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

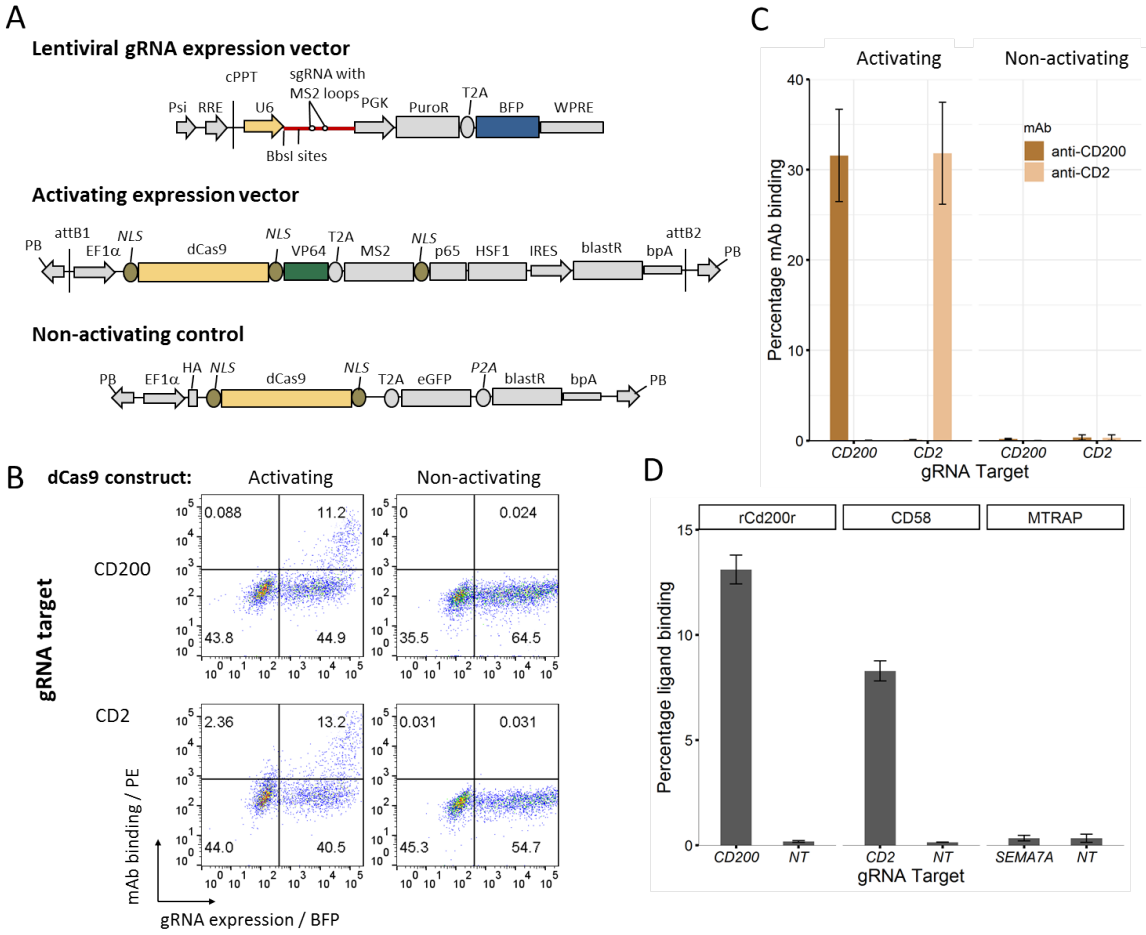

Figure 3.1 CRISPR activation of CD200 and CD2 is specific and allows ligand binding. A) Schematic of expression vectors used. gRNAs were expressed under a U6 promoter along with BFP as a fluorescent marker. dCas9-VP64 and MS2-p65-HSF1 fusion proteins were expressed as a single transcript separated by 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 - λ recombination attachment sites, Ef1α - Human elongation factor-1 α promoter, NLS - Nuclear Localisation Signal, IRES - Internal Ribosomal Entry Site, bpA - Bovine growth hormone polyadenylation site. Psi - Viral packaging signal sequence, RRE - Rev response element, cPPT - Central Polypurine tract, PGK - Phosphoglycerate kinase promoter, WPRE - Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element. B) Flow plots showing clear upregulation of CD200 and CD2 in cells co-transfected with gRNA and dCas9-activators but not in cells co-transfected with gRNA and a non-activating control. Surface expression was quantified by antibody staining