1

TOXICOLOGY, ''OMICS'' TECHNOLOGIES, AND TOXICOGENOMICS: A PRIMER

Darrell R. Boverhof and B. Bhaskar Gollapudi

Toxicology and Environmental Research and Consulting, The Dow Chemical Company, Midland, MI

1.1 INTRODUCTION

The last half century has witnessed an explosion in the advancement of computational sciences that has fostered growth in numerous other disciplines. This has notably included advancements in biological and biochemical sciences that have been translated to applications in the field of toxicology. This chapter provides

Applications of Toxicogenomics in Safety Evaluation and Risk Assessment, First Edition.
Edited by Darrell R. Boverhof and B. Bhaskar Gollapudi.

a broad overview of the changing landscape of toxicology as influenced by these advancements. A general overview of the traditional foundations of toxicology will be presented, followed by a brief description of the dawn of the "omics" era and its many tools. In closing, this chapter will highlight the emergence and growth of toxicogenomics by illustrating how toxicologists are integrating the omics tools to advance the science of toxicology.

1.2 TOXICOLOGY

Toxicology can be defined, in very general terms, as the study of adverse effects of physical and chemical agents on living organisms (Eaton and Gilbert, 2008). It is a field that borrows and applies the concepts of the basic sciences: chemistry, biology, and mathematics. Therefore the discipline of toxicology has advanced as the foundational basic sciences have advanced by continually adopting their concepts to better understand toxicity. Traditionally toxicology has been much of an observational science; however, with developing knowledge in the basic sciences, there has been a continual movement (and desire for further movement) toward being a more mechanistically informed discipline.

The history of toxicology is rooted in poisons. In its earliest applications, toxicology can be traced back to the dawn of humanity where humans used plant extracts and snake venoms for hunting or poisonings. Since these early times, there have been various indications of a progressively increased understanding of toxins and toxicity as evidenced through the works of Hippocrates, Aristotle, and Dioscorides, who documented their understanding of poisons and even made descriptions and classifications of poisons into origins from plants, animals, and minerals (Gallo, 2008; Lane and Borzelleca, 2008). The understanding and use of poisons and the development of antidotes continued to progress through the middle ages. The sixteenth century was marked with numerous developments in the field of toxicology, many of which were attributed to Paracelsus (1493–1541) who is often referred to as the father of toxicology. Paracelsus was a physician-alchemist who tried to bring chemistry and the scientific method into medicine. In doing so, he established some of the earliest and most basic concepts of toxicology. Perhaps the most well-known of these is the famous statement:

> *All substances are poisons; there is none which is not a poison. The right dose differentiates poison from remedy*.

Paracelsus used this concept to defend his use of inorganic substances in medicine when his critics claimed that such substances were too toxic to be used as therapeutic agents (Borzelleca, 2000). This concept is today often stated simply as—*It is the dose that makes the poison*—which is a central tenet of toxicology. Other contributions from Paracelsus toward the field of toxicology included the concept that diseases and toxins localize to a particular organ, which contributed to the concept of target organs of toxicity (Borzelleca, 2000). Paracelsus was also instrumental in documenting the hazards associated with metalworking and mining. In

1567 he published, *On the Miner's Sickness and Other Diseases of Miners* (Gallo, 2008). This publication was integral in establishing the concepts for occupational toxicology and medicine.

Toxicology continued to evolve over the coming centuries with developments in the fields of medicine and chemistry. Its rapid development in the 1900s coincided with a dramatic increase in the production and marketing of pharmaceuticals, pesticides, and industrial chemicals. One view on the development of toxicology as a discipline is that it has expanded as a result of incidents of poisonings and the subsequent legislation that was put in place to address these occurrences (Gallo, 2008). Examples in the United States include the Wiley Bill introduced in 1906 to address several incidents of poisonings with patent medicines. This was followed by the Copeland Bill in 1938 which was in response to several deaths due to the use of ethylene glycol in sulfanilamide elixirs. In 1945 the Federal Insecticide and Rodenticide Act was signed into law and represented the first time in history that a substance that was neither a drug nor a food had to be shown to be safe and efficacious. Additional events that continued to fuel the development of the field of toxicology included the tragedy of thalidomide in late 1950s and early 1960s, in which several thousand children were born with birth defects, and Rachel Carson's publishing of *Silent Spring* in 1962, which raised concerns about environmental pollution. Events such as these led to an expanded role for the US Food and Drug Administration (FDA) and the establishment of the US Environmental Protection Agency (EPA) in December 1970. Later incidents such as Love Canal, in which health hazards were associated with the development of a community on a former chemical waste site, led to the development of the Toxic Substances Control Act, which regulates the manufacture, handling, and use of chemicals. These incidents and legislative acts represent only a small portion of the events of the 1900s that promoted the development of toxicology and the application of this discipline to safety evaluation and risk assessment (Gallo, 2008).

