Genomics

DATA OVERFLOW
progress in automating and streamlining data

The genomic world moves faster and faster. The pool of data has exploded exponentially in the past few years due to the growing knowledge about the relevance of genomic findings for medicine and the environment. Like all other developing industries, the trend in genomics is about producing more in a shorter amount of time with a higher quality. What genomics researchers need today are tools to effectively and efficiently turn this data into meaningful information. The processing of results and the learning curve from these results need to be more efficient. Intelligent bioinformatics and flexible methods are the most crucial factors for success in science these days.

By Peer Staehler

Modern sciences from different fields learned a lot about the impact of genes on natural processes and the feasibility of learning from these processes for diagnostics, therapeutics, agriculture and cleantech in the last decade. The past 10-15 years have been characterised by using increasingly powerful instrumentation and data management. This is the result of an explosion and overflow of data, to a degree that it has become almost unmanageable. The acceleration and growing pool of genetic information – mostly not specified in function or usability, yet, has meant inefficiencies in the handling of the data through manual processes. The minefield of options and opportunities can turn into mistakes when it comes to statistical interpretation. The numerous manual processes are prone to human errors. The development of more automated and high-throughput technologies is the logical consequence like in all other fast developing industries.

Facilitated by automation: manpower and machine power work together to produce intelligence

In parallel with the demand for larger technical capacities, researchers today also need greater intelligence to filter and derive qualified findings. Automation of processes is helping to free the lab technicians and researchers from tedious tasks, allowing them to work on value-added activities, such as interpreting results and transforming them into new, target-oriented experiments. The learning curve will rise due to access to additional resources. Facilitated by automation, researchers turn utopist dreams to feasible tasks by saving time and cost. Basic science spurs a wide spectrum of applied research. Biological systems are influenced by a multitude of uncontrolled factors; hence the reproducibility of analysis is always an issue. It is important to have good reproducibility in order to verify a systematic and robust procedure.

Automating processes provide more uniform conditions and thus yield more reliable and comparable results.

The biotech industry is situated in a steady contest to provide the genomic market with more capable tools to work on the results. Evolution of new technological solutions seeks to provide a high degree of flexibility, while focusing on high-throughput and miniaturisation, optimising user interfaces and precision. Speed and efficacy are also key success factors.

Human Genome Project

Since the Human Genome Project began, it involved the development of highly efficient methodic tools. This project was expected to last 15 years, but it was completed in 13 years as a result of rapid technological advances. Technological advances not only saved time but also cost. The project was estimated to cost $3 billion, but ended up costing less, about $2.7 billion. The explosion and overflow of data for researchers in the genomics field makes it a prime example that there is the need to automate and streamline data.

The project's goal included identifying all the approximately 20,000-25,000 genes in human DNA and determining the sequences of the three billion chemical base pairs that make up human DNA. While the project is finished, the analysis of data is yet to be complete. There is an unprecedented volume of data on human chromosomes and tens of thousands of genes (many associated with genetic disorders) residing in them. The result of this project provides a magnificent and unprecedented biological resource that will serve as the basis for research and discovery and lead to practical applications. Growing areas of research will focus on identifying important elements in the DNA sequence responsible for regulating cellular functions and providing the basis of human variation. The analysis and management of data gathered will provide a deeper and more comprehensive understanding of the molecular processes underlying life.

Microarrays

Prior to the technological advancements researchers have today, researchers would study one or a few genes or proteins at a time. This has not been the most effective method since life does not operate in isolation; rather a systemic or much grander scale view is needed. Most companies have also focused on improving the design of the microarrays to hold more samples, but not much effort into automating the testing process,

with a few exceptions. Microarrays are a widely used tool for generating expression profiles on a genomic scale. New advances in bioinformatics and high-throughput technologies such as microarray analysis are allowing scientists to understand the molecular mechanisms underlying normal and dysfunctional biological processes. It has provided scientists with a tool to investigate the structure and activity of genes on a global scale. These new advances also provide scientists efficient automation and maximum flexibility.

Microarrays have allowed the creation of data sets of molecular information to represent many systems. It may flag genes expressed under particular cellular conditions and offer clues to gene function and regulation. Biological systems are influenced by a multitude of uncontrolled factors; hence the reproducibility of analysis is always an issue. These tools search systematically for interesting sequences and for mutations within the genome on a DNA or RNA level. The

Figure 1

Figure 2

Genomics

RNA level gives information about the expressed part of the genome using Gene Expression Profiling experiments.

High-density microarrays offer the possibility to screen the whole human genome on one biochip, such as the one offered by Agilent Technologies. Microarray technologies may also screen pre-fabricated biochips which are limited in their content. The Geniom® (**Figure 1** and **Figure 2** of the Geniom® RT Analyzer) (developed by febit) is an advanced microarray-based DNA analysis and synthesis system. Integrating microarray production, hybridisation, and detection, researchers have an automated process and when combined with bioinformatics software, researchers may design and perform microarray-based experiments using sequence information derived from public databases[1] or resulting from their own research covering non-published data even from newly discovered organisms.

