

Stanford University
318 Campus Drive
Clark Center Room W352
Stanford, CA 94305-5444

SeeingScience

BY KATHARINE MILLER

VISUALIZING HUMAN GENOME VARIATION

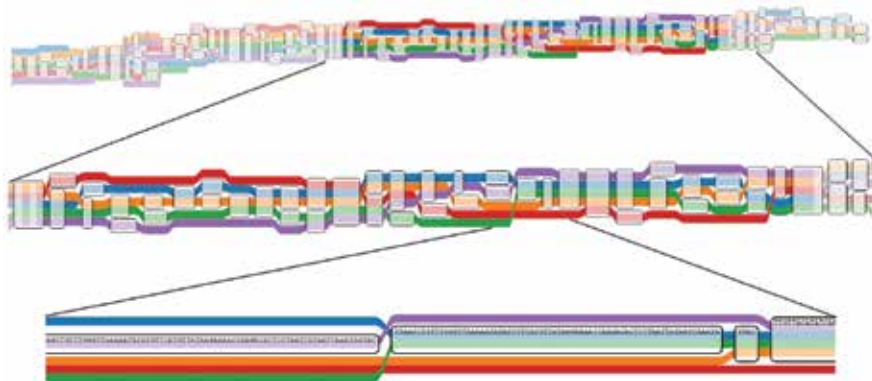
Humans share 99.5 percent of their DNA sequence, but that still leaves plenty of variation to go around. To get a handle on which variations contribute to health or disease, researchers typically compare individuals' genomes to a single "reference" genome that represents an assemblage of very high quality human genome sequences.

But now researchers are envisioning a better way to think

about reference genomes by building a genome graph that represents not just a single linear genome but also known variation. "The graph is this comprehensive representation of human variation that allows us to have a discourse that computers can understand about all of the different ways that humans vary," says **Benedict Paten, PhD**, associated research scientist at the **Big Data and Translational Genomics (BDTG) BD2K Center** at the University of California, Santa Cruz.

BDTG is developing tools that Paten says are "head and shoulders better" for understanding human genome variation than what can be done with just a linear reference genome. Alongside the applications that enable analysis of the human genome variation map, Paten's team is creating a visualization tool that shows the way different humans are explicitly represented within the graph. "These are pretty pictures that make intuitive sense," Paten says.

"The work started out largely theoretical but is really moving toward something that we think will have wide-scale practical application in the next few years," Paten says. □



*This prototype visualization of a genomic variation graph zooms in on portions of the NOTCH2 gene, an important gene for development. The colored bands represent 5 different variants of the gene, with rectangular shapes representing nodes (shared DNA sequences) and the colored ribbons between nodes representing paths/edges (not sequences). In the top panel, introns are shaded out (at right and left) while the solid colors represent exons 4 and 5. The exons are shown in increasingly greater detail in the bottom two panels. The visualization tool can also provide an intuitive graphical view of inversions, as shown in the green and red loops in the simulated example to the right. Images courtesy of **Wolfgang Beyer**, software developer for the Computational Genomics Laboratory at the University of California, Santa Cruz.*

