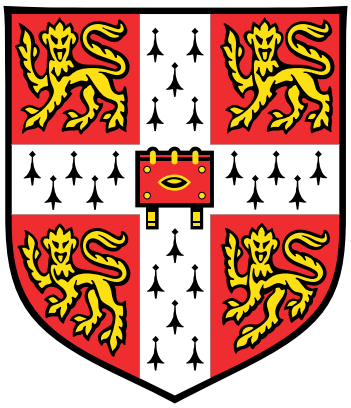
# Extracellular receptor-ligand interaction screening using CRISPR activation



**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 glycophosphatidylinositol (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

# LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

## 1.1 Extracellular receptor-ligand interactions

Extracellular protein-protein interactions are central to many biological processes including cell signalling, migration and fusion. Beyond interacting with proteins from the same organism, some membrane associated proteins are also exploited by pathogens to evade host immune recognition and enable successful reproduction in host tissue (Pizarro-Cerdá and Cossart, 2006). Therefore, determining extracellular receptor-ligand pairs essential to such processes is pertinent to gain a more complete understanding of signalling pathways that govern development and disease. In addition, cell surface proteins are attractive targets for drug discovery due to their accessibility to systemically administered therapeutics such as antibodies. The identification of conserved host-pathogen interactions at the cell surface has also helped to inform vaccine development (Ord et al., 2015).

Despite their importance in development and infection, interactions involving plasma membrane proteins are not well characterised due to their amphipathic nature and the specialised environment in which they function. This presents problems when using common biochemical methods for large-scale protein interaction screening, which are more often suited for studying intracellular complexes. As a result, interactions involving plasma membrane proteins tend to be underrepresented in large-scale proteomic interaction datasets (Futschik et al., 2007). Several methods have been developed to address the unique characteristics of extracellular interactions, but can be restricted to investigating certain classes

of membrane proteins or require large investments in terms of cost and equipment. Consequently, the development of more cost-effective and comprehensive approaches for large-scale extracellular interaction screening would facilitate the discovery of novel receptor-ligand pairs.

### 1.1.1 Challenges of studying extracellular protein-protein interactions

#### Solubilising membrane proteins with detergents interferes with mass spectrometry analysis

Intracellular proteins are often soluble in aqueous solutions, forming stable globular structures with solvent-exposed hydrophilic amino acid side chains facing outward and hydrophobic moieties hidden within the core of the protein. As a result, intracellular protein complexes can be easily solubilised for further characterisation without disrupting binding between members of the complex. In contrast, integral membrane proteins possess distinct hydrophilic and hydrophobic domains and may be dependent on insertion into the plasma membrane to maintain their 3-dimensional structure. Due to their amphipathic nature, membrane proteins are not as easily solubilised in aqueous solvents as intracellular proteins are. Detergents such as Triton X-100 and sodium dodecyl sulfate (SDS) can be helpful for solubilising proteins with hydrophobic domains, but may at the same time denature secondary or tertiary structures and therefore disrupt native protein complexes.

The use of detergents to solubilise membrane protein complexes is particularly relevant when using affinity purification in tandem with mass spectrometry (AP-MS) to characterise protein complexes. AP-MS involves the enrichment of protein complexes containing a specific protein, called a 'bait' protein, which is typically tagged and overexpressed in a cell line of interest. Enrichment of protein complexes from cell lysate is performed using a matrix with high affinity for the epitope tag. Purified complexes are denatured and size-separated by gel electrophoresis before being digested into peptides with trypsin. Rapid and sensitive identification of peptide sequences is achieved by tandem mass spectrometry and peptide fragments mapped to a database of possible fragments from known proteins in that species to elucidate protein identity and abundance (Huttlin et al., 2015). Although some detergent (0.05%) is generally used during protein extraction and purification to reduce non-specific interactions and loss of protein due to adsorption to surfaces, the higher concentrations of detergent (0.5 - 1%) needed to solubilise many membrane protein complexes are known to interfere with tryptic digest as well as mass spectrometry analysis of the resulting peptides

(Zhang and Li, 2004). In addition, membrane proteins are often glycosylated, which alters the mass of subsequent peptides in unpredictable ways, making them challenging to identify using mass spectrometry. Unsurprisingly, membrane protein complexes are therefore underrepresented in large-scale interactome studies using AP-MS (Huttlin et al., 2015).

### Low affinity interactors may be lost during multiple stringent wash steps

Cell surface receptors diffuse within the plasma membrane and are thought to sometimes form local concentrations of receptors together with accessory proteins required for functional signalling. Clusters of proteins also serve to increase the avidity of an interaction occurring by providing multiple binding sites for receptor-ligand interaction. A corollary of this is that extracellular interactions, especially those between two cell surface proteins, can have low monomeric affinities and still be biologically functional. For instance, the T-cell receptor co-regulatory complex CD55-CD97, has a very high equilibrium dissociation constant ($K_D$) of 86 x $10^{-6}$ M, indicating a very low affinity. In contrast, most antibodies have $K_D$ values in the nanomolar range ($10^{-7}$ to $10^{-9}$ M). Blocking the CD55-CD97 interaction with antibodies against the extracellular domains of either CD55 or CD97 results in a significant inhibition of T-cell proliferation, indicating that interactions of such low affinity can still be functionally important (Abbott et al., 2007). Therefore, it is critical that screening technologies aimed at identifying novel extracellular interactions should be sensitive enough to detect such low affinity interactions. Further to the use of detergents reducing the likelihood of detecting low-affinity binding partners, affinity purification requires multiple stringent wash steps to reduce background contaminants due to non-specific binding to the affinity matrix, which may cause further loss of low-affinity interactors. This makes AP-MS less than ideal for identifying extracellular receptor-ligand interactions.

### Detecting transient interactions with mass spectrometry is challenging

Low affinities facilitate the formation of extremely transient interactions needed for dynamic processes regulated by cell surface molecules. Leukocyte extravasation is a good example of an interaction that undergoes quick formation and dissociation as leukocytes roll and attach to endothelial cells under flow rates of 5-10 $\mu$m/s (van der Merwe and Barclay, 1994). This process is mediated by three lectins present on the endothelium (P-selectin and E-selectin) and on leukocytes (L-selectin), and is thought to be dependent on fast dissociation rates of the lectins from their respective binding partners. In addition to

dissociation rates, highly regulated expression of binding partners at the surface can also give rise to transient interactions. For instance, P and E-selectin are not expressed on the surface of quiescent endothelial cells but upregulated in response to pro-inflammatory factors like tumour necrosis factor (TNF)-$\alpha$ and interleukin (IL)-1. P-selectin is usually present in storage granules called Weibel–Palade bodies and is rapidly transported to the cell surface upon endothelial cell activation. This stimulates transcription of E-selectin from the *SELE* gene which is then trafficked to the plasma membrane. Maximal expression of E-selectin occurs around 6–12 hours after cytokine stimulation, with levels returning to baseline after 24 hours (Leeuwenberg et al., 1992). The short timeframe in which E-selectin is present on the surface highlights how dynamically regulated membrane proteins can form interactions only transiently.

To address the challenges of capturing low-affinity or transient extracellular interactions with mass spectrometry-based approaches, a chemically-defined crosslinking reagent, TRICEPS, has been developed to covalently link a ligand of interest to glycoproteins at the cell surface (Frei et al., 2012). TRICEPS is a trifunctional reagent that contains an N-hydroxysuccinimide (NHS)-ester for coupling to polypeptide ligands, a hydrazine moiety for capturing glycoprotein receptors on the cell surface, and a biotin moiety for affinity purification of the cross-linked receptor-ligand complex. The suggested workflow also involves performing the trypsin digest before rather than after affinity purification, which circumvents issues associated with purifying intact plasma membrane proteins, such as limited solubility and nonspecific interactions through exposed hydrophobic domains. This approach may enable the identification of receptors under near-physiological conditions; however it may result in high background signals due to nonspecific crosslinking to neighbouring glycoproteins as well as more complicated analysis of mass spectrometry data to account for the presence of the cross-linking agent in the sample.

Although AP-MS is a sensitive technique for detecting proteins, the purification of specific protein complexes before MS analysis necessitates a large amount of starting material, typically from cell lines, which can be easily expanded, or from tissue lysates provided the interaction of interest is fairly stable and abundant in the tissue. This poses a challenge for detecting interactions which occur in a short timeframe or between rare cell types as it could be difficult to obtain sufficient amounts of primary material for AP-MS.

### Lack of proper post-translational modifications hinders analysis by yeast-two hybrid

Many proteins undergo post-translational modifications crucial for their function. In eukaryotes, the majority of proteins synthesised in the endoplasmic reticulum (ER) are modified by the addition of carbohydrates in a process called glycosylation. Glycosylation results in the addition of a glycosyl group to either asparagine, hydroxylysine, serine, or threonine residues in a polypeptide. Secreted and membrane-bound proteins are synthesised in the ER before being trafficked to the surface. Consequently, most secreted and membrane-bound proteins are glycosylated, and the addition of these large, bulky oligosaccharide chains modulates binding properties as well as receptor recognition (Ulloa-Aguirre et al., 1999). In some cases, the sugar moieties themselves can function as adhesive molecules independent of the core protein they are attached to. This is exemplified by the role of glycosaminoglycans (GAGs) in infectious disease, where they have been shown to act as receptors for initial attachment of a wide variety of microbial pathogens (Jinno and Park, 2015). GAGs are formed by the sulfation of mannose and N-acetylglucosamine (GlcNAc) residues of a carbohydrate side chain in a process catalysed by sulfotransferases using 3'-phosphoadenosine-5'-phosphosulfate (PAPS) as a sulfuryl donor.

Yeast-two hybrid (Y2H) is a genetic system to detect direct binary interactions between proteins in a living cell. Y2H uses the yeast Gal4 transcription factor, which consists of distinct DNA-binding and transcriptional activation domains, to bind to a conserved Upstream Activation Sequence (UAS) to regulate the expression of galactose-induced genes. Through the expression of two hybrid proteins - a bait protein fused to the N-terminal DNA-binding domain of Gal4 and a prey protein fused to the C-terminal transcriptional activator domain - interactions between bait and prey can be detected by measuring transcription levels of a reporter gene with a UAS (Fields and Song, 1989). An adaptation of the Y2H system to study interactions involving membrane proteins (MYTH) utilises a split ubiquitin system where bait proteins are fused to a C-terminal fragment of ubiquitin, as well as a transcription factor, and prey proteins are fused to the N-terminal fragment of ubiquitin. Interaction between bait and prey proteins results in the formation of a full-length 'pseudoubiquitin' molecule, which can be recognised and cleaved by cytosolic deubiquitinising enzymes. Cleavage releases the transcription factor which then enters the nucleus and activates a reporter gene to provide a readout for the presence or absence of interaction (Snider et al., 2010). This method is a cheap, efficient strategy for detecting interactions between intracellular proteins, and has been performed at scale to detect binary interactions at a proteome level (Rolland et al., 2014; Snider and Stagljar, 2016). Crucially, Y2H and MYTH proteins are expressed in yeast cells,

which may not fully recapitulate the range of post-translational modifications needed for human membrane proteins. For instance, *Saccharomyces cerevisiae* proteins undergo both N- and O-linked glycosylation, but do not form complex carbohydrate structures (Tanner and Lehle, 1987).

In addition, Y2H is performed in the reducing environment of the cytoplasm, which does not favour the formation of disulfide bonds needed to provide extracellular proteins with the proper tertiary structure for interaction (Feige and Hendershot, 2011). Disulfide bonds are formed between sulfhydryl groups on cysteine residues, and due to their covalent nature form extremely stable structures crucial for ligand binding and cell adhesion. The presence of hydrophobic transmembrane domains in integral membrane proteins also precludes them from being expressed as both soluble and functional baits/preys for Y2H, unless truncated to express only hydrophilic domains. *In vivo*, secreted and membrane-associated proteins are synthesised in the ER, which maintains an oxidising environment that better resembles the extracellular milieu, along with chaperones for protein folding and insertion of integral membrane proteins into the ER membrane.

In summary, extracellular proteins and their interactions have unique attributes that allow them to function extracellularly in a membrane-embedded context. These attributes render such interactions unsuitable for investigation using standard biochemical techniques like AP-MS and Y2H, and require novel strategies for accurate and sensitive characterisation.

### 1.1.2 Interaction screening using soluble recombinant ectodomains

Several studies have addressed the challenge of extracellular interaction screening by performing plate-based screens using libraries of recombinant secreted and plasma membrane proteins from cell types known to interact (Bushell et al., 2008; Özkan et al., 2013; Wojtowicz et al., 2007) (Figure 1.1). In these screens, truncated ectodomains of membrane proteins are fused to various biochemical tags to form soluble probes for directly detecting binding between ectodomains or secreted proteins. Binding is generally indicated by the retention of an ectodomain fused to an enzymatic tag, which is detected by addition of a colourimetric substrate. In order to detect low-affinity interactions, ectodomains are also often oligomerised to increase local avidity during the binding assay.

The exact strategies for affinity capture onto plates, ectodomain oligomerisation, and interaction detection differ between studies. Wojtowicz *et al.* expressed the ectodomains of 92 isoforms of the highly polymorphic Drosophila Down Syndrome cell adhesion molecule

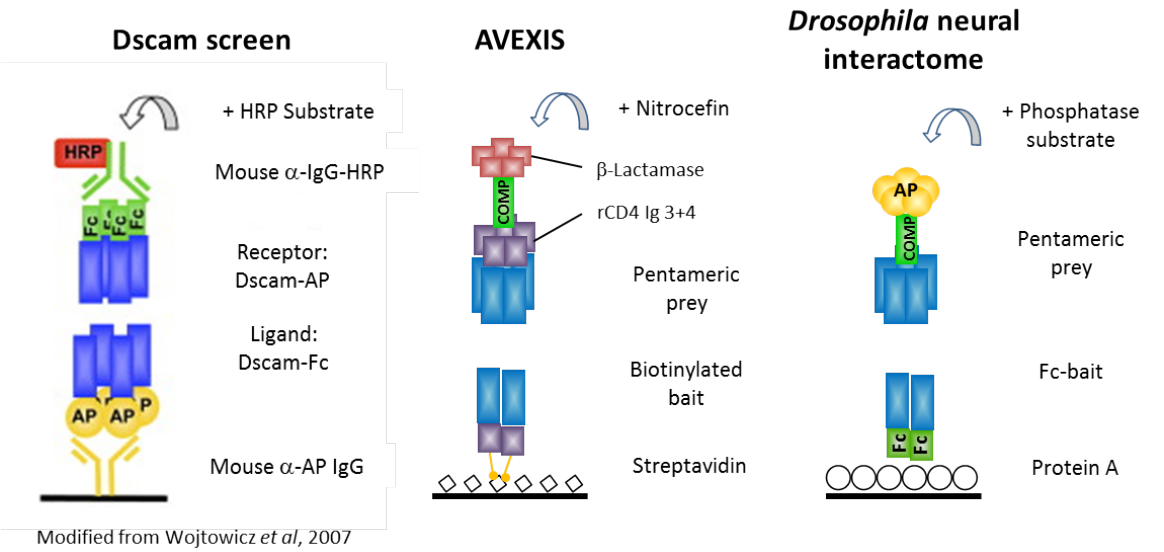



**Figure 1.1 Various strategies for plate-based interaction screening.** The ectodomains of cell surface receptors are produced as soluble recombinant form fused to biochemical tags for immobilisation, oligomerisation and detection by enzymatic activity. Immunoglobulin Fc-fusions (Fc) and the pentamerising domain of rat cartilage oligomeric matrix protein (COMP) mediate oligomerisation for increased avidity. Alkaline phosphatase (AP), horseradish peroxidase (HRP) and $\beta$-lactamase provide enzymatic activity for detection with the relevant substrates. AP and Fc tags also facilitate immobilisation onto $\alpha$-AP antibody or Protein A-coated plates. Biotinylation enables capture and clustering on streptavidin-coated plates.

(Dscam) gene as alkaline phosphatase (AP)-tagged and immunoglobulin Fc-tagged constructs (Wojtowicz et al., 2007). AP-tagged ectodomains were captured on an anti-AP antibody coated 96-well plate before incubation with Fc-tagged ectodomains. Fc-fusion proteins are generally expressed as homodimers due to the disulfide bond in the hinge region of the Fc domain, increasing its avidity as compared to a monomeric probe (Czajkowsky et al., 2012). After washing to remove unbound Fc-fusion proteins, any remaining Fc-tagged ectodomains were detected with an anti-human IgG antibody conjugated to horseradish peroxidase (HRP), and detected with an appropriate HRP substrate. This study showed that a large number of DSCAM isoforms interact in a homophilic manner, providing a mechanistic explanation for how they mediate self-avoidance during neural circuit formation.

In contrast, Bushell *et al.* expressed the ectodomains of 110 zebrafish immunoglobulin (Ig) proteins as biotinylated 'baits' and pentameric 'preys' (Bushell et al., 2008). Both constructs contained a fusion of the ectodomains to the third and fourth Ig domains of rat Cd4 as a carrier. Carrier domains fold independently of other domains and can boost ex-

pression of a recombinant protein (Brown and Barclay, 1994). Baits contained an additional 17-amino acid biotinylation peptide, which becomes biotinylated when co-expressed with biotin ligase (BirA) from *E. coli*, whilst preys were additionally fused to the pentamerisation domain of rat cartilaginous oligomeric matrix protein (COMP) and the enzyme $\beta$-lactamase, separated by a flexible protein linker sequence (-LDRNLPPLAPLGP-). Streptavidin-coated plates were used to capture biotinylated baits. Interactions were detected by incubating with pentameric preys and captured preys were detected using a nitrocefin hydrolysis assay. This strategy, termed Avidity-based extracellular interaction screening (AVEXIS), eliminates an antibody incubation and wash step needed in the Dscam screen, thereby shortening the screen duration and reducing the loss of low affinity interactors during wash steps. The use of pentamers rather than dimers for prey proteins could also facilitate the detection of low-affinity interactions, although no systematic comparison has been conducted.

More recently, Ozkan *et al.* (2013) investigated the interactions between 202 *Drosophila* neuronal proteins by expressing them as Fc-fusion baits and pentameric, AP-tagged preys. The 202 proteins included members of the immunoglobulin superfamily (IgSF), fibronectin type III (FnIII), and leucine-rich repeat (LRR) families. Fc-fusion baits were captured using Protein A-coated plates, whilst interactions were detected by assaying for phosphatase activity using colourigenic BluePhos phosphatase substrate. Again, the benefit of oligomerising both bait and prey proteins was not explicitly demonstrated, although it could ostensibly promote the detection of very low affinity interactions. However, it could be argued that capturing baits on streptavidin also clusters them, providing little benefit over producing biotinylated monomeric baits. This study uncovered a set of previously unknown interactions between a 20 member subfamily of defective-in-proboscis-response IgSF proteins, and showed that they selectively interact with an 11 member subfamily of previously uncharacterized IgSF proteins.

An important similarity between interaction screening studies described above is the investigation of protein families with single, contiguous ectodomains containing distinct structural domains. To produce soluble ectodomain fusions that retain the ability to bind their native ligands, these ectodomains should consist of a single polypeptide chain that is able to fold independently of the transmembrane region. This presents difficulties for investigating multimeric receptors or integral membrane proteins with more than one transmembrane domain, as ligand binding interfaces may be formed from non-contiguous ectodomains, even within the same protein. Integrins are a prime example of heterodimeric receptors, as they are composed of an $\alpha$ and $\beta$ chain. At least 18 $\alpha$ subunits and eight $\beta$ subunits have been identified in humans, which are able to generate 24 different integrins with distinct binding

specificities (Stupack and Cheresh, 2002). Although another study has tried to address this by constructing different vectors to express tagged ectodomains of heterodimeric receptors and proteins of different membrane topologies (Sun et al., 2012), identifying extracellular interactions involving multimeric or structurally unstable binding sites on a large scale remains a challenge.

### 1.1.3 Interaction screening using cDNA overexpression

Ectopic surface expression of functional multipass membrane proteins is routinely achieved by transient transfection of mammalian cells with full-length cDNA constructs. The expression of transmembrane proteins in their native context increases the likelihood of having endogenous glycosylation patterns as well as the proper extracellular domain conformation. Expression cloning is a method of identifying genes that code for proteins with a specific attribute, and has been used to identify the receptors of numerous hormones and other secreted ligands (Simonsen and Lodish, 1994). This strategy involves creating pools of cDNA clones from cell or tissue lysate, transfection into mammalian cells, and selection of pools based on binding to a ligand of interest. cDNA clones from that pool are expanded and split into further pools which undergo the same process until the binding phenotype can be attributed to one or a few transcripts which are identified by sequencing. Instead of pooling, cells with binding phenotype can also be selected by cell sorting, separation with magnetic beads, or 'panning' approaches. However, the construction of good cDNA libraries from cells is technically challenging and can be heavily biased towards shorter or incomplete inserts. In addition, cell surface receptors are generally not highly expressed in cells, and as a result large fractions of the cDNA library would be occupied by transcripts coding for cytosolic housekeeping proteins that were not of interest. Although strategies have been developed to deplete expression cloning libraries of unwanted transcripts, this is challenging and not always fully effective. This means not only that a lot of primary material would be needed to capture transcripts of interest, but also that a large number of cells have to be screened in order to detect one or two positive events (Simonsen and Lodish, 1994).

After the sequencing of the human genome, large collections of individual clones containing defined cDNA transcripts became available, avoiding the need to create cDNA libraries from transcripts expressed in particular cell lines or tissues. Transfected cell microarrays were developed to utilise this resource for interrogating the function of genes in a systematic, unbiased, and high-throughput fashion (Ziauddin and Sabatini, 2001). For extracellular interaction screening, clusters of cells expressing different cell surface receptors

are formed by printing nanolitre volumes of plasmids and transfection reagent on glass slides in a known layout, then covering the slides with mammalian cells in medium. The glass slides are then incubated with a fluorescently-labelled ligand of interest and examined for clusters of cells which show increased ligand binding. Since each cluster overexpresses a single receptor, the positions of ligand-binding cell clusters would indicate the identity of the receptor. This approach circumvents technical challenges associated with constructing cDNA libraries from RNA, including overrepresentation of cytosolic housekeeping proteins and bias towards shorter inserts. In addition, sensitivity is increased from concentrating the signal in a localised region as compared to pooled expression cloning selection where only few cells in a well might show increased ligand binding. Known multi-subunit receptors can also be screened using this method by co-transfection of plasmids encoding individual subunits. The use of transfected cell microarrays for identifying important receptors involved in infection and metabolism (Mullican et al., 2017; Turner et al., 2013) highlights the utility of cell-based overexpression assays for more comprehensive screening of proteins with complex membrane architectures.

Nonetheless, transfected cell microarrays have yet to be widely adopted for extracellular interaction screening in individual laboratories due to high set-up costs needed in terms of procuring a comprehensive set of plasma membrane protein cDNA clones and hardware needed for arraying reagents. Instead, transfected cell microarray screening is typically provided as a service by biotechnology or pharmaceutical companies. Thus, an approach for large-scale extracellular interaction screening that involves low set-up costs and little specialised equipment could provide more flexibility for researchers to fine-tune their assays for studying any ligand of interest. Such a screening platform should be able to test for interactions against ideally all cell surface receptors encoded in the human genome, and be sensitive enough to detect low-affinity interactions with micromolar $K_{\mathrm{D}}$.

## 1.2 CRISPR/Cas9-based transcriptional activation

Recent developments in CRISPR/Cas9 technology have allowed for specific transcriptional activation of target genes on a genome-wide scale. Cas9 is an RNA-guided DNA endonuclease functioning alongside Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) in bacteria to recognise and cleave foreign DNA elements (Sapranauskas et al., 2011). Since its discovery, this system has been adapted to manipulate eukaryotic genomes and transcriptomes by delivering wildtype or modified versions of Cas9, along with

a single guide RNA (gRNA), into eukaryotic cells. Each gRNA contains a 20 nucleotide (nt) long guide sequence complementary to a target genomic locus upstream of a Protospacer Adjacent Motif (PAM), and also encodes a scaffold for forming a complex with Cas9. Nuclease-deficient Cas9 (dCas9) fused to transcriptional activator domains can also complex with gRNA, and result in transcriptional upregulation when directed to the promoter regions of genes in a process termed CRISPR activation (CRISPRa).

For high-throughput gain-of-function screening, complex pools of different 20 nt guide sequences can be easily synthesised to form gRNA libraries targeting multiple promoter regions. This overcomes some of the technical challenges associated with using cDNA libraries for overexpression screening, such as restrictions on insert size. Moreover, each 20 nt guide sequence doubles up as a unique barcode that can be detected by next generation sequencing (NGS), and multiple gRNAs targeting the same promoter provide degeneracy in the library as well as a greater ability to distinguish between biological effects and gRNA-specific artefacts. Thus, the ability to overexpress virtually all receptors in the genome regardless of transcript length and assay for binding in the context of the plasma membrane makes CRISPRa an attractive potential platform for extracellular interaction screening.

### 1.2.1 Origins as a bacterial adaptive immune system

CRISPR loci were first discovered in bacteria and gained interest due to the unusually regular repetitions of DNA sequences separated by short, regularly-sized 'spacer' sequences, that occur throughout the loci. The importance of CRISPR arrays, along with flanking CRISPR-associated (Cas) genes, to the survival of prokaryotic cells is underscored by their presence in approximately 40% and 90% of sequenced bacterial and archaeal genomes respectively (Sorek et al., 2008). Spacer sequences were found to originate from bacteriophages or plasmids, and to mediate sequence-specific cleavage of these foreign genetic elements by Cas effector proteins (Mojica et al., 2005). Several types of CRISPR systems have been discovered and classified based on their effector modules and signature Cas proteins. Of these, Cas9 is a single effector protein characteristic of Class II, Type II CRISPR systems, which means that it performs gRNA binding, genomic DNA binding, and nuclease activity with a single protein. In contrast, Class I CRISPR systems are characterised by multi-subunit effector complexes where nucleic acid binding and nuclease activity are performed by separate Cas proteins (Makarova et al., 2015). Although the CRISPR/Cas9

system was not the first CRISPR system discovered, the simplicity of having a single effector like Cas9 has made it the system of choice for heterologous use.

For genome editing in mammalian cells, both Cas9 and at least one guide RNA (gRNA) must be introduced into the cell. The gRNA provides targeting specificity through complementary base-pairing between the target genomic loci and a 20 nt region within the gRNA scaffold. The rest of the scaffold forms a complex with Cas9 and samples PAM-containing sites in the genome by random diffusion/collision (Sternberg et al., 2014). PAMs differ between species, although the most commonly used Cas9 proteins come from *Streptococcus pyogenes* and recognise a 5'-NGG-3' motif. Upon reaching a site which is complementary to the guide sequence, Cas9 initiates a double strand break just upstream of the PAM (Figure 1.2A). In most cells, this double strand break is repaired by non-homologous end joining (NHEJ), resulting in small deletions or insertions ranging from 1-15 nt at the double strand break (Brandsma and Gent, 2012). Thus, targeting the Cas9-gRNA complex to coding exons of a gene often results in a frameshift mutation and functional knock-out of the encoded protein.

### 1.2.2 Transcriptional activation with CRISPR/Cas9

Cas9 contains two conserved nuclease sites, HNH and RuvC, for cleaving both strands of genomic DNA. The HNH domain is responsible for cleaving the target DNA strand (complementary to gRNA) whilst the RuvC domain cleaves the non-target DNA strand (Jinek et al., 2012). Inactivating mutations to either one of the nuclease domains create Cas9 nickases, which can only catalyse single-stranded DNA breaks and can be useful for promoting homologous recombination with a supplied donor template. A fully inactive Cas9 enzyme (dCas9) with mutations to both nuclease domains is devoid of nuclease activity but retains its RNA-guided DNA binding activity, transforming it into a nucleic acid-programmable DNA-binding protein. The *S. pyogenes* dCas9 (D10A/H840A) mutant is most commonly used. When recruited near a transcriptional start site (TSS), dCas9 blocks transcription and elongation, resulting in a reduction in gene expression. However, the addition of tethered activation domains like VP16 to dCas9 enabled transcriptional upregulation when targeted to regions upstream of the TSS (Maeder et al., 2013) (Figure 1.2B). VP16 is a transcription factor from the herpes simplex virus involved in the expression of viral immediate-early genes. *In vivo*, VP16 binds DNA indirectly through host factors Oct-1 and human factor C1 (HCF), and initiates transcription by recruiting appropriate factors through a C-terminal transcriptional activation domain. The VP16 activation domain

interacts with numerous host factors including basal transcription factors and the Mediator complex, which then recruits RNA polymerase II. VP16 also binds histone modifiers such as the SAGA and NuA complexes to promoters, causing chromatin decondensation detectable by fluorescence microscopy (Tumbar et al., 1999). In dCas9 fusion proteins, the VP16 activation domain is commonly present in four copies for enhanced activity, and referred to as VP64.

An issue with simple dCas9-VP64 CRISPRa systems was that many gRNAs did not induce activation on their own, but only when combined with other gRNAs targeting the same locus. Furthermore, targeting of endogenous genes gave modest levels of activation, with most displaying under ten-fold difference at mRNA level (Gilbert et al., 2013). An early attempt to boost gene induction by adding an N-terminal VP64 fusion, creating a dCas9 enzyme flanked by two VP64 domains, showed modest increase in efficacy of about four-fold compared to using single C-terminal dCas9-VP64 fusions for a single gene, Myod1 (Chakraborty et al., 2014). The need for multiple gRNAs targeting the same loci for significant mRNA upregulation severely limited the use of CRISPRa, especially for high-throughput pooled screening applications where each cell receives only one gRNA and large effect sizes are desirable for better signal-to-noise ratios.

