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.

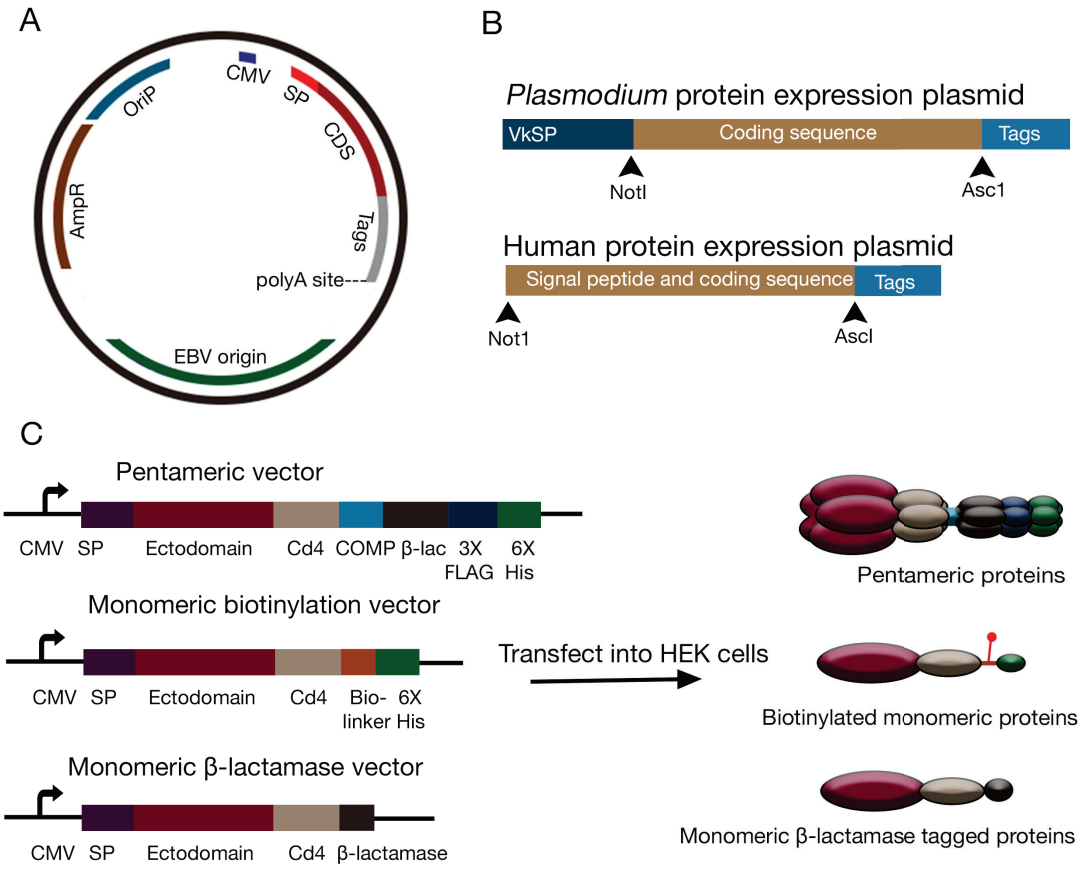

Fig. 2.1 Schematic diagram representing the plasmids used for protein production using the HEK-293-6E protein expression system. A. Generalized map of pTT3-based expression plasmids. EBV: Epstein-Barr virus, AmpR: Ampicillin resistance cassette, oriP: pBR322 origin of replication, CMV: Cytomegalovirus 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 1×10^6 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 $^\circ$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 $^\circ$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].