

Extracellular receptor-ligand interaction screening using CRISPR activation



Chong Zheng Shan

Wellcome Sanger Institute
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Homerton College

September 2018

To my family for their infinite patience and support.

DECLARATION

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 60,000 words including appendices, bibliography, footnotes, and tables.

Chong Zheng Shan
September 2018

ACKNOWLEDGEMENTS

First and foremost I would like to acknowledge the excellent guidance provided by my supervisor, Gavin Wright, who was always available to discuss ideas and provide valuable feedback both conceptual and technical. I thank Gavin also for the resources to undertake this project and for taking a chance on a late applicant.

My training has also benefitted greatly from sound and timely advice from my second supervisor, Kosuke Yusa, particularly with respect to plasmid construction and pooled screening.

Many past and present members of the Cell Surface Signalling Laboratory have generously shared their advice, reagents, and techniques with me, which I am immensely grateful for. In particular I should acknowledge Sumana Sharma for unselfishly providing troubleshooting tips and resources which really expedited setting up the CRISPRa screen. I am grateful also to Laura Wood who let me pick many cDNA clones from her impressive cDNA library. Thanks also to many other members of the lab who made it a fun and comfortable environment to work in.

I would also like to thank Ulrika Frising for many stimulating discussions about science and working in science, which I think has made me a more resourceful, meticulous and thoughtful scientist.

Finally, I am grateful to Valentine Svensson for all the interesting discussions about both work and life, for his patience in dealing with my questions about his field, as well as all the fun times spent together. Without Valentine, my PhD would not have been as enriching or enjoyable.

ABSTRACT

Interactions between cell surface proteins mitigate multiple signalling and cell recognition events during development and disease. Previous methods to screen for extracellular protein interactions have relied on recapitulating these interactions using recombinant ectodomains of membrane-associated proteins. This approach is well suited for studying single pass transmembrane and glycosylphosphatidylinositol (GPI)-linked proteins, which tend to have single chain ectodomains. However, it has limited capacity for identifying interactions involving proteins with multiple transmembrane domains, which tend to have complex extracellular regions composed of more than one extracellular loops. Cell-based methods can help overcome this limitation by providing a native environment for the presentation of multipass membrane ectodomains. The success of cell microarrays transfected with cDNA libraries encoding plasma membrane proteins in identifying surface receptors demonstrates the feasibility of upregulating cell surface proteins for studying extracellular interactions, but requires significant cost to perform at scale.

Recently, advances in transcriptome editing using CRISPR/Cas9 have enabled highly efficient and specific gene activation on a genome-wide scale. By systematically upregulating plasma membrane proteins in human cell lines using CRISPR activation (CRISPRa), I developed a screening approach to identify novel receptor-ligand interactions in a convenient, single tube format. I show that this approach detects known interactions with a low false positive rate and apply it to identify ligands for the adhesion G-protein coupled receptors. I found that myelin-associated inhibitory proteins, the Nogo receptors, interact with Brain angiogenesis inhibitor 1 (ADGRB1) and show that the interaction is mediated by the first three thrombospondin repeat domains of ADGRB1. Together, this suggests that pooled CRISPRa screening presents a sensitive and convenient approach for genome-scale extracellular receptor-ligand identification, avoiding costly and technically challenging preparation of cDNA or recombinant protein libraries.

CONTENTS

List of Figures	xvii
List of Tables	xxxi
1 Introduction	1
1.1 Extracellular receptor-ligand interactions	1
1.1.1 Challenges of studying extracellular protein-protein interactions	2
1.1.2 Interaction screening using soluble recombinant ectodomains	6
1.1.3 Interaction screening using cDNA overexpression	9
1.2 CRISPR/Cas9-based transcriptional activation	10
1.2.1 Origins as a bacterial adaptive immune system	11
1.2.2 Transcriptional activation with CRISPR/Cas9	12
1.2.3 Genome-scale gain-of-function screening using CRISPR activation	16
1.3 Aims and objectives	20
2 Materials and Methods	23
2.1 Plasmids and cloning	23

