

[Send to printer »](#)

## Genetic Engineering & Biotechnology News

---

Feature Articles : Jan 1, 2016 (Vol. 36, No. 1)

# The Next Next Thing in Sequencing

## Next-Gen Sequencing Is Preparing To Navigate New Molecular Codescapes

*Kathy Liszewski*

Though it may seem to be navigating by perceptibly unfixed stars, next-generation sequencing (NGS) is journeying ever more adventurously into the obscure, the rare, and the confoundingly heterogeneous domains within life's molecular codescapes.

NGS is already capable of producing billions of short reads, and it can do so quickly and economically. And NGS is reaching well beyond genomics. For example, it is revolutionizing transcriptomics through advances in RNA sequencing (RNA-Seq). Yet, despite this dazzling progress, a number of significant challenges remain.

These challenges were discussed at a recent Oxford Global event, the "Seventh Annual Next Generation Sequencing Congress". The event provided a window through which attendees could browse the NGS field's most daunting obstacles. It also displayed technologies that could allow these obstacles to be circumvented.

New capabilities and applications include the removal of toxic, unwanted transcripts from RNA-Seq libraries as well as the mapping of under-explored alternative splicing spaces. NGS is also making progress toward sequencing mitochondrial DNA. Mitochondria are increasingly recognized in disease development, yet sequencing the DNA from these organelles is complicated because mitochondria harbor considerable genetic complexity and heterogeneity.

Another area bedeviled by heterogeneity is tumor analysis. Fortunately, it appears that heterogeneous tumor samples could be subjected to sorting procedures that could isolate pure populations of cells. These populations would be more amenable to sequencing.

Finally, NGS is enhancing its single-molecule capabilities. One emerging approach involves coupling nanopore technology and mass spectrometry (MS).

## Singular Insights

"Nanopores may soon help revolutionize the fields of DNA and protein sequencing," asserts Derek M. Stein, Ph.D., associate professor of physics and engineering, Brown University. "My colleagues and I are developing a new single-molecule sequencing strategy that combines the processivity of solid-state nanopores with the sensitivity of MS. The idea is to sequentially cleave a nucleotide or amino acid from a single molecule after it

transits through a nanopore using photofragmentation, and then to identify it by determining its mass-to-charge ratio in an MS instrument.”

Dr. Stein’s team is starting by focusing on DNA sequencing. “Marrying nanopore technology with the electrospray ionization of MS allows rapid transfer of intact biomolecules from liquid into the vacuum,” he explains. “The technology could also accelerate the speed and precision of analysis. Traditional sequencing takes tens of milliseconds to visualize an individual base, but our approach can detect a base every microsecond or less.”

The scientists faced a number of challenges, including the need to construct their own instrumentation and components. “We knew that building and using a nanopore mass spectrometer would be an ambitious project,” recalls Dr. Stein. “So we first wanted to see if this was even feasible.”

Not only did they find that it was possible, they also developed a number of key improvements. “We miniaturized the droplets sprayed into the MS instrument. Conventional MS instruments lose 99.9% or more of the analyte molecules because they are injected into the instrument in large, charged droplets,” notes Dr. Stein. “When these droplets evaporate and explode, they send molecules flying all over.”

“We decided to make glass nanocapillaries that could miniaturize droplets to less than 10 nanometers, which is small enough that they can hold only a single ion—a quantum of charge,” he continues. “Thus, instead of the traditional ‘Coulomb explosions,’ we are creating ions one-by-one by ‘ion evaporation.’”

Dr. Stein believes that utilizing nanotechnology for protein sequencing presents even more exciting opportunities: “MS is unique in its ability to identify all 20 amino acids. The same nanotechnology DNA sequencing strategy also can be applied to proteins. That is, push proteins through a nanotube, photofragment them with a laser, and use an MS detector for accurate identification.”

Gazing into the nanotechnology crystal ball, Dr. Stein sees many futuristic applications: “Much like a needle draws blood from a small vein in the body, one could envision sometime in the future creating nanotips so small that they could penetrate into a single cell to remove a sample for analysis. Imagine what we could learn.”

## Removing Toxic Transcripts

Creating high-specificity RNA-Seq libraries remains an ongoing challenge. “It is critical to minimize the population of undesirable transcripts (often greater than 80% of a library) such as rRNA, globin, and other housekeeping species, while at the same time maintaining desirable transcripts from the original total RNA population,” advises Luke Sherlin, Ph.D., director of technical services, NuGEN Technologies.

