

Using high-throughput barcode sequencing to efficiently map connectomes

Ian D. Peikon^{1,2,†}, Justus M. Kebschull^{1,2,†}, Vasily V. Vagin^{2,†}, Diana I. Ravens², Yu-Chi Sun², Eric Brouzes^{3,4}, Ivan R. Corrêa, Jr⁵, Dario Bressan^{1,2,6} and Anthony M. Zador^{2,*}

¹Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA, ²Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA, ³Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794, USA, ⁴Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794, USA, ⁵New England Biolabs, Inc., Ipswich, MA 01938, USA and ⁶Cancer Research UK Cambridge Institute, Li Ka Shing Centre, University of Cambridge, Cambridge CB2 0RE, UK

Received February 11, 2017; Revised March 20, 2017; Editorial Decision April 10, 2017; Accepted April 13, 2017

ABSTRACT

The function of a neural circuit is determined by the details of its synaptic connections. At present, the only available method for determining a neural wiring diagram with single synapse precision—a ‘connectome’—is based on imaging methods that are slow, labor-intensive and expensive. Here, we present SYNseq, a method for converting the connectome into a form that can exploit the speed and low cost of modern high-throughput DNA sequencing. In SYNseq, each neuron is labeled with a unique random nucleotide sequence—an RNA ‘barcode’—which is targeted to the synapse using engineered proteins. Barcodes in pre- and postsynaptic neurons are then associated through protein-protein crosslinking across the synapse, extracted from the tissue, and joined into a form suitable for sequencing. Although our failure to develop an efficient barcode joining scheme precludes the widespread application of this approach, we expect that with further development SYNseq will enable tracing of complex circuits at high speed and low cost.

INTRODUCTION

The brain is extraordinarily complex, consisting of myriad neurons connected by even larger numbers of synapses. Disruption of these connections contributes to many neuropsychiatric disorders including autism, schizophrenia and depression. Understanding how the brain processes information and produces actions requires knowledge of both the structure of neural circuits, and of the patterns of neural activity. Sophisticated technology for recording ever-larger numbers of neurons is now widely available and is providing

unprecedented insight into the physiological responses of brain circuits (1,2). In contrast, circuit-mapping technologies with synaptic resolution remain very slow, expensive and labor intensive.

Mapping neural connectivity is traditionally viewed as a problem of microscopy. Electron microscopy (EM) allows direct imaging of synaptic contacts between neurons, so in principle circuit mapping with EM is trivial. In practice, however, it is complicated by a mismatch of scales. Imaging synapses requires nanometer resolution. In contrast, brain circuits span macroscopic distances, from millimeters in small organisms to tens of centimeters in humans. Circuit reconstruction using EM thus needs to bridge these scales, resulting in the requirement that thin axonal processes be traced across thousands of sections at an exceedingly low error rate. For example, for a 5 mm axon, and EM sections 50 nm thick, the required accuracy per single axon section would need, under simple assumptions, to exceed 99.999% in order to achieve a 36% chance of assigning a correct connection. Several major efforts are underway to increase the throughput and autonomy of EM and have resulted in impressive improvements of speed and scale (3–10). Unfortunately, most of these advancements require very expensive instruments, and the challenge of automatically tracing axonal processes through EM stacks remains unsolved.

Electrophysiological approaches allow probing the connectivity of pairs or small groups of nearby neurons (11–13). These efforts have uncovered elements of high-order structure within neural circuits, as well as spatially intertwined but non-interconnected networks (12,14). However, such physiological methods are labor-intensive, and cannot readily be scaled for the analysis of larger neural circuits or a full nervous system (see however ref (15)).

We have been developing high-throughput sequencing as a fast and efficient alternative to microscopy or physiology for probing neuroanatomical connectivity (16,17). To

*To whom correspondence should be addressed. Tel: +1 516 367 6950; Email: zador@cshl.edu

†These authors contributed equally to the work as first authors.

translate anatomical questions to a format amenable to sequencing, we label neurons uniquely with random nucleic acid sequences ('barcodes'). As a first proof of principle, we recently described MAPseq, a method for reading out long range projections with single neuron resolution (17). In MAPseq, we infect neurons with a pool of barcoded virus particles and thus uniquely label every infected neuron with the barcode sequence carried by the viral particle that infected the neuron. The barcode is then expressed as an mRNA and is transported into axons, where we detect the barcode mRNA by sequencing as a proxy for the axonal projection of every labeled neuron. MAPseq allows the simultaneous tracing of thousands and potentially millions of single neuron projections—presenting a speedup of up to five orders of magnitude over traditional, microscopy-based methods. While MAPseq provides information about area-to-area connectivity at single neuron resolution, it does not provide single-neuron information about neuron-to-neuron connectivity.

Here, we introduce SYNseq, a method for converting synaptic connections into a form suitable for readout by high-throughput DNA sequencing. SYNseq consists of four steps: neuronal barcoding, trafficking of barcodes to the synapses via tight association with engineered synaptic proteins, joining of barcodes into a form suitable for sequencing, and reconstruction of the network connectivity (Figure 1). Briefly, a pre-synaptic mRNA barcode is trafficked to the presynaptic terminal via association with an engineered version of the Neurexin1B (Nrx1B) protein. Likewise, the postsynaptic barcode is trafficked to the post-synaptic terminal via association with a modified Neuroligin1AB (Nlg1AB) protein. Across a synapse, the presynaptic SYNseq components are covalently linked to the postsynaptic SYNseq components and then immunoprecipitated for further biochemical manipulation to link the pre- and postsynaptic barcodes.

We develop a set of reagents that allow for synaptic trafficking of barcodes, specific crosslinking of carrier proteins across synapses, and recovery and joining of the barcode pairs that define a specific synapse. Unfortunately, however, the efficiency of the last step—barcode joining—is in our hands insufficient to provide for reliable synapse recovery while avoiding false positives. Despite this shortcoming, our work lays the foundation for using high-throughput DNA sequencing for circuit reconstruction. Future improvements upon SYNseq have the potential to enable high-throughput connectomics, opening up many new avenues of research in neuroscience.

MATERIALS AND METHODS

Cell culture

We grew HEK293 cells under standard conditions at 37°C with 5% CO₂ in DMEM (Thermo Fisher) supplemented with 10% heat inactivated FBS (Thermo Fisher) and 5% Pen-Strep (Thermo Fisher). We cultured the cells in tissue culture plates coated with poly-L-lysine (Sigma-Aldrich) to aid adherence to the plate.

We prepared primary cultures of dissociated mouse hippocampal neurons from E18 mice as previously described (18). Briefly, we isolated the brains and placed them into

cold HBSS. We dissected hippocampi and incubated them in HBSS + trypsin + DNase for 15 min at 37°C with gentle periodic agitation. We dissociated the cells and plated the cells at a density of 200 000 cells per well of a 12-well plate or 100 000 cells per side of a Xona Microfluidic Chamber (SND450, Xona Microfluidics). We cultured the neurons in Neurobasal (Thermo Fisher) + 2% B27 (Thermo Fisher) + 1% Pen-Strep (Thermo Fisher) + 0.5 mM L-glutamine (Thermo Fisher). For the first 4 days after dissociation, we added glutamate to the media at a final concentration of 125 μM. We incubated all neuronal cultures at 37°C with 5% CO₂ and conducted experiments after 14 days *in vitro*.

Plasmid construction

We constructed all plasmids used in this study by Gibson assembly (19) and standard cloning procedures and confirmed successful assembly by Sanger sequencing. All plasmids and annotated plasmid maps used in this study are available from Addgene. For accession numbers consult Supplementary Table S2.

We produced barcoded plasmid diversity libraries for both HEK cell transfection and Sindbis virus production as previously described (17) aiming for a minimum diversity of 10⁶ different plasmid sequences as estimated from colony counts.

RNA design

The presynaptic and postsynaptic RNA are based on GFP and mCherry coding sequences, respectively. In the 5'UTR of the preRNA we placed an anchor sequence (preHandle), a 100 bp spacer sequence, a NotI restriction site (for barcode cloning), the reverse complement of a qPCR tag denoting the library batch (A or B), a 30 nt barcode, the reverse complement sequence of the Illumina sequence P5-SBS3T, and an MluI restriction site for barcode cloning. In the 5'UTR of the postRNA we placed an anchor sequence (postHandle), a spacer sequence, an MluI restriction site (for BC cloning), the reverse complement of a qPCR tag denoting the library batch (A or B), a 30 nt barcode, the reverse complement sequence of the Illumina sequence P7-SBS8, and a NotI restriction site for barcode cloning. In the 3'UTR of both sequences, we placed four repeats of the boxB hairpin motif, 4xBoxB (20).

Protein design

We started with two interacting synaptic proteins – the presynaptic protein Neurexin1B (Uniprot accession Number: P0DI97, isoform 1b) and the post-synaptic protein Neuroligin1AB (Uniprot accession Number: Q99K10, isoform 1). In the final synPRE-P and synPOST-P proteins we fused MYC-CLIP (21) and HA-SNAP (22) to Nrx1B and Nlg1AB, respectively. In both cases, fusion was directly after the signal peptide sequence (after amino acid G46 in Nrx1B and K47 in Nlg1AB). We further fused a single copy of the nL domain (20) flanked by flexible linkers after amino acid S423 in Nrx1B and after T776 in Nlg1AB.

