LETTER

Multiplex single-molecule interaction profiling of DNA-barcoded proteins

Liangcai Gu¹, Chao Li², John Aach¹, David E. Hill^{1,3}, Marc Vidal^{1,3} & George M. Church^{1,2}

In contrast with advances in massively parallel DNA sequencing¹, high-throughput protein analyses²⁻⁴ are often limited by ensemble measurements, individual analyte purification and hence compromised quality and cost-effectiveness. Single-molecule protein detection using optical methods⁵ is limited by the number of spectrally non-overlapping chromophores. Here we introduce a single-molecularinteraction sequencing (SMI-seq) technology for parallel protein interaction profiling leveraging single-molecule advantages. DNA barcodes are attached to proteins collectively via ribosome display⁶ or individually via enzymatic conjugation. Barcoded proteins are assayed en masse in aqueous solution and subsequently immobilized in a polyacrylamide thin film to construct a random single-molecule array, where barcoding DNAs are amplified into in situ polymerase colonies (polonies)⁷ and analysed by DNA sequencing. This method allows precise quantification of various proteins with a theoretical maximum array density of over one million polonies per square millimetre. Furthermore, protein interactions can be measured on the basis of the statistics of colocalized polonies arising from barcoding DNAs of interacting proteins. Two demanding applications, G-protein coupled receptor and antibody-binding profiling, are demonstrated. SMI-seq enables 'library versus library' screening in a one-pot assay, simultaneously interrogating molecular binding affinity and specificity.

To analyse proteins in a massively parallel single-molecule format, we generated proteins that are molecularly coupled to a DNA bearing a barcoding sequence. One barcoding approach is to translate and display proteins on protein-ribosome-messenger-RNA-complementary-DNA (PRMC) complexes in vitro, 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 with purified components8 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 individual barcoding. Thus, fusion proteins were constructed with an engineered enzyme tag, HaloTag⁹, which mediates an efficient covalent conjugation to a HaloTag-ligandmodified double-stranded DNA (Fig. 1b). Our method is adaptable to a microtitre 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 of their barcodes (Fig. 2a). The proteins were immobilized into an ultrathin layer of crosslinked polyacrylamide 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 PCR, with two gel-anchored primers, was performed according to an adapted isothermal bridge amplification protocol¹⁰ in an assembled flow cell. This amplification showed a high efficiency of ~80% barcode detection (Extended Data Fig. 4a), and resulted in polonies of $\sim 1 \,\mu m$ diameter (Fig. 2b), similar to the clusters generated on an Illumina platform¹⁰. Polonies were identified by hybridization with fluorescent probes, single-base extension (SBE) or ligation-based sequencing¹¹.

To test the accuracy of our method, we selected nine immunoglobulin and non-immunoglobulin binding proteins and three antigens (for example, human, bacterial and viral proteins) of a molecular weight ranging 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 single-molecule quantification. Barcode detection efficiencies of different proteins were found to be almost identical at various concentrations (Extended Data Fig. 4). The *in situ* single-molecule quantification can avoid PCR amplification bias¹² and shows high reproducibility; the Pearson correlation coefficient *r* was above 0.99 when over 1,000 protein polonies were detected (Fig. 2c). Because proteins were highly diluted (at less than picomolar concentrations) before 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 extension¹⁵. Here, direct observation and counting of single-molecule protein complexes should



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 carboxy-terminal *Escherichia coli* SecM peptide. Displayed proteins bearing a C-terminal Flag tag are separated from the ribosomes by an *E. coli* TolA spacer domain. RBS, ribosomal binding site. **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.

¹Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, USA. ³Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA.

RESEARCH LETTER



Figure 2 | **Amplification and quantification of barcoding DNAs.** a, Schematic of *in situ* polony amplification and sequencing. Barcoded proteins were immobilized in a polyacrylamide (PAA) gel matrix attached to a Bind-Silane-treated glass slide. The slide was assembled into a flow cell, where barcoding DNAs were amplified *in situ* into polonies for DNA sequencing. b, Representative merged images of polonies hybridized with Cy5 (red), Cy3 (green) and fluorescein (blue)-labelled oligonucleotides (×20 objective magnification). c, Polony quantification of mixed protein binders and antigens. The Pearson correlation coefficient *r* was calculated for different coverages grouped by dotted lines.

