High throughput protein-based technologies and computational models for drug development, efficacy and toxicity

Leonidas G. Alexopoulos[1,2], Julio Saez Rodriguez[1,2], Christopher W. Espelin[3]

[1]Dept of Systems Biology, Harvard Medical School, Boston, MA, USA
[2]Dept of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[3]Pfizer Research Technology Center, Cambridge, MA, USA

‡Please address correspondence to:

Leonidas Alexopoulos, PhD
Department of Mechanical Engineering
National Technical University of Athens
Heroon Polytechniou 9
15780 Zografou, Athens, Greece
e-mail: leo@mail.ntua.gr

1. INTRODUCTION

Major advances in high-throughput assay capability coupled with increasingly sophisticated computational methods for systematic data analysis have provided scientists with tools to better understand the complexity of biological systems. This potent combination of novel experimental and analytical approaches should in turn lead to more effective therapeutic design. Most high-throughput experimental technologies can generally be categorized based on their readouts: genomics, transcriptomics, metabolomics and proteomics. The increasingly interdisciplinary nature of science has also given rise to complementary disciplines such as bioinformatics, systems biology, and computational biology, which are charged with incorporating and interpreting the vast amounts of experimental data and generating hypotheses of biological significance. That said, despite the fact that all these high-throughput assays are based on different readouts (DNA, mRNA, metabolites, or proteins), the biological information that they capture is highly overlapping. However, since each experimental assay represents a different approach, scientists should be able to understand the limitations and advantages of DNA, mRNA, metabolomic or proteomic approaches and use them appropriately depending on the knowledge that they want to obtain. Especially critical for the pharmaceutical industry is the coupling of experimental approaches with computational algorithms. These decisions can significantly impact discovery, reduce research and development costs, minimize drug failures by predicting drug efficacy and toxicity, and ultimately impact a company's competitiveness in the global market.

The primary objective of a therapeutic strategy is to selectively alter targeted protein(s) or pathway(s) within diseased cells in order to ameliorate an undesired phenotype (unrestrained cell proliferation, inflammatory cytokine release, etc.). Ideally, other pathways within the diseased cells, as well as all cellular functions in healthy cells, would remain unaffected by the therapeutic approach. Thus, target selection is a multi-faceted problem with several levels of complexity: 1) cellular pathway(s) or extracellular targets need to be selected 2) associated bio-markers need to be identified which distinguish healthy and disease phenotypes and ideally are capable of reflecting pharmacological efficacy and safety, 3) the risk-benefit of different therapeutic approaches should be considered: small molecule inhibitors, biologics, siRNA-derived therapeutics (transcriptional targets), etc., and 4) a selection criteria needs to be identified for screening compounds against a desired target. Addressing these initial questions, however, is just the beginning of the drug discovery journey; drug metabolism might render the pharmacologic approach ineffective, genetic and epigenetic person-to-person variability might cause *idiosyncratic toxicity* (i.e. a term that signifies the unknown cause of the adverse drug effects), and the overlapping cellular pathway architecture might result in even very selective on-

target compounds having unwanted effects on the same or different cells [1, 2].

In this complicated world of drug development, high-throughput technologies (genomics, transcriptomics, metabolomics and proteomics) have become invaluable tools for tackling the complexity of the biological systems and optimizing therapeutic strategies. In the following section, we will explore different technologies and their niches in the area of drug development, efficacy, and toxicity.

Genomics technologies:

Several diseases are strongly correlated with specific genomic mutations (Huntington's disease, sickle cell anemia, etc). As a general concept, cancer progression can be facilitated by gain-of-function mutations in oncogenes or loss-of-function mutations in tumor-supressor genes [3]. In certain instances, dozens of genomic aberrations are associated with the development of a single cancer type [4]. Specifically, in hepatocellular carcinoma (the most common type of liver cancer) genomic aberrations have been found in several proteins including TP53, TGFb, Ras, and Rb, EGFR, ERbB2 and members of the Wnt-signaling pathway [5-9]. Thus, understanding the genomic basis for development of a disease can significantly impact the focus of pharmaceutical therapies. It is now evident, that genomics have moved therapeutic strategies away from being phenotype specific (i.e. chemotherapy for highly proliferative cells), and focused them on pathway specific targets (i.e. anti-EGFR or anti-VEGF treatments etc).

Specific examples of the role that person-to-person genetic variability plays in drug efficacy have also been demonstrated in recent years. The epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor Gefitinib represents just a case. Gefitinib originally exhibited significant clinical response in only 10% of patients with non-small-cell lung cancer, a modest response rate [10]. However, when primary tumors from a group of patients was screened for EGFR gene mutations, eight of nine gefitinib-responsive patients had somatic mutations in the EGFR gene [11]. This example of genomic information coupled with drug mechanism underscores the importance of understanding underlying genetic anomalies in order to better predict whether a therapy will be effective. Individual genetic variability of more general biological mechanisms involved on drug metabolism, distribution, and clearance can also influence drug efficacy. For example, polymorphisms in the P450 drug-metabolizing enzyme among populations can generate significantly different profiles in drug metabolism [12, 13].

In order to better understand and predict an individual's response to a drug therapy, genomics approaches provide valuable information regarding:

- identification of critical genomic aberrations which strongly correlate with a specific disease, and in turn, implicate associated molecular pathways which may be appropriate for therapeutic intervention (i.e. targeted pathway(s) identification).
- identification of genetic variations among the population which indicate the likely success rate of a therapeutic intervention (i.e. treatment identification).
- identification of genetic variations in the biological machinery required for drug metabolism, distribution, and clearance (impact on efficacy and toxicity).

The main approaches utilized for genomic discovery are *whole genome sequencing* (i.e. Sanger sequencing, 454 Life Sciences) which entail sequencing the entire genome without bias towards particular gene sequences, and *genotype analysis* which is focused on specific genes or genomic locations. Further, genotype analysis may evaluate the entire raw sequence of particular genes or, alternatively, identification of Single Nucleotide Polymorphisms (SNPs) located at discrete genomic locations on a population level. New high-density DNA microarrays and bead-coupled universal Tag arrays have further enabled SNP analysis to move towards a whole-genome approach [14]. For further details regarding the application of genomic technologies to drug development, we direct the reader to several reviews [4, 15-18].