### Synergistic activation domains facilitate transactivation

Second generation CRISPRa systems sought to address this issue by using synergistic combinations of activator domains or chromatin modifiers. To mimic the coordinated recruitment of transcriptional machinery *in vivo*, Chavez *et al.* screened a series of candidate effectors for activation of a fluorescent reporter in HEK293 cells and selected the three most active activation domains (VP64, Rta, and p65) to create a tripartite activator fused directly to dCas9. Along with the respective gRNAs, this dCas9-VPR fusion was capable of inducing mRNA levels of more than 100-fold greater than that with dCas9-VP64 alone (Chavez et al., 2015). Rta (Replication and Transcription Activator) is an immediate-early gene product found in murine gammaherpesviruses. The endogenous function of Rta is to activate viral lytic genes by binding to a 27 bp RTA-responsive element (RRE) in the promoter regions of target genes. The transactivation domain of Rta has been compared to VP16 in competition assays and shown to activate genes by distinct mechanisms than VP16, suggesting that these domains can synergise (Hardwick et al., 1992). The third activator domain comes from p65 (also known as RelA), a human transcription factor and member of the NF-$\kappa$B family. Accordingly, p65 is involved in innate and adaptive immune responses, and can be

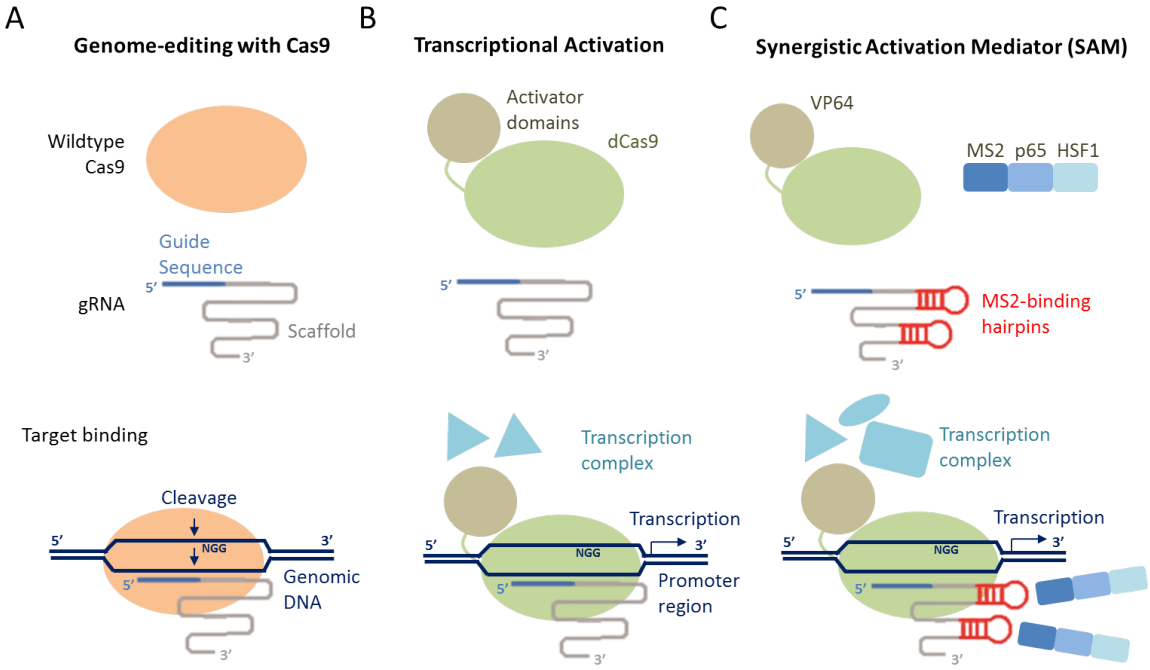



**Figure 1.2 Original and modified CRISPR/Cas9 systems.** A) Nuclease-active Cas9 complexes with gRNA and is targeted to genomic loci complementary to a 20 nt guide sequence at the 3' end of the gRNA and upstream of a PAM sequence (NGG). Upon binding, Cas9 catalyses a double strand break upstream of 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.

activated by proinflammatory factors TNF-$\alpha$ and IL-1. The transactivation domain of p65 interacts with members of the basal transcription complex such as TATA-binding protein and Transcription factor II B. Notably, p65 recruits transcription factors distinct from VP16, such as Activator protein 1 and the cAMP response element binding protein (CREB) family (van Essen et al., 2009).

In contrast, Konermann *et al.* opted for an engineered gRNA scaffold containing MS2-binding hairpin loops to recruit additional transactivator domains from p65 and HSF1 to the dCas9-VP64:gRNA complex (Figure 1.2C). The resulting system, called Synergistic Activation Mediator (SAM), demonstrated higher levels of transcriptional activation using a single guide compared to using multiple guides without the engineered scaffold (Konermann et al., 2015). HSF1 is a human transcription factor and the major mediator of the heat shock response. HSF1 induces transcription of heat shock proteins in response to environmental stress through a 150 amino acid long C-terminal region containing two transcriptional activator domains TAD1 and TAD2. TAD1 interacts with TATA-box binding protein-associated factor TAF9, as well as members of the histone remodelling SWI/SNF complex, whilst little is known about the binding partners of TAD2 (Dayalan Naidu and Dinkova-Kostova, 2017).

Rather than using different activator domains, Tanenbaum *et al.* sought to boost CRISPRa efficiency by recruiting multiple copies of VP64 to a repeating peptide array using single chain variable fragment (svFc) antibodies. A svFC consists of the epitope binding regions of the light and heavy chains of the antibody, fused to form a single polypeptide. Unlike conventional antibodies which generally do not fold properly in the cytoplasm, svFCs have been successfully expressed in soluble form in cells. dCas9 was fused to a peptide array containing 10 copies of a short peptide epitope, and expressed along with the corresponding svFC fused to VP64. This strategy was able to increase transcription of endogenous *CXCR4* by 10 - 50 fold of that inducd by dCas9-VP64 alone (Tanenbaum et al., 2014). Importantly, all three second generation CRISPRa sytems demonstrated the ability to achieve robust upregulation of mRNA abundance using a single gRNA. A systematic comparison of the three second generation CRISPR activation systems confirmed the increased efficiency of second generation systems over dCas9-VP64, and found that the SAM system from Konermann *et al.* was the most consistent in delivering high levels of upregulation across multiple genes in HEK293T cells (Chavez et al., 2016).

### Epigenetic modifiers promote transcription and increase deposition of active chromatin marks

Besides transactivating domains, CRISPRa has also been achieved by directly fusing chromatin modifiers to dCas9. Although transcriptional activators like VP16 and Rta recruit chromatin modifiers as part of their transactivating mechanism, fusion of a single histone acetyltransferase domain from p300 (p300core) to dCas9 was shown to induce a significant increase (between 7 to 200 fold) in gene expression of *MYOD*, *OCT4* and *HBD* from promoter regions, as well as both proximal and distal enhancers up to a distance of 46 kb from the TSS (Hilton et al., 2015). No increase in expression was observed when dCas9-VP64 was used and targeted to enhancer regions. Strikingly, the use of dCas9-p300core, but not dCas9-VP64, resulted in increased levels of histone acetylation, specifically of the $27^{th}$ lysine residue of histone 3 (H3K27ac), at promoter regions of the human $\beta$-globin locus when an associated enhancer was targeted. p300 is a transcriptional coactivator first described through its interactions with adenoviral protein E1A (Chan and La Thangue, 2001). This interaction causes a loss of cell cycle control, and inactivating mutations in p300 have been described in several types of cancer, indicating a tumour suppressor role. p300 contains a core histone acetyltransferase domain flanked by two transactivation domains (Chan and La Thangue, 2001). A bromodomain within the acetyltransferase domain recognises acetylated residues and facilitates the transfer of an acetyl group from acetyl-CoA to the $\varepsilon$-amino group of a lysine residue (Dancy and Cole, 2015). Histone acetylation is a key mechanism in regulating transcription and is generally associated with euchromatin and actively transcribed promoters.

## 1.2.3 Genome-scale gain-of-function screening using CRISPR activation

Second generation CRISPRa systems have been used for genome-scale gain-of-function screening to identify genetic factors influencing cell growth and sensitivity to Cholera-diphtheria Toxin (CTx-DTA) in human myeloid leukemia K562 cells, as well as resistance of the A375 melanoma cell line to BRAF inhibitor vemurafenib treatment (Gilbert et al., 2014; Konermann et al., 2015). In the cell growth screen, the fraction of cells expressing gRNAs and the CRISPRa system was stable over the course of the experiment, indicating that there was no general toxicity associated with the CRISPRa platform used, and that overexpression with CRISPRa was specific. In addition, many of the genes which inhibited growth in the screen had previously known functions in regulating cell cycle and differentiation, including tumour suppressor genes, transcription factor families

involved in differentiation, and genes involved in the negative regulation of mitosis. Similarly, BRAF inhibitor resistance screening identified a number of gene candidates that confirmed known resistance pathways from previous knockout and knockdown screens. For instance, reactivation of the ERK pathway is one of the more well-studied resistance mechanisms, and components of this pathway were enriched in the gain-of-function screen. Taken together, these results highlight the feasibility and utility of CRISPRa for genome-scale gain-of-function screening.

### Guide RNA library design

A key component of pooled CRISPRa screening is the gRNA library used. Since the gRNA contains a 20 nt guide sequence responsible for targeting the dCas9-activator complex, the library of gRNAs used defines the search space of the screen and can be designed to target all known promoters in the genome or focused on a subset of genes. Typically, five to ten guides are designed to target an individual locus, as guides vary in their efficiency and off-target profile (Haeussler et al., 2016). This degeneracy provides a buffer for ineffective guides that might have inadvertently and unknowingly been included in the library, and facilitates the differentiation of biological signal from technical artefacts generated by off-target effects of an individual guide. However, a trade-off between degeneracy and practicality exists because the scale of the screen increases with library complexity to maintain sufficient coverage of the library in a pooled screen. A meta-analysis of CRISPR knockout libraries suggests that six guides per gene represents the best trade-off as the amount of information provided by additional guides past six drops rapidly (Ong et al., 2017). As of yet, no similar analyses have been performed with CRISPRa libraries to determine if this can be generalised across different CRISPR screening modalities.

Clearly, any gRNA library would benefit from having a small number of highly active guides per promoter region. Therefore, it is essential to select the most efficient and specific guides to include in a library. The rules for gRNA design have been largely determined for CRISPR knockout systems, but research is underway to determine if these rules apply to CRISPRa as well. A small survey of off-target modifications using nuclease-active Cas9 and 26 gRNAs found that whilst the ratio of off-target modification to on-target modification frequency was generally low (<1%), sites that accounted for 88.3% of off-target modification contained up to four mismatches to the guide sequence (Haeussler et al., 2016). This suggests that when counting the number of potential off-target sites for a guide, sequences with up to four mismatches should be included. However, as dCas9 does not cause

double strand breaks and therefore toxicity, the number of potential off-target sites might be less relevant to CRISPRa as compared to whether they occur within the promoter region of another gene.

On-target gRNA efficiency is affected by gRNA expression levels and other criteria specific to the CRISPR modality used. gRNAs are transcribed by RNA polymerase III (RNAPIII), and in most expression vectors are driven by a U6 or H1 promoter. The murine U6 promoter favours transcription initiation on purine nucleotides (A and G) (Ma et al., 2014), so a common strategy for enhanced gRNA expression is to add a guanine nucleotide to the 5' end of 19 nt guide sequences. In addition, RNAPIII recognises a poly-T termination signal that causes catalytic inactivation and subsequent release of the polymerase from the transcript (Nielsen et al., 2013), suggesting that poly-T stretches of more than three thymine nucleotides within the guide sequence should be avoided.

In general, such factors affecting gRNA expression should be applicable to both CRISPR knockout and CRISPRa as both systems use similar gRNA expression vectors. However, other rules governing gRNA activity relate specifically to each CRISPR modality. For instance, targeting nuclease-active Cas9-gRNA complexes to early exons increases the likelihood of fully disrupting protein function and is thus favourable when designing CRISPR knockout gRNA, but is not as relevant for designing CRISPRa gRNA libraries. Instead, a machine learning algorithm trained on data from nine CRISPRa screens revealed that distance from the TSS and nucleosome positioning were major determinants of CRISPRa gRNA activity (Horlbeck et al., 2016). Fitting the positions of gRNA relative to TSSs using support vector regression showed a broad window of moderate activity between -500 to -50 bp, with a narrow peak between -100 and -250 bp where gRNAs are more likely to be highly active. Periodic patterns of highly active guides that were anti-correlated with nucleosome positioning also indicated an inhibitory effect of nucleosomes on the formation of CRISPRa complexes.

Consequently, designing gRNA libraries for genome-scale screening requires careful selection of parameters including the number of gene targets, number of guides per promoter, positional window for guide selection, and criteria for ranking or filtering guide sequences. Several genome-wide CRISPRa libraries are available and a quick review might provide some insight into the 'best practices' of gRNA library design (Table 1.1). First generation libraries were generally designed based on a combination of low numbers of off-target sites and proximity to the TSS (Gilbert et al., 2014; Konermann et al., 2015). Target gene lists used RefSeq predictions or APPRIS annotation to pick canonical isoforms. Better

| Library Name | No. genes targeted | No. guides per gene | Targeting window (bp relative to TSS) | No. of guides (Total) |
|---|---|---|---|---|
| SAM v1 | 23,430 | 3 | -200 to 0 | 70,290 |
| CRISPRa | 15,977 | 10 | -400 to -50 | 198,810 |
| CRISPRa-v2 | 18,916 | 5 or 10 | -550 to -25 | 209,080 |
| Calabrese | >18,000 | 3 or 6 | -150 to -75 | 113,238 |

**Table 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.

predictions of TSSs using FANTOM5 datasets were used in subsequent libraries, along with more sophisticated algorithms which take into account nucleosome positioning (Horlbeck et al., 2016). The most recent genome-wide CRISPRa library was designed following the same principles but selected guides from a much narrower window of 75-150 bp upstream of the TSS (Sanson et al., 2018).

In summary, CRISPRa is a promising alternative to cDNA overexpression for gain-of-function screening and has already been applied to investigate genetic factors underlying cell growth as well as resistance to toxins and drugs. Initial screens show no evidence of inherent toxicity associated with CRISPRa in human cancer cell lines and the identification of known factors suggest that the CRISPRa platform is active and specific. CRISPRa enables pooled screening without the need for barcoding individual constructs as short 20 nt guide sequences double up as barcodes which can be easily retrieved and counted using next generation sequencing (NGS). In addition, cDNA libraries are often limited by a maximum insert size, resulting in the omission of genes with long transcripts (> 3-4 kb) and isoform specification may be an issue for proteins that are not well studied. CRISPRa circumvents these issues by activating transcription from endogenous promoters, enabling upregulation regardless of transcript length and allowing cellular machinery to capture the full diversity of isoform expression. However, potential drawbacks of CRISPRa include limits on the levels of endogenous overexpression compared to cDNA overexpression, which is usually driven by a highly active CMV promoter. In addition, promoter regions of transcriptionally inactive genes may be packed away in chromatin and thus difficult for the CRISPRa complex to access.

Nonetheless, the advantages of CRISPRa screening merit its use alongside cDNA libraries for large-scale gain-of-function screening. This is supported by a side-by-side comparison of CRISPRa and cDNA overexpression screening that showed that both platforms yielded many common hits as well as distinct and complementary hits (Sanson et al., 2018).

## 1.3 Aims and objectives

As detailed above, extracellular interactions are challenging to identify and as a result are underrepresented in large protein interaction datasets. Nonetheless, the identification of key interactions governing important biological processes is of scientific and clinical interest. Current methods for large-scale screening of cell surface receptors to one or a few defined ligands are restricted to investigating certain classes of membrane proteins or require large investments in terms of cost and equipment. In particular, none of these methods enables genome-scale interrogation of all membrane proteins encoded in the human genome. For instance, the largest plate-based recombinant protein screen tested pairwise interactions of 249 proteins (Martin et al., 2010), whilst the largest available membrane protein cDNA library contains clones encoding 4,493 membrane proteins (Mullican et al., 2017), or an estimated 75% of the human surfaceome. Consequently, the development of more cost-effective and comprehensive approaches for large-scale extracellular interaction screening would facilitate the discovery of novel receptor-ligand pairs.

Recently, CRISPR/Cas9 technologies have provided a highly adaptable platform for genome-scale forward genetic screening. In fact, whole genome CRISPR knockout screening has been successfully applied to elucidate pathways required for cell surface signalling and detect novel extracellular interactions (Sharma et al., 2018). This strategy necessitates first screening of ligands against a panel of cell lines to identify a cell line which exhibits ligand-binding properties. In addition, although CRISPR knockout screening is designed to target all genes encoded in the human genome, it is in practice restricted to genes that are expressed in the particular cell line being used. CRISPR activation (CRISPRa), at least in principle, provides an attractive approach for screening against virtually all cell surface proteins in the human genome with a single cell line.

Therefore, the aim of this thesis was to adapt CRISPRa screening for extracellular receptor-ligand detection by establishing the best parameters to upregulate cell surface receptors using CRISPRa and constructing a CRISPRa gRNA library targeting membrane

proteins. To validate this approach I also apply it to screen for known antibody targets and endogenous interaction partners. Finally, to demonstrate the utility of the CRISPRa approach, I screened several members of the adhesion GPCR family and identified a set of novel interactions between Brain angiogenesis inhibitor 1 (ADGRB1) and three closely related myelin-associated inhibitory proteins (RTN4R, RTN4RL1 and RTN4RL2).

# 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

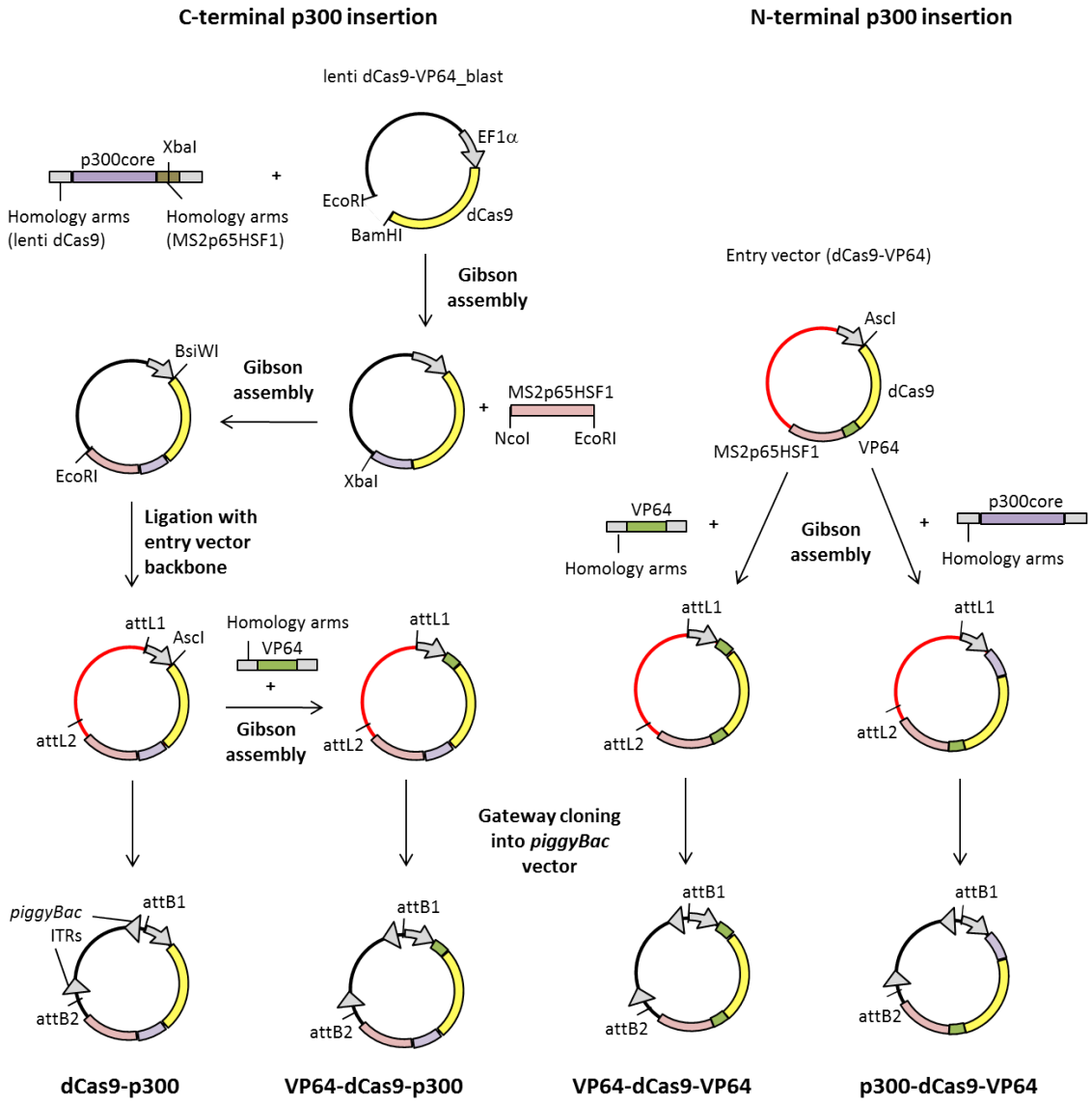



**Figure 2.1 Construction of dCas9-fusions containing additional p300core or VP64 domains** Separate strategies were used for N-terminal and C-terminal insertion of p300core or VP64 fragments. A lentiviral dCas9-VP64 vector from the Zhang lab was modified to include a C-terminal p300core fusion and expression 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 $\mu$L reaction volumes with 5 $\mu$g DNA, 5 $\mu$L 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 $\mu$L reaction volumes with 12.5 $\mu$L 2x Q5 Hotstart Hifi Master Mix (NEB), 1 $\mu$L each 10 mM sense and antisense primers, 1$\mu$L (1 $\mu$g) template DNA and 9.5 $\mu$L 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 $\mu$g 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 $\mu$L of diluted cDNA in a final reaction volume of 15 $\mu$L . 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 $\mu$M, 12.5 $\mu$L Q5 2x High Fidelity Hot-start Master Mix (NEB), and nuclease-free water to a final volume of 25 $\mu$L. 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 $\mu$g/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

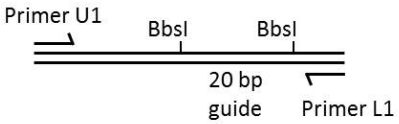



**Figure 2.2** The CRISPRa gRNA library was synthesised as a complex pool of 77 nt oligonucleotides with asymmetric sequences flanking two BbsI restriction sites. This generates three fragments of unequal lengths after digestion for size separation of guide sequences for cloning into the final lentiviral 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 ~1x$10^9$ 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 $\mu$g/mL G418 (Invitrogen) and 0.1% Kolliphor, but after single-cell cloning were maintained in Freestyle media supplemented with 50 $\mu$g/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 $\mu$g/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 $\mu$g of transfer plasmid, 9 $\mu$g ViraPower lentivirus packaging vectors (Invitrogen) using 36 $\mu$L Lipofectamine LTX and 12 $\mu$L 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 $\mu$L of 1 $\mu$g/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 $\mu$L of 0.1 $\mu$g/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

1x$10^5$ cells were washed 1x in PBS and 1x in binding buffer for Annexin V staining (Invitrogen), before being resuspended in 100 $\mu$L binding buffer. 5 $\mu$L Annexin V-FITC (eBioscience) was added to 100 $\mu$L 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 $\mu$L binding buffer for analysis. 5 $\mu$L 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 $\beta$-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 $\mu$g/mL G418 (Invitrogen) and 0.1% Kolliphor. Transfections were left for 5 days and supernatants were harvested and filtered through a 0.2 $\mu$m 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 $\mu$L 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 $\mu$L 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 $\mu$L 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 $\mu$L 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 $\beta$-lactamase enzymatic activity. Serial dilutions of prey supernatants were made in PBS-1% BSA and 20 $\mu$L of each dilution incubated with 60 $\mu$L of 125 $\mu$g/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 $\mu$L of 125 $\mu$g/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 $\mu$g/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 $\mu$g 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 $\mu$L of 1 $\mu$g/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 $\mu$g/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 $\mu$g/mL Proteinase K for 50 min at 55 °C followed by 10 min at 95 °C for inactivation. 10 $\mu$L of treated lysate was used as template for each 50 $\mu$L 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 $\mu$g gDNA was used as template for each 50 $\mu$L 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 $\mu$L 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.

# CHAPTER 3

# CRISPR ACTIVATION ENABLES RAPID AND STABLE OVEREXPRESSION OF CELL SURFACE PROTEINS

## 3.1 Introduction

CRISPR activation (CRISPRa) is a technique for eliciting targeted endogenous transactivation of genes using a synthetic, programmable transcription factor. CRISPRa systems are modified versions of the CRISPR/Cas9 adaptive immune system which mediate RNA-guided recognition of foreign genetic elements in bacteria and archaea. In mammalian cells, expressing a nuclease-inactive Cas9 (dCas9) fused to general transcriptional activator domains along with gRNA molecules containing a 20 nt guide sequence complementary to a target promoter region in the genome leads to increased transcription of the corresponding gene (Chavez et al., 2015; Gilbert et al., 2013; Konermann et al., 2015). However, the magnitude of upregulation varies widely between genes, from anywhere between less than 10-fold increase in mRNA abundance up to 10,000-fold increases for certain genes such as IL1RN. In addition, few studies have directly investigated the effect of CRISPRa on protein abundance, although phenotypic assays provide indirect evidence that changes in transcript abundance are generally reflected on the protein level.

For extracellular interaction screening, it is not only essential to ensure high levels of protein overexpression but also that overexpressed receptors are properly transported and presented on the cell surface. Plasma membrane proteins are generally synthesised in the endoplasmic reticulum (ER) where they are inserted into the lipid bilayer and trafficked through the general secretory pathway to the cell surface. However, some receptors that perform specialised functions may require chaperones only expressed in the relevant cell type for presentation at the surface. In addition, receptors which exhibit restricted tissue expression may be epigenetically silenced in other cell types and therefore inaccessible to transcriptional machinery.

In this chapter, I optimised the parameters for overexpressing plasma membrane proteins using CRISPRa and constructed a CRISPRa gRNA library targeting all membrane protein coding genes in the human genome. To do so I investigated the efficiency of different dCas9-fusions to upregulate a panel of cell surface receptors in HEK293 cells and generated a cell line constitutively expressing dCas9-activators. Using this cell line I also investigated the stability of overexpression over time. Importantly, I evaluated CRISPRa efficiency based on the levels of target proteins expressed on the surface using specific monoclonal antibodies (mAbs) rather than using mRNA abundance as a proxy for protein expression. I show that for a small number of proteins, an increase in mRNA levels did not necessarily result in an increase in cell surface expression. Finally I designed and cloned a gRNA library targeting all membrane protein coding genes in the human genome for large-scale extracellular interaction screening.

## 3.2 Results

### 3.2.1 Rapid upregulation of cell surface proteins using CRISPRa

To investigate if CRISPR activation could upregulate plasma membrane proteins, I designed a pool of guides targeting two cell surface receptors CD2 and CD200, which are not normally expressed in HEK293 cells. gRNA expressing constructs were co-transfected into HEK293 cells along with a plasmid encoding dCas9-VP64 and synergistic activation domains p65 and HSF1 (henceforth referred to collectively as 'dCas-activators') (Figure 3.1A). Expression of CD2 and CD200 was detected on the cell surface 36 h post transfection using mAbs against either protein and a suitable phycoerythrin (PE)-conjugated secondary antibody (Figure 3.1B). As the gRNA expression vector encodes BFP as a fluorescent marker,

the percentage of PE+/BFP+ cells out of the total number of BFP+ cells was quantified. CD2 was not detected on cells which were transfected with guides against CD200 and vice versa, and neither protein was detected on cells expressing the respective gRNA and dCas9 without any activator domains (Figure 3.1C).

To determine whether activated surface proteins were expressed at a sufficient level to detect interactions using recombinant protein, I produced soluble recombinant probes containing the ectodomains of CD58 and rat Cd200r, which are known to bind to CD2 and CD200, respectively (Selvaraj et al., 1987; Wright et al., 2000). Recombinant CD58 bound to cells expressing CD2, as did recombinant rCd200r to cells expressing CD200 (Figure 3.1D). I also tested another known interaction (SEMA7A/PfMTRAP) but did not observe binding (Josefin Bartholdson et al., 2012).

### 3.2.2 An additional histone deacetylase domain reduces CRISPRa activity

Next, I tested the general activation efficiency of CRISPRa with a panel of cell surface receptors representing different classes of membrane proteins. I selected 12 receptors based on the availability of established mAbs, lack of expression in HEK293-6E cells and endogenous expression in different tissues (Table 3.1). The last consideration was an attempt to pick gene targets with a range of chromatin availability, which would be the case during large-scale screening. To re-create large-scale screening conditions, I designed multiple guides targeting each gene and transduced cells with pools of lentivirally packaged gRNA targeting at a low MOI to ensure that each cell received only a single gRNA. I then transfected gRNA-expressing cells with the dCas9-activator construct and measured surface expression of each receptor 48 h post transfection. I detected surface expression of eight out of twelve proteins based on a cut-off derived from transfection of cells with the non-activating construct (Figure 3.2A), indicating broad applicability of using CRISPRa to overexpress cell surface proteins.

