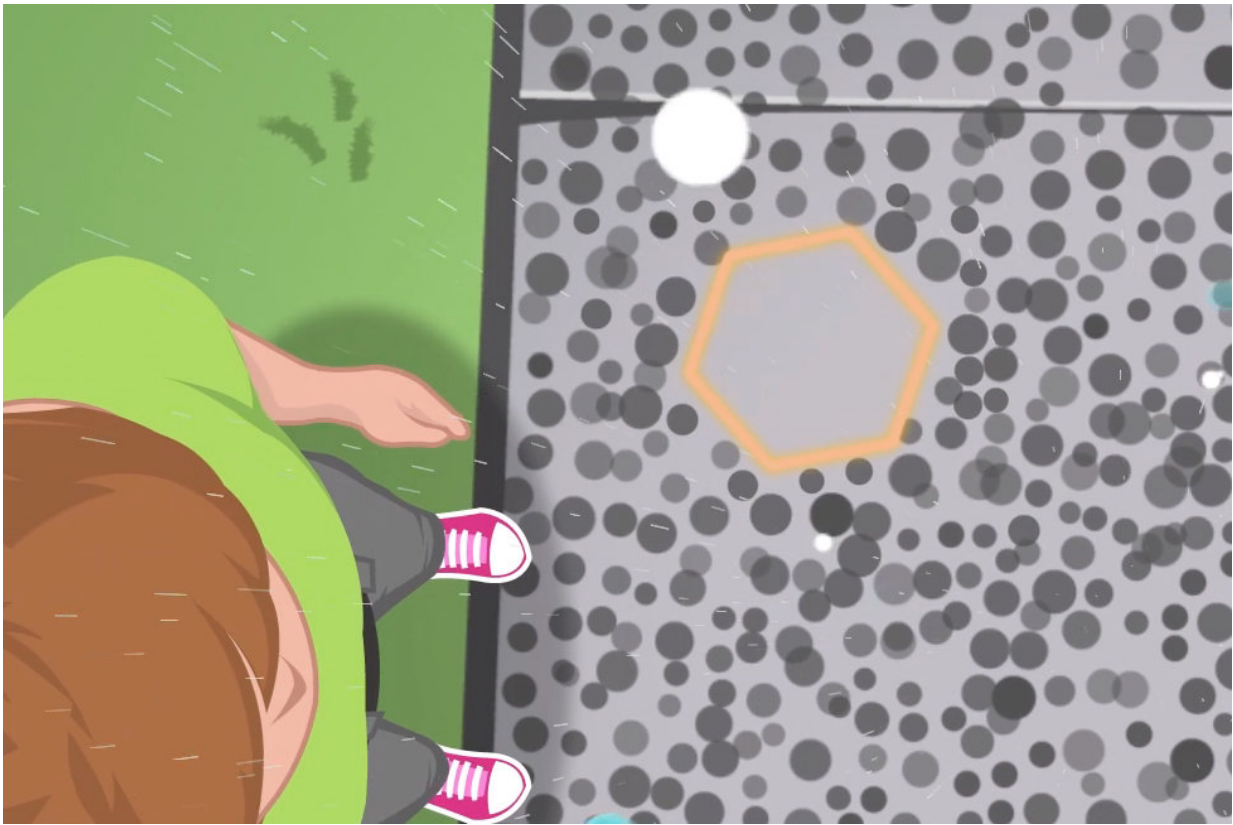# Big datasets pinpoint new regions to explore the genome for disease

December 10 2018

Researchers use the analogy of raindrops on the sidewalk to explain their new method to identify genetic variations that cause severe pediatric diseases. Credit: University of Utah Health

Imagine rain falling on a square of sidewalk. While the raindrops appear

to land randomly, over time a patch of sidewalk somehow remains dry. The emerging pattern suggests something special about this region. This analogy is akin to a new method devised by researchers at University of Utah Health. They explored more than 100,000 healthy humans to identify regions of our genes that are intolerant to change. They believe that DNA mutations in these "constrained" regions may cause severe pediatric diseases.

"Instead of focusing on where DNA changes are, we looked for parts of genes where DNA changes are not," said Aaron Quinlan, Ph.D., associate professor of Human Genetics and Biomedical Informatics at U of U Health and associate director of the USTAR Center for Genetic Discovery. "Our model searches for exceptions to the rule of dense genetic variation in this massive dataset to reveal constrained regions of genes that are devoid of variation. We believe these regions may be lethal or cause extreme phenotypes of disease when mutated."

While this approach is conceptually simple, only recently has there been enough human genomes available to make it happen. These new, invariable stretches may reveal new disease-causing genes and can be used to help pinpoint the cause of disease in patients with developmental disorders. The results of this study are available online in the December 10 issue of the journal *Nature Genetics*.

According to Quinlan, genes that have not previously been associated with disease often harbor one or more highly constrained regions. A mutation in these regions could cause disease.

"We are confident that these genes play a role in development of disease, but we currently know little about their role," said Quinlan, senior author on the paper. "That's where the exciting potential for discovery is."

Many of the most constrained regions are enriched for genes associated

with developmental disorders, including developmental delay, seizure disorders and congenital heart defects. This information gives the team confidence that the method is revealing truly constrained regions of genes.

His team created a [detailed map](#) of these constrained regions using more than 120,000 genomes obtained from the Genome Aggregation Database (gnomAD), a project that provides a massive catalog of human genetic variation detected in exome and genome sequencing data from a variety of large-scale sequencing projects.

The maps reveal both disease-causing variations and de novo mutations that underlie developmental disorders. This approach opens the door to identify new coding regions to study disease.

"A gene as a whole might be able to tolerate variation, but variation in one critical section [of the constrained [region](#)] could have serious developmental consequences," said James Havrilla, first author on the paper and graduate student in Quinlan's lab.

Quinlan cautions that the model is only powered to find extreme phenotypes, like [developmental disorders](#) responsible for intellectual disabilities, seizure disorders, facial dysmorphism and issues with heart development. The model is not adequate to identify regions of genes for common diseases, such as diabetes or coronary artery disease. In addition, the study is based primarily on individuals with European ancestry.

"The map we created will provide the community with a resource to study [genes](#) that heretofore had no disease association," Quinlan said. "The beauty and power of this approach is that, as we obtain more data from ever more human genomes, we can continue to improve the resolution of this map to pinpoint areas to study for disease."

Provided by University of Utah