Transcriptomic technologies:

Transcriptome analysis is based on the detection and quantification of mRNA transcripts. The main advantage transcript profiling provides, relative to genomic approaches, is the ability to capture the *dynamic* response of a biological system under different conditions by measuring the expression profile of thousands of mRNA transcripts. There are approximately 25,000 human genes, yet only a small percentage (~20%) is active at any given time. Transcriptome technology provides the researcher with a perspective closer to the functional

response of the biological system. The technological platforms for transcript profiling are very similar to those used for genomic analysis and similarly, can be subdivided into two categories: approaches that evaluate all mRNA without prior bias regarding the sequence (i.e. serial analysis of gene expression (SAGE)), and analysis that is focused on a predefined set of genes (i.e. DNA microarrays). Even thought SAGE and DNA microarrays are based on different technologies, they have been demonstrated to correlate well in terms of mRNA quantitation, especially in the case of highly expressed transcripts [19]. Current technologies allow high density DNA microarrays to contain ~45,000 unique transcripts and genes that span the entire genome, with several commercial DNA microarrays now available at low to moderate cost (Affymetrix, Agilent, GE Healthcare, Applied Biosystems etc). Results from high throughput transcript profiling approaches are frequently confirmed using lower throughput assays such as qRT-PCR or Northern blotting.

Similar to genomic approaches, transcriptomic technologies have facilitated the identification of putative targets and pathways, as well as provide predictions regarding drug efficacy and toxicity:

- transcript analysis can identify mRNA transcripts that are strongly correlated with a disease (and indicate associated molecular pathways suitable for therapeutic intervention).
- identification of on-target and off-target drug effects throughout the whole genome which may influence drug efficacy and toxicity
- transcriptomic profiling of the biological mechanisms engaged in drug metabolism, distribution, and clearance

For further information regarding current DNA microarray products and applications, the reader may refer to an excellent recent review on "Toxicogenomic Technologies to Predictive Toxicilogy" by the National Research Council of the National Academies [20]. In addition to genomic and transcriptomic profiling, several related areas of study show promise in furthering the understanding of drug efficacy but are beyond the scope of this chapter. These include technologies involved in detection of DNA methylation [21], microRNA technologies, and CpG microarrays [22].

Metabolomic technologies:

Metabolomics refers to the study of metabolites (i.e. products or small molecule intermediates of the biological processes) as they exist in the cell, tissue, organ, or animal as a whole. Biological samples including cellular supernatants, blood, plasma, saliva, urine, and stool provide the source material which is typically analyzed using nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry (MS). These technologies are capable of evaluating a wide range of metabolic components (for example disease-specific metabolome changes). Metabolites have also been used as biomarkers in order to quantify the toxic effects of drugs. For a complete review of metabolomic applications, the reader can refer to metabolomic focused reviews [23-25]

Proteomic and Protein activity technologies:

Proteins are the ultimate executors of cellular function, and thus are directly responsible for a biological phenotype [26-30]. Proteomics is the study of the expression, modification and activity of proteins in order to better understand a biological system. In comparison to genomic or transcriptome approaches, proteomic profiling involves several unique challenges: 1) screening of the entire proteome in a manner similar to whole genome or transcriptome is currently impossible/impractical 2) the proteome cannot be defined using a constrained list of proteins (similar to genome approaches) because of the wide range of post-translational protein modifications which in turn can produce infinite combinations that are ultimately responsible for the phenotype 3) protein modifications such as phosphorylation, ubiquitination, methylation, sulfation or proteolytic cleavage (which are the main modulators of protein activity) cannot simultaneously be measured from a single assay [31] 4) protein abundance from a single cell population can span more than six orders of magnitude, creating a bias towards high-abundance proteins, 5) as compared to DNA-based experiments, protein experiments lack a "protein-amplification" feature (similar to PCR for genes) that makes protein amount valuable, 6) the ability to broadly analyze protein binding affinity (the driving force in signal transduction) is not possible on a high-throughput scale.

As with the previously described genomic and transcriptomic techniques, protein analysis can be divided into two separate approaches: one that makes no *a priori* assumption about protein composition or

structure (i.e. 2D-PAGE and mass spectrometry (MS) technology), and another which is based on a predetermined set of target proteins (i.e. affinity-based approaches). Using a 2D-PAGE approach, control lysates are separated based on their physical properties (mass and charge). Diseased (or drug treated lysates) are identically processed and comparative analysis performed to evaluate protein expression between the two samples. The MS approach, which often complements the 2D-PAGE approach, entails protein digestion to generate a mixture of peptides which are then separated and analyzed using liquid-chromatography-coupled mass spectrometry (LC-MS). Depending on specific physical properties, each peptide generates a unique signature (i.e. MS-MS spectra) that can be identified by a semi-automated search against proteomic databases. This approach, known as shotgun proteomic analyses, enables the automated characterization of hundreds of proteins in a complex mixture. Both 2D-PAGE and MS approaches have the disadvantage of being biased towards high abundance proteins and can only process a very low number of samples at a time.

Affinity-based assays utilize capture entities (i.e. antibody, peptide, nucleotides, etc.) to preferentially bind target proteins of interest. In contrast to the 2D-PAGE and MS approaches, affinity based assays identify proteins that can be recognized and captured by high-affinity molecules. Thus, a well developed affinity assay provides a high degree of specificity. A critical feature of affinity-based approaches is the ability to quantitate both protein concentration and protein activity, information which is essential for computational models. Antibody-based detection approaches are among the most commonly used for high throughput protein analysis and therefore will be more thoroughly described later in this chapter.

High-throughput measurements of protein activity can capture the dynamic changes in intracellular signals, thus generating large amount of data that encompass high-quality protein-level information. This data carries valuable information regarding propagation of signals, a process known to be non-linear and highly dynamic. Time-dependent and non-linear processes have been well-studied in many fields of engineering, a discipline known as dynamical systems. It is anticipated that similar methods can be utilized to extract valuable information from protein activity measurements to describe the biological phenotype. Thus, the combination of high-throughput data under different conditions (perturbations) and mathematical modeling (see Figure 1) supported by bioinformatics tools is a paradigm of the way in which a systems biology approach can have an impact on drug development, efficacy and toxicity [32, 33].

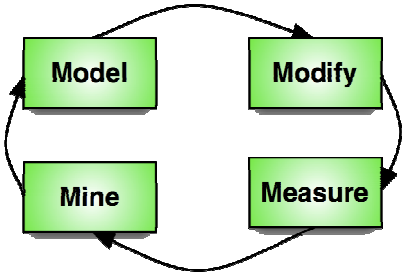

Figure 1: A systems biology approach for high-throughput protein-based datasets: The Modify-Measure-Mine-Model paradigm (Aldridge et al., 2006)

2. EXPERIMENTAL PLATFORMS FOR PROTEIN ACTIVITY QUANTIFICATION

This section describes high-throughput and multiplexed protein activity-based measurements in cells. First, we will introduce the basics for high throughput protein measurements. Then, we will focus on select platforms that are based on antibody detection and list their major advantages and limitations. This is not an exhaustive list of all proteomics platforms as the technology is constantly developing and new assays are continually emerging. Purposely, we will not discuss single-cell based approaches (FACS, high content screening/microscopy) since they are extensively covered in other chapters of this book. Finally, we will introduce computational tools which can be used for the analysis of protein signaling datasets obtained by high throughput approaches.