As the four proteins that were intractable to upregulation included several erythrocyte-restricted proteins, I hypothesised that the lack of activation might be due to epigenetic silencing of the promoter regions, making them inaccessible to dCas9-activator complexes. Hence, I fused a histone acetyltransferase (HAT) domain from p300 to the dCas9-activator construct, either C-terminal or N-terminal of dCas9, creating three types of dCas9 fusion proteins: p300-dCas9-VP64, dCas9-p300, and VP64-dCas9-p300 (Figure 3.2B). In addition, I generated another dCas9-activator variant with an additional N-terminal VP64 domain

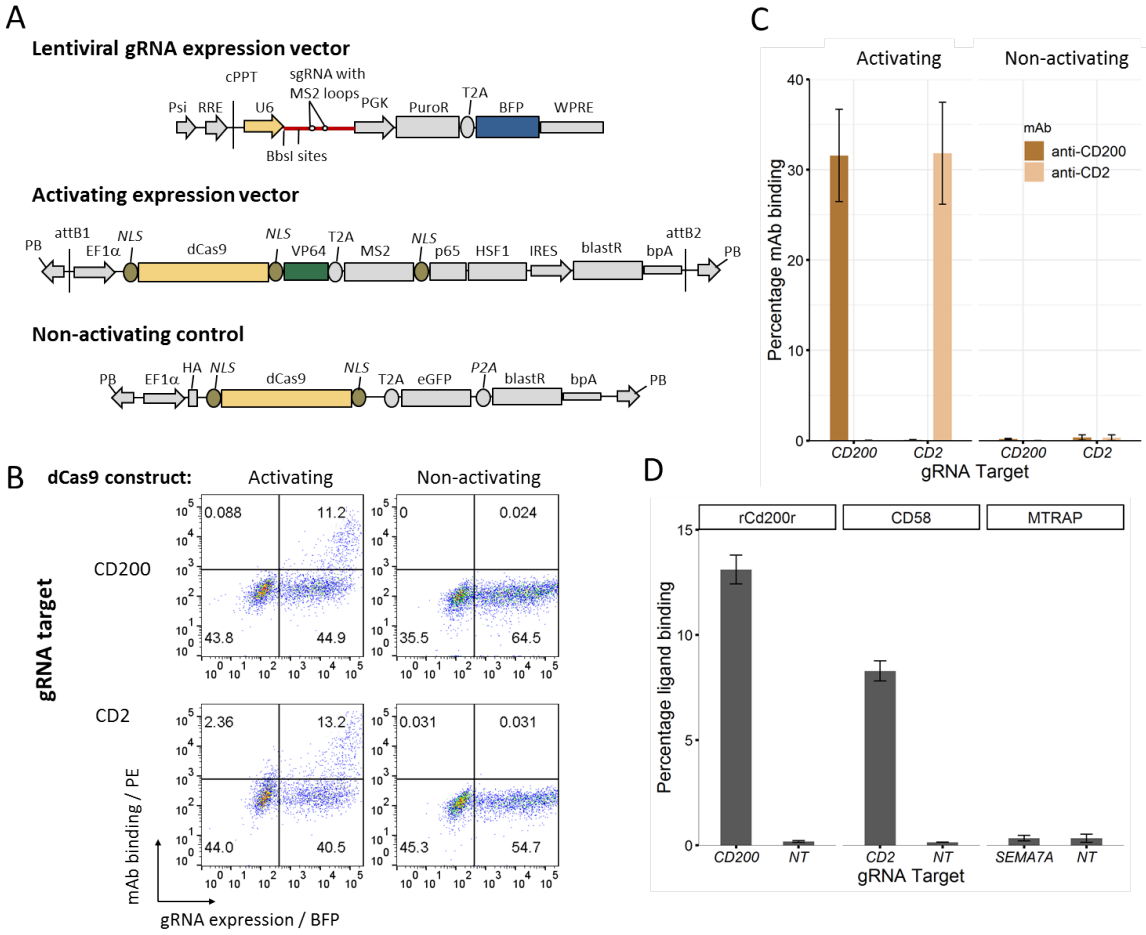



**Figure 3.1 CRISPR activation of CD200 and CD2 is specific and allows ligand binding.** A) Schematic of expression vectors used. gRNAs were expressed under a U6 promoter along with BFP as a fluorescent marker. dCas9-VP64 and MS2-p65-HSF1 fusion proteins were expressed as a single transcript separated by a T2A self-cleaving peptide. A non-activating vector expressing dCas9 without any transactivators was used as a negative control. PB - piggyBac inverted terminal repeats, attB1/B2 - $\lambda$ recombination attachment sites, Ef1$\alpha$ - Human elongation factor-1 $\alpha$ 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.

| Gene symbol | Protein type | Antibody clone | Expression in HEK293 | Endogenous expression |
|---|---|---|---|---|
| SEMA7A | GPI-anchored | MEM-150 | 0.407 | Ubiquitous |
| ENG | Single pass Type I | P3D1 | 0.055 | Endothelium |
| CD200 | Single pass Type I | OX-104 | 0.015 | Ubiquitous |
| P2RX7 | Multi-pass (2 TMD) | P2X7-L4 | 0.045 | Brain, immune tissues |
| CD2 | Single pass Type I | TS2/18.1.1 | 0.000 | T-lymphocytes |
| ICAM1 | Single pass Type I | P2A4 | 0.080 | Endothelium |
| PROM1 | Multi-pass (5 TMD) | HB#7/HC7 | 0.677 | Ubiquitous |
| VCAM1 | Single pass Type I | P3C4 | 0.000 | Myeloid cells |
| SELE | Single pass Type I | 1.2B6 | 0.000 | Activated endothelium |
| KEL | Single pass Type II | BRIC18 | 0.000 | Erythrocytes |
| SLC4A1 | Multi-pass (12 TMD) | BRAC18 | 0.018 | Erythrocytes |
| RHD | Multi-pass (11 TMD) | BRAD2 | 0.058 | Erythrocytes |

**Table 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.

(VP64-dCas9-VP64), as simply increasing the number of VP64 domains has increased activation efficiency in previous publications (Chakraborty et al., 2014; Tanenbaum et al., 2014). Unfortunately, addition of the p300 HAT domain did not induce expression of proteins that could not be upregulated by the original dCas9-activator construct, and even showed a decrease in activation efficiency for proteins that could be upregulated (Figure 3.2C).

### 3.2.3 Increased mRNA abundance does not necessarily lead to increased surface expression

Strikingly, none of the five dCas9-activator constructs were able to induce expression of SLC4A1, RHD, KEL or SELE. To investigate whether this was due to a lack of transcriptional activation, I measured the relative mRNA levels of the corresponding genes in cells transduced with targeting gRNAs compared to baseline levels in untransduced cells. Surprisingly, I found that the relative abundances of *SLC4A1*, *RHD*, and *SELE* mRNA showed significant increases in cells transduced with targeting gRNAs versus untransduced cells, with more than 1000-fold increase for *SLC4A1* (Figure 3.3A). To rule out the possibility that the antibodies I was using were not functional, I showed that antibodies against SLC4A1, RHD and KEL were able to bind human erythrocytes (Figure 3.3B), whilst anti-SELE was able to specifically recognise recombinantly produced E-selectin (Figure 3.3C). This indicates that the CRISPRa system is able to induce transcription of these genes, and that the lack of surface expression in these cases is likely to be attributed to a post-transcriptional process, for instance, the lack of specific chaperones for trafficking to the surface or other cell specific contextual effects.

### 3.2.4 Constitutive expression of dCas-activators enables stable overexpression of cell surface proteins

To reduce the variation in dCas9-activator expression between cells during large-scale screening, I generated a clonal cell line constitutively expressing dCas9-activators. I selected the two strongest activator constructs, dCas9-VP64 and VP64-dCas9-VP64, and transfected them into HEK293 cells along with PiggyBac transposase to allow integration into the genome. Transfections with the VP64-dCas9-VP64 construct yielded a blasticidin-resistant population of cells after five days of selection, whilst the dCas9-VP64 construct did not result in stably integrated, blasticidin resistant clones even after several transfections (data not shown). To generate clonal cell lines I isolated single cells from the blasticidin-resistant

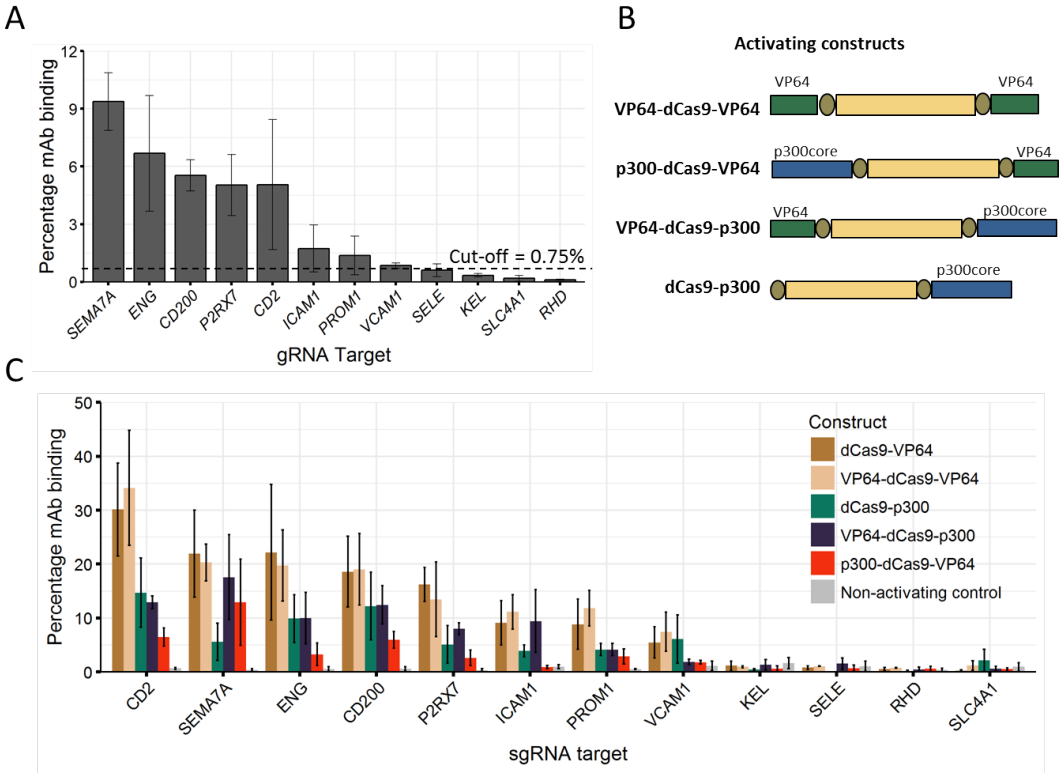



**Figure 3.2 Reduced CRISPRa efficiency using dCas9-activators with an additional p300 HAT domain.** A) Quantification of surface expression of 12 cell surface proteins after CRISPRa. Cells were transduced with lentiviruses carrying a pools of 8 gRNAs targeting each gene, before being transfected with 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 addtional 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.

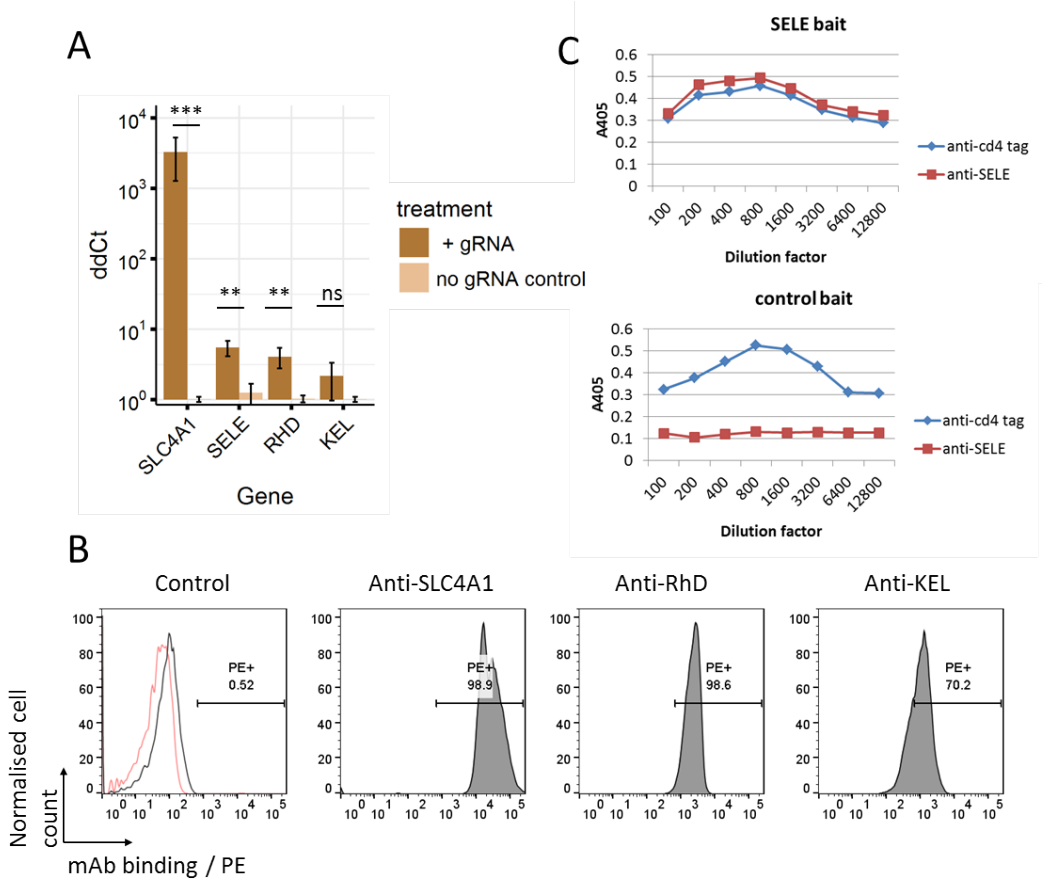



**Figure 3.3 Upregulation of mRNA transcripts does not necessarily lead to with an increase in surface protein levels.** A) Quantification of mRNA abundance by qRT-PCR of indicated target genes in cells 48 h post co-transfection with dCas9-VP64 and either targeting gRNA (+) or no gRNA control. Transcript abundance was normalised to *CYPA* expression; bars represent mean ± 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 3$^{rd}$ and 4$^{th}$ 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.

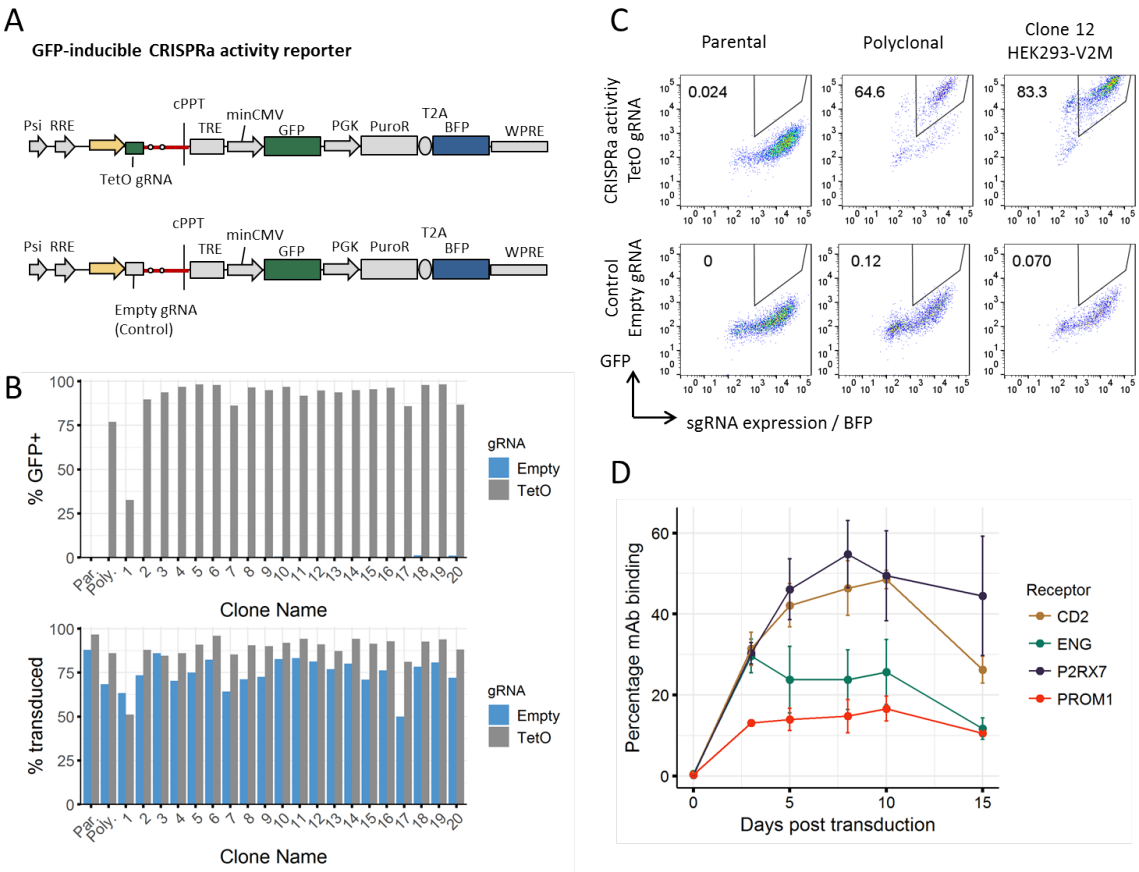



**Figure 3.4 A cell line constitutively expressing dCas9-activators allows sustained CRISPRa overexpression of cell surface proteins** A) Schematic of GFP-inducible CRISPRa activity reporter system. In both vectors, GFP expression is controlled by a TetO responsive element (TRE) consisting of several 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 tranduced 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 ± s.e.m; *n*= 3.

VP64-dCas9-VP64 line in media supplemented with 1% FBS. Isolation of single cells in unsupplemented Freestyle media did not yield any colonies. 20 colonies were picked and all but one showed higher levels of CRISPRa activity than the polyclonal line when transduced with a GFP-inducible reporter construct (Figure 3.4A,B). This is shown in detail for Clone 12, where more homogenous overexpression of GFP is observed within gRNA-expressing BFP+ cells as compared to the polyclonal population (Figure 3.4C) making the clonal line favourable for use in large-scale screening. This clonal cell line is subsequently named HEK293-V2M.

Next, I investigated the kinetics of receptor upregulation using CRISPRa to determine when receptor overexpression reaches its peak. To do so I measured the cell surface abundance of four receptors over two weeks after induction using CRISPRa. For all four proteins, I observed a rapid increase in surface expression between days one to three. This plateaued between day five and ten, with CD2 and ENG showing decreased levels of expression at day 15 (Figure 3.4D). Taken together, these results suggest that CRISPRa induces rapid and stable overexpression of surface receptors, making it a feasible strategy of overexpressing surface proteins for interaction screening.

### 3.2.5 A CRISPRa gRNA library targeting all human membrane proteins

For genome-wide extracellular screening, I designed and cloned a gRNA library targeting all membrane proteins encoded in the human genome (Figure 3.5). Using information from a public database of protein localisation, along with bioinformatic and mass spectrometric studies of the cell surface proteome (Bausch-Fluck et al., 2015; da Cunha et al., 2009; Thul and Lindskog, 2018), I compiled a list of genes encoding at least one predicted membrane-associated protein isoform. The criteria for inclusion was deliberately lenient so as to include as many candidate receptors as possible, and the final list contained 6, 213 genes. For each gene, 20 nt gRNA were designed to target within 400 bp of at most two unique Transcriptional Start Sites (TSS), with most TSSs being targeted by seven gRNAs. The final library contained 58,571 guides and was synthesised as a complex oligonucleotide pool.

Cloning guide sequences into the expression vector in a pooled fashion creates a bottleneck where specific guides may be lost due to random chance, or preferentially amplified. To reduce the chances of losing guides, I determined the coverage of the transformed library by plating out a small proportion of the transformation reaction on agarose plates and

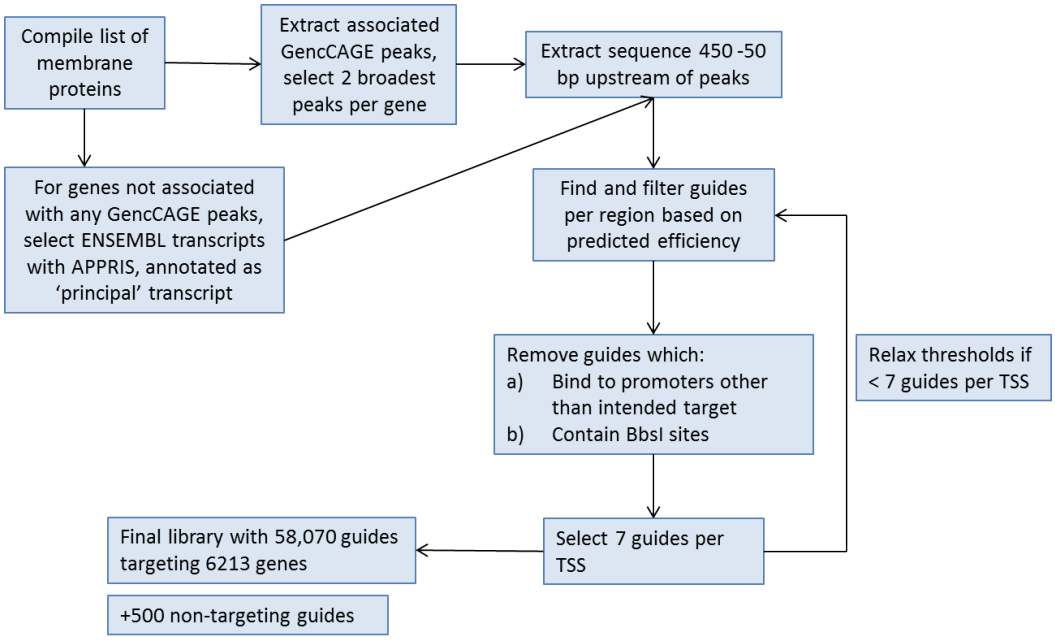



**Figure 3.5 Schematic showing workflow of membrane protein gRNA library design** A list of putative membrane proteins were compiled from public databases, mass spectrometry and bioinformatic studies. Unique TSSs were extracted based on a combination of Gencode gene models and CAGE-seq peaks (GencCAGE peaks); a maximum of 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.

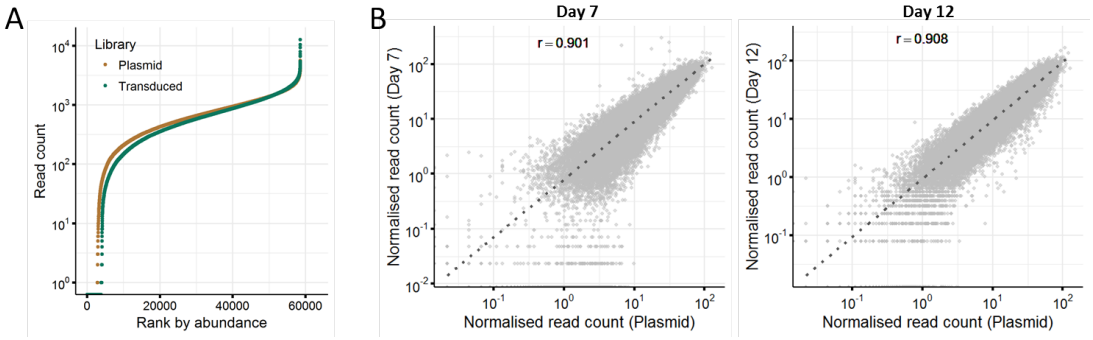



**Figure 3.6 Quality controls indicate complete and fairly even representation of guide sequences in the gRNA library** A) Ranked gRNA abundance in the plasmid library (brown) and cells transduced with the CRISPRa lentiviral library and cultured for seven days (green) as determined by raw read counts 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.

counting the number of colonies formed. I obtained an estimated 529, 875 colony forming units (9.05x coverage). Next, I measured the baseline distribution of gRNA using next generation sequencing (NGS) to take into account any dropouts or preferentially amplified gRNAs. Deep sequencing of the plasmid library detected 55,800 or 95.2% of guides designed, with 1.82% of reads mapping to the empty gRNA expression vector. In addition, I observed a fairly even distribution with 89% of guides having read counts within two orders of magnitude (Figure 3.6A). In cells that had been transduced with lentivirally packaged gRNA library the number of dropouts were slightly higher (3,888 rather than 2,770) and gRNA distribution was slightly more skewed. Libraries from cells after seven and 12 days of culture showed a high correlation with the original plasmid library (Figure 3.6B). Surprisingly, I did not observe systematic depletion of guides targeting particular genes after seven or 12 days of culture (data not shown), as would be expected with CRISPR knockout libraries where the knockout of essential genes reduces cell viability.

To determine the activity of the library I performed a small-scale validation experiment using 34 guides targeting four proteins. Each guide was individually cloned and transduced into HEK293-V2M cells before being assayed for cell surface expression of the target protein after 48 h. I found that 22 out of 34 guides (64.7%) induced target upregulation relative to a non-targeting gRNA control (Figure 3.7A). This included seven gRNAs targeting CD55, a receptor already highly expressed on HEK293 cells, indicating that CRISPRa can

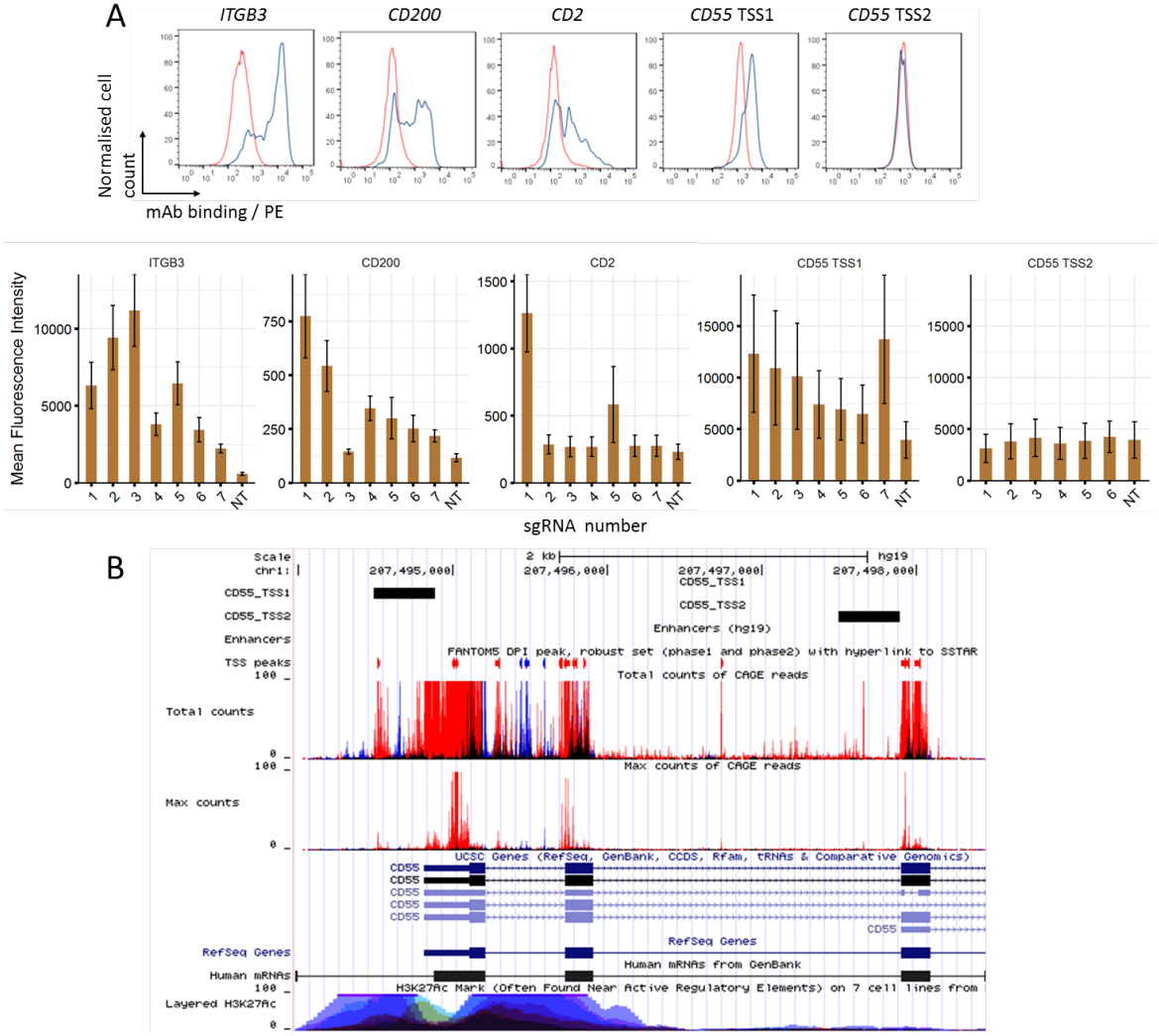



**Figure 3.7 Small scale validation of individual guides show that majority of guides are active** A) (Top) mAb binding histograms of HEK293-V2M cells transfected with the individual gRNAs targeting the promoter region of the named receptor genes (blue traces) compared to control non-targeting gRNAs (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.

further increase levels of cell surface proteins. *CD55* has two predicted TSSs, one of which produces a shorter transcript starting at the third exon of the canonical isoform which is unlikely to contain a signal peptide for trafficking to the surface (Figure 3.7B). Unsurprisingly, gRNAs targeting this TSS 2 did not result in any visible increase in surface expression of CD55.

## 3.3 Discussion

In this chapter, I demonstrate that CRISPRa can be used to overexpress cell surface receptors, with detectable increases in surface expression as early as 36 h after transfection. In addition, the two upregulated proteins, CD2 and CD200, are able to bind soluble ectodomains of their ligands, CD58 and rCd200r. This suggests that it is feasible to use soluble recombinant ectodomains to isolate a cell population expressing a receptor of interest after upregulation with CRISPRa.

Using a panel of 12 proteins I show that a broad range of cell surface receptors, including single pass, multi-pass, and GPI-anchored proteins, can be overexpressed using CRISPRa. Out of the four proteins which could not be upregulated, three (SLC4A1, RHD and KEL) are highly restricted to erythrocyte membranes whilst E-selectin is usually only expressed in endothelial cells after being induced by cytokines. ChIP-seq data from 7 cell lines (ENCODE, Broad Institute) showed reduced acetylation of histone 3 lysine residue 27 (H3K27ac) within the promoter regions of these four genes. H3K27ac is an epigenetic mark commonly found at the promoter regions of actively transcribed genes and its addition is catalysed by HATs like p300. However, fusion of the p300 HAT domain to dCas9-VP64 either had no effect or decreased activation efficiency. As a result, I chose to focus on fusions containing one or multiple copies of VP64.

Interestingly, I found that the inability to upregulate surface expression of SLC4A1 was not due to a corresponding inability to upregulate SLC4A1 at a transcriptional level. Instead, *SLC4A1* showed more than a 1000-fold increase in mRNA abundance when induced with CRISPRa. This indicates that post-translational processes might be responsible for the lack of surface expression. Consistent with this is the observation that SLC4A1 expression on the surface of erythrocytes is enhanced by co-expression of Glycophorin A (GYPA) (Young et al., 2000), which is not expressed in HEK293 cells. This suggests that CRISPRa

screening might not be able to detect interactions involving receptors that require specialised chaperones for transport to the surface.

For sustained expression of dCas9-activators, I generated a stable cell line using the VP64-Cas9-VP64 activator construct. Single cell cloning resulted in more homogenous CRISPRa activity as seen by an increase in the percentage of gRNA-expressing cells that have also upregulated GFP using a CRISPRa activity reporter construct. Homogenous expression of CRISPRa activity is essential for accurate estimations of library coverage during large-scale screening. Hence, I decided to use the single cell clonal lines for subsequent experiments.

Determining when surface receptors are most highly overexpressed is crucial to obtain the best separation between cells which have gained the ability to bind to a ligand of interest and cells that have not. Large cDNA libraries typically use expression vectors that support transient overexpression which increases over 48 h and peak between three to four days post transfection. By contrast, both dCas9-activator and gRNA expression constructs can be integrated into the genome and stably expressed under the appropriate antibiotic selection. I found that the peak of overexpression with CRISPRa is reached at day five post transduction and maintained up to day ten. This indicates that the best time to perform large-scale screening would be within that window.

To achieve genome-wide extracellular interaction screening, I designed and cloned a gRNA library targeting the promoters of 6,213 putative membrane proteins. Quality checks using deep sequencing indicated that the library was of sufficient quality for large scale screening as it did not contain a large percentage of dropouts or a significantly skewed guide distribution. In addition, small-scale validation of 34 individually cloned guides showed that 64.7% were working, providing some confidence that the automated design algorithm was able to select active gRNA. Of the guides that did not work, six were targeting an alternative TSS that was predicted to encode an isoform that was unlikely to be trafficked to the surface. This suggests that the fraction of active guides could be higher, at 78.6%.

