CHAPTER 2

MATERIALS AND METHODS

This chapter describes the general methods used in this project

2.1 Plasmids and cloning

2.1.1 Generation of different dCas9 fusion constructs

Expression vector pPB-R1R2_EF1adCas9VP64_T2A_MS2p65HSF1-IRESbsdpA and an entry vector carrying the same construct were provided by the Yusa Lab (Wellcome Sanger Institute) and expression vector pMCV-EF1a_grow_dCas9-GFP_Blast_pA was obtained from the Bradley Lab (Wellcome Sanger Institute). p300 core (histone acetyltransferase) domain was synthesised as several gBlock DNA fragments (Integrated DNA Technologies) containing homology arms for insertion into lenti dCAS-VP64_Blast from the Zhang lab (Addgene plasmid #61425) (Konermann et al., 2015). The p300 core domain sequence used was previously published in Hilton et al. (2015).

Expression constructs encoding different dCas9 fusion proteins were generated as shown in Figure 2.1. For C-terminal insertion (relative to dCas9) of the p300 core domain, vector lenti dCAS-VP64_Blast was linearised by digestion with BamHI and EcoRI. The linearised vector was assembled with corresponding gBlock DNA fragments using Gibson assembly and the resulting plasmids digested with XbaI. This allowed for insertion of a