2.1: Affinity-based assays

Affinity-based assays utilize molecules with high affinity and specificity for the capture and detection of targeted protein(s). The most commonly used high affinity molecules are antibodies, whose selectivity and affinity for the targeted protein are critical for the outcome of the assay. Despite the fact that antibody development and production is a well-matured industry, scientists should be aware that a significant percentage of antibodies often recognize more than one target and thus are not suitable or interchangeable for every assay. Assays are only as good as the antibodies available. In addition, when we discuss antibody-based measurements of "protein activity" we are generally referring to the phosphorylation state of the protein and not to the actual protein activity. This is especially relevant for kinase molecules, in which phosphorylation state of a substrate is a good surrogate of kinase activity. However, discrepancies between kinase phosphorylation status and activity are to be expected under varied conditions [34].

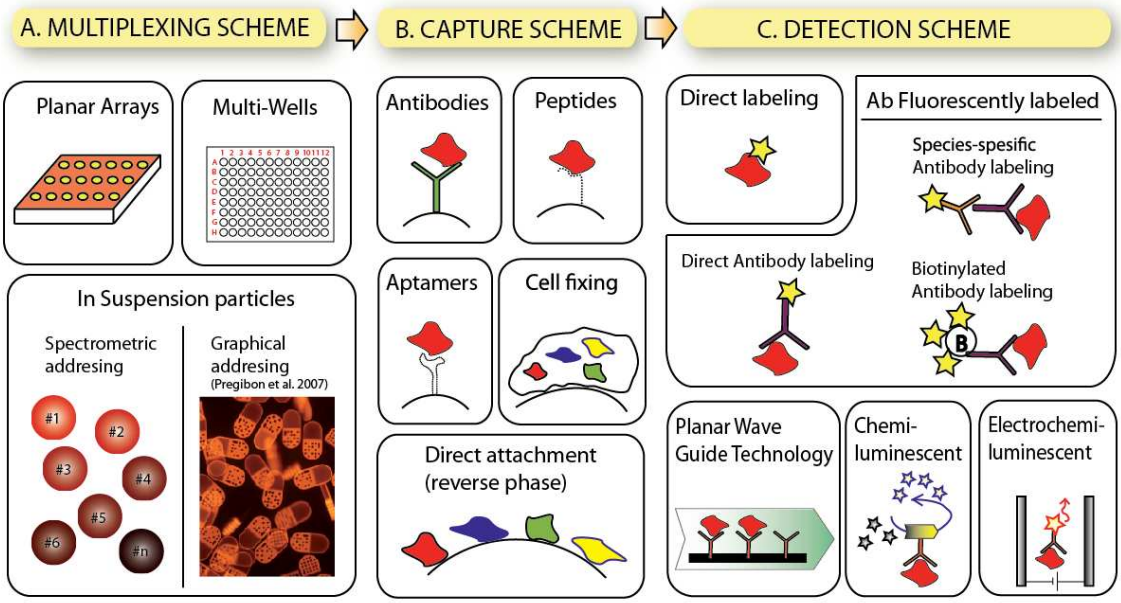

Figure 2: The three main characteristics of a high-throughput affinity based assay. A: A multiplexing scheme in which solid supports (e.g. microparticles [35], beads, spots, or wells) are obtain a unique identity. B: A capture scheme immobilizes and/or isolates the protein of interest. C: A detection scheme generates a readable output which is linearly correlated to the amount of the immobilized protein if interest.

Three are three main characteristics of every affinity based assay (Figure 2):

1. An *identification (multiplexing) scheme* in which a *solid support* (individual wells, spots on glass, beads etc) obtains a unique identity required for multiplexing. A solid support not only contributes to the multiplexability of the assay, but is also essential for all steps in the assay including washes, protein separation and coupling, measurements etc. This unique identity can be obtained by either:
   - *Spatially distributed 2D arrangements*, where the individual coordinates on a 2D plane correspond to individual conditions (e.g. well "B2" in a 96/384well plates, spot with coordinates (*n,m*) on a printed glass slide, or combination of both (5th spot on the B2 well))
   - *Microparticles in suspension* where each microparticle has a unique physical characteristic (i.e. emits a unique spectrometric signature, or carries a unique graphical signature – See figure 2A: In suspension particles)
2. A *capture scheme* that aims to immobilize the protein of interest to the uniquely addressed solid support. Three main approaches widely used for immobilization of a protein of interest are:
   - A capture antibody, peptide, or aptamer attached to a solid substrate (plate, slide, bead)
   - Cell lysate directly spotted on a chemically derived glass slide (usually known as *reverse phase* approach)
   - Cells that are directly fixed on a support (usually on a 96/384 well plate).

3. A *detection scheme* that aims to produce a signal which ideally is linearly proportional to the amount of the protein captured on the support. Depending on the multiplexing and capturing scheme there are at least four distinct categories of detection:
   - Fluorescent-labeled detection which can be either: (i) *Direct labeling*: samples are chemically labeled with a fluorophore. Even though this approach has successfully been applied to study cancer markers [36], it has not been widely adopted because it involves chemical modifications of the samples which may affect their biochemical properties (ii) *Single-antibody approach*: proteins that have been covalently immobilized on a substrate are recognized by a single antibody that is either directly labeled, biotinylated (for binding fluorescent Streptavidin) or recognized by a species-specific antibody (i.e. anti-goat fluorescent secondary antibody). (iii) *Double-antibody approach (or sandwich assay)*: the protein of interest is captured between an immobilized antibody and a secondary antibody which can be either directly labeled, biotinylated or can be recognized by a species-specific antibody.
   - Enzymatic labeled detection: typically a biotinylated secondary antibody is bound by a streptavidin-linked horseradish peroxidase which yields amplified chromogenic products.
   - Planar Wave Guide Technology: This technology, implemented by Zeptosens (www.zeptosens.com), uses guided light passed over a thin film which is located below the detection antibody. The electromagnetic field created by the propagation of the light can lead to measurements with 50 fold increase in sensitivity compared to regular fluorescent schemes.
   - Electrochemiluminescent detection: This technology, implemented by Meso Scale Discovery- MSD (www.meso-scale.com), is based on the electrical induction of an oxidation-reduction cycle resulting in emitted light. Similar to Planar Wave Guide Technology, only labels proximal to the captured antigen surface can be detected, resulting in minimal background signals.

A novel detection scheme recently reported is the proximity ligation procedure in which detection signals can be generated by a PCR- type reaction. This assay is based on pairs of primers that are separately coupled to individual antibodies which recognize closely related epitopes. Thus, when the antibodies bind their targets and come in close proximity, a PCR reaction can generate a DNA product which verifies the proximity and presence of the antibodies ([37-40]).