2.1.1	Generation of different dCas9 fusion constructs	23
2.1.2	Individual gRNA cloning	25
2.1.3	RNA isolation and q-RT-PCR	26
2.2	CRISPRa gRNA library construction	27
2.2.1	Computational selection of gRNAs	27
2.2.2	gRNA Library synthesis and cloning	28
2.3	Cell lines and culture	29
2.3.1	Generation of dCas9-V2M line	29
2.3.2	Lentiviral production and titering	30
2.3.3	CRISPRa GFP reporter assay	31
2.3.4	Cell binding assay and flow cytometry	31
2.3.5	cDNA transfections	32
2.3.6	Annexin V staining	32
2.4	Recombinant protein production	32
2.4.1	Ectodomain construct design	32
2.4.2	Expression and His-tag purification	33
2.4.3	SDS-PAGE and Western blot	34
2.4.4	Prey normalisation with nitrocefin hydrolysis assay	34
2.4.5	Avidity based extracellular interaction screen	34
2.5	CRISPRa extracellular interaction screen	35
2.5.1	Lentiviral transduction and cell library sort	35
2.5.2	Tetramerisation of biotinylated proteins	35

2.5.3	Probe normalisation using enzyme-linked immunosorbent assay . . .	35
2.5.4	Fluorescence activated cell sorting	36
2.5.5	Genomic DNA extraction and sequencing	36
2.5.6	CRISPRa screen analysis	37
3	CRISPR activation enables rapid and stable overexpression of cell surface proteins	39
3.1	Introduction	39
3.2	Results	40
3.2.1	Rapid upregulation of cell surface proteins using CRISPRa	40
3.2.2	An additional histone deacetylase domain reduces CRISPRa activity	41
3.2.3	Increased mRNA abundance does not necessarily lead to increased surface expression	44
3.2.4	Constitutive expression of dCas-activators enables stable overexpression of cell surface proteins	44
3.2.5	A CRISPRa gRNA library targeting all human membrane proteins .	48
3.3	Discussion	52
4	CRISPR activation screening detects known extracellular interactions	55
4.1	Introduction	55
4.2	Results	57
4.2.1	Establishing a pooled CRISPRa extracellular interaction screening approach	57
4.2.2	Less stringent sort threshold facilitates identification of GPI-anchored proteins using CRISPRa screening	59

4.2.3	Multiplexing selection probes enables detection of multiple interactions using a single screen	63
4.2.4	CRISPRa screening detects low-affinity endogenous interactions . .	66
4.3	Discussion	66
5	CRISPRa screening of adhesion GPCRs	73
5.1	Introduction	73
5.2	Results	77
5.2.1	A T/S→G mutation at the GPS site enables high level of recombinant ectodomain production	77
5.2.2	CRISPRa screening of adhesion GPCR ectodomains identifies known interactions	79
5.2.3	CRISPRa screening with ADGRB1 uncovers novel interactions with Nogo receptors	81
5.2.4	The first three thrombospondin repeats on ADGRB1 is sufficient for RTN4R binding	84
5.3	Discussion	84
6	General Discussion	89
6.1	Summary of results	89
6.2	Evaluation of CRISPR activation screening	90
6.2.1	CRISPR activation uses gRNA libraries that are cost-effective to produce and maintain	90
6.2.2	CRISPR activation allows genome-scale interaction screening . . .	91
6.2.3	CRISPR activation screening is not restricted to receptors that are already expressed by a cell line	91

6.2.4	CRISPRa screening is able to detect low affinity interactions	92
6.2.5	CRISPRa screening is unable to detect certain types of extracellular interactions	92
6.2.6	Potential explanations for false negatives arising from CRISPRa screening and suggestions for improvement	92
6.3	Implications of ADGRB1-RTN4R interactions	93
6.3.1	Knockout phenotypes complicate the assessment of ADGRB1-RTN4R function	95
6.3.2	Levels of downstream effectors can help determine if RTN4Rs are activating ligands of ADGRB1	95
6.4	Other possible applications of CRISPRa extracellular interaction screening	96
6.4.1	Uncovering novel viral receptors	96
6.4.2	Identifying extracellular interactions underlying cancer metastasis .	97
6.5	Conclusion	98
Appendix A		111
A.1	Table 1: gRNA sequences targeting 12 cell surface proteins	111
A.2	Table 2: PCR and sequencing primers, gblock sequences	115
A.3	Table 3: Primary and secondary antibodies	120

