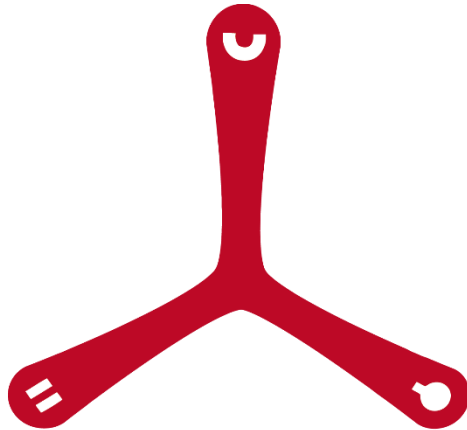


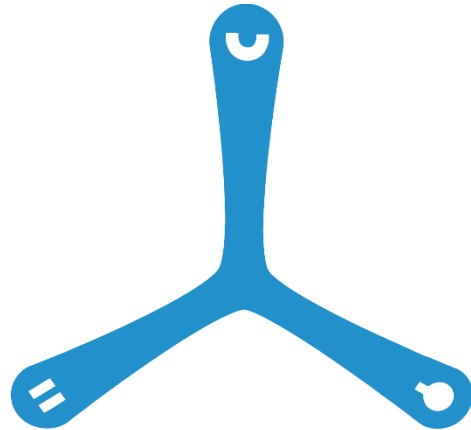
Flow-TriCEPS



Service & Kit

Cell selection and optimization of binding conditions for target identification

LRC-TriCEPS



Service

Identification of the targets and off targets at the cell surface on the living cells

Scratching the surface:

Technologies for targeting the cell surfaceome

Maria P. Pavlou¹ and Paul Helbling¹

1. Dualsystems Biotech AG, Schlieren, Switzerland - www.dualsystems.com

Contact

Dualsystems Biotech AG
Grabenstrasse 11a
CH-8952 Schlieren
Switzerland

maria.pavlou@dualsystems.com
paul.helbling@dualsystems.com

The development of methodologies that enable the study of the surface of a cell – the surfaceome – is pivotal to advancing the understanding of cellular differentiation, and through this will come the development of new, much-needed treatments.

The cell surface membrane – or plasma membrane (PM) – surrounds the cell providing necessary boundaries between the cytoplasm and the extracellular environment. This thin, semi-permeable membrane plays a vital role in protecting the integrity of the cell through selective movement of substances in and out. It also constitutes the base for the attachment of cytoskeleton and cell wall – for bacteria and plants – thereby providing and maintaining the shape of the cell. Moreover, PM allows cells to recognize one another and transmits signalling processes.

The building blocks of the cell membrane are lipids, proteins and their associated sugars. The composition and relative concentration of these molecules define the membrane function and vary among different organisms, cell types and cell states. Based on the fluid mosaic model introduced in 1972 by Singer and Nicolson, the PM is a mosaic of components – primarily phospholipids, cholesterol, proteins and associated carbohydrates – moving freely and fluidly in the plane of the membrane. Although it was thought that the distribution of components is uniform, current data suggest that the cell membrane is highly and tightly organised in heterogeneous microdomains: the maintenance of this heterogeneity is associated with a large energetic cost indicating its significance (1). In support of this hypothesis, perturbations to the lipid composition of the membrane that disrupt the proposed compartmentalisation drastically reduce the efficiency of signal transduction (1).

Introducing the surfaceome

Although lipids and glycans are key components of the PM, the focus of the present review is on the collection of proteins that resides at the cell surface or surfaceome. Surface proteins can be physically embedded in the lipid bilayer (integral), be anchored to the phospholipids or integral proteins at either side of the cell membrane (peripheral) or even associate to the membrane only under specific conditions. The surfaceome constitutes roughly 50% of the PM mass (2) and exhibits a wide variety of functions. These include transport, enzymatic activity, signal transduction, cell-cell interaction and attachment to the cytoskeleton or the extracellular matrix. Different classes of surface proteins carry out these tasks for example channel and carrier proteins, enzymes, receptors, cell recognition and cell adhesion proteins.