Modern-day toxicology is often described as consisting of three main categories: descriptive (or observational), mechanistic, and regulatory (Eaton and Gilbert, 2008). Each category has distinct aspects; however, they are all interrelated and contribute to the assessment of risk. Descriptive or observational toxicology is concerned with implementation and conduct of standardized testing approaches to identify and assess hazards posed by new and existing chemicals. This is done through the observation of toxic responses in animal or cellular test systems and using these data to assess human and environmental risk. The category of mechanistic toxicology is concerned with understanding the molecular and biochemical mechanisms by which chemicals or physical agents result in toxicity. This category involves a more basic research approach and can provide insight into human relevance of responses as well as information for the development of safer alternatives, Regulatory toxicology is involved in the review of both descriptive hazard assessment data along with any available mechanistic information to assess the risk posed by a substance to humans and the environment. Although each of these areas contributes to the risk assessment of a toxicant, a full assessment requires a very expansive set of information such that toxicologists often focus on more specific

aspects of hazard and risk. This call for specificity has given rise to many subdisciplines of toxicology including genetic toxicology, immunotoxicology, reproductive toxicology, developmental toxicology, and ecotoxicology. Each one of these subdisciplines has its own set of assays for assessing hazard and even differing approaches for the assessment of risk. As with the field of toxicology as a whole, the emergence and refinement of these subdisciplines is closely related to developments in their respective basic areas of biology (genetics, immunology, etc.).

Thus toxicology has a long history, but the discipline continues to grow in response to increases in our understanding of the complexity of its foundational basic sciences, namely chemistry, biology, and mathematics. Toxicology has evolved from the study of adverse effects of chemicals on living organisms to protecting human health and the environment by predicting the hazards of chemicals prior to human exposure, limiting the levels of exposure through risk assessment methodologies, and when needed, informing remediation methodologies to mitigate human exposure from contaminated legacy sites. This is reflected in the recently adopted definition by the Society of Toxicology, which defines toxicology as the study of the adverse effects of chemical, physical, or biological agents on living organisms and the ecosystem, including the prevention and amelioration of such adverse effects. Fundamental to this process has been the understanding of the relationship between exposures and hazards and the subsequent use of these data to define risk. Historically, risk assessments for chemicals have been largely developed on the basis of descriptive or observational data; however, with advancements in our understanding of biology, there has been a push to increase the relevance of risk assessments through increased use of mechanistic data. Therefore toxicology is continually looking to advancements in biology, chemistry, and mathematics that can be adapted to improve the overall assessment of risk. In line with this approach recent years have seen an intense focus on the application of "omics" technologies to toxicology.

1.3 THE GENOMICS ERA AND OMICS TECHNOLOGIES

Significant scientific breakthroughs in the past half-century have helped usher in a new era of biological science: the genomic era. The pinnacle event that facilitated all downstream events during this period was the elucidation of the physical structure of DNA by Watson and Crick in 1953. This fundamental discovery has provided the basis for all subsequent research in such areas as DNA replication, protein synthesis, and gene regulation. It is through our understanding of DNA that many of our most commonly used molecular biology techniques have arisen, from the cloning of genes to the amplification of DNA through polymerase chain reaction (PCR). Roughly 25 years after the discovery of the double helix, Frederick Sanger developed a quicker and more efficient method to sequence DNA strands, which would provide the basic principles used in high-throughput sequencers in the 1990s (Smith et al., 1986). Combined with the ability to amplify vast amounts of DNA using PCR, these high-throughput capillary sequencers would eventually set the

stage for the completion of the first draft of the human genome in early 2000s and define the start of the genomic era.

