Published in final edited form as:
Nature. 2014 November 27; 515(7528): 554–557. doi:10.1038/nature13761.

Multiplex single-molecule interaction profiling of DNA barcoded proteins

Liangcai Gu[1], **Chao Li**[2], **John Aach**[1], **David E. Hill**[1,3], **Marc Vidal**[1,3], and **George M. Church**[1,2]

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA.

[2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA.

[3]Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA.

Abstract

In contrast with advances in massively parallel DNA sequencing[1], high-throughput protein analyses[2-4] are often limited by ensemble measurements, individual analyte purification and hence compromised quality and cost-effectiveness. Single-molecule (SM) protein detection achieved using optical methods[5] is limited by the number of spectrally nonoverlapping chromophores. Here, we introduce a single molecular interaction-sequencing (SMI-Seq) technology for parallel protein interaction profiling leveraging SM advantages. DNA barcodes are attached to proteins collectively via ribosome display[6] or individually via enzymatic conjugation. Barcoded proteins are assayed en masse in aqueous solution and subsequently immobilized in a polyacrylamide (PAA) thin film to construct a random SM array, where barcoding DNAs are amplified into in situ polymerase colonies (polonies)[7] and analyzed by DNA sequencing. This method allows precise quantification of various proteins with a theoretical maximum array density of over one million polonies per square millimeter. Furthermore, protein interactions can be measured based on the statistics of colocalized polonies arising from barcoding DNAs of interacting proteins. Two demanding applications, G-protein coupled receptor (GPCR) and antibody binding profiling, were demonstrated. SMI-Seq enables “library vs. library” screening in a one-pot assay, simultaneously interrogating molecular binding affinity and specificity.

To analyze proteins in a massively parallel SM format, we generated proteins that are molecularly coupled to a DNA bearing a barcoding sequence. One barcoding approach is to *in vitro* translate and display proteins on protein–ribosome–mRNA–cDNA (PRMC) complexes, in which the cDNA contains a synthetic barcode at the 5’ end of protein open reading frames (ORFs) (Fig. 1a). Specifically, the ribosome display was performed by using mRNA–cDNA hybrids as templates and an *in vitro* translation (IVT) system reconstituted

Correspondence and requests for materials should be addressed to L.G. (liangcaigu@gmail.com) or G.M.C. (gchurch@genetics.med.harvard.edu)..

Author Contributions L.G. and G.M.C. conceived the technique; L.G. and C.L. performed experiments and analyzed data; J.A. built the mathematical model and assisted the colocalization analyses; D.H and M.V. assisted the production of barcoded proteins; L.G., J.A. and G.M.C. wrote the manuscript with help from the other authors.

with purified components[8] that was shown to stabilize PRMC complexes (Extended Data Fig. 1). PRMC complexes bearing full-length proteins of interest were enriched by Flag-tag affinity purification. Notably, this approach is applicable to a library of proteins of various sizes and size-related biases during decoding can be avoided by using uniformly sized barcoding DNAs. Alternatively, some proteins that can only be functionally expressed *in vivo* require to be individually barcoded. Thus, fusion proteins were constructed with an engineered enzyme tag, HaloTag[9], which mediates an efficient covalent conjugation to a HaloTag ligand-modified double stranded DNA (Fig. 1b). Our method is adaptable to a microtiter plate format for automated parallel protein production (Extended Data Fig. 2).

A complex mixture of barcoded proteins can be identified and quantified by in situ sequencing their barcodes (Fig. 2a). They were immobilized into an ultrathin-layer crosslinked PAA gel attached to a microscopic slide, and their barcoding DNAs bearing a 5'-acrydite modification (Fig. 1) were covalently crosslinked to the gel matrix to prevent template drifting (Extended Data Fig. 3). A solid-phase polymerase chain reaction (PCR), with two gel-anchored primers, was performed according to an adapted isothermal bridge amplification protocol[10] in an assembled flow cell. This amplification showed a high efficiency of ~80% barcode detection (Extended Data Fig. 4a), and resulted in polonies of ~1 μm diameter (Fig. 2b) similar to the clusters generated on an Illumina platform[10]. Polonies were identified by hybridization with fluorescent probes, single-base extension (SBE) or ligation-based sequencing[11].

To test the accuracy of our method, we selected nine immunoglobulin and non-immunoglobulin binding proteins and three antigens (e.g., human, bacterial and viral proteins) of a molecular weight range from 3.4 to 120 kDa (Extended Data Table 1). Mixed PRMC complexes were prepared in six barcoded dilutions, with concentrations spanning six orders of magnitude, pooled together and subjected to the SM quantification. Barcode detection efficiencies of different proteins were found to be almost identical at various concentrations (Extended Data Fig. 4). The in situ SM quantification can avoid PCR amplification bias[12] and shows high reproducibility, e.g., Pearson correlation coefficient R was above 0.99 when over 1,000 protein polonies were detected (Fig. 2c). Because proteins were highly diluted (e.g., at less than picomolar concentrations) prior to array deposition, protein monomers should be the predominant form.

Interacting barcoded proteins can be indirectly detected by joining their barcoding DNAs via ligation[13,14] or primer extension[15]. Here, direct observation and counting of SM protein complexes should be possible if barcoding DNAs of interacting proteins can be amplified into colocalized polonies. To test this, we generated DsRed, which naturally forms a tetramer, with monomers each bearing one of two different barcodes. To avoid dissociation of any complexes during the array analysis, we crosslinked them with an amine-reactive crosslinker, bis-*N*-succinimidyl-(pentaethylene glycol) ester ($BS(PEG)_5$). The crosslinking was shown to be efficient due to the presence of a lysine-rich TolA spacer domain (Fig 1a and Extended Data Fig. 5). It is evident that barcoding DNAs of the colocalized monomers ($DsRed^a$ and $DsRed^b$) were coamplified into overlapping polonies (Fig. 3a), providing a solid basis for further applications.

Unlike other methods that only detect affinity-enriched proteins (e.g., PLATO[16]), our approach simultaneously counts polonies of both unbound and bound proteins in a single solution. Thus, we sought to determine if it can provide a measure of protein binding affinities. We chose a model system, the GTP-dependent binding of human H-Ras (Ras) to Ras-binding domain of c-Raf-1 (Raf-RBD)[17]. A Raf-RBD polony colocalization ratio—the percentage of Raf-RBD polonies colocalized with Ras polonies—was measured for wild-type (WT) Ras and Raf-RBD and eight Raf-RBD mutants; the Ras protein concentration was titrated over three orders of magnitude (Fig. 3b). Although the colocalization ratio is dependent on protein concentration and crosslinking efficiency and can be affected by experimental variables (e.g., protein quality, crosslinking conditions, polony array density, etc.), all the proteins within a single assay are under the same reaction conditions. Given a similar proportion of active protein and crosslinking efficiency, polony colocalization ratios could be correlated with ratios of bound proteins at equilibrium and thus their binding affinities. To test this, the colocalization ratios were plotted against previously reported dissociation constants (K_Ds) ranging from nanomolar to micromolar[18,19] (Fig. 3b and Supplementary Table 1), and fitted by using a one-site-specific binding model (dashed curves). The fitted and observed average colocalization ratios show relatively high agreement ($R > 0.96$), except for the A85K mutant of significantly lower experimental values than predicted by the model, likely due to the disruption of Lys^{85}-mediated interactions by the crosslinking[19]. Therefore, this method could be useful for high-throughput screening of protein binding affinities.

As a first high-throughput screening application, we have looked at small molecule-mediated protein–protein interactions. An advantage of our method over traditional solid-phase techniques such as protein microarrays[3] is that we store and assay proteins in aqueous solution. To exploit it, we decided to address challenges in screening G-protein coupled receptors (GPCRs), the largest membrane protein family and premier drug targets[20]. Current GPCR–ligand screening techniques mainly rely on cell-based assays[21], which are subject to limitations such as the inhomogeneous nature of the samples, the presence of other cellular components that can cause false positives or negatives, and limited miniaturization and multiplexing capability (e.g., one receptor per assay). To prepare a homogenous SM GPCR sample compatible with our approach, receptors were stabilized in phospholipid bilayer nanodiscs[22] by assembling detergent-solubilized GPCRs, phospholipids and a membrane scaffold protein, MSP1E3D1, into GPCR–nanodisc complexes[23,24]. GPCR activation upon ligand binding can be functionally assessed by β-arrestin binding to activated receptors, which is a G-protein independent assay applicable to almost all GPCRs including orphan receptors[25].