2.2: Specific platforms for protein measurements

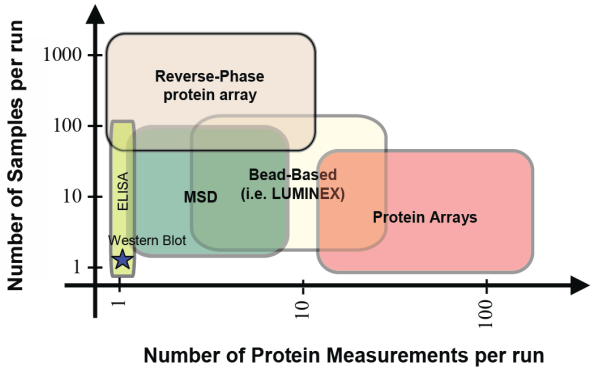

Figure 3: High throughput proteomics platform comparison based on the number of samples and the number of analytes that can be processed.

Protein Microarrays

Protein microarrays (or capture microarrays) employ a capture antibody (alternatively: peptide, nucleotide, etc.) which is covalently bound to a slide (or membrane) in an ordered manner; multiple distinct antibodies may be affixed at separate locations on the same slide. Their main advantage is the ability to measure dozens to hundreds of proteins in each sample although the total number of samples that can be processes is somewhat limited (Figure 3 and Table 1). Even spotting of the antibody is critical for proper interaction with analytes and subsequent reading of the slide. Following blocking of free microarray surfaces to reduce nonspecific binding (and improve signal to background ratio), lysate bathes the entire slide. Detection and quantification of an interaction between the capture antibody and analyte can be achieved either

by *direct labeling* or *sandwich assay*, as mentioned above. The advantage of direct labeling is the simultaneous measurement of many analytes and a requirement for only a single antibody (the capture antibody). However, uneven labeling of all proteins and the chemical alteration of the labeled protein is a source of false positive and high background signals. On the other hand, a sandwich assay employs two antibodies, a capture antibody and a detection antibody. The detection antibody binds to a protein which is already immobilized to the surface via the capture antibody. Subsequently, a fluorescent-labeled secondary antibody associates with the capture antibody allowing for detection and quantification of the signal. The sandwich assay avoids labeling of the sample and ensures more accurate and specific detection of a positive signal. In turn, a limitation is the requirement for two high quality antibodies. Identification of high quality antibodies is a common theme for all protein microarrays and poses a challenge beyond identifying antibodies suitable for Western Blotting which detects denatured proteins. In either case (direct-labeling or sandwich assay), slides are analyzed using readers which are fairly common equipment in many labs today. An alternative to fluorescent detection is the planar wave guided technology by Zeptosens which offers increased signal sensitivity and lower background. However, this methodology requires specialized instrumentation and the proprietary technology makes the assay more expensive.

Meso Scale Discovery (MSD)

Meso Scale Discovery (MSD) is a multi-array technology which utilizes electrochemiluminescent detection of antibody-analyte interactions (www.meso-scale.com). Its main advantage is the ability to measure a few proteins (up to 10) in large number of samples (96 per plate) (Figure 3, Table 1). Capture antibodies are bound to the surface of the wells of a 96-well plate. Each well can accommodate up to 10 distinct antibodies at defined locations (multi-spot). Lysate, supernatants, etc. are incubated in the well followed by the addition of a capture antibody containing a proprietary SULFO-Tag. Electrochemical stimulation is then initiated by carbon electrodes located in the base of the microplate. Activation of SULFO-Tags within close proximity to the electrode results in the emission of light which is read by the specialized MSD reader. The reader can process a plate in 1-3 minutes and does not involve any fluidics. One drawback is the inability to re-read plates, if desired, as the signal is significantly reduced after the plate has been read once. The assays are quite sensitive (to near 10 attamole for some analytes) and have a dynamic range of approximately 5 logs. However, because of the specialized nature of the MSD plates, the customer is limited to targets available from the company, or alternatively developing their own plates. The MSD reader, plates and reagents are quite specialized and come with appreciable cost.

xMAP thechnology

xMAP technology developed by the Luminex Corporation (www.luminexcorp.com) is a bead-based assay which allows for the simultaneous analysis of up to 100 different analytes from a single well. Microspheres (5.6 micron polystyrene beads) are internally dyed to generate up to 100 distinct spectral signatures. Each uniquely identified bead can in turn be coupled to a different capture antibody (or enzyme substrate, DNA, receptors, antigens, etc.). The distinctly conjugated beads can then be mixed (multiplexed) and incubated with a single sample. Next, a mixture of biotinylated detection antibodies is added which are able to interact with analytes to form a sandwich assay on the surface of the bead, and a fluorescently labeled reporter molecule (StreaptAvedin PhycoErythrin or SAPE) binds the detection antibody allowing for detection and quantification. A flow cytometer-based instrument equipped with two lasers and associated optics excite the dyes and allows for quantification of the fluorescent signal representing each analyte. The red diode (635nm) laser excites and identifies the bead as one of 100 distinct signatures. The green diode (532 nm) laser simultaneously excites the fluorescent reporter tag bound to the detection antibody, with the resultant amount of green fluorescence proportional to the amount of analyte captured in the assay.

The theoretical multiplex capabilities of xMAP technology are currently unparalleled. Theoretically, it can measure 100 proteins in each 96 samples (Figure 3, Table 1). In practice, however, there are several limitations with regard to the number of targets which simultaneously can be evaluated including antibody cross-reactivity, natural protein abundance, and antibody competition for protein complexes [41]. A

further advantage using bead suspensions in liquid relative to planar microarrays is the fast reaction kinetics and the high surface-to-volume ratio which leads to better washes, and homogeneous chemical reactions. As with other antibody-based assays, scientists are limited to commercially available targets, although several vendors currently provide a growing list of analytes. The vacuum-based protocol and plate-reading time is moderately time-consuming and limits the number of plates that can be processed. However, advances including the use of magnetic bead technology and automated assay/readers promise to couple the tremendous amount of data per well with the ability to increase the number of plates measured. The xMAP technology requires a specialized bead-reader, although a similar technology has also been developed by Becton Dickinson and Company (Cytometric bead array) which utilizes a more typical FACS instrument.

Reverse Phase

Reverse phase assays also utilize an antibody-based approach, but in contrast to capture microarrays, the cell lysate itself is immobilized on a solid support (typically a chemically-treated slide or membrane). A real strength of this approach is that multiple lysates (dozens to several hundreds) corresponding to different treatments or conditions can be arrayed and processed on a single slide (Figure 3, Table 1). The entire slide can be probed with a single antibody, or alternatively, distinct individual primary antibodies can be compartmentalized at discrete locations. A labeled secondary antibody binds the capture antibody allowing for detection and quantification. Reverse phase arrays only require a single antibody for detection, which is an advantage over the necessity to identify two reliable antibodies as is the case for sandwich assays. However, the signal which is generated by reverse phase represents the sum of specific and non-specific antibody binding, and thus is very dependent on the quality of the antibodies being used (an issue avoided by Western blotting, in which individual proteins are separated). The presence of all cellular proteins bound to a surface also carries with it issues of specificity (cross-reactive antibodies can result in false-positive signals) and decreased signal to noise ratio.