In summary, I have determined some important parameters for overexpressing cell surface receptors using CRISPRa with the objective of large-scale extracellular interaction screening. This includes the selection of a dCas9-activator construct with high activity, generating a cell line with more homogenous CRISPRa activity, and determining when the peak of overexpression occurs. In addition, I constructed a CRISPRa gRNA library for upregulating all putative membrane proteins in the genome, and performed quality checks to ensure that it is suitable for use in large-scale screening.

# CHAPTER 4

# CRISPR ACTIVATION SCREENING DETECTS KNOWN EXTRACELLULAR INTERACTIONS

## 4.1 Introduction

The use of short 20 nt guide sequences for targeted gene overexpression makes CRISPR activation (CRISPRa) an attractive alternative to using cDNA-based methods for large-scale gain-of-function screening. As compared to expression cloning or libraries of defined cDNAs, large numbers (40,000-100,000) of guide sequences can be synthesised as a complex pool relatively cheaply, and allow targeting of any gene regardless of transcript length. Consequently, several groups have developed gRNA libraries targeting the promoter regions of all known genes, and performed genome-scale screening to identify genetic factors underlying cancer cell viability, drug resistance and antiviral response (Gilbert et al., 2014; Heaton et al., 2017; Konermann et al., 2015). Here, we aim to adapt CRISPRa screening for identifying extracellular interactions.

As with all new screening approaches, the CRISPRa platform needs to be benchmarked against a set of known interactions to determine its false-positive and false-negative rates. The interaction between bacterial proaerolysin and human glycosylphosphatidylinositol (GPI)-anchored proteins seemed like a useful test of the membrane protein gRNA library and CRISPRa screening approach because proaerolysin binding is thought to be independent of core protein sequence and GPI-anchored proteins are a well-annotated class

of proteins. GPI-anchors are post-translational modifications added to the C-terminus of many eukaryotic membrane proteins which facilitates attachment to the membrane bilayer (Paulick and Bertozzi, 2008). The human genome is thought to encode approximately 139 GPI-anchored proteins. Proaerolysin is the inactive precursor of the channel-forming bacterial toxin aerolysin secreted by *Aeromonas sp.* The monomeric form of proaerolysin binds a subset of GPI-anchored proteins and is then cleaved to form aerolysin by furin proteases. FLAER is a fluorescently labelled inactive variant of aerolysin (T253C/A300C) which binds to GPI-anchored proteins but does not get cleaved, allowing convenient labelling of GPI-anchored proteins on the cell surface (Brodsky et al., 2000).

Low-affinity interactions ($K_D$ >1 $\mu$M) are frequently understudied due to the challenges of detecting them in large-scale interaction screening (Wright, 2009). Nonetheless, such interactions can have important biological functions. The CD55-ADGRE5 interaction is an example of a low-affinity interaction ($K_D$ = 86 $\mu$M) that promotes T-cell proliferation upon antigen stimulation (Abbott et al., 2007). CD55 is a GPI-anchored protein expressed in haematopoietic and endothelial cells, whilst ADGRE5 is a seven-transmembrane G-protein coupled receptor (GPCR) expressed in leukocytes. ADGRE5 is also upregulated in some cancers. The inhibitory T-cell interaction between CTLA4-CD86 ($K_D$ = 2.6 $\mu$M) is another example of a low-affinity interaction. Cell surface CTLA4 is increased upon T-cell activation, where it competes with activating receptor CD28 for ligands CD80 and CD86 (Sansom, 2000). CD86 is expressed abundantly on professional antigen-presenting cells such as dendritic cells, monocytes and activated B cells which interact with T-cells. This interaction serves to tightly regulate T-cell activation. CTLA4 is also a promising target for cancer immunotherapy as it is upregulated on cancer cells as a strategy to evade immune attack (Contardi et al., 2005).

In this chapter, I benchmarked the membrane protein gRNA library and CRISPRa screening approach using several sets of probes. I established a screening workflow using monoclonal antibodies to highly activated cell surface targets, assessed the effect of sort thresholds and FDR cut-offs for determining screening 'hits', and demonstrated the ability of CRISPRa screening to detect medium to low-affinity endogenous interactions.

## 4.2 Results

### 4.2.1 Establishing a pooled CRISPRa extracellular interaction screening approach

Previously, I demonstrated that CRISPRa overexpression peaked between five and ten days post transduction. Based on these results, I designed a workflow for enrichment screening for receptor-ligand interactions using CRISPRa (Figure 4.1). HEK-V2M cells transduced with the lentiviral gRNA library were expanded up to nine days post-transduction. Cells were then incubated with a fluorescently labelled selection probe. After incubation, cells were washed to remove unbound probe, and a fraction of cells with the highest fluorescence intensities were isolated by cell sorting. Guide abundance within this population of cells was quantified by next-generation sequencing and enrichment analysis performed with MAGeCK (Li et al., 2014).

Given that guides in the gRNA library targeting CD200 and ITGB3 were highly active, I first performed CRISPRa screening using antibodies targeting CD200 and integrin $\alpha$v$\beta$3. I also screened the fluorescently conjugated secondary antibody used in all the screens, to determine if any surface proteins (such as Fc receptors) bound to the secondary antibody. The screens were performed in triplicate to investigate variation between replicates. A comparison of normalised read counts across all nine screens showed clear enrichment of gRNAs targeting ITGB3 in all replicates performed using the antibody against integrin $\alpha$v$\beta$3, but not in screens performed using only secondary antibody (anti-ms), or unsorted libraries (Figure 4.2A). CD200-targeting guides were simlarly enriched in screens performed with the anti-CD200 antibody, but also showed some level of enrichment in most other screens besides plasmid or unsorted libraries (Figure 4.2B). Gene level enrichment analysis showed no genes significantly enriched in any secondary-only screens, which was expected (data not shown). Accordingly, the same analysis identified ITGB3 as the most highly enriched gene in all three screens using anti-$\alpha$v$\beta$3, whilst the results of anti-CD200 screens were less reproducible, with CD200 being identified as the most highly enriched in only one out of three replicates (Table 4.1).

Comparison of gRNA abundance between the first and second anti-CD200 replicate showed a large number of depleted guides in the second replicate (Figure 4.2C), indicating insufficient coverage of the library during screening, resulting in the remaining guides appearing enriched just by chance and dilution of CD200-targeting guide enrichment. Insufficient

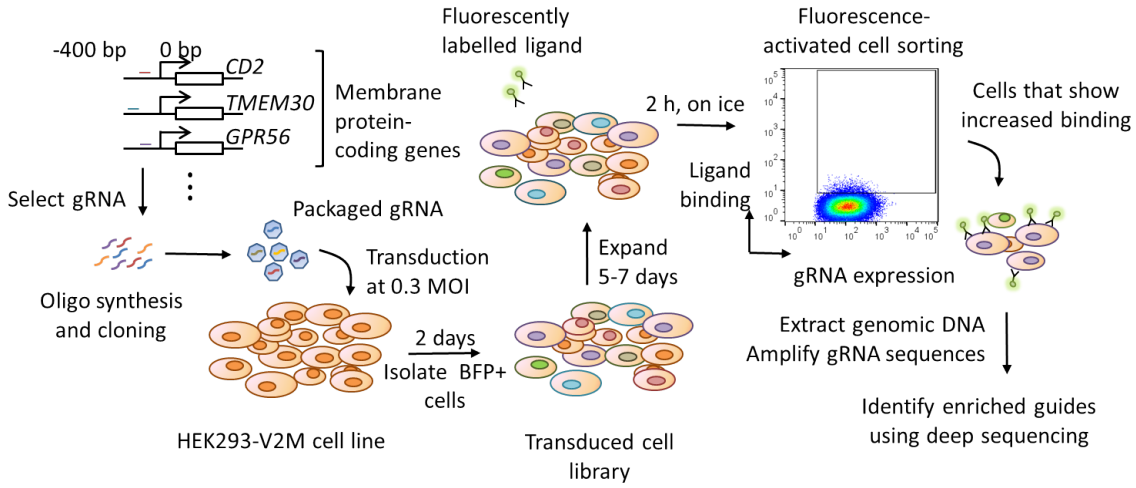



**Figure 4.1 Schematic of CRISPRa extracellular interaction screening.** A CRISPRa gRNA library targeting genes encoding membrane proteins was designed, cloned and packaged into lentiviruses for transduction. Transduction of a cell line constitutively expressing dCas9-activators at a low multiplicity 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 x $10^8$ 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.

| Reagent | Replicate | Target gene rank | Target FDR |
|---|---|---|---|
| Anti-$\alpha$v$\beta$3 | 1 | 1 | 0.00495 |
| | 2 | 1 | 0.00495 |
| | 3 | 1 | 0.00495 |
| Anti-CD200 | 1 | 1 | 0.00495 |
| | 2 | 604 | 0.936 |
| | 3 | 24 | 0.584 |

**Table 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 $> 0.05$).

coverage could be due to sorting line blockage during the sorting procedure resulting in fewer cells actually being collected than reported by the machine. To provide a buffer against unexpected cell loss, I doubled the number of cells screened from 5 x $10^7$ to 1 x $10^8$ and optimised my resuspension protocols to reduce the frequency of clumps or cell debris which might contribute to blockage during cell sorting. Screening with the improved protocol resulted in more robust enrichment of guides targeting CD200 and less variation of guide abundances overall (Figure 4.2D).

### 4.2.2 Less stringent sort threshold facilitates identification of GPI-anchored proteins using CRISPRa screening

To estimate the fraction of the membrane protein gRNA library that was active, I performed CRISPRa screening using fluorescently-labelled aerolysin (FLAER), which binds GPI-anchored proteins on the cell surface. In order to achieve a high signal-to-noise ratio, I initially selected a stringent sort threshold of 0.5% based on fluorescence intensity (Figure 4.3A), which had been used for previous antibody screens. However, sorting at this threshold returned very few significantly enriched genes (Table 4.2), although all six top-ranking genes were GPI-anchored (Table 4.2). As GPI-anchored proteins are highly expressed on HEK293 cells, I hypothesised that many guides targeting GPI-anchored proteins might have a small effect size, as seen previously with guides targeting CD55. The sort threshold of 0.5%

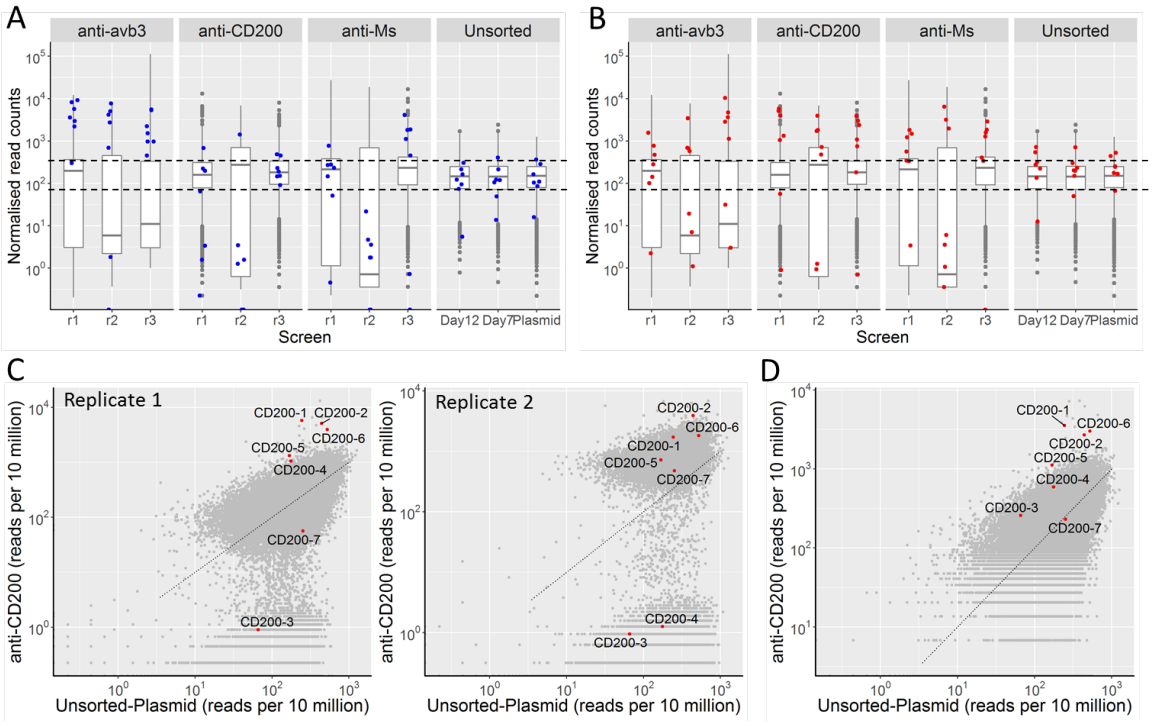



**Figure 4.2 Sufficient library coverage is required for robust receptor identification with CRISPRa screening.** A) ITGB3-targeting guides (blue dots) are enriched specifically in cell populations sorted for binding to anti-integrin $\alpha v\beta 3$. Box plots of normalised gRNA abundances are shown 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 no 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).

| Rank | Gene Symbol | FDR | LFC | GPI-linked |
|---|---|---|---|---|
| 1 | ULBP3 | 0.00165 | 6.1327 | Y |
| 2 | ULBP2 | 0.00165 | 5.8634 | Y |
| 3 | CD52 | 0.00165 | 5.5818 | Y |
| 4 | RTN4RL2 | 0.006188 | 1.6046 | Y |
| 5 | ENPP7 | 0.032673 | 0.57897 | Y |
| 6 | GFRA1 | 0.05198 | 1.4981 | Y |
| 7 | ANTXRL | 0.246975 | 1.7623 | |
| 8 | ALPPL2 | 0.246975 | 1.1281 | Y |
| 9 | PIGV | 0.246975 | 1.1261 | |

**Table 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.

might therefore have been too stringent to capture small increases in fluorescence, hence I performed the screen again at a higher threshold of 5% (Figure 4.3A). Sorting the top 5% of cells also increased the number of cells collected after sorting, resulting in higher coverage and therefore less variation in guide abundance, as can be seen from the increase in correlation of overall guide abundances with the plasmid library (Figure 4.3B).

Using the 5% sort threshold, I identified two to three times as many genes being enriched at similar false discovery rates (FDR), with twice as many genes annotated as being GPI-linked as compared to the first screen (Figure 4.3C, Table 4.3). This suggests that a sort threshold of 0.5% results in many false negatives that can be detected when using a less stringent sort threshold of 5%. To select an FDR cut-off for calling 'hits', I plotted the difference in number between true positives (GPI-anchored proteins; TP) and false positives (non-GPI-anchored proteins; FP) at different FDR cut-offs (Figure 4.3D). When going from an FDR cut-off of 0.05 to 0.1, the difference between true and false positives increases, indicating that the number of additional TPs detected at that cut-off outnumbers that of FPs. At higher FDR cut-offs, the difference either remains the same, indicating that the number of additional TPs equals that of additional FPs, or decreases drastically, suggesting that the number of additional FPs now outnumber TPs. Thus, I decided that an FDR of 0.1 represented a reasonable cut-off for calling 'hits'. Unfortunately, only 12 GPI-anchored proteins were identified at that cut-off, suggesting an extremely high false-negative rate of

| Rank | Gene Symbol | FDR | LFC | GPI-linked |
|---|---|---|---|---|
| 1 | ULBP3 | 0.000707 | 3.0472 | Y |
| 2 | CD52 | 0.000707 | 3.3022 | Y |
| 3 | ULBP2 | 0.000707 | 3.0274 | Y |
| 4 | GFRA1 | 0.000707 | 1.1937 | Y |
| 5 | RTN4RL2 | 0.000707 | 0.78178 | Y |
| 6 | ALPPL2 | 0.000707 | 1.7128 | Y |
| 7 | OR10A7 | 0.000707 | 1.1243 | |
| 8 | ULBP1 | 0.002475 | 0.56565 | Y |
| 9 | ENPP7 | 0.002475 | 0.64116 | Y |
| 10 | CD24 | 0.002475 | 0.13689 | Y |
| 11 | RAET1L | 0.005851 | 0.98166 | Y |
| 12 | ART3 | 0.006188 | 0.41892 | Y |
| 13 | PRND | 0.035415 | 1.4016 | Y |
| 14 | MAL | 0.036421 | 1.0057 | |
| 15 | OR10A4 | 0.053135 | 0.86797 | |
| 16 | CNTFR | 0.111696 | 0.18345 | Y |
| 17 | SPATA9 | 0.152994 | 0.81535 | |
| 18 | FOLR2 | 0.152994 | 0.33968 | Y |
| 19 | SPTSSA | 0.152994 | 0.86468 | |
| 20 | VKORC1L1 | 0.152994 | -0.06074 | |
| 21 | SGCG | 0.152994 | 0.14246 | |
| 22 | OR6P1 | 0.15414 | 0.55077 | |
| 23 | SLC28A3 | 0.171545 | 0.46685 | |
| 24 | GPR82 | 0.21019 | 0.79616 | |
| 25 | RPN2 | 0.222201 | -0.43292 | |
| 26 | GPC1 | 0.222201 | 0.49491 | Y |

**Table 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.

91.4%. However, increasing the FDR cut-off did not greatly reduce the false negative rate (Figure 4.3E). In addition, this estimation may be inflated due to the high baseline expression of GPI-anchored proteins on HEK293 cells and the existence of GPI-anchored proteins that do not bind aerolysin. On the other hand, the 0.1 FDR cut-off results in a high positive predictive value of 80.0%, indicating that CRISPRa screening can identify interactions with high confidence. All subsequent screens were performed using the improved protocol and 5% sort threshold, and interaction 'hits' were called using an FDR cut-off of 0.1.

### 4.2.3 Multiplexing selection probes enables detection of multiple interactions using a single screen

The number of samples that can be screened with the CRISPRa approach is largely limited by the number of cells required for each selection probe (1 x $10^8$ cells) and sorting time (3 h per library and a further 3 h per probe, not including setting up and shutting down of the machine). In an attempt to further increase the throughput of this approach, I performed a screen using a mixture of eight antibodies targeting surface receptors that had previously showed upregulation with CRISPRa. This screen identified seven significantly enriched hits, six of which were the expected antibody targets and the last being WNT3 (Figure 4.4A). A comparison of individual gRNA abundance between guides targeting WNT3 (false positive, FP) and those targeting the expected antigens (true positives, TP) show that all of the WNT3-targeting guides were not as highly enriched in the sorted population as guides targeting the six antigens (Figure 4.4B). This suggests that WNT3 might be a false positive detected by the enrichment test algorithm only because all five guides were slightly enriched by chance. Another explanation could be antibody cross-reactivity, which is not infrequently observed (Michel et al., 2009). Replicate screens would be able to address whether the enrichment of WNT3-targeting guides was due to cross-reactivity or random chance. Surprisingly, guides targeting two antigens (PROM1 and P2RX7) were not enriched, despite showing upregulation with a previous set of guides (Chapter 3.2.1). However, it is important to note that the guides used in the membrane protein gRNA library were different from those used in previous experiments, and it is possible that during gRNA design inefficient guides were selected for these two targets. The detection of multiple interactions from a complex pool of antibodies has implications for broader applications of the CRISPRa approach beyond interaction screening using single, defined ligands. For instance, sera from patients with autoimmune disease could be used to determine cell surface factors responsible for Ig-mediated autoimmunity.

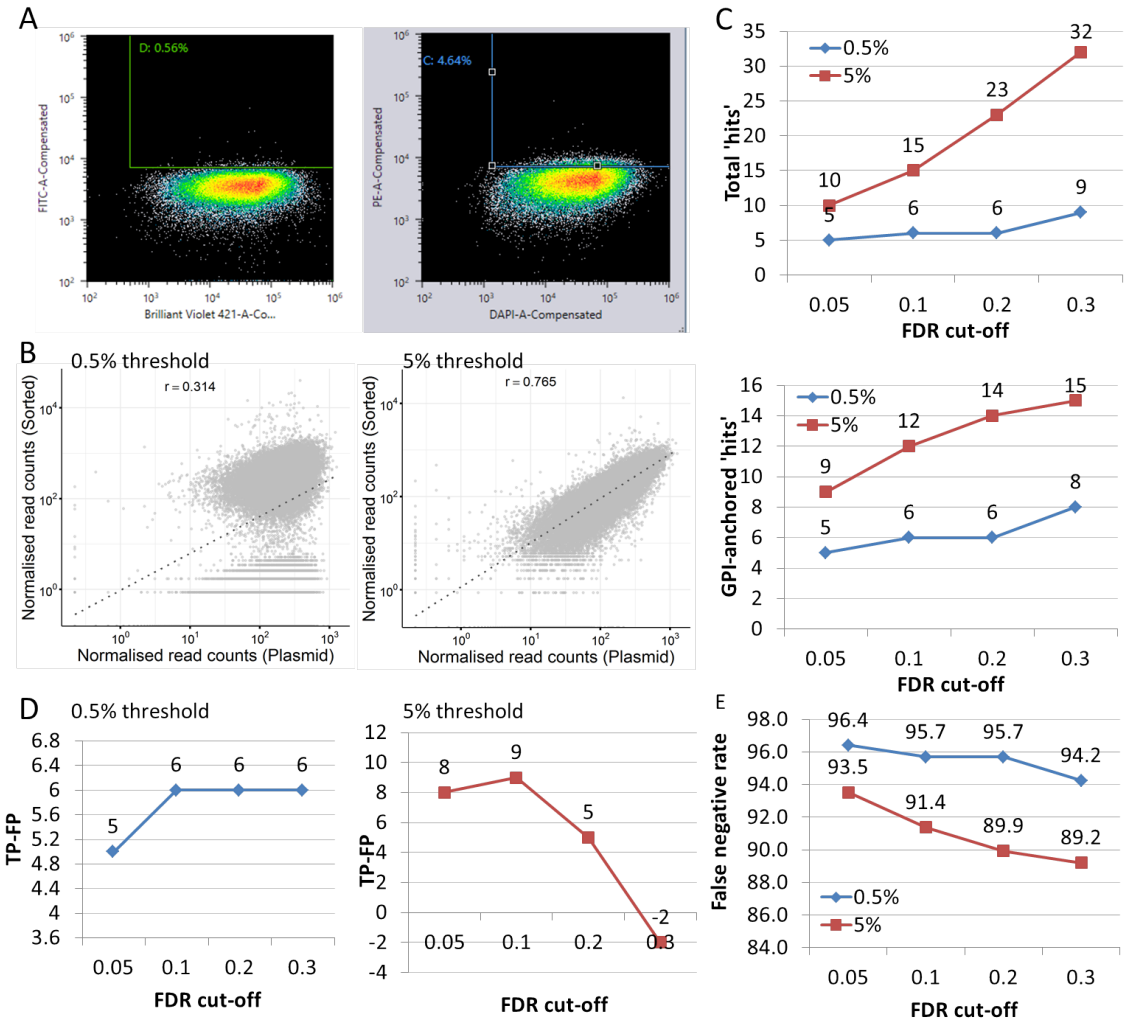



**Figure 4.3 A 5% sort threshold during CRISPRa screening reduces false negatives whilst an FDR cut-off of 0.1 limits false positives** A) Sort gates using during screening at a 0.5% (left) and 5% threshold (right). The percentage of cells in the gate fluctuates during sorting hence numbers in the image 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 correction (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).

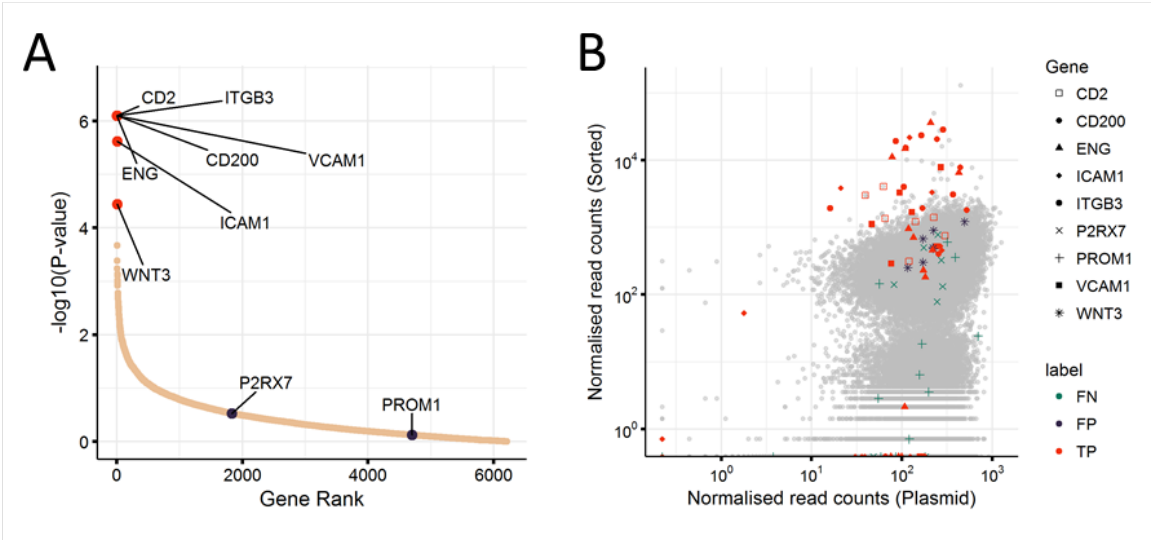



**Figure 4.4 CRISPRa screening simultaneously identifies multiple targets to a pool of antibodies.** A) Transduced cell libraries were sorted to isolate cells binding to an equimolar pool of eight antibodies, and gRNA abundance quantified by next-generation sequencing. Enrichment analysis indicated that 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).

### 4.2.4 CRISPRa screening detects low-affinity endogenous interactions

To determine the sensitivity of this CRISPRa approach for identifying low-affinity interactions, I produced a panel of soluble recombinant ectodomains of proteins with known cell surface binding partners. Recombinant ectodomains were produced as biotinylated monomers, and for increased avidity, were tetramerised around streptavidin molecules (Figure 4.5A). The streptavidin molecules were also conjugated to phycoerythrin (PE) for fluorescent detection during cell sorting. For each screen, the amount of recombinant protein was normalised to the amount needed to saturate binding of 2 $\mu$g fluorescently-labelled streptavidin. This was determined using competitive enzyme-linked immunosorbent assays (ELISAs) using streptavidin-coated plates to measure the amount of free biotinylated protein after incubation with 10 ng of PE-conjugated streptavidin (Figure 4.5B). The highest concentration of biotinylated monomers that resulted in no excess biotinylated protein was used.

Screening of the four recombinant protein tetramers (or 'baits') resulted in a total of six hits, all six of which were known endogenous binding partners out of nine previously reported interactions (Figure 4.6A). Importantly, CRISPRa screening detected the weakest interaction (CD55-ADGRE5, $K_D$ = 86 $\mu$M), in addition to higher affinity interactions such as EFNA1-EPHA2, CTLA4-CD80 and rCd200R-CD200 (Figure 4.6B). Interactions with known affinities are listed in Table 4.4. In general, screening results were replicable and each interaction was identified with a similar level of confidence in at least two replicates (Data not shown). Moreover, screening with EFNA1 identified multiple binding partners (EPHA2, EPHA4 and EPHA7), illustrating the utility of CRISPRa screening compared to loss-of-function or affinity-purification / mass spectrometry based approaches which would only identify binding partners expressed by the cell line being screened. Taken together, these results suggest that the improved protocol for pooled CRISPRa screening and an FDR cut-off of 0.1 is able to reliably and unambiguously identify endogenous interactions, even ones with micromolar affinities.

## 4.3 Discussion

In this chapter, I established a workflow for extracellular screening using CRISPRa. Using two antibodies to cell surface targets, I showed that CRISPRa screening can identify antibody targets and identified library coverage as an important parameter for reliably

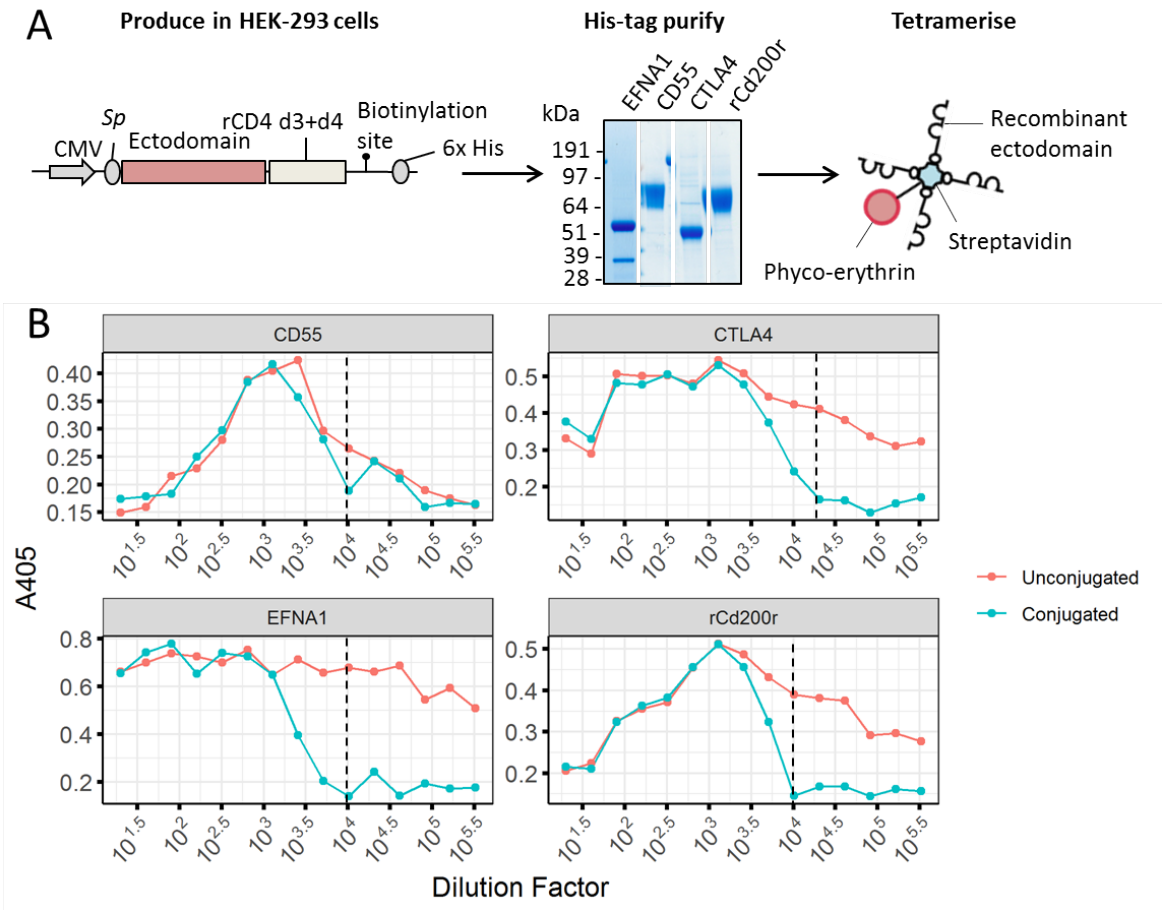



**Figure 4.5 Highly avid tetramers are produced from recombinant biotinylated ectodomains and normalised for use in CRISPRa screening** A) Schematic showing the production of tetramers from purified biotinylated monomers containing the full length ectodomain of four cell surface ligands. A construct 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 - 3$^{rd}$ and 4$^{th}$ Ig domains of rat CD4. B) The amount of recombinant protein used for screening is normalised using the amount needed to saturate 2 $\mu$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 $\mu$g.

