An Introduction to Functional Genomics and Systems Biology

Evelien M. Bunnik and Karine G. Le Roch*

Department of Cell Biology and Neuroscience, University of California, Riverside, California.

Karine G. Le Roch, PhD

Submitted for publication November 8, 2012. Accepted in revised form December 12, 2012.

*Correspondence: Institute for Integrative Genome Biology, Center for Disease Vector Research, Department of Cell Biology and Neuroscience, University of California–Riverside, 900 University Ave., Genomics Building Room 2121B, Riverside, CA 92521 (e-mail: karine.leroch@ucr.edu).

Abbreviations and Acronyms

AT = adenine + thymine

cDNA = complementary deoxyribonucleic acid

ChIP-Seq = chromatin immunoprecipitation coupled to next-generation sequencing

cRNA = complementary ribonucleic acid

ddNTP = dideoxy nucleotide triphosphate

DNA = deoxyribonucleic acid

FAIRE = formaldehyde-assisted isolation of regulatory elements

GeLC-MS = gel electrophoresis coupled to liquid chromatography and mass spectrometry

(*continued*)

Objective: Over the past decade, the development of high-throughput technologies for DNA and protein analysis has revolutionized the ways in which cells can be studied. Within a relatively short time frame, research has changed from studying individual genes and proteins to analyzing entire genomes and proteomes.
Approach: In this article, we summarize the technologies and concepts that form the basis of this functional genomics approach.
Results: Microarray and next-generation sequencing technologies have allowed researchers to investigate many different aspects of the cell, including DNA mutations, histone modifications, DNA methylation, chromatin structure, transcription, and translation on a genome-wide level. In addition, mass spectrometry technologies have undergone significant development and currently enable us to globally profile protein levels, protein–protein interactions, post-translational protein modifications, and metabolites.
Innovation and Conclusion: The integration of information from the various processes that occur within a cell provides a more complete picture of how genes give rise to biological functions, and will ultimately help us to understand the biology of organisms, in both health and disease.

INTRODUCTION

The field of functional genomics attempts to describe the functions and interactions of genes and proteins by making use of genome-wide approaches, in contrast to the gene-by-gene approach of classical molecular biology techniques. It combines data derived from the various processes related to DNA sequence, gene expression, and protein function, such as coding and noncoding transcription, protein translation, protein–DNA, protein–RNA, and protein–protein interactions. Together, these data are used to model interactive and dynamic networks that regulate gene expression, cell differentiation, and cell cycle progression.

Studying cells at a systems level has been facilitated by recent technological advancements, as well as the availability of complete genome sequences. Since the landmark publication of the first draft of the human genome in 2001,[1,2] the genomes of hundreds of organisms from all branches of the tree of life have been sequenced. This has lead to improved annotations of genes and their products, and has enabled genome-wide studies aimed at understanding interactions and molecular processes in the cell.

CLINICAL PROBLEMS ADDRESSED

This article will give a brief overview of high-throughput omics technologies and their applications, and how these powerful tools have expanded the possibilities for studying the complex biology of cells, organs, and full organisms.

ADVANCES IN WOUND CARE, VOLUME 2, NUMBER 9

DOI: 10.1089/wound.2012.0379

MATERIALS AND METHODS

DNA microarrays

DNA microarrays consist of thousands of microscopic DNA spots (probes) that are bound to a solid surface, such as glass or a silicon chip (Affymetrix) or microscopic beads (Illumina). Labeled single-stranded DNA or antisense RNA fragments from a sample of interest are hybridized to the DNA microarray under high-stringency conditions. Each probe is identified by its location on the DNA microarray, and the amount of hybridization detected for a specific probe is proportional to the level of nucleic acids from the corresponding genomic location in the original sample.