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 co-amplified into overlapping polonies (Fig. 3a), providing a solid basis for further applications.

Unlike other methods that only detect affinity-enriched proteins (for example, PLATO¹⁶), 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 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 (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_d values) ranging from nanomolar to micromolar values^{18,19} (Fig. 3b and Supplementary Table 1), and fitted by using a one-sitespecific binding model (dashed curves). The fitted and observed average



Figure 3 | Analyses of protein interactions via polony colocalization. a, Interaction of DsRed subunits resulted in colocalized polonies. Polonies of the differently barcoded subunits, DsRed^a and DsRed^b, were identified by SBE with, respectively, Cy5 (red) or Cy3 (green)-labelled dideoxynucleotide triphosphates (ddNTPs). b, Correlation between the polony colocalization ratios and K_d values of Ras-Raf-RBD complexes. Means of measurements at 100 imaging positions \pm 95% confidence level (CL; refer to Supplementary Table 1). Fitting equation, $R = R_{\text{max}} \times P/(K_{\text{d}} + P)$, where R is 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. WT, wild type. 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 β₂-adrenergic receptor with or without GRK2-mediated phosphorylation. Titration data of β -arr2 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, highlighted in red for agonists (refer to Supplementary Table 2). **P < 0.01, ***P < 0.001, one-tailed paired Student's t-test.

colocalization ratios show relatively high agreement (r > 0.96), except for the A85K mutant which displayed significantly lower experimental values than predicted by the model, probably owing to the disruption of Lys 85-mediated interactions¹⁹ by the crosslinking. Therefore, this method could be useful for high-throughput screening of protein binding affinities.

As a first high-throughput screening application, we investigated small molecule-mediated protein-protein interactions. An advantage of our method over traditional solid-phase techniques such as protein microarrays³ is that we store and assay proteins in an aqueous solution. To exploit this, we decided to address the challenges in screening G-protein coupled receptors (GPCRs), the largest membrane protein family and premier drug targets²⁰. Current GPCR-ligand screening techniques mainly rely on cell-based assays²¹, which are subject to limitations such as the heterogeneous nature of the samples, the presence of other cellular components that can cause false positives or negatives, and limited miniaturization and multiplexing capability (for example, one receptor per assay). To prepare a homogeneous single-molecule GPCR sample compatible with our approach, receptors were stabilized in phospholipid bilayer nanodiscs²² 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 receptors²⁵.

A compound library can be screened in multi-well 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; 3to 11-fold increases (largest P = 0.002) of the average colocalization ratios after phosphorylation were observed. 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 radiolabelled β -arr2²⁶.

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 $P < 2.7 \times 10^{-10}$). Moreover, different types of agonists can be distinguished by comparing their polony colocalization ratios, for example, 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; $P < 3 \times 10^{-6}$). 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 (for example, yeast two-hybrid systems²) for library versus library screening are cell-based and require pairing both genes from two libraries to identify positive clones by performing individual PCR reactions²⁷. To demonstrate this capability, we prototyped a test of a demanding application, the binding profiling of an antibody repertoire. The screening of natural or semisynthetic monoclonal antibody 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 test²⁸, 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 ribosomedisplayed single-chain variable fragments (scFvs) and 55 human proteins including cytokines, growth factors and receptors synthesized *in vitro* (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 microchip²⁹. We sequenced ~0.64 million polonies and measured the colocalization ratios for 11,000 scFv-target protein (probe) pairs at 100 imaging positions (Fig. 4a and Supplementary Table 3). Of 200 scFvs, 147 were found with the highest colocalization ratios, 95 of which are significantly above the second highest (P < 0.05), and thus the highest specificity, to their predicted



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 tag and bound to streptavidin magnetic beads to pull down human protein probes bearing a HaloTag, which can be labelled by Halo- tetramethylrhodamine (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.

targets; the others failed probably because the construction of scFv fragments and mutations inhibit target binding. Substantial cross-reactivity can be sensitively detected, for example, 3,474 scFv–probe pairs showed tenfold 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 probe protein bands (Fig. 4b). Moreover, to further assess multiplexing potential, we developed a mathematical model that integrated parameters including K_d values 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 quantifiably analysed within a single assay.

