

Targeted Resequencing Improves Precision

Ability to Focus on Specific Regions of Interest Pushes Technology to the Next Level

K. John Morrow Jr., Ph.D.

Next month in San Diego, CHI will present “Now Generation Sequencing” in which discussions about targeted resequencing will be front and center. This mixture of molecular biology and sophisticated computer analysis hones in on genomic regions of interest, interrogating multiple genetic sites. Enabling the detection of rare mutations and hard-to-reach corners of the genome, it takes advantage of low sample-input requirements. Moreover, these strategies allow areas of interest to be identified through genome-wide association studies while sequencing genes and candidate regions.

Using next-generation sequencing for clinical diagnostic purposes offers unique challenges due to the complexity of the technology and data analysis issues. Presentations at “Now Generation Sequencing” will discuss how to streamline technical processes and bioinformatics analyses to rapidly move this technology into the clinic.

The identification of the genes for extremely rare disorders was extremely difficult until the development of targeted approaches for sequencing partial components of the entire genome. Classic human gene mapping studies required large pedigrees of many affected families in which

the association between the disease and the candidate gene could be narrowed down through its linkage with such markers as single nucleotide polymorphisms and microsatellites.

Today the availability of powerful computer analysis combined with automated sequencing technology has allowed a brute-force approach to identification of the genetic region of interest. The large size of the entire human genome, however, requires a method of selective sequencing. Since the exome or coding regions comprise only 30 megabases, or 1% of the total human genetic makeup, a second-generation approach targeting this select class would make the task manageable and more cost-friendly.

Sarah Ng, a graduate student in genome sciences at the University of Washington, has been involved in a project to demonstrate proof of principle for targeted capture and massive parallel sequencing, using the Freeman-Sheldon Syndrome as a model. Success in the endeavor spurred the team on to nail down the location of the Miller Syndrome, a previously unmapped disorder.

“For the first time, we have pursued candidate gene identification through exome sequencing of a small number of unrelated affected individuals,” says Ng.

The team’s approach requires the enrichment of exomes by hybridization of DNA



febit biomed’s Geniom® RT Analyzer was designed for the automated processing of microarray experiments. It uses microfluidic Geniom Biochips for a wide range of applications.

from shotgun libraries to microarrays with synthesized DNA probes complementary to the human exome sequence. In this project, four libraries, each consisting of broken-up genomic DNA fragments from a subject with the condition, were hybridized to two custom 244 K Agilent (www.agilent.com) microarrays, and the captured exomic DNA from the affected individuals was entirely sequenced on the Illumina (www.illumina.com) Genome Analyzer platform.

This data revealed a large number of variants in each individual, which could be subjected to additional filtering techniques

to identify the sequences responsible for the disease. The candidate gene, which codes for an enzyme in pyrimidine biosynthesis, is referred to as DHODH. Mutations in this gene were then identified in a number of unrelated Miller Syndrome families.

Prior to the work of the University of Washington team, it was unclear whether Miller Syndrome was dominant or recessive since large pedigrees had never been observed. This unknown, combined with the decreased reproductive ability of individuals with this condition, greatly restricted the options available for its analysis.

“We have demonstrated the power, efficiency, and cost-effectiveness of this strategy,” says Ng. “This approach is likely to become a standard tool for the discovery of genes underlying rare monogenic disorders.”

