Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Approaches to identify extracellular receptor–ligand interactions

Laura Wood and Gavin J Wright

Thousands of secreted factors and plasma membrane-associated cell surface receptors are categorised into families that vary widely in their structures and functions. They often participate in extracellular binding events, but due to their unique physicochemical properties, their interactions are challenging to study. As lists of extracellular proteins become more complete and accurate, new methodologies are being developed to systematically identify how these proteins interact. Two main approaches have been used: direct binding between recombinant soluble receptor ectodomains and cell-based assays. Recent advances in chemoproteomic reagents, cDNA overexpression, and cell-based genetic approaches promote the identification of extracellular protein–protein interactions within the context of an intact plasma membrane in living cells and opens up the discovery of cell surface recognition events that were previously intractable.

Address
Cell Surface Signalling Laboratory, Wellcome Trust Sanger Institute, Cambridge, CB10 1SA, United Kingdom

Corresponding author: Wright, Gavin J (gw2@sanger.ac.uk)

Current Opinion in Structural Biology 2019, **56**:28–36

This review comes from a themed issue on **Sequences and topology**

Edited by **Anna Panchenko** and **Mónika Fuxreiter**

For a complete overview see the Issue and the Editorial

Available online 21st November 2018

https://doi.org/10.1016/j.sbi.2018.10.002

Introduction

Multicellular organisms are composed of many different cell types that co-ordinate their activities to form and maintain complex biological structures such as tissues and organs. As discrete units, cells must be able to recognise local cues and activate signalling pathways to alter cell behaviour in accordance with their immediate surroundings. To accomplish this, an array of structurally varied protein receptors embedded within the semi-permeable lipid bilayer of the plasma membrane serve as an interface between the external space and cell interior. These molecules usually require an N-terminal signal peptide for transport to the cell surface and can be tethered to the outer membrane through a glycosylphosphatidylinositol (GPI)-lipid modification, or contain one or more transmembrane domains. Regions that project out into the extracellular space are capable of binding soluble secreted factors or cell surface ligands exposed on adjacent cells and receptors that span the plasma membrane can relay instructional information to the cytoplasm to activate cell responses such as migration, differentiation, proliferation, cell growth and apoptosis. Cell surface receptor–ligand binding between cells (trans-interactions) are essential regulatory events coordinating many developmental and biological processes, and aberrant loss or gain of extracellular recognition can contribute to inappropriate changes in cell behaviour (e.g. cancer metastasis [1]). Receptors are also at the centre of host-pathogen interactions where protein binding is essential for influencing the pathology of infection which confers a major public health risk in relation to emerging diseases, most significantly by determining host tropism [2]. Motivations to study extracellular interactions are driven, not only by their wide ranging implications in development and disease, but also due to their accessibility to systematically delivered therapeutics, making these class of interactions tractable drug and vaccine targets. Currently ~70% of FDA-approved drugs target proteins containing transmembrane domains and/or signal peptides [3,4].

The unique biophysical properties of secreted and membrane spanning receptors make them a difficult subset of proteins to study. Firstly, oxidising environments, such as those found in the extracellular space, are required for disulphide bond formation between cysteine residues which are required for correct folding of protein ectodomains. Secondly, without an intact plasma membrane, solubilisation of full length functional receptors can be difficult to achieve. This is due to the amphipathic nature of transmembrane proteins which often contain both hydrophilic glycans, as well as stretches of hydrophobic amino acids which span the membrane. Finally, recognition events involving cell surface receptors are frequently low affinity (K_Ds in μM–mM range) and usually require localised clustering within the plasma membrane to increase binding avidity [5]. Physiologically, this means that protein–protein interactions can be easily reversed, allowing the cell to react quickly to continual changes in surrounding stimuli. Biochemically, this presents challenges and monovalent binding events with fast dissociation rates may not be detected with many high-throughput methods [5,6]. Certainly, interactions involving extracellular proteins were found to be underrepresented in commonly used protein–protein interaction screens

(e.g. Yeast-2-Hybrid (Y2H), Mammalian Protein–Protein Interaction Trap (MAPPIT), LUminescence-based Mammalian IntERactome (LUMIER)) [7] and binding events with proteins containing transmembrane helices or hydrophobic regions are also depleted in larger Y2H and co-fractionation studies [8]. Systematic interactome maps employing affinity purification–mass spectrometry (AP–MS) of stably expressed bait proteins do appear to identify plasma membrane binding partners, but whether these interactions represent trans-interactions on the cell surface is not clear [9,10].

Recent efforts to better classify the extracellular interactome ('secretome' and 'surfaceome') highlights the potential complexity of this interaction network [3,11–13]. Factors that are secreted into the extracellular space are predicted to form ~15% of the human proteome based on computational models to identify proteins containing signal peptides, but lacking transmembrane domains [3]. Similarly, the surfaceome is comprised of thousands of proteins with 1492 glycoproteins across 41 human cell lines accessible to chemoproteomic capture on the cell surface [12], while >1700 proteins have been confirmed to localise at plasma membranes based on immunocytochemistry observations [3,11]. Here, we aim to review some of the key methods used to systematically identify interactions between these groups of proteins. Particular emphasis has been placed on recent approaches that- first, take account of the biochemical challenges described above in identifying trans-interactions and second, have the potential to be developed, or are currently being applied, as high-throughput techniques. We discuss two broad experimental platforms: the use of heterologous expression systems in the production of large recombinant protein libraries consisting of soluble receptor ectodomains and approaches that use living cells to study cell surface interactions within the physiologically relevant microenvironment of the plasma membrane.

