Published in final edited form as:

J Mol Biol. 2015 October 23; 427(21): 3416–3440. doi:10.1016/j.jmb.2015.07.021.

Studying Cellular Signal Transduction with OMIC Technologies

Benjamin D. Landry¹, David C. Clarke², and Michael J. Lee^{1,3,*}

¹ Program in Systems Biology (PSB), University of Massachusetts Medical School, Worcester MA, USA

² Department of Biomedical Physiology and Kinesiology (BPK), Simon Fraser University, Burnaby BC, Canada

³ Program in Molecular Medicine (PMM); Department of Molecular, Cell, and Cancer Biology (MCCB); University of Massachusetts Medical School, Worcester MA, USA

Abstract

In the gulf between genotype and phenotype exists proteins and, in particular, protein signal transduction systems. These systems use a relatively limited parts list to respond to a much longer list of extracellular, environmental, and/or mechanical cues with rapidity and specificity. Most signaling networks function in a highly nonlinear and often contextual manner. Furthermore, these processes occur dynamically across space and time. Because of these complexities, systems and “OMIC” approaches are essential for the study of signal transduction. One challenge in using OMIC-scale approaches to study signaling is that the “signal” can take different forms in different situations. Signals are encoded in diverse ways such as protein-protein interactions, enzyme activities, localizations, or post-translational modifications to proteins. Furthermore, in some cases signals may be encoded only in the dynamics, duration, or rates of change of these features. Accordingly, systems-level analyses of signaling may need to integrate multiple experimental and/or computational approaches. As the field has progressed, the non-triviality of integrating experimental and computational analyses has become apparent. Successful use of OMIC methods to study signaling will require the “right” experiments and the “right” modeling approaches, and it is critical to consider both in the design phase of the project. In this review, we discuss common OMIC and modeling approaches for studying signaling, emphasizing the philosophical and practical considerations for effectively merging these two types of approaches to maximize the probability of obtaining reliable and novel insights into signaling biology.

I. INTRODUCTION

A cell's ability to perceive and understand its surroundings, respond to changes in the environment, and/or adapt to defend homeostasis is governed by complex signal transduction systems. In plain terms, the basic purpose of signaling is to relay information about the outside of the cell to the nucleus. This signaling often results in alterations to the transcriptional state of the cell as to modify cellular behavior [1]. The state of the signaling network creates the background “context” that is often missing in genomic studies of

* To whom correspondence should be michael.lee@umassmed.edu.

behavior/drug response/disease outcome, which has substantially limited our ability to leverage genomic or transcriptomic insight for therapeutic benefit. Decades of research have identified signaling pathways that control essentially every cellular function (growth, migration, death, differentiation, division, secretion, etc.). Owing to this fact, and because these systems typically include many “drug-able” components, signaling has emerged as a common target in modern molecular medicine. However, after a decade or so of optimism, it has become clear that the pharmacological targeting of signaling often results in unanticipated consequences, which manifest clinically as toxicity, acquired resistance, and/or limited efficacy. Progress in the area of targeted (or “*personalized*” or “*precision*”) medicine will likely require a more refined understanding of how signaling pathways function, how they are altered in disease states, and how they are augmented by drug therapies.

The complexity of signal transduction networks demands “systems-level” approaches, but answering this demand remains a major challenge. Signaling networks are densely interconnected, which facilitates signaling “crosstalk” (i.e., pathway activities influence each other), and most forms of crosstalk are incompletely understood. Furthermore, the “architecture” of a signaling process (i.e., the connectivity or wiring of the signaling circuit) is regulated dynamically and can change through time. This rewiring or reprogramming of signaling networks is a concept that is only now being explored in a systematic way [2,3]. Finally, the information content (i.e., the signal) takes different forms in different situations. Thus, no single OMIC method is appropriate for the comprehensive study of signaling. In canonical examples, the signal is encoded and transmitted by the absolute levels of a protein or post-translationally modified form of a protein. Additionally, signals are also encoded by the location of a protein, the interactions between proteins, the relative stoichiometry of positive and negative regulators, and even the rate, duration or patterns of change between these states (Figure 1). And if the future continues to reaffirm recent precedent, these “complex” forms of signaling will prove to be commonplace.

Because signaling has long been acknowledged as being complex, many systems biology approaches have been applied to their study. In the early days, the typical approach was to generate a large amount of data using the latest OMIC technology (e.g., microarray), apply whatever computational approach was currently the “hottest”, followed by believing that new insights would inevitably emerge from the modeling. This philosophically problematic approach was exacerbated by the practical difficulties of these two steps often being performed by different people in different labs, whose communication was likely impaired by differences in field-specific jargon and culture as well as more fundamental differences in intellectual goals. Since these early efforts – which we call *version 1.0* of systems biology of signal transduction – the field has matured considerably. Systems approaches featuring “OMIC” data and computational modeling are and will continue to be instrumental in the study of signaling. However, it is clear that their successful implementation requires careful planning and consideration of the respective strengths and weaknesses of the experimental and computational strategies both in isolation but also in relation to each other. In this review, we will cover different ways that a systems biologist can collect and model signaling data. We will cover some recent examples from the literature and discuss in detail some

common pitfalls and practical considerations for successfully integrating OMIC data with computational approaches in signaling research.

II. METHODS FOR HIGH-THROUGHPUT COLLECTION OF QUANTITATIVE SIGNALING DATA

The creation of robust and precise models relies on the generation of accurate data sets that measure how signals in the cell are conveyed. High-throughput methodologies have facilitated the generation of reliable large-scale datasets for a wider variety of signals. Because signals are encoded in different ways, many different technologies are necessary to measure the propagation of a signal throughout the cell. Methods include those that are based on “molecular profiling” of protein signaling states, and those based on “molecular perturbation” and inference [4]. In this section we will review various assays used to collect signaling data, discuss the advantages and limitations of each method, and also discuss practical considerations in the design of OMIC experiments for interrogating signaling. An overview of these approaches is also available in Table 1. This section will be formatted around four basic ways in which signals are encoded: the levels, interactions, localizations, and activities of various signaling proteins [5].

Measurement of Protein Expression and Post-translational Modification Levels

In the simplest situations, the signal is encoded by the mere existence or non-existence of a post-translationally modified form of a protein, or more rarely, the total expression level of a protein. Several methods exist for quantifying protein levels, but only one, mass spectrometry, can be argued to provide comprehensive OMIC-scale depth.

Mass Spectrometry—To many, the gold standard in proteomics is mass spectrometry. Unlike any other proteomic technologies, mass spectrometry can be used to measure complete proteomes and thousands of post-translationally modified proteins, rivaling the depth and richness of other genomic technologies. Furthermore, these measurements are precise and quantitative, and obviate the need to rely on the existence of specific antibodies (although antibodies are often used in mass spectrometry experiments to enrich for protein fractions of interest). By using cellular fractionation techniques, or enrichment of proteins by chromatographic or other methods, is it possible to comprehensively and quantitatively study the global protein landscape, specific subsets of proteins, or proteins belonging to certain cellular compartments. For example, mass spectrometry has been used to study the global ubiquitin proteome by immunoprecipitating ubiquitin [6], phospho-tyrosine proteomes by using pan phospho-tyrosine antibodies [7], and autophagy-regulatory proteins by precipitating proteins involved in vesicle trafficking [8]. These insights have provided a comprehensive parts list for important regulatory functions in biology.

One particularly exciting application of mass spectrometry in signaling research is the use of multiplex kinase-inhibitor beads (MIB) [9]. In MIB-MS, a cocktail of inhibitors that promiscuously bind to many kinases are conjugated to beads and mixed with protein lysate. Several inhibitors are used to target the majority of the expressed kinome. The binding of a particular kinase to the bead occurs as a function of the kinase's native affinity for the

inhibitor, its level of expression, and also the level of kinase activity since kinase inhibitors bind preferentially to activated forms of kinases. Thus, differential binding between experimental conditions can be inferred in most cases to reflect differences in kinase activity. Duncan et al. used MIB-MS to profile kinome activity in breast cancer cells revealing total kinome reprogramming following targeted MEK inhibition [2]. In addition to these multi-inhibitor conjugated beads, others have used covalent kinase inhibitors in a similar manner. Activity-based proteome profiling (ABPP) uses small molecule probes designed to react with nucleophilic residues, most typically cysteine, near the kinase active site to form irreversible covalent interactions. These activity-dependent interactions can be specific for target kinases, although recent research suggests that off-target binding to both kinases and non-kinase proteins can be problematic [10].

While powerful, some issues do potentially exist for the use of mass spectrometry in the study of signal transduction. For example, decoding mass spectra to reveal peptide identity is non-trivial in general, and this may be worsened in signaling studies because many signaling proteins share similar domains, some of which have very similar and sometimes identical peptide sequences. Excellent software exists to aid in peptide identification, however, many of these statistical tools may be underestimating false discovery rates, sometimes dramatically so [11,12]. Furthermore, mass spectrometry based proteomics is relatively slow, expensive, and does not scale well to the analysis of hundreds or thousands of samples which can easily emerge from studies that involve many ligand doses, experimental time points, and other perturbations.

Undoubtedly, many of these issues will ultimately be resolved. Mass spectrometry is a mature but ever-changing technology with many different specific iterations. For example, while mass spectrometry typically subsamples the input protein, recent studies report deep, even full, proteome depth [13]. Also, when proteomes are subsampled, comparisons between samples are made non-trivial, but these comparisons can be made using labeling methods like SILAC [14], the use of isobaric mass tags, or in label-free approaches with the aid of computational tools [15]. Recent advances in targeted mass spectrometry have increased the specificity and sensitivity of target identification and analysis [16].

Immunoblotting—Traditional methodologies based on western blotting continue to be tried and true tools for measuring total protein levels and levels of specific modified forms of proteins, assuming suitable antibodies exist (see note below on antibody-based proteomics). Using various biochemical separation and/or immunoprecipitation techniques, it is also possible to quantify protein localization to various subcellular compartments, or interactions between proteins and other proteins or biomolecules. These tools are often comforting to biological scientists both because of their familiarity and because visualization of a band of predicted size on a gel provides assurance that the antibody is reporting the correct protein. If paired with an appropriate analysis platform – one with a large dynamic range that is confirmed to be linear or whose non-linearities can be fit to a standard curve – these approaches can be quantitative[17]. Although the multiplexing capabilities on immunoblots are limited to two or three parallel measurements, the relatively low cost of these approaches allows throughput to be achieved through repetition, particularly when the lysate material is abundant. Also, modern commercially available platforms have dramatically shortened the

gel electrophoresis and transfer durations, such that it is feasible for a single operator to analyze hundreds of samples per day, each for dozens of specific analytes. Immunoblotting remains a common tool for validating and precisely studying single proteins or even small protein networks or as complementary methods to “fill in the gaps” left by other high-throughput techniques. For most applications, however, these approaches simply lack the throughput necessary for generating comprehensive OMIC-scale quantitative data sets.

Protein Lysate Microarrays—A higher throughput approach that uses the same intuitive design as immunoblotting are lysate microarrays [18]. In these approaches, antibodies are still needed to quantify the levels of proteins or modified protein species. These approaches utilize robotic tools to print material in regular patterns, typically on a smaller and denser scale that could be achieved by hand. Analysis of protein lysates in this miniaturized format requires much less biological material than standard approaches, and uses dramatically reduced amounts of expensive reagents like antibodies. Two basic formats exist: forward- and reverse-phase arrays. Forward-phase arrays are constructed by printing capture antibodies on a flat surface and flowing a cellular lysate over the printed antibody chip. In this approach, a second antibody is typically then applied to determine the level of protein captured by the immobilized antibody. While this approach negates the need to label the lysate material, multiple specific antibodies must exist that can bind to the target protein without competition.

In addition to the need for multiple target specific antibodies, forward-phase arrays also generally require relatively large amounts of protein lysate. Reverse-phase protein arrays (RPPAs) alleviate this issue. Here, small amounts of protein lysates (~ picogram range) are spotted and dried onto a rigid protein-binding surface like nitrocellulose. The protein lysate array can then be probed and measured in a manner akin to a traditional immunoblot. RPPA can rapidly measure many lysates in parallel, wherein “multiplexing” is achieved through repeated printing of lysates onto parallel arrays and probing with different antibodies. RPPA are likely the most cost effective and highest-throughput solution for experiments that include many samples (~100-10,000+) with relatively few target analytes of interest (dozens to hundreds). For example, a single array can be printed to include thousands (even 10,000-100,000+) of individual lysates, which may represent different cell types, drug treatments, time points, patient samples, etc. Once printed, RPPAs essentially function similarly to high-throughput “dot blots,” and they share many of the same benefits and limitations. For example, RPPAs omit the electrophoresis step used in immunoblotting. This omission reduces assay time but also can increase non-specific antibody binding, which limits dynamic range and assay sensitivity. Thus, antibodies used in RPPA need to be carefully validated [19,20]. A modification to the RPPA method that may improve sensitivity is the micro-western array [21]. The central difference between these methods is that cell lysates are printed onto a gel and undergo a short semi-dry electrophoresis step before being transferred to a membrane and probed with primary and secondary antibodies. The short electrophoresis step may separate competing epitopes and allow for resolution of the specific protein bands, assuming that non-specific binding proteins migrate differently (i.e., are of a different size) [22].

In spite of some obvious limitations, many antibodies have been carefully validated and hundreds of protein targets can currently be studied using RPPA [23]. Many researchers have used RPPA and other array-based methods to study signaling, particularly in cancer biology. For example, RPPAs are used to analyze primary human tumor samples as a part of the Tumor Cancer Genome Atlas (TCGA) [24]. Considering the limited amounts of patient material available for each sample and the relevance of signaling proteins in targeted medicine, RPPA plays a critical role in this analysis. In another example, Federici et al. recently used RPPA to interrogate signaling pathway activation states of all cell lines in the NCI-60 [25]. Their pathway activation mapping approach revealed several phenotypes that were common to subsets of the NCI-60 cell lines and transcended typical organ-type distinctions. By integrating these data with other OMIC data for coding and non-coding RNAs and genomic features, they were able to identify functional networks that significantly improved prediction of drug sensitivity compared to traditional approaches that use only genomic or transcriptomic features.