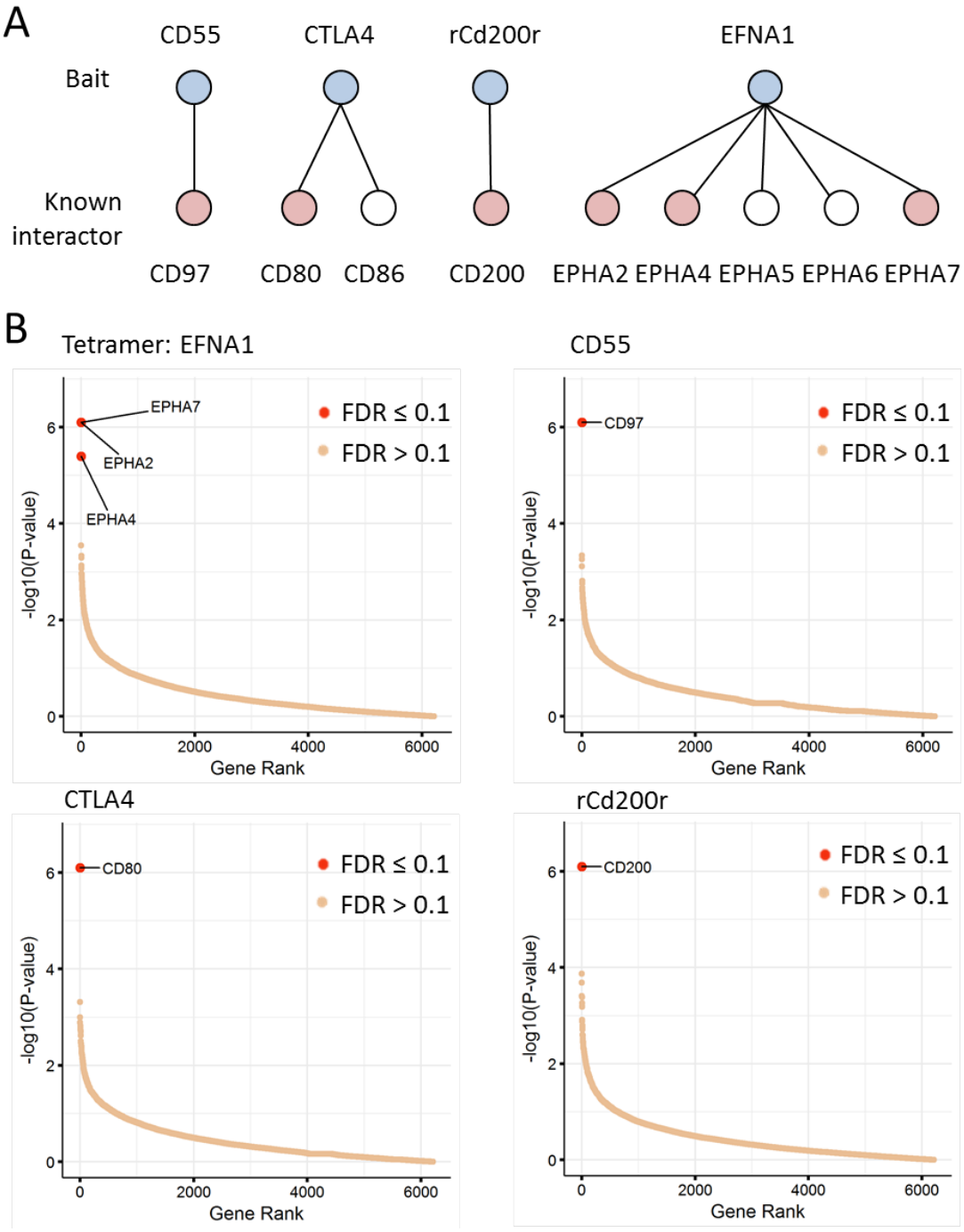



**Figure 4.6 CRISPRa screening unambiguously identifies low-affinity endogenous interactions** A) CRISPRa screening identifies six out of nine reported interactions involving CD55, CTLA4, rCd200 and EFNA1. Blue circles represent cell surface ligands used as tetramers for screening, pink circles represent 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.

| Interaction | $K_D$ ($\mu$M) | Reference | Enriched in screen |
|---|---|---|---|
| CD55 - CD97 | $86 \pm 1$ | Lin et al. (2001) | Y |
| CTLA4 - CD86 | 2.6 | Collins et al. (2002) | |
| Cd200r - CD200 | $0.59 \pm 0.07$ | Wright et al. (2003) | Y |
| EFNA1 - EPHA2 | $0.58 \pm 0.24$ | Lema Tomé et al. (2012) | Y |
| CTLA4 - CD80 | $0.42 \pm 0.06$ | van der Merwe et al. (1997) | Y |

**Table 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 $\mu$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.

detecting guide enrichment. Library coverage refers to the number of times a gRNA is represented, assuming a uniform distribution of gRNAs, and affects the variation in individual guide abundance that is due to random chance. Higher coverage leads to lower variation in guide abundance, particularly for lowly-represented guides, and therefore more confident estimations of guide enrichment. However, higher coverage represents a trade-off with screening practicalities as the number of cells needed for screening increases. I found that 1 x $10^8$ cells provided a feasible number of cells for screening whilst maintaining sufficient coverage to reliably detect guide enrichment.

Using proareolysin binding to GPI-anchored proteins, I determined a suitable FDR cut-off of 0.1 for calling 'hits' in future screens, based on the percentage of false positives detected at different FDRs. Using two thresholds for sorting, I also found that the less stringent threshold of 5% led to the detection of more GPI-anchored proteins under an FDR of 0.1. However, only 12 out of 139 or 8.63% of GPI-anchored proteins were detected. Although this suggests a very high false-negative rate, this estimate could be inflated due to the already high expression of GPI-anchored proteins on HEK293 cell surfaces and possible GPI-anchored proteins that do not bind aerolysin. In particular, a recent mass spectrometry study found that a number of GPI-anchored proteins did not bind aerolysin (Wuethrich et al., 2014). Additionally, GPI-anchors exhibit considerable variation in phospholipidinositol (PI) side chains such as inositol acylation, which refers to the presence of an ester-linked fatty acid attached to the C-2 hydroxyl of the inositol residue. This modification makes the anchor inherently resistant to the action of bacterial PI-specific Phosphoinositide phospholipase C (PLC), which recognises and cleaves GPI-anchors on mammalian cells (Ferguson et al.,

2017). Similarly, it is not unreasonable to expect that certain modifications to the GPI-anchor may render proteins refractory to staining with aerolysin. CD55 guides that showed activation when individually tested were not enriched in either aerolysin screen, despite CD55 being a known binder of proaerolysin, further suggesting that the results of these screens do not fully reflect the true fraction of working guides in the library.

In addition to detecting surface targets of single antibodies, I showed that CRISPRa screening can simultaneously identify multiple targets of a pool of antibodies. At a FDR cut-off of 0.1, I identified six out of eight antibodies targets along with an unexpected hit, WNT3. WNT3 is a member of the WNT family which is involved in oncogenesis, regulation of cell fate, and patterning during embryogenesis. It shows little sequence homology to any of the eight target receptors and guides targeting WNT3 are not as highly enriched as guides targeting the other six target receptors, suggesting that WNT3 might be a false positive. This highlights certain limitations of the gene enrichment algorithm in detecting true enrichment when all guides targeting a particular gene are only slightly enriched by chance. In such cases, reviewing individual gRNA abundance can be useful in determining how likely a hit is to be a true positive. Out of eight target receptors, two were not identified. Given that I have shown that these targets can be upregulated, and that the antibodies are able to detect them, the likely explanation for this result is that PROM1 and P2RX7-targeting gRNAs in the membrane protein library are ineffective at eliciting expression of the two proteins. This indicates one source of false negatives for CRISPRa screening. Nonetheless, detecting multiple interactions from a complex pool of antibodies has implications for broader potential applications of the CRISPRa approach beyond interaction screening using single, defined ligands. Blood serum contains a mixture of antibodies, which under certain circumstances may be self-reactive, causing a variety of inflammatory-related symptoms. In this regard, CRISPRa could potentially be used to identify cell surface autoantigens. Another possible application could be the characterisation of host cell surface factors interacting with secreted factors from bacteria or parasites.

Finally, I demonstrated that CRISPRa screening is able to identify endogenous interactions with high confidence. Importantly, I was able to detect the very weak CD55-ADGRE5 interaction, demonstrating the sensitivity of this approach. Screening a limited set of five proteins resulted in the detection of six out of nine previously reported interactions, suggesting a 30.3% false-negative rate. This rate is much lower than previously estimated with aerolysin/GPI-anchor protein screens, and is probably more accurate as it is not confounded by high baseline fluorescence levels or possible subsets of proteins that are refractory to staining. However, CRISPRa screening failed to detect low affinity CTLA4-CD86 interac-

tions. Aside from possible loss of the low-affinity interactors during multiple washing steps, this result could also be due to a number of other factors including ineffective gRNAs, TSS misprediction or targeting of alternative TSSs. For EFNA1, multiple binding partners from the same family were identified. This highlights the advantages of using a gain-of-function approach for extracellular interaction screening, as multiple interactors can be identified in a single screen regardless of their expression patterns in vivo.

In summary, I have adapted the CRISPRa screening platform for extracellular interaction detection and identified a few key paramters for interaction screening. This includes number of cells screened and sort threshold. In addition, I demonstrated that CRISPRa interaction screening is able to detect multiple binding partners to a pool of antibodies, as well as interactors to endogenous ligands with a low false-positive rate. The data in this chapter also highlights some of the possible mechanisms underlying false positives and false negatives from CRISPRa interaction screening and provides a suggestion for 'sanity checks' to perform in order to determine whether a hit is likely to be a false positive.

# CHAPTER 5

# CRISPR ACTIVATION SCREENING OF ADHESION GPCRS IDENTIFIES KNOWN AND NOVEL INTERACTIONS

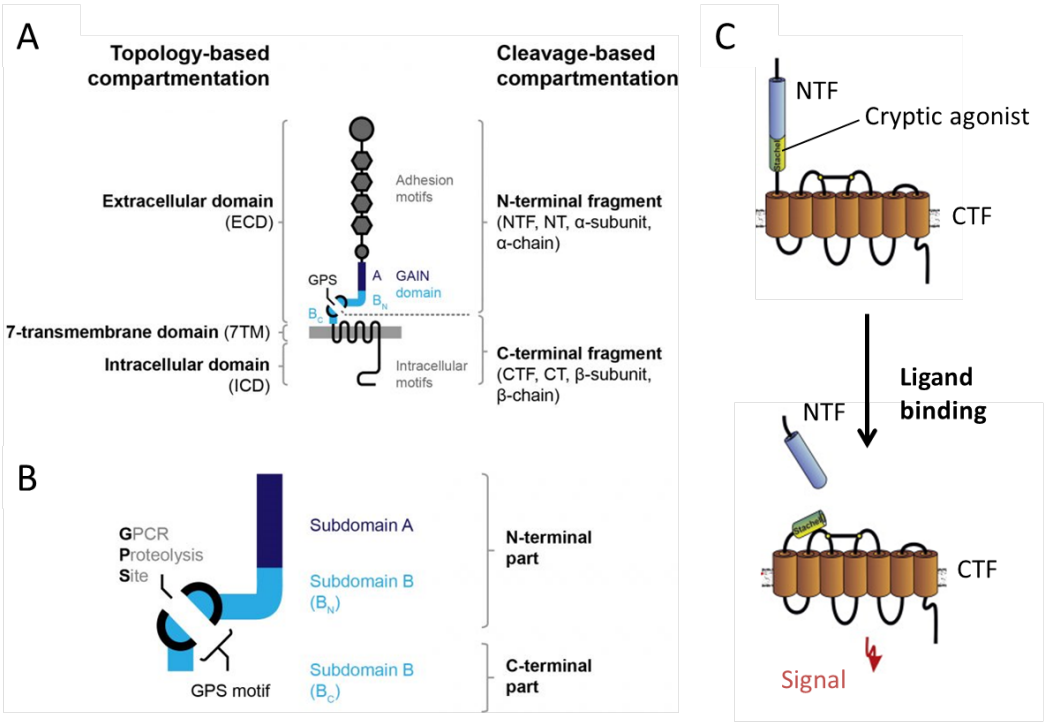
## 5.1 Introduction

The G-protein coupled receptor (GPCR) superfamily is one of the largest groups of transmembrane proteins, and is extensively targeted by therapeutic drugs. An estimated 34% of FDA-approved drugs target just 108 GPCRs (Hauser et al., 2018). GPCRs are characterised by seven transmembrane domains and cytosolic association with heterotrimeric G-proteins, which participate in a wide variety of downstream signalling cascades. However, many GPCRs do not have known endogenous ligands and as a result, there is great interest in understanding the biology of such GPCRs and identifying their ligands, a process called 'GPCR deorphanisation'. Adhesion GPCRs, a subfamily characterised by large N-terminal domains containing multiple adhesion-related motifs, contain the highest number of 'orphan' receptors which have no known endogenous ligand. Members of the adhesion GPCR family play important roles in immune regulation, central nervous system development, and angiogenesis (Bjarnadóttir et al., 2004). Knowledge of endogenous ligands would therefore shed some light on the molecular mechanisms of these receptors.

As previously mentioned, the large extracellular N-terminal regions of adhesion contain multiple protein domains involved in cell adhesion, as well as a conserved GPCR Proteolysis Site (GPS) located in the ectodomain almost adjacent to the first transmembrane domain (Figure 5.1A). During translation, the GPS is cleaved, forming a C-terminal fragment containing the seven transmembrane domains and an N-terminal fragment consisting most of the ectodomain (Figure 5.1B). Both fragments remain non-covalently attached during trafficking to the plasma membrane, and the prevailing theory of adhesion GPCR activation is that ligand binding to the N-terminal ectodomain causes a conformational change or complete dissociation of the N-terminal fragment revealing a cryptic tethered peptide agonist (Figure 5.1C) (Araç et al., 2012; Stoveken et al., 2015). As such, adhesion GPCRs are thought to bind cell surface or extracellular matrix (ECM) proteins to cause N-terminal fragment dissociation. This is supported by the fact that known binding partners include GPI-anchored, transmembrane, and fibrous proteins like collagen or laminin. The propensity for binding cell surface receptors along with large N-terminal ectodomains that can be produced in soluble recombinant form makes this family of proteins ideal for CRISPRa interaction screening.

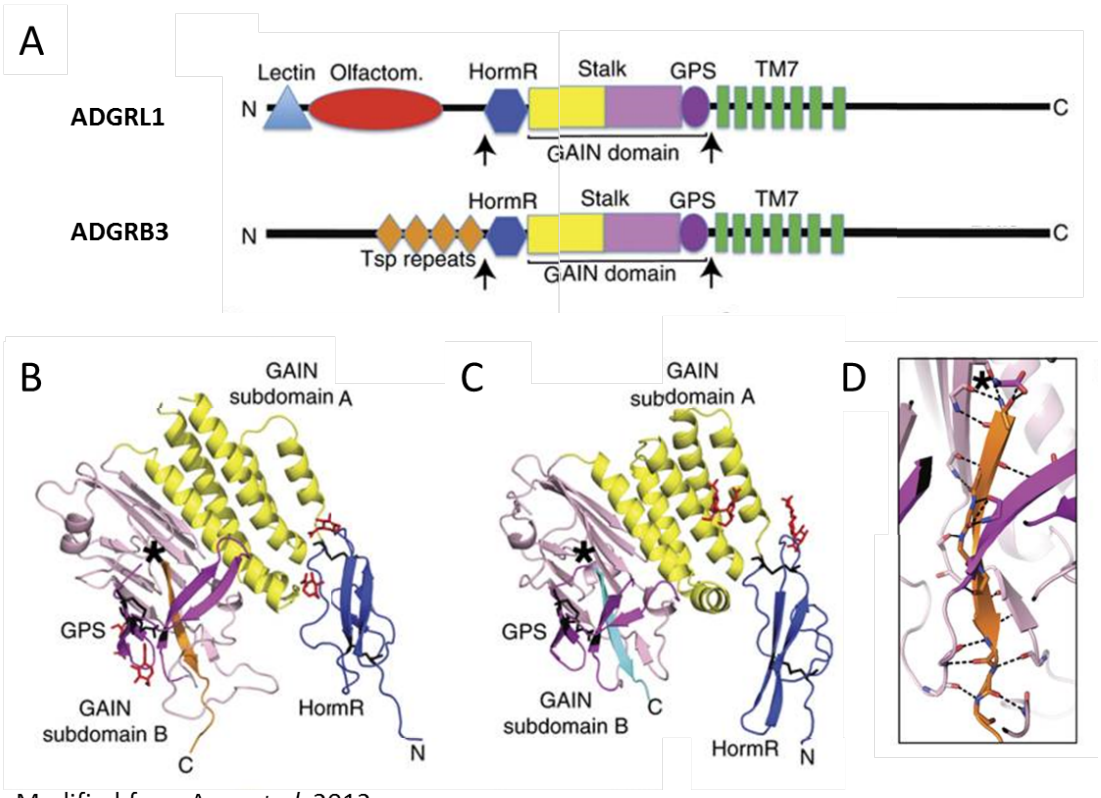
All 33 members of the adhesion GPCR family in humans harbour a conserved GPCR autoproteolysis-inducing (GAIN) domain which includes the GPS motif and is minimally required for proteolysis. The GPS motif consists of a conserved histidine, leucine and threonine or serine, with proteolysis occurring between the leucine and threonine/serine residues (H↓T/S). Structural studies indicate that the C-terminal region of the GAIN domain consists of a twisted $\beta$-sandwich including 13 $\beta$-strands and two small $\alpha$-helices (Figure 5.2A-C) (Araç et al., 2012). The GPS is located between the last two $\beta$-strands and cleavage results in the separation of the last $\beta$-strand, which is then kept in place by numerous hydrophobic interactions with the surrounding $\beta$-strands (Figure 5.2D). Mutagenesis of the GPS motif indicated that autoproteolysis is not required for surface transport, although certain mutations resulted in cytosolic retention of the receptor, possibly due to the steric changes affecting the structure of the GPS/GAIN domain (Araç et al., 2012).

In this chapter, I applied the CRISPRa screening approach to identify cell surface binding partners of adhesion GPCRs. To do so, I expressed the entire ectodomain of 13 adhesion GPCRs as soluble recombinant proteins by mutating the GPS site to prevent proteolysis. Screening identified several previously described interactions, as well as novel associations between brain angiogenesis inhibitor 1 (ADGRB1) and members of the myelin-associated Nogo receptor family. Using cDNA overexpression and Avidity-based extracellular interaction screening (AVEXIS), I confirmed that ADGRB1 binds all three




Modified from Langenhan *et al*, 2013 and Liebscher *et al*, 2014

**Figure 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.




Modified from Araç *et al*, 2012

**Figure 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 $\beta$ sheets and 2 $\alpha$ helices. D) After cleavage, NTF and CTF remain attached by numerous hydrogen bonds shown between the cleaved $\beta$-strand (orange) and the surrounding $\beta$-strands (purple) in ADGRL1. The cleavage site is indicated with a black star in B), C) and D)

members of the Nogo receptor family (RTN4R, RTN4RL1, RTN4RL2) and showed that the first three thrombospondin repeats on ADGRB1 are sufficient for binding.

## 5.2 Results

### 5.2.1 A T/S→G mutation at the GPS site enables high level of recombinant ectodomain production

To identify binding partners using CRISPRa screening, I sought to produce the ectodomains of members of the adhesion GPCR family in soluble, recombinant form. Using constructs encoding recombinant ADGRL4 and ADGRG1 that were already available in the lab, I transfected HEK293-6E cells and harvested culture supernatants six days post transfection. Western blotting of culture supernatant showed faint bands that did not correspond to the expected molecular weights, indicating a lack of proper protein expression (Figure 5.3A). These constructs contained the ectodomains of both receptors truncated at the GPS cleavage motif (Figure 5.3B). Given that the N- and C-terminal fragments of the receptors remain associated by numerous hydrophobic bonds (Figure 5.2D), I hypothesised that the GAIN domain might be unable to undergo proper folding without the last $\beta$-strand, which was not included in these constructs. Thus, I designed new constructs that consisted of the full length ectodomains of ADGRL4 and ADGRG1 up to the start of the first transmembrane domain, and included a T→G mutation at the GPS motif which had been previously shown to abolish proteolysis but not trafficking to the cell surface (Araç et al., 2012). Expression of the new constructs yielded a much higher level of recombinant protein that was expressed at the expected molecular weights (Figure 5.3C). Hence, I applied this strategy to design recombinant ectodomain constructs for expressing other members of the adhesion GPCR family.

After eliminating adhesion GPCRs with ectodomains exceeding the maximum length for gene synthesis, I designed a total of 21 constructs, 18 of which were successfully synthesised and 13 produced sufficient protein for CRISPRa screening (Figure 5.3D). To determine whether constructs harbouring the T/S→G mutation retained the ability to bind to their endogenous ligands, I performed plate-based interaction assays with recombinant CD55 and ADGRE5, and showed that recombinant ADGRE5 was able to bind to CD55 (Figure 5.3E).

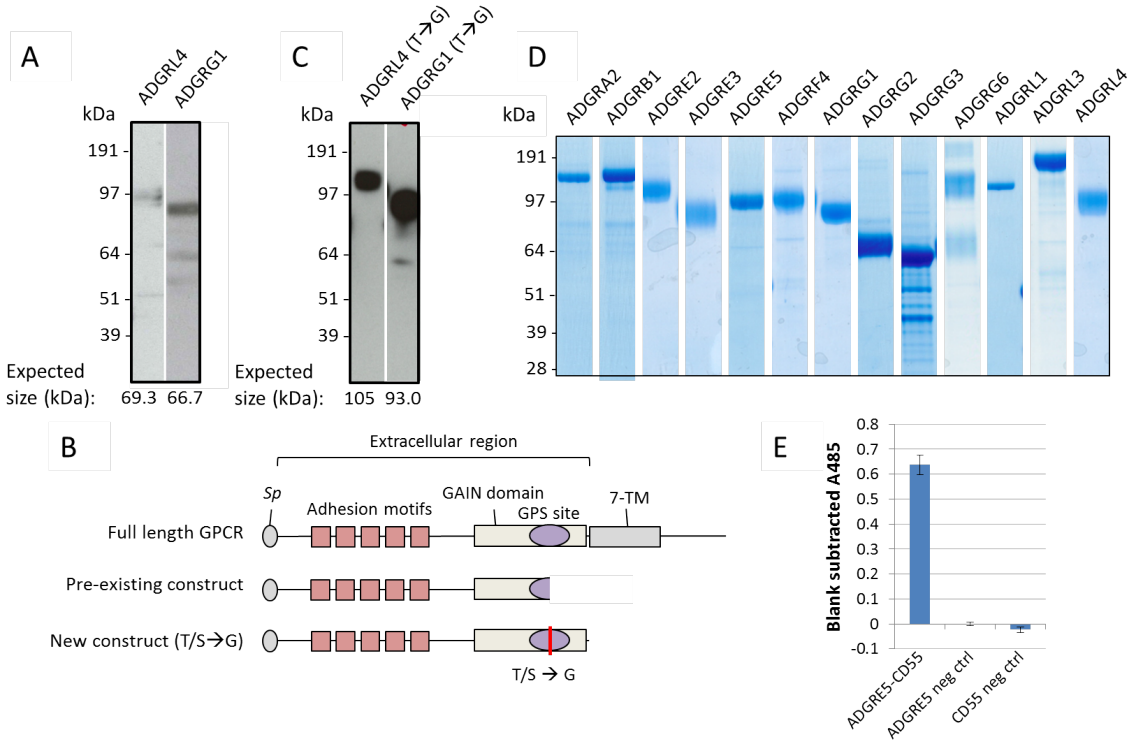



**Figure 5.3 T/S→G mutation at the GPS site enables the production of soluble recombinant adhesion GPCR ectodomains for CRISPRa screening.** A) Constructs expressing truncated ectodomains of ADGRL4 and ADGRG1 do not produce biotinylated proteins at the expected sizes as observed by western blotting. 10 $\mu$L of culture supernatant was loaded in each well. Detection of biotinylated proteins was performed by incubating blot with streptavidin conjuated 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 resistent 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 $Ni^{2+}$ 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 $\beta$-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.

### 5.2.2 CRISPRa screening of adhesion GPCR ectodomains identifies known interactions

For CRISPRa screening, I prepared highly avid tetramers from biotinylated ectodomains of 13 adhesion GPCRs according to the strategy described in (Chapter 4.2.4). Fluorescently labelled tetramers were screened using the improved CRISPRa screening protocol at a 5% sort threshold, and any genes enriched with an FDR of $< 0.1$ were considered 'hits'. CRISPRa screening using ADGRL1 and ADGRL3 tetramers detected previously reported interactions with members of the FLRT and TENM families (Figure 5.4A). ADGRL1 and ADGRL3 belong to the Latrophilin subfamily and are neuronal receptors for $\alpha$-latrotoxin, controlling neurotransmitter release and presynaptic calcium levels. In separate mass spectrometry based studies, ADGRLs have been found to interact with FLRTs (fibronectin leucine-rich transmembrane proteins) and TENMs (Teneurins) that are expressed on the neuronal cell surface and are implicated in controlling neurite outgrowth and patterning (O'Sullivan et al., 2012; Silva et al., 2011). Importantly, the detection of the ADGRL1-TENM3/4 interactions would likely not be possible using plate-based assays as TENMs are very large, Type II transmembrane proteins which might be difficult to produce in soluble recombinant form.

Additionally, a screen using ADGRA2 tetramers detected an enrichment of guides targeting syndecans (SDC1 and SDC2), a major family of heparan sulfate proteoglycans (Figure 5.4B). ADGRA2 is known to bind sulfated glycosaminoglycans (GAG) such as heparan and chondroitin sulfates (Vallon and Essler, 2006), and syndecans are composed mainly of such GAGs attached to a core protein. To determine if binding is dependent purely on sulfated glycosaminoglycans, I performed cell binding assays with a cell line lacking SLC35B2, a transporter of 3'-phosphoadenosine-5'-phosphosulfate (PAPS), which acts as a sulfate donor during GAG sulfation. SLC35B2 knockout lines retain expression of syndecans but are unable to produce sulfated GAGs like heparan sulfate. A complete loss of binding of ADGRA2 tetramers was observed with SLC35B2 knockout cells (Figure 5.4C), suggesting that GAGs form the major determinant of ADGRA2 binding rather than binding being specific to syndecans. Taken together, the detection of known interactions provide further validation of the CRISPRa screening approach for the identification of extracellular receptor-ligand pairs.

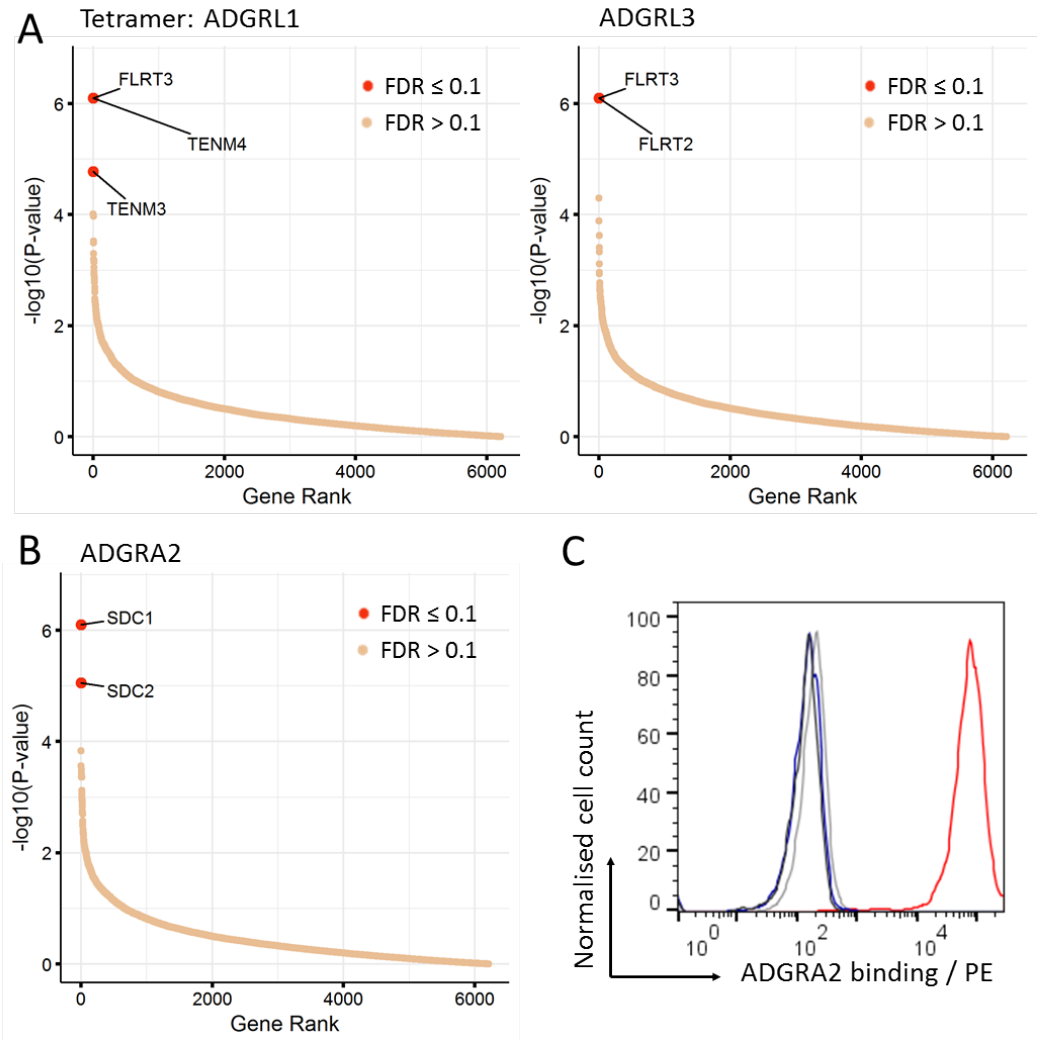



**Figure 5.4 CRISPRa screening identifies known interactions of ADGRL1 and ADGRL3, as well as glycosaminoglycan (GAG)-binding properties of ADGRA2.** A) Transformed gene enrichment P-values are plotted against a rank-ordered gene list for CRISPRa enrichment screens with cells selected using recombinant 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.