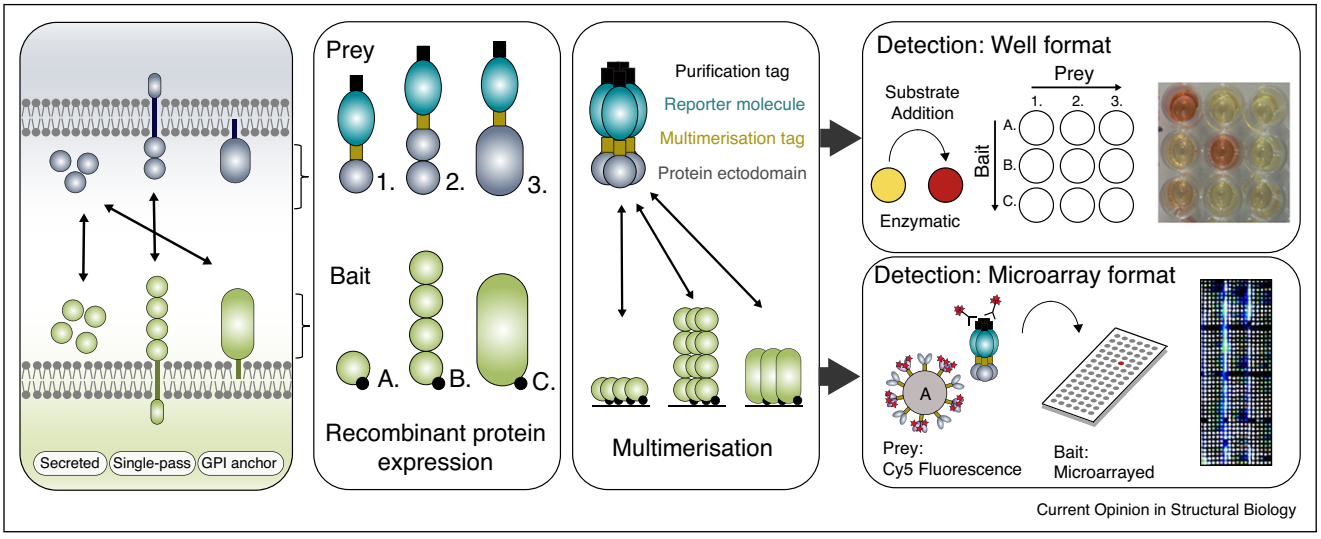
High-throughput detection: recombinant protein libraries

An important discovery in the development of high-throughput assays to detect cell surface interactions is that receptor ectodomains, when expressed as soluble recombinant proteins, can retain ligand binding functions [5,6]. Using mammalian and insect cell lines, recombinant proteins can be processed with appropriate post-translational modifications, including glycans and disulphide bonds, which are often critical for correct folding. Most methods utilise a 'prey' and 'bait' approach whereby ectodomains are immobilised on a solid surface and systematically probed for direct interactions with another recombinant protein (Figure 1) [14–18]. Multimerisation of proteins is an important step in this process as it functions to increase the binding avidities of transient interactions and occurs in two formats: localised conjugation in a microtitre well (baits) or using an oligomerisation tag (preys). This method has successfully discovered many ligand–receptor interactions, such as those essential for merozoite invasion of erythrocytes [19], neural guidance and interconnectivity [15,17], and studies focused within defined protein families [16,20]. More recently, this technique has been used for the first time to screen a library of single pass transmembrane receptor ectodomains (~1300) against recombinantly produced viral envelope proteins [21••]. By taking advantage of different multimerisation tags, Martinez-Martin *et al.* were able to produce trimers and pentamers of distinct glycoproteins found on the surface of the human cyto-megalovirus (HCMV) and found non-overlapping receptor binding partners for the two HCMV complexes. Of these, NRP2, a novel binding partner of the pentameric complex, was shown to play an important role in HCMV infection of epithelial and endothelial cells [21••].

Protein production is resource intensive and many studies limit the number of soluble ectodomains by concentrating on a subset of proteins, such as surface receptors expressed on a specific cell type (e.g. erythrocytes, platelets). To expand the number of proteins tested and reduce the amount of material required for each interaction assay, microarray technology enables the spotting of thousands of recombinant extracellular domains on a single slide and fluorescence signals, rather than enzymatic reactions, can be used to map protein–protein binding events (Figure 1) [22,23•]. The largest of these screened 40 500 binary events and identified 51 novel interactions between human receptors and an immunomodulatory protein on adenovirus family members [23•]. This methodology also has its problems, with potential issues in printing reproducibility between slides and long print runs that may compromise protein functionality if not maintained at low temperatures. Nucleic Acid Programmable Arrays (NAPPA) support an alternative microarray format whereby printed complementary DNAs (cDNAs) can be transcribed and translated *in vitro* directly on slides, and therefore eliminates the need for resource intensive protein purifications [24,25]. Combined with a microfluidic platform, Glick *et al.*, was able to utilise this technology to create arrays of ~2100 human membrane proteins and could successfully identify virus-receptor interactions [26]. As this is a cell-free system, it is still unclear as to what extent missing post-translation modifications and incorrect folding may affect receptor binding properties across the library.

The major drawback of recombinant expression libraries is that they only work for ectodomains that can be functionally expressed as a single contiguous region,

Figure 1



Extracellular protein interaction screening using soluble recombinant ectodomains.
Secreted factors, GPI anchored proteins and the ectodomains of single-pass transmembrane receptors can be expressed in heterologous expression systems where they are released into the cell media and either used directly in supernatants or concentrated with purification tags. Screening usually requires two protein libraries: a bait (A–C) and a prey (1–3). Prey recombinant proteins are typically oligomerised using a tag that promotes spontaneous multimerisation (e.g. dimers, trimers, pentamers), while bait proteins are conjugated to a solid substrate. In this schematic, bait proteins have been tagged with biotin and bound to a streptavidin coated surface, although other methods have been used [16,18]. In well-based detection, baits and preys are systematically screened against one another to account for all pairwise interactions. Prey proteins are fused with enzymatic reporter molecules so that binding to the bait library can be assessed using colorimetric measurement changes after substrate addition. Microarray technology is capable of spotting large recombinant bait libraries in a defined layout onto treated slides. Protein A microbeads coated with an unlabelled Fc-fusion prey protein and Cy5-labelled IgG can then be used to map extracellular interactions using localised fluorescence signals imaged using a microarray scanner [22,23•]. Binding to the arrayed bait library can also be detected using fluorescently labelled antibodies against a specific tag on the prey [65].