Therapy for genetically based diseases invariably requires a solid understanding of the molecular basis of the condition—phenylketonuria being a classic example. This knowledge can now be put to use, although the rareness of the condition may stand in the way of efforts to develop a workable treatment.

febit biomed (www.febit.com) is using its HybSelect™ automated technology to investigate new cancer markers as well as other issues pertaining to the isolation of disease-related genes.

febit’s targeted resequencing technology is a microarray-based sequence capture strategy in which the entire genome is scaled down to manageable portions that can be used in focused investigations. It consists of three steps—hybridization of a

See Targeted Resequencing on page 28

News GENOMICS & PROTEOMICS

TGen Gains Open Access to Horizon’s Gene-Engineering Platform

Translational Genomics Research Institute (TGen) forged an open-source relationship with Horizon Discovery (www.horizondiscovery.com) that will allow it to exploit the company’s Genesis™ gene-engineering platform and X-Man™ isogenic cell lines. Under terms of the deal, TGen will use Horizon’s technology to develop a panel of novel cancer cell lines that will be licensed back to the company. Horizon will then be able to use the cell lines in future licensing deals with academic, biotech, and pharma organizations.

Horizon says that the deal will allow TGen to create a key center of excellence for generating and applying predictive human disease models in drug discovery. The company also hopes to establish similar relationships with other centers of excellence for translational research.

NextGen to Develop Biomarker Assays for Takeda’s Research

NextGen Sciences (www.nextgensciences.com) will develop validated protein biomarker assays to be used in Takeda Pharmaceutical’s (www.takeda.com) pre-clinical and clinical studies, supporting its drive toward personalized medicine.

The biomarker assays will be designed to target specific protein isoforms found in the liver and muscle of multiple species for use in preclinical studies. The assays will also be validated for use in clinical studies. NextGen Sciences reports that its Multiple Reaction Monitoring protein assay platform is useful for studies of this kind, which involve multiple isoforms and multiple species.

Beckman Coulter and Genpathway Team Up to Provide ChIP-Seq Services

Beckman Coulter Genomics (www.beckmancoulter.com) and Genpathway (www.genpathway.com) established a strategic

partnership to offer customers whole-genome chromatin immunoprecipitation sequencing (ChIP-Seq) services. The alliance provides researchers investigating gene regulation and genetic pathways a single, inclusive source for services supporting genome-wide studies of transcription factor binding, gene transcription, and epigenetic events.

ChIP-Seq combines chromatin immunoprecipitation with ultra high-throughput next-generation DNA sequencing to enable the study of complex interactions between proteins and nucleic acids. Mapping protein-DNA complexes in vivo on a genome-wide scale with ChIP-Seq builds a picture of the intricate network of interactions that regulate genes and genetic pathways in a sample, the companies claim. Comparing ChIP-Seq results from different conditions—diseased versus normal cells, for example—can reveal distinguishing patterns of genetic regulation that may be biologically significant. n

K. John Morrow Jr., Ph.D. (jmorrow@genengnews.com), is president of Newport Biotech and a contributing editor for GEN. Web: www.newportbiotech.com.

Targeted Resequencing

Continued from page 26

genomic library to a Geniom biochip, washing and elution, and finally sequencing and analysis. Working with 10% of the genome (300 million bases) the company scales down further to focus on clinically relevant regions. "Currently our main interest is in the area of oncogenes and oncology," says Peer Stähler, CSO.

febit uses barcoding for the indexing of samples. This is accomplished by adding identifying base sequences onto the PCR primers, which allows samples to be identified and followed, making automation of sequencing possible.

"We are currently conducting a large-scale analysis of cancer-related genes," explains Stähler. "This includes the breast cancer genes BRCA-1 and -2. We can analyze more than 60 samples in one run."

Custom Arrays

According to Fred P. Ernani, Ph.D., senior product manager for emerging genomics applications at **Agilent Technologies** (www.agilent.com), "Agilent's SureSelect Target Enrichment System is unlike any other commercially available target-enrichment method." The platform is based on targeting

with a biotinylated RNA library, which is hybridized with a genomic DNA sample to retrieve coding regions (the exome), which can then be bound to streptavidin-coated magnetic beads. "This approach allows us to build a custom array for the individual researchers' needs."

The custom design of the system permits customers to expand their focus on specific genome regions to elaborate complex designs, including unusual organisms and specialized cell types, such as cancer cells and samples expressing rare human genetic mutations. Investigations of genetic diversity are a critical area of evolutionary and environmental studies; the SureSelect system is ideal for resolving significant questions in these disciplines.

