

# A High Through-put Platform for Recombinant Antibodies to Folded Proteins

Michael Hornsby<sup>#1</sup>, Marcin Paduch<sup>#2</sup>, Shane Miersch<sup>#3</sup>, Annika Sääf<sup>#2</sup>, Tet Matsuguchi<sup>1</sup>, Brian Lee<sup>1</sup>, Karolina Wypisniak<sup>1</sup>, Allison Doak<sup>1</sup>, Daniel King<sup>2</sup>, Svitlana Usatyuk<sup>2</sup>, Kimberly Perry<sup>2</sup>, Vince Lu<sup>2</sup>, William Thomas<sup>2</sup>, Judy Luke<sup>2</sup>, Jay Goodman<sup>2</sup>, Robert J. Hoey<sup>2</sup>, Darson Lai<sup>2</sup>, Carly Griffin<sup>3</sup>, Zhijian Li<sup>3</sup>, Franco J. Vizeacoumar<sup>5</sup>, Debbie Dong<sup>3</sup>, Elliot Campbell<sup>4</sup>, Stephen Anderson<sup>4</sup>, Nan Zhong<sup>6</sup>, Susanne Gräslund<sup>6</sup>, Shohei Koide<sup>2</sup>, Jason Moffat<sup>3</sup>, Sachdev Sidhu<sup>3\*</sup>, Anthony Kossiakoff<sup>2\*</sup>, and James Wells<sup>1\*</sup>

<sup>1</sup>Department of Pharmaceutical Chemistry University of California, San Francisco, CA;

<sup>2</sup>Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL;

<sup>3</sup>Donnelly Center for Cellular and Biomolecular Research, Department of Molecular Genetics, University of Toronto, Toronto, Canada; <sup>4</sup>Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ; <sup>5</sup>Saskatchewan Cancer Agency, University of Saskatchewan, Saskatoon, Canada; <sup>6</sup>Structural Genomics Consortium, Toronto, Canada

\*To whom correspondence should be addressed: James Wells, ([Jim.Wells@ucsf.edu](mailto:Jim.Wells@ucsf.edu)); Anthony Kossiakoff, ([koss@bsd.uchicago.edu](mailto:koss@bsd.uchicago.edu)); Sachdev Sidhu, ([sachdev.sidhu@utoronto.ca](mailto:sachdev.sidhu@utoronto.ca))

<sup>#</sup>These authors contributed equally to the work

Running title: Recombinant Antibody Platform for Folded Proteins

Keywords: recombinant antibodies; high through-put robotics, transcription factors; epigenetic factors; affinity reagents; chromatin remodeling

Abbreviations: Fab: Fragment antigen-binding domain, TEV: Tobacco Etch Virus Protease, TF: Transcription Factor, SEC: Size exclusion chromatography, CDR: complementary determining region, DSF: Differential Scanning Fluorimetry, Spiked-IP: Spiked Immunoprecipitation, pl: isoelectric point, YAD, yeast antigen display, IF: immunofluorescence, KD: knockdown, PrEST: Protein Epitope Signature Tags, HPA: Human Protein Atlas

**Summary:**

Antibodies are key reagents in biology and medicine, but commercial sources are rarely recombinant and thus do not provide a permanent and renewable resource. Here, we describe an industrialized platform to generate antigens and validated recombinant antibodies for 346 transcription factors (TFs) and 211 epigenetic antigens. We describe an optimized automated phage display and antigen expression pipeline that in aggregate produced about 3000 sequenced Fabs (Fragment antigen-binding domain) that had high affinity (typically EC<sub>50</sub><20nM), high stability (T<sub>m</sub>~80°C), good expression in *E. coli* (~5mg/L), and ability to bind antigen in complex cell lysates. We evaluated a subset of Fabs generated to homologous SCAN domains for binding specificities. These Fabs were mono-specific to their target SCAN antigen except in rare cases where they cross-reacted with a few highly related antigens. Remarkably, immunofluorescence experiments in six cell lines for 270 of the TF antigens, each having multiple antibodies, show that ~70% stain predominantly in the cytosol and ~20% stain in the nucleus which reinforces the dominant role that translocation plays in TF biology. These cloned antibody reagents are being made available to the academic community through our web site [recombinant-antibodies.org](http://recombinant-antibodies.org) to allow a more system-wide analysis of TF and chromatin biology. We believe these platforms, infrastructure, and automated approaches will facilitate the next generation of renewable antibody reagents to the human proteome in the coming decade.

**Introduction:**

Antibodies are crucial reagents for biological research and therapeutics. However, reproducibility for antibody reagents is a major concern, especially for polyclonals and even monoclonals where genetic drift of hybridoma stocks can be problematic[1, 2]. Moreover, some have estimated that less than half of the animal derived antibodies bind their cognate native proteins[3, 4]. The systematic generation of recombinant antibodies would provide a renewable collection of cloned and highly validated antibody genes and a permanent validation database[5, 6]. Recombinant antibodies also afford a biosynthetic tool kit for recombination and gene fusions to generate new sensors and functional modulators. Other efforts for renewable antibody reagents[7-9] have highlighted the need to design robotics and high-throughput platforms for antigen production, antibody selections, and characterization[10].

One area of need for renewable antibody reagents are proteins involved in chromatin biology including transcription factors (TFs) and epigenetic antigens. According to The Human Protein Atlas (HPA; [www.proteinatlas.org](http://www.proteinatlas.org)), there are commercially available antibodies to 362 of the estimated 1550 human TFs[11] and none are from recombinant sources ([www.antibodypedia.org](http://www.antibodypedia.org)). Thus, the absence of validated recombinant antibodies to profile specific TF interactions and their spatial distribution is needed. As a protein class, TFs have been particularly challenging because they contain multiple domains, often intrinsically disordered[12, 13], and thus difficult to express as full-length proteins. Hence the NIH Common Fund funded this

broad effort to generate renewable antibody reagents to this class of proteins (1U54HG006436). Such antibodies would be an important resource for biologists interested in understanding trafficking of TFs, their expression patterns in cells at the protein level, and ultimately their binding sites and partners during signaling.

Recombinant technologies for antigen production and antibody selections are poised for a large-scale effort to generate renewable antibodies to chromatin remodeling proteins. Recombinant antibody generation by phage display is not dependent on animal immunizations where control of the target protein is relinquished to the animal's immune system. Maintaining control of the protein status *in vitro* allows the user to customize selection conditions such as buffer, pH, temperature, and competitor proteins. *In vitro* methods eliminate antigen proteolysis, clearance, and auto-antigen anti-selection in an animal setting [14]. These bench-scale technologies are well-honed but we believe that by automating the *in vitro* selection technologies can fully realize their additional advantages of reducing the processing time from months to weeks, with much less antigen and reduced cost.

Here, we present an industrialized platform (**Figure 1A and B**) for generating high affinity renewable antibodies at large scale that is exemplified here for TFs and epigenetic factors, 557 total chromatin remodeling targets. Soluble domains of TFs were expressed in multiple high-throughput expression formats to ensure a soluble and intact antigen at a sufficient quantity for antibody selections (<1mg). We utilized a highly stable and diverse synthetic Fab scaffold, displayed on filamentous phage[15]. Robotic platforms were designed to rapidly select and characterize multiple Fabs of high affinity, specificity, stability, and expression in *E. coli* for each of 346 human TFs (representing

>18 protein domain folds) and 211 different epigenetic proteins. Remarkably, immunofluorescence with multiple antibodies per TF in six different cell lines showed that about two thirds of human transcription factors tested reside predominantly in the cytosol, but the precise distribution is cell-line dependent. These data highlight the importance of translocation in TF biology. Our studies describe the pitfalls and viable solutions for a high-throughput platform that we believe will greatly accelerate the process of producing renewable, high quality, and evolvable antibody reagents to folded proteins. These cloned antibodies are available to the academic community for research purposes through the recombinant-antibody.org.

### **Experimental Procedures:**

#### *Antigen sources:*

An NIH funded project was launched to rapidly develop and implement a high-throughput pipeline designed to generate high quality renewable recombinant antibodies to human transcription factors (U54 HG006436-01). TF antigens reported here were generated by Rutgers University, epigenetic antigens generated by the Structural Genomics Consortium (SCG), and in-house at the RAN facilities.

#### *Bacterial Strains:*

*E. coli* strain XL1-blue (Stratagene) or T1 phage resistant cells were used for phage propagation, whereas strains DH10B (Invitrogen), or other T1 phage resistant cells were used for Fab expression C43 (DE3) Pro+[16] or BL21 (DE3) were used. XL1-blue or T1 phage resistant cells were grown in the presence of 5ug/ml tetracycline (Tet) to ensure expression of the F' pilus that allowed for phage infection and DH10B cells were grown with 50ug/ml carbenicillin (Carb) for production of plasmid DNA. Fab

expression cells, C43 (DE3) Pro+, were grown with 50ug/ml Carb and 25ug/ml kanamycin (Kan). Where biotinylation was required (Antigen or Fab) recombinant BirA co-expressed or added during purification (Avidity, LLC.)

*Antigen Cloning and expression:*

Recombinant transcription factor domains were designed[17-19], genes synthesized (Life Technologies, Carlsbad CA), and were provided as *E. coli* expression constructs. In order to generate constructs for antigen expression by *in vitro* transcription and translation (IVTT), and *E. coli*, gene fragments were PCR amplified from transfer plasmids and recombined into their specific expression vectors by SLICE or Ligase Independent Cloning techniques[20, 21]. All expression plasmids were sequence verified prior to expression studies.

