Perspective

Application Guide of OMICs Approaches to Cell Signaling

Zhong Yao[1*], Julia Petschnigg[2*], Robin Ketteler[2], and Igor Stagljar[1,3,4 #]

[1] Donnelly Centre, University of Toronto, Ontario, Canada

[2] MRC-Laboratory for Molecular Cell Biology, University College London, UK

[3] Department of Molecular Genetics, University of Toronto, Ontario, Canada

[4] Department of Biochemistry, University of Toronto, Ontario, Canada

* equal contribution

[#]corresponding author:

Igor Stagljar

Phone: 1-416-946-7828

Fax: 1-416-978-8287

E-mail: igor.stagljar@utoronto.ca

Abstract

Research in signal transduction aims to identify the functions of different signaling pathways in physiological and pathological states. Traditional techniques using biochemical, genetic or cell biological approaches have made important contributions to our understanding of cellular signaling. However, the single-gene approach does not take into account the whole complexity of cell signaling. With the availability of OMICs-techniques, great progress has been made in understanding signaling networks. OMICs approaches can be classified into two categories: "molecular profiling", including genomic-, proteomic-, post-translational modification- and interactome-profiling; and "molecular perturbation", including genetic and functional perturbations.

Due to the ever-growing field of method development to characterize cellular processes on genomic and proteomic levels and in many other dimensions, it has become a challenge to select and apply the appropriate methods suitable for addressing a specific biological question. In this perspective, we will describe some selected OMICs-techniques and discuss their applicability to cell signaling research.

Introduction

Cell signaling (signal transduction) relays extracellular and internal signals to different cellular compartments, and regulates various cellular reactions in response to environmental and intracellular changes. It plays essential roles in almost all cellular functions and is carried out by multiple parallel pathways.

In general, despite diverse and complex pathways, signal transduction proceeds through five distinct steps: (i) recognition of a signal – typically in the form of a receptor-ligand engagement; (ii) conversion of a signal into biochemical imprints such as phosphorylation or other post-translational modifications (PTMs); (iii) relay of a signal by engagement of binding partners and subsequent transmission of information through either changes in conformation or biochemical imprints; (iv) signal processing to maintain the robustness of information transmission and signal integration to synchronize multiple signal inputs; (v) conversion of the signal into a biological response such as the transcription of target genes or the synthesis of proteins and metabolites (**Fig. 1**).

Understanding each of these steps lies at the heart of cell signaling research and requires distinct technologies in order to achieve this.

Cells use different mechanisms such as (a) PTMs, (b) protein-protein interactions (PPIs), and (c) changes in localization (translocation) to relay and process signals. Phosphorylation is one of the most common PTMs[1]. It functions through inducing conformational change in a target protein, or changing the accessibility between enzyme and interacting protein[2]. Such mechanisms allow for a change of enzyme activity, or changes in interactions between host proteins and other molecules and hence, provide a tuneable process for signal propagation and termination. Ubiquitylation, sumoylation or acetylation are other important PTMs[3-5], each leading to a very distinct response such as regulating protein stability, enzyme activity or modifying protein-protein interactions. The most common techniques to monitor PTMs are based on mass spectrometry (MS) as outlined in more detail below.

PPIs are often achieved through specialized domains and play a crucial role in signal transduction[6, 7]. PPIs can regulate enzyme activity through allosteric effect or direct interaction with active sites. Interaction of a protein with scaffold/adaptor proteins brings signaling molecules close to each other, or targets them to the right cellular location. Sequestering signaling molecules from other signaling components or locations secures

signaling specificity and reduces signaling noise. Different types of PPIs with various physical characteristics such as strong/weak and stable/transient interactions all play roles in signaling and have different biological advantages. Various PPI-assays amenable to large-scale OMICs measurement will be discussed below.

Translocation of signaling molecules, required for accurate and efficient signal processing, is usually carried out by facilitated/active transportation machineries[8]. Changes in localization are usually visualized by microscopy-based techniques such as high-content screening.

Although most signaling pathways display features of serial processing, they are usually not linear[9]. It is common for several upstream pathways to converge into one common downstream pathway, or one upstream pathway to diverge into several downstream pathways. Existence of protein isoforms or different molecules with similar functions creates pathway redundancy. Signaling circuits such as feedforward- and feedback-loops are also present in many pathways. Moreover, parallel pathways can communicate with each other through lateral crosstalk. These features enable a complex, robust, and flexible signaling network, making it resistant to signal noise but able to respond to environmental and cellular alterations. **Box 1** highlights some major signaling pathways.

Applying OMICs techniques to identify cell signaling networks in health and disease can help identify sites for therapeutic intervention and define potential drug targets. In general, OMICs approaches are defined as high-throughput technologies that aim to generate a comprehensive view of molecular cell components. They generally aim at the universal detection of genes (genomics), mRNA (transcriptomics), proteins (proteomics), metabolites (metabolomics) or lipids (lipidomics) in a specific biological sample.

OMICs approaches comprise molecular profiling techniques that allow the capture of protein and gene components and their interactions. They are often complemented by perturbation methods that probe the gain or loss of such components in order to understand their functional role in signaling pathways. A third approach, the computational reconstruction of networks, allows assessment of signal integration and processing[10, 11].

In this perspective, we will mainly discuss OMICs approaches applied to measure PPIs and protein-based cell signaling. This perspective does not aim to cover all OMICs techniques; instead, we only try to introduce the more recently developed approaches as well as some already existing techniques that had important impacts on cell signaling

research. As this is a very complex research area, we recommend other review articles that specifically cover other aspects of this fast growing research field[11-13]. Many techniques are now available that address a specific aspect in the signaling cascade and we provide a guide for the reader to identify which method for profiling/perturbation is best suited for the question they want to address and additionally, which key properties and intrinsic limitations each approach has (**Table 1** and **2**).

Molecular Profiling: Assessing protein networks

Proteomic profiling based on mass spectrometry (MS)

MS is an important proteomics tool and is widely used for protein, PTM and protein complex profiling in near genomics-wide scales. For example, tandem affinity purification (TAP) has achieved great success in protein complex identification[14]. An impressive development is quantitative MS, which allows in-depth analysis of quantitative traits of signaling processes. Quantification can be accomplished through isotopic labeling or a label-free approach. In the first approach, stable isotopes are introduced into proteins/peptides through metabolic labelling (SILAC) or direct chemical labelling such as ICAT and iTRAQ[15] **(Fig. 2a).** However, flaws include relatively narrow dynamic range and high expense.

A label-free MS, selected reaction monitoring (SRM), has been used for analysis of small molecules for decades (**Fig. 2a**). It was adapted to quantitative MS analysis of the proteome[16]. The instrument setting of SRM is composed of HPLC followed by triple quadrupoles. Targeted peptides are selected by the first quadrupole, followed by fragmentation in the second quadrupole and fragment selection in the third. Thus, the MS signal intensities are recorded for each selected transition (precursor/fragment ion pair) and the abundance of each precursor peptide is calculated by integrating the intensities of correspondent ions to the retention peak. Several features make it an attractive method for quantitative studies: high selectivity, high sensitivity at subfemtomolar levels, a wide dynamic range with 4-5 orders of magnitude, high reproducibility, and multiplexity. Multiple reaction monitoring (MRM) indicates the same technology, but emphasizes its multiplex capability. Compared to normal stochastic sampling MS, SRM focuses only on preselected target proteins, and therefore favors hypothesis-driven studies. SRM/MRM has successfully tackled some important signaling issues, which have previously been difficult to address by

classical methods. Recently, SRM/MRM was used to characterize the kinetics of receptor complexes during EGFR signaling processes[17, 18]. Nevertheless, assay development for SRM is time- and labour-intensive.

SWATH is a label-free quantitative MS method[19] (**Fig. 2a**). It resembles SRM with MS/MS settings. However, SRM processes samples through a data-independent acquisition (DIA) approach. In contrast, SWATH scans and fragments all the precursor peptides within a range from the first MS. The entire peptide ions are recorded in the second MS. The target peptides can be extracted with an algorithm similar to SRM by correlating them to a known peptide set. Thus, SWATH complements SRM and traditional MS. On one side, it keeps similar in-depth features of SRM such as sensitivity, reproducibility and dynamic range. On the other side, it harbors the same coverage as traditional MS. Furthermore, assay development required for SRM is not necessary for SWATH. The power of SWATH was exemplified by two studies on the dynamics of 14-3-3 complex[20] and several other complexes[21].