The company developed a technology to do just that. During library construction, the Insert-Dependent Adaptor Cleavage (InDA-C) approach employs specific enzymatic steps to deplete any unwanted transcript sequences from the final library. This approach contrasts with hybridization capture methods, which can alter the original RNA populations. “InDA-C is a flexible approach that can be easily applied across a variety of species such as human, mouse, rat, *Drosophila*, and *Arabidopsis* species.”

The initial InDA-C step in the workflow adds forward and reverse library adaptors during untargeted strand selection. Next, the targeted transcripts are depleted by annealing specific InDA-C primers that are tiled across the unwanted sequences. A reverse adaptor, which can form a unique cleavage site, is then added. The final step eliminates unwanted species from the pool in the subsequent PCR step by enzymatically cutting the cleavage site in the reverse adaptor.

According to Dr. Sherlin, “While the InDA-C technology is a relatively new approach, it provides a unique way to construct unbiased RNA-Seq libraries from all RNA populations from any species. It is also easily customizable.”

Dr. Sherlin says that NuGEN supports any investigator wishing to design primer-specific transcripts for depletion free of charge. The company expects to continue leveraging the InDA-C technology for other sequencing applications, such as generation of libraries from the single-cell level.

## Alternative Splicing and RNA-Seq Data

RNA-Seq technology also provides an invaluable tool for deciphering the extensive alternative splicing of the transcriptome. Using this shuffling process, genes can code for multiple forms of the same protein. Alternative splicing creates two to potentially thousands of variants and occurs in more than 90% of human genes. This RNA processing mechanism, however, also plays a major role in multiple genetic disorders.

“The Human Genome Project created an initial map of splice variations more than 10 years ago,” notes Liliana Florea, Ph.D., assistant professor, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine. “But the map remains largely incomplete. There is still no repository for all alternative splicing events.”

“The challenge of cataloging every splice variant is daunting,” she continues. “But RNA-Seq experiments can survey the transcriptome of cells and tissues in great depth, providing a means to characterize alternative splicing in greater detail. But the main drawback to RNA-Seq continues to be the difficulty of accurately piecing together the short reads generated by the technology into long isoforms.”

Dr. Florea reports that she and her colleagues decided to develop a new computational system to apply to the abundance of RNA-Seq data: “We wanted to develop a user-oriented software to find the sweet spot that meets all one needs to do. We developed a suitable algorithm to identify secondary variations that cannot be found with other programs.”

Using this tool, she expects to be able to address how much alternative splicing occurs and how it varies among cell types. Dr. Florea’s group previously built a comprehensive catalog of alternative splicing for each tissue utilizing the Illumina Body Map dataset. The catalog spans 16 tissues and features more than 2.5 billion sequences as a test case. It has been used to generate insights on the basis of comparisons across tissue types.

Such comparisons are massive tasks. To carry out these comparisons with the necessary accuracy, Dr. Florea’s group relied on SpliceBox, an in-house software suite. (SpliceBox software and related information is available at the SourceForge and Zenodo repositories.) The group found that more than 60% of events were novel, involving new exons, new introns, or both.

“This collection will complement existing annotation databases and also provide key insights into the mechanisms involved and the evolution of alternative splicing,” insists Dr. Florea. “But many more RNA-Seq analyses need to be assessed. We expect numerous applications such as clinical sequencing, basic molecular biology, cancer research, and even plant genomics.”

## Mitochondrial DNA Sequencing

Mitochondria, those amazing cellular powerhouses, not only create an energy resource by generating ATP, they also impact one’s chances for developing cancer, heart disease, and diabetes. That’s the conclusion reached by

new studies led by Ravi Sachidanadam, Ph.D., assistant professor of oncological Sciences at Mount Sinai Hospital.

“Eukaryotic cells harbor two genomes, nuclear DNA (nDNA) and mitochondrial DNA (mtDNA),” says Dr. Sachidanadam. “While mitochondrial activity depends on more than one thousand proteins, primarily coded by nDNA, proteins (~13 of them) coded by the mitochondrial genome also play critical roles.”

Sequencing of mtDNA is no small feat because mitochondria harbor significant complexity. That is, each mitochondrion carries multiple mitochondrial genomes (5–10), and each cell possesses hundreds to thousands of mitochondria.

“Accurately cataloging the diversity of mtDNA has been a challenge,” Dr. Sachidanadam admits. “Although mtDNA represents less than 1% of total DNA, it varies intercellularly. Also, interpretation of sequencing data is confounded by mtDNA pseudogenes in nDNA.”