The antigen production group at Rutgers University or at the SGC Consortium provided purified human biotinylated human transcription factor domain antigens as N-terminal tagged AviTag fusions. For Rutgers antigens, expression constructs were synthesized (GenScript USA Inc., Piscataway, NJ), cloned into an expression vector containing an N-terminal AviTag, HisTag, and TEV cleavage site, and expressed in *E. coli* Tuner (DE3) harboring a pLysSRARE2 helper plasmid (EMD Millipore, Billerica, MA). Cells were harvested, lysed by sonication, and clarified lysate was loaded onto HisTrap HP (GE Healthcare, Piscataway, NJ) columns for purification. If the protein of interest was predominantly found in inclusion bodies, pellets were lysed in solubilization buffer (binding buffer containing 6M urea), bound to the HisTrap column, and subjected to a slow refolding gradient into binding buffer without urea prior to elution. Pooled

elution fractions were *in vitro* biotinylated (Avidity LLC, Aurora, CO), concentrated, and applied to a HiLoad 16/600 Superdex 75pg (GE Healthcare, Piscataway, NJ) column to remove unreacted biotin and ATP. Fractions corresponding to the monomeric peak were pooled and concentrated before being snap frozen in liquid N<sub>2</sub> and stored at -80°C until shipment (Campbell, E. Anderson, S. in preparation). SGC produced antigens were expressed in *E. coli*, purified by Immobilized Metal Affinity Chromatography, and tested for biotinylation prior to entry into the Fab selection pipeline [22, 23]

Protein expression system comparisons consisted of base-line TF expression in *E. coli* as intracellular GST fusion constructs (His-GST-Avi-TEV-Antigen) driven from the pTac promoter (**Figure SI 1C**). *E. coli* based expression of soluble GST-TF fusion antigens was conducted by standard protein expression methods. Briefly *E. coli* cultures were grown in Terrific Broth with antibiotics to an OD<sub>600</sub> of 0.6-0.8 with 1mM IPTG for 3-6 hours at 37°C, then harvested by centrifugation. Cells were resuspended, lysed by sonication, and clarified prior to purification with GST Sepharose [24]. As a second system we expressed the TFs as His-Avi-TEV N-terminal fusion driven from a T7 promoter using an *E. coli* *in vitro* transcription-translation system (IVTT) that was kindly provided by Sutro Biosciences[25] (**Figure SI 1D**). The cell-free reaction mixture contained an ATP regeneration system supplemented with chaperones and redox enzymes. Briefly, 10µg of IVTT plasmid constructs were combined with 1ml cell free extract containing recombinant BirA protein and 50µM biotin and allowed to shake at 650rpm for 2 hour in an Eppendorf Thermomixer Expressed and biotinylated protein was purified over NiNTA agarose (Qiagen) with PBS based buffer system and eluted with PBS + 400mM Imidazole. Protein samples were buffer exchanged with Amicon

spin concentrators (EMD Millipore), concentrations determined by Bradford BSA assay (Thermo Fisher), and analyzed for purity by SDS-PAGE prior to entering the selection pipeline[25, 26]. In the third system, we displayed the antigen for selection on the surface of yeast[27, 28]. To generate constructs for YAD, gene fragments were PCR amplified from gene synthesis transfer plasmids and then recombined into YAD plasmid. YAD constructs utilized a N-terminal Avi-6xHis-Aga2-TEV protease fusion partner with a C-terminal V5-His tag. The YAD strain, EBY100, was engineered to overexpress agglutinin AGA1 under the galactose inducible Gal1-10 promoter, which formed a disulfide bonds with the antigen-Aga2 protein in the extracellular space. Constructs were biotinylated during secretion by a cytoplasmic or ER localized BirA, which allowed for efficient capture to streptavidin magnetic beads (Promega) [29-31]. The TF domain was expressed with a C-terminal V5 epitope tag to monitor expression level (**Figure SI 1E and SI 4**). The proteins generated from these three systems were then tested for the ability to generate antibodies by phage display.

#### *Fab-phage Selections:*

All phage selections were done according to previously established protocols[24, 26] with several modifications as outlined below. Briefly, Fab-phage selections were automated allowing for multi-parallel processing of biotinylated protein antigens either generated in-house or obtained from collaborators. Up to four rounds of soluble phage selections were conducted with biotinylated protein antigens bound to streptavidin magnetic beads (Promega). The antigen concentration on the bead was systematically decreased with successive selection round as follows: 100nM Round 1, 50nM Round 2,

10nM Rounds 3 and 4. To maximize throughput and reduce the amount of phage library used, up to 8 antigen-streptavidin beads complexes were combined into one selection well for round 1. Antigen-streptavidin beads were incubated with  $1 \times 10^{13}$  Fab-phage particles from either Library E or F for one hour with gentle mixing on a King Fisher Flex magnetic bead separator (Thermo Scientific) followed by 3 washes in Phosphate Buffered Saline supplemented with 0.05% Tween-20 and 0.2% Bovine Serum Albumin (PBST+BSA). To reduce the deleterious affects of non-specific binding phage, we employed a “catch and release” strategy, where specific antigen binding Fab-phage were selectively eluted from the magnetic beads by the addition of 2ug/well TEV Protease for 10 minutes. Once liberated from the streptavidin beads, Fab-phage were introduced to 5ml (Selection Round 1) or 100ul (Selection Rounds 2-4) of exponentially growing *E. coli* XL1-Blue or T1 phage resistant cells and propagated overnight at 37°C with shaking.

Once propagated, Fab-phage were recovered from culture medium with Protein A magnetic beads (EMD Millipore) on the King Fisher Flex. Briefly, 20-50 $\mu$ l of a Protein A magnetic bead slurry was incubated with up to 1ml culture supernatant for 60 minutes then beads were collected and washed prior to elution of Fab-phage with 100 $\mu$ M Acetic acid for 10 minutes. Tris-base pH11 buffer was added to neutralize the purified Fab-phage prior to further processing. The initial Fab-phage library and amplified eluents can contain a small subpopulation of streptavidin binding Fab-phage, therefore prior to each selection round Fab-phage pools were incubated with 50ul streptavidin-beads for 10 minutes then the beads were removed with the King Fisher Flex to deplete the library

of any non-specific binding Fab-phage. Plate based selections were done by immobilization of antigen on protein binding plates as described previously [24].

*Competition ELISA:*

Pools of phage from selection round 3 or 4 were purified and diluted in PBS and allowed to infect 100ul *E. coli* cells ( $OD_{600}$  0.6-0.8) for 20 minutes then plated on a LB +50ug/ml Carb Omnitray (Nunc) so least 200 individual infected colonies would grow across the plate. Phage for individual antigens were plated per plate and 96 individual colonies were picked with a K6-2 Colony Picker (KBiosystems) or manually into 96-deep well blocks containing 0.5ml 2xYT +50ug/ml Carb + $1\times10^9$  KO7 helper phage/ml (New England Biolabs). Culture plates were grown 16-18 hours at 37°C shaking at 600-900 RPM prior to processing the competition ELISA.

Two 96-well immunoassay plates (Corning) coated with 50ng/well neutravidin (Thermo Fisher) were used for each plate of picked colonies as follows to generate Competition ELISA data: Direct Binding: Biotinylated protein antigen was diluted to 20nM in PBST + BSA, 50ul dispensed per well, incubated for 20 minutes, and Fab-phage containing culture supernatant was added for 15 minutes. Unbound Fab-phage was washed away by 3 additions of PBST using a BioTek EL406 or similar plate washer followed by the addition of 50 $\mu$ l anti-M13 phage-HRP antibody (diluted 1:5000 in PBST + BSA) and 30 minute incubation with shaking. Incubated plates were again washed 3 times with PBST or PBS then developed with 50ul TMB substrate (KPL Labs) and analyzed with a Tecan M1000 or similar plate reader monitoring  $OD_{650}$  for up to 5 minutes. Competition ELISA: Biotinylated protein antigen was diluted to 20nM in

PBST+BSA, 50ul dispensed per well, and incubated 20 minutes. In a separate microtiter plate, 20nM antigen was mixed with the phage containing culture supernatant and incubated for 15 minutes with shaking. After incubation, the phage-antigen complex was added to the washed immunoassay plate and incubated for 15 minutes before washing final washing and development as described above for Direct Binding Plate.

By plotting the OD<sub>650</sub>/min (or OD<sub>450</sub>) data from the Direct Binding plate on the Y-axis then calculating a ratio of Direct Binding/Competition and plotting on the X-axis, Fab-phage with at least 20nM affinity were separated from Fab-phage with poor binding affinities. Desired Fab-phage have a competition ratio of <0.5 with direct binding signal >0.005 units and these phage were subjected to DNA sequence analysis to determine unique Fab sequences.

Fab-phage passing the competition ELISA assay were prepared for sequencing by PCR amplification of the Fab gene directly from phage supernatant. PCR reactions in either 96- or 384-well format were conducted in a Roche LC480 or similar thermocycler. PCR products were analyzed by gel electrophoresis or by Sybr Green melting curve analysis, treated with 0.025 units of exonuclease (Affymetrix) and 0.25 units of shrimp alkaline phosphatase (Affymetrix), prior to sequencing.

Analysis of the sequencing results was automated by the generation of several scripts where the sequences flanking CDR's were recognized and both DNA and amino acid translations were returned for the 4 CDR's of interest (LC3, HC1, HC2, and HC3). Duplicate Fab-phage sequences were removed and only unique Fab-phage clones passed into the cloning pipeline.

*Fab Expression and Purification:*

C43 (DE3) Pro+ *E. coli* containing expression plasmids were grown in Terrific Broth supplemented with 0.5% glycerol, Carb, Kan, Chlor, and 5uM Biotin to an OD<sub>600</sub> of 0.6-0.8 at 37°C then Fab expression was induced by the addition of 1mM IPTG. At the time of induction incubation temperature was reduced to 30°C and allowed to shake for 16-18 hours at which time cells were harvested by centrifugation. Recombinant Fabs were purified by Protein A chromatography and buffer exchanged into PBS buffer for subsequent storage and validation assays.

*Differential Scanning Fluorimetry:*

To assess the stability of the recombinant Fabs, we employed Differential Scanning Fluorimetry (DSF) in which Sypro Orange (Invitrogen) binds to hydrophobic regions of partially or fully denatured proteins giving a high fluorescent signal[32]. DSF was conducted either on Roche LC480 Lightcycler or similar qRT-PCR instrument in either 96- or 384-well formats. Briefly, purified recombinant Fab was diluted to 2uM in DSF buffer containing Sypro Orange 4x and PBS then subjected to a temperature gradient (0.5°C/30 seconds) from 50 to 95°C. Data was acquired as often as possible depending on instrument at an ~490nm and ~575nm (excitation and emission wavelengths) then processed to generate 1<sup>st</sup> derivative curves where the curve peak corresponds to the melting temperature of Fab.

*Spiked Immunoprecipitation:*

To assess Fab specificity, we utilized an immunoprecipitation method where target Fab was bound to Streptavidin magnetic beads (Dynabeads M-280, LifeTechnologies) then incubated with 50 nM biotinylated antigen for 30 minutes in the presence or absence of a cellular extract corresponding to OD<sub>280</sub> 2.5 and OD<sub>260</sub> 3.68 (HEK293 cell extract) followed by washing in buffer, incubation with an Streptavidin-Fluorophore, staining with a fluorescently labeled and analysis of binding profiles by Flow cytometry. Details of this assay will be presented elsewhere (Koide et. al. in preparation).