including type I and type II single pass transmembrane proteins, GPI anchored proteins and secreted polypeptides. Multi-pass membrane proteins and co-factor complexes are usually excluded from these expression lists, although integrin receptor combinations do appear to maintain binding specificities when co-expressed as α and β subunits [27]. In addition to this, large recombinant protein libraries are beyond the scope of many laboratories that may only be interested in identifying the interaction partners for one or a small group of proteins. Cell-based assays can serve as an alternative approach by simply using the endogenous receptor repertoire as an existing bait library, or utilising the cell machinery to exogenously overexpress cell surface proteins. Provided the plasma membrane remains intact, this can function as a platform to study cell surface recognition events that were previously biochemically intractable.

High-throughput detection: cell-based assays

Ligand-receptor capture proteomics

Advances in mass spectrometry sensitivity and quantitation, together with improved isolation techniques, have been instrumental in the identification of large protein interactomes from complex mixtures (e.g. crude cell extracts) [9,10,28]. To gain access to the internal cell proteome, plasma membranes must be disrupted. In some instances, this can lead to non-functional receptors since membrane-spanning polypeptides may no longer be able to maintain their native conformation. Larger insoluble plasma membrane fragments may also pellet during early isolation steps, causing membrane-associated proteins to be depleted from downstream analysis. Although classical AP–MS studies have had success in identifying host receptors for a number of virus glycoproteins, including HCMV and Herpes simplex virus (HSV) [29,30], extracellular protein interactions can be transient in nature ($t^1/_2$ <1 s) [5] and washes containing salts and detergents to remove non-specific binders may also exclude weak cell surface interactions. To overcome these challenges, the Wollscheid group synthesised a trifunctional compound (TRICEPS) that utilises the glycan rich coat displayed by many cell surface proteins to capture receptor interactions on intact living

cells [31,32]. TRICEPS is covalently conjugated to the primary amines of a ligand of interest, while a second functional group facilitates crosslinking to glycosylated binding partners on the surface of cells cultured in an oxidising environment. A final biotin group is bound by streptavidin for the affinity purification of peptides containing *N*-glycosylation motifs, and glycan cleavage followed by mass spectrometry analysis reveals enriched ligand interactors. Secreted proteins, therapeutic antibodies, peptides and virus particles were all shown to act as successful probes in the capture of known cell surface recognition events [31]. Notably, they were able to verify binding of a peptide (apelin-17) with its target G-protein-coupled receptor (Apelin receptor), highlighting that protein interactions with multi-spanning cell surface receptors can indeed be identified by techniques that take into account the integrity of the cell membrane. Further development of this concept has led to the creation of a trifunctional crosslinker called ASB (aldehyde-reactive aminooxy group, a sulfhydryl, and a biotin) [33] and more recently HATRIC-based ligand receptor capture (HATRIC-LRC) [34•]. In the latter, experiments can be performed in a physiologically relevant environment (pH 7.4), opening up the discovery to pH-sensitive cell surface interactions. The use of azide click chemistry to label glycoproteins for affinity isolation also means that peptides within the full length protein can be used for mass spectrometry identification (not just *N*-glycosylated peptides) reducing the need for large amounts of starting material—an important technical consideration for cell lines that are difficult to grow in culture [34•]. Although TRICEPS, ASB and HATRIC possess many advantages when it comes to studying receptor binding events in their natural states, components of the cell surface must be glycosylated and therefore a subset of glycan-free proteins may be missing from these interaction lists.