LIST OF FIGURES

- 1.1 **Various strategies for plate-based interaction screening.** The ectodomains of cell surface receptors are produced as soluble recombinant form fused to biochemical tags for immobilisation, oligomerisation and detection by enzymatic activity. Immunoglobulin Fc-fusions (Fc) and the pentamerising domain of rat cartilage oligomeric matrix protein (COMP) mediate oligomerisation for increased avidity. Alkaline phosphatase (AP), horseradish peroxidase (HRP) and β -lactamase provide enzymatic activity for detection with the relevant substrates. AP and Fc tags also facilitate immobilisation onto α -AP antibody or Protein A-coated plates. Biotinylation enables capture and clustering on streptavidin-coated plates. 7
- 1.2 **Original and modified CRISPR/Cas9 systems.** A) Nuclease-active Cas9 complexes with gRNA and is targeted to genomic loci complementary to a 20 nt guide sequence at the 3' end of the gRNA and upstream of a PAM sequence (NGG). Upon binding, Cas9 catalyses a double strand break upstream of the PAM sequence. B) Nuclease-inactive Cas9 (dCas9) is fused to an activator domain like VP16 and retains gRNA-mediated targeting to genomic loci upstream of a coding gene. Binding of dCas9-activator results in recruitment of transcription initiation complex and subsequently transcriptional activation of downstream gene. C) Second generation dCas9-activator system SAM utilises synergistic activation domains from VP64, p65 and HSF1 to achieve high levels of transactivation from a single gRNA. MS2-fused p65 and HSF1 transactivator domains are recruited to MS2-binding hairpin loops engineered into the gRNA scaffold. 14

-
- 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 synergistic activators MS2-p65-HSF1 by gibbon 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 gibbon assembly before transfer into a PiggyBac expression vector by Gateway cloning. 24
- 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. 29

3.1 CRISPR activation of CD200 and CD2 is specific and allows ligand binding.

A) Schematic of expression vectors used. gRNAs were expressed under a U6 promoter along with BFP as a fluorescent marker. dCas9-VP64 and MS2-p65-HSF1 fusion proteins were expressed as a single transcript separated by a T2A self-cleaving peptide. A non-activating vector expressing dCas9 without any transactivators was used as a negative control. PB - piggy-Bac inverted terminal repeats, attB1/B2 - λ recombination attachment sites, Ef1 α - Human elongation factor-1 α promoter, NLS - Nuclear Localisation Signal, IRES - Internal Ribosomal Entry Site, bpA - Bovine growth hormone polyadenylation site. Psi - Viral packaging signal sequence, RRE - Rev response element, cPPT - Central Polypurine tract, PGK - Phosphoglycerate kinase promoter, WPRE - Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element. B) Flow plots showing clear upregulation of CD200 and CD2 in cells co-transfected with gRNA and dCas9-activators but not in cells co-transfected with gRNA and a non-activating control. Surface expression was quantified by antibody staining of CD200 and CD2 respectively. gRNA expression indicated by expression of BFP C) Quantification of CD200 and CD2 expression shows no cross-reactivity with either antibody and is dependent on transactivators expressed by the activating construct. mAb binding was calculated as a percentage of total BFP+ cells D) CRISPRa of CD200 and CD2 is sufficient to induce gain-of-binding of known ligands rCd200r and CD58. Soluble recombinant ectodomains of rCd200r and CD58 were produced as highly avid FLAG-tagged pentameric proteins and detected by fluorescently-labelled anti-FLAG antibody. Data points in C) and D) represent mean \pm s.e.m; $n=3$

- 3.2 Reduced CRISPRa efficiency using dCas9-activators with an additional p300 HAT domain.** A) Quantification of surface expression of 12 cell surface proteins after CRISPRa. Cells were transduced with lentiviruses carrying a pools of 8 gRNAs targeting each gene, before being transfected with dCas9-VP64 activator construct. Surface expression was assessed by antibody binding 48 h post transfection B) Schematic of dCas9-transactivator variants generated, with either an additional p300 HAT domain (p300core) or VP64 domain. These constructs were transferred into the PiggyBac expression vector in Figure 3.1A for co-expression with MS2-p65-HSF activators. C) Comparison of all five dCas9-activator variants indicate that constructs with a p300core domain achieved lower levels of surface receptor overexpression than constructs with one or two VP64 domains. Data points in A) and C) represent mean \pm s.e.m; $n=3$ 45
- 3.3 Upregulation of mRNA transcripts does not necessarily lead to with an increase in surface protein levels.** A) Quantification of mRNA abundance by qRT-PCR of indicated target genes in cells 48 h post co-transfection with dCas9-VP64 and either targeting gRNA (+) or no gRNA control. Transcript abundance was normalised to *CYP4A* expression; bars represent mean \pm s.e.m; $n=6$. P-values calculated using a Student's t-test, ns $P > 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$ B) Antibodies against SLC4A1, RHD and KEL bind to erythrocytes. mAb binding histograms show that incubation of erythrocytes with only fluorescently conjugated secondary (red trace) results in similar profile to unstained erythrocytes (black trace), whilst staining with mAbs against SLC4A1, RHD and KEL result in a rightward shift in fluorescence intensity. C) Anti-SELE antibody specifically recognises recombinant E-selectin but not a control protein (Cd200r) in an enzyme-linked immunosorbent assay. The ectodomains of SELE and Cd200r are produced as soluble biotinylated proteins fused to the 3rd and 4th Ig domains of rat Cd4. Recombinant SELE and Cd200r are captured on streptavidin coated plates and detected by incubation with anti-SELE or anti-rCd4, followed by a secondary anti-mouse antibody conjugated to alkaline phosphatase. Binding is quantified by absorption at 405 nm of a hydrolysis product of a phosphatase substrate. Binding of anti-rCd4 reflects relative amounts of SELE or Cd200r captured on the plate. 46

