Grand challenge: accelerating discovery through technology development

Health & Medicine

As biologists we know that organisms are composed of subunits called cells, that sequential DNA trinucleotides encode the amino acid sequences of proteins, and that RNA transfers genetic information from DNA to the ribosome. But what we may not often reflect on is that these facts, the very tenets of modern biology, represent the fruits of technological innovation. From the early microscopes through which Hooke and van Leeuwenhoek saw cells for the first time, to the DNA sequencing technologies that have allowed decoding of entire genomes, the development of new instruments and assays has always been key to making discovery possible. This relationship between technological innovation and biological discovery is as important today as it ever was, and we must therefore continue to push the boundaries of current methodologies in order to accelerate the pace of science and to translate the knowledge gained into practical applications.

An excellent example of the power of technical advances to drive discovery comes from the rise of the field of genomics. With the realization that DNA contains genetic information, it became clear that biologists would need to be able to determine the nucleotide sequences of DNA molecules in order to understand genetics at the molecular level. This first became widely possible in the early 1970s with Sanger and Coulson's introduction of the " plus and minus" sequencing technique (Sanger and Coulson, 1975), followed by the Maxam and Gilbert method(Maxam and Gilbert, 1977). The application of these methods led to significant new findings, including the complete sequence of the phi X174 phage genome (Sanger et al., 1977a), but both were plagued by complex protocols and were limited to sequencing only short stretches of DNA at a time. In 1977 Sanger and colleagues introduced a

DNA sequencing method based on chain-terminating dideoxy nucleotides (Sanger et al., 1977b) which went on to become the method of choice due to its relative simplicity and ability to deliver longer sequence reads. After further improvement over the years and eventual automation (Prober et al., 1987), this method was employed to sequence the entire genomes of many different species, leading to unprecedented views of the structure of genomes from all kingdoms of life.

In addition to analyzing the genome itself, large-scale sequencing of cDNA libraries and the use of methods such as serial analysis of gene expression (SAGE; Velculescu et al., 1995) gave the first glimpses into the landscape of the transcriptome. These methods allowed more precise mapping of genes within a genome and provided insights into differences in their expression levels. With the advent of microarray technology (Schena et al., 1995), routine and accurate analysis of global gene expression profiles became possible in individual laboratories. In subsequent years, coupling of the chromatin immunoprecipitation (ChIP) assay with microarray analysis allowed mapping of the locations of many different gene regulatory factors across the genome (Ren et al., 2000), leading to the birth of epigenomics and a new era in the study of gene regulation. With the arrival of the latest generation of massively parallel sequencing technologies (Morozova and Marra, 2008) it has now become possible to sequence entire genomes, to map the global distribution of chromatin-associated proteins and DNA methylation, and to quantitatively measure the transcriptome very rapidly and at an ever-decreasing cost. This technological revolution in genomics

has led to a veritable explosion of new insights into the workings of biological systems.

From such humble beginnings, the evolution of technologies for analyzing DNA has provided answers to the previously intractable questions of what a genome looks like, how it works, and how it changes at large scale over evolutionary time. Yet we still have a long way to go toward understanding all of the operating principles of the genome, characterizing the molecular circuitry controlling cell physiology and behavior, and elucidating the interactions between cells that allow the development of a multicellular organism from a single cell. These challenges will require further advances in instrumentation and methodology in all fields of biological science.

Like genomics, the high-throughput analysis of proteins has also seen great technical advances in recent years. Important goals in this area are to fully define the proteomes of different tissues, cells, and organelles and to reliably quantify differences in the composition of these proteomes. Progress along these lines has been driven in large part by the improvement in mass spectrometry (MS) instruments along with development of clever new assays and sample preparation methods. Techniques such as multidimensional protein identification technology (MudPIT; Washburn et al., 2001), which combines liquid chromatography and tandem MS, have driven the number of proteins that can be identified in a single sample into the thousands. Further, quantitative proteomics methods such as stable isotope labeling by amino acids in cell culture (SILAC; Ong et al., 2002) and isobaric tags for relative and absolute quantitation (iTRAQ; Ross et al., 2004) have improved

quantitative comparisons between protein samples. At this point one major challenge in the field is to increase the number of proteins that can be identified simultaneously in a sample.

