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 μM) 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 μM) 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 μM) 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 αvβ3. 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 αvβ3, 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-αvβ3, 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

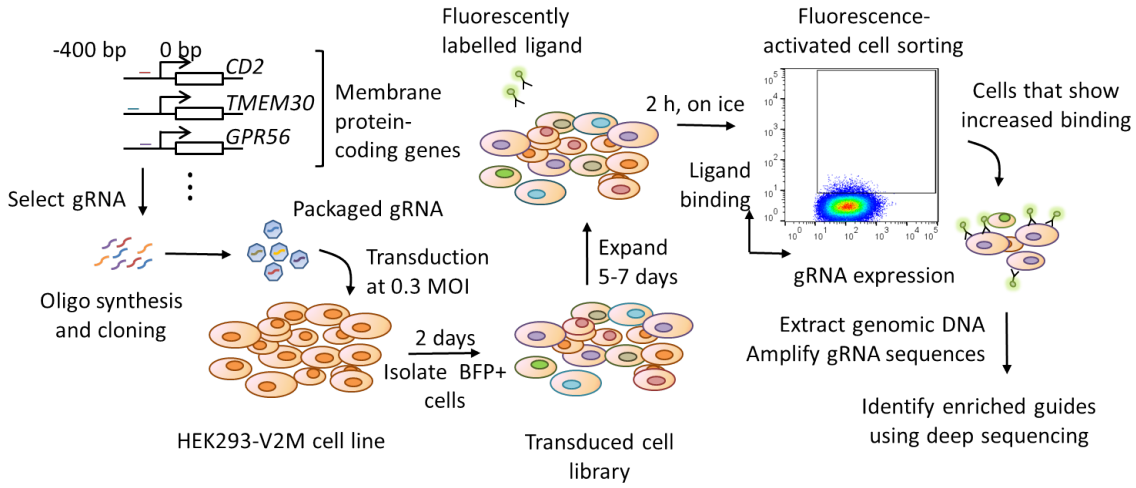

Figure 4.1 Schematic of CRISPRa extracellular interaction screening. A CRISPRa gRNA library targeting genes encoding membrane proteins was designed, cloned and packaged into lentiviruses for transduction. Transduction of a cell line constitutively expressing dCas9-activators at a low multiplicity of 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-αvβ3	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 αvβ3 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%