Clients can do their own custom design using E-array, a web-based tool that has been expanded to SureSelect. The system takes advantage of barcoding in which identifying DNA bases are inserted into the gene fragments. This permits highly targeted sequencing on a statistically relevant number of samples.

According to Dr. Ernani, SureSelect was originally designed to work with the

The RDT 1000 from RainDance Technologies is the centerpiece of the company's sequence-enrichment solution for the targeted genomic sequencing market.



Illumina Genome Analyzer end-sequencing protocol but its functions have been expanded for compatibility with the Illumina Genome Analyzer paired-end sequencing and the Applied Biosystems™ SOLiD System.

Translational Medicine

RainDance Technologies ([\[tech.com\]\(http://tech.com\)\) takes advantage of microdroplet-based solutions to empower its targeted-sequencing technology for the discovery of rare variants, according to Jeremy Lambert, product manager. "Our RainStorm™ technology produces millions of picoliter-volume droplets per hour in a PCR format," he explains. Each droplet behaves as a test tube that can contain a](http://www.raindance</p>
</div>
<div data-bbox=)

Evaluation of Enrichment Technologies for Diagnostic Applications



Researchers at Radboud University are evaluating Applied Biosystems' SOLiD and Roche's 454 GS FLX Titanium next-generation sequencing equipment for routine clinical diagnostic applications.

Sophia Fox

Targeted sequencing has obvious potential in clinical diagnostics, particularly for complex disorders that may hinge on either one or a combination of up to perhaps hundreds of genes. The approach, however, requires enriched templates for sequencing on next-generation platforms.

To address this issue and help validate the use of parallel targeted resequencing approaches for fast and cost-effective diagnostic applications, researchers in the department of human genetics at the Radboud University Nijmegen Medical Centre are evaluating a number of enrichment technologies, including **Roche NimbleGen's** (www.nimblegen.com) sequence capture microarray platform, **Agilent's** (www.agilent.com) SureSelect Target Enrichment System, and **RainDance Technologies'** (www.raindancetech.com) RDT 1000 instrument and Sequence Enrichment

platform. For sequencing, the team is using Roche's 454 system, as well as the Applied Biosystems SOLiD sequencer from **Life Technologies**.

"The use of sequencing for diagnostics is a field in development, and there is no one ideal system at this moment," comments Joris Veltman, Ph.D., associate professor. "We started research in this area using the NimbleGen arrays (in combination with Roche 454 sequencing), which are ideal for enriching small gene sets that will possibly be required for complex monogenic diagnostic applications. We are also evaluating the RainDance system for this targeted approach."

"For our research on larger-scale exome enrichment we are comparing the NimbleGen system to the Agilent SureSelect system (in combination with SOLiD sequencing). Results to date have confirmed that target enrichment provides high specificity and coverage and a high accuracy of SNP/mutation detection."

Professor Veltman and a team of colleagues led by Alexander Hoischen, Ph.D., recently published validation of an array-based sequence-capture method for identifying genetically heterogeneous disorders. They selected autosomal recessive ataxia as a model disorder, and performed enrichment by hybridization on NimbleGen sequence-capture arrays, followed by sequencing with Roche 454 Titanium shotgun sequencing. The approach was evaluated on five subjects carrying known mutations and two unaffected controls.

The method involved targeting the genomic sequences of seven disease genes, together with two control loci, on a 2 Mb sequence-cap-

ture array. After enrichment, the patients' DNA samples were analyzed using one quarter Roche GS FLX Titanium sequencing run, resulting in an average of 65 Mb of sequence data per patient, the researchers report.

This was sufficient for an average 25-fold coverage/base in all targeted regions. Enrichment showed high specificity, and on average, 80% of uniquely mapped reads were on target. Importantly, the authors claim, that the reported approach enabled automated detection of deletions and hetero- and homozygous point mutations for 6/7 mutant alleles, and greater than 99% accuracy for known SNP variants.