of CD200 and CD2 respectively. gRNA expression indicated by expression of BFP C) Quantification of CD200 and CD2 expression shows no cross-reactivity with either antibody and is dependent on transactivators expressed by the activating construct. mAb binding was calculated as a percentage of total BFP+ cells D) CRISPRa of CD200 and CD2 is sufficient to induce gain-of-binding of known ligands rCd200r and CD58. Soluble recombinant ectodomains of rCd200r and CD58 were produced as highly avid FLAG-tagged pentameric proteins and detected by fluorescently-labelled anti-FLAG antibody. Data points in C) and D) represent mean $\pm$ s.e.m; n= 3.

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

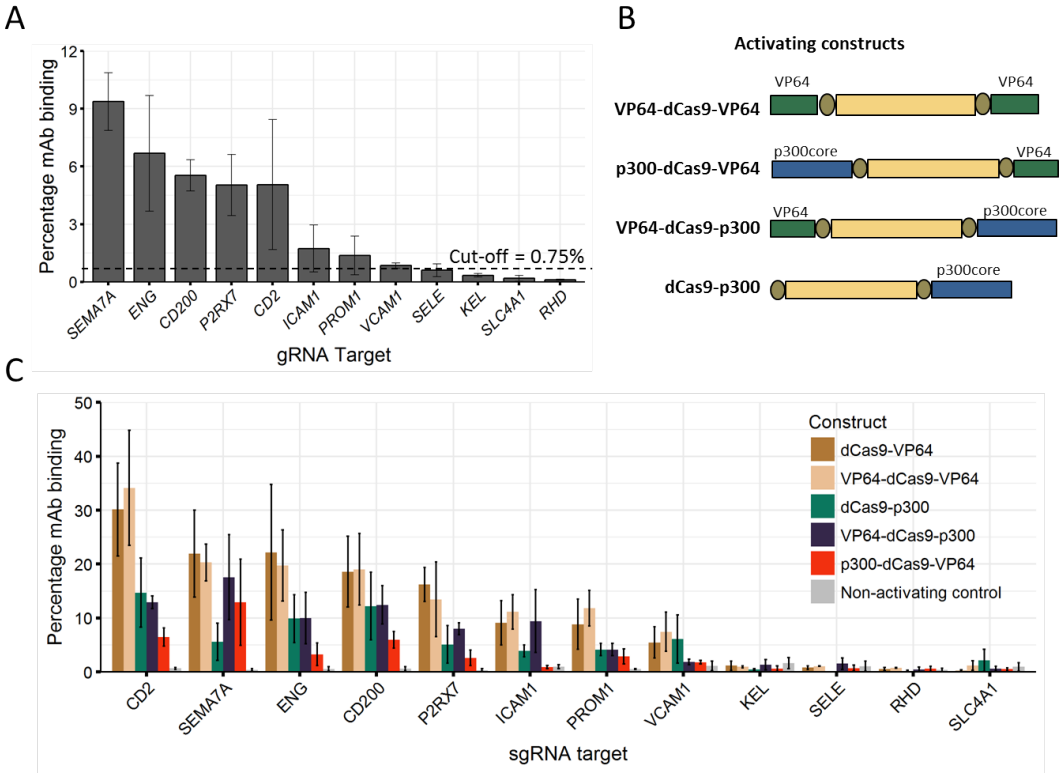

Figure 3.2 Reduced CRISPRa efficiency using dCas9-activators with an additional p300 HAT domain. A) Quantification of surface expression of 12 cell surface proteins after CRISPRa. Cells were transduced with lentiviruses carrying a pools of 8 gRNAs targeting each gene, before being transfected with dCas9-VP64 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.