Flow cytometry studies multiple molecular events simultaneously at the single cell level. Traditionally, it is facilitated by fluorescent labelling. Nevertheless, fluorescent flow cytometry is confined by the limitation of the number of fluorescent labellings due to spectral overlap. Mass cytometry, or cytometry time-of-flight (CyTOF), is a new type of cytometry[22] (**Fig. 2b**). In contrast to fluorescent cytometry, it uses MS as readout. For this purpose, cells are labelled with multiple antibodies tagged with different transition element isotopes. Each labelled cell is nebulized and the metals are ionized. The abundance of each metal is analyzed and quantified by time-of-flight MS. The striking feature of this method is its considerable multiplexity derived from high resolution in a relatively wide range of measurement. Theoretically, it can reach close to 100 measurements for each sample, whereas practically, 30-40 measurements have been achieved. Another advantage is low background, because the tagged transition elements are usually absent in cells. In contrast, various levels of autofluorescence, dependent on cell type, usually elevate the signal baseline in fluorescent cytometry. Mathematical compensation is also needed in fluorescent cytometry when using multiple labelling due to spectral overlap, which is not necessary for mass cytometry. Mass cytometry produces a large amount of data and thereby needs

extensive bioinformatics analysis. It has been successfully used in different studies such as single cell characterization, signaling and the impacts of small molecular regulators[23, 24].

Protein-small molecule profiling: Chemoproteomics

Small molecule compounds have proven very useful to perturb or probe signal transduction pathways and have the potential to become therapeutic agents as activators or inhibitors of such pathways. Chemical proteomics, or chemoproteomics, studies such protein/small molecule interaction at the proteomics level (**Fig. 2c**) and is a combination of both profiling and perturbation methods. There are two chemoproteomics approaches, activity-based protein profiling (ABPP)[25] and compound-centric chemical proteomics (CCCP)[26]. ABPP utilizes active site-directed probes toward a specific enzyme group. In the CCCP approach, bioactive small molecules such as drugs are immobilized to a solid matrix. For both approaches, the target proteins are identified by quantitative MS. Compared to traditional techniques, chemoproteomics excels with its unbiasedness and proteome-wide coverage. However, its limitations should be considered. First, it is difficult to immobilize some chemicals without affecting their biological function. Second, some targets are underrepresented due to their physical properties or their low abundance in cells. Third, nonspecific interaction is a major challenge caused by abundant proteins or interactions with unrelated parts of probes or compounds. Competition by pre-incubation with soluble compounds can partially overcome this problem. Competition can also be used for drug screening, or for IC50 estimation[27].

Chemoproteomics is usually employed for target identification. This was perfectly exemplified by a recent study[28] that identified human mutT homologue MTH1 as the target for a previously mechanistically elusive antitumor compound SCH51344. Since MTH1 has been previously demonstrated to be involved in preventing reactive oxygen species-induced DNA damage, this study revealed the role of DNA repair in the maintenance of cancer state in some tumors. Chemoproteomics can also directly serve as a tool for studying signal transduction, especially for characterizing some subsets of signaling enzymes such as the kinome and phosphatome. This can be achieved by an ABPP approach using probes targeting specific enzymes. For targets such as kinases, whose general activity probes are difficult to develop, multiple kinase inhibitors have been used as affinity reagents with high coverage of the kinome[29]. Using such a platform, a recent study investigated the dynamic

changes of the kinome in response to MAPK pathway inhibition in triple-negative breast cancer cells during kinase inhibitor treatment[30].

Interactome profiling: PPI-approaches

Most cellular processes, including cell signaling, rely on the formation of protein-complexes and the crosstalk of proteins within the same or another complex. Defining the interaction patterns of proteins can give valuable insight into the function of a protein and can define new drug targets.

To date, many interactome-studies rely on biochemical MS-based approaches[31], which directly address PPIs in protein complexes, but require extensive optimization and under-represent weak and transient interactions. An alternative is the classical yeast-two-hybrid system[32], which is a robust method for monitoring PPIs reconstituted in yeast cells. One limitation is that interactions are forced to take place in the nucleus and many mammalian proteins cannot be assayed in yeast. Over the years, a plethora of PPI-techniques have emerged, which led to improvement of throughput and accessibility to study interactions that have been difficult to study, such as between membrane proteins, post-translationally modified proteins and transient interactions[7].

The challenge of next-generation PPI-studies is to move from a static view of protein complexes to a dynamic overview of proteins that change in response to various stimuli or are deranged in diseases. As most aberrant signaling pathways are the result of network rewiring, *i.e.* proteins interacting with different effector proteins, or a change in PTMs such as phosphorylation, this stresses the need for methods that can probe PPIs specific to disease-states or that can be perturbed by drugs. Here, we highlight some of the PPI techniques that allow for uncovering disease-related interactions that occur in a phosphorylation-dependent manner, can be modified by various conditions, or that allow for defining drug-protein interactions.

Genetic systems such as protein fragment complementation (PCA) assays and their variants such as split-TEV (**Fig. 3a**) have been used to detect ligand-dependent interactions of ErbB family receptors, ligand-induced G protein–coupled receptor (GPCR) activation and hormone-induced ErbB-heterodimerization[33-35]. The principle of PCAs exploits the spontaneous folding of two split-halves of enzymes once the proteins they are fused to

interact[36]. A versatile variant of PCA is BiFC (bimolecular fluorescence complementation, **Fig. 3a**), which uses halves of fluorescent proteins and has been applied to functional investigation of TGF–β and insulin signaling pathways[37]. It has the advantage of allowing direct visualization of the site of protein interactions. Recently, the irreversible nature of fluorescent PCAs has been overcome by the development of a novel PCA based on *Deinococcus radiodurans* infrared protein IFP1.4, which has been successfully applied to reconstitute the temporal EGF-dependent interaction between Shc1 and Grb2 and allows for detection of spatio-temporal dynamics of PPIs, also at low abundances[38].

Another powerful PPI method is the mammalian protein-protein interaction trap (MAPPIT) (**Fig. 3b**), which is based on the premise that a dysfunctional JAK-STAT signaling pathway is restored upon a specific bait and prey interaction. Since its development, MAPPIT has been further developed to allow for identification of modification-dependent interactions (heterotrimeric MAPPIT) or identification of small molecule compounds that specifically disrupt interactions (reverse MAPPIT)[39, 40]. Another recently developed two-hybrid-based method, which shares similarities with the classical MAPPIT assay, is KISS (kinase substrate sensor) (**Fig. 3b**), which allows for *in situ* analysis of interactions in response to physiological challenges[41]. In KISS, the bait protein is fused to a kinase-containing portion of TYK2 and the prey is coupled to a gp130 cytokine receptor fragment. Upon bait and prey interaction, TYK2 phosphorylates STAT3 docking sites on the prey chimera, resulting in reporter gene activation. KISS has so far been applied to determine external stimuli, such as agonist-dependent interactions between GPCRs and β-arrestins. Furthermore, KISS allows for assaying pharmacological disruption of PPIs. In the three-hybrid KISS set-up small molecules can be presented as baits inside cells and can thus be assayed for their interaction with target prey proteins, which offers a potential novel drug-screening platform.

The mammalian membrane two-hybrid (MaMTH)[42] is a recently developed PCA-variant based on split-ubiquitin reconstitution and is derived from the original membrane-yeast two-hybrid system (MYTH)[43, 44] (**Fig. 3c**). It uses split-halves of ubiquitin attached to an integral membrane protein-bait (in addition to a transcription factor) and a prey. Bait-prey interaction results in pseudo-ubiquitin formation that can be cleaved by de-ubiquitinating

enzymes. MaMTH detects reporter activity once the transcription factor has been cleaved off the bait, which only occurs if bait and prey interact. MaMTH adds novel features to existing *in vivo* PPI-techniques. It can be used to identify phosphorylated residues on integral membrane proteins such as receptor tyrosine kinases (RTKs) that confer interactions with adaptor/effector proteins. To date, phosphorylated sites of ErbB receptors have only been detectable using MS-based and anti-phosphotyrosine antibody-based methods. MaMTH was applied to identify the increased activation state of various onco-genic ErbB family members via measurement of adaptor-protein interaction. This adaptor-protein recruitment as a sensor for activity status of RTKs also allows for testing novel drugs that can specifically inhibit oncogenic, hyperactive signatures of RTKs.