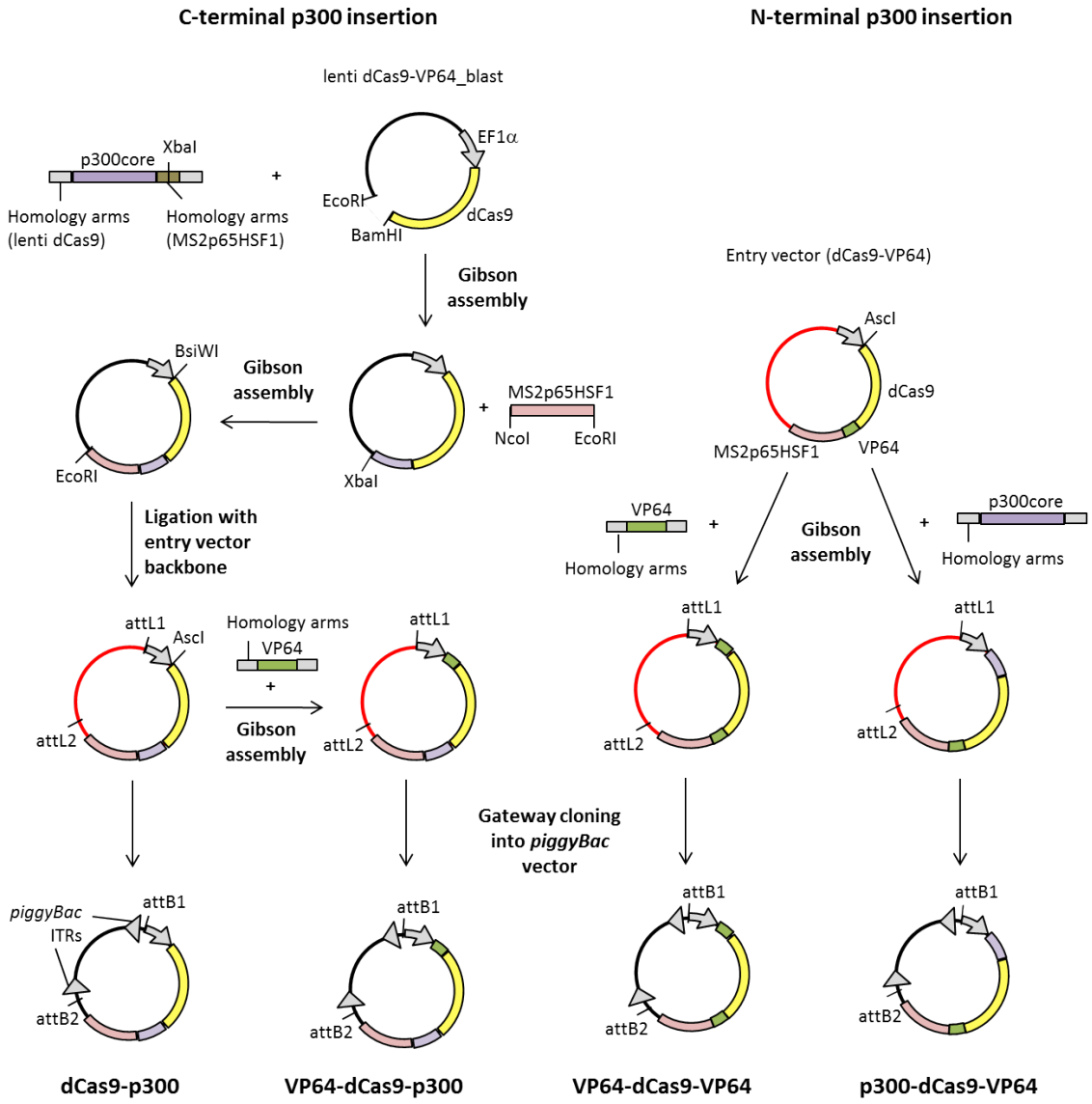

Figure 2.1 Construction of dCas9-fusions containing additional p300core or VP64 domains Separate strategies were used for N-terminal and C-terminal insertion of p300core or VP64 fragments. A lentiviral dCas9-VP64 vector from the Zhang lab was modified to include a C-terminal p300core fusion and expression of synegystic activators MS2-p65-HSF1 by gibson assembly before transfer into a PiggyBac expression vector by Gateway cloning. An additional N-terminal insertion of VP64 was made to create a VP64-dCas9-p300 fusion construct. A Gateway entry vector containing dCas9-VP64 constructed by Kosuke Yusa lab was modified to include an N-terminal insertion of either p300core or VP64 by gibson assembly before transfer into a PiggyBac expression vector by Gateway cloning.

MS2p65HSF1 construct obtained by digesting vector pPB-R1R2_EF1adCas9VP64_T2A_MS 2p65HSF1-IRESbsdpA with NcoI and EcoRI. The resulting dCas9-fusion-T2A-MS2p65HSF1 constructs were subsequently cloned into a kanamycin-resistant entry vector using restriction enzyme digest with BsiWI and EcoRI (NEB). VP64 or p300 domains were inserted upstream of dCas9 coding sequence by Gibson assembly with the entry vector digested with AscI. Gateway cloning was performed to transfer the dCas9-fusion-T2A-MS2p65HSF1 constructs into the final ampicillin-resistant expression vector pPB-R1R2-IRESbsdpA. VP64 fragments were PCR amplified from lenti dCAS-VP64_Blast. NEBuilder HiFi DNA Assembly Cloning Kit (NEB) was used for all Gibson assembly reactions and conducted according to manufacturer's protocol. Final dCas9 fusion constructs were Sanger sequenced (Eurofins Genomics) to ensure correct positions of activator domains as well as confirm sequences of PCR-amplified and synthesised fragments.

All enzymatic digestions were performed in 50 μL reaction volumes with 5 μg DNA, 5 μL 10x digestion buffer, 10-20 units of each restriction enzyme, and incubated at 37 °C for at least 6 h. 5' dephosphorylation was achieved by incubation with Antartic Phosphatase (NEB) for 30 min at 37 °C followed by inactivation for 5 min at 80 °C. Digested products were separated by gel electrophoresis on a 1.2% agarose gel and the desired fragments purified using Qiagen Gel Purification kit (Qiagen).

PCR reactions for generating VP64 and MS2p65HSF1 fragments were performed in 25 μL reaction volumes with 12.5 μL 2x Q5 Hotstart Hifi Master Mix (NEB), 1 μL each 10 mM sense and antisense primers, 1μL (1 μg) template DNA and 9.5 μL nuclease-free water (Ambion). The PCR reactions were performed in a Tetrad 2 Thermal Cycler (Bio-Rad) and cycling conditions were as follows: 30 s at 95 °C for initial denaturation, followed by 25 cycles of 30 s at 95 °C for denaturation, 30 s at 60 °C for annealing, 90 s at 72 °C for extension, and 5 min at 72 °C for the final extension. PCR product clean-up was performed with Qiagen PCR Purification kit (Qiagen). All sequencing primers, PCR primers and gBlock sequences are listed in Appendix A.

2.1.2 Individual gRNA cloning

For the panel of 12 cell surface receptors, potential guides were identified and ranked using CRISPR-ERA. CRISPR-ERA ranked sequences based on an on-target S score based on distance to the transcriptional start site (TSS), and an off-target E score based on number of off-target sites (Liu et al., 2015). Eight non-overlapping guides most proximal

to the TSS of the longest RefSeq isoform were chosen for each gene. Guides targeting the same gene were cloned as a pool using One Shot TOP10 Chemically Competent *E. coli* (Invitrogen) and propagated in liquid culture. In all other experiments, guides were cloned individually and sequence verified before lentiviral production or transfection into cells. The sequences of individual guides mentioned in this thesis are listed in Appendix A.