- 3.4 **A cell line constitutively expressing dCas9-activators allows sustained CRISPRa overexpression of cell surface proteins** A) Schematic of GFP-inducible CRISPRa activity reporter system. In both vectors, GFP expression is controlled by a TetO responsive element (TRE) consisting of several TetO repeats and a minimal CMV promoter. BFP is constitutively expressed under a PGK promoter. The reporter construct expresses a gRNA targeting the TetO sequence and when expressed in a cell with dCas9-activators results in increased GFP expression. A control construct expresses an 'empty' gRNA which does not target TetO and acts as a control to measure baseline levels of GFP expression. B) Quantification of GFP expression in the parental (Par.) HEK293 cell line before transfection with dCas9-activator, the polyclonal (Poly.) line generated after a week of blasticidin selection, as well as 20 single cell clonal lines. Percentage transduced cells was determined by BFP+ cells and show that cell lines were transduced with both reporter and control constructs to a comparable level. Data points are from a single experiment. C) Flow plots showing baseline GFP expression in parental HEK293 line as well as in cells transduced with control construct. The polyclonal line exhibits some level of CRISPRa activity (64.6% GFP+/BFP+) and single-cell cloning results in increased CRISPRa activity (83.3% GFP+/BFP+) D) Percentage of cells expressing the indicated cell surface receptors as determined by mAb staining after transduction of the cloned activator cell line, HEK293-V2M, with appropriate pooled gRNAs. Data points represent mean \pm s.e.m; $n=3$. 47
- 3.5 **Schematic showing workflow of membrane protein gRNA library design** A list of putative membrane proteins were compiled from public databases, mass spectrometry and bioinformatic studies. Unique TSSs were extracted based on a combination of Gencode gene models and CAGE-seq peaks (GencCAGE peaks); a maximum of two unique TSSs were chosen per gene. In the absence of predicted GencCAGE peaks, the APPRIS annotation pipeline was used instead. All 19 nt sequences adjacent to a 5'-NGG-3' PAM upstream of the peaks were found and filtered based on GC content and distance from the peak. Guides with off-target sites or BbsI sites were discarded. Most TSSs were targeted by 7 guides. The final library contained 58,570 guides targetign 6,213 genes. 49

- 3.6 **Quality controls indicate complete and fairly even representation of guide sequences in the gRNA library** A) Ranked gRNA abundance in the plasmid library (brown) and cells transduced with the CRISPRa lentiviral library and cultured for seven days (green) as determined by raw read counts from deep sequencing of PCR-amplified products. B) The gRNA library complexity is maintained in transduced cells. A comparison of the gRNA read count abundance from products amplified from the plasmid library and cells seven and twelve days post transduction. Pearson's coefficient of correlation (r) of libraries on day 7 and day 12 post transfection with the original plasmid library was calculated. 50
- 3.7 **Small scale validation of individual guides show that majority of guides are active** A) (Top) mAb binding histograms of HEK293-V2M cells transfected with the individual gRNAs targeting the promoter region of the named receptor genes (blue traces) compared to control non-targeting gRNAs (red traces) and stained with the respective mAbs. sgRNA number 1 is shown for each target gene. (Bottom) Each gRNA targeting the promoter region of the named receptor proteins were numbered and individually tested and their ability to upregulate cell surface protein expression quantified by FACS compared to a non-targeting (NT) control. B) Screenshot of UCSC Genome browser showing the *CD55* locus. Regions containing gRNAs targeting *CD55* TSS1 and TSS2 are indicated as black bars. Other tracks show FANTOM5 CAGE-seq peaks, predicted gene models, and H3K27ac data. The signal peptide sequence of *CD55* is encoded in exon 1 of the transcript produced from *CD55* TSS1, whilst *CD55* TSS2 starts at the third exon of the same transcript and thus does not contain a signal peptide for trafficking to the surface. 51