#### *ELISA Affinity Measurement:*

To assess the affinity or specificity of Fab's we utilized EC<sub>50</sub> or single point ELISA assays conducted in 96- or 384-well ELISA plates pre-coated with neutravidin. Briefly, 20nM biotinylated antigen was bound to ELISA plate and incubated for 20 minutes prior to washing with PBST. Purified Fab protein was either diluted in a 2- or 3-fold dilution series (EC<sub>50</sub>) or at 20nM (specificity assay) then transferred to prepared plates for 30 minutes prior to washing and development with anti-Flag-HRP antibody (1:5000 dilution; Sigma Aldrich) and TMB substrate.

#### *Immunofluorescence:*

96-well cell carrier microtiter plates (Perkin Elmer) were seeded with 5X10<sup>3</sup> A2780Cis, 3345A, A2780, HPAFII, HEPC2 or 293T cells in individual wells after which they were allowed to adhere by incubating overnight at 37°C. After incubation, media was removed and cells washed 1X with ice-cold PBS pH 7.4, then fixed by incubating

with ice-cold methanol at -20°C. Fixed cells were then permeabilized with 0.25% Triton X-100 in PBS, washed 3X with 200 mL PBS pH 7.4 and blocked with 1% goat serum in PBS for 1h at RT, before the addition of a 10mg/mL solution of primary Fab diluted in 1% goat serum in PBS. Cells were incubated with primary Fab for one hour at RT then washed 3x with PBS prior to the addition of and 30 min incubation with a 1:2000 dilution of FITC-labeled anti- human Fab secondary antibody protected from light.

Immuno-stained images were collected for Fab clones against 6 cell lines with an *Opera* High Content Screening System (Perkin Elmer) at identical incident light intensity and photomultiplier sensitivity and gain. Analysis was performed by visual inspection to determine localization and recorded using a limited number of commonly used GO terms to facilitate analysis and prioritization of downstream testing. Scanning multiple fields of view to determine nuclear or cytoplasmic staining localization was done in order to generate a nuclear/cytoplasmic ratio that could be visualized as a heat map.

#### *siRNA Knockdown:*

The day before transfection, A2780cis cells (Sigma-Aldrich) were seeded at ~5-7  $\times 10^4$  cells per well of a 96-well plate (Perkin Elmer) in 100  $\mu$ l of RPMI with 10% FBS without antibiotics. For each transfection well, *Silencer Select* Pre-designed siRNA (Ambion Life Technologies) was diluted into 25  $\mu$ l Opti-MEM reduced serum medium for final concentration of 10 nM per well. In a separate tube, 0.5  $\mu$ l of Lipofectamine RNAiMAX Transfection Reagent was added per well and diluted into 25  $\mu$ l Opti-MEM reduced serum medium and mix. After a 5-minute incubation the tubes were combined then incubated at room temperature for 30 minutes to allow siRNA-liposome complexes

to form. Once complexed, the mix was added to cells (growth media removed), and then diluted with growth media to 100ul final volume. Cells are then incubated with siRNA for 18 hours at 37°C in a CO<sub>2</sub> incubator, at which time the mixture was removed and replaced with complete growth medium. Cells were screened for Fab binding after 48 hours of siRNA addition.

## Results:

### *Antigen expression*

Production of soluble intact protein is important for successful selection of antibodies to native proteins (7, 9). Unstable or mis-folded protein antigens typically result in poor binding reagents from phage display. For example, in small test sets we observed that proteins that are significantly proteolyzed or expressed in low levels typically did not produce antibodies of high quality whereas intact and soluble proteins do generate high quality antibodies (**Figure SI 2**). Therefore, to avoid pitfalls with potentially disordered[12] multi-domain[13] TF antigens, we focused on those regions of TFs (10 to 25kDa) that had clear domain architecture based on structural and bioinformatic considerations[17-19] and thus a good chance of expressing a soluble protein (**Table SI 1**). Virtually all TF antigen domains were expressed in *E. coli* as fusion proteins to facilitate purification (His tag), antigen capture via a single biotinylated site (Avi tag), and tobacco etch viral (TEV) mediated protease cleavage to elute bound phage (**Figure SI 1A-C** [9, 22, 23, 26]). Purity and integrity of the TF to proteolysis was assessed by SDS-PAGE prior to entry into the antibody selection pipeline. Size exclusion chromatography (SEC) and thermostability measurements were occasionally

conducted as a check of antigen quality; however these metrics were not usually integrated into the antigen expression pipeline due to the large number of antigens and the relatively low throughput nature of standard SEC.

### ***Fab-phage library and robotic selections***

We chose a Fab phage display format where the heavy chain constant domain-1 and variable domain is fused to the gene III and the light chain is co-expressed using a phagemid system[33, 34]. The Fabs assemble in the periplasm of *E. coli* through disulfide bond formation as a fusion protein with gene III that are subsequently packaged and displayed in bivalent format on the phage coat[35]. The Fab scaffold is derived from the 4D5 anti-Her2 antibody, a very stable IgG1 ( $T_m$  80° C) for which several approved antibody drugs are based[36-38]. The Fab library is a synthetic and codon restricted library that was designed to focus on the most important amino acids in four of the six complementary determining (CDR) loops[15]. The CDRs were mutated in a combinatorial fashion based upon structural bioinformatic analysis of antibody-antigen structures in the PDB as previously described[15, 39, 40]. This diverse library contains approximately  $3 \times 10^{10}$  members and has been used successfully in manual selections for scores of protein antigens with varying folds[41], altered conformations[26, 42, 43], and nucleic acids[44].

To industrialize the process we designed a pipeline approach augmented by custom assembled robotics to process antigens (**Figure 1A and B**). This pipeline can accommodate various antigen formats through phage selection and primary validation of Fab affinity, sequence, and *E. coli* Fab expression. We typically used a “catch and

“release” proteolysis process for both soluble antigens and yeast cell-displayed antigen. To begin, biotinylated antigens are preferentially bound to streptavidin coated magnetic beads and excess antigen is washed away using a King Fisher Flex magnetic bead separator (**Figure 1B**)[26, 39]. The bead-based immobilization methods using streptavidin capture preserves the native fold of the antigen and reduces the possibility of localized antigen unfolding that sometimes occurs when antigens are adsorbed directly to plastic immunoassay plates[45]. In some cases direct binding of antigen onto microtiter plates was conducted for selections, but this precludes the use of the King Fisher Flex device.

The Fab-phage library was mixed with the antigen-loaded magnetic beads and unbound phage was washed away (**Figure 1C**, Steps 1- 3). Beads were treated with TEV protease and eluted phage was recovered and propagated in *E. coli* for subsequent selection rounds (**Figure 1C**, Steps 4-7). Unlike standard elution by low pH glycine elution from bare plastic plates[46], the proteolytic catch and release format ensures clean and efficient recovery of antigen bound Fab-phage regardless of antigen affinity. The catch and release approach, which is highly amenable to automation, reduces the recovery of non-specific phage bound to the streptavidin bead or otherwise absorbed or aggregated onto the surface.

Several other improvements were made to the standard Fab-phage protocols. We found it useful to enrich phage particles that display functional Fab-phage with Protein A magnetic beads because only about 10% of phagemid have a Fab displayed. In addition, we exposed the phage to streptavidin magnetic beads to reduce the number of non-specific or avidin binding Fab-phage. Lastly, the stringency of the selection was

increased each round by systematically lowering the antigen concentration from 100nM to 10nM. From experience with hundreds of antigen selections, we found that approximately 1nmol of antigen is needed for the selection process (~25ug for a 25 kDa antigen). This is considerably below what is typically used for animal immunizations. We monitored enrichment by positive binding Fab-phage versus the starting library and found that after 3 to 4 rounds of selection the Fab-phage pool was optimally enriched for antigen binding clones.

### ***Fab-Phage Pool Deconvolution and primary validation***

Usually 50-100 individual Fab-phage clones from the final selection pools were plated and grown for validation. These were tested in a high-throughput format (96- or 384-plate) for direct and competitive binding using a robotic ELISA assay (**Figure SI 3**). We measured the OD<sub>650</sub>/min for direct binding of the Fab-phage to biotinylated antigen bound to a neutravidin or streptavidin coated ELISA plate (or an antigen coated ELISA plate) that was then probed with an anti-M13 phage antibody (**Figure 1D**, Y-axis). This high throughput assay provided a composite estimate of functional Fab-phage affinity and expression. We estimated the relative affinity by a single point Fab-phage competition binding ELISA. Here the Fab-phage was mixed with 20nM soluble antigen and then allowed to bind to the same antigen plate (**Figure 1D**, X-axis) where OD<sub>650</sub>/min was monitored and used to calculate a competition ratio. Competition ratio is defined as the OD<sub>650</sub>/min of competitive binding divided by the OD<sub>650</sub>/min of direct binding signal. Fab-phage clones were considered to have passed when they were shown to have expressed at OD<sub>650</sub> >0.01 (or OD<sub>450</sub> >0.1) and were competed by soluble

antigen ratio <0.5 (**Figure 1D**, upper left green quadrant). Representative graphs from ELISA validation show typical selection results for passing antigen (**Figure 1D**; top right and lower left) and an unsuccessful antigen (**Figure 1D**; bottom right). This simple ELISA triage step for 96 Fab-Phage generally requires less than 1nmol of antigen. The most promising clones (typically ~48) were picked and prepared for DNA sequencing. Fab sequences are aligned and unique representative clones chosen for subsequent isolation, expression, and purification.

For expression of soluble monovalent Fab fragments, the Fab-phage clones were PCR amplified and subcloned into standardized *E. coli* expression plasmids, either Avi- or non-Avi-tagged (recombinant-antibodies.org/protocols), by SLICE or Ligase independent cloning methods[20, 21]. Sequence verified plasmids were then transformed into C43 Pro+ *E. coli*[16] for expression studies. Expressed Fabs were purified by semi-automated Protein A chromatography that typically recovered 100-500ug purified Fab proteins. Not all Fab frameworks bind to Protein A, but the Herceptin (4D5) framework chosen for our Fab-phage library binds efficiently to Protein A[47].

