Next-Generation Data Analysis Software Helps Researchers to Understand the Link Between Genetics and Disease

Although modern gene expression studies are proving invaluable to the study of human biology, the amount of data that is produced by this kind of research is enormous. As a result, it is impossible to derive any real biological meaning from these findings unless sophisticated analysis methods are used to help interpret the data effectively.

To address this issue, a team of scientists at the Institute of Human Genetics of the Christian-Albrechts-University in Kiel, Germany, is currently using sophisticated data analysis software on a number of national and international projects in order to study the epigenetic alterations related to several cancers, including malignant lymphoma, colorectal cancer, and hepatocellular carcinoma, as well as developmental disorders and other diseases.

"Fortunately, new technological advances in this area are making it much easier for scientists to compare the vast quantity of data generated by epigenetic studies, to test different hypotheses, and to explore alternative scenarios within seconds"

Author Details:

Carl-Johan Ivarsson
President of Qlucore
Tel: +46 46 286 3110
Email: carl-johan.ivarsson@qlucore.com
Web: http://www.qlucore.com

The team of scientists at the Institute of Human Genetics of the Christian-Albrechts-University in Kiel, Germany, led by Dr Ole Ammerpohl, is currently studying the raw data obtained from array-based DNA methylation analysis, including Illumina's HumanMethylation27k Bead Chip, which provides semi-quantitative data from more than 27,000 CpG loci. Along with his team of four scientists and three technicians, Dr Ammerpohl, is examining how alterations in the DNA methylation pattern contribute to a specific phenotype or disease like cancer, as well as the way in which tumor induction introduces epigenetic alterations, which support tumor growth and progression.

"Because of the amount of data being analysed, conducting microarray analysis has always been a hassle," said Dr Ammerpohl. "Larger studies, especially those which include multiple samples that need to be analysed on comprehensive array platforms, have traditionally been very time-consuming, and have also required a considerable amount of computer power. Plus, in most cases, it is mandatory to define the exact conditions before beginning the intended analysis, and just this process alone can take hours. Worse still, any recalculations – even if they are just minor adjustments to these initial conditions – would mean that the whole process would have to be restarted from the beginning."

Fortunately, new technological advances in this area are making it much easier for scientists to compare the vast quantity of data generated by epigenetic studies, to test different hypotheses, and to explore alternative scenarios within seconds. As a result, the latest generation of data analysis software is helping scientists to regain control of this analysis, and to realise the true potential of research in this area.

Unlocking the Mysteries of Human Disease

Epigenetics, the name given to the study of changes in phenotype (appearance) or gene expression caused by mechanisms other than changes in the underlying DNA sequence, is helping to provide a link between the (static) genome and the unsettled environment, and therefore to establish a gene expression profile that will enable the survival of the cell in certain environmental conditions.

Epigenetic variations are involved in many processes, starting with transgeneration heritable effects, but also in areas like stress tolerance, behavior, psychological problems, drug addiction, and severe developmental disorders and diseases. Since epigenetic modifications are more responsive to drug treatment than genetic alterations, understanding epigenetic alterations might provide new therapeutic options in the future.

"The methylation of the cytosine residue in a CpG dinucleotide is probably the best characterised epigenetic modification in the human epigenome," explains Dr Ammerpohl. "DNA methylation is essential for normal development; it is not only a key player in gene expression control and parental imprinting, but it also assures genomic integrity by preventing repetitive sequences from recombination. At the same time, it also silences parasitic sequences, such as retroviral sequences which have been integrated into the genome."

Making Sense of Important Data

Until now, most of the software that has been designed to study areas like array-based DNA methylation has mainly focused on the ability to handle increasingly vast amounts of data, which means that the role of the scientist/researcher has been largely set aside. As a result, a lot of data analysis has been passed on to bioinformaticians and biostatisticians.

Campus in Kiel, Germany

Photo courtesy of The University Medical Centre, Schleswig Holstein.

A new generation of data analysis software is helping to redress the balance, however, and is already playing a key role in unveiling important new discoveries, since it allows the actual researchers involved to study the data and to look for patterns and structures, without having to be a statistics or computer expert.

At the same time, the overall performance of data analysis software has been optimised significantly over the past three years. With key actions and plots now displayed within a fraction of a second, researchers can increasingly perform the research they want and find the results they need instantly.

"For us, a big advantage of the latest data analysis software is its speed," said Dr Ammerpohl. "We are using an application called Qlucore Omics Explorer, and this software makes it very easy to assign samples to defined groups, to change the applied statistical methods, to create new groups, and to modify the thresholds for items such as variance and p-value in real time, with results returned immediately."

New Software Makes Data Analysis Easier

The latest data analysis software is now making it possible for scientists to analyse proteomic, genomic and microarray data with a combination of statistical methods and visualisation techniques such as Heat maps and Principal Component Analysis (PCA).

As a result, scientists studying DNA methylation analysis and other genomic data can now analyse all of this important information in real-time, by themselves, directly on their computer screen, since the software can provide instant user feedback on all actions, as well as an intuitive user interface that can present all data in 3D.

By using data analysis software in this way, Dr Ammerpohl is able to apply different statistical approaches, and to keep track of the effects in a PCA or cluster analysis. As such, subgroups in the sample collection – comprised of specific groups of genes – can be identified intuitively. All relevant statistics (together with the corresponding variable and sample list) can be exported, as well, so that they can be easily integrated into publications or presentations.