- 4.1 **Schematic of CRISPRa extracellular interaction screening.** A CRISPRa gRNA library targeting genes encoding membrane proteins was designed, cloned and packaged into lentiviruses for transduction. Transduction of a cell line constitutively expressing dCas9-activators at a low multiplicity of infection (MOI) ensures majority of cells receive one gRNA per cell, however this means that only around 30% of cells are transduced. Removal of untransduced cells is achieved by sorting for BFP+ cells. Transduced cells are expanded for 5 - 7 days to provide libraries for screening multiple ligands. For each screen, 1×10^8 cells are incubated with a fluorescently labelled ligand or antibody, and cells which gain an ability to bind to the ligand of interest are sorted by fluorescence-activated cell sorting (FACs). Sorted cells are lysed and gRNA sequences amplified for quantification by next-generation sequencing. Analysis of guide enrichment in the sorted population as compared to the plasmid library allows identification of receptor candidates. 58
- 4.2 **Sufficient library coverage is required for robust receptor identification with CRISPRa screening.** A) ITGB3-targeting guides (blue dots) are enriched specifically in cell populations sorted for binding to anti-integrin $\alpha v \beta 3$. Box plots of normalised gRNA abundances are shown for screens using anti-integrin $\alpha v \beta 3$, anti-CD200, and anti-Ms secondary. Three replicates (r1/2/3) were performed for each antibody. Unsorted controls include the plasmid library and cell libraries cultured for 7 or 12 days post transduction. Dotted lines indicate one order of magnitude around the median of unsorted samples. B) In contrast, CD200-targeting (red dots) are enriched in screens using anti-CD200 but also in several other screens. C) Replicate 2 of the anti-CD200 screens shows a high level of guide depletion, possibly due to insufficient coverage. Dotplots of gRNA abundance in cells sorted with anti-CD200 against that of the plasmid library for replicate 1 (left) and replicate 2 (right) show that in replicate 1, majority of gRNAs having a similar abundance to that in the plasmid library, and cluster around the dotted line where $x=y$. In replicate 2, some gRNAs appear highly abundant whilst others are depleted. As a result CD200-targeting guides (red, labelled by gRNA number) do not appear enriched in statistical enrichment tests. D) A CRISPRa screen using anti-CD200 with increased library coverage and optimised resuspension protocols result in better baseline correlation with the plasmid library (dotted line denoting $x=y$) and robust enrichment of all seven CD200-targeting guides (red, labelled by gRNA number). 60

- 4.3 **A 5% sort threshold during CRISPRa screening reduces false negatives whilst an FDR cut-off of 0.1 limits false positives** A) Sort gates using during screening at a 0.5% (left) and 5% threshold (right). The percentage of cells in the gate fluctuates during sorting hence numbers in the image are not exactly 0.5% or 5%. B) Sorting at a 5% threshold results in a more consistent baseline gRNA distribution as seen from increased Pearson's correlation (r) of 0.765 as compared to 0.314. Scatterplots of gRNA abundance between sorted and plasmid samples also show increased clustering around the line $x=y$ (dotted). Majority of gRNAs should not have an effect and therefore should be present in similar relative abundance in both sorted and plasmid libraries. C) 5% sort threshold (red squares) results in an increased number of 'hits' (top) as well as number of GPI-anchored proteins identified (bottom) at different FDR cut-offs as compared to a 0.5% threshold (blue diamonds). D) An FDR cut-off of 0.1 provides a balance between identifying additional GPI-anchored proteins (true positives, TP), and detecting false positives (FP) at both sort thresholds. Plotting the difference (number of TP-FP) shows an increase when going from a cut-off of 0.05 to 0.1, but not for higher FDR cut-offs. E) False negative rates do not decrease drastically at higher FDR cut-off rates at either sort threshold. False negative rate was calculated by taking the percentage of GPI-anchored proteins that were not identified at that FDR cut-off out of 139 (total number of GPI-anchored proteins). 64
- 4.4 **CRISPRa screening simultaneously identifies multiple targets to a pool of antibodies.** A) Transduced cell libraries were sorted to isolate cells binding to an equimolar pool of eight antibodies, and gRNA abundance quantified by next-generation sequencing. Enrichment analysis indicated that guides targeting six out of eight cell surface targets were enriched in sorted cells at an FDR of less than 0.1 (red dots, labelled with gene symbol). WNT3 was also identified under than FDR cut-off but at a lower significance. Guides targeting P2RX7 and PROM1 were not enriched in the screen (blue dots). B) Visualising enrichment at an individual gRNA level shows that WNT3-targeting guides (dark blue asterisks, FP) are not highly enriched, unlike guides targeting the six cell surface targets (red, various shapes, TP). Guides targeting P2RX7 and PROM1 are not enriched at all (green crosses/pluses, FN). 65

