In the first half of the 20th century, the emerging field of molecular biology was met with suspicion by established biochemists: after all, didn’t they already work on biomolecules? And was it appropriate to think of biology in terms of information transfer? The research community has often been slow to embrace emerging scientific disciplines that use new tools to ask new or reformulated questions. In more recent times, many cell biologists have been wary of the development of systems biology. Don’t they already work on biological systems? Does generating long lists of genes and proteins even count as “proper” science?

Doubts have been voiced about systems biology ever since the field’s beginnings in the 1960s. One biologist, writing in the journal Science in 1968, admitted that “system-theoretic ideas seem somewhat strange, and perhaps just a little frightening, to the present generation of structurally-oriented biologists.” Early systems biologists were therefore at pains to explain the relevance of their techniques to existing biological problems (Rosen, 1968).

Forty years on, high-throughput omics technologies have driven the wider emergence of systems biology, with attendant increases in funding and highly cited publications (see box). Ironically, those same technologies only served to increase skepticism in many quarters. Early attempts to map protein interactions on a large scale by mass spectrometry and yeast two-hybrid assays were conceptually important and provided a lot of new data, says cell biologist Tony Pawson (Samuel Lunenfeld Research Institute, Toronto, Canada). “But they suffered from false positives and negatives that gave the whole field a bad rap and made it easier to dismiss. Cell biologists don’t want to worry about what they can and can’t believe when looking at a large data set.”

“Omics can really expand the scope of a project—it’s no longer a case of following up on one or two top hits.”

-Ruedi Aebersold

These early datasets left cell biologists unsatisfied for other reasons too. Many early omics papers focused more on data acquisition than on answering specific biological questions; providing exhaustive parts lists of the proteins and RNAs present within particular cells or organelles. Often, only a handful of those parts were validated, whereas the rest lay unconfirmed in supplemental data tables. Such endeavors were dismissed by some as “stamp-collecting.” Most cell biologists seem to prefer a different pastime: “We like to solve puzzles,” explains Pawson. “It’s what we do and it remains incredibly exciting and rewarding.”

The two camps remained largely separate, following their apparently different pursuits: traditional cell biologists studying individual proteins or processes intently but in isolation from the rest of the cell, and technology-driven omicists generating large amounts of data with little additional interpretation.

But as cell biologists’ puzzles involve more and more molecular pieces that fit together in increasingly complex ways, researchers have begun to appreciate that reductionist approaches can’t fully explain a cell’s behavior. Omics technologies, meanwhile, now produce more reliable and quantitative data, revealing all the molecular pieces of a particular puzzle.
And thanks to advances in data analysis and statistical modeling of their dynamic processes, systems biologists are beginning to arrange these pieces in ways that hint at what the final picture will look like. The two camps have increasingly intersected as systems biology has become more hypothesis driven, and cell biology has begun to think more globally.

A SYSTEMATIC APPROACH

Ultimately, just like traditional cell biological strategies, systems biology aims to link the genotype of a cell to its phenotype—to explain how molecular function generates cellular function. What makes systems biology unique is its focus on understanding the “emergent properties” of a particular system: the behaviors that arise from the virtually innumerable interactions and dynamics of all the components of that system. These interactions and behaviors are much more complex than the simple, linear pathways commonly depicted in cell biology textbooks. John Aitchison, an associate director of the Institute for Systems Biology in Seattle, WA, explains that this is where the systems approach comes into its own: “The dynamic interactions of components can lead to behaviors that are thresholded, that amplify or repress noise, or where an initial response plateaus,” Aitchison says. These synergistic effects on the entire system are almost impossible to predict or explain using the traditional, reductionist approach of studying just a few molecules at a time.

The more holistic approach of systems biology was largely confined to computation and theory until the 1990s, when omics technologies capable of generating large-scale, global datasets began to develop. For a time, many people considered systems biology to be synonymous with these high-throughput methods. Although this “discovery-driven” systems biology may not have immediately excited all cell biologists, these early efforts revealed the huge number of components that function in a particular process and the daunting complexity of their dynamic interactions. “These technologies opened doors for us,” says Aitchison. “But our brains aren’t big enough to handle the incredible complexity that we can now see. In many ways, we’ve opened up Pandora’s box.”

