have been sequenced.

There is a growing awareness of the antibody reproducibility crisis, and concerted efforts to define a global standard for antibody-validation criteria are being made[29, 30]. Our study demonstrates the broad usefulness of protein-array-based specificity analysis as a standard approach for antibody validation.

ONLINE METHODS

Validating and curating clones used in HuProt

Proteins spotted on HuProt were purified from yeast transformed with expression vectors encoding the said ORF[31]. We used Blast+ to align the ORF sequence to multiple public databases (UniProt, CCDS, RefSeq, and Ensembl) to generate an integrated alignment score for each of our clones. If a clone covered the entire sequence of a known protein, we considered the clone as full length (F), whereas partial matches were regarded as indicative of truncated (TRUNC) clones. Because our source clones included ORFs containing untranslated regions, unannotated splice variants, and single-nucleotide polymorphisms, we categorized our clones into groups ranging from perfect matches to the known protein-coding transcriptome to yet unreviewed potential protein-coding ORFs. A detailed breakdown of this classification, along with our threshold parameters, can be accessed at https://collection.cdi-lab.com/public.

To systematically check and confirm the identity of each of the ORFs, we carried out bidirectional Sanger sequencing, Blast+-aligned to the expected record ORF sequence and also to the CCDS database to find the best matches. We considered a pair of reads as passng if at least one of the reads matched the reference in our records, with less than three substitutions, and no deletions or insertions. Surprisingly, about 8% of our collection contained inconsistencies as indicated by sequencing and was therefore omitted from any further HuProt analysis.

Target selection

In addition to the high-confidence TFs annotated previously[1], the list of proteins with TF-like activity was assembled on the basis of relevant GO annotations such as "regulation of transcription, DNA-dependent." Other genes that were well-established transcriptional coregulators or epigenetic modifiers according to the literature but were not classified as such by GO were added manually. Unconventional sequence-specific DNA-binding proteins that showed clear evidence of being able to bind specific DNA sequences in electrophoretic mobility shift assays[16] were also added. Finally, we identified TF-associated proteins from protein–protein interaction databases and included a subset of these, prioritizing only those considered to be both of high interest to the scientific community and lacking high-quality commercially available reagents.

Antigen production

TF antigens used for immunization, in the forms of domains or full-length proteins, were produced in two expression systems. Defined TF domains were expressed as AviTag-labeled proteins from a T7-containing plasmid in *E. coli* (Tuner(DE3)/pLysSRARE2) grown in Luria–Bertani (LB) medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) with appropriate antibiotics. When culture density reached an OD_{600} of 0.8, protein expression was induced by addition of IPTG to 0.2 mM. Expression was continued for 16 h at 17 °C, after which cells were collected by centrifugation and lysed by sonication in 50 mM Tris-HCl, 6 M urea, 10 mM imidazole, 500 mM NaCl, 1 mM TCEP, 0.02% NaN_3, pH 7.5. Lysates were clarified by centrifugation (30 min at 10,000*g*, 4 °C), and the soluble fraction was used for HisTrap-mediated purification. Bound proteins were refolded on-column using a reverse gradient from 6 M to 0 M urea, and then bound proteins were eluted with an imidazole gradient. Eluted proteins were dialyzed in 20 mM HEPES, 100 mM NaCl, 400 mM L-arginine, 20% glycerol, 1 mM TCEP, 0.02% NaN_3, pH 7.5. Purity was assessed by SDS-PAGE and MALDI-TOF-MS, and if necessary, proteins were further purified by gel-filtration chromatography in the same buffer. At early stages of the program, antigens were suspended in PBS to reach the desired concentration for immunization. To prevent footpad necrosis, antigens were concentrated, dialyzed three times against PBS containing 10–20% glycerol, and in some cases supplied as a mixed suspension of soluble and precipitated material. Antigens were then stored frozen at −80 °C until needed for immunization. Full-length TFs were expressed as GST-fused proteins as described[13].

Animal work and immunization

Animals were housed in the animal facilities of Ponce Health Sciences University, which has an approved OLAW Animal Welfare Assurance (#A3585-01) on file with the Office for

Protection from Research Risks of the NIH. For immunization, antigens were adjusted to a concentration of 2–4 mg/mL by spin concentration or by dilution with PBS. For footpad immunization, 25–30 μg of protein in 12.5 μL was mixed with an equal volume of TiterMax adjuvant, vortexed, and injected into one rear footpad of a 6–8-week-old female BALB/c or CD-1 mouse. In some cases, mice were killed 12–14 d later and popliteal lymph nodes were aseptically collected. In other cases, 12 d after footpad immunization, an equal amount of antigen/adjuvant was injected into the hock area above the same foot. Three days later these mice were killed and their popliteal lymph nodes were aseptically harvested. For i.p. immunizations, 100 μg of antigen was mixed with an equal volume of Sigma Adjuvant system adjuvant, vortexed, and injected. On days 14, 28, and 42, the same amounts of antigen/adjuvant were administered by i.p. injection, after which mice were killed on day 46 and spleens were aseptically harvested. For a small cohort, a combination of i.p. and intravenous injection was used. On day 0, 100 μg of antigen plus adjuvant (Sigma Adjuvant system) was injected i.p., and on days 14 and 21, 50 μg of antigen mixed with adjuvant was injected i.p. On days 28, 29, and 30, 50 μg of antigen without adjuvant was injected into the tail vein. On day 31 the mice were killed and their spleens were aseptically harvested for hybridoma production. As per recommendations of the American Veterinary Medical Association, a mixture of ketamine (7.5–12 mg per 100 g) and xylazine (1.0–1.5 mg per 100 g) was used for anesthesia before bleeding and collection of spleen and lymph node cells.

Hybridoma generation

Immune lymph nodes and spleens were teased into single-cell suspensions in DMEM containing 100 U/mL penicillin and 0.1 mg/mL streptomycin (P–S), and cell numbers and viability were determined. Immune cells and Sp2/Ag0 myeloma cells (ATCC) were washed twice with DMEM– P-S, combined at a 5:1 ratio, and fused with 50% buffered PEG (Hybri-Max; Sigma), according to standard procedures[28]. Fused cells were plated in DMEM with 20% FBS, P-S, 1× HAT, 1× HFCS, 2% methylcellulose, and 1 μg/mL goat anti-mouse IgG–Alexa Fluor 488. After growth for 8–10 d at 37 °C, 5% CO_2, plates were scanned with a fluorescence dissecting microscope (Olympus MVX10), and positively staining colonies were transferred with microcapillary pipets into individual wells of 96-well plates. Colonies were expanded in DMEM with 10% FBS, P-S, 1× HT, at which point supernatant was assessed for the presence of IgG by ELISA. Supernatants from IgG-secreting clones were tested by GST-specific ELISA to identify and eliminate clones producing antibodies that bound to the fusion tag.

Protein arrays

We created microarrays on glass slides by spotting the immunogens in a 2 × 7 sub-array pattern (‘minichips’) on epoxide-coated slides using an ArrayIt NanoPrint LM210 microarray printer. Slides were fitted with 2 × 7 format gaskets, blocked with PBS/BSA, and IgG-positive culture supernatants were added to individual sub-array wells. Binding of an antibody to the spotted proteins was revealed by incubation of all wells with goat anti-mouse IgG (H+L)–Alexa Fluor 647, washing, and array scanning with a Molecular Devices GenePix 4000B scanner supported with GenePix Pro 7 software. When GST-tagged full-length proteins were administered, ELISA-based counter-screening was performed to exclude mAbs that recognized GST. A single replicate was performed for each mAb tested,

and positive signal was considered as any spot significantly brighter than all other spots in the same well when examined by eye. Clones secreting antigen-binding antibodies were expanded to T-600 flasks, and the cells were collected by centrifugation, suspended in growth medium plus 10% DMSO, and frozen in liquid nitrogen. The associated culture supernatants were used for purification of IgG with protein G–Sepharose. To determine the relative level of target-protein binding specificity, we screened the purified IgGs individually on HuProt. A single replicate was performed for each mAb tested. Antibody binding to a cognate target on the HuProt was scored using the *z*- and *S*-scores as described previously[13]. Briefly, the *z*-score represents the strength of a signal that a mAb (in combination with a fluorescently tagged secondary antibody to IgG) produces when binding to a particular protein on the HuProt array. *z*-scores are described in units of s.d. above the mean value of all signals generated on that array. If targets on HuProt are arranged in descending order of the *z*-score, the *S*-score is the difference (also in units of s.d.) between the *z*-scores of consecutive targets. *S*-score therefore represents the relative target specificity of a mAb to its intended target. A mAb is considered specific to its intended target if it has an *S*-score of at least 2.5. In some cases, HuProt assays were denatured by pretreatment with 9 M urea and 5 mM DTT before analysis.