4.5 Highly avid tetramers are produced from recombinant biotinylated ectodomains and normalised for use in CRISPRa screening

A) Schematic showing the production of tetramers from purified biotinylated monomers containing the full length ectodomain of four cell surface ligands. A construct encoding the recombinant protein is transfected into HEK293 cells. After six days recombinant protein is harvested and purified using nickel affinity beads which bind a 6x histidine tag on the C-terminus of the protein. Tetramers are formed by incubating recombinant protein with fluorescently labelled streptavidin (streptavidin-PE). CMV - human cytomegalovirus immediate-early promoter; Sp - Signal peptide; rCD4 d3+d4 - 3rd and 4th Ig domains of rat CD4. B) The amount of recombinant protein used for screening is normalised using the amount needed to saturate 2 μ g of streptavidin-PE. Different dilutions of purified proteins are conjugated to 10 ng streptavidin-PE overnight and the remaining free biotinylated proteins are captured on a streptavidin-coated plate. Captured protein is detected with an antibody targeting rCD4 d3+d4 followed by an appropriate alkaline phosphatase-conjugated secondary. Absorbance at 405 nm indicates the amount of free protein remaining after conjugation and is shown for the four ligands CD55, CTLA4, EFNA1 and rCd200r. The highest concentration of biotinylated protein that resulted in no excess protein was determined (dotted lines) and scaled linearly to derive the amount needed to saturate 2 μ g.

67

4.6 CRISPRa screening unambiguously identifies low-affinity endogenous interactions

A) CRISPRa screening identifies six out of nine reported interactions involving CD55, CTLA4, rCd200 and EFNA1. Blue circles represent cell surface ligands used as tetramers for screening, pink circles represent binding partners identified by CRISPRa screening and white circles are binding partners that were not detected. B) Endogenous binding partners are identified with high confidence as seen in the gene level enrichment analysis of each screen. In all four screens, at least one binding partner is detected below an FDR of 0.1 (red dots) with no other genes showing significant enrichment at that cut-off.

68

- 5.1 **Structure of a typical adhesion GPCR** A) Adhesion GPCR structure can be compartmentalised with reference to topology or cleavage at the GPCR proteolytic site (GPS). All adhesion GPCRs consist of a tripartite structure consisting of an extracellular domain (ECD), a seven transmembrane domain (7TM), and an intracellular (ICD). Some adhesion GPCRs undergo autoproteolysis at the GPS to produce an N-terminal (NTF) and C-terminal fragment (CTF). B) The GAIN domain is a complex fold that mediates autoproteolysis and subsequent attachment of cleaved NTF and CTFs. It is divided into two subdomains, A and B. Subdomain B contains and is cleaved at a conserved sequence of residues (HL↓T/S) located within the GPS motif. C) Ligand binding to the NTF is thought to induce intracellular signalling by causing structural changes or complete dissociation of the NTF to reveal a cryptic tethered agonist which then binds to and activates the receptor. 75
- 5.2 **The GPS is not an autonomously folded domain but is part of a larger domain.** A) Diagram of ADGRL1 and ADGRB3 showing the domains suggested by the SMART protein domain prediction server. The GPS is defined as a separate domain in the Pfam database (dark purple). B) Structures of the GPCR autoproteolysis-inducing (GAIN) domain of ADGRL1 and C) ADGRB3 by Araç et al. (2012) show that the GPS motif is part of a more complex fold comprising 13 β sheets and 2 α helices. D) After cleavage, NTF and CTF remain attached by numerous hydrogen bonds shown between the cleaved β -strand (orange) and the surrounding β -strands (purple) in ADGRL1. The cleavage site is indicated with a black star in B), C) and D) . 76