A compound library can be screened in multiwell plates, and in each well one compound is assayed with many barcoded GPCRs and a β-arrestin-2 (β-arr2) protein bearing a well-position-associated barcode (Fig. 3c). All the samples were pooled and deposited on one slide, and GPCR agonists were detected by measuring GPCR polony colocalization with corresponding β-arr2 polonies. Our efforts to obtain functional GPCRs using IVT systems were not successful, so they were expressed in baculovirus-infected insect cells, purified in nanodiscs and individually barcoded (Fig. 1b). To establish assay conditions, we examined

β-arr2 binding to an agonist (isoproterenol) saturated β_2-adrenergic receptor (ADRB2), with and without GPCR kinase 2 (GRK2)-mediated receptor phosphorylation and under varied β-arr2 protein concentrations (Fig. 3d). The colocalization ratios were measured at 50 imaging positions on the array for statistical analysis. As expected, coupling the receptor phosphorylation to the assay improves the β-arr2 binding, e.g., a ~3 to 11-fold increase (largest $P = 0.002$) of the average colocalization ratios after the phosphorylation. The fitting of β-arr2 titration data for the phosphorylated receptor yielded an apparent K_D of 0.95 nM, which is close to the K_D of 0.23 nM obtained from traditional binding assays using radiolabeled β-arr2[26].

To test the screening performance, we assayed three GPCRs, ADRB2, M1 and M2 muscarinic acetylcholine receptors (CHRM1 and CHRM2) with six compounds including full, partial, subtype selective and non-selective agonists and two antagonists (Fig. 3e and Supplementary Table 2). The colocalization statistical analysis based on measurements of ~13,000-17,000 polonies for each receptor precisely identified the full agonists (isoproterenol and carbachol) from the others (largest $P < 2.7\times10^{-10}$). Moreover, different types of agonists can be distinguished by comparing their polony colocalization ratios, e.g., the full and partial agonists of ADRB2 (isoproterenol and pindolol, respectively; $P < 0.004$), and the orthosteric and allosteric agonists of CHRM1 (carbachol and xanomeline, respectively; $P < 3\times10^{-6}$). Thus, our method could allow parallel GPCR screening and compound profiling.

An intriguing feature of this approach is the ability to screen two barcoded libraries in a single binding assay. Established techniques (e.g., yeast two-hybrid (Y2H) system[2]) for library vs. library screening are cell-based and require pairing both genes from two libraries to identify positive clones by performing individual PCR reactions[27]. To demonstrate this capability, we prototyped a test of a demanding application, the binding profiling of antibody repertoire. The screening of natural or semisynthetic monoclonal antibody (mAb) libraries essentially includes binding affinity selection and specificity profiling which have to be conducted separately with current techniques. The traditional specificity profiling is costly, usually requiring at least one protein chip for a single antibody test[28], and thus has only been commercially applied to therapeutic antibodies. However, both processes could be integrated on our platform by screening an antibody library with a target protein library.

Specifically, we performed a one-pot assay containing 200 ribosome-displayed single-chain variable fragments (scFvs) and 55 *in vitro* synthesized human cytokines, growth factors and receptors (Extended Data Table 2). Twenty scFvs were derived by random mutagenesis from each of ten scFvs, the genes of which were previously synthesized from a programmable DNA microchip[29].

We sequenced ~0.64 million polonies and measured the colocalization ratios for 11,000 scFv–probe pairs at 100 imaging positions (Fig. 4a and Supplementary Table 3). 147 out of 200 scFvs were found with highest colocalization ratios, 95 of which are significantly above the second highest ($P < 0.05$), and thus the highest specificity, to their predicted targets; the others failed probably because the construction of scFv fragments and mutations inhibit target binding. Substantial cross-reactivity can be sensitively detected, e.g., 3474 scFv–

probe pairs showed 10-fold higher polony colocalization than random distribution ($P < 0.01$). ScFv mutants of a same scFv, grouped by their numbers, exhibit similar but not identical binding patterns to the probes. Next, we confirmed the results of 40 scFv–probe pairs by individual immunoprecipitation assays and the colocalization statistics were consistent with relative fluorescence intensities of the protein bands (Fig. 4b). Moreover, to further assess multiplexing potential, we developed a mathematical model that integrated parameters including K_Ds of protein–probe complexes to be detected and numbers of proteins and probes that can be assayed simultaneously (Supplementary Notes). The model suggests that tens of thousands of proteins and probes can be quantitatively analyzed within a single assay.

The protein barcoding requirement imposes limitations on SMI-Seq. First, it cannot directly analyze proteins from biological samples. However, non-barcoded proteins can be detected by using barcoded antibodies or aptamers similarly as a proximity ligation assay[13,14]. In addition, PRMC complexes are susceptible to nuclease contamination, thus limiting the choice of IVT systems. Finally, barcoding DNA can non-specifically bind to proteins bearing nucleic acid-binding domains. Although in the present study DNA templates were individually barcoded, a large library can be prepared by introducing millions of chip-synthesized[29] or random barcoding sequences to an ORF library by a single PCR reaction and later matching them to ORF sequences by next-generation sequencing (NGS). SMI-Seq enables in situ SM counting of proteins and complexes, fundamentally improving sensitivity, accuracy and multiplexity (Extended Data Table 3 and Supplementary Discussion), and thus the demonstrated applications are difficult or impossible for other high-throughput techniques (e.g., PLATO). It is readily adaptable to industrial NGS platforms and translated into many applications. In addition to natural and recombinant proteins, it will be applicable to de novo proteins (e.g., with unnatural amino acids or modifications), nucleic acids and barcoded small molecules[30].

METHODS

DNA construction

Protein coding sequences were synthesized by Genewiz and Integrated DNA Technologies, PCR amplified from plasmids or genomic DNA, or transferred from Gateway-adapted human ORF clones[31] (refer to Supplementary Table 4 for DNA sources, sequences and construction methods) and inserted into a multiple cloning site or Gateway recombination sites of expression vectors (refer to Supplementary Fig. 1 for plasmid construction and Supplementary Table 5 for plasmid and primer sequences) for *in vitro* or *in vivo* protein expression.

Ribosome display-based protein barcoding

To barcode protein libraries of relatively small size (e.g., ≤ 200 in this work), synthetic barcoding sequences were introduced to DNA templates via individual PCRs. Barcoded linear DNA templates were pooled and *in vitro* transcribed using a HiScribe T7 kit (New England Biolabs). To generate mRNA–cDNA hybrids, cDNAs were synthesized by incubating ~0.10 μM mRNA, 1 μM 5'-acrydite and desthiobiotin-modified primer, 0.5 mM

each dNTP, 10 U/μL SuperScript III, 2 U/μL RNaseOUT (Invitrogen) and 5 mM dithiothreitol (DTT) in a buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, and 5 mM $MgCl_2$) at 50°C for 30 min. Synthesized mRNA–cDNA hybrids serve as templates for ribosome display using a PURExpress Δ ribosome kit (New England Biolabs). Typically, a 100 μL IVT reaction with ~0.40 μM mRNA–cDNA hybrids and ~0.30 μM ribosome was incubated at 37°C for 30 min, quenched by addition of 100 μL ice-cold buffer HKM (50 mM HEPES, pH 7.0, 250 mM KOAc, 25 mM $Mg(OAc)_2$, 0.25 U/mL RNasin (Promega), 0.5 mg/mL chloramphenicol, 5 mM 2-mercaptoethanol and 0.1% (v/v) Tween 20) and centrifuged (14,000 g, 4°C) for 10 min to remove insoluble components. PRMC complexes, always kept on ice or in cold room, were subjected to two-step Flag tag and desthiobiotin tag affinity purification to enrich full-length and barcoded target proteins. In brief, proteins were sequentially purified using anti-Flag M2 magnetic beads (Sigma-Aldrich) and streptavidin-coated magnetic beads (Dynabeads M-270 Streptavidin, Life Technologies) blocked with the buffer HKM supplemented with 0.1 mg/mL yeast tRNA and 10 mg/mL BSA. The bound proteins were eluted with the buffer HKM containing 0.1 mg/mL Flag peptide or 5 mM biotin, and their barcoding DNAs were quantitated by real-time PCR.

Protein expression and purification and HaloTag-based barcoding

His-tagged HaloTag–TolA, HaloTag–DsRed–TolA, HaloTag–mCherry–TolA, Ras–TolA–HaloTag and β-arr2–TolA–HaloTag were expressed in *E. coli* or using an *E. coli* IVT system. Proteins were expressed in an OverExpress C41(DE3) strain (Lucigen) with 1 mM isopropyl-β-D-galactopyranoside (IPTG) induction at 30°C for 8-10 h, and purified using immobilized metal affinity chromatography (IMAC) at 4°C. In brief, harvested cells were resuspended in 50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 10 mM imidazole and 20% glycerol, and disrupted by French press. Supernatants of cell lysates were loaded on a 5 ml HisTrap column (GE Healthcare) and non-specifically bound components were washed off with 50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 20 mM imidazole and 10% glycerol. Proteins were eluted with 50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 250 mM imidazole and 10% glycerol, concentrated with Amicon Ultra-15 centrifugal filter units (Millipore), buffer exchanged to a storage buffer (50 mM HEPES, pH 7.0, 150 mM KOAc and 20% glycerol) using a PD10 desalting column (GE Healthcare), flash frozen in 100-500 μL aliquots by liquid N_2 and stored at −80°C. Relatively small amounts of proteins were typically synthesized in an *E. coli* crude extract (RTS 100 *E. coli* HY, 5 PRIME) at 30°C for ~4 h, and similarly purified with His-tag magnetic beads (Dynabeads His-tag, Life Technologies).