Guide RNA with an improved scaffold (Chen et al., 2013) and MS2-binding hairpin loops were expressed from a U6 promoter on an expression vector provided by the Yusa Lab. Individual guides were synthesised as 24 bp oligomers (Sigma Aldrich and IDT) containing complementary overhangs to those generated by BbsI digestion of the gRNA expression vector. These oligomers underwent 5' phosphorylation by treatment with T4 PNK (NEB) for 30 min at 37 °C prior to annealing in 1x T4-ligation buffer (NEB) by incubating for 50 min at 95 °C before slowly decreasing the temperature to 25 °C at 0.1 °C/s. Annealed oligos were ligated into the lentiviral sgRNA vector by incubating with T4 DNA Ligase (NEB) for 4 h at 16 °C. All gRNA sequences used except those from the membrane protein gRNA library are listed in Appendix A.

2.1.3 RNA isolation and q-RT-PCR

Relative mRNA expression levels were quantified by reverse transcription and quantitative PCR (qPCR). Total RNA was isolated from approximately 5 x 10^6 cells per sample using TRIzol Reagent (Ambion) according to manufacturer's protocol. 1 μg total RNA was reverse transcribed using SuperScript III First-Strand Synthesis Kit (Invitrogen), and remaining RNA was removed by incubation with RNase H for 20 min at 37 °C. The resultant cDNA was diluted 30-fold in nuclease-free water. qPCR was performed using Sensimix SYBR Low-Rox Kit (Bioline) with 5 μL of diluted cDNA in a final reaction volume of 15 μL . Samples were prepared in 384-well format with two technical replicates for every RNA sample and cycled on a LightCycler 480 Instrument II. Cycling parameters were as follows: 10 min at 95 °C for polymerase activation, followed by 40 cycles of 15 s at 95 °C for denaturation, 15 s at 55 °C for annealing, and 15 s at 72 °C for extension. A melt-curve analysis (from 25 °C to 95 °C) was performed at the end of the run to check for the presence of primer-dimers or other unwanted products.

Primers targeting *GAPDH*, Cyclophilin A have been previously published and were used as housekeeping controls (Hellebrekers et al., 2006; Hilton et al., 2015). All other primers were designed using Primer-BLAST (Ye et al., 2012), with the exception of IL1RN

primers which have also been previously published in Cheng et al. (2013). All qPCR primers used are listed in Appendix A. Threshold cycle (Ct) values were determined by the number of cycles needed to reach an arbitrary fluorescence threshold set just above baseline. Relative mRNA expression was determined using the $2^{\Delta\Delta Ct}$ method where target Ct values were first normalized to *GAPDH* and Cyclophilin A Ct values, which are not expected to change between samples (Livak and Schmittgen, 2001). Fold changes in target gene mRNA levels were then determined by comparing to mock-transfected experimental controls. Student's t-test was performed in R.

2.2 CRISPRa gRNA library construction

2.2.1 Computational selection of gRNAs

Genes encoding membrane proteins were compiled from five sources: a mass-spectrometry derived Cell Surface Atlas (Bausch-Fluck et al., 2015), a bioinformatic construction of the surfaceome (da Cunha et al., 2009), a manually curated list of proteins with experimentally verified cell surface localisation (Laura Wood, Wellcome Sanger Institute, personal communication), the transmembrane protein cDNA set sold by Origene, and the Human Protein Atlas (filtered for location: plasma membrane) (Uhlén et al., 2015).

The final number of genes targeted was 6,213. TSS predictions were selected from Gencode v19 TSS stratified by strict Fantom5 CAGE clusters, and the two broadest peaks per gene were selected (Harrow et al., 2012). For genes that were not associated with a CAGE peak, ENSEMBL transcripts annotated as 'principal' in the APPRIS database were selected instead (Rodriguez et al., 2013). Where no transcripts with this criterion were found, all RefSeq transcripts with NM accession numbers were selected.

Promoter region sequences (450 – 50 bp upstream of each TSS) were obtained from the human assembly hg19 in Ensembl using the `BiomaRt` package (Durinck et al., 2009). All 19 nucleotide sequences adjacent to an NGG protospacer adjacent motif (PAM) within these sequences were identified. Guides with <30% or >75% GC content, polyT sequences, or BbsI restriction sites were discarded, and the resulting guides were ranked according to proximity to the TSS peak.

Each guide was mapped using BLAT (Kent, 2002) to all promoter regions targeted and guides with exact matches to promoters other than their intended target were removed, with the exception of those targeting genes with shared promoter regions or gene families with similar promoter sequences. As far as possible, 7 guides were selected per transcript/peak. Where a gene had 6 guides or fewer, rules concerning GC content and polyT stretches were relaxed such that every transcript had at least 2 guides, with only 8 genes having 2 guides per gene. To ensure a high level of transcription by the U6 promoter, a guanine nucleotide was added to the 5' end of all guide sequences (Ma et al., 2014).

