

THE AUTHOR FILE

Evan Eichler

By mapping the underbelly of human genomes, researchers open new questions.

Sequencing machines are faster than ever, but they are ill-equipped to ‘illuminate’ regions referred to as dark matter: common insertions not reflected in the reference genome, certain repetitive sequences and ‘tough spots’ on chromosomes. That is not good enough

for Evan Eichler of the University of Washington. “I can’t accept people classifying these regions as inaccessible,” he says.

Eichler’s solution, capturing long chunks of human DNA and replicating them in bacteria for sequencing, hearkened back to the early days of the Human Genome Project. “It was considered a little bit old school,” he recalls.

However, Eichler’s

team had a new way to home in on uncharacterized insertions. They used fosmid, phage-processed plasmids with the peculiarly precise size of 40 kilobases. The researchers could sequence one end of the fosmid and, by looking 40 kilobases away on the human reference genome, predict what the other end should be. Pairs of end sequences that did not match the prediction indicated something interesting in the middle.

Even though they sequenced only the ends of most fosmids, the task was a huge undertaking: it required a million captures for each genome along with all the concomitant analysis and debugging. For example, the human cell lines used for analysis had been immortalized through a process that involved insertion of the Epstein Barr virus; making sure viral artifacts were excluded was frustrating, time-consuming and essential. “These are not the Eureka moments,” says Eichler. “They tell you why many people don’t work through these problems.”

On page 365 of this issue, Eichler and colleagues report an analysis of DNA from nine individuals that uncovers nearly 2,400 new insertion sequences corresponding to over 700 loci. Though next-generation sequencing had missed, misassigned or fragmented many of these sequences, Eichler’s complete resequencing of some of the loci revealed

both new exons and conserved noncoding regions, many of which correspond to differences among Africans, Asians and Europeans. “When we started to see how much variation there was in different populations, it sort of blew us away,” says Eichler. These new sequences offer fuel for intriguing and testable hypotheses. For example, most sequences from Africans include a 3.9 kilobase insertion in an untranslated region of the lactase gene; the insertion is largely absent in Europeans and occurs in a similar genetic background as a single-nucleotide polymorphism associated with the ability to digest dairy products.

Eichler says the goal of his team is both to pursue how such genetic variations relate to phenotypes and to open up these kinds of questions for others. “We’re mapping all this stuff that’s not in the reference genome. By pushing it all the way to the finished sequence, there’s so much more that we [the scientific community] can do,” Eichler says. “This paper focuses on the insertion because it’s something that you can’t get easily from any other technology. The quality of product we’re generating is something that people could use ten or fifteen years from now.”

The point of digging deep into individual variation is not simply to find new insertion sequences, says Eichler, but to build better models for population genetics and disease. It is unclear whether the differences he is starting to uncover are the result of drift or selection, he says, but these differences are nonetheless informative. “It tells us something fundamental about the sequences that are missing from the reference. They are relatively young. If there are regulatory elements, then it follows that there are different regulatory elements in different people and in some cases different populations,” Eichler says.

There is much more to do. Eichler is enthusiastic about the 1000 Genomes project, a large international effort to catalog human genetic variation. Eichler’s published analysis covered only nine individuals, and his research team is already finding interesting patterns. Eichler concludes: “The hardest decision was when to stop and write the paper. This is a rabbit hole that goes down, down, down.”

Monya Baker

Kidd, J.M. *et al.* Characterization of missing human genome sequences and copy-number polymorphic insertions. *Nat. Methods* 7, 365–371 (2010).

“I can’t accept people classifying these regions as inaccessible.”

—Evan Eichler

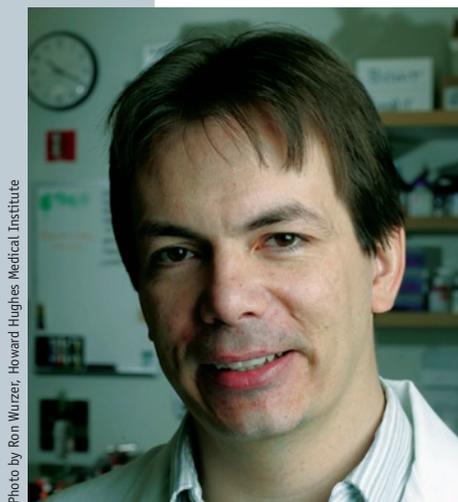


Photo by Ron Wurzer, Howard Hughes Medical Institute