The protein barcoding requirement imposes limitations on SMIseq. First, it cannot directly analyse proteins from biological samples. However, non-barcoded proteins can be detected in a similar fashion by using barcoded antibodies or aptamers as part of 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-synthesized²⁹ or random barcoding sequences to an open reading frame (ORF) library by a single PCR reaction and later matching them to ORF sequences by next-generation sequencing. SMI-seq enables single-molecule counting of proteins and complexes in situ, fundamentally improving sensitivity, accuracy and multiplexity (Extended Data Table 3 and Supplementary Discussion), and thus the demonstrated applications are difficult or impossible to perform with other highthroughput techniques (for example, PLATO). It is readily adaptable to industrial next-generation sequencing platforms and translated into

many applications. In addition to natural and recombinant proteins, it will be applicable to *de novo* proteins (for example, with unnatural amino acids or modifications), nucleic acids and barcoded small molecules³⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 February; accepted 8 August 2014. Published online 21 September 2014.

- 1. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nature Biotechnol.* **26**, 1135–1145 (2008).
- Dreze, M. et al. High-quality binary interactome mapping. Methods Enzymol. 470, 281–315 (2010).
- MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760–1763 (2000).
- Gavin, A. C. et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- 5. Weiss, S. Fluorescence spectroscopy of single biomolecules. *Science* **283**, 1676–1683 (1999).
- Hanes, J. & Pluckthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl Acad. Sci. USA* 94, 4937–4942 (1997).
- Mitra, R. D. & Church, G. M. *In situ* localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res.* 27, e34–e39 (1999).
 Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components.
- Nature Biotechnol. 19, 751–755 (2001).
 Los G, V. et al. HatoTag: A novel protein labeling technology for cell imaging and
- Los, G. V. et al. HatoTag: A novel protein labeling technology for cell imaging and protein analysis. ACS Chem. Biol. 3, 373–382 (2008).
- Bentley, D. R. et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59 (2008).
- Shendure, J. et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309, 1728–1732 (2005).
- 12. Aird, D. et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, R18 (2011).
- Fredriksson, S. et al. Protein detection using proximity-dependent DNA ligation assays. Nature Biotechnol. 20, 473–477 (2002).
- Hammond, M., Nong, R. Y., Ericsson, O., Pardali, K. & Landegren, U. Profiling cellular protein complexes by proximity ligation with dual tag microarray readout. *PLoS ONE* 7, e40405 (2012).
- 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.* 39, e102 (2011).
- Zhu, J. et al. Protein interaction discovery using parallel analysis of translated ORFs (PLATO). Nature Biotechnol. **31**, 331–334 (2013).
- Vetter, I. R. & Wittinghofer, A. The guanine nucleotide-binding switch in three dimensions. Science 294, 1299–1304 (2001).
- 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. Nature Struct. Biol.* 3, 244–251 (1996).

- Kiel, C. et al. Improved binding of Raf to Ras GDP is correlated with biological activity. J. Biol. Chem. 284, 31893–31902 (2009).
- Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? Nature Rev. Drug Discov. 5, 993–996 (2006).
- Zhang, R. & Xie, X. Tools for GPCR drug discovery. Acta Pharmacol. Sin. 33, 372–384 (2012).
- Denisov, I. G., Grínkova, Y. V., Lazarides, A. A. & Sligar, S. G. Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size. J. Am. Chem. Soc. 126, 3477–3487 (2004).
- Leitz, A. J., Bayburt, T. H., Barnakov, A. N., Springer, B. A. & Sligar, S. G. Functional reconstitution of β₂-adrenergic receptors utilizing self-assembling nanodisc technology. *Biotechniques* 40, 601–612 (2006).
- Whorton, M. R. et al. A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. Proc. Natl Acad. Sci. USA 104, 7682–7687 (2007).
- Luttrell, L. M. & Lefkowitz, R. J. The role of β-arrestins in the termination and transduction of G-protein-coupled receptor signals. J. Cell Sci. 115, 455–465 (2002).
- Gurevich, V. V. et al. Arrestin interactions with G-protein-coupled receptors Direct binding-studies of wild-type and mutant arrestins with rhodopsin, β₂-adrenergic, and m2-muscarinic cholinergic receptors. J. Biol. Chem. 270, 720–731 (1995).
- Yu, H. et al. Next-generation sequencing to generate interactome datasets. Nature Methods 8, 478–480 (2011).
- Michaud, G. A. et al. Analyzing antibody specificity with whole proteome microarrays. Nature Biotechnol. 21, 1509–1512 (2003).
- Kosuri, S. et al. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. Nature Biotechnol. 28, 1295–1299 (2010).
- McGregor, L. M., Jain, T. & Liu, D. R. Identification of ligand-target pairs from combined libraries of small molecules and unpurified protein targets in cell lysates. J. Am. Chem. Soc. **136**, 3264–3270 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements 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.