Multiplexed Flow Cytometry, Mass Cytometry, and Bead-Based Assays—For many signaling proteins, the population expression or activity levels measured in bulk lysates can be deceiving. For example, at the single-cell level, some signaling systems only exist in “on” or “off” states and graded population responses actually reflect differences in the percent of cells that are “on,” not the degree to which the pathway is “on.” This scenario is obviously true for binary signals or “all-or-none” processes like cell death [26], but also true for other processes that may intuitively appear more graded than binary. In fact, even in uni-modal distributions, cell-to-cell heterogeneities can be substantial and underlie divergent behaviors, including sensitivity and resistance to drugs [27,28]. Flow cytometry is probably the most widely used assay to quantitatively study signal transduction with single-cell resolution. The basic workflow is similar to other antibody-based techniques, although cells are first fixed as individual cellular units, before permeabilization and labeling with specific antibodies. Labeled cells are passed in single file through a flow cell where fluorescence intensity, cell size, and cellular internal complexity can be measured. In OMIC applications, flow cytometry suffers from two basic weaknesses. First, this is not a high-throughput method, since sample preparation and read time can be quite extensive, particularly for adherent cells. Second, multiplexing of measurements is limited by the available fluorescent spectra. Using advanced machinery and careful fluorescence compensation, one can multiplex more than a dozen measurements [29], but most users measure 1-12 signals per sample.

To improve upon the multiplexing limitations of traditional flow cytometry, Scott Tanner and colleagues coupled the single-cell profiling capabilities of flow cytometry to the exquisite sensitivity of mass spectrometry [30]. Mass cytometry uses antibodies conjugated to rare-earth metals, typically those in the lanthanide series, instead of fluorophores [31]. Samples are first processed using flow cytometry techniques to create a stream of single cells, which are subsequently vaporized into single-cell droplets. The amount and identity of the bound antibody is quantified using a mass spectrometer to detect differences in mass corresponding to the different atomic masses of the rare-earth metals. Using mass cytometry, one could theoretically analyze more than 100 unique analytes in a single cell, but currently 30-50

distinct signals are feasibly analyzed simultaneously. Mass cytometry is still nascent technology but has already transformed our understanding of immune cell development. In a recent publication, Nolan and colleagues used mass cytometry to label 34 parameters in human hematopoietic cells, including surface and intercellular proteins [32]. Their analyses revealed previously unknown relationships between precise signaling states and well characterized hematopoietic cell subtypes and revealed a more refined view of hematopoietic stem cell developmental trajectories.

While mass cytometry offers impressive multi-plexing capabilities, other flow-based approaches can expand multiplexing capability by an order of magnitude, although not with single-cell resolution. Multiplexed-bead based assays (like those sold commercially by Luminex) combine antibody labeling with laser-based cytometry. Here, polystyrene or magnetic beads are filled with specific ratios of two or three dyes, creating a colorimetric signature analogous to a barcode used in genomic or RNA sequencing methods. Each uniquely colored bead is then conjugated to an antibody against a specific protein such that measured fluorescence levels can be attributed to the specific protein of interest. This technology allows for the concurrent measurement of 500 simultaneous signals per sample, although antibody cross-reactivity typically limits the multiplexing capability well below this threshold. Using this method, Du et al. recently profiled nearly all tyrosine kinases in hundreds of cancer cell lines, each treated with dasatinib, an inhibitor of the tyrosine kinase, Src [33]. A similar strategy was also used to increase the throughput of transcriptome profiling. In this case, antibodies were replaced with oligonucleotide probes for each of 1000 “landmark” genes, which allowed inference of genome wide transcriptional states at a fraction of the cost of traditional array or sequencing methods [34].

Caveats in Antibody-Based Proteomics—Essentially all of the above mentioned approaches rely on the existence of high-fidelity antibodies to detect the specific epitopes of interest, and a discussion of their use must consider the main limitation: the lack of reliable antibodies for most protein targets [35]. Also, with the exception of immunoblotting, the other methods all attempt to bind to a target epitope without an electrophoresis step to remove other competing off-target binding sites. Thus, the proportion of available antibodies that produce reliable signals in these high-throughput platforms is only a fraction of the antibodies that produce reliable signals in immunoblots. Estimates vary, but generally between 1-10% of antibodies that work in immunoblots will work in the other methods we discussed. Also, because most methods do not produce a visible band on a gel, additional measures should be taken to validate that the signal is reporting the levels of the correct protein. When validating antibodies for RPPA, for example, we previously used a two-step protocol. We began first by testing antibodies in RPPA and identifying those that reported a dynamic range of at least 1.5-fold across 90 generic control conditions, which were designed to feature positive and negative controls for all targets of interest. In a second phase, we retested that small subset of antibodies in immunoblot format to confirm that the change in signal occurred in a protein of correct predicted size and that the two methods produced correlated data [3]. Additionally, antibodies that are accurate when used in isolation may not produce reliable signals when multiplexed with other antibodies due to cross-reactivity, so

the reliability of measurements derived from highly multiplexed platforms should be rigorously tested and validated.

Methods to Measure Protein-Protein Interaction

Many post-translational modifications or activated signaling proteins cause conformational changes or the creation of new binding domains, resulting in regulated protein-protein interactions. Thus, measuring the extent to which proteins interact can indicate signaling pathway activity, and identifying new binding partners may reveal new proteins involved in signal regulation. Below is an overview of common tools used to measure protein-protein interaction in high throughput.

Yeast 2-Hybrid (Y2H)—The yeast-2-hybrid system is a classic approach to identify protein-protein interactions [36]. This approach takes advantage of the modularity of many transcription factors, which have two separable functional domains that confer DNA-binding and transcription-activating functions. In Y2H, a protein of interest is fused to the DNA-binding domain while a protein or library of proteins to be screened are fused to the corresponding transcriptional activation domain. Protein-protein interaction can then drive transcription of a reporter gene [37]. There are several caveats to this method. Non-physiological false positive interactions can occur either due to overexpression of the screened proteins or because proteins are forced to coexist in the nucleus that may not typically reside in a shared subcellular compartment. Also, false negatives can occur. Y2H will often miss dynamically regulated interactions, or interactions that require a post-translational modification that may be lacking in yeast or not properly regulated in the assay. Also, since the readout for binding occurs in the nucleus and only proteins that can accommodate nuclear localization can be screened.

Y2H screening can be performed in high throughput, even on a comprehensive scale. Using Y2H, Vidal and colleagues have recently published the first full proteome scale map of the human protein-protein interaction network [38]. This systematic analysis has doubled the list of known/potential binary interactions, and strikingly, their analysis has revealed that the interactions identified by systematic Y2H are free from a bias towards well-studied proteins that exists in the established interactome. Thus, this dataset is certain to include clues into many unexplored functional relationships.

MYTH, MaMTH, and other protein complementation assays—The nuclear localization of traditional Y2H assays can be a severe limitation for some aspects of signal transduction. Transmembrane proteins, for example, play important roles in signaling as the primary signal relay between the cell exterior and interior, and regulated protein-protein interaction among transmembrane proteins is a major mechanism by which signals are encoded at the receptor layer of a signaling pathway. For example, the primary means by which receptor tyrosine kinase autophosphorylation transmits a signal is through the recruitment of proteins containing phospho-tyrosine binding domains. Determining binding interactions in high throughput for membrane proteins is challenging since they contain large stretches of hydrophobic residues and cannot be readily expressed in aqueous solutions. Membrane yeast two hybrid (MYTH) and a modified version that works in human

cells, mammalian membrane two hybrid (MaMTH), were developed to circumvent this issue and allow for high-throughput screening of membrane protein interactions [39,40]. Both assays rely on the concept of split-ubiquitin binding. A transmembrane protein of interest is tagged with both the C-terminal half of ubiquitin and a transcription factor that is used to drive transcription of a colorimetric or fluorescent reporter protein. Proteins that are to be screened for binding are tagged with the N-terminal half of ubiquitin. Binding events between the membrane protein of interest and the screened protein yield the formation of a complete ubiquitin molecule, which is recognized by endogenous deubiquitinating enzymes, releasing the tethered transcription factor to drive reporter expression. Using MaMTH, Petschnigg et al. studied EGFR signaling in non-small cell lung cancers containing EGFR mutations [40]. By screening binding partners of the mutated receptor and comparing them to a wild-type EGFR signature, the investigators were able to identify a novel binding protein that promotes constitutive signaling. MaMTH is a new technology that will certainly be used in coming years for the study of receptor signaling, and drug mechanism of action in human disease.

In addition to these approaches, several other methods exist that rely on the complementation of a split reporter protein, which can regain function when the two halves are in proximity. In addition to the use of ubiquitin or transcription factors, other similar strategies have used fragments fluorescent proteins (Bimolecular Fluorescence Complementation or BiFC) [41], of the tobacco etch virus protease (Split TEV) [42], dihydrofolate reductase (DHFR) [43], beta-lactamase [44], luciferase [45], LacZ [46], and many other proteins, demonstrating the versatility and generalizability of this technique.

LUMIER—Another commonly used method for detecting protein-protein interactions is co-affinity purification. To adapt this generic strategy to high-throughput analyses, the Wrana Lab developed LUMIER (luminescence-based mammalian interactome mapping). In this method, a protein of interest is fused to luciferase and co-expressed in mammalian cells with target proteins fused to an affinity purification tag. After signal stimulation and cell lysis, interaction/co-immunoprecipitation can be inferred through luminescence intensity. The main advantage of this approach over other co-IP based methods is throughput, especially since these assays are sensitive enough to be performed in micro-titer plates. LUMIER was originally used to investigate transforming growth factor- β (TGF- β) signaling in human cells. Wrana and colleagues screened over 500 potential TGF- β binding partners and identified links between the TGF- β pathway and various other processes, such as epithelial tight junction regulation, which contextually help regulate TGF- β signaling [47]. More recently, Lindquist and colleagues applied the LUMIER technique to comprehensively study the Hsp90 interactome [48], and to study the specificity of kinase inhibitors in living cells [49]. In the latter example, the authors leveraged the fact that binding of small molecules to kinases results in thermodynamic stabilization, and thus loss of interaction with chaperone proteins. This study revealed critical targets of kinase inhibitors, including many that were unexpected.

Protein Chips—In addition to protein lysate microarrays described above, many other types of arrays can be built to study protein biology, some of which are powerful tools for

studying protein-protein interaction rather than expression levels [50]. Protein chips are engineered by arraying and immobilizing purified proteins or protein domains, instead of antibodies or lysates. Protein chips have many uses. For example, a protein, lipid, nucleic acid, or small molecule of interest could be profiled against a near-proteome-wide protein chip to reveal all potential binding partners[51]. Huang et al. used this approach to identify new components of mTOR signaling [52], and Jones et al. used protein chips featuring nearly all human Src homology 2 (SH2) and phosphotyrosine binding (PTB) domains to comprehensively study protein recruitment to activated ErbB receptors [53].

Quantitative Measurement of Protein Localization or Protein Activity

For many signaling nodes, the activation status cannot be inferred from the levels of a post-translationally modified protein or from the interactions between proteins. In these cases, it may be necessary to directly measure enzymatic activity or localization of the signaling protein of interest. In general, available methods can be quantitative but are not easily performed in high-throughput. Here, we briefly survey quantitative methods for studying protein activity and the dynamics of protein localization.

Quantitative Microscopy—The transmission of a signal is often driven by a translocation event of a protein, for example from the cytoplasm to the nucleus, or the clustering of a protein to the plasma membrane. Thus, for some signaling nodes, quantitative microscopy can be a reliable approach to monitor pathway activation. Additionally, microscopy may be the best approach when studying rapid signals or for those that occur in an oscillatory manner, as is the case for NF-kappa B signaling [54] and p53 signaling [55,56]. Recent analyses of NF-kappa B signaling by Covert and colleagues revealed substantial heterogeneity at the single-cell level. These quantitative single-cell insights were instrumental in revealing the use of both digital and analog features to differentially control activation of specific NF-kappa B target genes [54]. Single-cell resolution has also been valuable for the study of signaling processes that are initiated stochastically. For example, a critical event in apoptotic cell death signaling activation is the formation of pores that permeabilize the mitochondria (called mitochondrial outer membrane permeabilization or MOMP). The timing of MOMP after cell death stimulation is variable, but can be directly measured using fluorescence resonance energy transfer (FRET) sensors, fluorescent reporter constructs, or membrane potential dyes that localize to the mitochondria [26,57]. Single-cell analyses of MOMP dynamics have revealed a nearly invariant “switch-like” activation that is masked by the graded population level responses, and computational modeling of these data have revealed critical features that contribute to this binary response as well as to activation stochasticity [58].

Quantitative, automated microscopy is becoming increasingly available as microscopes have become better and cheaper. Image processing and quantification, however, remain major challenges in the use of quantitative microscopy. Several freely available image analysis programs exist to aid in quantitative image analysis, including CellProfiler [59], PhenoRipper [60], ImageRail [61], and ImageJ [62].