Next-generation sequencing technologies

Three main next-generation sequencing (NGS) platforms are widely used: the Roche 454 platform (Roche Life Sciences),[3] the Applied Biosystems SOLiD platform (Applied Biosystems),[4] and the Illumina (formerly known as Solexa) Genome Analyzer and HiSeq platforms (Illumina).[5] For these three NGS platforms, template DNA is fragmented, bound to adaptors, amplified by polymerase chain reaction, and subsequently immobilized on beads or on an array where clusters consisting of identical DNA fragments are formed. These clusters are read by sequential cycles of nucleotide incorporation, washing, and detection, where the number of cycles eventually determines the read length (Fig. 1). A fourth DNA sequencing technology has been recently developed by Ion Torrent. The Ion Torrent technology takes advantage of the hydrogen ion that is released as a byproduct of the incorporation of a nucleotide into a DNA strand by polymerase. The sequencer directly senses the ions produced by template-directed DNA polymerase synthesis on a massive parallel semiconductor-sensing device that directly transforms this chemical signal to digital information.[6]

Over the years, sequencing pipelines have greatly improved in throughput and costs for instruments and reagents, along with improvements in computational power, data storage, and bioinformatics tools that facilitate the analysis of the growing quantities of sequence reads. Together, these advancements have caused a dramatic drop in sequencing costs, down to about $0.09 (U.S.) per megabase in early 2012.[7] Several new companies, such as Helicos Biosciences, Pacific Biosciences, and Oxford Nanopore Technologies, are currently developing novel, so-called third generation sequencing techniques that do not require amplification of template DNA, but are able to read the sequence of single DNA molecules.[8,9] These technologies could significantly advance the sequencing field by greatly reducing the cost for reagents and improving the throughput, while simultaneously eliminating any bias introduced during the template amplification step of the NGS protocol.

Mass spectrometry

A mass spectrometer consists of three components: an ion source to convert a gas-phase sample into ions, a mass analyzer to separate the ions by means of an electromagnetic field, and a detector. The development of ionization techniques that enable the transfer of proteins and peptides into the gas phase without substantial degradation has been crucial for the application of mass spectrometry (MS) in large-scale proteomic studies. The most commonly used ionization techniques are matrix-assisted laser desorption ionization and electrospray ionization. These ionization techniques can be combined with various types of mass analyzers that separate ions based on the mass-to-charge ratio by either trapping ions in an electrical field (trapping mass spectrometers) or by accelerating ions through an

Abbreviations and Acronyms (*continued*)