Given the range of functions carried out by surface proteins, it is not surprising that roughly 30% of predicted open reading frames in a typical genome encode membrane and PM proteins (2).

The surfaceome content differs among cell types and changes during developmental and disease states. Therefore, it contains unique markers that can be used to distinguish cellular phenotypes and disease states. These properties along with the fact that cell surface proteins are readily available make the surfaceome a rich source of phenotypic, diagnostic, prognostic and therapeutic targets that can be used in a variety of fields including oncology, immunology and stem cell research.

Overall, the development of methodologies that enable the study of the surfaceome is pivotal to advance our understanding regarding cellular differentiation and development, host-pathogen interactions and metastatic processes, and will lead to the development of new treatments.

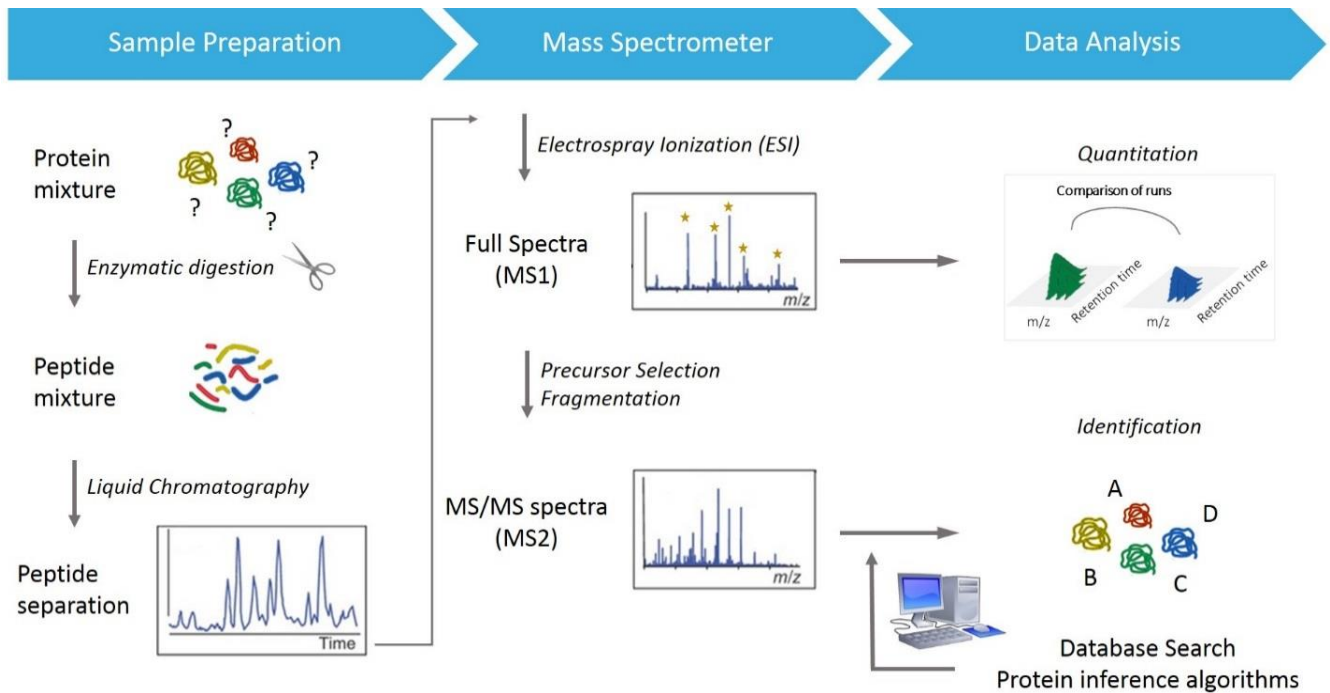
Discovery-based approaches

To explore the surfaceome content, one must consider discovery-based approaches.

