

# Genomic Answers for Kids Team Releases Thousands of Genomes to Gain Insights Into Rare Diseases

By Andrea Anderson

NEW YORK – Investigators with the Children’s Mercy Research Institute in Kansas City’s Genomic Answers for Kids (GA4K) project have released genome sequence data for thousands of participants in an effort to better understand pediatric rare diseases. But this first batch of data only represents a fraction of what they expect to be able to share in coming years as the program continues.

“Genomic Answers for Kids is our approach to change the outcomes for rare disease families, primarily in the greater Kansas City area, where Children’s Mercy’s main catchment area is,” explained Tomi Pastinen, director of the Children’s Mercy Genomic Medicine Center and a researcher with the University of Missouri-Kansas City School of Medicine. While the program accepts patients from other institutions, as well, more than 90 percent of patients come from the area, he added.

“By building a large database from tens of thousands of individuals in our area, we will disseminate the data and enable rare disease discovery throughout the country and beyond, globally,” he noted.

The GA4K project is being done through the Children’s Mercy Research Institute, which opened this year, following a large fundraising effort. The team has been doing related sequencing work for two years and has already enrolled thousands of pediatric patients and their family members. The sequencing analyses include one or both parents, when possible, although a subset of the rare disease patients are tested as singletons or analyzed with the help of sequences from unaffected siblings.

In addition to increasing research into pediatric rare diseases, the program is intended to boost the coverage of genomic medicine services and increase access to genomic testing when appropriate, Pastinen noted, particularly for families that are not being reached by existing genetics or genomics programs.

“We see disparities across our catchment region and, as usual with health disparities, these touch mostly minorities and inner-city populations, as well as some rural populations,” he said.

Pastinen emphasized that the sequencing and analytical work is offered for free to families seeking rare disease diagnoses. Participating patients and their families donate their data, which is subsequently shared with other scientists in a de-identified manner to encourage rare disease research (the GA4K team has pledged that it will not sell participant data).

Although the diagnoses through GA4K have led to treatment or

clinical management changes for a small subset of patients so far, Pastinen explained, genetic testing can frequently provide families with an explanation for a child’s mysterious symptoms, bringing the often stressful, time-consuming, and frustrating processes of seeking a diagnosis to a close.

That was the case for Celia Steele, a girl from Wichita, Kansas, who began experiencing developmental delays, movement problems, epilepsy, and other symptoms at around the time that her twin brother began to crawl and walk. Her family had been looking for answers since she was a year or two old.

“It’s been a long journey. We first got to Children’s Mercy in the genetics department probably six years ago,” recalled Celia’s mother Teresa Cruz-Steele. “We tried everything that we could locally here in Wichita to figure out what was happening. Nobody here in Wichita knew — they sent us to all kinds of different doctors, thinking it was GI issues and all kinds of different things.”

A local neurologist referred Celia to a genetics department, where tests repeatedly came back negative. “Nothing’s ever really come back to say anything about why she has a movement disorder, why she’s delayed,” Cruz-Steele said. “She doesn’t walk, she doesn’t talk.”

The family started to get some answers when a doctor at Children’s Mercy told them that Celia had dystonia — sometimes painful, involuntary muscle contractions. Around 2019, a neurologist told them about GA4K and they thought “why not”? Celia’s parents both had blood drawn for DNA testing and participated in interviews. After that, though, they largely forgot about it, as Celia’s immediate health concerns took precedent.

While staying close to home during the COVID-19 pandemic last fall, Cruz-Steele got a call that brought tears of relief: Exome and genome sequencing in parallel through the GA4K program showed that Celia carries an altered version of the PDE2A gene that has been linked to many of the same symptoms she experiences.

The long-awaited genetic diagnosis has not changed Celia’s day-to-day care. But now, her mother explained, “we know the reason why she’s delayed. We know why she has dystonia.”

“It’s just, honestly, a relief,” Cruz-Steele said. “It’s one of those things that was always there, and we never knew why. Now at least we have some idea what’s going on.”

The diagnosis also makes it possible to keep tabs on PDE2A-related research and treatment advances in the future, she noted,

and to find out more about the symptoms or medical requirements that other children or adults with the condition experience at different points in their lives.

So far, the GA4K investigators have been able to make nearly 600 genetic diagnoses for Celia and other children, starting with exome sequencing and expanding their analyses to whole-genome sequencing and long-read whole-genome sequencing for patients who do not get diagnostic answers from protein-coding sequences alone.