Limitations for MaMTH and MAPPIT/KISS are that they are reporter-based systems. Thus, relaying the signal from the site of interaction to reporter gene activation reflects accumulation of luciferase over time, rather than real-time situations. Amplification of signal through the luciferase reporter read-out poses an advantage, as weak and transient interaction can still be assayed.

It should be kept in mind that proteins in all above-mentioned approaches are mostly overexpressed, which might be useful for weak or transient interactions, but can make these assays prone to false positives.

Proximity assays include optical methods based on resonance-energy transfer, such as between fluorescent or bioluminescent proteins fused to interacting proteins (FRET, fluorescence or BRET, bioluminescence resonance energy transfer methods), which have been widely used to study GPCR signaling[45]. The proximity ligation assay, PLA[46] (**Fig. 3d**), is an *in situ* method that allows for detection of PPIs, PTMs and protein-nucleic acid interactions in fixed cells and tissues. The technique allows for detection of proteins on a single-cell level and can visualize compartmentalization of PPIs in a very sensitive manner.

A pair of proximity probes (primary antibodies with a conjugated oligonucleotide) targets the proteins of interest. Then, a connector oligonucleotide is added and connects both proximity probes upon interaction (or close proximity) of the proteins. The connector functions as a bridge for the enzymatic ligation reaction of the oligonucleotides on the proximity probes. This produces a new DNA molecule that serves as a template for rolling-

circle amplification (RCA), replicating the DNA multifold, which can then be visualized by attachment of fluorescent oligonucleotides. This results in powerful signal amplification[47].

Multiplex PLA allows for parallel visualization of various protein complexes *in situ*[48]. This improved PLA method can detect partners within protein complexes and their subcellular localization. *In situ* PLA has been successfully applied to detect interactions between ErbB family members and to probe anticancer drugs for their potential to disrupt ErbB homo- and heterodimer formation and to influence their phosphorylation status. In a recent study, PLA was further developed to measure EGFR-associated signaling complexes from patient-derived materials[49]. PLA can be used on cell lysates and tissue samples, which is of great importance for clinical studies. A limitation is that the assay requires cell permeabilization/fixing and thus does not reflect *in vivo* situations.

Proximity-dependent biotin identification (BioID) is an alternative approach to affinity-purification MS (AP-MS) that exploits proximity-dependent *in vivo* protein biotinylation[50]. Bait proteins are fused to a promiscuous biotin ligase, BirA, which conjugates biotin to proximal proteins in living cells. Interaction partners of BirA-baits can be enriched by using streptavidin-affinity purification followed by MS-analysis (**Fig. 3e**).

Through combination of AP-MS and BioID approaches a high-confidence map of the Hippo-pathway was recently generated[51]. The study further showed that components of the Hippo pathway are modulated by phosphatase inhibition. BioID usually results in larger interactomes than AP-MS analysis and allows for detection of significantly lower-abundance prey proteins. Plus, it allows for identification of interactions between chromatin-associated or membrane-associated proteins, which is difficult to address by AP-MS.

One limitation for both PLA and BioID is that they are proximity assays that measure close distance rather than direct physical interactions.

Molecular Perturbations: Modulating signaling events

Functional Genomics

Functional Genomics typically consists of cellular perturbation and subsequent recording of a change in phenotype. Gene overexpression and knockdown or knockout are common gene-level perturbation strategies. This can be achieved by administration of large-

scale short interfering (si) or short hairpin (sh) RNA, complementary (c) DNA expression libraries, peptide approaches and, more recently, genome editing technologies such as CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas9 (CRISPR-associated protein 9)[52] (**Fig. 4**). Recording of a phenotype can be via biochemical assays, reporter genes, high-content screening or other OMICs technologies such as proteomic or genomic profiling as discussed elsewhere in this article.

Prior to the availability of arrayed cDNA expression libraries, so called "million clone" libraries from cDNA samples were prepared by a number of labs. Such libraries have been in use since the late 1980s, but one challenge has always been over- or under-representation of specific genes as well as missing estimates of coverage. Today, arrayed libraries are available from a variety of commercial sources, although access is often limited due to their high costs. It should be noted that these libraries do not take into account alternative ways of decoding[53], nor are they inclusive of all transcript variants and therefore, may represent a cell type-specific transcriptome rather than a genome-wide coverage. Nonetheless, expression cloning and arrayed cDNA expression have led to numerous breakthrough discoveries such as de-orphanizing GPCRs, the identification of tumour necrosis factor receptor, the regulatory p85 subunit of PI-3-kinase, the TYK2 kinase and the identification of multiple CD antigens, amongst others[54]. Similar to siRNA/shRNA experiments, one of the challenges is delivery into cells that are difficult to transfect. Another problem is that overexpression may lead to mislocalization in the cell. This can be circumvented by validation of cDNA screening hits with complementary knockout or knockdown experiments such as siRNA/shRNA or genome editing using zinc finger nucleases, TALENs (Transcription activator-like effector nucleases) or the CRISPR-Cas9 system. Furthermore, cDNA expression approaches will fail if multiple proteins must co-ordinate to achieve a function. Nonetheless, cDNA expression screening is particularly powerful for the identification of ligand-receptor interactions and can be used for cell-based protein interaction methods as discussed in the previous section.

The field of functional genomics has benefitted from two major breakthrough discoveries made at the start of this century. First, the completion of the human genome project allowed researchers to catalogue most coding (and non-coding) sequences in the

human genome. Second, the discovery of siRNAs and the generation of genome-wide knockdown libraries allowed for interrogating loss of function on a genomic scale. SiRNA can be provided as synthetic oligonucleotides or as shRNA embedded in plasmid vectors. Genome-wide siRNA-mediated knockdown has been successfully used in a large variety of processes and has contributed to seminal discoveries in signal transduction, including the identification of synthetic lethal interactions with the Ras oncogene[55], novel regulators of mitosis[56], kinases involved in clathrin-mediated endocytosis[57] or proteins involved in virus infection[58].

The biggest challenges in the use of siRNA/shRNA are a) delivery and b) off-target effects. Methods for delivery of siRNA/shRNA include lipofection-based transfection, electroporation and lentiviral transduction[59, 60]. Most cell types can be transfected reasonably well with one of these methods. Off-target effects are more difficult to tackle. For instance, it has been noted in a genome-wide screen for regulators of Parkin translocation that around 30% of identified hits show off-target recognition of the Parkin kinase PINK1[61]. In general, the phenomenon of off-target effects has been studied extensively and explained reasonably well[62-64], and improvements to existing genome-wide libraries include optimization of the siRNA design algorithms and chemistry of the siRNA oligonucleotides resulting in reduced off-target effects, as well as computational methods to deconvolute false positives[65]. Another problem is reproducibility of siRNA screening results. For example, in multiple HIV virus studies, minimal overlap of hit genes that regulate HIV infection was noted[66]. Poor reproducibility can also arise from differences in growth patterns of cells in tissue culture that can be overcome by computational methods[67]. For this reason, most researchers are eagerly looking for alternative methods of genome-wide loss of function screening, for example using the recently developed CRISPR-Cas9 technology[52].

Genome editing technologies are based on the targeted disruption of a gene locus, most commonly through the use of endonucleases that recognize and cut specific gene sequences. There are four main methods for genome editing, differing in the choice of the nuclease and target recognition mode: zinc finger nucleases, TALENs, homing meganucleases and CRISPR. To date, only the CRISPR-Cas9 technology has advanced to a

stage where genome-wide loss of function screening is possible, although in principle TALENs may be adaptable to generate large-scale libraries[68].