In addition to proteome analysis, the large-scale qualitative and quantitative measurement of metabolites has become possible in recent years through the use of nuclear magnetic resonance (NMR) and MS. These technologies have been used to identify a wide range of primary and secondary metabolites in plants (Fiehn et al., 2000), including quantitative analysis of hormones in plant extracts (Birkemeyer et al., 2003 ; Pan et al., 2010). As in the case of proteomics, only a relatively small number of molecular species can be identified in a sample at one time, and this is a challenge that will need to be addressed in the years ahead. A second problem common to current methods for high-throughput analysis of proteins and small molecules is that they require relatively large amounts of starting material. This is particularly acute in the case of proteomics and metabolomics, where amplification of molecules in the sample is not possible as it is for nucleic acids. In the future this limitation must be overcome in order to allow the study of small numbers of cells or even single cells. Solutions to the problems of depth of coverage and starting sample size will likely come in the form of improvements to instrumentation and further development of sample preparation and fractionation methods prior to analysis.

To date, the majority of genomic, proteomic, and metabolomic studies in plants have been conducted on whole plants or selected tissues, which gives an output representing an amalgamation of signals from different cell types.

However, in attempting to understand the biology of multicellular organisms we must ultimately aim our experimental measurements at individual cell types, given that each is clearly specialized for a specific function and therefore has a unique physiology. Historically this has been a difficult task, but several methods have been developed that allow individual cell types to be isolated from plant tissue and studied. One of these is laser capture microdissection (LCM), in which specific cells from a tissue section are cut out with a laser and captured (Nakazono et al., 2003). Fluorescence activated cell sorting (FACS) is another technique that can be used to isolate specific cell types from a tissue. In this method a fluorescent protein is expressed in the desired cell type, the tissue is digested with cell wall-degrading enzymes to produce free protoplasts, and the protoplasts of interest are separated from the others based on their fluorescence (Birnbaum et al., 2003 , 2005). In addition, a recently developed method called isolation of nuclei tagged in specific cell types (INTACT) relies on cell type-specific expression of an affinity-tagged nuclear envelope protein. Nuclei bearing the tag can be affinity purified from a total nuclei preparation in order to study gene expression and other nuclear processes that take place in the cell type of interest (Deal and Henikoff, 2010 , 2011). Each of these methods has been applied successfully to studies of specific cell types in plants, but each approach has its virtues and drawbacks. For example, LCM has the ability to isolate specific cells directly from tissue without the use of transgenic plants, but only small numbers of cells can be isolated by this method. On the other hand, FACS and INTACT deliver larger quantities of target cells and nuclei, respectively, but require cell type-specific promoters

for transgenic expression of the proteins on which each purification method is based. A challenge for the future is the development of new techniques that will allow any desired cell type to be isolated in quantity without prior knowledge of cell type-specific promoters.

Another key to understanding cell and developmental biology is the use of imaging methods. Great progress has been made in this area, particularly in microscopic imaging of individual cells. A major breakthrough in studying the behavior of proteins and cell structures came with the discovery of the green fluorescent protein (GFP), which could be fused to proteins of interest to allow their visualization (Chalfie et al., 1994). Since that time additional fluorescent proteins with different spectral properties, such as YFP and RFP, have been discovered and employed for simultaneous tracking of multiple proteins in a cell. The use of these FPs has allowed not only the subcellular location of proteins to be examined, but also studies of protein dynamics through assays such as fluorescence recovery after photobleaching (FRAP; Jacobson et al., 1976 ; Dundr and Misteli, 2003). Further, protein interactions within a living cell can now be examined with the use of techniques like bimolecular fluorescence complementation (BiFC; Hu et al., 2002) and Forster resonance energy transfer (FRET; Clegg, 1995). In recent years we have seen the advent of many genetically encoded protein tags, as well as a new generation of chemical tags (Beatty, 2011). These new tools are diversifying the approaches to protein visualization as well as the number and types of proteins that can be monitored simultaneously. While there are now many tools for high resolution imaging of proteins and nucleic acids, an exciting new area is the development of genetically encoded sensors for

metabolites and ions (Frommer et al., 2009). In addition, MS-based methods have recently been developed for the imaging of other types of biomolecules, including lipids and carbohydrates in tissues (Goto-Inoue et al., 2011 ; Lunsford et al., 2011). Although the limited resolution provided by standard light microscopes has been a hindrance to determining protein location with high precision, a recent breakthrough by Tsien and colleagues has produced a genetically encoded protein tag that can be visualized by both light and electron microscopy (Shu et al., 2011). Also, super-resolution optical microscopy methods have provided views of subcellular structures at unprecedented resolution, down to the single molecule level (Huang, 2010). Future work should include improvements to automated microscopy instrumentation and software in order to allow high-throughput analyses to be conducted easily (Wollman and Stuurman, 2007). There is also a need for new affinity reagents, such as aptamers (Liang et al., 2011), for diversifying the types of molecules that can be imaged within cells.