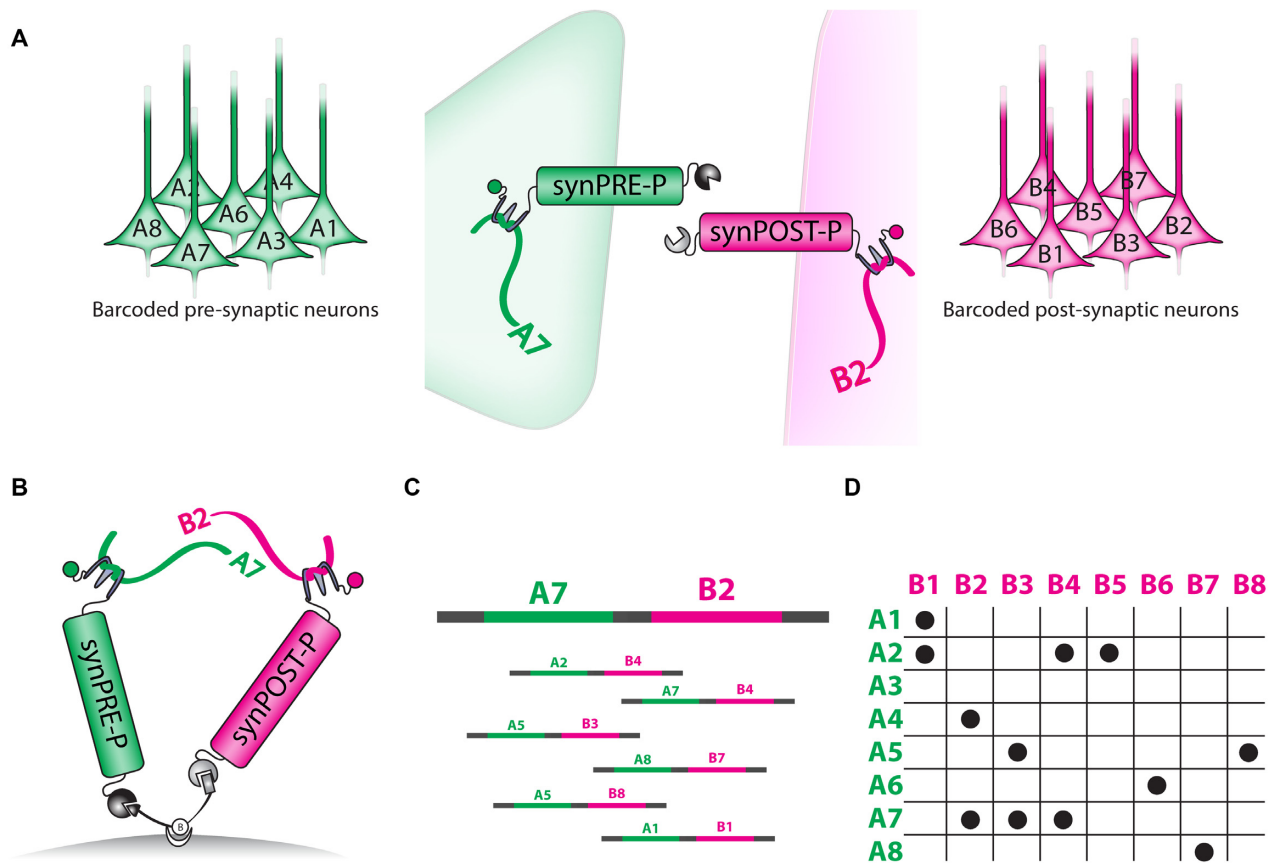


Figure 1. Overview of SYNseq. (A) First, we express random mRNA barcodes and modified synaptic proteins in a separate population of pre- and postsynaptic neurons, so that each neuron is uniquely labeled with a different barcode sequence (compare ref (17,18)). The modified synaptic proteins specifically bind the mRNA barcodes via an RNA binding domain, thereby trafficking barcodes to the pre- (*left*) or postsynaptic (*right*) compartments, respectively. The proteins meet at the synapse. (B) Next, the pre- and postsynaptic proteins are cross-linked *in situ* using a custom synthetic bivalent linker, which covalently joins the extracellular domains of the pre- and postsynaptic proteins. The resulting complexes, consisting of covalently bound pre- and postsynaptic protein pairs bound via the RNA binding domain to their respective barcodes, are then purified via immunoprecipitation (IP). (C) The associated barcode pairs, which represent pairs of connected neurons, are joined, amplified, and sequenced. (D) Sequencing data then allows the reconstruction of a connectivity matrix.

Proximity ligation assay

We fixed neurons grown in Xona microfluidic chambers in 4% PFA in 0.1 M PBS for 15 min at room temperature, washed 2× in PBS, and quenched excess PFA by incubating 4 mM glycine in PBS for 5 min.

We then performed PLA according to the manufacturer's instructions (Duolink PLA kit; Sigma-Aldrich). For the detection of the pre-synaptic proteins, we used goat anti-MYC antibody ab9132 (Abcam), which recognizes the MYC tag on the extracellular domain of the presynaptic proteins. For detection of the post-synaptic proteins, we used rabbit anti-HA antibody ab9110 (Abcam), which recognizes the HA tag on the extracellular domain of the postsynaptic proteins. We used Duolink secondary antibodies MINUS Probe Donkey anti-Rabbit (Sigma DUO92005) and PLUS Probe Donkey anti-Goat (Sigma DUO92003) and used the Duolink *in situ* far red detection reagents (Sigma DUO92013) for detection. All imaging was done using 20× objective on an LSM710 confocal microscope (Zeiss). One representative z-section is shown.

Sindbis virus production and titering

We produced and titered all Sindbis viruses used in this study as previously described (18).

Briefly, we linearized the genomic plasmid and the DH(26S)5'SIN or DH-BB(5'SIN;TE12ORF) helper plasmid (18) by digestion with PacI or XhoI (New England Biolabs), respectively. We *in vitro* transcribed the linearized digestion products using the SP6 mMessage mMachine *in vitro* transcription system (Thermo Fisher) and transfected genomic and helper RNA into BHK cells using Lipofectamine 2000 (Thermo Fisher). We harvested virus 40 h after transfection and concentrated it by ultracentrifugation. We determined the titer of all viruses by qRT-PCR against a plasmid standard.

Chemical syntheses of bifunctional crosslinkers

We prepared BG-PEG-Biotin-PEG-BC (BC = benzylcytosine; BG = benzylguanidine) and BG-PEG-(S-S)-biotin-PEG-BC by reacting BG-PEG-BC (23) with commercially available biotin amidohehexanoic acid NHS (Sigma-

Aldrich) and NHS-S-S-dPEG[®]₄-biotin (Quanta Bio design), respectively. Briefly, we dissolved BG-PEG-BC (5.0 mg, 4.8 μ mol) in anhydrous DMF (1.0 ml). We added biotin amidohexanoic acid NHS (2.3 mg, 5.0 μ mol) or NHS-S-S-dPEG[®]₄-biotin (3.8 mg, 5.0 μ mol), followed by triethylamine (1.0 μ l, 7.2 μ mol). We then stirred the reaction mixture overnight at room temperature. We removed the solvent under vacuum and purified the products by reversed-phase HPLC on a VYDAC 218TP series C18 column (22 \times 250 mm, 10 μ m particle size) at a flow rate of 20 ml/min using a water/acetonitrile gradient (0–95% acetonitrile over 45 min). Yields: BG-PEG-biotin-PEG-BC (44%), ESI-MS m/z 1387.6916 (calc. for C₆₄H₉₄N₁₈O₁₅S⁺, m/z 1387.6940); BG-PEG-(S-S)-biotin-PEG-BC (21%), ESI-MS m/z 1684.7621 (calc. for C₇₄H₁₁₄N₁₉O₂₀S₃⁺, m/z 1684.7644). We recorded high resolution mass spectra by electrospray ionization (ESI) on an Agilent 6210 time-of-flight (TOF) instrument.

Tagging with BG/BC derivatives

We obtained all BG- and BC-functionalized derivatives, including the bifunctional cross-linkers (synthesis see above), CLIP-Surface488 and SNAP-Surface488 from New England Biolabs (NEB) and used them according to the manufacturer's instructions. Briefly, we prepared a stock of 1 mM tag in DMSO and labeled cells expressing cell-surface CLIP- or SNAP-fusion proteins by incubating them in 5 μ M tag in full media for 30 min, followed by three washes. We performed all imaging of PFA fixed, surface labeled cells using a 63 \times objective on an LSM710 confocal microscope (Zeiss). 3D reconstructions of the images are shown to illustrate surface staining.

We slightly altered the above protocol for crosslinking cultured hippocampal neurons. Briefly, we incubated the cells with 2.5 μ M bifunctional crosslinker in complete neuron media for 30 min. We then blocked all unreacted SNAP and CLIP epitopes by incubating the cells in 10 μ M each of SNAP-cell block and CLIP-cell block (NEB S9106S and S9220S) in full neuron media for 30 min. We then washed the cells three times with full neuron media and finally in PBS.

Protein and RNA IPs from HEK cells

After tagging with biotinylated linkers, we lysed cells in 1 ml of PLB3 lysis buffer (20 mM Tris-HCl, 20 mM NaHCO₃, 150 mM NaCl, 4 mM KCl, 2.5 mM MgCl₂, 2.5 mM CaCl₂, 10% glycerol, 1% CHAPS, 1% TritonX-100, 1 \times Complete EDTA-free protease inhibitor (Roche) (24)), per 12-well well of HEK cells or Xona microfluidic chamber. For RNA IPs we added RNase inhibitors RNasin (Promega N2515) to PLB3 at 120 U/ml of lysis buffer.

We incubated the crude cell lysates on ice for 30 min, pelleted cell debris by centrifugation at 21 000 \times g for 30 min at 4°C and saved a fraction of the resulting lysate for total protein analysis via western blot or other means. We loaded the remainder of the lysate onto 100 μ l of pre-washed Dynabeads M280 Streptavidin magnetic beads (Thermo Fisher) and incubated with gentle rotation overnight at 4°C. We

then washed the beads with PLB3 with ascending salt concentrations (2 \times with 150 mM NaCl, 2 \times with 300 mM NaCl and 2 \times with 500 mM NaCl) for 10 min each.

We finally eluted the RNA or protein samples depending on the downstream application.

For emulsion RT-PCR we eluted protein samples from Dynabeads via addition of 15 mM DTT for 1 h at 37°C.

For western blotting, we boiled the dynabeads in SDS containing loading buffer (New England Biolabs, B7709S) for 10 min.

For qPCR, we resuspended the beads in 400 μ l Proteinase K buffer (100 mM Tris-HCl pH 7.5, 12.5 mM EDTA, 150 mM NaCl, 1% SDS) with 2 μ l of GlycoBlue (Thermo Fisher) and 20 μ l of Proteinase K (20 μ g/ μ l; Roche 03115879001). We incubated the samples at 65°C for 1 h with gentle shaking and extracted the RNA using acid phenol:chloroform (Thermo Fisher) according to the manufacturer's protocol. We then DNase treated the RNA (RQ1 RNase-free DNase, Promega) according to the manufacturer's instructions. Finally, we reverse transcribed the RNA using oligo dT primers and SuperscriptIII (Thermo Fisher) according to the manufacturer's instructions.

We performed qPCR using primers 5'-GAC GAC GGC AAC TAC AAG AC-3' and 5'-TAG TTG TAC TCC AGC TTG TGC-3' for *gfp*, primers 5'-GCT TCA AGT AGT CGG GGA TG-3' and 5'-CCT GTC CCC TCA GTT CAT GT-3' for *mcherry* and primers 5'-CGC GAG AAG ATG ACC CAG AT-3' and 5'-ACA GCC TGG ATA GCA ACG TAC AT-3' for human β -actin in power SYBRgreen Master mix (Applied Biosystems) according to the manufacturer's instructions.

We calculated enrichment in qPCR by first calculating the Δ ct values for IP-Total for a housekeeping gene (β -actin) and the barcode containing transcripts. We then calculated the Δ Δct by subtracting the actin Δ ct from the barcode Δ ct (25).

Double-IP from neuronal cultures

We applied the cell lysate to 100 μ l of anti-HA magnetic beads (Thermo Scientific 88837) that we pre-blocked by overnight incubation with mouse brain lysate. We incubated the lysates on beads at 4°C for 8–12 h with rotation and washed in PLB3 in ascending salt concentrations as above. We then rinsed the beads in standard PBL3 and eluted the precipitated protein with rotation in 100 μ l of 2 μ g HA peptide (Thermo Fisher 26184) in PBL3 at 37°C for 30 min. We collected 10% of the eluate for western blot analysis and incubated the rest with pre-blocked anti-FLAG M2 beads (Sigma Aldrich A2220) at 4°C for 8–12 h with rotation. We washed the beads 6 \times in PLB3 with ascending salt concentrations as above, rinsed with standard PBL3 and eluted the captured proteins by two sequential incubations with 25 μ l of 0.5 μ g of 3xFLAG peptide (Sigma F4799) in PBL3 at 4°C with rotation for 2 h. We collected 10 μ l of the combined eluate for western analysis and snap froze the rest in liquid nitrogen for droplet overlap RT-PCR.

Western blotting

We used anti-MYC tag antibody, clone A46 (Millipore) and secondary anti-mouse IgG1 HRP-linked antibodies (Cell

Signaling 7076s) for the detection of the pre-synaptic proteins and anti-HA tag antibodies, HA.11 (Covance) or anti-HA-HRP 3F10 (Roche 12013819001), for the detection of the post-synaptic proteins. We visualized western blots using either Odyssey fluorescent detection according to the manufacturer's instructions or SuperSignal Western Femto Maximum Sensitivity Substrate (Thermo Scientific 34095).