The researchers did observe far higher coverage for the coding sequences of the disease genes (31-fold), compared to noncoding sequences (24-fold). They suggest that such differences are most likely due to a greater level of uniqueness in coding sequences, rather than array design issues. The local sequence architecture also appeared to have a strong effect on the efficiency of DNA enrichment for individual exon, with those located near repeat-rich regions generally showing lower coverage.

Also, a comparison of SNP genotypes from massively parallel sequencing with SNP genotyping microarray studies suggested that SNP detection accuracy in next-generation sequencing data is mainly limited by sequence coverage and sequence misalignment, rather than by sequencing errors.

"Our results indicate that 15-fold sequence coverage (comparable to a PHRED quality score of 20) is required for reliable detection of heterozygous mutations, which was achieved for 93.3% of all coding exons in this proof of principle study," researchers conclude.

The reported study also aimed to evaluate general characteristics of the enrichment proce-

dures and, consequently, a large genomic locus, as well as the total genomic content of seven disease genes, was placed on the sequence-capture array. This resulted in a total of about 2 Mb of genomic DNA for which enrichment was performed and sequencing data was obtained. For diagnostic procedures, however, the researchers stress that it should be possible to limit analysis to just the coding and upstream noncoding exons of known disease genes. This will significantly reduce the amount of sequence to be enriched and sequenced, thereby increasing the coverage required for reliable mutation detection, they stress.

"For diagnostic applications we will develop a second-generation enrichment array containing targets for the coding sequence of all 70 genes that have been linked to ataxia. The coding sequence of these genes will still only occupy 320 Kb. The density of array probes per target can therefore be increased, resulting in even higher enrichment specificity. Because of the reduced target region, one-eighth Roche 454 sequence run will be sufficient to reach even better coverage compared to the current study while simultaneously lowering costs per patient."

To further develop targeted resequencing approaches for routine clinical diagnostic applications, Professor Veltman's group is involved in the EU-funded 7th Framework Program TECHGENE project (technological innovation of high-throughput molecular diagnostics of clinically and molecularly heterogeneous genetic disorders).

An industry-academic partnership, the three-year TECHGENE is focused on the development of massively parallel sequencing technologies that will be able to extend genetic diagnostics from simple monogenic disorders to more complex, genetically heterogeneous disease. Key criteria to be met include accuracy, throughput, completeness, low cost per

single molecule, reaction, or cell.

“Targeted resequencing provides the opportunity to assay both common and rare variants,” says Lambert. “There is increasing interest by translational medicine researchers in understanding their contribution in complex diseases.”

The company has designed an approach to targeted sequencing that relies on microdroplet technology. Briefly, regions of interest in the genome are identified, and forward and reverse primers for these many DNA segments are synthesized. The aqueous PCR primers for each gene segment of interest are then encapsulated in an inert carrier oil with a co-polymer surfactant, so each droplet behaves as a tiny test tube. Then genomic DNA fragments from the target are prepared, mixed together with the PCR reactants, and formed into their own microdroplets.

The primer droplet is merged with the genomic DNA droplet, one for one. This is the equivalent of 1.5 million individual PCR reaction tubes running in parallel in a thermocycler all pooled in a single 0.2 mL reaction tube. The amplicons are then released, and the individual reactions purified and sequenced. Thus a massively high-throughput performance is achieved.

RainDance’s first commercial application for the RDT 1000 instrument platform is sequence enrichment, attacking relevant target-specific regions on which researchers want to focus for resequencing, for example as a follow-up to genome-wide association studies. The company’s depth of coverage provides complete resolution and supports all sequencing platforms. “We will shortly announce extensions of our platform into DNA methylation and deep-resequencing applications, which we believe will be of particular interest to researchers studying cancer and immunology,” says Lambert.