Author Contributions L.G. and G.M.C. conceived the technique; L.G. and C.L. performed experiments and analysed data; J.A. built the mathematical model and assisted the colocalization analyses; D.E.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.

Author Information MATLAB scripts for imaging analyses, colocalization statistics and mathematical modelling can be found at http://arep.med.harvard.edu/SMI-Seq/. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.G. (liangcaigu@gmail.com) or G.M.C. (gchurch@genetics.med.harvard.edu).

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 clones³¹ (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 (≤ 200 in this work), synthetic barcoding sequences were introduced to DNA templates via individual PCRs. Barcoded linear DNA templates were pooled and transcribed in vitro 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 of each deoxynucleotide triphosphate (dNTP), 10 U µl⁻¹ SuperScript III, 2 U µl⁻¹ RNaseOUT (Invitrogen) and 5 mM dithiothreitol (DTT) in a buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, and 5 mM MgCl₂) 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 μ I IVT reaction with \sim 0.40 μ M mRNAcDNA hybrids and \sim 0.30 μ M ribosome was incubated at 37 $^{\circ}$ 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⁻¹ RNasin (Promega), 0.5 mg ml⁻¹ chloramphenicol, 5 mM 2-mercaptoethanol and 0.1% (v/v) Tween 20) and centrifuged (14,000g, 4 °C) for 10 min to remove insoluble components. PRMC complexes, always kept on ice or in a 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⁻¹ yeast transfer RNA and 10 mg ml^{-1} BSA. The bound proteins were eluted with the buffer HKM containing 0.1 mg ml^{-1} Flag peptide or 5 mM biotin, and their barcoding DNAs were quantified 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 nonspecifically 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 N₂ 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 purification³⁴. In brief, GPCR genes were synthesized and inserted into a pBac-NFlagHA vector for the expression of the fusion proteins bearing amino-terminal Flag and hemagglutinin (HA) epitope tags and a HaloTag domain. Cells were harvested at 2 days after transfection, homogenized in a 50 mM Tris-HCl, pH 7.4, 50 mM NaCl and 1 mM EDTA with a protease inhibitor 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 CaCl₂, 5 mM MgCl₂, 2 mM EDTA, 10% glycerol, 1% N-dodecyl- β -D-maltopyranoside (DDM) and a protease inhibitor cocktail (Set III, EMD Biosciences), and centrifuged at 15,000g 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 CaCl₂, 5 mM MgCl₂, 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 $^{-1}$ Flag peptide.

Human proteins were synthesized *in vitro* with a human IVT kit (Thermo Scientific). Proteins were individually translated at 30 $^{\circ}$ 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 per 50 µl reaction, Life Technologies). To semi-quantifiably analyse HaloTag fusion proteins, their HaloTag domains were covalently labelled with a fluorescent reporter Halo-TMR (Promega) and analysed by SDS–PAGE and the fluorescent gel imaging with a Typhoon Trio Imager (GE Healthcare).