HEK cell emulsion RT-PCR

We encapsulated "transsynaptic" complexes from HEK cells using a custom made setup.

We fabricated microfluidic chips as PDMS/glass hybrids using soft-lithography as previously described (26). We treated the microfluidic channels with a fluorinated tri-chloro silane reagent (heptadecafluoro-1,1,2,2-tetrahydrodecyl) trichlorosilane (Gelest) diluted at 1% weight in FC3280 oil (3M).

To generate droplets, we actuated the fluids by a set of pressure controllers (MPV1, Proportion Air), which were controlled by a Labview (National Instruments, TX, USA) application via a microprocessor (Arduino). We generated 13 pl (29 μ m diameter) droplets at 1 kHz, using a 15 μ m deep \times 20 μ m wide hydrodynamic focusing nozzle. We stabilized the droplets with a PEG-Krytox based surfactant (27) dissolved at 1% weight in HFE7500 fluorinated oil (3M). Stability through thermocycling was further increased by adding Tetronic 1307 (BASF) at 1% weight to the RT-PCR mixture (28). We collected droplets into 0.2 ml PCR tubes and removed the bulk oil phase.

We performed one-step overlap RT-PCR in emulsions as previously described (29). We made a 100 μ l RT-PCR mixture containing: IP products (variable), 25 μ l OneStep MasterMix, 5 μ l primer 5'-CAG CTC GAC CAG GAT GGG CA-3' (10 μ M), 5 μ l Primer 5'-TTC AGC TTG GCG GTC TGG GT-3' (10 μ M), 5 μ l primer 5'-TAT TCC CAT GGC GCG CCG CTG GTC GGT ACG GTA ACG GA-3' (1 μ M), 5 μ l primer 5'-GGC GCG CCA TGG GAA TAC GGA CGA TGC CGT CCT CGT A-3' (1 μ M), 20 μ l 5% Tetronic and H₂O to 100 μ l. Briefly, we performed reverse transcription for 30 min at 55°C, followed by 2 min at 94°C. We performed PCR amplification with the following thermocycling conditions. 1 cycle of: 94°C for 30 s, 50°C for 30 s, 72°C for 2 min; followed by 4 cycles of: 94°C for 30 s, 55°C for 30 s and 72°C for 2 min; followed by 22 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 2 min; followed by a final extension step for 7 min at 72°C. After thermal cycling, we visually inspected the emulsion to ensure stability. We broke the emulsions with perfluorooctanol, and collected the aqueous phase for subsequent purification using a spin-column clean-up kit (Promega Wizard SV PCR Cleanup Kit).

We performed a nested PCR amplification (Accuprime Pfx Supermix (Thermo Fisher)) in a total volume of 100 μ l using 10 μ l of emulsion product as template with 100 nM primers 5'-AAT GAT ACG GCG ACC ACC GA-3' and 5'-CAA GCA GAA GAC GGC ATA CGA-3' under the following conditions: 2 min 94°C; followed by 30 cycles of 94°C for 30 s, 62°C for 30 s and 72°C for 30 s; followed by a final extension at 72°C for 7 min.

We purified the final product by gel electrophoresis and sequenced using the PE36 protocol on the MiSeq platform (Illumina).

qPCR to measure crossover rate

We measured the crossover rate in emulsion PCR samples using a quantitative qPCR assay for A-A, B-B and A-B and B-A barcode pairs. We used primers 5'-CAC ATA AGA CGT GTC CAC CGG TT-3' and 5'-GGT GCC AGC ATT TTC GGA GGT T-3' to detect A-A barcode pairs, primers 5'-GCG CAG AGG AAC GCC CAT TTA G-3' and 5'-GCA GTG GTC GGT GCT CTA AAC T-3' to detect B-B barcode pairs, primers 5'-CAC ATA AGA CGT GTC CAC CGG TT-3' and 5'-GCA GTG GTC GGT GCT CTA AAC T-3' to detect A-B barcode pairs and primers 5'-GCG CAG AGG AAC GCC CAT TTA G-3' and 5'-GGT GCC AGC ATT TTC GGA GGT T-3' to detect B-A barcode pairs. We performed all qPCR reactions using Power SYBRgreen master mix (Applied Biosystems) and quantified each of the different barcode pairs against a standard.

Neuron emulsion RT-PCR

To increase throughput of droplet generation, we switched from the custom build setup to a commercially available solution by Dolomite Microfluidics. We produced droplets with 20 μ m diameter using a 1R DE Chip with 20 μ m etch depth and fluorophilic coating (Dolomite Microfluidics, catalogue number 1864021), pumping fluids using the MitoS P-Pump Remote Basic unit (Dolomite Microfluidics, catalogue number 3200177). We used commercially available oil to encapsulate the droplets (Droplet Generation Oil for probes; Bio-Rad, catalog number 186-3005) and used the compatible One-Step RT-ddPCR Advanced Kit for Probes (Bio-Rad) with 2% Tetronic 1307 (BASF) for the emulsion RT-overlap PCR reaction. Per 20 μ l RT-PCR reaction, we used 4 μ l of 10% Tetronic 1307, 5 μ l RT-ddPCR master mix, 2 μ l Reverse transcriptase, 1 μ l 300mM DTT, 1 μ l sample, 1 μ l of a 10 μ M mixture of 5'-CAG CTC GAC CAG GAT GGG CA-3' and 5'-TTC AGC TTG GCG GTC TGG GT-3' and 1 μ l of a 1 μ M mixture of 5'-TAT TCC CAT GGC GCG CCG CTG GTC GGT ACG GTA ACG GA-3' and 5'-GGC GCG CCA TGG GAA TAC GGA CGA TGC CGT CCT CGT A-3'.

After droplet generation, we performed reverse transcription at 50°C for 60 min, followed by PCR (initial enzyme activation, 95°C for 10 min; 40 cycles of 95°C for 30 s, 60°C for 1 min; followed by 98°C for 10 min). Finally, we broke the emulsions using three rounds of phenol-chloroform (Thermo Fisher) cleanup.

We prepared sequencing libraries from the droplet RT-PCR products by performing a nested PCR reaction using Accuprime Pfx Supermix (Thermo Fisher) and primers 5'-AAT GAT ACG GCG ACC ACC GA-3' and 5'-CAA GCA GAA GAC GGC ATA CGA-3' at 68°C annealing temperature. We gel purified the final product and sequenced it using the PE36 protocol on the MiSeq platform (Illumina).

Analysis of sequencing data

We error corrected the barcode sequencing reads as previously described (17) and filtered for barcode pairs that contained perfect matches to the zip-codes hardcoded into the barcodes that allowed us to measure crossover rates. We then set an minimum read count for sequences to be analysed to remove template switched and erroneous reads (30). Finally, we calculated the crossover rate of presynaptic barcodes from one experiment joined to postsynaptic barcodes from another, and vice versa, and constructed a connectivity matrix. All sequencing files are accessible on the SRA under accessions SRS1899968 (ZL039: HEK cell emulsions), SRS1899969 (ZL110: neuron emulsions) and SRS1902304 (ZL107: neuron emulsions, high concentration). Code to process these datasets is available in the Supplementary materials.

RESULTS

In what follows, we first describe the design of the pre- and postsynaptic proteins and demonstrate their function in HEK cell culture. We then demonstrate the synaptic localization of the proteins in primary neuronal culture and describe the isolation of the transsynaptic complex formed by the proteins and their associated RNA barcodes. Finally, we describe our efforts to join the pre- and postsynaptic barcodes of each complex by emulsion RT-PCR.

Neuronal barcodes

We have previously developed a Sindbis-based system for over-expression of neuronal barcodes, which was used to map projections (MAPseq (17)). This work revealed that unique barcodes can be efficiently targeted to individual neurons, and amplified through the Sindbis virus system. In this expression system each virus particle carries a unique RNA barcode, which is extensively replicated after infection by the machinery of the Sindbis virus, and is flanked by sequences that allow for efficient recovery and sequencing. In this work, we sought to extend the MAPseq system by developing proteins that could specifically target barcodes to the pre- and post-synaptic density, and enable recovery of barcode pairs across individual synapses.

Protein design

In order to enable recovery of synaptic connectivity from neuronal barcodes, we developed a pair of SYNseq carrier proteins that conform to three requirements. First, we required that they traffic to the pre- and postsynaptic terminals, respectively. Second, we required that the pre- and postsynaptic components be in a form that allowed for synapse-specific covalent cross-linking. Finally, we required that they bind the barcode mRNA tightly and specifically.

To design proteins that fulfill these requirements, we started with the presynaptic protein Nr_x1B and the postsynaptic protein Nl_g1AB. Both are relatively simple, single pass membrane proteins that traffic naturally to the appropriate side of a mammalian synapse.

To enable these proteins to carry barcode mRNAs, we inserted the RNA binding domain n_λ into the cytoplasmic tail

of the two proteins. The n_λ domain is a 22 amino acid peptide (derived from the λ_N protein of λ bacteriophage) that specifically and strongly binds to a particular 15 nt RNA sequence, termed boxB (20). We also added the corresponding boxB sequence to the barcode mRNA, causing the association of SYNseq proteins and barcode mRNAs.

To permit crosslinking of the proteins across the synapse, we fused the self-labeling proteins CLIP and SNAP to the extracellular domains of Nr_x1B and Nl_g1AB, respectively. CLIP and SNAP are engineered versions of a DNA repair enzyme that specifically react and form a covalent bond with benzylcytosine (BC) and benzylguanine (BG) (21,22), respectively. To crosslink the CLIP and SNAP domains, we synthesized a bi-functional small molecule that contains benzylcytosine on one end, and benzylguanine on the other, and is thus capable of forming a covalent bond between our pre- and postsynaptic proteins. In addition, this crosslinker contains a biotin functional group that allows efficient immunoprecipitation (IP). A second version of the crosslinker further allows for elution from beads after IP for biotin via a cleavable disulfide bridge (Supplementary Figure S1). The two versions of the crosslinker (BG-PEG-biotin-PEG-BC and BG-PEG-(S-S)-biotin-PEG-BC) behave largely identically (data not shown) and we used them interchangeably in this study, unless elution was necessary. Finally, we fused a Myc epitope tag to Nr_x1B, and an HA epitope tag to Nl_g1AB, to aid in biochemistry and/or visualization.

The design of proteins that fulfilled all of the above requirements necessitated considerable troubleshooting. We tested various positions, RNA binding domains, and linker structures before finding a pair of proteins in which our modifications to Nr_x1B and Nl_g1AB did not disrupt the proteins' endogenous trafficking pattern (Supplementary Figures S2-S5 and Table S1). For ease of use and fast turnaround, we performed initial troubleshooting and benchmarking of different constructs by expression and co-expression in HEK cell culture. We then tested promising proteins for the more stringent criteria of synaptic localization and RNA binding in neuronal culture.

Barcode mRNA design

The barcode mRNAs necessary for SYNseq underwent few changes from the initial design. Both pre- and postsynaptic barcode RNA carry a random 30 nt barcode sequence in their 5' UTR, and four copies of the boxB sequence, which allows tight association of the barcode RNA to the SYNseq proteins, in the 3' UTR. Finally, we introduced the coding sequences for GFP and mCherry into the presynaptic and postsynaptic barcode mRNAs, respectively.

SYNseq proteins traffic to the membrane, can be crosslinked and bind barcode mRNA in HEK cell culture

After troubleshooting (Supplementary Table S1), we focused on two proteins, Myc-CLIP-Nr_x1B-1xn_λ (synPRE-P) and HA-SNAP-Nl_g1AB-1xn_λ (synPOST-P), for further testing. Each contained a single repeat of n_λ surrounded by long flexible linker sequences in their cytoplasmic tail, in a position that was previously reported not to disrupt the pro-

teins' endogenous trafficking (31–34). The matching barcode mRNAs encode for GFP and mCherry, respectively, contain a random 30 nt barcode sequence in their 5' UTR, and have four repeats of the boxB stem loop in their 3' UTR (Figure 2A and B).