Expression cloning using cDNA expression libraries

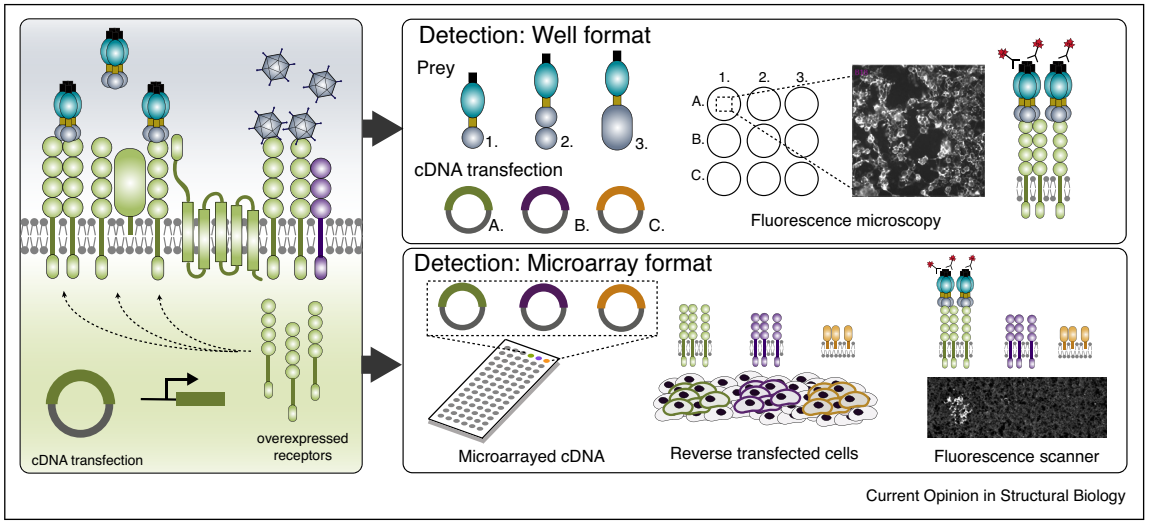
In expression cloning, a library of complementary DNAs (cDNAs) is transfected into cultured cell lines and screened for a phenotype of interest. Multiple subdividing rounds filter 'positive' and 'negative' cDNA pools until a single expression plasmid is recovered [35]. In regards to extracellular interactions, a common readout would be cells that have gained the ability to bind a recombinant ligand of interest. In the past this technique was crucial in the discovery of a number of growth factor receptors [36–38]. Adapted protocols have been used to identify interactions between Hepatitis C virus and multi-spanning transmembrane proteins CD81 [39], Claudin-1 [40] and Occludin [41] and more recently the low affinity binding (*K*D of ~12 μM) between egg and sperm surface proteins during fertilisation [42]. In recent years, genome sequencing and gene annotation have enabled projects such as the Mammalian Gene Collection [43] and the Orfeome collaboration [44•] to compile sequence-verified plasmid libraries. This innovation supports the replacement of pooled approaches with individual cDNA transfections as it eliminates the need for iterative rounds of selection and instead reveals an immediate binary interaction (Figure 2). As an example, Lin *et al*. expressed ~400 surface receptors in COS7 cells and probed with a similar library of ectodomain-Fc fusions reporting a single positive hit, the Netrin-G1 and NGL-1 interaction [45]. Recent efforts to make this more high-throughput have seen this technique develop into a 'microarray'-based screen. The Sabatini group first showed that spotted cDNAs (120–150 μm in diameter) could be 'reverse transfected' into HEK293T cells grown over the top of slides, with subsequent protein expression only seen in localised areas where cells had come in direct contact with the patterned plasmids [46]. Large libraries of expression plasmids encoding for full length plasma membrane receptors can be arrayed on slides, reverse transfected into cells, and the overexpressed receptors used as a surface to probe for extracellular interactions (Figure 2). EMP1, a *Plasmodium falciparum* (*Pf*) protein exposed on infected erythrocytes, was shown to bind to cells locally overexpressing the endothelial protein C receptor (EPCR) using a commercially available adaption of this technology [47]. Thousands of interactions can be probed in a single experiment, with the most recent attempts screening 3559 and 4493 cell surface proteins against Hom-1 virus particles [48] and a recombinant growth factor (GDF15) [49••], respectively.

CRISPR and haploid genetic screens

CRISPR/Cas9 technology and haploid genetic screens allow genome scale study of loss of function phenotypes [50–53]. In the pooled CRISPR/Cas9 approach, a library of guide RNAs are targeted to essentially all protein-coding genes within the genome so that libraries of cells, each deficient in a non-essential gene, can be created [50,51]. In haploid cell screens, retroviral gene-traps integrate into the genome and inactivate single alleles through insertional mutagenesis [52,53]. By selecting for cell populations that are refractory to pathogen infection, these gene disruption techniques have been particularly successful in the discovery of virus-host cell entry factors [54]. Known virus recognition events with human cell surface receptors have been corroborated using both CRISPR and haploid genetic screens: Hepatitis C Virus (Occludin, CD81, Claudin-1 [55]), Zika virus (AXL [56]), HIV (CD4 and CCR5 [57]), Poliovirus (PVR [58]) and Coxsackie virus B1 (CXADR [58]) and these approaches have also been used to identify novel host receptors for viruses [59–62,63•].

Fluorescently conjugated protein extracellular domains that bind to endogenous receptors on the surface of cells can be used as an effective marker during fluorescence-

Figure 2



Extracellular protein interaction screens using cDNA transfection protocols. cDNAs encoding full length receptors are introduced into cells using commercially available transfection reagents. The cell utilises its own machinery to overexpress receptors on the cell surface and this in turn is used as a platform to study extracellular protein–protein interactions with ligands such as recombinant protein ectodomains and virus particles. In well-based detection, cDNAs are transfected individually into cells grown in microtitre wells (A–C). Prey recombinant proteins (1–3) are then incubated with cells and interactions detected using fluorescently labelled antibodies and fluorescence microscopy imaging. Alternatively, cDNAs can be spotted onto slides with a microarrayer and reverse transfected into a lawn of cells. Cells that locally take up cDNAs will express the specific receptors on their cell surface and detection of extracellular interactions can be mapped using fluorescent antibodies against a specific tag on the recombinant prey.

activated cell sorting (FACS) [64••]. By selecting cells that exhibit a 'loss of binding' phenotype seven low affinity ligand–receptor interactions, and a novel binding event, were reported using pooled CRISPR/Cas9 technology. Multi-pass transmembrane proteins were discovered in two instances; first as a direct receptor for Syncytin-1, and second as a cell surface chaperone for basigin, the receptor of *Pf*RH5. The latter highlights one of the main advantages of this technique—the ability to explore contributions from genes in pathways critical for receptor presentation on the cell surface–and includes components such as transcription factors, trafficking proteins and post-translational modification enzymes. For example, CD59 surface detection with an antibody was found to depend on GPI anchor biosynthesis pathway components [64••], while plasma membrane localisation of CCR5, an essential receptor for HIV infection, requires factors that attach sulfates to key tyrosine residues [57]. Genome-wide approaches can also be used to identify complex carbohydrate-based receptors, such as surface displayed heparan sulfates. When evaluating genes responsible for loss of *Pf*RH5 cell binding, both basigin and heparan biosynthesis components were identified as high confidence hits [64••]. A protein receptor containing immunoglobulin-like domains and heparan biosynthesis enzymes were also enriched in genetic screens studying adeno-associated virus infection [61] and highlights that multiple cell surface receptors can be detected in a single screen and these can be protein or non-protein-based molecules. Therefore, a crucial advantage of this technique is that no prior assumptions need to be made regarding the molecular nature of the receptors involved.