Non-targeting gRNA sequences were selected from gRNA sequences previously published in Wang et al. (2015), and were designed to have no binding sites in the human genome (up to two mismatches).

2.2.2 gRNA Library synthesis and cloning

58,570 gRNA sequences were synthesised as a complex pool of 77-mer single-stranded DNA oligos (Twist Biosciences). Each 77-mer oligo contained the guide sequence as well as asymmetrical flanking regions for primer annealing and amplification (Figure 2.2). Double-stranded DNA was amplified from 40 ng of ssDNA oligos using primer pair 77-mer_U1 and 77-mer_L1 in 40 PCR reactions. Each reaction contained 1 ng ssDNA, 1.25 uL of each primer at 10 μM, 12.5 μL Q5 2x High Fidelity Hot-start Master Mix (NEB), and nuclease-free water to a final volume of 25 μL. Cycling conditions were as follows: 30 s at 98 °C for enzyme activation, followed by 8 cycles of 10 s at 98 °C for denaturation, 15 s at 63 °C for annealing, 15 s at 72 °C for extension, and a final extension for 2 min at 72 °C.

PCR products were purified using Qiagen Nucleotide Removal kit (Qiagen) and digested with BbsI (NEB) overnight. Digested fragments were separated on a 20% TBE PAGE gel (Invitrogen) at 200 V for 1.5 h and the guide-containing 24 bp fragment excised and purified using the crush-and-soak method in 0.3 M NaCl overnight, followed by ethanol precipitation and resuspension in TE (Sambrook and Russell, 2006). DNA bands in polyacrylamide gels were visualised by incubating the gel in 0.5 μg/mL ethidium bromide for 10 min followed by ultraviolet light exposure on a transilluminator.

Ligation of the membrane protein gRNA library into the pKLV2-U6gRNA_SAM(BbsI)-PGKpuroBFP-W expression vector was performed at a 1:5 insert to vector ratio with T4 DNA Ligase for 2 h at 25 °C and transformed into One Shot TOP10 Chemically Competent *E. coli* (Invitrogen) by heat shock transformation at 42 °C. The total number of colony-forming

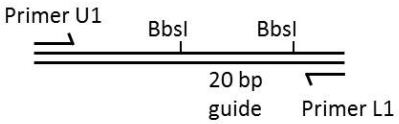

Figure 2.2 The CRISPRa gRNA library was synthesised as a complex pool of 77 nt oligonucleotides with asymmetric sequences flanking two BbsI restriction sites. This generates three fragments of unequal lengths after digestion for size separation of guide sequences for cloning into the final lentiviral expression vector. Primers U1 and L1 were used to amplify the oligonucleotide pool before BbsI digestion for cloning.

units was estimated by plating dilutions of the transformed cells, to be 11x the complexity of the library. Transformants were cultured in a liquid culture and DNA preparation performed using a PureLink HiPure Plasmid Filter Maxiprep Kit (Invitrogen), according to manufacturer's instructions.

To determine the distribution of gRNA in the plasmid library, 10 ng of plasmid (approximately ~1x10^9 copies) was used for Illumina sequencing as described in Section 2.5.5 and reads were mapped to the original gRNA sequences using MAGeCK (Li et al., 2014).

2.3 Cell lines and culture

2.3.1 Generation of dCas9-V2M line

HEK293-6E cells which are adapted to suspension growth and serum-free conditions were initially cultured in Freestyle media (Invitrogen) supplemented with 25 μg/mL G418 (Invitrogen) and 0.1% Kolliphor, but after single-cell cloning were maintained in Freestyle media supplemented with 50 μg/mL G418 and 1% FBS (Invitrogen). Cells cultured without FBS were refractory to colony formation after single-cell sorting. Cells were maintained in suspension in shaking incubators at 125 rpm and passaged every two to three days.

To generate constitutively dCas9-expressing cell lines for screening, HEK293-6E cells were transfected with pPB-R1R2_EF1adCas9VP64_T2A_MS2p65HSF1-IRESbsdpA or pPB-R1R2_EF1aVP64dCas9VP64_T2A_MS2p65HSF1-IRESbsdpA (encoding dCas9-

VP64 with or without an N-terminal VP64 fusion, along with a hyperactive *piggyBac* transposase (hyPBase) in a 1:5 ratio of transposase to transposon vector (Yusa et al., 2011). Selection with Blasticidin S (TOKU-E) at 5 μg/mL was initiated 48 h post transfection. Only cells transduced with pPB-R1R2_EF1aVP64dCas9VP64_T2A_MS2p65HSF1-IRESbsdpA and hyPBase gave rise to stable cell lines and were single-cell sorted into 96-well plates with a BD Influx cell sorter (BD Biosciences). This cell line is henceforth referred to as HEK-6E-V2M, where V2M stands for dCas9 with 2 x VP64 and MS2p65HSF1. Clonally derived lines were expanded and the clone with the highest CRISPRa activity as evaluated with a CRISPRa GFP reporter assay was expanded. All cell lines used in this project were tested and found negative for mycoplasma contamination (Surrey Diagnostics).