In vitro Kinase and Other In vitro Enzyme Assays—In many kinases, phosphorylation on the activation loop can safely be used to indicate kinase activity. Unfortunately, this feature is occasionally misinterpreted to assume that all phosphorylation events on kinases modulate their activity. Most phosphorylation on kinases does not alter their kinase activity and many kinases do not require activation loop phosphorylation to become active. Furthermore, for many kinases, good activation-loop antibodies do not exist. Thus, the only universal method for determining the activity of all kinases is to measure it directly. *In vitro* kinase assays have been traditionally performed using purified proteins or cell lysates incubated with radioactive isotopes and purified substrate proteins or peptides. These assays are reasonably straight-forward, do not require much specialized equipment, and can be performed in medium- to high-throughput with very little material [63]. Because many kinases are promiscuous, these assays often need to be conducted in the presence of inhibitors or following enrichment or purification. Accordingly, these assays likely cannot be performed at a comprehensive OMIC scale. Thus, most practical applications will use *in vitro* enzyme assays to study single proteins, very small networks, or as a complementary assay in conjunction with other approaches. For example, Janes et al. used *in vitro* kinase assays in conjunction with immunoblotting and antibody arrays to study contextual dependencies in cytokine-induced apoptosis [64]. Recently, the Gygi lab has developed high-throughput methods to directly monitor kinase activity against synthetic substrates *in vitro* using mass spectrometry [65]. This approach, called KAYAK (kinase activity assay for kinome profiling) uses synthetic peptides featuring motifs for known kinases. Arrayed or pooled peptides are mixed with cell lysates and the acquisition of peptide phosphorylation is quantified by mass spectrometry following some enrichment steps. This approach was used to quantify the activities of 90 kinases across various treatment conditions in many different cell lines.

Of course, kinases are not the only enzymes and most enzymes can be purified and studied biochemically in the cell-free *in vitro* environment. For example, *in vitro* biochemical assays are a common method for studying the activity of GTPases, which do not leave a physical “mark” on their substrates. Methods that have sufficient throughput and coverage to be considered “OMIC” approaches are rare, but some exist and certainly others are in development. For example, assays have been developed to facilitate high-throughput quantification of phospho-protein specific phosphatase activity [66], as well as matrix metalloprotease activity [67].

For most *in vitro* enzyme assays, the main limitation to throughput is the need to purify specific proteins. Activity-based probes and chemical sensors provide an excellent alternative when available. One example are sensors that exploit the chelation enhanced fluorescence of sulfonamido-oxine (Sox) chromophores, which increase in fluorescence when bound to Mg^{2+} [68]. These chromophores can be conjugated to peptides containing phosphorylation motifs for a kinase of interest, and engineered in such a way that Mg^{2+} -binding requires peptide phosphorylation. In addition to Sox chromophores, other fluorophores and dyes exist whose fluorescence properties are sensitive to the micro-environment in which they are bound [69,70]. These reagents are particularly valuable for studying signals transmitted through G proteins, which transmit signals only by virtue of a

small conformational change upon GTP binding. Because their activation cycles in vivo are short and their interactions with substrates relatively weak, this large and important class of signaling proteins has traditionally been difficult to study in the context of living cells. FRET-based activity probes were used by Klaus Hahn and colleagues to visualize the location and dynamics of RhoA activation in living cells [71]. These studies revealed distinct spatiotemporal patterns of RhoA activity in response to different extracellular cues, revealing a previously unappreciated level of richness in signaling diversity.

Inference of Pathway Activity Through Transcription or Functional Genomics

When collecting signaling data, it is not always possible (and sometimes not necessary) to view the signal at each node as it is transmitted through a network. In these cases, an end-point measurement of network activation may be a sufficient substitute. Most stimuli-dependent signaling events result in a change in transcriptional states, whether to activate or repress certain genetic programs. A transcriptional readout can be used to infer the signaling state of a cell using microarray, RNA sequencing, or microfluidic technologies that allow for high-throughput qPCR [34]. These approaches are quantitative, high throughput, and can be performed on a comprehensive scale. The main limitation is the difficulty in relating a particular transcriptional state to the activation or inhibition of a discrete signaling protein. Some computational approaches exist to aid in this effort. The most widely used would be GSEA (gene set enrichment analysis) [72]. GSEA uses a rank-ordered list of differentially expressed genes to identify enrichments or depletions of *a priori* defined genetic “signatures” (i.e., cohorts of genes that are somehow related to each other or to a biological process). This approach is not formally restricted to the study of signal transduction, but its application to signaling research is possible by using signatures that report signaling states. Additionally, the Califano lab has invented several approaches for reverse engineering regulatory networks using transcriptional profiles [73-75]. Most of these algorithms focus on transcription factor activity rather than on the activity of upstream signaling proteins, but in many cases the “master regulators” identified are proteins with defined roles in a signaling circuit. Inference algorithms that use proteome data rather than transcriptome data also exist. The Linding lab has developed several approaches that use mass spectrometry data to infer the connectivity of signaling networks [76].

In addition to transcription- and mass spectrometry-based methods, functional genomics can be used to infer signaling mechanisms. Several screening approaches are common in biological research, including the use of specific drugs or small molecules, shRNA/RNAi, or clustered-regularly-interspaced-short-palindromic-repeat-based genome editing (CRISPR) [77,78]. These perturbations can be paired with simple measures of pathway activation (even at a single node) to reveal functional signaling networks. Screens can be performed in a high-throughput manner but hits typically need to be validated by an independent method due to high potential for off-target effects. With all inference-based approaches, observed phenotypes will be a mixture of direct, indirect, and compensatory effects. Deconvoluting these effects can be challenging, particularly because signaling proteins typically have multiple roles in different contexts and are shared in many different processes, creating extensive crosstalk. Functional genomics are ideally suited for revealing a molecular “parts list.” However, these approaches do not typically result in sufficient understanding of

pathway structure, which may be needed to make reliable predictions about how cells will behave following pathway stimulation or inhibition at various nodes. Thus, these approaches are best used in combination with more direct measures of signaling activity and in close partnership with computational approaches.

Experimental Design for OMIC Experiments: Basic Principles and Practical Considerations

Experimental design is an important issue in all areas of research, but some special considerations are needed in OMIC experiments due to their size and complexity. Additionally, OMICs experiments are expensive, with a single experiment often costing thousands of dollars. These experiments therefore demand effective and efficient designs and error-free execution because they are not feasibly redone. Several features that are common to OMIC experiments, like the use of multi-well plates, can create strong batch effects [79]. Proper experimental design is necessary to minimize and remove, or at least identify, these problems. In the early days of “big data” science, some practitioners held the view that OMIC experiments could be held to a different, maybe lower, standard of experimental rigor (leaving out some controls, not performing biological or technical replicates, etc.). These decisions were often made based on cost feasibility and also with the idea that data generated were just “a screen” and would be carefully validated afterwards. If these ideas did not seem crazy at the time, they certainly should now. Large signaling datasets are not being used solely as unvalidated resources, but also by computational biologists to aid in model parameterization or by pharmaceutical companies to motivate drug development programs [80-82]. Implementing the basic principles of sound experimental design and including all necessary controls are essential steps for minimizing bias and maximizing interpretability and reliability of OMIC experiments. Experimental design principles for proteomics experiments have been reviewed in detail elsewhere [83-86] and Glass has presented a valuable discussion of the various types of controls needed to interpret biological experiments [87]. In this section, we briefly highlight some key concepts.

A major challenge of biological experimentation is separating signal from noise. In statistical terms, the experimental signals represent biological sources of “variance,” i.e., changes in the measured signal caused by the experimental treatments, which the experimenter seeks to quantify. The noise of an experiment will likely arise from two principal sources: bias and random error. Bias is the systematic deviation of the measured effect from the “true effect”. These deviations can be attributed to both technical and biological sources, and presumably the experimenter only cares about the latter. Random error refers to the unsystematic errors that enter the experiment due to unknown sources, which reduce the precision of the measurements and the ability to detect treatment effects. In OMIC experiments, because the operator cannot optimize each analyte, data point, or assay individually, the signal-to-noise requirements will likely be more stringent than for a low-throughput assay that can be more easily fine tuned per experiment.

The three universal principles of sound experimental design are randomization, blocking, and replication. Randomization is a process that is applied in two contexts: first, experimental units are allocated to treatment groups with equal probability and second, experimental units are treated (and the resulting samples processed) in a random order.

Randomization helps to reduce bias by minimizing the probability that experimental units with similar characteristics are assigned to the same treatment group. Randomization also helps satisfy the necessary condition of independence, which is a common assumption in most statistical tests.

Blocking refers to the allocation of treatment groups into experimental units or “blocks.” Blocks are subgroups of experimental units that exhibit less variance within groups than between groups. An example of an experimental block in an OMIC experiment is each 96-well assay plate within an experiment that features multiple plates or each sequencing lane in a RNA sequencing experiment. In multi-well plates, even the well location represents a potentially confounding block. Despite the experimenter's best efforts to treat them equally, each block is processed at a different time, which may introduce a systematic variance. These block or “batch” effects are prevalent in OMICs experiments [79]. Defending against batch effects requires the use of blocking strategies in which replicates from the same treatment group are distributed across different plates (and preferably in different locations on a plate), rather than loaded together on a single plate, in order to ensure the batch effect does not confound the biological effect.

Replication is the repeated treatment and/or assaying of multiple experimental units. Replication is needed for precise and accurate results, for generalizing the results of the sample to the population from which it was drawn, and for generating statistically powerful tests. Two replicate types are commonly featured in OMICs experiments: biological and technical. Some ambiguity exists in the community for what constitutes a biological versus technical replicate [88], but in most instances, a biological replicate represents an experimental unit that was treated independently of the other units. Examples include each animal of a mouse experiment or each dish of cells in a cell-culture-based experiment. Technical replicates consist of samples drawn from the same experimental unit that are measured independently. While technical replicates are useful for assessing the variability of the assay, for most applications, biological replicates should be emphasized because they incorporate both the biological and technical sources of variance. Furthermore, replication should occur across blocks that reflect principal technical sources of variance in order to avoid bias and to maximize generalizability. For example, an experiment should be purposefully performed in separate batches, even if not logistically required to do so, in order to obtain a more accurate effect estimate. The use of inappropriate replication strategies is called “pseudoreplication” and is a common problem in many reported studies [89]. An added complexity common to many OMIC approaches is the use of nested data, which are data featuring multiple observations derived from a single sample, as in the use of highly multiplexed signaling measurements. These nested data invalidate the presumption of independence. Recent reports have discussed methods to address some of these issues through optimal experimental design and altered strategies for replication [90,91].

Additional design issues include the decision of whether to pool samples, which is usually considered when costs must be kept low, and in sample size calculation. These figures are required to estimate the number of replicates that the experiment must have to detect meaningful treatment effects with acceptable probability, if they in fact exist. Often is the

case, however, that sample size is fixed due to logistics (e.g., cost). In such cases, strategies exist to pick optimal type I and type II error pairs [92].

Conceiving OMIC Experiments: Hypotheses Or Discovery?

An additional consideration is whether the OMIC approach will be used to facilitate a hypothesis-driven experiment or strictly for discovery. Since the inception of systems biology, the validity of discovery science has been a tough sell to grant panels and thesis committees, who generally expect hypotheses to be tested [93]. However, specifying sensible hypotheses for OMIC experiments is sometimes challenging, which has led to revisiting the debate that science is only valid when using Popper's hypothetico-deductive logical framework [93,94]. In practice, scientific advances are made in many ways, including deduction (hypothesis falsification), induction (data-driven discovery) [94], the question/model cycle [93], and strong inference (propose multiple competing hypotheses and design experiments to discriminate them) [95,96]. Discovery science should therefore be considered valid, at least in principle. Even so, all forms of science, from those that are purely discovery driven to those that are hypothesis driven, do need to start with a cogent and well conceived scientific question [97], and the lack of a good question, not the lack of a hypothesis, likely caused the limitations of many studies that were based in discovery.

Considerations in Discovery Science—We have observed that discovery science has pitfalls that can render experiments uninformative, thus wasting time, money, and effort. First, effective discovery experiments require just as much, if not more, careful thought in conception and design as do those seeking to test a hypothesis. The designs of these experiments are too often given insufficient thought, the reasons for which include a lack of training among biomedical researchers in experimental design and analysis [98], as well as the strong temptation to just “do the experiment” because “we’ll find something.” The latter attitude likely arises due to the desire to publish outcompeting the desire to seek the “truth.”

Second, effective discovery experiments demand that the measurement technology measures most if not all of the system in an unbiased manner, such that the results can be reproduced and generalized. For sequencing genomes, this ideal is arguably met [99], whereas it is almost certainly not met for the other OMIC methods that we discussed above, all of which are widely used in systems research. The data from such experiments may therefore be idiosyncratic to the specific experiment and not safely generalizable.

Third, having a well justified hypothesis is important for determining the prior probability of a result being true, which in turn determines the false positive and false negative probabilities of the result once obtained [100]. Accordingly, OMICs-based measurements based on weak or non-existent hypotheses should be expected to be rife with false results, which in some cases may exacerbate the problem of irreproducibility of biomedical science [101,102].

Given these pitfalls, it is worth highlighting that systems biology was not, and is not, meant to be a discovery science [103]. Although the term “systems biology” was coined at least fifty years ago, the modern era of systems biology is typically said to begin with the Human Genome Project [104]. The Human Genome Project, and other endeavors that purely

practice discovery science are commonly grouped with systems biology due to the use of similar OMIC technologies. While these fields are related, they are not the same. A relationship exists because discovery science aims to define the elements, whereas systems biology aims to define the functional relationships among these elements. But in another way, discovery science and systems biology are very different. Discovery science and hypothesis-driven science philosophically contrast each other, but the same discordance does not exist between hypothesis-driven science and systems biology. Since its initial definition, it was proposed that “The integration of these two approaches, discovery and hypothesis-driven science, is one of the mandates of systems biology” [104]. Thus, a systems-level study is meant to use OMIC-scale data to perform predictive, hypothesis-driven science. Generating OMIC-scale hypotheses is challenging, and systems biologists continue to struggle with casting their work in a logically satisfying intellectual framework.