Conclusions and perspectives

In this review, we present an overview of the latest high-throughput techniques used to study extracellular protein–protein interactions. Although efforts have been made to identify genes that encode proteins destined for the cell surface or extracellular space, a definitive list that takes into account the localisation patterns of canonical and alternatively spliced isoforms, as well as genetic variants, is currently not complete. This level of complexity is a major research challenge and even if full coverage of the extracellular proteome is achieved, any one method is unlikely to detect all interaction networks since each have their own advantages and disadvantages (Table 1). With the dropping costs of gene synthesis and improved yields in heterologous protein expression systems, producing large recombinant libraries is not as daunting as it once was; however, the technique is still limited by its inability to identify receptor interactions involving multiple co-factors. Cell-based assays provide an opportunity to bridge this gap by ensuring that protein interactions are studied on intact cell surfaces. In genome

Table 1

Advantages and disadvantages of extracellular protein–protein interaction screens

	Biochemical assays	Cell-based assays		
	Recombinant protein production	Chemoproteomic reagents	cDNA expression libraries	Genome wide loss-of-function screens
Advantages	Ectodomains from secreted factors, single-pass and GPI linked proteins can be expressed and solubilised	Potential to identify interactions with endogenously expressed single-pass, GPI linked, multi-pass and multi-subunit receptors	Potential to identify interactions with overexpressed single-pass, GPI linked, multi-pass and multi-subunit receptors	Potential to identify interactions with protein receptors, non-protein receptors (e.g. Heparan sulfate) and upstream pathway components
	Suitable post-translational modifications may increase the likelihood of correct folding	Full length functional receptors are studied within the context of the cell surface microenvironment	Full length functional receptors are studied within the context of the cell membrane, although the surface microenvironment is altered due to forced overexpression of a receptor	Full length functional receptors are studied within the context of the cell surface microenvironment
	Multimerisation strategies increase the binding avidity of low affinity cell surface interactions	Variety of ligands can be used to probe cell surface interactions— peptides, viruses, proteins	Variety of ligands can be used to probe cell surface interactions— proteins, viruses	Variety of ligands can be used to probe cell surface interactions and the readout is phenotype driven. Can study processes such as pathogen invasion and cell survival
	Recombinant proteins can be concentrated using purification tags and protein activities normalised for robust readouts in downstream assays	HATRIC-LRC: Can detect pH sensitive interactions and requires low amounts of starting material	Sequence validated Open reading frame (ORF) expression clones are readily available. cDNA libraries are stable and can be frozen for long term storage	No prior assumptions on the nature of the receptor need to be made
Disadvantages	Post-translational modifications may be missed (e.g. under glycosylated). A fraction of the protein may be misfolded.	Depends on the endogenous levels of a receptor. Non-expressing or low abundance receptors may be missed	Depends on the cell's ability to overexpress and transport receptors to the cell surface	Depends on the endogenous levels of a receptor. Non-expressing or low abundance receptors may be missed
	Protein production is costly and resource intensive. Low expressers can be difficult to obtain in sufficient amounts.	Requires receptor to be glycosylated	High numbers of transient transfections performed for every individual screen. Potential variation in transfection efficiency	Difficult to identify essential genes as the cells with mutations in these genes are likely to drop out of the population causing under sampling
	Long-term storage difficult— multiple freeze–thaws may cause protein denaturation	Amine conjugation may mask ligand binding sites (e.g. protein binding domains containing lysine residues)	Large cDNA libraries can be difficult to compile and organise.	The use of loss-of-function approaches makes it difficult to identify functionally redundant receptors
	Multi-pass membrane proteins and multi-subunit receptors are often biochemically intractable	Mild chemical oxidation of living cells may inhibit some cell surface interactions	Co-transfection probably limited to multi-subunit complexes with 2–3 components	Large numbers of cells are required for statistically significant results. This leads to long cell sorting times, reducing the throughput of interaction screens

wide genetic screens, the use of intact living cells enables a phenotype-driven approach for evaluating protein binding events, and may have a greater potential for revealing physiologically relevant interactions, such as those described for virus infection models. One of the most interesting outcomes is that genome wide loss-of-function screens can identify, not only direct ligand binding partners, but also upstream pathways critical for receptor presentation on the cell surface. Highlighting multiple targets at different points along the same pathway is likely to be helpful in drug discovery pipelines. All of these methods have the ability to uncover independent and overlapping protein binding events and this is determined, in part, by the biochemical nature of the receptor in question. With the potential for increased scalability and sensitivity, integration of these interaction networks will be necessary for the construction of a comprehensive and accurate map of the extracellular interactome.

Declarations of interest

None.

Acknowledgements

This work was supported by the Wellcome Trust grant 206194. The funder had no input into the collection, analysis, interpretation or writing of this manuscript. We thank Sumana Sharma for valuable discussions and advice.

References and recommended reading[1]

Papers of particular interest, published within the period of review, have been highlighted as