2.3.2 Lentiviral production and titering

HEK293-FT cells were used as a packaging cell line for lentivirus production. HEK293-FT cells were maintained in DMEM with GlutaMAX supplemented with 10% FBS (Invitrogen) and 1% penicillin-streptomycin, and passaged every two to three days. For virus production, 5 x 10^6 cells were seeded in per 10 cm plate at day 0 and transfected with 3 μg of transfer plasmid, 9 μg ViraPower lentivirus packaging vectors (Invitrogen) using 36 μL Lipofectamine LTX and 12 μL PLUS reagent diluted in Opti-MEM I (Invitrogen) transfection media. Cells were incubated for 4 h at 37 °C in transfection media before changing to DMEM with 10% FBS. Viral supernatant was harvested two days later, filtered, aliquoted and stored at -80 °C. Transduction of other cell lines was performed by incubating with a defined volume of virus overnight at 37 °C. Viral titers were determined by transducing HEK293-6E cells with a serial dilution of viral supernatant and quantifying the percentage of BFP+ cells on Day 2 post-transduction by flow cytometry.

Before performing pooled screens, viruses were titered to achieve a multiplicity of infection (MOI) of 0.3. Transduction at 0.3 MOI was to ensure that the majority of infected cells receive one virus per cell (Ellis and Delbrück, 1939). However, it was found that performing small-scale infections in 96-well plates did not scale up linearly, resulting in a higher level of infection than calculated. Instead, 1 x 10^7 HEK293-6E-V2M cells were transduced with 3 different volumes of library virus by overnight incubation at 37 °C. Cells were analysed two days post-transduction by flow cytometry, with BFP as a marker for successful transduction, and the volume of virus which resulted in 25-30% BFP+ cells was chosen. This process was repeated with each batch of virus produced.

To determine the effect of gene activation on cell growth, 6 x 10^7 cells (1000x library coverage) were sampled seven and 12 days post-transduction. To compare the distribution of gRNAs in the transduced library with that the original plasmid library, as well as between different virus preparations, 6 x 10^7 cells were sampled on Day 7 post-transduction with either virus preparation. Extraction of gDNA and sequencing were performed as described in Section 2.5.5 and mapping of guides was performed as described in Section 2.5.6.

2.3.3 CRISPRa GFP reporter assay

Reporter constructs pKLV2-U6gRNASAMg(TetO)-TREGFP-PGKpuroBFP-W and pKLV2-U6gRNASAMg(Empty)-TREGFP-PGKpuroBFP-W were constructed by Kosuke Yusa and obtained from the Yusa Lab. HEK293-6E cells expressing various dCas9 and MS2p65HSF1 fusion proteins were transduced with lentiviruses carrying either reporter, and GFP/BFP expression was analysed 72 h post transduction by flow cytometry on a BD LSRFortessa flow cytometer (BD Biosciences) as a measure of activation efficiency.

2.3.4 Cell binding assay and flow cytometry

Hybridoma supernatants were obtained from either the International Blood Group Reference Laboratory (National Health Service, UK) or the Developmental Studies Hybridoma Bank (University of Iowa, USA). Purified antibodies were purchased from either Abcam, Merck Millipore, or Biolegend. All antibodies used for flow cytometric analysis, along with their provenance, are listed in Appendix A.

For immunofluorescent staining, 100 μL of 1 μg/mL primary antibody was incubated with 5 x 10^5 cells for 1 h at 4 °C. Cells were then washed 1x in PBS-1%BSA before incubation with 100 μL of 0.1 μg/mL Phycoerythrin (PE)-conjugated secondary for 1 h at 4 °C. Finally, cells were washed 1x with PBS-1%BSA before resuspension in PBS without carrier protein and analysis by flow cytometry. Resuspension in PBS-1%BSA increased the occurance of instrument blockage, causing fluctuations in fluorescence intensity during aquisition. Samples were analysed on a LSRFortessa flow cytometer (BD Biosciences) and the resulting data were analysed using FlowJo (BD Biosciences).