- 5.3 **T/S→G mutation at the GPS site enables the production of soluble recombinant adhesion GPCR ectodomains for CRISPRa screening.** A) Constructs expressing truncated ectodomains of ADGRL4 and ADGRG1 do not produce biotinylated proteins at the expected sizes as observed by western blotting. 10 μ L of culture supernatant was loaded in each well. Detection of biotinylated proteins was performed by incubating blot with streptavidin conjugated to HRP and visualised with chemiluminescent peroxidase substrate. B) Diagram of truncated ectodomain encoded by the original constructs and full length ectodomain encoded in the new constructs. Full length ectodomains are resistant to cleavage at the GPS by a T/S→G mutation (red line) adjacent to the cleavage site. C) Constructs expressing full length ectodomains produce higher levels of biotinylated recombinant protein at the expected sizes. The same amount of culture supernatant as in A) was loaded in all wells and masses listed include predicted glycosylation. D) 13 adhesion GPCRs ectodomains were produced as biotinylated, His-tagged recombinant proteins and purified using Ni²⁺ affinity beads. Purified protein corresponded to their expected sizes as determined by SDS-PAGE and Coomassie staining. E) Recombinant ADGRE5 interacts with its endogenous ligand CD55. Increased absorbance at 485 nm indicate retention of β -lactamase-tagged CD55 prey in wells coated with recombinant ADGRE5 bait. Negative controls were performed with an unrelated protein, rCd200, which did not interact with either recombinant ADGRE5 or CD55. Bars represent blank subtracted mean \pm s.d.; $n=3$ 78
- 5.4 **CRISPRa screening identifies known interactions of ADGRL1 and ADGRL3, as well as glycosaminoglycan (GAG)-binding properties of ADGRA2.** A) Transformed gene enrichment P-values are plotted against a rank-ordered gene list for CRISPRa enrichment screens with cells selected using recombinant tetramers for ADGRL1 (left), ADGRL3 (right) and B) ADGRA2. An FDR cut-off of 0.1 was used to determine which genes were considered significantly enriched (red dots). C) Cell surface binding assays with SLC35B2 knockout (KO) HEK293 cells suggest that ADGRA2 binding is GAG-dependent. Fluorescently labelled ADGRA2 tetramers bound to wildtype HEK293 (red trace) but not SLC35B2 KO cells (blue trace). Unstained wildtype HEK293 cells (black trace) or cells incubated with streptavidin-PE (grey trace) were used as negative controls. A representative of three independent experiments is shown. 80

- 5.5 CRISPRa screening identifies novel interactions between ADGRB1 and Nogo receptors.** A) Guides targeting RTN4RL1 and 2 are enriched in a population of cells sorted for gain-of-function binding to ADGRB1 tetramers. A plot of normalised gRNA read counts in the sorted population against that of the plasmid library show increased abundance of RTN4RL1 and 2-targeting guides (left). Transformed gene enrichment P-values plotted against a rank-ordered gene list for a screen performed with ADGRB1 tetramers show that RTN4RL1 and 2 are the only genes found to be significantly enriched under an FDR of 0.1 (right). B) ADGRB1 tetramers stained cells transfected with cDNAs encoding full-length RTN4R, RTN4RL1, RTN4RL2 (blue lines) but not mock-transfected cells compared to a control ADGRL1 tetramer (orange line), or streptavidin-PE alone (red line). A representative of four independent experiments is shown. C) RTN4R-targeting guides were not enriched in the CRISPRa screen using ADGRB1 tetramers. Normalised read counts of all 5 RTN4R-targeting guides were similar between the ADGRB1-sorted population and the plasmid gRNA library (left). Transformed gene enrichment P-values plotted against gene rank also show that RTN4R was not highly ranked in terms of enrichment. D) Transfection of cells with cDNAs encoding full-length RTN4R, RTN4RL1, RTN4RL2 did not cause an increase in the levels of cell surface phosphatidylserine, a known ligand of ADGRB1, as determined by Annexin V staining of cells in comparison to mock-transfected cells. 83