• of special interest

•• of outstanding interest

1. Massague J, Obenauf AC: **Metastatic colonization by circulating tumour cells**. *Nature* 2016, **529**:298-306.
2. Douam F, Gaska JM, Winer BY, Ding Q, von Schaewen M, Ploss A: **Genetic dissection of the host tropism of human-tropic pathogens**. *Annu Rev Genet* 2015, **49**:21-45.
3. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A *et al.*: **Tissue-based map of the human proteome**. *Science* 2015, **347**:1260419.
4. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J: **DrugBank: a comprehensive resource for in silico drug discovery and exploration**. *Nucleic Acids Res* 2006, **34**:D668-672.
5. Wright GJ: **Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions**. *Mol Biosyst* 2009, **5**:1405-1412.
6. Wright GJ, Martin S, Bushell KM, Sollner C: **High-throughput identification of transient extracellular protein interactions**. *Biochem Soc Trans* 2010, **38**:919-922.
7. Braun P, Tasan M, Dreze M, Barrios-Rodiles M, Lemmens I, Yu H, Sahalie JM, Murray RR, Roncari L, de Smet AS *et al.*: **An experimentally derived confidence score for binary protein-protein interactions**. *Nat Methods* 2009, **6**:91-97.
8. Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R *et al.*: **A proteome-scale map of the human interactome network**. *Cell* 2014, **159**:1212-1226.
9. Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J, Buchholz F *et al.*: **A human interactome in three quantitative dimensions organized by stoichiometries and abundances**. *Cell* 2015, **163**:712-723.
10. Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H *et al.*: **Architecture of the human interactome defines protein communities and disease networks**. *Nature* 2017, **545**:505-509.
11. Thul PJ, Lindskog C: **The human protein atlas: a spatial map of the human proteome**. *Protein Sci* 2018, **27**:233-244.
12. Bausch-Fluck D, Hofmann A, Bock T, Frei AP, Cerciello F, Jacobs A, Moest H, Omasits U, Gundry RL, Yoon C *et al.*: **A mass spectrometric-derived cell surface protein atlas**. *PLoS One* 2015, **10**:e0121314.
13. da Cunha JP, Galante PA, de Souza JE, de Souza RF, Carvalho PM, Ohara DT, Moura RP, Oba-Shinja SM, Marie SK, Silva WA Jr *et al.*: **Bioinformatics construction of the human cell surfaceome**. *Proc Natl Acad Sci U S A* 2009, **106**:16752-16757.
14. Bushell KM, Sollner C, Schuster-Boeckler B, Bateman A, Wright GJ: **Large-scale screening for novel low-affinity extracellular protein interactions**. *Genome Res* 2008, **18**:622-630.
15. Gao X, Metzger U, Panza P, Mahalwar P, Alsheimer S, Geiger H, Maischein HM, Levesque MP, Templin M, Sollner C: **A floor-plate extracellular protein-protein interaction screen identifies draxin as a secreted Netrin-1 antagonist**. *Cell Rep* 2015, **12**:694-708.
16. Ozkan E, Carrillo RA, Eastman CL, Weiszmann R, Waghray D, Johnson KG, Zinn K, Celniker SE, Garcia KC: **An extracellular interactome of immunoglobulin and LRR proteins reveals receptor-ligand networks**. *Cell* 2013, **154**:228-239.
17. Visser JJ, Cheng Y, Perry SC, Chastain AB, Parsa B, Masri SS, Ray TA, Kay JN, Wojtowicz WM: **An extracellular biochemical screen reveals that FLRTs and Unc5s mediate neuronal subtype recognition in the retina**. *eLife* 2015, **4**:e08149.
18. Wojtowicz WM, Wu W, Andre I, Qian B, Baker D, Zipursky SL: **A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains**. *Cell* 2007, **130**:1134-1145.
19. Crosnier C, Bustamante LY, Bartholdson SJ, Bei AK, Theron M, Uchikawa M, Mboup S, Ndir O, Kwiatkowski DP, Duraisingh MT *et al.*: **Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum**. *Nature* 2011, **480**:534-537.
20. Sollner C, Wright GJ: **A cell surface interaction network of neural leucine-rich repeat receptors**. *Genome Biol* 2009, **10**:R99.
21. •• Martinez-Martin N, Marcandalli J, Huang CS, Arthur CP, Perotti M, Foglierini M, Ho H, Dosey AM, Shriver S, Payandeh J *et al.*: **An unbiased screen for human cytomegalovirus identifies Neuropilin-2 as a central viral receptor**. *Cell* 2018, **174**:1158-1171 e1119.

In this paper, optimisation of the avidity-based extracellular interaction screen (Avexis) is used to probe ~1500 single-pass transmembrane recombinant ectodomains against two distinct human cyto-megalovirus (HCMV) envelope protein complexes. This highly sensitive, high-throughput screen successfully identified independent HCMV host receptors for fibroblast and epithelial/endothelial infection.

22. Tom I, Lewin-Koh N, Ramani SR, Gonzalez LC: **Protein microarrays for identification of novel extracellular protein-protein interactions**. *Curr Protoc Protein Sci* 2013:23. Chapter 27: Unit 27.
23. • Martinez-Martin N, Ramani SR, Hackney JA, Tom I, Wranik BJ, Chan M, Wu J, Paluch MT, Takeda K, Hass PE *et al.*: **The extracellular interactome of the human adenovirus family reveals diverse strategies for immunomodulation**. *Nat Commun* 2016, **7**:11473.

This large host-pathogen extracellular interaction screen uses microarray printing technology to spot >1500 recombinant human ectodomains on slides and probe with immunomodulatory proteins from adenovirus family members to identify 51 novel binding events.

24. Ramachandran N, Raphael JV, Hainsworth E, Demirkan G, Fuentes MG, Rolfs A, Hu Y, LaBaer J: **Next-generation high-density self-assembling functional protein arrays**. *Nat Methods* 2008, **5**:535-538.
25. Carlson ED, Gan R, Hodgman CE, Jewett MC: **Cell-free protein synthesis: applications come of age**. *Biotechnol Adv* 2012, **30**:1185-1194.
26. Glick Y, Ben-Ari Y, Drayman N, Pellach M, Neveu G, Boonyaratanakornkit J, Avrahami D, Einav S, Oppenheim A, Gerber D: **Pathogen receptor discovery with a microfluidic human membrane protein array**. *Proc Natl Acad Sci U S A* 2016, **113**:4344-4349.
27. Dundas K, Shears MJ, Sun Y, Hopp CS, Crosnier C, Metcalf T, Girling G, Sinnis P, Billker O, Wright GJ: **Alpha-v-containing integrins are host receptors for the Plasmodium falciparum sporozoite surface protein, TRAP**. *Proc Natl Acad Sci U S A* 2018, **115**:4477-4482.
28. Wan C, Borgeson B, Phanse S, Tu F, Drew K, Clark G, Xiong X, Kagan O, Kwan J, Bezginov A *et al.*: **Panorama of ancient**

[1] We would also like to highlight two very relevant studies that were published whilst this manuscript was in review. First, a review focusing on the surfaceome interaction network [66]; and second, a paper using an in silico machine learning strategy to further define the composition of the surfaceome [67].

metazoan macromolecular complexes. *Nature* 2015, **525**:339-344.