The sequencing of the human genome could not have been completed without the parallel advancements in the computational world. Roughly around the same time as the discovery of DNA, groundbreaking achievements were also being realized in the in the field of computers. Just three years after Watson and Crick's discovery, the integrated circuit was developed. The integrated circuit is analogous to DNA in that it is the basic building block in every electronic device that we use today. In 1965, Intel co-founder Gordon Moore speculated that the number of transistors on an integrated circuit would double every two years. During the 1970s, this statement was generally applied to the rate of increase in computational power and is commonly referred to as Moore's Law (Moore, 1995). As we fast forward into the 1990s, supercomputers are able to process incredible volumes of information and personal computers are commonplace in the average Western household.

The parallel achievements in scientific research and computer technology reached a crossroads around the 1990s. Automated sequencers were rapidly being developed to handle large-scale sequencing projects, which helped establish a handful of sequencing centers around the globe. These large-scale sequencing efforts generated a huge flow of data that was previously unprecedented and required advanced computational methods for analysis and informatics infrastructure for the data management. The unique coupling of these two fields paved the way for the successful completion of the initial draft of the human genome project in 2000. This accomplishment was achieved in only four years, over 10 years quicker than originally planned and arguably began the genomic era (Venter et al., 2001). Advancements in both fields are still following Moore's Law and have resulted in deep-sequencing technologies that have successfully sequenced the human genome in a fraction of the time and cost of the initial sequencing project. Deep-sequencing will no doubt exponentially increase the number of completed sequenced genomes and will spur on the rapid and continued growth of genomics (Rogers and Venter, 2005; Morozova and Marra, 2008).

Other notable advancements during this period, each with their own historical accounts, that were integral to the establishment of omic technologies were in the areas of chemistry and engineering. From a chemistry perspective, advancements in chemical fluorescence allowed for detection strategies to move away from radioactivity with increased specificity and sensitivity. This technology was not only important for modern sequencing technologies but has also played a vital role in labeling and detection strategies used by other omic technologies. Engineering advancements included the increased resolution, throughput, and specificity of various analysis instruments including microscopy, mass spectrometry, and nuclear magnetic resonance (NMR).

All of these advancements helped pave the way for the omics era. The core omics technologies are natural extensions on the study of the various levels of the central dogma of molecular biology. This central dogma involves the general organization and flow of information from DNA to RNA to protein (Crick, 1970). An additional level that can be added to this are the resultant metabolites that

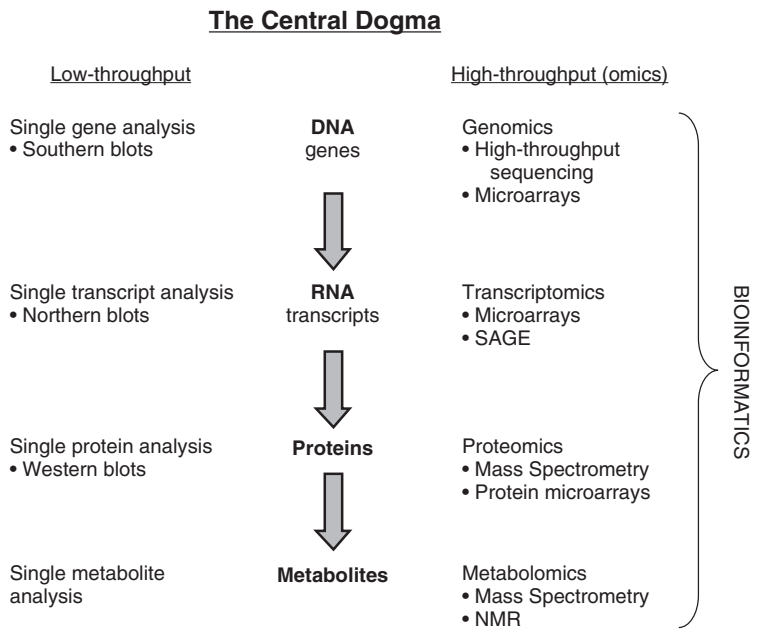

Figure 1.1 The relationship of the central dogma of molecular biology to omics technologies. The central dogma involves the organization and flow of information from DNA to RNA to proteins and the resultant metabolites. Developing an understanding of the interrelations of these levels is the focus of biological research. Omic technologies represent high-throughput and expansive approaches to the study of each of these levels. Bioinformatics is a key component to the interpretation and integration of the complex data-sets generated by omic technologies.