Transcriptomics coupled to in-silico predictions

Deciphering the contents of the surfaceome can be approached in different ways. One genome-based approach is based on the *in-silico* prediction of surface proteins followed by transcriptional profiling of a given cell type to measure the expression of these genes (3, 4). Although whole transcriptome analysis is readily available by next generation deep sequencing, even at the single-cell level (5), the mRNA expression does not correlate fully to protein expression. The correlation between mRNA and protein levels is even lower for certain protein classes, including surface proteins (6). Even in instances when mRNA expression can accurately reflect protein expression, information regarding protein abundance, location, post-translational modifications and isoforms cannot be extracted. The above-mentioned properties impact the functions and signalling potential of proteins directly, however. This observation highlights the need for systematic protein-centric analyses of the surfaceome.

Figure 1: Typical workflow of a bottom-up MS-based proteomics experiment



Extracted proteins are digested into peptides with proteases such as trypsin. Enzymatic digestion of a complex sample can generate an enormous number of peptides that cannot be analyzed by MS directly. Hence, the liquid chromatography (LC) separation is usually performed to reduce the sample complexity before the MS analysis. The LC is coupled online to the MS and upon elution peptides are ionized. After ionization, peptide precursor ions are introduced into the MS, which records both their mass-to-charge (m/z) ratio and intensity (MS1). Next, the most intense precursors are selected for further fragmentation to generate tandem mass spectra (MS/MS). During data analysis, the combination of precursor m/z and its tandem mass spectrum along with protein database searches are used to determine peptide sequences, and then proteins are inferred from the identified peptides. Finally, peptides and proteins are quantified based on MS1 intensity to estimate protein abundance. Comparison of MS1 intensities of the same peptide/protein across different samples can reveal the differential abundance of the protein.

Flow cytometry antibody-based

Flow cytometry (FC) is a simple and robust method to assess cell surface protein expression on a single-cell level. FC utilizes fluorescently-tagged monoclonal antibodies to label surface proteins on living cells and has contributed immensely in the phenotypic characterization of multiple cell types. Typically, only a handful of proteins can be simultaneously analysed by FC but, recently a high-throughput FC screen has been developed composed of almost 370 fluorescent antibodies arrayed in 96 well plates (7). The prevalent advantage of this methodology is the ability to perform cell surface analysis of selected subpopulations within complex cell mixtures in a specific, sensitive, reproducible and rapid manner. Nevertheless, it entirely depends on the availability of high quality monoclonal antibodies that apparently

cover only a small subset of the estimated 2000-2500 cell surface proteins.

Mass spectrometry-based

An unbiased and global protein study can be predominately performed through mass spectrometry (MS) based proteomics. The evolution of MS-based proteomics over the last two decades allows the analyses of many characteristics of proteins such as sequence, quantity, modifications, structure and macromolecular organisation in unprecedented depth (8).

In MS-based approaches, proteins can be studied as intact entities – top-down proteomics – or as peptides – bottom-up proteomics. For experimental and computational reasons the bottom-up approach is followed the most. During a bottom-up experiment, proteins are extracted from the sample of interest and

digested into peptides by sequence-based proteases such as trypsin. The peptide mixture is then separated by reverse-phase chromatography coupled online to electrospray ionization. The resulting peptide ions are transferred into the mass spectrometer where their mass-to-charge ratio is detected and recorded. These ions are then fragmented to generate ions that contain the necessary information to identify the peptide sequence. The resulting data are analysed by sophisticated algorithms to provide information regarding the protein identity and quantity in the sample (Figure 1).

Although MS-based proteomics allows for almost complete characterisation of proteomes (8) the properties of surface proteins still pose a technological challenge. First, the inherent hydrophobicity of these (multi)-transmembrane proteins confers them as poorly soluble in the aqueous solutions used for sample preparation prior to MS analysis. Additionally, the overall low cellular abundance in combination with the highly abundant cytoplasmic proteins brings surface proteins close to the limit of detection of MS. Finally, separating the PM proteins from intracellular membrane proteins is challenging. Therefore relative enrichment of the cell surface proteins is necessary.