Strong Hypotheses For OMIC-Scale Systems Biology

Three basic types of hypotheses apply to OMIC experiments in signaling research:

- 1) the “discovery” hypothesis that at least one of the analytes will be affected, i.e., the levels of a protein or its post-translationally modified forms will increase or decrease in response to at least one of the experimental perturbations,
- 2) a systems-level hypothesis pertaining to a property of the *system* or to a mathematical or computational model of the system. For example, the hypothesis can focus on the network topology (e.g., that it is scale-free or that some importance can be inferred from the structure), the estimation of parameter values, and/or model discrimination, i.e., distinguishing whether one model explains the data better than alternative models.
- 3) specific (reductionist) hypotheses about one or more of the measured analytes, for example, that c-Jun-N-terminal kinase will be phosphorylated in response to TNF- α treatment.

The first type of hypothesis is the most prevalent in the systems biology literature and the third the least prevalent. This ordering is likely due to the burden for finding evidence *a priori* to support the hypothesis. In the case of the first two types of hypotheses, relatively little evidence is needed for justification. In the case of the third type, one would have to justify distinct hypotheses for each analyte, which to do properly would require systematic reviewing the relevant literature and databases for each analyte in order to determine how it is likely to respond to perturbation(s) in the experiment. The advantage of doing this initial work prior to conducting the experiment is that specific elements of the design might be better resolved. Furthermore, it may help to alleviate the bias that could be introduced by doing the same type of search after the experiment to rationalize the “hits.”

Each of these hypotheses is acceptable as long as they are exploited for specifying the experimental design and analysis, which is ultimately the *raison d'être* of hypotheses. In crafting a strong hypothesis, no matter the type, the variables must be operationalized, assumptions explicitly stated or considered, and thought given to the extent to which the conclusions can be generalized. Articulating hypotheses that contain these details facilitate

decision-making regarding the design of the experiment, the statistical analyses, and the appropriate interpretation of the data. In practice, bridging the gap between OMICs-driven discovery science and question- or hypothesis-driven science often requires some form of computational or mathematical modeling. As with the experimental approaches discussed above, these methods each have distinct strengths, limitations, and design principles.

III. APPROACHES TO MODELING SIGNAL TRANSDUCTION PROCESSES

For most signaling processes, a substantial amount of information has been accumulated already about the proteins involved (the parts list or “nodes”) and the functional relationships or connections between these proteins (the “edges”). Thus it is worth considering whether adding another node or another edge will improve our understanding of the system. One such example is shown in Figure 2, a wiring diagram of the integrated signaling pathways that control cell death, including DNA damage, cytokine, growth-factor, and other stress-response signals. The edges shown are all high confidence, meaning they “probably” all exist. Even so, these “probably” do not exist in all contexts, or at all time points, and certainly not all with the same levels of importance. The dense interconnectivity and temporal/contextual dependence, impairs our intuitive understanding of how the system operates. This situation worsens with each new insight, whether it is new nodes or new edges. Thus, once OMIC data are collected for a signaling process of interest, additional tools are typically necessary to facilitate the transition from “data” to “understanding.” Highly quantitative data of the type discussed above can be modeled using various approaches (Figure 3). Different approaches might be preferred in different scenarios. Considerations include the nature of the data available (amount, specificity, level of detail, depth of information within a signaling “pathway,” breadth of information across parallel pathways), the size of the process being modeled (i.e., the number of nodes and edges), and the question being asked. Other more authoritative and comprehensive reviews exist for each of the methods that we will discuss [105-108]. Here, we survey common approaches that can be used, with an emphasis on statistical approaches that may be optimal for studying large, complex, and imperfectly understood networks.

Mechanistic and Semi-Mechanistic Approaches

The most commonly used framework for modeling signal transduction processes is kinetic modeling, which are typically expressed as ordinary or partial differential equations (ODEs or PDEs). ODEs are attractive because they are a natural framework for dynamical systems such as signaling [106]. Each ODE represents the rate of change for one molecular component, which is a function of the rates of each biochemical transformation that influences the level of that component (i.e., enzyme-catalyzed reactions, protein-protein interactions, and transport between compartments). The rates of the biochemical transformations are functions of the concentrations of the reactants (the molecular components) and intrinsic kinetic parameters. The concentrations of the molecular components change over time but the kinetic parameter values are typically held constant. Because the molecular components are linked by their overlapping participation in a common set of biochemical reactions, the equations for the individual molecular components are organized into a system of inter-dependent ODEs. In most applications to

date, the models were calibrated to experimental data consisting primarily of time series of protein levels, measured using immunoblotting, in response to experimental perturbations such as stimulation by receptor ligands (e.g., hormones or cytokines). These models can be used to simulate the behaviors of a system over time and are particularly useful for testing conditions that would otherwise be infeasible to test (due to lack of reagents, cost, etc.). Kinetic modeling can be used to generate testable predictions, rule out possible mechanistic explanations or competing hypotheses, or identify the system parameters to which the model behavior is particularly sensitive or insensitive.

The dynamic nature of signal transduction makes kinetic modeling the ideal framework. Indeed, some flexibility exists such that the approaches can be used to model spatial information [109], stochastic or deterministic biological processes, and even responses that occur on very different time scales (e.g., cell division, which occurs on the timescale of hours, in response to hormone signaling, facets of which occur on shorter timescales). However, these approaches are most effective for signaling systems that are relatively well understood and for which the rate constants are known or can be easily measured. These two factors typically represent major obstacles to implementation because our understanding of signaling network biochemistry is far from complete. In particular, our understanding of crosstalk is poor for most signaling processes, likely due to our long history of studying signaling using reductive approaches. Furthermore, even for systems in which crosstalk is known to exist, these connections are often context specific [110], thus causing an added complication that is only recently being explicitly explored in a systematic fashion [111]. These problems scale with the size of the modeled system, such that in practical terms, kinetic modeling is typically best suited for the study of relatively small signaling circuits, comprised of tens of proteins, not hundreds or thousands.

A reasonable compromise for many signaling systems is the use of logical models. These models lack the highly mechanistic insight generated from a kinetic model but do not require any rate parameters to be measured or estimated. As a result, logic models can generally be used to study much larger signaling networks, even in cases where connectivity is contextual or imperfectly understood, as is often the case in signaling biology. In a logic model the signaling pathway is reconstructed as a series of logical operators – “AND,” “OR,” “NOR” – that describe the relationship between the activities of protein intermediates. This approach happens to be similar to the way signaling biologists have been intuitively drawing wiring diagrams since the advent of signal transduction as a research topic, both conceptually and with respect to model's output. In the simplest forms, such as those that use Boolean “on/off” (0,1) rules, logic models also suffer from some of the same shortcomings as static signal wiring diagrams. For example, the influences between proteins are equal in magnitude and binary, which would typically not be the case in a real biological system. Furthermore, these models do not traditionally provide as output temporal dynamics of signaling. In more sophisticated approaches, both of these shortcomings are addressed, and logic modeling frameworks now exist that encode graded influences between pathway nodes and the time dependencies of signaling activity states [112,113]. While these methods are still nascent, we predict that they will serve as important tools in the signaling biologist's toolkit.

Data-Driven Modeling Approaches

Both kinetic and logic modeling methods are valuable and intuitive platforms, however, both methods require a reasonably mature understanding of the circuitry of a signaling process, the so-called “prior knowledge network” (PKN). Because our existing knowledge of network circuitry is incomplete, particularly in the area of signaling crosstalk, research in signal transduction can still benefit greatly from approaches that require no prior knowledge and instead are completely empirical and data-driven. Integration of empirical insights with prior knowledge can often form a strong foundation to infer function or regulatory relationships. Data-driven approaches generally produce more abstract insights than mechanistic modeling but are nevertheless well suited to enhance understanding about the basic topology (i.e., wiring) of a signaling network, the influences and connectivity between network components, and the likely routes of information flow through a network. These approaches share the common underlying assumption that the relationships between signaling proteins (i.e., their influences and interactions) manifest as statistically observable trends within the data [114]. For example, the activities of signaling proteins that control a cellular response like proliferation or apoptotic cell death would be predicted to covary with these responses when measured across many diverse conditions or system perturbations.

Many data-driven modeling methods exist and are becoming widely used in signaling research. For example, graphical network maps (protein-protein interaction networks, gene regulatory networks) have long been used to define the abstract structure of a system [115]. The edges that connect nodes in a graph can represent essentially any relationship and need not be mechanistic or physical in nature. Thus, networks can be a useful method for integrating disparate types of data [116,117], as is often necessary in signaling research. When considered in the context of ontology or known functional groups, these networks can be used to make predictions about function or co-regulation.

Clustering approaches (hierarchical, K means, K-nearest neighbors, etc.) are commonly used to identify subsets of the data that appear to share a relationship. Co-clustering might imply that the entities are themselves parts of a functional unit, like a multi-subunit complex, or that they are all regulated by the same upstream protein, or that they are all simply parts of a shared regulatory network. When these approaches are integrated with variables such as time and experimental perturbations, more mechanistic insight can often be inferred about why the clusters emerged and/or about the function of certain components within a cluster [118]. In this way, clustering can be a powerful hypothesis-generating tool, but as is true with most data-driven approaches, these hypotheses should be validated using other techniques, both computational and experimental [119].

Finally, several multivariate regression and dimensionality reduction methods exist, including principal component analysis, partial-least-squares regression (PLSR), stepwise regression, Least Absolute-Shrinkage-and-Selection-Operator (LASSO) and elastic-net regression, linear discriminant analysis, factor analysis, etc. These are becoming common tools in biological research because of their power for describing signaling networks and their ease of use [105]. These methods take advantage of the fact that many signaling nodes carry redundant information or are highly covariant in response to a given stimulus. Dimensionality reduction approaches like PCA can be used to simplify the data without

losing much information. These techniques dramatically improve our ability to visualize complex networks and obtain a more intuitive understanding of how signals flow through the network. Other approaches, like PLSR, which also uses eigenvector decomposition, offer the ability to explicitly identify relationships between a set of input signals and output cellular responses.

One important caveat in the use of data-driven modeling is the selection of data to be modeled. Data-driven approaches are unbiased (or at least, they are not biased by prior knowledge), thus they offer an opportunity to identify novel and even counterintuitive or paradoxical information, provided that those signals were chosen to be measured. This advantage is counterbalanced by the limitation that spurious relationships may be identified. This limitation is not applicable if the data generated are truly comprehensive, but such data are rare in signaling research, particularly for methods that rely on antibodies as described above. Thus, some thought or strategy should be applied to select targets for analysis that are both likely to be involved but not already known. One effective strategy is to combine multiple OMIC approaches, some that allow comprehensive analysis, to identify and prioritize certain nodes for study, and others that allow more detailed study of signaling activation states. For example, global transcriptomic analysis could be used to identify processes that are changed in response to stimuli, or activities or sub-networks that appear to be enriched in the transcriptional signature. Based on pathway informatics, the activity of some or all proteins involved in those pathways could be directly measured in response to stimuli over time and these data could be modeled to identify which signaling behaviors were most predictive of a cell fate of interest. This approach was used in our recent work, which studied mechanisms of drug synergy in triple-negative breast cancer [3]. Another point of caution with data-driven modeling approaches is that variance scaling is often used to make comparable disparate measurements. This can lead to artifactual “stretching” of invariant data (i.e., data with no changes across all conditions), such that the model will essentially be fitting or overemphasizing experimental noise.

Many excellent resources exist that discuss the mathematics that underlie each of these modeling formalisms and for describing how each can be used [63,106-108,120,121]. In the remainder of this review, we discuss considerations for applying these modeling methods on the types of OMIC data described and how their integration advances the study of signaling.

Considerations for Modeling Signaling Networks

One overarching goal of systems biology is to build predictive models of cell signaling networks. Achieving this goal requires special considerations with respect to what to measure and when or how to measure it, in addition to the basic issues described above for experimental design. In general, the quality of existing models of cell signaling networks is deficient. These models feature incomplete and uncertain topologies and large uncertainties in the parameter values [122]. Importantly, these models are often still useful, and in many cases have made reliable predictions that led to biological insight.

Parameter Estimation—As we stated above, most mathematical models of signaling are kinetic models in which signaling is a function of the concentrations of the proteins and their

intrinsic rates of production and decay. For most systems, the concentrations of most molecular components are measurable, but typically the kinetic parameter values are not and must be inferred from the data using model optimization methods (see Sun et al. [108] for a comprehensive review of optimization methods in systems biology). Any OMICs techniques applied for the purposes of calibrating or validating models of signal transduction pathways will have to follow the same principles.

Two related but distinct concepts in parameter estimation are parameter identifiability and parameter uncertainty. The former refers to the theoretical ability to learn a model parameter's true value given infinite data. Mathematically, an identifiable parameter will have finite confidence intervals [123]. Conversely, non-identifiability implies that the true values of one or more of a model's parameters can never be known. Two types of parameter non-identifiability exist: structural and practical. Structural non-identifiability refers to redundant off-setting parameters in the model whereas practical non-identifiability refers to the general lack of data given the number of parameters [124]. Structural non-identifiability can arise in several ways such as the model having excessive parameters, inadequate experimental design (choice of perturbations and sampling strategies), or the lack of measured observables. Some of these issues can be resolved through model reduction techniques, but the lack of measurements can only be resolved by collecting qualitatively different data, either by applying different perturbations or by measuring additional observables. Practical non-identifiability is alleviated by collecting data of sufficient quantity and quality. Specifically, these data could be qualitatively different, or be collected from additional or different sampling times, from additional replicates, or using more reliable assays with less variance. Assuming the parameters are identifiable, uncertainties in their estimated values will still exist. These uncertainties are typically expressed as confidence intervals. Parameter uncertainties lead to uncertainties in the predictive power of the model, although it is possible for a model's predictions to be well constrained even though the parameters are not [125]. As with the solutions to non-identifiability, parameter uncertainties can be reduced with data of higher quality and quantity.