Lucas Pelkmans, an assistant professor at the ETH, Zurich in Switzerland, agrees. “Whenever we do a new experiment, be it an RNAi screen or a proteomics analysis, we realize that there’s revealed the huge number of components that function in a particular process and the daunting complexity of their dynamic interactions. “These technologies opened doors for us,” says Aitchison. “But our brains aren’t big enough to handle the incredible complexity that we can now see. In many ways, we’ve opened up Pandora’s box.”

Lucas Pelkmans, an assistant professor at the ETH, Zurich in Switzerland, agrees. “Whenever we do a new experiment, be it an RNAi screen or a proteomics analysis, we realize that there’s revealed the huge number of components that function in a particular process and the daunting complexity of their dynamic interactions. “These technologies opened doors for us,” says Aitchison. “But our brains aren’t big enough to handle the incredible complexity that we can now see. In many ways, we’ve opened up Pandora’s box.”

Those hypotheses might be about individual interactions within the network, testable by traditional cell biology techniques. As systems biologists continue to improve their methods however, they are beginning to infer the emergent properties of entire systems. High-throughput data now provide more quantitative and dynamic information, and different types of data can be integrated and analyzed with improved statistical models. The hypotheses produced by these approaches can suggest a great deal about a system’s
overall behavior, and can be tested with further omics-type experiments. It’s an iterative process in which cell biologists are needed at every step; their expertise is vital to both formulate the hypotheses and to test them, whether by traditional or high-throughput methods.

Scientists like Pawson see huge potential in fostering such collaborations between systems and cell biologists. Together with systems biologist Ruedi Aebersold (ETH, Zurich) he recently organized a Keystone symposium in Breckenridge, CO, called “Omics Meets Cell Biology.” “We thought it would be interesting to consider how far omics has gone beyond simple data acquisition and ask how cell biologists can make use of that,” explains Pawson. That interest was widely shared by the meeting’s participants, from researchers dipping their toes in new, high-throughput technologies to computational biologists looking to make more biological sense of their complex networks and lengthy gene lists.

INTEGRATING DATA

A key development in systems biology in recent years has been the integration of multiple types of data to draw a more comprehensive picture of cellular behavior (see box). For example, high-throughput measurements of a cell’s mRNA and protein levels can be integrated with data on post-translational modifications of those proteins and on the interactions between them. Phenotypic data arising from large scale RNAi and genetic screens can also be added to the mix. Two genes that give rise to similar phenotypes when knocked out can be mapped onto a network just as two proteins that interact with one another can, and combining this with other data allows interactions to be linked to their functional consequences.

“Our brains aren’t big enough to handle the incredible complexity... we’ve opened up Pandora’s box.”
- John Aitchison

One of the biggest effects that high-throughput omics technologies have had on cell biology is to facilitate functional genomics such as RNAi screens. Researchers increasingly use systems approaches to extend the results of their screens beyond simple gene lists, from which only the top few hits are selected for further study. Data from other omics experiments are integrated with screen results to generate maps of the networks controlling entire cellular processes. Combining multiple types of data can both reveal the full extent of a biological network and increase confidence that functional links between individual components are genuine.

For example, Jennifer Mummery-Widmer, a student in Jürgen Knoblich’s laboratory (IMBA, Vienna, Austria) performed a tissue-specific inducible RNAi screen for genes regulating the Notch signaling pathway in fly external sensory organ development. Mummery-Widmer combined her results with information available in several databases to construct a functionally validated interaction network. Further analysis revealed multiple, highly interconnected clusters within the network. Such interaction clusters usually correspond to particular protein complexes or cellular pathways. Mummery-Widmer could therefore predict that cellular activities not previously linked to Notch signaling were actually key to mediating the pathway’s effects in vivo. She validated this for clusters of nuclear import factors and COP9 signalosome components (Mummery-Widmer et al., 2009).