MAINE = micrococcal nuclease-assisted isolation of nucleosomes

mRNA = messenger ribonucleic acid

MS = mass spectrometry

MudPIT = multidimensional protein identification technology

NGS = next-generation sequencing

NMR = nuclear magnetic resonance

RNA = ribonucleic acid

RNA-Seq = high-throughput ribonucleic acid sequencing

SAGE = serial analysis of gene expression

SDS-PAGE = sodium dodecyl sulfate–polyacrylamide gel electrophoresis

TAP = tandem affinity purification

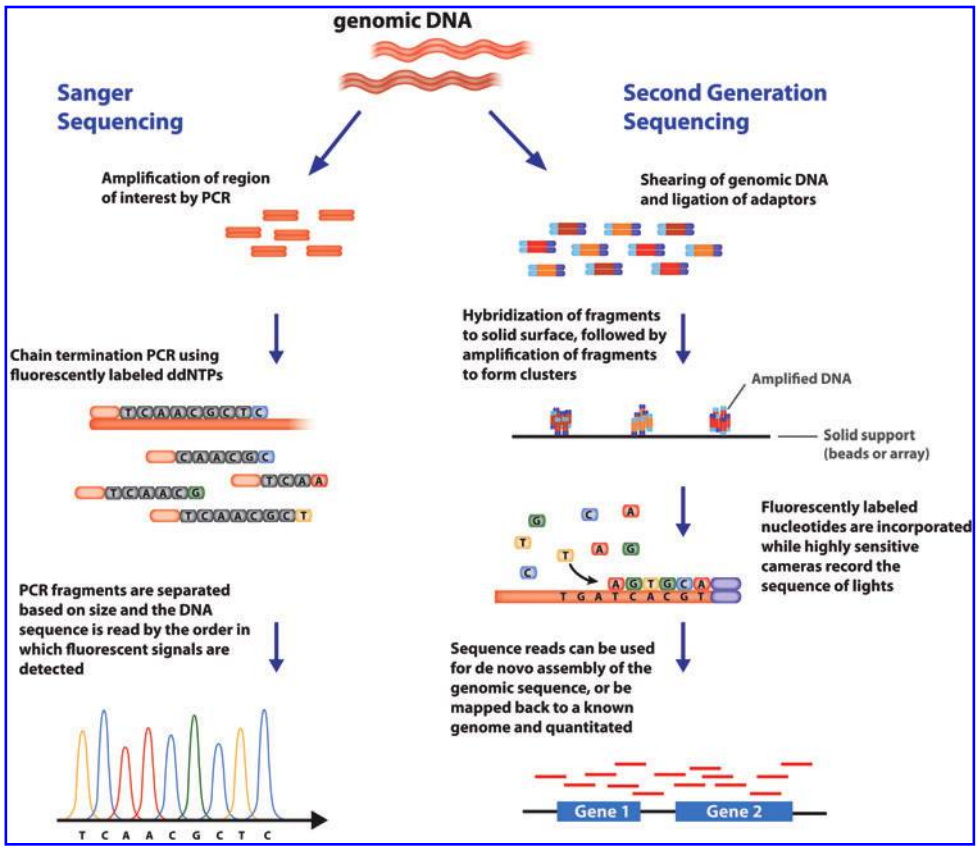

Figure 1. Comparison between Sanger sequencing and next-generation sequencing (NGS) technologies. Sanger sequencing is limited to determining the order of one fragment of DNA per reaction, up to a maximum length of ~700 bases. NGS platforms can sequence millions of DNA fragments in parallel in one reaction, yielding enormous amounts of data. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/wound

electrical field and measuring the time-of-flight. A comparison of instrument configurations that are most commonly used in proteomics is provided elsewhere.[10] The most advanced mass spectrometer available to date is the Orbitrap, which has a high resolution, a high mass accuracy, and a large dynamic range that make it suitable for a wide range of proteomics and metabolomics applications.

The most common strategy for proteomic studies is a bottom-up approach, in which a protein sample is first enzymatically digested into smaller peptides, followed by separation of the peptides by charge, hydrophobicity, or a combination of these characteristics, and then injected into the mass spectrometer. Individual peptide spectra are used to indirectly identify complete proteins that were present in the original sample.

RESULTS

Genomics

For almost 30 years, sequencing of DNA has largely been dependent on the first-generation Sanger dideoxy sequencing method. Sanger sequencing requires each sequence read to be amplified and read individually (Fig. 1). Despite considerable improvements in automation and throughput, Sanger sequencing remains relatively expensive and labor intensive. For whole-genome sequencing, it is dependent upon bacterial cloning, which is time-consuming and can introduce biases as a result of, for example, difficulties in cloning AT-rich fragments or genes that are toxic to bacteria. Since 2005, several NGS technologies have become commercially available, which have transformed the field of whole-genome sequencing. The amount of data generated in parallel from small amounts of DNA is enormous, and currently reaches up to 6 billion short reads or 600 gigabase per instrument run. This has greatly facilitated the sequencing of the complete genome of organisms to identify DNA mutations, ranging from single-nucleotide polymorphisms to large gene deletion or duplication events. In addition, these technologies have enabled a range of novel applications, including genome-wide analysis of epigenetic mechanisms, such as DNA methylation, location of

histone modifications, transcription factors binding events, and nucleosome positioning, as well as profiling of gene expression (see Transcriptomics section).

Many of these applications are based on mapping short reads of DNA obtained from a particular sample to a reference genome and analyzing the distribution of these reads (Fig. 2).[11] For example, for determining the locations of a particular histone modification, chromatin is sheared into mononucleosomal fragments. Chromatin fragments containing the histone modification of interest are immunoprecipitated and the corresponding DNA fragments are sequenced (ChIP-Seq). Since only the 5′ end of DNA fragments are sequenced, the sequence reads obtained in this experiment will map to the outer side of the nucleosome. However, the midpoint of the histone can be determined by extrapolation of the distribution of sequence reads from either side of the nucleosome. This type of experimental setup and data analysis yields highly accurate positioning of modified histones. Novel experimental procedures are continuously being applied to achieve even a higher resolution, recently yielding single-base pair resolution for both transcription factor binding sites[12] and nucleosome positioning.[13]

NGS has greatly improved our ability to study the various genetic and epigenetic mechanisms, with unprecedented detail and specificity. This information has provided us with enormous insight into gene regulation and cell cycle control, as well as the roll of mutations and epigenetic mechanisms in pathogenesis.