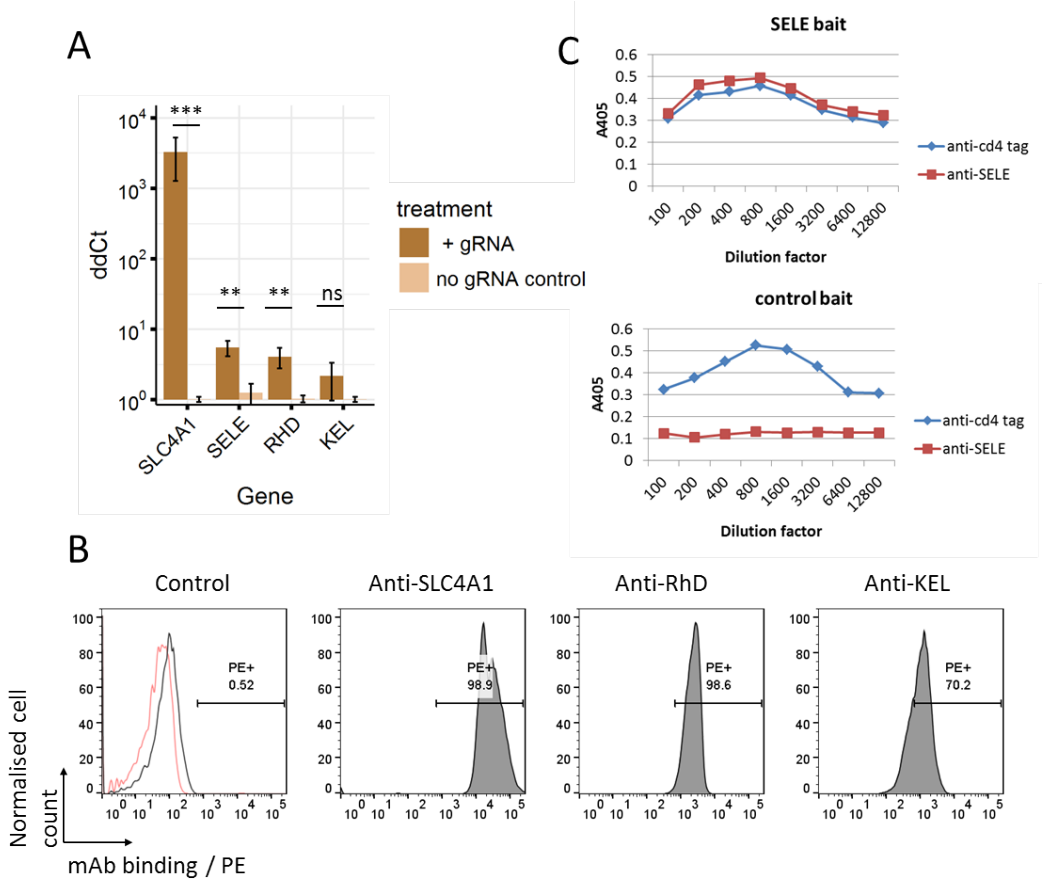

Figure 3.3 Upregulation of mRNA transcripts does not necessarily lead to with an increase in surface protein levels. A) Quantification of mRNA abundance by qRT-PCR of indicated target genes in cells 48 h post co-transfection with dCas9-VP64 and either targeting gRNA (+) or no gRNA control. Transcript 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 3rd and 4th Ig domains of rat Cd4. Recombinant SELE and Cd200r are captured on streptavidin coated plates and detected by incubation with anti-SELE or anti-rCd4, followed by a secondary anti-mouse antibody conjugated to alkaline phosphatase. Binding is quantified by absorption at 405 nm of a hydrolysis product of a phosphatase substrate. Binding of anti-rCd4 reflects relative amounts of SELE or Cd200r captured on the plate.