29. Kabanova A, Marcandalli J, Zhou T, Bianchi S, Baxa U, Tsybovsky Y, Lilleri D, Silacci-Fregni C, Foglierini M, Fernandez-Rodriguez BM *et al.*: **Platelet-derived growth factor-alpha receptor is the cellular receptor for human cytomegalovirus gHgLgO trimer**. *Nat Microbiol* 2016, **1**:16082.

30. Satoh T, Arii J, Suenaga T, Wang J, Kogure A, Uehori J, Arase N, Shiratori I, Tanaka S, Kawaguchi Y *et al.*: **PILRalpha is a herpes simplex virus-1 entry coreceptor that associates with glycoprotein B**. *Cell* 2008, **132**:935-944.

31. Frei AP, Jeon OY, Kilcher S, Moest H, Henning LM, Jost C, Pluckthun A, Mercer J, Aebersold R, Carreira EM *et al.*: **Direct identification of ligand-receptor interactions on living cells and tissues**. *Nat Biotechnol* 2012, **30**:997-1001.

32. Frei AP, Moest H, Novy K, Wollscheid B: **Ligand-based receptor identification on living cells and tissues using TRICEPS**. *Nat Protoc* 2013, **8**:1321-1336.

33. Tremblay TL, Hill JJ: **Biotin-transfer from a trifunctional crosslinker for identification of cell surface receptors of soluble protein ligands**. *Sci Rep* 2017, **7**:46574.

34. • Sobotzki N, Schafroth MA, Rudnicka A, Koetemann A, Marty F, Goetze S, Yamauchi Y, Carreira EM, Wollscheid B: **HATRIC-based identification of receptors for orphan ligands**. *Nat Commun* 2018, **9**:1519.

This work describes improvements in Ligand-based receptor capture (LRC) through the development of HATRIC, a chemoproteomic reagent that can be used to detect extracellular protein interactions with glycoylated receptors that are sensitive to changes in pH.

35. Nakayama N, Yokota T, Arai K: **Use of mammalian cell expression cloning systems to identify genes for cytokines, receptors, and regulatory proteins**. *Curr Opin Biotechnol* 1992, **3**:497-505.

36. Davis S, Gale NW, Aldrich TH, Maisonpierre PC, Lhotak V, Pawson T, Goldfarb M, Yancopoulos GD: **Ligands for EPH-related receptor tyrosine kinases that require membrane attachment or clustering for activity**. *Science* 1994, **266**:816-819.

37. Lin HY, Wang XF, Ng-Eaton E, Weinberg RA, Lodish HF: **Expression cloning of the TGF-beta type II receptor, a functional transmembrane serine/threonine kinase**. *Cell* 1992, **68**:775-785.

38. Mathews LS, Vale WW: **Expression cloning of an activin receptor, a predicted transmembrane serine kinase**. *Cell* 1991, **65**:973-982.

39. Pileri P, Uematsu Y, Campagnoli S, Galli G, Falugi F, Petracca R, Weiner AJ, Houghton M, Rosa D, Grandi G *et al.*: **Binding of hepatitis C virus to CD81**. *Science* 1998, **282**:938-941.

40. Evans MJ, von Hahn T, Tscherne DM, Syder AJ, Panis M, Wolk B, Hatziioannou T, McKeating JA, Bieniasz PD, Rice CM: **Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry**. *Nature* 2007, **446**:801-805.

41. Ploss A, Evans MJ, Gaysinskaya VA, Panis M, You H, de Jong YP, Rice CM: **Human occludin is a hepatitis C virus entry factor required for infection of mouse cells**. *Nature* 2009, **457**:882-886.

42. Bianchi E, Doe B, Goulding D, Wright GJ: **Juno is the egg Izumo receptor and is essential for mammalian fertilization**. *Nature* 2014, **508**:483-487.

43. Team MGCP, Temple G, Gerhard DS, Rasooly R, Feingold EA, Good PJ, Robinson C, Mandich A, Derge JG, Lewis J *et al.*: **The completion of the Mammalian Gene Collection (MGC)**. *Genome Res* 2009, **19**:2324-2333.

44. • Collaboration OR: **The ORFeome Collaboration: a genome-scale human ORF-clone resource**. *Nat Methods* 2016, **13**:191-192.

This report summarises the work of the ORFeome Collaboration, an initiative that aims to compile Open Reading frame (ORF) expression vectors covering the whole genome.

45. Lin JC, Ho WH, Gurney A, Rosenthal A: **The netrin-G1 ligand NGL-1 promotes the outgrowth of thalamocortical axons**. *Nat Neurosci* 2003, **6**:1270-1276.

46. Ziauddin J, Sabatini DM: **Microarrays of cells expressing defined cDNAs**. *Nature* 2001, **411**:107-110.

47. Turner L, Lavstsen T, Berger SS, Wang CW, Petersen JE, Avril M, Brazier AJ, Freeth J, Jespersen JS, Nielsen MA *et al.*: **Severe malaria is associated with parasite binding to endothelial protein C receptor**. *Nature* 2013, **498**:502-505.

48. Sosnovtsev SV, Sandoval-Jaime C, Parra GI, Tin CM, Jones RW, Soden J, Barnes D, Freeth J, Smith AW, Green KY: **Identification of human junctional adhesion molecule 1 as a functional receptor for the Hom-1 Calicivirus on human cells**. *mBio* 2017, **8**.

49. •• Mullican SE, Lin-Schmidt X, Chin CN, Chavez JA, Furman JL, Armstrong AA, Beck SC, South VJ, Dinh TQ, Cash-Mason TD *et al.*: **GFRAL is the receptor for GDF15 and the ligand promotes weight loss in mice and nonhuman primates**. *Nat Med* 2017, **23**:1150-1157.