For analysis of *Xist* RNA binding to HuProt, Cy5-UTP-incorporated cRNA probes of *Xist* produced by T7-directed transcription were a kind gift from Eric Lander's lab (MIT). Labeling and developing of the HuProt was performed as described previously[32].

Antibody purification

Antibody from culture supernatant was affinity purified with Protein G beads according to the manufacturer's instructions (Sigma Aldrich; P7700). For concentration of the eluted IgG samples and buffer exchange, we added 3 column volumes of PBS (Corning Cellgro; 46-013-CM), 30% glycerol (Sigma Aldrich; G5516), and 0.02% sodium azide (Sigma Aldrich; S2002) to a Spin-X UF 6-mL centrifugal concentrator with a 50,000 MWCO membrane (Corning, 431485). Samples were reduced to a volume of 1 mL, and antibody concentration was determined with a NanoDrop Lite spectrophotometer (Thermo Fisher Scientific).

Cell culture for secondary validation

Adherent 3G-Tet-ON HeLa cells (Clontech; 631183) were maintained in DMEM (high-glucose; Gibco) supplemented with 10% FBS (tetracycline-free; Clontech; 631106), 4 mM glutamine (Gibco) and 100 units/ml penicillin–streptomycin (Invitrogen, Carlsbad, CA). Tet-On HEK293$_{LD}$ suspension cells[33, 34] were grown in Freestyle 293 medium (Gibco) supplemented with 4 mM L-glutamine and 1% FBS (tetracycline-free; Clontech; 631106) in 8% CO_2. Suspension cells were maintained between 0.3 and 3 × 10^6 per ml in disposable baffled and vented PETG shake flasks (Thermo Fisher) at 165 r.p.m. using a Celltron orbital shaker (25-mm orbital diameter; INforsHT). Suspension cells were always seeded at a volume of up to 30–40% of the culture bottle's capacity.

Plasmids and transfection

We used a Gateway-compatible and tetracycline-inducible human expression vector (HuEV-A)[26] for expression of the intended target of a test antibody. The HuEV-A (Addgene plasmid #68342) construct appends an N-terminal fusion of 3×Flag peptide, a V5 epitope tag followed by Venus fluorescent protein (YFP variant) to the expressed protein. Transfection of expression constructs in the adherent HeLa cells was done with FugeneHD (Promega) per the manufacturer's instructions. For suspension cells, transfection was carried out with polyethylenimine 'Max' high-potency linear PEI (Polysciences) as described[35]. We observed a target-dependent sensitivity of certain proteins that led them to form aggregates when overexpressed after induction. By titrating down the doxycycline levels to reduce induction of the protein, and by reducing the incubation time post-transfection, we overcame the majority of such aggregation artifacts.

Immunoprecipitation and immunoblotting

16–24 h post-transfection, cells were split into six fractions; one fraction was left uninduced as a control, two fractions were induced with low levels of doxycycline (1–50 ng/ml), and the remaining three fractions were induced with high levels of doxycycline (20–1,000 ng/μl). Each fraction corresponded to ~5×10^5 adherent HeLa cells or ~2×10^6 $HEK293_{LD}$ suspension cells. After 16–48 h of induction, we confirmed expression of target protein by screening for YFP expression. After confirming YFP expression, we collected cells in PBS, pelleted them, and stored them at −80 °C until further use. On the day of IP, respective pellet fractions of a given target were vortexed in lysis buffer (100 mM Tris-HCl, 150 mM NaCl, 25 mM NaF, 50 μM $ZnCl_2$, 15% glycerol, 1% Triton X-100) supplemented with protease inhibitors (Roche; 11697498001) and BitNuclease (Biotools; B16003). After a 30-min incubation at 4 °C, ~80% of the supernatant was collected after centrifuging at 4,000 r.p.m. for 20 min at 4 °C. An aliquot of the supernatants was stored as input for the IP.

Supernatants from samples induced with high levels of doxycycline were next incubated with 5 μg of either mouse IgG sera (sc-2025; Santa Cruz) or the test CDI monoclonal antibody, or with 1 μg of the positive control monoclonal M2 anti-Flag (F1804; Sigma). Additionally, supernatants from samples induced with low levels of doxycycline were also incubated with 5 μg of the test CDI monoclonal antibody, to test the sensitivity range of the antibody in IP experiments. Incubation was carried out with gentle rocking overnight at 4 °C. Next, the antibody–protein complexes were pulled down in a 2-h incubation with Protein G Dynabeads (Life Technologies; 10004D) at 4 °C, washed three times with lysis buffer, and eluted in 35 μl of LDS sample loading buffer supplemented with β-mercaptoethanol as a reducing agent (Life Technologies; NP0008). Input lysate and the IP samples were resolved in an SDS-PAGE gel (Life Technologies; WG1403A and EP04808) and immunoblotted with both rabbit anti-Flag (1:1,000; 2368; Cell Signaling) and the test monoclonal antibodies from CDI. Anti-rabbit IRDye680RD (LiCor; 925-68071), anti-mouse IRDye800CW (LiCor; 926-32210), and light-chain-specific anti-mouse Alexa Fluor 790 (Jackson Immunolabs; 115-655-174) secondary antibodies were used for visualization of bands, and the blots were imaged with an infrared fluorescence imager (LiCor Clx).

Competitive immunoprecipitation protocol

The IP protocol described above was modified for testing of commercially sourced antibodies (Supplementary Table 5) for specific recognition of their intended target. $HEK293_{LD}$ suspension cells were transfected individually with either the intended target or off-targets identified by the HuProt screens (Supplementary Table 6). 24 h after transfection, expression of constructs was induced by doxycycline (20 ng/μl), and the constructs were collected 24 h after induction in aliquots containing ~2×10^6 cells each. Off-targets with the strongest signal on HuProt (HuProt rank) were selected for this assay, provided the off-targets were sufficiently different than the intended target in molecular weight and showed robust expression when transfected in mammalian cell lines (see notes in Supplementary Table 5). For each antibody tested, aliquots were lysed, and one-third of the lysates each of the intended target and the two chosen off-targets were combined together. The combined lysates were then tested for IP using 5 μg of mouse IgG sera (sc-2025; Santa Cruz), a selected CDI monoclonal antibody, or the commercially sourced antibody, or 1 μg of the positive control M2 anti-Flag mAb (F1804; Sigma). Particulars such as clone ID, catalog number, and lot number for both CDI and commercially sourced mAbs are detailed in Supplementary Table 4. The three input lysates, along with the IP samples, were resolved by SDS-PAGE gel and immunoblotted with both rabbit anti-Flag (1:1,000; 2368; Cell Signaling) and 5 μg of the commercially sourced test mAbs.

ChIP-seq

Exponentially growing cells from ENCODE tier 1 and tier 2 cell lines were cross-linked in 1% formaldehyde and harvested in batches of 20 million cells per replicate. Prepared nuclei were resuspended in RIPA buffer, and chromatin shearing was done on a Bioruptor Twin instrument (Diagenode). Antibody (5 μg per reaction) was coupled to M-280 Dynabeads (Thermo Fisher Scientific; 11201) overnight before the addition of sheared chromatin. A small aliquot of the sheared chromatin was reserved as an input control. After the incubation and wash steps, captured chromatin and input control were reverse-cross-linked overnight at 65 °C and recovered with the QIAquick PCR purification system (Qiagen; 28104). ChIP-seq libraries were prepared and run on an Illumina HiSeq 2500 instrument. ChIP-seq biological replicates represent IPs from two independent growths of the same cell line. Sequencing reads were aligned to the human hg19_Female genome using Bowtie2. We used the MACS2 algorithm to find significant peak signals above background (input control), and Homer to find enriched known and *de novo* motifs. A detailed protocol can be found at the ENCODE Data Portal: https://www.encodeproject.org/documents/6ecd8240-a351-479b-9de6-f09ca3702ac3/@@download/attachment/ChIP-seq_Protocol_v011014.pdf.