Table 1: A partial list of commercial and generic high-throughput assays for protein-based measurements

ASSAY	MULTIPLEXING SCHEME	CAPTURE SCHEME	DETECTION SCHEME	Size of SAMPLES (per run)	Size of SIGNALS (per run)
ELISA (several brands)	Multi-wells (96 or 384)	Sandwich antibodies	Fluorescent or Enzymatic	100 / plate	1 per well
In-Cell Western *LICOR*	Multi-wells (usually 96)	Fixed Cell and single antibody	Fluorescently labeled Antibody	100 / plate	1-2 per well
Protein Arrays (several brands)	Planar Array slides/membranes	Sandwich antibodies	Fluorescently labeled Antibody	1 / slide	~100 per slide (Ab pair limited)
Protein Arrays on multi-wells (several brands)	Planar Array slides/membranes on multi-wells (usually 96)	Sandwich antibodies and different sample per well	Fluorescently labeled Antibody	100 / plate	~10 spots/well (spot size limited)
Protein Arrays Reverse Phase (several brands)	Planar Array slides/membranes	Spotting and single antibodies	Fluorescently labeled Antibody	1000s / slide	Usually 1 per slide
Protein Arrays Reverse Phase on multi-wells (several brands)	Planar Array slides/membranes on multi-wells (usually 96)	Spotting and single antibodies per well	Fluorescently labeled Antibody	~10 spots/well	~100 per plate

Protein Arrays Meso Scale Discovery	Planar Array on multi-wells (usually 96)	Sandwich antibodies and different sample per well	Electrochemi-luminescent labeled Antibody	100 / plate	~10 spots/well (spot size limited)
Protein Arrays *Zeptosens*	Planar Array on slides	Antibody or Peptides	Antibody w/ Planar Wave Technology	32 / chip	up to 6 per chip
Protein Arrays Reverse Phase *Zeptosens*	Planar Array on slides	Direct attachment (spotting)	Antibody w/ Planar Wave Technology	200 / chip but scalable 1000s / run	1 per chip
Bead-Based (Luminex or Cytometric Bead Array)	In suspension bead particles on multi-wells (usually 96)	Sandwich antibodies	Fluorescently labeled antibody	100 / plate	~30 per well (Ab pair limited)
Flow Cytometry (several brands)	Individual vials or wells	Fixed/Live cell	Fluorescently labeled antibody	1 per vial or 100 per well	up to 4 signals on 1000s of single cells.

3. ORGANIZING AND ANALYZING DATA

The tremendous amount of data obtained through these high-throughput approaches necessitates an organized system of data storage and handling. Context specific information (e.g. conditions under which the data was obtained, protocols used, cell type, etc) should accompany the experimental data itself. A popular approach to storing data (widely used in genomics) is relational database management systems (RDBMS). In a relational database, the subdivision of data and its storage follows a predefined schema, which allows one to identify and maintain links between disparate pieces of information. However, this approach comes at the price of limited flexibility: it is difficult for a relational database to accommodate frequent changes in data formats and to incorporate unstructured information. Such changes and adaptation may be common during the experimental process and thus should be considered when deciding on system for data storage. There are a number of RDBMS type databases adapted for proteomics data, such as SBEAMS—http://www.sbeams.org—or Bioinformatics Resource Manager ([42]). In order to overcome the limitations of RDBMS, we have developed *DataRail*, a free and open-source (http://code.google.com/p/sbpipeline/) MATLAB (http://www.mathworks.com/) toolbox for data management which stores, processes and visualizes experimental data. *DataRail* supports both scripting and GUI-based interaction, and incorporates a variety of data processing algorithms (normalization, discretazation, scaling, etc.) and visualization routines. The information derived from a set of experiments is organized into a structure called a compendium, which consists of multiple n-dimensional data arrays. The algorithms and parameters used during data processing are stored with each array to maintain a record of the provenance of the data [43].

Besides its flexibility, the strength of *DataRail* is its the ability to link the data to mathematical models. Data is imported and exported using a MIDAS format (Minimal Amount of Information for Data Analysis in Systems Biology), a derivative of the MIACA (Minimum Information About a Cellular Assay, http://miaca.sourceforge.net/) format. In addition, import of the data generated by disparate devices (e.g. Luminex reader, ELISA, etc.), as well as export to specific modeling tools (*CellNetAnalyzer* ([44]) and the differential-equation based modeling package *PottersWheel*-http://www.PottersWheel.de/) are possible (see Figure 4).

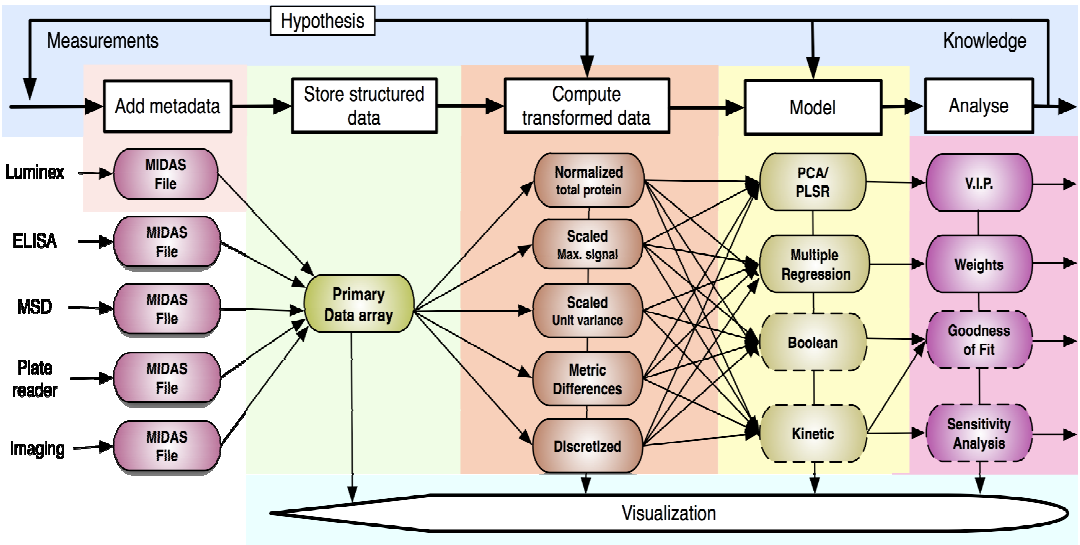

Figure 4: DataRail Workflow. DataRail helps to close the iterative loop between measurements, modeling and generation of hypothesis characteristic of systems biology (see Figure 3). Experimental measurements are first converted into a MIDAS format and then used to assemble a multi-dimensional primary data array. Different algorithms transform the data to create new data arrays (orange) that can then be modeled using internal (full-line boxes) or external routines (dashed boxes) [SaezRodriguez2008].