Human ADRB2, CHRM1 and CHRM2 were expressed in baculovirus-infected Sf9 cells (Life Technologies), solubilized with detergents as previously described[32,33], and assembled into GPCR–nanodisc complexes followed by affinity purification[34]. Briefly, GPCR genes were synthesized and inserted into a pBac-NFlagHA vector for the expression of the fusion proteins bearing N-terminal Flag and HA tags and a HaloTag domain. Cells were harvested at two days after transfection, homogenized in a 50 mM Tris-HCl, pH 7.4, 50 mM NaCl and 1 mM EDTA with a protease inhibitor (PI) cocktail (Roche), and centrifuged to collect membrane fractions. The membranes were solubilized in 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$, 5 mM $MgCl_2$, 2 mM EDTA, 10% glycerol, 1% *n*-dodecyl-*β*-D-

maltopyranoside (DDM) and a PI cocktail (Set III, EMD Biosciences), and centrifuged at 15,000 g for 15 min; the supernatants were subjected to a bicinchoninic acid assay (Thermo Scientific) to determine the protein concentration. The nanodiscs were assembled by incubating 90 μM MSP1E3D1 (Sigma-Aldrich), 8 mM POPC, 40 mM DDM and 180 μg total membrane protein in 50 mM Tris pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$, 5 mM $MgCl_2$, 2 mM EDTA and 2% glycerol on ice for 45 min, followed by removal of the detergent using Bio-Beads SM-2 (Bio-Rad)[34]. GPCR–nanodisc complexes were bound to anti-Flag M1 agarose resin (Sigma-Aldrich) and eluted with a conjugation buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM EDTA and 5% glycerol) in the presence of 0.2 mg/mL Flag peptide.

Human proteins were *in vitro* synthesized with a human IVT kit (Thermo Scientific). Proteins were individually translated at 30°C for 2 h and purified with the anti-Flag M2 or His-tag magnetic beads. Membrane proteins were stabilized by addition of preassembled nanodiscs (2 μL MembraneMax reagent/50 μL reaction, Life Technologies). To semi-quantitatively analyze HaloTag fusion proteins, their HaloTag domains were covalently labeled with a fluorescent reporter Halo-TMR (Promega) and analyzed by SDS-PAGE and the fluorescent gel imaging with a Typhoon Trio Imager (GE Healthcare).

220-bp barcoding DNAs were prepared in parallel by adding barcoding sequences via PCR with a universal template and barcoded primers, and introducing the modifications by a secondary PCR with the modified primers (Integrated DNA Technologies). A HaloTag ligand was conjugated to the primer by incubating 100 μM amino modified oligo and 10 mM succinimidyl ester (O4) ligand (Promega) in 50 mM Na_2HPO_4, pH 8.0, 150 mM NaCl and 50% formamide at room temperature for 1 h; the modified oligo was purified by reverse-phase high-performance liquid chromatography using a Zorbax Eclipse XDB-C18 column (5 μm, 9.4×250 mm, Agilent Technologies) and an elution gradient of 5-70% CH_3CN/H_2O (0.1 M triethylammonium acetate). To generate protein–DNA conjugates, we typically incubated ~0.5-2 μM modified barcoding DNAs and ~2-5 μM HaloTagged proteins in the conjugation buffer with gentle shaking at room temperature for 2-4 h; the conjugates were purified with the anti-Flag M2 or His-tag magnetic beads and the streptavidin magnetic beads. Barcoded proteins were eluted in assay buffers (see below) in the presence of 5 mM biotin.

Ras–Raf-RBD binding assay

Prior to the barcoding, the *E. coli* expressed and purified Ras protein was saturated with a non-hydrolyzable GTP analog, Gpp(NH)p, by EDTA-enhanced nucleotide exchange as previously described[35]. 2 nM mixed Raf-RBD WT and mutants displayed on PRMC complexes were incubated with different concentrations of barcoded Ras in 50 mM HEPES, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1% (v/v) Tween 20 and 0.5 mM Gpp(NH)p for 1 h. After reaching equilibrium, Ras–Raf-RBD complexes were crosslinked with 0.5 mM $BS(PEG)_5$ at 4°C for 1 h. The reaction was quenched by adding Tris-HCl, pH 8.0 to a final concentration of 50 mM. As unbound Ras can contribute to random Raf-RBD polony colocalization, it was removed by HA-tag affinity purification to enrich HA-tagged Raf-RBD and bound Ras. Thus, the samples were incubated with anti-HA magnetic beads (Thermo Scientific) at 4°C for ~2 h and eluted with an array deposition buffer (20 mM

HEPES, pH 7.0, 50 mM KOAc, 6 mM $Mg(OAc)_2$, 0.25 U/mL RNasin (Promega) and 0.1% Tween 20) in the presence of 2 mg/mL HA peptide.

GPCR profiling assay

Mixed barcoded GPCRs were assayed with 100 μM alprenolol, pindolol, isoproterenol, atropine and carbachol (Sigma-Aldrich) and 100 nM xanomeline (Tocris Bioscience). The GPCR–β-arr2 binding assay was performed by adding a ligand to ~1 nM GPCR–nanodisc complexes in 20 mM HEPES, pH 7.5, 50 mM KOAc, 2 mM EDTA and 5 mM $MgCl_2$, followed by addition of 10 nM GRK2 (Life Technologies), 0.1 mM ATP, 10 nM G protein $\beta_1\gamma_2$ subunits (KeraFAST) and 5 nM barcoded β-arr2 to a total volume of 25 μL. Compounds were assayed in parallel, and reactions were incubated at 30°C for 30 min followed by the crosslinking and the HA-tag affinity purification described above. Proteins from multiple wells were pooled and analyzed on a single array.

ScFv binding profiling and immunoprecipitation assay

To diversify scFvs, error-prone PCR was performed for the ten scFv genes by using a random mutagenesis kit (Clontech Laboratories) under the condition of 3.5 mutations per 1,000 bp. Twenty mutants for each scFv were randomly picked and barcoded to construct a scFvs library. Ribosome display of the scFv library was specifically performed with the PURExpress Δ ribosome kit supplemented with disulfide bond enhancers (New England Biolabs, 4 μL of the enhancer 1 and 2 per 100 μL reaction). The binding assay was performed by incubating 200 scFvs (~5.5 nM in total) and 55 barcoded human proteins (~1.8 μM in total) in an assay buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$ and 0.1% (v/v) Tween 20) at 4°C for 4 h. Similarly as above, samples were subjected to the crosslinking and the HA-tag affinity purification.

For the immunoprecipitation assay, selected scFv genes were subcloned into pEco-CSBP to express scFv fusions bearing a C-terminal SBP tag. Proteins were *in vitro* synthesized using a PURExpress IVT kit supplemented with the disulfide bond enhancers. In each binding assay, a 10 μL IVT reaction typically containing 0.1-0.4 μM translated scFvs was incubated with 2 μL human proteins (4.6-9.5 nM) labeled by Halo-TMR in the assay buffer at 4°C for 4 h. Bound human proteins were pulled down with the streptavidin magnetic beads and analyzed by SDS-PAGE and the fluorescent gel imaging.

Array deposition

Barcoded proteins were diluted with the deposition buffer to a 10-fold deposition concentration between 0.1 and 1 nM. Because the presence of oxygen can inhibit the gel polymerization, a gel-casting solution (6.66% acrylamide/bis-acrylamide (19:1, molecular grade, Ambion) and two 5'-acrydite-modified bridge amplification primers (278 μM each) in the deposition buffer) were degassed with argon and mixed with diluted proteins by a 9:1 volume ratio in an anaerobic chamber (Coy Lab). To form a gel layer of less than 5-μm thickness, ≥ 20 μL gel-casting mix, immediately after addition of 0.1% (v/v) TEMED and 0.05% (w/v) ammonium persulfate, was applied to a glass slide surface pretreated with Bind-Silane (GE Healthcare)[7,36], and a coverslip was placed on the top of the liquid and tightly pressed to form a liquid layer evenly spreading over the glass surface. The gel was

polymerized in the chamber for 4 h. After removal of the coverslip, the slide was washed with Milli-Q H_2O and dried by a quick spin.