Transcriptomics

Regulation of gene expression is fundamentally important for cell development and differentiation. Profiling the abundance of transcripts in different cell types and under various conditions increases

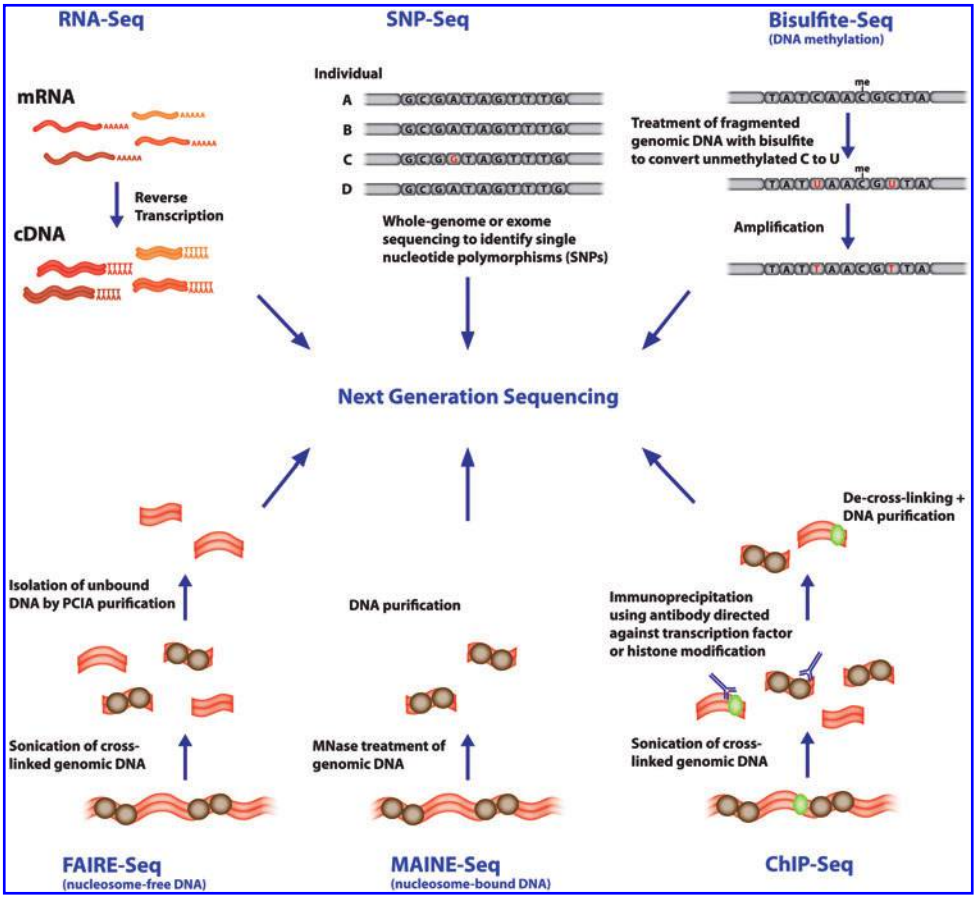

Figure 2. Applications of NGS. The types of experiments that can be performed using NGS are many fold and are certainly not limited to the applications listed here. Applications include sequencing the complete genome or exome (all coding regions of the genome) to identify single-nucleotide polymorphisms (SNP-Seq) or other DNA mutations, profiling the genome-wide locations of methylated cytosines (Bisulfite-Seq), investigating various aspects of chromatin structure and regulation of gene expression by determining nucleosome positioning (MAINE-Seq and FAIRE-Seq), histone modifications or transcription factor binding (ChIP-Seq), and determining mRNA levels to study gene expression and its regulation (RNA-Seq). To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/wound

our knowledge about gene function and regulatory pathways. In the past, RNA transcripts have been analyzed using Northern blotting or reverse transcription polymerase chain reaction, which are restricted to limited numbers of known transcripts. Serial analysis of gene expression (SAGE) was developed in 1995, and consists of sequencing small tags that correspond to the 3′ fragments of messenger RNA (mRNA).[14] This allows for a highly quantitative analysis by simply counting the number of tags that map to a particular gene. Despite several improvements to the original protocol, SAGE is no longer widely used as it is very labor intensive and relatively low-throughput compared to newly developed NGS applications.