We first examined the localization of the two proteins. Both synPRE-P and synPOST-P traffic to the HEK cell membrane, as demonstrated by surface staining (Figure 2C and D). This indicates that at least rudimentary trafficking of the proteins is uninterrupted by the addition of crosslinking and RNA binding domains.

We next tested whether we could specifically crosslink synPRE-P and synPOST-P in a simple scenario. We incubated HEK cells expressing synPRE-P and/or synPOST-P with the crosslinker, and performed an IP for the biotin on the crosslinker (Figures 1C and 2E). We only detect a 'crosslinked' synPRE-P—synPOST-P band (Figure 2E, arrow) in addition to the signals of the individual synPRE-P and synPOST-P proteins (Figure, 2E star), when both synPRE-P and synPOST-P proteins are expressed in the cells, and the crosslinker is added before cell lysis (Figure 2E, lane 5). Importantly, we do not detect this crosslinked band when only one of the two proteins is expressed (Figure 2E, lanes 1 and 2) or when the second protein was added post-lysis (Figure 2E, lanes 3 and 4). Moreover, the crosslinker is inactive in the lysis buffer (Figure 2E, lane 6) and does not result in nonspecific bands (Figure 2E, lane 7). These results indicate that we can form covalent synPRE-P—synPOST-P protein complexes, and that post-lysis association of synPRE-P and synPOST-P is low.

Finally, we confirmed that synPRE-P and synPOST-P specifically bind to barcode mRNA. After expression of either synPRE-P or synPOST-P in HEK cell culture and incubation with the biotinylated crosslinker, we find that barcode RNA is specifically enriched >32-fold by a pulldown for the biotinylated crosslinker (Figure 2F). Importantly, the protein-RNA interaction depends on the presence of both the RNA binding domain $n\lambda$ and its RNA recognition sequence boxB.

From these experiments, we confirm that synPRE-P and synPOST-P fulfill basic requirements for SYNseq proteins in HEK cell culture: membrane trafficking, efficient crosslinking and barcode mRNA binding. To further confirm these properties, we turned to more sensitive assays using neuronal cell culture.

SYNseq proteins traffic to synapses in neuronal cell culture and can be crosslinked into transsynaptic complexes

Using neuronal cell culture, we next set out to characterize the behavior of synPRE-P and synPOST-P in neurons, including tests for synaptic localization and protein interaction using a series of increasingly stringent measures.

We first tested whether our protein modifications still permitted proper trafficking of each protein to the cell membrane in neurons. We used a double promoter Sindbis virus to express both the synPRE-P or synPOST-P protein and the pre- or postsynaptic barcode mRNA from a single virus (Figure 3A and B), ensuring that SYNseq protein expression is always coupled with barcode mRNA expression. Sindbis virus is a positive strand RNA virus, character-

ized by a large (up to 6 kb (35)) payload as well as rapid and strong expression of transgenes. It therefore allows for rapid turnaround and iteration of constructs during troubleshooting. In addition, we have previously shown that Sindbis virus can be used to produce high diversity barcode libraries and can uniquely label neurons (17). The Sindbis virus was thus well-suited to act as an expression vehicle for assessing SYNseq feasibility. Surface staining for Sindbis-expressed SYNseq proteins with CLIP-Surface488 (synPRE-P; Figure 3C) and SNAP-Surface488 (synPOST-P; Figure 3D) confirmed that our protein modifications did not interfere with proper trafficking of each protein to the cell membrane in neurons.

We then used the proximity ligation assay (PLA) to screen for protein pairs that interact across synapses. PLA is a double antibody stain that results in a highly amplified signal only if the two probed epitopes are within 40 nm of each other (Figure 4A–D) (36). This interaction length is comparable to the dimensions of a synaptic cleft (20 nm (37)). The presence of a PLA signal between a pre- and a postsynaptic protein expressed in separate, but synaptically coupled populations of neurons thus provided a necessary but not sufficient criterion for establishing the synaptic localization of the two proteins (Figure 4E, Supplementary Figures S2–S5, Table S1). These experiments thus served as a screen for candidate protein pairs to be later tested in more stringent assays that restrict the interaction distance to <5 nm.

To achieve expression of synPRE-P and synPOST-P in separate, but synaptically coupled populations of neurons, we grew cultured primary hippocampal neurons in a XONA microfluidic device (38). In a XONA, two chambers containing primary neurons are connected via long, thin microfluidic grooves. The setup ensures fluidic isolation of the two neural populations, allowing independent viral infection of the two populations, while permitting the growth of axons between chambers through the grooves (Figure 4F) (39).

We infected one population of cells with Sindbis virus expressing the pre-synaptic components and the other population of cells with Sindbis virus expressing the post-synaptic components. Using PLA we detected interaction between the synPRE-P and synPOST-P proteins preferentially where presynaptic (GFP positive) axons reach the postsynaptic chamber, suggesting that our proteins come within at least 40 nm of each other across separate but synaptically coupled cell populations (Figure 4G).

PLA was a rapid and convenient screening tool to troubleshoot protein design (see Supplementary Figures S2–S5, Table S1). However, its interaction distance of 40 nm is too large to confirm synaptic co-localization of synPRE-P and POST. In contrast, the bifunctional crosslinker designed to create the trans-synaptic synPRE-P—synPOST-P complex is only 4 nm long (Supplementary Figure S1). Thus, the creation of a synPRE-P—synPOST-P complex across synaptically coupled neuronal populations necessarily implies that the two proteins came within 4 nm of each other, satisfying even the strictest distance-based definitions of a synapse (4,40) (Figure 5A). Again expressing synPRE-P and synPOST-P in synaptically coupled but separate neuronal populations using a XONA device, we found that we could specifically crosslink and immunoprecipitate

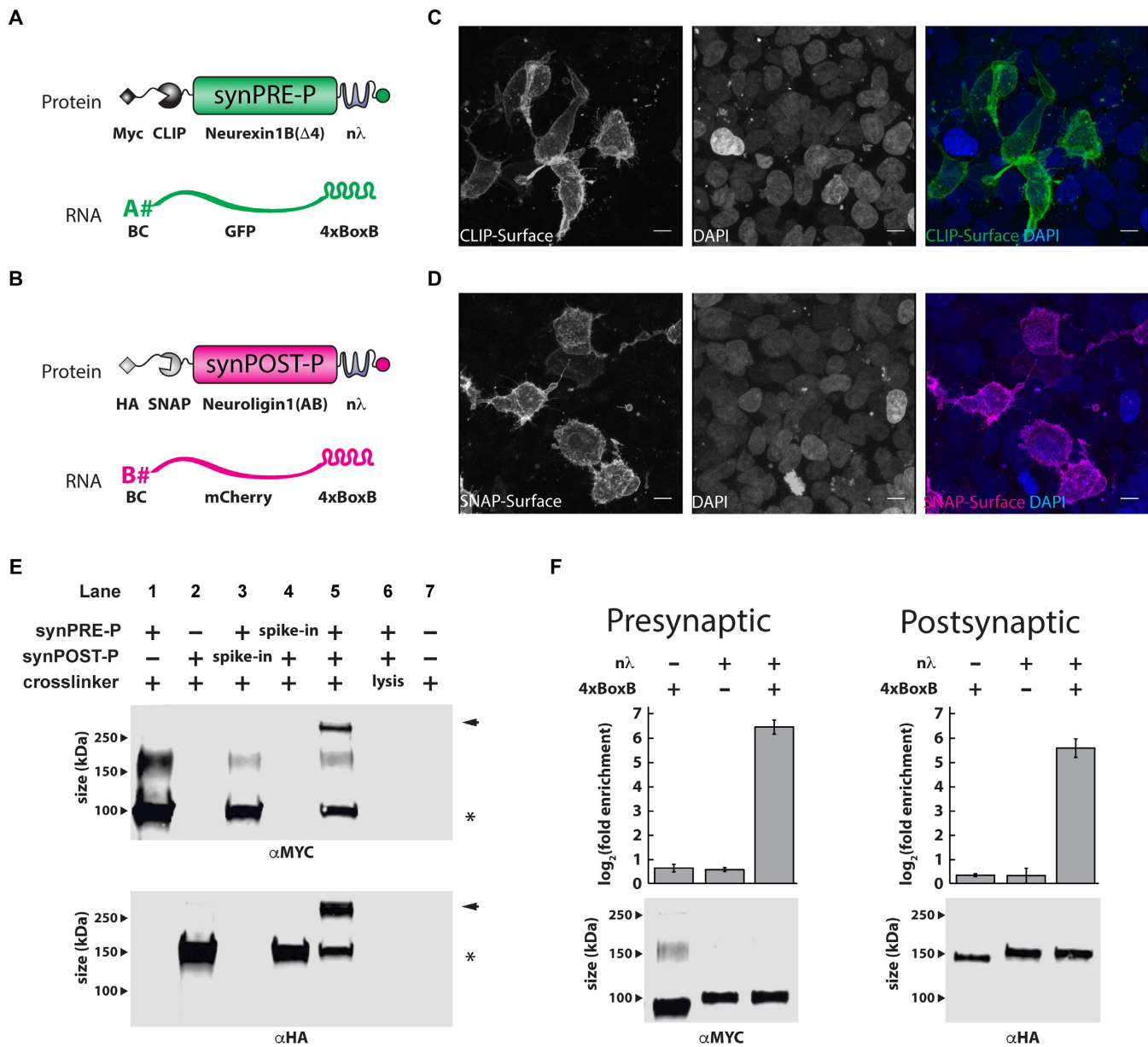


Figure 2. Optimization of SYNseq components in HEK cells. (A) The presynaptic components of the SYNseq system, consisting of CLIP-Nrx1B-1xnλ and a GFP encoding barcode RNA. (B) The postsynaptic components of SYNseq, consisting of SNAP-Nlg1AB-1xnλ and a mCherry encoding barcode RNA. (C and D) A clear membrane staining of synPRE-P and synPOST-P can be observed after staining (C) synPRE-P expressing HEK cells with CLIP-Surface488 and (D) synPOST-P expressing HEK cells with SNAP-Surface488. Scale bar = 5 μm. (E) Western blot analysis shows that synPRE-P and synPOST-P can be specifically crosslinked by addition of a small molecule BG-PEG-Biotin-PEG-BC crosslinker. A crosslinked product is only produced when both synPRE-P and synPOST-P were expressed in HEK cells and the crosslinker was added before lysis (lane 5). Arrow = crosslinked band; star = uncrosslinked synPRE-P or synPOST-P. (F) synPRE-P and synPOST-P specifically and strongly bind to their respective barcode mRNAs as evident in RNA-IPs from transiently transfected HEK cells, after membrane tagging with BG-PEG-biotin-PEG-BC. We show qRT-PCR analysis of three independent RNA-IP experiments and western blot analysis of a representative sample.

synPRE-P and synPOST-P across the two populations of neurons, and by that logic, synapses (Figure 5B). Note that, to simplify biochemistry from neuronal culture, we used a version of synPRE-P that contains a 3xFLAG epitope tag in addition to the myc epitope tag in this and all subsequent experiments.