High throughput data sets are inherently difficult to interpret solely based on inspection and intuition. Thus, mathematical analyses can help to extract information which is not readily apparent. There are two main approaches: by strictly studying the data itself in a hypothesis-free manner using 'data-driven models' ([45]), or by comparing the data to *a priori* biological knowledge encoded in a mathematical framework ([32]).

3.1. Data-driven models

Data-driven approaches to data analysis encompass several methods derived mainly from machine learning and statistics. Methods of machine learning follow one of two paradigms: supervised or unsupervised learning [46]. In supervised learning, such as the support vector machine (SVM) algorithm, a set of objects (e.g. drugs) with certain properties (e.g. their effect on intracellular signals) are assigned to different groups (e.g. toxic or non-toxic). Rules are constructed which link the properties to the groups. These rules can then be used to classify objects whose class is unknown. For example, one could train a classification system with information (e.g. based on intracellular signals) from drugs known to be either toxic or non-toxic, and then use this system to screen new compounds with unknown toxicity [47].

In unsupervised learning, there is no information about which group each object (e.g. drug) belongs to, and therefore, one tries to define such groups (called clusters) based on similarities. One application of clustering is to organize signaling responses based on their similarities ([45]). Another method of unsupervised learning is based on the computation of principal components, which transforms the values of a single measurement into combinations of measurements capturing the maximum of data variability. As an example, instead of considering individual phosphorylation of particular proteins [48], one considers the combined phosphorylation of groups of proteins [45]. This approach has the additional feature that it reduces the dimension of the data. Principal Component Analysis (PCA) can be helpful to obtain biological insight. Using signaling profiles, PCA was employed to qualitatively discriminate apoptotic cell fates ([45]). Partial Least Square Regression (PLSR) is another technique similar to PCA. However, in PLSR, the data is structured into independent and dependent variables (inputs and outputs), whereas in PCA there is only one set of data. PLSR

reduces the inputs and outputs to their principal components and then identifies a linear solution that relates the former to the latter ([45]). PLSR not only provides biological insight but can also be used to predict the results of new experiments. Multiple Linear Regression (MLR) is a technique used to extract correlations between inputs and outputs and can be viewed as simplified PLSR. In MLR, the linear solution is computed directly between the measured variables. It does not reduce the components as PLSR does, but on the other hand, the resulting correlations are links between experimentally measured variables, and therefore the results are easier to interpret. For example, the results of MLR can be visualized as a pathway map that connects phosphoprotein activity to cytokine release [41]. Therefore, MLR can be used as a means to reconstruct the network topology from the experimental data. There are also more sophisticated methods to construct pathways maps out of data that constitute a field known as reverse engineering ([49]), which has been extensively applied to gene regulatory networks ([49]), and to a lesser extent to signaling networks [50, 51]).

In addition to *DataRail*, a number of tools are available to perform data-driven analyses, including the free open source systems R/Bioconductor [46], as well as commercial software such as MATLAB specific toolboxes.

3.2. Topology-based models

The methods described above make use of only the data itself, and in some cases, a classification scheme (categorization of objects into classes in the case of supervised learning). Knowledge accumulated from decades of research and gathered in the form of thousands of scientific publications is not utilized. In order to take advantage of this vast resource, an *in-silico* replica of the current knowledge of the signaling system under consideration can be generated. This model can then be interrogated with respect to its ability to reproduce the experimental data in order to obtain mechanistic insight. There are several mathematical formalisms which can be used to describe signaling networks, varying in the level of detail they incorporate. An extensive review is out of the scope of this chapter, but we will describe the most popular approaches and refer to reviews for further reading.

Probably the simplest description of a signaling network is what is known in mathematics as a graph: each species (typically proteins) is represented as a node, and the nodes are linked with edges ("lines" representing the interactions). Using this simple description, one can unravel important structural properties of the network such as the presence of clusters of proteins (potentially involved in common biological functions), or to identify elements which are highly connected (known as hubs) representing central elements of a signaling network [52, 53].

Directionality and sign (positive or negative) of the arrows (defining the effect between nodes) leads to interaction graphs which capture the direct dependencies among species (see Figure 5). This description allows a useful analysis in terms of examining networks and data: One can compute the paths from any species A to any other species B. Four possibilities exist: the paths are all positive, all negative, mixed positive and negative, or there is no path. Accordingly, the species A can be classified with respect to B as activator, inhibitor, ambiguous, or non-affecting. By comparing the *possible* connections (positive, negative, ambiguous) between two proteins with the experimental data, one can identify consistencies and discrepancies, thereby building a model which best represents the cellular system [54]. For example, in the case of an intervention in which a particular protein A is blocked (e.g. with an inhibitor), if the activity of a different protein B increases, A cannot have a direct positive effect on B.

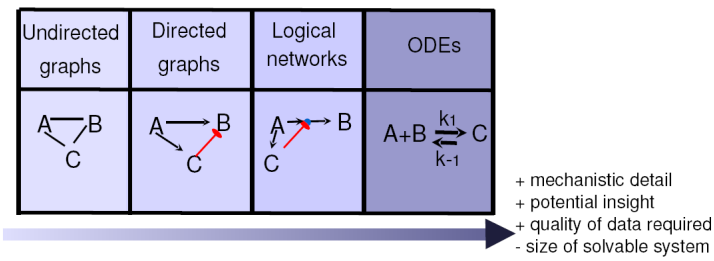

Figure 5. Different representations of pathway topology in mathematical terms.

In cellular networks, an interaction (edge) often represents a relationship among more than two species (nodes). In an interaction graph, the interpretation of such cases can be ambiguous: if two kinases both have a positive effect on a third C, does it mean that either of them can activate C, or rather that both together are required? The logical gates AND (for the latter case) and OR (for the former) can be used to distinguish the possibilities. Thus, by including some additional information about the logic of the signal propagation, one obtains a more accurate description. This type of refinement of interaction graphs is called interaction hypergraphs and is the framework for Boolean (or logical) networks (see Figure 5), in which all states have a discrete value (in the simplest case, either ON or OFF) [54].

Within this framework, it is possible to test the consistency between the network topology and experimental data across different cell types and conditions. Measurements of protein activation under different conditions can automatically be compared with the predictions of a model based on a certain topology. By comparing the models with data from different cell types, one can uncover significant differences in the signaling networks. Furthermore, experimental data that can not be reconciled with the *a priori* knowledge encoded in the maps suggests gaps in our current knowledge that point to potential new connections [41].