Around the same time, the first DNA microarrays were developed for measuring the expression levels of large numbers of genes.[15,16] Transcripts isolated from a sample of interest are converted into cDNA or cRNA, are labeled, and are subsequently hybridized to the DNA microarray. The amount of hybridization detected for a specific probe is proportional to the transcript level of the corresponding gene. Comparing transcript levels between various cell types or conditions can be used to identify genes that are involved in cell differentiation or in responses to certain environmental changes. Cluster analysis is often employed to characterize genes that have similar expression profiles and are therefore likely to have similar biological functions. DNA microarrays are still in use today and continue to provide valuable biological information, although it is to be expected that gene expression profiling will shift more and more toward the use of NGS tools.

NGS technologies have opened the door for a broad range of genome-wide analyses related to gene expression and transcript profiles, which are collectively known as RNA-Seq (Fig. 2). Sequence reads derived from an RNA sample of interest are mapped to a reference genome, where the number of reads that map to a certain gene corresponds to the expression level of that gene. Besides profiling gene expression levels, RNA-Seq can be used to analyze transcript boundaries and intron/exon junctions and to discover novel transcripts and novel alternative splice variants. In addition, it can be applied to, for example, profiling of noncoding RNA, nascent transcripts, and ribosome-associated mRNA, and has the potential to immensely increase our understanding of the different roles of RNA and of the various levels of regulation of gene expression. RNA-Seq provides a combination of high-throughput, large sequencing depth, and genome-wide coverage, which is not offered by any other tool used for gene expression analysis in the past. An additional advantage of RNA-Seq over DNA microarrays is that it is not dependent on the availability of a microarray for the species of interest and can therefore be implemented for all organisms.

Proteomics

Proteins are one of the functional units of the cell, and it is therefore essential to understand how proteins function to completely understand biological processes. Since transcript levels do not necessarily correlate with protein levels, quantitation of proteins is required to unequivocally determine their abundance. In addition, many proteins are post-translationally modified, adding an extra level of complexity to their structure and function.

Analysis of the protein content of cells can be performed by two-dimensional gel electrophoresis, where proteins are separated first by size, and then by charge, followed by MS. Another relatively straightforward proteomics approach is geLC-MS, where proteins are first separated by one-dimensional gel electrophoresis (SDS-PAGE). Each gel lane is then divided into equally sized sections, and the proteins from each section are digested, separated by liquid chromatography, and analyzed by MS. More recently, a high-throughput technology has been developed that is more suitable for identifying large numbers of proteins from complex mixtures. Using the multidimensional protein identification technology (MudPIT), proteins are digested into peptides that are then separated by means of two-dimensional chromatography, based on both charge and hydrophobicity, and are subsequently analyzed by MS.[17] The signals of each peptide obtained using MS can then be compared to a database of previously sequenced proteins or to a database of predicted proteins based on the genome sequence to identify the protein from which the peptide was derived. MudPIT allows for a highly sensitive detection of proteins and has over the last decade been applied to a broad range of cells and organisms. It has successfully been used to profile organelle and membrane proteins, identify post-translational modifications, dissect protein complexes, and analyze protein expression.

Interactomics

Determining the abundance and localization of a protein is not sufficient to understand its function. Many molecular processes in the cell are performed by complexes of proteins that are organized by protein–protein interactions. Such functional interactions are found in signal transduction,

transcriptional regulation, metabolic pathways, and many other biological functions. Deciphering these interactions is crucial to understanding the interactive pathways and networks that form the basis of many cellular processes.

Protein–protein interactions can be studied using a variety of methods. The two-hybrid system has been used for the first time in 1989[7] and has since then been modified to allow proteome-scale screening.[18,19] In the two-hybrid method, one protein of interest is fused to a DNA binding domain, while another protein of interest is fused to an activation domain. Both fusion proteins are then expressed in the same cell, which could in theory be any living cell, although yeast and bacterial cells are most widely used. If the proteins interact, a reporter gene is transcriptionally activated, which will change the phenotype of the cell and allow for an easy readout. In addition to the original two-hybrid system, which requires proteins to be present in the nucleus, the cytotrap yeast two-hybrid tool has been developed for detection of protein–protein interactions in the cytoplasm. The two-hybrid technique is relatively straightforward and can be used as a first screen to identify interacting protein partners. However, the rate of false positives is relatively high, and interactions found by the two-hybrid technique should always be validated using other tools.