### 5.2.3 CRISPRa screening with ADGRB1 uncovers novel interactions with Nogo receptors

In addition to known interactions, CRISPRa screening identified previously unreported binding partners of ADGRB1. Guides targeting RTN4RL1 and RTN4RL2 were significantly enriched in a screen performed with ADGRB1 tetramers (Figure 5.5A). ADGRB1 (Brain Angiogenesis Inhibitor 1) is a phosphatidylserine receptor on professional phagocytes (Park et al., 2007), and is enriched in the postsynaptic density in neurons where it regulates excitatory synapse formation in hippocampal and cortical cultures (Duman et al., 2013) but has no documented ligands in the nervous system. On the other hand, RTN4RL1 and RTN4RL2 belong to the Nogo receptor family which are GPI-linked membrane proteins and are known to be involved in regulating axon growth and synapse formation, most notably through interactions between Nogo receptor 1 (RTN4R) and the myelin-associated inhibitor, Nogo-66 (Liu et al., 2002).

Given that guides targeting RTN4RL1 and RTN4RL2 were highly enriched in the sorted population, along with the fact that all three proteins were expressed in the brain, I decided to validate this interaction using cDNA overexpression in HEK293 cells. Transfection of HEK293 cells with constructs expressing the full-length cDNAs of RTN4RL1 and RTN4RL2 conferred an increased affinity for ADGRB1 tetramer binding relative to an untransfected control (Figure 5.5B). Surprisingly, overexpression of RTN4R also resulted in increased binding, even though guides targeting RTN4R were not significantly enriched in the initial screen (Figure 5.5C). The gain of binding phenotype was not due to recombinant protein tags as a control protein bearing the same tags did not show any increase in binding to transfected cells. To determine if the increase in binding was due to an indirect effect of upregulating extracellular phosphatidylserine, I performed Annexin V staining and did not see a significant increase in levels of extracellular phosphatidylserine in cells overexpressing Nogo receptors relative to an untransfected control (Figure 5.5D). These data suggest that ADGRB1 binds to all three members of the Nogo receptor family.

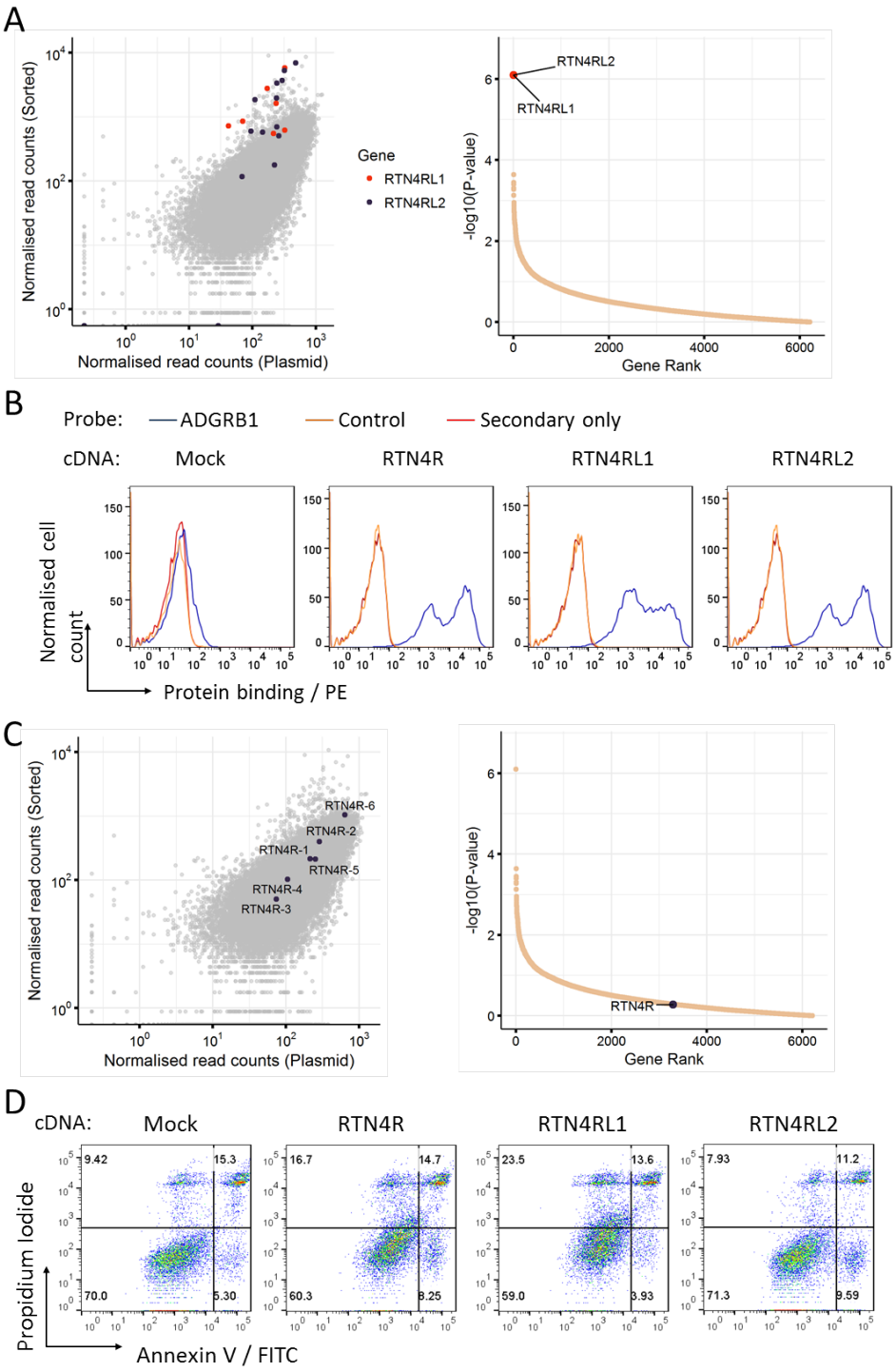

A
Normalised read counts (Sorted)
Normalised read counts (Plasmid)
Gene
RTN4RL1
RTN4RL2
-log10(P-value)
Gene Rank
B
Probe:
ADGRB1
Control
Secondary only
cDNA:
Mock
RTN4R
RTN4RL1
RTN4RL2
Normalised cell count
Protein binding / PE
C
RTN4R-6
RTN4R-2
RTN4R-1
RTN4R-5
RTN4R-4
RTN4R-3
D
Propidium Iodide
Annexin V / FITC
9.42
15.3
70.0
5.30
16.7
14.7
60.3
8.25
23.5
13.6
59.0
3.93
7.93
11.2
71.3
9.59

**Figure 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.

### 5.2.4 The first three thrombospondin repeats on ADGRB1 is sufficient for RTN4R binding

To determine direct binding between ADGRB1 and RTN4Rs, I performed plate-based interaction assays using recombinant ectodomains of all four receptors as previously described in Bushell et al. (2008). Briefly, the full length ectodomains of ADGRB1 and RTN4R, RTN4RL1 and RTN4RL2 were produced either as biotinylated monomers ('baits') or as pentameric, $\beta$-lactamase tagged proteins ('preys'). Biotinylated baits were captured on streptavidin coated plates and overlaid with pentameric preys before washing to remove unbound preys. A nitrocefin hydrolysis assay was then used to detect the presence of remaining preys. Colourimetric readouts indicated that ADGRB1 interacted with all three Nogo receptors (Figure 5.6A). This interaction was shown in both orientations, with ADGRB1 as either a bait or prey. RTN4RL2 could not be expressed in pentameric form, but when produced as a bait showed binding to ADGRB1. Importantly, none of the Nogo receptors interacted with ADGRB2, a closely related receptor to ADGRB1. Both ADGRB1 and ADGRB2 ectodomains contain several thrombospondin repeats (TSRs) as well as a hormone binding domain (HRM), but clearly exhibit different binding properties. This suggests that Nogo receptors interact specifically with ADGRB1.

To investigate the minimal requirements for ADGRB1-RTN4R binding, I produced truncated versions of the ADGRB1 ectodomain and tested their ability to interact with RTN4R. I expressed just the first three TSRs (TSR1-3), all five TSRs (TSR1-5) or just the HRM and GAIN domains (HRM+GAIN) of ADGRB1. Plate-based interaction assays using these constructs indicated that the first three TSRs were sufficient for binding of RTN4R (Figure 5.6B). Accordingly, the TSR1-5 fragment was also able to interact with RTN4R, but not the HRM+GAIN fragment. This indicates that the HRM and GAIN domains do not play a role in the interaction between ADGRB1 and RTN4Rs. In summary, ADGRB1 interacts directly and specifically with all three Nogo receptors through the first three TSRs on its ectodomain.

## 5.3 Discussion

In this chapter, I applied CRISPRa extracellular interaction screening to identify cell surface binding partners of adhesion GPCRs. To do so I designed expression constructs to produce full length ectodomains with a mutation at the GPS site to prevent cleavage. The

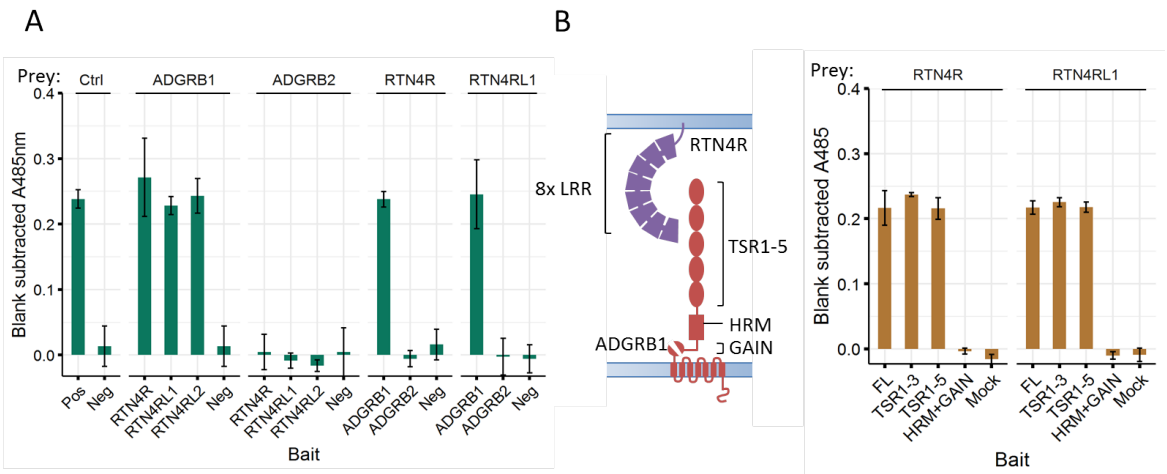



**Figure 5.6 ADGRB1 specifically and directly interacts with Nogo receptors through the first three thrombospondin repeats (TSRs) in its ectodomain.** A) The ectodomains of ADGRB1 and RTN4R family members directly interact. The extracellular regions of the named receptors were expressed as soluble biotinylated bait proteins, captured in individual wells of a streptavidin-coated plate and probed for interactions with pentameric $\beta$-lactamase-tagged prey proteins. Binding is quantified by absorbance at 485 nm of a hydrolysis product of the colourimetric $\beta$-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.

success of this strategy over producing the N-terminal fragment of the ectodomain after proteolytic cleavage indicates that the presence of the extracellular portion of the C-terminal fragment is essential for proper folding of the receptor. Furthermore, the ability to produce fragments containing just the TSRs of ADGRB1 suggest that this dependence is specific to the GAIN domain, and that other adhesion motifs within the ectodomain are able to fold independently.

CRISPRa screening of adhesion GPCRs identified known interactions involving members of the Latrophilin subfamily and detected GAG-binding properties of ADGRA2. Interestingly, although CRISPRa screening identified TENM3 and TENM4 as binding partners for ADGRL1, these interactions were not observed in a cell binding assay using cDNA overexpression of members of the Teneurin family (Silva et al., 2011). This discrepancy could be due to the use of a specific isoform for cDNA overexpression which may be non-functional, whilst CRISPRa allows endogenous splicing decisions which might have resulted in the expression of an isoform capable of binding ADGRL1. This highlights an advantage of the CRISPRa platform over cDNA overexpression, particularly for receptors which are poorly annotated or have many isoforms. However, CRISPRa screening failed to identify several binding partners of both ADGRLs including FLRT1 and other members of the TENM family. Given that ADGRL tetramers were able to interact with their endogenous receptors, this is likely due to an inability of CRISPRa to upregulate FLRT1 and other TENMs. This could be due to inefficient guides, or targeting of alternative TSSs, and should be improved with better knowledge of guide design principles as well as TSS annotation.

For the remaining nine adhesion GPCRs, no hits were obtained at an FDR cut-off of 0.1. Reducing the threshold to FDR $< 0.25$ did not reveal any additional hits, suggesting that no genes were enriched in these screens. This could be due to a number of reasons, including the possibility that the tetramers for these adhesion GPCRs were unable to bind their endogenous partners, or that the adhesion GPCRs do not have cell surface ligands. Instead, some adhesion GPCRs may only bind extracellular matrix proteins. For instance, collagen type III is an activating ligand of ADGRG1 during cortical development and laminationLuo et al. (2011) whilst ADGRG6 interacts with laminin-211 and collage type IV to regulate Schwann cell development and peripheral nerve development respectively (Paavola et al., 2014; Petersen et al., 2015). This highlights a limitation of the CRISPRa screening platform, as this approach is unable to detect extracellular interactions with soluble secreted factors or extracellular matrix proteins.

CRISPRa screening of ADGRB1 tetramers detected novel interactions with members of the Nogo receptor family, RTN4RL1 and RTN4RL2. Although it did not appear as a hit in the initial screen, subsequent validation experiments showed that ADGRB1 was also able to bind to the third Nogo receptor, RTN4R. The expression of all four proteins are enriched in the brain, and all have documented functions in the regulation of neurite growth and synapse formation both *in vitro* and *in vivo* (Duman et al., 2013; Wills et al., 2012). Importantly, no neuronal ligands of ADGRB1 have been identified, and no common ligands have been identified which bind to all three RTN4Rs. The Nogo receptor family is known to function redundantly with regards to regulating neuronal growth *in vivo* (Wills et al., 2012), and therefore the discovery of common binding partners may provide an explanation for this functional redundancy.

In summary, I demonstrated the utility of CRISPRa screening by applying it to a family of GPCRs with few known ligands. CRISPRa screening of adhesion GPCRs identified previously reported interactions, which provides confidence that this screening strategy works, and also identified novel interactions between ADGRB1 and the Nogo receptor family. These interactions were validated using cDNA overexpression and plate-based interaction assays and were shown to be mediated by TSRs on ADGRB1.

# CHAPTER 6

# GENERAL DISCUSSION

## 6.1 Summary of results

Extracellular interactions form the basis of how cells sense and respond to their environment. As such, these interactions are involved in a variety of biological processes including development, immune regulation and pathogen invasion. Such interactions also make attractive drug targets as they are readily accessible to systemically delivered drugs, such as therapeutic monoclonal antibodies. However, the identification of key receptor-ligand pairs, especially low-affinity cell adhesion interactions, can be technically challenging to perform at scale. CRISPR activation (CRISPRa) provides an attractive potential strategy for genome-wide extracellular interaction screening as it enables in principle upregulation of virtually any cell surface receptor in the genome and investigation of extracellular interactions in the context of a cell membrane. CRISPRa screening also circumvents the need for the production of large recombinant protein or cDNA libraries which can be costly and resource-intensive.

In this thesis I have investigated some parameters affecting cell surface protein upregulation using CRISPRa, and established a workflow for extracellular interaction screening using a gRNA library targeting the promoter regions of genes encoding all putative membrane proteins in the human genome. Using both antibodies and endogenous ligands, I show that CRISPRa screening can detect interactors with high confidence, even those which

bind with low affinity. Finally, I applied the CRISPRa screening approach to members of the adhesion GPCR family and identified novel cell surface ligands for ADGRB1.

## 6.2 Evaluation of CRISPR activation screening as an approach for receptor identification

Currently available strategies for identifying extracellular interactions include affinity purification with mass spectrometry (AP-MS), plate-based interaction assays using soluble recombinant ectodomains, arrayed cDNA overexpression screening, and CRISPR/Cas9 knockout screening. Each approach has its advantages and limitations, but some key advantages and limitations of CRISPRa screening are discussed below.

### 6.2.1 CRISPR activation uses gRNA libraries that are cost-effective to produce and maintain

Improvements in oligonucleotide synthesis methods have greatly reduced the cost of producing complex pools of oligonucleotides with good accuracy. As such, gRNA libraries capable of targeting thousands of genes can be synthesised at a fraction of the cost of a comparably-sized library of full-length cDNA or recombinant protein expression constructs. Furthermore, a large plasmid or lentiviral library preparation can be used for numerous screens, reducing the need for maintenance of cDNA stocks or repeated protein production. In this study I designed and cloned a gRNA library targeting 6,213 genes encoding all putative membrane proteins. By contrast, the largest plate-based recombinant protein screen tested pairwise interactions of 250 proteins (Martin et al., 2010), whilst the largest available membrane protein cDNA library contains clones encoding 4,493 membrane proteins (Mullican et al., 2017), or an estimated 75% of the human surfaceome. One caveat with the membrane protein gRNA library is that I was unable to obtain an accurate estimate of the fraction of proteins that could be successfully upregulated. However, this is difficult to assess without having access to a large number of antibodies and cloning hundreds of individual gRNAs, or performing single-cell experiments which can be costly and technically challenging.

### 6.2.2 CRISPR activation allows genome-scale interaction screening

Aside from using cost-effective gRNA libraries to screen a large number of cell surface proteins at once, endogenous overexpression using CRISPRa also circumvents the restrictions of maximum insert lengths associated with conventional cloning into plasmid or virus-based expression vectors. By contrast, cDNA libraries tend to be biased towards smaller transcripts, with most plasmid-based expression vectors exhibiting reduced cloning efficiencies for inserts exceeding 7,000 - 8,000 bp. In this study, CRISPRa screening of ADGRL1 resulted in the enrichment of two members of the Teneurin family, TENM3 and TENM4, which are large proteins with a coding region of at least 8,097 and 7,209 bp respectively. This highlights the utility of endogenous overexpression for studying receptors with large domains. Furthermore, CRISPRa allows screening of multi-pass membrane proteins, which is difficult to achieve using recombinant protein approaches. This because non-contiguous ectodomains may not be able to fold independently of the transmembrane domains when produced recombinantly. In this study I demonstrated that CRISPRa can upregulate other multi-pass membrane receptors such as P2RX7 and ADGRE5. Thus, compared to currently available methods, CRISPRa is arguably the closest to achieving genome-scale extracellular interaction screening.

### 6.2.3 CRISPR activation screening is not restricted to receptors that are already expressed by a cell line

Furthermore, gain-of-function studies enable systematic testing of interactions without being restricted by endogenous expression in the screening cell line. In contrast, AP-MS and CRISPR knockout studies rely on pre-existing expression of a receptor candidate, possibly missing other interactors which are expressed in different tissues or under different contexts. In addition, CRISPR knockout screening might have problems identifying multiple, co-expressed receptors if knocking out one receptor on the cell does not reduce ligand binding (Sharma et al., 2018). On the other hand, CRISPRa can identify multiple interaction partners provided they can be expressed on the cell surface. For example, CRISPRa screening using EFNA1 identified EPHA2, EPHA4 and EPHA7.

### 6.2.4 CRISPRa screening is able to detect low affinity interactions

Importantly, I show that CRISPRa screening can identify endogenous interactions of medium to low affinity, particularly the CD55-ADGRE5 interaction. This is important as low-affinity interactions are often understudied, and can be difficult to detect with certain methods such as AP-MS, where stringent wash steps and the use of detergents for solubilisation of protein complexes can cause low-affinity binders to be lost (Wright, 2009). I was unable to detect the other low-affinity interaction tested (CTLA4-CD86) using CRISPRa screening, despite showing that the CTLA4 ectodomain probe was active and able to bind to a second receptor, CD80. However, the failure to detect the CTLA4-CD86 interaction may not be only due to low affinity and other explanations such as alternative TSSs are explored in Section 6.2.6. Ideally, more screens should be performed to gain a more accurate estimate of the sensitivity of this approach. However, as the cost of running each screen is not trivial, I limited the number of proteins tested to demonstrate that CRISPRa screening works.

### 6.2.5 CRISPRa screening is unable to detect certain types of extracellular interactions

One limitation of CRISPRa screening is that it cannot detect interactions between secreted proteins, or those requiring heteromeric receptors, unless the other subunits are already expressed in HEK293 cells. An example of the latter is integrins, which are formed from one $\alpha$ and one $\beta$ subunit and exhibit specific binding patterns depending on the combination of subunits. CRISPRa screening using an antibody against integrin $\alpha$v$\beta$3 detected enrichment of the gene encoding the $\beta$-subunit, ITGB3, but not the $\alpha$ subunit, suggesting that the $\alpha$-subunit is already expressed in HEK293 cells and able to form additional heterodimers with upregulated ITGB3. Additionally, interactions between soluble secreted factors or with extracellular matrix proteins tend to be of higher affinity than that between cell surface molecules and hence might be more amenable to mass spectrometry-based approaches.

### 6.2.6 Potential explanations for false negatives arising from CRISPRa screening and suggestions for improvement

One issue with CRISPRa screening is that it has a high false negative rate of an estimated 33.3%. This estimate is based on a relatively low number of interactions (nine). In addition, some known adhesion GPCR interactions like ADGRL1-TENM2 and ADGRL3-

FLRT1 were not identified with CRISPRa screening. Similarly, the ADGRB1-RTN4R interaction was not detected in the initial screen even though it was later shown using cDNA overexpression assays and AVEXIS. In the above examples of known false negatives, the selection probe was active as other cell surface binding partners were identified.

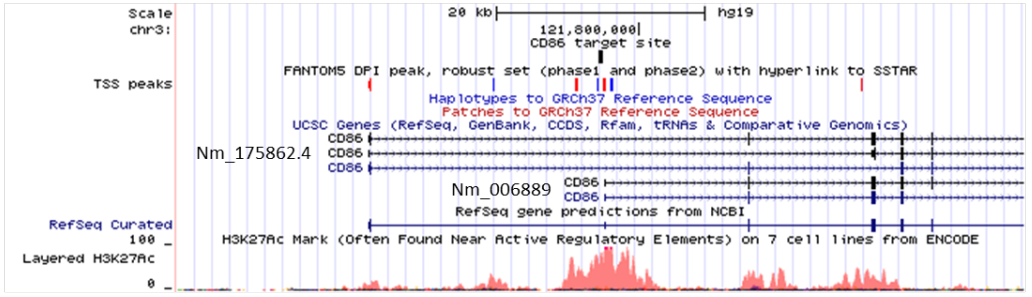
Potential reasons for the occurrence of false negatives including inactive guides, mis-annotation of transcription start sites (TSSs), targeting of alternative TSSs, or the lack of specific chaperones for surface transport and other post-transcriptional contextual effects. For instance, screening with CTLA4 ectodomains was able to identify a known binding partner CD80, but not a lower affinity interaction with CD86. Mapping of the CD86 gRNA targeting region alongside predicted gene models shows that gRNA in the library target a non-cannonical isoform of CD86 with a shorter signal peptide (Figure 6.1). This could result in a lack of expression of CD86 during pooled screening.

Whilst the inability to upregulate surface expression due to inactive guides or mis-targeting of TSSs is unique to CRISPRa, a lack of chaperones or other context-specific effects should be shared with cDNA-based overexpression approaches and might be remedied by using a different cell line for screening. Regarding gRNA design, there is ongoing research to study the parameters affecting guide effectiveness specifically for CRISPRa, as well as looking at more sophisticated ways to determine canonical TSSs. In fact, new genome-wide libraries have been published with improved selection algorithms that consider nucleosome positioning (Horlbeck et al., 2016), or with improved TSS predictions (Sanson et al., 2018). Although experimental comparisons are needed to determine whether more sophisticated design algorithms do indeed improve CRISPRa efficiency, future gRNA libraries might reduce the number of false negatives from CRISPRa screening that are due to the failure of inefficient or mistargeted gRNAs to upregulate receptors at the cell surface.

## 6.3 Implications of ADGRB1-RTN4R interactions and suggestions for further investigation

In this study, CRISPRa screening identified and validated a set of interactions between ADGRB1 and all three members of the Nogo receptor family (RTN4R, RTN4RL1, and RTN4RL2). All four proteins are implicated in synaptogenesis and neurite outgrowth, although no neuronal ligands for ADGRB1 have yet been identified. Nogo receptors on the other hand are known to regulate synaptogenesis through interactions between RTN4R

A




B

```
NM_006889       1   ------MGLSNILFVMAFLLSGAAPLKIQAYFNETADLPCQFANSQNQSLSELVVFWQDQ   54
NM_175862.4     1   MDPQCTMGLSNILFVMAFLLSGAAPLKIQAYFNETADLPCQFANSQNQSLSELVVFWQDQ   60
                    ************************************************************

NM_006889      55   ENLVLNEVYLGKEKFDSVHSKYMGRTSFDSDSWTLRLHNLQIKDKGLYQCIIHHKKPTGM  114
NM_175862.4    61   ENLVLNEVYLGKEKFDSVHSKYMGRTSFDSDSWTLRLHNLQIKDKGLYQCIIHHKKPTGM  120
                    ************************************************************

NM_006889     115   IRIHQMNSELSVLANFSQPEIVPISNITENVYINLTCSSIHGYPEPKKMSVLLRTKNSTI  174
NM_175862.4   121   IRIHQMNSELSVLANFSQPEIVPISNITENVYINLTCSSIHGYPEPKKMSVLLRTKNSTI  180
                    ************************************************************

NM_006889     175   EYDGIMQKSQDNVTELYDVSISLSVSFPDVTSNMTIFCILETDKTRLLSSPFSIELEDPQ  234
NM_175862.4   181   EYDGIMQKSQDNVTELYDVSISLSVSFPDVTSNMTIFCILETDKTRLLSSPFSIELEDPQ  240
                    ************************************************************

NM_006889     235   PPPDHIPWITAVLPTVIICVMVFCLILWKWKKKRPRNSYKCGTNTMEREESEQTKKREK  294
NM_175862.4   241   PPPDHIPWITAVLPTVIICVMVFCLILWKWKKKRPRNSYKCGTNTMEREESEQTKKREK  300
                    ************************************************************

NM_006889     295   IHIPERSDEAQRVFKSSKTSSCDKSDTCF                               323
NM_175862.4   301   IHIPERSDEAQRVFKSSKTSSCDKSDTCF                               329
                    *****************************
```

**Figure 6.1 CRISPRa library guides target a non-cannonical isoform of CD86**A) 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.

and Nogo, an inhibitory molecule expressed on myelin (Fournier et al., 2001). Myelin-associated glycoprotein (MAG) is another inhibitory myelin-associated molecule that binds to RTN4R and RTN4RL2, but not RTN4RL1 (Robak et al., 2009). RTN4RL1 does not bind any members of the reticulon family, however all three receptors functionally compensate for each other *in vivo* (Wills et al., 2012). Thus, a common binding partner of all three RTN4Rs might help explain functional redundancy.

### 6.3.1 Knockout phenotypes complicate the assessment of ADGRB1-RTN4R function

The function of an interaction can be easily demonstrated when knocking out one binding partner results in a phenocopy of the other knockout (KO). However, knockout phenotypes of ADGRB1 and RTN4Rs do not exactly coincide. Whilst the knockout of either ADGRB1 and RTN4Rs have effects on synaptogenesis both *in vitro* and *in vivo*, ADGRB1 knockout mice show reduced synaptogenesis in the hippocampus whilst a triple KO of Nogo receptors in mice resulted in abnormally elevated synaptogenesis in the same region (Wills et al., 2012; Zhu et al., 2015). ADGRB1 deficient mice have additional deficits in synaptic plasticity and spatial learning (Zhu et al., 2015). The phenotype of Nogo receptor deficient mice has been attributed to the interaction between RTN4R and myelin-associated inhibitors like MAG and Nogo, but this does not explain the fact that individual knockdowns of RTN4RL1 and RTN4RL2 result in increased numbers of excitatory synapses in neuronal cultures. The difference in KO phenotypes between ADGRB1 and RTN4Rs suggest that the interaction could be inhibitory, with binding preventing normal signalling of one of the receptors, or that additional receptors are involved. This makes it more challenging to elucidate the exact function of these interactions *in vivo*.

### 6.3.2 Levels of downstream effectors can help determine if RTN4Rs are activating ligands of ADGRB1

ADGRB1 is known to signal through G-protein $\alpha$-12/13, which is coupled to the activation of small GTPase RhoA (Stephenson et al., 2013). Additionally, ADGRB1 is known to signal through Rac1 GTPases independently of G-protein activation and couples with different Rac1-guanine nucleotide exchange factor modules during synaptogenesis and phagocytosis (Duman et al., 2013). To determine if Nogo receptors are activating (or inhibitory) ligands of ADGRB1, one could monitor levels of activated RhoA and Rac1

in cells expressing ADGRB1 and exposed to soluble ectodomains of RTN4Rs. Although I performed some preliminary experiments investigating RhoA activation in response to ADGRB1-RTN4R binding, I failed to see any activation of RhoA in HEK293 cells (data not shown). However, a more relevant cell line, such as the neuroblastoma-derived SHSY-5Y or primary hippocampal cultures, might provide a more accurate assessment of ADGRB1 activation/inhibition. Importantly, previous studies have shown that transfection of full length ADGRB1 into HEK293 cells results in a slight increase in RhoA activation above baseline (Stephenson et al., 2013), raising the possibility that inhibitory ligands for ADGRB1 could exist.

One explanation for the lack of evidence for ADGRB1 activation could be that ADGRB1-RTN4R interactions are purely adhesive and do not trigger downstream signalling via either receptor. Non-activating ligands of adhesion GPCRs have been previously described (Safaee et al., 2013), although it is difficult to conclude with certainty if an interaction is non-activating, if the subsequent response was below the detection threshold, or if signalling is taking place through alternative pathways.

## 6.4 Other possible applications of CRISPRa extracellular interaction screening

CRISPRa extracellular interaction screening is not restricted to the testing of single defined ligands to identify cell surface receptors. The membrane protein gRNA library and pooled screening approach can be used together with more physiologically relevant selection assays or reagents to identify relevant cell surface interactions involved in a number of biological processes.

