

```
Drupal.behaviors.print = function(context) {window.print();window.close();}>
```



UW Team Improves MIP-Based Method to Sequence Gene Panels in Large Sample Sets

December 18, 2012

UW Team Improves MIP-Based Method to Sequence Gene Panels in Large Sample Sets

By Julia Karow

Researchers at the University of Washington have improved a multiplexed targeted sequencing method based on molecular inversion probes that allows them to capture and sequence dozens of candidate genes in very large numbers of samples at low cost.

The improved method, described in a [recent paper](#) in *Science*, could be "broadly useful for the rapid and economical resequencing of candidate genes in extremely large cohorts," for example to validate rare variants or *de novo* mutations in genetically complex disorders, according to the authors.

The method has advantages over other candidate gene resequencing approaches, they say, in that it requires small amounts of DNA, can easily be optimized by rebalancing the probes, allows for the target composition to be changed, offers a streamlined workflow, and has low costs per sample when amortized over large sample sets.

According to Brian O'Roak, the lead author of the article and a senior fellow in the groups of Jay Shendure and Evan Eichler in the department of genome sciences at UW, the improved method is based on a MIP-based high-throughput exon capture and sequencing approach that the Shendure group published three years ago ([IS 4/21/2009](#)).

In that publication, the researchers used MIPs that had been synthesized on microarrays to target either 13,000 or 55,000 exons in HapMap samples followed by direct Illumina sequencing, and captured between 91 percent and 98 percent of the targets.

A major technical drawback of that method is the poor uniformity of coverage, O'Roak said – about 60 percent of targets were captured within a 10-fold range, and 90 percent within a 100-fold range – as well as the fact that the array-synthesized oligos need to be pre-amplified in order to have enough material, which is cumbersome and can make their amounts uneven.

He and his colleagues set out to modify the approach in order to make it suitable for validating results from exome sequencing projects, specifically from an autism trio study where they

identified many *de novo* mutations in different genes, in large cohorts. "We really wanted a method where we could do a reasonable number of targets across thousands of samples," O'Roak told *In Sequence*.

One change they made is to column-synthesize the oligonucleotide probes individually rather than synthesizing them simultaneously on arrays. Thus, they obtain enough material to analyze "millions of samples" and they can rebalance pools of probes by swapping out individual poor performers, improving uniformity.

Having individual probes available also allows them to change the number of genes they want to sequence in each project. "We can add in new genes, take out genes that we are no longer interested in or that are poor performers," O'Roak said.

Another major improvement is the use of a new algorithm that better predicts the performance of probes, based on the researchers' experience from prior studies.

Rebalancing the probe pool "makes a huge difference" to uniformity and performance, O'Roak said, noting that one round of rebalancing enabled them to cover 95 percent of the target regions on average, with most of the targets being captured within a 10-fold range.

To make the method applicable to large numbers of samples, the researchers also simplified the workflow. They combined three steps – probe hybridization, polymerase gap-filling, and ligation – into a single one, which "makes setup much easier," he said. They were also able to reduce the enzyme concentrations significantly, lowering costs and helping, for some reason, to increase the capture efficiency.

The protocol then requires two additional steps, treating the samples with exonuclease to degrade non-circular DNA and adding Illumina adaptors and barcodes. All steps are amenable to liquid handling automation, O'Roak said.

The approach also uses only small amounts of DNA, requiring about 50 nanograms of genomic DNA as input, according to the paper.

For their published proof-of-concept study, the scientists captured and sequenced 44 candidate genes, covering about 100 kilobases of target DNA, in almost 2,500 autism spectrum disorder probands from the Simons Simplex Collection.

Overall, they discovered 27 *de novo* mutations in 16 of the 44 genes in the autism samples, and recurrent disruptive mutations in six genes. Of those mutations, six were not contained in exome sequencing data that were available for 82 of the individuals, "consistent with an increased sensitivity for MIP-based resequencing," the authors wrote.