Polony amplification, linearization and blocking

A protein-deposited slide was assembled in a FC 81 transmission flow cell containing a 1.85-mm-thick polycarbonate flow channel (BioSurface Technologies) for polony amplification, linearization and blocking. Flow cell components including the channel, a coverslip and tubing were cleaned by sonication in 5% Contrad 70 and Milli-Q H_2O, and air dried in an AirClean PCR hood. Prior to the amplification, samples containing mRNA −cDNA hybrids can be digested with 10 U/mL RNase H (New England Biolabs) in 50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$ and 0.1% (v/v) Triton X-100 at 37°C for 20 min. Polony amplification, linearization and blocking were performed by using an adapted cluster generation protocols[10]. In brief, immobilized barcoding DNAs were subjected to 32-35 cycles of isothermal bridge amplification at 60°C. For each cycle, the flow cell was washed with deionized formamide (Ambion) and an amplification buffer (20 mM Tris-HCl, pH 8.8, 10 mM ammonium sulfate, 2 mM magnesium sulfate, 0.1% (v/v) Triton X-100, 1.3% (v/v) DMSO and 2 M betaine) and incubated with 200 μM dNTPs and 80 U/mL Bst polymerase (New England Biolabs) in the amplification buffer for 5 min. Resulted double-stranded polonies were linearized by adding 10 U/mL USER enzyme (New England Biolabs) in a linearization buffer (20 mM Tris-HCl, pH 8.8, 10 mM KCl, 10 mM ammonium sulfate, 2 mM magnesium sulfate and 0.1% (v/v) Triton X-100) followed by incubating the flow cell at 37°C for 1 h. Excised strands were eluted with a wash buffer W1 (1×SSC and 70% formamide). Exposed 3'-OH ends of polonies and primers were blocked by adding 250 U/mL terminal transferase (New England Biolabs) and 10 μM ddNTPs in a blocking buffer (20 mM Tris-acetate, pH 7.9, 50 mM KOAc, 10 mM $Mg(OAc)_2$ and 0.25 mM $CoCl_2$) followed by incubating the flow cell at 37°C for 10 min.

DNA sequencing

Linearized and 3'-OH blocked polonies were analyzed by hybridization with fluorescently labeled oligos, SBE or sequencing-by-ligation as previously described[11,36]. The assays can be performed within the flow cell or a gasket chamber assembled with the polony slide taken out of the flow cell and a microarray gasket slide (Agilent Technologies). Polonies were probed with oligos or dideoxynucleotides (PerkinElmer) labeled by fluorescein/FAM, Cy3/Ty563 or Cy5/Ty665 and subjected to three-color fluorescence imaging. In brief, the hybridization was performed by incubating polonies with oligos (2 μM each) in a hybridization buffer (5×SSC and 0.1% (v/v) Tween 20) at 60°C for 10 min followed by decreasing the temperature to 40°C and washing off unbound oligos with a wash buffer W2 (0.3×SSC and 0.1% (v/v) Tween 20). The SBE was performed by incubating primer-bound polonies with fluorescently labeled ddNTPs (1 μM each) and 0.32 U/μl Thermo Sequenase (GE Healthcare) in 26 mM Tris-HCl, pH 9.5, 6.5 mM $MgCl_2$ and 0.05% (v/v) Tween 20 at 60°C for 5 min, followed by washing off excess ddNTPs with the wash buffer W2. For each sequencing-by-ligation cycle, polonies were hybridized to a sequencing primer and probed with a query oligo set (fluorescent nonamers, 2 μM each subpool) in a ligation buffer (50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM ATP and 5 mM DTT) in the presence of 30 U/μl T4 DNA ligase (Enzymatics). The ligation was incubated at room temperature for 20

min followed by increasing the temperature and maintaining it at 35°C for 40 min. Prior to a next cycle, hybridized primers were stripped with the buffer W1 at 60°C. To facilitate the deconvolution of colocalized polonies from two protein libraries, each library was separately sequenced using a different sequencing primer.

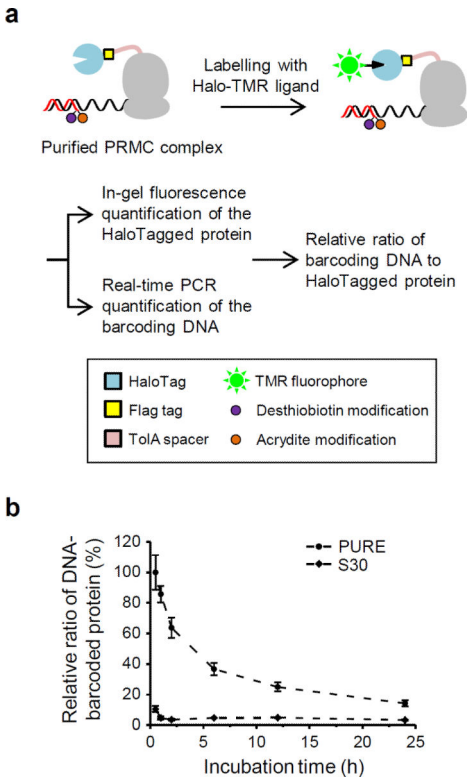
Image acquisition, processing and base calling

Fluorescence imaging was conducted with a Leica AM TIRF MC system including a DMI6000 B inverted microscope, a motorized scanning stage and a Hamamatsu C9100-02 electron multiplying CCD camera (1000 × 1000 pixels, Hamamatsu Photonics). Polony images were acquired under an epi-illumination mode by using a 20× objective (HCX PL Fluotar L, N.A. 0.40, Leica) or 40× (HCX PL APO, N.A. 0.85, Leica) and from three channels (fluorescein, Cy3 and Cy5) using 488, 561 and 635 nm lasers and excitation–emission filter pairs (490/20–525/50, 552/24–605/65 and 635/10–720/60, respectively). Raw images were exported by LAS AF Lite software (Leica) and processed using ImageJ and MATLAB (R2011a) scripts to remove background fluorescence and exclude small-size impurities and large-scale structures. Image analyses and base calling were conducted similarly as previously described[11]. In brief, MATLAB scripts were applied to identify polony coordinates by finding local maxima or weighted centroids, construct a reference image containing all detected polonies by super-imposing images taken in the first cycle, and then align images from later cycles to the reference image. Thus, a set of fluorescence values for each acquisition cycle as well as the coordinates were obtained for polony identification and the colocalization analysis. No more than 5 sequencing cycles were required to analyze each library used in this work.

Colocalization analysis and statistics

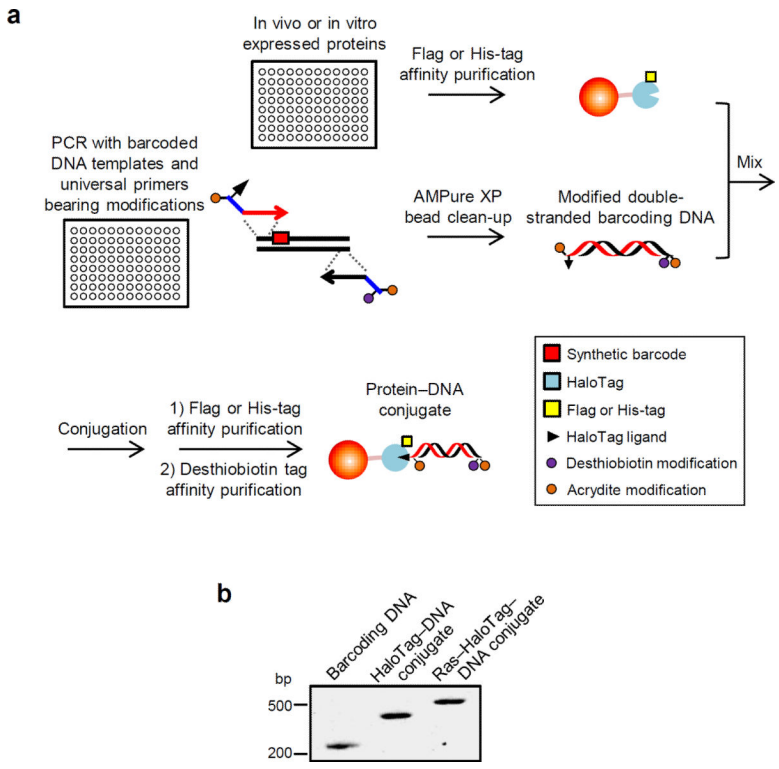
To align reference images for protein and probe libraries, polonies were hybridized with both sequencing primers labeled by Cy3 or Cy5, and then their images were super-imposed to generate a cross-library reference. MATLAB scripts calculated the offset of reference images generated from two sequencing rounds, measured distances between all polony positions identified from the two libraries, and compared them to a defined threshold to determine the colocalization. We considered a polony exclusion effect usually observed for competitive coamplification of colocalized templates[37,38], and set an optimized threshold distance to be 0.7 μm. Total and colocalized polony numbers were computed for each paired polony species at each imaging position. Colocalization statistics were calculated using Student's t-tests based on measurements at all imaged positions. In addition, a pair cross-correlation function (PCCF) statistic[39] was applied to analyze polony colocalization patterns (Supplementary Notes).