### 6.4.1 Uncovering novel viral receptors

Viruses are intracellular obligate parasites which rely on host machinery to replicate (Alkhatib, 2009). This involves breaching the cell membrane to insert viral DNA within a host cell. Many viruses show specific host cell tropism, suggesting that specific host factors play a part in enabling viruses to invade host cells. Well-established examples are the CD4 receptor which, along with CXCR4 and CCR5 co-receptors, interact with Human Immunodificiency Virus (HIV) coat protein gp120 to initiate HIV fusion with the host cell membrane (Alkhatib,

2009). HIV co-receptor CCR5 has subsequently become a key target for the development of antiviral drugs and immunotherapy (Lopalco, 2010). CRISPRa screening could be used to uncover novel viral receptors by selecting for cells which gain an ability to be infected by fluorescently-labelled viral or virus-like particles from a population of cells transduced with the membrane protein gRNA library. In fact, other studies have demonstrated the use of CRISPRa for identifying host cell factors involved in influenza infection using genome-wide gRNA libraries (Heaton et al., 2017). A focused gRNA library like the membrane protein gRNA library provides a smaller library size whilst focusing on identifying extracellular factors that can be easily targeted to prevent viral entry.

### 6.4.2 Identifying extracellular interactions underlying cancer metastasis

The pooled rather than arrayed format of CRISPRa screening also provides opportunities for use with other types of selection assays, such as *in vivo* models of cancer metastasis. Metastasis refers to the spread of cancer cells from the primary tumour to surrounding tissues and is estimated to be the primary cause of cancer mortality as well as relapse (Seyfried and Huysentruyt, 2013). As such, the discovery of the mechanisms of primary tumour cell extravasation, immune evasion, and subsequent establishment in different tissues is paramount to developing more effective anticancer therapy. Established metastatic models in mice involve the injection or engraftment of cancer cells in host mice and subsequent review of the resulting metastases. Such models could be used to select for receptors that confer an advantage or disadvantage for cancer cells to undergo metastasis by injecting mice with a pool of cancer cells transduced with gRNA targeting membrane proteins. However, a challenge of performing *in vivo* studies is that the maximum number of cells injected or engrafted is typically quite low, thus requiring a large number of animals to provide sufficient coverage for screening with pooled libraries. One solution is to use smaller libraries targeting subsets of genes likely to contribute to metastasis, and pre-select highly active gRNA. Another important consideration would be the pre-existing metastatic ability of the cancer cell line used for screening. A cell line that spontaneously forms many metastatic colonies might make it difficult to detect enrichment of gRNAs causing gain-of-function metastatic ability over baseline.

## 6.5 Conclusion

In conclusion, I have shown that CRISPRa is a feasible approach for overexpression of cell surface receptors and established a CRISPRa screening workflow to identify novel extracellular interactions. CRISPRa screening provides a complimentary approach to currently available techniques for identifying extracellular interactions, with the advantage of allowing cost-efficient genome-scale screening. However, it is limited to detecting cell surface binding partners and false negatives may occur from inactive gRNAs. Nonetheless, CRISPRa screening can be used to detect novel interactions and could serve as an approach for identifying multiple receptors to a defined ligand or pool of ligands in a single screen.

# BIBLIOGRAPHY


Abbott, R. J. M., Spendlove, I., Roversi, P., Fitzgibbon, H., Knott, V., Teriete, P., McDonnell, J. M., Handford, P. A., and Lea, S. M. (2007). Structural and functional characterization of a novel T cell receptor co-regulatory protein complex, CD97-CD55. *J. Biol. Chem.*, 282(30):22023–22032.

Alkhatib, G. (2009). The biology of CCR5 and CXCR4. *Curr. Opin. HIV AIDS*, 4(2):96–103.

Altman, J. D., Moss, P. A., Goulder, P. J., Barouch, D. H., McHeyzer-Williams, M. G., Bell, J. I., McMichael, A. J., and Davis, M. M. (1996). Phenotypic analysis of antigen-specific T lymphocytes. *Science*, 274(5284):94–96.

Araç, D., Boucard, A. A., Bolliger, M. F., Nguyen, J., Soltis, S. M., Südhof, T. C., and Brunger, A. T. (2012). A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. *EMBO J.*, 31(6):1364–1378.

Bausch-Fluck, D., Hofmann, A., Bock, T., Frei, A. P., Cerciello, F., Jacobs, A., Moest, H., Omasits, U., Gundry, R. L., Yoon, C., Schiess, R., Schmidt, A., Mirkowska, P., Härtlová, A., Van Eyk, J. E., Bourquin, J.-P., Aebersold, R., Boheler, K. R., Zandstra, P., and Wollscheid, B. (2015). A mass Spectrometric-Derived cell surface protein atlas. *PLoS One*, 10(4):e0121314.

Bjarnadóttir, T. K., Fredriksson, R., Höglund, P. J., Gloriam, D. E., Lagerström, M. C., and Schiöth, H. B. (2004). The human and mouse repertoire of the adhesion family of g-protein-coupled receptors. *Genomics*, 84(1):23–33.

Brandsma, I. and Gent, D. C. (2012). Pathway choice in DNA double strand break repair: observations of a balancing act. *Genome Integr.*, 3(1):9.

Brodsky, R. A., Mukhina, G. L., Li, S., Nelson, K. L., Chiurazzi, P. L., Buckley, J. T., and Borowitz, M. J. (2000). Improved detection and characterization of paroxysmal nocturnal hemoglobinuria using fluorescent aerolysin. *Am. J. Clin. Pathol.*, 114(3):459–466.

Brown, M. H. and Barclay, A. N. (1994). Expression of immunoglobulin and scavenger receptor superfamily domains as chimeric proteins with domains 3 and 4 of CD4 for ligand analysis. *Protein Eng.*, 7(4):515–521.

Bushell, K. M., Söllner, C., Schuster-Boeckler, B., Bateman, A., and Wright, G. J. (2008). Large-scale screening for novel low-affinity extracellular protein interactions. *Genome Res.*, 18(4):622–630.

Chakraborty, S., Ji, H., Kabadi, A. M., Gersbach, C. A., Christoforou, N., and Leong, K. W. (2014). A CRISPR/Cas9-based system for reprogramming cell lineage specification. *Stem Cell Reports*, 3(6):940–947.

Chan, H. M. and La Thangue, N. B. (2001). p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J. Cell Sci.*, 114(Pt 13):2363–2373.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., Ter-Ovanesyan, D., Braff, J. L., Davidsohn, N., Housden, B. E., Perrimon, N., Weiss, R., Aach, J., Collins, J. J., and Church, G. M. (2015). Highly efficient cas9-mediated transcriptional programming. *Nat. Methods*, 12(4):326–328.

Chavez, A., Tuttle, M., Pruitt, B. W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S. J., Cecchi, R. J., Kowal, E. J. K., Buchthal, J., Housden, B. E., Perrimon, N., Collins, J. J., and Church, G. (2016). Comparison of cas9 activators in multiple species. *Nat. Methods*, 13(7):563–567.

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155(7):1479–1491.

Cheng, A. W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T. W., Rangarajan, S., Shivalila, C. S., Dadon, D. B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.*, 23:1163.

Collins, A. V., Brodie, D. W., Gilbert, R. J. C., Iaboni, A., Manso-Sancho, R., Walse, B., Stuart, D. I., van der Merwe, P. A., and Davis, S. J. (2002). The interaction properties of costimulatory molecules revisited. *Immunity*, 17(2):201–210.

Contardi, E., Palmisano, G. L., Tazzari, P. L., Martelli, A. M., Falà, F., Fabbi, M., Kato, T., Lucarelli, E., Donati, D., Polito, L., Bolognesi, A., Ricci, F., Salvi, S., Gargaglione, V., Mantero, S., Alberghini, M., Ferrara, G. B., and Pistillo, M. P. (2005). CTLA-4 is constitutively expressed on tumor cells and can trigger apoptosis upon ligand interaction. *Int. J. Cancer*, 117(4):538–550.

Crosnier, C., Bustamante, L. Y., Bartholdson, S. J., Bei, A. K., Theron, M., Uchikawa, M., Mboup, S., Ndir, O., Kwiatkowski, D. P., Duraisingh, M. T., Rayner, J. C., and Wright, G. J. (2011). Basigin is a receptor essential for erythrocyte invasion by plasmodium falciparum. *Nature*, 480(7378):534–537.

Czajkowsky, D. M., Hu, J., Shao, Z., and Pleass, R. J. (2012). Fc-fusion proteins: new developments and future perspectives. *EMBO Mol. Med.*, 4(10):1015–1028.

da Cunha, J. P. C., Galante, P. A. F., de Souza, J. E., de Souza, R. F., Carvalho, P. M., Ohara, D. T., Moura, R. P., Oba-Shinja, S. M., Marie, S. K. N., Silva, Jr, W. A., Perez, R. O., Stransky, B., Pieprzyk, M., Moore, J., Caballero, O., Gama-Rodrigues, J., Habr-Gama, A., Kuo, W. P., Simpson, A. J., Camargo, A. A., Old, L. J., and de Souza, S. J. (2009). Bioinformatics construction of the human cell surfaceome. *Proc. Natl. Acad. Sci. U. S. A.*, 106(39):16752–16757.

Dancy, B. M. and Cole, P. A. (2015). Protein lysine acetylation by p300/CBP. *Chem. Rev.*, 115(6):2419–2452.

Dayalan Naidu, S. and Dinkova-Kostova, A. T. (2017). Regulation of the mammalian heat shock factor 1. *FEBS J.*, 284(11):1606–1627.

Duman, J. G., Tzeng, C. P., Tu, Y.-K., Munjal, T., Schwechter, B., Ho, T. S.-Y., and Tolias, K. F. (2013). The adhesion-GPCR BAI1 regulates synaptogenesis by controlling the recruitment of the Par3/Tiam1 polarity complex to synaptic sites. *J. Neurosci.*, 33(16):6964–6978.

Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomart. *Nat. Protoc.*, 4(8):1184–1191.

Einhauer, A. and Jungbauer, A. (2001). The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods*, 49(1):455–465.

Ellis, E. L. and Delbrück, M. (1939). THE GROWTH OF BACTERIOPHAGE. *J. Gen. Physiol.*, 22(3):365–384.

Feige, M. J. and Hendershot, L. M. (2011). Disulfide bonds in ER protein folding and homeostasis. *Curr. Opin. Cell Biol.*, 23(2):167–175.

Ferguson, M. A. J., Hart, G. W., and Kinoshita, T. (2017). Glycosylphosphatidylinositol anchors. In Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H., editors, *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

Fields, S. and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246.

Fournier, A. E., GrandPre, T., and Strittmatter, S. M. (2001). Identification of a receptor mediating nogo-66 inhibition of axonal regeneration. *Nature*, 409(6818):341–346.

Frei, A. P., Jeon, O.-Y., Kilcher, S., Moest, H., Henning, L. M., Jost, C., Plückthun, A., Mercer, J., Aebersold, R., Carreira, E. M., and Wollscheid, B. (2012). Direct identification of ligand-receptor interactions on living cells and tissues. *Nat. Biotechnol.*, 30(10):997–1001.

Futschik, M. E., Chaurasia, G., and Herzel, H. (2007). Comparison of human protein-protein interaction maps. *Bioinformatics*, 23(5):605–611.

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M., and Weissman, J. S. (2014). Genome-Scale CRISPR-Mediated control of gene repression and activation. *Cell*, 159(3):647–661.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2):442–451.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J.-S., and Concordet, J.-P. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.*, 17(1):148.

Hardwick, J. M., Tse, L., Applegren, N., Nicholas, J., and Veliuona, M. A. (1992). The Epstein-Barr virus R transactivator (rta) contains a complex, potent activation domain with properties different from those of VP16. *J. Virol.*, 66(9):5500–5508.

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B. L., Barrell, D., Zadissa, A., Searle, S., Barnes, I., Bignell, A., Boychenko, V., Hunt, T., Kay, M., Mukherjee, G., Rajan, J., Despacio-Reyes, G., Saunders, G., Steward, C., Harte, R., Lin, M., Howald, C., Tanzer, A., Derrien, T., Chrast, J., Walters, N., Balasubramanian, S., Pei, B., Tress, M., Rodriguez, J. M., Ezkurdia, I., van Baren, J., Brent, M., Haussler, D., Kellis, M., Valencia, A., Reymond, A., Gerstein, M., Guigó, R., and Hubbard, T. J. (2012). GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.*, 22(9):1760–1774.

Hauser, A. S., Chavali, S., Masuho, I., Jahn, L. J., Martemyanov, K. A., Gloriam, D. E., and Babu, M. M. (2018). Pharmacogenomics of GPCR drug targets. *Cell*, 172(1-2):41–54.e19.

Heaton, B. E., Kennedy, E. M., Dumm, R. E., Harding, A. T., Sacco, M. T., Sachs, D., and Heaton, N. S. (2017). A CRISPR activation screen identifies a pan-avian influenza virus inhibitory host factor. *Cell Rep.*, 20(7):1503–1512.

Hellebrekers, D. M. E. I., Castermans, K., Viré, E., Dings, R. P. M., Hoebers, N. T. H., Mayo, K. H., Oude Egbrink, M. G. A., Molema, G., Fuks, F., van Engeland, M., and Griffioen, A. W. (2006). Epigenetic regulation of tumor endothelial cell anergy: silencing of intercellular adhesion molecule-1 by histone modifications. *Cancer Res.*, 66(22):10770–10777.

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., and Gersbach, C. A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.*, 33(5):510–517.

Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R., and Stüber, D. (1988). Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Biotechnology*, 6:1321.

Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., Fields, A. P., Park, C. Y., Corn, J. E., Kampmann, M., and Weissman, J. S. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife*, 5.

Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., Gygi, M. P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., Dong, R., Guarani, V., Vaites, L. P., Ordureau, A., Rad, R., Erickson, B. K., Wühr, M., Chick, J., Zhai, B., Kolippakkam, D., Mintseris, J., Obar, R. A., Harris, T., Artavanis-Tsakonas, S., Sowa, M. E., De Camilli, P., Paulo, J. A., Harper, J. W., and Gygi, S. P. (2015). The BioPlex network: A systematic exploration of the human interactome. *Cell*, 162(2):425–440.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096):816–821.

Jinno, A. and Park, P. W. (2015). Role of glycosaminoglycans in infectious disease. *Methods Mol. Biol.*, 1229:567–585.

Josefin Bartholdson, S., Bustamante, L. Y., Crosnier, C., Johnson, S., Lea, S., Rayner, J. C., and Wright, G. J. (2012). Semaphorin-7A is an erythrocyte receptor for p. falciparum Merozoite-Specific TRAP homolog, MTRAP. *PLoS Pathog.*, 8(11):e1003031.

Kent, W. J. (2002). BLAT—The BLAST-Like alignment tool. *Genome Res.*, 12(4):656–664.

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., Hsu, P. D., Habib, N., Gootenberg, J. S., Nishimasu, H., Nureki, O., and Zhang, F. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, 517(7536):583–588.

Leeuwenberg, J. F., Smeets, E. F., Neefjes, J. J., Shaffer, M. A., Cinek, T., Jeunhomme, T. M., Ahern, T. J., and Buurman, W. A. (1992). E-selectin and intercellular adhesion molecule-1 are released by activated human endothelial cells in vitro. *Immunology*, 77(4):543–549.

Lema Tomé, C. M., Palma, E., Ferluga, S., Lowther, W. T., Hantgan, R., Wykosky, J., and Debinski, W. (2012). Structural and functional characterization of monomeric EphrinA1 binding site to EphA2 receptor. *J. Biol. Chem.*, 287(17):14012–14022.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M. I., Zhang, F., Irizarry, R. A., Liu, J. S., Brown, M., and Liu, X. S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.*, 15(12):554.

Lin, H. H., Stacey, M., Saxby, C., Knott, V., Chaudhry, Y., Evans, D., Gordon, S., McKnight, A. J., Handford, P., and Lea, S. (2001). Molecular analysis of the epidermal growth factor-like short consensus repeat domain-mediated protein-protein interactions: dissection of the CD97-CD55 complex. *J. Biol. Chem.*, 276(26):24160–24169.

Liu, B. P., Fournier, A., GrandPré, T., and Strittmatter, S. M. (2002). Myelin-associated glycoprotein as a functional ligand for the nogo-66 receptor. *Science*, 297(5584):1190–1193.

Liu, H., Wei, Z., Dominguez, A., Li, Y., Wang, X., and Qi, L. S. (2015). CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation. *Bioinformatics*, 31(22):3676–3678.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods*, 25(4):402–408.

Loignon, M., Perret, S., Kelly, J., Boulais, D., Cass, B., Bisson, L., Afkhamizarreh, F., and Durocher, Y. (2008). Stable high volumetric production of glycosylated human recombinant IFNalpha2b in HEK293 cells. *BMC Biotechnol.*, 8:65.

Lopalco, L. (2010). CCR5: From natural resistance to a new Anti-HIV strategy. *Viruses*, 2(2):574–600.

Luo, R., Jeong, S.-J., Jin, Z., Strokes, N., Li, S., and Piao, X. (2011). G protein-coupled receptor 56 and collagen III, a receptor-ligand pair, regulates cortical development and lamination. *Proc. Natl. Acad. Sci. U. S. A.*, 108(31):12925–12930.

Ma, H., Wu, Y., Dang, Y., Choi, J.-G., Zhang, J., and Wu, H. (2014). Pol III promoters to express small RNAs: Delineation of transcription initiation. *Mol. Ther. Nucleic Acids*, 3:e161.

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H., and Joung, J. K. (2013). CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods*, 10(10):977–979.

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., Backofen, R., and Koonin, E. V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.*, 13(11):722–736.

Martin, S., Söllner, C., Charoensawan, V., Adryan, B., Thisse, B., Thisse, C., Teichmann, S., and Wright, G. J. (2010). Construction of a large extracellular protein interaction network and its resolution by spatiotemporal expression profiling. *Mol. Cell. Proteomics*, 9(12):2654–2665.

Michel, M. C., Wieland, T., and Tsujimoto, G. (2009). How reliable are g-protein-coupled receptor antibodies? *Naunyn. Schmiedebergs. Arch. Pharmacol.*, 379(4):385–388.

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, 60(2):174–182.

Mullican, S. E., Lin-Schmidt, X., Chin, C.-N., Chavez, J. A., Furman, J. L., Armstrong, A. A., Beck, S. C., South, V. J., Dinh, T. Q., Cash-Mason, T. D., Cavanaugh, C. R., Nelson, S., Huang, C., Hunter, M. J., and Rangwala, S. M. (2017). GFRAL is the receptor for GDF15 and the ligand promotes weight loss in mice and nonhuman primates. *Nat. Med.*, 23(10):1150–1157.

Nam, J.-W., Rissland, O. S., Koppstein, D., Abreu-Goodger, C., Jan, C. H., Agarwal, V., Yildirim, M. A., Rodriguez, A., and Bartel, D. P. (2014). Global analyses of the effect of different cellular contexts on microRNA targeting. *Mol. Cell*, 53(6):1031–1043.

Nielsen, S., Yuzenkova, Y., and Zenkin, N. (2013). Mechanism of eukaryotic RNA polymerase III transcription termination. *Science*, 340(6140):1577–1580.

Ong, S. H., Li, Y., Koike-Yusa, H., and Yusa, K. (2017). Optimised metrics for CRISPR-KO screens with second-generation gRNA libraries. *Sci. Rep.*, 7(1):7384.

Ord, R. L., Rodriguez, M., and Lobo, C. A. (2015). Malaria invasion ligand RH5 and its prime candidacy in blood-stage malaria vaccine design. *Hum. Vaccin. Immunother.*, 11(6):1465–1473.

O'Sullivan, M. L., de Wit, J., Savas, J. N., Comoletti, D., Otto-Hitt, S., Yates, 3rd, J. R., and Ghosh, A. (2012). FLRT proteins are endogenous latrophilin ligands and regulate excitatory synapse development. *Neuron*, 73(5):903–910.

Özkan, E., Carrillo, R. A., Eastman, C. L., Weiszmann, R., Waghray, D., Johnson, K. G., Zinn, K., Celniker, S. E., and Garcia, K. C. (2013). An extracellular interactome of immunoglobulin and LRR proteins reveals receptor-ligand networks. *Cell*, 154(1):228–239.

Paavola, K. J., Sidik, H., Zuchero, J. B., Eckart, M., and Talbot, W. S. (2014). Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. *Sci. Signal.*, 7(338):ra76.

Park, D., Tosello-Trampont, A.-C., Elliott, M. R., Lu, M., Haney, L. B., Ma, Z., Klibanov, A. L., Mandell, J. W., and Ravichandran, K. S. (2007). BAI1 is an engulfment receptor for apoptotic cells upstream of the ELMO/Dock180/Rac module. *Nature*, 450(7168):430–434.

Paulick, M. G. and Bertozzi, C. R. (2008). The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. *Biochemistry*, 47(27):6991–7000.

Petersen, S. C., Luo, R., Liebscher, I., Giera, S., Jeong, S.-J., Mogha, A., Ghidinelli, M., Feltri, M. L., Schöneberg, T., Piao, X., and Monk, K. R. (2015). The adhesion GPCR GPR126 has distinct, domain-dependent functions in schwann cell development mediated by interaction with laminin-211. *Neuron*, 85(4):755–769.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, 8(10):785–786.

Pierleoni, A., Martelli, P. L., and Casadio, R. (2008). PredGPI: a GPI-anchor predictor. *BMC Bioinformatics*, 9:392.

Pizarro-Cerdá, J. and Cossart, P. (2006). Bacterial adhesion and entry into host cells. *Cell*, 124(4):715–727.

Quail, M. A., Otto, T. D., Gu, Y., Harris, S. R., Skelly, T. F., McQuillan, J. A., Swerdlow, H. P., and Oyola, S. O. (2011). Optimal enzymes for amplifying sequencing libraries. *Nat. Methods*, 9(1):10–11.

Robak, L. A., Venkatesh, K., Lee, H., Raiker, S. J., Duan, Y., Lee-Osbourne, J., Hofer, T., Mage, R. G., Rader, C., and Giger, R. J. (2009). Molecular basis of the interactions of the nogo-66 receptor and its homolog NgR2 with myelin-associated glycoprotein: development of NgROMNI-Fc, a novel antagonist of CNS myelin inhibition. *J. Neurosci.*, 29(18):5768–5783.

Rodriguez, J. M., Maietta, P., Ezkurdia, I., Pietrelli, A., Wesselink, J.-J., Lopez, G., Valencia, A., and Tress, M. L. (2013). APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res.*, 41(Database issue):D110–7.

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S. J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., Kamburov, A., Ghiassian, S. D., Yang, X., Ghamsari, L., Balcha, D., Begg, B. E., Braun, P., Brehme, M., Broly, M. P., Carvunis, A.-R., Convery-Zupan, D., Corominas, R., Coulombe-Huntington, J., Dann, E., Dreze, M., Dricot, A., Fan, C., Franzosa, E., Gebreab, F., Gutierrez, B. J., Hardy, M. F., Jin, M., Kang, S., Kiros, R., Lin, G. N., Luck, K., MacWilliams, A., Menche, J., Murray, R. R., Palagi, A., Poulin, M. M., Rambout, X., Rasla, J., Reichert, P., Romero, V., Ruyssinck, E., Sahalie, J. M., Scholz, A., Shah, A. A., Sharma, A., Shen, Y., Spirohn, K., Tam, S., Tejeda, A. O., Trigg, S. A., Twizere, J.-C., Vega, K., Walsh, J., Cusick, M. E., Xia, Y., Barabási, A.-L., Iakoucheva, L. M., Aloy, P., De Las Rivas, J., Tavernier, J., Calderwood, M. A., Hill, D. E., Hao, T., Roth, F. P., and Vidal, M. (2014). A proteome-scale map of the human interactome network. *Cell*, 159(5):1212–1226.

Safaee, M., Clark, A. J., Ivan, M. E., Oh, M. C., Bloch, O., Sun, M. Z., Oh, T., and Parsa, A. T. (2013). CD97 is a multifunctional leukocyte receptor with distinct roles in human cancers (review). *Int. J. Oncol.*, 43(5):1343–1350.

Sambrook, J. and Russell, D. W. (2006). Isolation of DNA fragments from polyacrylamide gels by the crush and soak method. *CSH Protoc.*, 2006(1).

Sansom, D. M. (2000). CD28, CTLA-4 and their ligands: who does what and to whom? *Immunology*, 101(2):169–177.

Sanson, K. R., Hanna, R. E., Hegde, M., Donovan, K. F., Strand, C., Sullender, M. E., Vaimberg, E. W., Goodale, A., Root, D. E., Piccioni, F., and Doench, J. G. (2018). Up, down, and out: optimized libraries for CRISPRa, CRISPRi, and CRISPR-knockout genetic screens.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The streptococcus thermophilus CRISPR/Cas system provides immunity in escherichia coli. *Nucleic Acids Res.*, 39(21):9275–9282.

Selvaraj, P., Plunkett, M. L., Dustin, M., Sanders, M. E., Shaw, S., and Springer, T. A. (1987). The T lymphocyte glycoprotein CD2 binds the cell surface ligand LFA-3. *Nature*, 326(6111):400–403.

Seyfried, T. N. and Huysentruyt, L. C. (2013). On the origin of cancer metastasis. *Crit. Rev. Oncog.*, 18(1-2):43–73.

Sharma, S., Bartholdson, S. J., Couch, A. C., Yusa, K., and Wright, G. J. (2018). Genome-scale identification of cellular pathways required for cell surface recognition. *Genome Res.*

Silva, J.-P., Lelianova, V. G., Ermolyuk, Y. S., Vysokov, N., Hitchen, P. G., Berninghausen, O., Rahman, M. A., Zangrandi, A., Fidalgo, S., Tonevitsky, A. G., Dell, A., Volynski, K. E., and Ushkaryov, Y. A. (2011). Latrophilin 1 and its endogenous ligand lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities. *Proc. Natl. Acad. Sci. U. S. A.*, 108(29):12113–12118.

Simonsen, H. and Lodish, H. F. (1994). Cloning by function: expression cloning in mammalian cells. *Trends Pharmacol. Sci.*, 15(12):437–441.

Snider, J., Kittanakom, S., Curak, J., and Stagljar, I. (2010). Split-ubiquitin based membrane yeast two-hybrid (MYTH) system: a powerful tool for identifying protein-protein interactions. *J. Vis. Exp.*, (36).

Snider, J. and Stagljar, I. (2016). Membrane yeast Two-Hybrid (MYTH) mapping of Full-Length membrane protein interactions. *Cold Spring Harb. Protoc.*, 2016(1):db.top077560.

Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR–a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.*, 6(3):181–186.

Stephenson, J. R., Paavola, K. J., Schaefer, S. A., Kaur, B., Van Meir, E. G., and Hall, R. A. (2013). Brain-specific angiogenesis inhibitor-1 signaling, regulation, and enrichment in the postsynaptic density. *J. Biol. Chem.*, 288(31):22248–22256.

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease cas9. *Nature*, 507(7490):62–67.

Stoveken, H. M., Hajduczok, A. G., Xu, L., and Tall, G. G. (2015). Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. *Proc. Natl. Acad. Sci. U. S. A.*, 112(19):6194–6199.

Stupack, D. G. and Cheresh, D. A. (2002). Get a ligand, get a life: integrins, signaling and cell survival. *J. Cell Sci.*, 115(Pt 19):3729–3738.

Sun, Y., Gallagher-Jones, M., Barker, C., and Wright, G. J. (2012). A benchmarked protein microarray-based platform for the identification of novel low-affinity extracellular protein interactions. *Anal. Biochem.*, 424(1):45–53.

Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S., and Vale, R. D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*, 159(3):635–646.

Tanner, W. and Lehle, L. (1987). Protein glycosylation in yeast. *Biochim. Biophys. Acta*, 906(1):81–99.

Thul, P. J. and Lindskog, C. (2018). The human protein atlas: A spatial map of the human proteome. *Protein Sci.*, 27(1):233–244.

Tumbar, T., Sudlow, G., and Belmont, A. S. (1999). Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J. Cell Biol.*, 145(7):1341–1354.

Turner, L., Lavstsen, T., Berger, S. S., Wang, C. W., Petersen, J. E. V., Avril, M., Brazier, A. J., Freeth, J., Jespersen, J. S., Nielsen, M. A., Magistrado, P., Lusingu, J., Smith, J. D., Higgins, M. K., and Theander, T. G. (2013). Severe malaria is associated with parasite binding to endothelial protein C receptor. *Nature*, 498(7455):502–505.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., and Pontén, F. (2015). Proteomics. tissue-based map of the human proteome. *Science*, 347(6220):1260419.

Ulloa-Aguirre, A., Timossi, C., Damián-Matsumura, P., and Dias, J. A. (1999). Role of glycosylation in function of follicle-stimulating hormone. *Endocrine*, 11(3):205–215.

Vallon, M. and Essler, M. (2006). Proteolytically processed soluble tumor endothelial marker (TEM) 5 mediates endothelial cell survival during angiogenesis by linking integrin alpha(v)beta3 to glycosaminoglycans. *J. Biol. Chem.*, 281(45):34179–34188.

van der Merwe, P. A. and Barclay, A. N. (1994). Transient intercellular adhesion: the importance of weak protein-protein interactions. *Trends Biochem. Sci.*, 19(9):354–358.

van der Merwe, P. A., Bodian, D. L., Daenke, S., Linsley, P., and Davis, S. J. (1997). CD80 (b7-1) binds both CD28 and CTLA-4 with a low affinity and very fast kinetics. *J. Exp. Med.*, 185(3):393–403.

van Essen, D., Engist, B., Natoli, G., and Saccani, S. (2009). Two modes of transcriptional activation at native promoters by NF-kappaB p65. *PLoS Biol.*, 7(3):e73.

Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., Lander, E. S., and Sabatini, D. M. (2015). Identification and characterization of essential genes in the human genome. *Science*, 350(6264):1096–1101.

Wills, Z. P., Mandel-Brehm, C., Mardinly, A. R., McCord, A. E., Giger, R. J., and Greenberg, M. E. (2012). The nogo receptor family restricts synapse number in the developing hippocampus. *Neuron*, 73(3):466–481.

Wojtowicz, W. M., Wu, W., Andre, I., Qian, B., Baker, D., and Zipursky, S. L. (2007). A vast repertoire of dscam binding specificities arises from modular interactions of variable ig domains. *Cell*, 130(6):1134–1145.

Wright, G. J. (2009). Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions. *Mol. Biosyst.*, 5(12):1405–1412.

Wright, G. J., Cherwinski, H., Foster-Cuevas, M., Brooke, G., Puklavec, M. J., Bigler, M., Song, Y., Jenmalm, M., Gorman, D., McClanahan, T., Liu, M.-R., Brown, M. H., Sedgwick, J. D., Phillips, J. H., and Barclay, A. N. (2003). Characterization of the CD200 receptor family in mice and humans and their interactions with CD200. *J. Immunol.*, 171(6):3034–3046.