While the two-hybrid system is limited to screening the interaction between two proteins at a time, affinity purification methods may be more suitable to study the organization of proteins into complexes. This technique entails fusing a tag to a protein of interest, which is subsequently used to isolate this protein together with all proteins that are bound. The bound proteins are then analyzed by MS. In 2002, this technique was first performed in yeast and revealed thousands of protein–protein interactions, many of which had not been described before.[20,21] Since then, the tandem affinity purification (TAP) strategy has become increasingly popular. TAP involves two rounds of affinity purification that provide a high specificity, but may on the other hand result in the loss of transient or very weak protein–protein interactions. In combination with other tools, such as protein microarray and phage display, these technologies have vastly increased our understanding of interactive protein networks.

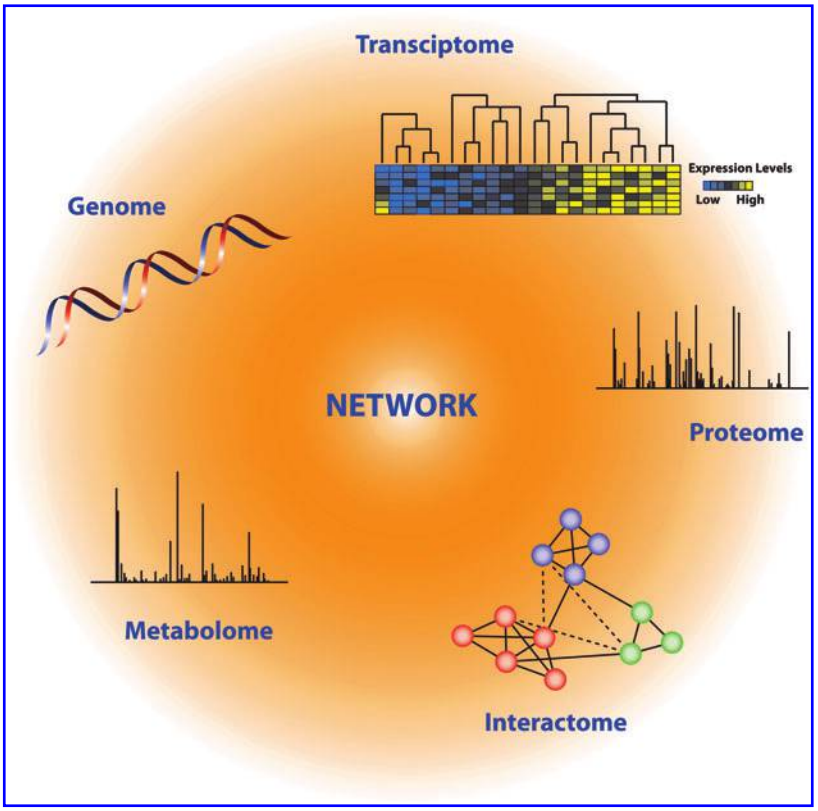

Figure 3. Schematic overview of network analysis. Integration of information from different aspects of the cell, such as genome, transcriptome, proteome, interactome, and metabolome, will increase our understanding of how these components are interconnected and how these interactions determine biological functions. To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/wound

Metabolomics

Metabolites are small molecules, such as amino acids, sugars, and fatty acids, that are chemically transformed by enzymes during metabolism and that play critical roles in various biological processes. Metabolite levels correlate more directly with a cellular phenotype than genes or proteins, and therefore provide an accurate functional readout of the state of a cell. Researchers have long been interested in profiling metabolites on a global level, but only recently technologies have emerged that enable these types of studies. The tools most widely used for global metabolomics approaches are nuclear magnetic resonance (NMR) and liquid chromatography coupled to MS. The main advantages of NMR are its high reproducibility and ease of sample preparation. However, the sensitivity of MS-based techniques is higher compared to NMR, and allows the detection of most metabolites present in a cell. Information from metabolomics studies will increase our understanding of complex cellular metabolism, characterize new metabolic pathways, and identify new targets for therapeutic intervention in, for example, cancer.

KEY FINDINGS

- NGS technologies enable a range of applications for studying various cellular processes related to DNA, chromatin structure, transcription, and translation on a genome-wide level.
- Advances in MS allow large-scale studies into proteins, protein–protein interactions, post-translational protein modifications, and metabolites.
- Integration of genome-wide data by network analyses will improve our understanding of cellular biology.