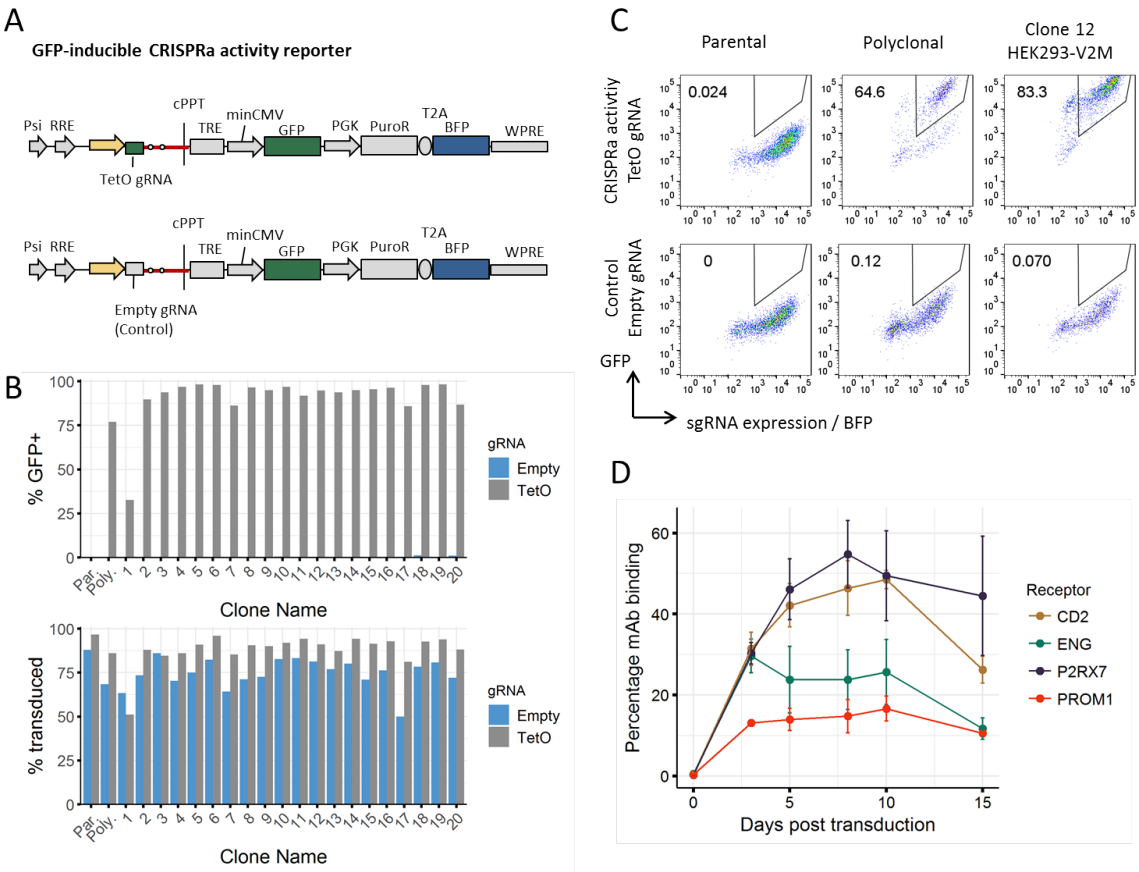

Figure 3.4 A cell line constitutively expressing dCas9-activators allows sustained CRISPRa overexpression of cell surface proteins A) Schematic of GFP-inducible CRISPRa activity reporter system. In both vectors, GFP expression is controlled by a TetO responsive element (TRE) consisting of several TetO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