Extended Data



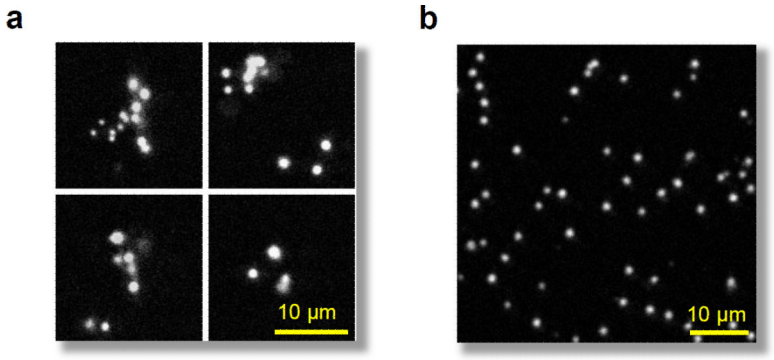
Extended Data Figure 1. Improved stability of PRMC complexes generated in a reconstituted *E. coli* IVT (PURE) system

a, Analysis of PRMC complex stability by measuring the relative ratio of barcoding DNA to HaloTagged protein. **b**, Comparison of PRMC complex stabilities in the PURE and an *E. coli* crude extract (S30) IVT system (5 PRIME). Nucleic acid degradation or ribosome dissociation can result in the loss of barcoding DNAs. IVT reactions were performed at 37°C for 30 min and PRMC complexes were further incubated at room temperature for indicated periods of time before the affinity purification and the stability analysis. Means of three independent experiments ± standard deviations.



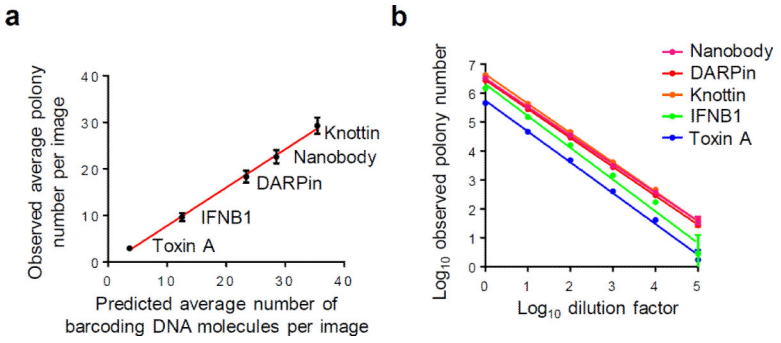
Extended Data Figure 2. HaloTag-based protein–DNA conjugation

a. Schematic of the individual barcoding method adaptable to an automatic platform. Fusion proteins bearing an N or C-terminal HaloTag and the affinity tags were purified and conjugated to a barcoding DNA bearing three modifications. **b.** Agarose gel electrophoresis of the barcoding DNA and selected protein–DNA conjugates.



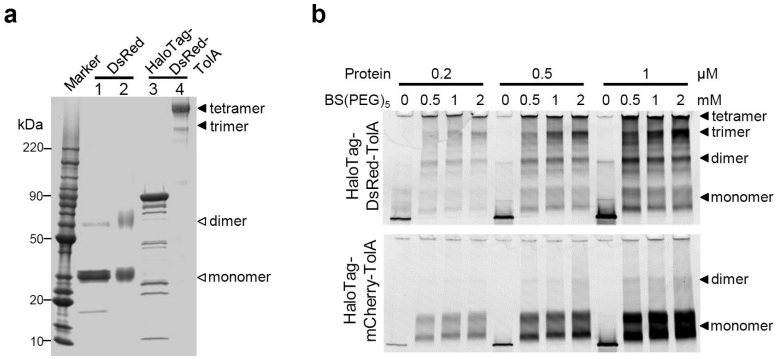
Extended Data Figure 3. Covalent immobilization of barcoding DNAs is required for in situ polony amplification

Representative images of polonies amplified from barcoding DNA templates without (**a**) or with (**b**) 5’-acrydite modifications. Oversized polonies or polony clusters shown in (**a**) were resulted from template-drifting-induced multiple seeding events during the amplification.



Extended Data Figure 4. Polony quantification of various barcoded proteins

a, Plot showing the average number of polonies detected at a single imaging position vs. the average number of barcoding DNA templates predicted by real-time PCR quantification. Data represent mean values of 100 measurements; error bars, 95% CL. **b,** Log–log plot of total numbers of polonies detected vs. dilution factors. Data represent mean values of two technical replicates.



Extended Data Figure 5. Crosslinking efficiency of DsRed is improved by a lysine-rich TolA domain

a, SDS-PAGE analysis of purified DsRed (Clontech Laboratories) and HaloTag–DsRed–TolA proteins before (lanes 1 and 3) and after (lanes 2 and 4) the crosslinking. 10 μM purified proteins were crosslinked by 1 mM $BS(PEG)_5$ in 20 mM HEPES buffer, pH 8.0, 150 mM KOAc at 4°C for 1 h. Proteins were stained with Coomassie blue. Only a minor band of the crosslinked dimer was observed for DsRed (lane 2); in contrast, HaloTag–DsRed–TolA was all crosslinked as a tetramer or a trimer (lane 4). Co-purified *E. coli* proteins (some protein bands below the major band in the lane 3), likely bound to TolA during the purification, and degradation products (due to the hydrolysis of an acylimine bond in a DsRed chromophore) were efficiently crosslinked to HaloTag–DsRed–TolA. **b,** Comparison of HaloTagged DsRed and mCherry, a monomeric fluorescence protein, crosslinked at different conditions. Proteins labeled with Halo-TMR were analyzed by fluorescent gel imaging. Only a minor fraction of HaloTag–mCherry–TolA, a control to show non-specific crosslinking, was crosslinked at increased protein concentrations. Intramolecularly crosslinked proteins show multiple bands or smears corresponding to different quaternary structures of the multidomain proteins stabilized by crosslinking. Bands of the crosslinked trimers show an increased intensity at higher $BS(PEG)_5$ concentrations

probably because the primary amine groups on surface are more quickly modified by $BS(PEG)_5$, thus preventing the further crosslinking to form the tetramer.

Extended Data Table 1

Polony quantification of titrated binder proteins and antigens

Name	Details	Protein length (a. a.)	Observed polonies for different dilutions					
			10×	100×	1000×	10,000×	100,000×	1000,000×
Scfv	Single-chain variable fragment (Oportuzumab)	254	3,457,566	354,491	37,034	3,343	342	43
			3,783,673	384,958	40,638	3,595	415	41
Nanobody	Single-domain antibody (Caplacizumab)	260	3,139,771	319,974	33,354	2,932	394	31
			3,419,354	349,061	36,785	3,257	345	49
Adnectin	Engineered the tenth fibronectin type III domain, CT-322	103	3,386,196	346,682	36,263	3,238	378	28
			3,722,036	379,665	39,761	3,745	391	30
Affibody	Engineered Z domain of protein A, ABY-025	59	3,904,925	398,140	41,749	3,773	373	41
			4,307,520	440,476	45,790	4,335	509	25
DARPin	Engineered ankyrin repeat proteins, MP0112	136	2,632,662	267,989	28,038	2,563	265	30
			2,893,576	295,016	30,697	2,990	327	25
Anticalin	Engineered lipocalin, PRS-050	155	2,579,188	263,339	28,315	2,520	279	26
			2,830,314	289,320	30,149	2,704	269	46
Knottin	Engineered cystine-knot miniprotein, 2.5D	34	3,990,730	406,311	42,426	3,885	468	41
			4,407,769	448,907	47,177	4,174	452	26
TUBEs	Tandem ubiquitin-binding entities, ubiquilin-1	252	2,311,215	235,068	24,739	2,339	288	18
			2,525,796	258,409	27,117	2,429	286	31
HB36	Computationally designed hemagglutinin binding protein	93	2,391,775	242,272	25,598	2,322	255	37
			2,622,536	267,086	28,051	2,662	275	23
GP120	The outer domain of HIV envelope glycoprotein GP120	192	1,693,019	172,836	18,255	1,671	169	14
			1,854,037	188,939	19,322	1,878	182	10
IFNB1	Human interferon beta 1	166	1,396,290	142,750	15,082	1,390	151	8
			1,527,030	155,242	16,298	1,492	189	1
Toxin A	*Clostridium difficile* toxin A receptor-binding domain	1048	441,523	45,073	4,717	439	37	3
			461,514	47,540	4,784	364	46	1