Wright, G. J., Puklavec, M. J., Willis, A. C., Hoek, R. M., Sedgwick, J. D., Brown, M. H., and Barclay, A. N. (2000). Lymphoid/neuronal cell surface OX2 glycoprotein recognizes a novel receptor on macrophages implicated in the control of their function. *Immunity*, 13(2):233–242.

Wuethrich, I., Peeters, J. G. C., Blom, A. E. M., Theile, C. S., Li, Z., Spooner, E., Ploegh, H. L., and Guimaraes, C. P. (2014). Site-specific chemoenzymatic labeling of aerolysin enables the identification of new aerolysin receptors. *PLoS One*, 9(10):e109883.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13:134.

Young, M. T., Beckmann, R., Toye, A. M., and Tanner, M. J. (2000). Red-cell glycophorin a-band 3 interactions associated with the movement of band 3 to the cell surface. *Biochem. J*, 350 Pt 1:53–60.

Yusa, K., Zhou, L., Li, M. A., Bradley, A., and Craig, N. L. (2011). A hyperactive piggybac transposase for mammalian applications. *Proc. Natl. Acad. Sci. U. S. A.*, 108(4):1531–1536.

Zhang, N. and Li, L. (2004). Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation. *Rapid Commun. Mass Spectrom.*, 18(8):889–896.

Zhu, D., Li, C., Swanson, A. M., Villalba, R. M., Guo, J., Zhang, Z., Matheny, S., Murakami, T., Stephenson, J. R., Daniel, S., Fukata, M., Hall, R. A., Olson, J. J., Neigh, G. N., Smith, Y., Rainnie, D. G., and Meir, E. G. V. (2015). BAI1 regulates spatial learning and synaptic plasticity in the hippocampus. *J. Clin. Invest.*, 125(4):1497–1508.

Ziauddin, J. and Sabatini, D. M. (2001). Microarrays of cells expressing defined cDNAs. *Nature*, 411(6833):107–110.

# APPENDIX A

## A.1 Table 1: gRNA sequences targeting 12 cell surface proteins

| ID | Gene symbol | Transcript | sgRNA sequence | Chr | Location |
|---|---|---|---|---|---|
| SLC4A1-sg1 | SLC4A1 | NM_000342 | GGGTTTGCGAGCTGCCCCTG | 17 | 42345519 |
| SLC4A1-sg2 | SLC4A1 | NM_000342 | GCCAGTGGGGCGGGCAGATT | 17 | 42345561 |
| SLC4A1-sg3 | SLC4A1 | NM_000342 | AAGAGATAACTCTGTTTACT | 17 | 42345629 |
| SLC4A1-sg4 | SLC4A1 | NM_000342 | ACTCACAGCTGTCCAGATGT | 17 | 42345653 |
| SLC4A1-sg5 | SLC4A1 | NM_000342 | GACTCTTCCTTTGTGGATGA | 17 | 42345743 |
| SLC4A1-sg6 | SLC4A1 | NM_000342 | GTTTGATCGCTCTGTCCTCA | 17 | 42345765 |
| SLC4A1-sg7 | SLC4A1 | NM_000342 | GGGAACTGCTCAGCACTCAC | 17 | 42345964 |
| SLC4A1-sg8 | SLC4A1 | NM_000342 | AGTCTGGATCAAGGAGGGGA | 17 | 42346002 |
| RHD-sg1 | RHD | NM_001127691 | GCCTGAGATAAGGCCTTTGG | 1 | 25598918 |
| RHD-sg2 | RHD | NM_001127691 | TCCGTGTTAACTCCATAGAG | 1 | 25598875 |
| RHD-sg3 | RHD | NM_001127691 | GCACAGCAGGAACCTGTAAC | 1 | 25598775 |
| RHD-sg4 | RHD | NM_001127691 | GGATTATGTTTGGGTGTCAA | 1 | 25598727 |
| RHD-sg5 | RHD | NM_001127691 | CATTGTTGTTAAGAGCTCAC | 1 | 25598559 |
| RHD-sg6 | RHD | NM_001127691 | GCCCTCTCTGTCATGTAGTA | 1 | 25598584 |
| RHD-sg7 | RHD | NM_001127691 | TGGTTGTGCTGGCCTCTCTA | 1 | 25598890 |
| RHD-sg8 | RHD | NM_001127691 | ATTTCAACTGTGTAACTATG | 1 | 25598456 |

| | | | | | |
|---|---|---|---|---|---|
| P2RX7-sg1 | P2RX7 | NM_002562 | GTTTATCACAGCCACATGTG | 12 | 121570578 |
| P2RX7-sg2 | P2RX7 | NM_002562 | GGTGAGGTCATCTGCCAGCC | 12 | 121570533 |
| P2RX7-sg3 | P2RX7 | NM_002562 | ACCATCTTTGTGTAGGCATC | 12 | 121570496 |
| P2RX7-sg4 | P2RX7 | NM_002562 | GACCAAAAAAGTGAAAGGAA | 12 | 121570433 |
| P2RX7-sg5 | P2RX7 | NM_002562 | CTCATGTCTCTTGGGAGAAA | 12 | 121570414 |
| P2RX7-sg6 | P2RX7 | NM_002562 | AGTCCTTTTCTGAGGCATAA | 12 | 121570355 |
| P2RX7-sg7 | P2RX7 | NM_002562 | AGCGCCAAGTCCTACGGGCC | 12 | 121570554 |
| P2RX7-sg8 | P2RX7 | NM_002562 | GCATCTGGGGGAGGCCAGCT | 12 | 121570511 |
| ENG-sg1 | ENG | NM_000118 | CCACCCAGTGACAAAGCCCG | 9 | 130617056 |
| ENG-sg2 | ENG | NM_000118 | AGCCTTGGAGAGGGTGGGAT | 9 | 130617155 |
| ENG-sg3 | ENG | NM_000118 | GGCCCCCTGAAAGTTCCCCT | 9 | 130617235 |
| ENG-sg4 | ENG | NM_000118 | GGAACTACTTTAGCCAAGAC | 9 | 130617256 |
| ENG-sg5 | ENG | NM_000118 | ATGGGATCAGTGAGCTCAGG | 9 | 130617294 |
| ENG-sg6 | ENG | NM_000118 | AACCAGTGATCTCAACACAT | 9 | 130617340 |
| ENG-sg7 | ENG | NM_000118 | ATGCCCGACAAGACGTGAAG | 9 | 130617400 |
| ENG-sg8 | ENG | NM_000118 | GTCAACTGCACTTAGTAGGC | 9 | 130617428 |
| CD2-sg1 | CD2 | NM_001767 | AGGAACTGAAGTGAGACTGG | 1 | 117297049 |
| CD2-sg2 | CD2 | NM_001767 | GAGGCACGTGGTTAAGCTCT | 1 | 117297013 |
| CD2-sg3 | CD2 | NM_001767 | ACTGTAAAAGATGTAAAGAG | 1 | 117296994 |
| CD2-sg4 | CD2 | NM_001767 | GGCAAAGGAGCACATCAGAA | 1 | 117296920 |
| CD2-sg5 | CD2 | NM_001767 | AATTCTCACACAAAAAAATT | 1 | 117296857 |
| CD2-sg6 | CD2 | NM_001767 | ACTCATAAACACATCTGCTT | 1 | 117296899 |
| CD2-sg7 | CD2 | NM_001767 | AGAGGCTAAGTAGATCACTA | 1 | 117296792 |
| CD2-sg8 | CD2 | NM_001767 | AGTATACCTAAGTGGATAAA | 1 | 117296725 |
| VCAM1-sg1 | VCAM1 | NM_001078 | CTTCCAAGACTATAAAATAC | 1 | 101185088 |
| VCAM1-sg2 | VCAM1 | NM_001078 | TCCTCATCTTCGACTCCAAA | 1 | 101185061 |
| VCAM1-sg3 | VCAM1 | NM_001078 | TATCTTTACTGGAAAGATAA | 1 | 101185037 |
| VCAM1-sg4 | VCAM1 | NM_001078 | GAATCCAATGTGGGTTAAGG | 1 | 101184985 |
| VCAM1-sg5 | VCAM1 | NM_001078 | GAAGCTTTCTGAATCCAATG | 1 | 101184995 |
| VCAM1-sg6 | VCAM1 | NM_001078 | TTCTACTCTGGTTTTTGAAC | 1 | 101184929 |
| VCAM1-sg7 | VCAM1 | NM_001078 | TGAAGCTCCTCTCTCTGTCC | 1 | 101185123 |
| VCAM1-sg8 | VCAM1 | NM_001078 | TGAAATTGCTGCCAAAACAA | 1 | 101184660 |
| PROM1-sg1 | PROM1 | NM_001145847 | GACTGAGGCAGATCCCCACG | 4 | 16085635 |
| PROM1-sg2 | PROM1 | NM_001145847 | ATCAGAGTGCGTCCAGGGCT | 4 | 16085676 |
| PROM1-sg3 | PROM1 | NM_001145847 | GCGTTGCAAGAAGGGAGTGC | 4 | 16085790 |
| PROM1-sg4 | PROM1 | NM_001145847 | ATTCTAAGTAAGGGACTCTG | 4 | 16085830 |
| PROM1-sg5 | PROM1 | NM_001145847 | CAGAAGGGTCTAATGCGGCC | 4 | 16085886 |
| PROM1-sg6 | PROM1 | NM_001145847 | GAGGCGCAAGCGTTGCAAGA | 4 | 16085799 |
| PROM1-sg7 | PROM1 | NM_001145849 | GCAAGGCCTCCAGCCTAATC | 4 | 16077832 |

| | | | | | |
|---|---|---|---|---|---|
| PROM1-sg8 | PROM1 | NM_001145849 | GCGTGTAAACTGCCTGCACC | 4 | 16077776 |
| SEMA7A-sg1 | SEMA7A | NM_001146029 | CGCTTGGGTCTGCCTGCGGC | 15 | 74726312 |
| SEMA7A-sg2 | SEMA7A | NM_001146029 | AGGCGAGAAAAGGCTGCGAG | 15 | 74726380 |
| SEMA7A-sg3 | SEMA7A | NM_001146029 | AGCGAGAGCGGAACTGCTGG | 15 | 74726414 |
| SEMA7A-sg4 | SEMA7A | NM_001146029 | GAACCTTCGCCACCCTCTCC | 15 | 74726436 |
| SEMA7A-sg5 | SEMA7A | NM_001146029 | GCTTTCCCCGTAGAGTTGCC | 15 | 74726458 |
| SEMA7A-sg6 | SEMA7A | NM_001146029 | AGTCTGGCTTGTCCGCAGCT | 15 | 74726480 |
| SEMA7A-sg7 | SEMA7A | NM_001146029 | GATTGGCGTAGAAGTCGTGG | 15 | 74726575 |
| SEMA7A-sg8 | SEMA7A | NM_001146029 | ACCTCTCTCCAAGGGCGCAG | 15 | 74726598 |
| CD200-sg1 | CD200 | NM_001004196 | GACAGCCTCCGCTCCTGTGA | 3 | 112051884 |
| CD200-sg2 | CD200 | NM_001004196 | GAGCGGAGGCTGTCTGTGTG | 3 | 112051878 |
| CD200-sg3 | CD200 | NM_005944 | GAGAAAGGAAATGAGGTGGG | 3 | 112051716 |
| CD200-sg4 | CD200 | NM_001004196 | CACTTTGTCAGTTTCCCCAG | 3 | 112051787 |
| CD200-sg5 | CD200 | NM_001004196 | AGCTCTTGATGTAGTGGCAA | 3 | 112051667 |
| CD200-sg6 | CD200 | NM_001004196 | CAGTCCAGGTAGCAGGAAAA | 3 | 112051735 |
| CD200-sg7 | CD200 | NM_001004196 | ATCCTCATCATTAATGCAAG | 3 | 112051485 |
| CD200-sg8 | CD200 | NM_001004196 | AGAATTGATCACATCATGAA | 3 | 112051526 |
| ICAM1-sg1 | ICAM1 | NM_000201 | ACTTAATAAACCGCTTAGCG | 19 | 10381479 |
| ICAM1-sg2 | ICAM1 | NM_000201 | GAGGCCTGCGTAAGCTGGAG | 19 | 10381346 |
| ICAM1-sg3 | ICAM1 | NM_000201 | ATAACAGTCTCCACTCTCCG | 19 | 10381435 |
| ICAM1-sg4 | ICAM1 | NM_000201 | GTTCGGACCCCCTCGCAGCC | 19 | 10381379 |
| ICAM1-sg5 | ICAM1 | NM_000201 | GCTCATCCACTCGATTAAAG | 19 | 10381327 |
| ICAM1-sg6 | ICAM1 | NM_000201 | GGGAGCCCGGGGAGGATTCC | 19 | 10381285 |
| ICAM1-sg7 | ICAM1 | NM_000201 | ACGTCCACACCTAGCTGACA | 19 | 10381235 |
| ICAM1-sg8 | ICAM1 | NM_000201 | ATCCCTCAGTGGAGGGAGCC | 19 | 10381272 |
| SELE-sg1 | SELE | NM_000450 | AAGCAATCCCTCCTATAAAA | 1 | 169703243 |
| SELE-sg2 | SELE | NM_000450 | AATATCCTCCTATTATTCAC | 1 | 169703266 |
| SELE-sg3 | SELE | NM_000450 | ATTGTCCACATCCAGTAAAG | 1 | 169703283 |
| SELE-sg4 | SELE | NM_000450 | GAAAGTTTTTGGATGCCATT | 1 | 169703314 |
| SELE-sg5 | SELE | NM_000450 | GATATTCCCGGGAAAGTTTT | 1 | 169703325 |
| SELE-sg6 | SELE | NM_000450 | GCATATACGATATAAAGGCA | 1 | 169703410 |
| SELE-sg7 | SELE | NM_000450 | ATTAGAATTTCAGAAACAGA | 1 | 169703548 |
| SELE-sg8 | SELE | NM_000450 | GATTTCCTCTTTACTGGATG | 1 | 169703291 |
| KEL-sg1 | KEL | NM_000420 | CAGCTTCTCAGGGGAGAAGA | 7 | 142659512 |
| KEL-sg2 | KEL | NM_000420 | GACCAAGGGCAAGATTGCTT | 7 | 142659553 |
| KEL-sg3 | KEL | NM_000420 | AATACAGAAGAAATGAGAGA | 7 | 142659638 |
| KEL-sg4 | KEL | NM_000420 | GGGGAGCACCAGACCGACAA | 7 | 142659723 |
| KEL-sg5 | KEL | NM_000420 | ACACTAAACCTTTGTCGGTC | 7 | 142659734 |
| KEL-sg6 | KEL | NM_000420 | GAGTCACAGTGCAAGACAAA | 7 | 142659580 |

| | | | | | |
|---|---|---|---|---|---|
| KEL-sg7 | KEL | NM_000420 | TTAGAAATAAAGGAACTTCA | 7 | 142659606 |
| KEL-sg8 | KEL | NM_000420 | GTCTTTGGCTTTGTTTGCCT | 7 | 142660016 |

## A.2 Table 2: PCR and sequencing primers, gblock sequences

| Name | Sequence (5' - 3') |
|---|---|
| **p300 core domain gBlock sequences** | |
| C-terminal insertion 5' fragment 1 (replace VP64) | aagaggaaggtggcgggaggtggaagcggaggaATTTTCAAACCAGAA GAACTACGACAGGCACTGATGCCAACTTTGGAGGCAC TTTACCGTCAGGATCCAGAATCCCTTCCCTTTCGTCAA CCTGTGGACCCTCAGCTTTTAGGAATCCCTGATTACTTT GATATTGTGAAGAGCCCCATGGATCTTTCTACCATTAA GAGGAAGTTAGACACTGGACAGTATCAGGAGCCCTGG CAGTATGTCGATGATATTTGGCTTATGTTCAATAATGC CTGGTTATATAACCGGAAAACATCACGGGTATACAAA TACTGCTCCAAGCTCTCTGAGGTCTTTGAACAAGAAAT TGACCCAGTGATGCAAAGCCTTGGATACTGTTGTGGCA GAAAGTTGGAGTTCTCTCCACAGACACTGTGTTGCTAC GGCAAACAGTTGTGCACAATACCTCGTGATGCCACTTA TTACAGTTACCAGAACAGGTATCATTTCTGTGAGAAGT GTTTCAATGAGATCCAAGGGGAGAGCGTTTCTTTGGGG GATGACC |
| C-terminal insertion 5' fragment 2 (after VP64) | gacctggacatgctgggaggtggaagcggaggaggaggtggaagcggaggaATTT TCAAACCAGAAGAACTACGACAGGCACTGATGCCAAC TTTGGAGGCACTTTACCGTCAGGATCCAGAATCCCTTC CCTTTCGTCAACCTGTGGACCCTCAGCTTTTAGGAATC CCTGATTACTTTGATATTGTGAAGAGCCCCATGGATCT TTCTACCATTAAGAGGAAGTTAGACACTGGACAGTATC AGGAGCCCTGGCAGTATGTCGATGATATTTGGCTTATG TTCAATAATGCCTGGTTATATAACCGGAAAACATCACG GGTATACAAATACTGCTCCAAGCTCTCTGAGGTCTTTG AACAAGAAATTGACCCAGTGATGCAAAGCCTTGGATA CTGTTGTGGCAGAAAGTTGGAGTTCTCTCCACAGACAC TGTGTTGCTACGGCAAACAGTTGTGCACAATACCTCGT GATGCCACTTATTACAGTTACCAGAACAGGTATCATTT CTGTGAGAAGTGTTTCAATGAGATCCAAGGGGAGAGC GTTTCTTTGGGGGATGACC |

| | |
|---|---|
| C-terminal insertion common 3' fragment | TCTTTGGGGGATGACCCTTCCCAGCCTCAAACTACAAT<br>AAATAAAGAACAATTTTCCAAGAGAAAAAATGACACA<br>CTGGATCCTGAACTGTTTGTTGAATGTACAGAGTGCGG<br>AAGAAAGATGCATCAGATCTGTGTCCTTCACCATGAGA<br>TCATCTGGCCTGCTGGATTCGTCTGTGATGGCTGTTTA<br>AAGAAAAGTGCACGAACTAGGAAAGAAAATAAGTTTT<br>CTGCTAAAAGGTTGCCATCTACCAGACTTGGCACCTTT<br>CTTGAGAATCGTGTGAATGACTTTCTGAGGCGACAGAA<br>TCACCCTGAGTCAGGAGAGGTCACTGTTAGAGTAGTTC<br>ATGCTTCTGACAAAACCGTGGAAGTAAAACCAGGCAT<br>GAAAGCAAGGTTTGTGGACAGTGGAGAGATGGCAGAA<br>TCCTTTCCATACCGAACCAAAGCCCTCTTTGCCTTTGA<br>AGAAATTGATGGTGTTGACCTGTGCTTCTTTGGCATGC<br>ATGTTCAAGAGTATGGCTCTGACTGCCCTCCACCCAAC<br>CAGAGGAGAGTATACATATCTTACCTCGATAGTGTTCA<br>TTTCTTCCGTCCTAAATGCTTGAGGACTGCAGTCTATC<br>ATGAAATCCTAATTGGATATTTAGAATATGTCAAGAAA<br>TTAGGTTACACAACAGGGCATATTTGGGCATGTCCACC<br>AAGTGAGGGAGATGATTATATCTTCCATTGCCATCCTC<br>CTGACCAGAAGATACCCAAGCCCAAGCGACTGCAGGA<br>ATGGTACAAAAAAATGCTTGACAAGGCTGTATCAGAG<br>CGTATTGTCCATGACTACAAGGATATTTTTAAACAAGC<br>TACTGAAGATAGATTAACAAGTGCAAAGGAATTGCCT<br>TATTTCGAGGGTGATTTCTGGCCCAATGTTCTGGAAGA<br>AAGCATTAAGGAACTGGAACAGGAGGAAGAAGAGAG<br>AAAACGAGAGGAAAACACCAGCAATGAAAGCACAGA<br>TGTGACCAAGGGAGACAGCAAAAATGCTAAAAAGAAG<br>AATAATAAGAAAACCAGCAAAAATAAGAGCAGCCTGA<br>GTAGGGGCAACAAGAAGAAACCCGGGATGCCCAATGT<br>ATCTAACGACCTCTCACAGAAACTATATGCCACCATGG<br>AGAAGCATAAAGAGGTCTTCTTTGTGATCCGCCTCATT<br>GCTGGCCCTGCTGCCAACTCCCTGCCTCCCATTGTTGA<br>TCCTGATCCTCTCATCCCCTGCGATCTGATGGATGGTC<br>GGGATGCGTTTCTCACGCTGGCAAGGGACAAGCACCT<br>GGAGTTCTCTTCACTCCGAAGAGCCCAGTGGTCCACCA<br>TGTGCATGCTGGTGGAGCTGCACACGCAGAGCCAGGA<br>CGAGGGCAGAGGAAGTCTCCTAACATGCGGTGACGTG<br>GAGGAGAATCCTGGCCCAggaggtatggcttcaaactttactcagtTCT<br>AGAtgtctcctGTACATGAGgaattccgatatcaagcttatcgg |

| | |
|---|---|
| p300 core full | tcgtggGAAGCTTGGggccaccatgATTTTCAAACCAGAAGAA<br>CTACGACAGGCACTGATGCCAACTTTGGAGGCACTTTA<br>CCGTCAGGATCCAGAATCCCTTCCCTTTCGTCAACCTG<br>TGGACCCTCAGCTTTTAGGAATCCCTGATTACTTTGAT<br>ATTGTGAAGAGCCCCATGGATCTTTCTACCATTAAGAG<br>GAAGTTAGACACTGGACAGTATCAGGAGCCCTGGCAG<br>TATGTCGATGATATTTGGCTTATGTTCAATAATGCCTG<br>GTTATATAACCGGAAAACATCACGGGTATACAAATAC<br>TGCTCCAAGCTCTCTGAGGTCTTTGAACAAGAAATTGA<br>CCCAGTGATGCAAAGCCTTGGATACTGTTGTGGCAGAA<br>AGTTGGAGTTCTCTCCACAGACACTGTGTTGCTACGGC<br>AAACAGTTGTGCACAATACCTCGTGATGCCACTTATTA<br>CAGTTACCAGAACAGGTATCATTTCTGTGAGAAGTGTT<br>TCAATGAGATCCAAGGGGAGAGCGTTTCTTTGGGGGA<br>TGACCCTTCCCAGCCTCAAACTACAATAAATAAAGAA<br>CAATTTTCCAAGAGAAAAAATGACACACTGGATCCTG<br>AACTGTTTGTTGAATGTACAGAGTGCGGAAGAAAGAT<br>GCATCAGATCTGTGTCCTTCACCATGAGATCATCTGGC<br>CTGCTGGATTCGTCTGTGATGGCTGTTTAAAGAAAAGT<br>GCACGAACTAGGAAAGAAAATAAGTTTTCTGCTAAAA<br>GGTTGCCATCTACCAGACTTGGCACCTTTCTAGAGAAT<br>CGTGTGAATGACTTTCTGAGGCGACAGAATCACCCTGA<br>GTCAGGAGAGGTCACTGTTAGAGTAGTTCATGCTTCTG<br>ACAAAACCGTGGAAGTAAAACCAGGCATGAAAGCAAG<br>GTTTGTGGACAGTGGAGAGATGGCAGAATCCTTTCCAT<br>ACCGAACCAAAGCCCTCTTTGCCTTTGAAGAAATTGAT<br>GGTGTTGACCTGTGCTTCTTTGGCATGCATGTTCAAGA<br>GTATGGCTCTGACTGCCCTCCACCCAACCAGAGGAGA<br>GTATACATATCTTACCTCGATAGTGTTCATTTCTTCCGT<br>CCTAAATGCTTGAGGACTGCAGTCTATCATGAAATCCT<br>AATTGGATATTTAGAATATGTCAAGAAATTAGGTTACA<br>CAACAGGGCATATTTGGGCATGTCCACCAAGTGAGGG<br>AGATGATTATATCTTCCATTGCCATCCTCCTGACCAGA<br>AGATACCCAAGCCCAAGCGACTGCAGGAATGGTACAA<br>AAAAATGCTTGACAAGGCTGTATCAGAGCGTATTGTCC<br>ATGACTACAAGGATATTTTTAAACAAGCTACTGAAGAT<br>AGATTAACAAGTGCAAAGGAATTGCCTTATTTCGAGG<br>GTGATTTCTGGCCCAATGTTCTGGAAGAAAGCATTAAG<br>GAACTGGAACAGGAGGAAGAAGAGAGAAAACGAGAG<br>GAAAACACCAGCAATGAAAGCACAGATGTGACCAAGG<br>GAGACAGCAAAAATGCTAAAAAGAAGAATAATAAGA<br>AAACCAGCAAAAATAAGAGCAGCCTGAGTAGGGGCAA<br>CAAGAAGAAACCCGGGATGCCCAATGTATCTAACGAC<br>CTCTCACAGAAACTATATGCCACCATGGAGAAGCATA<br>AAGAGGTCTTCTTTGTGATCCGCCTCATTGCTGGCCCT |

**PCR primers for VP64 amplification**

| | |
|---|---|
| VP64 Forward | TCGTGGGAAGCTTGGGGCCACCATGGACGCATTGGAC |
| VP64 Reverse | TCATGGTGGCCGTACATCCAGAACCTCCACCCAGCATGTCCAGGTC |

**Sequencing primers for dCas9-activator constructs**

| | |
|---|---|
| p300 F1 | TGCCTCCCATTGTTGATCCT |
| p300 F2 | TTGTGAAGAGCCCCATGGAT |
| p300 F3 | CCCTTCCCAGCCTCAAACTA |
| p300 F4 | TGCTTCTGACAAAACCGTGG |
| N term dCas9 F1 | TCTCAAGCCTCAGACAGTGG |
| N term dCas9 R1 | TGTACTCGTCGGTGATCACG |
| C term dCas9 F1 | ACTTTGACACCACCATCGAC |
| C term dCas9 R1 | CGTCACCGCATGTTAGGAGA |
| post T2A F1 | AGTCTCCTAACATGCGGTGA |
| post T2A R1 | CACACCGGCCTTATTCCAAG |

**Primers for initial amplification of gRNA library**

| | |
|---|---|
| 77-mer_U1 | GCAGATGGCTCTTTGTCCTA |
| 77-mer_L1 | GCGACGAGAAGACTAAAAC |

**Primers for q-RT-PCR**

| | |
|---|---|
| SLC4A1 F | GGGCTCAGATCACCGTAGAC |
| SLC4A1 R | AGGAGGACAGTACCCTTGGT |
| RHD F | AGGATCAAAAGGGGCTCGTG |
| RHD R | TGTTCATGTGGTAGTCTGTGTTG |
| SELE F | GCCTGCAATGTGGTTGAGTG |
| SELE R | ATTCATGTAGCCTCGCTCGG |
| KEL F | CTGATAAGCAGGCTCCACCC |
| KEL R | CTGGAGTGCTCTCTTGGCTC |

**Primers for amplification of RTN4R family ectodomains**

| | |
|---|---|
| RTN4R F | CCAAGTTTAAACTGCGGCCGCCACCATGAAGAGGGCGTCCGCTG |
| RTN4R R | TGGAGGTCGACGGCGCGGGCGCGCCTGAGCCTTCTGAGTCACCAG |
| RTN4RL1 F | CCAAGTTTAAACTGCGGCCGCCACCATGCTTCGCAAAGGGTGCT |
| RTN4RL1 R | TGGAGGTCGACGGCGCGGGCGCGCCGCTGGGGGCACG |
| RTN4RL2 F | CCAAGTTTAAACTGCGGCCGCCACCATGCTGCCCGGGCTCA |
| RTN4RL2 R | TGGAGGTCGACGGCGCGGGCGCGCCGGAGTCCGGGGGCGCCTGG |

| | |
|---|---|
| **Primers for Illumina library preparation and sequencing** | |
| **First PCR** | |
| SAMlibrary-HiSeq_50bp-F1 | ACACTCTTTCCCTACACGACGCTCTTCCGATATATCTTGTGGAAAGGACGAAACA |
| SAMlibrary-HiSeq_50bp-R1 | TCGGCATTCCTGCTGAACCGCTCTTCCGATCTCAGACTGCCTTGGGAAAAG |
| **Second PCR** | |
| HiSeq-PE 1.0 | AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T |
| HiSeq-iPCRTag-11mer | CAAGCAGAAGACGGCATACGAGATNNNNNNNNNNNGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T |
| * indicates phosphorothioate | |
| **Sequencing primer** | |
| U6-SAMlibrary-Illumina-seq (19bp-SE) | TATATCTTGTGGAAAGGACGAAACACCG |

## A.3 Table 3: Primary and secondary antibodies

**Primary monoclonal antibodies**

| Clone | Source | Cat No. | Antigen gene symbol | Antigen gene name |
|---|---|---|---|---|
| BRAC18 | IBGRL | | SLC4A1 | Solute Carrier Family 4 Member 1 |
| BRAD2 | IBGRL | | RHD | Rh blood group D antigen |
| BRIC18 | IBGRL | | KEL | Kell blood group glycoprotein |
| P2X7-L4 | DSHB | | P2RX7 | P2X purinoceptor 7 |
| P3D1 | DSHB | | ENG | Endoglin |
| TS2/18.1.1 | DSHB | | CD2 | Cluster of differentiation 2 |
| P3C4 | DSHB | | VCAM1 | Vascular cell adhesion molecule 1 |
| HB#7/HC7 | DSHB | | PROM1 | Prominin 1 |
| MEM-150 | Abcam | ab26012 | SEMA7A | Semaphorin 7A |
| OX-104 | Bioline | 329202 | CD200 | Cluster of differentiation 200 |
| P2A4 | DSHB | | ICAM1 | Intercellular Adhesion Molecule 1 |
| 1.2B6 | Santa Cruz | sc-18852 | SELE | E-selectin |
| OX68 | from hybridoma | | rCD4 (d3/4) | Rat CD4 (Domains 3 and 4) |
| LM609 | Merck Millipore | MAB1976 | avb3 | Integrin avb3 |

IBGRL - International Blood Group Reference Laboratory
DSHB - Developmental Studies Hybridoma Bank

**Secondary antibodies and other flow cytometry reagents**

| **Antibody description** | **Source** | **Cat No.** |
|---|---|---|
| anti-mouse-PE | Abcam | ab7002 |
| anti-mouse-AP | Sigma Aldrich | A4656 |
| Streptavidin-PE | Biolegend | 405204 |
| Annexin V-FITC | eBioscience | BMS306F1 |