Additional insight can be unlocked by constructing and subsequently analyzing detailed, mechanistic, kinetic models. Here, one considers the individual chemical processes underlying signaling events, and defines the reactions consuming and producing species, leading to balances of their concentrations. This is typically encoded as a set of ordinary differential equations (ODEs) which describes the time-dependent concentration of the protein species as a function of kinetic parameters and initial concentration [Aldridgeetal2006]. However, due to the large number of unknown parameter values and entities (enzymatic activities, binding constants, etc), modeling a very large network in a mechanistic manner is an arduous task. If large amounts of data are available, one can try to calibrate the model by finding parameter values so that the model optimally describes the data. This is, however, a very challenging problem [55, 56], so that fully calibrated models are typically available only for small systems or subsets of systems.

A wide range of methods are suitable for the analysis of ODE models: analysis of nonlinear behavior (e.g. oscillations), calculation of sensitivities (the variations of the activation level of certain proteins upon changes in the model parameters, which reveal key parameters in the signaling network, of importance for drug discovery) or finite-time Lyapunov exponents (which discern how initial transients in a signaling network determine alternative cell fates) [32, 56].

Many computational tools are now available which allow scientists to establish, simulate and analyze models. These include a number of tools devoted to graph-based methods (e.g. Cytoscape, visANT, etc.; see Aittokallio for reviews [52]), others to analyze biochemical systems from a qualitative perspective (GNA:[57],GINsim:[58],CAN:[44]), and a larger number are suitable for kinetic modeling ([59, 60]). Most of these tools provide support to set up and analyze the models using user/interfaces and/or programming of scripts. The standard formats CellML and specially SBML ([61]) are widely used to allow an exchange of data between different programs. Information to begin to establish specific models are widely available in the literature and on the web including vast resources about signaling pathways, interactions between molecules (binding, substrate/enzyme relationships, etc.) and cell-specific information. Data mining tools provide an invaluable resource for facilitating the retrieval of this information (see chapter on data mining).

4. CONCLUSION

Knowledge is power. Scientists have long sought to extract as much information as possible from their experiments in order to maximize understanding of the systems they study. This effort is demonstrated by the wide range of genomic screening approaches taken by academics and industry such as the much celebrated Human Genome Project and the ongoing International HapMap Project. Several computational tools have evolved around those datasets that seek a global view of biological systems.

When it comes to protein-based measurements, scientists have been reluctant to follow similar approaches as generating large proteomic datasets represents a tremendous (if not impossible) task using standard biological assays such as Western blots or ELISAs [62, 63]. However, the recent high-throughput protein platforms such as the ones described in this chapter have provided scientists with the opportunity to use computational models to harvest the power of protein and phospho-protein based measurements [41, 47, 48].

That said, despite the fact that current protein datasets lack the synoptic scope of whole genome sequencing, phosphoprotein measurements incorporate knowledge much closer to a cell's biological function.

The pharmaceutical industry has a vested interest in applying a Systems Biology approach to the drug discovery process. For target identification, the interconnectivity of biological pathways can be incorporated into models and optimal novel targets identified. Scientist can investigate whether a single target approach is preferential to multi-targeting approaches which may be carried out either by co-drugging schemes or "dirty" inhibitors. For lead compound discovery, cell-based protein screens can complement chemical screens for understanding the effects of "on-target" inhibition on the rest of the cellular network. For toxicity studies, compounds can be classified as toxic or non-toxic based on their effect on the intracellular and extracellular protein space. The impact of such knowledge on the success or failure of a drug therapy is self-evident.

As the size of protein-based datasets increase, it is apparent that computational tools are becoming more and more invaluable for understanding cell biology. However, combination of a mathematical model with protein datasets is not a trivial task and certain limitations should be considered prior to the implementation of a model or to the generation of an expensive dataset. For example, correlation algorithms such as PCA, PLSR or MLR cannot handle the time space as efficiently as an ODE model. ODE models in turn cannot handle large topologies (because they are computationally expensive) and require time-dense, high quality experimental data. Boolean models on the other hand, can handle large topologies but with limited description of the time domain. Other algorithms, such as SVM, can optimally handle classification problems (e.g. toxic vs. non-toxic drugs) but require a sizable training set.

For a given proteomic platform, two are the main reasons that limit the size of protein datasets: Bench time and cost. On the other hand, they are many reasons for a large dataset: time points requirements, dosing concentrations, measured signals, many stimuli, inhibitors (especially for compound screening), and number of replicates. This is actually a cost/benefit optimization problem where the several dimensions of the dataset (time, inhibitors etc) should be designed to work in harmony with the computational model keeping the total experimental cost down. The complexity of both experimental approaches and computational algorithms also highlights the importance of interdisciplinary collaborations.

The coupling of high throughput proteomic approaches with advanced computational models promises to vastly increase our understanding of biological systems and their behavior. It will be up to the pharmaceutical industry to implement this knowledge in order to develop more efficient and effective drug development strategies.

REFERENCES

[1] Csermely P, Agoston V, Pongor S. The efficiency of multi-target drugs: the network approach might help drug design. Trends Pharmacol Sci 2005;26:178-82.

[2] Araujo RP, Liotta LA, Petricoin EF. Proteins, drug targets and the mechanisms they control: the simple truth about complex networks. Nature Reviews Drug Discovery 2007;6:871-80.

[3] Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100:57-70.

[4] Koch WH. Technology platforms for pharmacogenomic diagnostic assays. Nat Rev Drug Discov 2004;3:749-61.

[5] Teufel A, Staib F, Kanzler S, Weinmann A, Schulze-Bergkamen H, Galle PR. Genetics of hepatocellular carcinoma. World J Gastroenterol 2007;13:2271-82.

[6] Wilkens L, Bredt M, Flemming P, Klempnauer J, Heinrich Kreipe H. Differentiation of multicentric origin from intra-organ metastatic spread of hepatocellular carcinomas by comparative genomic hybridization. J Pathol 2000;192:43-51.

[7] Bosch FX, Ribes J, Borras J. Epidemiology of primary liver cancer. Semin Liver Dis 1999;19:271-85.

[8] Abou-Alfa GK. Hepatocellular carcinoma: molecular biology and therapy. Semin Oncol 2006;33:S79-83.

[9] Thorgeirsson SS, Grisham JW. Molecular pathogenesis of human hepatocellular carcinoma. Nature Genetics 2002;31:339-46.

[10] Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004;350:2129-39. Epub 004 Apr 29.

[11] Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy. Science 2004;304:1497-500.

[12] Ingelman-Sundberg M, Oscarson M, McLellan RA. Polymorphic human cytochrome P450 enzymes: an opportunity for individualized drug treatment. Trends Pharmacol Sci 1999;20:342-9.

[13] Desta Z, Zhao X, Shin JG, Flockhart DA. Clinical significance of the cytochrome P450 2C19 genetic polymorphism. Clin Pharmacokinet 2002;41:913-58.

[14] Syvanen AC. Toward genome-wide SNP genotyping. Nat Genet 2005;37:S5-10.

[15] Bumol TF, Watanabe AM. Genetic information, genomic technologies, and the future of drug discovery. Jama 2001;285:551-5.