Extended Data Table 2

Lists of scFvs and human proteins used in the binding profiling

ScFv #	Original mAb	Construction method	Binding target
1-20	Anrukinzumab	V_H-linker-V_L	IL13
21-40	Ustekinumab	V_H-linker-V_L	IL12B
41-60	Canakinumab	V_H-linker-V_L	IL1B
61-80	Conatumumab	V_H-linker-V_L	TNFRSF10B
81-100	Ibalizumab	VH-linker-VL	CD4
101-120	Oportuzumab	V_H-linker-V_L	CD326
121-140	Otelixizumab	V_H-linker-V_L	CD3E

ScFv #	Original mAb	Construction method	Binding target
141-160	Ranibizumab	V_H-linker-V_L	VEGFA
161-180	Siltuximab	V_H-linker-V_L	IL6
181-200	Tanezumab	V_H-linker-V_L	NGF

Human protein #	Entrez Gene symbol	Entrez Gene ID	Protein length (a.a.)
1	IL1B	3553	270
2	IL2	3558	154
3	IL3	3562	153
4	IL4	3565	154
5	IL6	3569	212
6	IL7	3574	178
7	IL8	3576	100
8	IL9	3578	145
9	IL10	3586	179
10	IL12B	3593	328
11	IL13	3596	147
12	IL15	3600	163
13	IL17A	3605	156
14	IL17B	27190	181
15	IL17F	112744	164
16	IL18	3606	194
17	IL20	50604	177
18	IL25	64806	178
19	IL26	55801	172
20	IL28A	282616	201
21	IL29	282618	201
22	IL31	386653	165
23	CD1A	909	312
24	CD1E	913	334
25	CD3E	916	208
26	CD4	920	459
27	CD7	924	241
28	CD27	939	261
29	CD38	952	301
30	CD40	958	278
31	CD52	1043	62
32	CD58	965	241
33	CD59	966	129
34	CD79B	974	230
35	CD74	972	233
36	CD80	941	289
37	CD207	50489	329
38	CD247	919	165

Human protein #	Entrez Gene symbol	Entrez Gene ID	Protein length (a.a.)
39	CD274	29126	291
40	CD302	9936	171
41	CD320	51293	283
42	CD326	4072	314
43	CNTF	1270	201
44	CSF2	1437	145
45	CSF3	1440	201
46	NGF	4803	241
47	NODAL	4838	348
48	OSM	5008	253
49	THPO	7066	350
50	VEGFA	7422	237
51	FASLG	356	282
52	LTA	4049	206
53	TNF	7124	234
54	TNFRSF10B	8795	440
55	TNFRSF13B	23495	248

Extended Data Table 3

Comparison of protein interaction profiling technologies based on nucleic acid barcoding and high-throughput sequencing.

	QIS-Seq and Stitch-Seq	Ig-seq	PhIP-Seq	PLATO and IVV-HiTSeq	ProteinSeq and IDUP	SMI-Seq
Principle and references*	Y2H coupled with NGS[27,52]	B cell-based antibody screeningcoupled with NGS[92]	Phage display coupled with NGS[51]	Ribosome and mRNA displays, respectively, coupled with NGS[16,53]	Proximity ligation and extension assays, respectively, coupled with NGS[30,50]	In situ polony sequencing of DNA-barcoded proteins
Choice of proteins	Wide	Immunoglobulin	Limited	Limited	Wide	Wide
Barcodes	CDS	CDS	CDS	CDS	Synthetic (non-CDS)	Synthetic (non-CDS)
Barcoding method	Compartmentation	Compartmentation	Compartmentation	Molecular attachment	Molecular attachment	Molecular attachment
Difficulty of preparing protein library	Moderate	Simple	Moderate	Moderate	High	Moderate to high
Interaction context	Intracellular	Probe fluorescently labelled or immobilized	Bait immobilized	Bait immobilized	In solution	In solution
Separation of interacting and non-interacting proteins	Selection in medium or flow cytometry sorting	Flow cytometry sorting or affinity enrichment	Affinity enrichment	Affinity enrichment	No	No
Protein fraction(s) to be detected	Bound (positive clones)	Bound	Bound	Bound	Bound (ligated)	Bound (crosslinked) and unbound
Barcode amplification	PCR	PCR	PCR	PCR	PCR	In situ SM amplification
Barcode sequencing	NGS	NGS	NGS	NGS	NGS	In situ polony sequencing
Quantitative measure of binding affinity	Difficulties in measuring effective protein concentration	Difficulties in measuring effective protein concentration	Difficulties in quantifying bound and free proteins	Difficulties in quantifying bound and free proteins	Difficulties in quantifying bound and free proteins	Polony colocalization
Cost per binding assay	Low medium	Low medium	Low medium	Low	Low	Low
Library vs. library screening	Limited throughput due to individual barcode pairing[27]	No	No	No	Demonstrated at a scale of 5 by 5[14]	Demonstrated at a scale of 200 by 55**
Applications in compound screening and references*	Demonstrated with unlabeled[54,55] and labeled compounds[56]	No	Demonstrated with barcoded peptides[57,58]	Demonstrated with bead-immobilized compounds[16]	Demonstrated with unlabeled[60] and barcoded compounds[30]	Demonstrated at a scale of 6 by 3 with unlabeled compounds

CDS: protein coding sequence; NGS: next-generation sequencing

*References can be found in Supplementary Information

**Mathematical model suggests that $\geq 10^5 \times 10^5$ interactions may be analyzed in a single binding assay (refer to Supplementary Notes)

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a grant from the US Department of Energy (DE-FG02-02ER63445) to G.M.C. and a grant from the NIH NHGRI (HG001715) to M.V. and D.E.H. L.G. was supported by a postdoctoral fellowship from the Jane Coffin Childs Fund for Medical Research and a grant from Harvard Origins of Life Initiative. We thank W. Harper, L. Pontano Vaites, S. Elledge and J. Zhu for providing plasmids, F. Vigneault for comments on the manuscript, J. Lai and D. Breslau for assistance on imaging setup, and members of the Church and Vidal groups for constructive discussions.

References

1. Shendure J, Ji H. Next-generation DNA sequencing. Nat. Biotechnol. 2008; 26:1135–1145. [PubMed: 18846087]
2. Dreze M, et al. High-quality binary interactome mapping. Methods Enzymol. 2010; 470:281–315. [PubMed: 20946815]
3. MacBeath G, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. Science. 2000; 289:1760–1763. [PubMed: 10976071]
4. Gavin AC, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. 2002; 415:141–147. [PubMed: 11805826]
5. Weiss S. Fluorescence spectroscopy of single biomolecules. Science. 1999; 283:1676–1683. [PubMed: 10073925]
6. Hanes J, Pluckthun A. In vitro selection and evolution of functional proteins by using ribosome display. Proc. Natl. Acad. Sci. U.S.A. 1997; 94:4937–4942. [PubMed: 9144168]
7. Mitra RD, Church GM. In situ localized amplification and contact replication of many individual DNA molecules. Nucleic Acids Res. 1999; 27:e34. [PubMed: 10572186]
8. Shimizu Y, et al. Cell-free translation reconstituted with purified components. Nat. Biotechnol. 2001; 19:751–755. [PubMed: 11479568]
9. Los GV, et al. HatoTag: A novel protein labeling technology for cell imaging and protein analysis. ACS Chem. Biol. 2008; 3:373–382. [PubMed: 18533659]
10. Bentley DR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008; 456:53–59. [PubMed: 18987734]
11. Shendure J, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 2005; 309:1728–1732. [PubMed: 16081699]
12. Aird D, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 2011; 12
13. Fredriksson S, et al. Protein detection using proximity-dependent DNA ligation assays. Nat. Biotechnol. 2002; 20:473–477. [PubMed: 11981560]
14. Hammond M, Nong RY, Ericsson O, Pardali K, Landegren U. Profiling cellular protein complexes by proximity ligation with dual tag microarray readout. PLoS One. 2012; 7
15. Lundberg M, Eriksson A, Tran B, Assarsson E, Fredriksson S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. Nucleic Acids Res. 2011; 39:e102. [PubMed: 21646338]
16. Zhu J, et al. Protein interaction discovery using parallel analysis of translated ORFs (PLATO). Nat. Biotechnol. 2013; 31:331–334. [PubMed: 23503679]
17. Vetter IR, Wittinghofer A. Signal transduction-The guanine nucleotide-binding switch in three dimensions. Science. 2001; 294:1299–1304. [PubMed: 11701921]
18. Block C, Janknecht R, Herrmann C, Nassar N, Wittinghofer A. Quantitative structure-activity analysis correlating Ras/Raf interaction in vitro to Raf activation in vivo. Nat. Struct. Biol. 1996; 3:244–251. [PubMed: 8605626]