To better understand the global biophysical properties of the Fabs generated from this pipeline we analyzed a subset of primary validated Fabs for expression level, stability, and affinity. Expression levels in *E. coli* of over 700 monomeric Fabs showed that more than 90% express in the range of 1 to 10 mg/L, as assessed by Bradford assay or A<sub>280</sub> measurements (**Figure 2A**). Given that commercial antibody reagents are generally provided in 100μg aliquots, simple 50-100ml-shake flasks culture provide access to purified Fabs for many applications. To test the stability of the Fab antibodies we measured the thermostability of 96 randomly chosen Fabs by Differential Scanning

Fluorimetry (DSF) facilitated by robotic cycling PCR instruments[48]. The  $T_m$  values of the Fabs varied over a narrow range, 75-85°C (**Figure 2B**). Detailed affinity measurements for ~200 Fabs showed EC<sub>50</sub> values that range from high pM to 50nM with the average around 10nM (**Figure 2C**). The high expression level *E coli*, the robust Fab stability, and high affinity encouraged us to scale up the process. To date, 537 TF antigens have been processed through the robotic Fab-phage selection pipeline and 435 antigens (81%) have produced about 3000 primary validated Fab antibodies that were moved into a secondary validation pipeline.

### **Secondary validation and protein properties that correlate with success**

As a next level of antibody validation, we sought to determine how the Fabs would bind to their parent TF in a cellular lysate. Kelley and co-workers have shown that human antibodies that are rapidly cleared in serum exhibit high levels of non-specific binding in assays containing cellular extract[49]. Non-specific binding was estimated by the reduction in the direct binding (ELISA EC<sub>50</sub>) of antibody to antigen in the presence of non-cognate protein from baculovirus extract. We developed a similar assay termed Spiked Immunoprecipitation (Spiked-IP) designed to rapidly screen all Fabs to identify any sticky or non-specific binding Fabs[50]. This allowed us to assess the ability of the Fab to bind and pull-down exogenously added TF antigen in the presence of complex human cell lysates compared to buffer alone. Briefly, biotinylated Fab is bound to streptavidin magnetic beads and mixed with antigen for binding in the presence or absence of a HEK cell lysate as monitored by flow cytometry (**Figure 3A**). The assay

was also run in reverse mode where biotinylated antigens were bound to beads and allowed to bind Fab in the presence of the HEK cell lysate or in buffer alone.

In order to pass the Spiked-IP assay, lysate to buffer binding signals had to be within two-fold of each other. Those Fabs having lysate signals more than two-fold lower than the buffer condition were considered to have too much non-specific or sticky binding to other proteins in the lysate as shown in an example for 3 antigens with 3 to 4 Fabs each (**Figure 3B**). While this assay does not directly address specificity of the Fab to its target antigen, we believe this assay robustly screens for Fabs with a high probability of being specific. About 56% of the antibodies that passed the primary validation steps (affinity, stability, and expression) passed the Spiked-IP secondary validation test (**Table 1**). The endogenous levels of TFs are typically more than 100 fold lower than we estimate on the bead so it is unlikely we would discard passing antibodies from specific endogenous TF competition. Nonetheless, false negatives are possible, and we did not evaluate them further.

The ability to successfully generate at least one high affinity Fab that passed the primary and secondary validation depended on TF domain type (**Table 1**). For example, antibody selections for SCAN domains succeeded 88% of the time, while Zinc finger domains were only successful 26% of the time (**Figure 3C**). It is possible these differences reflect the stability of the domains or that antigens form complexes with a molecule in the lysate (e.g. DNA) in such a way that prevents Fab binding.

We also tested a subset of Fabs (n=18) that were generated to SCAN domain-containing antigens in a single point pairwise ELISA analysis to address specificity of the Fabs to the protein of interest. In almost all cases the Fabs generated here mono-

specific (**Figure 4**, left panel). In the few cases where the Fab was poly-specific, for example RAB-S181, S145, and S169, the target antigens were >93% identical and only differ in amino acid composition by up to 5 residues (**Figure 4**, right panel). This degree of selectivity was remarkable given we did not impose a counter selection against the closely related antigen. We also tested a Fab that failed the Spiked-IP assay as a negative control and this indeed showed a higher degree of poly-specificity (**Figure 4**, left panel). This suggests that the Spiked-IP assay is useful to differentiate Fabs that are mono- vs poly-specific unless the target antigens are closely related to one another.

We also noted that the number of unique Fabs obtained per antigen increased as antigen size increased over 14kDa (**Figure 5A**). This is not surprising, as the surface area for binding increases. We also find that the number of Fabs identified per antigen systematically decreases as the isoelectric point (pl) of the antigen increased from pl 4 to 10 (**Figure 5B**). This could reflect the anionic nature of the phage coat to bind cationic proteins [51]. These general parameters help to define antigen clones with high likelihood of success.

Overall, 56% of all TF antigens screened generated Fabs that passed primary and secondary validation criteria. We believed some of this failure could be from inefficient protein folding in *E. coli*. Therefore, we sought to develop additional high through-put antigen expression systems to supplement *E. coli* expression system. To systematically compare expression methods and selections from them, we studied the ability to produce Fabs from 32 different TF antigens expressed in three different formats: *E. coli*, in vitro transcription translation, and yeast antigen display.

We put these 32 antigens through the automated Fab selection pipeline then scored resulting Fabs where we utilized the primary validation criteria. Indeed there was considerable variation in the success of obtaining primary validated Fabs for the 32 TF antigens across the different expression formats (**Figure 5C and Table SI 2**). While the three formats in aggregate produced at least one primary validated Fab-phage per antigen, IVTT antigens were somewhat more successful, 66%, with *E. coli* constructs and Yeast antigen display (YAD) averaging 41% and 47% success, respectively. Some antigens produced many passing antibodies (>45 Fabs per 96 tested) and others produced only a few. We conclude that having multiple antigen formats increases the likelihood of successful selection by providing back-up systems.

#### ***Immunofluorescence validation of Fabs and location of antigens in the cell***

Transcription factors can reside in almost any compartment in the cell from the cell membrane, to cytosol to nucleus[52]. Having an unprecedented collection of renewable antibodies to TFs allows us to begin to probe the cellular distribution of these TFs by immunofluorescence (IF). We chose six human cell lines that are known to express detectable levels of mRNA based on RNA sequence data (not shown) for each of 270 TFs which we had multiple passing Fabs. We used a high-throughput IF assay to screen 1017 Fabs in permeabilized cells (**Table SI 3**). Columbus software associated with an Opera imager (Perkin Elmer) was utilized to quantify the fluorescence intensities across multiple fields for each antibody to generate a heat map showing a ratio of nuclear intensity to cytoplasmic intensity (**Figure 6A**). Staining patterns were also scored by visual inspection using four standard GO-based descriptors typically applied

to categorize cellular localization: cytoplasmic, nuclear, mixed cytoplasmic/nuclear, and miscellaneous. Interestingly, TF proteins appear to have different localization patterns depending on cell type and growth conditions. These results suggest that the Fabs produced here were specific for their target TF and could be used to monitor TF trafficking. Examples of each of these categories are shown in **Figure 6B**. When analyzing the binding patterns of all Fabs tested by IF, we found that on average ~70% of the Fabs tested localized predominantly to the cytosol and ~20% to the nucleus (**Figure 6C**). There was remarkable consistency between these two scoring methods and across all cell lines in terms of the proportion in the nucleus vs. cytosol (20% vs 70%). Moreover, some TFs were found in different locations depending on the cell type likely reflecting their activation state differences. We also compared our IF data with those in common from the Protein Atlas, which generated affinity purified polyclonal antibodies against segments of TFs. Remarkably, we found that of the 55 antigens where an overlapping antibody has been developed, 50 (90.1%) show identical localization patterns (**Table SI 5**).

An example siRNA knockdown (KD) validation was done using the BATF Fab/antigen combination and showed significant reduction in staining intensity when cells were transfected with antigen specific siRNA and compared to the anti-flag secondary antibody only control, that showed no non-specific signal due to secondary antibody binding (**Figure SI 5**). Quantitation was done by ImageJ analysis and showed a significant decrease in staining with BATF ( $p<0.0001$ ) antibodies in siRNA-transfected cell in contrast to control siRNA-transfected cells. Although outside the scope and resources of this project, further studies such as identifying external stimulus required to

alter TF trafficking from the cytosol to the nucleus. Fabs shown in Table SI 4 are currently available to the scientific community for further study.

### **Fabs to epigenetic proteins**

In a parallel project, we collaborated with the Structural Genomics Consortium (SGC; <http://www.thescg.org/>) to generate Fabs to an extensive group of chromatin remodeling proteins (**Table SI 1 and 4**). Similar to the TFs, these chromatin-remodeling proteins are large multi-domain molecules that are difficult to express and are prone to aggregation in their full-length forms. Thus, as was the case for the TFs, the design of the antigens was based on identifying domain types that could be expressed as independent and stable entities. In many cases we had direct structural information to guide the domain expression experiments. From the phage display selections against 211 antigens we obtained 334 Fabs that passed primary affinity, and secondary Spiked-IP or IP-MS validation. Marcon and co-workers[53], developed an IP-MS method and tested 1154 Fabs to 154 of the epigenetic antigens for binding their endogenous target in HEK293 cells. They found that 452 of these antibodies pulled down the endogenous target for 98 antigens where the antigen (and in some cases known binding partners) were among the top 3 proteins identified. The other 54 targets were presumed not expressed at high enough endogenous levels to detect. This success rate from primary to secondary validation was roughly the same as the Spiked-IP method. While the IP-MS assay is more labor intensive than Spiked-IP, it provides direct mass data for antigen capture in the presence of complex cellular lysate like the Spiked-IP assay. The overall success rate for antigens producing primary and secondary passing Fabs was

68% (123 out of 200 tested). The generally higher success rate for the epigenetic targets was probably due to the fact that the expressed domains benefited from knowledge of the actual structures of many of the domains, which had been determined previously in the SGC.

### **Discussion:**

There is a recognized need for creating tool antibodies to the entire human proteome[54, 55]. The first systematic effort to generate proteome-wide antibody reagents began with the Human Protein Atlas (HPA) over a decade ago focused on generating polyclonal antibodies in rabbits[3, 56]. To increase antigenicity, antigens were produced to 100-residue segments, called Protein Epitope Signature Tags (PrEST). PrESTs were chosen from the human target predominantly to be highly non-homologous to the rabbit homolog and without consideration of domain structure or fold. This large-scale effort generated affinity purified rabbit polyclonal antibodies that reportedly react with PrESTs to over 80% of expressed human genes. Validation data is publically available through the HPA web site ([proteinatlas.org](http://proteinatlas.org)). Many of these antibodies are now available through commercial vendors, but not yet for open source distribution given the expense of production. In comparing our IF data to those in the Protein Atlas we found 55 targets in common for staining the same cells. Remarkably, 50 of the 55 targets had the same staining patterns despite the one being a polyclonal and the other a recombinant monoclonal as well them having been raised to different regions of the common transcription factor. Polyclonal antibodies can have advantages over monoclonal antibodies because they can bind many epitopes thus ensuring greater

probability of reacting with proteins in various formats. However the antibody epitopes and binding mechanisms are challenging to assess, as they are mixtures. Animal derived polyclonal antibodies suffer from the fact that they are not renewable and thus expensive validation data and production yields are lot dependent. Moreover, the human antigen can lead to self-selection against conserved proteins or epitopes an animal host and thus can preclude identification of functionally relevant epitopes[57].