Immunohistochemistry

IHC on human tissue was performed on 5-μm-thick sections cut from formalin-fixed, paraffin-embedded human tumor specimens and mounted on charged slides (Mercedes Medical). The slides were deparaffinized in SlideBrite, washed with 100% ethanol, and rehydrated in a graded ethanol series (95%, 70%, and 50%). We blocked endogenous peroxidase by immersing slides in 3% hydrogen peroxide for 5 min and then washing them in distilled water. For antigen retrieval, we heated slides in 10 mM citrate buffer, pH 6.0, in a

pressure cooker (Decloaking Chamber, Biocare Medical) at 125 °C for 20 min. Tissue slides were then cooled to 80 °C and washed in water. All primary antibodies were diluted to 2 μg/ml in 10 mM PBS, pH 7.4, containing 3% BSA and 0.05% a zide. Slides were incubated for 30 min with the diluted antibodies and then rinsed with PBS containing 0.05% Tween-20. Tissue sections underwent goat anti-mouse horseradish peroxidase polymer detection (ScyTek Laboratories) for 15 min at room temperature and then were rinsed in PBS as before. Sections were finally incubated in 3,3′-diaminobenzidine (Scytek Laboratories) for 5 min. Counterstaining was done with hematoxylin. Negative controls consisting of diluent with no antibody were used in all experiments.

IHC on mouse retina and brain tissue was carried out as described previously[36, 37]. Briefly, tissue was collected after animal perfusion, post-fixed after harvest in 4% paraformaldehyde for 1 h at 4 °C or fresh-frozen in Clear Frozen Section Compound (VWR), and stored at −80 °C until use for cryo-sectioning. Glass-slide-mounted tissue sections (15–20-μm sections) were first blocked with standard blocking buffer and then incubated with primary antibodies (1:200) overnight at 4 °C. After a wash with 1× PBST (1× PBS + 0.05% Triton X-100), secondary antibodies conjugated to Alexa Fluor 488 (1:500) were incubated for 2 h at room temperature in the dark. After being washed and counterstained with the nuclear stain DAPI (1:5,000), slides were mounted with VectaShield (Vector Laboratories) and recorded with a confocal and/or BZ-X700 Keyence fluorescence microscope. Each antibody was tested once, with 4–8 sections of the same source tissue analyzed in each experiment.

Image processing and quantification

The software was developed to allow faster and more automated processing and analysis of gel images. It consists of a web-based interface and server-hosted Python and Perl scripts for data organization and image processing. Source code for the gel-analysis software is freely available at https://github.com/FenyoLab/cdilab, and the fully working software is hosted at http://cdilabs.fenyolab.org. The left-hand pane of the interface provides a tree view of all uploaded files and allows for the creation of folders for organization. The right-hand pane is the analysis window. The initial input for the software is a high-dynamic-range image (HDRI) in red/green channels, for which the software will convert to a format that can be displayed on screen, and a text file with input parameters. The software contains a custom algorithm to detect and separate each gel in the collage image. The interface provides a proposed sectioning scheme, which can be adjusted by the user and separated into individual images for further processing.

For each gel, the software will automatically detect bands in the marker lane and associated masses, and their location will be indicated on screen for the user to adjust manually, as needed. These locations will be saved and used later in the processing pipeline to create a labeled image. The software also allows for adjustment of the clarity of the image by allowing the user to interactively set the minimum and maximum display pixels for the conversion from HDRI. Further, for densitometry calculation, relevant bands are detected by a custom algorithm and band intensity is measured from the original HDRI image pixels. The signal is calculated the sum of the intensity minus the background, where background is the median intensity over a 3-pixel-wide rectangle surrounding the band. The user may

adjust the outline defining the detected bands interactively. All densitometry calculations are then saved to an output text file. The software also produces labeled images for catalog use based on the input parameter file. These labeled images are made available for download as a .zip file via a link in the web interface.

Scoring and threshold criteria for mAbs

For secondary validation, antibodies were evaluated against preset and standardized thresholds for quality and efficiency in IP and IB applications. For IP-grade mAbs, after densitometric analysis of signal intensity in the immunoblots, test antibodies with >0.15% IP recovery, as compared with that for the total input sample, were considered as passers. Results for each antibody were manually inspected by at least two independent investigators before a determination was made as to passing/non-passing status. As long as the target protein at the expected molecular weight was observed in the input and corresponding IP lanes, the reagent was considered as having passed. The ability of the antibody to IP-recover the degraded fragments was not evaluated or considered with respect to the efficiency of the IP recovery.

All constructs expressing the target ORF were verified for integrity by Sanger sequencing of the construct expression cassette. However, we observed up to 20-kDa differences relative to the expected molecular weight for certain target proteins in our SDS-PAGE gels. As the protein was not visible in the lanes containing lysates of uninduced cells, we considered such cases to be under the normal purview of aberrant migration reported for many proteins in the literature.

Because we used Tet-ON HeLa or HEK cell lines in our IP experiments, we expected to observe no expression of the transfected protein in the absence of tetracycline/doxycycline treatment. However, we occasionally observed leaky expression of the target protein even in the absence of doxycycline, a known property of the older generation of Tet-ON expression cassettes used in our cell lines. We considered the experiment to be acceptable if the levels of target protein visibly increased at least twofold compared with those in the uninduced and highest-induction samples in the input lanes.

For unknown reasons, in a small subset of our experiments we observed that the rabbit anti-Flag mAb (Cell Signaling) recognized nonspecific signal(s) in all three input lanes. This band intensity did not change as a function of doxycycline treatment, and was not recovered by IP with the IgG or test antibodies. Such nonspecific signal was not considered relevant for experimental evaluation of the antibody. We observed that certain target proteins had an increased propensity to form insoluble/sticky protein aggregates when expressed at high levels. This feature was readily visible as a strong signal in the IgG-IP lanes of our experiments. By serially titrating the level of induction by doxycycline, we were able to prevent this unwanted aggregation of target proteins with concomitant loss in signal intensity of the IgG-IP lanes. However, in certain scenarios a weak but visible signal might be present in the IgG lane. If the densitometric signal observed in the IP lane (lane 6) using the test mAb was found to be at least twofold higher than the cognate signal in the IgG-IP lane (lane 4), we considered the antibody to be a passing reagent.

In a substantial fraction of our experiments, we observed failure of recovery in our positive control with mouse anti-Flag IP (lane 7). As the Flag tag was an N-terminal fusion to the target protein, a likely explanation for the absence of recovery is epitope masking/unavailability. Therefore, absence of a signal in the positive control was not considered a reason to fail the test antibody if all other results for that antibody met the criteria for passing. For evaluation of the performance of mAbs in IB applications, a reagent was considered to pass if the signal was observed at the expected size in either input (lane 2 or 3) or IP lanes (lane 5, 6, or 7), but not in the uninduced (lane 1) and IgG-IP lane (lane 4). The presence of a band at an unexpected size in either input or IP lanes was not considered negatively in evaluations of the reagent. Such background signal in IB can probably be eliminated through standard techniques but must be evaluated by the end user of these reagents.