2.3.5 cDNA transfections

Plasmids expressing full-length RTN4R, RTN4RL1 and RTN4RL2 were obtained from Origene. Full-length cDNA constructs were transfected into HEK293-6E cells with Linear polyethylenimine (PEI) at either a 2:1 ratio of PEI to DNA.

2.3.6 Annexin V staining

1x10^5 cells were washed 1x in PBS and 1x in binding buffer for Annexin V staining (Invitrogen), before being resuspended in 100 μL binding buffer. 5 μL Annexin V-FITC (eBioscience) was added to 100 μL cell suspension and incubated at room temperature for 10 min. Cells were then washed 1x with 2 mL of binding buffer and resuspended in 200 μL binding buffer for analysis. 5 μL of propidium iodide was added just before analysis by flow cytometry.

2.4 Recombinant protein production

2.4.1 Ectodomain construct design

Members of the adhesion GPCR (aGPCR) family that were selected for expression possessed clear signal peptide sequences as predicted by SignalP 4.0 (Petersen et al., 2011), lacked known extracellular cleavage sites other than the GPCR proteolysis site (GPS), and had extracellular domains (ECDs) of less than 2,000 amino acids. The entire extracellular region (with exception of the signal peptide) up to the beginning of the first transmembrane domain was produced. Where the HLT/S cleavage sequence was conserved, a T/S→G mutation was introduced to prevent self-cleavage.

ECDs were synthesised (GeneArt Gene Synthesis, Invitrogen) and cloned into bait and prey expression vectors pMero-Cd4d3+4-BioLHis and pMero-Cd4d3+4-COMP-blac-FLAGHis. Both vectors contained an exogenous signal peptide that facilitates protein secretion, domains 3 and 4 of rat Cd4 to boost protein expression, and a polyhistidine-tag for purification (Brown and Barclay, 1994; Crosnier et al., 2011; Hochuli et al., 1988). In addition, the bait expression vector contained a biotinylation sequence that can be enzymatically biotinylated by *E. coli* biotin ligase (BirA), whilst the prey expression vector contained a

pentamer-forming domain of rat Cartilege Oligomeric Matrix Protein (COMP), a β-lactamase catalytic domain, and a FLAG tag (Bushell et al., 2008; Einhauer and Jungbauer, 2001). These tags were used for relative quantification and normalisation of proteins, as well as forming oligomers for increased avidity.

ECDs of RTN4R, RTN4RL1 and RTN4RL2 were amplified from full length cDNA constructs. The site of proteolytic cleavage and GPI-anchor attachment was predicted with PredGPI (Pierleoni et al., 2008) and the entire extracellular domain, including the endogenous signal peptide, up to the predicted cleavage site was amplified. The amplified fragment was cloned into bait and prey expression vectors pTT3-Cd4d3+4-BLH and pTT3-Cd4d3+4-COMP-blac-FLAGHis, which lack an exogenous signal peptide.

2.4.2 Expression and His-tag purification

All constructs were sequence verified and produced using a mammalian expression system by transfecting HEK293-6E cells with an expression construct (Loignon et al., 2008). For bait proteins, expression constructs were co-transfected with a plasmid encoding BirA in a 9:1 ratio. Linear polyethylenimine (PEI) was used for all transfections at either a 2:1 or 3:1 ratio of PEI to DNA. HEK-6E cells were maintained in Freestyle medium (Invitrogen) supplemented with 25 μg/mL G418 (Invitrogen) and 0.1% Kolliphor. Transfections were left for 5 days and supernatants were harvested and filtered through a 0.2 μm filter. Supernatants containing prey proteins were used neat or diluted without purification whilst those containing bait proteins were subjected to His-tag affinity purification.

Supernatants containing biotinylated bait proteins were incubated with Ni-NTA agarose beads (Jena Bioscience) overnight at 4°C with constant rotation. 100 μL of beads with a binding capacity of 0.5 mg was used for every 50 mL of supernatant. Polypropylene columns (Qiagen) were equilibrated with 2 mL binding buffer (20 mM sodium phosphate buffer, 0.5 mM NaCl, 40 mM imidazole) before addition of the bead-supernatant mixture. Beads were washed with 5 mL binding buffer and proteins eluted in 500 μL of elution buffer (20 mM sodium phosphate buffer, 0.5 mM NaCl, 400 mM imidazole) by incubating for 30 min at room temperature.