From the perspective of parameter identification, OMIC technologies will likely play an important role in creating better models of signal transduction. OMICs can help address issues of structural non-identifiability by adding observables. OMICs can help to constrain model topologies by identifying protein-protein interactions in the system of interest. The role of OMICs in addressing practical non-identifiability and parameter uncertainties is less clear because these entities are a strong function of the experimental design and the quality rather than quantity of the data. In this regard, one solution may be to use many perturbations and sampling times, which limits the applicability of certain OMICs techniques (such as mass spectrometry) due to the prohibitive cost of their use in measuring hundreds or thousands of samples. In such cases, many of the antibody-based “medium-throughput” proteomic techniques described above may represent a good compromise between the numbers of perturbations and observables. Accordingly, multiplexed bead-based immunoassays, reverse-phase protein arrays, and micro-western arrays – each of which can be applied to hundreds of samples – have been used to supply data for computational models of signaling networks [3,21,107,112,126].

Choice of Perturbations, Observables, and Sampling Times—The power and efficiency of experimental designs for calibrating and validating models of cell signaling depend on the choices of what to measure and how to measure them, namely the experimental perturbations (and perturbation sources), the observables, and the sampling times [84]. Of these, particularly important are the perturbations used to activate or deactivate signaling pathways. Well-chosen perturbations can reveal spurious relationships, and ultimately allows the parameters associated with their logical and/or biochemical relationships to be estimated and refined. Furthermore, most topology-estimation methods leverage the correlations between the changes in node levels in response to perturbation. Accurate estimation of these correlations depends on the sufficiency and quality of the perturbations. The combinatorial application of perturbations is also critical for detecting interactions (i.e., synergism or antagonism), which can serve as evidence for connectivity [127]. We believe that most early systems biology efforts used insufficient numbers of perturbations and many did not apply perturbations in a combinatorial manner. Common perturbations include naturally occurring variation (e.g., different cell types); receptor ligands such as hormones or cytokines; genetic perturbations such as transient, stable, or inducible overexpression or knockdown of protein; and small molecules, which pharmacologically activate or inhibit protein activities. The concentrations or doses of these variables represent experimental design choices, which are often useful for contributing variation in the observables, thus enabling effective parameter estimation.

Observables (i.e., what is measured) are typically the levels of total proteins and/or their post-translationally modified forms (e.g., phosphorylated, acetylated, sumoylated, etc.), as well as levels of coding and non-coding RNAs that are regulated by signaling pathways. Less common observables include the levels of protein complexes (e.g., measured by immunoprecipitation) and localization of proteins (e.g., protein levels in nuclear extracts). Observables are usually chosen on the basis of logistics, such as the availability and/or cost of detection reagents, but should more commonly be based on the biology (i.e., how is the signal encoded). Sampling times are chosen according to the time scale on which the signaling system operates as well as its duration of action.

Optimizing Experimental Design for Modeling—All logistically feasible combinations of the above variables are possible, such that it is exceedingly difficult to choose the ones to use in a rational manner. In the past, these choices were made *ad hoc*, based on tradition (e.g., three replicates per treatment condition) or feasibility (e.g., availability of antibodies). Today, computational methods are available to optimize experimental designs devoted to generating data for modeling signal transduction networks [84]. Work on optimal experimental designs for signaling models is still in its infancy. Early work used sensitivity analysis to prioritize experiments for investigating the dynamics of TNF α /NF- κ B signaling [128]. Casey et al. demonstrated the use of optimal design for constraining the variability of predictions in a model of EGFR signaling [129], and more recently, Tidor and colleagues have demonstrated that optimal selection of experiments can lead to accurate parameter estimates in models of EGF and NGF signaling in PC12 cells [130,131]. It is currently unclear as to whether computational optimization of experimental

design is feasible for OMICs-sized models. In principle, these methods should be scalable but in practice may be limited by computational power.

While considerable important progress has been made, these methods of optimization have some important limitations. First, the methods are nontrivial to apply. The calculations must be customized to the experimental question, which precludes automation and requires specialized expertise [84]. Second, the topology of the model is typically assumed to be known [130], which is almost never the case in practice. Third, uncertainties in the input functions representing the perturbations must be accounted for because they can propagate into uncertainties in the parameter values [132]. These uncertainties come in many forms. For example, the degree to which a siRNA can knockdown protein levels varies widely, and this information is generally unknown and not included in the optimization. Also, it is well established that the potencies of receptor ligands such as hormones and cytokines vary over time but many signaling models do not incorporate this knowledge. In early models of transforming growth factor- β signaling, for example, the level of TGF- β was assumed to be constant [133], whereas it was later found that TGF- β depletion is a major factor in downstream signaling kinetics [134]. Finally, experimental design optimization algorithms assume that only random error is present in the data, despite bias being a major source of variance. Some biases, such as batch effects, can be eliminated by combining proper experimental design (blocking and randomization) with data normalization techniques.

Data Pre-processing, Normalization, and Statistical Analysis—All proteomic methods provide quantitative readouts that are assumed to be proportional to the abundance or activity of the signaling proteins of interest. In the case of mass spectrometry, the readout can be spectral counts or spectral intensities. For antibody-based proteomics, the readouts are typically fluorescence (e.g., the median fluorescence intensity of each bead type in a multiplexed bead-based immunoassay), chemiluminescence, or optical density (e.g., from plate-based colorimetric assays such as ELISA). To provide interpretable data, these raw readouts must be translated into quantities that reflect the absolute or relative abundances of total or modified proteins, which are then further analyzed to infer the biology of the system.

Translating raw data into interpretable quantities typically involves transformation, normalization, and scaling, while the downstream analyses involve statistical or mathematical modeling. Log transformation is a common means to render the data more interpretable and to stabilize the variance of the signal as a function of the signal magnitude. The data are then commonly normalized to facilitate fair comparisons across samples or batches. Normalization is used to remove unwanted systematic variance from the data, typically that which is associated with technical sources of variance.

Numerous normalization methods have been developed for OMIC data. Many of these are applicable to any type of high-throughput data, but some are idiosyncratic to a specific proteomic assay. Normalization methods can be classified into three basic types: 1) biological methods, in which one or more separate measurements are made in order to interpret the measurement of interest, such as comparing the signal measure to that of an invariant “housekeeping” protein, 2) unsupervised methods in which models are used to specify putative sources of variance that are not related to the variables describing the study,

and 3) supervised methods, which utilize study-specific variables including batch and treatment effects [135]. Different normalization techniques have been evaluated for mass spectrometry [136], multiplexed bead-based flow cytometric immunoassays [126], reverse-phase protein arrays [137,138], 2-D gel electrophoresis [139], antibody microarrays, immunoblotting [140], and flow cytometry [141,142]. Data normalization methods continue to be developed, such that it would be premature to declare that one approach is best, but we emphasize that normalization is a critical step in the analysis of OMIC data that should be carefully considered.

In addition to normalization, analysis of large signaling data sets also often requires scaling to compare disparate analytical measurements. The need for scaling is particularly critical for antibody-based proteomics, because the assay signal intensity will be a function of both the protein expression level *and* antibody-epitope binding affinity (i.e., how “good” is the antibody), and the latter must be accounted for in the analysis. One common approach is to unit variance scale the data [143], but this procedure can substantially alter the data. Specifically, the procedure maintains of the general “shape” of the data, but disregards its magnitude. Situations exist in which this approach is beneficial but for others it is counterproductive. Variance rescaling can lead to misinterpretations if the source data was of low variance to begin with. Proper quantification, transformation, normalization, and scaling are all required to gain reliable insight using statistical or data-driven models and the “right” ways to do each of these depends greatly on the nature of the data collected. Another more complex issue is the integration of very heterogeneous types of data [144]. It will likely be necessary to integrate genomic, proteomic, or clinical data of different types, some of which may be categorical, continuous or discontinuous, and have different sizes, formats, and parameters. Some efforts are underway to improve standardization and improve the ability to compare disparate data across different assay platforms [145]. This issue is complex and represents a contemporary challenge for systems biology to resolve. In the meantime, efforts involving the integration of OMIC datasets will benefit from carefully considering experimental design issues and planning devoted to merging experimental and computational approaches in order to maximize the potential for new insights.

IV. CONCLUSION

The Philosophy Of OMIC Experiments: Will They Get Us Closer To The Truth?

To some, OMIC-based science represents an important way forward in generating comprehensive and unbiased knowledge, ultimately promising to reveal predictive cause-and-effect relationships about nature that allow us to diagnose, manage, treat, and prevent disease [146]. To others, the OMIC revolution has been a disappointment. Sydney Brenner, for example, has famously called systems biology a “low-input, high-throughput, no-output” science [147]. In light of these conflicting views, and the roughly two decades of OMIC history on which to reflect, it is worth considering how far we have come and where we need to go moving forward.

As we stated in the introduction, there was indeed a nascent form of systems biology – a *version 1.0* – in which the apparent idea was to do science as it had always been done, but to simply do it on a larger scale: rather than make a knockout, make all knockouts, rank order

these and take the best one to study. This approach quickly morphed into a realization that these mountains of data could be “hypothesis generating” rather than “hypothesis driven.” It is perhaps a valid criticism that this approach produced relatively few successes [102]. There are good reasons for this. For one, OMIC technologies, in and of themselves, did not address many of the pitfalls of biomedical science, and in fact may have exacerbated some of these issues [102]. Also, early OMIC experiments were limited by cost feasibility, which occasionally resulted in “atypical” experimental designs missing some controls and/or replication. Furthermore, researchers had often fallen prey to the mistaken assumption that measuring more variables might overcome the limitations of these feasibility-limited experimental designs. Instead, OMIC experiments require just as much, if not more, care in experimental conception, design, and analysis than do traditional assays. As we suggested throughout this review, planning of experimental approaches and computational strategies should be integrated, paired with an understanding of the benefits and weaknesses of the approaches, and considered before any data is collected to ensure that the right types of data are collected in the right formats. An explosion of OMIC research has occurred in all fields of biological study, including in signal transduction. This explosion has been driven in part by successes in the OMIC study of genomics and transcriptomics. As we highlighted herein, unique challenges exist for applying OMIC approaches to understanding signal transduction. Importantly, however, common challenges exist that all OMIC researchers face in producing reliable, meaningful, and practical insights. We expect that many of these challenges will be resolved by virtue of an expanding new generation of scientists who are fluent in both experimental and computational matters, and indeed the recent additions to the systems literature show demonstrable progress in improving our understanding and treatment of human disease [148]. OMIC approaches or computational tools are not a replacement for carefully thought out questions or hypotheses, and well-designed experiments. The idea that one should “measure more for the sake of measuring more,” or “model more for the sake of modeling more” has not worked well in the past and likely will not work well in the future. In many (well lit) corners of biology, a frustration has been expressed with the relative lack of benefit brought by resource-intensive systems and OMIC approaches. A clear task exists for those of us in the *version 2.0* of systems biology: prove its value. We will be judged moving forward by our ability to make insightful predictions, particularly ones that are either counter-intuitive or could not have been made in the absence of OMIC or systems tools. To meet this mandate, we will have to be better at developing appropriate OMIC-scale questions or hypotheses. Each one of us must have more dexterity in the use and integration of various OMIC scale experimental approaches. We will have to be more fluent in the proper use of computational tools. And importantly, we will have to be more careful in marrying these experimental and computational approaches appropriately.

ACKNOWLEDGEMENTS

MJL and DCC would like to thank the personnel and leadership of the MIT Center for Cell Decision Processes (CDP) and the MIT Integrative Cancer Biology Program (ICBP) for funding and intellectual support. In particular, these groups provided a forum for critical discussion and open sharing of ideas that was instrumental in advancing the ideas presented in this review. Additionally, BDL is funded by a training grant from the NCI (T32 CA130807). DCC receives funding support from the Natural Sciences and Engineering Research Council of Canada. MJL receives funding support from the Richard and Susan Smith Family Foundation and from the Breast Cancer Alliance.

REFERENCES

1. Downward J. The ins and outs of signalling. *Nature*. 2001; 411:759–62. doi:10.1038/35081138. [PubMed: 11459043]
2. Duncan JS, Whittle MC, Nakamura K, Abell AN, Midland AA, Zawistowski JS, et al. Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. *Cell*. 2012; 149:307–21. doi:10.1016/j.cell.2012.02.053. [PubMed: 22500798]
3. Lee MJ, Ye AS, Gardino AK, Heijink AM, Sorger PK, Macbeath G, et al. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell*. 2012; 149:780–94. doi:10.1016/j.cell.2012.03.031. [PubMed: 22579283]
4. Yao Z, Petschnigg J, Ketteler R, Stagljar I. Application guide for omics approaches to cell signaling. *Nat Chem Biol*. 2015; 11:387–97. doi:10.1038/nchembio.1809. [PubMed: 25978996]
5. Lee MJ, Yaffe MB, Cantley LC, Hunter T, sever R, thorne J. Protein Regulation in Signal Transduction. *Signal Transduction Principles, Pathways, and Processes*. 2014:31–50. doi:10.1101/cshperspect.a005918.
6. Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, et al. A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol*. 2003; 21:921–6. doi:10.1038/nbt849. [PubMed: 12872131]
7. Huang PH, Mukasa A, Bonavia R, Flynn RA, Brewer ZE, Cavenee WK, et al. Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc Natl Acad Sci USA*. 2007; 104:12867–72. doi:10.1073/pnas.0705158104. [PubMed: 17646646]
8. Behrends C, Sowa ME, Gygi SP, Harper JW. Network organization of the human autophagy system. *Nature*. 2010; 466:68–76. doi:10.1038/nature09204. [PubMed: 20562859]
9. Bantscheff M, Eberhard D, Abraham Y, Bastuck S, Boesche M, Hobson S, et al. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol*. 2007; 25:1035–44. doi:10.1038/nbt1328. [PubMed: 17721511]
10. Lanning BR, Whitby LR, Dix MM, Douhan J, Gilbert AM, Hett EC, et al. A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors. *Nat Chem Biol*. 2014; 10:760–7. doi:10.1038/nchembio.1582. [PubMed: 25038787]
11. White FM. The potential cost of high-throughput proteomics. *Sci Signal*. 2011; 4:pe8–pe8. doi:10.1126/scisignal.2001813. [PubMed: 21325204]
12. Noble WS. Mass spectrometrists should search only for peptides they care about. *Nat Meth*. 2015; 12:605–8. doi:10.1038/nmeth.3450.
13. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, et al. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol*. 2011; 7:548–8. doi:10.1038/msb.2011.81. [PubMed: 22068331]
14. Ong S-E, Foster LJ, Mann M. Mass spectrometric-based approaches in quantitative proteomics. *Methods*. 2003; 29:124–30. [PubMed: 12606218]
15. Evans C, Noirel J, Ow SY, Salim M, Pereira-Medrano AG, Couto N, et al. An insight into iTRAQ: where do we stand now? *Anal Bioanal Chem*. 2012; 404:1011–27. doi:10.1007/s00216-012-5918-6. [PubMed: 22451173]
16. Ludwig C, Claassen M, Schmidt A, Aebersold R. Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. *Mol Cell Proteomics*. 2012; 11:M111.013987. doi:10.1074/mcp.M111.013987. [PubMed: 22101334]
17. Janes KA. An analysis of critical factors for quantitative immunoblotting. *Sci Signal*. 2015; 8:rs2–rs2. doi:10.1126/scisignal.2005966. [PubMed: 25852189]
18. Wilson B, Liotta LA, Petricoin E. Monitoring proteins and protein networks using reverse phase protein arrays. *Dis Markers*. 2010; 28:225–32. doi:10.3233/DMA-2010-0705. [PubMed: 20534907]
19. Sevecka M, Wolf-Yadlin A, Macbeath G. Lysate Microarrays Enable High-throughput, Quantitative Investigations of Cellular Signaling. *Mol Cell Proteomics*. 2011; 10:M110.005363–3. doi:10.1074/mcp.M110.005363. [PubMed: 21296872]