Historically, PM has been isolated by subcellular fractionation using centrifugation. During differential centrifugation, the components of cell lysates are separated based on size differences, with larger particles found at the bottom. Alternatively, centrifugation can take place in a density gradient resulting in higher resolution. Relatively low yields and limited specificity – due to cross-contamination from other organelles with similar density – are key limitations of centrifugation.

Biotinylation of surface proteins has been a popular alternative to surfaceome isolation for subsequent analysis. In principle, surface proteins are labelled covalently with a biotinylation reagent and purified using avidin/streptavidin interaction for subsequent MS analysis. Although the highly stable interaction of biotin and avidin/streptavidin offers high selectivity it also poses a major challenge for elution of the attached proteins.

Modern chemoproteomic strategies can now provide a comprehensive view of the surfaceome. The cell surface capture (CSC) methodology in particular developed by Wollscheid *et al.* (9), takes advantage of the prediction that >90% of surface proteins are glycosylated (10) and provides a workflow for surfaceome analysis with high selectivity and broad

applicability. Experimentally, extracellular sugars on living cells are mildly oxidised to generate aldehyde groups that are targeted by a membrane impermeable bi-linker that is covalently linked to the aldehydes, resulting in hydrazone formation. Cells are then lysed, proteins are digested and the resulting biotinylated peptides are enriched through streptavidin beads. Glycopeptides are enzymatically released from the beads using PNGase F treatment and de-glycosylated peptides are analysed by MS. CSC is performed on living cells and provides information regarding not only surface protein identity but also transmembrane orientation and site of glycosylation (11).

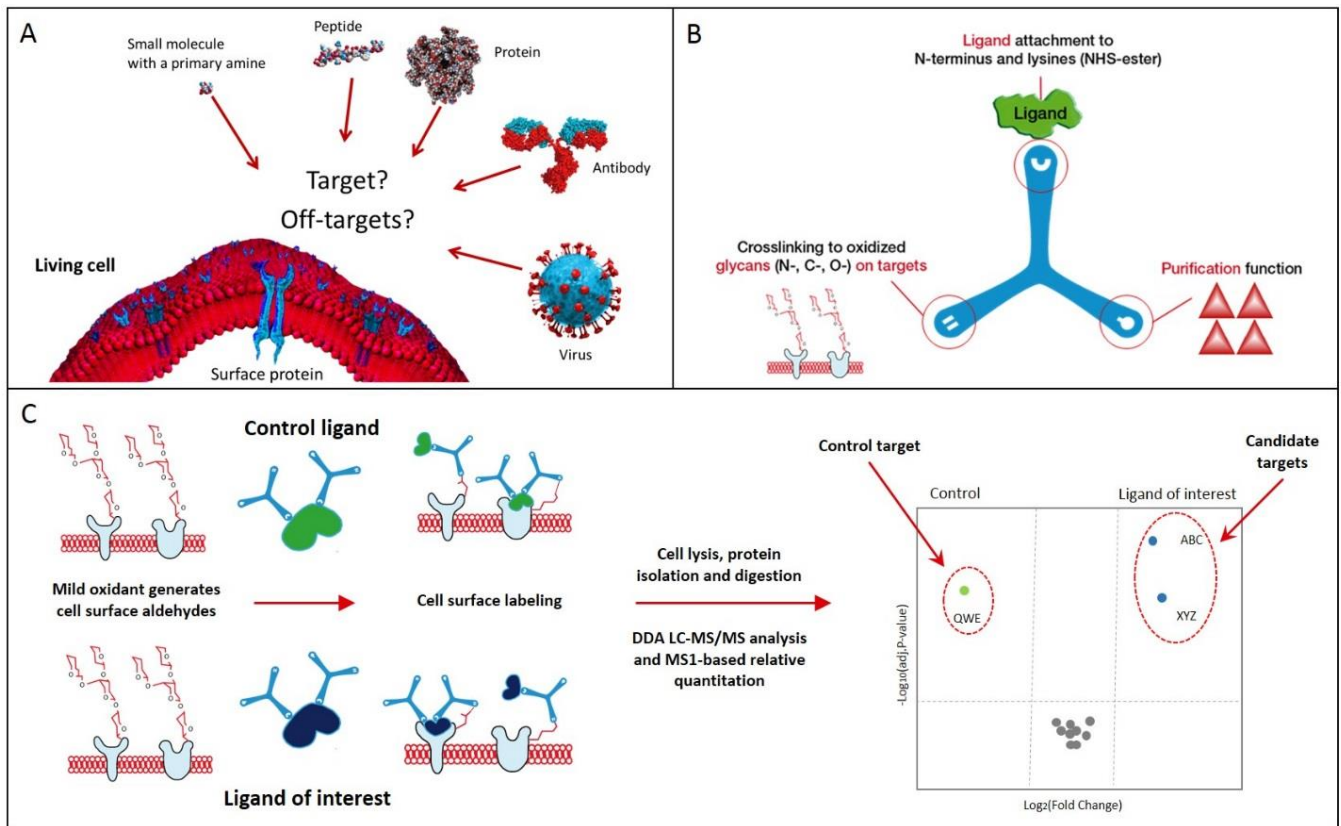
Overall MS-based analyses of the surfaceome have significantly expanded our knowledge regarding surfaceome content and quantity in a multitude of cell types. Noteworthy MS-based methodologies provide an average measure of the entire sample unlike the single cell resolution of flow cytometry.