Reverse transection microarray technology was used to screen recombinant GDF15 against ~4500 full length human receptors. GFRAL was discovered as a direct receptor and has important implications in controlling weight gain through control of appetite.

50. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, Heckl D, Ebert BL, Root DE, Doench JG *et al.*: **Genome-scale CRISPR-Cas9 knockout screening in human cells**. *Science* 2014, **343**:84-87.

51. Wang T, Wei JJ, Sabatini DM, Lander ES: **Genetic screens in human cells using the CRISPR-Cas9 system**. *Science* 2014, **343**:80-84.

52. Carette JE, Guimaraes CP, Varadarajan M, Park AS, Wuethrich I, Godarova A, Kotecki M, Cochran BH, Spooner E, Ploegh HL *et al.*: **Haploid genetic screens in human cells identify host factors used by pathogens**. *Science* 2009, **326**:1231-1235.

53. Carette JE, Guimaraes CP, Wuethrich I, Blomen VA, Varadarajan M, Sun C, Bell G, Yuan B, Muellner MK, Nijman SM *et al.*: **Global gene disruption in human cells to assign genes to phenotypes by deep sequencing**. *Nat Biotechnol* 2011, **29**:542-546.

54. Martinez-Martin N: **Technologies for proteome-wide discovery of extracellular host-pathogen interactions**. *J Immunol Res* 2017, **2017** 2197615.

55. Marceau CD, Puschnik AS, Majzoub K, Ooi YS, Brewer SM, Fuchs G, Swaminathan K, Mata MA, Elias JE, Sarnow P *et al.*: **Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens**. *Nature* 2016, **535**:159-163.

56. Savidis G, McDougall WM, Meraner P, Perreira JM, Portmann JM, Trincucci G, John SP, Aker AM, Renzette N, Robbins DR *et al.*: **Identification of Zika Virus and Dengue Virus dependency factors using functional genomics**. *Cell Rep* 2016, **16**:232-246.

57. Park RJ, Wang T, Koundakjian D, Hultquist JF, Lamothe-Molina P, Monel B, Schumann K, Yu H, Krupzcak KM, Garcia-Beltran W *et al.*: **A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors**. *Nat Genet* 2017, **49**:193-203.

58. Staring J, von Castelmur E, Blomen VA, van den Hengel LG, Brockmann M, Baggen J, Thibaut HJ, Nieuwenhuis J, Janssen H, van Kuppeveld FJ *et al.*: **PLA2G16 represents a switch between entry and clearance of Picornaviridae**. *Nature* 2017, **541**:412-416.

59. Orchard RC, Wilen CB, Doench JG, Baldridge MT, McCune BT, Lee YC, Lee S, Pruett-Miller SM, Nelson CA, Fremont DH *et al.*: **Discovery of a proteinaceous cellular receptor for a norovirus**. *Science* 2016, **353**:933-936.

60. Haga K, Fujimoto A, Takai-Todaka R, Miki M, Doan YH, Murakami K, Yokoyama M, Murata K, Nakanishi A, Katayama K: **Functional receptor molecules CD300lf and CD300ld within the CD300 family enable murine noroviruses to infect cells**. *Proc Natl Acad Sci U S A* 2016, **113**:E6248-E6255.

61. Pillay S, Meyer NL, Puschnik AS, Davulcu O, Diep J, Ishikawa Y, Jae LT, Wosen JE, Nagamine CM, Chapman MS *et al.*: **An**

essential receptor for adeno-associated virus infection. *Nature* 2016, **530**:108-112.

62. Raaben M, Jae LT, Herbert AS, Kuehne AI, Stubbs SH, Chou YY, Blomen VA, Kirchhausen T, Dye JM, Brummelkamp TR *et al.*: **NRP2 and CD63 are host factors for Lujo Virus cell entry**. *Cell Host Microbe* 2017, **22**:688-696 e685.

63. • Staring J, van den Hengel LG, Raaben M, Blomen VA, Carette JE, Brummelkamp TR: **KREMEN1 is a host entry receptor for a major group of enteroviruses**. *Cell Host Microbe* 2018, **23**:636-643 e635.

Example of a genome wide haploid screen identifying KREMEN1 as an entry receptor for coxsackie virus A10. Interestingly, KREMEN1 was found to function as a receptor for a distinct branch of type A enteroviruses and holds promise as an interesting therapeutic target.

64. •• Sharma S, Bartholdson SJ, Couch ACM, Yusa K, Wright GJ: **Genome-scale identification of cellular pathways required for cell surface recognition**. *Genome Res* 2018, **28**:1372-1382 http://dx.doi.org/10.1101/gr.231183.117.

In this study, genome wide Cas9/CRISPR technology is used to detect receptors that are required for the binding of antibodies and soluble recombinant ectodomains to cells. In addition to direct receptor interactions, pathways that are required for the functional display of the receptor on the cell surface can also be identified using this method.

65. Sun Y, Vandenbriele C, Kauskot A, Verhamme P, Hoylaerts MF, Wright GJ: **A human platelet receptor protein microarray identifies the high affinity immunoglobulin E receptor subunit alpha (FcepsilonR1alpha) as an activating platelet endothelium aggregation receptor 1 (PEAR1) ligand**. *Mol Cell Proteomics* 2015, **14**:1265-1274.

66. Bausch-Fluck D, Milani ES, Wollscheid B: **Surfaceome nanoscale organization and extracellular interaction networks**. *Curr Opin Chem Biol* 2018, **48**:26-33.

67. Bausch-Fluck D, Goldmann U, Müller S, van Oostrum M, Müller M, Schubert OT, Wollscheid B: **The in silico human surfaceome**. *Proc Natl Acad Sci U S A* 2018 http://dx.doi.org/10.1073/pnas.1808790115.