Recombinant antibodies offer a renewable and stable source in the cloned monoclonal gene and hence there has been a call from the community to generate them systematically[5, 58]. The fact that selections can be run *in vitro* avoids problems with self-selection in the animal and can allow for isolation of very high-resolution antibody reagents ensured to bind the native protein and even specific conformations or post-translationally modified forms[16, 59-61]. Moreover, cloned genes can be further manipulated as genetic tools and plasmids are easily made publically available to researchers at low cost. Our intent here was to develop an industrialized platform empowered by robotics for the high-throughput production of renewable recombinant antibodies using phage display and high throughput antigen production. We demonstrate this in generating multiple high affinity, well expressed, and cell reactive antibodies to 346 transcription factors and 211 epigenetic proteins.

Production of high quality antigens is a key step in the process because denatured or unstable proteins typically do not reliably yield good monoclonal antibodies to native proteins[7]. We focused on expression of what we considered to be stable domains (10-20kDa) based on bioinformatics analysis. It is well-known in the X-ray crystallography community that one often needs to try different constructs in different

hosts to generate high purity and stable protein for crystallography[62]. Thus, we investigated three high throughput protein expression platforms to supplement the antigen pipeline. Indeed, there was considerable variability in success in generation of validated antibodies from the same antigen produced in these formats. We suspect the differences in selection success are due to antigen stability and solubility from the different formats as has been noted by McCafferty and coworkers (7). These different expression formats provide alternative pathways for generating antibodies to recalcitrant antigens.

The success rate for generating antibodies varied from 20-100% depending on the fold-type of the TF. These differences may reflect the relative stabilities of these TF domains. Indeed, it was noted that zinc finger domains were particularly problematic to express and purify and they were typically the least successful in generating validated Fabs. There is also a systematic correlation to molecular weight and charge. For example, small domains generated fewer passing Fabs. This may reflect the smaller surface area and hence fewer epitopes, or that they are intrinsically less stable. Also, highly basic proteins with pls above 8 to 9 are generally less successful than those with pl values between 4 and 7. This is possibly due to non-specific interactions with the anionic phage coat precluding selection of specific binders. These trends help us predict which antigens might require more attention in order to produce high quality binders. Currently we have expressed 557 TFs, and 435 yielded about 3000 primary validated antibodies (based on unique sequence, affinity, stability, expressibility). Of the Fabs tested by Spiked-IP, 56% have passed. We suspect that the higher success rate for

epigenetic antigens (68% of input) is because the antigens were chosen based on more accurate structural information.

Laboratory automation was critical to standardize the many routine and laborious activities ranging from phage sorting, ELISA, colony picking, PCR and stability measurements to facilitate large-scale generation of affinity reagents. The automated selection process greatly reduced the time from months in animals to weeks on the robot with reduced costs associated with antibody generation, and increased the throughput to up to 48 per batch of antibody selections. To take maximal advantage of the robotics it was necessary to make numerous adjustments in the bench scale phage and antigen protocols. For example, the automated magnetic bead selections coupled with the catch and release approach permitted clean elution of antigen-bound phage of very high affinity that can be problematic for low pH glycine elution. Moreover, protocols for pre-clearing the starting library and enriching selection pools for functional phage using Protein A purification have been added.

We attribute much of the success here to choosing a remarkably stable Fab library scaffold (average  $T_m$  80°C). The selected Fabs show consistently high expression by simple shake flask culturing in *E. coli*. They are easily purified in high-throughput using simple Protein A columns. This synthetic library built on a highly stable and well-expressed scaffold can provide more consistent stability and expression than for natural B-cell derived libraries (7), which have multiple scaffolds that vary widely in stability and recombinant expression.

The selected Fabs had EC<sub>50</sub> values that ranged from high pM to 50nM (average 20nM) without additional affinity maturation. Moreover, these are readily converted to a

Mab format that routinely affords a 10-50 fold boost in affinity due to avidity and can be expressed in HEK293 cells at levels comparable to those in *E. coli* (Sachdev Sidhu et al., unpublished results).

The Fabs typically showed good evidence for specificity in cell extracts by IP-MS [53] or here by Spiked-IP. While the latter assay and others like it (49) do not directly identify possible off-targets, these are quite stringent for removing antibodies that react with non- or poly-specificity. Fabs generated to highly homologous SCAN domains showed high mono-specificity. In cases where cross-reactivity was observed this was to highly related domains, typically having homology between 80-90%. This is remarkable in that we did not employ counter selections against the closest homologs and neither did we affinity mature the antibodies to increase selectivity. Such measures could be employed with these recombinant antibodies should cross-reactivity be observed.

TFs are extremely important players in biology. We have tested 1017 Fabs for their ability to bind 270 TFs in six different human cell lines. Remarkably, an average of 70% the Fabs recognized their target antigen predominantly in the cytosol and only about 20% in the nucleus. These data are similar to data presented at the HPA where a low percentage of TF antibodies bind only in the nucleus. Although we found the average distribution of TFs in the cytosol and nucleus were similar across the six cell lines, the TFs often were found in different compartments in different cells. The IF staining pattern represents a different bar code for each cell line and we believe this reflects differences in activation states in different cells. Our studies reinforce many individual studies showing that TFs transit from the cytosol to the nucleus upon

activation. We believe these reagents will be very useful for systems-wide studies to study how TF activation affects transport as well as other important studies.

The recombinant antibodies have characteristics of highly selective binding reagents based on the high throughput validation data presented: stringent selection by phage display, unique sequences for each Fab, high stability and affinity, high and soluble expression in *E. coli*, competitive binding with antigen, binding in the presence of complex cell lysates, and specific immunofluorescence binding patterns in cells. Although no systematic analysis of the suitability of the antibodies for Western blotting has been done, we do not anticipate high performance since the Fabs were intentionally raised to soluble protein domains. Nonetheless, for about 20 Fabs tested with collaborators, 5 showed a Western positive band corresponding to the TF suggesting a subset will work in Western blots (Gordon Whiteley, National Cancer Institute, unpublished results).

We have not systematically tested the performance of these antibodies in Chromatin immunoprecipitation (ChIP-Seq) experiments due to cost, but selective examples have been tested by Marcon and co-workers (53). Clearly much more validation is warranted for these and many other applications including their use for structural studies. Nearly 1000 of the genes and expression constructs for these Fabs are currently available to academic researchers along with validation data through the Recombinant Antibody Network ([recombinant-antibodies.org](http://recombinant-antibodies.org)) and the plasmid repository at DNASU (<http://tinyurl.com/lm69t3s>). More will be made available as they transit the pipeline. Although we have only scratched the surface, an industrial-scale platform is in place for the generation of renewable antibodies to the entire proteome.

**References:**

1. Pasqualini, R. and W. Arap, *Hybridoma-free generation of monoclonal antibodies*. Proc Natl Acad Sci U S A, 2004. **101**(1): p. 257-9.
2. Harlow, E. and D. Lane, *Using antibodies : a laboratory manual*. 1999, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. xiv, 495 p.
3. Berglund, L., et al., *A gene-centric Human Protein Atlas for expression profiles based on antibodies*. Mol Cell Proteomics, 2008. **7**(10): p. 2019-27.
4. Bordeaux, J., et al., *Antibody validation*. Biotechniques, 2010. **48**(3): p. 197-209.
5. Bradbury, A. and A. Pluckthun, *Reproducibility: Standardize antibodies used in research*. Nature, 2015. **518**(7537): p. 27-9.
6. Gloriam, D.E., et al., *A community standard format for the representation of protein affinity reagents*. Mol Cell Proteomics, 2010. **9**(1): p. 1-10.
7. Schofield, D.J., et al., *Application of phage display to high throughput antibody generation and characterization*. Genome Biol, 2007. **8**(11): p. R254.
8. Turunen, L., et al., *Automated panning and screening procedure on microplates for antibody generation from phage display libraries*. J Biomol Screen, 2009. **14**(3): p. 282-93.
9. Pershad, K., et al., *Generating a panel of highly specific antibodies to 20 human SH2 domains by phage display*. Protein Eng Des Sel, 2010. **23**(4): p. 279-88.
10. Colwill, K. and S. Graslund, *A roadmap to generate renewable protein binders to the human proteome*. Nat Methods, 2011. **8**(7): p. 551-8.
11. Vaquerizas, J.M., et al., *A census of human transcription factors: function, expression and evolution*. Nat Rev Genet, 2009. **10**(4): p. 252-63.
12. Spolar, R.S. and M.T. Record, Jr., *Coupling of local folding to site-specific binding of proteins to DNA*. Science, 1994. **263**(5148): p. 777-84.
13. Frankel, A.D. and P.S. Kim, *Modular structure of transcription factors: implications for gene regulation*. Cell, 1991. **65**(5): p. 717-9.
14. Bradbury, A.R., et al., *Beyond natural antibodies: the power of in vitro display technologies*. Nat Biotechnol, 2011. **29**(3): p. 245-54.
15. Persson, H., et al., *CDR-H3 diversity is not required for antigen recognition by synthetic antibodies*. J Mol Biol, 2013. **425**(4): p. 803-11.
16. Koerber, J.T., et al., *Nature-inspired design of motif-specific antibody scaffolds*. Nat Biotechnol, 2013. **31**(10): p. 916-21.
17. Xiao, R., et al., *The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium*. J Struct Biol, 2010. **172**(1): p. 21-33.
18. Acton, T.B., et al., *Preparation of protein samples for NMR structure, function, and small-molecule screening studies*. Methods Enzymol, 2011. **493**: p. 21-60.
19. Harrison, S.C., *A structural taxonomy of DNA-binding domains*. Nature, 1991. **353**(6346): p. 715-9.
20. Zhang, Y., U. Werling, and W. Edelmann, *SLiCE: a novel bacterial cell extract-based DNA cloning method*. Nucleic Acids Res, 2012. **40**(8): p. e55.
21. Aslanidis, C. and P.J. de Jong, *Ligation-independent cloning of PCR products (LIC-PCR)*. Nucleic Acids Res, 1990. **18**(20): p. 6069-74.