Approaches to explore protein-protein interactions at the cell surface

The original concept, depicting the PM as a homogeneous fluid bilayer with freely diffusing proteins (see section 2), has been evolved to another depicting a highly organised and crowded mosaic of interacting lipids and glycoproteins. This higher organisation modulates the biological processes occurring on the cell surface, exemplified by receptors being active only when they form dimers, heterodimers or higher order oligomers (12). The importance of protein interaction networks is now well-recognised and a large number of methodologies aiming to build comprehensive protein interactome maps have been developed. However, the protein interactions taking place at the cell surface are generally under-represented.

Imaging can provide valuable insights into the macromolecular organisation of the cell surface. Fluorescence microscopy in combination with genetically expressed fluorescent proteins can be applied to map protein interactions with typical spatial resolution below 500nm (13). In particular, Forster resonance energy transfer (FRET) is a widely used fluorescence technique to study bi-molecular interactions in living cells. FRET involves the non-radiative energy transfer from a fluorescent donor to an acceptor that can happen only when the two fluorophores are situated at distances less than 10nm.

Figure 2: Outline of the LRC-TriCEPS methodology



A. The LRC-TriCEPS methodology enables the identification of targets and off-targets of orphan ligands on the surface of living cells. Ligands can range from small molecules to whole viruses.

B. TriCEPS is a chemoproteomic reagent with three moieties; one that binds ligands containing an amino group, a second that binds glycosylated receptors on living cells and a function for purifying the receptor peptides for MS-based identification and quantitation.

C. In a typical LRC-TriCEPS experiment, at least two treatment arms are performed in parallel: one with the ligand of interest and a second with a control ligand (that is, a ligand with a known target). The first step of the LRC includes the conjugation of the two ligands to TriCEPS. The TriCEPS-ligand conjugates are then incubated with previously oxidized cells under near-physiological conditions. During this phase, transient and stable ligand-receptor interactions will result in covalent capture events between TRICEPS and nearby carbohydrates. After the receptor-capture reaction, the cells are lysed, proteins are isolated and processed for MS-based analysis. Upon identification, the relative abundance of cell surface proteins in the ligand samples are compared to those in the control sample using MS1-based label-free quantification. Randomly identified cell surface proteins are expected to have equal abundance in both samples, whereas the corresponding receptors are found enriched in the ligand sample. The results can be shown as a volcano plot. A volcano plot combines a measure of statistical significance with the magnitude of the change, enabling quick visual identification of proteins that display changes that are also statistically significant. The x-axis represents the mean ratio fold change (on a Log₂ scale). The y-axis represents the statistical significance p-value of the ratio fold change for each protein (on a -Log₁₀ scale). Proteins that are enriched in one of the samples, will plot either left or right of the x-axis origin, indicating in which sample that protein is enriched.