Pawson’s group applied a similar combination of techniques; performing an RNAi screen for genes affecting ephrin-mediated cell sorting, and a comprehensive analysis of the changes in phosphorylation resulting from ephrin/ephrin receptor signaling. Merging these two datasets with coimmunoprecipitation results produced a global network for both forward and reverse ephrin signaling.

Combining genetic perturbations (for example, in enhancer and repressor screens, or in synthetic lethal studies) further increases the information provided by functional genomics. Julie Ahringer’s group at the Gurdon Institute, Cambridge, UK, is conducting suppressor screens to understand the development of polarity in early C. elegans embryos. Their screen of 29 temperature-sensitive mutants against a library of RNAi constructs will ultimately test around 66,000 interactions. Vitaly, they have streamlined the process to conduct one screen per week. Although primarily directed at gene discovery, further systems analysis and network building will suggest the modes of action of the polarity network as a whole.

MAKING CELL BIOLOGY COUNT

To maximize their usefulness, all of these different datasets—including the phenotypes—must be as quantitative and dynamic as possible. Cell biology remains a largely descriptive science although quantitative imaging techniques are beginning to change this, describing a cell’s properties with the same accuracy that microarrays describe gene expression. And, just like transcriptional profiling, quantitative microscopy is being adapted for high-throughput applications.

Many image-based RNAi screens make use of an open-source, image-analysis software package called CellProfiler. Initially developed by teams at the Whitehead Institute and Massachusetts Institute of Technology (Carpenter et al., 2006), CellProfiler can analyze complex cellular traits by measuring the sizes, shapes, and intensities seen in immunofluorescence images. Designed to be easily adaptable to all kinds of measurements and assays without...
requiring any computer-programming expertise, the software can even learn to recognize interesting phenotypes for itself after a period of “training” with a few sample images. Thus, when combined with automated microscopy, thousands of images can be analyzed to quickly identify positive hits from a large-scale screen (Moffat et al., 2006). This saves time and effort for the researcher when choosing interesting genes to pursue further, and hits can be classified by their precise phenotype (suggesting groups of genes that function together). Quantitative information contained in the images can be extracted and integrated with other data to generate a biological network and predict its properties.

Quantifying cell biology leads to a greater appreciation of the heterogeneity inherent within a population of cells. Even on a coverslip, cells exist in a range of states—different sizes, densities, stress levels, cell cycle stages, etc. Thus, information is lost if phenotypic data about single cells is averaged across a population. Pelkmans believes that population context has a huge influence on RNAi screens. By normalizing for different cellular states within a population, Pelkmans improves the correlation between screens performed in different cell lines and between the phenotypes caused by different siRNAs targeting the same gene.

To generate hypotheses about a biological network’s properties, high-throughput data must also be dynamic, providing information about the system’s response to different perturbations. In the case of proteomics, “selected reaction monitoring” can simultaneously analyze changes in hundreds of specific proteins or post-translational modifications under various conditions. This allows researchers to measure the effects of specific perturbations throughout an extensive biological network, and to correlate this with the phenotypic endpoint. Aebersold’s group in Zurich knocked out almost every kinase and phosphatase in yeast and quantitatively compared thousands of phosphopeptides in every strain. This approach enabled them to identify new substrates and cellular processes regulated by the enzymes, and to construct a comprehensive kinase-substrate network for yeast.

BUILDING COLLABORATIONS

Faced with the almost limitless possibilities presented by omics technologies, knowing which perturbations to perform, and which molecular changes to monitor is more critical than ever. The diversity of the participants’ backgrounds at the Breckenridge meeting reflected Aebersold’s belief that collaborations

Systems cell biology: From scanning mutagenesis to game theory

Recent years have demonstrated the many ways that systems-based approaches can be applied to cell biology. Here are a few highlights that caught our attention:

DATA INTEGRATION

The shift to hypothesis-driven systems biology has coincided with the generation of higher quality global datasets. Batada et al. (2006) assembled a high-confidence yeast protein interaction network from multiple sources, prompting a reexamination of the importance of highly connected proteins to the network’s overall configuration and function. A key advance has been the integration of the different data types produced by omics technologies. Ptacek et al. (2005) identified over 4,000 phosphorylation events catalyzed by yeast protein kinases and integrated their results with protein interaction and transcription factor binding data to uncover regulatory modules within the phosphorylation network. Alber et al. (2007) combined biochemical and morphological information on individual nucleoporins into a comprehensive structure of the nuclear pore complex, providing new functional predictions for cell biologists to investigate.