20. Wolf-Yadlin A, Sevecka M, Macbeath G. Dissecting protein function and signaling using protein microarrays. *Current Opinion in Chemical Biology*. 2009; 13:398–405. doi:10.1016/j.cbpa.2009.06.027. [PubMed: 19660979]
21. Ciaccio MF, Wagner JP, Chu C-P, Lauffenburger DA, Jones RB. Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nat Meth*. 2010; 7:148–55. doi:10.1038/nmeth.1418.
22. Hughes AJ, Spelke DP, Xu Z, Kang C-C, Schaffer DV, Herr AE. Single-cell western blotting. *Nature Publishing Group*. 2014; 11:749–55. doi:10.1038/nmeth.2992.
23. Tibes R, Qiu Y, Lu Y, Hennessy B, Andreeff M, Mills GB, et al. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Molecular Cancer Therapeutics*. 2006; 5:2512–21. doi: 10.1158/1535-7163.MCT-06-0334. [PubMed: 17041095]
24. Ng PKS, Werner HMJ, Shahmoradgoli M, Zhang F, Ju Z, Liu W, et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nature Communications*. 2014; 5:1–14. doi:10.1038/ncomms4887.
25. Federici G, Gao X, Slawek J, Arodz T, Shitaye A, Wulfkühle JD, et al. Systems analysis of the NCI-60 cancer cell lines by alignment of protein pathway activation modules with “-OMIC” data fields and therapeutic response signatures. *Mol Cancer Res*. 2013; 11:676–85. doi: 10.1158/1541-7786.MCR-12-0690. [PubMed: 23635402]
26. Albeck JG, Burke JM, Aldridge BB, Zhang M, Lauffenburger DA, Sorger PK. Quantitative Analysis of Pathways Controlling Extrinsic Apoptosis in Single Cells. *Mol Cell*. 2008; 30:11–25. doi:10.1016/j.molcel.2008.02.012. [PubMed: 18406323]
27. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*. 2009; 459:428–32. doi:10.1038/nature08012. [PubMed: 19363473]
28. Cotari JW, Voisinne G, Dar OE, Karabacak V, Altan-Bonnet G. Cell-to-cell variability analysis dissects the plasticity of signaling of common γ chain cytokines in T cells. *Sci Signal*. 2013; 6:ra17–7. doi:10.1126/scisignal.2003240. [PubMed: 23482665]
29. Perfetto SP, Chattopadhyay PK, Roederer M. Seventeen-colour flow cytometry: unravelling the immune system. *Nat Rev Immunol*. 2004; 4:648–55. doi:10.1038/nri1416. [PubMed: 15286731]
30. Bandura DR, Baranov VI, Ornatsky OI, Antonov A, Kinach R, Lou X, et al. Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. *Anal Chem*. 2009; 81:6813–22. doi:10.1021/ac901049w. [PubMed: 19601617]
31. Bodenmiller B, Zunder ER, Finck R, Chen TJ, Savig ES, Bruggner RV, et al. Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat Biotechnol*. 2012; 30:857–66. doi:10.1038/nbt.2317.
32. Bendall SC, Simonds EF, Qiu P, Amir E-AD, Krutzik PO, Finck R, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*. 2011; 332:687–96. doi:10.1126/science.1198704. [PubMed: 21551058]
33. Du J, Bernasconi P, Clauser KR, Mani DR, Finn SP, Beroukhi R, et al. Bead-based profiling of tyrosine kinase phosphorylation identifies SRC as a potential target for glioblastoma therapy. *Nat Biotechnol*. 2008; 27:77–83. doi:10.1038/nbt.1513. [PubMed: 19098899]
34. Peck D, Crawford ED, Ross KN, Stegmaier K, Golub TR, Lamb J. A method for high-throughput gene expression signature analysis. *Genome Biol*. 2006; 7:R61. doi:10.1186/gb-2006-7-7-r61. [PubMed: 16859521]
35. Bradbury A, Pluckthun A. Standardize antibodies used in research. *Nature*. 2015
36. Fields S, Song O. A novel genetic system to detect protein protein interactions. *Nature*. 1989; 340:245–6. doi:10.1038/340245a0. [PubMed: 2547163]
37. Fields S. High-throughput two-hybrid analysis. *FEBS Journal*. 2005 doi:10.1111/j.1742-4658.2005.04973.x.
38. Rolland T, Tasan M, Charlotiaux B, Pevzner SJ, Zhong Q, Sahni N, et al. A Proteome-Scale Map of the Human Interactome Network. *Cell*. 2014; 159:1212–26. doi:10.1016/j.cell.2014.10.050. [PubMed: 25416956]

39. Lam MHY, Stagljar I. Strategies for membrane interaction proteomics: No mass spectrometry required. *Proteomics*. 2012; 12:1519–26. doi:10.1002/pmic.201100471. [PubMed: 22610515]
40. Petschnigg J, Groisman B, Kotlyar M, Taipale M, Zheng Y, Kurat CF, et al. The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells. *Nat Meth*. 2014; 11:585–92. doi:10.1038/nmeth.2895.
41. Hu C-D, Chinenov Y, Kerppola TK. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell*. 2002; 9:789–98. [PubMed: 11983170]
42. Wehr MC, Laage R, Bolz U, Fischer TM, Grünwald S, Scheek S, et al. Monitoring regulated protein-protein interactions using split TEV. *Nature Publishing Group*. 2006; 3:985–93. doi: 10.1038/nmeth967.
43. Tarassov K, Messier V, Landry CR, Radinovic S, Serna Molina MM, Shames I, et al. An in vivo map of the yeast protein interactome. *Science*. 2008; 320:1465–70. doi:10.1126/science.1153878. [PubMed: 18467557]
44. Remy I, Ghaddar G, Michnick SW. Using the β -lactamase protein-fragment complementation assay to probe dynamic protein–protein interactions. *Nature Protocols*. 2007; 2:2302–6. doi: 10.1038/nprot.2007.356. [PubMed: 17853887]
45. Li Y-C, Rodewald LW, Hoppmann C, Wong ET, Lebreton S, Safar P, et al. A Versatile Platform to Analyze Low-Affinity and Transient Protein-Protein Interactions in Living Cells in Real Time. *CellReports*. 2014; 9:1946–58. doi:10.1016/j.celrep.2014.10.058.
46. Rossi F, Charlton CA, Blau HM. Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc Natl Acad Sci USA*. 1997; 94:8405–10. doi:10.1016/j.celrep.2014.10.058. [PubMed: 9237989]
47. Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, et al. High-throughput mapping of a dynamic signaling network in mammalian cells. *Science*. 2005; 307:1621–5. doi: 10.1126/science.1105776. [PubMed: 15761153]
48. Taipale M, Tucker G, Peng J, Krykbaeva I, Lin Z-Y, Larsen B, et al. A Quantitative Chaperone Interaction Network Reveals the Architecture of Cellular Protein Homeostasis Pathways. *Cell*. 2014; 158:434–48. doi:10.1016/j.cell.2014.05.039. [PubMed: 25036637]
49. Taipale M, Krykbaeva I, Whitesell L, Santagata S, Zhang J, Liu Q, et al. Chaperones as thermodynamic sensors of drug-target interactions reveal kinase inhibitor specificities in living cells. *Nat Biotechnol*. 2013; 31:630–7. doi:10.1038/nbt.2620. [PubMed: 23811600]
50. Zhu H, Snyder M. Protein chip technology. *Current Opinion in Chemical Biology*. 2003; 7:55–63. doi:10.1016/S1367-5931(02)00005-4. [PubMed: 12547427]
51. Fan B, Lu K-Y, Raymond Sutandy FX, Chen Y-W, Konan K, Zhu H, et al. A human proteome microarray identifies that the heterogeneous nuclear ribonucleoprotein K (hnRNP K) recognizes the 5' terminal sequence of the hepatitis C virus RNA. *Mol Cell Proteomics*. 2014; 13:84–92. doi: 10.1074/mcp.M113.031682. [PubMed: 24113282]
52. Huang J, Zhu H, Haggarty SJ, Spring DR, Hwang H, Jin F, et al. Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. *Proc Natl Acad Sci USA*. 2004; 101:16594–9. doi:10.1073/pnas.0407117101. [PubMed: 15539461]
53. Jones RB, Gordus A, Krall JA, Macbeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature*. 2006; 439:168–74. doi:10.1038/nature04177. [PubMed: 16273093]
54. Tay S, Hughey JJ, Lee TK, Lipniacki T, Quake SR, Covert MW. Single-cell NF- κ B dynamics reveal digital activation and analogue information processing. *Nature*. 2010; 466:267–71. doi: 10.1038/nature09145. [PubMed: 20581820]
55. Batchelor E, Mock CS, Bhan I, Loewer A, Lahav G. Recurrent Initiation: A Mechanism for Triggering p53 Pulses in Response to DNA Damage. *Mol Cell*. 2008; 30:277–89. doi:10.1016/j.molcel.2008.03.016. [PubMed: 18471974]
56. Purvis JE, Karhohs KW, Mock C, Batchelor E, Loewer A, Lahav G. p53 Dynamics Control Cell Fate. *Science*. 2012; 336:1440–4. doi:10.1126/science.1218351. [PubMed: 22700930]

57. Montero J, Sarosiek KA, DeAngelo JD, Maertens O, Ryan J, Ercan D, et al. Drug-Induced Death Signaling Strategy Rapidly Predicts Cancer Response to Chemotherapy. *Cell*. 2015; 160:977–89. doi:10.1016/j.cell.2015.01.042. [PubMed: 25723171]
58. Albeck JG, Burke JM, Spencer SL, Lauffenburger DA, Sorger PK. Modeling a snap-action, variable-delay switch controlling extrinsic cell death. *Plos Biol*. 2008; 6:e299.
59. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol*. 2006; 7:R100. doi:10.1186/gb-2006-7-10-r100. [PubMed: 17076895]
60. Rajaram S, Pavie B, Wu LF, Altschuler SJ. PhenoRipper: software for rapidly profiling microscopy images. *Nature Publishing Group*. 2012; 9:635–7. doi:10.1038/nmeth.2097.
61. millard B, Niepel M, Menden MP, menden, Muhlich JL, muhli, et al. Adaptive informatics for multifactorial and high-content biological data. *Nat Meth*. 2011:487–92.
62. Collins T. ImageJ for microscopy. *Biotech*. 2007; 43:S25–S30. doi:10.2144/000112517.
63. Janes KA, Albeck JG, Peng LX, Sorger PK, Lauffenburger DA, Yaffe MB. A high-throughput quantitative multiplex kinase assay for monitoring information flow in signaling networks: application to sepsis-apoptosis. *Mol Cell Proteomics*. 2003; 2:463–73. doi:10.1074/mcp.M300045-MCP200. [PubMed: 12832460]
64. Janes KA, Albeck JG, Gaudet S, Sorger PK, Lauffenburger DA, Yaffe MB. A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science*. 2005; 310:1646–53. doi:10.1126/science.1116598. [PubMed: 16339439]
65. Kubota K, Anjum R, Yu Y, Kunz RC, Andersen JN, Kraus M, et al. Sensitive multiplexed analysis of kinase activities and activity-based kinase identification. *Nat Biotechnol*. 2009; 27:933–40. doi: 10.1038/nbt.1566. [PubMed: 19801977]
66. Bose AK, Janes KA. A High-throughput Assay for Phosphoprotein-specific Phosphatase Activity in Cellular Extracts. *Molecular & Cellular Proteomics*. 2013; 12:797–806. doi:10.1074/mcp.O112.024059. [PubMed: 23233447]
67. Miller MA, Barkal L, Jeng K, Herrlich A, Moss M, Griffith LG, et al. Proteolytic Activity Matrix Analysis (PrAMA) for simultaneous determination of multiple protease activities. *Integr Biol*. 2011; 3:422–38. doi:10.1039/C0IB00083C.
68. Shults MD, Imperiali B. Versatile fluorescence probes of protein kinase activity. *J Am Chem Soc*. 2003; 125:14248–9. [PubMed: 14624552]
69. Han W-G, Liu T, Himo F, Touthkine A, Bashford D, Hahn KM, et al. A Theoretical Study of the UV/Visible Absorption and Emission Solvatochromic Properties of Solvent-Sensitive Dyes. *ChemPhysChem*. 2003; 4:1084–94. doi:10.1002/cphc.200300801. [PubMed: 14596006]
70. Goguen BN, Loving GS, Imperiali B. Development of a fluorogenic sensor for activated Cdc42. *Bioorganic & Medicinal Chemistry Letters*. 2011; 21:5058–61. doi:10.1016/j.bmcl.2011.04.051. [PubMed: 21549598]
71. Pertz O. Designing biosensors for Rho family proteins -- deciphering the dynamics of Rho family GTPase activation in living cells. *Journal of Cell Science*. 2004; 117:1313–8. doi:10.1242/jcs.01117. [PubMed: 15020671]
72. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005; 102:15545–50. doi:10.1073/pnas.0506580102. [PubMed: 16199517]
73. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera R, et al. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics*. 2006; 7:S7–15. doi:10.1186/1471-2105-7-S1-S7. [PubMed: 16723010]
74. Lefebvre C, Rajbhandari P, Alvarez MJ, Bandaru P, Lim WK, Sato M, et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol Syst Biol*. 2010; 6:1–10. doi:10.1038/msb.2010.31.
75. Wang K, Saito M, Bisikirska BC, Alvarez MJ, Lim WK, Rajbhandari P, et al. Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat Biotechnol*. 2009; 27:829–37. doi:10.1038/nbt.1563. [PubMed: 19741643]