In FRET experiments, the membrane proteins of interest are tagged with genetically encoded donor and acceptor fluorescent proteins and energy transfer occurs only when the two proteins are close to each other (13). Notably, imaging-based methodologies like FRET provide information on proximity rather than direct protein interactions and require genetic manipulation.

Affinity-based approaches such as immunoprecipitation coupled to MS-based analysis have been widely utilised to decipher protein-protein interactions. However, these approaches commonly use cell lysates and are not optimal for surface proteins, given that the spatial organisation of the proteins is disrupted. For this reason, proximity-dependent labelling methodologies have been developed over the past 5 years to investigate the membrane protein interactions on living cells (14). These methodologies use enzymes which are either fused to the protein of interest genetically or brought to the protein of interest via antibodies. These enzymes are then used to attach a tag covalently on proximal proteins, which allows their subsequent purification and identification by MS. Examples of such techniques include selective proteomic proximity labelling using tyramide (SPPLAT) (15), ascorbate peroxidase (APEX) (16), biotinylation by antibody recognition (BAR) (17) and proximity-dependent biotin identification (BioID) (18).

Although these novel methodologies are continuously advancing our understanding of surfaceome organisation, they do not necessarily provide information regarding the extracellular interactions taking place on the cell surface. Identifying the targets of key ligands provides valuable mechanistic information about signal transduction, drug action or off-target effects. For instance, pathogen or growth factor interactions are important for developing novel therapies. Additionally, numerous ligands exist – both biologics and small molecules – involved in biological functions mediated at the cell surface through still unknown protein targets. Recognising this unmet need, the Wollscheid group, building on the CSC approach, developed the Ligand Receptor Capture (LRC) methodology (Figure 2A) (19).

The key component of the LRC methodology is TriCEPS, a chemoproteomic reagent with three moieties; one that binds ligands containing an amino group, a second that binds glycosylated receptors on living cells and a function for purifying the receptor peptides for MS-based identification and quantitation

(Figure 2B). In a typical LRC-TriCEPS experiment, at least two treatment arms are performed in parallel: one with the ligand of interest and a second with a control ligand (that is, a ligand with a known target). The first step of the LRC includes the conjugation of the two ligands to TriCEPS. The TriCEPS-ligand conjugates are then incubated with previously oxidized cells under near-physiological conditions. During this phase, transient and stable ligand-receptor interactions will result in covalent capture events between TriCEPS and nearby carbohydrates. After the receptor-capture reaction, the cells are lysed, TriCEPS-labelled N-glycopeptides – similar to CSC – or whole TriCEPS-labelled proteins (Sobotzki *et al*, under review) are isolated and processed for MS-based analysis. Upon identification, the relative abundance of cell surface proteins in the ligand samples are compared to those in the control sample using MS1-based label-free quantification. Randomly identified cell surface proteins are expected to have equal abundance in both samples, whereas the corresponding receptors are found enriched in the ligand sample (Figure 2C).

The LRC methodology is unique in several ways. First, it can be applied in a multitude of ligands ranging from small molecules, to peptides, proteins, antibodies and even whole viruses. Second, it identifies the targets of these ligands on the surface of living cells under physiological conditions. This point is critical for cell surface proteins that typically need to be embedded in their natural environment – living cells – to exhibit their characteristic binding properties. Third, it does not require any genetic manipulations and therefore can be applied on a multitude of cell lines – including primary cells – and even tissues. Fourth, it is hypothesis-free meaning that no previous knowledge or speculation on the target is required. Finally, it allows the discovery of not only one-to-one but also one-to-many and many-to-many interactions in a single experiment.

