Next-Generation Sequencing & Bioinformatics

New technologies for rapid sequencing of DNA and RNA, with solutions in bioinformatics to handle a flood of genomic data, pave the way for research and clinical breakthroughs.

Our life code, DNA, contains some 22,000 genes made up of three billion pairs of letters. Reading this 'book' is the key to a vast treasure trove of answers to questions about our health. Modern sequencing technologies are increasingly bringing these answers to light for patients and doctors. Even so, obstacles to routine clinical use of next-generation sequencing remain – including high costs, fragmented workflows and the difficulty of analyzing and interpreting high volumes of complex genetic data.

In recent years a deeper understanding of our genetic makeup has been raising expectations among scientists, doctors and patients. The hope is that new and effective approaches to detecting and treating diseases such as cancer or Alzheimer's can be developed as the genetic code – including specific components of DNA, RNA and proteins – becomes known through advances in technologies. These tools also have applications in fields such as forensics.

Next-generation sequencing (NGS) technologies have given great impetus to this scientific progress. In NGS, human DNA is broken down into millions of tiny fragments that are read simultaneously down to the last base. Sequencing applications have become much faster and less expensive: While reading DNA cost more than $5,000 per 1 million raw elements (bases) in 2001, determination of a sequence of letters of the same size is now possible for less than 10 cents, and the cost of sequencing an entire genome is approaching $1,000. The process takes only a few days rather than several years as before.

The potential uses of the NGS technologies are spreading from purely academic research to clinical applications, such as the diagnosis of rare diseases or the reading of gene mutations to determine the best treatments for each patient. Market observers see a golden future for NGS applications: The market for NGS-related products and services – currently around $1 billion – is expected to rise to as much as $7 billion in the next decade.

1 Costs of sequencing a human genome have fallen dramatically since the introduction of NGS in 2005. Source: NIH 2015.

2 At the same time, value creation is likely to shift increasingly towards analysis and interpretation of sequenced data.
Further NGS expansion hampered by complexity

While next-generation sequencing has the potential to help laboratories create valuable insights, many barriers to adoption persist – including high costs, fragmented workflows and challenges in analyzing and interpreting the data. Thus, there is an urgent need for a simpler, more cost-effective and efficient way for laboratories to take advantage of NGS technology and improve outcomes.

To achieve the potential of NGS, users see a clear need to overcome major challenges in the healthcare system that are currently impeding the spread of NGS technology in clinical practice. The issues stem from the fact that existing NGS systems were primarily developed as tools for academic research, where user requirements are entirely different. Diagnostics experts are particularly skeptical of the lack of standardization, automation and flexibility of conventional NGS systems, as well as the immense complexity of the subsequent data analysis.

Comparison to established platforms such as QIAGEN's QIAsymphony RGQ shows just how much the existing NGS systems lag behind in usability for widespread routine use in clinical applications. The modular real-time PCR platform is easy to use, supports continuous loading of samples and allows users to perform several procedures on one batch of samples without reloading reagents. It also tracks and logs samples through the workflow and allows electronic transfer of testing results into the laboratory information system via a dedicated software interface.

By contrast, the workflow of existing NGS systems is divided into many more individual steps, which often require manual intervention and are therefore susceptible to operator errors. To ensure cost-efficient use, patient samples generally need to be gathered first, specially marked at the molecular level and then processed together. In this time the systems are blocked and cannot process any more samples.

Data overload complicates analysis and interpretation

By far the biggest challenge for laboratories is the processing, analysis and interpretation of the raw data generated by NGS sequencing. Since the advent of NGS technologies in 2005, the volume of digital sequence information generated each year has risen exponentially and is expected to reach the equivalent of 1.5 billion DVDs by the end of this year. Typically, sequencing data is processed in three phases: After the signals registered by the NGS device have initially been translated into digital information (primary analysis), the DNA fragments encoded must be merged into a connected sequence and analyzed for variants in relation to a human reference genome (secondary analysis). In the third and final step (tertiary analysis), the identified variants are interpreted in the context of a specific clinical picture. The insights from that interpretation must then be reported in a concise format to healthcare professionals.

What sounds relatively easy poses major challenges for laboratories due to the vast abundance and complexity of the data generated by NGS. The sequence information of a whole human genome equates to some 3 billion data points. However, reading errors inevitably arise during
sequencing, which is why – depending on the application – the necessary genetic information is read out more than 1,000 times in parallel, and the results are statistically averaged to determine the DNA sequence. Doing this for diagnostic purposes for a single human genome is not possible, even with the current state of the art. But even if only individual genes or gene combinations are examined, as is common in clinical practice, the technical complexity is immense.

When this process is successfully completed, all identified gene variants – often several hundred thousand in a complete genome – must be examined for relevance to the clinical situation and interpreted in order to ultimately condense the results to a few variables and derive recommendations for action for the treating doctor. To this end, the dataset must be compared with the most recent results of clinical and scientific studies, as our knowledge of the genetic foundations of many diseases and the interactions between individual gene variants is expanding rapidly.

In extreme cases, interpretation can take several months if each gene variant needs to be researched and evaluated manually. Another challenge is that new variants or combinations of variants with unknown clinical relevance are constantly emerging. NGS users may toil away with fragmented, uncoordinated software solutions that can only be used effectively with considerable expertise in bioinformatics. In diagnostic laboratories, however, time is of the essence and core competencies generally lie in areas other than bioinformatics.

**QIAGEN's next-generation sequencing initiative**

To overcome the obstacles that plague many NGS workflows today, QIAGEN has developed the GeneReader NGS System. The world’s first complete Sample to Insight NGS solution provides a simpler, more cost-effective way to take advantage of NGS technology, covering all steps of a NGS workflow from primary sample to a final report.

The GeneReader NGS System, which is currently available for research use only, includes QIAGEN’s new Actionable Insights Tumor Panel, the first member of the family of GeneRead QIAact Panels, which targets 12 clinically relevant genes of interest that are often analyzed in most prevalent types of cancer, including breast, ovarian, colorectal, lung and melanoma.

Both the platform and the panel integrate seamlessly with QIAGEN Clinical Insight (QCI), the bioinformatics solution for clinical labs interpreting and reporting on genomic variants identified using NGS. QCI supports labs by increasing the accuracy and speed of analyzing and interpreting NGS results.

To add further value for labs that adopt the GeneReader NGS System, QIAGEN plans to introduce additional cancer-related gene panels, broadening the coverage in oncology research, and also expects to expand the NGS content menu to indications beyond that medical branch. The platform also will evolve, as necessary, to address new applications and market segments, with current development programs targeting throughput expansion, increased output per flow cell, and an expanded range of sample types, including non-invasive liquid biopsies. QIAGEN bioinformatics solutions are continuously updated, including updates of the Knowledge Base.
In addition, QIAGEN markets a broad range of universal research products compatible with any major sequencing platform that addresses key bottlenecks of existing NGS workflows, including dedicated solutions for nucleic acid extraction, target enrichment, library preparation, and data analysis and interpretation.

With its current NGS solutions and ongoing innovation, QIAGEN is addressing major weak points of existing procedures, and helping users in clinical research to use NGS technology effectively and efficiently. By doing this, QIAGEN is playing a key role in unlocking the wealth of data contained in the human genome to improve the future of healthcare.

---

1 Leerink Swann 2013: The Future of DNA Sequencing
3 Cisco Systems, GenBank and Alliance Bernstein.