76. Linding R, Jensen LJ, Ostheimer GJ, van Vugt MATM, Jørgensen C, Miron IM, et al. Systematic Discovery of In Vivo Phosphorylation Networks. *Cell*. 2007; 129:1415–26. doi:10.1016/j.cell.2007.05.052. [PubMed: 17570479]
77. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. 2014; 159:647–61. doi: 10.1016/j.cell.2014.09.029. [PubMed: 25307932]
78. Silva JM, Li MZ, Chang K, Ge W, Golding MC, Rickles RJ, et al. Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet*. 2005; 1–8. doi:10.1038/ng1650. [PubMed: 15624007]
79. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*. 2010; 11:733–9. doi:10.1038/nrg2825.
80. Collins FS, Tabak LA. NIH plans to enhance reproducibility. *Nature*. 2014
81. Chervitz SA, Deutsch EW, Field D, Parkinson H, Quackenbush J, Rocca-Serra P, et al. Data standards for Omics data: the basis of data sharing and reuse. *Methods Mol Biol*. 2011; 719:31–69. doi:10.1007/978-1-61779-027-0_2. [PubMed: 21370078]
82. Mann M. Comparative analysis to guide quality improvements in proteomics. *Nat Meth*. 2009; 6:717–9. doi:10.1038/nmeth1009-717.
83. Cairns DA. Statistical issues in quality control of proteomic analyses: good experimental design and planning. *Proteomics*. 2011; 11:1037–48. doi:10.1002/pmic.201000579. [PubMed: 21298792]
84. Kreutz C, Timmer J. Systems biology: experimental design. *Febs J*. 2009; 276:923–42. doi: 10.1111/j.1742-4658.2008.06843.x. [PubMed: 19215298]
85. Oberg AL, Vitek O. Statistical design of quantitative mass spectrometry-based proteomic experiments. *J Proteome Res*. 2009; 8:2144–56. doi:10.1021/pr8010099. [PubMed: 19222236]
86. Rocke DM. Design and analysis of experiments with high throughput biological assay data. *Seminars in Cell and Developmental Biology*. 2004; 15:703–13. doi:10.1016/j.semcdb.2004.09.007. [PubMed: 15561590]
87. Glass DJ. Experimental design for biologists. 2007
88. Naegle K, Gough NR, Yaffe MB. Criteria for biological reproducibility: what does “n” mean? *Sci Signal*. 2015; 8:fs7–fs7. doi:10.1126/scisignal.aab1125. [PubMed: 25852186]
89. Lázic SE. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? *BMC Neurosci*. 2010; 11:5. doi:10.1186/1471-2202-11-5. [PubMed: 20074371]
90. Aarts E, Verhage M, Veenvliet JV, Dolan CV, van der Sluis S. A solution to dependency: using multilevel analysis to accommodate nested data. *Nat Neurosci*. 2014; 17:491–6. doi:10.1038/nn.3648. [PubMed: 24671065]
91. Altman N, Krzywinski M. Points of significance: Sources of variation. *Nat Meth*. 2015; 12:5–6. doi:10.1038/nmeth.3224.
92. Ioannidis JPA, Hozo I, Djulbegovic B. Optimal type I and type II error pairs when the available sample size is fixed. *Journal of Clinical Epidemiology*. 2013; 66:903–910. e2. doi:10.1016/j.jclinepi.2013.03.002. [PubMed: 23664493]
93. Glass DJ. A critique of the hypothesis, and a defense of the question, as a framework for experimentation. *Clin Chem*. 2010; 56:1080–5. doi:10.1373/clinchem.2010.144477. [PubMed: 20511448]
94. Kell DB, Oliver SG. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays*. 2004; 26:99–105. doi: 10.1002/bies.10385. [PubMed: 14696046]
95. Platt JR. Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*. 1964; 146:347–53. doi:10.1126/science.146.3642.347. [PubMed: 17739513]
96. Beard DA, Kushmerick MJ. Strong inference for systems biology. *PLoS Comput Biol*. 2009; 5:e1000459. doi:10.1371/journal.pcbi.1000459. [PubMed: 19714210]
97. Glass DJ, Hall N. A Brief History of the Hypothesis. *Cell*. 2008; 134:378–81. doi:10.1016/j.cell.2008.07.033. [PubMed: 18692458]

98. Collins FS, Tabak LA. NIH plans to enhance reproducibility. *Nature*. 2014 doi:10.1111/j.1365-2796.2011.02498.x.
99. Ioannidis JP, Khoury MJ. Improving validation practices in “omics” research. *Science*. 2011 doi:10.1126/science.1214532.
100. Ioannidis JPA. Why Most Published Research Findings Are False. *Plos Med*. 2005; 2:e124–6. doi:10.1371/journal.pmed.0020124. [PubMed: 16060722]
101. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature*. 2012; 483:531–3. doi:10.1038/483531a. [PubMed: 22460880]
102. Ioannidis JPA. Expectations, validity, and reality in omics. *Journal of Clinical Epidemiology*. 2010; 63:945–9. doi:10.1016/j.jclinepi.2010.04.002. [PubMed: 20573481]
103. Chuang H-Y, Hofree M, Ideker T. A decade of systems biology. *Annu Rev Cell Dev Biol*. 2010; 26:721–44. doi:10.1146/annurev-cellbio-100109-104122. [PubMed: 20604711]
104. Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet*. 2001; 2:343–72. doi:10.1146/annurev.genom.2.1.343. [PubMed: 11701654]
105. Janes KA, Yaffe MB. Data-driven modelling of signal-transduction networks. 2006; 7:820–8. doi:10.1038/nrm2041.
106. Aldridge BB, Burke JM, Lauffenburger DA. Physicochemical modelling of cell signalling pathways. *Nature Cell*. 2006; 8:1195–203. doi:10.1038/ncb1497.
107. Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, et al. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol*. 2009; 5:1–19. doi:10.1038/msb.2009.87.
108. Sun J, Garibaldi JM, Hodgman C. Parameter Estimation Using Metaheuristics in Systems Biology: A comprehensive review. *Computational Biology and Bioinformatics*. 2012; 9:185–202. [PubMed: 21464505]
109. Santos SDM, Wollman R, Meyer T, Ferrell JE. Spatial positive feedback at the onset of mitosis. *Cell*. 2012; 149:1500–13. doi:10.1016/j.cell.2012.05.028. [PubMed: 22726437]
110. Litvin O, Schwartz S, Wan Z, Schild T, Rocco M, Oh NL, et al. Interferon α/β Enhances the Cytotoxic Response of MEK Inhibition in Melanoma. *Mol Cell*. 2015; 57:784–96. doi:10.1016/j.molcel.2014.12.030. [PubMed: 25684207]
111. Lee MJ. Let a Hundred Flowers Bloom: The Role of Context Dependence in Creating Phenotypic Diversity following Targeted Therapy. *Mol Cell*. 2015; 57:763–4. doi:10.1016/j.molcel.2015.02.030. [PubMed: 25747655]
112. Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA. Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. *PLoS Comput Biol*. 2011; 7:e1001099. doi:10.1371/journal.pcbi.1001099. [PubMed: 21408212]
113. MacNamara A, Terfve C, Henriques D, Bernabé BP, Saez-Rodriguez J. State–time spectrum of signal transduction logic models. *Phys Biol*. 2012; 9:045003–17. doi:10.1088/1478-3975/9/4/045003. [PubMed: 22871648]
114. Pe’er D, Hachohen N. Principles and strategies for developing network models in cancer. *Cell*. 2011; 144:864–73. doi:10.1016/j.cell.2011.03.001. [PubMed: 21414479]
115. Ideker T, Krogan NJ. Differential network biology. *Mol Syst Biol*. 2012; 8:1–9. doi:10.1038/msb.2011.99.
116. Vidal M, Cusick ME, Barabási A-L. Interactome Networks and Human Disease. *Cell*. 2011; 144:986–98. doi:10.1016/j.cell.2011.02.016. [PubMed: 21414488]
117. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L. The human disease network. *Proc Natl Acad Sci USA*. 2007; 104:8685–90. doi:10.1073/pnas.0701361104. [PubMed: 17502601]
118. Amit I, Citri A, Shay T, Lu Y, Katz M, Zhang F, et al. A module of negative feedback regulators defines growth factor signaling. *Nat Genet*. 2007; 39:503–12. doi:10.1038/ng1987. [PubMed: 17322878]
119. Naegle KM, Welsch RE, Yaffe MB, White FM, Lauffenburger DA. MCAM: Multiple Clustering Analysis Methodology for Deriving Hypotheses and Insights from High-Throughput Proteomic

Datasets. PLoS Comput Biol. 2011; 7:e1002119. doi:10.1371/journal.pcbi.1002119. [PubMed: 21799663]

120. Tentner AR, Lee MJ, Ostheimer GJ, Samson LD, Lauffenburger DA, Yaffe MB. Combined experimental and computational analysis of DNA damage signaling reveals context-dependent roles for Erk in apoptosis and G1/S arrest after genotoxic stress. *Mol Syst Biol.* 2012; 8:568. doi: 10.1038/msb.2012.1. [PubMed: 22294094]
121. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* 2012; 483:603–7. doi:10.1038/nature11003. [PubMed: 22460905]
122. Brown KS, Sethna JP. Statistical mechanical approaches to models with many poorly known parameters. *Physical Review E.* 2003 doi:10.1103/PhysRevE.68.021904.
123. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics.* 2009; 25:1923–9. doi:10.1093/bioinformatics/btp358. [PubMed: 19505944]
124. Raue A, Kreutz C, Maiwald T, Klingmüller U, Timmer J. Addressing parameter identifiability by model-based experimentation. *IET Systems Biology.* 2011; 5:120–30. doi:10.1049/iet-syb.2010.0061. [PubMed: 21405200]
125. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol.* 2007; 3:1871–8. doi: 10.1371/journal.pcbi.0030189. [PubMed: 17922568]
126. Clarke DC, Morris MK, Lauffenburger DA. Normalization and statistical analysis of multiplexed bead-based immunoassay data using mixed-effects modeling. *Mol Cell Proteomics.* 2013; 12:245–62. doi:10.1074/mcp.M112.018655. [PubMed: 23071098]
127. Natarajan M, Lin K-M, Hsueh RC, Sternweis PC, Ranganathan R. A global analysis of cross-talk in a mammalian cellular signalling network. *Nat Cell Biol.* 2006; 8:571–80. doi:10.1038/ncb1418. [PubMed: 16699502]
128. Cho K-H, Shin S-Y, Kolch W, Wolkenhauer O. Experimental Design in Systems Biology, Based on Parameter Sensitivity Analysis Using a Monte Carlo Method: A Case Study for the TNF α -Mediated NF- κ B Signal Transduction Pathway. *Simulation.* 2003; 79:726–39.
129. Gutenkunst RN, Myers CR, Casey FP, Baird D, Brown KS, Waterfall JJ, et al. Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model. *IET Systems Biology.* 2007; 1:190–202. doi:10.1049/iet-syb:20060065. [PubMed: 17591178]
130. Hagen DR, White JK, Tidor B. Convergence in parameters and predictions using computational experimental design. *Interface Focus.* 2013; 3:20130008. doi:10.1098/rsfs.2013.0008. [PubMed: 24511374]
131. Apgar JF, Witmer DK, White FM, Tidor B. Sloppy models, parameter uncertainty, and the role of experimental design. *Mol Biosyst.* 2010; 6:1890–900. doi:10.1039/b918098b. [PubMed: 20556289]
132. Schelker M, Raue A, Timmer J, Kreutz C. Comprehensive estimation of input signals and dynamics in biochemical reaction networks. *Bioinformatics.* 2012; 28:i529–34. doi:10.1093/bioinformatics/bts393. [PubMed: 22962477]
133. Vilar JMG, Jansen R, Sander C. Signal processing in the TGF-beta superfamily ligand-receptor network. *PLoS Comput Biol.* 2006; 2:e3. doi:10.1371/journal.pcbi.0020003. [PubMed: 16446785]
134. Clarke DC, Brown ML, Erickson RA, Shi Y, Liu X. Transforming growth factor beta depletion is the primary determinant of Smad signaling kinetics. *Molecular and Cellular Biology.* 2009; 29:2443–55. doi:10.1128/MCB.01443-08. [PubMed: 19223462]
135. Mecham BH, Nelson PS, Storey JD. Supervised normalization of microarrays. *Bioinformatics.* 2010; 26:1308–15. doi:10.1093/bioinformatics/btq118. [PubMed: 20363728]
136. Callister SJ, Barry RC, Adkins JN, Johnson ET, Qian W-J, Webb-Robertson B-JM, et al. Normalization approaches for removing systematic biases associated with mass spectrometry and