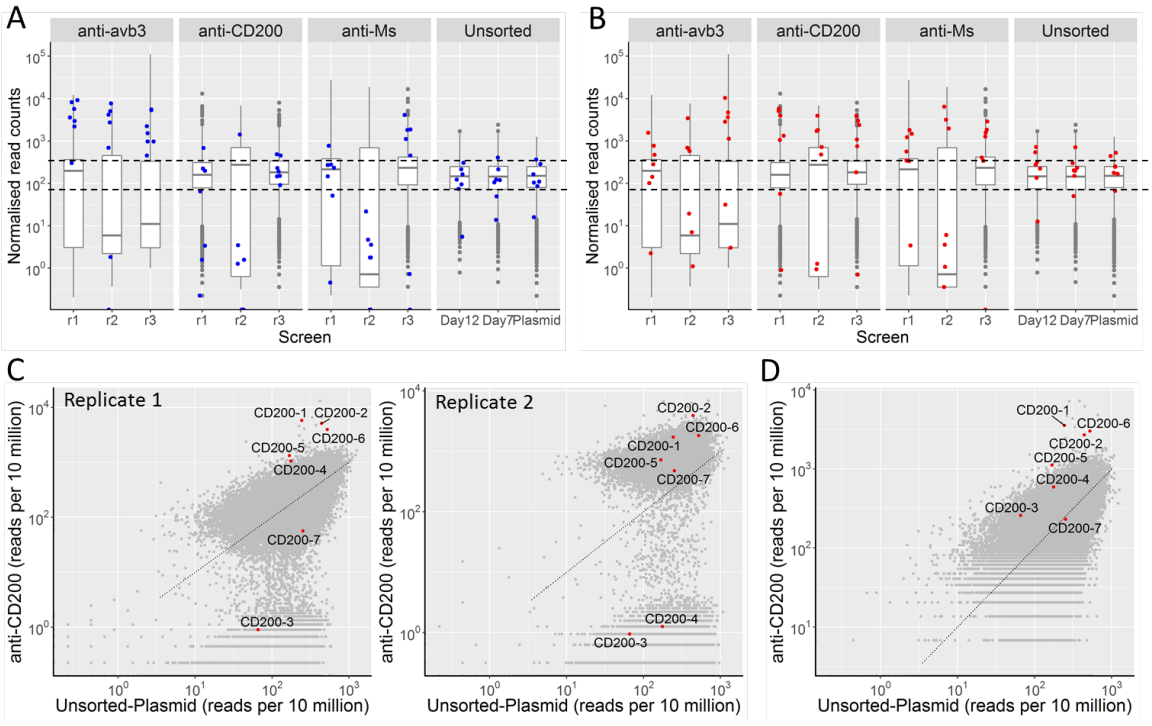

Figure 4.2 Sufficient library coverage is required for robust receptor identification with CRISPRa screening. A) ITGB3-targeting guides (blue dots) are enriched specifically in cell populations sorted for binding to anti-integrin $\alpha v\beta 3$. Box plots of normalised gRNA abundances are shown for 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.