Today's advances with microarray analysis provide efficient automation and maximum flexibili-

ty to researchers. Simplifying and automating the processes through microarray technology that combines microfluidics and *in situ* oligonucleotide synthesis for reading, writing and understanding the code of life is a solution for many genomics researchers these days. The impact of microarray technologies has proven tremendous because it has enabled researchers to progress from studying the expression of one gene in several days to hundreds of thousands of gene expressions in a single day[2].

This also becomes possible because of the deployment of two additional wild cards of modern biotechnology: multiplex experiments and the use of microfluidics. The modern biochip (**Figure 4**) allows high end analysis testing of up to eight samples in parallel to display all sequences the researcher might be interested in (**Figure 3**). The microfluidic channels provide a closed environment like a reaction tube – even the use of enzymes within the microarray experiment will become possible within the near future.

Next-generation sequencers

The growing capacity of the new sequencer technology microarrays will also be used in the future as a pre-selective tool for sequence capture. This reduces considerably the amount of DNA which has to be sequenced due to well directed research.

The development of so called next-generation sequencers brought down costs from $3 billion for the first human genome to $5 million using high-end services from GATC Biotech. The cost may be brought down even further for human genome sequencing, by using next-generation sequencing and high-end bioinformatics to make this a more realistic option for pharmaceutical research. High throughput DNA sequencers are a powerful way to understand how organisms evolve and adapt looking at the whole genomes. On a DNA level you can see how healthy and diseased organisms differ or how bacteria mutate to develop drug resistance.

Next-generation sequencing differs in that it provides an unprecedented amount of genomic information, but it comes to the same endpoint that researchers will need to know how to analyse and manage the data – turning it into meaningful information.

microRNA research and the impact of modern technology

Only a few years ago a completely unknown class of regulatory molecules rose like a phoenix from the ashes: microRNAs (miRNA), formerly despised as junk-RNA, little RNA molecules from only a few nucleotides of lengths have emerged as being important post-transcriptional regulators of gene expression, which they do by partial or full base-pairing with their target RNA. Initial studies indicate that miRNAs may regulate as much as 30% of all genes in the human genome[3]. Their presence is thought to be influential in cellular developmental pathways, proliferation, apoptosis and differentiation. Since then a large number of different non-protein-coding RNAs (ncRNA), like miRNA, snoRNA, piRNA, siRNA have been discovered. Non-protein-coding RNAs form a massive hidden network of regulatory information that directs the precise patterns of gene expression during growth and development. What has been dismissed as junk because it was not understood may well hold the key to understanding human complexity, as well as our idiosyncrasies and susceptibility to common diseases[5]. With miRNA research still in its infancy the door is open to the potential of innovative discoveries in basic research, diagnostics and therapeutics centred on these important molecules.
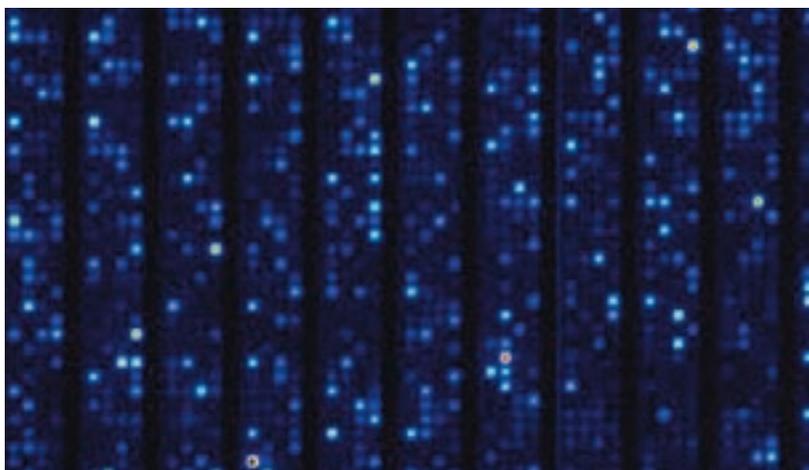

The miRBase, from the Welcome Trust Sanger Institute, reflects the development of the miRNA research: starting with only few sequences, the database contains in its current version 6,396 sequences of RNAs. The next version is expected soon. Research tools which can display these sequences immediately after their release are rare but already on the market, including the miRNA-Services at febit. Newly published sequence information on microarray biochips are available because of the on-demand production of biochips.