Optimization and maintenance of hybridoma production for passing antibodies

Hybridoma cultures were expanded in medium containing DMEM (Cellgro, 10-013-CM), 15% FBS (Cellgro, 35-016CV), and P-S (Mediatech, MT 30-002-CI) and subcultured every 48 h, while maintaining a minimum of 80% viability. Before freezing, cell lines were tested for mycoplasma contamination with a PCR mycoplasma detection kit (ABM; G238). If the result of the assay was negative, each cell line was expanded to a tissue culture flask of 175 cm^2 and incubated for another 48 h at 37 °C, 5% CO_2. Cell lines with a density of 5–8 × 10^5 cells/mL, and with over 90% viability, were collected by centrifugation. Culture supernatant was stored at −20 °C for further testing, and cells were suspended in 10% DMSO (Sigma-Aldrich; D2650), 20% FBS (Cellgro; 35-011CV) and P-S (Mediatech; MT 30-002-CI) and frozen. Isotyping of each hybridoma cell line was done with the Rapid Mouse Antibody Isotyping XpressCard (Antagen Pharmaceuticals, Inc., ISO-M8ac). Frozen hybridoma cell lines were shipped to the Developmental Studies Hybridoma Bank at the University of Iowa (Iowa City, IA, USA).

ELISA, SPR, and IP-MS experiments

Detailed protocols for these experiments can be accessed at https://proteincapture.org/protocols.

Meta-analysis of pipeline efficiency

We empirically determined *z*- and *S*-scores for mAbs that were tested on HuProt, and for a select subset we also measured affinity values by Octet and/or OIRD. We compared differences in procedures used to generate the mAbs and/or structural differences among the intended targets of individual mAbs. With the procedures used, our mAbs could be categorized into groups on the basis of four variables: (i) the type of antigen used (domains (D) versus full-length (F) protein), (ii) whether mAb raised against domains recognized only the immunized domain (D) or both the domain and its full-length (F) counterpart on the HuProt, (iii) the immunization protocol (i.p. versus f.p.) used, and (iv) whether a single antigen or a pool of antigens was used for immunization. Unless otherwise stated, we made comparisons to mAbs generated from both single and multiple immunogens. Furthermore, as detailed immunization records for a subset of targets were not adequately maintained during the initial phase of the project, these were excluded from further analysis. By

combining the remaining three variables, we categorized our mAbs and represented them using the abbreviations mentioned above. The letter "A" was used when all subgroups in a given variable were considered. For example, D-D-A represents all mAbs that were generated with a domain immunogen that recognizes only its cognate domain immunogen on HuProt by any of the immunization routes.

At the structural level, targets are hierarchically classified such that multiple targets are represented in each class/subclass, and multiple mAbs represent each target. A success rate calculated at the level of target, target subclass, or target class is defined as the fraction of corresponding mAbs that passed relative to the total number tested. Success rates were calculated at the stages of minichip, HuProt, secondary validation, and overall pipeline (by taking the multiplication product of the success rates for minichip, HuProt, and secondary validation).

We wished to generate comparisons by combining the procedural groups with structural classification. If a target was represented by mAbs generated by more than one type of procedure, we separated them into a category labeled "mixed." We restricted comparisons of groups with at least 5 mAbs and excluded a subset of comparisons manually to remain conservative in making conclusions. Two prominent examples of such exclusion are (1) mAbs generated from full-length antigen by i.p. injection, as this represented a very small subset, and (2) comparisons with mAbs in the 'mixed' category.

Antibody availability

Hybridomas and mAbs described in this study are available through the DSHB. Purified mAbs are available through NeoBiotechnologies, CDI Labs, and other third-party commercial suppliers via https://proteincapture.org/.

Statistical analysis

Unpaired two-tailed Student's *t*-test with Welch's correction was carried out with GraphPad Prism software. For the *P* values in Figure 5, the *t*-test function in statistical software R was used, and effect size was calculated using the cohensD function of package lsr. The Benjamini–Hochberg FDR was calculated via the R p.adjust function. We used the R pwr package (power.t.test) to check sample size. If the true difference in means is >2, then a sample size of 5 can reach a power of 0.8. Therefore, when using t.test to compare two groups, we excluded any group with a sample size of <5 and considered all categories with $n > 5$ for statistical comparison. For FDR calculations, we grouped hypotheses into two groups representing comparisons of parameters measured for individual mAbs (Supplementary Tables 9, 11, and 13) and at the level of targets (Supplementary Tables 10, 12, and 14). Additional details are available in the **Life Sciences Reporting Summary**.

Code availability

Source code for immunoblot processing software is freely available at https://github.com/FenyoLab/cdilab, and the fully working software is hosted at http://cdilabs.fenyolab.org.

Life Sciences Reporting Summary

Further information regarding experimental design is available in the **Life Sciences Reporting Summary**.

Data availability

PCRP antibody production and validation standard operating protocols and processed data are hosted publicly at https://proteincapture.org/. The raw and processed ChIP-seq data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus[38] and are accessible through GEO Series accession number GSE97661. Source data for Figures 2–4 and Supplementary Figures 1, 4 and 5 are available online.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Anand Venkataraman[1,iD], Kun Yang[2], Jose Irizarry[3], Mark Mackiewicz[4,iD], Paolo Mita[5,6,7,iD], Zheng Kuang[6,19,iD], Lin Xue[1], Devlina Ghosh[1], Shuang Liu[8], Pedro Ramos[3], Shaohui Hu[3], Diane Bayron Kain[3,19], Sarah Keegan[5,6], Richard Saul[8], Simona Colantonio[9], Hongyan Zhang[8], Florencia Pauli-Behn[4], Guang Song[8], Edisa Albino[3], Lillyann Asencio[3], Leonardo Ramos[3], Luvir Lugo[3], Gloriner Morell[3], Javier Rivera[3], Kimberly Ruiz[3], Ruth Almodovar[3], Luis Nazario[3], Keven Murphy[3], Ivan Vargas[3], Zully Ann Rivera-Pacheco[3], Christian Rosa[3], Moises Vargas[3], Jessica McDade[7], Brian S Clark[1], Sooyeon Yoo[1], Seva G Khambadkone[10], Jimmy de Melo[1], Milanka Stevanovic[1], Lizhi Jiang[1], Yana Li[11], Wendy Y Yap[3], Brittany Jones[12], Atul Tandon[12], Elliot Campbell[13,14], Gaetano T Montelione[13,14], Stephen Anderson[13,14], Richard M Myers[4], Jef D Boeke[5,6,7], David Fenyö[5,6], Gordon Whiteley[9], Joel S Bader[2], Ignacio Pino[3,iD], Daniel J Eichinger[3,iD], Heng Zhu[8,15,iD], and Seth Blackshaw[1,15,16,17,18,iD]

Affiliations

[1]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[2]Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[3]CDI Laboratories, Mayaguez, PR, USA

[4]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA

[5]Institute for System Genetics, NYU Langone Health, New York, New York, USA

[6]Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, New York, USA

[7]Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[8]Department of Pharmacology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[9]Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, Maryland, USA

[10]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[11]Eukaryotic Tissue Culture Facility, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[12]NeoBiotechnologies, Inc., Union City, California, USA

[13]Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey, USA

[14]Center for Advanced Biotechnology and Medicine, Piscataway, New Jersey, USA

[15]Center for Human Systems Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[16]Department of Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[17]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[18]Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Acknowledgments

This work was supported by the NIH Common Fund (awards U54HG006434 (to J.D.B., S.B., and H.Z.) and U01DC011485 (to S.A. and G.T.M.)). Cy5-UTP-incorporated cRNA probes of *Xist* produced by T7-directed transcription were a kind gift from E. Lander's lab (MIT, Cambridge, Massachusetts, USA).