may be formed through this process and in turn may have biological function (Figure 1.1). Developing an understanding of the interrelations of these levels and how they lead to function in biological systems is the focus of intense biological research. At the level of DNA, research has included developing an understanding of DNA sequence and the organization of individual genes. In this regard the entire genetic makeup of an organism is commonly referred to as the genome, with the study of the entire genome referred to as genomics. Application of such a naming scheme to each level of the central dogma has given rise to the terms transcriptome (the complete makeup of RNA transcripts), proteome (the complete makeup of proteins), and metabolome (the complete makeup of metabolites), with the respective research in these areas referred to as transcriptomics, proteomics, and metabolomics. Due to the extensive nature of each of these complete "ome" levels, omics research as been facilitated by the development of high-throughput and/or high density analysis approaches, also known as omic technologies.

Genomic technologies are those that facilitate the wide-scale study of individual genomes. These include high-throughput sequencing technologies such as

those used for the sequencing of the human genome. This area also includes genotype analysis approaches used to detect sequence variations between individual genomes in a population, often referred to as single nucleotide polymorphisms or SNPs. Assessment of SNPs has been conducted using high-throughput sequencing approaches but has also been assessed through the use of microarray technology in which thousands of gene sequences can be profiled through complementary hybridization and detection on a microscope slide or "chip." More recently analysis of the genome has expanded beyond the assessment of genome sequence into areas that monitor modifications to DNA that influence gene expression, such as methylation, which is commonly referred to as the field of epigentics. Approaches to characterize and monitor the epigenome are still evolving and include both sequencing- and microarray-based approaches (Suzuki and Bird, 2008).

Transcriptomic technologies are those that monitor the RNA transcript population of cells or tissues. Research into the trascriptome benefited immensely from the sequencing of various genomes as this helped to define its full complement of transcripts. As with the genomic technologies, transcriptomics research has taken advantage of both sequencing- and microarray-based monitoring approaches. Sequencing-based approaches include serial analysis of gene expression (SAGE) in which count data are determined for individual transcripts through sequencing approaches and used to determine their expression level. Microarray-based approaches analyze the expression level of transcripts using microscope slides or chips that contain the entire transcriptome of an organism as immobilized sequences. The expression level of the transcripts in a biological sample is then determined by fluorescent labeling of transcripts and by allowing for complementary binding or hybridization to their respective sequences immobilized on the chip. Laser scanning then provides an indication of the fluorescence intensity for each transcript that corresponds to its expression level in the sample (Schulze and Downward, 2001; Boverhof et al., 2006). Such an approach represented the first widely available technique for omic profiling. Transcriptomic profiling has the advantage that the transcripts exhibit the same stability and characteristics under common conditions, which facilitates profiling of the entire transcriptome simultaneously. This ease of handling, processing, and profiling, as well as between broad availability of the technology from a number of vendors, has made transcriptomics the dominant omic technology.

Proteomic technologies are those that are used to study the full collection of proteins in a biological sample, the proteome. However, study of the proteome has been more difficult compared to that of the genome or transcriptome because of the increasing complexity that is observed with proteins. For example, proteins can undergo numerous post-translational modifications that affect their form, function and stability. In addition there is no approach for the amplification of proteins, as exists for nucleic acid sequences, so the limits of detection in the complex milieu of a protein extract can become problematic. Furthermore protein instability can make analysis difficult as adducts and other forms of protein damage can complicate proteomic analysis approaches. However, as proteins are the main functional units of biology, understanding the protein complement of a cell or tissue can provide

insights into biological function that cannot be gleaned from genomics or transcriptomics. The dominant technology applied to the study of proteomics has been mass spectrometry based analysis of peptide fragments (Boverhof et al., 2006).

Metabolomic technologies are applied to study a wide spectrum of small molecules in a biological sample that are the products of metabolic processes (Nicholson et al., 1999; Goldsmith et al., 2009). Because these metabolites reflect any alterations in gene and protein expression and activity, metabolomics allows for potential insights into the functional state of cells and tissues. Metabolomic technologies have most commonly monitored endogenous metabolite changes of biofluids such as blood and urine. The advantage of these biological samples is that they can be sampled noninvasively and therefore have ready availability for clinical applications. The key instrumentation employed for metabolomic analysis has included NMR spectroscopy and mass spectrometry (Nicholson et al., 1999; Goldsmith et al., 2009). Metabolome profiling shares some of the difficulties of proteomics analysis, which are due to the complex and diverse nature of the metabolites in the biofluds and to our lack of understanding of the extent and function of the entire metabolome.

Application of omics technologies to study biology can generate massive amounts of data that can be difficult to organize and analyze in a manner that will provide meaningful interpretation within the biological context. Therefore bioinformatics is an integral component of the application of omic technologies to the study of biology. Bioinformatics is defined as a branch of computational biology focused on applying advanced computational techniques to the collection, management, analysis, and interpretation of numerical biological data (NRC, 2007). This includes the statistical analysis of the data, the development of standards and databases for conduct and storage of these data, as well as the integration of disparate data sets within and across the different omics technologies. Bioinformatics is also vital for data analysis and interpretation as special tools and software are typically required to visualize large omics data sets, to mine data sets for biological meaning through interfacing with other functionally annotated databases, and for the development of complex mathematical algorithms and classification approaches (Schmidt, 2003). It is useful to think of bioinformatics as a component of omics research as opposed to a separate element. This is because the need for such analysis approaches is inherent in collection and interpretation of omics experiments.

The omics technologies listed above can be described as high-throughput and/or high-density approaches to the study of molecular biology as an extension of traditional lower throughput approaches. In the case of DNA, high-throughput profiling of SNPs can be considered analogous to running thousands of Southern blots. High-throughput transcript and protein profiling can be considered analogous to running thousands of northern and western blots. Metabolomics represents a multiparameter high-throughput approach that is analogous to current approaches that profile single blood or urine metabolites. Advancements in these areas have been facilitated by technological and computational advancements and encouraged by the success of the human genome sequencing efforts. However, such technological advancements

in data acquisition do not necessarily bring about the coordinated understanding of the biological meaning of the collected data. Therefore the continued focus of omic research is not only on the generation of data but on data mining and integration to decipher biological meaning, and this requires bioinformatic tools (Figure 1.1). Key applications of these approaches have been in the areas of cancer research and tumor characterization and in the development, progression, and detection of disease where genomic technologies are rapidly evolving as powerful tools for discovery and hypothesis-driven research.

1.4 TOXICOGENOMICS

As indicated previously, toxicology is a discipline that borrows and applies concepts and technologies from a number basic sciences to better understand toxic responses in biological systems. Therefore it would seem logical that toxicologists would seek to adapt the omic technologies as tools to better understand toxicity. Toxicogenomics is defined as the application of omic technologies, including genomics, transcriptomics, proteomics and metabolomics, to the study of adverse effects of environmental and pharmaceutical chemicals on human health and the environment (NRC, 2007). More simply put, toxicogenomics is the application of omic technologies and endpoints to toxicology.

The emergence of toxicogenomics into the discipline of toxicology has come with great expectations and is part of a growing movement of toxicology away from a descriptive and observational science toward a discipline that is based on developing a deeper understanding of toxic mechanisms of action. In its basic concepts the endpoints of toxicogenomics are not new to the discipline. Toxicologists have long used information from gene sequences and polymorphisms and changes in transcript and protein levels to study mechanisms of toxicity and as biomarkers of susceptibility and toxicity. What is offered by the expansive nature of the omics technologies is the ability to probe more deeply, and in a higher throughput fashion, into the complexities of gene-environment interactions and the responses of biological pathways and networks to chemical perturbations. Most important is that this can be done without prior knowledge of which pathways or systems may be perturbed, so new discoveries and hypotheses can be made in an accelerated fashion. The ultimate goal is the generation of information or "omic signatures" that are more informative, discriminating, and predictive than the current approaches of toxicology. Some of the specific applications include hazard identification via the development of omic signatures of toxicity, deciphering the mechanisms of action of toxicants, identification of biomarkers of exposure, toxicity and susceptibility, development of a better understanding of the relevance of cross-species (e.g., animal to human) extrapolations, characterization of the toxic responses to complex mixtures, and a more detailed characterization of dose–response relationships and thresholds to toxicant exposure.

Although the promise is great, effective and productive communication and collaboration remain critical to establishing validated approaches for the interpretation and incorporation of toxicogenomics into quantitative risk assessment.

Along these lines, the consortia of International Life Sciences Institute–Health and Environmental Sciences Institute (ILSI-HESI), the Toxicogenomics Research Consortium, the MicroArray Quality Control (MAQC) Consortium, the InnoMed PredTox effort, and the Predictive Safety Testing Consortium have addressed issues such as array reproducibility, best practices for assays and analysis, biological relevance of microarray results relative to traditional endpoints, and robustness of statistical models on diverse data sets (Mattes, 2008). It is likely that the application of toxicogenomic approaches will continue to evolve in an iterative fashion as all stakeholders gain further experience with the emerging technologies (Boverhof and Zacharewski, 2006).

REFERENCES

Borzelleca, J. F. (2000). Paracelsus: herald of modern toxicology. *Toxicol Sci* **53**, 2–4.

Boverhof, D. R., Burt, J., and Zacharewski, T. R. (2006). Introduction to toxicogenomics. In *Fundamental Toxicology* (J. H. Duffus, and H. G. J. Worth, eds.), pp. 127–141. Royal Society of Chemistry, Cambridge.

Boverhof, D. R., and Zacharewski, T. R. (2006). Toxicogenomics in risk assessment: applications and needs. *Toxicol Sci* **89**, 352–360.

Crick, F. (1970). Central dogma of molecular biology. *Nature* **227**, 561–563.

Eaton, D. L., and Gilbert, S. G. (2008). Principles of toxicology. In *Casarett and Doull's Toxicology: The Basic Science of Poisons* (C. D. Klaassen, ed.), pp. 11–43. McGraw-Hill, New York.

Gallo, M. A. (2008). History and scope of toxicology. In *Casarett and Doull's Toxicology: The Basic Science of Poisons* (C. D. Klaassen, ed.), pp. 1–10. McGraw-Hill, New York.

Goldsmith, P., Fenton, H., Morris-Stiff, G., Ahmad, N., Fisher, J., and Prasad, K. R. (2009). Metabonomics: a useful tool for the future surgeon. *J. Surgl Res*. **160** (1), 122–132

Lane, R. W., and Borzelleca, J. F. (2008). Harming and helping through time: the history of toxicology. In: *Principles and Methods of Toxicology* (A. W. Hayes, ed.), pp. 3–43. CRC Press, New York.

Mattes, W. B. (2008). Public consortium efforts in toxicogenomics. (Clifton, N.J.) *Meth Mol Bio* **460**, 221–238.

Moore, G. M. (1995). Lithography and the future of Moore's law. *Proc SPIE* **2437**, 7.

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

Nicholson, J. K., Lindon, J. C., and Holmes, E. (1999). "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**, 1181–1189.

NRC (2007). *National Research Council of the National Academies: Applications of Toxicogenomic Technologies to Predicitive Toxicology and Risk Assessment*. National Academies Press, Washington, DC.

Rogers, Y. H., and Venter, J. C. (2005). Genomics: massively parallel sequencing. *Nature* **437**, 326–327.

Schmidt, C. W. (2003). Data explosion: bringing order to chaos with bioinformatics. *Environ Health Perspect* **111**, A340–345.

Schulze, A., and Downward, J. (2001). Navigating gene expression using microarrays: a technology review. *Nature Cell Biol* **3**, E190–195.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B., and Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674–679.

Suzuki, M. M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev* **9**, 465–476.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., 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, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al. (2001). The sequence of the human genome. *Science* **291**, 1304–1351.