CRISPR requires three components: tracrRNA, sgRNA and nuclease Cas9. The sequence of the sgRNA guides the Cas9 endonuclease to a specific site within the genome, thus allowing site-specific gene modification[69]. In combination with different Cas9 genes, it is possible to use CRISPR for diverse applications such as gene knockout, gene knockdown or gene modification. Furthermore, CRISPR has recently been used for other approaches such as chemical target identification[70, 71]. The simplicity of the sgRNA sequence allowed a rapid design of genome-wide libraries using custom array synthesis and Gibson assembly. The multiple libraries generated using this approach include libraries for gene-knockouts, as well as transcriptional activation or repression of genes[72].

One obvious concern is that the sgRNA target sequence may result in off-target effects similar to those seen in siRNA/shRNA libraries[73-75]. The level of off-target effects in CRISPR-Cas9 is not yet fully resolved. Unlike siRNA, the precise contribution of each nucleotide to target cleavage is not yet determined and off-target evaluations are mostly based on empirical methods.

The genome-wide libraries to date are pooled lentiviral vectors for use in positive selection screens[74, 75]. Pooled libraries have over- and under-representation of certain genes that complicate genomic analysis. Hence, the generation of arrayed libraries is underway in both academic labs and from reagent providers that will be important when determining gene function in assays not amenable to positive selection.

Another concern is that knockout of genes using CRISPR is based on insertions/deletions (indels) generated by DNA repair mechanisms that result in missense frameshifting. Such indels will be different for each experiment and the exact nature of gene modification in diploid or polyploid cell lines is very difficult and challenging to evaluate. In this regard, the generation of precisely defined cell lines with gene modification may be helpful, though only feasible through a concerted effort of a large consortium. There is considerable development in the creation of mammalian haploid cell lines that will overcome some limitations of this technology due to mixed genetic alterations in di- or poly-ploid cell lines[76].

Genetic Perturbation: Chemical Genetics

Chemical genetics is a powerful method for disruption of gene or protein function in many cell types and even whole organisms such as zebrafish[77, 78]. Strictly speaking, chemical genetics is not a genomic or genetic method. It is based on the administration of small molecule compounds that disrupt a genetic component of the cell. In many cases, the gene or protein target of the compound is not known, making target identification one of the biggest challenges. Efforts are underway to develop targeted chemical compound sets and some commercial vendors offer libraries that target crucial signal transduction pathways such as kinases, proteases or autophagy regulators. The use of small molecule compounds has several important advantages[79]: first, the effect of the compound is much quicker than that of gene loss, as a protein can be immediately inactivated. Thus, the method is better suited to identify direct effects than indirect effects. Second, small molecule-based inhibition is typically reversible, allowing easy rescue experiments by washing out the compound. Third, handling is easy, and dose responses by titration of the compound can be informative when assessing kinetics of signal transduction pathways. While delivery and transfection are not generally concerns, some compounds are not able to cross the cell membrane. However, with improved library design, this will be solved. Chemical genetics has been successfully used to understand novel mechanisms of cancer cell resistance to PI3K inhibitors[80], the identification of AMPKα2 substrates[81] and the identification of neuroactive compounds that modulate behavior[82], to name just a few, and is frequently used in combination with functional genomic approaches[83].

Functional protein perturbation

Protein interaction domains (PIDs) play important roles in the communication between various proteins. One protein can have multiple PIDs, which poses a major hurdle to specifically disrupt the function of one individual PID without disrupting the whole protein. RNA interference or gene knockout disrupts the whole protein, and mutagenesis-strategies can often influence protein expression levels. Thus, protein-based inhibitors are a promising alternative to perform a targeted perturbation of PPIs. Multiple methods based on small synthetic proteins exist, including monobodies, nanobodies, affibodies, stapled peptides and DARPins, just to name a few[84]. Antibodies are the most commonly used protein-based inhibitors, but they have caveats as they are large and contain disulfide-bonds, which excludes them from being correctly folded in the reducing environment of the cytoplasm[85].

Monobodies offer an attractive alternative and allow for targeted PPI-disruption at the level of a single PID. Monobodies are single-domain binding proteins, based on the human fibronectin type III domain (FN3), which is a highly stable β-sandwich protein[85]. They do not have disulfides and are thus suitable for cellular studies.

Monobodies have been successfully used to disrupt intramolecular and intermolecular interactions of Bcr-Abl[86]. Constitutively active tyrosine-kinase Bcr-Abl causes chronic myelogenous leukemia and can be successfully treated with the tyrosine kinase inhibitor imatinib (Gleevec)[87]. Nonetheless, the appearance of secondary, imatinib-resistance-conferring mutations in Bcr-Abl poses a major clinical problem[88]. FN3-based monobodies have been shown to disrupt the SH2-kinase domain interface of Bcr-Abl, resulting in inhibition of Bcr-Abl activity through induction of apoptosis[86]. In another study, monobodies were developed to target the SH2 of the protein SHP2, the SH2-domain containing phosphatase 2 that is required for Bcr-Abl dependent oncogenic transformation. Inhibition of the SHP2-SH2 domain efficiently attenuated tyrosine phosphorylation and blocked ERK activation[89]. There are 120 human SH2 domains, all with a highly conserved phosphopeptide binding pocket[90]. Thus, developing specific inhibitors for SH2 domains of different proteins poses a challenge that can be overcome by monobody development[91].

Monobodies prove useful tools for the functional perturbation of signaling networks, especially in cancer cells. As they are very specific, this allows selective disruption of any PID without affecting the rest of the targeted protein. An alternative to monobodies are non-antibody binding proteins based on synthetic protein scaffolds and developed through directed evolution to bind to a specific protein domain, thereby potentially disrupting PPIs[92].

OMICs – where do we go from here?

OMICs techniques have enabled major breakthroughs in the field of cell signaling, in large part due to the development of specific assays capable of dissecting the distinct steps of cellular signaling events. The major advantage of OMICs-techniques is their ability to uncover novel aspects through unbiased, large-scale approaches, which provide researchers with data on a previously unprecedented genome/proteome-wide scale. Careful use of individual OMICs techniques in complex with other assays, including orthogonal validation

and bioinformatics analysis, allows the production of detailed high-confidence maps of cell signaling events.

One of the primary challenges that come with high-throughput research is effectively translating the vast amounts of accumulated data into meaningful biological contexts. Combining computational methods with OMICs-data is a powerful synergy, and can be used to model and predict signaling events or to integrate new results with other existing datasets to identify high-fidelity hits whose proper biological context (and potential clinical application) can be more clearly ascertained[93]. Biological validation experiments are necessary to functionally confirm the identified candidates and characterize their roles in cellular mechanisms.

Another bottleneck of large-scale analyses is data-standardization and comparability among various technologies. While many OMICs-techniques complement each other, meaning that different techniques uncover different aspects of cell signaling, this also poses a challenge for developing a gold standard for OMICs-analyses, as results from different approaches can often not directly be compared with each other. Standardized reference datasets and confidence scores, as used in PPI-interaction mapping, can help address this difficulty, facilitating comparison of datasets produced by different techniques and aiding in the identification of promising interactions[94]. This has recently been done for functional genomics, where off-target effects of shRNA and CRISPR screens pose a major limitation for genome-manipulation and where error rates have been difficult to assess, to evaluate data quality of genome-scale fitness screens[95].

Recent years have seen the development of many new technologies allowing the identification of novel interactors, phosphorylation sites and PTMs involved in signaling cascades, as well as new tools to perturb signaling events on the level of genome-modification or interference with proteins. In order to improve OMICs approaches in the signaling field, technologies will have to be developed to investigate transient and weak interactions. Kinases and phosphatases are key regulators of signaling events and can determine whether signaling is turned on or attenuated, however, kinase-substrate and phosphatase-substrate interactions are mostly missing from current PPI-datasets. One promising approach is the yeast-based M-track method, which allows for detection of short-lived interactions[96]. This study showed that three-quarters of the proteome in human cancer cells is phosphorylated, and corroborated the key involvement of tyrosine

phosphorylation in signaling cascade events by demonstrating the importance of phosphotyrosine in stimulated relative to unstimulated cells[97].