Enrichment for transsynaptic protein complexes in IP

When we immunoprecipitated crosslinked neuron samples from XONA cultures using the biotin tag on the crosslinker, we can detect at least three major products. First, we recover the desired synPRE-P—synPOST-P complex (Figure 5B, arrow) formed by interaction of the proteins across the synaptic cleft. However, we also observe signals arising from

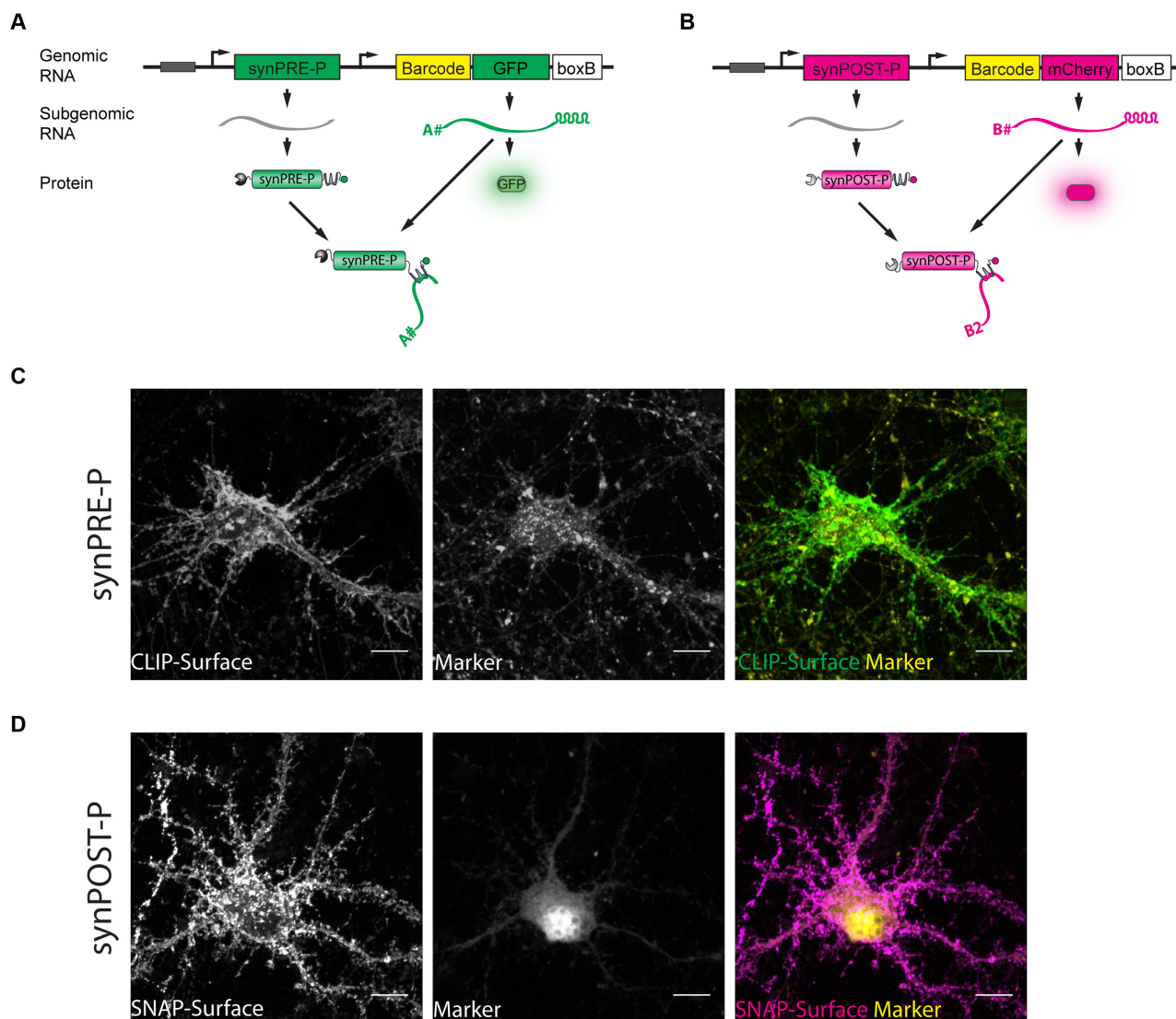


Figure 3. Viral expression and membrane trafficking of synPRE-P and synPOST-P in neurons. We use a double promoter Sindbis virus that expresses the (A) presynaptic or (B) postsynaptic components of SYNseq. (C and D) When expressed in neurons using Sindbis virus synPRE-P and synPOST-P show clear membrane trafficking as revealed by staining (C) synPRE-P with CLIP-Surface488 and (D) synPOST-P with SNAP-Surface488. Scale bar = 5 μ m.

single synPRE-P and synPOST-P proteins that reacted with the one side of the crosslinker, but which have no partner on the other side of the crosslinker (Figure 5B, star). Such unpaired single proteins occur from at least three processes: from incomplete crosslinking across a synapse, from synapses at which only one side expresses a SYNseq protein, as well as from expression of either synPRE-P or synPOST-P in regions of the neuronal membrane that are not part of a synapse. Unfortunately, these unpaired proteins present a challenge to many barcode mRNA joining strategies. We therefore sought to remove them from the IP reaction by performing a double IP, selecting first for the postsynaptic HA tag, then for the presynaptic 3xFLAG tag, so that only proteins that contain both synPRE-P and synPOST-P remain after the second IP step. This strategy efficiently removed all unpaired synPRE-P or synPOST-P proteins from

the IP and enriches for the transsynaptic protein complex (Figure 5C).

From these results, we conclude that synPRE-P and synPOST-P conform to all three requirements that we had set out for SYNseq proteins: They specifically bind to barcode mRNA (Figure 2F), traffic to synapses (Figures 4 and 5B) and can be specifically joined into a transsynaptic protein complex (Figure 5B).

Joining pre- and postsynaptic barcode mRNAs by emulsion RT-PCR

With the desired functional properties of synPRE-P and synPOST-P confirmed, we set out to develop a method for joining RNA barcodes specifically and efficiently. We tested a variety of methods (including splinted ligation of RNaseH trimmed barcode mRNAs by T4 RNA ligase, double-headed ribozymes drawing on the L21 and

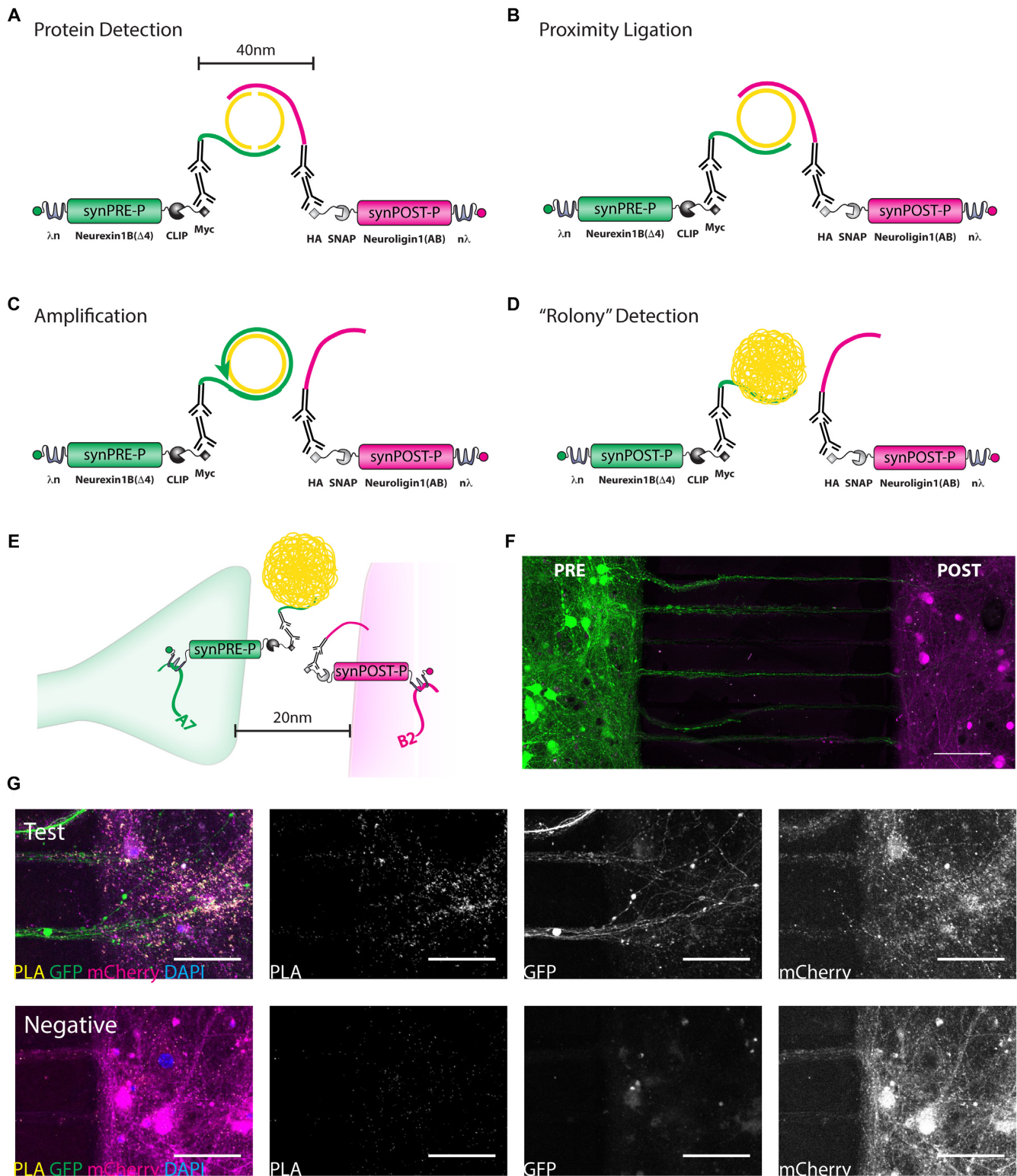


Figure 4. synPRE-P and synPOST-P co-localize at synapses as measured by PLA. PLA is a multistep procedure involving (A) protein detection by two primary antibodies, followed by incubation with oligo-conjugated secondary antibodies. Two bridging oligos (yellow) are hybridized and will form a circle only if the two probed for epitopes are within 40 nm of each other. (B) The bridging oligos are ligated into a circle, which (C) is then amplified by rolling circle amplification and (D) detected by hybridizing fluorescently labeled oligos to the DNA ball. (E) If a labeled presynaptic cell forms a synapse with a labeled postsynaptic cell and synPRE-P and synPOST-P are co-localized at the synapse this should result in a PLA signal. (F) Two populations of neurons grown in a XONA microfluidic chamber system can be independently infected with Sindbis virus, as illustrated here by application of GFP-expressing virus on the left and mCherry expressing virus on the right. Scalebar = 100 μ m. (G) In this system PLA signals are clearly visible when synPRE-P positive (GFP-labeled) axons come close to synPOST-P positive (mCherry-labeled) cells. The signal vastly exceeds the background staining observed in the absence of synPRE-P. Scale bar = 50 μ m.