The team added more than 2,300 of those genome sequences to the National Institutes of Health dbGAP database last month, and submits data to public databases such as ClinVar.

Along with collaborations with other groups working on rare undiagnosed diseases, the GA4K group has also set up a larger repository for registered researchers and clinicians that is updated each week with de-identified genetic data, prioritized variant information, and electronic medical record-based patient phenotypes.

“I think that’s a key thing for the future: to have a dynamic analysis of genetic variation, but also the evolution of a patient’s clinical picture in parallel,” Pastinen said. “That’s going to augment discoveries. Without this data sharing, [some] families will remain in limbo, because [many] families that are undiagnosed do have significant findings in their genome. We just can’t call them diagnostic as of today.”

The team also described a subset of the cases assessed by both exome sequencing and Pacific Biosciences long-read genome sequencing in a preprint posted to MedRxiv in mid-October. That study focused on 1,083 patients from 960 participating families and provided an opportunity to look at the additional data offered by long-read sequencing.

In that group, almost 35 percent of patients with no previous genetic testing received a diagnosis following the team’s sequencing-based testing and machine learning-based genetic variant prioritization efforts, as did 11 percent of the rare disease patients with genetic tests that came up negative in the past.

Even more patients — more than half of those who did not get a definitive diagnosis — carried variants of uncertain significance, the researchers reported. Within that set, they went on to find more than 150 promising candidate genes with the help of the GeneMatcher service from the Baylor-Hopkins Center for Mendelian Genomics.

“[T]he majority of unsolved cases in our cohort do have candidate genes and variants but lack sufficient evidence to assign pathogenicity due to a lack of replication (also known as the ‘n of 1’ problem), with hundreds of genes and variants currently followed through GeneMatcher,” the preprint authors wrote.

For the broader GA4K effort, the investigators are applying still other analyses to better understand the yet-to-be-diagnosed cases. For example, they hope to develop induced pluripotent stem cell lines that can be used to derive disease-relevant tissue models for functional, transcriptomic, and epigenomic analyses.

“Our current approach, which we’ve been working on with investigators at the genome center here, is to systematically

derive pluripotent stem cells from the patients. Then the stem cell transcriptome and epigenome already increases the ability to look at many disease genes as compared to blood,” Pastinen explained. “And then, of course, the stem cells down the road will give us abilities to look at other [derived] tissue types.”

The availability of RNA sequences and epigenetic profiles could also prove useful for interpreting structural variants detected with long-read sequences, he noted. Due to capacity limitations, long-read genome sequencing is currently being applied to cases that both remain negative after exome sequencing and have DNA from both parents available for analyses, though that may change in the longer term, depending on the diagnostic yield associated with the approach.

Among the 906 families included in the MedRxiv preprint, for example, the investigators detected more than four times as many of those rare structural changes with long-read genomes compared to genome sequences generated with short reads alone.

“We believe that there’s a proportion of variation that is due to genetic changes that are difficult to observe by current ways of clinical genome sequencing, particularly structural variants,” Pastinen explained. “And we have some evidence of that through the early stages of our program: 5 to 15 percent of missed diagnoses that are diagnosable with current criteria are due to these structural variants that are kind of in a blind spot of short-read sequencing, as well as clinical microarray technology.”

Nearly 6,300 individuals from more than 2,700 families have enrolled in the GA4K program so far, and the pediatric genomes that were recently released to public databases represent roughly half of the individuals who have been sequenced, Pastinen said. He noted that the registered access database contains more than 5,000 sequences, which could be released to dbGAP by as early as the end of this year.

The investigators expect to sequence some 30,000 children with rare diseases over the full seven years of the GA4K project. They have already secured around \$18 million from philanthropic sources in the Kansas City area, Pastinen said, and established relationships with some commercial firms.

An additional funding drive is underway, and the team is considering other options — from federal support or commercial collaborations that do not compromise the data sharing spirit of the project to new avenues for reimbursement, he added, all while streamlining testing and trying to bring the cost down.

“Over time, we hope we’ll be able to solve over 50 percent of the cases of suspected rare genetic disease. And through the tools that we develop, and data sharing, [we hope] that we are able to accelerate the current diagnostic odyssey — which is, in our jurisdiction, about four years per patient for a positive diagnosis,” Pastinen said. “We haven’t yet engaged with the payors on the new models of delivering genomic medicine, but obviously, as we go along, we need to get this engulfed in the traditional reimbursement for genomic medicine.”