Conclusion and outlook

If the plasma membrane is considered the gateway through which cells communicate and interact with their environment, then proteins associated with the surface – referred to throughout as surfaceome – can be seen as the gatekeepers. Studying surfaceome biology and function can contribute to the discovery of new therapeutic targets and new immunophenotypic markers. Because of the vital role of the surfaceome in all functions performed by the PM, numerous methodologies have been developed

to shed light on the content, quantity and interaction of these proteins. Methodologies that enable examining the surfaceome in its natural environment are advantageous since the organisation of the PM in living cells is tight. Further improvements in resolution and sensitivity may allow future examination of the distinct microenvironments existing on the PM of a single cell.

References

1. Machta BB. et al, Critical Casimir forces in cellular membranes, *Phys Rev Lett.* 109(13):138101, 2012
2. Leth-Larsen R. et al, Plasma Membrane Proteomics and Its Application in Clinical Cancer Biomarker Discovery, *Mol Cell Proteomics.* 9(7):1369-82, 2010
3. da Cunha J. P. et al, Bioinformatics construction of the human cell surfaceome, *Proc Natl Acad Sci U S A.* 106(39):16752-7, 2009
4. Town J. et al, Exploring the surfaceome of Ewing sarcoma identifies a new and unique therapeutic target, *Proc Natl Acad Sci U S A.* 113(13):3603-8, 2016
5. Wang, Z. et al, RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10 (1), 57-63, 2009
6. Schwanhäusser B et al, Global quantification of mammalian gene expression control, *Nature* 473(7347):337-42, 2011
7. Paterson J, Ailles LE, High Throughput Flow Cytometry for Cell Surface Profiling, *Methods Mol Biol.* 1678:111-138, 2018
8. Aebersold R and Mann M, Mass-spectrometric exploration of proteome structure and function, *Nature*, 537(7620):347-55, 2016
9. Wollscheid B et al, Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins, *Nat Biotechnol.* 27(4):378-86, 2009
10. Gahmberg CG and Tolvanen M, Why mammalian cell surface proteins are glycoproteins, *Trends Biochem Sci.* 21(8):308-11, 1996
11. Bausch-Fluck D et al, A mass spectrometric-derived cell surface protein atlas, *PLoS One.* 10(3):e0121314, 2015
12. Milligan G, G protein-coupled receptor dimerisation: molecular basis and relevance to function, *Biochim Biophys Acta.* 1768(4):825-35, 2007
13. Margineanu A, Screening for protein-protein interactions using Förster resonance energy transfer (FRET) and fluorescence lifetime imaging microscopy (FLIM), *Sci Rep.* 6:28186, 2016
14. Chen CL1 and Perrimon N, Proximity-dependent labeling methods for proteomic profiling in living cells, *Wiley Interdiscip Rev Dev Biol.* 6(4), 2017
15. Rees JS et al, Selective proteomic proximity labelling assay using tyramide (SPPLAT): A quantitative method for the proteomic analysis of localized membrane-bound protein clusters, *Curr Protoc Protein Sci.* 80:19.27.1-18, 2015
16. Lobingier BT et al, An approach to spatiotemporally resolve protein interaction networks in living cells, *Cell.* 169(2):350-360, 2017
17. Bar DZ et al, Biotinylation by antibody recognition - A novel method for proximity labelling, *Nat Methods.* 15(2):127-133, 2018
18. Roux KJ et al, BioID: a screen for protein-protein interactions, *Curr Protoc Protein Sci.* 5;74, 2013
19. Frei AP et al, Direct identification of ligand-receptor interactions on living cells and tissues, *Nat Biotechnol.* 30(10):997-1001, 2012