Key Data gets the 3D Treatment

Data analysis software like Qlucore Omics Explorer can generate PCA-plots between various sample data interactively and in real time, directly on the computer screen, and work with all annotations and other links in a fully integrated way, all at the same time. This approach

has helped to open up new ways of working with data analysis and, as a consequence, has helped the biologists to be more actively involved in the analysis process.

Qlucore Omics Explorer, for example, graphically represents the high dimensional data in form of 3-dimensional plots on the computer screen. This instant visualisation technique is then combined with powerful statistical methods and filters, all of which are handled with just a single mouse-click.

"As humans, we are all used to interpreting 3D pictures in our environment, and so our brain is able to find structures in complex 3D figures very quickly. Therefore, it's no wonder that a 3D presentation of complex mathematical/statistical coherences makes its interpretation much easier for us," Dr Ammerpohl added.

A Step-By-Step Approach to Data Analysis

For their research into array-based DNA methylation, Dr Ammerpohl and his colleagues have compared the DNA methylation pattern in normal liver tissue, in cirrhosis of the liver, which is thought being a precursor of hepatocellular carcinoma (HCC), and in HCC. In addition to the DNA methylation values obtained from the array analysis, the researchers have also included information on the tissue, clinical features or exogenous exposures like viral infection of the patient.

Afterwards, the analysis of the data could easily be performed. By selecting the appropriate test and adapting the applied statistical thresholds for p-value, false discovery rate, or the minimal variance by using sliders in the Qlucore software, groups of samples with similar characteristics could be identified easily by PCA.

Furthermore, a PCA of the variables (in this case, the DNA methylation values of the CpG loci) or a hierarchical cluster analysis is available. In the particular study, the team of researchers identified genes which acquire epigenetic alterations already in cirrhosis, which are maintained in HCC, or genes acquiring epigenetic alterations in HCC exclusively.

The Future

Dr Ammerpohl has been using Qlucore Omics Explorer to analyse numerous epigenomes of tumour entities including colorectal cancer, HCC, lymphomas as well as of developmental sex disorders or imprinting diseases. These studies have already resulted in valuable data that is helping Dr Ammerpohl and his team in their efforts to understand the epigenetic background of these diseases or disorders.

"The exceptional speed that Qlucore's software provides is very important for us. The fast analysis of the data highly contributes to the identification of subpopulations in a sample collection or a list of variables," said Dr Ammerpohl. "Without a doubt, these rapid results – and the way in which they are presented – prompted us to perform analyses that we would have never performed otherwise."

"We don't want to give away too much at the moment, as many of our major findings have yet to be published, but I can report that we are feeling very positive about our research in this area, and about the future of the entire project," he added. "It is fair to say that we have already obtained some very interesting results."

Fastest Available DNA Methylation Kit Speeds Epigenetic Mapping

Merck Millipore announced the introduction of a rapid kit for bisulfite conversion, which is a critical first step in mapping the differences in genomic DNA methylation patterns. Aberrant DNA methylation can result in inappropriate activation or silencing of specific genes and is associated with errors in embryonic development as well as the onset of diabetes, cancer, cardiovascular, and other diseases.

The new CpGenome™ Turbo kit reliably converts unmethylated cytosines to uracil in 90 minutes, which is twice as fast as commonly used bisulfite kits and reagents. By increasing the throughput of bisulfite modification, the CpGenome Turbo kit allows researchers to examine DNA from more samples in less time.

The CpGenome Turbo kit features a proprietary conversion reagent that reduces incubation times while retaining high efficiency, converting more than 99.9% of unmethylated cytosines to uracil. The kit is also highly sensitive, requiring as little as 1ng of DNA.

The resulting, modified DNA is pure and ready for epigenetic mapping by sequencing, microarray hybridisation, or quantitative PCR. To facilitate PCR analysis, Merck Millipore offers more than 20 different gene-specific CpG WIZ® kits containing the necessary reagents for PCR amplification of bisulfite reactions.

Circle no. 520

Simplifies Solubilisation and Purification of Membrane Proteins

GE Healthcare's Membrane Protein Purification Kit is designed for rapid and easy detergent screening of histidine-tagged membrane proteins, enabling simultaneous evaluation of solubilization and purification efficiency. Analysis and evaluation can be performed by a number of methods such as Western blot, gel filtration, or light scattering.

The Kit contains seven detergents in ready-to-use solution (purity ≥ 99% by HPLC analysis), His Mag Sepharose™ Ni, buffer stock solutions, and an easy-to-follow protocol, and is sufficient for two complete screening experiments.

Integral membrane proteins play major roles in fundamental biological processes, and are common targets in the development of pharmaceuticals. To be studied, these proteins must be dispersed in an aqueous solution. To avoid protein loss and deactivation during solubilization and concomitant purification of integral membrane proteins, the choice of detergent is a key factor. Therefore, a detergent screen is often necessary to determine the optimal detergent for each protein and purpose.

His Mag Sepharose Ni are magnetic beads designed for rapid, small-scale purification and screening of histidine-tagged proteins from different sources. The high density beads allow rapid capture by magnetic devices, while the visibility of the beads permits reliable collection of the bound histidine-tagged membrane proteins in the screening and purification procedure. Purification of target membrane proteins for a number of analytical or other downstream applications can be achieved within a few hours.

Circle no. 521