Barcoding DNAs of 220 bp in length 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 μ M amino-modified oligonucleotide and 10 mM succinimidyl ester (O4) ligand (Promega) in 50 mM Na₂HPO₄, pH 8.0, 150 mM NaCl and 50% formamide at room temperature for 1 h; the modified oligonucleotide was purified by reverse-phase high-performance liquid chromatography using a Zorbax Eclipse XDB-C18 column (5 μ m, 9.4 \times 250 mm, Agilent Technologies) and an elution gradient of 5–70% CH₃CN/H₂O (0.1 M triethylammonium acetate). To generate protein-DNA conjugates, we typically incubated \sim 0.5–2 μ M modified barcoding DNAs and \sim 2–5 μ M HaloTag-labelled 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-hydrolysable GTP analogue, Gpp(NH)p, by EDTA-enhanced nucleotide exchange as previously described³⁵. Mixed wild-type and mutant Raf-RBD (2 nM) displayed on PRMC complexes were incubated with different concentrations of barcoded Ras in 50 mM HEPES, pH 7.5, 100 mM NaCl, 10 mM MgCl₂, 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⁻¹ RNasin (Promega) and 0.1% Tween 20) in the presence of 2 mg ml⁻¹ 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, pH7.5, 50 mM KOAc, 2 mM EDTA and 5 mM MgCl₂, 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 analysed on a single array.

ScFv binding profiling and immunoprecipitation assay. To diversify scFvs, errorprone 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 disulphide 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, pH7.5, 100 mM NaCl, 10 mM MgCl₂ and 0.1% (v/v) Tween 20) at 4 °C for 4h. 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 streptavidin binding peptide tag. Proteins were synthesized *in vitro* using a PURExpress IVT kit supplemented with the disulphide 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) labelled 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 analysed by SDS–PAGE and the fluorescent gel imaging.

Array deposition. Barcoded proteins were diluted with the deposition buffer to a tenfold 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 μ M 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, $\geq 20\,\mu l$ 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 H₂O, and air dried in an AirClean PCR hood. Prior to the amplification, samples containing mRNA-cDNA hybrids can be digested with 10 U ml⁻¹ RNase H (New England Biolabs) in 50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM MgCl₂ 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 protocols¹⁰. 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 sulphate, 2 mM magnesium sulphate, 0.1% (v/v) Triton X-100, 1.3% (v/v) dimethylsulphoxide and 2 M betaine) and incubated with 200 µM dNTPs and 80 U ml⁻¹ 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 sulphate, 2 mM magnesium sulphate 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 imes 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 CoCl₂) followed by incubating the flow cell at 37 $^\circ\mathrm{C}$ for 10 min.

DNA sequencing. Linearized and 3'-OH blocked polonies were analysed by hybridization with fluorescently labelled oligonucleotides, SBE or sequencing-byligation 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 oligonucleotides or ddNTPs (PerkinElmer) labelled by fluorescein/FAM, Cy3/Ty563 or Cy5/Ty665 and subjected to three-colour fluorescence imaging. In brief, the hybridization was performed by incubating polonies with oligonucleotides (2 µM each) in a hybridization buffer (5 \times 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 oligonucleotides with a wash buffer W2 ($0.3 \times SSC$ and 0.1% (v/v) Tween 20). The SBE was performed by incubating primer-bound polonies with fluorescently labelled ddNTPs (1 µM each) and 0.32 U µl⁻¹ 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 $^\circ \rm C$ for 5 min, followed by washing off excess ddNTPs with the wash buffer W2. For each sequencingby-ligation cycle, polonies were hybridized to a sequencing primer and probed with a query oligonucleotide set (fluorescent nonamers, 2 µM each subpool) in a ligation buffer (50 mM Tris-HCl, pH 7.6, 10 mM MgCl₂, 1 mM ATP and 5 mM DTT) in the presence of 30 U μ l⁻¹ 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 (1,000 \times 1,000 pixels, Hamamatsu Photonics). Polony images were acquired under an epi-illumination mode by using a \times 20 (HCX PL Fluotar L, N.A. 0.40, Leica) or \times 40 objective (HCX PL APO, N.A. 0.85, Leica) and from three channels (fluorescein, Cy3 and Cy5) using 488, 561 and 635 nm lasers and excitationemission 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 Image] 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 described¹¹. 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 superimposing 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 five sequencing cycles were required to analyse 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 labelled by Cy3 or Cy5, and then their images were superimposed 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 co-amplification of colocalized templates^{37,38}, and set an optimized threshold distance to be 0.7 μ m. 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 statistic³⁹ was applied to analyse polony colocalization patterns (Supplementary Notes).