In the future, OMICs-technologies will be tasked with uncovering the role of other post-translational modifications in signaling events[1], such as ubiquitylation, sumoylation and glycosylation, requiring the development of new assays that can detect those modifications in a fast and cost-effective manner. Overall, OMICs-technologies represent powerful tools to uncover networks of signaling events and to perturb signaling cascades. In combination with orthogonal validation methods, in-depth bioinformatics analyses and careful functional studies, data generated by OMICs-techniques will help answer fundamental biological questions and define novel targets of therapeutic relevance to human health.

Acknowledgements

Figure Legends

Figure 1. Overview of signal transduction cascades.

Signal transduction follows a general principle, which includes five steps to convert an input signal, such as receptor-ligand-binding, into an output signal like the biochemical and genetic alteration of a cell.

Signal recognition typically takes place at the cell surface (step 1). The signal has to be transferred into the cell without the molecule itself entering. A receptor often performs the function of information transfer across the membrane by binding to a ligand. Once receptor-ligand binding takes place, this leads to conformational changes of the receptor, which leads to activation of the receptor, when the receptor itself is an enzyme, or modification of the activities of recruited enzymes. As chemical imprints and major signal carriers, PTMs such as phosphorylation are carried out by receptors or a series of downstream enzymes (step 2). Signal processing is also facilitated by binary PPI, protein complex formation, or signaling protein translocation to sites such as the nucleus. Multiple input signals can be synchronized and processed to generate a signal response (steps 3 and 4). Final biological outcomes are achieved by modification of different intracellular machineries such as transcription and translation (step 5).

Each of the individual five steps of signal transduction can be dissected into various branches, depending of what components are known and which information is needed. Profiling methods aim to gain information about the network surrounding a certain signaling component, whereas perturbation methods aim to disrupt signaling networks on proteomic or genomic levels in order to identify their functional role.

Figure 2. Molecular profiling by MS-based proteomics methods.

a) SRM (MRM) and SWATH. In SRM, digested peptides are separated by liquid chromatography. The targeted peptides are selected by the first quadrupole. They are fragmented in the second quadrupole and the consequent fragments are selected by the third. The peptides are quantified by integrating the intensities of corresponding ions to the retention peak. In contrast, in the SWATH mode, all possible peptides in a range are selected and recorded. Targeted proteins/peptides are selected and analyzed during the stage of data analysis.

b) Mass cytometry. The cells are labelled with antibodies tagged with different transition element isotopes. Single cells are separated by a nebulizer and each tagged isotope is analyzed by time-of-flight MS. The abundance of each element in each cell is recorded and the whole set of data is then processed by bioinformatics-analysis.

c) Chemoproteomics. This method is based on chemical probes, either activity-based probe (ABP) (in the ABPP approach) or small compounds (CCCP approach). The probes can be conjugated to a fluorescent group, which can be resolved in a gel. More commonly, the target proteins are enriched by conjugating the probe to solid phase and are then subjected to MS analysis. Pre-incubation of various compounds with lysates followed by quantitative MS allows characterizing the dynamic feature of protein-compound interaction such as the IC50.

Figure 3. Molecular profiling using PPI-approaches.

a) PCA. A reporter enzyme is split into two halves (F1 and F2) and fused to bait and prey proteins. Upon interaction the reporter enzyme is reconstituted, resulting in enzymatic conversion of a substrate into a detectable product. In split-TEV, interaction reconstitutes a tobacco etch virus (TEV) protease, which cleaves at TEV recognition sites, releasing a transcription factor (TF), which subsequently activates reporter gene expression. In bimolecular fluorescence complementation (BiFC) assays, a bait-prey interaction complements a fluorescent protein.

b) MAPPIT. A bait-protein is coupled to a signal-deficient cytokine receptor, which lacks STAT-recruitment sites but can recruit JAKs. The prey is tethered to another receptor-moiety harboring STAT-recruitment sites. Bait–prey interaction reconstitute a functional receptor. Ligand-binding leads to cross phosphorylation of JAKs, which phosphorylate the prey receptor-fragment, rendering the receptor accessible to STAT-docking. Recruited STATs are phosphorylated by JAKs, dimerize and enter the nucleus to activate reporter gene expression.

KISS. The bait is fused to a kinase-containing portion of TYK2, the prey is fused to a gp130 cytokine receptor fragment. Upon bait and prey interaction, TYK2 phosphorylates STAT3 docking sites on the prey chimera, resulting in reporter gene activation.

c) MaMTH. An integral membrane-bait is tagged with C-terminal half of ubiquitin (Cub) and a TF, and the prey is fused to N-terminal half of ubiquitin (Nub). Upon bait and prey

interaction, ubiquitin reconstitution occurs, resulting in cleavage by deubiquitinating enzymes (DUBs) and release of the TF, which activates the luciferase reporter.

d) PLA. Two proximity probes are targeted to two proteins and the oligonucleotides on the proximity probes are brought close together. These antibody-conjugated oligonucleotides can hybridize to two connector oligonucleotides that are ligated and form a circular DNA molecule, which is amplified by rolling circle amplification (RCA). The resulting single-stranded DNA-molecule is detectable through hybridization of fluorescently labeled complementary oligonucleotides.

e) BioID. The bait is coupled to a promiscuous biotin protein ligase harboring a mutation (BirA*). BirA* can catalyze the formation of activated biotin and dissociates quickly from the intermediate. The BirA*-tagged bait generates an activated biotin-cloud *in vivo*, which reacts with free primary amines of lysine residues. Interaction partners of BirA*-baits can be enriched by streptavidin-purification coupled to MS analysis.

Figure 4. Molecular perturbation methods.

a) Genomic perturbations. CRISPR-technologies allow modifications on gene/genome-levels. Genes can be endogenously tagged or knocked out or mutated and modified. Chemical genetics also allow modifications/disruptions of genes through application of small-molecule compounds. Whereas CRISPR allows for targeted gene-modification, gene-targets of chemical compounds are often random and cannot be selected.

siRNA/shRNA technologies target mRNAs of specific genes, consequently leading to reduction or disruption of protein expression. Targeted protein overexpression can be achieved through expression of arrayed cDNA libraries.

b) Proteomic perturbations. Direct targeting of candidate proteins can be achieved by applying chemical genetics-screens and monobodies. Chemical compounds can directly bind to proteins and lead to disruption of interactions with other proteins or loss of protein activity. Whereas monobodies also lead to PPI-disruption and loss of protein function, they are very specific and can be targeted towards specific protein domains.

Box 1. Cell signaling in a nutshell

Intracellular signal transduction is carried out by multiple parallel signaling pathways. It is usually initiated by receptors to extracellular or internal signals. Receptor tyrosine kinases (RTK) are receptors for many growth factors, peptide hormones and cytokines[98]. Upon engagement with their ligands, RTKs undergo dimerization, leading to subsequent activation of the intracellular tyrosine kinase domains. The activated kinases phosphorylate tyrosine residues, which then recruit proteins through their SH2 or PTB domains. The recruited proteins can be enzymes or scaffold proteins, which relay the signals to downstream pathways such as Ras/ERK and PI3K/Akt pathways by mechanisms including phosphorylation, PPIs/dissociation, and molecule translocation. The signaling events occur proximal to RTKs, or in distal cellular compartments such as the nucleus. In the nucleus, subsequent changes in transcription and epigenetic modification lead to alterations in cellular outcomes such as proliferation, growth, survival, differentiation and more.

In contrast to RTKs, GPCRs do not have any enzymatic activity[99]. Binding to their agonists induces a conformational change that allows GTP-loading to heterotrimeric G proteins. Subsequently, the α-subunit binds and activates various effectors such as enzymes and ion channels. The βγ-subunits dissociate from α and act on other effectors. The outcomes of GPCR signaling are very diverse. GPCRs can be phosphorylated by GRK upon stimulation. β-arrestin is recruited to phosphorylated receptors and regulates GPCR internalization and degradation/recycling. Interestingly, β-arrestin also activates MAPK pathways.

