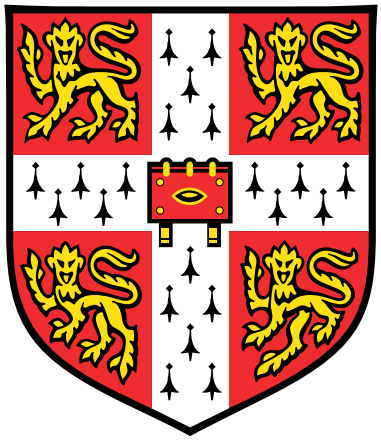
Genome-scale identification of cellular pathways required for cell surface recognition



Sumana Sharma

Wellcome Trust Sanger Institute

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Trinity College

December 2017

DECLARATION

I hereby declare that the contents of this thesis 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 thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. This thesis does not exceed the word limit set by the Faculty of Biology.

Sumana Sharma
December 2017

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my advisor Dr. Gavin Wright for the guidance, motivation, patience and the encouraging words all through my PhD. Thank you mostly for encouraging me to appreciate technology development related research! Thank you also to the members of my thesis committee, Dr. Julian Rayner and Dr. Kosuke Yusa for their insightful comments and encouragement, and also for providing a different perspective to my research.

Deepest gratitude to the members of the Wright Lab who have been a cherished combination of colleagues and friends. Special thank you to Josefin, who has provided instrumental support to my research throughout the PhD; Nicole, Shan, Laura and Kirsty for readily sharing their reagents and technical abilities; and Cecile and Enrica whose openness and scientific insight I always admire. I would also like to thank the members of the flow cytometry core-facility (Bee Ling Ng, Jennifer Grahm, Chistopher Hall and Sam Thompson) for their flow cytometry training and advice; Mandy Sanders for help with sequencing sample submission; and Sanger Institute Core Sequencing facility for sequencing.

Most of all, I would like to thank my parents, whose encouragement and love has made possible my journey from the hills of Rohote to Cambridge! Also, thanks to my beloved siblings, Rashmi and Prashant, for their guidance and unwavering support. Thank you to James- you have been amazingly patient and always been there to support me. Thanks to my friends, John, Katie and Tom who have made my time during my PhD a very enjoyable one.

ABSTRACT

A range of biochemically diverse molecules located in the plasma membrane—such as proteins, glycans, and lipids—mediate cellular recognition events, initiation of signalling pathways, and the regulation of processes important for the normal development and function of multicellular organisms. Interactions mediated by cell surface receptors can be challenging to detect in biochemical assays, because they are often highly transient, and membrane-embedded receptors are difficult to solubilise in their native conformation. The biochemical features of low-affinity extracellular protein interactions have therefore necessitated the development of bespoke methods to detect them.

Here, I develop a genome-scale cell-based genetic screening approach using CRISPR-Cas9 knockout technology that reveals cellular pathways required for specific cell surface recognition events. Using a panel of high-affinity monoclonal antibodies, I first establish a method from which I identify not only the direct receptor but also other required gene products, such as co-receptors, post-translational modifications, and transcription factors contributing to antigen expression and subsequent antibody-antigen recognition on the surface of cells. I next adapt this method to identify cellular factors required for receptor interactions for a panel of recombinant proteins corresponding to the ectodomains of cell surface proteins to the endogenous surface receptors present on a range of cell lines. In addition to finding general cellular features recognised by many ectodomains, I also identify direct interaction partners of recombinant protein probes on cell surfaces together with intracellular genes required for such associations.

Using this method, I identify IGF2R as a binding partner for the R2 subunit of GABAB receptors, providing a mechanism for the internalisation and regulation of GABAB receptor signalling. The results here demonstrate that this single approach can identify the molecular nature and cell biology of surface receptors without the need to make any prior assumptions regarding their biochemical properties.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Communication between cells is crucial for the normal development and functioning of a multicellular organism. Membrane-compartmentalised cells receive instructional information from their surroundings by extracellular signalling cues which are often initiated via specific binding events made by plasma-membrane embedded receptors. While some extracellular signals are received via soluble factors (e.g. secreted proteins, hormones, autacoids, and neurotransmitters), others act through direct cell-cell interactions between specific receptors exposed on the outer surface of apposing cells [1]. Understanding of the molecular basis of cellular recognition events has wide implications as a multitude of cellular processes such as differentiation, motility and proliferation of cells depend on inter-cellular communication mediated by membrane receptors. Interactions between our own cells (e.g. neural and immunological recognition) as well as between host cells and pathogens also rely on extracellular recognition and signalling events. This, together with the fact that membrane receptors are directly accessible to systematically delivered biological reagents such as monoclonal antibodies [2] and small-molecule antagonists [3], make the study of membrane receptors therapeutically attractive. In fact, cell surface receptors currently make up the largest group (44 %) of human drug targets [4]. Thus, elucidation of cellular recognition processes is of significant interest for advances in biomedical research.

1.1 Molecules mediating cell surface recognition

A wide range of biochemically diverse molecules located in the plasma membrane of cells such as proteins, carbohydrates, and lipids have the potential to mediate cell surface interactions that are required for a vast range of biological processes [5]. In this section, I will review some aspects of each of these in turn.

1.1.1 Proteins

Nearly a quarter of the human genome encodes for either secreted or membrane-bound proteins [6, 7]. An intregral membrane protein that is embedded within the lipid bilayer of the plasma membrane is intrinsically amphipathic in nature. The domain exposed to the exterior of the cell (ectodomain) is hydrophilic, often glycosylated and is responsible for binding to the ligand, whereas the transmembrane region is hydrophobic to allow the receptor to exist within the plasma membrane. The intracellular domain of membrane receptors is in direct contact with the components of the cellular cytoplasm, which allows some membrane proteins to function as 'signal transducers' to transmit extracellular signals across the membrane to influence cellular behaviour in a context (a cell or signal type) dependent manner [8].

Membrane proteins that behave as signal transducers can be broadly categorised into three groups: Ion channels, G-protein linked receptors, (GPCR) and enzyme-linked receptors. Ion channels are usually multimeric and form an aqueous pore in the plasma membrane that, upon specific perturbation, allow movement of inorganic ions in and out of the cells [9]. Voltage-gated ion channels respond to change in membrane potential whereas transmitter gated channels respond to binding of neurotransmitters [10, 11]. The second class of signal transducers, GPCRs, are one of the most diverse receptors that can transmit diverse extracellular signals ranging from peptides, lipids, neurotransmitters and nucleotides to light, Ca^{+} and odorants [12, 13, 14, 15]. There are approximately 800 identified members in the GPCR superfamily in the human genome. Despite the large diversity of GPCRs, they interact with a relatively small number of heterotrimeric (composed of α, β and γ subunits) G proteins to initiate intracellular signalling cascades [16]. GPCRs are usually characterised by the presence of seven-transmembrane regions and this includes a class of approximately 140 GPCRs called 'orphan GPCRs' whose sequence is known but the endogenous ligands to which they bind are still largely unknown [17]. The third class of signal transducers are enzyme linked receptors which, unlike the GPCRs, are usually single transmembrane proteins that, instead of recruiting a G-protein for signalling activity, usually contain intrinsic enzymatic activity in their cytoplasmic domain or associate directly with enzymes with signalling capabilities. Six different classes of enzyme linked receptors have been described: (1) receptor guanylyl cyclases, (2) receptor tyrosine phosphatases, (3) receptor serine/threonine kinases, (4)

receptor tyrosine kinases, (5) tyrosine kinase-associated receptors, and (6) histidine-kinase-associated receptors [18].

1.1.2 Glycans

The second class of molecules that mediate cellular interactions are complex carbohydrates or glycans. This is one of the most diverse classes of macromolecules found in nature. Mammalian glycomes are predominantly created out of numerous combinations of nine common sugars (Glucose, N-acetylglucosamine, Galactose, N-acetylgalactosamine, Mannone, Fucose, Glucosamine, (or isomer L-Iduronic acid), Xylose, and N-Acetylneuraminic acid). Glycans can be found displayed at cell surfaces, incorporated into the extracellular matrix and covalently attached to secreted glycoproteins. Most eukaryotic cells are surrounded with a dense coat consisting of glycans and glycoconjugates (glycocalyx) that is important not only for providing protective, organisational, and barrier functions to the cell but also for mediating cellular communication [19]. Glycans can act as direct receptors for glycan binding proteins (GBP), which can be broadly classified into two groups: lectins and glycosaminoglycan (GAG) binding proteins. The origins of lectin, a term derived from the Latin word “legere,” meaning “to select”, dates back to 1888 when Herrmann Stillmark first described the animal red blood cell agglutination properties of extracts of castor bean seeds. Subsequently, lectins were identified in almost every plant species and today the stock of known lectins has increased vastly to include those from viruses (hemagglutinins), bacteria (adhesins and toxins), invertebrates and vertebrates. Lectins are usually characterised by the presence of evolutionarily conserved “carbohydrate recognition domains” (CRDs) [20]. This is in contrast to GAG binding proteins that are evolutionarily unrelated to each other and rather than possessing a specific binding domain, rely on basic residues to mediate interaction with the negatively charged sulphated groups of GAGs. An important consideration for the interaction of GBPs with their glycan ligands is the principle of multivalency. The binding affinity of a single glycan to GBP is low; thus to achieve biologically significant interaction, glycans that bind GBPs contain a multiple repeating structure for increased avidity [21].

1.1.3 Lipids

The third class of macromolecules that mediate cellular recognition and signalling processes are lipids. Although lipids have been suggested to be as

diverse as proteins, we have a much poorer understanding of their functions. In the past, lipids were mainly studied in the context of structural components of the plasma membrane or intermediary metabolites. The initial studies delineating the pathophysiological roles of specific lipid molecules (prostaglandins and leukotrienes) first suggested a role for lipids as intracellular signalling molecules [22, 23]. Plasma membrane lipids are mainly composed of phospholipids (phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, and sphingomyelin) and cholesterol. Since the identification of inositol phospholipids as a plasma membrane lipid crucial for cellular signalling [24], much research has been devoted to understanding the mechanism by which lipid molecules carry out signalling functions in cells [25]. The cellular lipid pool also includes glycans and proteins that are covalently modified by lipids. Both plasma membrane and intracellular organelle membrane glycans are modified by lipids to form glycolipids (e.g, glyceroglycolipids, glycosphingolipids, sulfoglycosphingolipids and gangliosides), which serve both as signalling molecules, recognition sites for cell–cell interactions [26, 27, 28], and receptors of bacteria and bacterial toxins [29, 30, 31].

Similarly, proteins are also post-translationally modified with lipids (e.g, myristoylation, palmitoylation, geranylgeranylation, GPI (Glycosylphosphatidylinositol) anchored) [32]. Some of these lipid-modified proteins, specifically those with a GPI anchor, cholesterol-linked and palmitoylated proteins such as hedgehog, are particularly enriched in the specific micro-environment of the plasma membrane; these are termed ‘lipid rafts’. Lipid rafts that contain defined sets of proteins are known to be important for signal transduction processes in cells [33].

Here, I will mainly focus on extracellular protein-protein and protein-glycan interactions including those mediated by glycoproteins, glycolipids, proteoglycans and lipid modified proteins. I will first discuss some of the challenges in studying interactions mediated by membrane receptors. Next I will introduce the major methods that have been designed to address some of these challenges and discuss their applicability and their limitations. Finally, I will introduce the recent genetic loss-of-function screening approaches including those using the CRISPR-Cas9 technology and their potential for studying cell surface interactions.

1.2 The challenges of studying extracellular ligand-receptor interactions

Even though extracellular interactions mediated by membrane receptors have been recognised to be of biological and pharmacological importance, investigating such interactions remains technically challenging. The inherent biochemical properties of extracellular space exposed membrane proteins, such as their relatively low abundance (typically 10^4 to 10^5 copies per cell) and amphipathic nature causing poor solubility, make them difficult to isolate, purify, solubilise, and biochemically manipulate [34]. Advances in approaches such as (Immuno)affinity-based techniques combined with mass-spectometry (MS) [35, 36, 37], yeast two-hybrid-based methods (Y2H) [38], and in-vitro array-based technologies [39, 40], now allow interrogation of protein-protein interactions at a large scale. However, such approaches are generally considered unsuitable to study extracellular interactions as the affinities of binding between two membrane receptor proteins are usually weak (K_D in μM to mM range) with fast dissociation rate constants (half lives of the order of one second). This poses challenges in detecting such interactions in approaches that require stringent wash steps such as affinity purification and MS. Moreover, membrane proteins are usually post-translationally modified with structurally critical glycans and disulfide bonds; this limits the use of prokaryotic heterologous systems or cell-free systems that lack the oxidative environment and the cellular machinary to generate correctly folded and glycosylated recombinant membrane proteins [41, 42].

Historically, the study of interactions between glycans and proteins has also been challenging because of the high diversity of glycans and the difficulties in obtaining glycans in high quantities in homogenous form [43]. Common methods to detect the low-affinity monovalent interactions between a single carbohydrate ligand and a single binding domain of a protein, such as inhibition studies using soluble mono- and oligosaccharides, isothermal calorimetry, surface plasmon resonance (SPR), and enzyme-linked lectin assays (ELLA), require large quantities of purified glycans with precise structures, which are not always readily available [21]. Unlike proteins, glycans cannot be readily cloned as they are secondary gene products and are not encoded directly in the genome. The particularly high diversity of glycans is the result of biosynthetic enzymes (glycosyltransferases, glycosidases, sulfotransferases, etc.) that

generate them, which act together in numerous combinatorial possibilities to generate highly heterogenous glycan chains [44]. These enzymes are dynamic and respond to environmental cues to act in a context specific manner, which makes it difficult to predict the exact nature of the glycan generated by cells. Glycans have also been found to have different roles in cell cultures compared to whole organisms. In the relatively simple environment of a cell culture system, genetic defects in glycans usually do not have severe biological consequences; however, the same defects have been shown to have major phenotypic consequence in a complex multicellular organism, in which glycans are involved in mediating important cell-cell and cell-matrix interactions [45, 46]. This has complicated the interpretation of the role of glycans from genetic studies.

1.3 Methods to study extracellular protein-protein interactions

1.3.1 General overview

Historically, a range of biochemical and genetic methods have been used to identify specific extracellular receptor-ligand interactions. Almost four decades ago, the receptors for platelet-derived growth factor (PDGF), epidermal growth factor (EGF) and insulin were all isolated using similar biochemical methods that involved co-purification of chemically cross-linked receptors from cells that interacted with radiolabelled (Iodine-125) ligand [47, 48, 49, 50]. Upon isolation of the target receptor, common approaches were to either (i) microsequence parts of the protein to generate nucleic acid probes to screen a complementary DNA (cDNA) library generated from the tissue of interest to isolate the cDNA that encoded for the receptor of interest or, (ii) generate antibodies from the purified protein to carry out immunoscreening of cDNA expression libraries (reviewed in [51]). cDNA expression cloning was commonly used especially in the field of cell surface receptor characterisation of lymphocytes, where cDNA clones were transiently transfected into cells and screened with monoclonal antibodies to "pan" for cells that express the receptor [52, 53]. The use of interaction blocking monoclonal antibodies has also led to discovery of extracellular interactions especially in the context of immune cell interactions [54, 55]. Monoclonal antibodies have also contributed majorly to the identification of virus receptors for human immunodeficiency virus [56, 57] and rhinovirus [58, 59].

Early methods also included direct cell-cell binding assays using cell types that either expressed the interacting molecules in their endogenous forms or that had been transfected to express the receptor (or ligand) of interest. This was first demonstrated with T-lymphocytes expressing CD2 cell surface marker that formed “rosettes” with erythrocytes that expressed CD58, the receptor for CD2 [60]. A variation on this when studying interactions mediated by non-erythrocytic cells, was the cell aggregation assay, which has been used to identify both hereophilic [61] and homophilic [62] interactions. Finally, genetic approaches have also been mainly in the context of neuronal cell interactions to identify receptors using inferences from phenotypes of specific gene targeted mutants in model organisms (*Caenorhabditis elegans* and *Drosophila melanogaster*) [63, 64, 65].

The completion of the human genome project was quickly followed by multiple studies that mapped the human membrane proteome to annotate proteins that have the potential to participate in extracellular interactions [66, 67, 68]. While the basic principles of the methods used in the post-genomic era have not changed dramatically compared to the conventional methods, the knowledge of the cellular secretome has allowed these methods to be high-throughput. Here I will highlight the key aspects of some of these post-genomic methods and their utility in understanding molecular mechanisms of cellular interactions. These will exclude the genetic knockdown or knockout screening approaches, which will be discussed separately in section 1.6.

1.3.2 Cell-based binding assays

Cell-cell binding assays

Extracellular receptor-ligand interactions can be studied by investigating direct cell-cell adhesion, in which two cell types that express different sets of membrane proteins are mixed together and the binding is detected with microscopy or by labeling with a radioisotope or fluorochrome [69, 70, 71]. In the recent years, efforts have been made to improve such assays in terms of their sensitivity and quantifiability. For example, in a proof-of-principle demonstration of one of the recent approaches, the low-affinity interaction between the adhesion receptors, JAM-B and JAM-C was detected using a cell–cell mixing experiment with cDNA transfected CHO cells expressing the receptors in a GPI anchored form [71]. It was suggested that the addition of the GPI anchor to the surface receptors aids their lateral diffusion, thereby facilitating

ligand-induced clustering. The cells were differentially labelled with DiI or DiD phospholipid binding dyes, which could then be analysed by flow-cytometry, such that interactions could be readily detected based on the fraction of cells containing signals from both dyes. While such approaches provide platforms to study interactions in the context of the plasma membrane, their results can be confounded by the potential for non-specific cell–cell clustering usually mediated by endogenous adhesion receptors whose expression can be altered during the course of the experiment (for example by transfection or signalling derived from the exogenously expressed genes).

A slightly different approach in this context has been the Baculovirus (BV) display system, in which a library of BV particles expressing the membrane protein of interest is used to screen for interactions (reviewed in [72]). For this, insect cells are transduced with BV encoding for a given membrane receptor and as the viral particle buds off from the cell membrane it will incorporate the overexpressed membrane protein. The virions can be used as 'nanoparticles' to screen for interaction with whole cells that express the interacting receptor [73]. This approach has been validated for low-affinity interaction study by detecting the interaction between BV particles displaying CD58, CD40 and glucocorticoid-induced tumor necrosis factor (TNF) receptor family-related protein (GITR) and cells expressing the respective ligands, CD2, CD40 ligand (CD40L) and GITR-ligand (GITRL), using a flow-cytometry based binding assay.

Cell-recombinant protein interaction assays

Recombinant proteins are common tools used in many methods designed to study cell surface interactions. A common theme in any approach that uses recombinant proteins is the purposeful oligomerisation of proteins, which generates avid probes, allowing detection of low-affinity interactions. Common approaches for this include tagging proteins with coiled-coil sequence from rat cartilage oligomeric matrix protein (COMP)–which causes spontanous protein pentamerisation or with human immunoglobulin-γ (Fc)-fusion or alkaline phosphatase (AP)-fusion proteins– which allows for protein dimerisation. Further multimerization approaches, for example, conjugating Fc-fusion tagged protein to protein-A microbeads and conjugating AP to anti-AP antibody to generate tetramers are also used to achieve even higher avidity [42]. Avid probes generated in this way can be used to screen for interactions with endogenous or

over-expressed receptors (for example by cDNA transfection) on the intact cell surface.

Libraries of cDNAs encoding a broad range of receptors have been widely used to study extracellular receptor-ligand interactions in the context of immune regulation, axon guidance mechanisms, 'de-orphanising' GPCRs, egg-sperm interaction, and host-pathogen interactions [74, 75, 76, 77, 78, 79]. In this method, cDNA expression vectors, with each vector expressing one receptor protein, are used to transfect cells to generate an 'expression library', which is then screened with avid probes to identify the receptor-ligand interaction. The clones of interest are then recovered for further analysis. In one of the comparatively earlier studies, the interaction between a transmembrane protein netrin-G1 ligand (NGL-1) and netrin-G1 was identified by screening 300 secreted or transmembrane (TM) human proteins expressed as Fc-fusion proteins with 400 putative cell-surface human proteins that were transiently expressed in COS7 cells [74]. A focused screen investigating interactions between TNF ligands and TNF receptors in mouse and human has also used a similar approach, in which cDNA clones expressing TNF receptors linked to a C-terminal GPI anchor sequence were individually transfected into HEK293T cells and and binding with soluble recombinant Fc and FLAG tagged TNF ligands was investigated using a 96-well format flow cytometry [75]. A slightly different iterative cDNA cloning approach was used to identify Juno as the oocyte receptor for Izumo1. In this method, a normalised mouse oocyte cDNA library was transfected into HEK293T cells in a pooled format and probed with pentamerised Izumo1. Transfected cells that bound the probe were selected and cDNAs from the cells were extracted and re-transfected into the cells in an iterative manner until the single cDNA clone that expressed Juno was identified [79].

A commercial group (Retrogenix) has designed a cell microarray technology using a library of cDNA encoding approximately 4500 membrane proteins. The approach used here is to spot cDNA together with the transfection reagent in glass slides and overlay mammalian cells such that cells are 'reverse-transfected' to overexpress a wide range of membrane proteins. Such libraries have been probed with avid reagents to identify low-affinity interactions [77, 78].

1.3.3 Cell-free protein interaction approaches

In the recent years, methods that use cell-free binding assays have been commonly used to study extracellular protein interactions [80, 81, 82, 83]. The

ability to generate large libraries of recombinant proteins has allowed for the development of methods that can map interaction networks in a high-throughput manner. In these approaches, soluble recombinant proteins that recapitulate the ectodomain of the membrane proteins and retain their binding properties are used. Recombinant proteins can be produced either with cellular systems or with In Vitro Transcription and Translation (IVTT) system. Here I will be predominantly focusing on techniques that use mammalian and insect cell expression systems because these systems are better suited to studying extracellular proteins as they are better at achieving correct folding, glycosylation and secretion of recombinant proteins compared to IVTT systems.

AVEXIS

The Wright lab has developed a method for detecting of low affinity protein-protein interactions, termed 'avidity-based extracellular interaction screen' (AVEXIS). In AVEXIS, a mammalian expression system (with HEK293 cells) is used to produce libraries of soluble ectodomains of membrane proteins that can be used for interaction screening. The proteins are tagged either with biotin ('bait' protein) or with COMP alongside beta-lactamase tags ('prey' protein). Bait proteins are captured on streptavidin-coated plates and the binding between bait and prey is quantified using the colorimetric beta-lactamase substrate nitrocefin.The method has been used extensively to characterise interactions in zebrafish neural system interactions. It was first used to identify receptor–ligand pairs within the zebrafish immunoglobulin superfamily (IgSF) [84] and later expanded to screen for interactions mediated by leucine-rich repeat (LRR) proteins, secreted factors and proteins from floor-plate microenvironment in zebrafish [85, 86, 87].

AVEXIS has also been used to identify interactions between malaria causing parasite *P. falciparum* and the human red blood cell. In two studies, recombinant parasite proteins were used to probe a library of soluble ectodomains that represented the red blood cell surface protein repertoire to identify host-pathogen interactions between basigin (BSG)- reticulocyte-binding protein homolog 5 (RH5) and Semaphorin7A-merozoite thrombospondin-related anonymous protein (MTRAP) [88, 89]. The interaction between BSG and RH5 was found to be an essential interaction required by all tested strains of *P. falciparum* in mediating invasion of red-blood cells. Recently, a variation of AVEXIS has been used to characterise interdependencies of the other parasite proteins that interact with RH5 and BSG to make up the RH5 invasion complex [90].

Glass-slide based microarray techniques

Several studies have compiled lists of extracellular proteins to generate recombinant protein libraries using mammalian and insect based cell expression systems to carry out high throughput screens in glass-slide based microarray format, in which recombinant proteins are captured on a glass slide and interactions are detected using multivalent probes. In this regard, one of the first demonstrations of protein-microarray approach was the successful detection of the low-affinity protein interaction between CD200 coupled to multivalent microbeads and its receptor CD200R [91] immobilized on epoxysilane-coated glass slides [92, 93]. Subsequently, this approach was used to screen an array of 1,334 proteins with 89 immunoglobulin superfamily (IgSF) receptor baits using bait-Fc fusion proteins bound on Protein A coated microbeads [83]. Recently, this method was also used to describe a virus-host extracellular interaction map using an array of 1,500 human proteins and human adenovirus (HAdV) encoded immunomodulatory protein baits [94].

AVEXIS itself has also been miniaturised to a microarray format [95], in which biotinylated bait proteins are directionally arrayed on streptavidin coated glass slides and probed with pentamerised probes tagged with FLAG tag ('DYKDDDDK' epitope) for detection using anti-FLAG antibody. This approach has been utilised to identify FcϵR1α as a receptor for PEAR1 from a ectodomain library representing the secretome of the human platelet [96].

1.3.4 Mass-spectrometry-based methods

Mass spectrometry (MS) based proteomics is a powerful tool used for the identification and quantification of peptides, but its application in the analysis of cell surface receptors has been challenging mainly because of the difficulty in obtaining homogenous plasma membrane protein regions, low abundance of membrane proteins, and technical difficulties in identifying hydrophobic regions of membrane proteins [97, 98]. This has been significantly improved in the past decade as several approaches have been described, which has allowed the use of MS in the identification of cell surface proteins [97, 98, 99, 100, 101]. That said, there are still very few mass spectrometry based methods that not only identify the cell surface molecules but also directly investigate low affinity interactions mediated at the surface of cells.

One such approach that was recently described is the ligand-based, receptor-capture (LRC) technology using TRICEPS [102]. TRICEPS is a chemoproteomic reagent that consists of three distinct sites: an amine-reactive site

for non-specific amine conjugation of purified ligand of interest; a protected aldehydereactive site for covalent conjugation with carbohydrate groups on glycoproteins under oxidising conditions; and a biotin site purifying the receptor peptides for identification by quantitative mass spectrometry. In this approach, ligands conjugated to TRICEPS are added to cells that have been previously exposed to oxidising conditions. Upon a stable ligand-receptor interaction, TRICEPS is covalently captured with nearby carbohydrates. After the reaction, the cells are lysed, trypsinized and the captured glycopeptides are enriched using the biotin tag and identified using MS. This approach was first validated by detection of known and novel interaction mediated by extracellular ligands of diverse nature such as peptides, glycoprotein, therapeutic antibodies and intact viruses. It has subsequently been commercialised (LRC-TriCEPS; Dualsystems Biotech AG) and used in other studies to identify cell surface interactions between secreted proteins and membrane receptors [103].

1.4 Methods to study extracellular protein-glycan interactions

1.4.1 General overview

Studies involving protein-glycan interactions have utilised techniques from genetics, structural biology, biochemistry, organic, and analytical chemistry. In the past, plant lectins were crucial tools in the field of glycobiology mainly because of their high abundance, easy sourceability and ability to mediate high affinity interactions (K_D of nM range) with monosaccarides. Early studies utilised the agglutination properties of lectins to characterise blood group antigens, antigenic determinants of which are specified by terminal sugar residues that are recognised by specific lectins. Plant lectins such as concanavalin A, L-phytohemagglutinin lentil lectin (LCA), and *Maackia amurensis* leukoagglutinin (MAL) have been used extensively in affinity chromatography to isolate major glycan structures present in animal cells [104]. Over the years, a wide range of methods including monoclonal antibody blocking, expression of glycosyltransferases by transfection with cDNA encoding specific enzymes, chemical and enzymatic manipulation of carbohydrates and direct binding with immobilized (glycan arrays and SPR based methods) or soluble (NMR based methods) glycans have been used to study interactions between glycans and proteins [105].

1.4.2 Binding inhibition approaches

In the past, the main method to identify interactions mediated by glycans with the protein of interest was through binding inhibition studies that used soluble probes such as monoclonal antibodies and purified mono-(oligo-)saccharides. One of the best characterised functional evidence for protein-glycan interactions is the low affinity interactions that allow for leukocytes to roll along the vascular surface. The first indications of such interactions were studied using the monoclonal antibody (MEL14) that specifically blocked the binding of leukocytes to the high endothelial venules (HEV). Similarly, monoclonal antibodies were also used to identify ELAM 1 (endothelial-leukocyte adhesion molecule), which was shown to be expressed on stimulated but not unstimulated human endothelial cells and gp140, which was expressed in activated platelets. These molecules were termed as selectins and are known today as L-selectin (MEL 14 antigen), E-selectin (ELAM 1), and P-selectin (gp140). Subsequent binding inhibition studies with monoclonal antibodies identified sialyl-Lewis^x (siaLe^x) as the glycan necessary for the interaction of P- and E-selectins with both fucose and sialic acid residues required for binding (reviewed in [106, 107]).

Chemical or enzymatic manipulation of glycans presented on the cell surface has also been a valuable tool for identification of interactions mediated by glycans [108]. Tunicamycin, a compound that inhibits N-glycosylation, has been used extensively for studying the role of N-glycans in mediating interactions. One common approach when studying interactions mediated by sulfated receptors is the treatment of cells with chlorate, which has been shown to inhibit the production of the high-energy sulfate precursor 3'-phosphoadenosine 5' -phosphosulfate (PAPS). This leads to the generation of undersulfated glycoproteins and proteoglycans in cells [109]. The ligand for L-selectin was determined to be a sulfated version of siaLe^x based on the decrease in binding to undersulfated glycoproteins.

Enzymatic deglycosylation is also a common method to study the role of glycans in mediating interactions. Common enzymes include those that selectively remove N-linked oligosaccarides (PNGaseF), sialic acid residues in an O-glycan (sialidase), and core-O-glycans (O-glycosidase) [110]. The first prediction of the nature of endogenous ligands for L-selectin was made on the basis of loss of binding of lymphocytes to cells treated with sialidase [107]. Similarly, neuraminidase (a type of sialidase) treatment of red-blood cells, which abolished the binding of the malaria causing parasite *P. falciparum*

protein EBA175 to these cells led to the discovery of the first host receptor, glycophorin A, for this parasite [111, 112].

1.4.3 cDNA expression methods

Another approach to studying glycan-protein interactions is to express the cDNA encoding a glycotransferase in a cell culture system and to detect the binding mediated by the new glycans (neoglycans) on the cell surface. This method has been used in studying interactions mediated by selectins and siglecs (sialic acid-recognizing Ig-superfamily lectins). In the example of P-selectin, transfecting cDNA encoding a human α-1,3-fucosyltransferase together with cDNA encoding P-selectin glycoprotein ligand (PSGL-1) into nonmyeloid cell line (CHO cell line) was shown to confer high affinity binding of P-selectin to these cells [113]. In another example, one of the early studies in identifying a receptor for the B cell adhesion molecule CD22 used a cDNA expression cloning approach, which identified a sialyltransferase required for adhesion [114]. Although the original study suggested the sialyltransferase enzyme to be the direct receptor, further biochemical experiments soon demonstrated that CD22 bound to a sialic acid residues through its extracellular domain, the production of which required the enzyme. From there on, homology based studies led to recognition of a new family protein called siglecs, which mediate binding in a sialic acid dependent manner [115].

1.4.4 Glycan arrays

Many past methods investigated the glycan binding ability of proteins in material-intensive assays (e.g. inhibition assays using purified glycans, lectins and monoclonal antibodies). This was changed when technological advances in isolating 'natural' glycans from sources such as cells, tissues, and pathogens, and generating synthetic glycans from chemical and enzymatic methods, allowed for the generation of glycan arrays [43]. The two largest mammalian glycan libraries compiled from natural and synthetic sources are currently from Consortium for Functional Glycomics (CFG) ($\sim$ 609 glycans) and Feizi and coworkers ($\sim$ 830 glycans). These arrays have been used to probe for interactions with plant and microbial lectins, glycan-binding proteins involved in the innate and adaptive immune system, glycan-specific antibodies, virus particles, and whole cells (reviewed in [116]).

1.5 Limitations of the existing methods

While the methods described above have been successful in the past, they have several limitations:

1. One of the main drawbacks of methods that require large libraries of cDNA or recombinant proteins is that the initial compilation of libraries containing hundreds of molecules is resource intensive, and most researchers' interests are usually focussed on a single or small number of proteins rather than the networks of interactions within a larger family.

2. The cDNA transfection approach of membrane proteins requires all expressed proteins to be transported on the surface of cells; however cell surface targeting of transmembrane proteins is strictly regulated and often requires cellular accessory factors such as transporters, chaperones and correct oligomeric assembly. Thus, transfecting a single cDNA might not be enough to achieve cell surface expression in many cases.

3. Recombinant protein based methods require that the receptor binding function is retained when expressed by heterologous cells out of the context of the plasma membrane as a soluble recombinant protein. Whilst this is generally the case for proteins that span the membrane once (single-pass (type I, type II) or GPI-anchored), it is not so for receptor complexes and membrane proteins that span the membrane multiple times. Therefore, interactions made by these latter classes of complexes and proteins are usually underrepresented. In addition, although care can be taken to produce recombinant proteins in heterologous systems that increase the likelihood of structurally critical post-translational being added properly, this is not always achieved. An example of this is the under-glycosylation of HEK-293 expressed recombinant glycophorin A, which does not retain its property to interact with its known *P. falciparum* interaction partner EBA175 [117].

4. While the use of recombinant proteins makes the manipulation of membrane receptors biochemically tractable, it is challenging to investigate other cellular components required for cellular recognition at the surface of the cell when interactions are studied in isolation between purified probes rather than in the context of a cellular system. In a cell, interactions occur in a complex environment, which includes contributions from

a charged glycocalyx of carbohydrates and lipids displayed on a dynamic membrane [118, 119, 120]. As a consequence of taking the reductionist approach of studying cell surface molecules as soluble recombinant proteins, the contribution from cellular factors other than direct receptors that ultimately contribute in cellular recognition is usually ignored.

5. Mass-spectrometry based methods such as LRC do allow for the interrogation of cell surface interaction mediated by endogenous receptors in the context of the plasma membrane but they still require chemical manipulation of the cell surface (e.g, oxidation), which can alter the biochemical nature of the molecules present on the surface of the cells. In addition, this method relies on the biological properties of the ligand being preserved during the process of non-specific amine conjugation. This is difficult to achieve for proteins that rely on lysine residues (residues where amine conjugation is normally done) to mediate interactions. Moreover, the method requires the receptor to be glycosylated, which is not always the case.

6. The success of glycan arrays depends on the representation of the glycans they contain. While there exist multiple efforts to define the 'glycome' of an organism using mass-spectrometry, lectin and antibody array based approaches, due to the very high diversity of glycans, the size of mammalian glycan array libraries available today still pales in comparison to the DNA libraries; such mammalian glycan array libraries may contain a fraction of those glycans present in nature.

7. Methods that tether glycans to a surface such as SPR, glycan array or ELISA-based methods limit the number of ways in which glycans can be presented, which can change the binding properties of the isolated glycans. On the surface of cells, glycans are presented on a glycoprotein or a glycolipid scaffold in diverse conformations, which can be critical for ligand recognition.

8. Methods that use isolated glycans require the glycan receptors to independently mediate binding with the ligand. This is not always the case as many glycoprotein backbones also participate in the interaction. The best known example of this is the interaction mediated by P-selectin, whose interaction depends both on the glycan and the adjacent sulfotyrosine residues in the glycoprotein receptor PSGL1 [121].

1.6 Loss-of-function genetic approaches to study cellular recognition process

1.6.1 General overview

Genetic mutations can alter cellular processes, therefore a method to study a function of the gene is to investigate the effect of its absence. Genetic mechanisms of multiple biological processes can be studied by analysing loss-of-function (LOF) mutants, in which the altered gene product lacks the molecular function of the wild-type gene. The underlying principle of a LOF approach is to ablate the function of a gene by targeting DNA, RNA or protein. LOF approaches range from non-targeted chemical mutagenesis approaches to generate mutants, to systematic large-scale generation of mutant libraries for genetic screens using genome-editing technologies such as short interfering (si) or short hairpin (sh) RNA, and more recently the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 (CRISPR-associated protein 9) system (reviewed in [122]). These approaches can be applied both in a large-scale manner for the identification of novel factors in different cellular contexts and in a small-scale manner to explore the roles of few candidate genes usually involved in disease processes that are identified from large-scale screens. The pioneering works of genome-wide LOF screens carried out in *Caenorhabditis elegans* [123] and *Drosophila melanogaster* [124, 125] demonstrated how large-scale mutagenesis could be applied to assign gene function. These approaches have already been applied to many areas of biological and biomedical research including the investigation into cellular recognition events, which I will discuss in this section.

1.6.2 Study of naturally occurring mutants

Traditional loss-of-function analysis involved characterising naturally occurring mutants. For example, in the context of host-pathogen interactions, blood-group polymorphisms that occur in the human population were used to identify receptors required for invasion of host cells by malaria causing parasites *P. falciparum* and *P. vivax*. One of the best described examples of this is the identification of the mutation that abolishes the expression of the Duffy antigen receptor of chemokines (DARC) on the surface of red blood cells (the so-called 'Duffy-negative phenotype') in western and central Africa, which was shown to confer almost complete protection from *P. vivax* infection [126, 127]. This

led to the identification of DARC as a critical receptor for *P. vivax* invasion into red-blood cells and the subsequent identification of its parasite binding partner, Duffy binding protein (DBP). Untill today this receptor-ligand interaction remains the only known host-pathogen interaction in the context of *P. vivax* infection. Similarly, the studies characterising the host receptors for *P. falciparum* merozoite surface proteins have also utilised naturally occuring erythrocytes with mutations in blood group antigens. For example, the parasite-host interactions between EBA175 with glycophorin A (GYPA), EBA140 with glycophorin C (GYPC) and EBL1 with glycophorin B (GYPB) were studied using blood groups that lacked the respective receptors; En(a-) (absence of GYPA), Leach phenotype (absence of GYPC) and S-s-U-phenotype (absence of GYPB) (reviewed in [128]). Another well-known example of this is the study in individuals resistant to HIV infection, who were found to carry homozygous mutation in the cell surface receptor protein CCR5, which acts as the viral co-receptor [129]. While the study of natural mutants using biochemical or population genetics approach has been useful in the past, the difficulty in acquisition of mutants, the presence of rare alleles in a population, and the limited shelf life of biological materials makes this approach generally unsuitable for systematic investigation of cellular recognition processes.

1.6.3 Genetic screening approaches

Chemical and insertional mutagenesis

A different approach to study a gene function is to generate loss-of-function mutants using genome editing tools. The classical approach of mutant generation was by random mutagenesis through chemical mutagens (e.g, ethyl methanesulfonate (EMS)), that can introduce a variety of genetic lesions that are expressed as complete or partial loss of function of the gene product [130]. Such approaches were commonly carried out in *Saccharomyces cerevisiae*, which served as an ideal model organism because of its rapid generation time and its haploid genome, which permitted efficient generation of homozygous mutants for the study of recessive phenotypes. This approach served as a fast and effective method to generate a large number of mutants so that they could be screened for the phenotype of interest. Genetic screens in yeast using temperature sensitive mutants served as powerful tools to study cell essential mechanisms. The early works in yeast temperature sensitive mutants, which identified components of the SEC genes [131, 132], has been extremely valu-

able in the field of membrane protein biology as they have shaped our current understanding of the general secretory pathway that is required for the transport of the majority of membrane and secreted proteins. Similarly, chemical mutagenesis has also been used to mutate the genome of a hermaphrodite model organism *Caenorhabditis elegans*, which has allowed study into, among other cellular processes, the genetic control of neuronal development. For example, the molecular basis of UNC-6/Netrin signalling in axon guidance was first identified by characterisation of mutants generated through random mutagenesis in this organism (reviewed in [133]).

A different approach of random mutagenesis proceeds through insertional mutagenesis, which employs the strategy by which exogenous retroviral and transposable DNA can be inserted randomly into the genome such that if the insertions is in the coding or regulatory region of the gene, the gene product will be rendered non-functional. This approach has an advantage over chemical mutagenesis as it facilitates the identification of the mutated gene as the inserted DNA, whose sequence is known, serves as a molecular tag. While powerful, the mutagenesis techniques had limited use for the generation of homozygous mutants in mammalian cell culture systems mainly because of the diploid nature of mammalian cells, the lack of strategies to set up genetic crosses, and the low rates of homologous recombination. One of the strategies employed to address this limitation was to utilise mouse embryonic stem cell (ESC) lines that are deficient for Bloom's syndrome protein (BLM), in which cells exhibit a high rate of mitotic recombination to generate a genome-wide library of homozygous mutant cells from heterozygous mutations induced with insertional mutagenesis approach [134]. Furthermore, the discovery of haploid human cell lines, KBM7 [135, 136] and HAP1 (a derivative of KBM7) [137], and haploid mouse embryonic stem cells [138] further facilitated the use of insertional mutagenesis appraoch in mammalian cell culture system. The application of random mutations techniques in haploid cells can directly cause loss-of-function phenotypes, which can be studied in a high-throughput manner. Such approaches have been used in the context of host-pathogen interactions mediated by bacteria and viruses (reviewed in [139]). Examples with the gene trap[1] approach in haploid human cell lines include identification of host factors for bacterial toxins [140, 141, 142], intracellular receptor for the

[1] A type of insertional mutagenesis using a vector that contains a strong splice acceptor site, an efficient polyadenylation signal and a marker gene. The insertion of the vector into the intronic (or exonic) region leads to inactivation of the target gene.

Ebola virus [137], deciphering the glycosylation of the Lassa virus receptor α-dystroglycan (α-DG) [143], and receptor switching mediated by the Lassa virus upon virus internalisation [144]. More recently, the same approach has been used for the identification of the host receptor for adeno-associated virus (AAV) serotype 2 [145].

RNAi mediated approaches

During the time when the creation of homozygous mutant libraries for cells and organisms with diploid genomes was time- and capital-intensive and low throughput, RNA interference (RNAi) technology provided an alternative approach to allow high throughput gene silencing through sequence-specific targeting of mRNAs [146]. In this approach, RNAi reagents (such as synthetic siRNAs, siRNA precursors (short hairpin RNAs (shRNAs)), or long double-stranded RNAs (dsRNAs)) are introduced to cells or organisms via methods such as transduction, transfection, microinjection or, in case of *C.elegans*, by simply feeding organisms with *E.coli* expressing dsRNAs [147]. Once in the cells, siRNAs are incorporated directly into the RNA-induced silencing complex (RISC), whereas dsRNAs are processed by DICER to first generate siRNA. The siRNA-RISC complex can mediate gene silencing by cleaving the complementary mRNA (when the sequences are perfectly complementary) or by interfering with translation (when the sequences are partially complementary). The approach has allowed for the implementation of both small scale and genome-wide loss-of-function screens in human and Drosophila cell lines for the identification of genes and gene networks involved in signal transduction processes [148], identification of regulators of cell adhesion, [149] and extensively in host-pathogen interactions (e.g. colonisation of Drosophila cells by sinbad virus [150], dengue virus [151]; mammalian host factors required by hepatitis C virus (HCV) [152, 153], west nile virus [154], vaccinia virus [155] and multiple screens for host factors required by HIV (reviewed in [156])). Similar screens have also been carried out in hematopoietic progenitor cells that can be differentiated into erythroblasts, thereby allowing forward genetic screens to be carried out in the otherwise genetically intractable anucleate red-blood cells. A shRNA mediated genetic screen in such a system has identified CD55 as a crucial host receptor for *P. falciparum* invasion into red-blood cells, although the ligand on the parasite end is still unknown [157].

One of the biggest challenges of using RNAi as a tool to study gene function involves the sequence-specific off-target effects of siRNA. RNAi uses

an existing cellular pathway governed by endogenous microRNAs (miRNAs) that regulate cellular gene expression levels using incomplete complementarity between the miRNA and its target. Exogenously supplied siRNA can function as an endogenous miRNA and mediate target recognition by partial sequence complementarity, which can lead to translational repression and/or degradation of non targeted mRNAs. Such off target effects can confound the interpretation of the phenotypic effects and potentially create cellular toxicities [158]. In addition, the genetic perturbation using RNAi frequently results in incomplete silencing, which, combined with the off-target effects can lead to a decrease in sensitivity and inconsistent results. One such example of inconsistent results is the lack of overlap between three large-scale RNAi-mediated knockdown approaches conducted by independent laboratories for the identification of factors responsible for HIV infection, which have identified 842 putative factors out of which only 37 genes were identified in more than one screen and merely three genes identified in all three studies [159, 160, 161].

1.6.4 CRISPR-Cas9 approach

Introduction to CRISPR-Cas9

CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated) systems are adaptive immunity strategies developed by bacteria and archaea to protect themselves against foreign genetic elements such as viruses and plasmids [162, 163]. The system involves three key steps: (i) spacer acquisition, during which small DNA fragments from foreign DNA elements are inserted into the CRISPR locus; (ii) crRNA (CRISPR RNA) expression, during which the CRISPR locus is transcribed to generate a long primary CRISPR transcript (the pre-crRNA); (iii) target interference, during which the target is detected and degraded by the crRNA and Cas9 protein(s). Depending on the architecture of the interference molecules, the system can be divided to two main classes, consisting of six types (type I-VI) and 16 subtypes. The CRISPR-Cas9 technology is derived from type II CRISPR-Cas system, which utilises crRNA, tracrRNA (trans-activating crRNA[2]), and a single large multi-domain effector protein (Cas9) to mediate target recognition and cleavage. Under this system, the Cas9 protein, which is an endonuclease, is loaded with RNA duplex (tracrRNA:crRNA) and this riboprotein complex is directed to the target site (i.e., the site complementary to the guide sequence

[2]These are short sequences that are complementary to the corresponding crRNA. They are required for Cas9 to correctly recognise the target DNA.

from crRNA), enabling the Cas9 protein to introduce double stranded breaks (DSBs) in the DNA. Target recognition by Cas9 also requires the protospacer adjacent motif (PAM), which is a species-dependent short (2-6 nucleotide sequences (nts)) DNA sequence immediately following the target DNA sequence (reviewed in [164, 165]). The majority of the current CRISPR-Cas9 systems use the Cas9 protein from *Streptococcus pyogenes* (henceforth referred to as SpCas9), which uses 5′-NGG and 5′-NAG (although at less efficient rates) as the PAM sequence (figure 1.1).

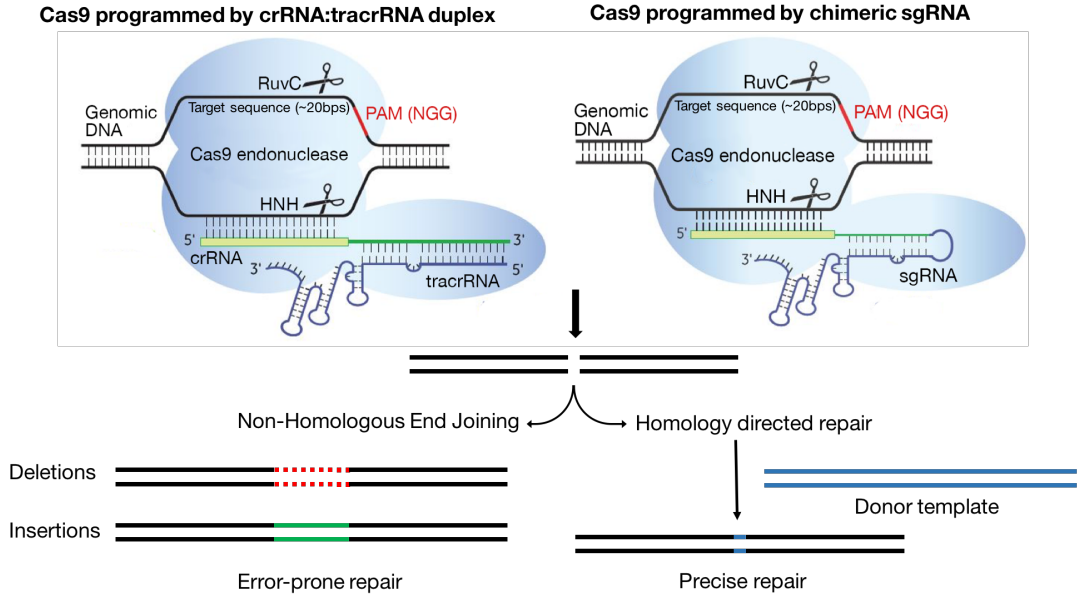


Fig. 1.1 Overview of type II CRISPR-Cas9 mediated gene editing. Cas9 can be programmed either with RNA duplex generated from crRNA and tracrRNA or with a single chimeric RNA, which mimics the RNA duplex. Once loaded with the RNA duplex or the chimeric RNA, the Cas9 protein is targeted to the region of the genome that is complementary to the 20 nts sequence of the crRNA (or the gRNA) where it mediates double stranded DNA cleavage. The Cas9 protein contains two endonuclease domains, the HNH and RuvC domains that either cleave the strand complementary to the guide sequence, or the strand matching the guide sequence. Once a double stranded break is generated the cellular DNA machinery repairs it either with the NHEJ pathway or the HDR pathway (if homologous donor template is present). NHEJ pathway is error-prone and can lead to generation of indel mutations, leading to inactivation of the gene [165]. Figure adapted from [166].

Gene targeting using the CRISPR-Cas9 system

One of the major findings in the field of CRISPR-mediated gene editing has been the observation that the tracrRNA:crRNA duplex can be engineered as a single piece of chimeric RNA (termed as single guide RNA (sgRNA) or gRNA) [167]. Immediately following this observation, CRISPR/Cas9 system

was adapted for genome engineering in human cells, where by changing the 20 nucleotide guide sequence of the gRNA, the DNA sequence of interest could be efficiently targeted by the Cas9 endonuclease to generate double stranded breaks (DSBs) [168, 169]. DSBs in cells are repaired using two major pathways: (i) the error-prone Non-Homologous End Joining (NHEJ) pathway; and (ii) the high-fidelity Homology Directed Repair (HDR) pathway. In higher organisms, the NHEJ pathway, which does not require specific sequence for ligation of DNA, is the major method for repairing DSBs and unlike HDR, this pathway is active during all stages of the cell cycle [170]. DNA repair using the NHEJ pathway is prone to insertion and/or deletion (indel) mutations at the junctional site, which in coding exons can introduce premature stop codons or frameshift mutations, leading to disruption of the targeted gene. This system therefore provides a convenient way to generate loss-of-function mutations in the mammalian genome.

Overview and basic principles of a knockout screen using the CRISPR-Cas9 system

Given the ease of generating libraries containing thousands of gRNAs that could be used to create large knockout collections, CRISPR-Cas9 technology quickly became the method of choice for genome-wide loss-of-function screening approaches. The first genome-wide loss-of-function screens were carried out in human [171, 172] and mouse [173] cells using cell growth as a phenotype and showed successful application with both positive and negative selection results. All the initial screens and the majority of the genome-scale screens that have been described up to now have used a pooled screening approach, the basic principles of which are summarised below.

In a pooled screening approach, gRNA oligonucleotides are synthesised as a pool and cloned to create plasmid library that is used for virus production. The viral library is then used to transduce Cas9-expressing cells[3] at a low multiplicity of infection (MOI) to generate a library of knockout cells. Ensuring a low MOI (usually MOI of 0.3) is a crucial step in pooled screens to reduce the probability of more than one gRNA being transduced and stably integrated into one cell. The mutant library thus generated is subjected to positive (e.g,

[3] In a single vector approach such as the LentiCRISPR used in [171], both Cas9 and gRNA are encoded from a single plasmid; thus the cells do not express Cas9 prior to transduction with the virus generated from this plasmid. In a dual vector approach, stable cell lines expressing Cas9 are first generated and then transduced with the gRNA library to generate knockout libraries.

drug, toxin resistance), negative (e.g. proliferation) or marker gene selections (expression of surface markers) and the cells of interest are recovered (will be discussed in detail below). The integrated virus serves as a molecular tag for each mutated gene and this can be read-out by isolating genomic DNA from the cell population, sequencing (using next-generation sequencing (NGS)) across the gRNA-encoding regions, and then mapping each sequencing read to a pre-compiled list of designed gRNA library. Computational analysis is then carried out to determine the differences in the abundance of gRNAs between the control and the phenotyped sample to identify the gene product involved [174].

A pooled screen can be carried out with negative, positive or marker gene selection. The goal of a negative selection screen is to identify perturbations that affect the survival or proliferation of cells, which cause the perturbed cells to be depleted during selection. The approach here is to transduce two sets of cell populations and subject one set to the selection while the other serves as a non-selected control. The gRNA abundance in both populations is then analysed to identify gRNAs that have been depleted because of the selection. One of the simplest form of a negative selection screen is the continued growth of cells for an extended amount of time to identify genes that are required for the proliferation of cells. Such screens have been used to identify essential genes for several cell lines [171, 175, 176, 177]. In a positive selection screen, a strong selective pressure is introduced such that the probability of cells being selected without the genetic perturbation is low. These screens have been used to identify perturbations that confer resistance to drug [171, 173], toxin [173, 178, 179], hypoxia stress [180], and pathogen infection [181, 182, 183]. Unlike negative selection screens, the signal for a positive selection is usually strong, as the abundance of relevant gRNAs in such screens increases relative to the rest of the gRNAs, which allows for easy detection of resistant cells [184]. A third type of selection is the marker gene selection in which the phenotype is not based on lethality of the cells but rather on mutations that change marker gene protein expression. In this type of screen, the marker gene is either endogenously-tagged with fluorescent proteins [185] or labelled with highly specific antibodies [186, 187] and cells with gRNAs that target genes whose perturbations contribute to the expression of marker gene are captured using fluorescence-activity cell sorting (FACS)-based approaches.

1.6.5 The scope of CRISPR-Cas9 knockout screening system in the context of cellular interactions

Over the last decade, technological advancements in gene annotation and gene synthesis have facilitated the generation of libraries of recombinant proteins to carry out numerous cell-based or cell-free assays, as discussed here (figure 1.2). The recent development of the genetic-screening method using the CRISPR-Cas9 system has the potential to utilise these libraries of recombinant proteins that could potentially be involved in cellular interaction within different biological contexts (e.g. neural and immunological recognition, and host-pathogen interactions). By using the 'binding' of recombinant proteins to cell lines as a measurable phenotype, a positive selection genome-scale knockout screen can be designed, in which cells within the knockout library that can no longer bind the protein they previously bound can be selected for the identification of the perturbations, which led to the loss in binding. Within those perturbations should, in principle, lie the receptor of the recombinant protein and any components that alter the expression of that receptor. This approach could be used for the study of cellular recognition events and has the following potential advantages over the previously discussed pre-existing methods:

1. Screening on cells that express a wide range of endogenous proteins in the context of a plasma membrane, which contain appropriate post-translational modifications, avoids the necessity of recombinantly generating a large number of proteins. This facilitates the investigation of proteins that are normally difficult to study with biochemical approaches (e.g, large proteins, multi-pass membrane receptors and protein complexes).
2. Genome-scale screens provide a unique platform to study interactions mediated by glycans as the identity of a glycan receptor can be readily inferred by identifying enzymes and intracellular transporters involved in their biosynthesis.
3. The use of an unbiased genome-wide approach to study receptor-ligand interactions on the cell surface facilitates the identification of cellular factors that are not necessarily direct receptors but are involved in their expression on the surface of cells.

4. One important advantage of this approach is that it should be able to identify all gene products required for extracellular interactions without the need to make any prior assumptions regarding the biochemical nature of the receptor.