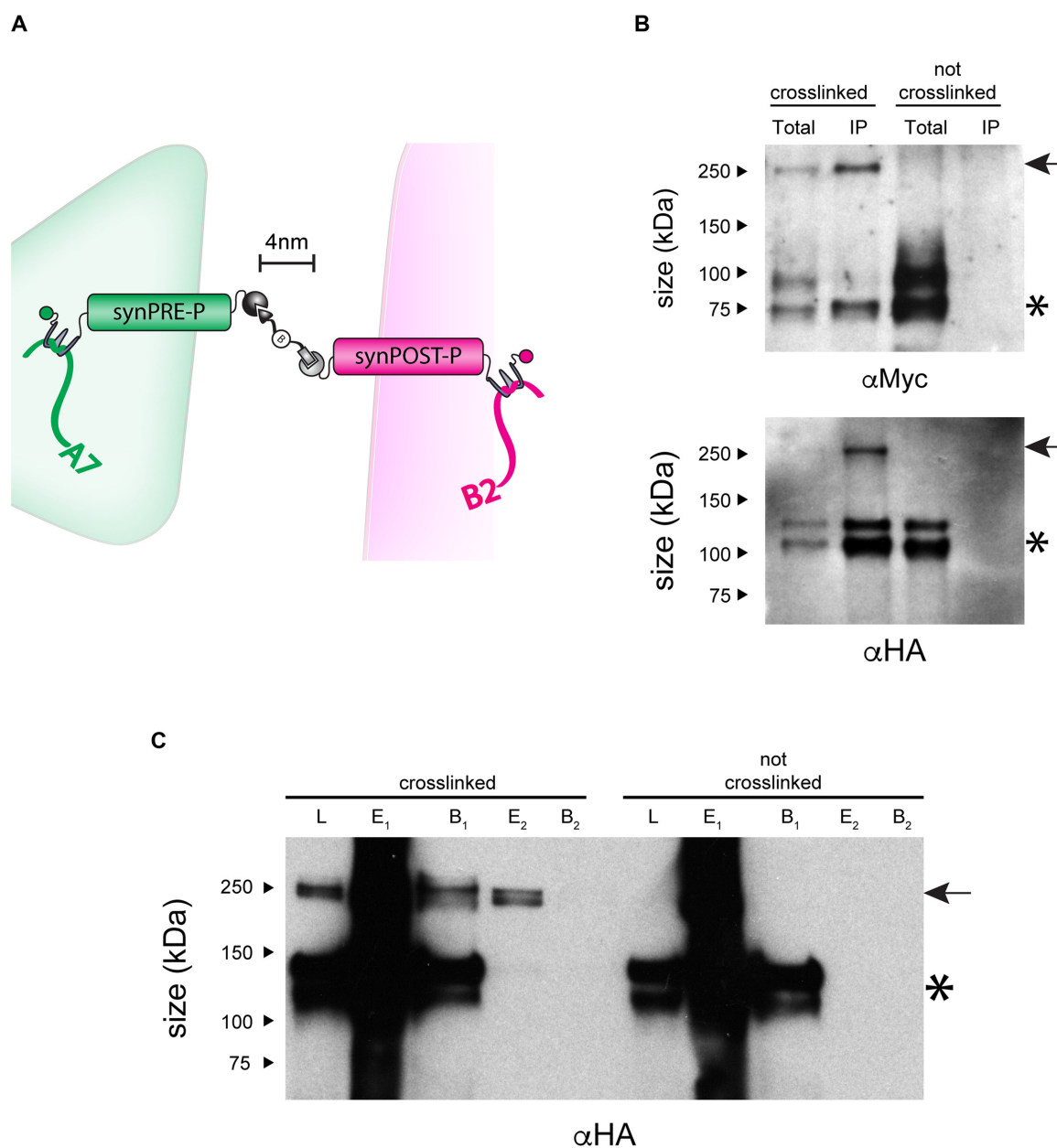


Figure 5. synPRE-P and synPOST-P can be crosslinked into a transsynaptic complex and bind barcode mRNA. (A) If synPRE-P and synPOST-P interact at a synapse, we should be able to crosslink them with our 4 nm long crosslinker. (B) A synPRE-P—synPOST-P crosslinked band (arrow) can be observed from streptavidin IPs of separately infected XONA cultures only after addition of crosslinker. Un-crosslinked single proteins (star) can also be detected. (C) Performing two consecutive IPs, first for HA, then for 3xFLAG efficiently removes un-crosslinked proteins (star) and purifies synPRE-P—synPOST-P complexes (arrow). L = lysate; E₁ = eluate off HA beads; B₁ = remaining protein on HA beads; E₂ = eluate off FLAG beads; B₂ = remaining protein on HA beads.

the Carinii ribozyme and barcode joining by reverse transcription where the two barcode RNAs are priming each other; data not shown) before settling on barcode joining by emulsion reverse transcription-PCR (RT-PCR) followed by overlap PCR, which had previously proven successful for joining immunoglobulin mRNAs (29). Briefly, we isolate synPRE-P—synPOST-P complexes by IP and emulsify the complexes using a microfluidic droplet system at a dilution such that each droplet contains no more than one complex. Inside each droplet, we then reverse transcribe the presynaptic

and postsynaptic barcode mRNAs and subsequently join them into a single DNA strand by overlap PCR (29). We then break the emulsion, isolate the barcode pairs and sequence them on an Illumina sequencing machine (Figure 6A, Supplementary Figure S6).

The success and failure of any method to join barcode mRNAs in SYNseq is measured not simply in the generation of barcode pairs, but critically in the ability of producing barcode pairs from mRNAs in the same complex, but not from mRNAs in different complexes or even unbound

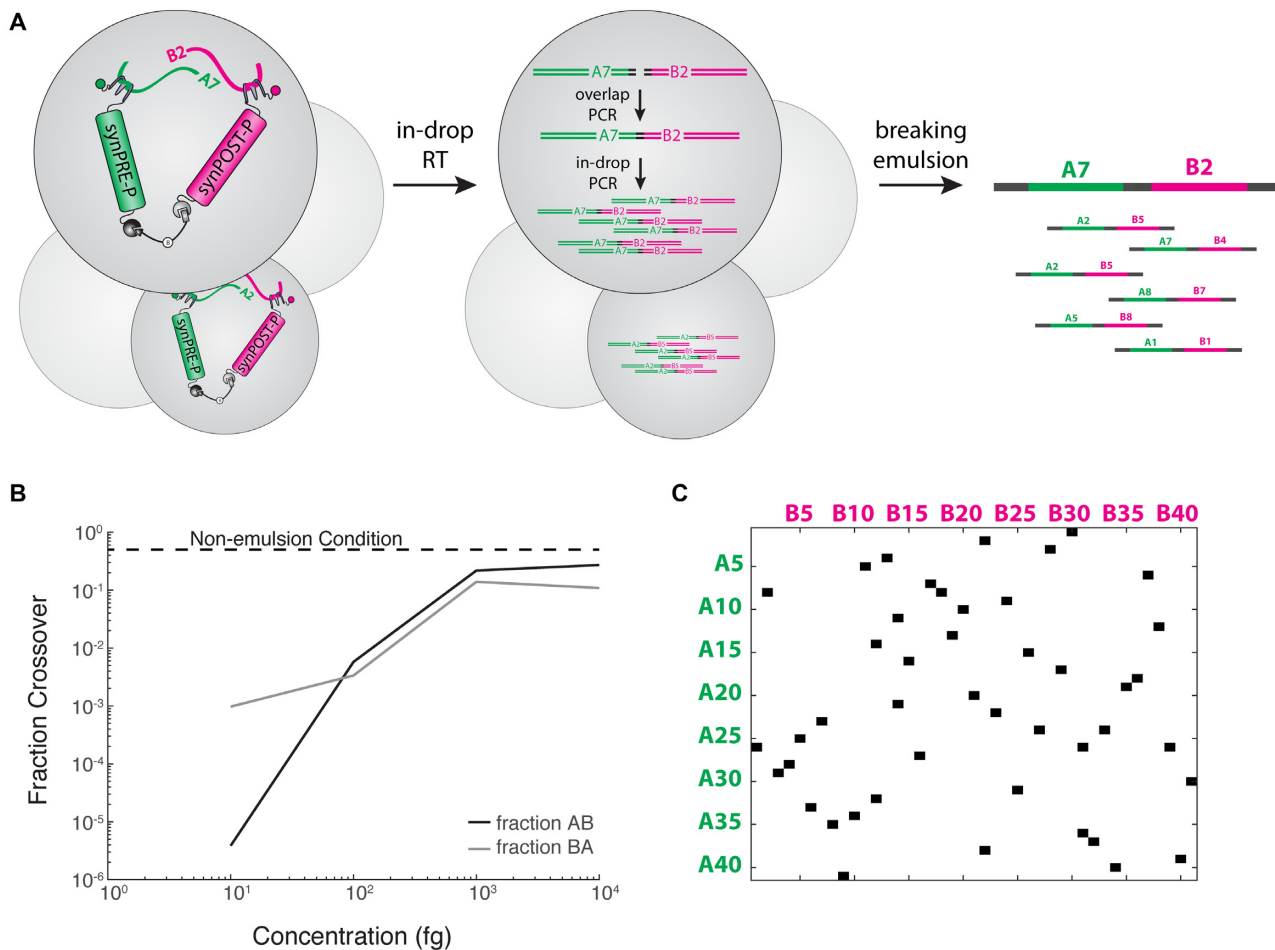


Figure 6. Emulsion RT-PCR can be used to join barcodes into barcode pairs. **(A)** Schematic of emulsion, in-drop RT and overlap PCR followed by breaking of the emulsion and purification of barcode pairs. **(B)** Dilution of input RNA decreases the false positive rate of emulsion PCR. Two separate barcode pairs (A–A and B–B) are subjected to uniform emulsions simultaneously to probe the concentrations at which overlap PCR becomes promiscuous. Fraction = 0.5 is equivalent to an aqueous (non-emulsified) PCR reaction. Here, RNA molecules containing both a pre- and post-segment were employed. **(C)** ‘Connectivity matrix’ obtained from emulsion RT-PCR followed by Illumina sequencing of synPRE-P–synPOST-P complexes from HEK cell culture, demonstrating that in principle SYNseq can be used to reconstruct synaptic circuits.

mRNAs. When using emulsion RT-PCR to join barcode mRNAs, such nonspecific joining occurs when more than one synPRE-P–synPOST-P complex, and thus more than one presynaptic or postsynaptic barcode mRNA are loaded into the same droplet. The fraction of unspecific joining events therefore ultimately depends on the concentration of complexes that are loaded into droplets and decreases with input concentration, as demonstrated by loading droplets with different concentrations of a test RNA that contains the sequences of both pre- and postsynaptic barcodes, but in inverted orientations relative to the final sequencing amplicon (Figure 6B, Supplementary Figure S7A).

To validate the emulsion RT-PCR approach, we first tested joining using protein-RNA complexes derived from HEK cell culture. To measure the rate of nonspecific joining of barcode mRNAs, we set up two experiments in parallel. Each experiment consists of one population of HEK cells that co-express synPRE-P, synPOST-P and barcode mRNAs. In these experiments we used an earlier version of the SYNseq proteins. The barcode mRNAs in

each experiment are tagged with an additional, experiment-specific sequence, a zip code (denoted here as A or B). We crosslinked and lysed each set of cells independently, and pooled the lysates. We then immunoprecipitated synPRE-P–synPOST-P complexes from the mixed lysate by streptavidin IP, released the complexes from the beads by cleaving the disulfide bridge on the crosslinker under mild reducing conditions (Supplementary Figure S7B) and finally joined barcode mRNAs using emulsion RT-PCR. Sequencing the resulting barcode pairs revealed 42 presynaptic and 42 postsynaptic barcodes, which were joined into 45 unique barcode pairs (Figure 6C).