2.4.3 SDS-PAGE and Western blot

To determine the purity and size of bait proteins, SDS-PAGE and Coomassie staining were performed with 10 μL of purified protein and SafeBLUE Protein Stain (NBS Biologicals). Proteins were first denatured by boiling for 10 min at 70 °C before gel electorphoresis using NuPAGE 4-12% Bis-Tris Gels (Invitrogen) and MOPS buffer. For detection by Western blot, 10 μL of undiluted supernatant was separated by SDS-PAGE and transferred to a nitrocellulose membrane. Blots were then blocked with PBS-2% BSA, probed with streptavidin conjugated to HRP (Sigma, 1:10000) for 1 h at room temperature. Bands visualised using Pierce ECL Western blotting substrate (Thermo Scientific) according to manufacturer's protocol.

2.4.4 Prey normalisation with nitrocefin hydrolysis assay

Prey proteins were normalised using a nitrocefin turnover assay to measure β-lactamase enzymatic activity. Serial dilutions of prey supernatants were made in PBS-1% BSA and 20 μL of each dilution incubated with 60 μL of 125 μg/mL nitrocefin (Calbiochem) at room temperature. Absorbance readings at 485 nm were taken once every minute for 20 minutes. Absorbance at 485 nm was plotted against time and the dilution which caused complete nitrocefin turnover at 10 min was selected.

2.4.5 Avidity based extracellular interaction screen

AVEXIS was performed essentially as described in Bushell et al. (2008). Different dilutions of bait proteins were captured on streptavidin-coated plates for 45 min at room temperature. Plates were washed in PBS-1% Tween 20 and normalised prey proteins were added for 1 h at room temperature. Excess prey protein was removed by washing gently with PBS-1% Tween 20 twice and 60 μL of 125 μg/mL nitrocefin was added to detect bait-prey interactions. Absorbance readings at 485 nm were taken 1 and 2 h after nitrocefin addition.

2.5 CRISPRa extracellular interaction screen

2.5.1 Lentiviral transduction and cell library sort

4 x 10^7 HEK293-6E-V2M cells were transduced to achieve between 25-30% BFP+ cells (approximately 0.3 MOI, 200x library coverage). To remove transduced cells, HEK293-6E-V2M cells were sorted for BFP+ expression two days post transduction. A minimum of 1.5 x 10^7 BFP+ cells were collected (250x library complexity) to ensure sufficient coverage of the gRNA library. Sorted cells were expanded in media supplemented with 2 μg/mL puromycin (Gibco) to maintain lentiviral construct expression.

2.5.2 Tetramerisation of biotinylated proteins

Bait protein concentrations were normalised to the amount of protein needed to saturate 2 μg of streptavidin conjugated to PE (BioLegend). Streptavidin contains four biotin-binding sites, allowing multiple biotinylated bait proteins to be clustered around a single molecule of streptavidin, thereby increasing the avidity of the oligomerised probe for potential binding partners (Altman et al., 1996). The concentration of biotinylated bait needed to saturate a fixed amount of streptavidin-PE was determined by enzyme-linked immunosorbent assay (ELISA).

2.5.3 Probe normalisation using enzyme-linked immunosorbent assay

Serial dilutions of each bait protein were incubated with or without 10 ng of streptavidin-PE overnight at 4°C. The remaining molecules of free biotinylated bait were captured on streptavidin-coated, flat-bottomed 96-well plates (Nunc) for 45 min at room temperature. Immobilised baits were detected by a primary incubation with monoclonal anti-rCd4 mouse IgG (OX68), which recognises a conformation-specific epitope on domains 3 and 4 of rat Cd4 present in the bait, followed by a secondary incubation with an alkaline phosphatase-conjugated anti-mouse IgG (Bethyl Laboratories). All incubations were performed for 1 h at room temperature and plates were washed 3x in PBS-0.1% Tween 20 and 1x in PBS between additions. 100 μL of 1 μg/mL alkaline phosphatase substrate (Sigma) dissolved in diethanolamine buffer (0.5 mM $MgCl_2$, 10% diethanolamine, pH 9.2) was added to wells, and substrate turnover after 15 min was quantified by measuring absorbance at

405 nm with a FLUOstar Optima plate reader (BMG Biotech). Absorbance at 405 nm was plotted against dilution factor for each bait protein and the highest concentration at which no free biotinylated bait remained after conjugation with streptavidin-PE was selected.

2.5.4 Fluorescence activated cell sorting

For GPI-linked protein screening, HEK293-6E-V2M cells transduced and sorted for BFP+ expression were assayed for overexpression of GPI-anchored proteins by incubation of 1 x 10^8 cells in 5 mL of 25 ng/mL Alexa Fluor 488 conjugated proaerolysin (Cedarlane) for 20 min at room temperature. Labelled cells were sorted using a SH800 cell sorter (Sony Biotechnology) and double positive BFP+AF488+ cells were collected. Two thresholds for defining AF488+ cells were compared and a final threshold of between 3-5% of cells was chosen for further experiments.