References

1. Bradbury A, Plückthun A. Reproducibility: standardize antibodies used in research. Nature. 2015; 518:27–29. [PubMed: 25652980]
2. Weller MG. Quality issues of research antibodies. Anal. Chem. Insights. 2016; 11:21–27. [PubMed: 27013861]
3. Pauly D, Hanack K. How to avoid pitfalls in antibody use. F1000Res. 2015; 4:691. [PubMed: 26834988]
4. Bordeaux J, et al. Antibody validation. Biotechniques. 2010; 48:197–209. [PubMed: 20359301]
5. Saper CB, Sawchenko PE. Magic peptides, magic antibodies: guidelines for appropriate controls for immunohistochemistry. J. Comp. Neurol. 2003; 465:161–163. [PubMed: 12949777]
6. Schonbrunn A. Antibody can get it right: confronting problems of antibody specificity and irreproducibility. Mol. Endocrinol. 2014; 28:1403–1407. [PubMed: 25184858]
7. Hornsby M, et al. A high through-put platform for recombinant antibodies to folded proteins. Mol. Cell. Proteomics. 2015; 14:2833–2847. [PubMed: 26290498]
8. Marcon E, et al. Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. Nat Methods. 2015; 12:725–731. [PubMed: 26121405]

9. Na H, et al. A high-throughput pipeline for the production of synthetic antibodies for analysis of ribonucleoprotein complexes. RNA. 2016; 22:636–655. [PubMed: 26847261]
10. Rhodes KJ, Trimmer JS. Antibodies as valuable neuroscience research tools versus reagents of mass distraction. J. Neurosci. 2006; 26:8017–8020. [PubMed: 16885215]
11. Uhlén M, et al. Tissue-based map of the human proteome. Science. 2015; 347:1260419. [PubMed: 25613900]
12. Blackshaw S, et al. The NIH Protein Capture Reagents Program (PCRP): a standardized protein affinity reagent toolbox. Nat. Methods. 2016; 13:805–806. [PubMed: 27684578]
13. Jeong JS, et al. Rapid identification of monospecific monoclonal antibodies using a human proteome microarray. Mol. Cell. Proteomics. 2012; 11 O111.016253.
14. Hu CJ, et al. Identification of new autoantigens for primary biliary cirrhosis using human proteome microarrays. Mol. Cell. Proteomics. 2012; 11:669–680. [PubMed: 22647870]
15. Hu S, et al. DNA methylation presents distinct binding sites for human transcription factors. eLife. 2013; 2:e00726. [PubMed: 24015356]
16. Hu S, et al. Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. Cell. 2009; 139:610–622. [PubMed: 19879846]
17. Newman RH, et al. Construction of human activity-based phosphorylation networks. Mol. Syst. Biol. 2013; 9:655. [PubMed: 23549483]
18. Cox E, et al. Identification of SUMO E3 ligase-specific substrates using the HuProt human proteome microarray. Methods Mol. Biol. 2015; 1295:455–463. [PubMed: 25820740]
19. UzomaI, , et al. Global identification of SUMO substrates reveals crosstalk between SUMOylation and phosphorylation promotes cell migration. Mol. Cell. Proteomics2008
20. Chu C, et al. Systematic discovery of *Xist* RNA binding proteins. Cell. 2015; 161:404–416. [PubMed: 25843628]
21. McHugh CA, et al. The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature. 2015; 521:232–236. [PubMed: 25915022]
22. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. 2009; 10:252–263. [PubMed: 19274049]
23. Greenspan NS. Cohen's Conjecture, Howard's Hypothesis, and Ptashne's Ptruth: an exploration of the relationship between affinity and specificity. Trends Immunol. 2010; 31:138–143. [PubMed: 20149744]
24. Steward MW, Lew AM. The importance of antibody affinity in the performance of immunoassays for antibody. J. Immunol. Methods. 1985; 78:173–190. [PubMed: 2580911]
25. Abdiche Y, Malashock D, Pinkerton A, Pons J. Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. Anal. Biochem. 2008; 377:209–217. [PubMed: 18405656]
26. Mita P, et al. Fluorescence ImmunoPrecipitation (FLIP): a novel assay for high-throughput IP. Biol. Proced. Online. 2016; 18:16. [PubMed: 27528826]
27. de Melo J, et al. Lhx2 is an essential factor for retinal gliogenesis and Notch signaling. J. Neurosci. 2016; 36:2391–2405. [PubMed: 26911688]
28. HarlowE, , LaneD. Using Antibodies: A Laboratory ManualCold Spring Harbor Laboratory Press; 1998
29. Roncador G, et al. The European antibody network's practical guide to finding and validating suitable antibodies for research. MAbs. 2016; 8:27–36. [PubMed: 26418356]
30. Uhlen M, et al. A proposal for validation of antibodies. Nat. Methods. 2016; 13:823–827. [PubMed: 27595404]
31. Zhu H, et al. Global analysis of protein activities using proteome chips. Science. 2001; 293:2101–2105. [PubMed: 11474067]
32. Rapicavoli NA, Poth EM, Zhu H, Blackshaw S. The long noncoding RNA Six3OS acts in trans to regulate retinal development by modulating Six3 activity. Neural Dev. 2011; 6:32. [PubMed: 21936910]

33. Taylor MS, et al. Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. Cell. 2013; 155:1034–1048. [PubMed: 24267889]

34. Dai L, Taylor MS, O'Donnell KA, Boeke JD. Poly(A) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation. Mol. Cell. Biol. 2012; 32:4323–4336. [PubMed: 22907758]

35. Longo PA, Kavran JM, Kim MS, Leahy DJ. Transient mammalian cell transfection with polyethylenimine (PEI). Methods Enzymol. 2013; 529:227–240. [PubMed: 24011049]

36. de Melo J, et al. Injury-independent induction of reactive gliosis in retina by loss of function of the LIM homeodomain transcription factor Lhx2. Proc. Natl. Acad. Sci. USA. 2012; 109:4657–4662. [PubMed: 22393024]

37. Lee DA, et al. Tanycytes of the hypothalamic median eminence form a diet-responsive neurogenic niche. Nat Neurosci. 2012; 15:700–702. [PubMed: 22446882]

38. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002; 30:207–210. [PubMed: 11752295]