19. Kiel C, et al. Improved binding of raf to Ras.GDP is correlated with biological activity. J. Biol. Chem. 2009; 284:31893–31902. [PubMed: 19776012]
20. Overington JP, Al-Lazikani B, Hopkins AL. Opinion-How many drug targets are there? Nat. Rev. Drug Disc. 2006; 5:993–996.
21. Zhang R, Xie X. Tools for GPCR drug discovery. Acta Pharmacol. Sin. 2012; 33:372–384. [PubMed: 22266728]
22. Denisov IG, Grinkova YV, Lazarides AA, Sligar SG. Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size. J. Am. Chem. Soc. 2004; 126:3477–3487. [PubMed: 15025475]
23. Leitz AJ, Bayburt TH, Barnakov AN, Springer BA, Sligar SG. Functional reconstitution of β_2-adrenergic receptors utilizing self-assembling Nanodisc technology. Biotechniques. 2006; 40:601–610. [PubMed: 16708760]
24. Whorton MR, et al. A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. Proc. Natl. Acad. Sci. U.S.A. 2007; 104:7682–7687. [PubMed: 17452637]
25. Luttrell LM, Lefkowitz RJ. The role of β-arrestins in the termination and transduction of G-protein-coupled receptor signals. J. Cell Sci. 2002; 115:455–465. [PubMed: 11861753]
26. Gurevich VV, et al. Arrestin interactions with G-protein-coupled receptors - Direct binding-studies of wild-type and mutant arrestins with rhodopsin, β_2-adrenergic, and m2-muscarinic cholinergic receptors. J. Biol. Chem. 1995; 270:720–731. [PubMed: 7822302]
27. Yu H, et al. Next-generation sequencing to generate interactome datasets. Nat. Methods. 2011; 8:478–480. [PubMed: 21516116]
28. Michaud GA, et al. Analyzing antibody specificity with whole proteome microarrays. Nat. Biotechnol. 2003; 21:1509–1512. [PubMed: 14608365]
29. Kosuri S, et al. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. Nat. Biotechnol. 2010; 28:1295–U1108. [PubMed: 21113165]
30. McGregor LM, Jain T, Liu DR. Identification of ligand-target pairs from combined libraries of small molecules and unpurified protein targets in cell lysates. J. Am. Chem. Soc. 2014; 136:3264–3270. [PubMed: 24495225]
31. Rual JF, Hill DE, Vidal M. ORFeome projects: gateway between genomics and omics. Curr. Opin. Chem. Biol. 2004; 8:20–25. [PubMed: 15036152]
32. Parker EM, Kameyama K, Higashijima T, Ross EM. Reconstitutively active G protein-coupled receptors purified from baculovirus-infected insect cells. J. Biol. Chem. 1991; 266:519–527. [PubMed: 1845979]
33. Kobilka BK. Amino and carboxyl terminal modifications to facilitate the production and purification of a G protein-coupled receptor. Anal. Biochem. 1995; 231:269–271. [PubMed: 8678314]
34. Mitra N, et al. Calcium-Dependent Ligand Binding and G-protein Signaling of Family B GPCR Parathyroid Hormone 1 Receptor Purified in Nanodiscs. ACS Chem. Biol. 2013; 8:617–625. [PubMed: 23237450]
35. John J, Frech M, Wittinghofer A. Biochemical properties of Ha-ras encoded p21 mutants and mechanism of the autophosphorylation reaction. J. Biol. Chem. 1988; 263:11792–11799. [PubMed: 3042780]
36. Mitra RD, et al. Digital genotyping and haplotyping with polymerase colonies. Proc. Natl. Acad. Sci. U.S.A. 2003; 100:5926–5931. [PubMed: 12730373]
37. Mitra RD, Shendure J, Olejnik J, Edyta Krzymanska O, Church GM. Fluorescent in situ sequencing on polymerase colonies. Anal. Biochem. 2003; 320:55–65. [PubMed: 12895469]
38. Aach J, Church GM. Mathematical models of diffusion-constrained polymerase chain reactions: basis of high-throughput nucleic acid assays and simple self-organizing systems. J. Theor. Biol. 2004; 228:31–46. [PubMed: 15064081]
39. Philimonenko AA, Janacek J, Hozak P. Statistical evaluation of colocalization patterns in immunogold labeling experiments. J. Struct. Biol. 2000; 132:201–210. [PubMed: 11243889]

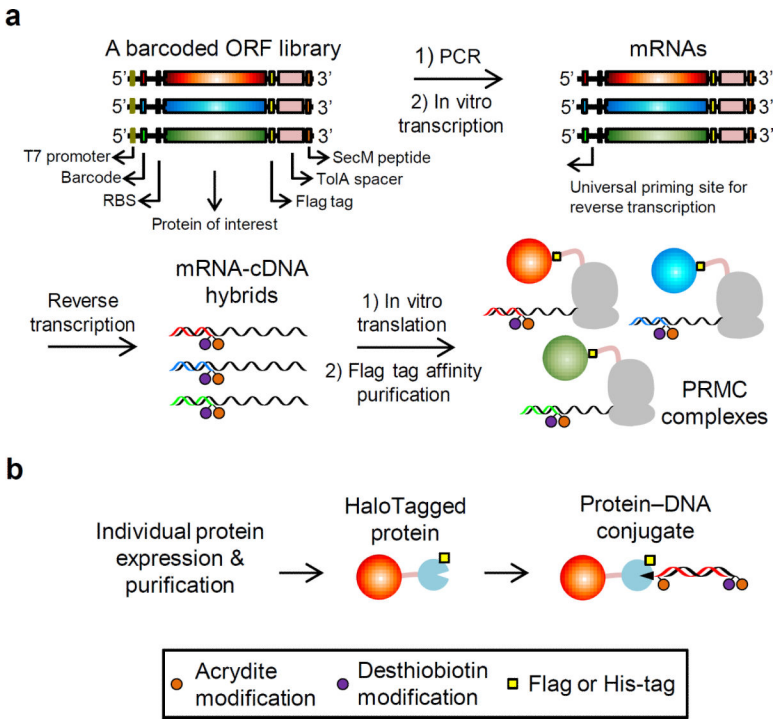


Figure 1. Schematics of protein barcoding methods

a, Collective barcoding via ribosome display. A short synthetic barcoding sequence is joined to the 5' end of DNA templates via PCR. PRMC complexes are formed via ribosome stalling triggered by a C-terminal *E. coli* SecM peptide. Displayed proteins bearing a C-terminal Flag tag are separated from the ribosomes by an *E. coli* TolA spacer domain. **b**, Individual barcoding via a HaloTag-mediated conjugation of proteins to a 220-base pair (bp) double-stranded barcoding DNA with a HaloTag ligand modification (black triangle). Modifications are introduced to barcoding DNAs by PCR with modified primers.

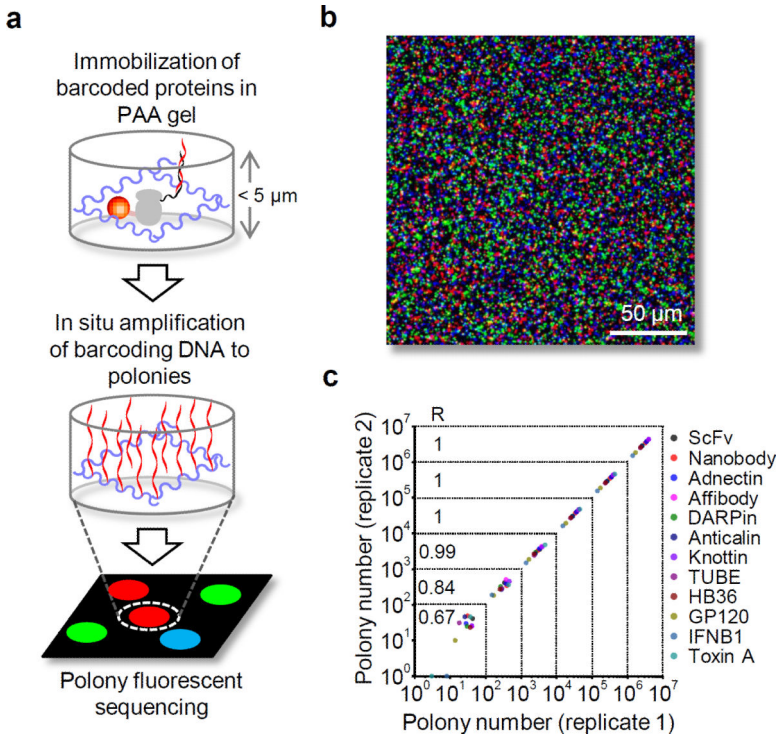


Figure 2. Amplification and quantification of barcoding DNAs

a, Schematic of in situ polony amplification and sequencing. Barcoded proteins were immobilized in a PAA gel matrix attached to a Bind-Silane treated glass slide. The slide was assembled into a flow cell, where barcoding DNAs were in situ amplified into polonies for DNA sequencing. **b**, Representative merged images of polonies hybridized with Cy5 (red), Cy3 (green) and fluorescein (blue)-labeled oligos (20× objective). **c**, Polony quantification of mixed protein binders and antigens. Pearson correlation coefficient R was calculated for different coverages grouped by dotted lines.