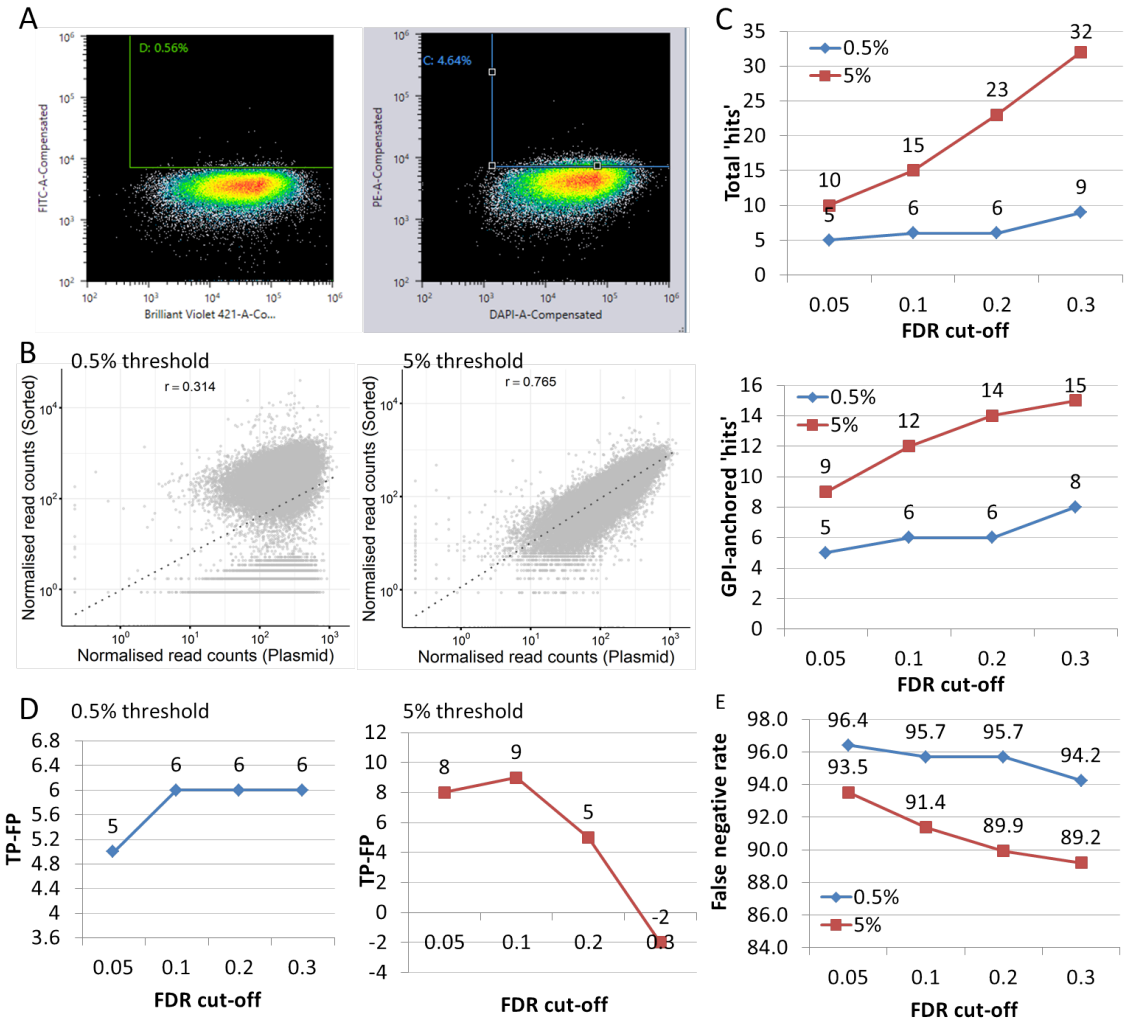

Figure 4.3 A 5% sort threshold during CRISPRa screening reduces false negatives whilst an FDR cut-off of 0.1 limits false positives A) Sort gates using during screening at a 0.5% (left) and 5% threshold (right). The percentage of cells in the gate fluctuates during sorting hence numbers in the image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

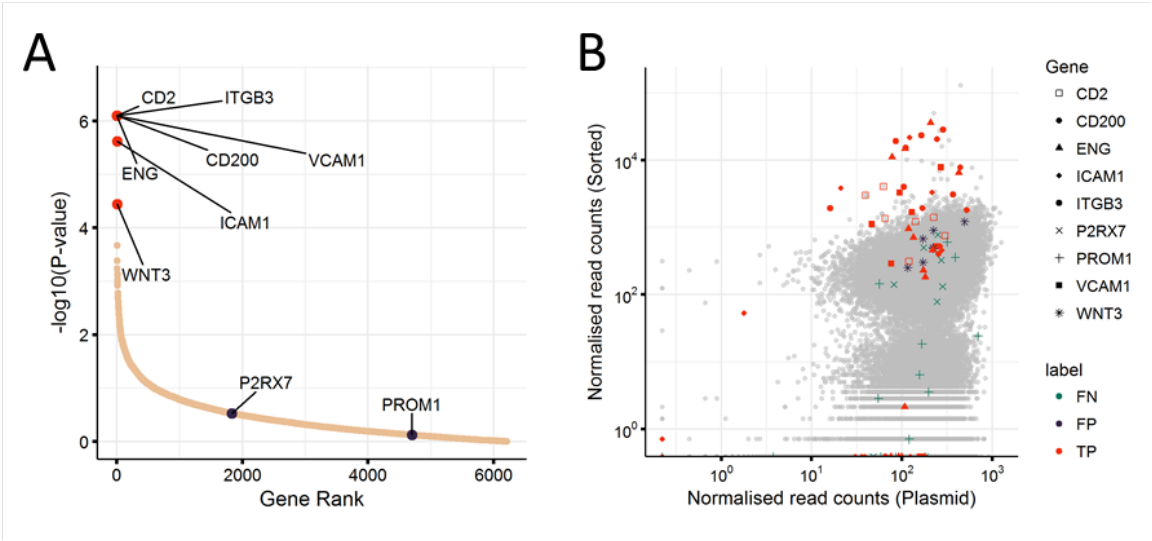

Figure 4.4 CRISPRa screening simultaneously identifies multiple targets to a pool of antibodies. A) Transduced cell libraries were sorted to isolate cells binding to an equimolar pool of eight antibodies, and gRNA abundance quantified by next-generation sequencing. Enrichment analysis indicated that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