Saleem et al. (2008), meanwhile, analyzed 249 yeast phosphatase and kinase deletion strains by FACS and microscopy, integrating their results with interaction data to model the signaling networks that regulate different stages of peroxisome biogenesis. König et al. (2008) also incorporated functional genomics with additional data, integrating an siRNA screen with interaction, expression, and gene ontology data to identify the many different interactions between HIV and host cells during viral infection. Multiple data types can be integrated to correctly anticipate the phenotypes of entire animals, as exemplified by Yang et al. (2009) to predict mouse genes that cause obesity. The interaction between computational and experimental biology is an iterative process: Hess et al. (2009) integrated multiple data types to predict new proteins and synthetic interactions involved in mitochondrial biogenesis. Validation of these candidates led to a second round of computational predictions that identified even more proteins, whose mutation caused subtle mitochondrial defects that might not have been noticed in undirected high-throughput screens.
between technologists and cell biologists are vital to fully integrating the two fields: “There are very few research groups who have both the biological expertise and the capability to do high-throughput measurements in the lab, so working together is crucial.” “Experts in one area aren’t necessarily experts in another,” agrees Pawson. “People at the meeting were very excited to forge new collaborations.”

While partnerships between individual laboratories are an important way forward, larger-scale collaborations are also under way that are changing how cell biologists obtain and share data. Mathias Uhlen (Royal Institute of Technology, Stockholm, Sweden) heads a global effort to produce a Human Protein Atlas by generating specific antibodies against every human protein to analyze their expression and localization. The numbers involved are stunning: teams in Sweden and India generate around 150 new antibodies every week, and 17,000 antigens have been sent for immunization so far. Each antibody is characterized at the tissue and cellular level, and the results are made available online (http://www.proteinatlas.org).

Edward Dennis (UC San Diego, CA) is involved in what may be an even greater task: an initiative to identify and quantify all the lipid metabolites present in human cells. The workload of the “Lipid MAPS” consortium spreads across multiple US laboratories because the complete lipidome is thought to encompass more than 10^6 different lipid species.

Cheryl Arrowsmith (Ontario Cancer Institute, Toronto, Canada) describes her research as “omics with purified proteins.” At the Keystone meeting, she presented her group’s work on ubiquitin ligase complexes and on how chromatin-binding proteins recognize histone modifications. These efforts are part of a wider initiative called the Structural Genomics Consortium. Teams in Canada, Sweden, and the UK are working to solve the structures of disease-related human proteins. Different groups focus on different gene families, and over 600 structures have been produced so far. Arrowsmith stresses that

NEW INSIGHTS FROM MODELING

Mathematical modeling can suggest unexpected ways in which biological systems function. Models can be based on traditional cell biology experiments, such as Cai et al.’s (2008) time-lapse microscopy of the yeast transcription factor Crz1, which ultimately revealed how the frequency, rather than duration, of nuclear localization “bursts” are important for gene regulation. Janes et al. (2005, 2006), on the other hand, revealed that cells respond to the proinflammatory cytokine TNF via an autocrine cascade, after modeling almost 8,000 time-dependent changes in various signaling molecules in response to different stimuli. Quantitative modeling can also be a starting point that leads to experimentation, such as Ben-Zvi et al.’s (2008) proposal and subsequent demonstration of a “shutting-based” mechanism for setting up a gradient of BMP activity in Xenopus embryos. For simpler organisms, such as Halobacteria, models and their predictions can be remarkably accurate: Bonneau et al. (2007) successfully predicted the transcription levels of 1,929 Halobacterium salinarum genes to a variety of novel perturbations.