- Yang, X. et al. A public genome-scale lentiviral expression library of human ORFs. Nature Methods 8, 659–661 (2011).
- Parker, E. M., Kameyama, K., Higashijima, T. & Ross, E. M. Reconstitutively active G protein-coupled receptors purified from baculovirus-infected insect cells. *J. Biol. Chem.* 266, 519–527 (1991).
- Kobilka, B. K. Amino and carboxyl terminal modifications to facilitate the production and purification of a G protein-coupled receptor. *Anal. Biochem.* 231, 269–271 (1995).
- 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. 8, 617–625 (2013).
- John, J., Frech, M. & Wittinghofer, A. Biochemical properties of Ha-ras encoded p21 mutants and mechanism of the autophosphorylation reaction. J. Biol. Chem. 263, 11792–11799 (1988).
- Mitra, R. D. et al. Digital genotyping and haplotyping with polymerase colonies. Proc. Natl Acad. Sci. USA 100, 5926–5931 (2003).
- Mitra, R. D., Shendure, J., Olejnik, J., Edyta Krzymanska, O. & Church, G. M. Fluorescent *in situ* sequencing on polymerase colonies. *Anal. Biochem.* 320, 55–65 (2003).
- Aach, J. & Church, G. M. Mathematical models of diffusion-constrained polymerase chain reactions: basis of high-throughput nucleic acid assays and simple self-organizing systems. *J. Theor. Biol.* **228**, 31–46 (2004).
- Philimonenko, A. A., Janacek, J. & Hozak, P. Statistical evaluation of colocalization patterns in immunogold labeling experiments. J. Struct. Biol. 132, 201–210 (2000).
- Lewis, J. D. et al. Quantitative interactor screening with next-generation sequencing (QIS-Seq) identifies Arabidopsis thaliana MLO2 as a target of the Pseudomonas syringae type III effector HopZ2. BMC Genomics 13, 8 (2012).
- Georgiou, G. et al. The promise and challenge of high-throughput sequencing of the antibody repertoire. Nature Biotechnol. 32, 158–168 (2014).
- Larman, H. B. et al. Autoantigen discovery with a synthetic human peptidome. Nature Biotechnol. 29, 535–541 (2011).
- Fujimori, S. *et al.* Next-generation sequencing coupled with a cell-free display technology for high-throughput production of reliable interactome data. *Sci. Rep.* 2, 691 (2012).
- 44. Darmanis, S. et al. ProteinSeq: high-performance proteomic analyses by proximity ligation and next generation sequencing. PLoS ONE 6, e25583 (2011).
- Young, K. et al. Identification of a calcium channel modulator using a high throughput yeast two-hybrid screen. Nature Biotechnol. 16, 946–950 (1998).
- Nishihara, T. et al. Estrogenic activities of 517 chemicals by yeast two-hybrid assay. J. Health Sci. 46, 282–298 (2000).
- Chidley, C., Haruki, H., Pedersen, M. G., Muller, E. & Johnsson, K. A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nature Chem. Biol.* 7, 375–383 (2011).
- Wrighton, N. C. et al. Small peptides as potent mimetics of the protein hormone erythropoietin. Science 273, 458–463 (1996).
- Lowman, H. B. Bacteriophage display and discovery of peptide leads for drug development. Annu. Rev. Biophys. Biomol. Struct. 26, 401–424 (1997).
- Leuchowius, K. J. *et al.* High content screening for inhibitors of protein interactions and post-translational modifications in primary cells by proximity ligation. *Mol. Cell. Proteomics* 9, 178–183 (2010).



Extended Data Figure 1 | Improved stability of PRMC complexes generated in a reconstituted *E. coli* IVT system. a, Schematic of PRMC complex stability analysis by measuring the relative ratio of barcoding DNA to HaloTag-labelled protein. b, Comparison of PRMC complex stabilities in the *E. coli* recombinant-factor-reconstituted (PURE) and an *E. coli* crude extract (S30) IVT system. 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.

а



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 different 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. a**, **b**, Representative images of polonies amplified from barcoding DNA templates without (**a**) or with

b



(**b**) 5'-acrydite modifications. Oversized polonies or polony clusters shown in **a** 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 against 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 against dilution factors. Data represent mean values of two technical replicates.