Another unique class of signaling pathways is involved in innate immunity and inflammation[100], and is initiated by receptors such as Toll-like-receptors (TLRs), Rig-I-like receptors (RLRs) and Nod-like receptors (NLRs), which respond to extracellular or intruding pathogen molecules, or the tumor-necrosis-factor-receptor (TNFR1), which responds to cytokines such as TNF. Binding to these molecules triggers the formation of intracellular complexes, including adaptor proteins and enzymes around the receptors, and finally leads to activation of NFκB pathway and MAPK pathways. Subsequent transcriptional activation produces a variety of inflammation mediators. These pathways are also crucial regulators of cell death. TNFR1 induces apoptosis through a series of caspase-mediated proteolysis-steps, and most of these receptors control necroptosis through RIPK1, RIPK3 and the downstream regulator MLKL. The full picture of necroptosis is still largely unknown. The exact

mechanisms of switching between inflammatory reaction, apoptosis and necroptosis also remain fragmental. Beside the above three examples, many other signaling pathways such as Wnt/βcatenin, TGFβ, Notch, Hedgehog, Hippo/MST pathways, also play diverse roles in various cellular processes.

References

1. Olsen, J.V. & Mann, M. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol. Cell Proteomics* **12**, 3444-3452 (2013).
2. Serber, Z. & Ferrell, J.E., Jr. Tuning bulk electrostatics to regulate protein function. *Cell* **128**, 441-444 (2007).
3. Choudhary, C., Weinert, B.T., Nishida, Y., Verdin, E. & Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat. Rev. Mol. Cell Biol.* **15**, 536-550 (2014).
4. Geiss-Friedlander, R. & Melchior, F. Concepts in sumoylation: a decade on. *Nat. Rev. Mol. Cell Biol.* **8**, 947-956 (2007).
5. Hunter, T. The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol. Cell* **28**, 730-738 (2007).
6. Pawson, T. & Nash, P. Protein-protein interactions define specificity in signal transduction. *Genes De.v* **14**, 1027-1047 (2000).
7. Petschnigg, J., Snider, J. & Stagljar, I. Interactive proteomics research technologies: recent applications and advances. *Curr. Opin. Biotechnol.* **22**, 50-58 (2011).
8. Teruel, M.N. & Meyer, T. Translocation and reversible localization of signaling proteins: a dynamic future for signal transduction. *Cell* **103**, 181-184 (2000).
9. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450-461 (2007).
10. Berger, B., Peng, J. & Singh, M. Computational solutions for omics data. *Nat. Rev. Genet.* **14**, 333-346 (2013).
11. Joyce, A.R. & Palsson, B.O. The model organism as a system: integrating 'omics' data sets. *Nat. Rev. Mol. Cell Biol.* **7**, 198-210 (2006).
12. Kolch, W. & Pitt, A. Functional proteomics to dissect tyrosine kinase signalling pathways in cancer. *Nat. Rev. Cancer* **10**, 618-629 (2010).
13. Kristensen, V.N. et al. Principles and methods of integrative genomic analyses in cancer. *Nat. Rev. Cancer* **14**, 299-313 (2014).
14. Rigaut, G. et al. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030-1032 (1999).
15. Ong, S.E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252-262 (2005).
16. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, 222 (2008).
17. Bisson, N. et al. Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nat. Biotechnol.* **29**, 653-658 (2011).
18. Zheng, Y. et al. Temporal regulation of EGF signalling networks by the scaffold protein Shc1. *Nature* **499**, 166-171 (2013).
* Together with the paper by Bisson *et al.*, this study analyzes the signaling complexes proximal to EGFR during EGF signaling with SRM. It reveals the dynamics of different complexes, which play distinct roles in EGF signaling.
19. Gillet, L.C. et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111 016717 (2012).

20. Collins, B.C. et al. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat. Methods* **10**, 1246-1253 (2013).

* This is the first successful example of using SWATH to address a biological question. It displays the dynamics of 14-3-3 protein complexes in response to IGF1 stimulation.

21. Lambert, J.P. et al. Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nat. Methods* **10**, 1239-1245 (2013).

22. Bandura, D.R. et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.* **81**, 6813-6822 (2009).

23. Bendall, S.C. et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687-696 (2011).

* An interesting paper to use mass cytometry to simultaneously characterize a single cell with 34 parameters.

24. Bodenmiller, B. et al. Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat. Biotechnol.* **30**, 858-867 (2012).

25. Nomura, D.K., Dix, M.M. & Cravatt, B.F. Activity-based protein profiling for biochemical pathway discovery in cancer. *Nat. Rev. Cancer* **10**, 630-638 (2010).

26. Rix, U. & Superti-Furga, G. Target profiling of small molecules by chemical proteomics. *Nat. Chem. Biol.* **5**, 616-624 (2009).

27. Bantscheff, M. et al. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat. Biotechnol.* **29**, 255-265 (2011).

28. Huber, K.V. et al. Stereospecific targeting of MTH1 by (S)-crizotinib as an anticancer strategy. *Nature* **508**, 222-227 (2014).

* An excellent example of using chemoproteomics for target identification and drug discovery. MTH1 was identified as the target for antitumor small molecule SCH51344. Furthermore, (*S*)-enantiomer of crizotinib has been identified as a selective inhibitor of MTH1.

29. Daub, H. et al. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol. Cell* **31**, 438-448 (2008).

30. Duncan, J.S. et al. Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. *Cell* **149**, 307-321 (2012).

* This study uses the CCCP chemoproeomics approach to characterize cancer signaling. It describes how the signaling network responds to the treatment with a MEK inhibitor.

31. Savas, J.N., Stein, B.D., Wu, C.C. & Yates, J.R., 3rd Mass spectrometry accelerates membrane protein analysis. *Trends Biochem. Sci.* **36**, 388-396 (2011).

32. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246 (1989).

33. Stefan, E. et al. Quantification of dynamic protein complexes using Renilla luciferase fragment complementation applied to protein kinase A activities in vivo. *Proc Natl Acad Sci U S A* **104**, 16916-16921 (2007).

34. Wehr, M.C. et al. Monitoring regulated protein-protein interactions using split TEV. *Nat. Methods* **3**, 985-993 (2006).

35. Wehrman, T.S. et al. A system for quantifying dynamic protein interactions defines a role for Herceptin in modulating ErbB2 interactions. *Proc Natl Acad Sci U S A* **103**, 19063-19068 (2006).

36. Michnick, S.W., Remy, I., Campbell-Valois, F.X., Vallee-Belisle, A. & Pelletier, J.N. Detection of protein-protein interactions by protein fragment complementation strategies. *Methods Enzymol.* **328**, 208-230 (2000).

37. Kerppola, T.K. Visualization of molecular interactions using bimolecular fluorescence complementation analysis: characteristics of protein fragment complementation. *Chem. Soc. Rev.* **38**, 2876-2886 (2009).

38. Tchekanda, E., Sivanesan, D. & Michnick, S.W. An infrared reporter to detect spatiotemporal dynamics of protein-protein interactions. *Nat. Methods* **11**, 641-644 (2014).

* This paper describes the development of a split-PCA variant based on a fluorescent infrared protein IFP. This novel PCA is reversible and allows detection of signaling complexes at low expression levels.

39. Lemmens, I. et al. Heteromeric MAPPIT: a novel strategy to study modification-dependent protein-protein interactions in mammalian cells. *Nucleic Acids Res.* **31**, e75 (2003).

40. Lemmens, I., Lievens, S., Eyckerman, S. & Tavernier, J. Reverse MAPPIT detects disruptors of protein-protein interactions in human cells. *Na. Protoc.* **1**, 92-97 (2006).

41. Lievens, S. et al. KISS, a mammalian in situ protein interaction sensor. *Mol. Cell. Proteomics* (2014).

42. Petschnigg, J. et al. The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells. *Nat. Methods* **11**, 585-592 (2014).

* Novel split-ubiquitin-based PCA-variant that allows for detection of PPIs of full-lenth-integral membrane proteins under context-dependent conditions, such as interactions that are conferred by phosphorylation, inhibited by drugs or stimulated by hormones/ligands.

43. Deribe, Y.L. et al. Regulation of epidermal growth factor receptor trafficking by lysine deacetylase HDAC6. *Sci. Signal.* **2**, ra84 (2009).

44. Stagljar, I., Korostensky, C., Johnsson, N. & te Heesen, S. A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. *Proc Natl Acad Sci U S A* **95**, 5187-5192 (1998).