To detect gain-of-function binding to recombinant protein probes or antibodies, 1 x 10^8 HEK293-6E-V2M cells were assayed between seven to ten days post-transduction. Cells were washed once in PBS-1% BSA, then incubated with 5 mL normalised prey proteins or 1 μg/mL primary antibodies for 2 h on ice. Cells were washed again with PBS-1% BSA and then incubated in secondary, PE-conjugated antibodies for 1 h on ice. Cells were washed a final time in PBS-1%BSA before cell sorting. All primary and secondary antibodies used are listed in Appendix A. For pre-conjugated bait proteins which had been oligomerised around streptavidin-PE, only a single incubation was performed for 2 h on ice. Labelled cells were resuspended in PBS and sorted using a SH800 cell sorter (Sony Biotechnology). Double positive BFP+PE+ cells were collected and stored at -20 °C before gDNA extraction and sequencing.

2.5.5 Genomic DNA extraction and sequencing

For samples with fewer than 1 x 10^6 cells, cells were resuspended in nuclease-free water at 8 x 10^5 cells/mL and lysed for 10 min at 95 °C. Lysates were treated with 2 μg/mL Proteinase K for 50 min at 55 °C followed by 10 min at 95 °C for inactivation. 10 μL of treated lysate was used as template for each 50 μL PCR reaction.

For samples with 1 - 2 x 10^6 cells and 5 - 6 x 10^7 cells, column-based purification of genomic DNA (gDNA) was performed with DNeasy Blood and Tissue kit (Qiagen) and

Blood and Cell culture DNA maxi kit (Qiagen), respectively. DNA concentration in eluate was quantified with a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific) and 1-2 μg gDNA was used as template for each 50 μL PCR reaction. Multiple reactions (8-36) were performed to achieve sufficient coverage of the library.

A 298 bp fragment containing the guide RNA sequence was amplified from gDNA. Illumina adapters and barcodes were added in two successive PCR reactions. Cycling conditions for both reactions were as follows: 30 s at 98 °C for enzyme activation, followed by a number of cycles of 10 s at 98 °C for denaturation, 15 s at 61 °C or 66 °C for primer annealing (first and second reactions respectively), 15 s at 72 °C for extension, and a final extension for 2 min at 72 °C. Depending on the type of input (column-purified gDNA or cell lysate), either 25 cycles or 30 cycles were run for the first PCR reaction. PCR products from the first reaction were purified using Qiagen PCR purification kit and 1 ng of purified product used as template in the second reaction. The second PCR reaction involved 15 cycles of amplification, after which PCR products were size-selected using solid phase reversible immobilization with Agencourt AmPure XP beads (Beckman Coulter) in a 0.7 v/v ratio of beads to sample. 5 μL of PCR product was analysed with gel electrophoresis on a 2% agarose gel to confirm for quantity and size after each reaction. No template controls were performed to monitor possible contamination from other sources.

Primers containing Illumina adaptors along with 11 bp barcodes were used to allow for multiplexing of up to 10 samples in a single run (Quail et al., 2011). 19 bp sequencing was performed with a custom sequencing primer on a HiSeq 2500 in rapid run mode. Library multiplexing and sequencing were performed by the Illumina Bespoke Sequencing team from the Wellcome Sanger Institute. All primers used for Illumina library preparation and sequencing are listed in Appendix A.

2.5.6 CRISPRa screen analysis

Raw sequencing reads were converted from CRAM to FASTQ format using the `fasta` function in SAMTools 1.3 (https://sourceforge.net/projects/samtools/files/samtools/1.3/). The 19 bp reads were then aligned to gRNA sequences using the `count` function in MAGeCK. MAGeCK is a statistical package built for model-based analysis of CRISPR screens and uses a mean-variance function to estimate a null negative binomial distribution for individual gRNA counts. For testing of gene level enrichment, MAGeCK employs a modified Robust-

Rank Aggregation approach to evaluate the likelihood that perturbing a particular gene is having an effect in a pooled CRISPR screen (Li et al., 2014).

Counts were normalised by total number of reads to account for differences in sequencing depth. Enrichment testing was performed using the `test` function in MAGeCK without further normalisation and with gRNAs grouped by gene rather than TSS. The sequenced plasmid library (as described in Section 2.2.2) was used as the control sample for all tests. Using sequences from unsorted libraries at Day 7 or Day 12 as the control sample gave similar results. All genes with a False Discovery Rate (FDR) below 0.1 were considered candidate receptors and secondary validation performed with individually cloned gRNA or overexpression with full length cDNA endcoding the targeted receptor.