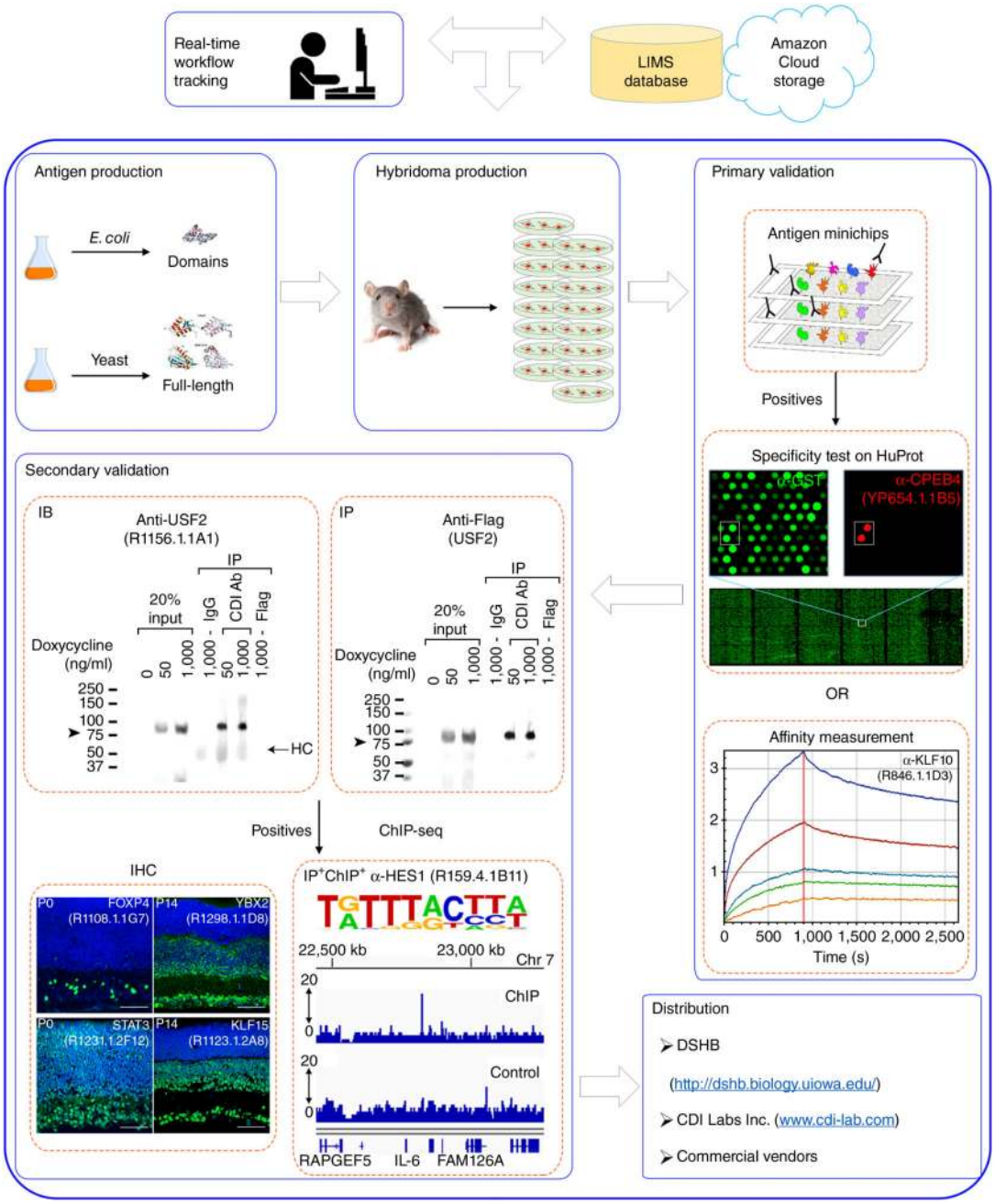

Figure 1.

Pipeline description. The critical steps involved in the generation and distribution of high-quality PCRP mAbs include antigen production; hybridoma production; primary validation using protein arrays; secondary validation by immunoprecipitation (IP), immunoblotting (IB), ChIP-seq, and immunohistochemistry (IHC); and finally distribution as a community resource. Antibody clone IDs are indicated in parentheses where applicable. LIMS, laboratory information management system; Ab, antibody; HC, heavy chain.

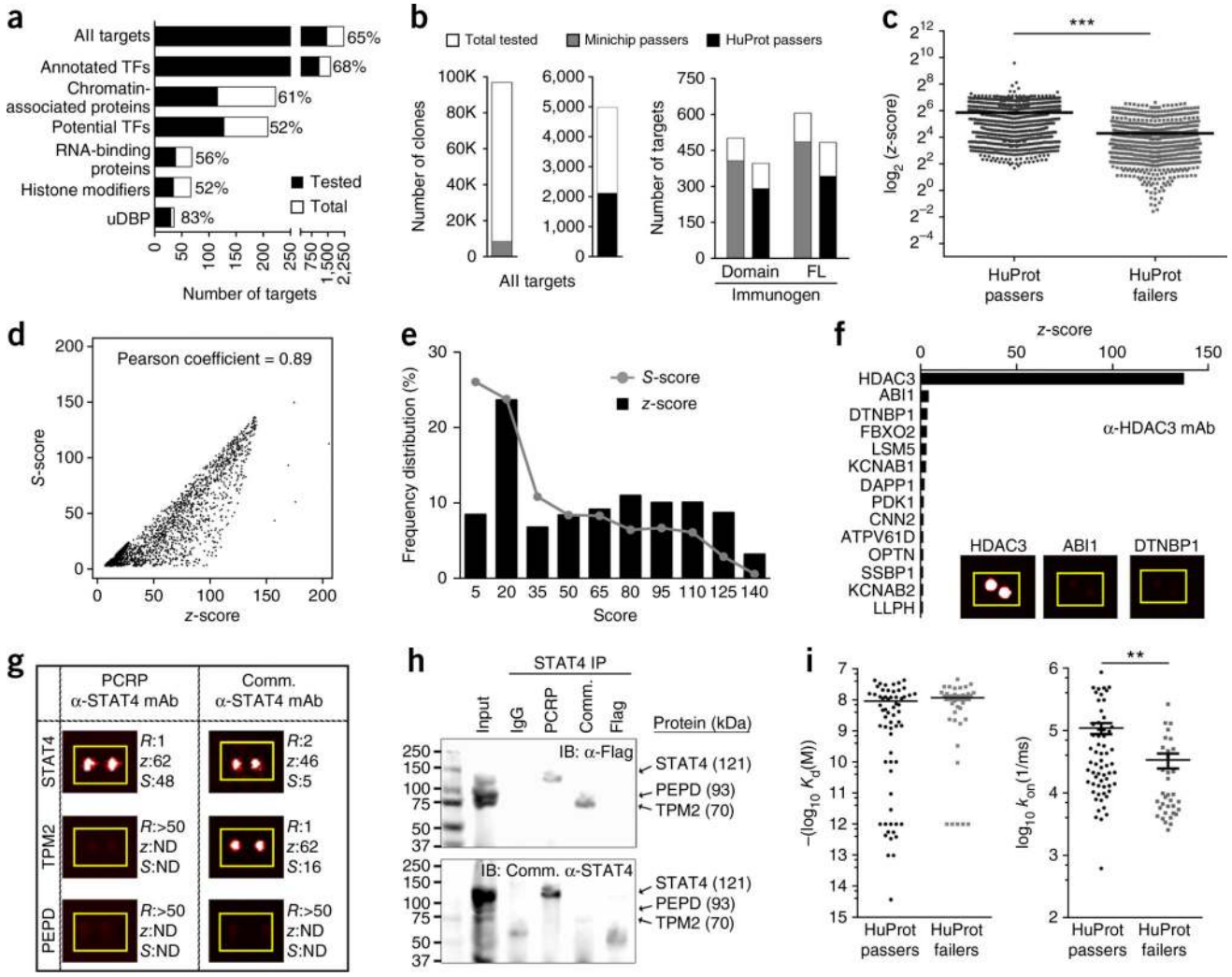

Figure 2.

Primary validation of PCRP reagents. (**a**) Targets of different functional classes tested for immunization. The values next to each bar indicate the percentage of the total number of targets tested in this study. uDBP, unconventional sequence-specific DNA-binding protein. (**b**) Number of mAbs and targets tested and identified as passing at minichip and HuProt stages. FL, full-length protein. (**c**) Comparison of the HuProt *z*-scores of passing (*S*-score > 2.5; *z*-score, 58.51 ± 0.962, mean ± s.e.m.; $n = 2{,}006$) versus failing mAbs (*z*-score, 19.66 ± 0.816, mean ± s.e.m.; $n = 640$). ***$P < 0.0001$, unpaired two-tailed *t*-test with Welch's correction. (**d**) Scatter plot showing the correlation between *z*- and *S*-scores for mAbs that passed HuProt ($n = 2{,}256$). (**e**) Frequency distribution of *z*- and *S*-scores among mAbs that passed the HuProt validation ($n = 2{,}082$). (**f**) *z*-scores for targets recognized by anti-HDAC3 (clone YP775.1.3C9). Insets show the HuProt signal observed for the indicated targets. (**g**) Micrographs representing the HuProt signal observed with either PCRP or commercial (Comm.; LSBio, LS-B5539, lot 67941) anti-STAT4 for the intended target (STAT4) and off-targets (TPM2 and PEPD). The corresponding rank (*R*), *z*-score (*z*) and *S*-score (*S*) for each target is shown next to the micrographs. ND, not determined. (**h**) IP experiment with mouse

pan-IgG sera, PCRP anti-STAT4 mAb (clone JH46.1.2A2), or commercial anti-STAT4 mAb (recommended for IP and IB). IP was carried out on a mixture of three lysates from HEK cells transfected with *STAT4, PEPD*, and *TPM2*. IB was carried out with both rabbit anti-Flag (top) and commercial mouse anti-STAT4 mAbs (bottom). (**i**) Comparison of affinity values (K_d and k_{on}) for mAbs that passed HuProt ($n = 62$; $K_d = 8.97 \times 10^{-9} \pm 1.43 \times 10^{-9}$; $k_{on} = 110{,}728 \pm 22{,}141$ (mean ± s.e.m.)) and for those that failed HuProt ($n = 41$; $K_d = 1.22 \times 10^{-8} \pm 1.86 \times 10^{-9}$; $k_{on} = 29{,}818 \pm 8{,}000$ (mean ± s.e.m.)). $^{**}P = 0.001$, unpaired two-tailed *t*-test with Welch's correction.