They also estimated reagent costs, which came out to \$14.60 per sample for their study, which is "much cheaper than anything I know of for this kind of target scale," O'Roak said.

The MIP library was the most expensive single item, totaling about \$14,400 for the 44 genes they analyzed. However, amortized over 3,000 samples, MIPs only amounted to \$4.80 per sample, while sequencing reagents were \$7.23 per sample, and capture reagents and consumables totaled \$2.57 per sample. Per gene and sample, reagent costs were about \$0.33. "If you are going to do any reasonably sized set of samples, the oligo cost starts to become fairly trivial per sample," O'Roak said.

According to Olivier Harismendy, an assistant adjunct professor at the University of California, San Diego, who has experience with several targeted sequencing methods, the MIP approach has a "clear cost advantage" over other methods and does not require any special instrumentation, like RainDance or Fluidigm.

However, the steep upfront cost of the probe library only makes sense when thousands of samples will be sequenced, he said. "For all other studies, in the few dozen to hundreds of samples, it is important to estimate the financial tradeoff."

In clinical sequencing of tumors, for example, typically small batches of samples are sequenced in order to return results in a timely manner, "so you need to think twice whether you want to order material for thousands of samples up front, and take the risk [of having] an outdated panel in six months or a year," he said.

Also, cancer samples require deep sequencing in order to compensate for contaminating normal tissue or to find subclonal mutations. "For this reason, the most uniform commercial targeted capture assay, [even though] more costly on a per-sample basis, will still be preferred for clinical sequencing," he said.

"For all other large academic projects [that] have the skills and expertise of the Shendure lab, MIPs are definitely a very wise option," he added.

Harismendy said the main drawback of the method is still a lack of uniformity. As sequencing costs go down, however, that might not be much of a problem because one can compensate for it by deeper sequencing.

Since their autism pilot, the UW researchers have applied their approach in a number of collaborations, O'Roak said, each involving sets of 20 to 30 genes, for example candidate genes for epilepsy and Joubert syndrome, and several hundred to thousands of samples.

They are also planning to expand their autism resequencing study to a larger set of genes and a cohort of about 4,000 individuals.

In addition, they are collaborating with researchers at the Puget Sound Blood Center to develop an assay for sequencing hemophilia genes in order to provide patients with a genetic diagnosis.

O'Roak and his colleagues are working on further optimizing the method, for example by updating the design algorithm in a way that takes advantage of the longer read length now available for the Illumina platforms. Right now, they generate 100 base paired-end reads on the HiSeq and 250-base paired-end reads on the MiSeq. O'Roak said the team is very interested in finding ways to increase the size of the gap fill while at the same time maintaining performance.

He said he is not aware of any plans by companies to offer their method as a commercial service, noting that the intellectual property landscape around MIPs, also known as padlock probes, is complicated.

Agilent Technologies has been offering a related method, called HaloPlex, both as a custom kit for up to 500 kilobases of target DNA and as pre-designed gene panels. That method, which is based on so-called selector probes, was developed by Halo Genomics, a Swedish startup that Agilent acquired a year ago ([*IS 12/6/2011*](#)).



Julia Karow tracks trends in next-generation sequencing for research and clinical applications for GenomeWeb's *In Sequence* and *Clinical Sequencing News*. E-mail [Julia Karow](mailto:Julia.Karow@genomeweb.com) or follow her GenomeWeb Twitter accounts at [@InSequence](https://twitter.com/InSequence) and [@ClinSeqNews](https://twitter.com/ClinSeqNews).

Related Stories

- [In Print: Last Week's Sequencing-Related Papers of Note March 19, 2013 / In Sequence](#)
- [In Print: Last Week's Sequencing-Related Papers of Note March 5, 2013 / In Sequence](#)
- [In Print: Last Week's Sequencing-Related Papers of Note February 26, 2013 / In Sequence](#)
- [In Print: Last Week's Sequencing-Related Papers of Note February 19, 2013 / In Sequence](#)
- [In Print: Last Week's Sequencing-Related Papers of Note February 12, 2013 / In Sequence](#)

footer