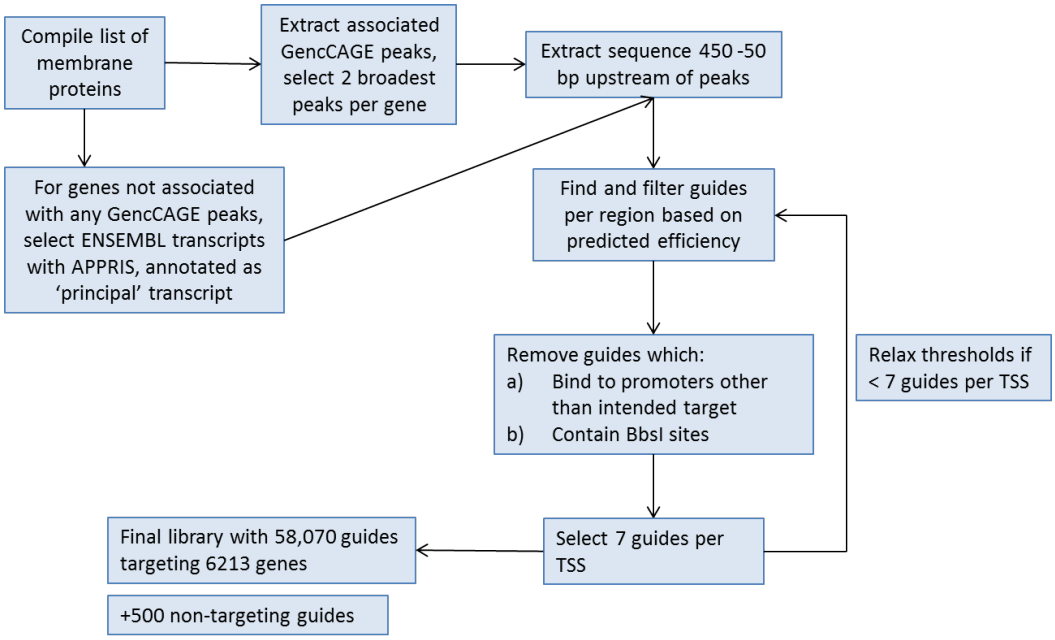

Figure 3.5 Schematic showing workflow of membrane protein gRNA library design A list of putative membrane proteins were compiled from public databases, mass spectrometry and bioinformatic studies. Unique TSSs were extracted based on a combination of Gencode gene models and CAGE-seq peaks (GencCAGE peaks);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

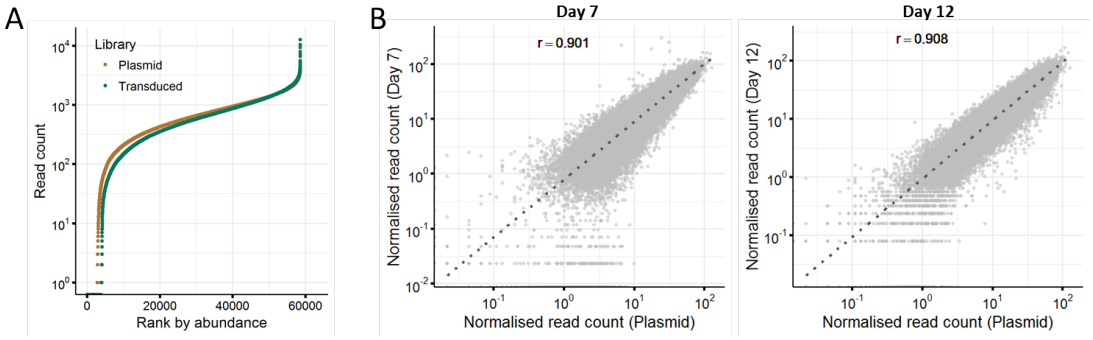

Figure 3.6 Quality controls indicate complete and fairly even representation of guide sequences in the gRNA library A) Ranked gRNA abundance in the plasmid library (brown) and cells transduced with the CRISPRa lentiviral library and cultured for seven days (green) as determined by raw read counts from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