Based on the experimental setup, barcode pairs should show matching zip codes (i.e. A–A pairs or B–B pairs), as those are the only biologically possible combinations. Any mixed barcode pairs (A–B or B–A) must arise from ‘crossover’ events and therefore represent the rate of false-positives in recovery of the correct set of synaptic connections by SYNseq. In our HEK cell experiment, we find a significant crossover rate of 13.3% (barcodes from one ex-

periment joined to barcodes of another experiment). Extrapolated to include undetectable A–A and B–B crossover events, this indicates an overall false-positive rate of 26.6%. This false-positive rate, importantly, only measures false-positives arising from biochemical procedures post-lysis (i.e. protein–RNA dissociation, insufficient dilution for emulsions, etc.). It cannot quantify the false-positive barcode pairs that result from aberrant localization of the proteins or any other *in vivo* effects. The false-positive rate of this experiment appears to be dominated by the emulsion input concentration (Supplementary Figure S7C), suggesting that we could reduce the false positive rate by further dilution of the complexes before emulsion.

We next tested emulsion RT-PCR for barcode joining in samples from neuronal cultures. We infected XONA devices with barcoded synPRE-P and synPOST-P Sindbis in the left and right chamber, respectively. Again, we set up two replicate experiments in parallel using experiment specific zip codes to differentiate the barcode mRNAs. We crosslinked, lysed and mixed the lysates from the two sets of XONA devices and performed a double immunoprecipitation to isolate synaptic synPRE-P–synPOST-P complexes. We then performed emulsion RT-PCR on a very dilute sample of the IP products to minimize the rate of crossover events.

Sequencing the resulting barcode pairs, we identified only 2 unique combinations of pre- and postsynaptic barcodes, albeit no false positive cross-experiment barcode pairs. While it is likely possible to increase this number of joined barcode pairs by increasing the number of droplets made, or by slightly increasing the concentration of emulsion input, we found ourselves in a regime far from the optimal necessary to achieve high throughput connectivity mapping by SYNseq. Taken together, our emulsion results suggest, that while SYNseq can be used to read out synaptic connectivity, barcode joining is currently a very low efficiency process, which will need to be optimized further before SYNseq can be used for circuit mapping.

DISCUSSION

We here present SYNseq, a scheme for converting synaptic connectivity into a form that can be deciphered using modern sequencing technology. SYNseq represents a significant step in the development of a method to map neuronal circuits with high throughput. We established each of the required biochemical components for SYNseq, namely barcoding (17), protein-joining, and RNA barcode joining (Figure 1), but further work—in particular on barcode joining—will be needed to enable the robust tracing of neuronal circuits. We expect that an efficient and high-throughput method for determining both the local and long-range synaptic connectivity of neuronal populations would represent a transformative technology for research on neural circuits.

DNA sequencing for neuroanatomy

Historically, the method of choice for inferring neuroanatomical connectivity has been microscopy. The specific form of microscopy used is determined by the resolution required to resolve the smallest important feature of the

circuit in question. This implies that the smaller the structure of interest, the greater the resources and effort needed to perform the requisite imaging. Imaging therefore renders high-resolution neuroanatomic studies low throughput and limits the ability to scale to larger circuits. Because of these technical limitations, we currently have detailed circuit information about only a very limited number of mammalian circuits (6,41–43).

By applying modern high-throughput sequencing technology to neuroanatomical problems, we are attempting to overcome the imaging and tracing bottlenecks in high-resolution circuit analysis. DNA sequencing has undergone tremendous improvements in speed, throughput and cost in the last decade: while sequencing a human genome cost 3 billion dollars in 2000, it now costs <1000 dollars, and prices continue to fall at a rate exceeding Moore's law (44,45). The exceptional speed, parallelization and flexibility of modern sequencing technology thus motivate our attempts to translate the circuit features of interest into a format that is accessible to sequencing.

We recently presented MAPseq, the first application of DNA sequencing to neuroanatomy. MAPseq permits the simultaneous read out of thousands and potentially millions of single neuron long-range projections in a single mouse brain (17). With SYNseq, we now provide a foundation for the considerably more challenging problem of determining synaptic connectivity. SYNseq takes inspiration from GRASP/mGRASP (46–48), chromosome confirmation capture (49,50) and synaptic brainbow (51). With further development, SYNseq may allow the routine mapping of complete circuits or even whole brains, on a time scale of weeks and at a cost measured in thousands rather than millions of dollars. The equipment needed for SYNseq is routinely available in many laboratories, and does not require major capital expenditures (with the ability to leverage existing core facilities and/or CROs for sequencing expertise). The high-throughput of SYNseq may make it possible to compare circuits across areas, species and treatments, and may lead to high-throughput screens for drugs or targets relevant to neurological disorders.

Challenges ahead

In the present study, we attempted to use emulsion overlap RT-PCR to join the pre- and postsynaptic RNAs held in one complex. Although we were able to show that this is technically possible (Figure 6), we were unable to recover many barcode pairs after joining, rendering SYNseq in its current form unusable for answering biological questions. We concluded by a process of elimination that it is barcode joining that is limiting our ability to detect synaptic connections by SYNseq. Barcoding of neurons and HEK cells is a high efficiency process (17). We further demonstrated that barcode RNA and SYNseq proteins interact strongly and specifically (Figure 2), and that we can produce a significant number of trans-synaptic synPRE-P–synPOST-P complexes (Figure 5). Despite this, we are unable to generate a comparable number of joined barcode pairs, which we therefore blame on inefficiencies of the droplet based joining strategy. We have not attempted to directly de-convolve the inefficiencies of reverse transcription vs. overlap PCR, but

preliminary evidence place our in-droplet RT efficiency on the order of 1%. Such low RT efficiency will have a dramatic impact on our ability to produce joint barcode pairs, as both pre and postsynaptic barcode mRNAs need to be successfully reverse transcribed in each successfully loaded droplet (this will happen in only $0.01^2 = 10^{-4} = 0.01\%$ of all properly loaded droplets). Given these considerations, we predict reverse transcription, not overlap PCR, which has a chance to join the two barcodes at every cycle of PCR, to be limiting the reaction. Although we were unable to achieve efficient and specific joining of barcodes in droplets, it is possible that with further optimization a droplet-based strategy, perhaps in combination with barcoded beads (52), might succeed. Alternatively, joining strategies based on RNA ligation, gap filling or immobilization of complexes on beads, chips or in gels might emerge as more efficient strategies. Finally, sequencing assays relying on the joining of barcodes (albeit in DNA form) are being developed for other applications (53–55), suggesting that in the future new technologies for barcode joining may emerge. Thus, although our present inability to achieve efficient and specific barcode joining limits the practical application of this approach, we expect the approach we have presented will form that foundation for an efficient method for deducing synaptic connections.

Another concern is that even brief (<48 h) expression of the synPRE-P and synPOST-P proteins, which are derived from synaptic proteins involved in synapses stabilization (56), could lead to formation of spurious synaptic connections. To circumvent this, the Nr1B–Nlg1AB interaction domains could be eliminated (47), or similar approaches could be applied to other synaptic proteins that do not interact. Finally, all data presented here are based on using Sindbis virus to deliver barcodes and carrier proteins to neurons. Sindbis was a convenient choice because of its large (>6 kb) payload, rapid (<48 h) expression and the ease of engineering. However, for some applications it may be advantageous to use a less toxic DNA virus amenable to cre-dependent control, such as adeno-associated virus (AAV) or herpes simplex virus (HSV) amplicons (57).

Before SYNseq can be used for circuit mapping procedures in the future, it will need to be validated. While ‘ground truth’ for synaptic connectivity is difficult to achieve, we envision a number of avenues for validation. First, statistical comparisons of circuit features between SYNseq and EM or physiology derived datasets can act as an initial check. For a more direct comparison, we it may be possible to determine the pairwise connectivity of barcoded neurons by electrophysiology, extract the cell body barcodes from the patched neurons, preform SYNseq on the remaining tissue, and then map the physiologically derived subset of connections on the SYNseq dataset.

Space and neuroanatomy

In its simplest incarnation, SYNseq fails to preserve the spatial relationships among neurons and their connections. Spatial resolution can, however, be recovered in a variety of ways, including dissection (16,17). Most promisingly, emerging *in situ* sequencing methods developed by our lab and others (Chen *et al.*, in preparation) (58) allow us to

combine the advantages of high throughput sequencing with the spatial resolution required to ask many biological questions.

Future directions

We envision that SYNseq will serve as a platform for future development toward encoding neural connectivity, and/or other biological variables—including cellular activity and/or identity—into a form that can be read by high-throughput sequencing, a technology which already operates at the scale of the complexity of neural circuits. Improvements to this method that allow for the co-expression of all of the components within the same cells—for example, by employing two separate RNA binding domains (Supplementary Figure S5)—will allow for bi-directional and local-circuit mapping. Transgenic techniques, combined with an *in vivo* barcoding scheme based on recombinases or CRISPR cas9 (59–61) will allow for scaling this technology to mapping full brains.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Author contributions: I.D.P., J.M.K., D.B. and A.M.Z. conceived the study. I.D.P., J.M.K., V.V.V., D.I.R. and Y.S. performed the experiments. I.D.P., E.B. and D.I.R. performed the emulsion experiments. I.R.C. designed and synthesized the bifunctional crosslinkers. J.M.K. and A.M.Z. wrote the paper. A.M.Z. supervised the project.

FUNDING

National Institutes of Health (NIH) [5R01NS073129 to A.M.Z., 5R01DA036913 to A.M.Z., 5U01MH109113 to A.M.Z., R01CA181595 to E.B.]; Brain Research Foundation [BRF-SIA-2014-03 to A.M.Z.]; IARPA [MICrONS to A.M.Z.]; Simons Foundation [382793/SIMONS to A.M.Z.]; Paul Allen Distinguished Investigator Award [to A.M.Z.]; PhD fellowship from the Boehringer Ingelheim Fonds to J.M.K.; PhD fellowship from the Genentech Foundation to J.M.K. Funding for open access charge: NIH [5R01NS073129].

Conflict of interest statement. I.R.C. is employee of New England Biolabs, Inc., which commercializes SNAP-tag and CLIP-tag reagents, some of which were used in this research. The remaining authors declare no competing financial interests.

REFERENCES

- Ahrens, M.B., Orger, M.B., Robson, D.N., Li, J.M. and Keller, P.J. (2013) Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nat. Methods*, **10**, 413–420.
- Sofroniew, N.J., Flickinger, D., King, J. and Svoboda, K. (2016) A large field of view two-photon mesoscope with subcellular resolution for *in vivo* imaging. *Elife*, **5**, 264–266.
- Kleinfeld, D., Bhardwaj, A., Blinder, P., Bock, D.D., Briggman, K.L., Chklovskii, D.B., Denk, W., Helmstaedter, M., Kaufhold, J.P., Lee, W.C.A. *et al.* (2011) Large-scale automated histology in the pursuit of connectomes. *J. Neurosci.*, **31**, 16125–16138.