DISCUSSION

With a plethora of information emerging from various omics studies, the main challenge in systems biology is to integrate these data into a single network and to find out how genes, transcripts, proteins, and metabolites interact to regulate the biological processes that determine cell function and cell cycle progression (Fig. 3). The availability of large amounts of data has led to the development of more robust computational methods for network analysis. These tools can, for example, be used to predict protein function by means of guilt-by-association analysis. This type of analysis is based on the principle that the function of a protein is likely to resemble the function of proteins with which it interacts or is coexpressed. In addition, multiple tools are available that support pathway analyses to determine whether certain pathways or gene ontologies are over-represented in certain biological processes.

To obtain accurate and complete cell models, network analysis should not only be based on experiments performed in model organisms under standard laboratory conditions. In contrast, using information obtained from multiple species, cell types, or under various environmental conditions will allow differentiation between relatively static housekeeping genes and the dynamic processes involved in response to stress or other external and internal signals.[22] This will ultimately lead to building improved models of biologically relevant interactions between all components of a cell.

INNOVATION

Novel technologies developed over the past decade allow a systems biology approach to studying the complex processes that shape cells, organs, and organisms. Instead of focusing on single genes or proteins, NGS platforms and MS applications provide the opportunity to study genes, transcripts, proteins, and their interactions on a genome-wide level. Ultimately, the integration of this information will result in an improved understanding of how genes give rise to biological functions.

ACKNOWLEDGMENTS AND FUNDING SOURCES

E.M.B. is supported by the Human Frontier Science Program (grant LT00507/2011-L) and K.G.R. is supported by the National Institutes of Health (grant R01 AI85077-01A1).

AUTHOR DISCLOSURE AND GHOSTWRITING

No competing financial interests exist. The content of this article was expressly written by the authors listed. No ghostwriters were used to write this article.

ABOUT THE AUTHORS

Dr. Evelien Bunnik is a post-doctoral fellow and **Dr. Karine Le Roch** is an associate professor at the University of California, Riverside, CA. They use functional genomics approaches, such as proteomics and high-throughput sequencing technologies to elucidate critical regulatory networks driving the malaria parasite life cycle progression and to identify novel drug targets.

REFERENCES

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ; International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. Nature 2001; **409:** 860.

2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, and Zhu X: The sequence of the human genome. Science 2001; **291:** 1304.

3. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, and Rothberg JM: Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005; **437:** 376.

4. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, and Church GM: Accurate multiplex polony sequencing of an evolved bacterial genome. Science 2005; **309:** 1728.

5. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara E Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L,

Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, and Smith AJ: Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008; **456:** 53.

6. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, and Bustillo J: An integrated semiconductor device enabling non-optical genome sequencing. Nature 2011; **475:** 348.
7. Wetterstrand KA: DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: www.genome.gov/sequencingcosts (accessed June 1, 2012).
8. Braslavsky I, Hebert B, Kartalov E, and Quake SR: Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci USA 2003; **100:** 3960.
9. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, and Turner S: Real-time DNA sequencing from single polymerase molecules. Science 2009; **323:** 133.
10. Yates JR, Ruse CI, and Nakorchevsky A: Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng 2009; **11:** 49.
11. Wold B and Myers RM: Sequence census methods for functional genomics. Nat Methods 2008; **5:** 19.
12. Rhee HS, and Pugh BF: Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 2011; **147:** 1408.
13. Brogaard K, Xi L, Wang JP, and Widom J: A map of nucleosome positions in yeast at base-pair resolution. Nature 2012; **486:** 496.
14. Velculescu VE, Zhang L, Vogelstein B, and Kinzler KW: Serial analysis of gene expression. Science 1995; **270:** 484.
15. Schena M, Shalon D, Davis RW, and Brown PO: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995; **270:** 467.
16. Schena M, Shalon D, Heller R, Chai A, Brown PO, and Davis RW: Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 1996; **93:** 10614.
17. Washburn MP, Wolters D, and Yates JR 3rd.: Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 2001; **19:** 242.
18. Fields S and Song O: A novel genetic system to detect protein-protein interactions. Nature 1989; **340:** 245.
19. Phizicky E, Bastiaens PI, Zhu H, Snyder M, and Fields S: Protein analysis on a proteomic scale. Nature 2003; **422:** 208.
20. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreault M, Muskat B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sørensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, and Tyers M: Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 2002; **415:** 180.
21. Gavin AC, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Höfert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, and Superti-Furga G: Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002; **415:** 141.
22. Ideker T and Krogan NJ: Differential network biology. Mol Syst Biol 2012; **8:** 565.