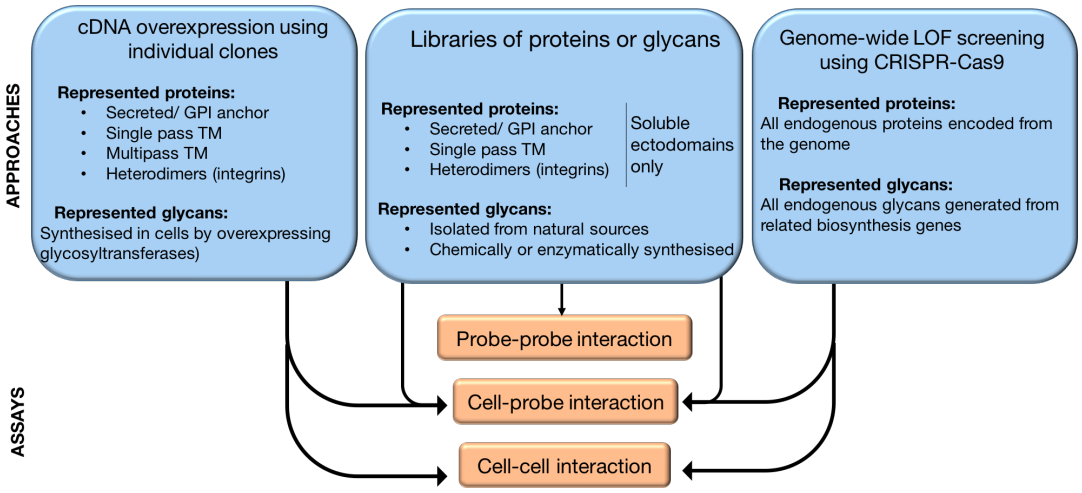


Fig. 1.2 Outline of the major approaches that are utilised for the study of extracellular interactions. Current high throughput approaches utilise compiled libraries of cDNA clones or protein/glycan libraries to investigate protein interactions. Gain-of-function approaches such as cDNA-mediated over-expression strategies have been used in cell-cell binding assays or cell-probe binding assays for the identification of novel extracellular interactions ('probe' in this context refers to avid recombinant proteins or glycans). A genome-scale loss-of-function screening approach has the potential to utilise the avid recombinant proteins or glycans in a cell-binding assay to complement the pre-existing methods to identify direct receptor-ligand interactions. In addition, a genome-scale knockout approach would, in principle, also allow for the identification of intracellular pathways contributing to the biology of the receptor.

1.7 Thesis aims

In this work, I explore whether genome-scale cell-based CRISPR-Cas9 knockout screens can be used to determine the molecular basis of cell surface recognition events. The aim is to develop an approach that utilises the commonly used tools in receptor discovery, monoclonal antibodies and recombinant proteins, to reveal direct receptor-ligand interactions, and assess the feasibility of such an approach to identify other important cellular factors required for the interaction, such as posttranslational modifications, co-receptors, and cytoplasmic proteins involved in receptor trafficking.

CHAPTER 2

MATERIALS AND METHODS

2.1 Buffers/Media/Solutions

The composition of the buffers and solutions used in this study are described as follows:

- TAE buffer-10x: 2 M Tris, 57.1 mL acetic acid (100%), 100 mL EDTA (0.5 M, pH 8.0) in MiliQ water
- LB media: 1% tryptone, 0.5% Yeast extract, 1% NaCl in MiliQ water.
- Diethanolamine buffer, pH 9.2: 10% diethylamine and 0.5 mM $MgCl_2$ in MliQ water
- HBS buffer (10x) pH 7.4: 1.5 M NaCl and 200 mM HEPES in MiliQ water
- PBS buffer (10x), pH 7.4 : 80 g NaCl, 2 g KCl, 14.4 g Na_2HPO_4 and 2.4 g KH_2PO_4 in MiliQ water
- Nitrocefin solution: 5 mg Nitrocefin was dissolved in 500 μL DMSO and added to 39.5 mL PBS to reach 40 mL final volume (concentration: 242 μM); filtered through 0.2 μm filter and stored in dark.
- RPMI culture media: RPMI 1640 (Life Technologies) supplemented with 10% heat-inactivated (50°C for 20 minutes) FBS, 1 mM sodium pyruvate (Life Technologies), 10 mM D-glucose (Sigma) and 1% penicillin–streptomycin (100 units/mL).
- DMEM/F12 culture media: DMEM/F12 media supplemented with 1% penicillin-streptomycin (100 units/mL) and 10% heat inactivated FBS
- IMDM culture media: IMDM media supplemented with 1% penicillin-streptomycin and 10% heat inactivated FBS

- Sodium phosphate buffer (80mM stock), pH-7.4 : 7.1 g $Na_2HPO_4.2H_2O$, 5.55 g NaH_2PO_4
- His-tag purification- binding buffer: 20 mM Sodium Phosphate Buffer, 0.5 M NaCl and 20 mM Imidazole, filtered and degassed
- His-tag purification- elution buffer: 20 mM Sodium Phosphate Buffer, 0.5M NaCl and 400 mM Imidazole, filtered and degassed

2.2 Generation of expression plasmids

All expression plasmids used for protein production via transient transfection of EBNA1 expressing HEK-293-6E cells was based on the pTT3 plasmid backbone, which features an EBV origin of replication that allows increased expression of transgenes in these cells [188] (figure 2.1A). While the endogenous signal peptide was included in the insert for the expression of mammalian proteins, an exogenous N-terminal signal peptide from a mouse immunoglobulin κ-light chain (VkSP) was cloned into the vector backbone for the expression of *Plasmodium falciparum* proteins (figure 2.1B). Additional processing of *Plasmodium* ectodomain coding sequences included trimming of the endogenous signal peptide, codon-optimisation for mammalian cell expression and mutation of N-linked glycosylation sequences from NXS/T to NXA (where X is any amino acid except Proline) in order to prevent inappropriate glycosylation.

All ectodomains were chemically synthesised together with the flanking NotI and AscI sites and cloned into either of the following vector(s) (also see figure 2.1C).

1. the 'pentameric vector' containing N-terminal signal peptide, and C-terminal tags in the following order: rat Cd4 domains 3 and 4 tag (henceforth referred to as the Cd4 tag), a pentamerization sequence from the cartilage oligomeric matrix protein (COMP), sequence for β-lactamase enzyme, 3$\times$ FLAG tag and a terminal hexa-his tag. All proteins generated via this plasmid were expressed as pentamers.
2. the 'monomeric biotinylation vector' containing N-terminal signal peptide and a C-terminal Cd4 tag followed by an *E.coli* BirA biotin ligase sequence motif and a hexa-his tag. This was used for the production of his-tagged monomeric protein with a mono-biotinylation on a specific lysine residue.

3. 'monomeric β-lactamase vector' containing N-terminal signal peptide and a C-terminal Cd4 tag followed by sequence for β-lactamase enzyme.

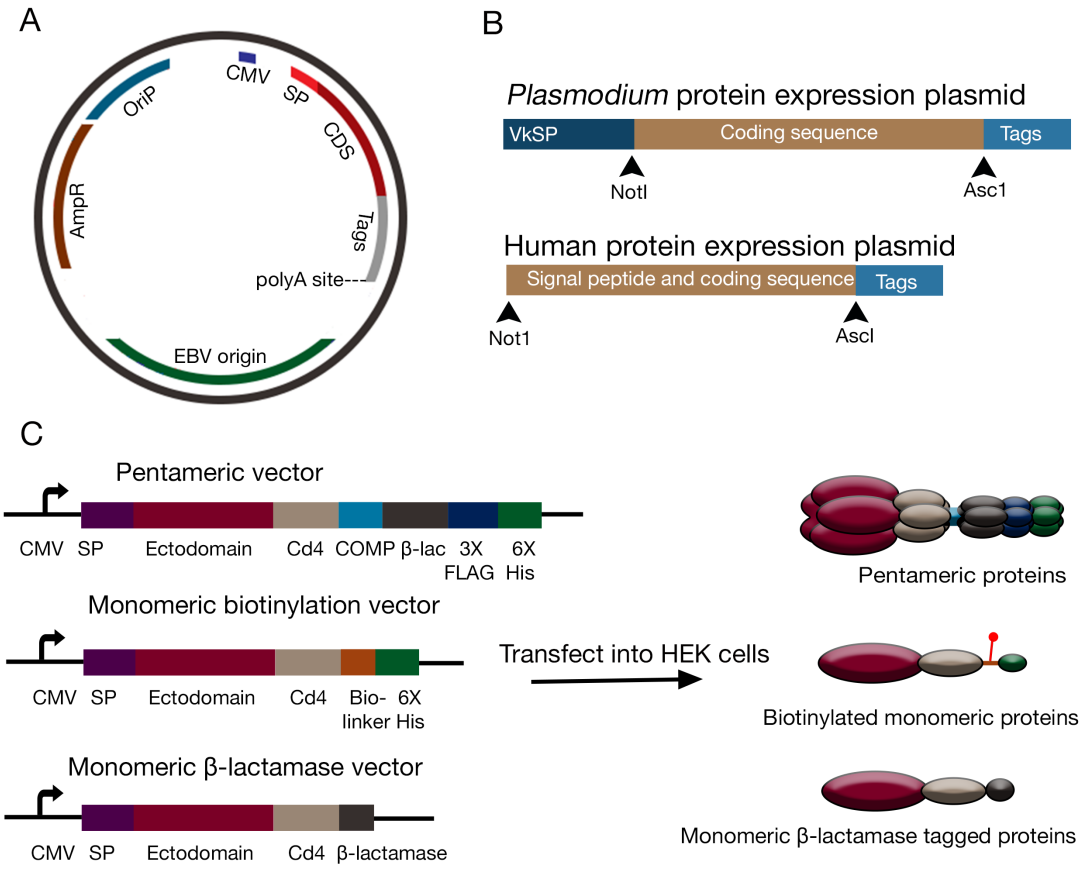


Fig. 2.1 Schematic diagram representing the plasmids used for protein production using the HEK-293-6E protein expression system. A. Generalized map of pTT3-based expression plasmids. EBV: Epstein-Barr virus, AmpR: Ampicillin resistance cassette, oriP: pBR322 origin of replication, CMV: Cytomegalovirus promoter, SP: signal peptide, CDS: protein ectodomain-coding sequence, polyA: polyadenylation site. **B.** The arrangement of restriction cloning sites were different in respect to the start of the signal peptide for *Plasmodium* and human proteins. **C.** All vectors contain a cytomegalovirus (CMV) promoter driving the transcription of the desired transgene. The pentameric vector construct contains the ectodomain followed by the Cd4 tag, a region from the cartilage oligomeric matrix protein (COMP) that forms pentamers, a β-lactamase enzymatic tag, a 3 $\times$FLAG-tag for immunological detection and a C-terminal 6 $\times$His tag for purification. The monomeric biotinylation vector contains the ectodomain followed by the CD4 tag, a biotinylation sequence and a his-tag (BLH), whereas the monomeric β-lactamase tagged vector is the same as the biotinylation vector but consists of β-lactamase tagged instead of the BLH tag.

All plasmids except those relating to IGF2R expression were obtained from the laboratory database. Transformation of the recombinant DNA was done with 1 μg of DNA using chemically competent *Escherichia coli* (*E. coli*)

cells ('Top10', Invitrogen) using manufacturers recommendation. Maxi-prep kit (Invitrogen) was used to purify the DNA and the concentration was adjusted to 1 mg/mL.

2.2.1 Cloning of IGF2R expression construct

PCR amplification of ectodomain of IGF2R

A plamid for the expression of full length IGF2R was obtained from Origene (SC300143). Region representing M1-2306V (the ectodomain) was amplified from this construct using IGF2R-fwd and IGF2R-rev primers (refer to appendix table A.1) with Q5 Hot Start High-Fidelity 2$\times$Master Mix. NotI and AscI restriction sites were added to forward and reverse primers correspondingly, in order to facilitate sub-cloning into the expression vectors. The PCR reaction was performed in a thermocycler with the following set up: Initial denaturation for 5 minutes at 95 °C; 25 cycles of denaturation at 95 °C for 20 s, elongation at 68 °C for 7 minutes and annealing at 72 °C for 90 s; and final elongation for 10 minutes at 72 °C. After finishing the run, the samples were cooled to 4 °C.

DNA samples (5 μL/sample) from the PCR reaction were analysed on a 1 % (w/v) agarose gel prepared in TAE buffer. After verifying the size of the amplified fragment, the remaining PCR product was cleaned using Qiagen PCR purification kit. In the final elution step, DNA was eluted in 30 μL EB buffer.

Restriction digest and ligation

The purified PCR fragments were digested with NotI and AscI enzymes. For each sample, a 50 μL reaction mixture was set up with the following components with the indicated volumes and concentrations: 30 μL purified PCR product, 1 μL NotI enzyme (10000U/mL), 1 μL AscI enzyme (20000U/ mL), 5 μL CutSmart (NEB) buffer (10$\times$) and 13 μL MiliQ water. The reaction mixture was incubated for 24 hours at 37 °C following which all 50 μL from each sample was run on a 1% agarose gel. The corresponding fragments were excised from the gel and purified using Qiagen gel extraction kit. DNA was eluted in 20 μL EB buffer and the concentration was measured using spectrophotometer.

The insert (restriction digested, purified and gel extracted PCR product) was next ligated with the desired vector backbone (also restriction digested, purified and gel extracted) using T4 DNA ligase in the presence of T4 DNA ligase buffer containing ATP (NEB). Ligation mixture was incubated for 16 hours at room temperature and transformed into chemically competent *E.coli*

(Top 10, Invitrogen) according to manufacturers recommendation and plated on a LB-agar plate supplemented with 100 μg/mL ampicillin.

Plasmid extraction

Randomly selected bacterial colonies from the LB plate were inoculated in 8 mL Falcon tubes with 2 mL LB media supplemented with ampicillin (100 μg/mL) and cultured overnight at 37 $^{\circ}$C and 225 rpm for up to 24 hours. Plasmid from the bacterial culture was isolated using miniprep from Qiagen miniprep kit using manufacturer's recommendation. The insert sequence was verified by sanger sequencing using sequencing primers OL497 and OL4006 (refer to table A.1, appendix section). The DNA obtained from clones with the correct insert were re-transformed and re-purified using maxi-prep kit (Invitrogen) and the DNA concentration was adjusted to 1 mg/mL.

2.3 Recombinant protein production using HEK-293-6E cells

All proteins were expressed in HEK293-6E cells maintained in Freestyle media (Life Technologies) supplemented with 50 μg/mL G418 and 0.1 % Kolliphor P188. For routine culture, 2.5 x 10^7 cells were seeded in 500 ml Erlenmeyer culture flasks containing 100 mL culture media and cultured in a shaker set at 37 $^{\circ}$C, 5 % CO_2, 70% humidity and 125 rpm. To maintain a logarithmic growth phase, cells were diluted into fresh media every 2-3 days when the cell density reached approximately 2 $\times 10^6$/mL.

The cells were transfected with plasmids carrying the desired inserts using polyethylenimine (PEI) as the transfection reagent. The cell culture was prepared 48 hours prior to transfection by seeding cells at a density of 5 $\times$ 10^5 cells/mL in 125 mL, 250 mL or 500 mL Erlenmeyer culture flasks with 25 mL, 50 mL or 100 mL culture media respectively. For the production of biotinylated protein, the cells were grown in culture media supplemented with D-biotin (100 μM) [80]. On the day of transfection, cells were counted to ensure that the desired density of 1.5 -2 $\times$ 10^6 cells/mL was reached.

Table 2.1 summarises the amounts of each component added to form DNA/PEI complexes. For each indicated transfection scale reaction, two separate tubes containing transfection media (unsupplemented Freestyle media) were prepared. Next, plasmid DNA was added to the one tube and PEI was added to the other. When expressing biotinylated proteins, plasmid encoding a secreted *E.coli* BirA enzyme was added together with the plasmid DNA in

1:10 ratio. Both tubes were immediately vortexed and the PEI solution was added to the DNA solution. The contents were incubated for 3 minutes at room temperature and the whole mixture was added to the pre-prepared cultures. The cells were grown for 5 days in order to achieve maximum protein yield.

Table 2.1 Summary of reagent quantities for transfection of HEK-293-6E cells.

Transfection scale/volume	Volume of transfection media	Amount of Expression plasmid	Amount of PEI	Amount of BirA plasmid
Small (25 mL)	1.25 mL + 1.25 mL	25 μg	50 μg	2.5 μg
Medium (50 mL)	2.5 mL + 2.5 mL	50 μg	100 μg	5 μg
Large (100 mL)	5 mL + 5 mL	100 μg	200 μg	10 μg

As all the constructs contained either an exogenous or an endogenous signal peptide, all recombinant proteins produced by the cells were released into the culture media. The proteins were harvested by spinning down the culture for 20 minutes at 3320 $\times$g and filtering the cell supernatant through a 0.2 μm filter. Supernatants containing pentameric recombinant proteins were stored at 4 °C until further use.

2.4 Protein purification and quantification

2.4.1 Immobilized metal ion affinity chromatography

The six-times his-tagged proteins were enriched from cell culture supernatants either on HisTrap HP columns (GE Healthcare) using an ÄKTAxpress (GE Healthcare) or on unpacked nickel-sepharose beads. In both case the same binding and elution buffer was used (see section 2.1)

Enrichment of his-tagged proteins using ÄKTAxpress

The culture supernatant containing the proteins were supplemented with 500 mM NaCl and 20 mM imidazole prior to enrichment to decrease non-specific protein binding. The HisTrap column was equilibrated with 20 mL of binding buffer and the protein was slowly loaded at the rate of 1 mL/min. The proteins were eluted from the column with 10 mL elution buffer and all 20 elution fractions were collected (0.5 mL/fraction). The readout from the machine was an absorbance curve based on the 280 nm absorbance of each eluted fraction. The fraction with the highest absorbance was collected to achieve maximum protein yield.

Enrichment of his-tagged proteins using Nickel beads

The culture supernatant containing the his-tagged proteins were enriched using Ni^{2+}-NTA sepharose beads using gravity-flow chromatography with Polypropylene columns (Qiagen). In brief, beads were first added to the supernatant in a 1:1000 ratio (i.e. 50 μL of 50% sepharose slurry into 50 mL supernatant) and binding to the beads was carried out overnight at 4 $^{\circ}$C or for at least for two hours at room temperature on a rotating platform. The beads were washed once with wash buffer before the samples were eluted with 300 to 500 μL of elution buffer.

2.4.2 Determination of protein expression and quality

Determination of protein concentration

The 280 nm absorbance of the enriched proteins was measured with a bench-top spectrophotometer. The extinction coefficients of the proteins were calculated using Snapgene software and the protein concentration was determined using the Beer-Lambert law. This method of concentration determination was carried out only for protiens that had been buffer exchanged into PBS. Buffer exchange of purified proteins were carried out using PD-10 desalting columns (GE-healthcare) using the manufacturers 'Gravity' protocol.

SDS-PAGE

The expression of recombinant proteins were detected by SDS-PAGE under reducing conditions to confirm their size and integrity. Culture supernatants (10-15 μL) or purified protein samples (5-10 μL) were mixed with NuPAGE LDS Sample Buffer (4$\times$) sample buffer NuPAGE Sample Reducing Agent (10$\times$) (Thermo-Fisher) and heat denatured by boiling the mix at 95 $^{\circ}$C for 10 minutes. The entire sample together with pre-stained gel marker were loaded on a 4-12% pre-cast gradient gels (Thermo-Fisher). The gel chamber was filled with NuPAGE running buffer and the separation was carried out for an hour at 200 V. The proteins on the gel were stained with InstantBlue Protein Stain (Expedeon) using manufacturers recommendation.

Western blotting

Following SDS-PAGE, the proteins were transfered to a PVDF (GE Healthcare) membrane using a wet transfer method with NuPage transfer buffer (Life Technologies) supplemented with 10 $\%$ methanol, an XCell II blot module (Novex) and 30 V voltage for an hour at room temperature. The membrane

was blocked with blocking buffer (PBS supplemented with 2.5% milk powder) for 1 hour in order to avoid spurious binding and was then probed with 200 ng/mL of appropriate HRP-conjugated antibody diluted in blocking buffer (anti-FLAG-HRP for pentamers and Streptavidin-HRP for the biotinylated monomers) for 1 hour. Finally, the blot was washed 3 times with PBST (PBS with 0.1% Tween 20). The signals from the proteins were detected on Hyperfilm (GE Healthcare) in the presence of SuperSignal West Pico enhanced chemiluminiescent HRP substrate (Thermo-Fisher). The reagent was used according to manufacturer's recommendation.

Quantification of pentameric proteins with β-lactamase tag

Prey protein expression was quantified by measuring the turnover rate of nitrocefin solution by the β-lactamase enzyme activity associated with the expressed proteins. First, 20 μL culture supernatant containing the prey protein was aliquoted on a 96-well plate. Next, 60 μL of 125 μg/mL nitrocefin solution (Calbiochem) was added to each well. Absorbance readings were taken 15-20 minutes post-substrate addition at 485 nm on a Pherastar plus (BMG laboratories). The proteins were normalised to enzyme activity corresponding to approximately 1 nmol/min, which corresponds to complete hydrolysis of 14.5 nmol nitrocefin in approximately 15 minutes.

Quantification of monomeric proteins with biotin tag with ELISA

Biotinylated bait proteins were detected by enzyme-linked immunosorbent assay (ELISA). Proteins were captured on 96-well streptavidin-coated plates (NUNC) for one hour before adding 10 μg/mL primary antibody recognising the rat Cd4 tag (mouse anti-rat Cd4, clone OX68) common to all recombinant proteins, for another hour. Plates were washed 3x in PBS/0.1% Tween-20 (PBST) before adding 100 μL of an anti-mouse alkaline phosphatase conjugate (Sigma) at 0.2 μg/mL. Plates were washed 3$\times$ PBST and 1$\times$ PBS before adding 100 μL p-nitrophenyl phosphate (Sigma 104 alkaline phosphatase substrate) at 1 mg/mL in diethanolamine buffer. Absorbance readings were taken 15-20 minutes post-substrate addition at 405 nm on a Pherastar plus (BMG laboratories).

2.4.3 Plate-based direct protein interaction assay

A biotinylated 'bait' protein consisting of the entire ectodomain of GABBR2 and controls were first immobilised in a well of a streptavidin-coated 96-well microtitre plate (NUNC) at a concentration that saturated the biotin binding ca-

pacity of the well and probed for direct interactions with the entire ectodomain of IGF2R expressed as a β-lactamase-tagged 'prey'. The plate was washed 2$\times$ in PBST after which normalised β-lactamase-tagged 'prey' (IGF2R and controls) proteins were added to the wells for one hour. Following another wash step (2$\times$ with PBST and final wash with only PBS), 100 μL of 125 μg/mL nitrocefin was added and prey capture was quantified by measuring the absorbance of nitrocefin hydrolysis products at 485 nm on a Pherastar plus (BMG laboratories). Biotinylated Cd4 tag alone was used as a negative control bait and a biotinylated anti-Cd4 monoclonal antibody (anti-prey) used as a positive control as required. Where soluble monosaccharides were used in blocking experiments, prey proteins were first incubated with a range of concentrations (10 mM- 0.04 mM) of mannose-6-phosphate or mannose for one hour, prior to incubation with bait proteins. To remove N-linked glycans from soluble recombinant GABBR2, 1500U of PNGaseF (New England Biolabs) were added to 10 μg of GABBR2 and incubated for duration ranging from 1-16 hours at 37 °C.

2.5 Human Cell line culture

All cell lines except HEL, HEK-293-E and HEK-293-6E were obtained from the Sanger Institute cell line database. HEK-293-E/6E cells were obtained from Yves Durocher [188] and HEL cells were purchased from DSMZ. All cell lines were tested and found to be mycoplasma free. The growth conditions for all cell lines are listed in table 2.2. For some lines, Cas9 expressing versions were made and therefore were grown with supplemented Blasticidin to the indicated concentration.

All cell lines were maintained in a static incubator at 37 °C in humidified atmoshere with and 5% CO_2. All suspension lines were passaged every 2-3 days or when the density reached 1.5 million/mL by diluting the confluent culture with fresh media to obtain a final density of 0.4 million/mL. To passage the adherent cell lines, the cells were first briefly rinsed with 1 $\times$ PBS. Prewarmed (37 °C) Trypsin solution (TrypLE, Gibco) was next added for 5 – 7 minutes and the plate was incubated at 37 °C. Once the cells were lifted off from the plate, Trypsin was deactivated by adding equal volume of complete growth medium. The cells were then seeded at a final density of 0.1 million/mL.

Table 2.2 Growth condition for cell lines used in this study

Cell-line	Origin	Media*/Growth condition	Blasticidin concentration
NCI-SNU-1	Human gastric carcinoma	RPMI- Suspension	20 μg/mL
NCI-SNU-16	Human gastric carcinoma	RPMI-Suspension	20 μg/mL
COLO-320-HSR	Human gastric carcinoma	RPMI-Suspension	20 μg/mL
KBM7	Human Chronic myelogenous leukemia (CML)	IMDM- Suspension	10 μg/mL
HEL	Human erythroleukemia	RPMI- Suspension	15 μg/mL
HL-60	Human promyelocytic leukemia	IMDM- Suspension	15 μg/mL
HepG2	Human liver hepatocellular carcinoma	DMEM/F12- Adherent	20 μg/mL
HEK-293-E	Human embryonic kidney cells 293	Freestyle 293- Suspension	20 μg/mL
SK-MEL-1	Human melanoma	DMEM/F12-Suspension	10 μg/mL
Lu-65	Human lung carcinoma	RPMI-Suspension	15 μg/mL

*The full composition of each media is listed in section 2.1

2.6 Flow cytometry based cell binding assay

Binding assay with streptavidin-PE conjugated proteins

To increase binding avidity, biotinylated monomeric Cd4-tagged proteins were multimerized around streptavidin-phycoerythrin (PE). To ensure all biotin binding sites on the streptavidin were occupied and yet to minimise the presence of excess monomer, serial dilutions of biotinylated protein samples were titrated against a fixed concentration of streptavidin-PE (100 μL of 0.1 μg/mL) for 20 minutes at room temperature before transferring to a streptavidin-coated plate and assaying for the capture of any excess biotinylated Cd4-tagged proteins using the OX68 ELISA. The minimal dilution at which all biotinylated Cd4-tagged protein was captured was subsequently used to create tetramers. A 10$\times$ tetramer staining solution was prepared using 4 μg/mL streptavidin-PE and the appropriate biotinylated protein dilution by incubating for 30 minutes at room temperature, then diluted to 1$\times$ and 100 μL added to 0.5-1$\times 10^6$ cells in an U-bottomed 96-well microtiter plates and incubated for one hour at room temperature. Where the anti-BSG antibody was used in blocking experiments, cells were first incubated with 10 μg/mL antibody (or isotype matched control) for one hour, prior to incubation with the RH5-streptavidin-PE complex. Cells

were washed once with wash buffer (PBS with Ca^{2+}/Mg^{2+} (Hyclone, Sigma) supplemented with 1 % BSA) and analysed by flow cytometry.

Binding assay with pentameric proteins

The 3$\times$ FLAG and β-lactamase-tagged pentameric proteins were quantified directly from supernatants and normalised to approximately 1 nmol/min using the β-lactamase enzyme activity. Next, 100 μL of diluted proteins were added to 0.5-1$\times10^6$ cells in a U-bottomed 96-well microtiter plates for 1 hour at room temperature. Following a wash with the wash buffer, 100 μL PE-conjugated anti-FLAG antibody (0.5 μg/mL, Abcam) was added to the samples and incubated for 1 hour. The cells were again washed once in wash buffer and analysed by flow cytometry.

Binding assay with antibody specific to the cell surface proteins

For antibody staining of cell surface proteins, 50 μL of 1 μg/mL primary antibody was incubated with 1$\times10^6$ cells in 96-well U bottom plates. The cells were washed after 1 hour of primary antibody incubation after which 100 μL of an appropriate secondary antibody, also conjugated to PE, was used at 0.1 μg/mL.

Binding assay with transiently transfected cells

Human IGF2R was expressed on the surface of transfected cells using an expression construct in which its cytoplasmic region was replaced by eGFP, as previously described [85]. NCI-SNU-1 cells, which do not have detectable levels of plasma membrane IGF2R expression, were transiently transfected with either IGF2R-TM-eGFP or CD200R-TM-eGFP as a control, and probed for binding interactions with either GABBR2 ectodomain presented as a tetramer around streptavidin-PE or with an anti-IGF2R mAb.

All flow cytometry was performed on a Becton-Dickinson (BD) LSR Fortessa flow cytometer, collecting between 10,000 to 30,000 events; live cells were gated using forward and side scatter. PE was excited at a wavelength of 561 nm and emission detected using a 582/15 band pass filter; BFP was excited at 405 nm and the emission detected using a 450/50 band pass filter. Analysis was performed using FloJo software (Treestar Inc.)

2.7 Genome-wide screening and validation

2.7.1 Construction of gRNA expression vector

The Human Improved Genome-wide Knockout CRISPR Library v1 consisting 90,709 sgRNAs targeting 18,010 human genes (Addgene: 67989), lentiviral Cas9 reporter plasmids : pKLV2-U6gRNA5(gGFP)-PGKBFP2AGFP-W (Addgene: 67980) and pKLV2-U6gRNA5(Empty)-PGKBFP2AGFP-W (Addgene: 67979), lentiviral vector expressing Cas9 fused with the Blasticidin resistant gene at the C-terminus pKLV2-EF1a-Cas9Bsd-W (Addgene: 68343) and lentiviral CRISPR gRNA expression vector pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W (Addgene: 67974) were obtained from Kosuke Yusa [176].

To target a particular gene of interest, 20 base pair constructs were cloned into the lentiviral CRISPR gRNA expression vector. For each of the targeting sequence, BbsI sites were introduced and the oligos were ordered from Sigma (table A.2, appendix section). Sense and antisense 24-nt oligonucleotides (10 mM each) were mixed in oligo annealing buffer (10 mM Tris-HCl (pH8.0) and 5 mM $MgCl_2$) in a total volume of 100 μL. The mixture was incubated at 95 $^\circ$C for 5 min and cooled to room temperature. The double stranded oligonucleotides were then cloned into the BbsI site of the backbone vector.

2.7.2 Lentiviral production

Lentivirus was produced by transfection of HEK-293-FT cells. These cells were cultured in RPMI 1640 media supplemented with 10% FBS. The cells were seeded one day prior to transfection such that they would be approximately 80% confluent on the day of transfection. For the generation of lentivirus from transfection of cells in 10 cm dish, 3 μg of a lentiviral vector, 9 μg of ViraPower Lentiviral Packaging Mix (Invitrogen) and 12 mL of the PLUS reagent were added to 3 mL of OPTI-MEM and incubated for 5 min at room temperature. 36 mL of the LTX reagent was then added to this mixture and further incubated for 30 min at room temperature. The transfection complex was added to the cells and incubated for at least 4 hours after which the media was replaced with fresh media. Media was again refreshed after 18 hours and finally 48 hours later, the viral supernatant was harvested and stored at –80 $^\circ$C.

2.7.3 Lentivirus transduction

Lentivirus transduction for stable Cas9 line generation

All Cas9-expressing human cell lines were selected following transduction of cells with lentivirus prepared from the pKLV2-EF1a-Cas9Bsd-W plasmid. Polybrene (8 μg/mL) was added for the transduction of all cell lines expect HEK293-E cell lines. Cells were selected using the indicated blasticidin concentration (see table 2.2) two days following transduction. Clonal high Cas9 activity cell lines were established by sorting individual blasticidin-resistant cells into wells of 96-well plates (MoFlo XDP) which were further expanded and tested for Cas9 activity using the GFP-BFP system [176]. In brief, cells were transduced with lentivirus encoding GFP, BFP and a gRNA targeting GFP (pKLV2-U6gRNA5(gGFP)-PGKBFP2AGFP-W) or the same construct with an 'empty' gRNA (pKLV2-U6gRNA5(Empty)-PGKBFP2AGFP-W) as a negative control. High activity Cas9 stable cell lines were selected by examining the ratio of BFP only to GFP-BFP-double positive cells transduced by the two lentiviruses. These clonal cell lines were expanded and further tested by targeting an endogenous gene encoding the BSG cell surface protein using lentivirus prepared using a plasmid encoding puromycin, BFP and a gRNA targeting *BSG* (pKLV2-U6gRNA5(BbsI-gBSG)-PGKpuro2ABFP-W). The surface expression of BSG was quantified by flow cytometry using an anti-BSG mAb (MEM-M6/6) eight days post-transduction to validate high Cas9 efficiency.

Lentivirus trasduction of HEK-293-E cells for knock-out library generation

A genome-scale 'knockout' library of HEK-293-E-Cas9 cells was produced by transducing 3×10^7 cells such that $\sim$30% of the total cell population were transduced to increase the chances that each cell just received a single gRNA. The transduced (BFP positive) cells were harvested three days after transduction using a cell sorter (MoFlo-XDP), and libraries containing at least 5x10^6 cells were selected. The libraries were cultured in media containing 2 μg/mL puromycin to remove the non-transduced cells and at every passage, at least 10$\times$ the initial library (starting cell number on day three) for each library were seeded into new culture flasks, thereby generating a sampling size of at least 50x10^6 cells for each screen. Phenotyping screens for cell surface binding events were carried out between 9 and 16 days post transduction.

Lentivirus transduction of other human cell lines for knock-out library generation

A spinoculation protocol was used to infect other human cells as they were more difficult to transduce than HEK-293-E cells. For this, 2 mL of 5$\times$ 10^6 cells/mL were aliquoted in 8$\times$ 15 mL Falcon tubes, and mixed with lentivirus together with 8 μg/mL polybrene and incubated at room temperature for 30 minutes followed by centrifugation for 100 min at 800$\times$g at 32 $^\circ$C. The supernatant was removed, and the cells from each Falcon tube were resuspended in 50 mL culture media. As with HEK-293-E cells, cells were sorted on day three post transduction to generate control and sample libraries and grown further in media supplemented with 1 μg/mL puromycin.

Lentivirus trasduction of human cells for a targeted gene knock-out

For targeted gene knockout in cancer cells, 1$\times$ 10^6 cells were seeded in one well of a 6-well plate. 100 μL virus was added to the cells for at least 5 hours. Polybrene (8 μg/mL) was added to all cells except HEK-293-E cells during virus infection. The virus containing media was removed after the infection period and fresh culture media was added to the cells. 48 hours post infection, the cells were selected with puromycin (2 μg/mL for HEK-293-E cells and 1 μg/mL for all other cell lines). Polyclonal lines were used for initial validation of the screen results. For the generation of clonal knockout lines of *SLC35B2* and *IGF2R*, the transduced cells were left to grow under selection for additional 10 days after which, the cells were single-cell sorted into 96-well plates (BD-MoFlo-XDP). Clonally derived lines of IGF2R were analysed for gene knockout using antibody staining with an anti-IGF2R antibody, from which clones lacking surface staining of IGF2R were chosen. Gene disruption of *SLC35B2* was verified by first isolating genomic DNA from 500,000 cells using a commercial kit (Blood and tissue miniprep kit, Qiagen), and then amplifying approximately 300 base pairs on either side of the expected cut site (gRNA targeted region) with the following primers: forward primer- 5‘ TGCTGCAGGAAAGAGTGATGACC3’; reverse primer- 5‘GCATGGGCAGCAAACTCACT3’. PCR products were sequenced by Sanger sequencing. Insertions and deletions were analysed by sequence alignment and Tracking of Indels by DEcomposition (TIDE) [189].

2.7.4 Cell surface phenotyping, selection and amplification of selected gRNAs

Genome-scale knockout libraries were phenotyped by cell surface staining using flow cytometry between 9 and 16 days post-transduction. The mutant library was divided into two parts: at least 5×10^7 cells from the mutant library were collected as 'control' population for later analysis whereas 5×10^7 - 15 $\times10^7$ cells from the library were stained with appropriate reagent (recombinant protein or antibody) using the binding assay protocol as described above with minor modifications: cells (5×10^6 cells/mL) were stained in 15 mL Falcon tubes with gentle rotation (6 rpm), and the stained cells were then analysed using an flow sorter (BD-MoFlo-XDP) and the BFP+/PE- cells were collected. The percentage of the total library population that was collected in each screen varied between 0.2% to 4%. Sorting threshold and the approximate number of cells sampled and collected for each screen is listed in appendix table A.3. All genetic screens performed in this study were carried out once.

Amplification of gRNA from control mutant cells

The human genome-wide lentiviral gRNA library has a high complexity as it contains approximately 91,000 different sequences. To capture every guide in the library, each mutant library was composed of at least 50 million transduced cells (500x of the gRNA library). However, it is not feasible to extract gRNA from all of 50 million cells therefore, gRNAs from at least 10 million cells (approximately 72 μg) were sampled from the control library. The genomic DNA from the cells were isolated with a commercial kit (Blood and tissue Maxi kit, Qiagen) using manufacturer's recommendation. gRNA was then amplified using the isolated genomic DNA (2 μg DNA/PCR reaction), primers (U1 and L1, refer to table A.4, appendix section) and Q5 Hot Start High-Fidelity 2 $\times$ Master Mix. The detailed protocol for PCR amplification of gRNAs from genomic DNA is provided in the appendix section A.1. The PCR products were cleaned with Qiagen PCR purification kit.

Amplification of gRNA from screened cells

For samples where the sorted cell number was less than 100,000, a cell lysate protocol was used to isolate guides prior to PCR enrichment. Cell lysates were prepared from sorted cells by boiling samples (10,000 cells/sample) at 95 °C with 25 μL water for 10 minutes. Next, 5 μL of 2 mg/mL freshly diluted Proteinase K was added to each well for 1 hour and incubated at 56 °C, after

which the enzyme was inactivated by boiling the sample for 10 minutes at 95 °C. The gRNAs were then amplified using 10 μL cell lysates/PCR reaction. For larger cell number, genomic DNA was isolated using a commercial kit (Blood and tissue miniprep kit, Qiagen) using manufacturer's recommendation. 100 ng DNA was used for each PCR reaction and the guides were amplified with the same L1 and U1 primers.

Illumina Hi-Seq sequencing of gRNA

The PCR amplified guides were diluted to 40 pg/μL in EB and tagged with illumina index primers using second round of PCR with using 200 pg template, PE 1.0 as forward primer, appropriate index tags as reverse primers and KAPA HiFi HotStart ReadyMix polymerase (also refer to appendix section A.1 for detailed protocol and table A.4 for primer information). The PCR products were cleaned using SPRI beads (Agencourt AMPure XP beads). When the samples were multiplexed, qPCR was conducted by the Sanger DNA pipelines to load equivalent molar quantities of differently tagged samples. 4 nM of multiplexed sample was prepared and loaded Hi-Seq 2500 rapid run sequencing machines and single-end sequencing (19bps) was performed with the custom sequencing primer, 5'-TCTTCCGATCTCTTGTGGAAAGGACGAAACACCG-3'. Approximately 300 million reads were obtained from each sequencing run in two lanes of Hi-seq 2500. Between 10-15 samples were multiplexed in each sequencing run, providing a read depth of at least 20 million for each sample.

2.7.5 Data analysis

The read count for each gRNA and gene level enrichment analysis was carried out using the MAGeCK statistical package (version, v0.5.5) [190]. Briefly, the software package was used to first convert the fastq files obtained from sequencing machines into a count file (csv format) using the "–count" function, which maps the 19 bps sequencing reads to the sequences of each gRNA. Thus obtained count files contained the read counts for each gRNA, which were used in the "–test" function to perform the gRNA and the gene ranking for both negative and positive selections. In this step, MAGeCK first performs a median normalisation of the read counts from treatment sample (sorted population in this case) and control sample. Next, a negative binomial (NB) model is applied to investigate the significant differences in the gRNA abundance between the control and treatment samples. The individual gRNAs are ranked based on the on P-values calculated from the NB model and the enriched

genes are identified using a modified robust ranking aggregation (RRA) algorithm which investigates whether gRNAs targeting the genes are consistently ranked higher than expected. Pathway analysis was also carried out using the RRA algorithm of the MAGeCK software using the ranked gene list with default settings and KEGG annotated pathways [191]. Full documentation regarding installation of MAGeCK and application of each function is provided in `https://sourceforge.net/p/mageck/wiki/usage/`. All further analysis was carried out using R [192]. Gene annotations were obtained from Uniprot [193].

CHAPTER 3

DEVELOPMENT OF A CRISPR-CAS9 BASED KNOCK-OUT SCREEN FOR CELLULAR RECOGNITION

3.1 Introduction

This chapter describes the design of a genome-scale knockout (KO) screening approach using the CRISPR-Cas9 system in human cell lines. The approach is subsequently used to screen a panel of monoclonal antibodies targeting structurally diverse membrane receptors to assess the potential of the approach and to determine the experimental parameters to identify cellular factors mediating cellular recognition events at the cell surface.

3.1.1 Monoclonal antibodies are ideal probes for the study of cell surface recognition

To first assess the feasibility of genome-scale KO screens for the identification of factors required for cell surface recognition events, I focused initially on the interactions mediated by monoclonal antibodies (mAbs). mAbs have already been established as powerful tools to investigate various aspects of receptor biology as exemplified by their use in the past to isolate, localise, and biochemically characterise membrane receptors, as described in Chapter 1. As mAbs are selected to bind specifically to a single epitope within a receptor with a high avidity (K_{Ds} in the nM to pM range), the binding of mAbs on a cell surface that abundantly expresses the corresponding antigen often leads to a bright staining with a high signal to noise ratio. In a CRISPR-Cas9 screen, this can be the basis for Fluorescence Activated Cell Sorting (FACS)-based phenotypic selection, allowing identification of factors, which if disrupted causes a decrease in the cell surface expression display of the mAb epitope. Strong phenotypic selections are usually desired in a pooled screening approach as they allow for the identification of candidate genes with

high confidence, thus the use of mAbs as probes will be ideal to determine the parameters of CRISPR-KO screens.

3.1.2 Genome-scale knockout screening approach has the potential to determine the specificity of mAbs.

The use of mAbs in a genome-scale KO screening approach in itself has the potential to become a novel means of determining the specificity of a given antibody. High quality antibodies should recognise their targets specifically but the batch-to-batch variability of the commercially available antibodies can lead to poor specificity, and has caused concerns in the research field [194, 195, 196, 197]. The need for target validation of mAbs is even higher now as there is a growing use of mAbs as therapeutics in human diseases such as cancer, rheumatoid arthritis and Crohn's disease [198]. Common methods to characterise antibody specificity involve immunoassays such as immunohistochemistry (IHC), immunoprecipitation (IP), Western blotting (WB), and mass spectrometry (MS) analysis [199, 200]. However, such approaches are not always applicable, especially if the mAbs are raised against the correctly folded protein epitopes such that they do not react to denatured or detergent-solubilised antigens. The genetic strategies discussed earlier in Chapter 1 to identify binding partners, such as the gain-of-function approach using a cDNA library and the loss-of-function screens using RNAi-mediated knockdown, have been used to identify targets of mAbs that recognise folded epitopes that are correctly expressed on the surface of cells [201, 202, 203]. However, the high off-target effects of RNAi systems that often lead to inconsistent results and the resource- and time-intensive nature of expression cloning approaches using cDNA libraries pose potential challenges in their applications. Genetic screening methods using the CRISPR-Cas9 KO system holds potential for identification of receptors targeted by mAbs, which could be valuable in the field of antibody characterisation. This will be explored in detail in this chapter.

3.1.3 Considerations for knockout screening approach to identify directly interacting receptors

One of the considerations for the use of genome-scale KO screens to identify directly interacting receptors is the possibility that cellular processes that contribute to general protein transport, would dominate the identified genes in all the screens and decrease the likelihood of identifying the direct receptor. Some of the cellular factors that transport plasma membrane-destined membrane

receptors from the endoplasmic reticulum (ER) lumen via the Golgi to the cell surface include: proteins of the signal recognition particle (SRP)-dependent protein targeting pathway, which mediates co-translational targeting of newly synthesised polypeptides into the ER; enzymes within the ER, which modify the polypeptides through signal peptide cleavage and initial glycosylation (also called 'core glycosylation'); and proteins that package folded polypeptides into specific vesicles for intracellular transport between organelles [204]. The disruption of these factors would affect the membrane expression of a large number of membrane proteins and therefore these factors are expected to be identified in the majority of the screens. However, many genes involved in these pathways (specifically the genes encoding the SRP-dependent protein translocation pathway proteins and the core glycosylation pathway proteins) are also known to be essential for the cells, and multiple negative selection CRISPR-KO based screens have shown that gRNAs targeting these genes usually drop-out of the mutant pool as the cells are continually grown [175, 176, 177, 205]. This would instead decrease the likelihood of these genes dominating the identified hits in the screen. In this chapter, I investigate how the experimental parameters in terms of the timing and the stringency of selection can influence the identification of genes that generally contribute to receptor expression versus genes that encode the directly interacting receptor.

3.1.4 Knockout approach used in this study

A dual vector approach is used in this study to generate genome-wide mutants. In this approach, stable Cas9-expressing cell lines are first created and then transduced with the lentiviral knockout library to generate the cell mutant library. The lentiviral library is generated from a plasmid pool obtained from Kosuke Yusa ('Yusa library'). This library consists of 90,709 gRNA targeting 18,009 genes (approximately five gRNAs/gene). The library was designed with features that have been shown to improve gRNA efficacy, such as the improved scaffold (iscaffold) on the gRNA. The conventional scaffold of a chimeric gRNA consists of a stretch of thymidines (T), which acts as a pause signal for RNA polymerase III that can potentially reduce the transcription efficiency [206]. The improved scaffold on the Yusa library is longer than the conventional scaffold by five nucleotides and avoids T stretches by mutating a single T residue within the stretch. The library has already been used for the study of essential genes in cells for the identifcation of genetic vulnerabilities in acute myeloid leukemia [176].

3.1.5 Scope of this chapter

In this chapter, I first describe the generation of a 'toolkit' required for setting up a genome-scale KO screening technology. This toolkit consists of: (a) human cell lines expressing highly efficient Cas9, and (b) a quality-controlled genome-scale library generated from the Yusa plasmid library. I next proceed to describe the optimal strategy for FACS-based phenotypic selection, before going on to describe multiple screens carried out with mAbs targeting structurally diverse membrane receptors, to identify cellular factors required for mAb binding to the cell lines. Finally, I will summarise the lessons learnt from this initial study.

3.2 Results

3.2.1 Generation of stable cell lines expressing Cas9

Determination of protein turnover time

To carry out CRISPR-Cas9 mediated KO screens, I first generated a stable Cas9-expressing HEK-293-E cell line using lentiviral transduction. I next designed a method to measure the Cas9 efficiency by making a gRNA construct expressing a single gRNA targeting *BSG*, which could be introduced to Cas9-expressing cells via lentiviral transduction (scheme for vectors used in figure 3.1A). Using this system, the decrease in cell surface expression of BSG in the transduced cells is readily established by flow cytometry using an anti-BSG mAb, which is then used as the measure of Cas9 efficiency. This flow-cytometric approach allows for rapid assessment of the efficiency of the individual cells in a pool rather than efficiency of the 'bulk' population. An important consideration in using this system is to wait long enough to allow complete protein turnover so that the previously transcribed mRNA and the translated protein is degraded. To determine the earliest timepoint where loss of the protein could be observed, polyclonal Cas9 cells infected with gRNA targeting *BSG* were checked for cell surface BSG expression on days 6, 8,10, 15 and 20, post transduction. While only a very small loss of cell surface BSG was observed at day six post infection, a clear population lacking the surface staining for BSG was observed from day eight onwards, demonstrating the expression of functional Cas9 in the cell line. However, approximately 25% cells retained surface expression of BSG on day eight and this knockout refractory population did not change even when examined at day 20 (figure 3.1B).

To assess whether eight days was a typical turnover time for other surface receptors, three more guides targeting *SDC1*, *CD55*, and *CD44*, as well as an empty construct with no targeting gRNA were cloned in the BbsI site of the reporter vector and was used to infect the Cas9-expressing HEK-293-E cells. The level of the respective surface proteins was also tested on day eight post transduction. A clear loss of the surface proteins were observed in all cases. Cells refractory to gene knockout were also observed in all cases (figure3.1C). Collectively, day eight was determined to be the earliest time-point that could be used to test the efficiency of Cas9 using the endogenous gene knockout system.

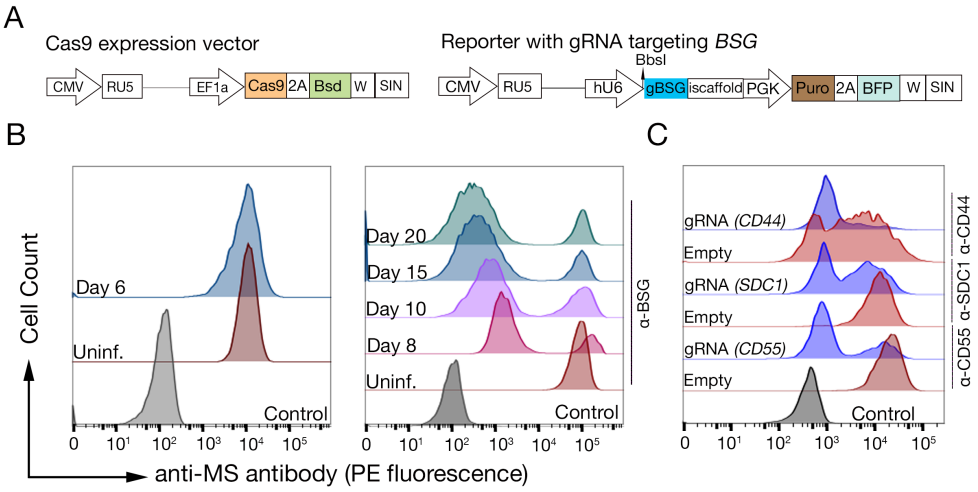


Fig. 3.1 Time- dependent decrease in cell surface expression of membrane receptors is observed in Cas9-expressing cells that are transduced with gene specific gRNAs. **A.** Schematics of lentiviral vector encoding Cas9 together with the blasticidin resistant gene at the C-terminus used to create stable Cas9 expressing lines (left panel) and Cas9 activity reporter with an improved scaffold and puro/BFP markers targeting *BSG* (right panel). **B.** Viruses with the activity reporter were used to quantify genome editing efficiency in Cas9-expressing HEK-293-E cell line. A small decrease in the surface expression of BSG was seen on transduced cells on day six post transduction. From day eight onwards, a very clear double population was observed. Approximately 25% of cells retained BSG expression on the cell surface by day eight and the frequency of this population did not change even at day 20. **C.** The surface expression of three other cell surface receptors (CD55, SDC1, and CD44) also decreased on day eight. Control: parental cell line stained with secondary anti-mouse(ms)-PE alone; Uninf: parental cell line stained with anti-BSG antibody, Empty: parental cell line stained with the indicated primary antibody. Both **B** and **C** are representative of two technical replicates. CMV: Cytomegalovirus promoter, RU5: 5' long terminal repeat lacking the U3 region, EF1a: intron-containing human elongation factor 1a promoter, Cas9: codon-optimised *Streptococcus pyogenes* Cas9, Bsd: Blasticidin resistant gene, W: Woodchuck Hepatitis Virus posttranscriptional regulatory element, SIN: ΔU3RU5 (self-inactivating 3' LTR); hU6: human U6 promoter, gBSG: gRNA targeting *BSG*, Iscaffold: Improved scaffold, 2A: self cleavage peptide, PGK: PGK1 promoter, puro: puromycin-resistant gene, BFP: blue fluorescent protein.

Cloning Cas9-expressing cell lines increases Cas9 cleavage efficiency

Having established the protocol to measure Cas9 efficiency, I proceeded to make stable Cas9 lines for a cell line originating from a different tissue source (colorectal carcinoma, NCI-SNU-1). The level of BSG expression on the surface of cells was analysed eight days post-infection by flow cytometry using an anti-BSG mAb. While the Cas9-mediated gene inactivation was evident from loss of BSG from the cell surface upon infection with gRNA targeting *BSG*, approximately 30% of cells that retained BSG expression were again observed

(figure 3.2B). One of the reasons for this incomplete knockout of *BSG* is likely to be the variability in expression of efficient Cas9-nuclease in a polyclonal cell population. In a polyclonal Cas9 line, it has been shown that some cells acquire mutations in the proviral Cas9-coding sequence with APOBEC3 mutational signature that can lead to Cas9 inactivation and decrease the overall efficiency [176]. Such inactivations within a population can potentially cause problems in genome-scale screens where Cas9 is assumed to be 100% efficient and the incorporation of a gRNA in a cell is equated to the complete loss of the targeted gene product. To reduce the heterogeneity and to select the cell with the highest Cas9 efficiency, single cells of the polyclonal line were sorted using a MoFlo-XDP sorter into 96-well tissue culture plates. Colonies appeared approximately two weeks later and as an initial test, 23 clones were tested. All except one clone had an increased population of cells that had lost surface BSG compared to the polyclonal line (five best clones depicted in figure 3.2B). The clone with the largest refractory population (clone 22) showed a broad surface expression profile with no clear population lacking cell surface BSG (figure 3.2C). In a polyclonal Cas9 line, cells with lower cleavage efficiency would decrease the overall Cas9 efficiency.

The loss of cell surface BSG expression as an indicator of Cas9 activity is useful for estimating efficiency for targeting endogenous genes but is time-consuming requiring steps involving antibody staining, and waiting for at least eight days to ensure complete protein turnover. A faster method (a GFP-BFP system) to check Cas9 efficiency has been developed using an exogenous system in which cells are transduced either with a construct expressing GFP with a gRNA targeting *GFP* or an empty gRNA as a control (vector schematics in 3.2 D) [176]. The expression of GFP can be monitored as early as three days post-transduction to determine the nuclease cleavage efficiency. To directly compare the *BSG* KO method with the *GFP* KO method, the five best clones of NCI-SNU-1 cells selected for the lowest refractory population of BSG surface expression were checked with the GFP-BFP system. All five clones showed a very high efficiency of *GFP* targeting compared to the polyclonal line (figure 3.2D). The fraction of cells from the polyclonal line that remain refractory to gene KO was lower using this system compared to the BSG KO system (6% compared to 30%). That said, the clone with the lowest refractory population under the *BSG* KO system (clone 4) also had the lowest refractory population with the GFP-BFP system.

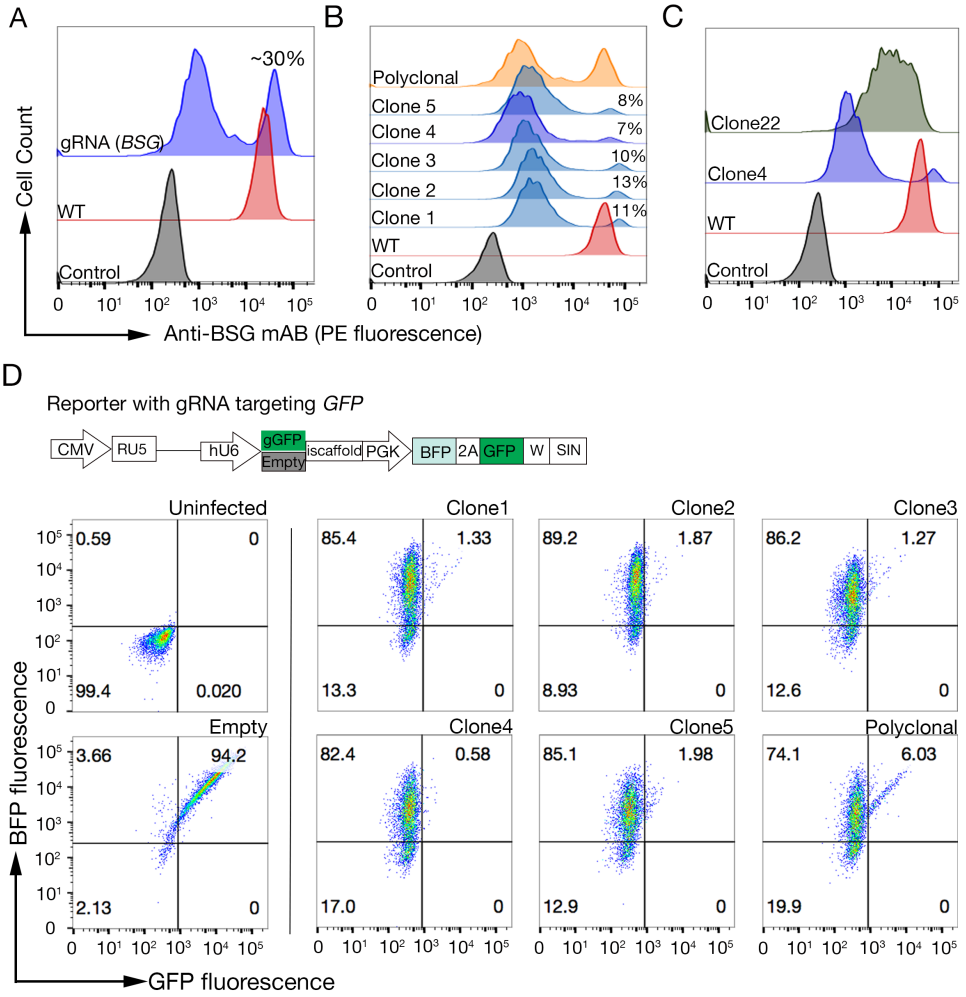


Fig. 3.2 Selecting clonal cell lines with high Cas9 activity for efficient genome-scale genetic screening. **A.** The Cas9 activity of polyclonal Cas9-expressing NCI-SNU-1 was tested by lentiviral transduction with a *BSG*-KO reporter: approximately 70% decrease in surface expression of BSG was observed. **B.** Clones of stable Cas9-NCI-SNU-1 showed variation in the fraction of the population that remained refractory to *BSG* inactivation. All five clones depicted a population that was approximately threefold lower compared to that from the polyclonal line. **C.** Comparison between the best and the worst clones from the same polyclonal line (23 clones tested). Clone 22 showed a small decrease in surface expression of BSG with no evident negative population as observed in clone 4. Control: parental NCI-SNU-1 cell line stained with secondary anti-mouse-(ms-) PE alone; WT: NCI-SNU-1 cell line stained with anti-BSG antibody. **D.** Viruses with a gRNA-targeting plasmid encoded GFP or without ('empty') (schematic depicted) were used to quantify genome editing efficiency of both polyclonal and cloned lines of Cas9-expressing NCI-SNU-1 cell line. The cells were tested for both BFP and GFP expression by flow cytometry after infection with lentivirus, or left uninfected. BFP expression marks transduced cells and the loss of GFP expression was used to quantify Cas9 activity. The profile for uninfected and 'empty' infected cells looked similar for all clones; representative profiles are depicted in the left panel. All five clones of the NCI-SNU-1 cell line show a higher loss of GFP compared to the polyclonal line (right panel), with clone 4 having the lowest refractory population.

Taken together, the GFP-BFP system used to assay Cas9 efficiecny correlated well with the *BSG* KO system, but a direct comparison between the assays suggested that the GFP-BFP system overestimates the Cas9 efficiency, whereas the *BSG* KO system provides a more realistic estimation of the efficiency of the targeting of endogenous genes. Taking this into consideration, the strategy I subsequently employed to create other stable Cas9-expressing lines was to initially use the GFP-BFP system to clone the high efficiency lines, and then to check the clone with the highest GFP cleavage efficiency with the *BSG* KO system to obtain a more accurate estimate of Cas9 efficiency. Five clones of HEK-293-E and HEL cells were each tested, and the clone with the lowest refractory population expressing GFP upon infection with gRNA targeting *GFP* was chosen to re-test with the *BSG* KO system (figure 3.3A). The chosen clones of HEL and HEK-293-E cell lines exhibited approximately 91% and 92% loss of BSG from the cell surface respectively, when transduced with a gRNA targeting *BSG* (figure 3.3B).

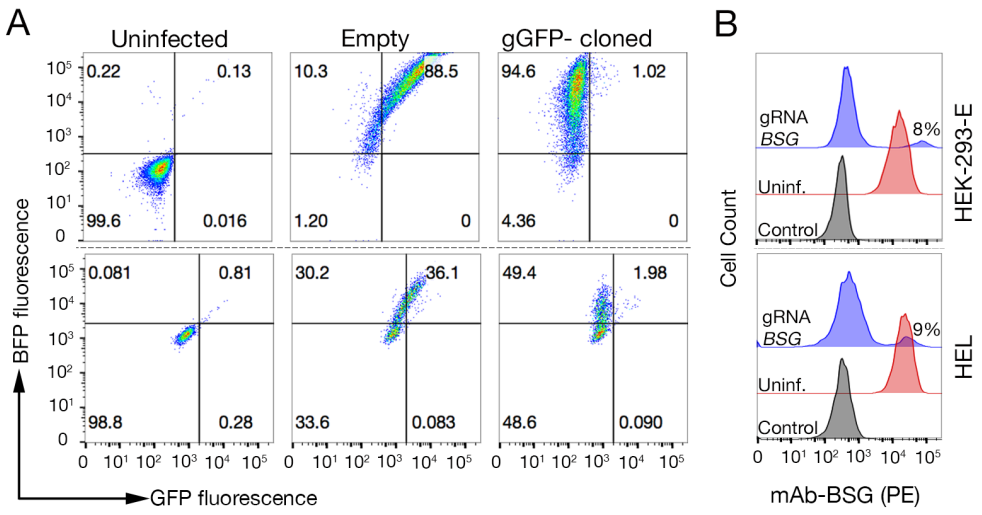


Fig. 3.3 Cas9-expressing human cell lines generated in this study demonstrate high *GFP* and *BSG* cleavage efficiency. A. Cloned Cas9-expressing HEK-293-E and HEL cell lines were tested for both BFP and GFP expression by flow cytometry after transduction with GFP-BFP reporter viruses, or alternatively left uninfected. BFP expression represents transduced cells (note the transduction rate for HEL cells was lower compared to the HEK-293-E cells). As before, the loss of GFP expression is used to quantify Cas9 activity; the line exhibiting the greatest loss of GFP expression out of at least five clones tested are shown. **B.** The cloned cell lines selected using the GFP-BFP assay were additionally tested for their ability to target an endogenous locus by infecting them with a lentivirus encoding a gRNA targeting *BSG* and quantifying expression of BSG on the cell surface with an anti-BSG mAb; both cell lines exhibited loss of BSG from >90% of the population. Control: parental cell line stained with secondary anti-ms-PE alone; Uninf: parental cell line stained with anti-BSG antibody.

3.2.2 Quality control of the genome-scale mutant cell library

Reproducible mutant cell libraries with a uniform coverage of the gRNAs can be created from plasmid gRNA library

I next set out to explore the optimal parameters to perform a pooled KO screen using the Cas9 cell lines that were generated. In the intended genome-scale screening setup, in which approximately 91,000 gRNAs are used, the library complexity plays a role in determining the success of a screen. The population of cells that are to be screened should be large enough to capture all gRNAs in the library. Usually, libraries that have high complexity (500-1000-fold representation of each guide) are used to screen for the desired phenotype. To obtain this representation, it is common to transduce 30-100 million cells at MOI of 0.3 to reduce the chance of more than one gRNA per cell [207]. The transduced cells make up the initial library. As a guiding principle, a library at least five fold larger than this library should be maintained while passaging the cells over the time that is required for the complete turnover of the proteins (at least eight days).

To ensure that gRNA representation is maintained using the cellular KO library preparation protocol, I first quantified the individual gRNA abundances in at least 5×10^7 cells on different days after transduction from two independent NCI-SNU-1-Cas9 libraries and one library of both HEK-293-E-Cas9 and HEL-Cas9 cells. The gRNA abundances in the libraries were quantified by deep sequencing and a high correlation between the biological replicates of NCI-SNU-1-Cas9 libraries and amongst the three different cell line libraries at equivalent time points was observed (figure 3.4A). This indicated that using this protocol, it was possible to reproducibly create mutant cell libraries with good representation of all gRNAs. Compared to the correlation between the different mutant cell libraries, the correlation between the original plasmid population and the mutant cell libraries was lower. This suggested that the most appropriate control for gRNA enrichment analysis was the cell line on that particular day rather than the population of gRNAs in the original plasmid population.

To further investigate the difference between the plasmid pool and the mutant cells in terms of their gRNA abundance, I next analysed the distribution of all gRNAs in each cell line and compared it to the distribution in the plasmid pool. In the plasmid library, while a small fraction of gRNAs were under- or over-represented, approximately 82% of gRNAs were uniformly distributed

with only 8-fold increase in abundance between the 10^{th} and 90^{th} percentiles. The mutant libraries generated from this plasmid library also showed a uniform coverage, but a small drop in the overall representation of the gRNA library was observed in all cell lines; this decrease was more evident for libraries on day 16 compared to those on day 9. (figure 3.4B).

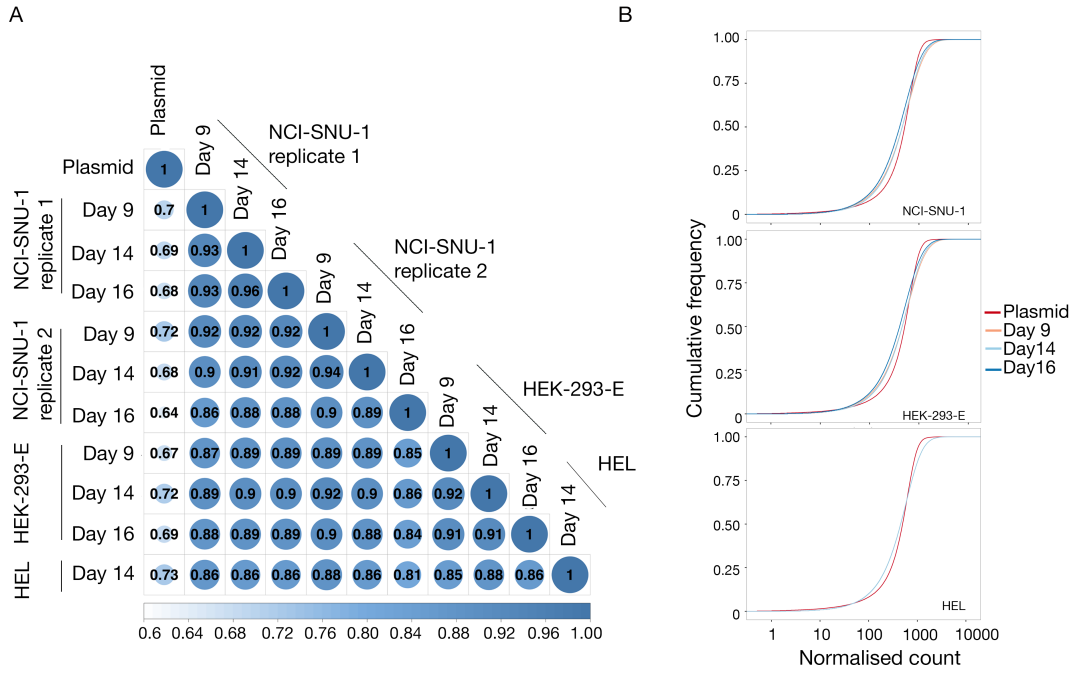


Fig. 3.4 Mutant cell libraries can be created reproducibly by transduction with the lentiviral gRNA library. A. Matrix plot depicting the correlation between normalised read counts of all gRNAs in the original plasmid library and gRNAs in the indicated cellular mutant libraries on the indicated days. Correlation between biological replicates of NCI-SNU-1 and between any cell line on any day was higher than when compared to the original plasmid pool. **B.** Cumulative distribution function plots comparing the gRNA abundance in the plasmid library to the mutant libraries of HEK-293-E and NCI-SNU-1 cells on day 9, 14 and 16 days post transduction. The differences in the curves from the mutant cell populations represent the depletion in a subset of gRNAs compared to the plasmid library.

Negative selection screens reveal essential genes

In a negative selection screen, in which the mutant cells are continually grown for an extended period of time, the cells that carry mutations in the genes that are essential for cell proliferation will deplete, which makes the corresponding gRNA to 'drop-out'. Therefore, when comparing the mutant gRNA abundance to the plasmid gRNA abundance, gRNAs that target fundamental cellular processes should, in principle, be depleted. These screens are sensitive to the library representation as losing the representation of gRNAs due to technical

reasons- such as the need to passage the library because of cell growth, which may create population restriction points, reducing gRNA library representation. Thus, investigating the depleted genes under negative selection can provide valuable insights into the quality of the mutant library.

To check whether the depleted gRNAs targeted essential genes, I next carried out a gene-level negative selection enrichment analysis to identify genes that were depleted in the mutant library compared to the plasmid library. As a quality control, I first analysed ribosomal genes (annotations from KEGG-Ribosome), which are known to be essential and are often identified robustly in similar negative selection screens [175, 176, 177]. Reassuringly, the majority of the ribosomal genes were amongst the most significantly depleted genes (False discovery rate (FDR) <10%) in all three days and in both HEK-293-E and NCI-SNU-1 cell lines (figure 3.5A). Next, I carried out pathway analysis using KEGG annotated pathways; among the most depleted pathways were essential biological processes such as the spliceosome, cell cycle, proteasome and DNA replication processes. The number of relevant genes ('hits') in these pathways was similar between the biological replicates and also between cell lines (figure 3.5B). These results provided further confidence that the cellular mutant libraries generated using the protocol retained their gRNA complexity and could be used for genome-scale screening.

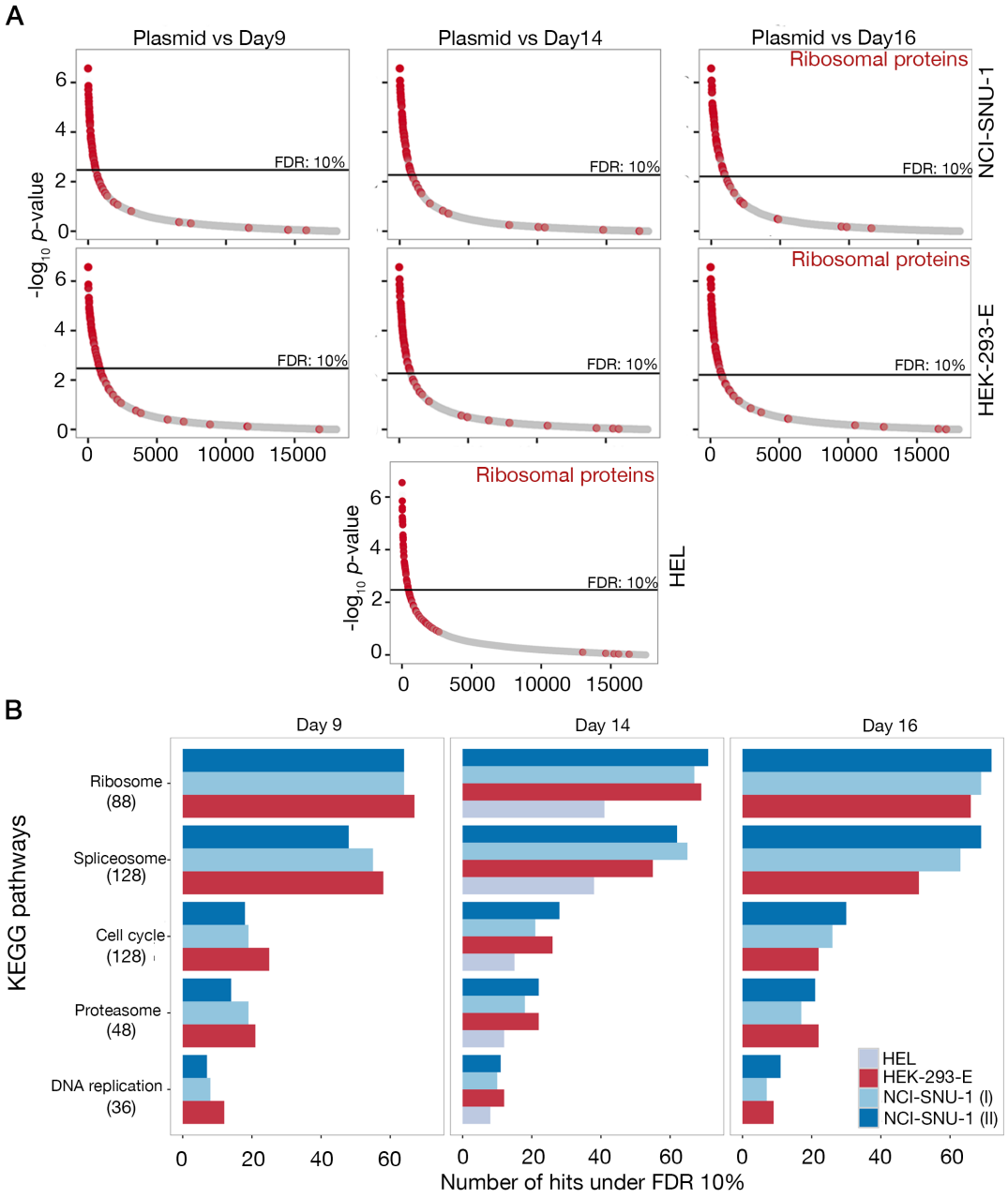


Fig. 3.5 gRNAs targeting genes involved in essential biological processes were depleted during culture of the mutant cell library. A. Gene-level enrichment analysis revealed genes involved in ribosomal biosynthesis to be among the most significantly depleted genes in NCI-SNU-1, HEL and HEK-293-E cell lines on all three time-points, days 9,14 and 16. For NCI-SNU-1, representative analysis from one of the two experiments is shown. **B.** Five KEGG-annotated pathways known to be involved in essential cell processes that were significantly depleted in all the samples are shown. The numbers under each pathway represents the total number of genes in the group. The number of genes involved in each pathway identified as being significant was similar for the replicates of NCI-SNU-1 and between the HEK-293-E, HEL and NCI-SNU-1 cell lines.

3.2.3 Genome-scale screens using monoclonal antibodies

Design of a genome-scale KO screening platform to investigate cellular recognition events at the cell surface

To identify factors required for molecular recognition events at the cell surface in the context of a plasma membrane, I initially designed a genome-scale KO screening system in human cell lines. A flow-cytometry based binding assay was initially used to identify a high-efficiency Cas9 expressing cell line that stained brightly with a mAb. I then created a genome-scale mutant cell library for the chosen cell line using a library of lentiviruses, each encoding a single gRNA from a pool of 90,709 individual gRNAs (gRNA expression vector depicted in figure 3.6A). Transduced cells that had lost the antibody epitope at the cell surface were isolated through fluorescence-activated cell sorting (FACS) and the genes responsible for this loss of binding were identified by comparing the relative abundance of the different gene-specific gRNAs present in the sorted cells compared to the total unsorted population using deep sequencing of PCR products and enrichment analysis (schematics depicted in figure 3.6).

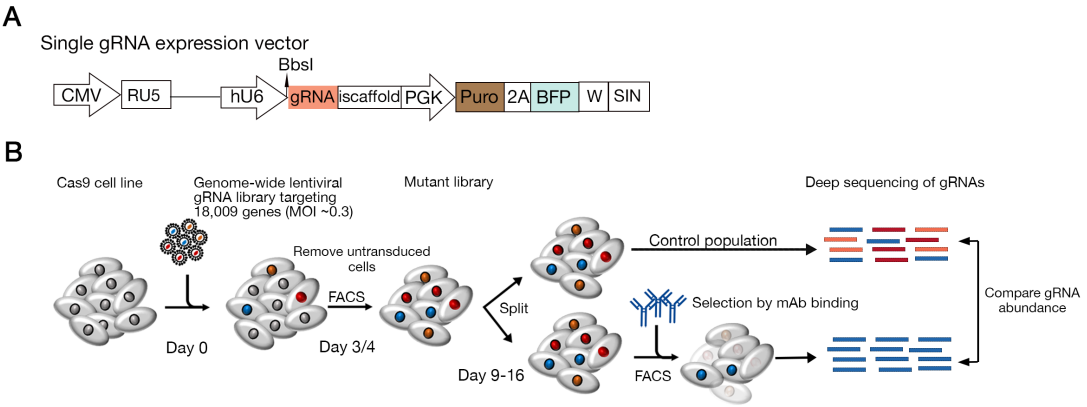


Fig. 3.6 Schematic of the genetic screening approach. A. Schematic of lentiviral single gRNA expression vector with BFP and puromycin selection cassettes with gRNAs cloned into the BbsI site. The backbone is the same as the reporter vector with gRNA targeting *BSG* described earlier (figure 3.1A). **B.** Strategy used to perform genome-scale KO screens using mAbs as selection reagents. The initial mutant library is generated by infecting cells at a low MOI of 0.3 and the infected cells expressing BFP are selected by FACS. The sorted library is kept in culture between 9 and 16 days. The total cell population on the screen day is divided into two; one half is kept as control population and the other half is selected for mAb binding. The cells from the mutant library that are refractory to binding are sorted using FACS and the abundance of gRNAs in the sorted population is compared to the control population to identify genes contributing to binding of the mAb to the cells.

A proof-of-principle screen using anti-BSG antibody identifies BSG and a chaperone required for BSG trafficking to the plasma membrane.

To test the use of the genome-scale screening approach, I carried out a proof-of-principle screen on HEK-293-E cells with an anti-BSG mAb as the selection reagent. BSG was a good model to test this system as the antibody provided a clear bright signal on the Cas9-expressing HEK-293-E cell line with a very high signal to noise ratio (figure 3.7A). The mutant library was screened 16 days post transduction and 0.22% of the population that expressed gRNA (BFP+) and lacked BSG on the surface (PE-) was sorted (figure 3.7B). Unsorted cells also from day 16 post-infection were used as a controls. gRNAs from both sorted and control populations were isolated and deep sequenced to quantify their abundance.

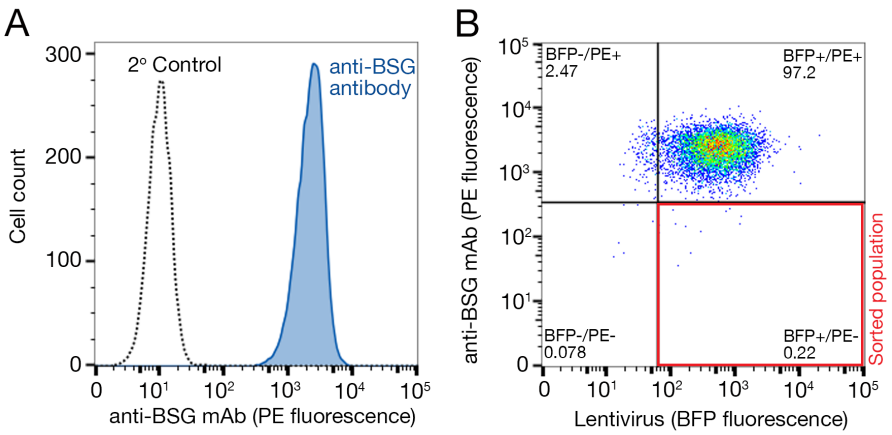


Fig. 3.7 Cell sorting strategy for a proof-of-principle genome-scale screening for recognition of BSG in HEK-293-E cells by an anti-BSG mAb A. BSG was highly expressed in HEK-293-E cells as indicated by the clear separation of cells stained with anti-BSG antibody versus secondary-only stained samples. **B.** Gating strategy for screen carried out with anti-BSG mAb in HEK-293-E cells. A very small fraction of PE-negative and BFP-positive cells were sorted. Approximately 60 million cells were sampled and 40,000 cells were collected during this screen.

A preliminary analysis using the raw abundance of gRNAs in the sorted population revealed a high enrichment of gRNAs targeting *BSG*, as expected, but also gRNAs targeting a gene encoding a monocarboxylate transporter, *SLC16A1* (or *MCT1*). Within the top ten most abundant gRNAs, four targeted *BSG* and four targeted *SLC16A1*—which is a known chaperone required for maturation and surface expression of BSG (3.8A). BSG is an unusual protein for a single transmembrane protein as it contains a charged glutamic acid residue within the transmembrane region. It has been suggested that this charged residue is important for the lateral association of BSG with the multi-

pass protein MCT1 on the plasma membrane (Schematics in figure 3.8B) [208, 209].

I next carried out a enrichment analysis using the MAGeCK software and identified four genes (*BSG, SLC16A1* and two novel factors *SPPL3* and *WDR48*) that were significantly enriched (FDR <0.05) in the sorted population compared to the control population (figure 3.8C). SPPL3 is a Golgi-resident intramembrane-cleaving protease that has been shown to regulate the activity of N-glycosylation related glycosyltransferases such as MGAT5, B3GNT5 and B4GALT1 [210]. The three N-glycosylation sites (N44Q, N152Q, and N186Q) of BSG are known to contain complex-type carbohydrate groups generated in the Golgi by the action of multiple glycosyltransferases, including MGAT5, thus it is possible that the loss of SPPL3 causes the generation of abnormal carbohydrate chains in BSG, which could lead to disregulated surface expression [211]. WDR48 is a regulator of deubiquitylating complexes, and, in this case also, there has not been a reported direct association of this protein with BSG. However, it has been suggested that BSG recruitment to the surface of the cell membrane is induced upon K63-linked ubiquitylation [212]. Therefore, it is possible that WDR48 has an effect on the pathway leading to BSG ubiquitylation thereby affecting its surface localisation. Further validation experiments are required to investigate these hypothesised roles.

This proof-of-principle study demonstrated that genome-scale KO screening approach was a suitable method to identify not only the gene encoding the antibody epitope, but also reveal factors such as chaperones that are important in expression of the receptors. However, one limitation of the method was that very few genes were identified as being significantly enriched. This was surprising because I was expecting that genes encoding proteins within the secretory pathway would be enriched in the sorted non-binding population as they would affect the surface expression of BSG. While a few members of the general secretory pathway are known to be essential, and by the day the screen was carried out, cells with mutations in those genes would have dropped-out of the mutant pool, it was surprising that none were identified. A likely explanation for the low number of significant genes could be the stringent gating threshold and the low number of cells ($\sim$40,000) that were collected, resulting in lower number of gRNA being represented in the sorted population. This could be observed by comparing the gRNA abundance in the control and sorted sample, which revealed a clear loss of many gRNAs in the sorted sample (figure 3.8D). Upon a closer look, only approximately 16,300 gRNAs

had representation (count above 10) in the sorted population compared to the 86,000 in the control population. Within the represented gRNAs, the abundance of gRNAs targeting the four significant hits were clearly higher in the sorted population compared to the control population. This suggested that using a highly stringent sorting threshold would result in the identification of few genes that have strong effects on binding loss. However, such a strategy might not be ideal for the identification of genes that have comparatively weaker size effects.

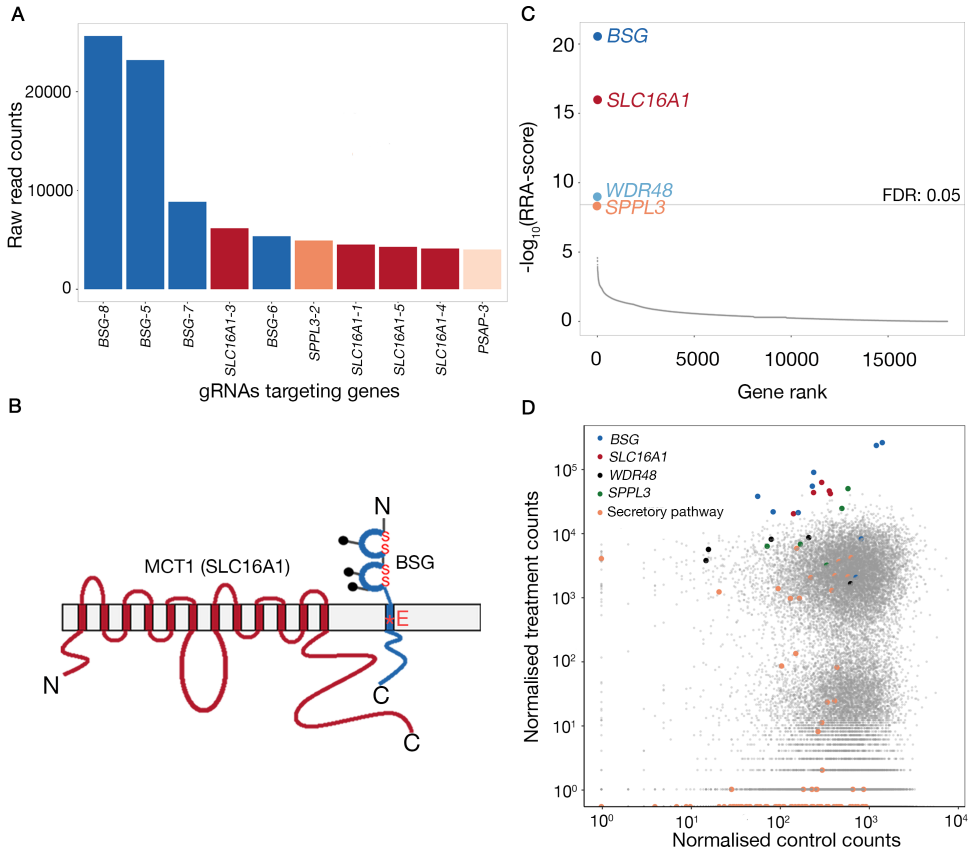


Fig. 3.8 A positive selection screen using anti-BSG mAb demonstrates the successful application of a CRISPR-based loss-of-function screen to identify factors required for epitope recognition by a mAb A. Top ten gRNAs with the highest raw read counts in the sorted non-mAb binding population. Four out of nine gRNAs targeting *BSG* and four out of five gRNAs targeting *SLC16A1* were represented in the top ten list. **B.** Schematic representation of cell surface interaction between 12-transmembrane transporter protein MCT1 and the single pass type I protein BSG. **C.** Enrichment analysis using MAGeCK. The y-axis represents the Robust Rank Aggregation (RRA) scores of genes calculated by comparing the gRNA abundance in sorted vs control cells. Genes are ordered by their RRA-score. **D.** Comparison between normalised read counts in control versus treatment (sorted) population. Individual gRNAs targeting the four significantly enriched genes but no gRNAs targeting the secretory pathway genes (KEGG-protein export) were enriched in the sorted population.

An improved screen identifies cellular pathways required for receptor expression in addition to genes directly encoding for the mAb epitope

I next explored ways to improve the screen to increase its sensitivity and identify not only the direct receptor, but also the pathways required for the expression of the receptor on the cell membrane. To test if the stringency of sorting was the reason for the low number of significant genes being identified in the screen, I aimed to decrease the stringency and to increase the mutant library size, in order to increase the number of sorted cells, thereby increasing the representation of the gRNAs. I carried out the screen using these parameters with an anti-CD59 mAb antibody. CD59 was chosen to test this system for various reasons. First, the anti-CD59 provided a bright stain on Cas9 expressing HEK-293-E cells (figure 3.9A); second, as CD59 is a GPI-anchored protein, its expression on the plasma membrane depends on the components of the cellular GPI anchor biosynthesis pathway [213]. The GPI anchor pathway is known to be non-essential for cell lines and has been robustly identified in genome-scale loss-of-function screens [173]. For this experiment, I screened 100 million cells from the mutant library and sorted 0.45% of the non-binding and BFP expressing cells (figure 3.9B). For increased representation of the library in the sorted population, I aimed to collect up to ten times more cells in the non-binding population compared to the screen with anti-BSG mAb.

When comparing the overall abundance of gRNAs in the sorted population to that from the control population, I immediately observed a strong enrichment for individual gRNAs targeting *CD59*, the GPI anchor biosynthesis pathway and the secretory pathway in the sorted population (figure 3.10A). Next, I applied gene-level enrichment analysis using the MAGeCK software and observed that *CD59* itself and 25 genes directly related or contributing to the GPI anchor biosynthesis pathway were significantly enriched (FDR <0.05) in the sorted population (figure 3.10B, GPI anchor biosynthesis schematic depicted in figure 3.10C).

Upon a closer look at the secretory pathway genes, only one member of the protein export pathway, *SLC61A1*, was identified as being significantly enriched in the sorted population. All other genes corresponded to the initial stages of the N-linked glycosylation pathway or the 'core glycosylation' pathway (identified genes in figure 3.11). The early steps in N-linked glycosylation involves the generation of a unique 14- monosaccaride ($Glc_3Man_9GlcNAc_2$) structure, which is a conserved feature of eukaryotic cells and is used as

a signalling molecule in the protein folding pathway for a wide range of N-glycosylated proteins such as CD59[1] [215]. Thus, the genes in this pathway are likely to have contributed to the proper folding of CD59 ensuring its transport to the surface of the cells.

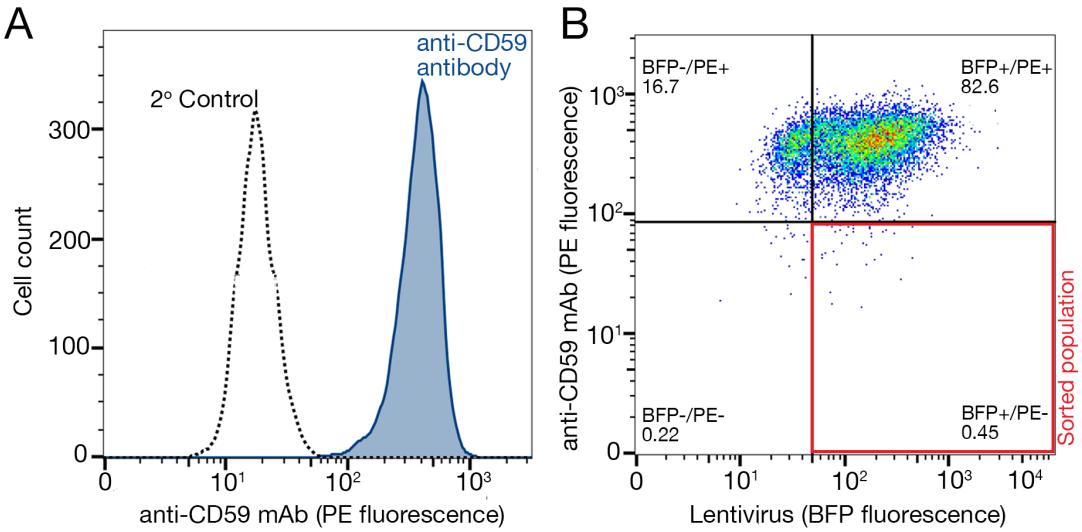


Fig. 3.9 Cell sorting profile for flow-cytometry based CRISPR-Cas9 screen in HEK-293-E cells with an anti-CD59 antibody. A. CD59 was highly expressed on the surface of HEK-293-E cells as indicated by the clear separation of cells stained with antibody compared to the secondary only stained sample. **B**. Gating strategy used in a genome-scale screen carried out using an anti-CD59 mAb. Approximately 100 million cells were sampled and 350,000 cells were collected during this screen. The mutant cell library has a clear BFP negative population as the library was made with only BFP sort and no puromycin was added to remove the contaminating non-transduced cells. The screen was carried out 15 days post infection with the gRNA lentiviral library.

The data here demonstrated that sampling enough cells from a high complexity library increases the sensitivity and allows for the identification of general and protein specific pathways required for transport of receptor to the surface in addition to the receptor itself. Increasing the sorting threshold did not influence the confidence with which the target receptor was identified as *CD59* was the still the most enriched gene in the sorted population.

[1] CD59 has two N-linked glycosylation sites and it mainly consists of a family of biantennary complex-type structures with and without lactosamine extensions and outer arm fucose residues. It has been shown that CD59 transport to the surface of the cells does not rely on this glycan being present on the protein [214]. Consistent with this, no genes in the later processing steps, which make these complex type structures were identified in the screen.

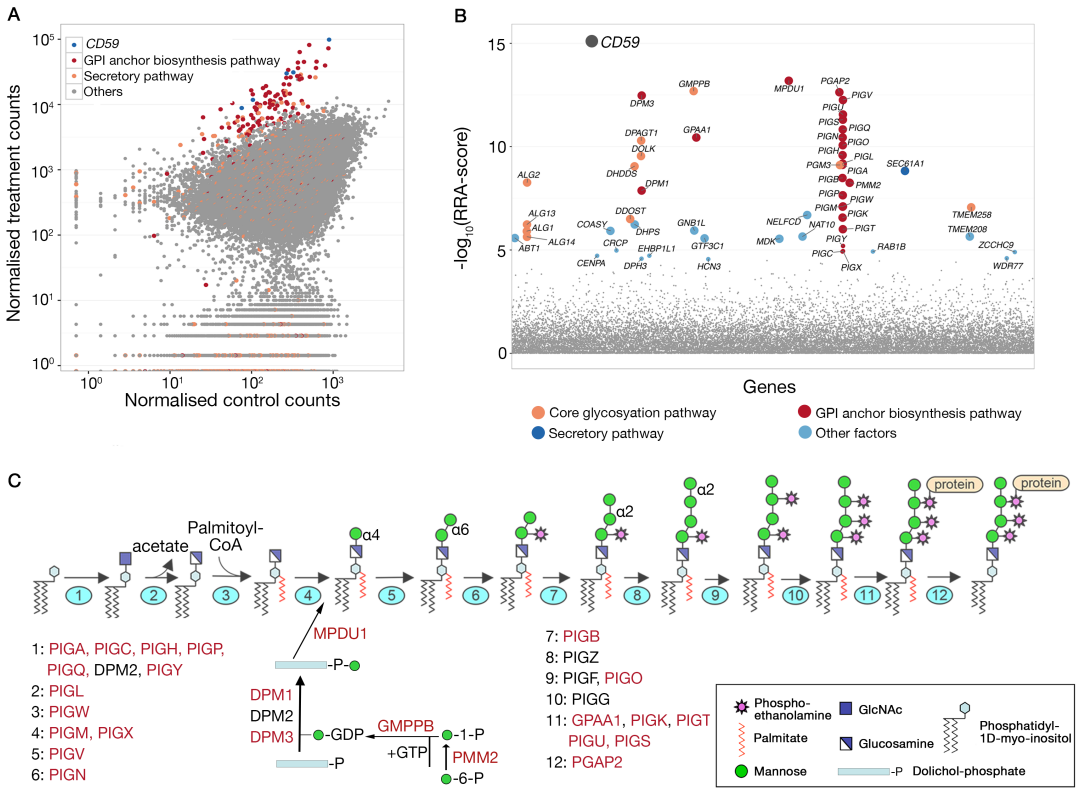


Fig. 3.10 An Improved CRISPR-mediated forward genetic screen identifies the genes required for the trafficking of the receptor in addition to the gene encoding the antibody epitope. A. Comparison of gRNAs abundance in sorted samples versus the control samples. Multiple gRNAs targeting the GPI-anchor pathway and *CD59* were clearly enriched in the sorted population. **B.** The enrichment of gRNAs targeting each gene (ordered alphabetically) is quantified as the RRA-score. Each circle represents a specific gene and the size corresponds to the FDR (large circle: FDR<1%, small circle: 1% < FDR < 5%). Only genes with FDR < 5% are labelled and the colors represent genes related by function. For better clarity, enlarged version of **B.** is also depicted in Appendix section figure A.1. **C.** Schematics of cellular GPI-anchor biosynthesis pathway (adapted from [216]). The genes that were significantly enriched are highlighted in red. Key enzymes along the pathway are labelled. PMM2 catalyses the isomerisation of mannose 6-phosphate to mannose 1-phosphate, which is a precursor to GDP-mannose necessary for the synthesis of dolichol-P-mannose by the members of dolichol-phosphate mannose (DPM) synthase complex (DPM1, DPM2 and DPM3). MPDU1 is required for the utilisation of the mannose donor dolichol-P-mannose by mannosyltransferase PIGM. Figure modified from [216].

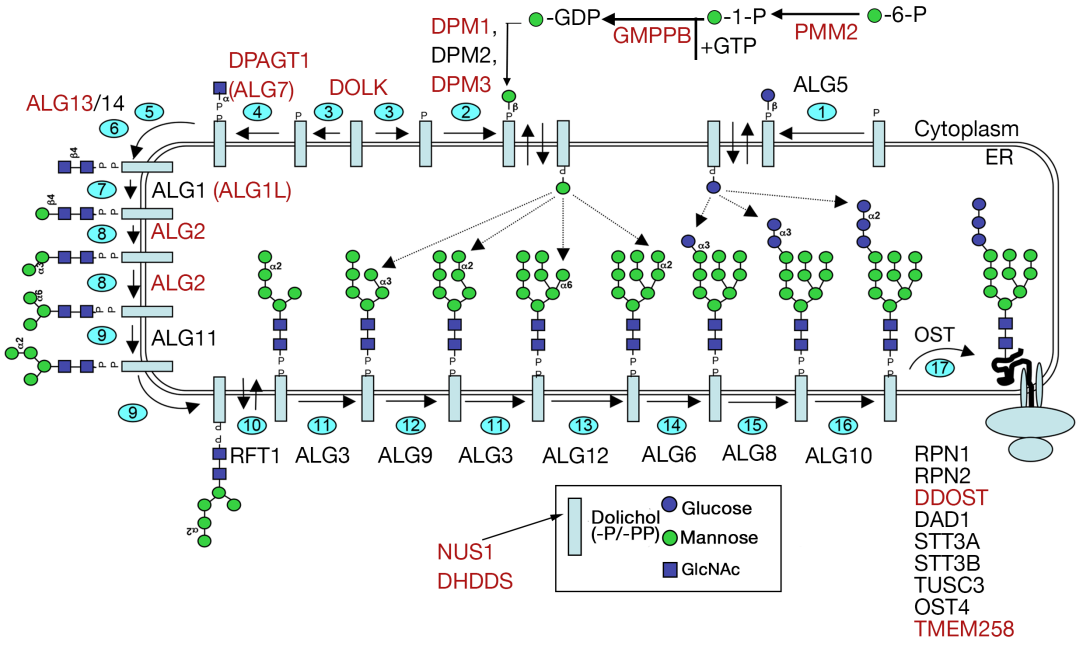


Fig. 3.11 Several genes involved in the early stages of N-glycan biosynthesis pathway were identified in the screen using an anti-CD59 mAb. Pathway depicting the core glycosylation that occurs in the ER to generate a 14-sugar-glycan precursor (dolichol-P-P-$GlcNAc_2Man_9Glc_3$), which is transferred *'en-bloc'* to the N-glycosylation consensus sequence (Asn-X-Ser/Thr) by the oligosaccharyltransferase (OST). The first committed step of the core glycan assembly is catalysed by a well-conserved essential gene DPAGT1 (or ALG7) [217]. The enzymes that catalyze each step in the biosynthesis have been identified mainly from studies carried out in mutants of the yeast *Saccharomyces cerevisiae*. The gene affected by each yeast mutation is known as an ALG gene (for altered in glycosylation). NUS1 together with DHDDS, forms the dehydrodolichyl diphosphate synthase (DDS) complex, an essential component of the dolichol monophosphate (Dol-P) biosynthetic machinery. Inhibition of protein N-glycosylation has been shown to cause cellular stresses that disrupt the ER leading to the induction of an unfolded protein response (UPR) and ER-assisted degradation (ERAD) [218]. Figure modified from [216]

Application of the screening method on a panel of antibodies reveals cellular factors required for receptor expression.

To determine whether the high sensitivity of the genome-scale method could be exploited to understand the other components of receptor biology, I next selected six mAbs targeting structurally diverse cell surface proteins expressed on either HEK-293-E or HEL cells and applied the method (summarised in table 3.1). All screens were carried out with libraries created from independent lentiviral transductions and for all screens, approximately 0.8-1% cells (250,000-400,000 cells) that had lost the ability to bind to the mAbs were collected.

Enrichment analysis using MAGeCK was carried out by comparing the gRNAs in the sorted population and the control population.

Table 3.1 Summary of mAbs used for genome-scale loss-of-function screening.

Cell line/Screen day	Antibody	Target receptor	Receptor feature
	BRIC125	CD47	
HEK-293-E cells (9 days p.i)	BRIC126	CD47	5-TM membrane protein
	B6H12	CD47	
	BRIC5	CD58	Single-Pass type I
HEK-293-E cells (10 days p.i)	P16	GP1a/IIa	(αIIβ1) Integrin heterodimer
HEL- (14 days p.i)	BRIC216	GYPA	O-glycosylated protein

Genome-scale screens carried out using mAbs identify SRP dependent ER translocation pathway. Three screens, using independently created knockout libraries, were performed with three different monoclonal antibodies targeting CD47 (BRIC125, BRIC126 and B6H12) on day nine post mutant library generation using lentiviral transduction. These screens were designed to (i) gain insights into the repeatability and robustness to biological variation in using the KO screening approach, and (ii) to assess the effect of screen day on genes that are identified in a genome-scale screen.

Enrichment analysis of the sorted population in all three screens revealed *CD47* among the most significantly enriched genes (FDR <0.01 in all cases), (figure 3.12A, B and C), demonstrating the robustness of the genetic screening method using the CRISPR-Cas9 system in identifying the gene encoding the direct receptor. In all three cases, multiple members of the general protein export pathway were also identified within FDR <0.05 (summarised in table 3.2). Excluding *CD47*, 16 genes were identified in at least two of three screens and nine of these common genes encoded the proteins involved in the general protein export pathway, specifically the components of the ER translocon protein complex (*SEC61A1, SEC61G*), the signal recognition particle (SRP) proteins (*SRP19, SRP14, SRP54, SRP68, SRPR*), and the members of the signal peptidase complex (*SPCS2* and *SPCS3*) (figure 3.12 D).

The general protein export pathway is the cellular machinery by which the majority of newly synthesised membrane and secreted proteins are folded and transported, thus in this context, their identification is to be expected. That said, it was interesting that these genes were mainly identified repeatedly in this set of screens, and not in the one with anti-CD59 antibody (where only *SLC61A1* was identified). A likely explanation for this is the day the screens

were carried out— day 9 here, versus day 15 post transduction in case of anti-CD59 mAb screen. Many identified genes along the general protein export pathway, including the ones identified here, such as *SLC61A1, SRPR, SPCS3, SRP54*, have been described as 'core fitness genes' that are essential for cellular proliferation in every cell [175]. Thus, the gRNAs targeting these genes are very likely to 'drop-out' as the cells are continually grown. Hence, the chance of identifying them could be time-dependent, as the longer the mutant library is kept, the higher the chance that the cells with mutations in essential genes will be non-viable and drop-out of the mutant pool, causing the signal to dilute.

Table 3.2 Genes involved in protein export (KEGG annotation) identified in screens carried out using anti-CD47 mAb on day nine post mutant library generation

Description	Protein	Number of genes(s) identified using mAb
Translocation channel constituents	SEC61A1	BRIC125, BRIC126, B6H12
	SEC61G	BRIC125, BRIC126
Components of SRP	SRP14	BRIC125, BRIC126, B6H12
	SRP19	BRIC125, BRIC126, B6H12
	SRP54	BRIC125, BRIC126, B6H12
	SRP68	BRIC125, BRIC126
	SRP72	B6H12
SRP-ribosome complex receptor	SRPR	BRIC126, B6H12
	SRPRB	BRIC126
Signal peptidase complex	SPCS2	BRIC125, BRIC126
	SPCS3	BRIC125, BRIC126

The other gene that was identified in all three screens was *ASCC3*, which has been recently identified to regulate gene expression after UV-induced DNA damage [219]. The role of ASCC3 in relation to CD47 expression on the cell surface is not known, but this gene has also been identified in CRISPR mediated KO screen carried out for host factors required by DENV virus [220], so it is possible that it plays a regulatory role for general protein expression rather than a CD47-specific role. Further studies would have to be done to confirm this.

The data presented here demonstrate that this method is reproducible and can tolerate biological variation especially for the identification of the direct target of the mAb. In addition, it can also repeatedly reveal both essential and non-essential cellular factors that contribute to the expression of the target receptor on the surface of the cell.

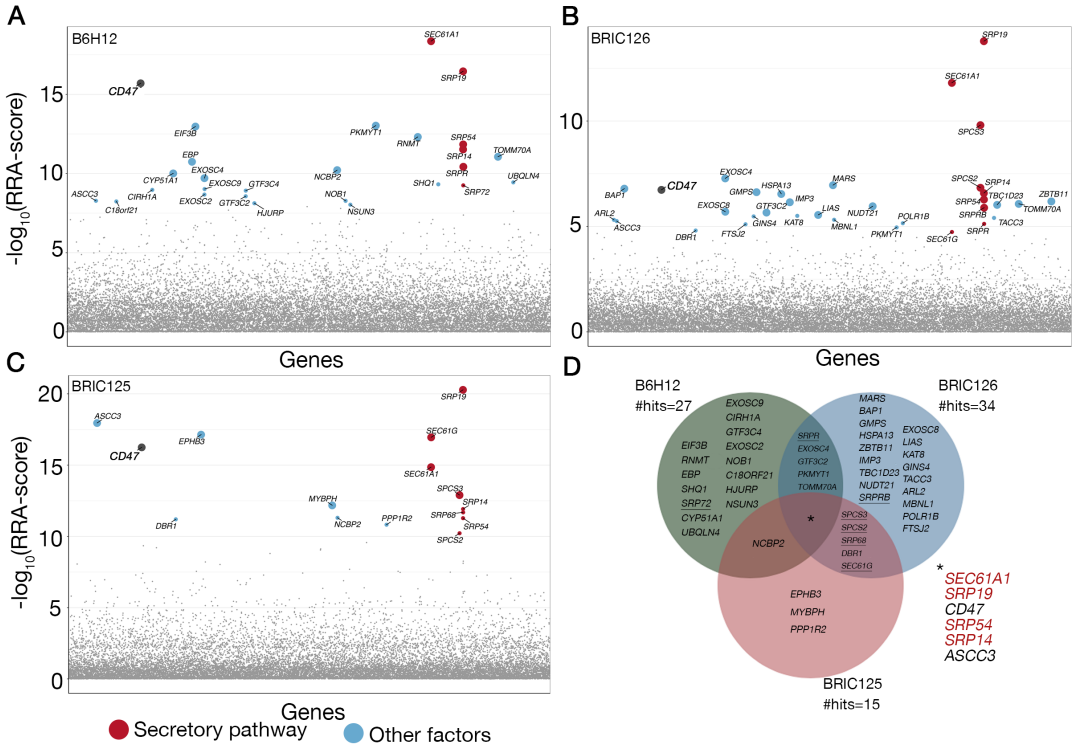


Fig. 3.12 Genome-scale CRISPR-Cas9 screen using anti-CD47 mAbs targeting cell surface protein reveal members of SRP dependent ER protein translocation pathway. Gene-level RRA-score calculated using MAGeCK in a genome-scale KO screens using anti-CD47 mAbs: B6H12 mAb (**A**), BRIC 125 (**B**) and BRIC126 (**C**). Each circle represents a specific gene and the size corresponds to the FDR (large circle: FDR<1%, small circle: 1% < FDR < 5%). Only genes with FDR < 5% are labeled and the colors represent genes related by function. In all cases, CD47 was identified amongst the most enriched genes, as expected; however, in the case of mAb clone BRIC125, an additional gene encoding for a different cell surface protein, EPHB3, was also identified with same the FDR as CD47. EPHB3 was not enriched for two other anti-CD47 mAbs, suggesting the BRIC125 epitope was present on both CD47 and EPHB3. **D.** Venn-diagram representing the genes identified with FDR < 5 % in the three screens. Six genes were common to all screens. These included genes of the SRP-dependent protein export pathway, *CD47* and *ASCC3*. All three screens were performed nine days post infection with the gRNA lentiviral library. For better clarity, an enlarged version of this figure is also depicted in Appendix section figure A.2.

Genome-scale screen using an anti-GYPA mAb reveals post-translational modifications present on GYPA required for the mAb epitope. Genome-scale screening methods provide a unique potential to study the enzymes required for cell surface post-translational modifications such as glycans. With the method described here, it should, in principle, be possible to identify genetic pathways leading to the biosynthesis and the subsequent positioning of the receptor glycans on proteins or lipids, thereby facilitating their identification.

The human erythrocyte protein, GYPA, serves as a good model to study post translational glycan modifications as it is one of the most heavily glycosylated proteins with 15 O-linked glycans, which consists predominantly of disiaylotetrasaccharide linked to a serine or threonine residue and one N-linked glycan [221]. To investigate this further, an anti-GYPA antibody (BRIC256), which recognises the extracellular epitope amino acids 41-58 on GYPA where three O-linked glycans are present (position 44, 47 and 50) [222] was used as a probe to carry out a genetic screen. The expression of GYPA is restricted to the cells of erythroid lineage, thus in this case, the screen could not be carried out in HEK-293-E cells. For this purpose, I selected the HEL cell line as these cells resemble erythroblasts (nucleated immature erythrocyte) and are known to express GYPA [223].

The gene-level enrichment analysis of the gRNAs in the cells that had lost binding the anti-GYPA antibody revealed *GYPA* as one of the most-enriched genes, as expected (figure 3.13A). In addition, genes required for the generation of the core-O-glycan structure (*C1GALT1, C1GALT1C1*) and genes involved in the addition of terminal N-acetyl neuraminic acid (Neu5Ac) modifications (*CMAS, SLC35C1*) were also identified. Schematics of the predominant O-glycan found on GYPA and the roles of the identified enzymes in generating this structure are depicted in Figure 3.13B. This suggests that the loss of GYPA itself or the loss of enzymes required for the generation of the disialotetrasaccaride present on GYPA leads to the loss of the mAb epitope. This could either happen if GYPA expression on the surface of the cells depended on the presence of O-linked glycans on the protein or if the O-linked glycans together with the protein backbone form the antibody binding epitope. As BRIC256 recognises the region of the protein where O-linked glycosylation sites are present, the latter seems more likely in this case. These results demonstrate that in situations where the interaction is mediated by glycans, the screen is able to identify the cellular pathways that are required for the biosynthesis of the glycan, thereby providing clues to the identify of the glycan.

The screen additionally identified an erythroid-specific transcription factor GATA1, which is presumably required for GYPA transcription in HEL cells. In addition, two genes, *CDIPT* and *SACM1L* were also identified. Both of these genes are required for metabolism of phosphatidylinositol (PtdIns). CDIPT protein is required for biosynthesis of phosphatidylinositol whereas SACM1 is a phosphoinositide lipid phosphatase. GYPA has been shown to be associated with PtdIns(4,5)P_2 in the cytosol [224], which suggests that the loss of phos-

phoinositides would adversely affect its membrane positioning. Additionally, the identification of *TNNT3*, a gene that encodes for the tropomyosin-binding subunit of troponin, was intriguing as tropomyosin has been previously shown to associate with the major proteins of the erythrocyte membrane skeleton (spectrin, actin, and 4.1R (human erythrocyte protein 4.1)) to form a membrane complex that includes glycophorins [225]. Further studies will have to be carried out to verify the role of TNNT3 in regulating GYPA expression.

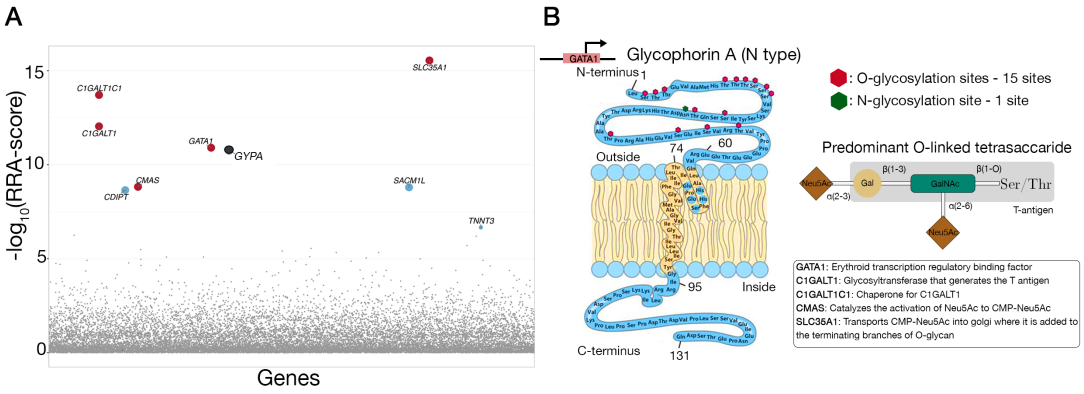


Fig. 3.13 Genome-scale KO screen using an anti-GYPA antibody reveals mAb epitope and factors required for the cell surface GYPA display. **A.** Enrichment analysis using MAGeCK revealed nine genes that were enriched in the sorted population compared to the control population (FDR< 5%). The number of cells sampled in this screen was lower compared to the previous screens carried out with HEK-293-E cells (50 million vs 100 million). The screen was performed on day 14 post infection with the gRNA lentiviral library. Genes relating to the post-translational modification present on GYPA and the gene encoding for erythroid specific transcription GATA1 are labelled in red. **B.** GYPA on the cell surface exits in either the M or N form with the M phenotype characterised by Ser-1, Gly-5 and the N-phenotype by Leu-1 and Gly-5. Schematics of the N-type GYPA is depicted in the left panel and the detailed structure of the predominant O-linked tetrasaccaride present in the protein is depicted in the upper right panel. The function of the GYPA specific genes identified in (A) are detailed in the lower right panel.

A Genome-scale screen using an anti $\alpha 2\beta 1$ mAb identifies the subunit encoding the antibody epitope and the critical requirement of actin regulation. A genome-scale screen was carried out in HEK-293-E cells using a monoclonal antibody (P16) targeting integrin $\alpha 2\beta 1$ (ITGA2/ITGB1) ten days post mutant library generation. Enrichment analysis revealed the gene encoding only the $\beta 1$ chain (*ITGB1*) of the integrin heterodimer in the sorted population demonstrating the epitope of this mAb is located within the $\beta 1$ and not $\alpha 2$ chain. In addition, genes of the general secretory pathway (SEC translocon complex and SRP components) were also significantly enriched in the sorted population and similar to the screen with anti-CD59 mAb, genes

required for early steps of N-linked glycosylation in the ER were also identified (figure 3.14A).

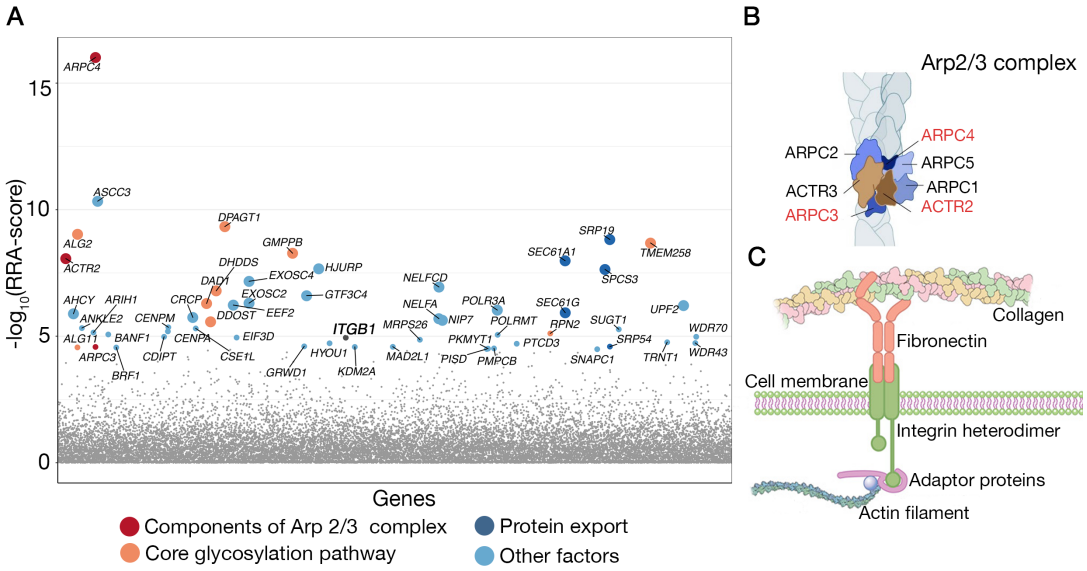


Fig. 3.14 Genome-scale loss-of-function screen using an anti-integrin αIIβ1 mAb identifies the subunit encoding the antibody epitope and components of the cytoplasmic Arp2/3 complex. **A.** Gene-level enrichment analysis on a screen carried out using an anti-αIIβ1 antibody. Only the β1-subunit encoding gene, *ITGB1*, was identified to be significantly enriched in the sorted population (FDR= 0.024). **B.** Schematics depicting the ARP2/3 complex, which consists of seven members, three of which were identified in the screen to be statistically significant (FDR $<$0.05) (highlighted in red, figure adapted from [226]). Depending on the combinations of α and β subunits, the extracellular domains of integrin heterodimer interact with the components of the extracellular matrix (ECM). This link between the intracellular cytoskeleton to the ECM via integrins has been shown to be important in cell motility, cell polarity, cell growth and survival (figure adapted from [227]). The screen was performed ten days post infection with the gRNA lentiviral library. For better clarity, an enlarged version of **A.** is also depicted in Appendix section figure A.3.

An interesting observation in this screen was also the identification of three genes relating to the members of Arp2/3 complex (*ARPC4, ACTR2, APRC3*). Arp2/3 protein complex consists of seven members and it is essential in the nucleation and assembly of branched actin filaments (figure3.14B) [226]. The cytoplasmic tails of β-subunits have been shown to be necessary and sufficient to link integrins to the actin cytoskeleton [228]. Actin binding proteins such as α-actinin, filamin and talin bind to actin filaments and mediate their link with cell surface integrin heterodimers (schematic depicted in figure 3.14C). Such interactions at the cytoplasmic domains of integrins have been shown to induce conformational changes in integrin extracellular domains (from a

'bent' (inactive) to 'extended' (active) conformation) that result in increased affinity for ligand in a process described as 'inside-out signalling' [229]. The screen here demonstrated that the components of the Arp2/3 complex were important for the recognition of ITGB1 at the cell surface by the mAb. This finding can be interpreted in two ways: (i) the expression of ITGB1 on the surface of cells depends on its interaction with the members of the Arp2/3 complex or (ii) the interaction with Arp2/3 complex with the cytoplasmic domain of ITGB1 changes the integrin subunit conformation that is required for the mAb epitope binding. The latter scenario is consistent with the presence of other mAbs for integrins that have been previously described to specifically recognise epitopes that are only present in the active conformation [230]. It is therefore likely that P16 falls into this class of integrin mAbs.

Summary of the genetic screens

All enrichment analysis here was done using MAGeCK software, which can use a single dataset to estimate a read count variance to determine the significantly enriched genes. I used a FDR of less than 0.05 as the cut-off point and estimated the biologically relevant genes that could be identified within that threshold. In all the screens that were carried out, the target receptor was identified with a very low FDR and well within the threshold (summarised in table 3.3).

Table 3.3 FDR of identification of the genes encoding direct receptor in a genome-scale screening approach using monoclonal antibodies targeting cell surface receptors.

Antibody	Day of screen	Target receptor	Target FDR	Genes (FDR<5%)
MEM6/6	Day 16	BSG	0.0012	4
BRIC222	Day 15	CD59	0.0002	58
BRIC125	Day 9	CD47	0.001	15
B6H12	Day 9	CD47	0.001	27
BRIC126	Day 9	CD47	0.004	35
P16	Day 10	ITGB1	0.024	52
BRIC256	Day 14	GYPA	0.001	9
BRIC5	Day 9	CD58*	0.001	14

*Full screen result for this screen is available in appendix section figure A.4

The number of significantly enriched genes identified in each screen depended on the quality of the screen performed in terms of the library size and the representation of the gRNA in the sorted cells. In the example of an

anti-BSG antibody where a stringent sorting threshold was used, very few genes were identified. This was improved when the number of sorted cells was increased in the subsequent screens.

Apart from the identification of the direct target receptor, specific cellular pathways required for the ligand recognition such as the GPI-anchor pathway (in anti-CD59 screen) and the O-glycosylation pathway (in anti-GYPA screen) were also identified. In addition, a pathway analysis of all enriched genes that were shared between antibody selections encoded proteins required for protein secretion and glycosylation, as expected, but also identified housekeeping pathways such as ribosome biosynthesis and RNA metabolism; genes identified in these pathways were grouped and labelled as 'other factors' (figure 3.15, also refer to table A.5 in the appendix section for specific 'other factors' that were enriched in at least two out the seven screens with FDR<0.05). I observed that the representation of these general pathways was often reduced when selections were performed several days later (day 15-16 rather than day 9), suggesting these genes are required for long-term cell viability in culture, and that antibody staining was reduced on moribund cells.

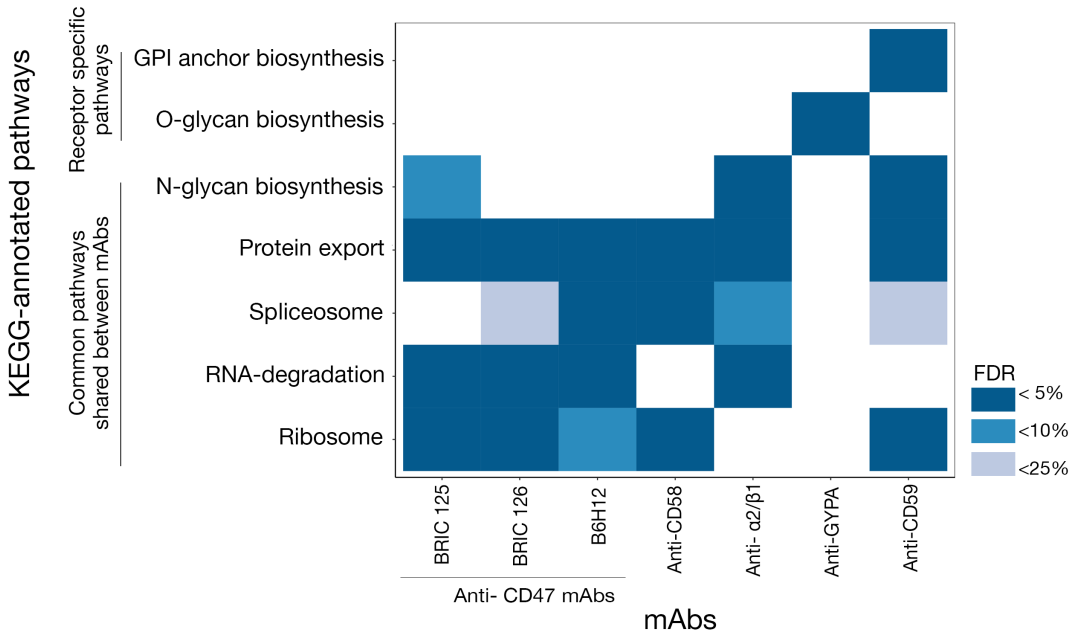


Fig. 3.15 Genes required for protein export, N-glycan biosynthesis and general housekeeping functions were enriched in, and shared between, cells selected for the loss of cell surface mAb staining. Pathway analysis on genes enriched in cells selected with the indicated mAbs identified some pathways shared between more than one antibody including protein export, N-glycan biosynthesis as well as more general housekeeping functions. No significantly enriched pathway was identified in the anti-BSG screen.

3.3 Discussion

In this chapter, I have demonstrated the use of a cell-based genome-scale CRISPR screening approach to identify genes encoding proteins required for extracellular recognition. I first generated high efficiency Cas9 cell lines by isolating stable clones grown from single cells within a polyclonal Cas9-expressing population. This approach allowed for the removal of cells which contained Cas9 with decreased efficiency resulting presumably from the acquisition of mutations that led to the inactivation of Cas9. I utilised two different methods to measure the Cas9 efficiency: (i) an endogenous gene KO method in which endogenous *BSG* was targeted, and (ii) a rapid exogenous method in which *GFP*, together with gRNA targeting *GFP*, was introduced to the cells and the efficiency of *GFP* targeting was measured. The exogenous approach provided a rapid means of determining the presence of functional Cas9 in the cells and was particularly useful during the single cell cloning steps, during which many clones had to be tested. However, targeting the single integrated GFP can be considerably more efficient than targeting both alleles of an endogenous gene (as exemplified by the lower efficiency observed for the same clonal line for *BSG* KO compared to *GFP* KO), so this approach usually represents the best-case-scenario for targeted gene KO. Additionally, endogenous targeting also depends on other factors, such as the protein-turnover time, the copy number of the target gene in a cell, and the targeting efficiency of the gRNA. To account for these factors, I opted to estimate the 'realistic' Cas9 efficiency based on the decrease in cell surface expression of BSG upon transduction with gRNA targeting *BSG*, rather than the *GFP* targeting approach.

In this work, I utilised the Yusa library to conduct genetic screens. There are several genome-wide libraries that have been described to date, which can be used to perform genome-wide KO screens [175, 176, 177, 205, 231, 232, 233]. Of these, four libraries are considered to be second generation libraries as they feature improved gRNA efficacy [234]. These are the 'Human V1 library' or the 'Yusa library' [176], the 'Whitehead library' [177], the 'Brunello library' [205] and the 'Toronto knockout libray version 3.0 (TKO3.0) library' [233]. The Whitehead library is the largest consisting of 10 gRNA/gene (182,134 gRNAs) compared to the 5 gRNAs/gene of the Yusa library (90,709 gRNAs) and 4 gRNAs/gene of the Brunello library (77,441 gRNAs) and the TKO3.0 library (71,090 gRNAs). The Yusa library differs from the other libraries as it uses the improved gRNA scaffold to avoid the T stretches as discussed in section 3.1.4. The design

of the gRNAs in the Yusa library, unlike those in the other three libraries, is not based on an on-target prediction for gRNA selection. However, this does not seem to affect the screening results as a recent comparison between the performance of these libraries in negative selection screens has shown that that these libraries identify genes with similar false discovery rates (between 14 % and 23 %) [234]. The same study has also suggested that while the higher number of gRNAs per gene, for example that of the Whitehead library, allows for better statistical confidence, this can also cause over-sensitivity of the analysis programs such as MAGeCK to call genes that have lower fold change values as statistical hits. In addition, as the number of gRNAs increases, the complexity of the library also increases, maintainance of which over an extended period of time can be practically challenging. This could instead lead to poor screen outcome. For second generation libraries it has been suggested that libraries with 6 gRNAs per gene are likely to be optimal for genome-wide CRISPR-KO screens [234]. While I only used the Yusa library in the study here, given the similarity of performance with the other libraries, it is unlikely that the other libraries would yield vastly differing results.

The genetic approach described in this chapter provides a valuable alternative to existing biochemical methods which must account for the largely insoluble nature of membrane-embedded receptors. An advantage of this method over existing methods is its ability to reveal the receptor protein at the cell surface which directly interacts with a presented ligand, but also identify other gene products that are required for the cell biology of the receptor. The screens carried out with the panel of mAbs demonstrated the ways in which the method can be used to identify the protein/glycan receptors and cellular factors such as chaperones, transcription factors, and cytoskeletal elements that are involved in the expression or correct positioning of the receptors on the surface of the cells.

The application of a genome-scale screening approach in the identification of monoclonal antibody targets can be a valuable tool for monoclonal antibody characterisation. A similar approach for this purpose has been described recently by others [235], where the cells refractory to antibody binding were enriched using two rounds of sorting and few gRNAs present in this enriched population were analysed using Sanger sequencing to identify the target receptor. Here I have extended the use of the genome-scale screening approach by using a single sorting approach that can not only capture gRNAs targeting genes encoding the antibody epitope, but also the gRNA targeting

genes contributing to the cell biology of the antibody epitope. In the example of an anti-integrin mAb, this included identifying the specific subunit of the heterodimer encoding the antibody epitope and components of the cytoplasmic Arp2/3 complex and for an anti-Glycophorin A mAb, a required role for enzymes involved in O-linked glycosylation in antibody binding.

In the genetic screens carried out here, I often identified genes that are required for the 'house-keeping' of receptors, including those involved in the secretory and glycosylation pathways. However, this did not greatly influence the confidence with which the receptor was identified. The target receptor was revealed in every case attempted within the FDR threshold of 0.05, with seven out of eight screens identifying the receptor with a very low FDR of under 0.005. The highest FDR of 0.024 was observed in the screen with the anti-ITGA2/B1 integrin antibody where admittedly, genes of the secretory and glycosylation pathway were enriched more than the target receptor. But even in such a scenario, assuming that the target of this antibody was unknown, the observation that (i) *ITGB1* was the only cell surface receptor encoding gene identified in the screen within the given threshold of 0.05; (ii) components of the Arp2/3 complex that fit with the integrin biology were identified among the most enriched genes and; (iii) the fact that genes that contribute to general receptor house-keeping can be ruled out as specific factors, suggest that it would have been possible to identity *ITGB1* as the target receptor.

One of the challenges of loss-of-function screen studies is the investigation of the effect of the genes, which are essential for cell growth and viability, on the phenotype being tested. Comprehensive studies carried out out using CRISPR–Cas screens on multiple cell lines have identified approximately 2000 'core-genes' that are designated to be 'essential' for optimal growth of the cells [175, 176, 177, 205]. Identifying the role of cell-essential genes on cellular recognition can be challenging as cells that contain mutations in essential genes become non-viable and are no longer represented in the mutant library. However, it has been suggested that the number of genes thought to be 'core-essential' could be overestimated as this includes genes that not only affect viability of the cells but also moderately affect cell growth [236, p. 357]. CRISPR-Cas9 KO screens carried out recently in the context of virus-host interactions have shown that it is possible to identify the effect of genes that have been categorised as core-essential genes in genome-scale KO screens (for example, the role of oligosaccharyl transferase complex (OST) in dengue virus infection [220]). The data here demonstrated that such screens can

indeed identify the core-essential genes involved in general protein export and glycosylation pathways, but unlike the direct receptors, genes in these categories were not identified in every case (e.g., not identified in anti-BSG and anti-GYPA screens). The timing (early time points rather than later) and the quality of the screen parameters in terms of the day post transduction the screening was carried out, the number of cells in the sorting population and the sorting stringency seemed to influence the successful identification of essential genes in such screens.

In summary, I have developed and applied a cell-based genetic method based on CRISPR-KO technology using mAbs to identify genes that mediate high affinity interactions at the cell surface. The method is able to identify the direct receptors at the cell surface robustly with high confidence and often also identifies cellular components that are related to the biology of the receptor.

CHAPTER 4

APPLICATION OF THE GENETIC SCREENING APPROACH TO IDENTIFY INTERACTIONS MEDIATED BY RECOMBINANT PROTEINS

4.1 Introduction

This chapter describes the application of the genome-scale KO screening approach that was developed in Chapter 3 to identify interactions between soluble recombinant proteins and the receptors on the surface of human cell lines.

While high affinity monoclonal antibodies are useful research tools and important therapeutic reagents, I sought next to determine whether this approach could be used to identify receptors for recombinant proteins. The use of recombinant proteins in the genetic screening approach designed here would open up the possibility to identify directly interacting receptors in wide range of biological contexts. In addition, a genome-scale approach could also reveal cellular factors required for the receptor to be correctly presented on the cell surface allowing novel insights into the biology of the receptor.

The interactions mediated by recombinant proteins, unlike mAbs, are usually of low-affinity. Therefore, to detect the interactions mediated by recombinant proteins with the cell surfaces, it is important to use oligomeric proteins that have increased avidity compared to their monomeric forms. In this work, I utilised two approaches for oligomerisation of the recombinant probes: (i) proteins were produced in monomeric biotinylated forms and conjugated to streptavidin-PE to generate fluorescent tetrameric avid probes; (ii) proteins were produced as pentamers using the COMP tag. The pentameric proteins also carried a FLAG-tag, which could be detected using an anti-FLAG-PE antibody.

To adapt the screening system to use recombinant proteins as screening probes, I utilised the protein resources that were already available in our laboratory. Below I will highlight the two main classes of protein libraries from which I selected primary candidates for the screening approach.

4.1.1 *P. falciparum* 'merozoite' protein library

P. falciparum is an obligate intracellular parasite that causes malaria in humans. The parasite has a complex life cycle that involves two organisms (humans as 'hosts' and mosquitoes as 'vectors') and three life-cycle stages ('sporozoites', 'merozoites' and 'gametocytes'). The blood stage of the infection, during which the merozoite invades erythrocytes, accounts for the majority of the symptoms and pathology of malaria. As the merozoites are briefly exposed to the host immune system between invasion cycles, merozoite ligands responsible for invasion are considered an attractive target for a vaccine intervention. The merozoite contains more than 100 proteins on its surface and intracellular vesicular organelles, most of which have been suggested to be involved in binding to the erythrocyte surface for mediating invasion [237]. Despite significant research efforts, less than ten of these interactions between merozoite protein ligands and human erythrocyte receptors have been described to date (figure 4.1).

Our laboratory has compiled a list of more than 60 proteins representing the ectodomains of abundant cell surface and secreted merozoite proteins of the 3D7 strain of *P. falciparum* [238, 239]. This protein library has already been used to identify some of the host-pathogen interactions such BSG-RH5, Semaphorin 7A- MTRAP, and P-selectin- MSP-7 [88, 128, 240]. Here, I mainly focused on 11 merozoite proteins from this library (highlighted in figure 4.1A). The proteins were selected on the basis of (a) known biology (i.e. previous reports indicating that they could be involved in host-parasite interactions; this includes proteins mainly located in the micronemes and rhoptries [241]); (b) proteins that interact with the invasion-related proteins (CyRPA, which is known to interact with RH5 to form an invasion complex [90, 242]); (c) proteins encoded by blood-stage essential genes (e.g. MSP1, which is also the most abundant protein on the surface of the parasite, [237]); and (d) parasite proteins against which in-vitro invasion blocking antibodies have been raised: MSRP5, SERA9 [243].

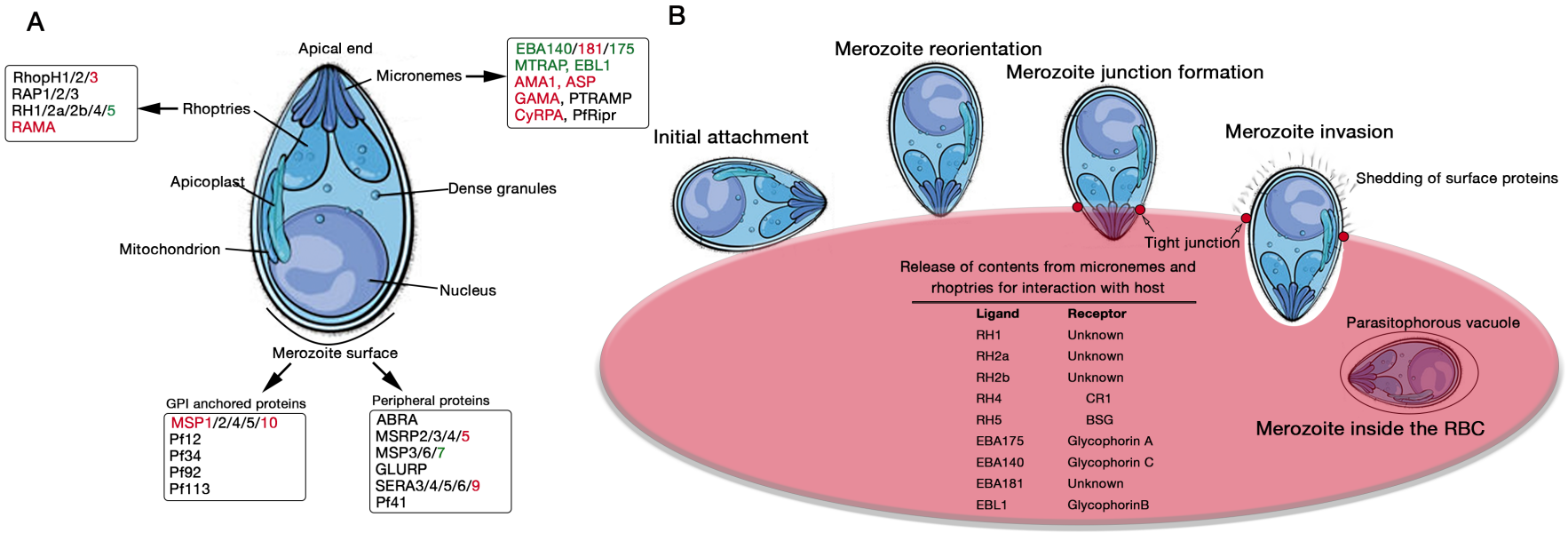


Fig. 4.1 Cellular organization and invasion process of *P. falciparum* merozoite. **A.** Schematics of cellular organisation of the merozoite. The merozoite is eukaryotic Apicomplexan parasite consisting of eukaryotic organelles such as a nucleus, a single mitochondrion, and an apicoplast together with specialised apicomplexan secretory organelles such as rhoptries, micronemes and dense granules. The secretory organelles contain distinct sets of proteins as labelled. The surface of the merozoite consists of a large number of invasion-related proteins making up the 'fuzzy surface coat'. The proteins labelled in red were investigated in this study, whereas the proteins labelled in green have been studied in the past and their host receptors already identified. **B.** Schematics of the process through which the merozoite enters a red blood cell (RBC). At the start of invasion, the merozoite interacts with the RBC in a reversible process called 'primary attachment', during which the surface proteins from any surface of the merozoite mediate low affinity interactions with the surface of RBC. The process is followed by 're-orientation', during which the parasite juxtaposes the apical surface towards the erythrocyte surface and releases the contents of its micronemes and rhoptries to mediate host cell surface interactions. Many proteins from these internal secretory organelles have been shown to be important for the host-pathogen interactions (some of the known interaction partners and proteins that are believed to have a host receptor are listed). This strong attachment leads to a non-reversible commitment towards invasion and formation of a 'tight junction'. As the parasite utilises the actin/myosin motor to force its way into the RBC, the tight junction moves along the length of the parasite. The surface proteins of the parasite are also cleaved by parasite proteases and shed off during the process. Eventually, the merozoite is engulfed by the erythrocyte and the parasite resides within the parasitophorous vacuoles [241]. Figure A is adapted from [244].

4.1.2 The platelet receptor library

Platelets are small anucleated cells in blood that play an important role in regulating hemostasis and thrombosis. Platelet-mediated thrombosis requires a large number of proteins present both on the platelet plasma membrane and within the secretory vesicles (also known as 'platelet granules'). Platelet functions are mediated by the important cell-cell and cell-subendothelial matrix interactions carried out by these receptors [245]. A protein library representing the cell surface receptor and secretome of the human platelet has also been compiled in our laboratory. This library includes roughly 200 proteins, which includes a diverse set of receptors such as integrins, leucine-rich repeats receptors, selectins, and immunoglobulin superfamily receptors. Many proteins in this library are not restricted only to platelets but are also found on many different cell types such as erythrocytes, leukocytes, and endothelial cells [96]. This library will be referred to as the 'human protein library'.

4.1.3 Scope of this chapter

In this chapter, I will first describe a successful proof-of-principle demonstration for the use of recombinant proteins as screening probes for the CRISPR-Cas9-based KO screening system. This demonstration uses the known interaction between RH5 and BSG. I will then describe the attempts made for the identification of host receptors for other merozoite proteins and optimisation steps made in the method to address some of the generic binding behaviour exhibited by recombinant proteins. Finally, I will use the human protein library to demonstrate how this method can be applied to robustly identify specific receptors and non-receptor cellular factors contributing to cellular recognition.

4.2 Results

4.2.1 Proof-of-concept study: BSG and RH5 interaction

Recombinant RH5 binds to BSG and an unknown factor on the cell surface of HEK-293-E cells

To assess whether recombinant proteins could be used as screening probes in this system, I initially selected the interaction between RH5 and its host receptor BSG. This served as a good model system because it is a low affinity interaction (K_D $\sim$1μM); is biochemically and structurally well characterised; and BSG is highly expressed on the Cas9-expressing HEK-293-E cell line [88, 246]. In a flow-cytometry based binding assay, I observed direct binding of the avid RH5 reagent (streptavidin-PE conjugated biotinylated RH5) to the surface of HEK-293-E cells, as expected (figure 4.2A). To test whether this binding was specific to BSG on the cell surface, I pre-incubated the cells with an anti-BSG antibody; however, the antibody, which has previously been shown to completely block this interaction in an in-vitro ELISA-based assay, did not prevent all RH5 binding on the cell surface even at 15 μg/mL final concentration. This additional binding was not due to a subfraction of inactive protein in the RH5 preparation since all binding could be prevented by heat treatment of the recombinant protein (figure 4.2B, figure 4.2C depicts the RH5 preparation that was used in the experiment). This observation suggested that there was an additional receptor/s for RH5 on HEK-293-E cells.

Genome-wide screens reveals BSG and heparan sulphate as independent receptors of RH5 on the cell surface.

To identify the receptor/s for RH5 other than BSG in the HEK-293-E cell context, I carried out a genetic screen with the avid-RH5 probe and compared the genes required for RH5 binding with those necessary for surface expression of BSG (data from the anti-BSG antibody screen that was carried out earlier- figure 4.3A). The enriched gRNAs common to both selections beyond those targeting general secretory pathway genes, corresponded to *BSG*, and the chaperone *SLC16A1*, as expected. In the case of RH5, the most highly enriched gene in the cells sorted using RH5 compared to anti-BSG mAb was *SLC35B2* (solute carrier family 35 member B2), which encodes a protein that transports 3'-phosphoadenosine-5'-phosphosulfate (PAPS), from the cytosol into the lumen of the Golgi apparatus where sulfotransferases use it as a universal

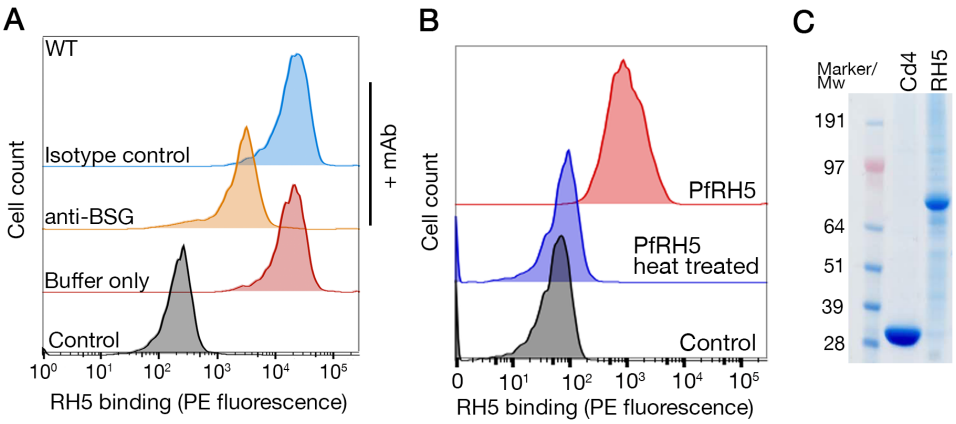


Fig. 4.2 RH5 binding to HEK293 was not completely dependent on BSG but was heat labile. Biotinylated RH5 was clustered around a streptavidin-PE conjugate and binding to HEK293 cells analysed by flow cytometry. **A.** RH5 binding was only partially blocked by an anti-BSG mAb relative to controls. **B.** Heat treatment (80°C for 10 minutes) of biotinylated RH5 abrogated all binding back to a negative control. In both cases a representative of three independent experiments is shown. Control represents binding of biotinylated protein tags (Cd4 alone) clustered around streptavidin-PE (henceforth referred to as Cd4-Strep-PE) to the HEK-293-E cell line. **C.** A coomassie stained (under reducing conditions) gel depicting the RH5 and the Cd4 used in **A** and **B**. Both bands are observed at the expected sizes (RH5- 88 kDa and Cd4- 28 kDa).

sulfuryl donor for the sulfation of major constituents of the glycocalyx including glycoproteins, glycolipids and glycosaminoglycans (GAGs) [247]. A broader pathway analysis using KEGG-annotated pathways identified the heparan sulphate (HS) but not chondroitin sulphate (CS)[1] biosynthesis pathway as significantly enriched (FDR < 0.05). Consistent with this, when I increased the FDR threshold to 0.25, I could identify two more genes (*NDST1* and *EXTL3*) encoding for enzymes that are critical for HS biosynthesis but not the genes critical for CS biosynthesis (*CSGALNACT1* and *CSGALNACT2*) (schematics depicted in figure 4.3B). Furthermore, RH5 binding to HEK-293-E cells could be inhibited to a threshold value by heparin[2], but not by CS (figure 4.3C). This is in agreement with the reported presence of heparin binding motifs in RH5 and its ability to bind heparin-coated agarose [248]. The role of *SLC35B2* and HS in RH5 binding was independent of BSG as the surface expression of BSG was not affected in cells in which genes required for GAG biosynthesis were targeted (figure 4.3C).

[1]HS and CS are types of GAGs. The HS biosynthesis pathway will be discussed in detail in figure 4.3B and again in section 4.2.2

[2]Heparin is commonly used as a model compound for the sulfated, protein-binding regions of HS. It is generally easier to obtain in higher quantities than HS.

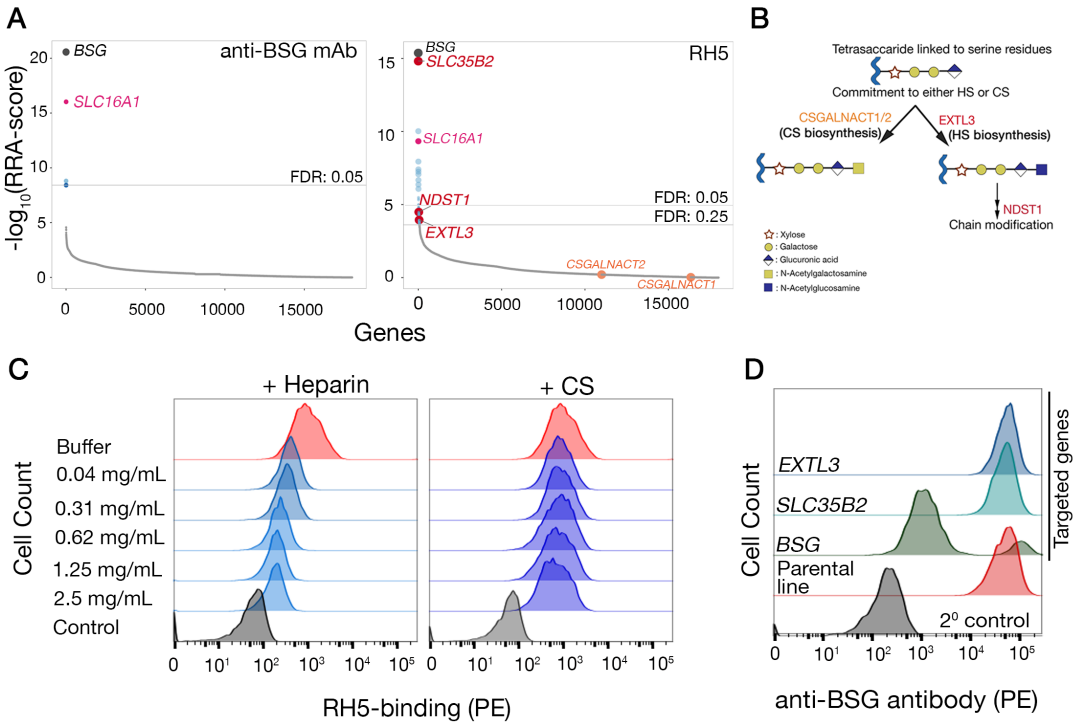


Fig. 4.3 Cell-based genetic screens identified BSG and heparan sulfate as independent receptors for *P. falciparum* RH5 on HEK293 cells. A. Rank-ordered genes identified from gRNA enrichment analysis required for cell surface display of an anti-BSG mAb (left panel) and RH5 binding (right panel). Significantly enriched genes with a FDR<0.05 and FDR<0.25 are marked separately (note that there were no additional genes identified in the anti-BSG screen with the increased threshold so the higher threshold value is not marked). Genes contributing to the HS-biosynthesis pathway were identified only in the RH5 screen. The full list of enriched genes for the RH5 screen is available in the appendix section table A.6. **B.** Schematic depicting the general GAG biosynthesis pathway with the relevant genes mapped to the corresponding steps. The general GAG biosynthesis pathway bifurcates into HS and CS biosynthesis pathway after the formation of a linkage tetrasaccharide structure [249] **C.** Clustered RH5 binding probe was pre-incubated with the indicated concentrations of either heparin or chondroitin sulfate (CS) prior to presentation to HEK-293-E cells and binding quantified by flow cytometry. Preincubation of RH5 with heparin showed a dose-dependent inhibition of binding up to a threshold; preincubation with CS showed no inhibition of RH5 binding, even at the maximum concentration of 2.5 mg/mL. The control represents binding of Cd4-Strep-PE to the cell line. A representative of three independent experiments is shown. **D.** Cells transduced with lentivirus-encoding gRNAs targeting enzymes in the heparan sulphate biosynthesis pathway show no alteration in surface BSG expression. Cell surface levels of BSG were quantified by flow cytometry on parental HEK-293-E cells or those transduced with lentivirus encoding single gRNAs targeting *SLC35B2* or *EXTL3* (genes required for HS biosynthesis), or *BSG*, as a control. Cells were stained with an anti-BSG mAb; control cells are stained with secondary antibody alone. A representative of two independent experiments is shown.

RH5 binding to HS was additive rather than co-dependent on BSG

To investigate further the role of SLC35B2 and HS in RH5 binding, I first tested the binding of an RH5 probe to cells targeted either for genes required for HS biosynthesis (*SLC35B2* and *EXTL3*), or for the known receptor *BSG*. In all three scenarios, I could only observe a partial reduction in RH5 binding (figure 4.4A). The residual binding in the case of cells lacking *SLC35B2* was specifically due to BSG because it could be completely blocked by the anti-BSG antibody (figure 4.4B). Whereas, in the case of cells lacking *BSG*, the residual binding was specifically due to HS because soluble heparin, but not soluble CS, could block all RH5 binding (figure 4.4C). This suggested that the RH5 binding to HS was additive rather than co-dependent on BSG.

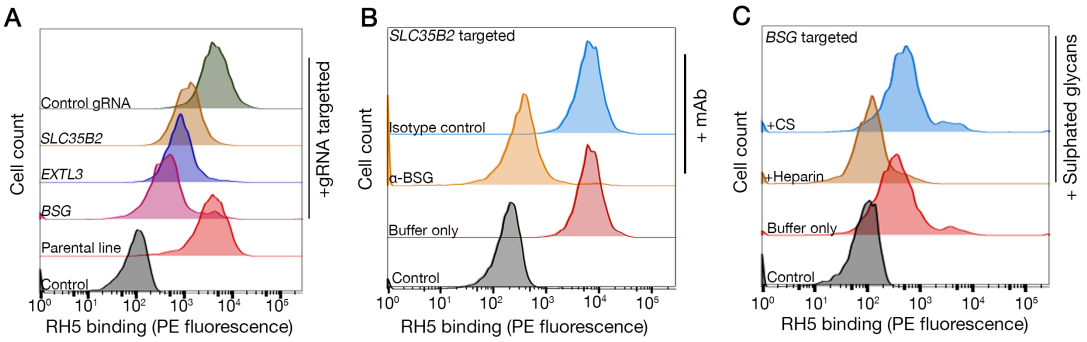


Fig. 4.4 The total observed binding of RH5 to HEK-293-E cell surface is the sum of independent contributions from BSG and HS. A. Binding of RH5 to cells is partially reduced when transduced with lentiviruses encoding gRNAs targeting either the receptor (*BSG*) or enzymes required for HS synthesis (*SLC35B2, EXTL3*) relative to controls. Transduced polyclonal lines were used for this experiment. **B.** RH5 binding to *SLC35B2*-targeted HEK-293-E cells could be completely blocked if preincubated with an anti-BSG mAb but not an isotype-matched control. **C.** RH5 binding to BSG-targeted HEK-293-E cells could be completely blocked if preincubated with 200 μg/mL heparin but not 200 μg/mL CS. A representative of three technical replicate experiments is shown in all three cases.

These experiments revealed a role for HS within the glycocalyx for interactions at the cell surface. Using the genetic approach, I was able to identify both the direct receptor of low affinity ligands (including the chaperone required for the receptor) and the contribution from HS in a single experiment. Further investigation showed that the binding contributions from the independent receptors could also be separated experimentally. I next proceeded to apply this approach to a panel of merozoite recombinant proteins with an aim to identify novel receptors and associated factors contributing to the receptor biology.

4.2.2 Heparan sulphates serve as common factors for cellular recognition

Initial screen of the protein library to identify candidates for screening

A pre-requisite for the screening approach designed here is the ability of a soluble probe to bind a cell line. Therefore, to first identify candidate proteins that bound to cell lines, I first tested the binding of a total of 11 avid monomeric biotinylated recombinant merozoite proteins clustered around PE (together with RH5 as a control) to a panel of six cell lines originating from different tissue sources (see figure A.2 in appendix section). From this initial screening list, I short-listed three merozoite proteins (SERA9, EBA181 and MSRP5), based on their bright staining on multiple cell lines tested (HEK-293-E, NCI-SNU-1, and KBM7).

The binding of multiple merozoite proteins to the cell surface is dependent on the cell surface HS

To identify the cellular factors mediating the binding of the merozoite proteins to the cell lines, I carried out the cell-based genome-scale KO screens in HEK-293-E cells using the avid monomeric protein probes. Notably, gRNAs targeting genes encoding for GAG (specifically HS) biosynthesis enzymes were found to be highly enriched in all three screens (figure 4.5A-EBA181, 4.5B-SERA9, 4.5C-MSRP5). I next mapped the genes identified in these screens to the HS biosynthesis pathway, from which I could clearly identify the majority of the pathway genes to be significantly enriched (figure 4.5D). The biogenesis of HS begins with the generation of a tetrasaccharide linkage on serine residue of the protein backbone through the sequential addition of four monosaccharide residues by glycosyltransferase enzymes. Commitment towards the HS pathway occurs via the EXTL3 enzyme, which adds the N-acetylglucosamine (GluNAc) residue to the existing polysaccharide chain. Within the Golgi, two enzymes EXT1 and EXT2 catalyse the initial chain polymerisation, during which multiple GluNAc and Glucuronic acid (GluUA) residues are added. A series of modifications to the growing polymer includes epimerization of GluUA to Iduronic acid (IdoA) by GLCE; N-sulfation by NDST family of sulfotransferases; and O-sulfation by and 2-O, 3-O and 6-O sulfotransferases [249]. Except for *GLCE*, genes encoding for the majority of these steps were identified in every screen. In all three cases, genes encoding for a candidate cell surface receptor proteins were not identified even at FDR<0.05.

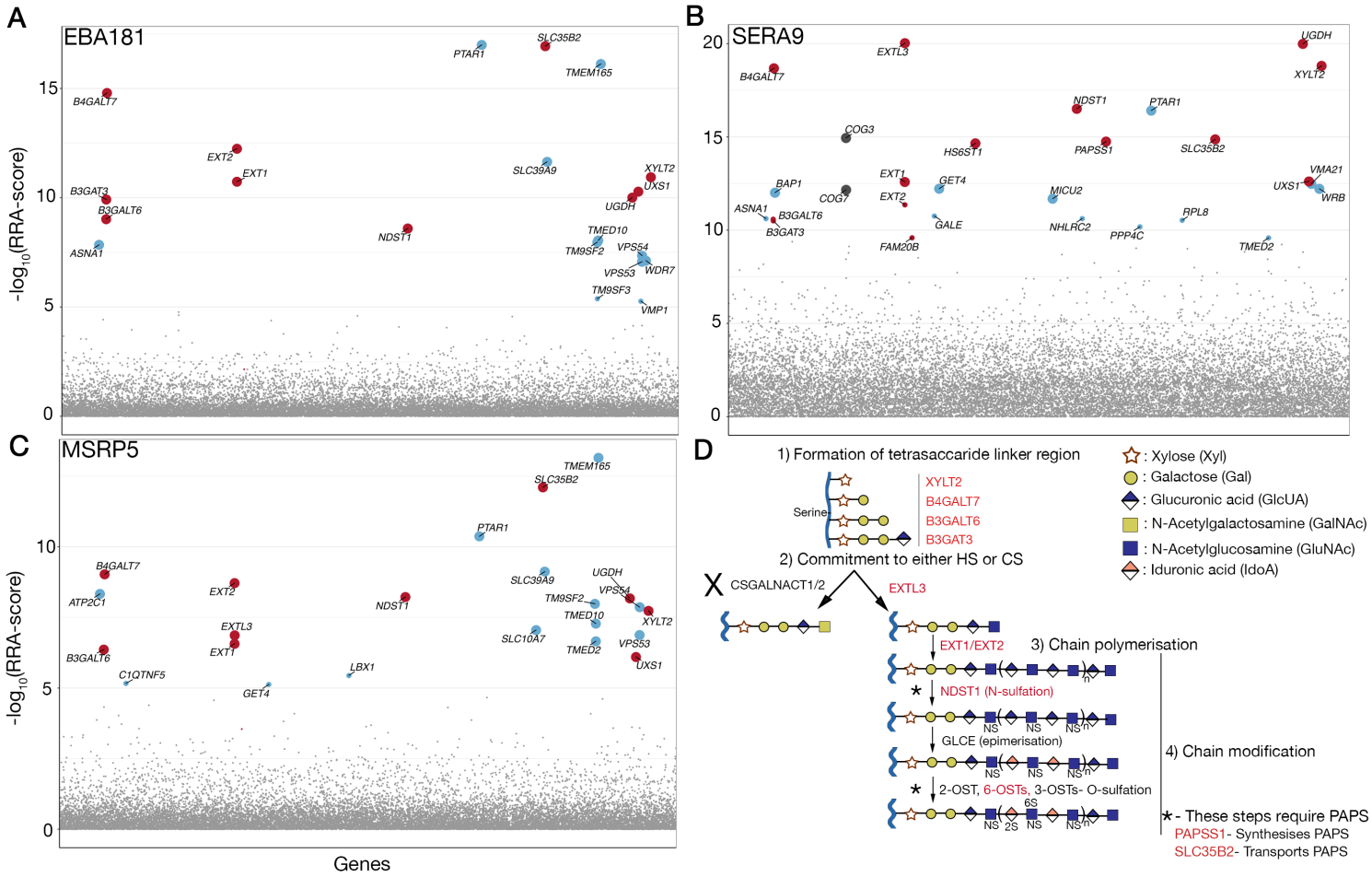


Fig. 4.5 All three genome-wide screens using merozoite proteins revealed the role of HS-biosynthesis pathway in mediating cell surface interactions. Gene level enrichment analysis on sorted mutant cells refractory to binding recombinant proteins- **A**-EBA181, **B**-SERA9 and **C**-MSRP5. Only genes with FDR<0.05 are labelled and the genes are ordered alphabetically. The highly enriched genes in all three cases correspond to the genes involved in HS-biosynthesis (labelled in red). This includes genes encoding enzymes directly catalysing each step of the pathway (pathway schematics with the corresponding enzymes depicted in **D** with the identified gene products from the screens highlighted in red); genes encoding enzymes that generate the activated monosaccharide precursors required for the generation of the initial tetrasaccharide structure, specifically UDP-xylose (*UXS1*) and UDP-glucuronate (*UGDH*); and genes encoding for a kinase that regulates the amount of mature GAG chains in a cell (*FAM20B*).

To validate the screen findings, I next investigated the extent to which the binding of the recombinant proteins depended on the presence of sulfated glycans on the surface of cells. I tested the binding of the three proteins in cell lines where *SLC35B2* was inactivated as *SLC35B2* was amongst the most enriched genes in all screens (FDR< 0.01). Unlike the case with RH5, in which *SLC35B2* inactivation led to partial loss of binding, in this case, there was a complete loss of binding of all three tested probes (figure 4.6A). This seemed to be a recurring theme in the interactions mediated by merozoite proteins as binding of two more merozoite ectodomains, CyRPA and RAMA, to HEK-293-E cells could also be completely abrogated by inactivating *SLC35B2* (figure 4.6B).

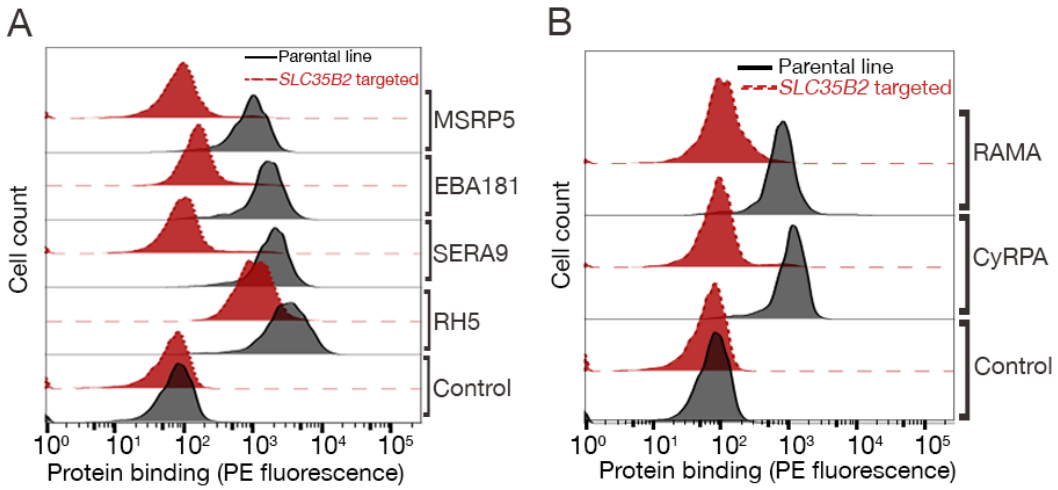


Fig. 4.6 The binding of multiple merozoite proteins to HEK-293-E cell surface can be completely abrogated by inactivating *SLC35B2*. Biotinylated ligands oligomerised around streptavidin-PE were tested for binding to an unmodified parental cell line (black histograms) or polyclonal *SLC35B2*-targeted cells (red histograms). **A.** All three proteins used for screening showed a complete loss in binding when tested on *SLC35B2*-targeted cells. In the case of RH5, a partial loss in binding was observed, as expected. **B.** RAMA and CyRPA demonstrated binding to HEK-293-E cells in the initial cell-binding assay. The binding of these two proteins were also completely dependent on *SLC35B2*.

Heparan sulphate proteoglycans are highly negatively charged biopolymers that have been known to bind to many ligands (e.g., growth factors, extracellular matrix proteins, chemokines, morphogenes, and cell surface proteins) usually via the sulfated domains within the HS chains. The interactions are largely electrostatic, with the brush-like negatively charged surface HS forming salt bridges with surface-exposed basic residues, and are generally thought to provide a suitable scaffold to present ligands to receptors in an appropriate

manner by regulating their orientation, oligomerisation and establishing local concentration gradients [21]. In the cell binding assay designed here, the contribution of HS towards binding was identified to be additive rather than co-dependent on other receptors, which suggested that HS may represent a factor responsible for cell surface binding for a range of ligands, even in the absence of another receptor. Given the observation with the merozoite proteins in which targeting *SLC35B2* was sufficient to inhibit all binding to the cell line, the presence of another receptor for these proteins in this cell line was unlikely. This posed a challenge in the screening strategy as binding to the cell line alone could no longer be the only pre-requisite for the screen as this binding could be due to proteins adsorbing into HS without binding to a specific receptor.

Development of a pre-screening approach to determine the fractional contribution of HS towards binding

To discriminate between proteins that only adsorb into HS versus those that interact with a specific receptor (with or without the contribution from HS), I took advantage of the *SLC35B2*-targeted cell line to rapidly determine the fractional contribution of HS adsorption to cell staining by comparing ligand binding events between the *SLC35B2*-targeted and the parental line (summarised in figure 4.7).

While using *SLC35B2*- targeted cells provides a rapid and consistent way to determine fractional contribution from HS, it is important to consider that SLC35B2 is involved in sulfation of other glycans, lipids and proteins in a cell, and thus its removal could have adverse effects to molecules other than HS. To address this, the approach I took was to first quickly assess binding on a cell line in which *SLC35B2* was targeted and upon observing a loss in binding confirm that this was specifically due to HS by either re-testing binding on cell lines where *EXTL3* (gene encoding for HS-specific enzyme) was targeted or with blocking experiments using soluble heparin. For the merozoite proteins for which I had not carried out the genetic screens (CyRPA, RAMA), I took the latter biochemical approach, in which I pre-incubated the recombinant proteins with a range of soluble heparin and demonstrated that concentrations above 200 μg/mL of soluble heparin led to a complete loss in binding to the parental cell line (appendix figure A.3). This indicated that all the five tested merozoite proteins from the initial screening list adsorbed into HS and this genetic approach would not be feasible to identify receptors for them.

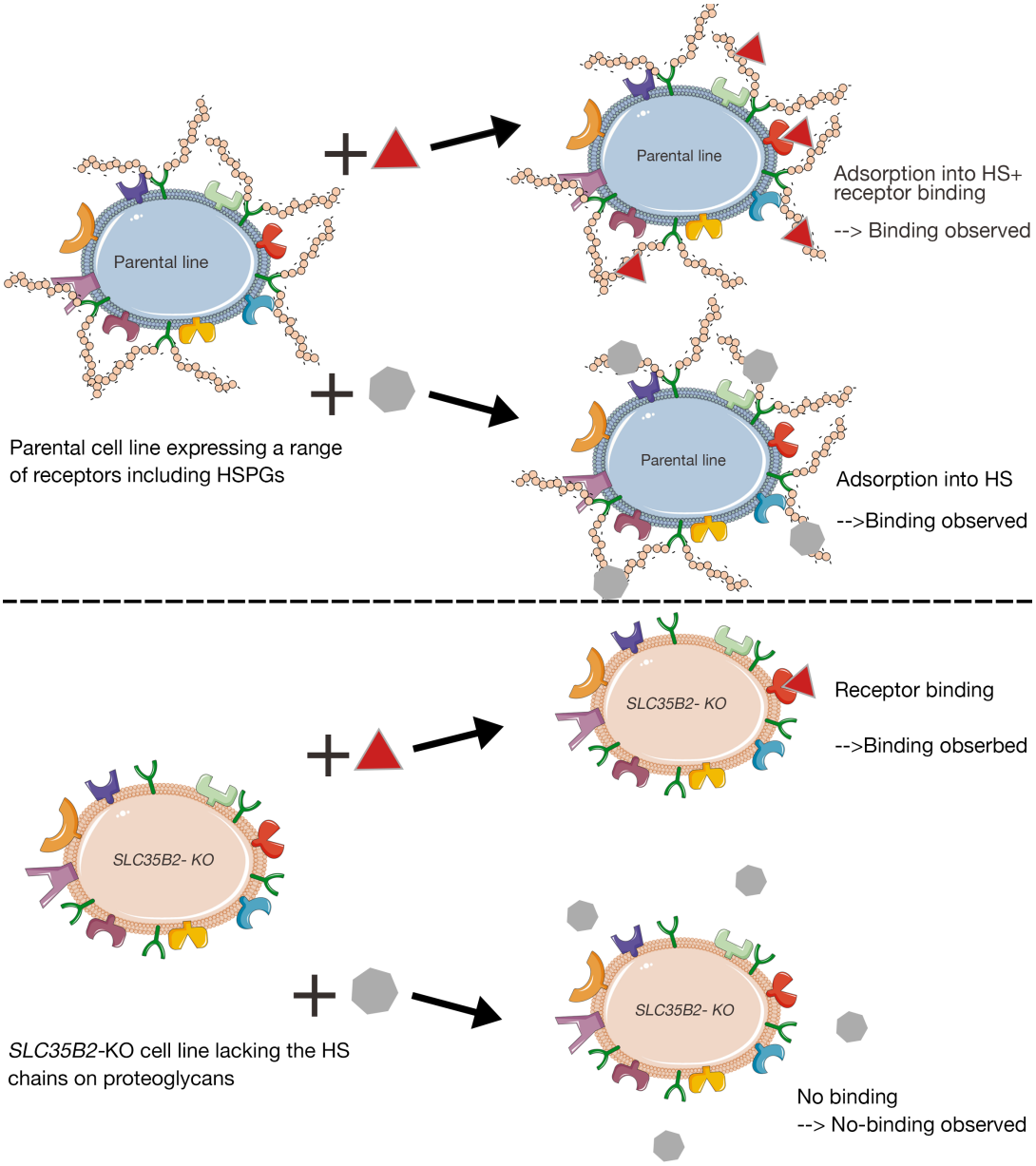


Fig. 4.7 Schematics for the approach to determine the fractional contribution of HS towards binding. Cell surfaces contain a range of receptors including heparan sulphate proteoglycans (HSPGs) such as syndecans and glypicans (represented in green) [250]. In the parental line, an observation of binding can be misleading as both proteins that bind to a specific receptor (represented in red) and proteins that adsorb into HS (represented by the chains emerging from HSPGs) display the same binding phenotype. However, the binding of a protein to the cell line that contains its specific receptor will be unaffected or only partially affected by targeting *SLC35B2* (as observed with RH5- represented by red triangles), whereas the binding of a protein which does not contain a specific receptor but adsorbs into HS will be predominantly lost in *SLC35B2* targeted cell line (as seen with all other tested merozoite proteins except RH5- represented by grey polygons). This approach would provide a basis for choosing candidates for the genome-scale screening system to identify receptors beyond HS. The genetic screen itself would be carried out on the parental cell line and not on the *SLC35B2* targeted line.

4.2.3 Investigating extracellular interactions mediated by human proteins

Application of the pre-screening approach to a panel of human proteins identifies the contribution of *SLC35B2* in binding.

To assess whether the pre-screening approach on cell lines lacking the GAG-biosynthesis enzymes would provide a rapid way to identify proteins that bind specifically to cell surface receptors, I next tested the binding of a panel of 51 human proteins, produced as recombinant pentamers from the platelet library, with the parental HEK-293-E cell line and HEK-293-E cell line in which *SLC35B2* was targeted (figure 4.8). From this panel of proteins, I identified six proteins that demonstrated binding to the parental cell line. Next, I categorised the six proteins into three classes, based upon their binding behaviour to cells with inactivated *SLC35B2* as compared with their binding behaviour to wild-type cells. The three categories were proteins with: (1) a small loss in binding ($<$ 20%, CD226, EPHB1, LPHN1) ; (2) a moderate loss in binding (>50%, G6B); and (3) a severe loss in binding (>80%, APLP2, APP).

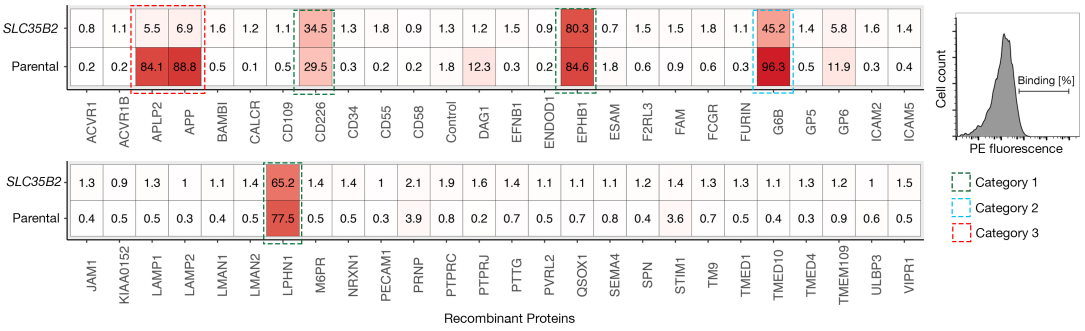


Fig. 4.8 The pre-screening strategy identifies human proteins whose binding is not solely dependent on SLC35B2. Pentameric ligands were tested for binding to either the unmodified parental cell line or to the polyclonal *SLC35B2*-targeted cells. The number in each grid represents the percentage of cells that fell within the 'binding' gate, which was drawn on the histogram obtained from control protein binding to the parental HEK-293-E cell line (depicted in the right panel). From the panel of 51 proteins, only six exhibited a clear binding ('binding' population higher than 25%). While CD226, EPHB1 and LPHN1 (category 1) retained more than 80% binding in the *SLC35B2*-targeted cells, APP and APLP2 (category 3) almost completely lost the binding. G6B (category 2) exhibited an intermediate phenotype in which more than 50% binding was lost.

To assess whether the loss of binding observed in the pre-screening step translated to the identification of the HS pathway in a genome-scale screen, I next carried out screens to identify the factors contributing to the binding of

proteins that exhibited a moderate or a severe loss (APLP2, APP, G6B) upon *SLC35B2* inactivation (figure 4.9A). A brief introduction to the investigated proteins is provided in table 4.1. In all three scenarios, the enriched genes contained enzymes required for HS biosynthesis as predicted (figure 4.9B). Other than the GAG biosynthesis pathway, a number of genes identified in these screens also overlapped with the screens carried out with merozoite proteins before (appendix figure A.4, genes identified in more than one screen are labelled in figure 4.9B as 'overlapping factors'). This included genes involved in processes such as the general secretory pathway, core glycosylation pathway, vesicular transport pathway, subunits of V-type ATPases, general transcription factors, elongation factors, and mRNA processing enzymes. In addition, genes encoding the proteins TMEM165 and PTAR1, the loss of both of which has been suggested to affect global glycosylation in cells [251, 144], were also identified in the majority of the screens. No specific cell surface receptor was identified in any of the screens. This demonstrated the applicability of the pre-screening step in rapidly determining HS contribution towards binding and showed that when the binding of the protein is mostly dependent on HS, this approach may have limited ability to identify a specific receptor.

Table 4.1 Background of the ligands that demonstrated dependency on *SLC35B2* for binding to HEK-293-E cells.

Ligand	Protein function	Known cell surface receptors	Ref.
APP and APLP2	Amyloid-beta precursor proteins (APPs) consist of APP, APLP1 and APLP2. All three are type I transmembrane proteins that are cleaved by secretases to form a number of peptides. The cleavage of APP leads to generation of the Aβ peptide, which is the major component of amyloid plaques found in the brains of Alzheimer's patients.	Both APP and APLP2 have been suggested to interact with HSPGs. APP interaction with HS on glypican 1 has been demonstrated in-vitro but the function of the interaction in-vivo is unclear. An ectodomain of APP, which is generated by cleavage with β-secratase has also been shown to bind to death receptor 6 (DR6) to activate intracellular caspases in axons.	[252, 253]
G6B	Cell surface receptor of the immunoglobulin superfamily that has been implicated to function in cellular recognition and signal transduction. It is expressed in platelets where it acts as a negative regulator of platelet function.	The ectodomain of G6B interacts with soluble heparin; however, it has been suggested to have a specific binding partner on the cell surface.	[254, 255]

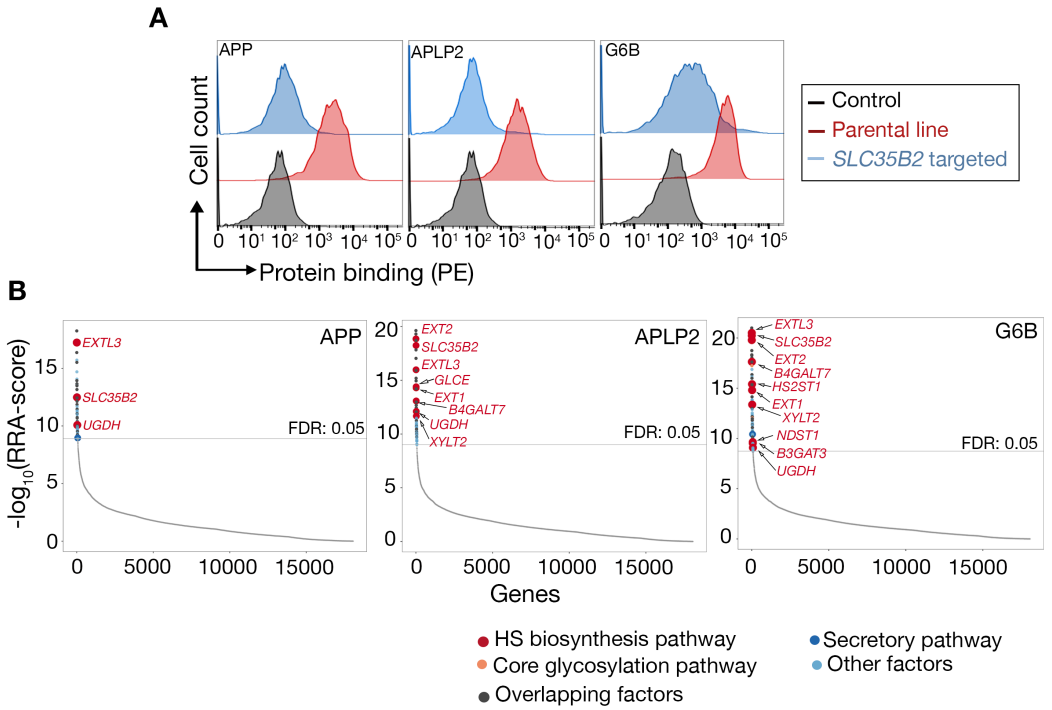


Fig. 4.9 Genome-scale screens identify HS-biosynthesis pathway when ligands that lose majority of binding upon targeting *SLC35B2* are used as screening probes. A. Pentameric ligands were tested for binding to unmodified parental cell lines or to polyclonal *SLC35B2*-targeted cells. The HEK293 cell line was used for all proteins. **B.** RRA-score rank-ordered genes identified from gRNA enrichment analysis from sorted cells which had lost binding to APP, APLP2 and G6B; in all three cases, genes encoding the cellular factors required for HS-biosynthesis pathway were identified. Multiple genes identified in this screen were also identified in screens carried out before with merozoite proteins. Overlapping factors represent genes that have been identified in at least any two out of the six screens, which have identified the HS-biosynthesis pathway.

Genome-scale cell-based CRISPR screens using recombinant protein probes identify directly interacting receptors

I next carried out genetic screens with ligands, which showed no (CD226, EPHB1) or merely fractional (<10-20%) decrease in binding upon targeting *SLC3B2*, to evaluate whether the screening approach would be able to identify specific receptors on the cell surface (figure 4.10A). Where the overall binding of the ligand had been established to have no contribution from *SLC35B2*, the gene with the most enriched gRNAs corresponded to a known receptor (refer to table 4.2) in every case: *EFNB2* was the top-ranked gene when selected with the EPHB1 ligand and *PVR* for CD226 (figure 4.10B). In the case of LPHN1, the binding of which had a partial contribution from SLC35B2, the most enriched genes included a known receptor - *TENM4* - as well as *SLC35B2* (figure

4.10C). These experiments demonstrated that the pre-screening approach is a useful way to rapidly establish whether the protein contains a receptor on the cell line, and that the directly interacting receptor in such cases can be determined with the genome-scale screening approach.

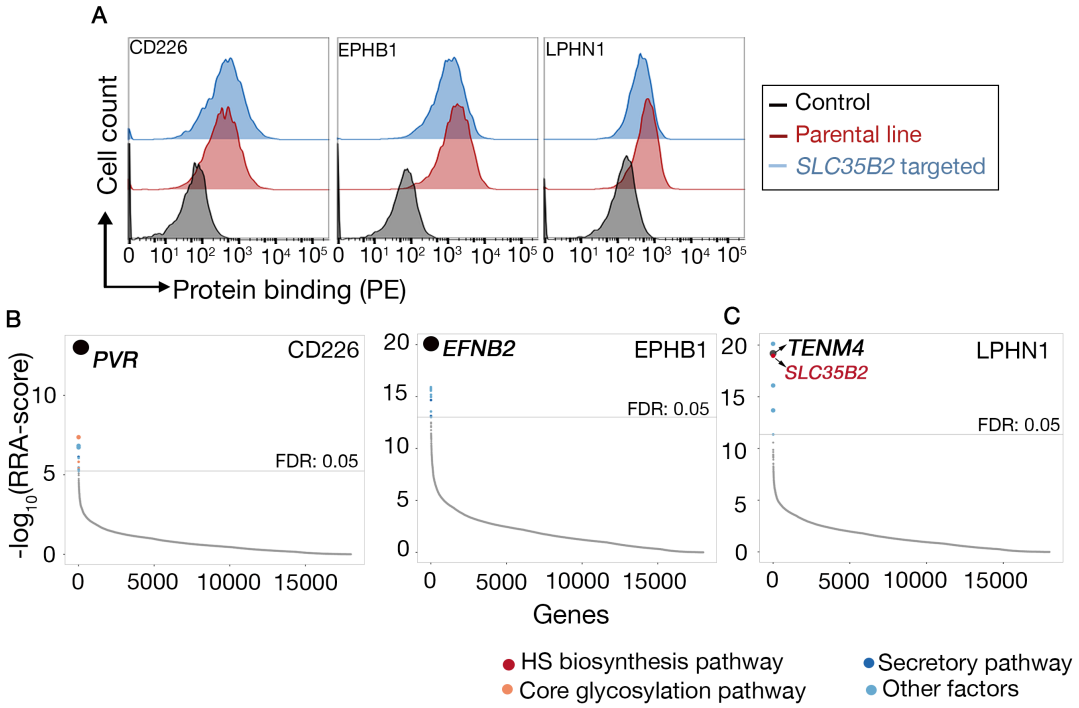


Fig. 4.10 Cell surface receptors were identified using cellular genetic screens. **A.** Pentameric ligands were tested for binding to the unmodified parental cell line, or to the polyclonal *SLC35B2*-targeted cells. The HEK293 cell line was used for all proteins. **B.** RRA-score rank-ordered genes identified from gRNA enrichment analysis from sorted cells which had lost binding to CD226 and EPHB1; in both cases, the gene encoding the known receptor was identified as the most significantly enriched gene with no contribution from HS, as expected. **C.** Rank-ordered genes identified from gRNA enrichment analysis from sorted cells which had lost binding to LPHN1. The top four genes were identified with the same FDR: these included a gene encoding a known receptor *TENM4*, *SLC35B2*, and two genes relating to global glycosylation in cells (*PTAR1* and *TMEM165*, which were also identified previously in the screens that identified the HS-biosynthesis pathway). The list of all identified genes with FDR<0.05 in each case is provided in appendix section table A.6.

Table 4.2 Background of the identified receptor-ligand interaction partners. The three detected receptor ligand-pairs in this genetic screening approach were identified in the past using distinct biochemical and genetic methods. The interactions represent both low (CD226-PVR) and high (EPHB1-EFNB2 and LPHN1-TENM4) affinity interactions.

Ligand	Receptor	FDR	Approach used in the past	Reported K_D	Ref.
CD226	PVR	0.005	Expression cloning with cDNA library using CD226 conjugated to Fc domain as a probe.	2.3x10^{-7} M	[256]
EPHB1	EFNB2	0.005	Ephrin B2 (EFNB2) was initially identified as an ephrin ligand as it shared sequence homology with then known ephrin ligand, EFNA2. Identification of EPHB1 as one of the binding partners was done through cDNA overexpression and cell binding assay.	0.78 x10^{-9} M	[257]
LPHN1	TENM4	0.005	Affinity purification from rat brain with the ectodomain of LPHN1 fused to the Fc domain followed by mass-spectrometry.	1.07x 10^{-9} M	[258]

Having established a system from which I could pre-screen for candidate proteins that bind to specific receptors on the cell surface, I next established a screening pipeline (summarised in figure 4.11) through which I could 'feed in' recombinant proteins to identify candidates that could be used in a genome-scale screening approach to identify novel receptors and non-receptor cellular factors that contribute to the biology of the receptor.

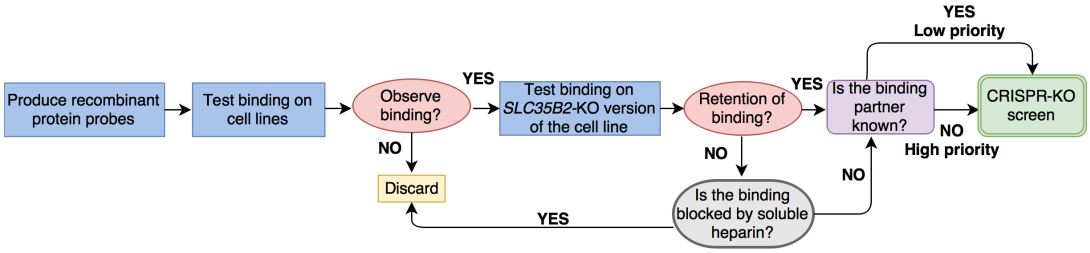


Fig. 4.11 Strategy for genetic screening using recombinant proteins. The strategy is based on the retention of binding on *SLC35B2*-targeted cell lines. A conclusive test for HS-adsorption is carried out with a blocking assay with soluble heparin. Where the receptor was already known, genetic screens were carried out to identify additional information regarding the biology of the receptor. For this purpose, *SLC35B2*-deficient versions of six different cell lines (HEK-293-E, HEL, NCI-SNU-1, KBM7, HL-60, HepG2) were produced and an additional ∼80 human proteins were screened. The binding of the merozoite proteins were also re-tested on the *SLC35B2*-KO versions of the cell lines generated here and in all cases, binding was completely abrogated (data not shown).

Genome-scale cell-based screens using recombinant protein probes also identify intracellular pathways required for receptor expression on the surface of cells

One of the ligands from the screening pipeline that I next investigated was TNFRSF9 (also known as CD137 or 4-1BB-receptor), which demonstrated *SLC35B2*-independent binding on the NCI-SNU-1 cell line (human colon cancer cell line). TNFRSF9 belongs to the tumor necrosis receptor family primarily present in T cells where its expression is up-regulated upon receiving antigen-specific signals. TNFRSF9 is a well-characterised co-stimulatory molecule on T cell and is known to interact with TNFSF9 (or CD137L), which is primarily expressed on antigen-presenting cells (APCs). The stimulation of TNFRSF9 has been shown to suppress tumour growth in murine models of sarcoma, mastocytoma, and glioma, which has made this antigen an attractive target for cancer immunotherapy [259, 260]. The expression of both TNFRSF9 and its ligand TNFSF9 is not restricted to the cells of the immune system: a number of non-lymphoid cell lines have also been shown to express these proteins on their surface. Human colon carcinoma lines have been shown to constitutively express varying levels of TNFSF9 that is able to functionally interact with TNFRSF9 on anti-CD3 activated T-cells [261].

To investigate the factors contributing to the interaction of TNFRSF9 with the NCI-SNU-1 cell line, I carried out a genome-scale screening approach using a monomeric biotinylated TNFRSF9 clustered around streptavidin-PE as the selection ligand. Enrichment analysis using cells refractory to binding the ligand revealed genes corresponding to the known interaction partner TNFSF9 as the most enriched gene. In addition, several genes involved in the p53 pathway (*CDKN2A, CDC37, STK11* and *DYRK1A*) and *TP53* itself were also enriched in the non-binding population, suggesting a role for the p53 pathway in presenting TNFSF9 in a ligand-binding form on the cell surface (figure 4.12B-see figure 4.12C for the relationship of these genes toTP53). I validated this by independently targeting *TP53*, which resulted in a decrease in the binding of the TNFRSF9 ligand (figure 4.12D).

These experiments further demonstrated that where binding is not largely dependent on adsorption into HS, the directly interacting receptor and the cellular pathways responsible for the cell biology of the receptor can be determined using this method.

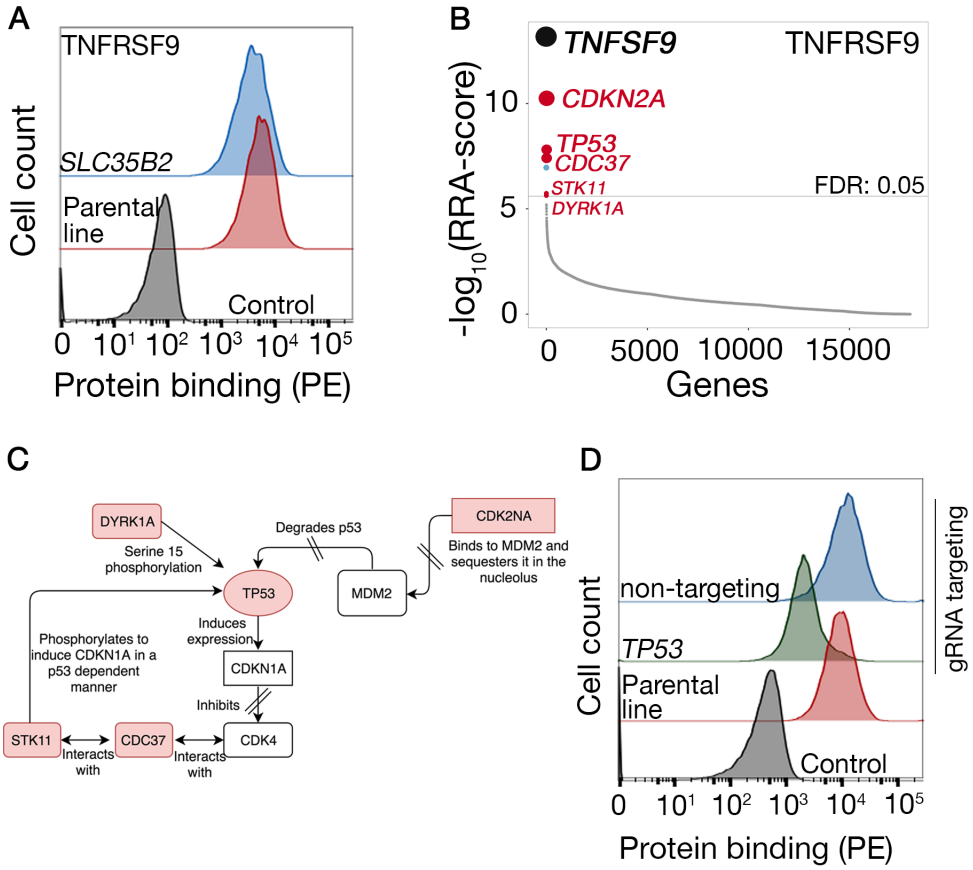


Fig. 4.12 Genome-scale screen using TNFRSF9 ectodomain as a sorting ligand identifies the interaction partner along with the p53 pathway. A. Monomeric biotinylated TNFRSF9 conjugated to streptavidin-PE binds to NCI-SNU-1 cell line; this binding was not affected by targeting of *SLC35B2*. **B.** RRA-score for genes that were identified to be enriched in the sorted cells that were refractory to binding to the TNFRSF9 probe. Genes are ranked according to the RRA-score. The known interaction partner *TNFSF9* and genes related to the TP53 pathway (labelled in red) were identified in the screen. **C.** The relationship of genes identified in **B** with TP53. Identified gene products are highlighted in red. **D.** TNFRSF9 binding to NCI-SNU-1 cells was reduced in *TP53*-targeted cells relative to a non-targeting control; targeted cells were maintained as polyclonal lines. **A** and **C** show representative experiments of three technical replicates. Refer to table A.6 in the appendix section for identity of all genes identified with FDR<0.05.

4.3 Discussion

In this chapter, I demonstrated how the genome-scale screening approach can be used to investigate interactions with cell lines mediated by recombinant protein probes. In the proof-of-principle demonstration using RH5 ectodomain as a screening probe, I was able to clearly identify, in a single experiment, both the direct receptor BSG (and the associated chaperone SLC16A1) and the contribution from the component of the cellular glycocalyx, especially HS. Further experiments with RH5 interaction with the surface of cells showed that the contributions to RH5-binding by its specific receptor and HS, at least in this context, were independent and could be experimentally separated.

When the approach was used subsequently to identify receptors for five other merozoite proteins, I consistently observed complete dependence of the proteins on the cellular HS-biosynthesis pathway to mediate binding to the cell lines. The role of sulfated glycans in the context of host-parasite interactions has been studied in the past: the addition of heparin to in-vitro cell cultures of *P. falciparum* has been shown to block the invasion of merozoites into the red-blood cells. It has been suggested that heparin like molecules could be involved in the initial attachment of the parasite to the host cell [262, 263, 264], as has been shown for the attachment of various viruses, bacteria and other parasites to the host cells [265, 139, 266, 267]. Heparin or heparin-like molecules have been shown to interact directly with recombinant or native merozoite proteins [268, 269, 248]. One such study, which used pull-down experiments with heparin affinity chromatography, has even suggested that almost all of the erythrocyte-binding proteins of *P. falciparum* (for example EBA140, RH2, RH4, and RH5) have the capacity to bind to heparin-like molecules [270]. It is not always easy to predict whether proteins have the ability to interact with HS, as there does not exist a specific protein fold or recognizable amino acid sequence patterns that determines the binding of proteins to HS[3]. Rather, the majority of the HS interactions are mediated by the negatively charged sulfated groups on the polysaccharides with the positively charged amino acid residues (such as lysine and arginine or protonated histidine residues at low

[3]Attempts to identify HS-binding domains in proteins have been made with suggestions of XBBXBX or XBBBXXBX, where B is lysine or arginine and X is any other amino acid as potential sequences. Such sequences can be identified in multiple merozoite proteins, including RH5; however, it is now suggested that such sequences merely imply a possible interaction with heparin and should not be taken as a proof for interactions under physiological conditions [271].

pH values) present on the external surface of the folded proteins [271]. Based on the observations here and from the studies from before, it is possible that a number of merozoite proteins possess such charged surfaces enabling them to interact with HS.

However, the question of whether the merozoite invasion ligands actually bind to heparin-like molecules in a physiological context remains to be answered. Some of the proteins that have been shown to bind heparin, such as the RH-family proteins (RH2, RH4, RH5), reside in the intracellular vesicular compartments of the parasite, are only released for a very short time (approximately 30s) during the invasion process, and are thus unlikely to encounter heparan sulfates. Furthermore, the proteins that have been shown to interact with heparin have also been shown to have a specific receptor (for example, BSG as the receptor of RH5 [88], and Complement receptor 1 (CR1) as the receptor of RH4 [272]). This suggests that the ability of the many merozoite proteins to interact with HS is a generic property of the proteins, and there possibly exist specific cell surface receptors for some of these proteins other than HS. Using the example of RH5, I demonstrated here that the contribution from HS and the specific receptor can be experimentally separated. Therefore, to use this approach for identifying specific receptors for merozoite proteins, it is important to first identify a cell line where the binding of the proteins is not completely dependant on the presence sulfated glycans, as it is likely that it will be in those cell lines there exists a specific receptor for these proteins. In this work I was unable to find such a cell line for the 11 merozoite proteins that were tested.

The HS-binding behaviour was not limited to merozoite proteins as some human proteins such as APP, APLP2 and G6B were also found to interact with HS. All of these human proteins have been shown to bind to soluble heparin in the past (see table 4.1), but the relevance of such binding behaviour in physiological condition is still not understood. There are over 100 proteins that have been described in literature to bind HS to carry out diverse cellular functions such as cell adhesion, migration, regulation of enzyme activity, and protection of proteins against degradation [271]. Members of the fibroblast growth factor (FGF) family have been shown to interact with heparin with high-affinity (K_D of nM range) and this association has been shown to be physiologically important for the signalling via FGF-FGF receptor(FGFR) complex, which is required during development [273]. It has been suggested that HS chains provide a linear domain over which growth factors can diffuse thereby increasing the local

concentration of the ligand so as to facilitate receptor-ligand interactions [271]. While such biologically relevant HS-mediated interactions do exist, there are suggestions that the specificity of HS-mediated interactions in such situations is determined by spatial and temporal expression of proteoglycans carrying the HS chains; if the ligand is not exposed to HS proteoglycan, it cannot interact even if it has that ability in principle [271]. Using the genetic screen approach on cell lines, in which proteins that have the ability to interact with HS were introduced to cell surfaces that bear HS, it was difficult to assess to what extent such bindings were relevant in biological contexts. Therefore, I decided not to focus on proteins that solely depended on HS for interaction at the cell surface.

To identify and eliminate proteins that only adsorbed to HS without binding to a specific receptor, I designed a strategy to rapidly establish the extent to which sulfated GAGs contributed to the observed binding events. The approach I took was to test binding on an *SLC35B2*-deficient cell line (used as a proxy for cells that lack HS), which provided clues as to the presence of a specific receptor on the surface of the chosen cell line. In situations where the binding was not largely dependent on the presence of *SLC35B2*, I was able to identify the specific receptor in every case attempted. In the case of TNFRSF9, I was also able to identify components of the p53 pathway to be important for the interaction of this probe with TNFSF9. It is possible that TP53 acts as the transcriptional factor responsible for the expression for TNFSF9 in these cell lines and this hypothesis is consistent with a previous report demonstrating the presence of TP53 binding sites in the promoter region of TNFSF9 [274]. This example showed the promise of this method in identifying the intracellular pathways that can contribute to interactions occurring at the surface of the cells.

In summary, in this chapter, I adapted the genome-scale screening system to identify receptors and non-receptor cellular factors contributing to the interaction of cells with recombinant protein probes. Below I will highlight some of the lessons learnt from adapting this strategy to use recombinant protein probes:

- Recombinant protein probes can be used in this screening system to identify both low and high affinity extracellular protein-protein interactions. The method is also able to identify intracellular pathways that contribute to the biology of the receptor.

- It is important to be aware that observation of binding on a parental cell line does not equate to the presence of a specific receptor as some proteins tend to adsorb into HS present at the cell surface. The binding behaviour of some proteins to HS in this system is not an indication that such interactions also occur in physiological conditions.
- If a protein binds to a specific receptor but also adsorbs into HS, both components can be identified using this method. In the example here, such binding events were found to be additive and not dependent on each other.
- The main limitation of this technique is that genes required for cell surface interactions that are genetically redundant might be refractory to identification. In the example of the screens in which heparan sulphate was identified, specific proteoglycans that carry the HS-chains in the cells were not identified. The reason for this could be the redundancy in these classes of molecules. Cell surface HSPGs are composed of the GPI-linked glypicans and the transmembrane syndecans. There are four syndecans (SDC1-4) and six glypicans (GPC1-6) in mammals [250], which can function equivalently on the cell line to provide HS chains; this could have precluded their identification. That said, such screens do still provide a wealth of information about the biology of the receptor, which can then be used in subsequent follow-up studies to identify the precise nature of the ligand-receptor interaction.

CHAPTER 5

IDENTIFICATION AND CHARACTERISATION OF IGF2R AS AN ENDOSOMAL RECEPTOR FOR GABBR2

5.1 Introduction

This chapter describes the application of the genome-scale screening approach to identify the interaction between a mannose-6 phosphate receptor (IGF2R) on the surface of HEK-293-E cells and the recombinant protein corresponding to the ectodomain of the B2 subunit of the gamma-aminobutyric acid (GABA) type B receptor (GABBR2 or $GABA_{B2}$). The interaction was subsequently validated using biochemical approaches.

5.1.1 Introduction to GABA-B receptors

Communication between neurons occurs via the release of small chemical molecules called neurotransmitters into the junction between two neuronal cells (called synapses). Neurotransmitters are of two types (i) excitatory, which increase the excitability of neurons, and (ii) inhibitory, which reduce the excitability of neurons. Gamma-aminobutyric acid (GABA) is the major inhibitory neurotransmitter of the mammalian central nervous system (CNS) [275]. $GABA_B$ receptors are widely expressed metabotropic transmembrane receptors for GABA that mediate slow inhibitory neurotransmission in the CNS to control the excitability of neurons [276]. The $GABA_B$ (or GABAB) receptor is composed of two subunits, $GABA_{B1}$ and $GABA_{B2}$, both of which are members of the GPCR family and contain the characteristic GPCR seven transmembrane (7-TM) regions. Both subunits also contain an extracellular domain called the 'venus fly trap domain' (VFTD), which acts as the binding site for GABA. Functional $GABA_B$ receptors are generated through the combination of either of the isoforms of the $GABA_{B1}$ subunit ($GABA_{B1a}$ or $GABA_{B1b}$) together with the $GABA_{B2}$ [277].

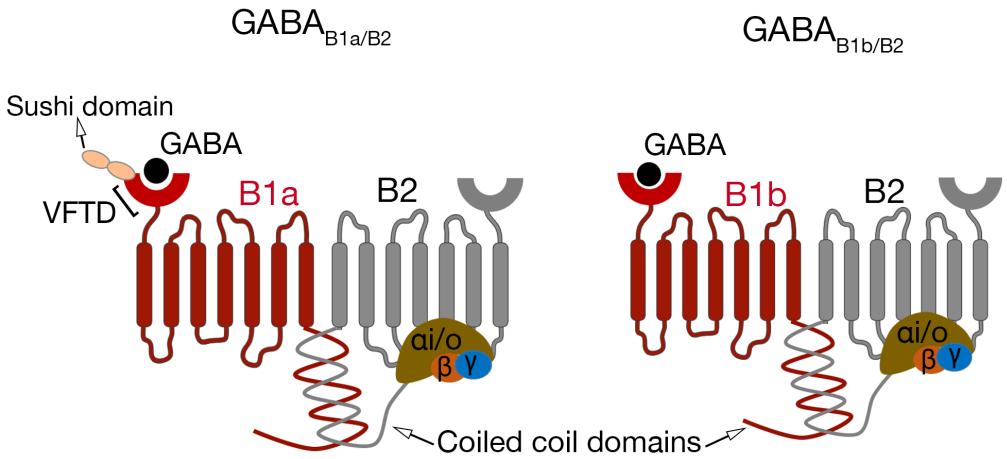


Fig. 5.1 Schematics of GABA$_B$ receptors. Formation of functional GABA$_B$ receptors requires dimerisation of the B1 and B2 subunits [278]. There are two isomers of B1 subunits that differ by the presence of two sushi domains on the B1a but not on the B1b subunit. GABA$_{B1a/B2}$ are preferentially found presynaptically whereas GABA$_{B1b/B2}$ are mainly localised postsynaptically [279]. The subunits interact via the coiled coil domains present on the C-terminus of the proteins. This interaction has been shown to be required for the transport of the GABBR1 subunit to the plasma membrane. The GABBR1 subunit contains an ER retention signal which is masked when it interacts with the GABBR2 subunit, thereby allowing it to reach the cell surface. The binding of GABA is medaited by the GABBR1 subunits whereas the GABBR2 subunit couples with the subset of heterotrimeric G-proteins (pertussis toxin sensitive Gi/o family) to regulate voltage-gated Ca(2+) (Ca(V)) channels, G-protein activated inwardly rectifying K(+) (GIRK) channels, and adenylyl cyclase activity [280].

Internalisation of GABA$_B$ receptors

The balanced expression of excitatory and inhibitory receptors in neurons is crucial for normal brain function. Disruption to such balances has been been implicated in a wide range of neurological disorders including anxiety, depression, epilepsy and neuropathic pain [281]. Internalisation of GPCRs receptors has been most comprehensively studied in the context of β-adrenergic receptors, the expression of which on the surface of cells is tightly controlled by agonist desensitization. In this model, upon a prolonged stimulation of GPCRs by an agonist, the agonist-bound GPCRs are phosphorylated by G protein-coupled receptor kinases (GRKs), which leads to decreased effector coupling and recruitment of arrestins and clathrins, in turn leading to the endocytosis of receptors followed by recycling or degradation [282, 283]. However, unlike the other GPCRs, GABAB are not GRK substrates [284, 285] and thus the mechanisms by which the plasma membrane expression of GABA$_B$ receptors are regulated have been suggested to follow other mechanism of internalisa-

tion rather than the classical pattern of agonist-induced desensitisation and internalisation[1]. While initial studies suggested that $GABA_B$ receptors are stably expressed on the surface of cells with very low rate of constitutive internalisation [285, 290, 284], it is now generally accepted that $GABA_B$ receptors are highly mobile and can undergo constitutive internalisation in both heterologous expression systems (including HEK293 cells) and cultured neurons even in the absense of an agonist [289, 291, 292, 288]. The internalisation of the receptor heterodimer has been reported to be carried out via clathrin- and dynamin-dependent pathways. Internalised receptors are either targeted to endosomes [289, 288] or degraded in the lysosomes [289]. Receptors destined for endosomes have been reported to be recycled back to the surface of the cells.

Several mechanisms mediated by post-translational modifications of heterodimer subunits have been proposed for the regulation of the expression of GABAB receptors on the surface of cells. It has been suggested that sustained activation of the N-methyl-D-aspartate (NMDA) receptor by glutamate causes the activation of the AMP-dependent protein kinase (AMPK) and the protein phosphatase 2A (PP2A). This in turn leads to the phosphorylation of serine 783 (S783) residue on the C-terminus of GABBR2 subunit, which causes the change in the fate of internalised GABAB receptors—rather than being recycled back to the surface of the cells, the receptors are instead directed towards lysosomal degradation [293]. Another post-translational mechanism that is known to contribute to receptor internalisation and regulation is ubiquitination. The role of ubiquitination in GPCR internalisation is well-studied in case of β-adrenergic receptors in which there exists an agonist-dependent ubiquitination of both β-arrestin and the receptor. It has been shown that ubiquitination on β-arrestin and on β-adrenergic receptors serve distinct signals. While the ubiquitination of β-arrestin is required for receptor internalisation, that of the receptor is required for receptor degradation [294, 295, 296, 297]. Although it is known that GABAB receptors do not engage β-arrestins, Lysine-63-linked ubiquitination of the GABBR1 subunit and its interaction with USP14 (ubiquitin-specific protease 14) has recently been reported to play a role in lysosomal targeting of GABAB receptors [298, 299]. While these studies provide evidence for ubiquitination mediated regulation of an amount of receptors present

[1]Studies regarding the effect of an agonist are conflicting as there are some studies that have shown agonist-induced internalisation [286, 287] and others that show no influence of an agonist on internalisation [288, 289].

on the plasma membrane, the mechanism by which GABAB receptors are constitutively internalised using clathrin-dependent endocytosis pathway still remains unknwown. The GABBR2 subunit of the GABAB receptors has also been suggested to play a regulatory role for mediating cell surface receptor stability of GABAB receptors, but the precise mechanism for this is also not known [292].

5.1.2 Introduction to IGF2R

IGF2R (also known as cation-independent mannose-6-phosphate receptor (CI-MPR) or CD222) is a multifunctional membrane-bound glycoprotein. It is mainly localised in the trans-Golgi network (TGN) and endosomes (90-95%) with 5–10% of the receptor expressed on the plasma membrane [300]. IGF2R, together with CD-MPR (cation-dependent mannose-6-phosphate receptor), makes up the class of 'P-type' lectins that primarily function in sorting mannose-6-phosphate- (M6P-) bearing glycoproteins from TGN to endosomes and lysosomes (figure 5.2A, the schematic representation of IGF2R is shown in figure 5.2B) [301].

Non-lysosomal targets of IGF2R

Apart from the transport of enzymes destined for lysosomes, IGF2R also functions in regulating other cellular functions by interacting with multiple non-lysosomal proteins. One of the first identified non-lysosomal protein bound by IGF2R was the insulin growth factor II (IGF-II, hence the name IGF2R). IGF2R has been suggested to be a 'sink' for excess IGF-II in the extracellular fluid [301]. The interaction with IGF2R leads to internalisation and subsequent degradation of IGF-II in the lysosomes. IGF-II does not contain M6P modifications and binds to a site different to the M6P binding sites of IGF2R. As IGF-II is important for cell growth, survival, and migration, maintenance of correct levels of IGF-II in the extracellular space has been shown to be crucial for normal growth and development of cells [302]. In addition to IGF-II, IGF2R interacts with multiple non-lysosomal ligands in both a M6P-dependent (such as Transforming growth factor-β precursor (TGFβ1), Leukemia inhibitory factor (LIF), Proliferin, Granzyme B, CD26, Herpes simplex virus glycoprotein D (HSV-glyD)) and -independent (such as Retinoic acid (RA), uPAR: urokinase-type (plasminogen activator) receptor, Plasminogen) manner. A summary of the cellular context in which the some of the interactions

with IGF2R occur, as well as the consequences of the interactions, is provided in table 5.1 (reviewed in [300]).

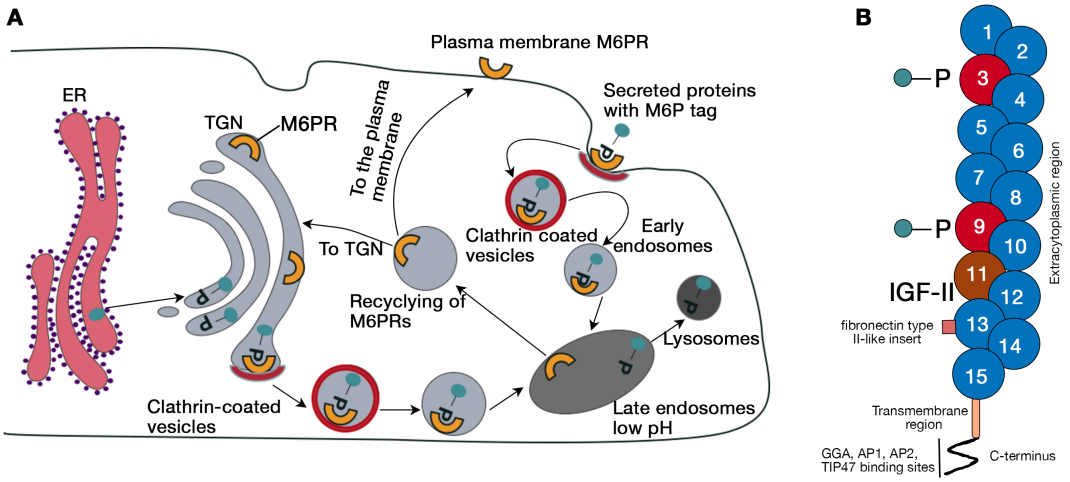


Fig. 5.2 Schematics of mannose-6-phosphate cellular transport pathway and structure of IGF2R. A. Newly synthesised enzymes such as soluble acid hydrolases that are destined for the acidic compartments in cells are post-translationally modified to contain M6P residues on their N-linked oligosaccharides in the golgi network (represented with -P). These residues are recognised by cellular M6P receptors (CI-MPR and CD-(cation-dependent)-MPR), which causes the receptors and their ligands to cluster into clathrin-coated transport vesicles at the trans-Golgi-network (TGN). Clathrin coated vesicles bud-out of the TGN and travel to an acidic late endosomal compartment in which the low pH causes dissociation of the receptor–ligand complex. The unbound M6PRs are then trafficked back to the TGN or trafficked to the plasma membrane. In some cases, lysosomal enzymes that carry the sorting tag escape the transport to the endosomes and are instead secreted. M6PRs (mainly IGF2R) present on the surface of cells are required for the 'secretion-recapture' pathway where such escapees are captured and are brought back into the cell via the clathrin dependent endocytosis pathway. These proteins are also eventually transported to the late endosomes (schematics is based on [303]). **B.** IGF2R is a 300 kDa glycoprotein that contains 15 repetitive extracytoplasmic domains each with 147 amino acids that share $\sim$14 % - 38 % sequence similarities. Monomeric receptors are found on the membrane surfaces; however, weak dimeric complexes formed upon ligand binding have also been reported. IGF-II binds to a site (domain 11) separate from the M6P binding sites (domain 3 and 9) of IGF2R. The c-terminus of CI-MPR contains important amino acid sorting signals (for example, GGA, Golgi-localized, gamma-ear-containing, ADP-ribosylation factor-binding protein TIP47, tail-interacting protein; AP1 and AP2, clathrin adaptor proteins) [303].

Table 5.1 Non-lysosomal interaction partners of IGF2R

Ligand	Context of interaction	Ref
	M6P bearing ligands	
TGF-β	The interaction leads to plasmin-mediated proteolytic activation of the precursor to generate active growth factors at the cell surface	[304]
LIF	Endocytosis and degradation in lysosomes	
Proliferin	Interaction induces endothelial cell migration and angiogenesis	[305, 306]
CD26	Internalisation of CD26 and role in T cell activation	[307]
Granzyme B	Internalisation and induction of apoptosis	[308]
HSV-glyD	Facilitation of viral entry into cells and transmission between cells	[309]
	Non-M6P bearing ligands	
IGF-II	Internalisation and degradation in lysosomes	[310, 311, 312]
RA	Growth inhibition and/or induction of apoptosis	[313]
uPAR	Binds to IGF2R to convert plasminogen into plasmin	[314, 315]
Plasminogen	Activated by uPAR to form plasmin and consequently activate the IGF2R bound TGF-β	[315]

5.1.3 Scope of this chapter

In this chapter I will describe the interaction that I identified between IGF2R and GABBR2 with the genome-scale screening approach developed in chapter 4 using an avid GABBR2 (ectodomain of GABBR2 conjugated to streptavidin-PE) as a screening probe. The interaction was subsequently validated using targeted gene knockout and biochemical approaches.

5.2 Results

5.2.1 IGF2R is required for the binding of GABBR2 to HEK-293-E cells

One of the proteins from the screening pipeline that was identified as binding HEK-293-E cells in a *SLC35B2* independent manner was the ectodomain of gamma-aminobutyric acid (GABA) type B receptor subunit 2 (GABBR2) (figure 5.3A). To identify the cellular components contributing to this interaction, a genome-scale screening approach was carried out and mutant cells that had lost the ability to bind the GABBR2 ectodomain were sorted (figure 5.3B). Enrichment analysis revealed a clear enrichment of gRNAs targeting *IGF2R* together with genes involved in endosomal function and trafficking in the sorted population (figure 5.3C). IGF2R is a known cargo receptor which transports mannose-6-phosphate (M6P)-modified proteins between the TGN, endosomes,

pre-lysosomal compartments, and the plasma membrane; it might therefore provide a mechanism for the known internalisation and lysosomal degradation of GABAB receptors through interactions with the GABBR2 subunit.

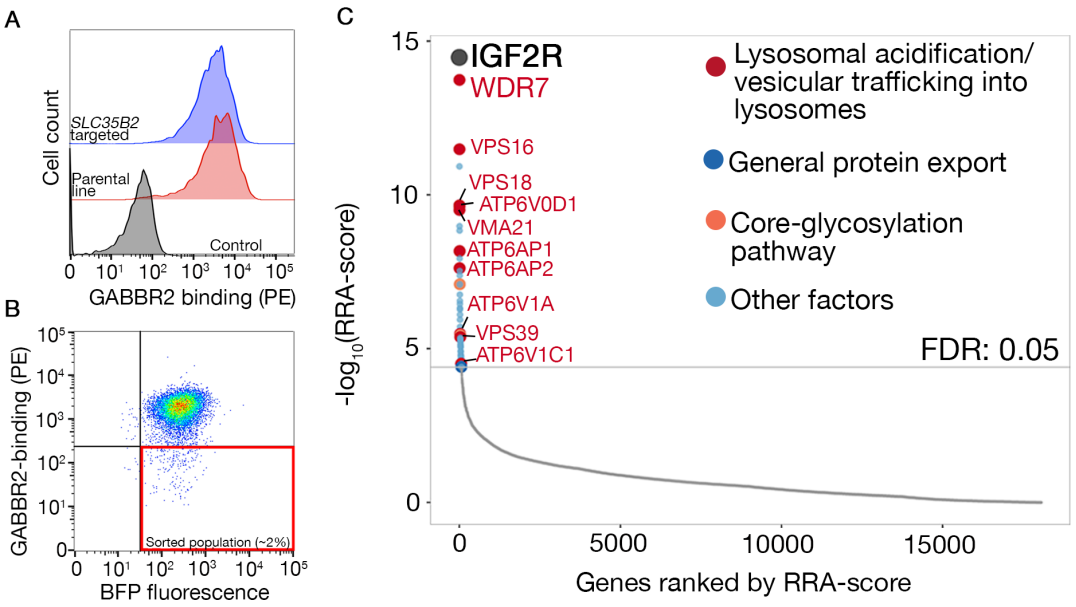


Fig. 5.3 A genome-scale screen using the ectodomain of GABBR2 identifies IGF2R and genes involved in endosomal function and trafficking. A. Monomeric biotinylated GABBR2 ectodomain conjugated to streptavidin-PE binds to HEK-293-E cells in a *SLC35B2*-independent manner. **B.** Cells lacking the surface staining from GABBR2 ecotodomain but expressing BFP fluorescence (from lentiviral transduction) were collected. Approximately 800,000 cells were collected during this sort. **C.** RRA-score rank-ordered genes identified from gRNA enrichment analysis from sorted mutant cells that had lost GABBR2 binding activity. Enriched genes encoded the IGF2R receptor and proteins involved in endo/lysosomal biology.

To investigate this further, I first validated the screen results using individual gRNAs targeting *IGF2R*, which resulted in the loss of binding of the GABBR2 ectodomain (figure 5.4A). IGF2R expression on the surface of cells is known to be dependent on lysosomal acidification, as cells treated with compounds that increase lysosomal pH cause IGF2R to accumulate in endosomes, with a consequent loss from the cell surface. This provides an explanation for why genes known to be required for endosomal function (such as the components or associated factors of the vacoular-type ATPases that pump protons into the acidic compartments of the cells: *ATP6VOD1, ATP6AP1, ATP6AP2, ATP6V1A1, ATP6V1C1, VMA21* [316], and genes encoding for endosomal trafficking proteins: *VPS16, VPS18 and VPS39* [317]) were also enriched [318]. One of the genes (*WDR7*) identified in this screen encodes for a poorly-

characterised protein implicated to be involved in lysosomal acidification [319]. Consistent with the suggested role, *WDR7* targeted HEK-293-E cells also demonstrated a decrease in the surface level expression of IGF2R (figure 5.4B) and a corresponding partial loss in GABBR2 ectodomain binding (figure 5.4A) .

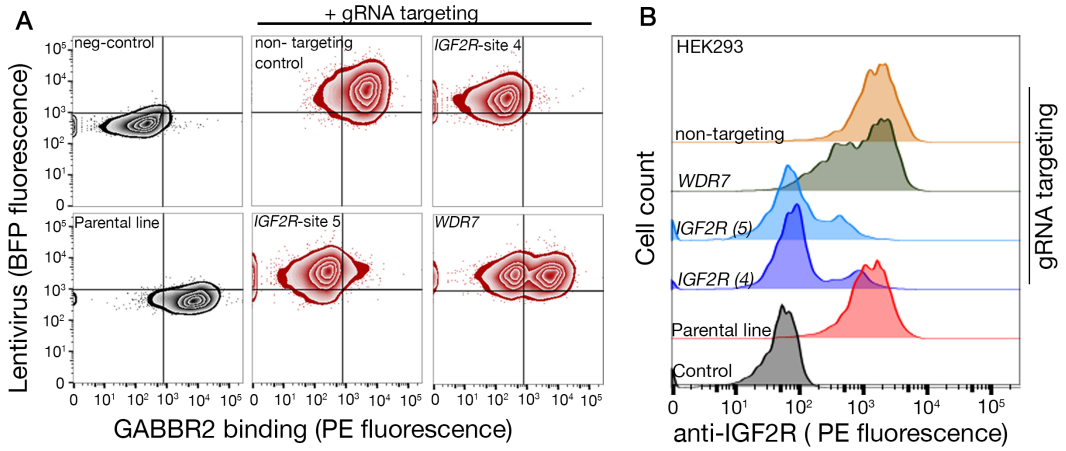


Fig. 5.4 Targeting *IGF2R* and *WDR7* on HEK-293-E cells leads to the loss of binding of GABBR2 ectodomain. A. Binding of GABBR2 was quantified on HEK-293-E cells transduced with two gRNAs targeting different exons of *IGF2R* and one gRNA targeting *WDR7*. A near complete loss of binding was observed on *IGF2R*-targeted cells and a partial loss on *WDR7*-targeted cells; targeted cells were maintained as polyclonal lines. **B.** Cells transduced with lentiviruses encoding individual gRNAs targeting *IGF2R* and *WDR7* show reductions in cell surface IGF2R levels. Cells were transduced with lentiviruses encoding the indicated gRNAs and stained 12 days later with a mouse anti-human IGF2R mAb. Cells stained with secondary antibody alone were the negative control and WT indicates untransduced cells as a positive control. In both cases, a representative of three technical replicates is shown.

5.2.2 Plasma membrane expression of IGF2R is required for the binding of GABBR2 ectodomain

I next set out to explore whether IGF2R was involved directly in the binding of the GABBR2 ectodomain, or whether, it was playing an indirect protein trafficking function in the cell, thereby assisting some other protein to the surface of the cells to mediate the binding. To test whether the ectodomain of IGF2R itself was required for the interaction with GABBR2, a IGF2R expression plasmid was constructed in which the cytoplasmic region of the protein, which contains the important amino acid sorting signals, was replaced with an eGFP reporter protein and then displayed at the surface of human cells by transiently transfecting the NCI-SNU-1 cell line. NCI-SNU-1 cells were chosen as they

exhibited very low levels of plasma membrane IGF2R compared to the HEK-293-E cells and did not bind the GABBR2 ectodomain (figure 5.5A). The NCI-SNU-1 cells transfected with the IGF2R-eGFP fusion protein but not the control-eGFP fusion protein showed surface expression of IGF2R as indicated by the presence of double-positive (GFP^+/PE^+) population when tested with an anti-IGF2R mAb (figure 5.5B, left panel). Correspondingly, the binding of the GABBR2 ectodomain could be conferred to the NCI-SNU-1 cells that expressed the IGF2R-eGFP fusion protein but not the control-eGFP fusion protein on the surface of cells (figure 5.5B). These data suggest that GABBR2 binding to cells was mediated by the ectodomain of IGF2R on the cell surface.

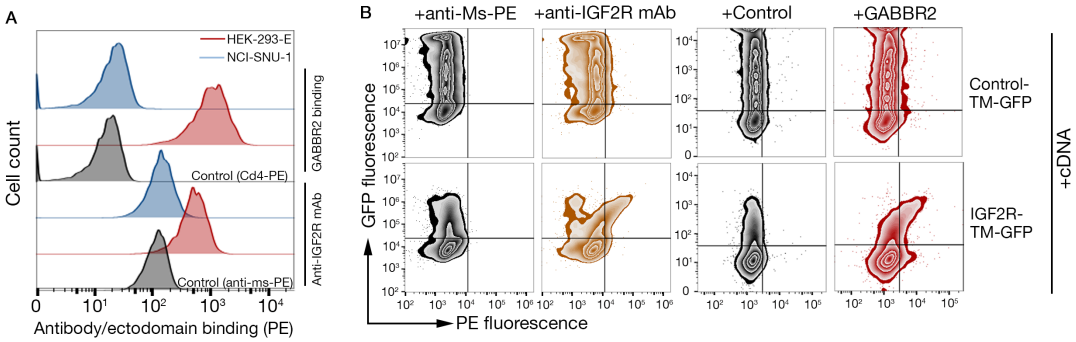


Fig. 5.5 Binding of GABBR2 can be conferred to NCI-SNU-1 cell line that do not display IGF2R on their surface. A. NCI-SNU-1 cell line was tested for IGF2R expression and GABBR2 ectodomain binding using an anti-IGF2R antibody and the avid GABBR2 probe respectively; neither the expression of IGF2R nor the binding to the ectodomain was observed. HEK-293-E cells were used as a positive control. **B.** Gain of IGF2R-eGFP expression (left panel) and GABBR2 binding (right panel) in cells transfected with a cDNA encoding IGF2R ectodomain. NCI-SNU-1 cells were transfected with either a cDNA construct encoding the entire ectodomain of IGF2R fused to a transmembrane (TM) region and an intracellular eGFP or a control TM-GFP tagged receptor and tested for their ability to bind a fluorescently labelled GABBR2 binding probe; only the IGF2R-GFP positive cells and not the GFP negative or control-TM-GFP bound GABBR2. In the case of anti-IGF2R antibody staining, a small fraction of cells that were GFP positive were PE-negative, suggesting that not all expressed IGF2R was displayed at the surface of the cells. In all cases, data is representative of three technical replicates.

5.2.3 IGF2R ectodomain and GABBR2 ectodomain directly interact

To further demonstrate that IGF2R and GABBR2 directly interact, I next expressed the entire ectodomain of IGF2R as a soluble beta-lactamase-tagged 'prey' and tested whether it could be captured specifically by a biotinylated GABBR2 ectodomain 'bait' in a plate-based ELISA-style assay (schematic

depicted in figure 5.6A). In the assay, biotinylated bait proteins are first immobilised on a streptavidin coated mictotitre plate and β-lactamase tagged prey proteins are added to the wells. Following wash steps, if an interaction with the bait has occurred, the prey proteins will be captured in the corresponding wells and this can be detected by addition of a β-lactamase substrate 'nitrocefin' the hydrolysis products of which absorb at 485 nm. Using this assay, I observed a specific capture of the IGF2R ectodomain by the GABBR2 ectodomain (figure 5.6B), demonstrating that the interaction between IGF2R and GABBR2 is direct.

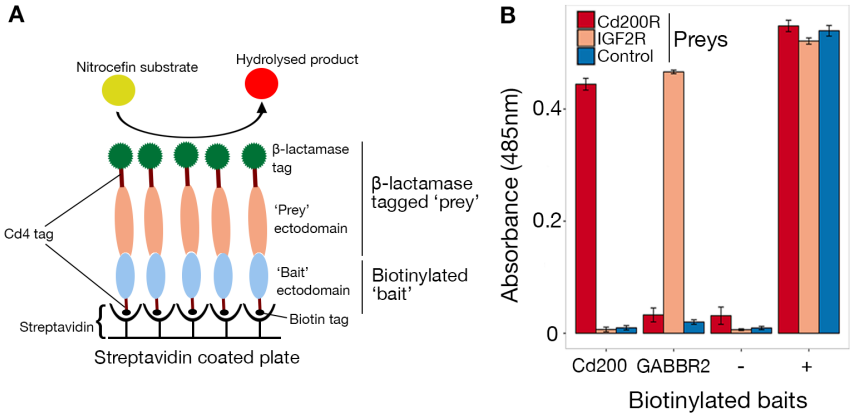


Fig. 5.6 Interaction between IGF2R and GABBR2 is direct. A. Schematic of plate-based ELISA-style assay to test direct binding between ectodomains. The biotinylated ectodomains are captured on streptavidin-coated plates and tested for direct binding using a beta-lactamase-tagged 'prey' ectodomain. **B.** In the plate based assay, the GABBR2 ectodomain was used as a bait and IGF2R ectodomain was used as a prey. Positive control was the Cd200-Cd200R interaction; control 'prey' is an unrelated ectodomain; positive (+) represents total capture of all preys with an anti-prey antibody and negative (-) represents a tag only bait control. Data points are mean $\pm$ sem, n=3.

5.2.4 GABBR2 interacts with IGF2R in a M6P-dependent manner

Given the known function of IGF2R to interact with multiple proteins in a M6P-dependent manner, I next investigated whether the interaction that was identified here also depended on the presence of M6P residues on GABBR2. In a plate-based assay, the interaction between the ectodomains of GABBR2 and IGF2R could be prevented completely by the preincubation of IGF2R with soluble M6P but not with a related monosaccharide, mannose (figure 5.7A). This suggested that the GABBR2 interaction with IGF2R was mediated by the M6P binding domains of IGF2R. Furthermore, GABBR2 binding to IGF2R was also dependent on the N-linked glycans as the interaction was abolished by treating the GABBR2 ectodomain with PNGaseF (an enzyme that

specifically removes N-glycans in a protein) (figure 5.7B). Together these data demonstrate that the interaction between IGF2R and GABBR2 is dependent on the M6P-modified N-linked glycans of GABBR2.

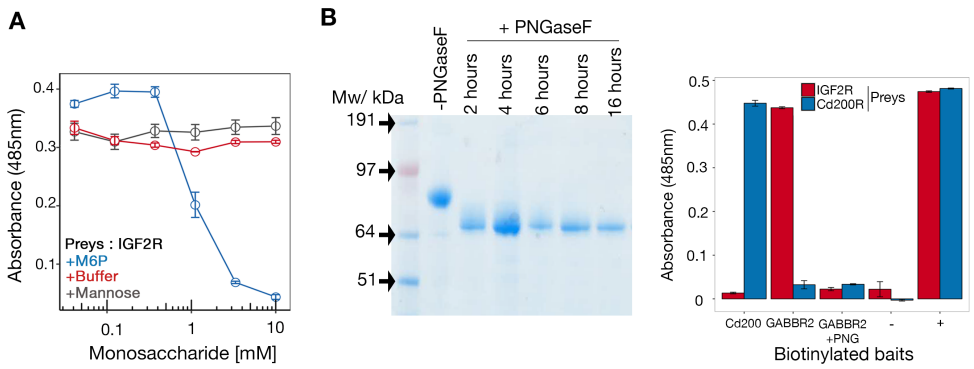


Fig. 5.7 Interaction between IGF2R and GABBR2 is dependent on the M6P-modified N-linked glycans of GABBR2. A. The interaction between IGF2R and GABBR2 can be completely inhibited by soluble M6P. The binding dependency on M6P was established by tested by adding serial dilutions of mannose, M6P, or buffer alone. **B.** Treating GABBR2 with PNGaseF ablates the interaction with IGF2R. Left panel: The purified ectodomain of GABBR2-Cd4-bio was treated with PNGaseF for the indicated times at 37 °C before aliquots were taken, resolved by SDS-PAGE under reducing conditions and stained with Coomassie Blue. No further reductions in molecular mass due to PNGaseF treatment occurred after eight hours incubation, suggesting that the vast majority of glycans had been removed from the protein. Right panel: The indicated biotinylated bait proteins were immobilised on a streptavidin-coated microtitre plate and probed for interactions with the beta-lactamase-tagged prey proteins. Prey binding was quantified by measuring the absorbance of the hydrolysis products of nitrocefin — a β-lactamase substrate — at 485 nm. The known interaction between Cd200-Cd200R was used as the positive control in the assay. Negative control (-) was prey capture by a biotinylated Cd4 tag only control and positive control (+) was total capture of all preys with an anti-prey antibody. Bars represent means $\pm$ sem, n=3. In **A** and **B**, data is representative of three technical replicates.

These data provide evidence that IGF2R directly interacts with the ectodomain of the GABBR2 receptor subunit via the M6P residues present on GABBR2; they may therefore provide mechanistic insights for the known constitutive internalisation of GABAB receptors.

5.3 Discussion

In this chapter, I have demonstrated the utility of the genome-wide screening approach developed in Chapter 4 by identifying IGF2R as a binding partner for GABBR2 receptor. GABAB receptors are expressed abundantly in almost all types of neurons and glia throughout the central nervous system and mediate slow-acting control of neuron excitability by inhibiting neurotransmitter release. This expression pattern overlaps with that of IGF2R, which is also widely distributed throughout the nervous system with particular enrichment in cortical areas, hippocampus and cerebellum [320]. The regulation of the surface level of GABAB receptors by endocytosis is an important mechanism to attenuate signal strength and can be modelled in HEK293 cells, where GABAB receptors have been shown to rapidly and constitutively internalise by the clathrin-dependent pathway to endosomes [289, 292]. The finding that IGF2R can interact directly with the GABBR2 subunit of the GABAB receptor complex provides a mechanism for the internalisation because IGF2R is itself constitutively endocytosed and trafficked to the endosomal compartments through clathrin-mediated uptake via 'YSKV' motifs in its cytoplasmic region [321]. This is also consistent with the regulatory role for the GABBR2 subunit in the uptake of the GABAB receptor complex, and the fact that antibodies directed against the extracellular region, but not receptor agonists, can inhibit GABAB receptor endocytosis [289]. A suggested physiological role of this interaction is depicted in figure 5.8.

A similar role for IGF2R in interacting with M6P-bearing ligands to down-regulate receptors from the cell surface has been shown for CD26 in activated T-cells [307]. CD26 expressed on the surface of activated T cells (activated with phytohemagglutinin), but not expressed on resting T cells was found to be mannose-6-phosphorylated. One of the key questions, which has not been addressed in this work, is the possible physiological mechanism by which GABBR2 would be mannose-6-phosphorylated. To address this, it will be important to first establish whether GABBR2 is constitutively mannose-6-phosphorylated or whether it is a regulated process. One way of addressing this experimentally would be to use the ectodomain of IGF2R to 'pull-down' endogenous GABBR2 from mouse (or rat) brain lysates. If GABBR2 exists in a constitutively mannose-6-phosphorlated state, it should, in principle, be possible to detect the interaction of IGF2R with endogenous GABBR2 using such an approach.

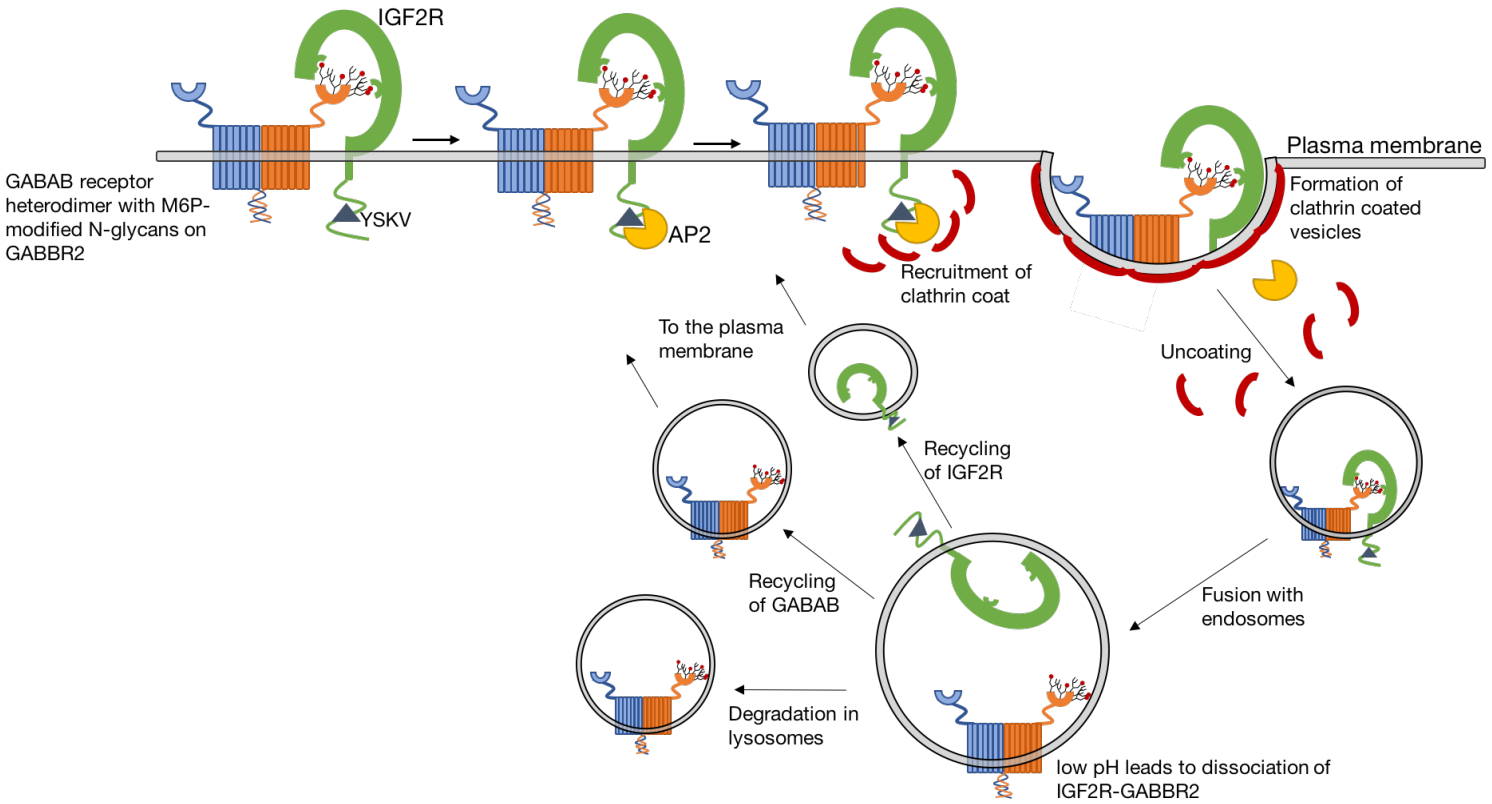


Fig. 5.8 Proposed mechanism of IGF2R mediated GABAB receptor internalisation. In this model, cell surface IGF2R associates with M6P-modified N-glycans of GABBR2 subunit on the plasma membrane. The 'YKSV' motif on IGF2R then recruits adaptor protein complex 2 (AP2 complex), which is a plasma membrane-localised clathrin adaptor composed of α, β2, μ2, and σ2 adaptin subunits [321]. This is consistent with the reported co-localisation of GABAB receptors with AP2 subunits in HEK293 cells [289, 322]. This is followed by clustering of IGF2R and its cargo (GABAB in this case) into clathrin-coated vesicles (CCVs). Following internalisation and uncoating of the clathrin coat proteins, the vesicles are fused with endosomes. The low pH of endosomes induces dissociation of cargo from IGF2R, thereby releasing GABAB receptors into endosomal lumen. IGF2R is recycled back to TGN or to the cell surface, whereas, GABAB receptors are either recycled back to the cell surface membrane or trafficked into lysosomes for degradation. This model is consistent with the observations in HEK293 cells where constitutively internalised GABAB receptors have been shown to be targeted to endosomes from where they are either recycled back to the cell surface or degraded in the lysosomes [289].

The genome-scale approach here has once again demonstrated how this method can be utilised to identify not only the receptor that directly interacts with the recombinant protein probe but also provide valuable information regarding the receptor biology. In the example here, I was able to identify cellular components involved into intracellular vesicle acidification and transport into endo/lysosomal compartments consistent with the known biology of the IGF2R receptor. In addition, I was also able to identify *WDR7* as an important factor mediating the surface expression of IGF2R. Recently, it has been shown that cells with RNAi mediated knock-down of WDR7 are unable to re-acidify intracellular vesicles during recovery from compounds that reversibly block acidification (bafilomycin A1), suggesting a role for regulation of vesicuar acidification for WDR7 [319]. The results here are consistent with this observation and further reveal the role of WDR7 in surface expression of mannose-6-phosphate receptors.

CHAPTER 6

DISCUSSION AND FUTURE DIRECTIONS

In this work, I have developed a genome-scale screening strategy using the CRISPR-Cas9 knockout system to identify cellular components involved in cellular recognition. The initial screens I carried out with mAbs as screening probes provided a means to define the methodological parameters for the genome-scale KO screening approach, so that the approach could be used not only to identify the directly interacting receptors on the surface of the cell, but also to reveal valuable information regarding the cell biology of those specific receptors (such as essential roles for chaperones, enzymes required for specific posttranslational modifications, and transcription factors). I then adapted the screening system in order that the approach could also identify both low- and high-affinity receptors of solubilised ectodomains of cell surface proteins. Finally, I demonstrated the utility of the screening approach by identifying IGF2R as a binding partner for GABAB receptors, providing a mechanism for the internalisation and regulation of GABAB receptor signalling.

In the recent years, a number of genetic screens using the CRISPR-Cas9 system have been described to investigate complex biological questions, topics of which ranging from gene essentiality, drug and toxin resistance, the hypoxia response, and host factors required for invasion of viruses and bacteria (refer to table 1 of [174]). In the context of cellular recognition, CRISPR-Cas9 based genetic screens have already demonstrated their utility by the identification of receptors and non-receptor host factors that are required for the entry or survival of bacteria and viruses [181, 323, 324, 220, 183, 182]; host factors required for resistance from pathogen toxins [173, 179]; and molecular mechanisms that control expression of secretory and membrane proteins [186, 187] (see table 6.1). The work here widens the applicability of a KO screening system by utilising recombinant proteins that represent a broad range of cellular contexts (both host and pathogen) to identify novel receptor-ligand interactions and additional intracellular genes required for such associations.

Table 6.1 Summary of genome-scale KO screens using the CRISPR-Cas9 approach to study cellular recognition events

Cell line	Phenotype/screen type	Genes screened	Primary finding(s)	Ref
		Virus-host interactions		
RAW264.7 cells	Host factors required for Murine norovirus (MNV) (lethality screen)	87,987 sgRNAs targeting 19,150 mouse genes	Identification of CD300lf (also known as CLM2) as the host receptor.	[324]
Huh7.5.1 cells	Host factors required by DENV serotype 2 and hepatitis C virus (HCV) (lethality screen)	122,417 sgRNA targeting 19,052 human genes	Host OST complex is hijacked by the DENV virus to mediate viral RNA replication. HCV screen revealed genes encoding for previously known host receptors (*CD81, OCLN, CLDN1*) and a novel role of genes involved in FAD biogenesis (*RFK* and *FLAD1*) for intracellular virus replication.	[220]
293T cells	Host factors required by West nile virus (WNV) (lethality screen)	122,411 sgRNAs targeting 19,050 human genes	Initial identification of 12 genes involved in endoplasmic reticulum-associated functions (carbohydrate modification, protein translocation and signal peptide processing, protein degradation, and heat shock response). Further validation in other flaviviruses identified a requirement of host signal peptide processing protein SPCS1 for flavivirus protein processing and infection.	[183]
CD4+ T cell line (GXRCas9)	Host factors required by CCR5-tropic HIV-1 strain (lethality screen)	87,536 sgRNAs targeting 18,543 human genes (+1,504 non-targeting control sgRNAs)	Identification of genes encoding for known receptors (*CD4* and *CCRC5*) and three novel host dependency factors (*TPST2*, *SLC35B2* and *ALCAM*). Tyrosine sulfation of CCR5 by SLC35B2 and TPST2 is crucial for the cell surface expression of this HIV co-receptor. ALCAM mediated cell aggregation and its loss confers strong protection against cell-to-cell HIV transmission.	[182]

Table 6.1 – *Continued on next page*

		Toxin resistance screens		
*M*ouse JM8 ESCs	*Clostridium septicum* alpha-toxin resistance (lethality screen)	87,897 sgRNAs targeting 19,150 mouse genes	Increased resistance to the toxin by inactivation of *B4galt7* and textitExt2 (enzymes required for Heparan sulphate (HS) biosynthesis.	[173]
U937 cell line	*Staphylococcus aureus* alpha hemolysin toxin (lethality screen)	120,000 sgRNAs targeting 19,050 human genes and 1864 miRNAs	Identification of previously known receptor ADAM10 and three novel components (SYS1, ARFRP1 and TSPAN14) that regulate the expression of ADAM10 on the cell surface.	[179]
		Bacteria-host interactions		
HT-29 cell line	Cytotoxicity of *Vibrio parahaemolyticus* Type III secretion system (lethality screen)	74,700 sgRNAs targeting 18,675 human genes	Removal of host cell sulfation reduces bacterial adhesion to cells and delays T3SS1-associated cytotoxicity; fucosylation of surface glycans is required for T3SS2 mediated killing.	[181]
		Cellular factors contributing to protein expression		
BMDCs from Cas9 expressing mouse	Processes regulating induction of tumor necrosis factor (Tnf) by bacterial lipopolysaccharide (LPS) (FACS-based screen)	125,793 sgRNAs targeting 21,786 genes + miRNAs + 1,000 non-targeting sgRNA	Identification of the components of the OST complex, Paf complex and other factors (e.g., *Tti2, Ruvbl2, Tmem258, Midn, Ddx39b, Stat5b and Pdcd10*) as important regulators of Tnf expression.	[186]
Pancreatic cell line BxPC-3	Constitutive and induced cell surface PD-L1 expression (FACS-based screen)	10-sgRNAs/gene targeting 20,500 human genes.	Identification of CMTM6 as the master regulator of PD-L1	[187]

ALCAM:CD166 antigen, ARFRP1: ADP Ribosylation Factor Related Protein 1, BMDC: Bone marrow derived dendritic cells, PD-L1: Programmed death-ligand 1, CMTM6: CKLF-like MARVEL transmembrane domain containing protein 6, CLDN1: claudin 1, FLAD1: FAD synthase; OCLN: occludin, OST: oligosaccharyltransferase, SLC35B2: adenosine 3′-phospho 5′-phosphosulfate transporter 1; SPCS: signal peptidase complex subunit, SYS1: Golgi trafficking protein, TPST2: protein tyrosine sulfotransferase, T3SS: Type III secretion system, TSPAN14: Tetraspanin 14

For the application of a genome-scale screening approach to identify receptors and receptor-specific pathways using soluble probes, the initial stages of this work focused on testing a range of parameters with respect to sorting thresholds, size of the gRNA mutant library, and the day of phenotypic selections, which facilitated a better understanding of how the results from KO screens could be influenced by the experimental design. Below, I highlight some of the considerations for setting the screening parameters for genome-scale CRISPR-Cas9 knockout screens.

6.1 Overview for screening parameters

6.1.1 Sorting strategy

All phenotypic selections in this work were carried out using a flow-cytometry-based sorting approach. FACS-based screens are generally thought to have an advantage over lethality screens for the identification of genes that have intermediate phenotypes, as the quantitative nature of flow-cytometry allows for the selection of cells with mutations that result in a partial as well as a complete phenotype [325]. That said, the sorting threshold used during phenotypic selections can influence the genes that are identified in this approach. When determining the FACS sorting cutoff, it is important not only to consider the stringency at which cells that show 'true' phenotypic changes are captured, but also to collect sufficient cells from the 'non-binding' population such that a wide range of gene perturbations that cause both weak and strong phenotypic effects can be identified. The use of highly stringent sorting gates (e.g. 0.1% of total population) can lead to a notable enrichment of very few genes with strong effect sizes in the sorted population (as seen on the screen using an anti-BSG mAb), but risks having insufficient representation of gRNAs to detect genes with low effect sizes.

The use of a CRISPR-Cas9 based screening approach for the identification of genes that have both strong and weak effects in regulating the expression of a protein has previously been demonstrated in the context of expression of tumor necrosis factor (Tnf) in mouse-derived primary dendritic cells upon lipopolysaccharide (LPS) stimulation [186]. A FACS-based sorting strategy (5% sorting threshold) was designed in the study in which cells within the mutant population that were either refractory to Tnf expression or that induced Tnf more strongly, were differentially collected and assessed for the cellular factors that mediated the regulatory response. A large number of genes were identified

in the screen as important for the regulation of Tnf expression upon LPS stimulation; the highest ranking 176 (112 positive and 64 negative) candidate regulators were chosen for targeted gene KO validation. Of the tested genes, 57/112 positive but only 4/64 negative regulators were correctly validated. To reduce the high number of false-positive genes, the authors opted for a secondary validation with a focused library, in which up to 10 gRNAs/gene were designed for the top ranked 2,569 genes and a secondary pooled screen was carried out, which was shown to have improved the specificity and sensitivity of a pooled screen. This study demonstrated the way in which FACS-based CRISPR-KO screens can be used to carry out comprehensive dissection of genetic pathways contributing to protein expression in a defined cellular context (for example, LPS induction). However, the screening approach also highlighted the importance of secondary validation steps that are usually required for high-confidence identification of regulatory genes when permissive sorting thresholds are used. In the work described here, I sought to establish the sorting parameter that would allow for a balance between identification of genes that have weak effects and genes that could be identified with high confidence without necessarily performing a secondary pooled screen. Based on the data obtained from the antibody screens in this study, I determined that collecting at least 300,000 to 500,000 cells at a 0.5-1% stringency threshold from a high complexity library (500-1000$\times$ per gRNA) is generally appropriate for the high-confidence identification of the directly interacting receptor and additional genes related to the biology of the receptor from a single experiment.

In the screens carried out in this work, the phenotype selections were carried out with a single sort. In the mAb screens described in chapter 3, I observed that sorting the selected population just once rather than multiple times is sufficient to identify both the epitope target and the cell pathways responsible for surface expression. A more stringent approach to sorting proceeds through iterative selections in which the mutant cells displaying the phenotype of interest are enriched through multiple rounds of sorting. Such approach have also been applied successfully in multiple loss-of-function genome-scale screens [187, 235, 145]. Multiple rounds of selections are usually desirable either when the signal-to-noise ratio of the desired phenotype is low or when the aim of the screen is to identify mutants that have strong phenotypes. When using an iterative selection approach for FACS-based screens, it is important to consider that often the sorting process can cause cell death (mainly caused by sheer force of the sorter); thus, not all collected

cells will be represented in the next round of sorting. Again, this might not greatly affect identification of genes with strong effect sizes, but this type of highly stringent sorting approach might not be ideal when the screen is to be used to investigate genetic pathways, where the representation of mutant cells displaying weaker phenotype is low initially.

6.1.2 The timing of phenotypic selections

The timing of phenotypic selections can influence the genes that are identified in genome-scale KO screens. As the number of days post transduction with the lentivirus (for the generation of the KO library) increases, the overall representation of the gRNAs targeting genes required for general cellular proliferation is likely to decrease from the total mutant population. The longer the mutant library is kept in culture, the more difficult it can be to investigate the role of essential genes. In the screens carried out using mAbs, I specifically observed the influence of screen timings in the identification of genes of the SRP-dependent protein export pathway. The majority of proteins destined for the plasma membrane are initially targeted to the endoplasmic reticulum by the SRP-dependent protein translocation machinery [204]. Thus, genes relating to this pathway are expected to be identified in screens designed to study cellular recognition events. However, a number of genes in this pathway are also known to be core-essential, which means that the likelihood of cells lacking those genes being non-viable increases, the longer the mutant library is cultured. This often led to the reduced representation of general protein export pathway when selections were performed at late time points (day 15-16) compared to early time points (day 9). This can be taken into consideration while designing similar screens in the future; if the effect of genes required for proliferation and viability is to be investigated in the context of cellular recognition process, carrying out screens at an early time-point (day 9 post-transduction) would be generally appropriate. On the other hand, if the approach is to be used to identify few targets with strong size effects rather than general cellular pathways, it might be appropriate to perform screens at a later time point (day 15-16 post transduction).

In loss-of-function screens, when the iterative selection approach is used to enrich for mutant cells displaying the phenotype of interest over an extended period of time, it also leads to the reduction in the representation of gRNA targeting genes indispensable for cell proliferation and viability. An iterative selection approach was recently described to identify host factors important for

HIV infection [182]. In the study, a naturally susceptible T-cell line was serially infected with HIV and the mutant cells refractory to infection were enriched over a course of six weeks. Using this approach, a restrictive set of five host factors (three novel factors; ALCAM, SLC35B2 and TPST2 and two known receptors; CD4 and CCRC5) critical to the survival of HIV in host cells was identified. All the identified factors were found to be dispensable for host cell viability; thus the authors suggested that these factors could be attractive targets for therapeutic intervention. This exemplifies how the sorting approach in terms of stringency and screening end points can influence the genes that are identified from a KO screen.

6.2 Potential of genome-scale KO screens using mAbs for the study of receptor biology

The genome-scale KO screens carried out in this work using mAbs as screening probes demonstrated how such screens can be used to study the biology of cell surface receptors. Similar use of mAbs in CRISPR-Cas9 KO screens to investigate the cellular factors required for cell surface expression of membrane proteins has been very recently applied by others for the identification of a CKLF-like MARVEL transmembrane domain containing protein 6 (CMTM6) as a critical regulator of programmed death-1 (PD-1) ligand 1 (PD-L1) in a broad range of cancer cells [187]. In the study, a KO screening approach using an anti-PD-L1 mAb as a screening probe was carried out, and cells refractory to antibody staining of PD-L1 were found to be enriched in gRNAs targeting *CMTM6* in addition to *PD-L1* itself. Further studies revealed the association of CMTM6 with PD-L1 at the plasma membrane and in recycling endosomes, where CMTM6 was found to protect PD-L1 from being targeted for lysosomal degradation. This study further exemplifies how a genome-scale approach using mAbs can be a valuable means to investigate the biology of cell surface receptors.

In this work, the genome-scale KO screens carried out using mAbs in some cases revealed potentially interesting novel factors (e.g., *SPPL3* and *WDR48* in the screen with anti-BSG mAb; *TNNT3* in the screen with anti-GYPA mAb) that have not been previously reported to be associated with the expression of the corresponding receptors. The precise roles of these factors were not investigated in this work, but before further research is carried out, it is important to first validate these genes using targeted gene KO approaches.

Based on the known functions of these factors (e.g. SPPL3 as a regulator of N-glycosylation [326]; WDR48 as a regulator of deubquitylating complex [327, 328]; and TNNT3 as a factor known to bind tropomyosin, which is a member of the erythrocytic membrane skeleton [225]), further studies can be designed to investigate their roles in membrane expression (or antibody epitope presentation) of the receptors.

6.2.1 Potential for the study of receptor biology in a high-throughput manner

The screening strategy described here using mAbs as screening probes can be adapted to be carried out in a high-throughput manner. From a single lentiviral transduction of 80-100 million cells, I was usually able to generate mutant libraries of sufficient size to carry out screens using 8-10 different probes. This, combined with the ability to multiplex up to 15-20 samples in a single sequencing (Hi-seq 2500 platform) run, will allow for up-scaling the throughput of this screening platform. This can be further facilitated by the Cancer Cell Line project and the cGAP facility at the Sanger Institute, which has generated approximately 400 Cas9-expressing cancer cell lines. Most of these cell lines are genomically well-characterised, which can facilitate the selection of a panel of transcriptionally diverse cell lines to conduct genetic screens to study the biology of a wide range of receptors. Using these resources, a systematic approach can be devised, in which cell lines that express the receptor of interest can be identified using transcriptomic analysis and screens can be carried out using mAbs to identify cellular factors that are important for the expression of the cell surface receptors. The approach could be carried out for cell surface receptors, which are currently targets of antibody therapy, or small molecule inhibitors in order to understand how the expression of the therapeutically important receptors are regulated at the surface of cells.

6.3 Assessment of the approach to identify receptors of soluble protein ectodomains

In Chapter 4, I demonstrated the utility of a genome-scale CRISPR-Cas9 screening approach for the identification of directly interacting receptors on the surface of cells for a panel of recombinant protein probes. I also found that a general factor involved in several cell surface recognition events was the role played by glycosaminoglycans (GAGs) (specifically HS), which form

a major part of the cellular glycocalyx. An important discovery in this regard was the observation that many recombinant proteins adsorb into HS without necessarily binding to a specific receptor. Additionally, in the context of the cellular binding assay described here, HS were also found to contribute to binding in an additive rather than the specific receptor-dependent manner for proteins such as RH5, and the individual contribution made by either the specific receptor or HS could be dissected at a molecular level. When the fractional contribution from adsorption into HS (as determined by comparing the binding to the parental line with the binding on *SLC35B2*-KO version of the line) was low (<20%), in every case I was able to identify the corresponding interaction partner using this approach. This led to the development of a two-step approach, in which every observed binding event on a parental line was first tested to be mostly retained on a *SLC35B2*-KO version of the line before proceeding to the genome-scale screening step. I believe that this will provide extremely useful guidance to others using this approach in the future.

To further develop this approach, I have now generated six different cell lines (HEK-293-E, NCI-SNU-1, HL-60, KBM7, HepG2 and HEL) that lack the *SLC35B2* gene. These cell lines can be used for the pre-screening step to determine ligands suitable for this approach to identify novel receptors. Currently, this approach is actively being used in the laboratory to screen a panel of approximately 70 recombinant proteins, representing diverse ligands from both human and parasite surface proteins (e.g., megakaryocyte proteins, immune regulatory proteins, sporozoite-stage proteins from *P. falciparum*, surface proteins of *Leishmania donovani*). These ligands will be tested for binding on the *SLC35B2*-KO cell lines and their corresponding parental versions; the candidates that retain their binding on the *SLC35B2* will be used as screening probes to carry out genetic screens to identify receptors.

The adaptation of this technique to identify cell surface receptors for recombinant protein ligands in a high-throughput manner is limited by the number of recombinant proteins that can be used as screening ligands. The limiting step in the method is the pre-screening criterion in which proteins are only chosen if they do not largely depend on HS for binding to the cell lines. Currently, it takes up to two months to recombinantly produce 100 proteins and screen them on the available cell lines (six cell lines and the corresponding versions that lack *SLC35B2* are available). However, binding to a given parental line is not a common event in the first place, and the observation up to now has been that it is more common to observe bindings dependent on *SLC35B2* than

without. Parasite proteins also seem to have a higher propensity of binding to cells in a HS-dependent manner than without. In this work, I identified multiple proteins of the *P. falciparum* merozoite that adsorb into HS. Similarly, recent pre-screens carried out in the lab by Zheng Shan Chong and Amalie Couch using 20 proteins of *Leishmania donovani* on six lines showed that seven bound to at least one parental line but all seven out of seven bindings were completely dependent on HS (as determined by testing binding on *SLC35B2*-KO line and with soluble heparin blocking experiments). Of the approximately 200 proteins that have been screened, only 35 have shown binding to at least one parental line and of these only 10 proteins have retained their binding on the *SLC35B2*-KO version of the parental line. Of the 10 proteins that have shown retention of binding, I have carried out genetic screens on five proteins (CD226, EPHB1, TNFRSF9, LPHN1 and GABBR2); in every case, I was able to identify the directly interacting receptor. Thus, this method is not necessarily a high-throughput method but with the controls in place, it has a very high success rate. In addition, because it does not require any *a priori* assumptions to be made regarding the biochemical properties of the cell surface receptor, and additionally identifies genetic pathways important for the cell biology of membrane-associated proteins (for example, the function of endosomal acidification in the transport of IGF2R and the role of p53 in expression of TNFSF9), it has an advantage over biochemical or cDNA-based gain-of-function approaches to study protein-protein interactions.

One potential way to improve this approach would be to screen the recombinant proteins in the 'biologically relevant' cell lines. In this work I mainly utilised HEK-293-E cells, because they are easy to transduce with lentiviruses and can be grown in suspension culture, thus avoiding the need for enzymatic or mechanical dissociation, which can alter the receptors on the surface of cells. This makes HEK-293-E cells, and easily transducible suspension cell lines in general, well-suited for genetic screens designed to study cellular recognition processes. In addition, the initial study into the interaction between RH5 and BSG suggested that physiologically relevant interactions, which occur at the red-blood cell surface, can still be identified in this cell line. However, a better approach would perhaps be to match the proteins to the cell line of related biology. For example, the proteins of the macrophage-invading *Leishmania donovani* parasite could be tested on macrophage-related cell lines such as U937 and THP1, and the proteins of hepatocyte-invading sporozoites could be tested on hepatocyte derived lines such as HepG2 and Huh7. The high activity

Cas9-expressing cell lines and the corresponding *SLC35B2*-KO versions that I have now generated represent cell lines from different origins, but this can be further expanded, depending on the proteins that are to be tested. Testing binding on biologically relevant cell lines in which there is possibly a higher chance of a specific receptor being present could also decrease the number of proteins that are identified to depend largely on HS for binding.

6.4 Interaction between IGF2R and GABBR2

Using the genetic screening approach developed in this work, I was able to identify a novel interaction between the ectodomain of GABBR2 receptor and a known cargo receptor IGF2R. Although the constitutive internalisation of GABAB receptors using clathrin/dynamin dependent pathways has been demonstrated in multiple studies [289, 292, 329, 330], the precise mechanism of internalisation has remained unknown. The GABBR2 subunit itself contains a known structural motif —Yxxϕ, where x can be any amino acid and ϕ is a bulky hydrophobic residue)—that is required for recruitment of adaptor proteins for clathrin-dependent endocytosis [331]; however, the mutation in the critical tyrosine (Y) residue together with the hydrophobic leucine residues has been shown to have no effect on internalisation [332]. In addition, the C-terminus truncation of either the GABBR1a subunit or the GABBR2 subunit has also been shown not to have an effect on the rates and extent of the GABAB heterodimer internalisation in live HEK293 cells. Based on these observations, researchers have suggested that the clathrin recruitment of GABAB receptors is likely to be mediated by other regions of the proteins [332]. The finding in this study that IGF2R, a known cargo receptor with a well-defined internalisation sequence for recruitment of clathrin coats, directly interacts with the ectodomain of GABBR2 potentially provides the missing molecular explanation for the observed GABAB internalisation.

To validate the hypothesised function of IGF2R in internalisation of GABAB receptors, further experiments can be carried out to address the following topics:

- Effect of *IGF2R*-KO on internalisation of GABAB receptors: As HEK293 cells have been used as models in the past to study the internalisation of GABAB receptors, I would conduct the preliminary experiments to test the hypothesised function of IGF2R also in the HEK-293-E cell system. To this end, I have generated an *IGF2R*-KO version of a HEK-293-E cell

line. To validate that GABAB on the surface of cells is internalised by IGF2R, parental HEK-293-E cells and cells lacking *IGF2R* can be transfected with both components of the heretodimers and their constitutive rate of internalisation can be compared. A number of endocytosis assays that utilise antibody labelling of extracellular epitopes on receptors, receptor biotinylation, and fluorescent-tagging techniques have already been applied to study the mechanism of internalisation of GABAB receptors in HEK293 cells [289, 292, 329, 330, 333]. Similar approaches could be designed to assess the effect of *IGF2R*-KO on constitutive internalisation of GABAB receptors.

- Regulation of mannose-6-phosphorylation on GABBR2: What causes GABBR2 to be mannose-6-phosphorylated? Is it a constitutive or a regulated process? I have conducted initial experiments in this regard and prepared lysates from mouse brains. I intend to carry out a pull-down experiment from this lysate using a recombinant biotinylated IGF2R ectodomain that is conjugated to streptavidin-coated paramagnetic beads for increased avidity. If the endogenous GABBR2 is constitutively mannose-6-phosphorylated, it should in principle be possible to observe this interaction using the biochemical pull-down approach. This experiment would rely on the ability of the mouse GABBR2 to interact with human IGF2R, but given that mouse GABBR2 is 98% identical to the human version of the protein, it is likely that the interaction is conserved.

Based on the initial experiments in heterologous cells, further long-term experiments can be designed to address how IGF2R functions in receptor internalisation of neuronal GABAB receptors. Experiments with primary cultured neurons have also shown rapid internalisation of GABAB receptors in the absence of an agonist, which is in line with the findings in this work. GABAB receptor internalisation is understood to be the main mechanism by which signalling through this important class of inhibitory neurological regulators is controlled; future work on this may therefore suggest new ways of neurotransmission regulation, which may be useful in treating a wide range of neurological disorders, including epilepsy and depression.

6.5 Concluding remarks

In this work, I have described a genome-scale CRISPR KO approach to investigate the molecular basis of cell surface receptor biology and recognition events. I have demonstrated how this method can be applied to identify directly interacting receptors for commonly used probes such as monoclonal antibodies and recombinant proteins, in addition to revealing the genetic pathways important for the cell biology of membrane-associated proteins. It is a generally applicable approach that can be used to explore cellular signalling and recognition processes in a wide range of different biological contexts, including between our own cells (e.g. neural and immunological recognition), as well as between host cells and pathogen proteins. Perhaps most importantly, because this technique does not require any prior assumptions to be made regarding the biochemical nature or cell biology of the receptors and provides an opportunity to study interactions mediated by cell surface receptors of unique biology, such as glycans, glycolipids, and phospholipids, it has great potential to make completely unexpected discoveries which would otherwise be very difficult to achieve.

BIBLIOGRAPHY

[1] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, *Overview of Extracellular Signaling*. W. H. Freeman and Company, 2000.

[2] G. J. Weiner, "Building better monoclonal antibody-based therapeutics," *Nat. Rev. Cancer*, vol. 15, pp. 361–370, June 2015.

[3] M. R. Arkin and J. A. Wells, "Small-molecule inhibitors of protein-protein interactions: progressing towards the dream," *Nat. Rev. Drug Discov.*, vol. 3, pp. 301–317, Apr. 2004.

[4] M. Rask-Andersen, M. S. Almén, and H. B. Schiöth, "Trends in the exploitation of novel drug targets," *Nat. Rev. Drug Discov.*, vol. 10, pp. 579–590, 1 Aug. 2011.

[5] B. K. Brandley and R. L. Schnaar, "Cell-surface carbohydrates in cell recognition and response," *J. Leukoc. Biol.*, vol. 40, pp. 97–111, July 1986.

[6] J. Liu and B. Rost, "Comparing function and structure between entire proteomes," *Protein Sci.*, vol. 10, pp. 1970–1979, Oct. 2001.

[7] M. Diehn, R. Bhattacharya, D. Botstein, and P. O. Brown, "Genome-scale identification of membrane-associated human mRNAs," *PLoS Genet.*, vol. 2, p. e11, Jan. 2006.

[8] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *General Principles of Cell Communication*. Garland Science, 2002.

[9] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Ion Channels and the Electrical Properties of Membranes*. Garland Science, 2002.

[10] W. A. Catterall, "Structure and function of Voltage-Gated ion channels," *Annu. Rev. Biochem.*, vol. 64, pp. 493–531, 1 June 1995.

[11] J.-L. Galzi and J.-P. Changeux, "Neurotransmitter-gated ion channels as unconventional allosteric proteins," *Curr. Opin. Struct. Biol.*, vol. 4, pp. 554–565, 1 Jan. 1994.

[12] J. Bockaert and J. P. Pin, "Molecular tinkering of G protein-coupled receptors: an evolutionary success," *EMBO J.*, vol. 18, pp. 1723–1729, 1 Apr. 1999.

[13] A. D. Howard, G. McAllister, S. D. Feighner, Q. Liu, R. P. Nargund, L. H. Van der Ploeg, and A. A. Patchett, "Orphan G-protein-coupled receptors and natural ligand discovery," *Trends Pharmacol. Sci.*, vol. 22, pp. 132–140, Mar. 2001.

[14] D. K. Vassilatis, J. G. Hohmann, H. Zeng, F. Li, J. E. Ranchalis, M. T. Mortrud, A. Brown, S. S. Rodriguez, J. R. Weller, A. C. Wright, J. E. Bergmann, and G. A. Gaitanaris, "The G protein-coupled receptor repertoires of human and mouse," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, pp. 4903–4908, 15 Apr. 2003.

[15] R. Fredriksson and H. B. Schiöth, "The repertoire of g-protein-coupled receptors in fully sequenced genomes," *Mol. Pharmacol.*, vol. 67, pp. 1414–1425, May 2005.

[16] W. M. Oldham and H. E. Hamm, "Heterotrimeric G protein activation by g-protein-coupled receptors," *Nat. Rev. Mol. Cell Biol.*, vol. 9, pp. 60–71, Jan. 2008.

[17] X.-L. Tang, Y. Wang, D.-L. Li, J. Luo, and M.-Y. Liu, "Orphan G protein-coupled receptors (GPCRs): biological functions and potential drug targets," *Acta Pharmacol. Sin.*, vol. 33, pp. 363–371, Mar. 2012.

[18] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Signaling through Enzyme-Linked Cell-Surface Receptors*. Garland Science, 2002.

[19] A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds., *Essentials of glycobiology*. Cold Spring Harbor Laboratory Press, 2009.

[20] H. Ghazarian, B. Idoni, and S. B. Oppenheimer, "A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics," *Acta Histochem.*, vol. 113, pp. 236–247, May 2011.

[21] R. D. Cummings and J. D. Esko, "Principles of glycan recognition," in *Essentials of Glycobiology* (A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds.), Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 20 Mar. 2010.

[22] U. S. von Euler, "On the specific vaso-dilating and plain muscle stimulating substances from accessory genital glands in man and certain animals (prostaglandin and vesiglandin)," *J. Physiol.*, vol. 88, pp. 213–234, 6 Nov. 1936.

[23] W. Feldberg, H. F. Holden, and C. H. Kellaway, "The formation of lysocithin and of a muscle-stimulating substance by snake venoms," *J. Physiol.*, vol. 94, pp. 232–248, 14 Nov. 1938.

[24] M. J. Berridge and R. F. Irvine, "Inositol trisphosphate, a novel second messenger in cellular signal transduction," *Nature*, vol. 312, no. 5992, pp. 315–321, 1984.

[25] Y. A. Hannun and L. M. Obeid, "Principles of bioactive lipid signalling: lessons from sphingolipids," *Nat. Rev. Mol. Cell Biol.*, vol. 9, pp. 139–150, Feb. 2008.

[26] S.-i. Hakomori, "Cell adhesion/recognition and signal transduction through glycosphingolipid microdomain," *Glycoconj. J.*, vol. 17, pp. 143–151, Mar. 2000.

[27] K. Iwabuchi, S. Yamamura, A. Prinetti, K. Handa, and S.-i. Hakomori, "GM3-enriched microdomain involved in cell adhesion and signal transduction through carbohydrate-carbohydrate interaction in mouse melanoma B16 cells," *J. Biol. Chem.*, vol. 273, pp. 9130–9138, 10 Apr. 1998.

[28] R. L. Schnaar, "Glycolipid-mediated cell-cell recognition in inflammation and nerve regeneration," *Arch. Biochem. Biophys.*, vol. 426, pp. 163–172, 15 June 2004.

[29] V. Cabiaux, C. Wolff, and J. M. Ruysschaert, "Interaction with a lipid membrane: a key step in bacterial toxins virulence," *Int. J. Biol. Macromol.*, vol. 21, pp. 285–298, Dec. 1997.

[30] C. Rodighiero, Y. Fujinaga, T. R. Hirst, and W. I. Lencer, "A cholera toxin b-subunit variant that binds ganglioside G(M1) but fails to induce toxicity," *J. Biol. Chem.*, vol. 276, pp. 36939–36945, 5 Oct. 2001.

[31] K. Bock, M. E. Breimer, A. Brignole, G. C. Hansson, K.-A. Karlsson, G. Larson, H. Leffler, B. E. Samuelsson, N. Strömberg, C. S. Edén, and J. Thurin, "Specificity of binding of a strain of uropathogenic escherichia coli to gal alpha 1-4gal-containing glycosphingolipids," *J. Biol. Chem.*, vol. 260, pp. 8545–8551, 15 July 1985.

[32] M. D. Resh, "Covalent lipid modifications of proteins," *Curr. Biol.*, vol. 23, pp. R431–R435, 20 May 2013.

[33] K. Simons and D. Toomre, "Lipid rafts and signal transduction," *Nat. Rev. Mol. Cell Biol.*, vol. 1, pp. 31–39, Oct. 2000.

[34] G. J. Wright, S. Martin, K. M. Bushell, and C. Söllner, "High-throughput identification of transient extracellular protein interactions," *Biochem. Soc. Trans.*, vol. 38, pp. 919–922, Aug. 2010.

[35] T. Glatter, A. Wepf, R. Aebersold, and M. Gstaiger, "An integrated workflow for charting the human interaction proteome: insights into the PP2A system," *Mol. Syst. Biol.*, vol. 5, p. 237, 20 Jan. 2009.

[36] Y. V. Miteva, H. G. Budayeva, and I. M. Cristea, "Proteomics-based methods for discovery, quantification, and validation of protein-protein interactions," *Anal. Chem.*, vol. 85, pp. 749–768, 15 Jan. 2013.

[37] N. C. Hubner, A. W. Bird, J. Cox, B. Splettstoesser, P. Bandilla, I. Poser, A. Hyman, and M. Mann, "Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions," *J. Cell Biol.*, vol. 189, pp. 739–754, 17 May 2010.

[38] T. Rolland, M. Taşan, B. Charloteaux, S. J. Pevzner, Q. Zhong, N. Sahni, S. Yi, I. Lemmens, C. Fontanillo, R. Mosca, A. Kamburov, S. D. Ghiassian, X. Yang, L. Ghamsari, D. Balcha, B. E. Begg, P. Braun, M. Brehme, M. P. Broly, A.-R. Carvunis, D. Convery-Zupan, R. Corominas, J. Coulombe-Huntington, E. Dann, M. Dreze, A. Dricot, C. Fan, E. Franzosa, F. Gebreab, B. J. Gutierrez, M. F. Hardy, M. Jin, S. Kang, R. Kiros, G. N. Lin, K. Luck, A. MacWilliams, J. Menche, R. R. Murray, A. Palagi, M. M. Poulin, X. Rambout, J. Rasla, P. Reichert, V. Romero, E. Ruyssinck, J. M. Sahalie, A. Scholz, A. A. Shah, A. Sharma, Y. Shen, K. Spirohn, S. Tam, A. O. Tejeda, S. A. Trigg, J.-C. Twizere, K. Vega, J. Walsh, M. E. Cusick, Y. Xia, A.-L. Barabási, L. M. Iakoucheva, P. Aloy, J. De Las Rivas, J. Tavernier, M. A. Calderwood, D. E. Hill, T. Hao, F. P. Roth, and M. Vidal, "A proteome-scale map of the human interactome network," *Cell*, vol. 159, pp. 1212–1226, 20 Nov. 2014.

[39] N. Ramachandran, J. V. Raphael, E. Hainsworth, G. Demirkan, M. G. Fuentes, A. Rolfs, Y. Hu, and J. LaBaer, "Next-generation high-density self-assembling functional protein arrays," *Nat. Methods*, vol. 5, pp. 535–538, June 2008.

[40] M. Meier, R. Sit, W. Pan, and S. R. Quake, "High-performance binary protein interaction screening in a microfluidic format," *Anal. Chem.*, vol. 84, pp. 9572–9578, 6 Nov. 2012.

[41] P. A. van der Merwe and A. N. Barclay, "Transient intercellular adhesion: the importance of weak protein-protein interactions," *Trends Biochem. Sci.*, vol. 19, pp. 354–358, Sept. 1994.

[42] G. J. Wright, "Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions," *Mol. Biosyst.*, vol. 5, pp. 1405–1412, Dec. 2009.

[43] O. Oyelaran and J. C. Gildersleeve, "Glycan arrays: recent advances and future challenges," *Curr. Opin. Chem. Biol.*, vol. 13, pp. 406–413, Oct. 2009.

[44] A. Varki and N. Sharon, "Historical background and overview," in *Essentials of Glycobiology* (A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds.), Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 20 Mar. 2010.

[45] H. H. Freeze and H. Schachter, "Genetic disorders of glycosylation," in *Essentials of Glycobiology* (A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds.), Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 20 Mar. 2010.

[46] J. D. Esko and P. Stanley, "Glycosylation mutants of cultured cells," in *Essentials of Glycobiology* (A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds.), Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 20 Mar. 2010.

[47] E. W. Raines and R. Ross, “Platelet-derived growth factor. i. high yield purification and evidence for multiple forms,” *J. Biol. Chem.*, vol. 257, pp. 5154–5160, 10 May 1982.

[48] D. F. Bowen-Pope and R. Ross, “Platelet-derived growth factor. II. specific binding to cultured cells,” *J. Biol. Chem.*, vol. 257, pp. 5161–5171, 10 May 1982.

[49] S. Cohen, R. A. Fava, and S. T. Sawyer, “Purification and characterization of epidermal growth factor receptor/protein kinase from normal mouse liver,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 79, pp. 6237–6241, Oct. 1982.

[50] S. Jacobs, E. Hazum, Y. Shechter, and P. Cuatrecasas, “Insulin receptor: covalent labeling and identification of subunits,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 76, pp. 4918–4921, Oct. 1979.

[51] H. Simonsen and H. F. Lodish, “Cloning by function: expression cloning in mammalian cells,” *Trends Pharmacol. Sci.*, vol. 15, pp. 437–441, 1 Dec. 1994.

[52] A. Aruffo and B. Seed, “Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 84, pp. 8573–8577, Dec. 1987.

[53] B. Seed and A. Aruffo, “Molecular cloning of the CD2 antigen, the t-cell erythrocyte receptor, by a rapid immunoselection procedure,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 84, pp. 3365–3369, May 1987.

[54] P. S. Linsley, E. A. Clark, and J. A. Ledbetter, “T-cell antigen CD28 mediates adhesion with B cells by interacting with activation antigen B7/BB-1,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 87, pp. 5031–5035, July 1990.

[55] K. Kato, M. Koyanagi, H. Okada, T. Takanashi, Y. W. Wong, A. F. Williams, K. Okumura, and H. Yagita, “CD48 is a counter-receptor for mouse CD2 and is involved in T cell activation,” *J. Exp. Med.*, vol. 176, pp. 1241–1249, 1 Nov. 1992.

[56] A. G. Dalgleish, P. C. Beverley, P. R. Clapham, D. H. Crawford, M. F. Greaves, and R. A. Weiss, “The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus,” *Nature*, vol. 312, no. 5996, pp. 763–767, 1984.

[57] D. Klatzmann, E. Champagne, S. Chamaret, J. Gruest, D. Guetard, T. Hercend, J.-C. Gluckman, and L. Montagnier, “T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV,” *Nature*, vol. 312, no. 5996, pp. 767–768, 1984.

[58] R. J. Colonno, P. L. Callahan, and W. J. Long, “Isolation of a monoclonal antibody that blocks attachment of the major group of human rhinoviruses,” *J. Virol.*, vol. 57, pp. 7–12, Jan. 1986.

[59] J. M. Greve, G. Davis, A. M. Meyer, C. P. Forte, S. C. Yost, C. W. Marlor, M. E. Kamarck, and A. McClelland, "The major human rhinovirus receptor is ICAM-1," *Cell*, vol. 56, pp. 839–847, 10 Mar. 1989.

[60] M. L. Plunkett, M. E. Sanders, P. Selvaraj, M. L. Dustin, and T. A. Springer, "Rosetting of activated human T lymphocytes with autologous erythrocytes. definition of the receptor and ligand molecules as CD2 and lymphocyte function-associated antigen 3 (LFA-3)," *J. Exp. Med.*, vol. 165, pp. 664–676, 1 Mar. 1987.

[61] R. G. Fehon, P. J. Kooh, I. Rebay, C. L. Regan, T. Xu, M. A. Muskavitch, and S. Artavanis-Tsakonas, "Molecular interactions between the protein products of the neurogenic loci notch and delta, two EGF-homologous genes in drosophila," *Cell*, vol. 61, pp. 523–534, 4 May 1990.

[62] M. Guttinger, F. Sutti, M. Panigada, S. Porcellini, B. Merati, M. Mariani, T. Teesalu, G. Consalez, and F. Grassi, "Epithelial v-like antigen (EVA), a novel member of the immunoglobulin superfamily, expressed in embryonic epithelia with a potential role as homotypic adhesion molecule in thymus histogenesis," *J. Cell Biol.*, vol. 141, pp. 1061–1071, 18 May 1998.

[63] S. S. Chan, H. Zheng, M. W. Su, R. Wilk, M. T. Killeen, E. M. Hedgecock, and J. G. Culotti, "UNC-40, a C. elegans homolog of DCC (deleted in colorectal cancer), is required in motile cells responding to UNC-6 netrin cues," *Cell*, vol. 87, pp. 187–195, 18 Oct. 1996.

[64] K. Keino-Masu, M. Masu, L. Hinck, E. D. Leonardo, S. S. Chan, J. G. Culotti, and M. Tessier-Lavigne, "Deleted in colorectal cancer (DCC) encodes a netrin receptor," *Cell*, vol. 87, pp. 175–185, 18 Oct. 1996.

[65] K. Brose, K. S. Bland, K. H. Wang, D. Arnott, W. Henzel, C. S. Goodman, M. Tessier-Lavigne, and T. Kidd, "Slit proteins bind robo receptors and have an evolutionarily conserved role in repulsive axon guidance," *Cell*, vol. 96, pp. 795–806, 19 Mar. 1999.

[66] H. F. Clark, A. L. Gurney, E. Abaya, K. Baker, D. Baldwin, J. Brush, J. Chen, B. Chow, C. Chui, C. Crowley, B. Currell, B. Deuel, P. Dowd, D. Eaton, J. Foster, C. Grimaldi, Q. Gu, P. E. Hass, S. Heldens, A. Huang, H. S. Kim, L. Klimowski, Y. Jin, S. Johnson, J. Lee, L. Lewis, D. Liao, M. Mark, E. Robbie, C. Sanchez, J. Schoenfeld, S. Seshagiri, L. Simmons, J. Singh, V. Smith, J. Stinson, A. Vagts, R. Vandlen, C. Watanabe, D. Wieand, K. Woods, M.-H. Xie, D. Yansura, S. Yi, G. Yu, J. Yuan, M. Zhang, Z. Zhang, A. Goddard, W. I. Wood, P. Godowski, and A. Gray, "The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment," *Genome Res.*, vol. 13, pp. 2265–2270, Oct. 2003.

[67] T. Battle, B. Antonsson, G. Feger, and D. Besson, "A high-throughput mammalian protein expression, purification, aliquoting and storage pipeline to assemble a library of the human secretome," *Comb. Chem. High Throughput Screen.*, vol. 9, pp. 639–649, Nov. 2006.

[68] M. S. Almén, K. J. V. Nordström, R. Fredriksson, and H. B. Schiöth, "Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin," *BMC Biol.*, vol. 7, p. 50, 13 Aug. 2009.

[69] J. Liu, V. T. K. Chow, and S. D. S. Jois, "A novel, rapid and sensitive heterotypic cell adhesion assay for CD2-CD58 interaction, and its application for testing inhibitory peptides," *J. Immunol. Methods*, vol. 291, pp. 39–49, Aug. 2004.

[70] S. D. Satyanarayanajois, S. Ronald, and J. Liu, "Heterotypic cell adhesion assay for the study of cell adhesion inhibition," *Methods Mol. Biol.*, vol. 716, pp. 225–243, 2011.

[71] D. M. E. Otto, M. A. Campanero-Rhodes, R. Karamanska, A. K. Powell, N. Bovin, J. E. Turnbull, R. A. Field, J. Blackburn, T. Feizi, and P. R. Crocker, "An expression system for screening of proteins for glycan and protein interactions," *Anal. Biochem.*, vol. 411, pp. 261–270, 15 Apr. 2011.

[72] A. R. Mäkelä and C. Oker-Blom, "The baculovirus display technology–an evolving instrument for molecular screening and drug delivery," *Comb. Chem. High Throughput Screen.*, vol. 11, pp. 86–98, Feb. 2008.

[73] T. Sakihama, T. Sato, H. Iwanari, T. Kitamura, S. Sakaguchi, T. Kodama, and T. Hamakubo, "A simple detection method for Low-Affinity membrane protein interactions by baculoviral display," *PLoS One*, vol. 3, p. e4024, 24 Dec. 2008.

[74] J. C. Lin, W.-H. Ho, A. Gurney, and A. Rosenthal, "The netrin-g1 ligand NGL-1 promotes the outgrowth of thalamocortical axons," *Nat. Neurosci.*, vol. 6, pp. 1270–1276, Dec. 2003.

[75] C. Bossen, K. Ingold, A. Tardivel, J.-L. Bodmer, O. Gaide, S. Hertig, C. Ambrose, J. Tschopp, and P. Schneider, "Interactions of tumor necrosis factor (TNF) and TNF receptor family members in the mouse and human," *J. Biol. Chem.*, vol. 281, pp. 13964–13971, 19 May 2006.

[76] S. V. Sosnovtsev, C. Sandoval-Jaime, G. I. Parra, C. M. Tin, R. W. Jones, J. Soden, D. Barnes, J. Freeth, A. W. Smith, and K. Y. Green, "Identification of human junctional adhesion molecule 1 as a functional receptor for the hom-1 calicivirus on human cells," *MBio*, vol. 8, 8 Mar. 2017.

[77] A. Salanti, T. M. Clausen, M. Ø. Agerbæk, N. Al Nakouzi, M. Dahlbäck, H. Z. Oo, S. Lee, T. Gustavsson, J. R. Rich, B. J. Hedberg, Y. Mao, L. Barington, M. A. Pereira, J. LoBello, M. Endo, L. Fazli, J. Soden, C. K. Wang, A. F. Sander, R. Dagil, S. Thrane, P. J. Holst, L. Meng, F. Favero, G. J. Weiss, M. A. Nielsen, J. Freeth, T. O. Nielsen, J. Zaia, N. L. Tran, J. Trent, J. S. Babcook, T. G. Theander, P. H. Sorensen, and M. Daugaard, "Targeting human cancer by a glycosaminoglycan binding malaria protein," *Cancer Cell*, vol. 28, pp. 500–514, 12 Oct. 2015.

[78] L. Turner, T. Lavstsen, S. S. Berger, C. W. Wang, J. E. V. Petersen, M. Avril, A. J. Brazier, J. Freeth, J. S. Jespersen, M. A. Nielsen, P. Magistrado, J. Lusingu, J. D. Smith, M. K. Higgins, and T. G. Theander, "Severe malaria is associated with parasite binding to endothelial protein C receptor," *Nature*, vol. 498, pp. 502–505, 27 June 2013.

[79] E. Bianchi, B. Doe, D. Goulding, and G. J. Wright, "Juno is the egg izumo receptor and is essential for mammalian fertilization," *Nature*, vol. 508, pp. 483–487, 24 Apr. 2014.

[80] K. M. Bushell, C. Söllner, B. Schuster-Boeckler, A. Bateman, and G. J. Wright, "Large-scale screening for novel low-affinity extracellular protein interactions," *Genome Res.*, vol. 18, pp. 622–630, Apr. 2008.

[81] J. J. Visser, Y. Cheng, S. C. Perry, A. B. Chastain, B. Parsa, S. S. Masri, T. A. Ray, J. N. Kay, and W. M. Wojtowicz, "An extracellular biochemical screen reveals that FLRTs and unc5s mediate neuronal subtype recognition in the retina," *Elife*, vol. 4, p. e08149, 2 Dec. 2015.

[82] E. Özkan, R. A. Carrillo, C. L. Eastman, R. Weiszmann, D. Waghray, K. G. Johnson, K. Zinn, S. E. Celniker, and K. C. Garcia, "An extracellular interactome of immunoglobulin and LRR proteins reveals receptor-ligand networks," *Cell*, vol. 154, pp. 228–239, 3 July 2013.

[83] S. R. Ramani, I. Tom, N. Lewin-Koh, B. Wranik, L. Depalatis, J. Zhang, D. Eaton, and L. C. Gonzalez, "A secreted protein microarray platform for extracellular protein interaction discovery," *Anal. Biochem.*, vol. 420, pp. 127–138, 15 Jan. 2012.

[84] K. M. Bushell, C. Söllner, B. Schuster-Boeckler, A. Bateman, and G. J. Wright, "Large-scale screening for novel low-affinity extracellular protein interactions," *Genome Res.*, vol. 18, pp. 622–630, Apr. 2008.

[85] C. Söllner and G. J. Wright, "A cell surface interaction network of neural leucine-rich repeat receptors," *Genome Biol.*, vol. 10, p. R99, 18 Sept. 2009.

[86] S. Martin, C. Söllner, V. Charoensawan, B. Adryan, B. Thisse, C. Thisse, S. Teichmann, and G. J. Wright, "Construction of a large extracellular protein interaction network and its resolution by spatiotemporal expression profiling," *Mol. Cell. Proteomics*, vol. 9, pp. 2654–2665, Dec. 2010.

[87] X. Gao, U. Metzger, P. Panza, P. Mahalwar, S. Alsheimer, H. Geiger, H.-M. Maischein, M. P. Levesque, M. Templin, and C. Söllner, "A Floor-Plate extracellular Protein-Protein interaction screen identifies draxin as a secreted netrin-1 antagonist," *Cell Rep.*, vol. 12, pp. 694–708, 28 July 2015.

[88] C. Crosnier, L. Y. Bustamante, S. J. Bartholdson, A. K. Bei, M. Theron, M. Uchikawa, S. Mboup, O. Ndir, D. P. Kwiatkowski, M. T. Duraisingh, J. C. Rayner, and G. J. Wright, "Basigin is a receptor essential for erythrocyte invasion by plasmodium falciparum," *Nature*, vol. 480, pp. 534–537, 22 Dec. 2011.

[89] S. J. Bartholdson, L. Y. Bustamante, C. Crosnier, S. Johnson, S. Lea, J. C. Rayner, and G. J. Wright, "Semaphorin-7A is an erythrocyte receptor for p. falciparum merozoite-specific TRAP homolog, MTRAP," *PLoS Pathog.*, vol. 8, p. e1003031, Jan. 2012.

[90] F. Galaway, L. G. Drought, M. Fala, N. Cross, A. C. Kemp, J. C. Rayner, and G. J. Wright, "P113 is a merozoite surface protein that binds the N terminus of plasmodium falciparum RH5," *Nat. Commun.*, vol. 8, p. 14333, 10 Feb. 2017.

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

[92] M. Letarte, D. Voulgaraki, D. Hatherley, M. Foster-Cuevas, N. J. Saunders, and A. N. Barclay, "Analysis of leukocyte membrane protein interactions using protein microarrays," *BMC Biochem.*, vol. 6, p. 2, 1 Mar. 2005.

[93] D. Voulgaraki, R. Mitnacht-Kraus, M. Letarte, M. Foster-Cuevas, M. H. Brown, and A. N. Barclay, "Multivalent recombinant proteins for probing functions of leucocyte surface proteins such as the CD200 receptor," *Immunology*, vol. 115, pp. 337–346, July 2005.

[94] N. Martinez-Martin, S. R. Ramani, J. A. Hackney, I. Tom, B. J. Wranik, M. Chan, J. Wu, M. T. Paluch, K. Takeda, P. E. Hass, H. Clark, and L. C. Gonzalez, "The extracellular interactome of the human adenovirus family reveals diverse strategies for immunomodulation," *Nat. Commun.*, vol. 7, p. 11473, 5 May 2016.

[95] Y. Sun, M. Gallagher-Jones, C. Barker, and G. J. Wright, "A benchmarked protein microarray-based platform for the identification of novel low-affinity extracellular protein interactions," *Anal. Biochem.*, vol. 424, pp. 45–53, 1 May 2012.

[96] Y. Sun, C. Vandenbriele, A. Kauskot, P. Verhamme, M. F. Hoylaerts, and G. J. Wright, "A human platelet receptor protein microarray identifies the high affinity immunoglobulin E receptor subunit α (FcϵR1α) as an activating platelet endothelium aggregation receptor 1 (PEAR1) ligand," *Mol. Cell. Proteomics*, vol. 14, pp. 1265–1274, May 2015.

[97] D. Josic and J. G. Clifton, "Mammalian plasma membrane proteomics," *Proteomics*, vol. 7, pp. 3010–3029, Aug. 2007.

[98] B. A. Macher and T.-Y. Yen, "Proteins at membrane surfaces-a review of approaches," *Mol. Biosyst.*, vol. 3, pp. 705–713, Oct. 2007.

[99] H. Zhang, X.-J. Li, D. B. Martin, and R. Aebersold, "Identification and quantification of n-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry," *Nat. Biotechnol.*, vol. 21, pp. 660–666, June 2003.

[100] C. C. Wu, M. J. MacCoss, K. E. Howell, and J. R. Yates, 3rd, "A method for the comprehensive proteomic analysis of membrane proteins," *Nat. Biotechnol.*, vol. 21, pp. 532–538, May 2003.

[101] H. Watarai, A. Hinohara, J. Nagafune, T. Nakayama, M. Taniguchi, and Y. Yamaguchi, "Plasma membrane-focused proteomics: dramatic changes in surface expression during the maturation of human dendritic cells," *Proteomics*, vol. 5, pp. 4001–4011, Oct. 2005.

[102] A. P. Frei, O.-Y. Jeon, S. Kilcher, H. Moest, L. M. Henning, C. Jost, A. Plückthun, J. Mercer, R. Aebersold, E. M. Carreira, and B. Wollscheid, "Direct identification of ligand-receptor interactions on living cells and tissues," *Nat. Biotechnol.*, vol. 30, pp. 997–1001, 16 Sept. 2012.

[103] Y. Li, T. Ozment, G. L. Wright, and J. M. Peterson, "Identification of putative receptors for the novel adipokine CTRP3 using Ligand-Receptor capture technology," *PLoS One*, vol. 11, p. e0164593, 11 Oct. 2016.

[104] R. D. Cummings and M. E. Etzler, "Antibodies and lectins in glycan analysis," in *Essentials of Glycobiology* (A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds.), Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 20 Mar. 2010.

[105] B. K. Brandley, S. J. Swiedler, and P. W. Robbins, "Carbohydrate ligands of the LEC cell adhesion molecules," *Cell*, vol. 63, pp. 861–863, 30 Nov. 1990.

[106] L. A. Lasky, "Selectin-carbohydrate interactions and the initiation of the inflammatory response," *Annu. Rev. Biochem.*, vol. 64, pp. 113–139, 1995.

[107] N. Sharon and H. Lis, "History of lectins: from hemagglutinins to biological recognition molecules," *Glycobiology*, vol. 14, pp. 53R–62R, 1 Nov. 2004.

[108] J. D. Esko and C. R. Bertozzi, "Chemical tools for inhibiting glycosylation," in *Essentials of Glycobiology* (A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds.), Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 20 Mar. 2010.

[109] D. E. Humphries and J. E. Silbert, "Chlorate: a reversible inhibitor of proteoglycan sulfation," *Biochem. Biophys. Res. Commun.*, vol. 154, pp. 365–371, 15 July 1988.

[110] H. H. Freeze and C. Kranz, "Endoglycosidase and glycoamidase release of n-linked glycans," *Curr. Protoc. Mol. Biol.*, vol. Chapter 17, p. Unit 17.13A, Jan. 2010.

[111] D. Camus and T. J. Hadley, "A plasmodium falciparum antigen that binds to host erythrocytes and merozoites," *Science*, vol. 230, pp. 553–556, 1 Nov. 1985.

[112] P. A. Orlandi, F. W. Klotz, and J. D. Haynes, "A malaria invasion receptor, the 175-kilodalton erythrocyte binding antigen of plasmodium falciparum recognizes the terminal Neu5Ac(alpha 2-3)gal- sequences of glycophorin a," *J. Cell Biol.*, vol. 116, pp. 901–909, Feb. 1992.

[113] F. Li, P. P. Wilkins, S. Crawley, J. Weinstein, R. D. Cummings, and R. P. McEver, "Post-translational modifications of recombinant p-selectin glycoprotein ligand-1 required for binding to P- and e-selectin," *J. Biol. Chem.*, vol. 271, pp. 3255–3264, 9 Feb. 1996.

[114] I. Stamenkovic, D. Sgroi, A. Aruffo, M. S. Sy, and T. Anderson, "The B lymphocyte adhesion molecule CD22 interacts with leukocyte common antigen CD45RO on T cells and alpha 2-6 sialyltransferase, CD75, on B cells," *Cell*, vol. 66, pp. 1133–1144, 20 Sept. 1991.

[115] P. R. Crocker and A. Varki, "Siglecs in the immune system," *Immunology*, vol. 103, pp. 137–145, June 2001.

[116] C. D. Rillahan and J. C. Paulson, "Glycan microarrays for decoding the glycome," *Annu. Rev. Biochem.*, vol. 80, pp. 797–823, 2011.

[117] M. Wanaguru, C. Crosnier, S. Johnson, J. C. Rayner, and G. J. Wright, "Biochemical analysis of the plasmodium falciparum erythrocyte-binding antigen-175 (EBA175)-Glycophorin-A interaction: IMPLICATIONS FOR VACCINE DESIGN," *J. Biol. Chem.*, vol. 288, pp. 32106–32117, 8 Nov. 2013.

[118] R. Phillips, T. Ursell, P. Wiggins, and P. Sens, "Emerging roles for lipids in shaping membrane-protein function," *Nature*, vol. 459, pp. 379–385, 21 May 2009.

[119] A. G. Lee, "How lipids affect the activities of integral membrane proteins," *Biochim. Biophys. Acta*, vol. 1666, pp. 62–87, 3 Nov. 2004.

[120] J. Oates and A. Watts, "Uncovering the intimate relationship between lipids, cholesterol and GPCR activation," *Curr. Opin. Struct. Biol.*, vol. 21, pp. 802–807, Dec. 2011.

[121] P. P. Wilkins, K. L. Moore, R. P. McEver, and R. D. Cummings, "Tyrosine sulfation of p-selectin glycoprotein ligand-1 is required for high affinity binding to p-selectin," *J. Biol. Chem.*, vol. 270, pp. 22677–22680, 29 Sept. 1995.

[122] B. E. Housden, M. Muhar, M. Gemberling, C. A. Gersbach, D. Y. R. Stainier, G. Seydoux, S. E. Mohr, J. Zuber, and N. Perrimon, "Loss-of-function genetic tools for animal models: cross-species and cross-platform differences," *Nat. Rev. Genet.*, vol. 18, pp. 24–40, Jan. 2017.

[123] S. Brenner, "The genetics of caenorhabditis elegans," *Genetics*, vol. 77, pp. 71–94, May 1974.

[124] M. Gans, C. Audit, and M. Masson, "Isolation and characterization of sex-linked female-sterile mutants in drosophila melanogaster," *Genetics*, vol. 81, pp. 683–704, Dec. 1975.

[125] C. Nüsslein-Volhard and E. Wieschaus, “Mutations affecting segment number and polarity in drosophila,” *Nature*, vol. 287, pp. 795–801, 30 Oct. 1980.

[126] L. H. Miller, S. J. Mason, D. F. Clyde, and M. H. McGinniss, “The resistance factor to plasmodium vivax in blacks. the duffy-blood-group genotype, FyFy,” *N. Engl. J. Med.*, vol. 295, pp. 302–304, 5 Aug. 1976.

[127] R. E. Howes, A. P. Patil, F. B. Piel, O. A. Nyangiri, C. W. Kabaria, P. W. Gething, P. A. Zimmerman, C. Barnadas, C. M. Beall, A. Gebremedhin, D. Ménard, T. N. Williams, D. J. Weatherall, and S. I. Hay, “The global distribution of the duffy blood group,” *Nat. Commun.*, vol. 2, p. 266, 2011.

[128] S. J. Bartholdson, C. Crosnier, L. Y. Bustamante, J. C. Rayner, and G. J. Wright, “Identifying novel plasmodium falciparum erythrocyte invasion receptors using systematic extracellular protein interaction screens,” *Cell. Microbiol.*, vol. 15, pp. 1304–1312, Aug. 2013.

[129] R. Liu, W. A. Paxton, S. Choe, D. Ceradini, S. R. Martin, R. Horuk, M. E. MacDonald, H. Stuhlmann, R. A. Koup, and N. R. Landau, “Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection,” *Cell*, vol. 86, pp. 367–377, 9 Aug. 1996.

[130] P. Anderson, “Mutagenesis,” *Methods Cell Biol.*, vol. 48, pp. 31–58, 1995.

[131] P. Novick, C. Field, and R. Schekman, “Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway,” *Cell*, vol. 21, pp. 205–215, Aug. 1980.

[132] P. Novick and R. Schekman, “Secretion and cell-surface growth are blocked in a temperature-sensitive mutant of saccharomyces cerevisiae,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 76, pp. 1858–1862, Apr. 1979.

[133] J. W. Ziel and D. R. Sherwood, “Roles for netrin signaling outside of axon guidance: a view from the worm,” *Dev. Dyn.*, vol. 239, pp. 1296–1305, May 2010.

[134] G. Guo, W. Wang, and A. Bradley, “Mismatch repair genes identified using genetic screens in blm-deficient embryonic stem cells,” *Nature*, vol. 429, pp. 891–895, 24 June 2004.

[135] M. Kotecki, P. S. Reddy, and B. H. Cochran, “Isolation and characterization of a near-haploid human cell line,” *Exp. Cell Res.*, vol. 252, pp. 273–280, 1 Nov. 1999.

[136] B. S. Andersson, V. P. Collins, R. Kurzrock, D. W. Larkin, C. Childs, A. Ost, A. Cork, J. M. Trujillo, E. J. Freireich, and M. J. Siciliano, “KBM-7, a human myeloid leukemia cell line with double philadelphia chromosomes lacking normal c-ABL and BCR transcripts,” *Leukemia*, vol. 9, pp. 2100–2108, Dec. 1995.

[137] J. E. Carette, M. Raaben, A. C. Wong, A. S. Herbert, G. Obernosterer, N. Mulherkar, A. I. Kuehne, P. J. Kranzusch, A. M. Griffin, G. Ruthel, P. Dal Cin, J. M. Dye, S. P. Whelan, K. Chandran, and T. R. Brummelkamp, "Ebola virus entry requires the cholesterol transporter Niemann-Pick C1," *Nature*, vol. 477, pp. 340–343, 24 Aug. 2011.

[138] M. Leeb and A. Wutz, "Derivation of haploid embryonic stem cells from mouse embryos," *Nature*, vol. 479, pp. 131–134, 7 Sept. 2011.

[139] S. Pillay and J. E. Carette, "Hunting viral receptors using haploid cells," *Annu Rev Virol*, vol. 2, pp. 219–239, Nov. 2015.

[140] J. E. Carette, C. P. Guimaraes, M. Varadarajan, A. S. Park, I. Wuethrich, A. Godarova, M. Kotecki, B. H. Cochran, E. Spooner, H. L. Ploegh, and T. R. Brummelkamp, "Haploid genetic screens in human cells identify host factors used by pathogens," *Science*, vol. 326, pp. 1231–1235, 27 Nov. 2009.

[141] J. E. Carette, C. P. Guimaraes, I. Wuethrich, V. A. Blomen, M. Varadarajan, C. Sun, G. Bell, B. Yuan, M. K. Muellner, S. M. Nijman, H. L. Ploegh, and T. R. Brummelkamp, "Global gene disruption in human cells to assign genes to phenotypes by deep sequencing," *Nat. Biotechnol.*, vol. 29, pp. 542–546, 29 May 2011.

[142] F. G. Tafesse, C. P. Guimaraes, T. Maruyama, J. E. Carette, S. Lory, T. R. Brummelkamp, and H. L. Ploegh, "GPR107, a g-protein-coupled receptor essential for intoxication by pseudomonas aeruginosa exotoxin a, localizes to the golgi and is cleaved by furin," *J. Biol. Chem.*, vol. 289, pp. 24005–24018, 29 Aug. 2014.

[143] L. T. Jae, M. Raaben, M. Riemersma, E. van Beusekom, V. A. Blomen, A. Velds, R. M. Kerkhoven, J. E. Carette, H. Topaloglu, P. Meinecke, M. W. Wessels, D. J. Lefeber, S. P. Whelan, H. van Bokhoven, and T. R. Brummelkamp, "Deciphering the glycosylome of dystroglycanopathies using haploid screens for lassa virus entry," *Science*, vol. 340, pp. 479–483, 26 Apr. 2013.

[144] L. T. Jae, M. Raaben, A. S. Herbert, A. I. Kuehne, A. S. Wirchnianski, T. K. Soh, S. H. Stubbs, H. Janssen, M. Damme, P. Saftig, S. P. Whelan, J. M. Dye, and T. R. Brummelkamp, "Virus entry. lassa virus entry requires a trigger-induced receptor switch," *Science*, vol. 344, pp. 1506–1510, 27 June 2014.

[145] S. Pillay, N. L. Meyer, A. S. Puschnik, O. Davulcu, J. Diep, Y. Ishikawa, L. T. Jae, J. E. Wosen, C. M. Nagamine, M. S. Chapman, and J. E. Carette, "An essential receptor for adeno-associated virus infection," *Nature*, vol. 530, pp. 108–112, 4 Feb. 2016.

[146] C. Fellmann and S. W. Lowe, "Stable RNA interference rules for silencing," *Nat. Cell Biol.*, vol. 16, pp. 10–18, Jan. 2014.

[147] L. Timmons and A. Fire, "Specific interference by ingested dsRNA," *Nature*, vol. 395, p. 854, 29 Oct. 1998.

[148] L. Pelkmans, E. Fava, H. Grabner, M. Hannus, B. Habermann, E. Krausz, and M. Zerial, "Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis," *Nature*, vol. 436, pp. 78–86, 7 July 2005.

[149] X. Huang, J. Y. J. Wang, and X. Lu, "Systems analysis of quantitative shRNA-library screens identifies regulators of cell adhesion," *BMC Syst. Biol.*, vol. 2, p. 49, 13 June 2008.

[150] P. P. Rose, S. L. Hanna, A. Spiridigliozzi, N. Wannissorn, D. P. Beiting, S. R. Ross, R. W. Hardy, S. A. Bambina, M. T. Heise, and S. Cherry, "Natural resistance-associated macrophage protein is a cellular receptor for sindbis virus in both insect and mammalian hosts," *Cell Host Microbe*, vol. 10, pp. 97–104, 18 Aug. 2011.

[151] O. M. Sessions, N. J. Barrows, J. A. Souza-Neto, T. J. Robinson, C. L. Hershey, M. A. Rodgers, J. L. Ramirez, G. Dimopoulos, P. L. Yang, J. L. Pearson, and M. A. Garcia-Blanco, "Discovery of insect and human dengue virus host factors," *Nature*, vol. 458, pp. 1047–1050, 23 Apr. 2009.

[152] A. W. Tai, Y. Benita, L. F. Peng, S.-S. Kim, N. Sakamoto, R. J. Xavier, and R. T. Chung, "A functional genomic screen identifies cellular cofactors of hepatitis C virus replication," *Cell Host Microbe*, vol. 5, pp. 298–307, 19 Mar. 2009.

[153] Q. Li, A. L. Brass, A. Ng, Z. Hu, R. J. Xavier, T. J. Liang, and S. J. Elledge, "A genome-wide genetic screen for host factors required for hepatitis C virus propagation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, pp. 16410–16415, 22 Sept. 2009.

[154] M. N. Krishnan, A. Ng, B. Sukumaran, F. D. Gilfoy, P. D. Uchil, H. Sultana, A. L. Brass, R. Adametz, M. Tsui, F. Qian, R. R. Montgomery, S. Lev, P. W. Mason, R. A. Koski, S. J. Elledge, R. J. Xavier, H. Agaisse, and E. Fikrig, "RNA interference screen for human genes associated with west nile virus infection," *Nature*, vol. 455, pp. 242–245, 11 Sept. 2008.

[155] J. Mercer, B. Snijder, R. Sacher, C. Burkard, C. K. E. Bleck, H. Stahlberg, L. Pelkmans, and A. Helenius, "RNAi screening reveals proteasome- and cullin3-dependent stages in vaccinia virus infection," *Cell Rep.*, vol. 2, pp. 1036–1047, 25 Oct. 2012.

[156] S. P. Goff, "Knockdown screens to knockout HIV-1," *Cell*, vol. 135, pp. 417–420, 31 Oct. 2008.

[157] E. S. Egan, R. H. Y. Jiang, M. A. Moechtar, N. S. Barteneva, M. P. Weekes, L. V. Nobre, S. P. Gygi, J. A. Paulo, C. Frantzreb, Y. Tani, J. Takahashi, S. Watanabe, J. Goldberg, A. S. Paul, C. Brugnara, D. E. Root, R. C. Wiegand, J. G. Doench, and M. T. Duraisingh, "Malaria. a forward genetic screen identifies erythrocyte CD55 as essential for plasmodium falciparum invasion," *Science*, vol. 348, pp. 711–714, 8 May 2015.

[158] A. L. Jackson and P. S. Linsley, "Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application," *Nat. Rev. Drug Discov.*, vol. 9, pp. 57–67, Jan. 2010.

[159] R. König, Y. Zhou, D. Elleder, T. L. Diamond, G. M. C. Bonamy, J. T. Irelan, C.-Y. Chiang, B. P. Tu, P. D. De Jesus, C. E. Lilley, S. Seidel, A. M. Opaluch, J. S. Caldwell, M. D. Weitzman, K. L. Kuhen, S. Bandyopadhyay, T. Ideker, A. P. Orth, L. J. Miraglia, F. D. Bushman, J. A. Young, and S. K. Chanda, "Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication," *Cell*, vol. 135, pp. 49–60, 3 Oct. 2008.

[160] A. L. Brass, D. M. Dykxhoorn, Y. Benita, N. Yan, A. Engelman, R. J. Xavier, J. Lieberman, and S. J. Elledge, "Identification of host proteins required for HIV infection through a functional genomic screen," *Science*, vol. 319, pp. 921–926, 15 Feb. 2008.

[161] H. Zhou, M. Xu, Q. Huang, A. T. Gates, X. D. Zhang, J. C. Castle, E. Stec, M. Ferrer, B. Strulovici, D. J. Hazuda, and A. S. Espeseth, "Genome-scale RNAi screen for host factors required for HIV replication," *Cell Host Microbe*, vol. 4, pp. 495–504, 13 Nov. 2008.

[162] F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, and E. Soria, "Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements," *J. Mol. Evol.*, vol. 60, pp. 174–182, Feb. 2005.

[163] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath, "CRISPR provides acquired resistance against viruses in prokaryotes," *Science*, vol. 315, pp. 1709–1712, 23 Mar. 2007.

[164] P. D. Hsu, E. S. Lander, and F. Zhang, "Development and applications of CRISPR-Cas9 for genome engineering," *Cell*, vol. 157, pp. 1262–1278, 5 June 2014.

[165] J. A. Doudna and E. Charpentier, "The new frontier of genome engineering with CRISPR-Cas9," *Science*, vol. 346, p. 1258096, 28 Nov. 2014.

[166] "CRISPR guide RNA | dharmacon." http://dharmacon.gelifesciences.com/gene-editing/crispr-cas9/crispr-guide-rna/. Accessed: 2017-9-18.

[167] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity," *Science*, vol. 337, pp. 816–821, 17 Aug. 2012.

[168] P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, and G. M. Church, "RNA-guided human genome engineering via cas9," *Science*, vol. 339, pp. 823–826, 15 Feb. 2013.

[169] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang, "Multiplex genome engineering

using CRISPR/Cas systems," *Science*, vol. 339, pp. 819–823, 15 Feb. 2013.

[170] Z. Mao, M. Bozzella, A. Seluanov, and V. Gorbunova, "DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells," *Cell Cycle*, vol. 7, pp. 2902–2906, 15 Sept. 2008.

[171] O. Shalem, N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. S. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench, and F. Zhang, "Genome-scale CRISPR-Cas9 knockout screening in human cells," *Science*, vol. 343, pp. 84–87, 3 Jan. 2014.

[172] T. Wang, J. J. Wei, D. M. Sabatini, and E. S. Lander, "Genetic screens in human cells using the CRISPR-Cas9 system," *Science*, vol. 343, pp. 80–84, 3 Jan. 2014.

[173] H. Koike-Yusa, Y. Li, E.-P. Tan, M. D. C. Velasco-Herrera, and K. Yusa, "Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library," *Nat. Biotechnol.*, vol. 32, pp. 267–273, 23 Mar. 2014.

[174] J. Joung, S. Konermann, J. S. Gootenberg, O. O. Abudayyeh, R. J. Platt, M. D. Brigham, N. E. Sanjana, and F. Zhang, "Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening," *Nat. Protoc.*, vol. 12, pp. 828–863, Apr. 2017.

[175] T. Hart, M. Chandrashekhar, M. Aregger, Z. Steinhart, K. R. Brown, G. MacLeod, M. Mis, M. Zimmermann, A. Fradet-Turcotte, S. Sun, P. Mero, P. Dirks, S. Sidhu, F. P. Roth, O. S. Rissland, D. Durocher, S. Angers, and J. Moffat, "High-Resolution CRISPR screens reveal fitness genes and Genotype-Specific cancer liabilities," *Cell*, vol. 163, pp. 1515–1526, 3 Dec. 2015.

[176] K. Tzelepis, H. Koike-Yusa, E. De Braekeleer, Y. Li, E. Metzakopian, O. M. Dovey, A. Mupo, V. Grinkevich, M. Li, M. Mazan, M. Gozdecka, S. Ohnishi, J. Cooper, M. Patel, T. McKerrell, B. Chen, A. F. Domingues, P. Gallipoli, S. Teichmann, H. Ponstingl, U. McDermott, J. Saez-Rodriguez, B. J. P. Huntly, F. Iorio, C. Pina, G. S. Vassiliou, and K. Yusa, "A CRISPR dropout screen identifies genetic vulnerabilities and therapeutic targets in acute myeloid leukemia," *Cell Rep.*, vol. 17, pp. 1193–1205, 18 Oct. 2016.

[177] T. Wang, K. Birsoy, N. W. Hughes, K. M. Krupczak, Y. Post, J. J. Wei, E. S. Lander, and D. M. Sabatini, "Identification and characterization of essential genes in the human genome," *Science*, vol. 350, pp. 1096–1101, 27 Nov. 2015.

[178] Y. Zhou, S. Zhu, C. Cai, P. Yuan, C. Li, Y. Huang, and W. Wei, "High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells," *Nature*, vol. 509, pp. 487–491, 22 May 2014.

[179] S. Virreira Winter, A. Zychlinsky, and B. W. Bardoel, "Genome-wide CRISPR screen reveals novel host factors required for staphylococcus

aureus α-hemolysin-mediated toxicity," *Sci. Rep.*, vol. 6, p. 24242, 12 Apr. 2016.

[180] I. H. Jain, L. Zazzeron, R. Goli, K. Alexa, S. Schatzman-Bone, H. Dhillon, O. Goldberger, J. Peng, O. Shalem, N. E. Sanjana, F. Zhang, W. Goessling, W. M. Zapol, and V. K. Mootha, "Hypoxia as a therapy for mitochondrial disease," *Science*, vol. 352, pp. 54–61, 1 Apr. 2016.

[181] C. J. Blondel, J. S. Park, T. P. Hubbard, A. R. Pacheco, C. J. Kuehl, M. J. Walsh, B. M. Davis, B. E. Gewurz, J. G. Doench, and M. K. Waldor, "CRISPR/Cas9 screens reveal requirements for host cell sulfation and fucosylation in bacterial type III secretion System-Mediated cytotoxicity," *Cell Host Microbe*, vol. 20, pp. 226–237, 10 Aug. 2016.

[182] R. J. Park, T. Wang, D. Koundakjian, J. F. Hultquist, P. Lamothe-Molina, B. Monel, K. Schumann, H. Yu, K. M. Krupzcak, W. Garcia-Beltran, A. Piechocka-Trocha, N. J. Krogan, A. Marson, D. M. Sabatini, E. S. Lander, N. Hacohen, and B. D. Walker, "A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors," *Nat. Genet.*, vol. 49, pp. 193–203, Feb. 2017.

[183] R. Zhang, J. J. Miner, M. J. Gorman, K. Rausch, H. Ramage, J. P. White, A. Zuiani, P. Zhang, E. Fernandez, Q. Zhang, K. A. Dowd, T. C. Pierson, S. Cherry, and M. S. Diamond, "A CRISPR screen defines a signal peptide processing pathway required by flaviviruses," *Nature*, vol. 535, pp. 164–168, 17 June 2016.

[184] E. Hartenian and J. G. Doench, "Genetic screens and functional genomics using CRISPR/Cas9 technology," *FEBS J.*, vol. 282, pp. 1383–1393, 1 Apr. 2015.

[185] R. DeJesus, F. Moretti, G. McAllister, Z. Wang, P. Bergman, S. Liu, E. Frias, J. Alford, J. S. Reece-Hoyes, A. Lindeman, J. Kelliher, C. Russ, J. Knehr, W. Carbone, M. Beibel, G. Roma, A. Ng, J. A. Tallarico, J. A. Porter, R. J. Xavier, C. Mickanin, L. O. Murphy, G. R. Hoffman, and B. Nyfeler, "Functional CRISPR screening identifies the ufmylation pathway as a regulator of SQSTM1/p62," *Elife*, vol. 5, 28 June 2016.

[186] O. Parnas, M. Jovanovic, T. M. Eisenhaure, R. H. Herbst, A. Dixit, C. J. Ye, D. Przybylski, R. J. Platt, I. Tirosh, N. E. Sanjana, O. Shalem, R. Satija, R. Raychowdhury, P. Mertins, S. A. Carr, F. Zhang, N. Hacohen, and A. Regev, "A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks," *Cell*, vol. 162, pp. 675–686, 30 July 2015.

[187] M. L. Burr, C. E. Sparbier, Y.-C. Chan, J. C. Williamson, K. Woods, P. A. Beavis, E. Y. N. Lam, M. A. Henderson, C. C. Bell, S. Stolzenburg, O. Gilan, S. Bloor, T. Noori, D. W. Morgens, M. C. Bassik, P. J. Neeson, A. Behren, P. K. Darcy, S.-J. Dawson, I. Voskoboinik, J. A. Trapani, J. Cebon, P. J. Lehner, and M. A. Dawson, "CMTM6 maintains the expression of PD-L1 and regulates anti-tumour immunity," *Nature*, vol. 549, pp. 101–105, 7 Sept. 2017.

[188] Y. Durocher, S. Perret, and A. Kamen, "High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells," *Nucleic Acids Res.*, vol. 30, p. E9, 15 Jan. 2002.

[189] E. K. Brinkman, T. Chen, M. Amendola, and B. van Steensel, "Easy quantitative assessment of genome editing by sequence trace decomposition," *Nucleic Acids Res.*, vol. 42, p. e168, 16 Dec. 2014.

[190] W. Li, H. Xu, T. Xiao, L. Cong, M. I. Love, F. Zhang, R. A. Irizarry, J. S. Liu, M. Brown, and X. S. Liu, "MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens," *Genome Biol.*, vol. 15, no. 12, p. 554, 2014.

[191] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation," *Nucleic Acids Res.*, vol. 44, pp. D457–62, Jan. 2016.

[192] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[193] The UniProt Consortium, "UniProt: the universal protein knowledgebase," *Nucleic Acids Res.*, vol. 45, pp. D158–D169, Jan 2017.

[194] N. L. Grimsey, C. E. Goodfellow, E. L. Scotter, M. J. Dowie, M. Glass, and E. S. Graham, "Specific detection of CB1 receptors; cannabinoid CB1 receptor antibodies are not all created equal!," *J. Neurosci. Methods*, vol. 171, pp. 78–86, 15 June 2008.

[195] B. C. Jensen, P. M. Swigart, and P. C. Simpson, "Ten commercial antibodies for alpha-1-adrenergic receptor subtypes are nonspecific," *Naunyn. Schmiedebergs. Arch. Pharmacol.*, vol. 379, pp. 409–412, Apr. 2009.

[196] C. B. Saper, "An open letter to our readers on the use of antibodies," *J. Comp. Neurol.*, vol. 493, pp. 477–478, 26 Dec. 2005.

[197] M. C. Michel, T. Wieland, and G. Tsujimoto, "How reliable are g-protein-coupled receptor antibodies?," *Naunyn. Schmiedebergs. Arch. Pharmacol.*, vol. 379, pp. 385–388, Apr. 2009.

[198] D. M. Ecker, S. D. Jones, and H. L. Levine, "The therapeutic monoclonal antibody market," *MAbs*, vol. 7, pp. 9–14, 2 Jan. 2015.

[199] G. Roncador, P. Engel, L. Maestre, A. P. Anderson, J. L. Cordell, M. S. Cragg, V. Č. Šerbec, M. Jones, V. J. Lisnic, L. Kremer, D. Li, F. Koch-Nolte, N. Pascual, J.-I. Rodríguez-Barbosa, R. Torensma, H. Turley, K. Pulford, and A. H. Banham, "The european antibody network's practical guide to finding and validating suitable antibodies for research," *MAbs*, vol. 8, no. 1, pp. 27–36, 2016.

[200] J. Bordeaux, A. Welsh, S. Agarwal, E. Killiam, M. Baquero, J. Hanna, V. Anagnostou, and D. Rimm, "Antibody validation," *Biotechniques*, vol. 48, pp. 197–209, Mar. 2010.

[201] H. A. Mannsperger, S. Uhlmann, C. Schmidt, S. Wiemann, O. Sahin, and U. Korf, "RNAi-based validation of antibodies for reverse phase protein arrays," *Proteome Sci.*, vol. 8, p. 69, 23 Dec. 2010.

[202] C. Stadler, M. Hjelmare, B. Neumann, K. Jonasson, R. Pepperkok, M. Uhlén, and E. Lundberg, "Systematic validation of antibody binding and protein subcellular localization using siRNA and confocal microscopy," *J. Proteomics*, vol. 75, pp. 2236–2251, 3 Apr. 2012.

[203] L. Zhao, L. Qu, J. Zhou, Z. Sun, H. Zou, Y.-Y. Chen, J. D. Marks, and Y. Zhou, "High throughput identification of monoclonal antibodies to membrane bound and secreted proteins using yeast and phage display," *PLoS One*, vol. 9, p. e111339, 29 Oct. 2014.

[204] D. Akopian, K. Shen, X. Zhang, and S.-O. Shan, "Signal recognition particle: an essential protein-targeting machine," *Annu. Rev. Biochem.*, vol. 82, pp. 693–721, 13 Feb. 2013.

[205] J. G. Doench, N. Fusi, M. Sullender, M. Hegde, E. W. Vaimberg, K. F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard, H. W. Virgin, J. Listgarten, and D. E. Root, "Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9," *Nat. Biotechnol.*, vol. 34, pp. 184–191, Feb. 2016.

[206] B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G.-W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, and B. Huang, "Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system," *Cell*, vol. 155, pp. 1479–1491, 19 Dec. 2013.

[207] A. Agrotis and R. Ketteler, "A new age in functional genomics using CRISPR/Cas9 in arrayed library screening," *Front. Genet.*, vol. 6, 2015.

[208] P. Kirk, M. C. Wilson, C. Heddle, M. H. Brown, A. N. Barclay, and A. P. Halestrap, "CD147 is tightly associated with lactate transporters MCT1 and MCT4 and facilitates their cell surface expression," *EMBO J.*, vol. 19, pp. 3896–3904, 1 Aug. 2000.

[209] A. A. Deora, N. Philp, J. Hu, D. Bok, and E. Rodriguez-Boulan, "Mechanisms regulating tissue-specific polarity of monocarboxylate transporters and their chaperone CD147 in kidney and retinal epithelia," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, pp. 16245–16250, 8 Nov. 2005.

[210] M. Voss, U. Künzel, F. Higel, P.-H. Kuhn, A. Colombo, A. Fukumori, M. Haug-Kröper, B. Klier, G. Grammer, A. Seidl, B. Schröder, R. Obst, H. Steiner, S. F. Lichtenthaler, C. Haass, and R. Fluhrer, "Shedding of glycan-modifying enzymes by signal peptide peptidase-like 3 (SPPL3) regulates cellular n-glycosylation," *EMBO J.*, vol. 33, pp. 2890–2905, 17 Dec. 2014.

[211] Y. Bai, W. Huang, L.-T. Ma, J.-L. Jiang, and Z.-N. Chen, "Importance of n-glycosylation on CD147 for its biological functions," *Int. J. Mol. Sci.*, vol. 15, pp. 6356–6377, 15 Apr. 2014.

[212] Z. Luo, X. Zhang, W. Zeng, J. Su, K. Yang, L. Lu, C. B. Lim, W. Tang, L. Wu, S. Zhao, X. Jia, C. Peng, and X. Chen, "TRAF6 regulates melanoma invasion and metastasis through ubiquitination of basigin," *Oncotarget*, vol. 7, pp. 7179–7192, 9 Feb. 2016.

[213] B. P. Morgan, "Isolation and characterization of the complement-inhibiting protein CD59 antigen from platelet membranes," *Biochem. J*, vol. 282 (Pt 2), pp. 409–413, 1 Mar. 1992.

[214] P. M. Rudd, B. P. Morgan, M. R. Wormald, D. J. Harvey, C. W. van den Berg, S. J. Davis, M. A. Ferguson, and R. A. Dwek, "The glycosylation of the complement regulatory protein, human erythrocyte CD59," *J. Biol. Chem.*, vol. 272, pp. 7229–7244, 14 Mar. 1997.

[215] C. Xu and D. T. W. Ng, "Glycosylation-directed quality control of protein folding," *Nat. Rev. Mol. Cell Biol.*, vol. 16, pp. 742–752, Dec. 2015.

[216] C. C. R. Center, "Integrated technology resource for biomedical glycomics." https://www.ccrc.uga.edu/~moremen/glycomics/. Accessed: 2017-6-10.

[217] M. A. Kukuruzinska and P. W. Robbins, "Protein glycosylation in yeast: transcript heterogeneity of the ALG7 gene," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 84, pp. 2145–2149, Apr. 1987.

[218] R. J. Kaufman, "Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls," *Genes Dev.*, vol. 13, pp. 1211–1233, 15 May 1999.

[219] L. Williamson, M. Saponaro, S. Boeing, P. East, R. Mitter, T. Kantidakis, G. P. Kelly, A. Lobley, J. Walker, B. Spencer-Dene, M. Howell, A. Stewart, and J. Q. Svejstrup, "UV irradiation induces a non-coding RNA that functionally opposes the protein encoded by the same gene," *Cell*, vol. 168, pp. 843–855.e13, 23 Feb. 2017.

[220] C. D. Marceau, A. S. Puschnik, K. Majzoub, Y. S. Ooi, S. M. Brewer, G. Fuchs, K. Swaminathan, M. A. Mata, J. E. Elias, P. Sarnow, and J. E. Carette, "Genetic dissection of flaviviridae host factors through genome-scale CRISPR screens," *Nature*, vol. 535, pp. 159–163, 7 July 2016.

[221] J. A. Chasis and N. Mohandas, "Red blood cell glycophorins," *Blood*, vol. 80, pp. 1869–1879, 15 Oct. 1992.

[222] IBGRL, "IBGRL-Research products," tech. rep., Bristol Institute for Transfusion Sciences, 1999.

[223] C. G. Gahmberg, M. Ekblom, and L. C. Andersson, "Differentiation of human erythroid cells is associated with increased o-glycosylation of the major sialoglycoprotein, glycophorin a," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 81, pp. 6752–6756, Nov. 1984.

[224] R. A. Anderson and V. T. Marchesi, "Regulation of the association of membrane skeletal protein 4.1 with glycophorin by a polyphosphoinositide," *Nature*, vol. 318, no. 6043, pp. 295–298, 1985.

[225] X. An, M. Salomao, X. Guo, W. Gratzer, and N. Mohandas, "Tropomyosin modulates erythrocyte membrane stability," *Blood*, vol. 109, pp. 1284–1288, 1 Feb. 2007.

[226] I. Rouiller, X.-P. Xu, K. J. Amann, C. Egile, S. Nickell, D. Nicastro, R. Li, T. D. Pollard, N. Volkmann, and D. Hanein, "The structural basis of actin filament branching by the arp2/3 complex," *J. Cell Biol.*, vol. 180, pp. 887–895, 10 Mar. 2008.

[227] "Introduction to essentials of cell biology | learn science at scitable." http://www.nature.com/scitable/ebooks/essentials-of-cell-biology-14749010. Accessed: 2017-8-17.

[228] K. Burridge and M. Chrzanowska-Wodnicka, "Focal adhesions, contractility, and signaling," *Annu. Rev. Cell Dev. Biol.*, vol. 12, pp. 463–518, 1996.

[229] C. Brakebusch and R. Fässler, "The integrin-actin connection, an eternal love affair," *EMBO J.*, vol. 22, pp. 2324–2333, 15 May 2003.

[230] A. Byron, J. D. Humphries, J. A. Askari, S. E. Craig, A. Paul Mould, and M. J. Humphries, "Anti-integrin monoclonal antibodies," *J. Cell Sci.*, vol. 122, pp. 4009–4011, 15 Nov. 2009.

[231] N. E. Sanjana, O. Shalem, and F. Zhang, "Improved vectors and genome-wide libraries for CRISPR screening," *Nat. Methods*, vol. 11, pp. 783–784, Aug. 2014.

[232] H. Ma, Y. Dang, Y. Wu, G. Jia, E. Anaya, J. Zhang, S. Abraham, J.-G. Choi, G. Shi, L. Qi, N. Manjunath, and H. Wu, "A CRISPR-Based screen identifies genes essential for West-Nile-Virus-Induced cell death," *Cell Rep.*, vol. 12, pp. 673–683, July 2015.

[233] T. Hart, A. H. Y. Tong, K. Chan, J. Van Leeuwen, A. Seetharaman, M. Aregger, M. Chandrashekhar, N. Hustedt, S. Seth, A. Noonan, A. Habsid, O. Sizova, L. Nedyalkova, R. Climie, L. Tworzyanski, K. Lawson, M. A. Sartori, S. Alibeh, D. Tieu, S. Masud, P. Mero, A. Weiss, K. R. Brown, M. Usaj, M. Billmann, M. Rahman, M. Constanzo, C. L. Myers, B. J. Andrews, C. Boone, D. Durocher, and J. Moffat, "Evaluation and design of Genome-Wide CRISPR/SpCas9 knockout screens," *G3*, vol. 7, pp. 2719–2727, Aug. 2017.

[234] S. H. Ong, Y. Li, H. Koike-Yusa, and K. Yusa, "Optimised metrics for CRISPR-KO screens with second-generation gRNA libraries," *Scientific Reports*, vol. 7, Aug. 2017.

[235] A. Zotova, I. Zotov, A. Filatov, and D. Mazurov, "Determining antigen specificity of a monoclonal antibody using genome-scale CRISPR-Cas9 knockout library," *J. Immunol. Methods*, vol. 439, pp. 8–14, Dec. 2016.

[236] A. S. Puschnik, K. Majzoub, Y. S. Ooi, and J. E. Carette, “A CRISPR toolbox to study virus-host interactions,” *Nat. Rev. Microbiol.*, vol. 15, pp. 351–364, June 2017.

[237] J. G. Beeson, D. R. Drew, M. J. Boyle, G. Feng, F. J. I. Fowkes, and J. S. Richards, “Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria,” *FEMS Microbiol. Rev.*, vol. 40, pp. 343–372, May 2016.

[238] C. Crosnier, M. Wanaguru, B. McDade, F. H. Osier, K. Marsh, J. C. Rayner, and G. J. Wright, “A library of functional recombinant cell-surface and secreted p. falciparum merozoite proteins,” *Mol. Cell. Proteomics*, vol. 12, pp. 3976–3986, Dec. 2013.

[239] Z. A. Zenonos, J. C. Rayner, and G. J. Wright, “Towards a comprehensive plasmodium falciparum merozoite cell surface and secreted recombinant protein library,” *Malar. J.*, vol. 13, p. 93, 12 Mar. 2014.

[240] A. J. Perrin, S. J. Bartholdson, and G. J. Wright, “P-selectin is a host receptor for plasmodium MSP7 ligands,” *Malar. J.*, vol. 14, p. 238, 5 June 2015.

[241] A. F. Cowman, D. Berry, and J. Baum, “The cellular and molecular basis for malaria parasite invasion of the human red blood cell,” *J. Cell Biol.*, vol. 198, pp. 961–971, 17 Sept. 2012.

[242] K. S. Reddy, E. Amlabu, A. K. Pandey, P. Mitra, V. S. Chauhan, and D. Gaur, “Multiprotein complex between the GPI-anchored CyRPA with PfRH5 and PfRipr is crucial for plasmodium falciparum erythrocyte invasion,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, pp. 1179–1184, 27 Jan. 2015.

[243] L. Y. Bustamante, G. T. Powell, Y.-C. Lin, M. D. Macklin, N. Cross, A. Kemp, P. Cawkill, T. Sanderson, C. Crosnier, N. Muller-Sienerth, O. K. Doumbo, B. Traore, P. D. Crompton, P. Cicuta, T. M. Tran, G. J. Wright, and J. C. Rayner, “Synergistic malaria vaccine combinations identified by systematic antigen screening,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 45, pp. 12045–12050, 2017.

[244] J. S. Richards and J. G. Beeson, “The future for blood-stage vaccines against malaria,” *Immunol. Cell Biol.*, vol. 87, pp. 377–390, July 2009.

[245] L. Brass, “Understanding and evaluating platelet function,” *Hematology Am. Soc. Hematol. Educ. Program*, vol. 2010, pp. 387–396, 2010.

[246] K. E. Wright, K. A. Hjerrild, J. Bartlett, A. D. Douglas, J. Jin, R. E. Brown, J. J. Illingworth, R. Ashfield, S. B. Clemmensen, W. A. de Jongh, S. J. Draper, and M. K. Higgins, “Structure of malaria invasion protein RH5 with erythrocyte basigin and blocking antibodies,” *Nature*, vol. 515, pp. 427–430, 20 Nov. 2014.

[247] S. Kamiyama, T. Suda, R. Ueda, M. Suzuki, R. Okubo, N. Kikuchi, Y. Chiba, S. Goto, H. Toyoda, K. Saigo, M. Watanabe, H. Narimatsu,

Y. Jigami, and S. Nishihara, “Molecular cloning and identification of 3’-phosphoadenosine 5’-phosphosulfate transporter,” *J. Biol. Chem.*, vol. 278, pp. 25958–25963, 11 July 2003.

[248] J. Baum, L. Chen, J. Healer, S. Lopaticki, M. Boyle, T. Triglia, F. Ehlgen, S. a. Ralph, J. G. Beeson, and A. F. Cowman, “Reticulocyte-binding protein homologue 5 - an essential adhesin involved in invasion of human erythrocytes by plasmodium falciparum,” *Int. J. Parasitol.*, vol. 39, pp. 371–380, Feb. 2009.

[249] J. D. Esko, K. Kimata, and U. Lindahl, “Proteoglycans and sulfated glycosaminoglycans,” in *Essentials of Glycobiology* (A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds.), Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 20 Mar. 2010.

[250] M. Bernfield, M. Götte, P. W. Park, O. Reizes, M. L. Fitzgerald, J. Lincecum, and M. Zako, “Functions of cell surface heparan sulfate proteoglycans,” *Annu. Rev. Biochem.*, vol. 68, pp. 729–777, 1 June 1999.

[251] F. Foulquier, M. Amyere, J. Jaeken, R. Zeevaert, E. Schollen, V. Race, R. Bammens, W. Morelle, C. Rosnoblet, D. Legrand, D. Demaegd, N. Buist, D. Cheillan, N. Guffon, P. Morsomme, W. Annaert, H. H. Freeze, E. Van Schaftingen, M. Vikkula, and G. Matthijs, “TMEM165 deficiency causes a congenital disorder of glycosylation,” *Am. J. Hum. Genet.*, vol. 91, pp. 15–26, 13 July 2012.

[252] A. Nikolaev, T. McLaughlin, D. D. M. O’Leary, and M. Tessier-Lavigne, “APP binds DR6 to trigger axon pruning and neuron death via distinct caspases,” *Nature*, vol. 457, pp. 981–989, 19 Feb. 2009.

[253] R. J. O’Brien and P. C. Wong, “Amyloid precursor protein processing and alzheimer’s disease,” *Annu. Rev. Neurosci.*, vol. 34, pp. 185–204, 2011.

[254] E. C. J. M. de Vet, S. A. B. Newland, P. A. Lyons, B. Aguado, and R. D. Campbell, “The cell surface receptor g6b, a member of the immunoglobulin superfamily, binds heparin,” *FEBS Lett.*, vol. 579, pp. 2355–2358, 25 Apr. 2005.

[255] S. A. Newland, I. C. Macaulay, A. R. Floto, E. C. de Vet, W. H. Ouwehand, N. A. Watkins, P. A. Lyons, and D. R. Campbell, “The novel inhibitory receptor G6B is expressed on the surface of platelets and attenuates platelet function in vitro,” *Blood*, vol. 109, pp. 4806–4809, 1 June 2007.

[256] S. Tahara-Hanaoka, K. Shibuya, Y. Onoda, H. Zhang, S. Yamazaki, A. Miyamoto, S.-I. Honda, L. L. Lanier, and A. Shibuya, “Functional characterization of DNAM-1 (CD226) interaction with its ligands PVR (CD155) and nectin-2 (PRR-2/CD112),” *Int. Immunol.*, vol. 16, pp. 533–538, Apr. 2004.

[257] A. D. Bergemann, H. J. Cheng, R. Brambilla, R. Klein, and J. G. Flanagan, “ELF-2, a new member of the eph ligand family, is segmentally

expressed in mouse embryos in the region of the hindbrain and newly forming somites," *Mol. Cell. Biol.*, vol. 15, pp. 4921–4929, Sept. 1995.

[258] A. A. Boucard, S. Maxeiner, and T. C. Südhof, "Latrophilins function as heterophilic cell-adhesion molecules by binding to teneurins: regulation by alternative splicing," *J. Biol. Chem.*, vol. 289, pp. 387–402, 3 Jan. 2014.

[259] I. Melero, W. W. Shuford, S. A. Newby, A. Aruffo, J. A. Ledbetter, K. E. Hellström, R. S. Mittler, and L. Chen, "Monoclonal antibodies against the 4-1BB t-cell activation molecule eradicate established tumors," *Nat. Med.*, vol. 3, pp. 682–685, June 1997.

[260] A. Makkouk, C. Chester, and H. E. Kohrt, "Rationale for anti-CD137 cancer immunotherapy," *Eur. J. Cancer*, vol. 54, pp. 112–119, Feb. 2016.

[261] H. R. Salih, S. G. Kosowski, V. F. Haluska, G. C. Starling, D. T. Loo, F. Lee, A. A. Aruffo, P. A. Trail, and P. A. Kiener, "Constitutive expression of functional 4-1BB (CD137) ligand on carcinoma cells," *J. Immunol.*, vol. 165, pp. 2903–2910, 1 Sept. 2000.

[262] A. Kulane, H. P. Ekre, P. Perlmann, L. Rombo, M. Wahlgren, and B. Wahlin, "Effect of different fractions of heparin on plasmodium falciparum merozoite invasion of red blood cells in vitro," *Am. J. Trop. Med. Hyg.*, vol. 46, pp. 589–594, May 1992.

[263] M. J. Boyle, J. S. Richards, P. R. Gilson, W. Chai, and J. G. Beeson, "Interactions with heparin-like molecules during erythrocyte invasion by plasmodium falciparum merozoites," *Blood*, vol. 115, pp. 4559–4568, 3 June 2010.

[264] G. E. Weiss, P. R. Gilson, T. Taechalertpaisarn, W.-H. Tham, N. W. M. de Jong, K. L. Harvey, F. J. I. Fowkes, P. N. Barlow, J. C. Rayner, G. J. Wright, A. F. Cowman, and B. S. Crabb, "Revealing the sequence and resulting cellular morphology of Receptor-Ligand interactions during plasmodium falciparum invasion of erythrocytes," *PLoS Pathog.*, vol. 11, p. e1004670, 27 Feb. 2015.

[265] S. M. de Boer, J. Kortekaas, C. A. M. de Haan, P. J. M. Rottier, R. J. M. Moormann, and B. J. Bosch, "Heparan sulfate facilitates rift valley fever virus entry into the cell," *J. Virol.*, vol. 86, pp. 13767–13771, Dec. 2012.

[266] D. M. Rosmarin, J. E. Carette, A. J. Olive, M. N. Starnbach, T. R. Brummelkamp, and H. L. Ploegh, "Attachment of chlamydia trachomatis L2 to host cells requires sulfation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, pp. 10059–10064, 19 June 2012.

[267] V. B. Carruthers, S. Håkansson, O. K. Giddings, and L. D. Sibley, "Toxoplasma gondii uses sulfated proteoglycans for substrate and host cell attachment," *Infect. Immun.*, vol. 68, pp. 4005–4011, July 2000.

[268] A. Barragan, V. Fernandez, Q. Chen, A. von Euler, M. Wahlgren, and D. Spillmann, "The duffy-binding-like domain 1 of plasmodium falciparum

erythrocyte membrane protein 1 (PfEMP1) is a heparan sulfate ligand that requires 12 mers for binding," *Blood*, vol. 95, pp. 3594–3599, 1 June 2000.

[269] K. Kobayashi, K. Kato, T. Sugi, H. Takemae, K. Pandey, H. Gong, Y. Tohya, and H. Akashi, "Plasmodium falciparum BAEBL binds to heparan sulfate proteoglycans on the human erythrocyte surface," *J. Biol. Chem.*, vol. 285, pp. 1716–1725, 15 Jan. 2010.

[270] K. Kobayashi, R. Takano, H. Takemae, T. Sugi, A. Ishiwa, H. Gong, F. C. Recuenco, T. Iwanaga, T. Horimoto, H. Akashi, and K. Kato, "Analyses of interactions between heparin and the apical surface proteins of plasmodium falciparum," *Sci. Rep.*, vol. 3, p. 3178, 11 Nov. 2013.

[271] J. D. Esko and R. J. Linhardt, "Proteins that bind sulfated glycosaminoglycans," in *Essentials of Glycobiology* (A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds.), Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 20 Mar. 2010.

[272] W.-H. Tham, D. W. Wilson, S. Lopaticki, C. Q. Schmidt, P. B. Tetteh-Quarcoo, P. N. Barlow, D. Richard, J. E. Corbin, J. G. Beeson, and A. F. Cowman, "Complement receptor 1 is the host erythrocyte receptor for plasmodium falciparum PfRh4 invasion ligand," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, pp. 17327–17332, 5 Oct. 2010.

[273] D. M. Ornitz, "FGFs, heparan sulfate and FGFRs: complex interactions essential for development," *Bioessays*, vol. 22, pp. 108–112, Feb. 2000.

[274] S. Rashi-Elkeles, R. Elkon, S. Shavit, Y. Lerenthal, C. Linhart, A. Kupershtein, N. Amariglio, G. Rechavi, R. Shamir, and Y. Shiloh, "Transcriptional modulation induced by ionizing radiation: p53 remains a central player," *Mol. Oncol.*, vol. 5, pp. 336–348, 1 Aug. 2011.

[275] D. R. Burt and G. L. Kamatchi, "GABAA receptor subtypes: from pharmacology to molecular biology," *FASEB J.*, vol. 5, pp. 2916–2923, Nov. 1991.

[276] N. G. Bowery, B. Bettler, W. Froestl, J. P. Gallagher, F. Marshall, M. Raiteri, T. I. Bonner, and S. J. Enna, "International union of pharmacology. XXXIII. mammalian gamma-aminobutyric acid(b) receptors: structure and function," *Pharmacol. Rev.*, vol. 54, pp. 247–264, June 2002.

[277] N. G. Bowery and S. J. Enna, "gamma-aminobutyric acid(b) receptors: first of the functional metabotropic heterodimers," *J. Pharmacol. Exp. Ther.*, vol. 292, pp. 2–7, Jan. 2000.

[278] T. Galvez, B. Duthey, J. Kniazeff, J. Blahos, G. Rovelli, B. Bettler, L. Prézeau, and J. P. Pin, "Allosteric interactions between GB1 and GB2 subunits are required for optimal GABA(B) receptor function," *EMBO J.*, vol. 20, pp. 2152–2159, 1 May 2001.

[279] R. Vigot, S. Barbieri, H. Bräuner-Osborne, R. Turecek, R. Shigemoto, Y.-P. Zhang, R. Luján, L. H. Jacobson, B. Biermann, J.-M. Fritschy, C.-M. Vacher, M. Müller, G. Sansig, N. Guetg, J. F. Cryan, K. Kaupmann, M. Gassmann, T. G. Oertner, and B. Bettler, "Differential compartmentalization and distinct functions of GABAB receptor variants," *Neuron*, vol. 50, pp. 589–601, 18 May 2006.

[280] M. Gassmann and B. Bettler, "Regulation of neuronal GABA(B) receptor functions by subunit composition," *Nat. Rev. Neurosci.*, vol. 13, pp. 380–394, 18 May 2012.

[281] S. Ghose, M. K. Winter, K. E. McCarson, C. A. Tamminga, and S. J. Enna, "The GABAβ receptor as a target for antidepressant drug action," *Br. J. Pharmacol.*, vol. 162, pp. 1–17, Jan. 2011.

[282] S. S. Ferguson, "Evolving concepts in G protein-coupled receptor endocytosis: the role in receptor desensitization and signaling," *Pharmacol. Rev.*, vol. 53, pp. 1–24, Mar. 2001.

[283] M. von Zastrow, "Mechanisms regulating membrane trafficking of G protein-coupled receptors in the endocytic pathway," *Life Sci.*, vol. 74, pp. 217–224, 5 Dec. 2003.

[284] B. P. Fairfax, J. A. Pitcher, M. G. H. Scott, A. R. Calver, M. N. Pangalos, S. J. Moss, and A. Couve, "Phosphorylation and chronic agonist treatment atypically modulate GABAB receptor cell surface stability," *J. Biol. Chem.*, vol. 279, pp. 12565–12573, 26 Mar. 2004.

[285] J. Perroy, L. Adam, R. Qanbar, S. Chénier, and M. Bouvier, "Phosphorylation-independent desensitization of GABA(B) receptor by GRK4," *EMBO J.*, vol. 22, pp. 3816–3824, 1 Aug. 2003.

[286] J. González-Maeso, A. Wise, A. Green, and J. A. Koenig, "Agonist-induced desensitization and endocytosis of heterodimeric GABAB receptors in CHO-K1 cells," *Eur. J. Pharmacol.*, vol. 481, pp. 15–23, 14 Nov. 2003.

[287] S. Laffray, K. Tan, J. Dulluc, R. Bouali-Benazzouz, A. R. Calver, F. Nagy, and M. Landry, "Dissociation and trafficking of rat GABAB receptor heterodimer upon chronic capsaicin stimulation," *Eur. J. Neurosci.*, vol. 25, pp. 1402–1416, Mar. 2007.

[288] K. J. Vargas, M. Terunuma, J. A. Tello, M. N. Pangalos, S. J. Moss, and A. Couve, "The availability of surface GABAB receptors is independent of γ-Aminobutyric acid but controlled by glutamate in central neurons," *J. Biol. Chem.*, vol. 283, pp. 24641–24648, 5 Sept. 2008.

[289] T. Grampp, K. Sauter, B. Markovic, and D. Benke, "Gamma-aminobutyric acid type B receptors are constitutively internalized via the clathrin-dependent pathway and targeted to lysosomes for degradation," *J. Biol. Chem.*, vol. 282, pp. 24157–24165, 17 Aug. 2007.

[290] S. Balasubramanian, J. A. Teissére, D. V. Raju, and R. A. Hall, "Hetero-oligomerization between GABAA and GABAB receptors regulates GABAB receptor trafficking," *J. Biol. Chem.*, vol. 279, pp. 18840–18850, 30 Apr. 2004.

[291] T. Grampp, V. Notz, I. Broll, N. Fischer, and D. Benke, "Constitutive, agonist-accelerated, recycling and lysosomal degradation of GABA(B) receptors in cortical neurons," *Mol. Cell. Neurosci.*, vol. 39, pp. 628–637, Dec. 2008.

[292] S. Hannan, M. E. Wilkins, E. Dehghani-Tafti, P. Thomas, S. M. Baddeley, and T. G. Smart, "Gamma-aminobutyric acid type B (GABA(B)) receptor internalization is regulated by the R2 subunit," *J. Biol. Chem.*, vol. 286, pp. 24324–24335, 8 July 2011.

[293] M. Terunuma, K. J. Vargas, M. E. Wilkins, O. A. Ramírez, M. Jaureguiberry-Bravo, M. N. Pangalos, T. G. Smart, S. J. Moss, and A. Couve, "Prolonged activation of NMDA receptors promotes dephosphorylation and alters postendocytic sorting of GABAB receptors," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, pp. 13918–13923, Aug. 2010.

[294] S. K. Shenoy, P. H. McDonald, T. A. Kohout, and R. J. Lefkowitz, "Regulation of receptor fate by ubiquitination of activated beta 2-adrenergic receptor and beta-arrestin," *Science*, vol. 294, pp. 1307–1313, Nov. 2001.

[295] S. Sarker, K. Xiao, and S. K. Shenoy, "A tale of two sites: How ubiquitination of a G protein-coupled receptor is coupled to its lysosomal trafficking from distinct receptor domains," *Commun. Integr. Biol.*, vol. 4, pp. 528–531, Sept. 2011.

[296] J. F. Nabhan, H. Pan, and Q. Lu, "Arrestin domain-containing protein 3 recruits the NEDD4 E3 ligase to mediate ubiquitination of the beta2-adrenergic receptor," *EMBO Rep.*, vol. 11, pp. 605–611, Aug. 2010.

[297] M. R. Dores and J. Trejo, "Atypical regulation of G protein-coupled receptor intracellular trafficking by ubiquitination," *Curr. Opin. Cell Biol.*, vol. 27, pp. 44–50, Apr. 2014.

[298] K. Zemoura, C. Trümpler, and D. Benke, "Lys-63-linked ubiquitination of γ-Aminobutyric acid (GABA), type b1, at multiple sites by the E3 ligase mind bomb-2 targets GABAB receptors to lysosomal degradation," *J. Biol. Chem.*, vol. 291, pp. 21682–21693, Oct. 2016.

[299] N. Lahaie, M. Kralikova, L. Prézeau, J. Blahos, and M. Bouvier, "Post-endocytotic deubiquitination and degradation of the metabotropic γ-Aminobutyric acid receptor by the ubiquitin-specific protease 14," *J. Biol. Chem.*, vol. 291, pp. 7156–7170, Mar. 2016.

[300] P. Ghosh, N. M. Dahms, and S. Kornfeld, "Mannose 6-phosphate receptors: new twists in the tale," *Nat. Rev. Mol. Cell Biol.*, vol. 4, pp. 202–212, Mar. 2003.

[301] A. Varki and S. Kornfeld, "P-type lectins," in *Essentials of Glycobiology* (A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds.), Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 20 Mar. 2010.

[302] J. Brown, E. Y. Jones, and B. E. Forbes, "Interactions of IGF-II with the IGF2R/cation-independent mannose-6-phosphate receptor mechanism and biological outcomes," *Vitam. Horm.*, vol. 80, pp. 699–719, 2009.

[303] N. M. Dahms and M. K. Hancock, "P-type lectins," *Biochim. Biophys. Acta*, vol. 1572, pp. 317–340, 19 Sept. 2002.

[304] P. A. Dennis and D. B. Rifkin, "Cellular activation of latent transforming growth factor beta requires binding to the cation-independent mannose 6-phosphate/insulin-like growth factor type II receptor," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 88, pp. 580–584, 15 Jan. 1991.

[305] F. Blanchard, S. Raher, L. Duplomb, P. Vusio, V. Pitard, J. L. Taupin, J. F. Moreau, B. Hoflack, S. Minvielle, Y. Jacques, and A. Godard, "The mannose 6-phosphate/insulin-like growth factor II receptor is a nanomolar affinity receptor for glycosylated human leukemia inhibitory factor," *J. Biol. Chem.*, vol. 273, pp. 20886–20893, 14 Aug. 1998.

[306] F. Blanchard, L. Duplomb, S. Raher, P. Vusio, B. Hoflack, Y. Jacques, and A. Godard, "Mannose 6-Phosphate/Insulin-like growth factor II receptor mediates internalization and degradation of leukemia inhibitory factor but not signal transduction," *J. Biol. Chem.*, vol. 274, pp. 24685–24693, 27 Aug. 1999.

[307] H. Ikushima, Y. Munakata, T. Ishii, S. Iwata, M. Terashima, H. Tanaka, S. F. Schlossman, and C. Morimoto, "Internalization of CD26 by mannose 6-phosphate/insulin-like growth factor II receptor contributes to T cell activation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, pp. 8439–8444, 18 July 2000.

[308] B. Motyka, G. Korbutt, M. J. Pinkoski, J. A. Heibein, A. Caputo, M. Hobman, M. Barry, I. Shostak, T. Sawchuk, C. F. B. Holmes, J. Gauldie, and R. C. Bleackley, "Mannose 6-Phosphate/Insulin-like growth factor II receptor is a death receptor for granzyme B during cytotoxic TCell–Induced apoptosis," *Cell*, vol. 103, pp. 491–500, 27 Oct. 2000.

[309] C. R. Brunetti, R. L. Burke, S. Kornfeld, W. Gregory, F. R. Masiarz, K. S. Dingwell, and D. C. Johnson, "Herpes simplex virus glycoprotein D acquires mannose 6-phosphate residues and binds to mannose 6-phosphate receptors," *J. Biol. Chem.*, vol. 269, pp. 17067–17074, 24 June 1994.

[310] R. G. MacDonald, S. R. Pfeffer, L. Coussens, M. A. Tepper, C. M. Brocklebank, J. E. Mole, J. K. Anderson, E. Chen, M. P. Czech, and A. Ullrich, "A single receptor binds both insulin-like growth factor II and mannose-6-phosphate," *Science*, vol. 239, pp. 1134–1137, 4 Mar. 1988.

[311] P. Y. Tong, S. E. Tollefsen, and S. Kornfeld, "The cation-independent mannose 6-phosphate receptor binds insulin-like growth factor II," *J. Biol. Chem.*, vol. 263, pp. 2585–2588, 25 Feb. 1988.

[312] D. O. Morgan, J. C. Edman, D. N. Standring, V. A. Fried, M. C. Smith, R. A. Roth, and W. J. Rutter, "Insulin-like growth factor II receptor as a multifunctional binding protein," *Nature*, vol. 329, no. 6137, pp. 301–307, 1987.

[313] J. X. Kang, Y. Li, and A. Leaf, "Mannose-6-phosphate/insulin-like growth factor-II receptor is a receptor for retinoic acid," *Proceedings of the National Academy of Sciences*, vol. 94, pp. 13671–13676, 9 Dec. 1997.

[314] A. Nykjaer, E. I. Christensen, H. Vorum, H. Hager, C. M. Petersen, H. Røigaard, H. Y. Min, F. Vilhardt, L. B. Møller, S. Kornfeld, and J. Gliemann, "Mannose 6-phosphate/insulin-like growth factor-II receptor targets the urokinase receptor to lysosomes via a novel binding interaction," *J. Cell Biol.*, vol. 141, pp. 815–828, 4 May 1998.

[315] S. Godár, V. Horejsi, U. H. Weidle, B. R. Binder, C. Hansmann, and H. Stockinger, "M6P/IGFII-receptor complexes urokinase receptor and plasminogen for activation of transforming growth factor-beta1," *Eur. J. Immunol.*, vol. 29, pp. 1004–1013, Mar. 1999.

[316] M. Forgac, "Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology," *Nat. Rev. Mol. Cell Biol.*, vol. 8, pp. 917–929, Nov. 2007.

[317] J. A. Solinger and A. Spang, "Tethering complexes in the endocytic pathway: CORVET and HOPS," *FEBS J.*, vol. 280, pp. 2743–2757, 1 June 2013.

[318] B. Reaves and G. Banting, "Vacuolar ATPase inactivation blocks recycling to the trans-golgi network from the plasma membrane," *FEBS Lett.*, vol. 345, pp. 61–66, 23 May 1994.

[319] M. Merkulova, T. G. Păunescu, A. Azroyan, V. Marshansky, S. Breton, and D. Brown, "Mapping the h(+) (V)-ATPase interactome: identification of proteins involved in trafficking, folding, assembly and phosphorylation," *Sci. Rep.*, vol. 5, p. 14827, 7 Oct. 2015.

[320] C. Hawkes and S. Kar, "The insulin-like growth factor-II/mannose-6-phosphate receptor: structure, distribution and function in the central nervous system," *Brain Res. Brain Res. Rev.*, vol. 44, pp. 117–140, Mar. 2004.

[321] M. Jadot, W. M. Canfield, W. Gregory, and S. Kornfeld, "Characterization of the signal for rapid internalization of the bovine mannose 6-phosphate/insulin-like growth factor-II receptor," *J. Biol. Chem.*, vol. 267, pp. 11069–11077, 5 June 1992.

[322] P. Ramoino, P. Bianchini, A. Diaspro, and C. Usai, "Visualizing GABAB receptor internalization and intracellular trafficking," in *Visualization Techniques: From Immunohistochemistry to Magnetic Resonance Imaging* (E. Badoer, ed.), pp. 71–95, Totowa, NJ: Humana Press, 2012.

[323] R. C. Orchard, C. B. Wilen, J. G. Doench, M. T. Baldridge, B. T. McCune, Y.-C. J. Lee, S. Lee, S. M. Pruett-Miller, C. A. Nelson, D. H. Fremont, and H. W. Virgin, "Discovery of a proteinaceous cellular receptor for a norovirus," *Science*, vol. 353, pp. 933–936, 26 Aug. 2016.

[324] K. Haga, A. Fujimoto, R. Takai-Todaka, M. Miki, Y. H. Doan, K. Murakami, M. Yokoyama, K. Murata, A. Nakanishi, and K. Katayama, "Functional receptor molecules CD300lf and CD300ld within the CD300 family enable murine noroviruses to infect cells," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, pp. E6248–E6255, 11 Oct. 2016.

[325] L. M. Duncan, R. T. Timms, E. Zavodszky, F. Cano, G. Dougan, F. Randow, and P. J. Lehner, "Fluorescence-Based phenotypic selection allows forward genetic screens in haploid human cells," *PLoS One*, vol. 7, p. e39651, 22 June 2012.

[326] P.-H. Kuhn, M. Voss, M. Haug-Kröper, B. Schröder, U. Schepers, S. Bräse, C. Haass, S. F. Lichtenthaler, and R. Fluhrer, "Secretome analysis identifies novel signal peptide Peptidase-Like 3 (SPPL3) substrates and reveals a role of SPPL3 in multiple golgi glycosylation pathways," *Mol. Cell. Proteomics*, vol. 14, pp. 1584–1598, 1 June 2015.

[327] M. A. Cohn, Y. Kee, W. Haas, S. P. Gygi, and A. D. D'Andrea, "UAF1 is a subunit of multiple deubiquitinating enzyme complexes," *J. Biol. Chem.*, vol. 284, pp. 5343–5351, 20 Feb. 2009.

[328] M. A. Cohn, P. Kowal, K. Yang, W. Haas, T. T. Huang, S. P. Gygi, and A. D. D'Andrea, "A UAF1-containing multisubunit protein complex regulates the fanconi anemia pathway," *Mol. Cell*, vol. 28, pp. 786–797, 14 Dec. 2007.

[329] M. E. Wilkins, X. Li, and T. G. Smart, "Tracking cell surface GABAB receptors using an alpha-bungarotoxin tag," *J. Biol. Chem.*, vol. 283, pp. 34745–34752, 12 Dec. 2008.

[330] S. Hannan, M. E. Wilkins, and T. G. Smart, "Sushi domains confer distinct trafficking profiles on GABAB receptors," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, pp. 12171–12176, 24 July 2012.

[331] J. S. Bonifacino and L. M. Traub, "Signals for sorting of transmembrane proteins to endosomes and lysosomes," *Annu. Rev. Biochem.*, vol. 72, pp. 395–447, 6 Mar. 2003.

[332] S. B. Hannan, *Cell Surface Mobility of GABAB Receptors*. PhD thesis, University College London, 2011.

[333] I. L. Arancibia-Cárcamo, B. P. Fairfax, S. J. Moss, and J. T. Kittler, "Studying the localization, surface stability and endocytosis of neurotransmitter

receptors by antibody labeling and biotinylation approaches," in *The Dynamic Synapse: Molecular Methods in Ionotropic Receptor Biology* (J. T. Kittler and S. J. Moss, eds.), Boca Raton (FL): CRC Press/Taylor & Francis, 5 Jan. 2011.

APPENDIX A

Table A.1 Primers used for cloning IGF2R into protein expression vectors Fwd=Forward, Rev=Reverse

Name	Sequence 5'-3' (restriction sites are indicated in bold)	Purpose	Restriction site
IGF2R-Fwd	GACT**GCGGCCGC**CACCATG GGGGCCGCCGCCGGCCG-GAG	Fwd primer for amplification of IGF2R ectodomain	NotI
IGF2R-Rev	CAGT**GGCGCGCC**GACTGCC TGGCTCCGTTCTGAC	Rev primer for amplification of IGF2R ectodomain	AscI
OL497	TGAGATCCAGCTGTTGGGGT	Fwd sequencing primer	-
OL4006	TCCCTGCAGGCTTTCCTCT CCAAGGTTGAG	Rev sequencing primer	-

Table A.2 Sequences of guide RNAs used in this study

Gene	Oligos: Sense (S) and antisense (AS)- 5'-3'	gRNA ID from Human v1 library
BSG	S: CACCGTCGTCAGAACACATCAACG	BSG_CCDS12032.1_ex2_19:580670
	AS: AAACCGTTGATGTGTTCTGACGAC	-580693:+_9-4
SDC1	S: CACCGATGAGACCTCAACCCCTGC	SDC1_CCDS1697.1_ex2_2:20403694
	AS: AAACGCAGGGGTTGAGGTCTCATC	-20403717:-_5-3
CD55	S: CACCGGGCCGTACAAGTTTTCCCG	CD55_CCDS31006.1_ex1_1:207495775
	AS: AAACCGGGAAAACTTGTACGGCCC	-207495798:+_5-1
CD44	S: CACCGTACAGCATCTCTCGGACGG	CD44_CCDS31455.1_ex1_11:35198176
	AS: AAACCCGTCCGAGAGATGCTGTAC	-35198199:+_5-2
EXTL3	S: CACCGGTCCATCCGTCCGCCAGTG	EXTL3_CCDS6070.1_ex0_8:28574365
	AS: AAACCACTGGCGGACGGATGGACC	-28574388:-_5-3
SLC35B2	S: CACCGTCGGCGCAGCTACGAACAC	SLC35B2_CCDS34462.1_ex0_6:44223027
	AS: AAACGTGTTCGTAGCTGCGCCGAC	-44223050:-_5-1
TP53	S: CACCGCAGTCACAGCACATGACGG	TP53_CCDS11118.1_ex6_17:7578415
	AS: AAACCCGTCATGTGCTGTGACTGC	-7578438:-_5-1
IGF2R(4)	S: CACCGCGGTCACTACGCATTCCAG	IGF2R_CCDS5273.1_ex16_6:160468883
	AS: AAACCTGGAATGCGTAGTGACCGC	-160468906:-_5-4
IGF2R(5)	S: CACCGACAACGACGGATACAGACC	IGF2R_CCDS5273.1_ex19_6:160477535
	AS: AAACGGTCTGTATCCGTCGTTGTC	-160477558:+_5-5
WDR7	S: CACCGTGCGGAATGAATCACTAGC	WDR7_CCDS11962.1_ex7_18:54358952
	AS: AAACGCTAGTGATTCATTCCGCAC	-54358975:-_5-5

Table A.3 Summary of the screening parameters for all screens carried out in this study. The table indicates the total number of cells screened within each library, the day the selection was performed, the number of collected cells, and the sorting threshold each for each indicated staining probe/cell line pairing used.

Cell line	Staining probe	Day of selection	Number of sorted cells in the library (x10^6)	Number of collected cells	Sorting threshold (% of non-binding cells selected)
HEK-293-E	BRIC229	15	100	380,000	0.5
HEK-293-E	BRIC126	9	80	280,000	1
HEK-293-E	B6H12	9	80	200,000	1
HEK-293-E	MEM6/6	16	60	40,000	0.2
HEK-293-E	BRIC5	9	50	120,000	0.8
HEK-293-E	P16	9	60	180,000	1
HEK-293-E	RH5	9	80	200,000	0.7
HEK-293-E	EBA181	14	70	420,000	1
HEK-293-E	MSRP5	14	70	390,000	1
HEK-293-E	APP	9	60	320,000	1.1
HEK-293-E	APLP2	9	70	330,000	1
HEK-293-E	G6B	15	100	800,000	0.75
HEK-293-E	CD226	9	70	340,000	1.7
HEK-293-E	EPHB1	9	70	300,000	1.4
HEK-293-E	LPHN1	14	70	1,200,000	4.3
HEK-293-E	GABBR2	14	125	780,000	2.3
NCI-SNU-1	SERA9	16	50	180,000	1
NCI-SNU-1	TNFRSF9	14	80	600,000	1
HEL	BRIC256	14	50	300,000	1.7

A.1 Protocol for PCR amplification of gRNAs from lentivirally transduced cells for Illumina sequencing.

First PCR: This set up is for analysis of gRNAs from high complexity samples (e.g. control samples).

Step 1: Set up the following PCR master mix and aliquot 50 μL to 36 wells.

Reagent	Volume per reaction	Master mix (x38)
Q5 Hot Start High-Fidelity 2× Master Mix	25 μL	950 μL
Primer (L1/U1) mix (10 μM each)	1 μL	38 μL
Genomic DNA (1 mg/mL)	2 μL	76 μL
H_2O	up to 50 μL	up to 1900 μL

Step 2: Run the following PCR program.

Cycle number	Denature	Annealing	Extension
1	98 °C, 30 s		
2-26	98 °C, 10 s	61 °C, 15 s	72 °C, 20 s
27			72 °C, 2 min

Step 3: Collect 5 μL PCR product from each well (180 μL in total) and purify using Qiagen PCR purification kit. Elute DNA into 50 μL EB buffer and measure DNA concentration.

Second PCR

Step 4: Dilute the first PCR product to 40 pg/μL and set up one PCR reaction per sample. The forward primer (PE 1.0) is common to all samples, whereas the reverse primers are sample specific index primers (refer to table A.4).

Reagent	Volume per reaction
KAPA HiFi HotStart ReadyMix	25 μL
Primer (PE1.0/index primer) mix (5 μM each)	2 μL
First PCR product (40 pg/μL)	5 μL
H_2O	18 μL

Step 5: Run the following PCR program.

Cycle number	Denature	Annealing	Extension
1	98 °C, 30 s		
2-15	98 °C, 10 s	66 °C, 15 s	72 °C, 20 s
16			72 °C, 5 min

Table A.4 Primer sequences for amplifying gRNA and NGS.

Name	Sequence 5'-3' (restriction sites are indicated in bold, * represents phosphorothioate)	Purpose
U1	ACACTCTTTCCCTACACGACGCTCTT-CCGATCTCTTGTGGAAAGGACGAAACA	Forward primer for gRNA amplification
L1	TCGGCATTCCTGCTGAACCGCTCTT-CCGATCTCTAAAGCGCATGCTCCAGAC	Reverse primer for gRNA amplification
PE1.0	AATGATACGGCGACCACCGAGATCTAC-ACTCTTTCCCTACACGACGCTCTTCCGATC*T	Forward primer for Illumina library preparation
iPCRTag1	CAAGCAGAAGACGGCATACGAGATAAC-GTGATGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation
iPCRTag2	CAAGCAGAAGACGGCATACGAGATAAA-CATCGGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation
iPCRTag3	CAAGCAGAAGACGGCATACGAGATATG-CCTAAGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation
iPCRTag4	CAAGCAGAAGACGGCATACGAGATAGT-GGTCAGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation
iPCRTag5	CAAGCAGAAGACGGCATACGAGATACC-ACTGTGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation
iPCRTag6	CAAGCAGAAGACGGCATACGAGATACA-TTGGCGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation
iPCRTag7	CAAGCAGAAGACGGCATACGAGATCAG-ATCTGGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation
iPCRTag8	CAAGCAGAAGACGGCATACGAGATCAT-CAAGTGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation
iPCRTag9	CAAGCAGAAGACGGCATACGAGATCGC-TGATCGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation
iPCRTag10	CAAGCAGAAGACGGCATACGAGATACA-AGCTAGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation
iPCRTag11	CAAGCAGAAGACGGCATACGAGATCTG-TAGCCGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation
iPCRTag12	CAAGCAGAAGACGGCATACGAGATAGT-ACAAGGAGATCGGTCTCGGCATTCCTG-CTGAACCGCTCTTCCGATC*T	Reverse indexing primer for Illumina library preparation

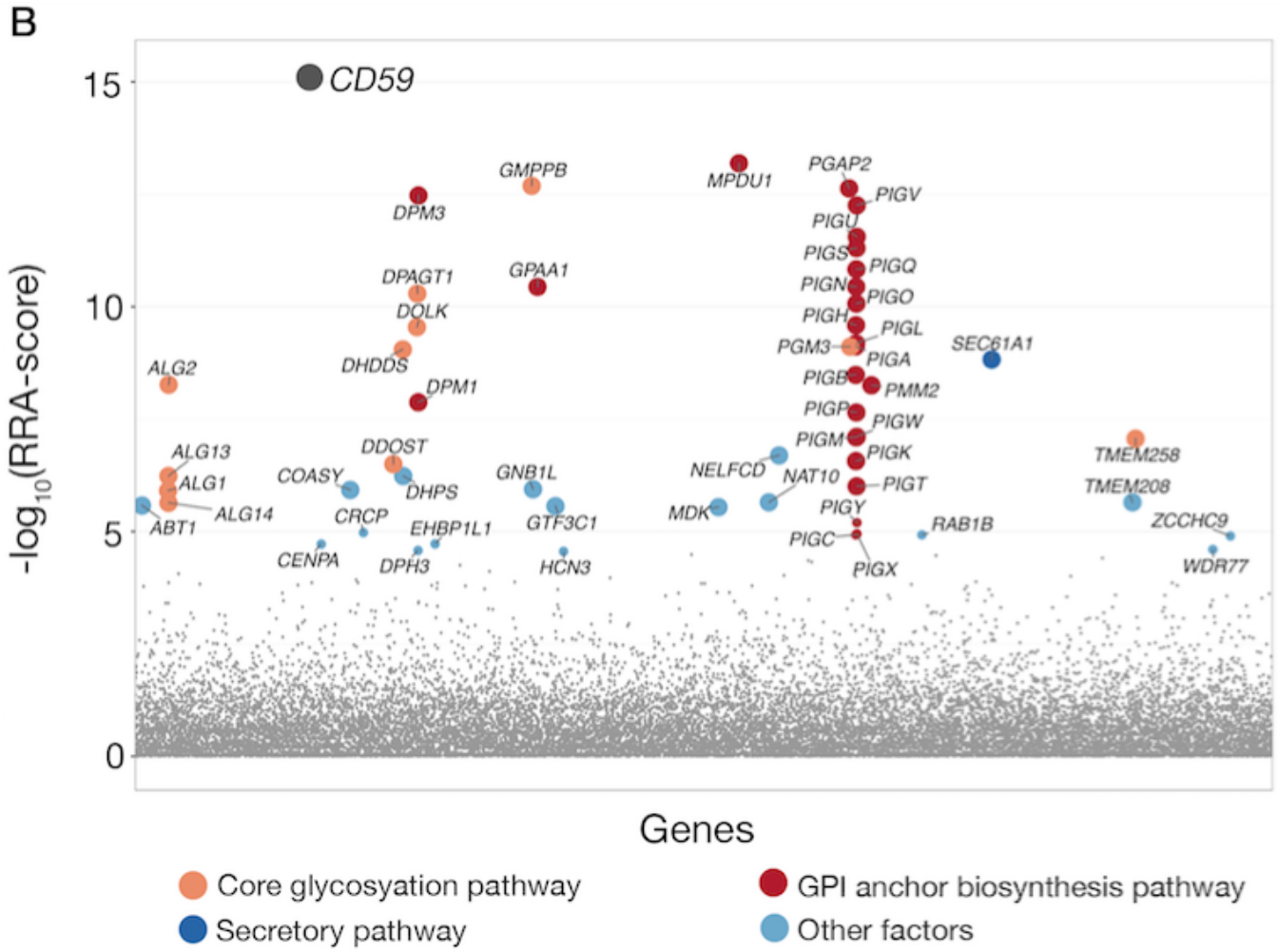


Fig. A.1 Enlarged version of figure 3.10B for better clarity of gene names

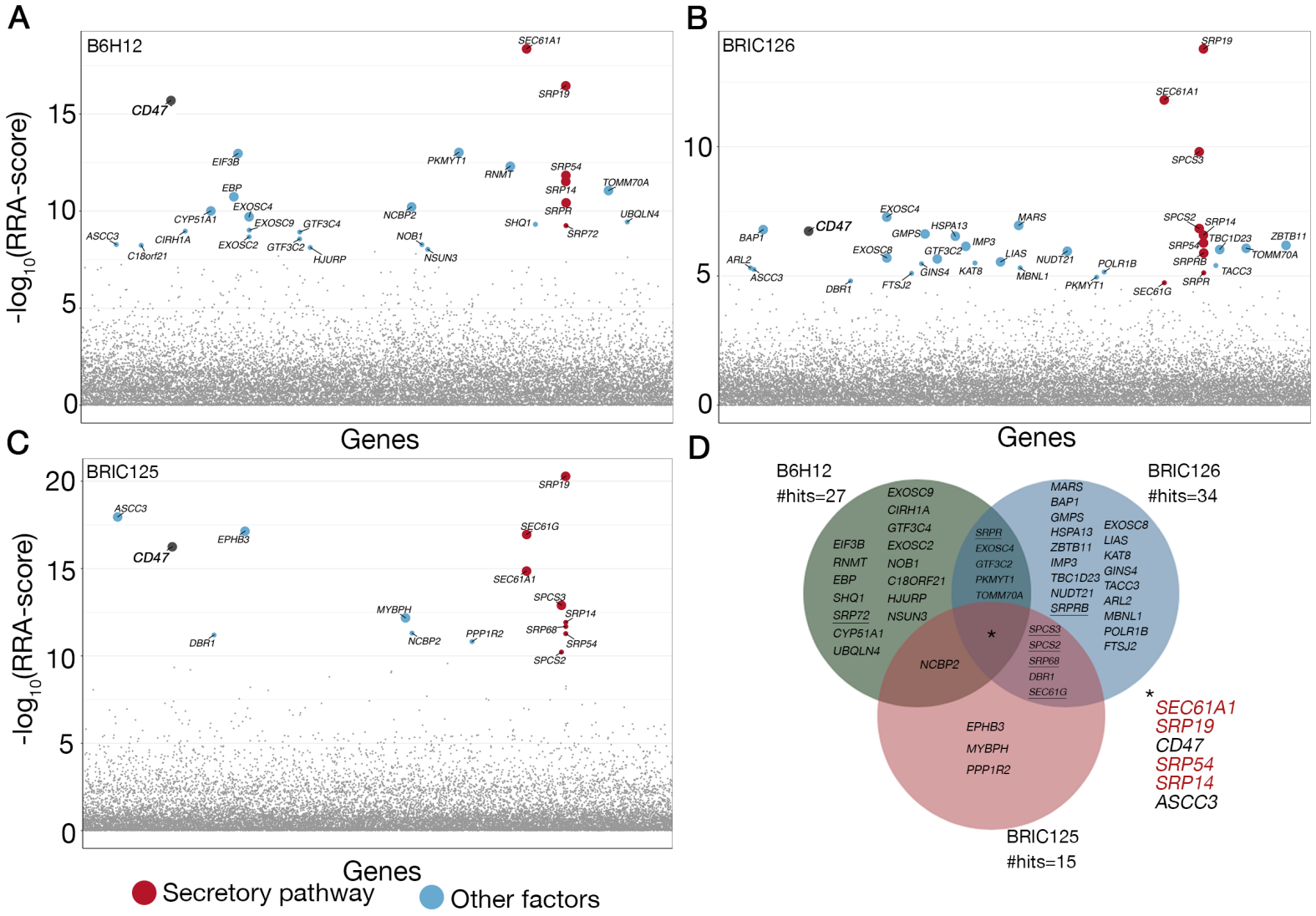


Fig. A.2 Enlarged version of figure 3.12 for better clarity of gene names

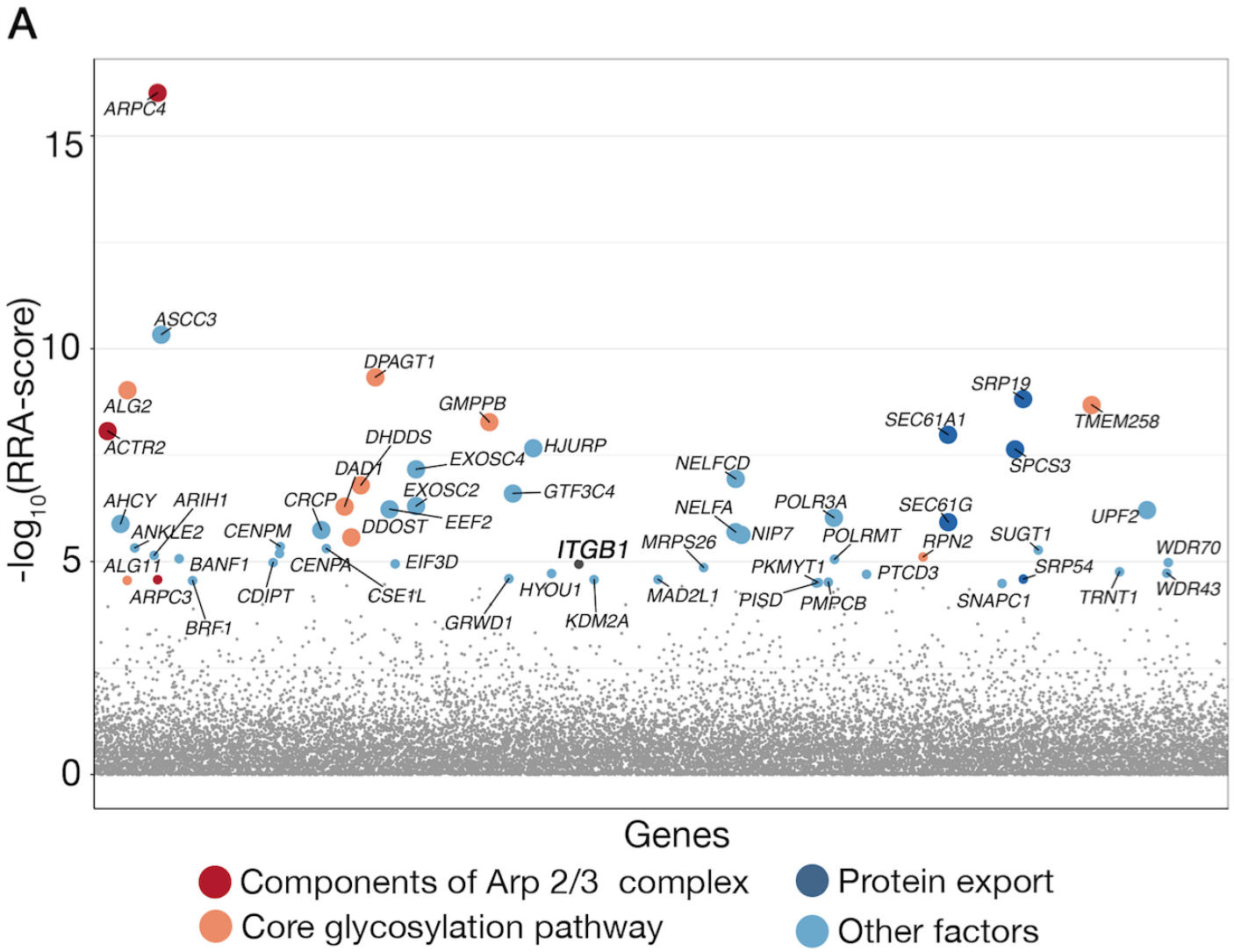


Fig. A.3 Enlarged version of figure 3.14A for better clarity of gene names

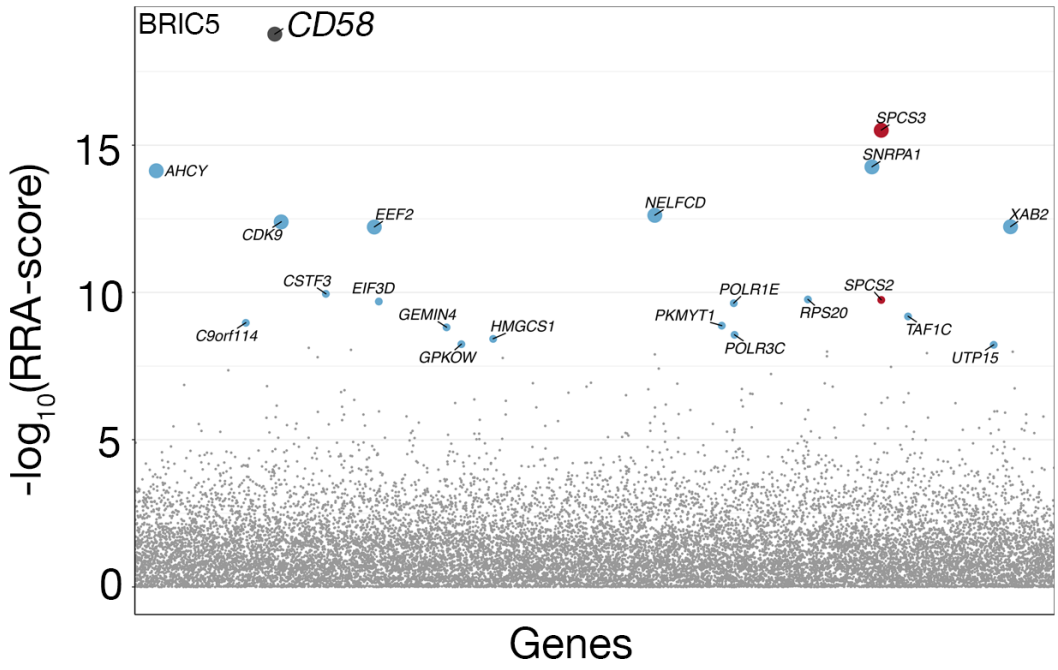


Fig. A.4 Genome-scale KO screen using an anti-CD58 mAb identifies *CD58* as the highest enriched gene in the sorted population. Gene-level enrichment analysis on a screen carried out using an anti-CD58 mAb. *CD58* was identified as the highest enriched gene. Two genes relating to general secretory were also identified (highlighted in red). Only genes with FDR<0.05 are depicted.

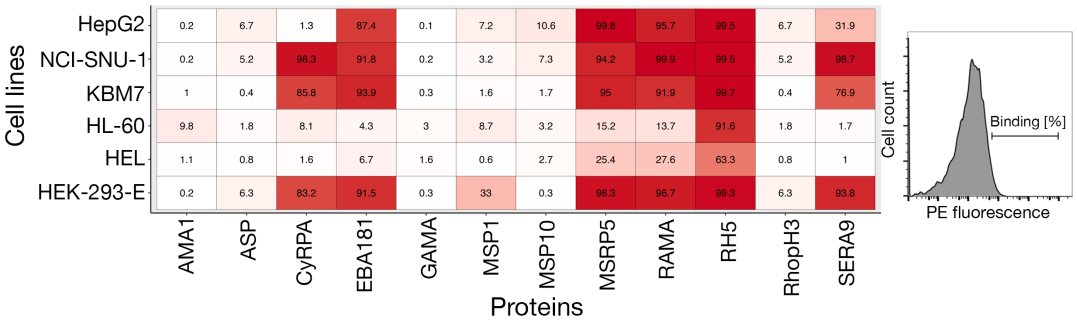


Fig. A.5 Primary screen for binding of 11 merozoite proteins to six cell lines. Biotinylated monomeric merozoite ectodomains conjugated to streptavidin-PE were tested for binding six cell lines. The number in each grid represents the percentage of cells that fell within the 'binding' gate, which was drawn on the histogram obtained from control protein binding to the parental cell lines (example depicted in the right panel). RH5 bound to all the tested cell lines, consistent with the expression of BSG in these lines (as determined from antibody staining with anti-BSG antibody- data not shown). HL-60 and HEL cells demonstrated restrictive binding profiles with clear binding only to RH5. Five ligands, SERA9, EBA181, MSRP5 CyRPA and RAMA bound at least three out of the six cell lines tested.

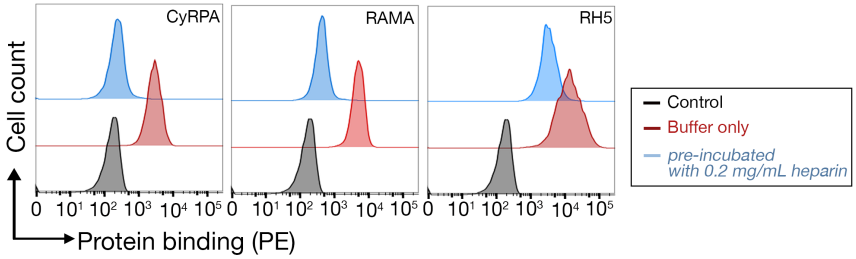


Fig. A.6 Binding of merozoite proteins RAMA and CyRPA to HEK-293-E cells can be completely blocked by soluble heparin. RAMA and CyRPA were identified in the preliminary screen to bind to HEK-293-E cells in a *SLC35B2*-dependent manner. Secondary validation with pre-blocking with soluble heparin (0.2 mg/mL) demonstrates complete inhibition of binding by soluble heparin. RH5 is used as a control; binding of RH5 is only partially blocked when pre-incubated with equivalent amount of heparin.

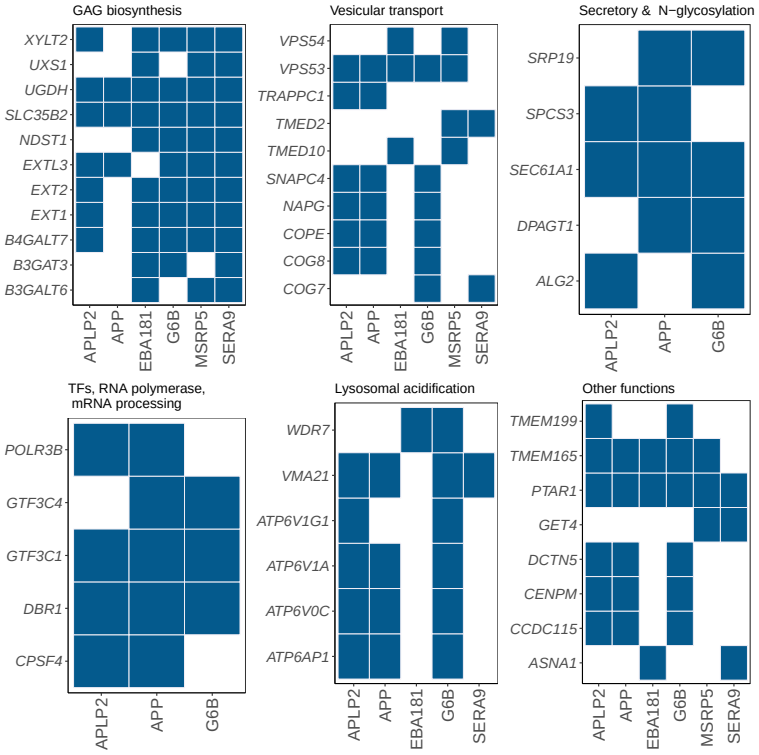


Fig. A.7 Summary of 'overlapping factors' identified in at least two out of six screen that identified the HS-biosynthesis pathway. Apart from the identification of multiple genes involved in biosynthesis of HS, these screens also revealed genes involved in protein transport and secretion, cellular house-keeping genes, genes involved in vesicular acidification, factors involved in global glycosylation in cells (*TMEM165, PTAR1*), and genes encoding for proteins involved in tail anchoring pathway (*ASNA1, GET4*). The screens with EBA181, MSRP5 and SERA9 were conducted on day 15/16 whereas the screens with G6B, APP and APLP2 were conducted on day 9: note that the representation of 'core-essential' genes (*SLC61A1, SRP19, ALG2, DPAGT1, GTF3C1, ATP6V1A, ATP6V0C, DBR1, CPSF4*) on screens carried out on later days is lower compared to the early time point screens.

Table A.5 Summary of 'other factors' identified with FDR<0.05 in at least two out of the seven different screens carried out using mAbs. Gene annotations are obtained from Uniprot.

Gene	Function	Identified in mAb targeting
ASCC3	Member of a family of helicases that are involved in the ATP-dependent unwinding of nucleic acid duplexes	CD47 (BRIC125, BRIC126, B6H12), ITGB1 (P16)
AHCY	Adenosylhomocysteinase; important for transmethylation reactions	CD58 (BRIC5), ITGB1 (P16)
HJURP	Centromeric protein that plays a role in the incorporation and maintenance of histone H3-like variant CENPA at centromeres	CD47 (B6H12), ITGB1 (P16)
CENPA	Centromere protein A- Histone H3-like variant	CD59 (BRIC 229), ITGB1 (P16)
CDIPT	Catalyzes the biosynthesis of phosphatidylinositol (PtdIns)	GYPA (BRIC 256), ITGB1 (P16)
EEF2	Eukaryotic Translation Elongation Factor 2	CD58 (BRIC5), ITGB1 (P16)
EIF3D	Eukaryotic Translation Initiation Factor 3 Subunit D	CD58 (BRIC5), ITGB1 (P16)
GTF3C2	General Transcription Factor IIIC Subunit 2	CD47 (B6H12, BRIC126)
GTF3C4	General Transcription Factor IIIC Subunit 4	CD47 (B6H12), ITGB1 (P16)
NELFCD	Essential part of the NELF complex- repress transcriptional elongation by RNA polymerase II	CD59 (BRIC229), ITGB1 (P16), CD58 (BRIC5)
NCBP2	Nuclear Cap Binding Protein Subunit 2	CD47 (BRIC125, B6H12)
PKMYT1	Member of the serine/threonine protein kinase family, which acts as a negative regulator of entry into mitosis	CD47 (BRIC126, B6H12), CD58 (BRIC5), ITGB1 (P16),
EXOSC4	Exosome component required for RNA processing and degradation activities	CD47 (BRIC126, B6H12), ITGB1 (P16)
EXOSC2	Exosome component required for RNA processing and degradation activities	CD47 (B6H12), ITGB1 (P16)
DBR1	RNA lariat debranching enzyme that facilitates ribonucleotide turnover	CD47 (BRIC125, BRIC126)

Table A.6 Genes identified in KO screens (FDR<0.05) using recombinant proteins as screening probes. The genes are ranked according to the RRA-score. Common gene refers to the genes identified repeatedly in screens using mAb (refer to table A.5). Gene annotations are obtained from Uniprot and annotations on core-essentiality are obtained from `http://www.hart-lab.org/Data/CEGv2.txt`.

Genes	Function	Rank	Notes
	Recombinant biotinylated RH5 conjugated to streptavidin-PE		
BSG	Known receptor for RH5	1	-
SLC35B2	Transporter of PAPS, required for GAG-biosynthesis	3	-
DBR1	RNA lariat debranching enzyme that facilitates ribonucleotide turnover	4	Common gene
SLC16A1	MCT1 transporter known to be required for transport of BSG	4	-
C9orf114	Required both for chromosome alignment and for association of the centrosomes	5	Core-essential gene
CCT3	Molecular chaperone required for protein folding	6	Core-essential gene
RPL27A	Encodes a ribosomal protein	7	Core-essential gene
SIN3A	Transcriptional regulatory gene	8	-
PSMG4	Promotes assembly of the 20S proteasome	9	-
FTSJ3	Probable methyltransferase	10	Core-essential gene
ARPC4	Component of Arp2/3 complex	11	-
XRCC5	Involved in NHEJ DNA repair	12	-
SRPRB	SRP Receptor Beta Subunit, general protein export	13	-
SRP19	Signal Recognition Particle 19, general protein export	14	Core-essential gene
ATRIP	Essential component of the DNA damage checkpoint	15	-
GNB2L1	Translational regulation	16	Core-essential gene
SART3	Encodes for RNA-binding nuclear protein, recycling factor of the splicing machinery	17	Core essential gene
CENPM	Centromere protein M	18	-
	Recombinant pentameric CD226		
PVR	Known receptor for CD226	1	-
DPAGT1	Core-glycosylation in the ER	2	Core-essential gene

Table A.6 – *Continued on next page*

Genes	Function	Rank	Notes
	Recombinant pentameric CD226 (*continued*)		
PIK3C3	Phosphatidylinositol 3-Kinase Catalytic Subunit Type 3, involved in multiple membrane trafficking pathways	3	-
TOP1	Topoisomerase	4	Core-essential gene
SEC61A1	General protein transport pathway	5	Core-essential gene
PIK3R4	Phosphoinositide-3-Kinase Regulatory Subunit 4, involved in multiple membrane trafficking pathways	6	-
GMPPB	Core-glycosylation pathway	7	Core-essential gene
SPDYC	Promotes progression through the cell cycle	8	-
ALG2	Core-glycosylation pathway	9	Core-essential gene
CHST8	Sulfotransferase: Transfer of sulfate N-acetylgalactosamine (GalNAc) residues in both N-glycans and O-glycans	10	-
WDR62	invoved in centriole duplication	11	-
DBR1	RNA lariat debranching enzyme that facilitates ribonucleotide turnover	12	Common gene
	Recombinant pentameric EPHB1		
EFNB2	Known receptor for EPHB1	1 -	
NDUFAF7	Arginine methyltransferase involved in the assembly or stability of mitochondrial NADH:ubiquinone oxidoreductase complex	2	-
GTF2H3	General Transcription Factor IIH Subunit 3	3	-
DDX27	DEAD-Box Helicase 27, required for the formation of ribosomal 47S rRNA	4	-
NDUFA3	Accessory subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase	5	-
LUC7L3	LUC7 Like 3 Pre-MRNA Splicing Factor	6	Core-essential gene
DR1	Down-Regulator Of Transcription 1	7	-
SPCS3	Signal Peptidase Complex Subunit, general protein transport	8	-

Table A.6 – *Continued on next page*

Genes	Function	Rank	Notes
	Recombinant pentameric EPHB1 (*continued*)		
RBM14	RNA Binding Motif Protein 14, a splicing modulator	9	Core-essential gene
SEC61G	General protein transport	10	-
HSD17B7	Hydroxysteroid 17-Beta Dehydrogenase 7, biosynthesis of cholesterol	11	-
	Recombinant pentameric LPHN1		
TMEM165	Global glycosylation in cells	1	-
TENM4	Known receptor for LPHN1	2	-
SLC35B2	Transporter of PAPS, required for GAG-biosynthesis	3	-
PTAR1	Global glycosylation in cells	4	-
ACTR2	Component of ARP2/3 complex	5	-
MCMDC2	Minichromosome Maintenance Domain Containing 2	6	-
EBP	Emopamil Binding Protein (Sterol Isomerase)	7	-
	Recombinant biotinylated TNFRSF9 conjugated to streptavidin-PE		
TNFSF9	Known interaction partner of TNFRSF9	1	-
CDKN2A	Binds to MDM2 and blocks MDM2-induced degradation of TP53	2	-
TP53	Tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains	3	-
CDC37	Molecular chaperone that forms a complex with Hsp90	4	Core-essential gene
DOHH	Required for generation of hypusine, an essential post-translational modification only found in mature eIF-5A factor	5	-
STK11	Serine/Threonine Kinase 11, able to phosphorylate TP53	6	-
DYRK1A	Member of Dual Specificity Tyrosine Phosphorylation Regulated Kinase family	7	-