22. Keates, T., et al., *Expressing the human proteome for affinity proteomics: optimising expression of soluble protein domains and in vivo biotinylation*. N Biotechnol, 2012. **29**(5): p. 515-25.
23. Structural Genomics, C., et al., *Protein production and purification*. Nat Methods, 2008. **5**(2): p. 135-46.
24. Miersch, S., et al., *Scalable High Throughput Selection from Phage-display Synthetic Antibody Libraries*. J Vis Exp, 2015(95).
25. Yin, G., et al., *Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system*. MAbs, 2012. **4**(2).
26. Paduch, M., et al., *Generating conformation-specific synthetic antibodies to trap proteins in selected functional states*. Methods, 2013. **60**(1): p. 3-14.
27. Boder, E.T. and K.D. Wittrup, *Yeast surface display for screening combinatorial polypeptide libraries*. Nat Biotechnol, 1997. **15**(6): p. 553-7.
28. Weaver-Feldhaus, J.M., et al., *Yeast mating for combinatorial Fab library generation and surface display*. FEBS Lett, 2004. **564**(1-2): p. 24-34.
29. Kay, B.K., S. Thai, and V.V. Volgina, *High-throughput biotinylation of proteins*. Methods Mol Biol, 2009. **498**: p. 185-96.
30. Ackerman, M., et al., *Highly avid magnetic bead capture: an efficient selection method for de novo protein engineering utilizing yeast surface display*. Biotechnol Prog, 2009. **25**(3): p. 774-83.
31. Rakestraw, J.A., et al., *Secretion-and-capture cell-surface display for selection of target-binding proteins*. Protein Eng Des Sel, 2011. **24**(6): p. 525-30.
32. Koerber, J.T., M.J. Hornsby, and J.A. Wells, *An improved single-chain fab platform for efficient display and recombinant expression*. J Mol Biol, 2015. **427**(2): p. 576-86.
33. Garrard, L.J., et al., *Fab assembly and enrichment in a monovalent phage display system*. Biotechnology (N Y), 1991. **9**(12): p. 1373-7.
34. Bass, S., R. Greene, and J.A. Wells, *Hormone phage: an enrichment method for variant proteins with altered binding properties*. Proteins, 1990. **8**(4): p. 309-14.
35. Lee, C.V., S.S. Sidhu, and G. Fuh, *Bivalent antibody phage display mimics natural immunoglobulin*. J Immunol Methods, 2004. **284**(1-2): p. 119-32.
36. Los, M., J.M. Roodhart, and E.E. Voest, *Target practice: lessons from phase III trials with bevacizumab and vatalanib in the treatment of advanced colorectal cancer*. Oncologist, 2007. **12**(4): p. 443-50.
37. Schulman, E.S., *Development of a monoclonal anti-immunoglobulin E antibody (omalizumab) for the treatment of allergic respiratory disorders*. Am J Respir Crit Care Med, 2001. **164**(8 Pt 2): p. S6-11.
38. Hudis, C.A., *Trastuzumab--mechanism of action and use in clinical practice*. N Engl J Med, 2007. **357**(1): p. 39-51.
39. Fellouse, F.A., et al., *High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries*. J Mol Biol, 2007. **373**(4): p. 924-40.
40. Miller, K.R., et al., *T cell receptor-like recognition of tumor in vivo by synthetic antibody fragment*. PLoS One, 2012. **7**(8): p. e43746.
41. Kaneko, T., et al., *Superbinder SH2 Domains Act as Antagonists of Cell Signaling*. Sci Signal, 2012. **5**(243): p. ra68.

42. Gao, J., S.S. Sidhu, and J.A. Wells, *Two-state selection of conformation-specific antibodies*. Proc Natl Acad Sci U S A, 2009. **106**(9): p. 3071-6.
43. Thomsen, N.D., J.T. Koerber, and J.A. Wells, *Structural snapshots reveal distinct mechanisms of procaspase-3 and -7 activation*. Proc Natl Acad Sci U S A, 2013.
44. Ye, J.D., et al., *Synthetic antibodies for specific recognition and crystallization of structured RNA*. Proc Natl Acad Sci U S A, 2008. **105**(1): p. 82-7.
45. Schwab, C. and H.R. Bosshard, *Caveats for the use of surface-adsorbed protein antigen to test the specificity of antibodies*. J Immunol Methods, 1992. **147**(1): p. 125-34.
46. Smith, G.P., *Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface*. Science, 1985. **228**(4705): p. 1315-7.
47. Starovasnik, M.A., et al., *Antibody variable region binding by Staphylococcal protein A: thermodynamic analysis and location of the Fv binding site on E-domain*. Protein Sci, 1999. **8**(7): p. 1423-31.
48. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability*. Nat Protoc, 2007. **2**(9): p. 2212-21.
49. Hotzel, I., et al., *A strategy for risk mitigation of antibodies with fast clearance*. MAbs, 2012. **4**(6): p. 753-60.
50. Nishikori, S., et al., *Broad ranges of affinity and specificity of anti-histone antibodies revealed by a quantitative peptide immunoprecipitation assay*. J Mol Biol, 2012. **424**(5): p. 391-9.
51. Lamboy, J.A., et al., *Phage wrapping with cationic polymers eliminates nonspecific binding between M13 phage and high pI target proteins*. J Am Chem Soc, 2009. **131**(45): p. 16454-60.
52. Reich, N.C., *STATs get their move on*. JAKSTAT, 2013. **2**(4): p. e27080.
53. Marcon, E., et al., *Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation*. Nat Methods, 2015. **12**(8): p. 725-31.
54. Uhlen, M., et al., *Towards a knowledge-based Human Protein Atlas*. Nat Biotechnol, 2010. **28**(12): p. 1248-50.
55. Hober, S. and M. Uhlen, *Human protein atlas and the use of microarray technologies*. Curr Opin Biotechnol, 2008. **19**(1): p. 30-5.
56. Mersmann, M., et al., *Towards proteome scale antibody selections using phage display*. N Biotechnol, 2010. **27**(2): p. 118-28.
57. Lipman, N.S., et al., *Monoclonal versus polyclonal antibodies: distinguishing characteristics, applications, and information resources*. ILAR J, 2005. **46**(3): p. 258-68.
58. Konthur, Z., M. Hust, and S. Dubel, *Perspectives for systematic in vitro antibody generation*. Gene, 2005. **364**: p. 19-29.
59. Kehoe, J.W., et al., *Using phage display to select antibodies recognizing post-translational modifications independently of sequence context*. Mol Cell Proteomics, 2006. **5**(12): p. 2350-63.
60. Kierny, M.R., T.D. Cunningham, and B.K. Kay, *Detection of biomarkers using recombinant antibodies coupled to nanostructured platforms*. Nano Rev, 2012. **3**.

61. Schilling, J., J. Schoppe, and A. Pluckthun, *From DARPins to LoopDARPins: novel LoopDARPin design allows the selection of low picomolar binders in a single round of ribosome display*. *J Mol Biol*, 2014. **426**(3): p. 691-721.
62. Elsliger, M.A., et al., *The JCSG high-throughput structural biology pipeline*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 2010. **66**(Pt 10): p. 1137-42.
63. Letunic, I., T. Doerks, and P. Bork, *SMART: recent updates, new developments and status in 2015*. *Nucleic Acids Res*, 2015. **43**(Database issue): p. D257-60.
64. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in performance and usability*. *Mol Biol Evol*, 2013. **30**(4): p. 772-80.
65. Waterhouse, A.M., et al., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench*. *Bioinformatics*, 2009. **25**(9): p. 1189-91.
66. Zhang, H., et al., *EvolView, an online tool for visualizing, annotating and managing phylogenetic trees*. *Nucleic Acids Res*, 2012. **40**(Web Server issue): p. W569-72.

#### Footnotes:

We are grateful to useful advice from the Sidhu lab and TRAC at University of Toronto, the Kossiakoff and Koide labs at University of Chicago, James Wells lab at UCSF, and James Marks for providing yeast display vectors and advice, and Janet Huang and Tom Acton in the Anderson lab at Rutgers University for protein purification. Funding was generously provided by a U54 grant from the NIH common fund (1U54HG006436). Additional funding was provided to build the UCSF robotics by the NIH Office of the Director (1S10OD01237) and California Institute for Quantitative Biosciences (QB3) at UCSF.

#### Figure Legends:

**Figure 1.** Overview of the RAN recombinant Fab selection pipeline, selection process, and examples of primary validation. (A) Protein antigens expressed in *E. coli* or other formats (Left Panel) enter the Fab-phage pipeline as biotinylated or GST-fusion antigens. Three to four rounds of antibody-phage selection generate pools of Fab-

phage that are subsequently validated in a single point competition Fabphage ELISA and sequenced to identify unique binding sequences (Middle Panel). Unique Fabs that pass primary these validation tests proceed into various secondary validation assays including ELISA EC50, Spiked-IP, and Immunofluorescence (Right Panel). (B) UCSF robotic antibody production pipeline showing two liquid handling instruments for ELISA, KingFisher Flex magnetic bead separator for phage selections, and K6-2 colony picking robots. Similar robotics platforms are in-place at all three RAN locations. (C) Solution based phage selection was conducted in a 7-step process that was repeated up to 4 cycles; 1. Biotinylated antigens were bound to streptavidin magnetic beads; 2. Bead-antigen complex was transferred to Fab-phage library; 3. Fab-phage-antigen-bead complex was transferred to wash buffer where any unbound or weakly associated Fab-phage was removed; 4. Specific Fab-phage bound to antigen was eluted from magnetic beads by TEV protease cleavage; 5. The processed magnetic beads and non-specifically bound Fab-phage were removed from well and discarded; 6. Antigen specific phage was propagated in *E. coli*; 7. Propagated Fab-phage was purified and the process repeated. (D) Examples of Fab-phage that pass (green box) or fail the competition ELISA (red boxes) are shown. Direct binding ELISAs (Y-axis) were conducted to measure the composite of expression and binding capability and single point competition ELISAs (X-axis) were measured to verify antigen specificity. For ELISAs 96 individual *E. coli* colonies were tested for each antigen with anti-M13-HRP phage secondary antibody with TMB chromagen development that is monitored kinetically at OD650/min. Each FAB-Phage expressing colony is represented as a single spot on the graph with the calculated Competition Ratio (x-axis) and Total Fab-

phage Binding (y-axis). Two examples are shown for antigens selections that produced Fabs that pass and one that fail validation.

