

# Researchers use machine-learning system to diagnose genetic diseases

April 24 2019

---



An expert in genomic and systems medicine research, Dr. Kingsmore holds the world speed record for achieving the fastest molecular diagnosis using whole genome sequencing in 19.5 hours. Credit: RCIGM

Researchers at Rady Children's Institute for Genomic Medicine

(RCIGM) have utilized a machine-learning process and clinical natural language processing (CNLP) to diagnose rare genetic diseases in record time. This new method is speeding answers to physicians caring for infants in intensive care and opening the door to increased use of genome sequencing as a first-line diagnostic test for babies with cryptic conditions.

"Some people call this [artificial intelligence](#), we call it augmented intelligence," said Stephen Kingsmore, MD, DSc, President and CEO of RCIGM. "Patient care will always begin and end with the doctor. By harnessing the power of technology, we can quickly and accurately determine the root cause of genetic diseases. We rapidly provide this critical information to [intensive care](#) physicians so they can focus on personalizing care for babies who are struggling to survive."

A new study documenting the process was published today in the journal *Science Translational Medicine*. The workflow and research were led by the RCIGM team in collaboration with leading technology and data-science developers —Alexion, Clinithink, Diploid, Fabric Genomics and Illumina.

Dr. Kingsmore's team has pioneered a rapid Whole Genome Sequencing process to deliver genetic test results to neonatal and pediatric intensive care (NICU/PICU) physicians to guide [medical intervention](#). RCIGM is the research arm of Rady Children's Hospital-San Diego.

By reducing the need for labor-intensive manual analysis of genomic data, the supervised automated pipeline provided significant time-savings. In February 2018, the same team achieved the Guinness World Record for fastest diagnosis through whole [genome](#) sequencing. Of the automated runs, the fastest times—averaging 19 hours—were achieved using augmented intelligence.

"This is truly pioneering work by the RCIGM team—saving the lives of very sick newborn babies by using AI to rapidly and accurately analyze their whole genome sequence " says Eric Topol, MD, Professor of Molecular Medicine at Scripps Research and author of the new book *Deep Medicine*.

RCIGM has optimized and integrated several time-saving technologies into a rapid Whole Genome Sequencing (rWGS) process to screen a child's entire genetic makeup for thousands of genetic anomalies from a blood sample.

Key components in the rWGS pipeline come from Illumina, the global leader in DNA sequencing, including Nextera DNA Flex library preparation, whole genome sequencing via the NovaSeq 6000 and the S1 flow cell format. Speed and accuracy are enhanced by Illumina's DRAGEN (Dynamic Read Analysis for GENomics) Bio-IT Platform.

Other pipeline elements include Clinithink's clinical natural language processing platform CliX ENRICH that quickly combs through a patient's electronic medical record to automatically extract crucial phenotype information.

Another core element of the machine learning system is MOON by Diploid. The platform automates genome interpretation using AI to automatically filter and rank likely pathogenic variants. Deep phenotype integration, based on natural language processing of the medical literature, is one of the key features driving this automated interpretation. MOON takes five minutes to suggest the causal mutation out of the 4.5 million variants in a whole genome.

In addition, Alexion's rare disease and data science expertise enabled the translation of clinical information into a computable format for guided variant interpretation.

As part of this study, the genetic sequencing data was fed into automated computational platforms under the supervision of researchers. For comparison and verification, clinical medical geneticists on the team used Fabric Genomics' AI-based algorithms—VAAST and Phevor integrated into the clinical decision support software, OPAL (now called Fabric Enterprise)—to confirm the output of the automated pipeline. Fabric software is part of RCIGM's standard analysis and interpretation workflow.

The study titled "Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation," found that automated, retrospective diagnoses concurred with expert manual interpretation (97 percent recall, 99 percent precision in 95 children with 97 genetic diseases).

Researchers concluded that genome sequencing with automated phenotyping and interpretation—in a median 20:10 hours—may spur use in intensive care units, thereby enabling timely and precise medical care.

"Using machine-learning platforms doesn't replace human experts. Instead it augments their capabilities," said Michelle Clark, Ph.D., statistical scientist at RCIGM and the first author of the study. "By informing timely targeted treatments, rapid genome sequencing can improve the outcomes of seriously ill children with genetic diseases."

An estimated four percent of newborns in North America are affected by genetic diseases, which are the leading cause of death in infants. Rare genetic diseases also account for approximately 15 percent of admissions to children's hospitals.

The RCIGM workflow is engineered to speed and scale up genomic data interpretation to reduce the time and cost of whole genome sequencing.

The team's goal is to make rWGS accessible and available to any child who needs it.

Increased automation of the process removes a barrier to scaling up clinical use of WGS by reducing the need for time-consuming manual analysis and interpretation of the data by scarce certified clinical medical geneticists. There were fewer than 1,600 of these experts nationwide in 2017, according to the American Board of Medical Genetics and Genomics.

Rady Children's Institute began performing genomic sequencing in July 2016. As of the end of March 2019, the team had completed testing and interpretation of the genomes of more than 750 children. One-third of those children have received a genetic diagnosis with 25 percent of those benefitting from an immediate change in clinical care based on their diagnosis.

**More information:** M. Clark et al., "Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation," *Science Translational Medicine* (2019). [stm.sciencemag.org/lookup/doi/ ... scitranslmed.aat6177](https://stm.sciencemag.org/lookup/doi/.../scitranslmed.aat6177)

Provided by Rady Children's Institute for Genomic Medicine

Citation: Researchers use machine-learning system to diagnose genetic diseases (2019, April 24) retrieved 3 July 2023 from <https://medicalxpress.com/news/2019-04-machine-learning-genetic-diseases.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--