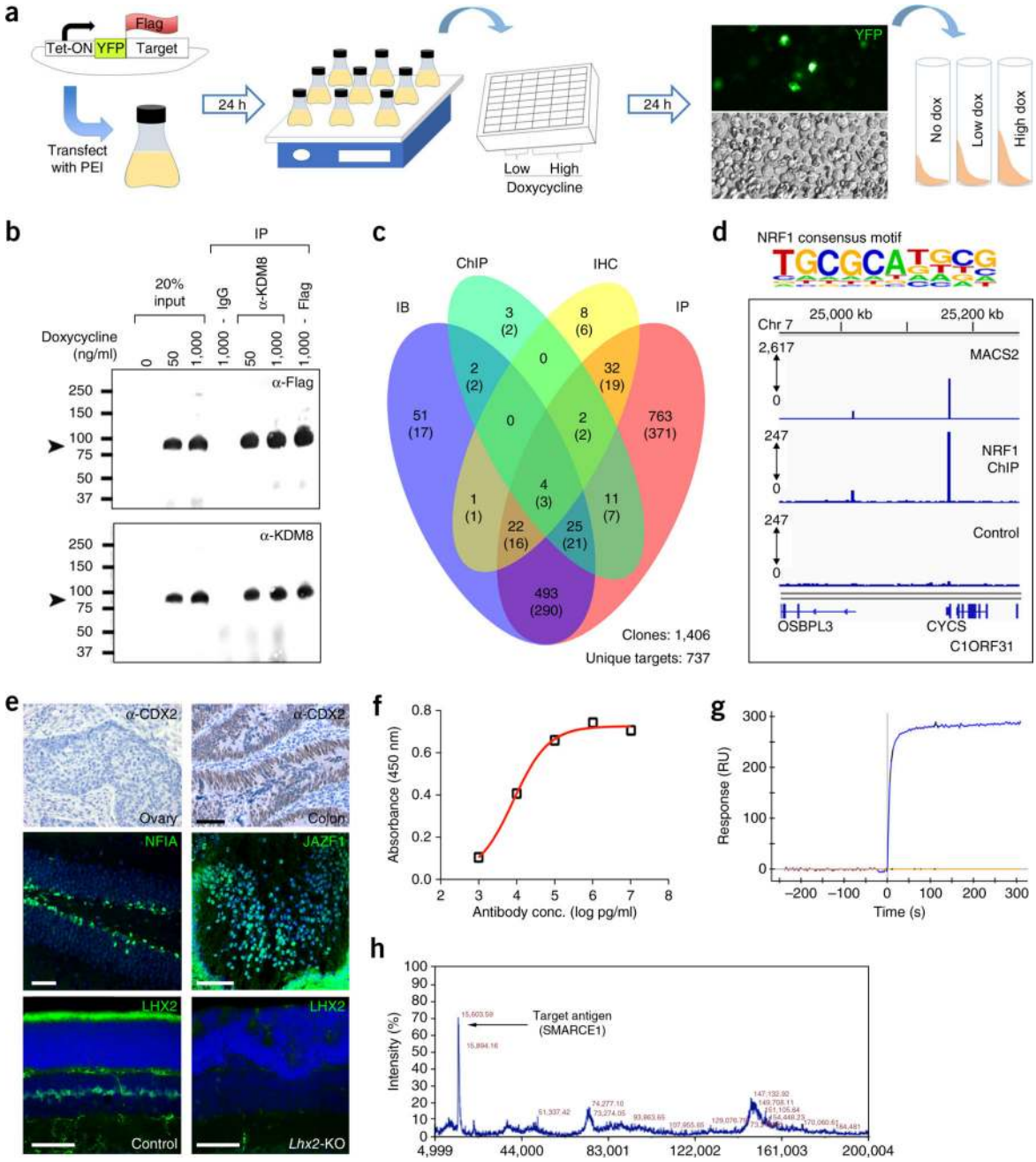

Figure 3.

Secondary validation of PCRP reagents. (**a**) The high-throughput pipeline for secondary validation included polyethylenimine (PEI)-mediated transfection of tetracycline-inducible constructs encoding an N-terminal YFP-tagged open reading frame of interest into $HEK293_{LD}$ suspension cells, scaled down to a 15-ml culture volume. Twenty-four hours after transfection, cells were induced with different doses of doxycycline in 48-deep-well plates and incubated for another 24 h. After confirmation of protein expression by visualization of the YFP fusion protein as a proxy, cell pellets were frozen and stored at −80 °C until use. (**b**) IP and IB data for an $IP^{+}IB^{+}$ anti-KDM8 mAb (clone YP50.1.1A2).

Arrowheads indicate the expected molecular weight of the expressed fusion KDM8 protein (84.6 kDa). (**c**) Venn diagram representing the distribution of mAbs and their corresponding targets (in parentheses) that were empirically determined to be applicable in IP, IB, ChIP, and/or IHC experiments. (**d**) ChIP-seq analysis using anti-NRF1. Top, a known consensus binding site for NRF1 was identified as the top enriched motif in the experiment. Bottom, a ChIP-seq peak detected in the promoter region of known NRF1 target CYCS. (**e**) Top, anti-CDX2 (clone R1435.1.1A3) tested for the detection of specific nuclear signal in human colon and ovarian tumors. Middle, NFIA and JAZF1 expression in adult mouse hippocampus (anti-NFIA clone R1356.1.2C6) and mouse embryonic (embryonic day 14) lens (anti-JAZF1 clone R913.1.1C2), respectively. Bottom, postnatal day 14 (P14) mouse pups of the *RaxCreER*^{T2}*;Lhx2*$^{lox/lox}$ background were injected with 0.2 mg of tamoxifen daily from P1 to P3 to induce conditional deletion of *Lhx2* in Müller glia (*Lhx2-KO*). *Lhx2*-KO pups and age-matched controls were processed for IHC with anti-LHX2 (clone R911.1.2G5). Each antibody was tested once, with 4–8 sections of the same source tissue analyzed in each experiment. Scale bars, 50 μm. (**f–h**) Representative examples of third-party validation of PCRP antibodies via (**f**) ELISA using anti-ANAPC2 mAb (clone R354.1.2F2). (**g**) SPR measurement for anti-HES1 mAb (clone R159.4.1B11), and (**h**) IP-MS using anti-SMARCE1 mAb (clone R330.2.2E9). Conc., concentration.