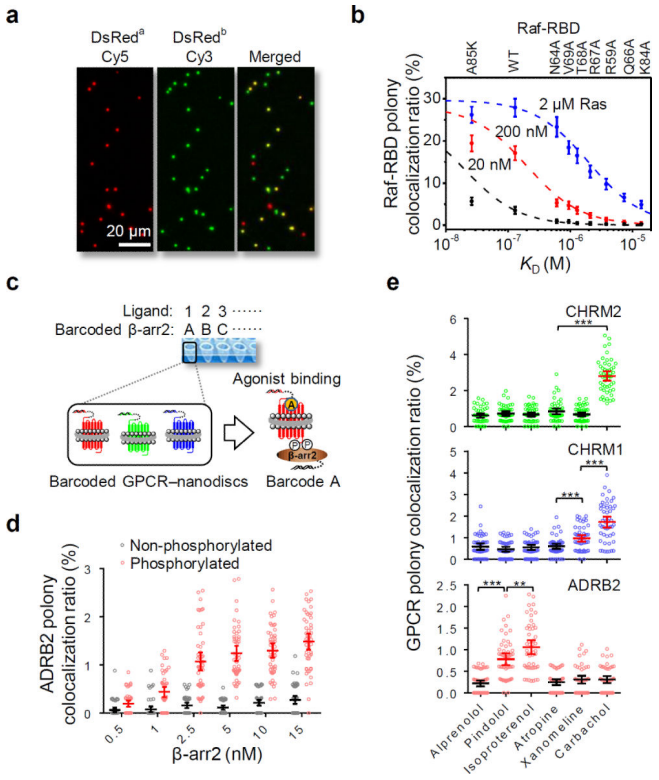


Figure 3. Analyses of protein interactions via polony colocalization

a, Interaction of DsRed subunits resulted in colocalized polonies. DsRed polonies were identified by SBE with Cy5 (red) or Cy3 (green)-labeled ddNTPs. **b,** Correlation between the polony colocalization ratios and K_Ds of Ras–Raf-RBD complexes. Means of measurements at 100 imaging positions ± 95% confidence level (CL; refer to Supplementary Table 1). Fitting equation, $R = R_{max} \times P / (K_D + P)$, where is R the predicted Raf-RBD polony colocalization ratio, R_{max} is the maximum polony colocalization ratio when Raf-RBD is saturated by Ras, and P is the Ras concentration. **c,** Schematic of multiplex GPCR screening and compound profiling by the binding assay of mixed barcoded GPCRs with barcoded β-arr2. **d**, Comparison of β-arr2 binding to isoproterenol-activated β_2-adrenergic receptor with or without GRK2-mediated phosphorylation. β-arr2 titration data were fitted by the one-site-specific model using GraphPad Prism 6. **e**, Parallel GPCR binding profiling. Data represent mean values of 50 measurements; error bars, 95% CL (refer to Supplementary Table 2). $^{**}P < 0.01$, $^{***}P < 0.001$, one-tailed paired Student's t test.

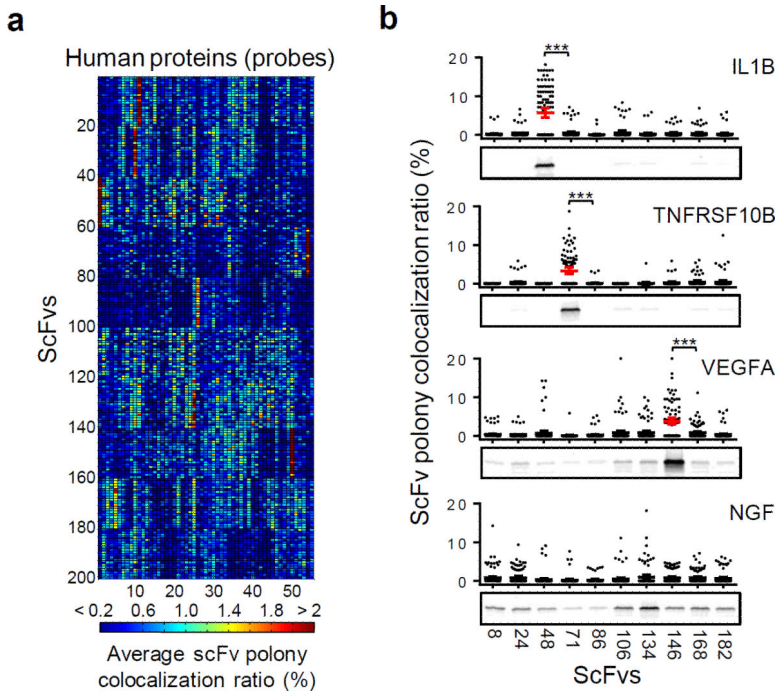


Figure 4. Parallel antibody binding profiling

a, Heat map of the mean colocalization ratios measured at 100 imaging positions (refer to Supplementary Table 3). ScFvs sharing the same origins were grouped by their numbers (Extended Data Table 2). **b,** Correlation between the polony colocalization statistics and the scFv immunoprecipitation results. For the immunoprecipitation assay, selected scFvs were fused to a C-terminal streptavidin binding peptide (SBP) tag and bound to streptavidin magnetic beads to pull down human protein probes bearing a HaloTag, which can be labeled by Halo-TMR for fluorescent gel imaging. Error bars, 95% CL, highlighted in red for specific scFv–probe binding. ***$P < 0.001$, one-tailed paired Student's t test.

Extended Data Table 3

Comparison of protein interaction profiling technologies based on nucleic acid barcoding and high-throughput sequencing.

	QIS-Seq and Stitch-Seq	Ig-seq	PhIP-Seq	PLATO and IVV-HiTSeq	ProteinSeq and IDUP	SMI-Seq
Principle and references*	Y2H coupled with NGS[27,52]	B cell-based antibody screeningcoupled with NGS[92]	Phage display coupled with NGS[51]	Ribosome and mRNA displays, respectively, coupled with NGS[16,53]	Proximity ligation and extension assays, respectively, coupled with NGS[30,50]	In situ polony sequencing of DNA-barcoded proteins
Choice of proteins	Wide	Immunoglobulin	Limited	Limited	Wide	Wide
Barcodes	CDS	CDS	CDS	CDS	Synthetic (non-CDS)	Synthetic (non-CDS)
Barcoding method	Compartmentation	Compartmentation	Compartmentation	Molecular attachment	Molecular attachment	Molecular attachment
Difficulty of preparing protein library	Moderate	Simple	Moderate	Moderate	High	Moderate to high
Interaction context	Intracellular	Probe fluorescently labelled or immobilized	Bait immobilized	Bait immobilized	In solution	In solution
Separation of interacting and non-interacting proteins	Selection in medium or flow cytometry sorting	Flow cytometry sorting or affinity enrichment	Affinity enrichment	Affinity enrichment	No	No
Protein fraction(s) to be detected	Bound (positive clones)	Bound	Bound	Bound	Bound (ligated)	Bound (crosslinked) and unbound
Barcode amplification	PCR	PCR	PCR	PCR	PCR	In situ SM amplification
Barcode sequencing	NGS	NGS	NGS	NGS	NGS	In situ polony sequencing
Quantitative measure of binding affinity	Difficulties in measuring effective protein concentration	Difficulties in measuring effective protein concentration	Difficulties in quantifying bound and free proteins	Difficulties in quantifying bound and free proteins	Difficulties in quantifying bound and free proteins	Polony colocalization
Cost per binding assay	Low medium	Low medium	Low medium	Low	Low	Low
Library vs. library screening	Limited throughput due to individual barcode pairing[27]	No	No	No	Demonstrated at a scale of 5 by 5[14]	Demonstrated at a scale of 200 by 55**
Applications in compound screening and references*	Demonstrated with unlabeled[54,55] and labeled compounds[56]	No	Demonstrated with barcoded peptides[57,58]	Demonstrated with bead-immobilized compounds[16]	Demonstrated with unlabeled[60] and barcoded compounds[30]	Demonstrated at a scale of 6 by 3 with unlabeled compounds

CDS: protein coding sequence; NGS: next-generation sequencing

*References can be found in Supplementary Information

**Mathematical model suggests that ≥ $10^5 \times 10^5$ interactions may be analyzed in a single binding assay (refer to Supplementary Notes)