4. Briggman, K.L., Helmstaedter, M. and Denk, W. (2012) Wiring specificity in the direction-selectivity circuit of the retina. *Nature*, **471**, 183–188.
5. Helmstaedter, M. (2013) Cellular-resolution connectomics: challenges of dense neural circuit reconstruction. *Nat. Methods*, **10**, 501–507.
6. Helmstaedter, M., Briggman, K.L., Turaga, S.C., Jain, V., Seung, H.S. and Denk, W. (2013) Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, **500**, 168–174.
7. Kasthuri, N., Hayworth, K.J., Berger, D.R., Schalek, R.L., Conchello, J.A., Knowles-Barley, S., Lee, D., Vázquez-Reina, A., Kaynig, V., Jones, T.R. *et al.* (2015) Saturated reconstruction of a volume of neocortex. *Cell*, **162**, 648–661.
8. Eberle, A.L., Mikula, S., Schalek, R., Lichtman, J., Knothe Tate, M.L. and Zeidler, D. (2015) High-resolution, high-throughput imaging with a multibeam scanning electron microscope. *J. Microsc.*, **259**, 114–120.
9. Kaynig, V., Vázquez-Reina, A., Knowles-Barley, S., Roberts, M., Jones, T.R., Kasthuri, N., Miller, E., Lichtman, J. and Pfister, H. (2015) Large-scale automatic reconstruction of neuronal processes from electron microscopy images. *Med. Image*, **22**, 77–88.
10. Bock, D.D., Lee, W.-C.A., Kerlin, A.M., Andermann, M.L., Hood, G., Wetzell, A.W., Yurgenson, S., Soucy, E.R., Kim, H.S. and Reid, R.C. (2012) Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, **471**, 177–182.
11. Song, S., Sjöström, P.J., Reigl, M., Nelson, S. and Chklovskii, D.B. (2005) Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.*, **3**, e68.
12. Yoshimura, Y. and Callaway, E.M. (2005) Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. *Nat. Neurosci.*, **8**, 1552–1559.
13. Jiang, X., Shen, S., Cadwell, C.R., Berens, P., Sinz, F., Ecker, A.S., Patel, S. and Tolias, A.S. (2015) Principles of connectivity among morphologically defined cell types in adult neocortex. *Science*, **350**, aac9462.
14. Jiang, X., Wang, G., Lee, A.J., Stornetta, R.L. and Zhu, J.J. (2013) The organization of two new cortical interneuronal circuits. *Nat. Neurosci.*, **16**, 210–218.
15. Kodandaramiah, S.B., Franzesi, G.T., Chow, B.Y., Boyden, E.S. and Forest, C.R. (2012) Automated whole-cell patch-clamp electrophysiology of neurons in vivo. *Nat. Methods*, **9**, 585–587.
16. Zador, A.M., Dubnau, J., Oyibo, H.K., Zhan, H., Cao, G. and Peikon, I.D. (2012) Sequencing the connectome. *PLoS Biol.*, **10**, e1001411.
17. Kechschul, J.M., da Silva, P., Reid, A.P., Peikon, I.D., Albeanu, D.F. and Zador, A.M. (2016) High-throughput mapping of single-neuron projections by sequencing of barcoded RNA. *Neuron*, **91**, 975–987.
18. Kechschul, J.M., da Silva, P. and Zador, A.M. (2016) A new defective helper RNA to produce recombinant sindbis virus that infects neurons but does not propagate. *Front. Neuroanat.*, **10**, 6439.
19. Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Publ. Gr.*, **6**, 343–345.
20. Daigle, N. and Ellenberg, J. (2007) $[\lambda]$ N-GFP: an RNA reporter system for live-cell imaging. *Nat. Methods*, **4**, 633–636.
21. Gautier, A., Juillerat, A., Heinis, C., Corrêa, I.R. Jr, Kindermann, M., Beaufils, F. and Johnsson, K. (2008) An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.*, **15**, 128–136.
22. Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H. and Johnsson, K. (2002) A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.*, **21**, 86–89.
23. Sun, X., Dusserre-Bresson, F., Baker, B., Zhang, A., Xu, P., Fibbe, C., Noren, C.J., Corrêa, I.R. and Xu, M.-Q. (2014) Probing homodimer formation of epidermal growth factor receptor by selective crosslinking. *Eur. J. Med. Chem.*, **88**, 34–41.
24. Bottos, A., Destro, E., Rissone, A., Graziano, S., Cordara, G., Assenzio, B., Cera, M.R., Mascia, L., Bussolino, F. and Aresè, M. (2009) The synaptic proteins neurexins and neuroligins are widely expressed in the vascular system and contribute to its functions. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 20782–20787.
25. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, **25**, 402–408.
26. Brouzes, E. (2012) Droplet microfluidics for single-cell analysis. *Methods Mol. Biol.*, **853**, 105–139.
27. Holtze, C., Rowat, A.C., Agresti, J.J., Hutchison, J.B., Angilè, F.E., Schmitz, C.H.J., Köster, S., Duan, H., Humphry, K.J., Scanga, R.A. *et al.* (2008) Biocompatible surfactants for water-in-fluorocarbon emulsions. *Lab Chip*, **8**, 1632–1639.
28. Tewhey, R., Warner, J.B., Nakano, M., Libby, B., Medkova, M., David, P.H., Kotsopoulos, S.K., Samuels, M.L., Hutchison, J.B., Larson, J.W. *et al.* (2009) Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat. Biotechnol.*, **27**, 1025–1031.
29. DeKosky, B.J., Ippolito, G.C., Deschner, R.P., Lavinder, J.J., Wine, Y., Rawlins, B.M., Varadarajan, N., Giesecke, C., Dörner, T., Andrews, S.F. *et al.* (2013) High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat. Biotechnol.*, **31**, 166–169.
30. Kechschul, J.M. and Zador, A.M. (2015) Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Res.*, **43**, e143.
31. Fairless, R., Masius, H., Rohlmann, A., Heupel, K., Ahmad, M., Reissner, C., Dresbach, T. and Missler, M. (2008) Polarized targeting of neurexins to synapses is regulated by their C-terminal sequences. *J. Neurosci.*, **28**, 12969–12981.
32. Slavoff, S.A., Liu, D.S., Cohen, J.D. and Ting, A.Y. (2011) Imaging protein-protein interactions inside living cells via interaction-dependent fluorophore ligation. *J. Am. Chem. Soc.*, **133**, 19769–19776.
33. Dresbach, T., Neeb, A., Meyer, G., Gundelfinger, E.D. and Brose, N. (2004) Synaptic targeting of neuroligin is independent of neurexin and SAP90/PSD95 binding. *Mol. Cell. Neurosci.*, **27**, 227–235.
34. Sara, Y., Biederer, T., Atasoy, D., Chubykin, A., Mozhayeva, M.G., Südhof, T.C. and Kavalali, E.T. (2005) Selective capability of SynCAM and neuroligin for functional synapse assembly. *J. Neurosci.*, **25**, 260–270.
35. Hahn, C.S., Hahn, Y.S., Braciale, T.J. and Rice, C.M. (1992) Infectious Sindbis virus transient expression vectors for studying antigen processing and presentation. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 2679–2683.
36. Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.-J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.-G. *et al.* (2006) Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods*, **3**, 995–1000.
37. Schikorski, T. and Stevens, C.F. (1997) Quantitative ultrastructural analysis of hippocampal excitatory synapses. *J. Neurosci.*, **17**, 5858–5867.
38. Taylor, A.M., Blurton-Jones, M., Rhee, S.W., Cribbs, D.H., Cotman, C.W. and Jeon, N.L. (2005) A microfluidic culture platform for CNS axonal injury, regeneration and transport. *Nat. Methods*, **2**, 599–605.
39. Taylor, A.M., Dieterich, D.C., Ito, H.T., Kim, S.A. and Schuman, E.M. (2010) Microfluidic local perfusion chambers for the visualization and manipulation of synapses. *Neuron*, **66**, 57–68.
40. Korogod, N., Petersen, C.C.H. and Knott, G.W. (2015) Ultrastructural analysis of adult mouse neocortex comparing aldehyde perfusion with cryo fixation. *Elife*, **4**, 241.
41. White, J.G., Southgate, E., Thomson, J.N. and Brenner, S. (1986) The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **314**, 1–340.
42. Jarrell, T.A., Wang, Y., Bloniarz, A.E., Brittin, C.A., Xu, M., Thomson, J.N., Albertson, D.G., Hall, D.H. and Emmons, S.W. (2012) The connectome of a decision-making neural network. *Science*, **337**, 437–444.
43. Takemura, S., Bharioke, A., Lu, Z., Nern, A., Vitaladevuni, S., Rivlin, P.K., Katz, W.T., Olbris, D.J., Plaza, S.M., Winston, P. *et al.* (2013) A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature*, **500**, 175–181.
44. Sheridan, C. (2014) Illumina claims \$1,000 genome win. *Nat. Biotechnol.*, **32**, 115.
45. Hayden, E.C. (2014) Technology: The \$1,000 genome. *Nature*, **507**, 294–295.
46. Feinberg, E.H., VanHoven, M.K., Bendsky, A., Wang, G., Fetter, R.D., Shen, K. and Bargmann, C.I. (2008) GFP reconstitution across synaptic partners (GRASP) defines cell contacts and synapses in living nervous systems. *Neuron*, **57**, 353–363.

47. Kim,J., Zhao,T., Petralia,R.S., Yu,Y., Peng,H., Myers,E. and Magee,J.C. (2011) mGRASP enables mapping mammalian synaptic connectivity with light microscopy. *Nat. Methods*, **9**, 96–102.
48. Yamagata,M. and Sanes,J.R. (2012) Transgenic strategy for identifying synaptic connections in mice by fluorescence complementation (GRASP). *Front. Mol. Neurosci.*, **5**, 18.
49. Dekker,J. (2002) Capturing chromosome conformation. *Science*, **295**, 1306–1311.
50. Kalhor,R., Tjong,H., Jayathilaka,N., Alber,F. and Chen,L. (2011) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.*, **30**, 90–98.
51. Mishchenko,Y., Ward,E., Ober,R., Peters,R. and Schlager,M. (2010) On optical detection of densely labeled synapses in neuropil and mapping connectivity with combinatorially multiplexed fluorescent synaptic markers. *PLoS One*, **5**, e8853.
52. Macosko,E.Z., Basu,A., Satija,R., Nemes,J., Shekhar,K., Goldman,M., Tirosh,I., Bialas,A.R., Kamitaki,N., Martersteck,E.M. *et al.* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, **161**, 1202–1214.
53. Hastie,A.R. and Pruitt,S.C. (2007) Yeast two-hybrid interaction partner screening through in vivo Cre-mediated binary interaction Tag generation. *Nucleic Acids Res.*, **35**, e141.
54. Jaffe,M., Sherlock,G. and Levy,S.F. (2017) iSeq: a new double-barcode method for detecting dynamic genetic interactions in yeast. *G3*, **7**, 143–153.
55. Yachie,N., Petsalaki,E., Mellor,J.C., Weile,J., Jacob,Y., Verby,M., Ozturk,S.B., Li,S., Cote,A.G., Mosca,R. *et al.* (2016) Pooled-matrix protein interaction screens using barcode fusion genetics. *Mol. Syst. Biol.*, **12**, 863.
56. Südhof,T.C. (2008) Neuroligins and neuroligins link synaptic function to cognitive disease. *Nature*, **455**, 903–911.
57. Neve,R., Neve,K., Nestler,E. and Carlezon,W. Jr (2005) Use of herpes virus amplicon vectors to study brain disorders. *Biotechniques*, **39**, 381–391.
58. Lee,J.H., Daugherty,E.R., Scheiman,J., Kalhor,R., Yang,J.L., Ferrante,T.C., Terry,R., Jeanty,S.S.F., Li,C., Amamoto,R. *et al.* (2014) Highly multiplexed subcellular RNA sequencing in situ. *Science*, **343**, 1360–1363.
59. Peikon,I.D., Gizatullina,D.I. and Zador,A.M. (2014) In vivo generation of DNA sequence diversity for cellular barcoding. *Nucleic Acids Res.*, **42**, e127.
60. Kalhor,R., Mali,P. and Church,G.M. (2016) Rapidly evolving homing CRISPR barcodes. *Nat. Methods*, **14**, 195–200.
61. McKenna,A., Findlay,G.M., Gagnon,J.A., Horwitz,M.S., Schier,A.F. and Shendure,J. (2016) Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907.