Late last year, RainDance Technologies and CLC bio (www.clcbio.com) inked a partnership to develop specialized software for the analysis of targeted-resequencing data generated by workflows incorporating RainDance’s Sequence Enrichment Solution and next-generation DNA sequencing platforms. RainDance believes that by partnering with CLC bio, it can supply researchers with a solution that leverages the power of its microdroplet-based technology to provide the highest-quality results in a simple yet powerful workflow.

As a first step, CLC bio will release an expansion of its CLC Genomics Workbench, enabling scientists to analyze data faster and more effectively when using the RainDance solution for large-scale targeted resequencing studies.

Healthcare Costs

Hanlee Ji, M.D., assistant professor of oncology at Stanford University, aims to lower the cost of diagnostics using his group’s resequencing strategy that employs powerful DNA sequencing approaches. The Stanford group is using its skills to design strategies to identify mutations in cancer and, ultimately, put the approach

into place for prospective clinical trials.

Clinical genetics and oncology are the principle areas in which targeted resequencing can lead to improved patient management. Thus it is hardly surprising that Nimblegen (www.nimblegen.com), a division of Roche, has focused its efforts on first wave applications of exome resequencing. Using its array-based platforms, the company has used capture oligonucleotide probes to query the entire comple-

ment of human exons for the purpose of disease gene discovery. But these array-based technologies are difficult to scale.

Dr. Ji cites example of mutation testing in colorectal cancer, from the point of view of the gastroenterologist. “I see a lot of GI cancers including many cases of colon cancer. We have found that K-ras mutation testing is predictive of patient response to specific therapeutics, but to reach the level of validation required for an acceptable diagnostic

test one needs a large study, beyond the reach of labor-intensive array screening.”

In the case of complex disorders with a polygenic determination, one must look at even more individuals in order to build a valid collection of data. Array-based assays are not scalable so the Stanford team is focused on in-solution approaches employing an aqueous reaction mixture with oligonucleotides in a single tube.

See Targeted Resequencing on page 30



No BULL

Swine, Chicken or Fish

With the pressure on to switch to **animal-free** media ingredients, look no further than **MARCOR**. We have the products you need to convert your current media formulations into “meatless” media.

MARCOR, with an extensive product line that includes **over 50 animal-free peptones, hydrolysates and extracts**, offers you an ideal, cost-effective method of converting to a totally “meatless” menu.

And that’s no bull.

Call today to speak to a sales representative, or visit our site at www.marcorclc.com for more information.



MARCOR
DEVELOPMENT
CORPORATION

341 MICHELE PLACE
CARLSTADT, NJ 07072
201-935-2111 • FAX: 201-935-5223
EMAIL: sales@marcorclc.com
www.marcorclc.com

Targeted Resequencing

Continued from page 29

Dr. Ji explains that it is practical to manipulate tubes in large-scale studies using 96-well plates, allowing analysis of thousands of patients.

As an example of the current dynamics of personalized medicine costs, Dr. Ji considers the test for the BRCA1 mutation, which predisposes to a risk of early-onset breast cancer. The cost of the BRCA test is over \$2,000 per patient. Dr. Ji believes that his approach could eventually, drastically

reduce the costs of this test to around \$20.

“With next-generation sequencing costs plummeting, we can foresee a time when we can screen 100 cancer-related genes for a few hundred dollars. We could analyze a patient’s tumor for less than \$100, and in this way, patients and physicians could avoid futile treatments. So we have a win-win situation for everyone, including the insurance providers, since they wouldn’t waste money on treatments that provide no benefit.”

A number of investigators from large genome centers are attempting to sequence the whole genome of an individual patient or tumor tissue in order to determine genetic profiles. Although genome sequencing costs have dropped precipitously in the last year, they are still too high, at around \$10,000, according to Dr. Ji. This is too much for a routine tumor genome screening, so it is essential that the costs be driven down more. He hopes to develop protocols that could be

done in any laboratory in the world.