Aside from cellular imaging, breakthroughs have also been made in the comprehensive imaging of developmental processes in plants. For example, Godin and colleagues tracked *Arabidopsis* flower development over time at cell resolution using a three-dimensional imaging and reconstruction method that allows the automated tracking of cell lineages (Fernandez et al., 2011). This work provides a generally applicable technique for analyzing the dynamics of developmental processes in terms of the behavior of cells individually and as a group. In addition, light sheet fluorescence microscopy has recently been used to follow *Arabidopsis* root development temporally at the whole-organ, single cell, and subcellular levels (Maizel et al., 2011).

These types of approaches and extensions of them will be essential for an integrated mechanistic understanding of plant development.

Our success in generating so much data of so many types has created new challenges in attaining a complete understanding of biological systems. Each day new data gush forth from the fields of genomics, proteomics, metabolomics, biochemistry, cellular biology, developmental biology, and genetics – but how do we put them all together? Integrating and synthesizing these diverse data types into a coherent understanding is one of the major problems that we currently face. Important steps have been taken in this direction, including the generation of new types of databases (Joung et al., 2009 ; Hamada et al., 2011) and development of computational approaches for integrating data across sets and types (Shannon et al., 2003 ; Tieri et al., 2011 ; Wiesinger et al., 2011). Also, literature mining methods are emerging as an excellent tool to collect all available data on a particular subject and to annotate large-scale experiments (Jensen et al., 2006). Clearly, we are in the early part of the learning curve in terms of data integration, and great challenges still lie ahead.

The topics discussed here do not, by any means, represent a comprehensive list of the technical challenges that we must rise to in the coming years. Every discipline in plant science has its own set of methodological hurdles that must be overcome, and there is also a need for general plant research tools including adaptations of existing methods for use in plants, as well as the development of new genetic tools and resources for non-model plants. While the task of technology development can be frustrating and time

consuming, it is indeed a worthwhile endeavor. The creation of new tools will continue to accelerate discovery by providing a bridge between what we know and what we would like to know but cannot yet tackle experimentally.

References

Beatty, K. E. (2011). Chemical strategies for tagging and imaging the proteome. *Mol. Biosyst.* 7, 2360–2367.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Birkemeyer, C., Kolasa, A., and Kopka, J. (2003). Comprehensive chemical derivatization for gas chromatography-mass spectrometry-based multi-targeted profiling of the major phytohormones. *J. Chromatogr. A* 993, 89–102.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Birnbaum, K., Jung, J. W., Wang, J. Y., Lambert, G. M., Hirst, J. A., Galbraith, D. W., and Benfey, P. N. (2005). Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nat. Methods* 2, 615–619.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Birnbaum, K., Shasha, D. E., Wang, J. Y., Jung, J. W., Lambert, G. M., Galbraith, D. W., and Benfey, P. N. (2003). A gene expression map of the Arabidopsis root. *Science* 302, 1956–1960.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science* 263, 802–805.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Clegg, R. M. (1995). Fluorescence resonance energy transfer. *Curr. Opin. Biotechnol.* 6, 103–110.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Deal, R. B., and Henikoff, S. (2010). A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev. Cell* 18, 1030–1040.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Deal, R. B., and Henikoff, S. (2011). The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana* . *Nat. Protoc.* 6, 56–68.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Dundr, M., and Misteli, T. (2003). Measuring dynamics of nuclear proteins by photobleaching. *Curr. Protoc. Cell Biol.* Chapter 13, 13. 5. 1–13. 5. 18.

Fernandez, R., Das, P., Mirabet, V., Moscardi, E., Traas, J., Verdeil, J.-L., Malandain, G., and Godin, C. (2011). Imaging plant growth in 4D: robust tissue reconstruction and lineaging at cell resolution. *Nat. Methods* 7, 547–553.

CrossRef Full Text

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R. N., and Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18, 1157–1161.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Frommer, W. B., Davidson, M. W., and Campbell, R. E. (2009). Genetically encoded biosensors based on engineered fluorescent proteins. *Chem. Soc. Rev.* 38, 2833–2841.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Goto-Inoue, N., Hayasaka, T., Zaima, N., and Setou, M. (2011). Imaging mass spectrometry for lipidomics. *Biochim. Biophys. Acta* doi: 10. 1016/j. bbalip. 2011. 03. 004. [Epub ahead of print].

