



# Evan Eichler's New Method for Next-Gen Identifies Thousands of New DNA Sequences

June 02, 2010

**Byline:** Christie Rizk

**Newsletter:** [Genome Technology](#)

[Genome Technology - June 2010](#)

Using a new DNA clone-based approach, the University of Washington's Evan Eichler and his team were able to find sequences that had been missing, fragmented, or misassigned in genome assemblies that were based solely on next-generation sequencing. They found 2,363 new DNA sequences corresponding to 720 genomic loci, as they reported in *Nature Methods* in April.

The researchers began by building a genomic library from a subset of nine people — four Africans, two people of European ancestry, two Asians, and one person of unknown ethnicity — and subcloned 40 KB segments of their DNA. The researchers then generated reads from both ends of each fragment and mapped the clones to the human reference genome.

They observed differences between the reference genome and the new sequences when some of the new sequences' end-pairs didn't map at one end or the other to the reference, Eichler says. "Those ends that didn't map corresponded to the contig of a new insertion sequence and that end that was mapping gave positional information of where it mapped," he adds. The researchers looked specifically for sequences where one end was anchored, and the other wasn't — so-called "one-armed bandits" — because they give clues as to where the new inserts belong in the genome, according to Eichler.

But more importantly, he says, this new approach proves that researchers who rely solely on next-gen sequencing technology end up missing a lot of information when they sequence a genome. "[The information] isn't being

properly anchored," he says. "It's of very little use to people if you don't know where this new sequence that carries a new exon, or a new promoter, or a new enhancer, where it actually maps with respect to a gene."

In comparing their results on the same individual genome to previous sequencing results from an Illumina SOAP *de novo* assembly, the researchers found that the older study was missing many sequences and that many of the new contigs were fragmented. In their Nature study, the scientists write that they found copy-number polymorphisms in 18 to 37 percent of the new insertions, including loci that are stratified among homogenous populations groups like Europeans, Asians, and Africans. But Eichler's new method accurately genotypes new insertions by mapping next-gen sequencing data sets to the breakpoint, which would allow researchers to characterize CNVs in regions they were previously unable to access.

Eichler characterizes the increased reliance of comparative genomics on next-gen sequencing — sometimes to the exclusion of other methods — as "a huge, huge problem," and recommends an approach that includes large-molecule characterization along with the next-gen sequencing, which he admittedly calls a "seductive" technology. "I would argue that a mix of new technologies that involve longer reads, larger insert libraries, and essentially larger molecules where you can get phase information is what you need to do a genome well," he says.

Genomeweb system

These settings are generally managed by the web site so you rarely need to consider them.

**Issue Order: 3**