Dr. Ji emphasizes that new DNA-based diagnostic tests can be moved rapidly into clinical use. “It’s astounding how fast the field is moving. We are planning for the day when any patient could come into an HMO and provide a family history, and then be tested for susceptibility to specific cancers, such as breast or colon cancer. That would prove to be enormously beneficial, as we identify individuals at high risk of cancer much earlier, which ultimately could be a lifesaver.”

Jurgen Vanhauwe, Ph.D., senior manager for sequencing market development at Applied Biosystems (www3.appliedbiosystems.com), part of Life Technologies, will talk about the SOLiD™ platform at the meeting. “This is a highly accurate, massively parallel sequencing platform that delivers a greater than 99.94% base identification accuracy.”

Accurate base reading is absolutely critical for the success of targeted resequencing strategies. If readings cannot be verified, then errors could be classified as mutations or SNPs, thus rendering the data useless. The SOLiD platform uses a sequencing methodology based on sequential ligation of dye-labeled oligonucleotide probes whereby each probe assays two base positions at a time (four fluorescent dyes encode 16 possible two-base combinations). This approach, in combination with the nature of the color code, will lead to a sequence of overlapping dimers that allows for error correction.

The SOLiD platform enables massive sequencing projects, heretofore impossible, Dr. Vanhauwe says. Dr. Peter J. Campbell and his colleagues at the Wellcome Trust and other institutes recently sequenced a small-cell lung cancer cell line, NCI-H209, exploring the mutations within 134 coding exons. They identified a tandem duplication in the gene referred to as CHD7, indicating that this gene may be critical to the transformation process. If it is determined that this alteration is an essential feature of the disease, this would represent an important landmark in our understanding of the mutations brought about by tobacco carcinogens and would point the way to therapeutic agents that could block the effect of these mutations.

“The exciting feature of this technology is its throughput and accuracy,” says Dr. Vanhauwe. “We could do whole genome sequencing the old way if we had 50 next-generation sequencing machines running in parallel, but with targeting resequencing we can focus the SOLiD platform on the regions of interest. We’re now able to sequence 100 patients in a single run.”

Instead of a huge, uncharted genetic wilderness, it now possible to focus on critical exons and portions of the genome that determine susceptibility to cardiovascular insult, malignancy, and autoimmune dysfunction. But this is only the beginning; within a short time (perhaps months) it should be possible to screen a patient for a large assortment of genetic disorders at a manageable cost.

Nanoscale Human Inflammation Cytokine ELISA

The Power of Nanoscale

Ultraminaturized & Ultrasensitive



Combining the power of Dip Pen Nanolithography™-deposited antibody arrays with a simple bench top protocol, Nanoink's Human Inflammation Cytokine Assay enables multiplexed detection of ten clinically relevant low-abundance cytokines: IL-1 α , IL-1 β , IL-2, IL-3, IL-4, IL-5, IL-6, IL-8, IFN γ and TNF α . Unlike conventional microarray platforms, DPN™ nanoarray deposition preserves the integrity of arrayed proteins. And DPN™ is capable of printing highly reproducible features of less than 4 μ m each. The result is single femtogram/ml assay sensitivity, even with the smallest of sample sizes.

Experience nanoscale now. Through April 2010 buy any Human Inflammation Cytokine Assay kit and get 50% off an equivalently priced kit.

Find out more: www.nanoink.net/nanobio
847-679-3432

© 2010 Nanoink, Inc. All rights reserved.
Nanoink, the Nanoink logo, Nano BioDiscovery, Dip Pen Nanolithography, and DPN are trademarks or registered trademarks of Nanoink, Inc.

Microarrays

125 μ features
5x objective
RFU = 132

Nanoarrays

4 μ features
10x objective
RFU = 44,169

10 pg/ml of IFN γ antigen with control subtracted



Nano BioDiscovery