

Feature Articles : Jun 15 2010 (Vol. 30, No. 12)

## Next-Generation Sequencing Moves to Next-Next Level

As Technology Proves Its Mettle and Advances Expand Usage, Applications Are Proliferating

Richard A. Stein, M.D., Ph.D.

The emergence of next-generation sequencing marked a historical moment for biomedical and social sciences, one that promised to settle long-standing questions and, concomitantly, opened a new set of challenges. Sequencing of the human genome revealed that unrelated individuals are more similar to each other than previously thought. At the same time, the large number of single nucleotide polymorphisms and copy number variants that were unveiled, opened the need to sequence individual genomes more reliably and in a more time-efficient and cost-effective manner.

Recent genetic and genomic advances revealed that mutations in one of several individual genes are often linked to the same condition or group of conditions, and new associations are constantly being unveiled.

At Select Biosciences' "Genomics Automation Congress" held recently in Boston, Birgit Funke, Ph.D., instructor of pathology at [Harvard Medical School](#) and associate director at the laboratory for molecular medicine at the [Partners HealthCare Center for Personalized Genetic Medicine](#), talked about testing for mutations associated with cardiomyopathies and using next-generation sequencing to implement new tests that would examine many of the genes known to harbor disease-associated changes.

Several classes of cardiomyopathies exist, and there are currently between 30 and 50 genes that have been individually linked to this group of conditions. "Testing for each of these genes using traditional Sanger sequencing technology is starting to be difficult to do in a cost-effective manner," explained Dr. Funke.

The laboratory for molecular medicine already expanded the number of tested cardiomyopathy genes by implementing a novel sequencing platform in 2007. Testing for all genes by next-generation sequencing represents the next goal that is already known to be technically feasible and, if introduced in the clinic, would offer great diagnostic benefits.

### Epigenetics

In addition to being a powerful tool to provide a vast amount of quantitative information about the genome, next-generation sequencing promises to re-shape the study of epigenetic modifications. "There is no doubt that next-generation sequencing is going to change epigenetic research measurements, and I suspect that in the next few years, everybody will be using next-generation platforms to explore DNA methylation and histone acetylation and modification," said Jean-Pierre Issa, M.D., professor in the department of leukemia at the [University of Texas MD Anderson Cancer Center](#) and co-director of the center for epigenetics.

Dr. Issa's group uses next-generation sequencing to study DNA methylation and histone-modification changes that occur during aging or in response to therapeutic agents used in cancer treatment. "Next-generation sequencing provides an accurate, deep, and quantitative approach to study aging, and

enables us to survey the genome to find the exact regions and genes that are affected and obtain information that was never available before.”

Recently, Dr. Issa and collaborators reported that DNA methylation can be used to predict progression-free survival in patients with myelodysplastic syndrome.

High-throughput DNA sequencing approaches promise to help the diagnosis and guide treatment decisions in many diseases and in many patients, and the combined analysis of the cancer genome and the epigenome promises to become a powerful diagnostic and therapeutic tool. “It is clear that the best way to study histone modifications across the genome is by next-generation sequencing. In terms of biology, this has been a revolution that transformed the field the same way as PCR did.”

## Electron Tunneling

At the meeting, Stuart Lindsay, Ph.D., the Edward and Nadine Carson professor of physics and chemistry at [Arizona State University](#), talked about recent advances that his group made in developing label-free sequencing by electron tunneling. It is known that, at moderate electrical fields, charged molecules such as DNA can translocate through single-wall nanotubes of approximately 2 nm in diameter, and the current generated by the nucleotide diffusion through the pore is sensitive to sequence, but the signal is averaged over several bases. An alternative is to measure the localized current owing to electron tunneling across the DNA.

Electron tunneling, a quantum mechanical effect, can be confined to a space as small as a single base. This powerful technique presents several challenges such as the interference by contaminating molecules and fluctuations in the current that are caused by changes in atomic positions.

To address these shortcomings, Dr. Lindsay and collaborators chemically functionalized one electrode with a reagent, 4-mercaptobenzoic acid, which binds the gold electrode through the sulphur atom to form monolayers, while its benzoic acid moiety is directed toward the solvent and forms hydrogen bonds with the passing nucleotides.

The coupling that results is sufficiently weak to allow DNA to pass through the tunnel, but at the same time it is strong enough to restrict the molecular orientation of the nucleotides and generate a signal of increased selectivity with higher signal-to-noise ratio.

As DNA enters the electron-gap tunnel, the current intensity indicates which nucleosides are passing and, as Dr. Lindsay and colleagues revealed, a distinctive signal can be generated for each of the four bases. “What we found is that we can generate distinct tunneling signatures for each of the four bases, and we can read the identity of a base with one read, which is pretty significant,” explained Dr. Lindsay.