To overcome these challenges, Dr. Sachidanadam and colleagues developed a mitochondrial sequencing technology called MSeek. “While isolating mtDNA has been quite difficult, our method involves enzymatically purifying mtDNA by depleting linear nDNA and inexpensively sequencing it,” he explains. “MSeek also yields highly pure mtDNA (>90%) and provides unprecedented sensitivity and specificity. This is a novel way to purify and sequence mtDNA.”

The key breakthrough for his group was identifying exonuclease V as the best enzyme to digest nDNA while leaving the small circular mtDNA intact. “We tried a lot of enzymes before we found the right one,” he reports.

Dr. Sachidanadam adds that his group performs the digestion step prior to sequencing on the Illumina MiSeq platform. His group also utilizes computational methods to discount any pseudogene content.

Using MSeek, the group made another remarkable discovery. “We not only confirmed the ubiquity of a process called heteroplasmy (the occurrence of multiple mtDNA haplotypes), we also found that heteroplasmy may provide a fingerprint to identify cell types,” details Dr. Sachidanadam. “Further, we also found that cells can interchange their mtDNAs, which has the effect of stabilizing the mtDNA content of individual cells.”

Dr. Sachidanadam says that his group has only scratched the surface of what MSeek might eventually reveal: “Although a current limitation is that MSeek requires at least 4 µg of intact total DNA, we expect that many applications will follow, not just in therapeutics, but also in areas such as forensics.”

## Isolating Pure Tumor Cell Populations from FFPE

Molecular analysis of tumors is often problematic because of their heterogeneity, according to Raimo Tanzi, Ph.D., chief commercial officer, Silicon Biosystems. “The introduction of digital analytical methods, such as NGS and digital PCR, has helped tease out minority populations,” Dr. Tanzi points out. “Ultimately, however, only a homogeneous sample provides for the clearest interpretation.”

To address this issue, Silicon Biosystems applied its new digital technology, called the DEPArray™, to the isolation and sorting of pure populations of cells from a heterogeneous tumor sample. “This semiconductor-based technology uses dielectrophoresis (DEP) to bind single cells to single microelectrodes on a CMOS chip,” explains Dr. Tanzi. “It leverages the absolute precision of semiconductors for sorting cells in absolute purity.”

“FFPE samples are initially transformed into cell suspensions, stained with appropriate markers, and then loaded onto a single-use microfluidic cartridge,” Dr. Tanzi continues. “Once inside the flowcell, each single cell is first captured by one of the 30,000 CMOS chip-controlled electrodes (that is, a DEP cage) and then scanned with a microscope to obtain fluorescent and bright field images.

“Based on these images, each cell is allocated in a specific category (tumor, stromal, etc.). Eventually, all cells belonging to one same category are delivered in 100% purity, moving via software control of the corresponding electrodes toward a recovery chamber in the chip.”

The new technology can isolate pools of tens to hundreds of pure cells available for up to whole-genome analysis. Pools may include epithelial-mesenchymal transition cells, tumor-infiltrating lymphocytes, and stromal cells.

“The DEPArray is the first automated technology that can isolate 100% pure populations of cells from heterogeneous FFPE samples,” asserts Dr. Tanzi. “This provides many advantages, some of which are particularly relevant for cancer research, such as uniformity of samples, clear identification of copy number variants, and loss of heterozygosity. Another possible use is for performing genetic analysis of very small samples such as fine needle aspiration samples and tumors with low cellularity.”

Clearly, investigators are making exciting progress in the NGS field. The future should feature faster, more precise, and novel technologies.

## Improving Library Preparation

Library preparation is a critical part of the next-generation sequencing workflow. Successful sequencing requires the generation of high-quality libraries of sufficient yield.

As sequencing technologies improve and capacities expand, boundaries are also being pushed on library construction. High performance is required from ever-decreasing input quantities and from samples of lower quality or those with extreme GC content.

At the same time, there is a need for protocols that perform reliably and do not compromise the quality of the libraries produced. Officials at New England Biolabs say they recently reformulated each of the reagents in their NEBNext Ultra DNA workflow to create the NEBNext Ultra II DNA Library Prep Kit for Illumina.

“This new kit utilizes a fast, streamlined, automatable workflow for high-yield production of superior quality libraries, with picogram to microgram input amounts of DNA of varying quality,” said Eileen Dimalanta, Ph.D., NEBNext development group leader.

According to Dr. Dimalanta, the reformulated reagents also enable users to overcome challenges previously associated with successful library preparation, such as the use of challenging sample types, e.g., FFPE DNA, uniform CG coverage of the sample, and the use of fewer PCR cycles.