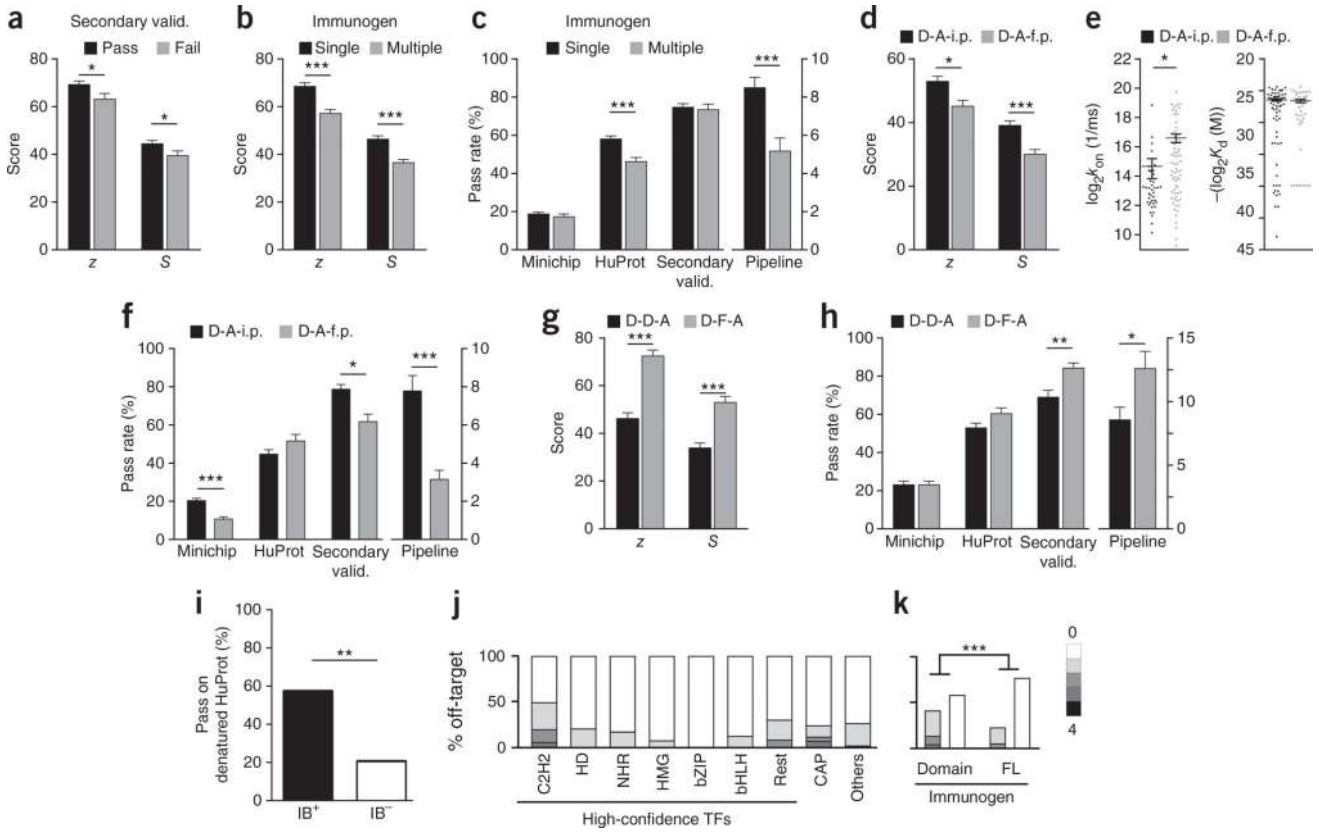

Figure 4.

Meta-analysis. (**a**) Comparison of *z*- and *S*-scores of mAbs that passed secondary validation (valid.) versus those that failed. $*P = 0.0162$, unpaired two-tailed *t*-test with Welch's correction. (**b**) Comparison of *z*- and *S*-scores for mAbs raised by immunization with a single antigen versus those for mAbs raised against multiple antigens. $***P = 6.5 \times 10^{-9}$ and 6.8×10^{-9} for *z*- and *S*-scores, respectively. (Statistical significance in **b–h** determined via the Benjamini-Hochberg FDR method.) (**c**) Comparison of success rates at different stages of the pipeline for mAbs generated by immunization with a single antigen versus for those raised against multiple antigens. $***P_{\text{HuProt}} = 2.45 \times 10^{-6}$; $P_{\text{pipeline}} = 7.18 \times 10^{-4}$. (**d**) Comparison of *z*- and *S*-scores of mAbs generated by i.p. and f.p. immunization. D, domain; A, all (see text for a detailed explanation of mAb group designations). $*P = 3.1 \times 10^{-3}$ $***P = 6.0 \times 10^{-5}$. (**e**) Comparison of k_{on} and K_{d} values for mAbs generated by i.p. and f.p. immunization. $*P = 6.7 \times 10^{-3}$. (**f**) Comparison of success rates at different stages of the pipeline for mAbs generated by f.p. and i.p. immunization. $*P = 7.11 \times 10^{-3}$ at secondary validation; $***P = 1.17 \times 10^{-6}$ and 1.32×10^{-5} at the minichip stage and in the overall pipeline, respectively. (**g**) Among mAbs raised against a single immunogen, comparison of *z*-and *S*-scores of mAbs that recognized their cognate full-length (F) target versus those of mAbs that recognized only their immunization domain. $***P = 1.83 \times 10^{-13}$ and 1.08×10^{-8} for *z*- and *S*-scores, respectively. (**h**) Among mAbs raised against a single-domain immunogen, comparison of success rates at different stages of the pipeline for mAbs that recognized their cognate full-length target versus those for mAbs that recognized only their immunization domain. $*P = 0.014$, $**P = 3.72 \times 10^{-4}$. In **c,f,h**, the *y*-axis label applies to

both vertical axes. (**i**) 27 of 47 IB^+ mAbs and 6 of 29 IB^- mAbs passed screening on denatured HuProt. $^{**}P = 0.002$, two-tailed Fisher's exact test. (**j**) The percentage of off-targets (rank 2–5 with *z*-score > 3) for each individual mAb that belong to the intended target's class/subclass. The fraction of mAbs for which all off-targets do not belong to the same class/subclass as the intended target is represented by the white portion of each bar. Light gray to black shading indicates fractions wherein one to all four off-targets belong to the same class/subclass as the intended target. (**k**) The distribution of off-target classes/subclasses based on the type of immunogen (domain or full-length) used to generate the mAb. $^{***}P = 2.8 \times 10^{-35}$, two-tailed Fisher's exact test. Error bars indicate s.d. except in a, where they indicate s.e.m. Details on means, error, and sample sizes for data presented in this figure can be found in Supplementary note 1.

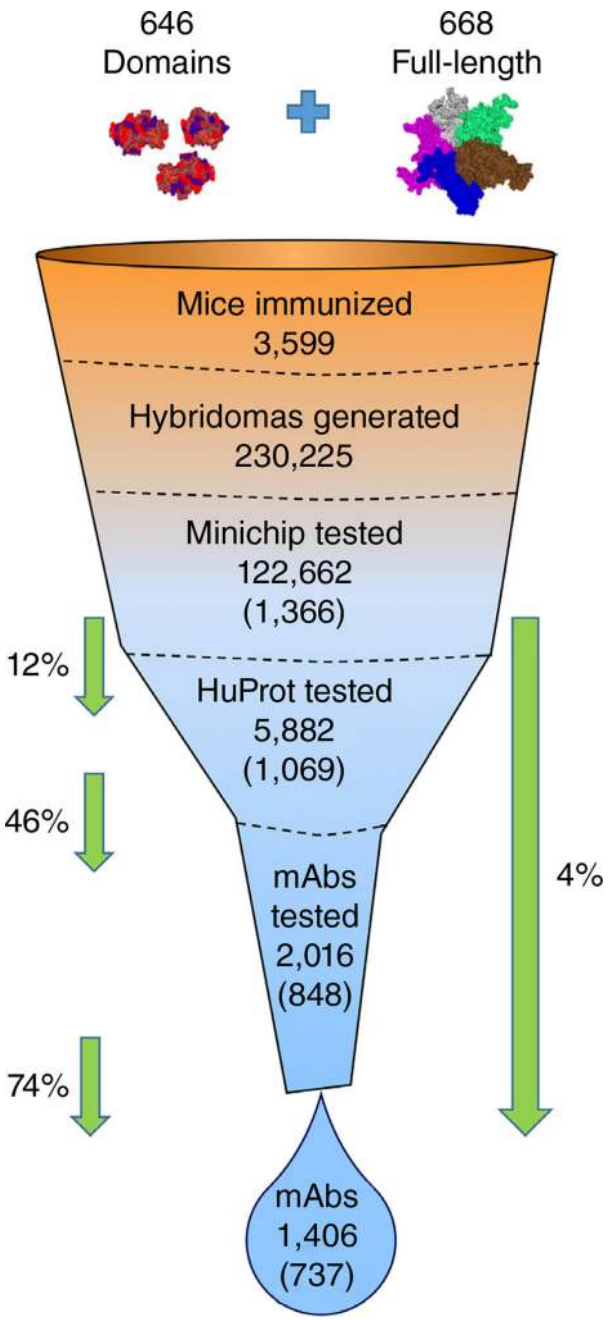

Figure 5.

The attrition funnel through which immunogens entering the PCRP pipeline ultimately generate high-quality mAbs. The number of mAbs and targets tested are indicated in parentheses for each relevant step. The success rates for the pipeline at the minichip, HuProt, and secondary-validation steps are indicated to the left of the funnel, and the overall pipeline success rate is indicated to the right of the funnel. Success rates were calculated at the level of target considering the fraction of corresponding mAbs that passed relative to the total number tested.