**Figure 2.** General parameters that affect successful Fab generation and characterization. Fabs expressed in *E. coli* were measured for antibody (A) expression, (B) stability, and (C) affinity. Fabs successfully passing primary validation were cloned into *E. coli* expression plasmids and are expressed and purified for further testing. Fabs derived from Library E and F typically expressed between 1 and 10mg/L of bacterial culture (n=720). Fabs were often tested for thermostability by DSF and as shown here produce melting peaks between 75°C and 85°C (n=96). EC50 ELISAs were conducted on Fabs to determine binding affinity values with a subset shown here (n=201) with ~85% of Fabs binding with an EC50 <20nM.

**Figure 3.** Overview of Spiked-IP with examples of data analysis and relative successes of two protein domains. (A) Diagram of Spiked IP showing Fabs immobilized on magnetic beads binding to biotinylated antigens (Left) and after transfer to fresh well (Middle) the antigen-Fab-bead complex was bound to fluorescently labeled streptavidin then binding was quantified by Flow Cytometry. Samples were analyzed for binding characteristics in both buffer and a complex Hek293T lysate that showed specificity of the Fab to its cognate antigen. This assay was designed run in forward or reverse mode where antigen could be immobilized on beads with Fabs in solution. (B) An example data set that shows 3 antigens with 3-4 Fabs each with both Passing and Failing Spiked-IP tests. Negative control (Blue) consists of beads without Fab in the presence of antigen then (Green) bars represent the assay run in the presence of buffer and (Red) lysate. To pass the assay, the differences in median fluorescence intensity

between buffer and lysate tests must not be  $>2x$ . Fabs that passed the Spiked-IP test were denoted with (\*). (C) Representative success of Fabs generated from SCAN ( $n=39$ ) and Zinc Finger ( $n=42$ ) domains through secondary Spiked-IP test.

**Figure 4. Assessment of the specificity of anti-SCAN domain Fabs.**

Left panel shows specificity heatmap of 18 Fabs tested against 17 closely related SCAN domains using the direct ELISA method. Fabs used in experiment are sorted according to their order in the phylogenetic tree and colors on the heatmap represent strength of the ELISA signal. anti-GFP Fab and eGFP protein was used as a positive control. An antibody raised again ZNF496 that failed Spiked-IP validation was used as a negative to show non-specific binding. Anticipated signals on diagonal represent interactions where Fabs recognize their intended antigens. Left panel shows pairwise identity heatmap of SCAN domain antigens tested. Multiple sequence alignment and phylogenetic tree of SCAN domains, as defined in SMART database, was constructed in MAFFT using L-INS-i and NJ methods respectively[63, 64]. All pairwise identities were calculated in Jalview [65]. Phylogenetic tree and heatmap annotations were visualized using EvolView [66].

**Figure 5. Analysis of Fab selection success rates.** Success rates were analyzed as a function of antigen (A) molecular weight or (B) isoelectric point. Greater Fab selection success was appreciated for antigens with a MW  $\geq 14\text{kDa}$  and  $\text{pI} < 8$  ( $n=447$  antigens;  $n=3343$  Fabs). (C) Multiple antigen expression pipelines can increase likelihood of obtaining high quality Fabs. Test antigens ( $n=32$ ) were expressed in three different expression formats: *E coli*, IVTT or YAD. Fab-phage selections were conducted for each antigen using the robotic Fab-phage selection pipeline. The number of primary

validated Fabs is shown for each antigen.

**Figure 6.** Photomicrographs and heat map showing Immunofluorescence staining patterns (cytosol in blue to nuclear in red) for 1017 Fabs directed to 270 TFs. (A) Expressed and purified Fabs were screened against 6 commonly used human cell lines and images were screened by the Columbus imaging software to generate signal intensities across multiple fields. Intensities were used go generate a heatmap that shows the ratio of intensities of nuclear to cytoplasmic localization of Fabs. (B) Representative photomicrographs for immunofluorescence staining of Fabs (n=1017) against TFs (n=270) in 6 different fixed and permeabilized mammalian cell lines known to express the TF. Scoring was represented as Cyto, Cyto/Nuc, Nuc or Misc. (C) Similar analysis was conducted by visually inspecting images and assessing staining patterns based on 4 GO terms; Nuclear (Nuc), Nuclear/Cytoplasmic (Nuc/Cyto), Cytoplasmic (Cyto), and Miscalenaous staining (Misc). Mean localization results were calculated and showed that ~67% of all Fabs labeled a protein in the cytoplasm compared to ~19% in the nucleus.

**Table 1.** Ability of antigens to produce primary and secondary validated Fabs as a function of antigen domain. Of the 537 TF antigens, representing at least 18 discrete domain types, 435 (81%) successfully generated sequence unique Fabs that passed competition ELISA. For 193 antigens with Fabs entering secondary validation, 108 (56%) passed the Spiked-IP validation test. The best performing domains are the Forkhead, SCAN, JMJ, and BTB domains with success rates of 100%, 88%, 100%, and 86% respectively.

Domain Type	Primary Validation			Secondary Validation		
	Total # of Antigens	Passing Antigens	Domain Success (%)	Total # of Antigens	Passing Antigens	Domain Success (%)
BRD	37	32	86	2	1	50
BTB	10	9	90	7	6	86
bZip	11	11	100	4	0	0
Cupin	3	3	100	NT	NT	
DUF	2	2	100	NT	NT	
Fork Head	5	4	80	3	3	100
HDAC	3	2	67	NT	NT	
HMG Box	10	8	80	5	2	40
Homeobox	6	5	83	15	5	33
Hormone Recep.	6	5	83	5	3	60
JMJ	12	11	92	3	3	100
Misc	141	122	87	50	34	68
PHD	15	12	80	3	0	0
PWWP	7	7	100	2	1	50
SCAN	46	46	100	40	35	88
SET	31	21	68	5	2	40
SIR2	4	4	100	3	1	33
ZNF	148	105	71	46	12	26
<b>Sum</b>	<b>537</b>	<b>435</b>	<b>81</b>	<b>193</b>	<b>108</b>	<b>56</b>

Figure 1:

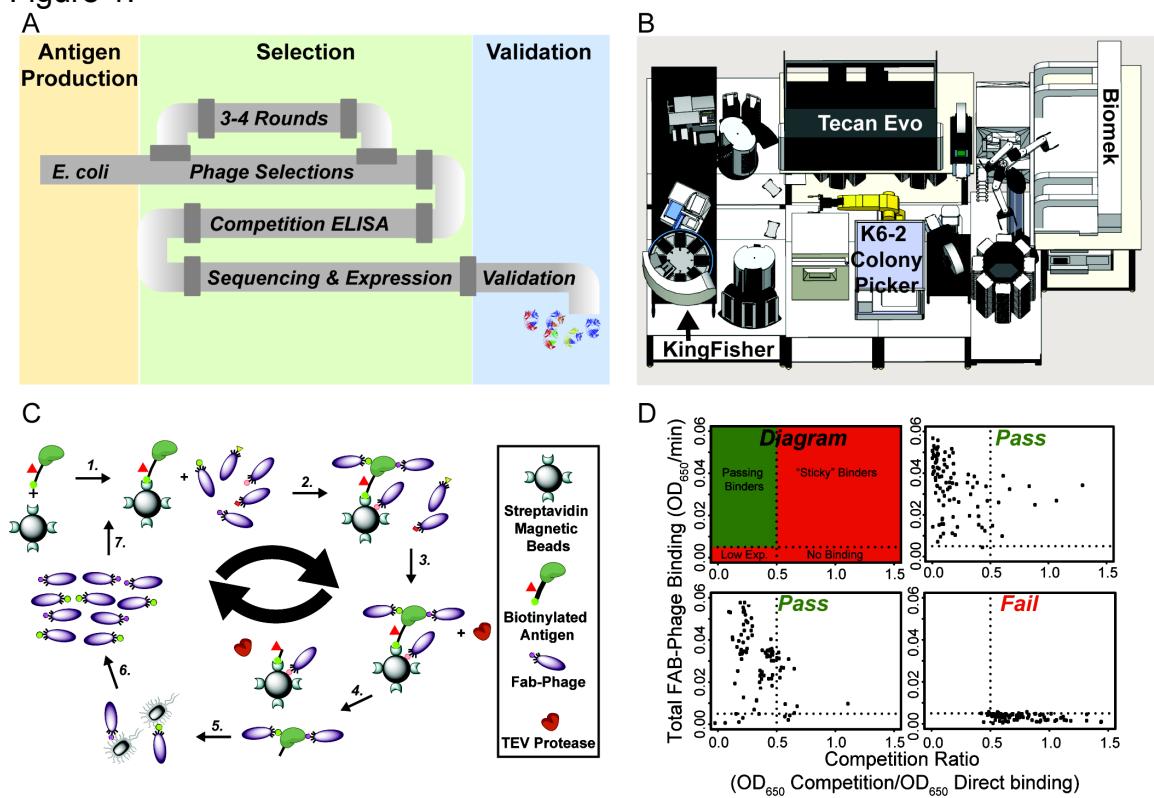


Figure 2:

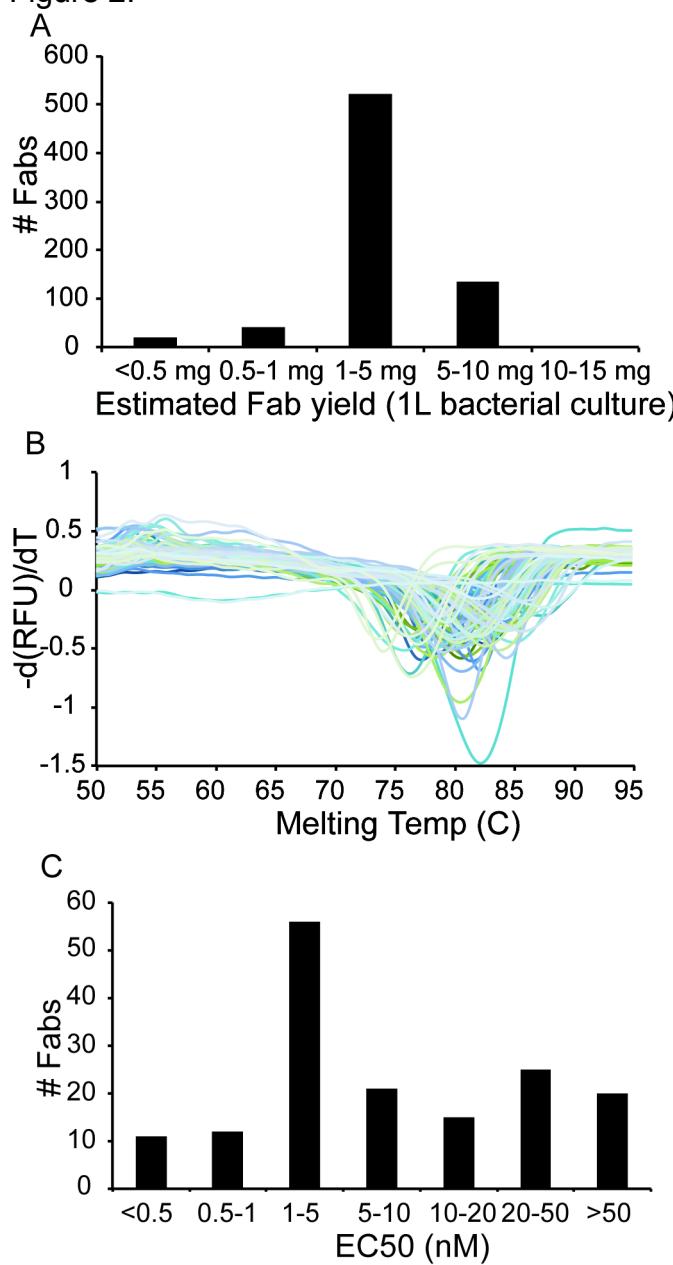


Figure 3:

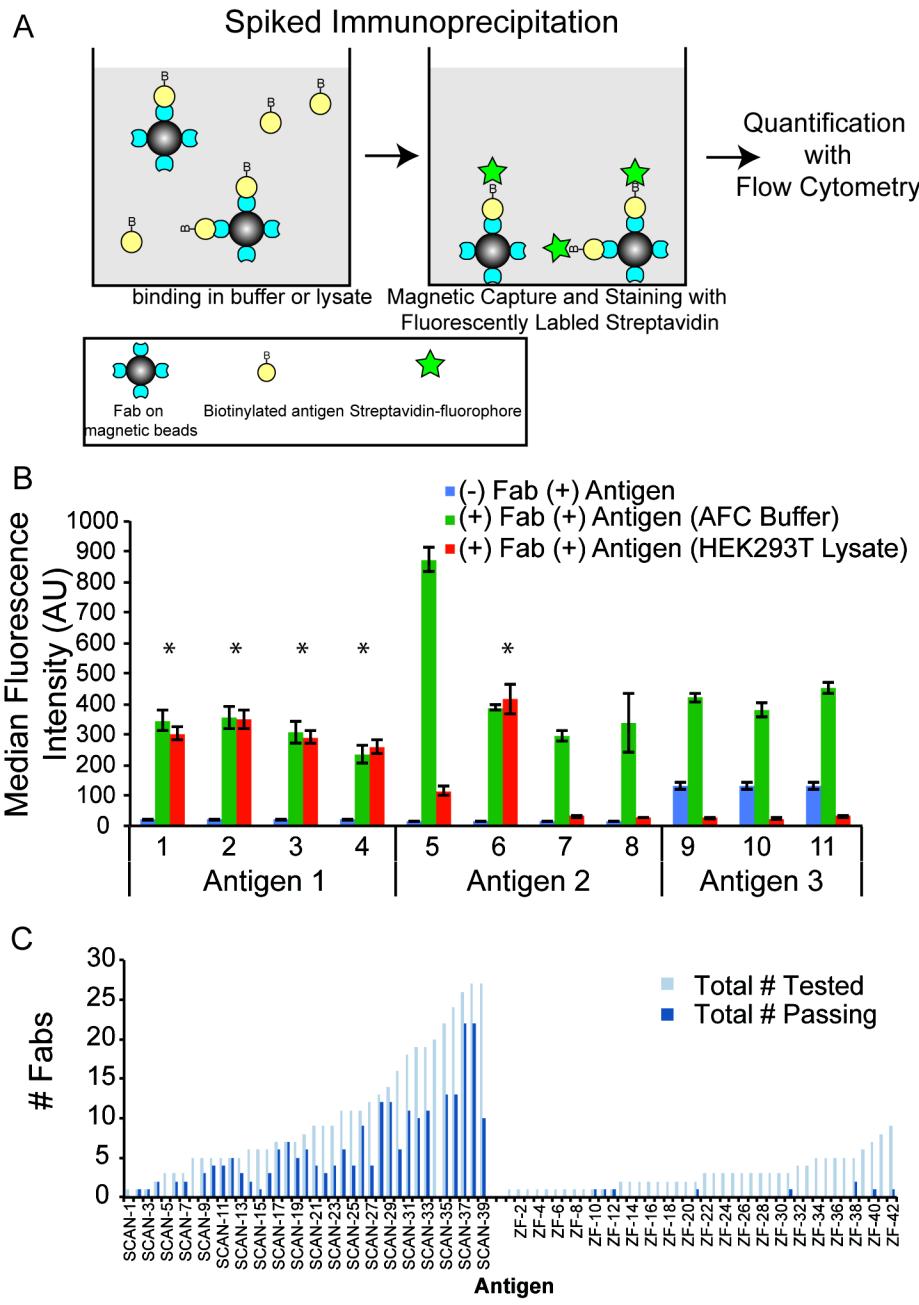


Figure 4:

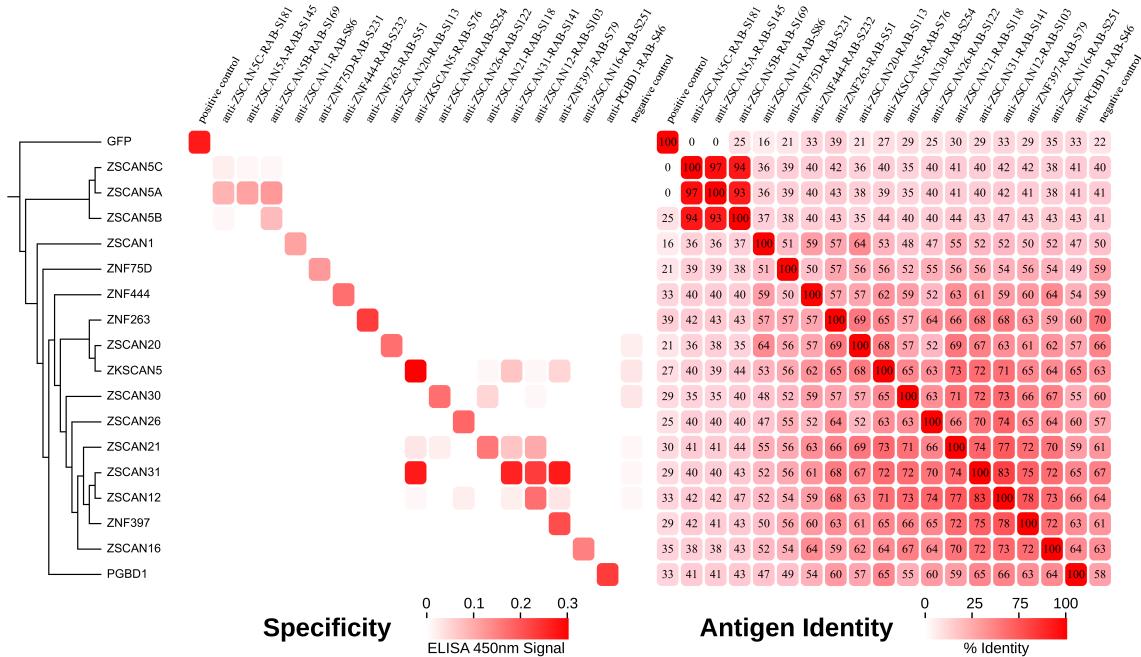
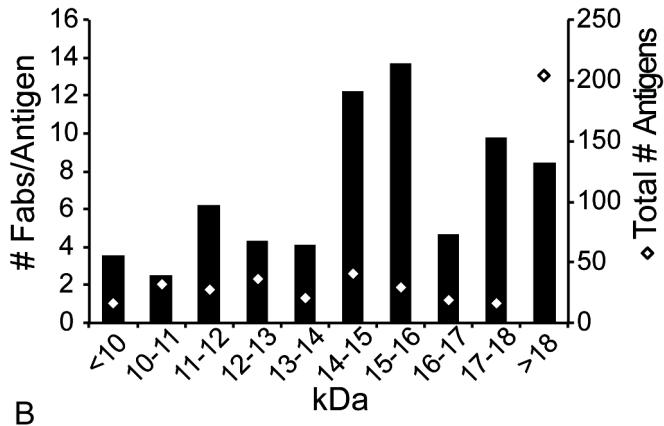
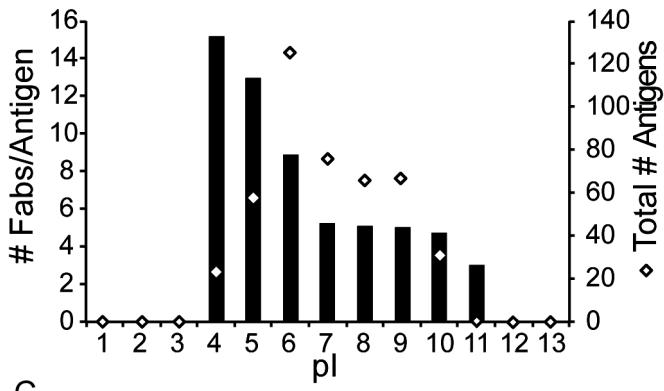


Figure 5:  
A



B



C

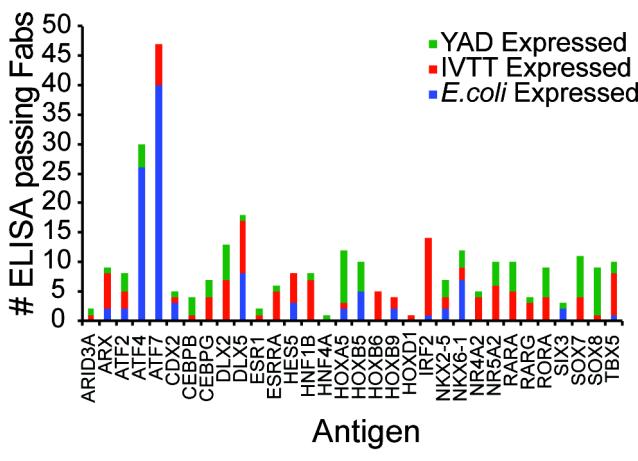


Figure 6:

