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].

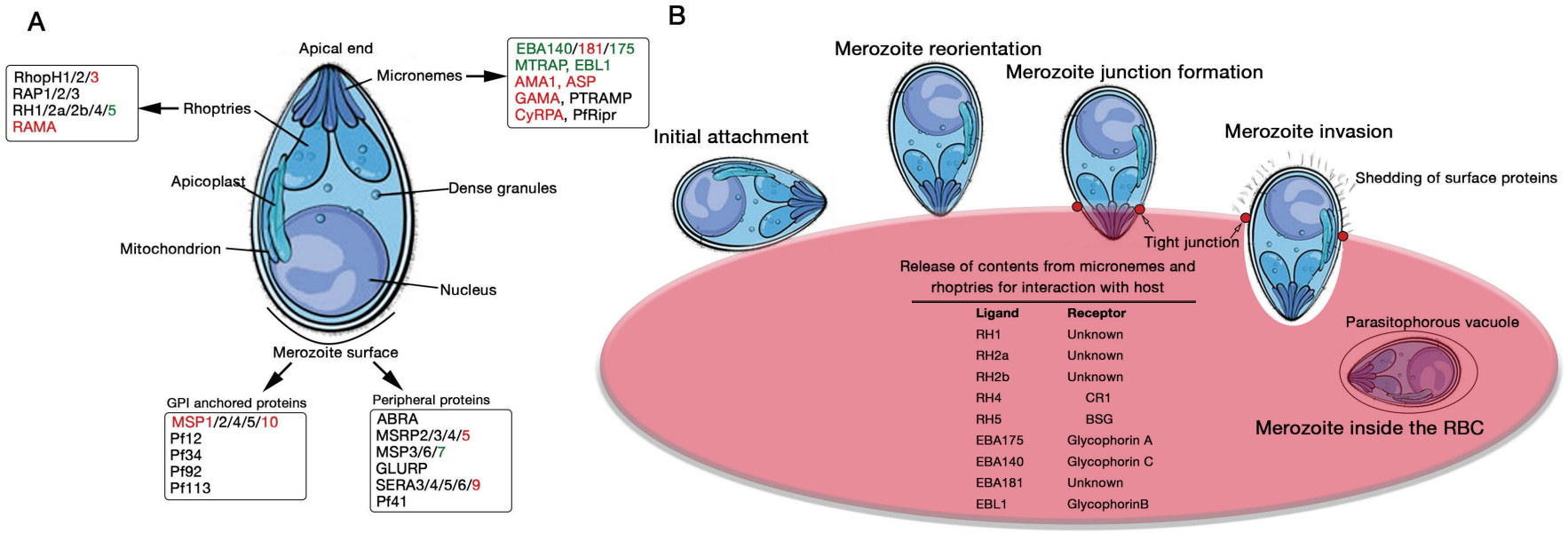

Fig. 4.1 Cellular organization and invasion process of *P. falciparum* merozoite. **A.** Schematics of cellular organisation of the merozoite. The merozoite is eukaryotic Apicomplexan parasite consisting of eukaryotic organelles such as a nucleus, a single mitochondrion, and an apicoplast together with 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

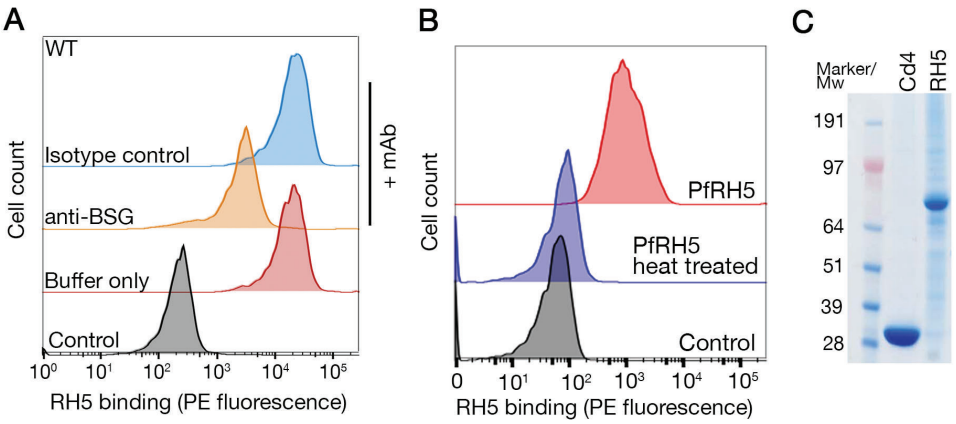

Fig. 4.2 RH5 binding to HEK293 was not completely dependent on BSG but was heat labile. Biotinylated RH5 was clustered around a streptavidin-PE conjugate and binding to HEK293 cells analysed by flow cytometry. **A.** RH5 binding was only partially blocked by an anti-BSG mAb relative to controls. **B.** 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.

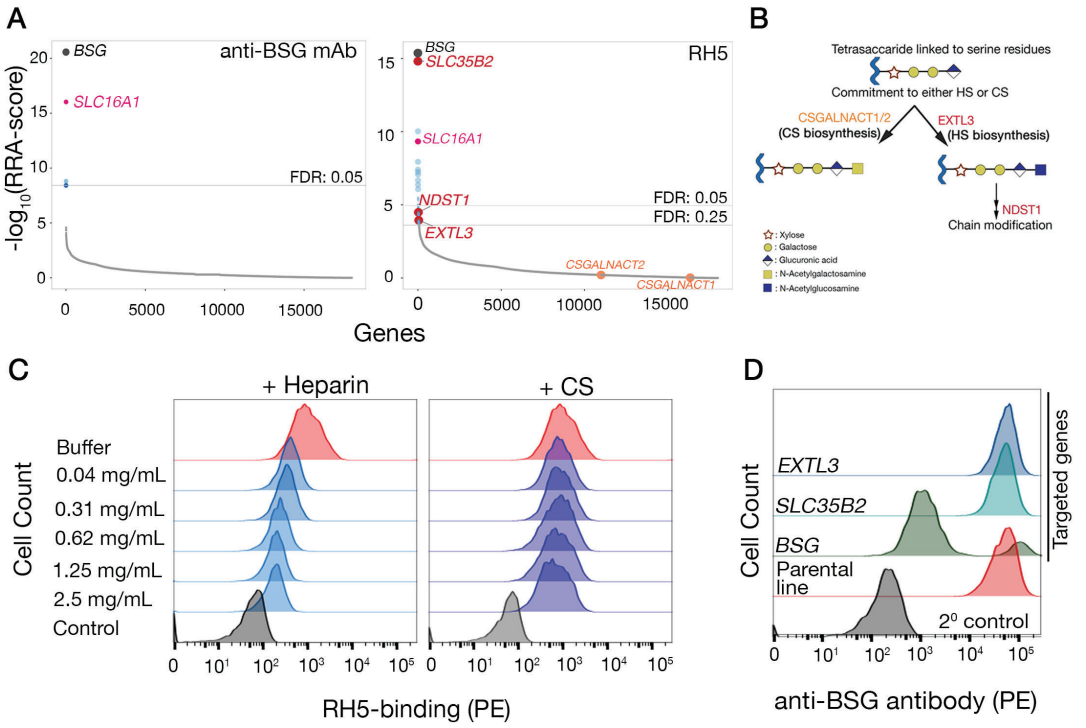

Fig. 4.3 Cell-based genetic screens identified BSG and heparan sulfate as independent receptors for *P. falciparum* RH5 on HEK293 cells. A. Rank-ordered genes identified from gRNA enrichment analysis required for cell surface display of an anti-BSG mAb (left panel) and RH5 binding (right panel). Significantly 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.

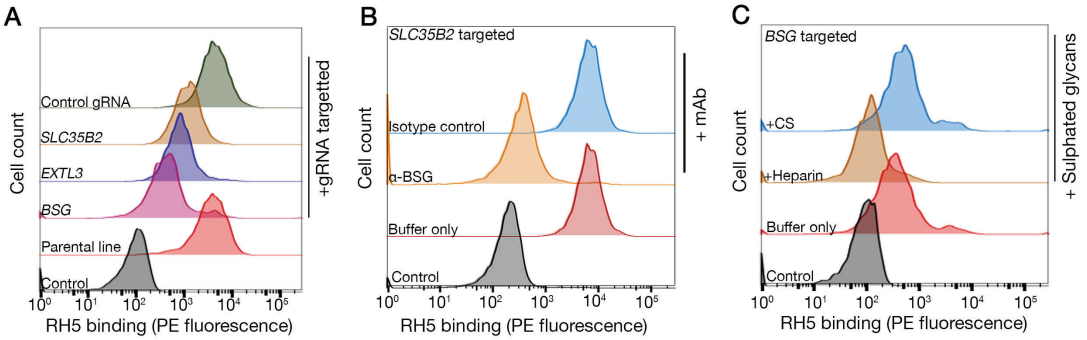

Fig. 4.4 The total observed binding of RH5 to HEK-293-E cell surface is the sum of independent contributions from BSG and HS. A. Binding of RH5 to cells is partially reduced when transduced with lentiviruses encoding gRNAs targeting either the receptor (*BSG*) or enzymes required for HS synthesis (*SLC35B2, 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.

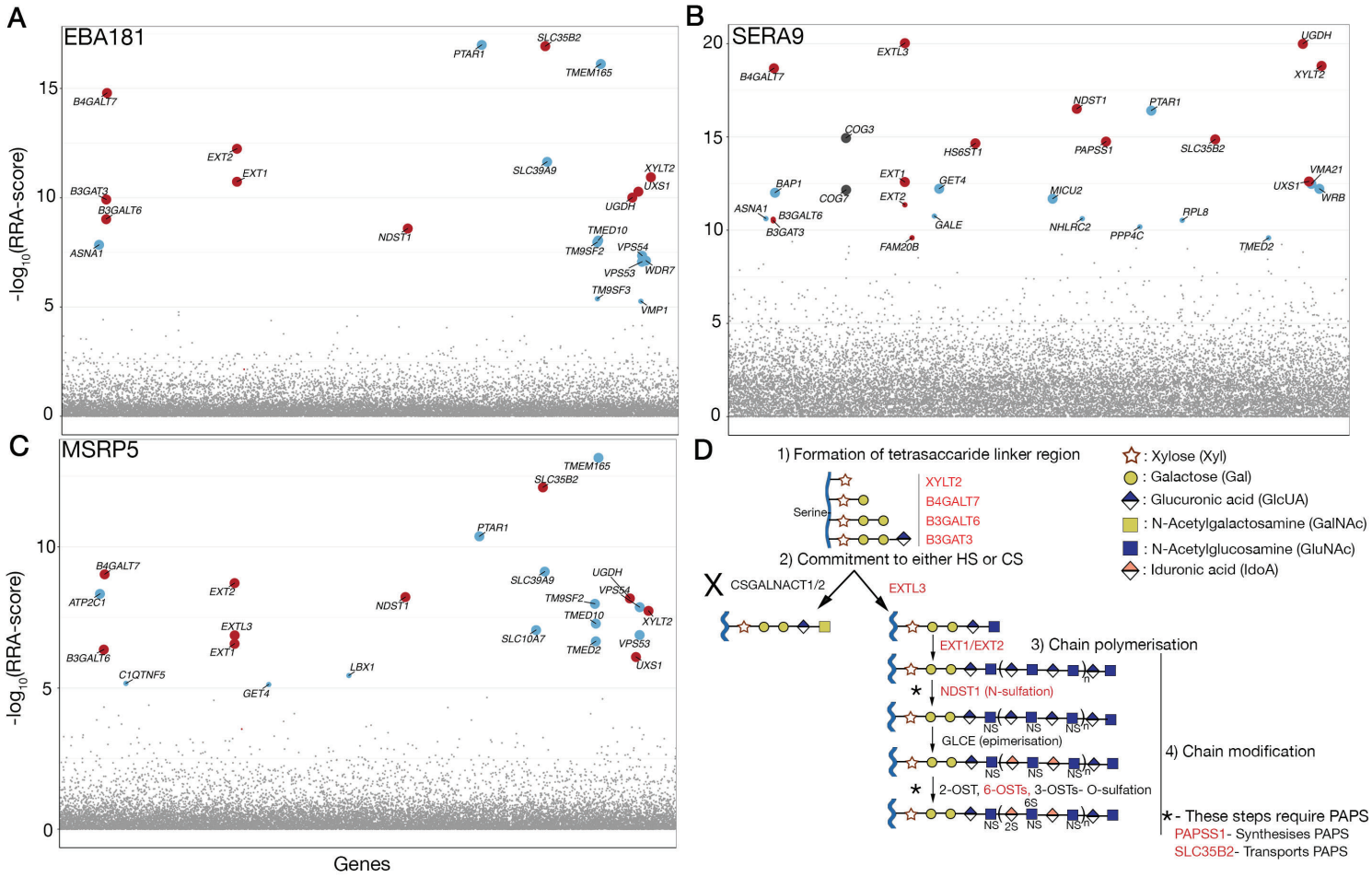

Fig. 4.5 All three genome-wide screens using merozoite proteins revealed the role of HS-biosynthesis pathway in mediating cell surface interactions. Gene level enrichment analysis on sorted mutant cells refractory to binding recombinant proteins- **A**-EBA181, **B**-SERA9 and **C**-MSRP5. Only genes 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).

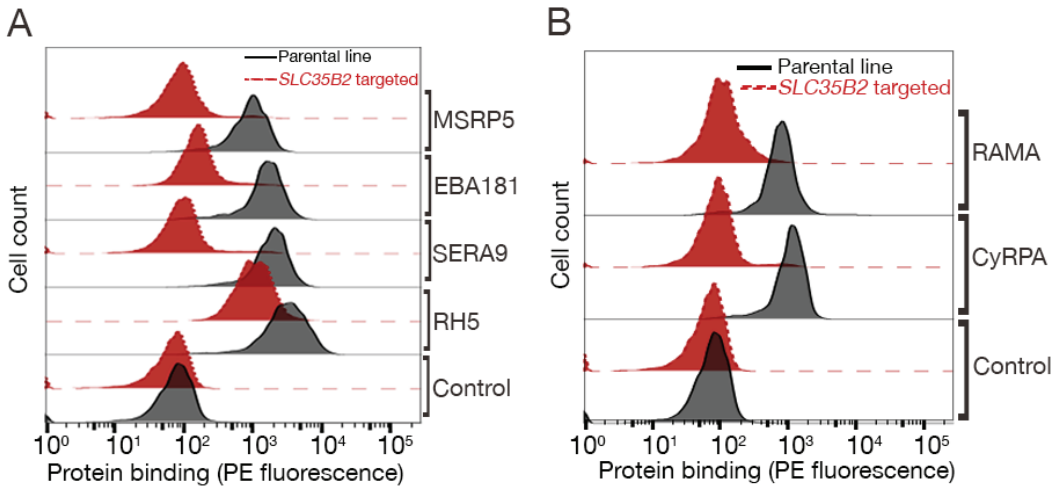

Fig. 4.6 The binding of multiple merozoite proteins to HEK-293-E cell surface can be completely abrogated by inactivating *SLC35B2*. Biotinylated ligands oligomerised around streptavidin-PE were tested for binding to an unmodified parental cell line (black histograms) or polyclonal *SLC35B2*-targeted 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.

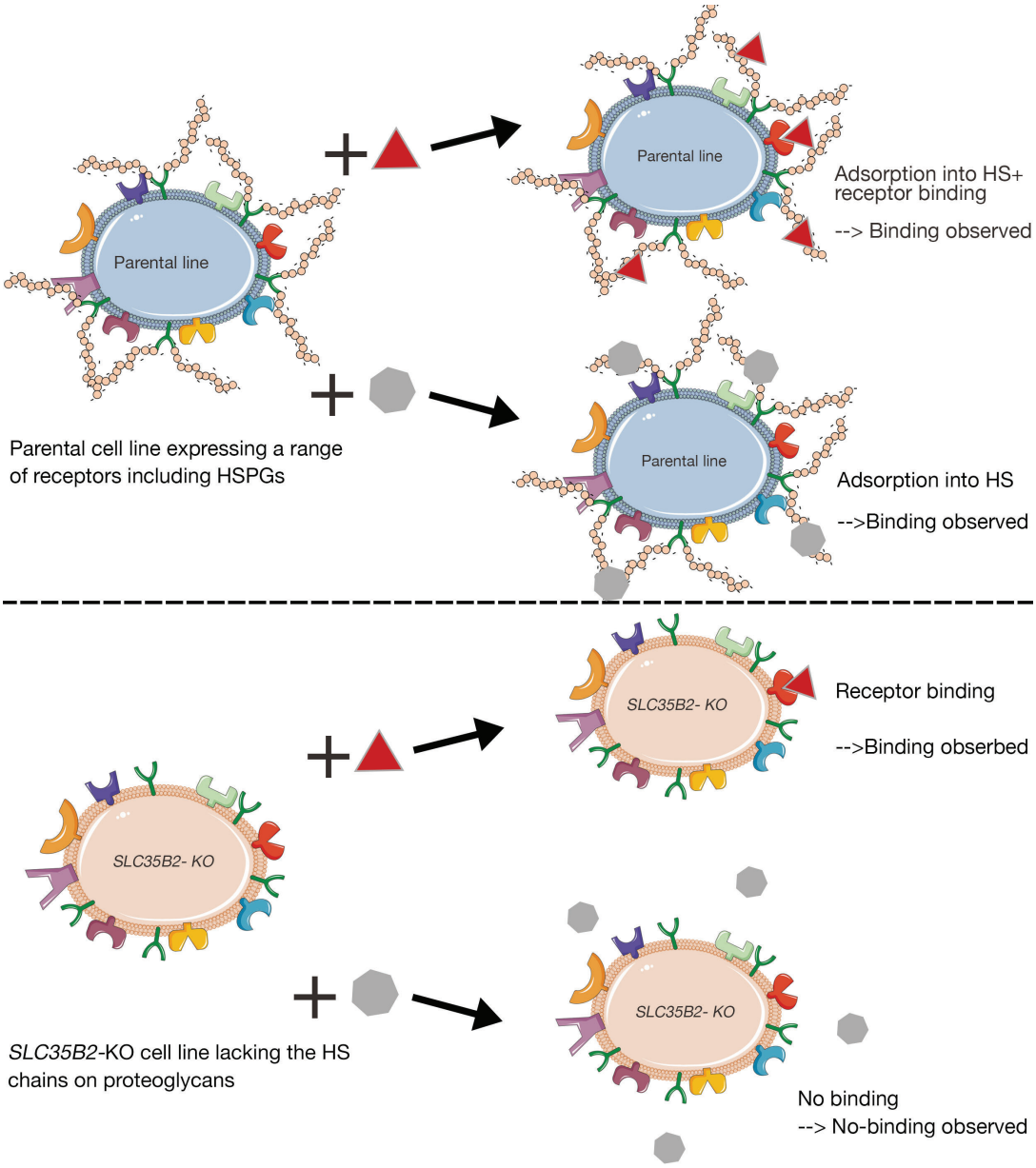

Fig. 4.7 Schematics for the approach to determine the fractional contribution of HS towards binding. Cell surfaces contain a range of receptors including heparan sulphate proteoglycans (HSPGs) such as syndecans and glypicans (represented in green) [250]. In the parental line, an observation of binding 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]

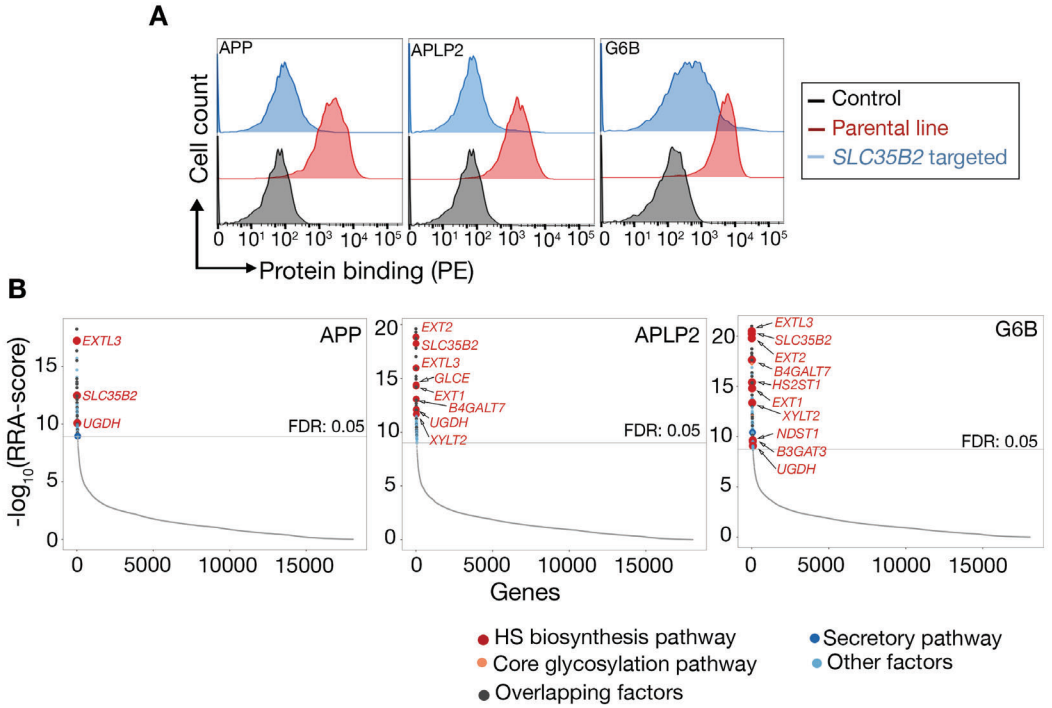

Fig. 4.9 Genome-scale screens identify HS-biosynthesis pathway when ligands that lose majority of binding upon targeting *SLC35B2* are used as screening probes. A. Pentameric ligands were tested for binding to unmodified parental cell lines or to polyclonal *SLC35B2*-targeted cells. The HEK293 cell line 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.

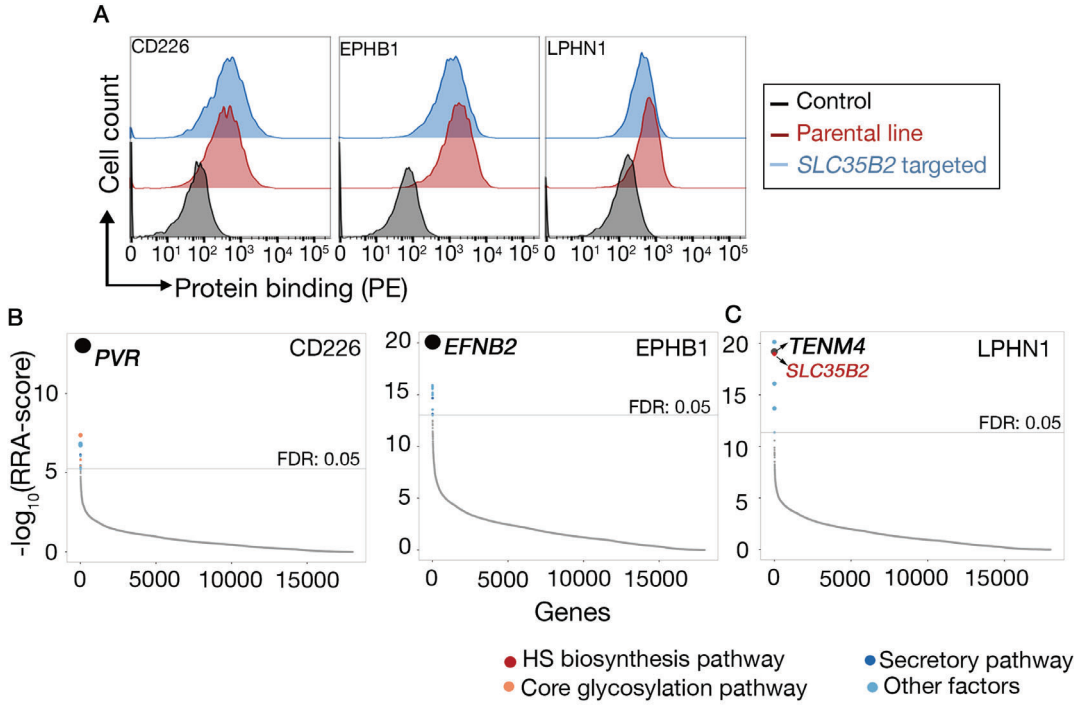

Fig. 4.10 Cell surface receptors were identified using cellular genetic screens. **A.** Pentameric ligands were tested for binding to the unmodified parental cell line, or to the polyclonal *SLC35B2*-targeted cells. The HEK293 cell line was used for all proteins. **B.** RRA-score rank-ordered genes identified 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.

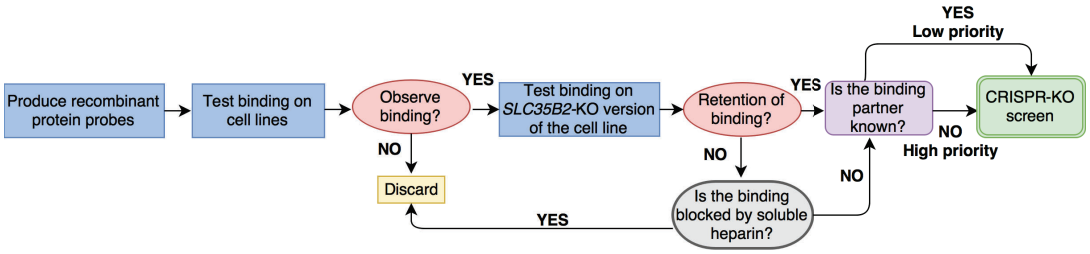

Fig. 4.11 Strategy for genetic screening using recombinant proteins. The strategy is based on the retention of binding on *SLC35B2*-targeted cell lines. A conclusive test for HS-adsorption is carried out with a blocking assay with soluble heparin. Where the receptor was already known, genetic screens 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.

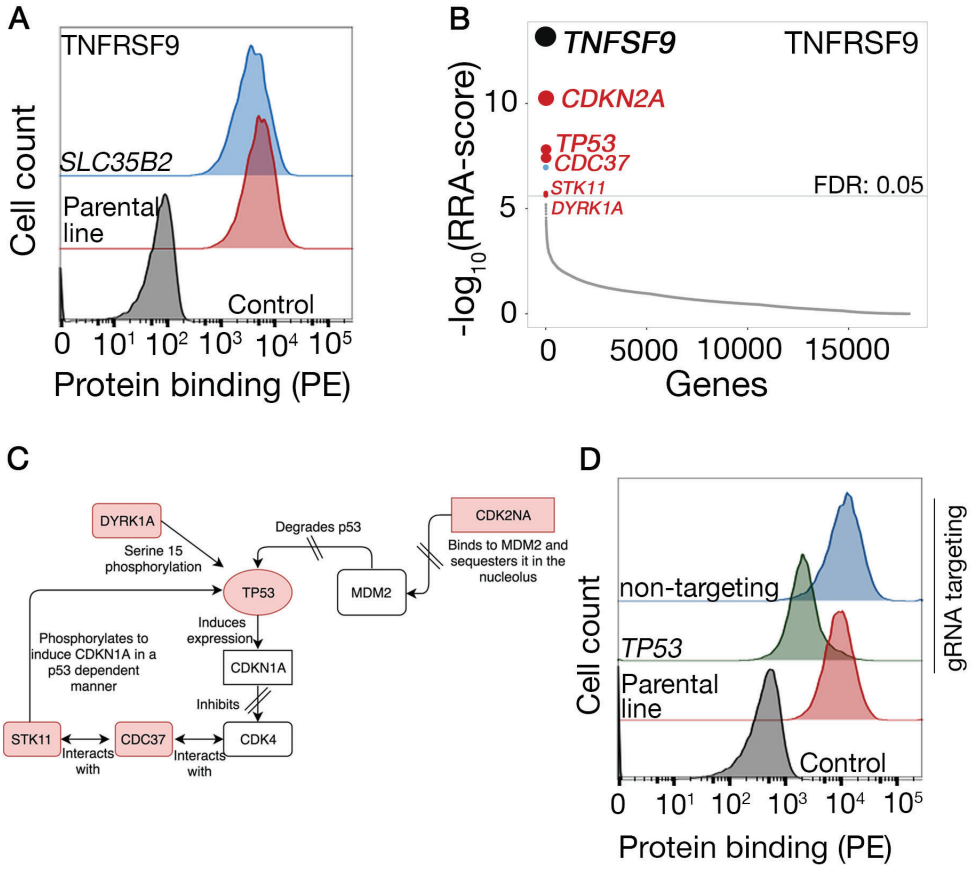

Fig. 4.12 Genome-scale screen using TNFRSF9 ectodomain as a sorting ligand identifies the interaction partner along with the p53 pathway. **A.** Monomeric biotinylated TNFRSF9 conjugated to streptavidin-PE binds to NCI-SNU-1 cell line; this binding was not affected by targeting of *SLC35B2*. **B.** RRA-score 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.