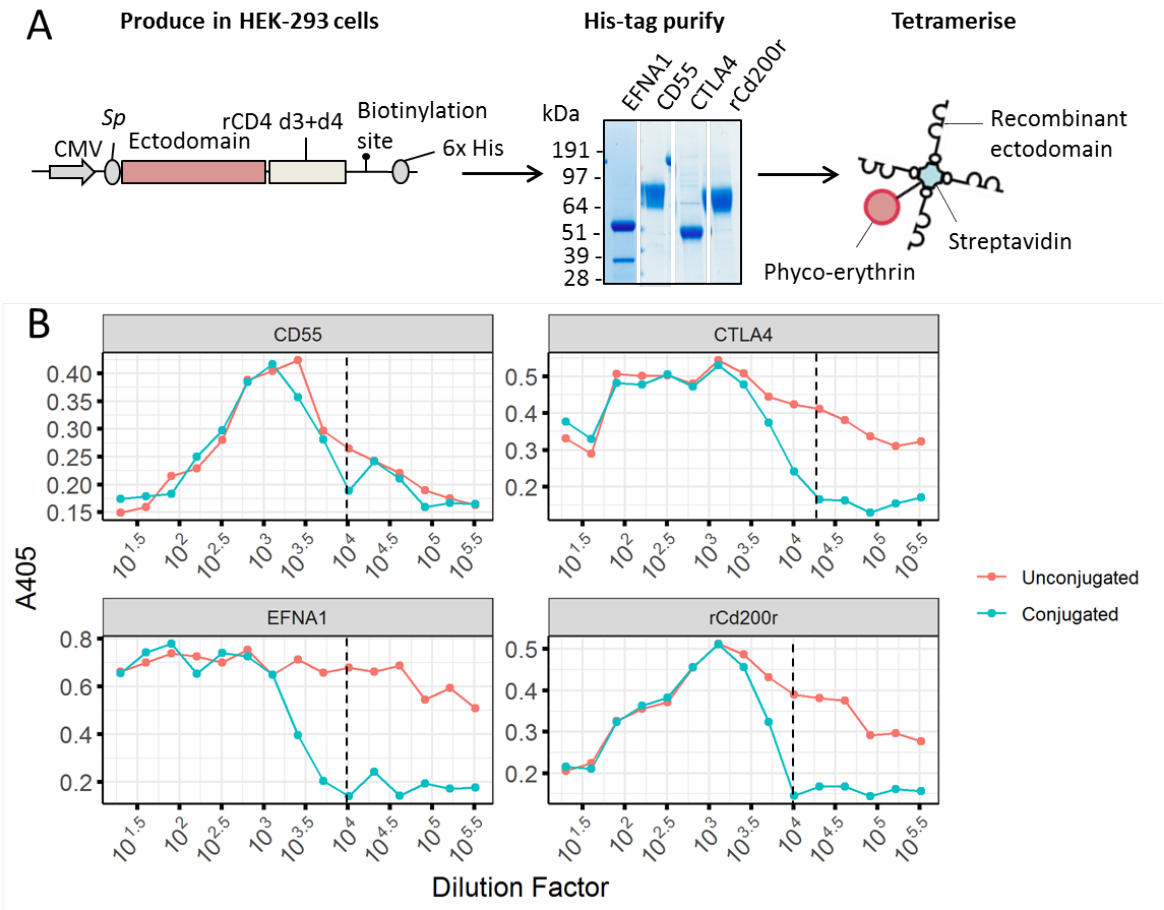

Figure 4.5 Highly avid tetramers are produced from recombinant biotinylated ectodomains and normalised for use in CRISPRa screening A) Schematic showing the production of tetramers from purified biotinylated monomers containing the full length ectodomain of four cell surface ligands. A construct encoding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