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Hamada, K., Hongo, K., Suwabe, K., Shimizu, A., Nagayama, T., Abe, R., Kikuchi, S., Yamamoto, N., Fujii, T., Yokoyama, K., Tsuchida, H., Sano, K., Mochizuki, T., Oki, N., Horiuchi, Y., Fujita, M., Watanabe, M., Matsuoka, M., Kurata, N., and Yano, K. (2011). OryzaExpress: an integrated database of gene expression networks and omics annotations in rice. *Plant Cell Physiol.* 52, 220–229.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Hu, C. D., Chinenov, Y., and Kerppola, T. K. (2002). Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* 9, 789–798.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Huang, B. (2010). Super-resolution optical microscopy: multiple choices. *Curr. Opin. Chem. Biol.* 14, 10–14.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Jacobson, K., Derzko, Z., Wu, E. S., Hou, Y., and Poste, G. (1976). Measurement of the lateral mobility of cell surface components in single, living cells by fluorescence recovery after photobleaching. *J. Supramol. Struct.* 5, 565(417)–576(428).

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Jensen, L. J., Saric, J., and Bork, P. (2006). Literature mining for the biologist: from information retrieval to biological discovery. *Nat. Rev. Genet.* 7, 119–129.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Joung, J.-G., Corbett, A. M., Fellman, S. M., Tieman, D. M., Klee, H. J., Giovannoni, J. J., and Fei, Z. (2009). Plant MetGenMAP: an integrative analysis system for plant systems biology. *Plant Physiol.* 151, 1758–1768.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Liang, Y., Zhang, Z., Wei, H., Hu, Q., Deng, J., Guo, D., Cui, Z., and Zhang, X. (2011). Aptamer beacons for visualization of endogenous protein HIV-1 reverse transcriptase in living cells. *Biosens. Bioelectron.* doi: 10. 1016/j. bios. 2011. 07. 031. [Epub ahead of print].

CrossRef Full Text

Lunsford, K. A., Peter, G. F., and Yost, R. A. (2011). Direct matrix-assisted laser desorption/ionization mass spectrometric imaging of cellulose and hemicellulose in populus tissue. *Anal. Chem.* doi: 10. 1021/ac2013527. [Epub ahead of print].

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Maizel, A., Von Wangenheim, D., Federici, F., Haseloff, J., and Stelzer, E. H. (2011). High resolution, live imaging of plant growth in near physiological bright conditions using light sheet fluorescence microscopy. *Plant J.* doi: 10. 1111/j. 1365-313X. 2011. 04692. x. [Epub ahead of print].

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Maxam, A. M., and Gilbert, W. (1977). A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U. S. A.* 74, 560–564.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Morozova, O., and Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255–264.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Nakazono, M., Qiu, F., Borsuk, L. A., and Schnable, P. S. (2003). Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. *Plant Cell* 15, 583–596.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics* 1, 376–386.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Pan, X., Welti, R., and Wang, X. (2010). Quantitative analysis of major plant hormones in crude plant extracts by high-performance liquid chromatography–mass spectrometry. *Nat. Protoc.* 5, 986–992.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A., and Baumeister, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* 238, 336–341.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J.,

Bell, S. P., and Young, R. A. (2000). Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell Proteomics* 3, 1154–1169.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M., and Smith, M. (1977a). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265, 687–695.

CrossRef Full Text

Sanger, F., Nicklen, S., and Coulson, A. R. (1977b). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–5467.

CrossRef Full Text

Sanger, F., and Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94, 441–448.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Shu, X., Lev-Ram, V., Deerinck, T. J., Qi, Y., Ramko, E. B., Davidson, M. W., Jin, Y., Ellisman, M. H., and Tsien, R. Y. (2011). A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues, and organisms. *PLoS Biol.* 9, e1001041. doi: 10. 1371/journal. pbio. 1001041

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Tieri, P., De La Fuente, A., Termanini, A., and Franceschi, C. (2011). Integrating omics data for signaling pathways, interactome reconstruction, and functional analysis. *Methods Mol. Biol.* 719, 415–433.

Pubmed Abstract | Pubmed Full Text

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science* 270, 484–487.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Washburn, M. P., Wolters, D., and Yates, J. R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 19, 242–247.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text

Wiesinger, M., Haiduk, M., Behr, M., De Abreu Madeira, H. L., Glockler, G., Perco, P., and Lukas, A. (2011). Data and knowledge management in cross-Omics research projects. *Methods Mol. Biol.* 719, 97–111.

Pubmed Abstract | Pubmed Full Text

Wollman, R., and Stuurman, N. (2007). High throughput microscopy: from raw images to discoveries. *J. Cell Sci.* 120, 3715–3722.

Pubmed Abstract | Pubmed Full Text | CrossRef Full Text