WIDE APPLICATION

Systems-based approaches can be applied to cell biology at a range of levels. Dai et al. (2008) performed high-throughput phenotyping of a library of histone mutants, mapping the contribution of each residue to nucleosome function. Our understanding of the genome is still increasing through new techniques such as RNA-Seq, developed by Nagalakshmi et al. (2008) to map all transcribed regions on yeast chromosomes. A systems-level appreciation of transcriptional regulation is also emerging, fully incorporating the role of microRNAs as in Marson et al.’s (2008) work on regulatory circuits in embryonic stem cells.

At the cellular level, nongenetic variability within clonal populations has a huge role in shaping a cell’s response to certain stimuli: Chang et al. (2008) determined that stochastic changes in gene expression levels affect cell fate decisions in hematopoietic progenitors, while Spencer et al. (2009) found that a cell’s decision to die in response to a pro-apoptotic signal is strongly influenced by natural fluctuations in protein levels. Finally, cooperation and competition between cells has also been modeled by Gore et al. (2009), who studied the interactions between wild-type yeast and freeloading mutants that rely on their wild-type neighbors for sucrose metabolism. The application of game theory to cellular behavior: it’s hard to imagine a more potent example of how systems approaches are revolutionizing cell biology.

Alber, F., et al. 2007. Nature. 450:683–694.
Batada, N.N., et al. 2006. PLoS Biol. 4:e317.
Ben-Zvi, D., et al. 2008. Nature. 453:1205–1211.
Bonneau, R., et al. 2007. Cell. 131:1354–1365.
Cai, L., et al. 2008. Nature. 455:485–490.
Chang, H.H., et al. 2008. Nature. 453:544–547.
Dai, J., et al. 2008. Cell. 134:1066–1078.
Gore, J., et al. 2009. Nature. 459:253–256.
Hess, D.C., et al. 2009. PLoS Genet. 5:e1000407.
Janes, K.A., et al. 2006. Cell. 124:1225–1239.
König, R., et al. 2008. Cell. 135:49–60.
Marson, A., et al. 2008. Cell. 134:521–533.
Nagalakshmi, U., et al. 2008. Science. 320:1344–1349.
Peace, J., et al. 2005. Nature. 438:679–684.
Saleem, R.A., et al. 2008. J. Cell Biol. 181:281–292.
Spencer, S.L., et al. 2009. Nature. 459:428–432.
Yang, X., et al. 2009. Nat. Genet. 41:415–423.
of collaboration, most researchers want to make their results fully available for other scientists to explore. Joan Brugge (Harvard Medical School, Boston, MA) and colleagues recently performed an RNAi screen for genes regulating cell migration (Simpson et al., 2008). Details of the screen, including the sequence and efficacy of every siRNA used and time-lapse movies of every identified hit, are available on the Cell Migration Consortium’s Web site (http://www.cellmigration.org), providing a valuable resource for other interested researchers.

MAKE ME A SUPER MODEL
But what about more global analyses of the results of RNAi and proteomic screens? Computational biologists are vitally important for identifying networks and analyzing their function. “Most cell biologists are not mathematically inclined,” says Aitchison. “Statistical and network analyses need mathematicians and engineers. Looking at the biology from an engineering perspective can really provide a lot of insights.”

And therein lies the main challenge facing systems biologists: how can real biological insight be generated from all of the data produced by omics approaches? Even presenting the complex mass of functional relationships revealed by these high-throughput experiments can pose significant problems. But free, open-source computer programs such as Cytoscape make the visualization and analysis of biological networks easier (Shannon et al., 2003). The program can take a list of genes, search databases for their interactions with one another, and construct a network between the components, much as Mummery-Widmer did with her Notch signaling screen. The resulting “hairball” diagrams contain nodes representing each of the system’s molecular components, linked according to the interactions proposed to exist between them. Those interactions may be physical or enzymatic, or may link genes with highly similar expression profiles or RNAi phenotypes.

The hairballs can still seem confusing, but further analysis often reveals sub-networks that correspond to more familiar cell biological concepts like signaling pathways or multi-protein complexes. This latter step can be important for placing a screen’s results into a biological context. One systems biologist admits that his colleagues often get lost in detailed network analyses, forgetting to apply their findings back to the underlying biology.