45. Ciruela, F. Fluorescence-based methods in the study of protein-protein interactions in living cells. *Curr. Opin. Biotechnol.* **19**, 338-343 (2008).

46. Soderberg, O. et al. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* **3**, 995-1000 (2006).

47. Gu, G.J. et al. Protein tag-mediated conjugation of oligonucleotides to recombinant affinity binders for proximity ligation. *N. Biotechnol.* **30**, 144-152 (2013).

48. Leuchowius, K.J. et al. Parallel visualization of multiple protein complexes in individual cells in tumor tissue. *Mol. Cell. Proteomics* **12**, 1563-1571 (2013).

49. Smith, M.A. et al. Annotation of human cancers with EGFR signaling-associated protein complexes using proximity ligation assays. *Sci. Signal.* **8**, ra4 (2015).

50. Roux, K.J., Kim, D.I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801-810 (2012).

51.   Couzens, A.L. et al. Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. *Sci. Signal.* **6**, rs15 (2013).

52.   Heintze, J., Luft, C. & Ketteler, R. A CRISPR CASe for high-throughput silencing. *Front. Genet.* **4**, 193 (2013).

53.   Ketteler, R. On programmed ribosomal frameshifting: the alternative proteomes. *Front. Genet.* **3**, 242 (2012).

54.   Seed, B. Developments in expression cloning. *Curr. Opin. Biotechnol.* **6**, 567-573 (1995).

55.   Luo, J. et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835-848 (2009).

56.   Neumann, B. et al. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**, 721-727 (2010).

57.   Pelkmans, L. et al. Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature* **436**, 78-86 (2005).

58.   Mercer, J. et al. RNAi screening reveals proteasome- and Cullin3-dependent stages in vaccinia virus infection. *Cell Rep.* **2**, 1036-1047 (2012).

59.   Marine, S. et al. High-throughput transfection of differentiated primary neurons from rat forebrain. *J Biomol. Screen.* **17**, 692-696 (2012).

60.   Moffat, J. et al. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283-1298 (2006).
* This paper describes the first generation of pooled and arrayed lentiviral shRNA libraries and enabled genome-wide loss-of-function screens in a variety of cell types.

61.   Hasson, S.A. et al. High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. *Nature* **504**, 291-295 (2013).

62.   Birmingham, A. et al. 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods* **3**, 199-204 (2006).

63.   Franceschini, A. et al. Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. *Proc Natl Acad Sci U S A* **111**, 4548-4553 (2014).

64.   Jackson, A.L. et al. Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* **12**, 1179-1187 (2006).

65.   Mohr, S.E., Smith, J.A., Shamu, C.E., Neumuller, R.A. & Perrimon, N. RNAi screening comes of age: improved techniques and complementary approaches. *Nat. Rev. Mol. Cell Biol.* **15**, 591-600 (2014).

66.   Bushman, F.D. et al. Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog.* **5**, e1000437 (2009).

67.   Snijder, B. et al. Single-cell analysis of population context advances RNAi screening at multiple levels. *Mol. Syst. Biol.* **8**, 579 (2012).

68.   Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* **30**, 460-465 (2012).

69.   Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
* First paper to describe the possibility of RNA-programmed genome editing using the CRISPR technology.

70.   Kasap, C., Elemento, O. & Kapoor, T.M. DrugTargetSeqR: a genomics- and CRISPR-Cas9-based method to analyze drug targets. *Nat. Chem. Biol.* **10**, 626-628 (2014).

71.   Smurnyy, Y. et al. DNA sequencing and CRISPR-Cas9 gene editing for target validation in mammalian cells. *Nat. Chem. Biol.* **10**, 623-625 (2014).

72.   Gilbert, L.A. et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661 (2014).

73.   Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267-273 (2014).

74.   Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87 (2014).

75.   Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80-84 (2014).

76.   Wutz, A. Haploid animal cells. *Development* **141**, 1423-1426 (2014).

77.   Kaufman, C.K., White, R.M. & Zon, L. Chemical genetic screening in the zebrafish embryo. *Nat. Protoc.* **4**, 1422-1432 (2009).

78.   Specht, K.M. & Shokat, K.M. The emerging power of chemical genetics. *Curr. Opin. Cell Biol.* **14**, 155-159 (2002).

79.   Alaimo, P.J., Shogren-Knaak, M.A. & Shokat, K.M. Chemical genetic approaches for the elucidation of signaling pathways. *Curr. Opin. Chem. Biol.* **5**, 360-367 (2001).

80.   Muellner, M.K. et al. A chemical-genetic screen reveals a mechanism of resistance to PI3K inhibitors in cancer. *Nat. Chem. Biol.* **7**, 787-793 (2011).

81.   Banko, M.R. et al. Chemical genetic screen for AMPKalpha2 substrates uncovers a network of proteins involved in mitosis. *Mol. Cell* **44**, 878-892 (2011).

82.   Kokel, D. et al. Rapid behavior-based identification of neuroactive small molecules in the zebrafish. *Nat. Chem. Biol.* **6**, 231-237 (2010).

83.   Gonsalves, F.C. et al. An RNAi-based chemical genetic screen identifies three small-molecule inhibitors of the Wnt/wingless signaling pathway. *Proc Natl Acad Sci U S A* **108**, 5954-5963 (2011).

84.   Weidle, U.H., Auer, J., Brinkmann, U., Georges, G. & Tiefenthaler, G. The emerging role of new protein scaffold-based agents for treatment of cancer. *Cancer Genomics Proteomics* **10**, 155-168 (2013).

85.   Koide, A., Wojcik, J., Gilbreth, R.N., Hoey, R.J. & Koide, S. Teaching an old scaffold new tricks: monobodies constructed using alternative surfaces of the FN3 scaffold. *J. Mol. Biol.* **415**, 393-405 (2012).

86.   Grebien, F. et al. Targeting the SH2-kinase interface in Bcr-Abl inhibits leukemogenesis. *Cell* **147**, 306-319 (2011).

87.   Deininger, M., Buchdunger, E. & Druker, B.J. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood* **105**, 2640-2653 (2005).

88.   Jabbour, E., Hochhaus, A., Cortes, J., La Rosee, P. & Kantarjian, H.M. Choosing the best treatment strategy for chronic myeloid leukemia patients resistant to imatinib: weighing the efficacy and safety of individual drugs with BCR-ABL mutations and patient history. *Leukemia* **24**, 6-12 (2010).

89.   Sha, F. et al. Dissection of the BCR-ABL signaling network using highly specific monobody inhibitors to the SHP2 SH2 domains. *Proc Natl Acad Sci U S A* **110**, 14924-14929 (2013).

90.   Campbell, S.J. & Jackson, R.M. Diversity in the SH2 domain family phosphotyrosyl peptide binding site. *Protein Eng.* **16**, 217-227 (2003).

91.   Wojcik, J. et al. A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nat. Struct. Mol. Biol.* **17**, 519-527 (2010).

92.   Tiede, C. et al. Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications. *Protein Eng. Des. Sel.* **27**, 145-155 (2014).

93.   Taylor, I.W. et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat. Biotechnol.* **27**, 199-204 (2009).

94.   Braun, P. et al. An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6**, 91-97 (2009).

95.   Hart, T., Brown, K.R., Sircoulomb, F., Rottapel, R. & Moffat, J. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol. Syst. Biol.* **10**, 733 (2014).

96.   Zuzuarregui, A. et al. M-Track: detecting short-lived protein-protein interactions in vivo. *Nat. Methods* **9**, 594-596 (2012).

97.   Sharma, K. et al. Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep.* **8**, 1583-1594 (2014).

98.   Lemmon, M.A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117-1134 (2010).

99.   Lefkowitz, R.J. Historical review: a brief history and personal retrospective of seven-transmembrane receptors. *Trends Pharmacol. Sci.* **25**, 413-422 (2004).

100.   Newton, K. & Dixit, V.M. Signaling in innate immunity and inflammation. *Cold Spring Harb. Perspect. Biol.* **4** (2012).