- label-free proteomics. *J Proteome Res.* 2006; 5:277–86. doi:10.1021/pr050300l. [PubMed: 16457593]
137. Liu W, Ju Z, Lu Y, Mills GB, Akbani R. A Comprehensive Comparison of Normalization Methods for Loading Control and Variance Stabilization of Reverse-Phase Protein Array Data. *Cin* 2014. doi:10.4137/CIN.S13329.
138. Neeley ES, Kornblau SM, Coombes KR, Baggerly KA. Variable slope normalization of reverse phase protein arrays. *Bioinformatics.* 2009; 25:1384–9. doi:10.1093/bioinformatics/btp174. [PubMed: 19336447]
139. Artigaud S, Gauthier O, Pichereau V. Identifying differentially expressed proteins in two-dimensional electrophoresis experiments: inputs from transcriptomics statistical tools. *Bioinformatics.* 2013; 29:2729–34. doi:10.1093/bioinformatics/btt464. [PubMed: 23986565]
140. Kreutz C, Bartolome Rodriguez MM, Maiwald T, Seidl M, Blum HE, Mohr L, et al. An error model for protein quantification. *Bioinformatics.* 2007; 23:2747–53. doi:10.1093/bioinformatics/btm397. [PubMed: 17768165]
141. Hahne F, Khodabakhshi AH, Bashashati A, Wong C-J, Gascoyne RD, Weng AP, et al. Per-channel basis normalization methods for flow cytometry data. *Cytometry A.* 2010; 77:121–31. doi:10.1002/cyto.a.20823. [PubMed: 19899135]
142. O'Neill K, Aghaeepour N, Spidlen J, Brinkman R. Flow cytometry bioinformatics. *PLoS Comput Biol.* 2013; 9:e1003365. doi:10.1371/journal.pcbi.1003365. [PubMed: 24363631]
143. Gaudet S, Janes KA, Albeck JG, Pace EA, Lauffenburger DA, Sorger PK. A compendium of signals and responses triggered by prodeath and prosurvival cytokines. *Mol Cell Proteomics.* 2005; 4:1569–90. doi:10.1074/mcp.M500158-MCP200. [PubMed: 16030008]
144. Hamid JS, Hu P, Roslin NM, Ling V, Greenwood CMT, Beyene J. Data Integration in Genetics and Genomics: Methods and Challenges. *Human Genomics and Proteomics.* 2009; 2009:1–13. doi:10.4061/2009/869093.
145. Braun P, Tasan M, Dreze M, Barrios-Rodiles M, Lemmens I, Yu H, et al. An experimentally derived confidence score for binary protein-protein interactions. *Nature Publishing Group.* 2008; 6:91–7. doi:10.1038/nmeth.1281.
146. Tian Q, Price ND, Hood L. Systems cancer medicine: towards realization of predictive, preventive, personalized and participatory (P4) medicine. *J Intern Med.* 2012; 271:111–21. doi: 10.1111/j.1365-2796.2011.02498.x. [PubMed: 22142401]
147. Cornish-Bowden A. Systems biology. How far has it come. *Biochemist.* 2011
148. Huber KVM, Salah E, Radic B, Gridling M, Elkins JM, Stukalov A, et al. Stereospecific targeting of MTH1 by (S)-crizotinib as an anticancer strategy. *Nature.* 2014; 508:222–7. doi:10.1038/nature13194. [PubMed: 24695225]

Figure 1. Common Mechanisms of Signal Transduction

A main complication in the study of signal transduction is the variety of ways that signals can be encoded, which include changes to the expression level, subcellular localization, conformation, enzymatic activity, or interactions between proteins. These also are often coupled changes. Conformational change of G proteins alters their binding; phosphorylation of receptor tyrosine kinases may alter conformation, activity, localization and binding interactions.

Figure 2. Information Overload

Most signaling systems now contain so many bona fide nodes and edges that an intuitive understanding of function has been lost. Shown is an example of a well studied signaling network, the cellular DNA damage response (DDR). High confidence nodes and edges were manually curated from published literature. NOTE: This is a dramatically simplified version of current “understanding.”

Figure 3. Methods of modeling signaling data

Common approaches to modeling signal transduction vary with respect to the specificity of insight generated and the level of prior knowledge required. Accordingly, modeling methods should be selected with specific questions in mind and experimental parameters should be considered that are well suited for the modeling and feasible to collect using the experimental platform. In general, all forms of modeling require specific types of information, and more is not better if the data is not of the right type. The schematic shown relates four commonly used modeling methods with regards to their requires for prior knowledge and the specificity/abstraction of the model's output. Each of these are discussed in detail. Kinetic modeling is the most precise, requires the most prior knowledge, and gives insight into the dynamic behaviors of a signaling pathway. Logic models are less precise, require less knowledge, and provide insight into the wiring of a pathway. Network inference and regression based models are the least precise and require essentially no prior knowledge. These provide understanding about the relationships among the measured variable, including shared regulation and input-output structure.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Experimental Methodology	Utility	Strengths of Method	Weaknesses of Method	Literature example	Experimental Question	Modeling Used	Major Result
Mass Spectrometry	E=protein expression level	D= easy to adapt to measure signaling dynamics	ND= not well suited for measuring signaling dynamics	Naegle et al. <i>Mol BioSyst</i> 2012	EGFR signaling network	Multiple Clustering Analysis Methodology (MCAM)	co-regulation of pY used to reveal interaction between EGFR and PDLIM1
	A=protein activity	SC= single cell resolution	Bulk= cannot measure single cell data				
	L=protein localization	C= comprehensive	NC= not comprehensive				
	I=interaction	HT= high throughput	LT= low throughput (slow, or cannot handle many samples)				
		\$= affordable for most labs	\$\$\$= expensive				
MIB-MS and ABPP		V= easy to validate	DV= difficult to validate	Kumar et al. <i>PLoS Comp Biol</i> /2007	Mechanisms of HER2 dependent migration/proliferation	PLSR	Identification of 9 signaling biomarkers that are necessary to predict HER2 mediated phenotypes
		Q= quantitative	NQ=not quantitative				
High-Throughput Immunoblotting				Matsuoka et al. <i>Science</i> 2007	Comprehensive analysis of the DNA Damage Response	network modeling	IP-MS used to identify targets of DDR kinases ATM and ATR
A specialized use of Mass Spec for reagent				Duncan et al. <i>Cell</i> 2012	Kinome dynamics following targeted therapy		extensive reprogramming of kinome following targeted therapy drives resistance
		C (kinase specific), Q	ND, Bulk, NC (kinase specific), LT, \$\$\$				
				Bantscheff et al. <i>Nat Biotechnol</i> 2007	True targets of kinase ABL kinase inhibitors		DDR1 and NQO2 identified as novel targets of imatinib
				Stuhlmiller et al. <i>Cell Rep</i> 2015	Kinome dynamics following lapatinib exposure	PCA	kinome reprogramming following HER2 inhibition mediated by BET family proteins
				Lanning et al. <i>Nat chem Biol</i>	Irreversible kinase inhibitor specificity		significant off-target interactions with non-kinase proteins
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
		D, HT, \$, V (band), Q	Bulk, NC, LT (lower throughput than other methods)				
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	Context-dependent roles for ERK in mediating survival or death
				Janes et al. <i>Cell</i> 2008	Information encoding in signaling	PLSR, breakpoint analysis	cells respond to changes in dynamic range, not signal magnitude
				Tentner et al. <i>Mol Syst Bio</i> . 2012	DDR signaling network architecture	PLSR, TI-SWR	

Experimental Methodology	Utility	Strengths of Method	Weaknesses of Method	Literature example	Experimental Question	Modeling Used	Major Result
Protein Lysate Microarrays (RPPA) and Protein Chips	E, A, L, I	D, SC (possible), HT, \$, Q	Bulk (typically), NC, DV	Ciaccio et al. <i>Nat Methods</i> 2010	EGF signaling network	Bayesian network inference; mutual information	micro-western arrays facilitate HT analysis of signaling networks
				Lee et al. <i>Cell</i> 2012	Combination therapy design	PCA, PLSR	Drug induced rewiring of signaling responses can potentiate chemosensitivity
				Lu et al. <i>Oncogene</i> 2011	Functional analysis of AKT signaling	network modeling	Novel feedback loops identified, including bi-directional interaction between Akt and GSK3
				Federici et al. <i>Mol Cancer Res</i> 2013 Huang et al. <i>PNAS</i> 2004	Proteomic analysis of NCI-60 mTOR network	clustering, network modeling	Tep Ip, Nir Ip, identified as components of TOR signaling
Flow and Mass Cytometry	E, A, I	D, SC, HT, Q	NC, \$\$\$	Bendall et al <i>Science</i> 2011	Hematopoietic stem cell development	SPADE (multivariate analysis)	development of CyTOF
				Bodenmiller et al. <i>Nat Biotech</i> 2012 Bandura et al. Anal. Chem 2009	Systems pharmacology development of CyTOF	SPADE, PCA	systems pharmacology proof of concept development of CyTOF
Bead Based Cytometry	E, A,	D, HT, Q	Bulk, NC, \$\$\$, DV	Lau et al. <i>Sci Signal</i> 2011	in vivo systems level analysis of responses to TNF using in vivo models of CRC	PLS-DA	
				Morris et al. <i>PLoS Comp Bio</i> 2011	Signaling downstream of TGFa and TNFa	Constrained Fuzzy Logic (cFL)	Development of cFL
				Saez-Rodriguez et al. <i>Cancer Res.</i> 2011	Network state in normal vs. transformed hepatocytes	Logic	Critical differences in pathway architecture between HCC and hepatocytes
Lumier	I	HT, \$, V, Q	ND, Bulk, NC,	Du et al. <i>Nat Biotechnol</i> 2008	Comprehensive RTK profiling; method development		SRC as a potential target in glioblastoma
				Barrios-Rodiles et al. <i>Science</i> 2005	TGF beta interactome	network analysis, clustering	ID of many new TGF beta interaction proteins
				Taipale et al. <i>Cell</i> 2014	Hsp90 interactome	network analysis	comprehensive study of Hsp90 cofactors
Yeast-2-Hybrid and Related Technologies	I	C, HT, \$, V,	ND, NQ	Taipale et al. <i>Nat Biotechnol</i> 2013	Kinase inhibitor specificity		Development of LUMIER with BACON; ID of several drug off targets
				Rolland et al. <i>Cell</i> 2014	Comprehensive PPI map	network analysis, clustering	known PPIs are heavily biased towards heavily studied genes

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Experimental Methodology	Utility	Strengths of Method	Weaknesses of Method	Literature example	Experimental Question	Modeling Used	Major Result
In vitro enzyme assay	A	D, \$, V, Q	Bulk, NC, LT,	Petschnigg et al. <i>Nat Meth</i> 2014	MaMYTH development	network analysis	identification of previously unknown EGFR interacting proteins
				Sun et a. <i>Cell</i> 2012	Modulators of HMEC transformation		BiFC used to confirm interaction between PTPN12 and EGFR
				Janes et al. <i>Science</i> 2005	Context-dependence in TNF mediated apoptosis	PLSR	autocrine feedback circuits connect stress signaling and survival signaling to augment cell death response
				Miller et al. <i>Integr Biol</i> 2011	MMP Activity in complex environments	PrAMA	Modeling of FRET-based measurement of MMP activity used to discern specific MMP activities from mixture
Quantitative Image analysis	E, A, L	D, SC, \$, Q	NC, LT,	Kubota et al. <i>Nat Biotechnol</i> 2009	inference of kinase activity by mass spec	KAYAK, clustering	Extension of KAYAK to multiplexed, pooled measurements
				Millard et al. <i>Nat Methods</i> 2011	Development of ImageRail		ImageRail software
				Batchelor et al. <i>Mol Cell</i> 2008; Purvis et al. <i>Science</i> 2012	p53 control of cell fate following DNA Damage	differential equations	pulsatile vs. sustained p53 signaling controls repair vs senescence
				Singh et al. <i>Mol Sys Biol</i> 2010	Heterogeneity in drug sensitivity	various multivariate analyses (MDS, PCA)	basal heterogeneity in signaling states predicts drug sensitivities
Signaling Network Inference	A	C, HT, Q	ND, Bulk (typically), \$\$\$, DV,	Lee et al. <i>Mol Cell</i> 1014	Heterogeneity in NF-kappaB activation following TNF exposure	differential equations	Fold change in nuclear NF-kappaB dictates transcriptional output
				Subramanian et al. <i>PNAS</i> 2005	Development of GSEA	GSEA	Development of GSEA
				Linding et al. <i>Cell</i> 2007	Development of NetworKIN heterogeneity in response to MEK inhibition	NetworKIN	Systematic analysis of kinase-substrate relationships
				Litvin et al. <i>Mol Cell</i> 2015		COSPER, clustering	Context dependence in signaling
Functional Genomics	A (inferred by LOF/GOF)	C, HT, \$ (some methods), SC (some methods)	ND, Bulk (some methods), \$\$ (some methods), DV, NQ	Pritchard et al. <i>Cancer Res</i> 2011	Chemosensitivity prediction	Clustering, PCA	microenvironment-specific interactions modulate chemotherapeutic response
				Wang et al. <i>Science</i> 2014; Shalem et al. <i>Science</i> 2014; Gilbert et al. <i>Cell</i> 2014	Genome scale analysis of LOF and GOF mechanisms of resistance to BRAF inhibition		use of CRISPR/Cas9 for diverse genome scale inquiries
				Sun et al. <i>Nature</i> 2014			SOX10 dependent adaptation controls adaptive resistance