- 5.6 **ADGRB1 specifically and directly interacts with Nogo receptors through the first three thrombospondin repeats (TSRs) in its ectodomain.** A) The ectodomains of ADGRB1 and RTN4R family members directly interact. The extracellular regions of the named receptors were expressed as soluble biotinylated bait proteins, captured in individual wells of a streptavidin-coated plate and probed for interactions with pentameric β -lactamase-tagged prey proteins. Binding is quantified by absorbance at 485 nm of a hydrolysis product of the colourimetric β -lactamase substrate, nitrocefin. Bars represent blank-subtracted mean \pm s.d.; $n=3$. ADGRE5-CD55 interaction was used as a positive control; negative control bait was the CD55 ectodomain. B) The Nogo receptor binding interface on ADGRB1 is composed of the N-terminal three TSR domains. Schematic of the Nogo receptor family and ADGRB1 proteins showing their domain organization (left). Binding of RTN4R and RTN4RL1 preys to fragments of ADGRB1 encompassing the full-length ectodomain (FL), thrombospondin repeats 1-3 (TSR1-3), TSRs 1-5, or the hormone receptor motif and GAIN domain (HRM+GAIN) is shown (right). Bars represent blank subtracted mean \pm s.d.; $n=3$ 85
- 6.1 **CRISPRa library guides target a non-cannonical isoform of CD86A)** Guides for CD86 were designed to target a region upstream of the TSS of NM_006889 as denoted by a black rectangle (CD86 target site). Although this site is associated with a predicted CAGE-seq TSS peak (FANTOM5 DPI peak) as well as epigenetic marks commonly associated with promoter regions (H3K27ac), the longer isoform encoded by NM_175862.4 is annotated as the cannonical isoform. B) Amino acid alignment of coding sequences of transcripts NM_006889 and NM_175862.4 showing only a difference of 6 amino acids within the signal peptide. 94

LIST OF TABLES

- 1.1 **Comparison of CRISPRa gRNA library specifications.** There are currently four genome-wide CRISPRa gRNA libraries available. SAM v1 and CRISPRa libraries were first generation libraries designed predominantly based on distance to TSS, whilst second generation libraries like CRISPRa-v2 and Calabrese utilised more complex algorithms taking into account nucleosome positioning and improved TSS prediction. SAM v1 targets all human RefSeq coding isoforms whilst CRISPRa targets a more restricted set of genes that are expressed in human K562 cell line. 19

- 3.1 **Properties of cell surface receptor panel selected for investigating CRISPRa efficiency.** 12 cell surface receptors of different membrane architectures were selected based on the availability of monoclonal antibodies, lack of expression in HEK293 cells (RPKM<2), and a mix of ubiquitous and restricted endogenous tissue expression. Expression in HEK293 cells are reported in reads per million kilobases (RPKM) derived from an RNA-seq dataset from Nam et al. (2014). TMD - transmembrane domain. 43

- 4.1 **Summary statistics of screens using antibodies against integrin $\alpha v \beta 3$ and CD200.** Gene rank and false discovery rate (FDR) of ITGB3 or CD200 for their respective screens after gene level enrichment analysis. Each replicate was analysed independently. ITGB3 is the top-ranking gene with an FDR of < 0.05 in all three replicate screens. CD200 is ranked first only in the first replicate screen but not in the other replicates, where it does not appear enriched (FDR > 0.05). 59

-
- 4.2 **Very few genes are enriched under a false discovery rate (FDR) of 0.25 for cells sorted at a 0.5% threshold** A total of nine 'hits' were identified at an FDR of 0.25. Seven are known to be GPI-anchored, whilst ANTXRL is a single-pass Type I protein and PIGV is a multi-pass transmembrane GPI mannosyltransferase involved in GPI-anchor biosynthesis. LFC - log fold change, GPI-linked - annotation based on UniprotKB/Swiss-prot database and literature. 61
- 4.3 **Top ranking genes using a sort threshold of 5%** With the increased sort threshold, a total of 26 'hits' were identified at an FDR of 0.25. 62
- 4.4 **Interactions detected by CRISPRa screening range from medium to low-affinity.** Published equilibrium dissociation constants (K_D) of several interactions tested, range from high nanomolar to micromolar. Low-affinity interactions are generally considered to have K_D s of above 1 μ M. CRISPRa screening identified the weakest interaction (CD55-CD97) but failed to detect the second weakest (CTLA4-CD86). The K_D s of interactions between EFNA1 and EPHA4/7 have not been published. 69