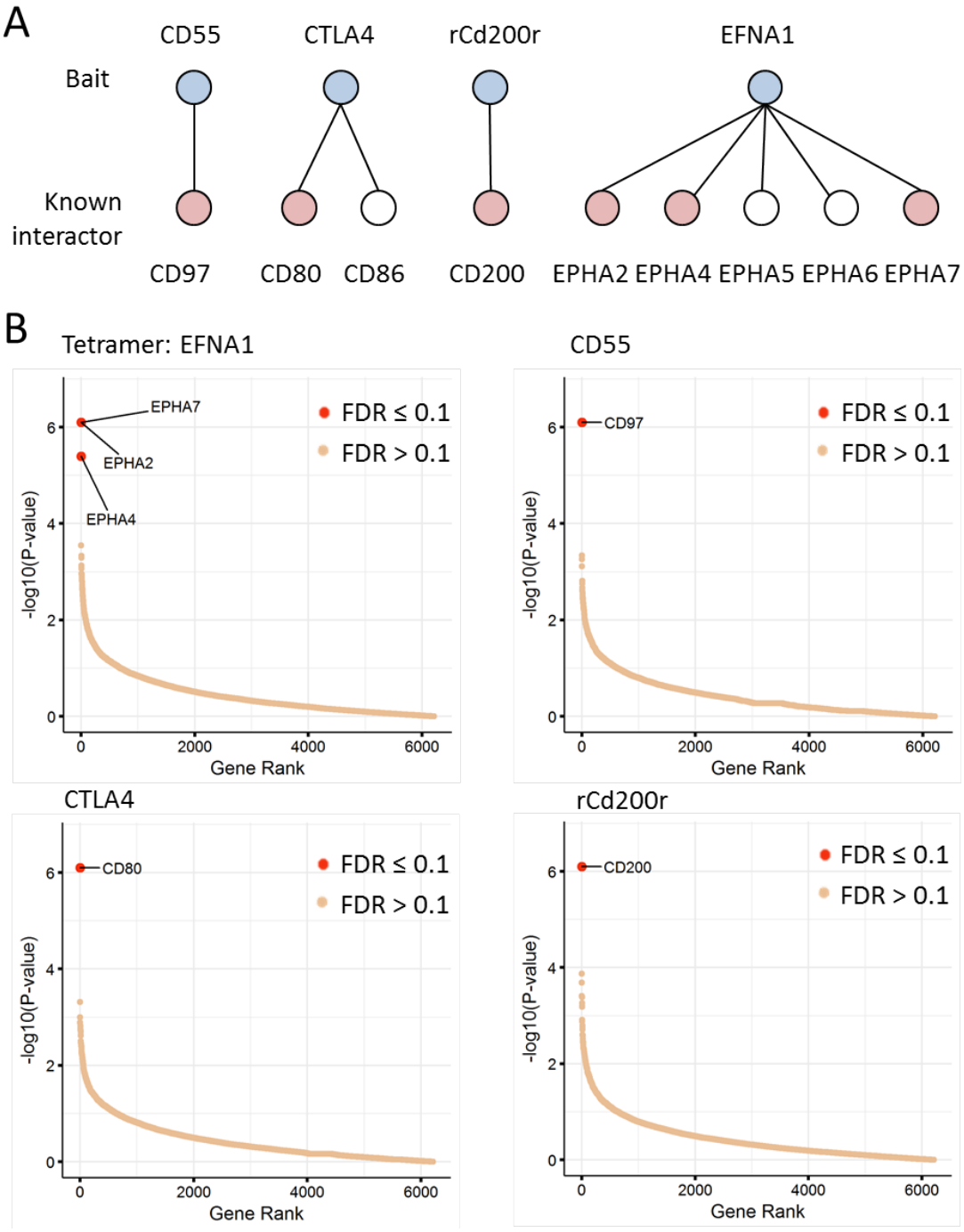

Figure 4.6 CRISPRa screening unambiguously identifies low-affinity endogenous interactions A) CRISPRa screening identifies six out of nine reported interactions involving CD55, CTLA4, rCd200 and EFNA1. Blue circles represent cell surface ligands used as tetramers for screening, pink circles represent 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 (μM)	Reference	Enriched in screen
CD55 - CD97	86 ± 1	Lin et al. (2001)	Y
CTLA4 - CD86	2.6	Collins et al. (2002)	
Cd200r - CD200	0.59 ± 0.07	Wright et al. (2003)	Y
EFNA1 - EPHA2	0.58 ± 0.24	Lema Tomé et al. (2012)	Y
CTLA4 - CD80	0.42 ± 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_Ds of above 1 μM. CRISPRa screening identified the weakest interaction (CD55-CD97) but failed to detect the second weakest (CTLA4-CD86). The K_Ds 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.