Meanwhile, thanks to the quantitative and dynamic data now available, computational biologists can statistically model how these biological networks function to produce the correct cellular outcome. Some of these complex systems are remarkably robust, maintaining constant output despite large fluctuations in the amounts or activities of individual components. Uwe Sauer (ETH Zurich), for example, finds that flux through metabolic pathways remains steady in different yeast mutant strains, despite large changes in the expression levels of metabolic enzymes and in the amounts of specific intermediate metabolites.

THE EMERGENT PROPERTIES OF CELL BIOLOGY
Systems biology studies like those discussed here rely equally on computational and experimental biology, and they require experts in each of these fields to understand and appreciate one another. “Cell biology students need to be taught more statistics and clustering algorithms,” says Aebersold. “That doesn’t mean everyone
needs to be a computer scientist, but you need a reasonable level of knowledge to be able to talk to the experts.”

Universities around the world are recognizing this and developing multidisciplinary systems biology programs for graduate students. But, Aebersold stresses, basic cell biological knowledge is still paramount: “Successful studies start with a clearly defined biological question. You need that knowledge to formulate the question and choose the best way to address it.”

Pawson, meanwhile, is impressed by some of the younger scientists in the field. “They move seamlessly between computational and cell biology,” he says. “I suspect they don’t think of them as separate skills; they’re just doing whatever they need to do to understand the biology.”

Aebersold thinks this trend will continue and even accelerate. “Most cell biologists will vastly expand the type of data that they seek and interpret,” he predicts. “Virtually everyone in my department in Zurich is thinking of branching out from their current research questions and making connections to other processes in the cell, looking at various levels of regulation. Omics can really expand the scope of a project—it’s no longer a case of following up on one or two top hits. The cell is increasingly seen as a tightly interlinked unit—you can’t just work on one process in isolation.”

Aitchison also sees huge changes ahead for cell biology. “We’ll be doing a lot of dynamic network analyses and I hope we’ll be able to describe cell biological phenomena at a higher level—in terms of the emergent properties of the underlying systems.”

“Technologies are moving so fast right now that the most successful integrated projects occur when biology- and technology-oriented groups work together,” he says. “Only when techniques have solidified—as in the case of transcript arrays—does it become a good idea to provide core facilities.” “These techniques and strategies will eventually move out of the hands of those with specialized expertise and access to the latest instruments,” agrees Pawson. “They’ll become much more accessible to the average cell biologist.”

Thus in the end, just as molecular biology was absorbed into numerous different fields, omics and systems biology will likely be assimilated, adding yet more tools to the cell biologist’s toolkit. In the meantime, Pawson and Aebersold are planning to organize another meeting of omics and cell biology two years from now. “It’s a matter of getting the omicists and cell biologists talking,” says Pawson. “We’re not exactly sure what the future holds, but the prospects are very exciting.”

But Aebersold has a note of caution for universities rushing to set up core omics facilities for their researchers. “Technologies are moving so fast right now that the most successful integrated projects occur when biology- and technology-oriented groups work together,” he says. “Only when techniques have solidified—as in the case of transcript arrays—does it become a good idea to provide core facilities.” “These techniques and strategies will eventually move out of the hands of those with specialized expertise and access to the latest instruments,” agrees Pawson. “They’ll become much more accessible to the average cell biologist.”

Thus in the end, just as molecular biology was absorbed into numerous different fields, omics and systems biology will likely be assimilated, adding yet more tools to the cell biologist’s toolkit. In the meantime, Pawson and Aebersold are planning to organize another meeting of omics and cell biology two years from now. “It’s a matter of getting the omicists and cell biologists talking,” says Pawson. “We’re not exactly sure what the future holds, but the prospects are very exciting.”

Carpenter, A.E., et al. 2006. Genome Biol. 7:R100.
Moffat, J., et al. 2006. Cell. 124:1283–1298.
Mummery-Widmer, J.L., et al. 2009. Nature. 458:987–992.
Rosen, R. 1968. Science. 161:34–35.
Shannon, P., et al. 2003. Genome Res. 13:2498–2504.
Simpson, K.J., et al. 2008. Nat. Cell Biol. 10:1027–1038.