## Environmental Influences

Next-generation sequencing provides a powerful tool to dissect the influences that environmental perturbations exert on the genome. In what became the first catalog of somatic mutations from a human cancer genome, Peter J. Campbell, Ph.D., group leader at the [Wellcome Trust Sanger Institute's Cancer Genome Project](#), and collaborators recently completed the sequencing of malignant melanoma and lymphoblastoid cell line genomes derived from the same person and, in a companion study, they examined the genomic mutational burden that accumulates in smokers who develop small-cell lung cancer.

“We set out to use next-generation sequencing to examine genomic changes at a detail that was never looked at before, and we wanted to identify the vast majority or all mutations present in those cancers,” explained Dr. Campbell. The causes of these two cancers were relatively well characterized—melanoma is linked to sunlight, while lung cancer is caused by exposure to cigarette smoke.

“We wanted to see if we are able to visualize the records of those exposures in the genome. And indeed that is what we found,” said Dr. Campbell. The

investigators identified tens of thousands of mutations in both cancers, and the vast majority provided a hallmark that can be used to establish a direct link between exposure and mutations in the respective cells, in addition to signs of DNA repair processes that were functional in the respective cells.

For example, over 22,000 somatic mutations, of which 134 were located within coding exons, were identified in small-cell lung cancer, providing a powerful signature that cigarette smoke leaves on the genome. In addition, the authors showed that over the lifetime of a cell clone that eventually becomes cancerous, one mutation would accumulate for approximately every 15 cigarettes smoked.

Next-generation platforms allow genetic variation to be characterized in a high-throughput way, at a fraction of the cost that we were previously used to. "In the future, patients with cancer and other disorders can have their entire genome sequenced with the aim of understanding what is going on in that individual patient to cause that disease, and hopefully that will lead us to much better targeted therapeutic approaches," added Dr. Campbell.

## Computational Challenges

One of the major challenges when examining whole-genome sequences is that particularly novel sequence insertions, by definition, are missing from the reference human genome, and their omission can lead to incomplete genome analysis.

Furthermore, many previous studies on structural variation avoid chromosomal repeat regions, despite the fact that these repeats constitute about 40% of the human genome. In addition to the large amount of information that can be missed by not analyzing these regions, it is also known that they harbor structural alterations more often than the rest of the chromosome.

"I think it is important to understand complex repeat regions in the chromosomes and to improve the algorithmic knowledge, because if we don't have the right methods, with the right mathematical tools behind it, it is difficult to arrive at the right conclusions," said S. Cenk Sahinalp, Ph.D., professor of computing science and director of the laboratory for computational biology at [Simon Fraser University](#).

To define the location and content of novel genomic sequence insertions, Dr. Sahinalp and collaborators recently developed and validated NovelSeq, a computational framework that uses sequencing data generated by next-generation sequencing platforms. By using this tool, the authors unveiled recurrent structural variants in several cell lines and cancer tissues from patients. "I really encourage researchers not to ignore repeat regions because there is a lot of interesting data in them. The lack of appropriate algorithmic tools can really hurt the progression of the field," he emphasized.

Dr. Sahinalp and collaborators are currently working on making the software even more user friendly and developing a web interface to enable more users to analyze and interpret their data.

## Mechanistic Details

An important application of next-generation sequencing methods is to understand the mechanism of action of small molecule drugs. An informative model system is provided by the budding yeast, which is inexpensive, easy to manipulate, and represents a well-characterized and powerful genetic tool. In addition, a complete deletion collection, where every single gene from the genome has been replaced with a unique barcode identifier, is available. It allows massively parallel experiments to be conducted by growing all the mutants simultaneously.

The consequence of a specific perturbation such as a therapeutic agent can be examined by PCR amplifying the barcode sequence and hybridizing the products to a barcode microarray that is complementary to the unique identifiers. By using this approach, Corey Nislow, Ph.D., and Guri Giaever, Ph.D., both assistant professors in the department of molecular genetics at the [University of Toronto](#), and collaborators, recently conducted massively parallel experiments and revealed that, in addition to the ability to interrogate 6,200 different mutants in one experiment, multiplexing is also associated with a great reduction in costs.

“In a sense, we are using next-generation sequencing as a simple and powerful molecular counter,” revealed Dr. Nislow. However, the quantity of data that is generated is not the only benefit, and this approach is attractive for several other reasons. “Instead of a time point, it is now possible to perform an entire drug titration, generate a complex time course, or even look at treatment combinations.” This aspect is significant, because it is becoming increasingly clear that most therapeutic agents are more effective in combination than as single agents.

In addition, Dr. Nislow’s group collaborates with Jason Moffat, Ph.D.’s lab at the University of Toronto to conduct a similar type of screening that uses mammalian cells in which, instead of deletions, gene function is knocked down by shRNA. “We are taking what we are learning from yeast and applying the findings to mammalian cell-based assays,” explained Dr. Nislow.

As these recent developments reveal, next-generation sequencing is witnessing exciting times. The ability to survey the extensive inter-individual genomic variations, understand the complex interplay between genetic and epigenetic modifications, and dissect the response to therapeutic agents are only some of the applications that place this technology at the forefront of clinical and research laboratories, where it promises important prophylactic, diagnostic, and therapeutic benefits.