LETTER RESEARCH



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 μ M 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), probably bound to TolA during the purification, and degradation products (due to the hydrolysis of an acylimine bond in the DsRed chromophore) were efficiently crosslinked to HaloTag–DsRed–TolA.

b, Comparison of HaloTag-labelled DsRed and mCherry, a monomeric fluorescence protein, crosslinked at different conditions. Proteins labelled with Halo-TMR were analysed 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	Observed polonies for different dilutions					
		length (a. a.)	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
Knattin	Engineered cystine-knot miniprotein, 2.5D	34	3,990,730	406,311	42,426	3,885	468	41
KIIOUIII			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
		202	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	1048	441,523	45,073	4,717	439	37	3
	receptor-binding domain		461,514	47,540	4,784	364	46	1

Polony counts of two technical replicates were shown. a.a., amino acids.

LETTER RESEARCH

Extended Data Table 2 \mid ScFvs and human proteins used in the one-pot binding profiling

ScFv #	Original mAb	Construction method	Binding target	Human	Entrez Gene	Entrez Gene	Protein length
1-20	Anrukinzumah	Vu-linker-V	13	22	31	386653	165
21-40	Ustekinumab	V _H -linker-V _L	II 12B	23	CD1A	909	312
41-60	Canakinumab	V _H -linker-V _L	II 1B	20	CD1F	913	334
61-80	Conatumumab	V _H -linker-V _L	TNERSE10B	25	CD3E	916	208
81-100	Ibalizumab	V _H -linker-V	CD4	26	CD4	920	459
101-120	Oportuzumab	V _H -linker-V _I	CD326	27	CD7	924	241
121-140	Otelixizumab	V _H -linker-V₁	CD3E	28	CD27	939	261
141-160	Ranibizumab	V _H -linker-V∟	VEGFA	29	CD38	952	301
161-180	Siltuximab	V _H -linker-V∟	IL6	30	CD40	958	278
181-200	Tanezumab	V _H -linker-V _L	NGF	31	CD52	1043	62
				32	CD58	965	241
Human	Entrez Gene	Entrez Gene	Protein length	33	CD59	966	129
protein #	symbol	ID	(a.a.)	34	CD79B	974	230
1	IL1B	3553	270	35	CD74	972	233
2	IL2	3558	154	36	CD80	941	289
3	IL3	3562	153	37	CD207	50489	329
4	IL4	3565	154	38	CD247	919	165
5	IL6	3569	212	39	CD274	29126	291
6	IL7	3574	178	40	CD302	9936	171
7	IL8	3576	100	41	CD320	51293	283
8	IL9	3578	145	42	CD326	4072	314
9	IL10	3586	179	43	CNTF	1270	201
10	IL12B	3593	328	44	CSF2	1437	145
11	IL13	3596	147	45	CSF3	1440	201
12	IL15	3600	163	46	NGF	4803	241
13	IL17A	3605	156	47	NODAL	4838	348
14	IL17B	27190	181	48	OSM	5008	253
15	IL17F	112744	164	49	THPO	7066	350
16	IL18	3606	194	50	VEGFA	7422	237
17	IL20	50604	177	51	FASLG	356	282
18	IL25	64806	178	52	LTA	4049	206
19	IL26	55801	172	53	TNF	7124	234
20	IL28A	282616	201	54	TNFRSF10B	8795	440
21	IL29	282618	201	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	lg-seq	PhIP-Seq	PLATO and IVV-HiTSeq	ProteinSeq and IDUP	SMI-seq
Principle and references*	Y2H coupled with NGS ^{27,40}	B cell-based antibody screening coupled with NGS ⁴¹	Phage display coupled with NGS ⁴²	Ribosome and mRNA displays, respectively, coupled with NGS ^{16,43}	Proximity ligation and extension assays, respectively, coupled with NGS ^{30,44}	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 and references*	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 ^{45,46} and labeled compounds ⁴⁷	No	Demonstrated with barcoded peptides ^{48,49}	Demonstrated with bead-immobilized compounds ¹⁶	Demonstrated with unlabeled ⁵⁰ and barcoded compounds ³⁰	Demonstrated at a scale of 6 by 3 with unlabeled compounds

Refer to Supplementary Discussion for a comparison of the above technologies. CDS, protein coding sequence; NGS, next-generation sequencing. *References 14, 16, 27, 30, 40–50 are cited in this table. **Mathematical model suggests that $\ge 10^5 \times 10^5$ interactions may be analysed in a single binding assay (refer to Supplementary Notes).