Table 1. Molecular profiling approaches and their key properties and limitations.
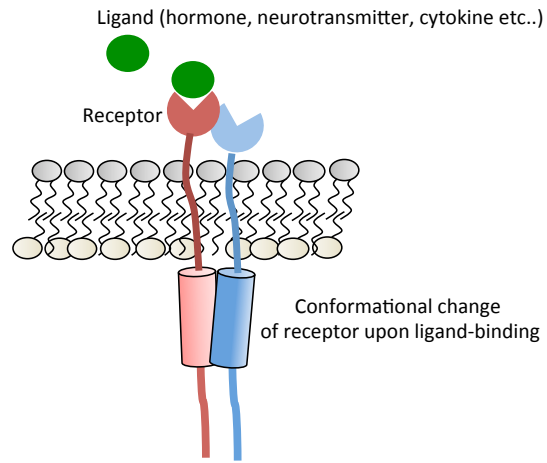
Method	Key properties	Limitations
SILAC	Simple labelling and fast quantitative mass spectrometry.	Limited to only living cells.
ICAT and iTRAQ	Different biological samples can be labelled.	Extra labelling step is needed.
SRM/MRM	High sensitivity, accuracy and reproducibility.	Only limited to targeted studies. Assay development is needed.
SWATH	Suitable for discovery study with wide proteome coverage. With similar sensitivity, reproducibility and dynamic range as SRM/MRM.	Highly dependent on computation.
Mass cytometry	Single cell measurement. High multiplexity, low background, and no spectral overlap compared to fluorescent cytometry.	Antibody labelling is required. Cannot sort the cells.
Chemoproteomics	Unbiased and proteome-wide study of protein-small molecule interaction. Potential for target identification.	Chemistry of immobilizing probes may affect the interaction. High background, and some interactions may be underrepresented.

PPI-profiling methods

Method	Key properties	Limitations
PCA	Live-cell assay. Fluorescent PCAs allow for visualization of site of interaction. Wide range of split-PCAs available (fluorophores, luciferase, TEV-protease, β-lactamase, etc.).	Most PCAs are irreversible, except infrared split-IFP. Overexpression artefacts possible.
MAPPIT/KISS	Signal amplification through additional reporter readout (weak and transient interactions can be detected). Live-cell assay, amenable to large-scale applications. KISS is compatible with full-length transmembrane proteins.	Overexpression artefacts. Indirect reporter readout excludes spatial and temporal PPI-analysis. The MAPPIT-interaction sensor is localized to the plasma membrane, thus excluding PPI analysis at their native localization. MAPPIT is incompatible with full-length transmembrane proteins.
MaMTH	Designed for assaying full-length integral membrane protein interactions. Live-cell assay, can detect phosphorylation-dependent interactions and can map phospho-sites. Signal amplification through additional reporter readout.	Overexpression artefacts possible (though lentiviral expression plasmid are available). Luciferase-readout only allows for arrayed interaction screen, no pooled screening yet possible. As for MAPPIT/KISS, indirect reporter readout excludes spatial and temporal PPI-analysis.
PLA	*In situ* assay. Allows for detection of PPIs in fixed cells and tissues. Can localize site of interaction.	Permeabilization or fixing is required. Proximity assay, which can detect adjacent proteins that might not be true interactors.
BioID	Can detect low abundant proteins (in contrast to conventional AP-MS).	Proximity assay can also detect adjacent proteins that might not be true interactors.

Table 2. Molecular perturbation approaches and their key properties and limitations.

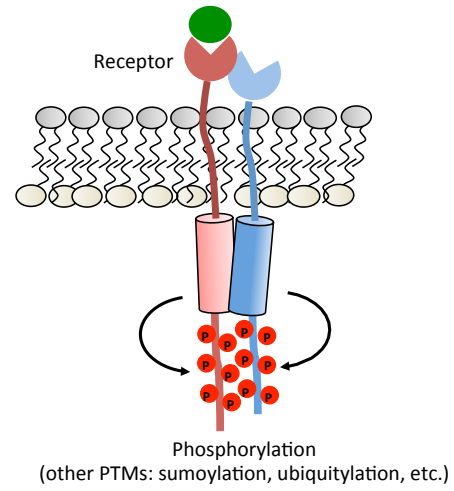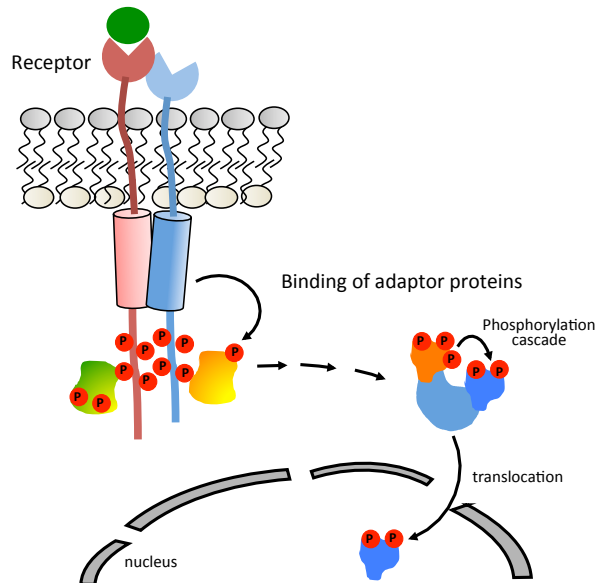| Method	Key properties	Limitations
cDNA libraries	Fast and high-throughput-amenable screening of overexpressed proteins. Full-length arrayed libraries are available from commercial vendors.	Overexpression artefacts. Classical cDNA libraries rarely contain full-length cDNAs, whereas commercially available, arrayed libraries often miss transcript variants.
siRNA/shRNA	Complements cDNA library screen. Fast and high-throughput-amenable screening of genome-wide knockdowns. Suitable for arrayed and pooled screens.	Off-target effects. Variable knockdown efficiencies. May fail in case of functionally redundant targets.
CRISPR	Knockout, knock-in and endogenous tagging possible. Activating or silencing possible.	Off-target effects remain to be evaluated. Limited (pooled) libraries available so far. May fail in case of functionally redundant targets. Knockout of essential genes can obscure functional readouts.
Chemical genetics	Small molecule compounds targeting a certain gene/protein in a reversible manner. Faster than si/shRNA knockdown or CRISPR knockout, as gene product can be directly targeted.	Some compounds cannot be delivered inside the cell. Problem of specificity. Often, cellular targets of the compounds are not known or compounds have off-target effects.
Functional protein perturbation using monobodies	Monobodies offer targeted disruption of domains mediating PPIs. No disulphide bonds allow use of monobodies in cellular studies in contrast to antibodies.	Relatively low-throughput. Time- and cost-consuming to develop reagents.

a

1. Signal recognition

Ligand (hormone, neurotransmitter, cytokine etc..)

Receptor

Conformational change
of receptor upon ligand-binding

b

2. Signal conversion

Receptor

Phosphorylation
(other PTMs: sumoylation, ubiquitylation, etc.)

c

3 & 4. Signal relay and processing

Receptor

Binding of adaptor proteins

Phosphorylation
cascade

translocation

nucleus

d

5. Cellular response

mRNA

AAAAAAAA

Translation

nucleus

Transcription of target genes

Figure 1

a

SRM

Q1
precursor selection

Q2
fragmentation

Q3
fragment selection

SWATH

Q1
precursor scanning

Q2
fragmentation

TOF
full scanning

targeted extraction

b

Nebulizer

ICP/MS elemental analysis

Cell 1
Cell 2
Cell 3
Cell 4

Mass

Data analysis

c

ABPP approach

CCCP approach

lysates

ABP labeling

fluorescent ABP labeling

capture of protein-compound interaction

compound preincubation

enrichment

competitive interaction

MS analysis

in-gel analysis

MS analysis

IC50 estimation

Figure 3

a

siRNA/shRNA ⊣ *Protein* ⊗

cDNA → *Protein*

mRNA

mRNA

mRNA

CRISPR
Chemical genetics ⊣

DNA

Gene-level perturbation

b

Chemical genetics
Monobodies ⊣ Disruption of protein activity

⊣ Protein interaction disruption
(inter-/intramolecular)

Proteins Protein-level perturbation

Figure 4