[16] Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. Science 1999;286:487-91.

[17] Evans WE, McLeod HL. Pharmacogenomics--drug disposition, drug targets, and side effects. N Engl J Med 2003;348:538-49.

[18] Phillips KA, Veenstra DL, Oren E, Lee JK, Sadee W. Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. Jama 2001;286:2270-9.

[19] Kim HL. Comparison of oligonucleotide-microarray and serial analysis of gene expression (SAGE) in transcript profiling analysis of megakaryocytes derived from CD34+ cells. Exp Mol Med 2003;35:460-6.

[20] Committee on Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment NRC. Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment. Washington, D.C.: The National Academies Press, 2007.

[21] Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, et al. High-throughput DNA methylation profiling using universal bead arrays. Genome Res 2006;16:383-93.

[22] Jabbari K, Bernardi G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene 2004;333:143-9.

[23] Raamsdonk LM, Teusink B, Broadhurst D, Zhang NS, Hayes A, Walsh MC, et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nature Biotechnology 2001;19:45-50.

[24] Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends in Biotechnology 2004;22:245-52.

[25] Haugen AC, Kelley R, Collins JB, Tucker CJ, Deng C, Afshari CA, et al. Integrating phenotypic and

expression profiles to map arsenic-response networks. Genome Biol 2004;5:R95. Epub 2004 Nov 29.
[26] Anderson NL, Anderson NG. The human plasma proteome - History, character, and diagnostic prospects. Molecular and Cellular Proteomics 2002;1:845-67.
[27] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003;422:198-207.
[28] Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 2001;292:929-34.
[29] Pandey A, Mann M. Proteomics to study genes and genomes. Nature 2000;405:837-46.
[30] Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotechnology 1999;17:994-9.
[31] Mann M, Jensen ON. Proteomic analysis of post-translational modifications. Nat Biotechnol 2003;21:255-61.
[32] Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK. Physicochemical modelling of cell signalling pathways. Nat Cell Biol 2006;8:1195-203.
[33] Albeck JG, MacBeath G, White FM, Sorger PK, Lauffenburger DA, Gaudet S. Collecting and organizing systematic sets of protein data. Nat Rev Mol Cell Biol 2006;7:803-12.
[34] 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.
[35] Pregibon DC, Toner M, Doyle PS. Multifunctional Encoded Particles for High-Throughput Biomolecule Analysis. Science 2007;315:1393-6.
[36] Vladimir Knezevic CLVEBJMWVVPJSGLALPJMEFPII. Proteomic profiling of the cancer microenvironment by antibody arrays. PROTEOMICS 2001;1:1271-8.
[37] Fredriksson S, Gullberg M, Jarvius J, Olsson C, Pietras K, Gustafsdottir SM, et al. Protein detection using proximity-dependent DNA ligation assays. Nat Biotech 2002;20:473-7.
[38] Gustafsdottir SM, Schlingemann J, Rada-Iglesias A, Schallmeiner E, Kamali-Moghaddam M, Wadelius C, et al. In vitro analysis of DNA-protein interactions by proximity ligation. Proceedings of the National Academy of Sciences 2007;104:3067-72.
[39] Schallmeiner E, Oksanen E, Ericsson O, Spangberg L, Eriksson S, Stenman U-H, et al. Sensitive protein detection via triple-binder proximity ligation assays. Nat Meth 2007;4:135-7.
[40] Soderberg O, Gullberg M, Jarvius M, Ridderstrale K, Leuchowius K-J, Jarvius J, et al. Direct observation of individual endogenous protein complexes in situ by proximity ligation. Nat Meth 2006;3:995-1000.
[41] Alexopoulos LG, Saez-Rodriguez J, Cosgrove B, Lauffenburger DA, Sorger PK. Context sensitive pathway maps of primary and transformed human hepatocytes reveal profound alterations in NfkB signaling in preparation 2008.
[42] Shah AR, Singhal M, Klicker KR, Stephan EG, Wiley HS, Waters KM. Enabling high-throughput data management for systems biology: The Bioinformatics Resource Manager. Bioinformatics 2007;23:906-9.
[43] Saez-Rodriguez J, Goldsipe A, Muhlich J, Alexopoulos LG, Millard B, Lauffenburger DA, et al. Flexible informatics for linking experimental data to mathematical models via DataRail. Bioinformatics 2008;24:840-7.
[44] Klamt S, Saez-Rodriguez J, Gilles E. Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC Systems Biology 2007;1:2.
[45] Janes KA, Yaffe MB. Data-driven modelling of signal-transduction networks. Nat Rev Mol Cell Biol 2006;7:820-8.
[46] Tarca AL, Carey VJ, Chen XW, Romero R, Draghici S. Machine learning and its applications to biology. PLoS Comput Biol 2007;3:e116.
[47] Alexopoulos LG, Saez-Rodriguez J, Lauffenburger DA, Sorger PK. Support vector machine approach for classicifation of drug toxicity based on their effects on the intracellular and extracellular signaling motifs. in preparation 2008.
[48] Espelin CW, Hendriks BS. Multidimensional Analysis of Context-dependent Production of TNF-alpha in U937 Macrophages. in preparation 2008.
[49] Cho KH, Choo SM, Jung SH, Kim JR, Choi HS, Kim J. Reverse engineering of gene regulatory networks. Iet Systems Biology 2007;1:149-63.

[50] Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Science 2005;308:523-9.
[51] Santos SDM, Verveer PJ, Bastiaens PIH. Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. Nat Cell Biol 2007;9:324-30.
[52] Aittokallio T, Schwikowski B. Graph-based methods for analysing networks in cell biology. Brief Bioinform 2006;7:243-55.
[53] Barabasi A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004;5:101-13.
[54] Klamt S, Saez-Rodriguez J, Lindquist JA, Simeoni L, Gilles ED. A methodology for the structural and functional analysis of signaling and regulatory networks. BMC Bioinformatics 2006;7:56.
[55] Jaqaman K, Danuser G. Linking data to models: data regression. Nat Rev Mol Cell Biol 2006;7:813-9.
[56] Kremling A, Saez-Rodriguez J. Systems biology--An engineering perspective. Journal of Biotechnology 2007;129:329-51.
[57] de Jong H, Geiselmann J, Hernandez C, Page M. Genetic Network Analyzer: qualitative simulation of genetic regulatory networks. Bioinformatics 2003;19:336-44.
[58] Gonzalez AG, Naldi A, Sanchez L, Thieffry D, Chaouiya C. GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. Biosystems 2006;84:91-100.
[59] Vacheva I, Eils R. Computational Systems Biology Platforms. it - Information Technology 2006;48:140-47.
[60] Alves R, Antunes F, Salvador A. Tools for kinetic modeling of biochemical networks. Nat Biotech 2006;24:667-72.
[61] Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 2003;19:524-31.
[62] 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:M500158-MCP200.
[63] 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.