Figure 3

Figure 4
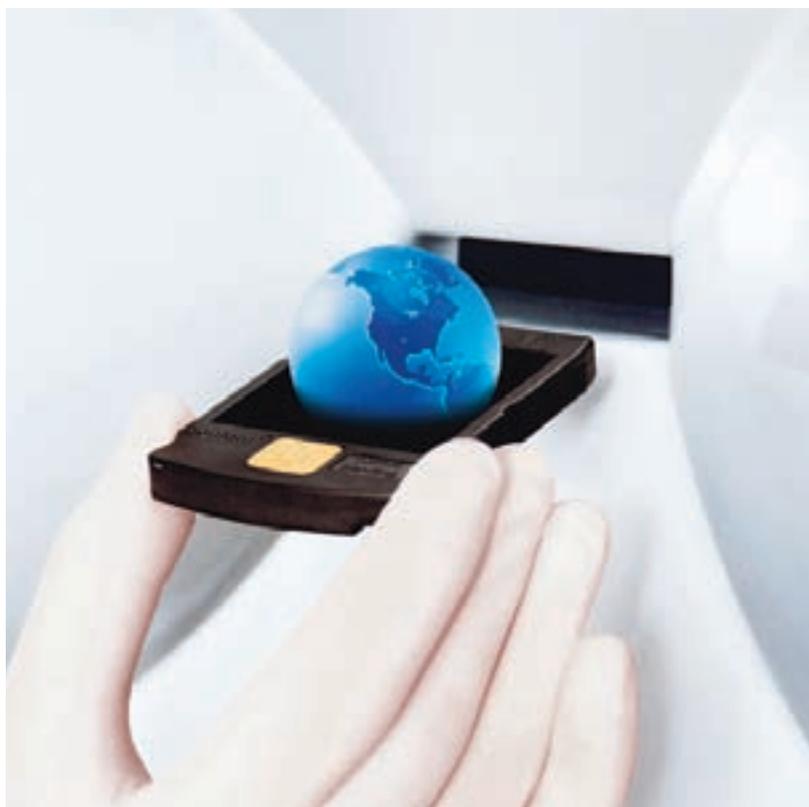
Genomics

References
1 Baum, M et. al. Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. Nucleic Acids Research, 2003, 31: 151.
2 Afshari, CA. Perspective: Microarray Technology, Seeing More Than Spots. Endocrinology, 2002, Vol. 143, No. 6: 1983-1989.
3 Vorwerk, S et al. The Geniom® One Platform Provides an Ideal Tool for High-Throughput Analysis of small non-coding RNAs. National Genome Research Network, November 10-11, 2007. Poster Session.
4 Meldrum, D. Automation for Genomics, Part Two: Sequencers, Microarrays and Future Trends. Genome Research, 2000, Vol. 10, Iss. 9: 1288-1303.
5 Mattick, J. The Human Genome as an RNA Machine. febit Science Lounge Webinar, June 12, 2008, www.febit.com.

Advances in new technology impacts genomics research and other fields

In addition to providing maximum automation, flexibility will also be an important factor for future laboratory devices. The amount of data raised by high throughput technologies will be followed by an increasing number of possibilities to use the results. And the human genome will be only one of the future focuses.

New pathogens for example are discovered each day. A human mutant of the bird-flu virus, H5N1, is expected every year, and the fear from artificial pathogens released by terrorists is still present. Scenarios like this stir up the demand for capable, but also flexible, technologies to detect new organisms within a short time.

The relevance for environmental research is growing as well: dwindling resources of oil has led to the development of new energy resources such as algae. Environmental pollution should be embanked by the use of newly designed bacteria. This leads to another new way of using the immense amount of data drawn from genomics: synthetic biology. Instead of using time consuming conventional cloning, new biological systems can be designed using a kind of biological lego. Bricks of genomic information with known functions can be newly combined to provide a complementary perspective to review, analyse and understand the living world.

Summary
New technological solutions help to automate and streamline data through its high degree of flexibility, speed and efficacy. The future for microarrays is bright for high-throughput screening of compound targets, diagnostic development and drug development. At the tremendous pace of genomics data output, researchers also need to adopt and implement the technological advances to support automation of processes, which will allow them to work on other value-added activities, such as interpreting results and transforming them into new, target-oriented experiments. New technologies for genomics are an interdisciplinary effort through contributions from various fields. Automating and streamlining processes provide an increase in efficiency, quality and reliability of data and intelligence, helping to reduce overall cost of research and development[5]. **DDW**

Peer Staehler is the co-founder and chief scientific officer at febit holding gmbh. Peer has a master's degree in molecular biology from Konstanz University with the main focus on genetics. As a research scientist at Max Planck Institute for Brain Research (MPIH) in Frankfurt/Main, Peer was involved in different DNA analysis projects in Europe and in the United States. Peer is an expert in the biochip sector and benefits from his many years of experience in the biotechnology industry.