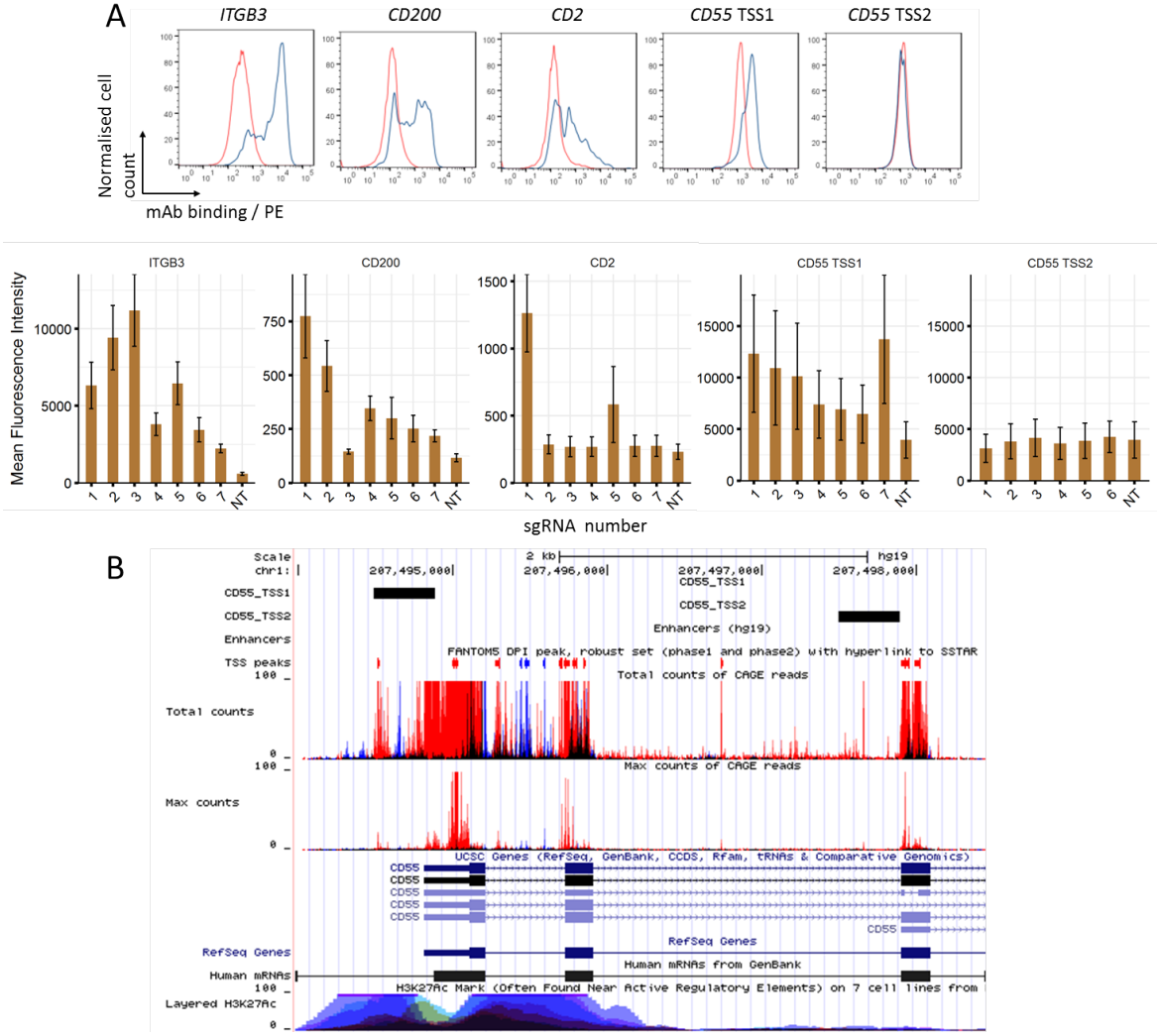

Figure 3.7 Small scale validation of individual guides show that majority of guides are active A) (Top) mAb binding histograms of HEK293-V2M cells transfected with the individual gRNAs targeting the promoter region of the named receptor genes (blue traces) compared to control non-targeting gRNAs (red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