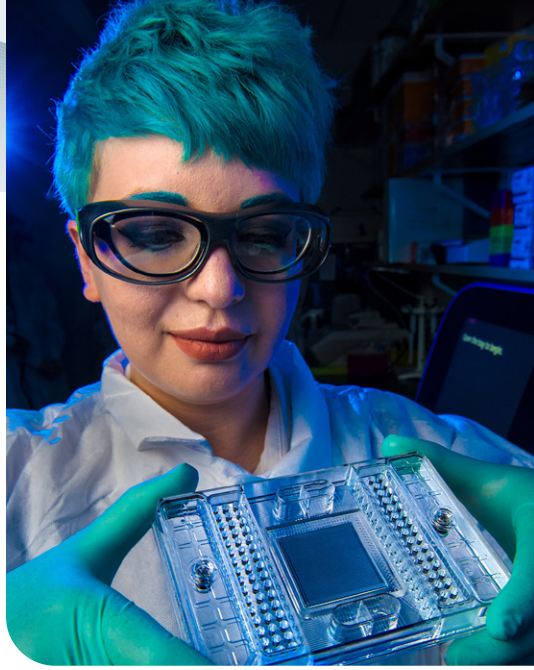
Advanced Molecular Detection

APPLYING PATHOGEN GENOMICS TO PUBLIC HEALTH

2014–2019

For six years, the Advanced Molecular Detection (AMD) program has worked to integrate the latest next-generation sequencing (NGS) technology with bioinformatics and epidemiology expertise across CDC and the nation. Today, pathogen genomics is part of almost every infectious disease program at CDC and has become central to U.S. public health system's efforts to identify, track, and stop infectious diseases.

OUTBREAK INVESTIGATIONS



A CDC public health scientist holds a plate used to amplify small segments of DNA from foodborne bacteria.

AMD methods help link *E. coli* O157:H7 outbreaks to contaminated romaine lettuce

In recent years, CDC has joined with other public health and regulatory officials to investigate troubling multistate outbreaks of *E. coli* infections linked to leafy greens. In one 2019 outbreak, 167 people in 27 states were infected with a strain of *E. coli* O157:H7. Whole-genome sequencing (WGS) showed that the bacteria from that outbreak were similar to the strain of *E. coli* O157:H7 from an outbreak linked to romaine lettuce a year earlier, in the fall of 2018, as well as a strain linked to leafy greens in late 2017. WGS data provided an early clue to investigators to consider leafy greens as a possible source of the 2019 outbreak. WGS revealed that samples taken from environments where lettuce was grown during some of these investigations also contained the same harmful strain of *E. coli* O157:H7 that was making people ill. These data led growers to adopt more stringent irrigation water standards to help reduce the risk of contamination in the future.

When combined with interview data from sick people and product traceback, the use of WGS data strengthens the U.S. food safety system by helping identify food vehicles and by providing information about how harmful bacteria like *E.coli* O157:H7 can contaminate food where it is grown. Since 2013, when PulseNet first began applying WGS in foodborne bacterial investigations, PulseNet and AMD have helped build national WGS capacity to detect and solve foodborne outbreaks with more precision than ever before. Lessons learned from this and other outbreak investigations will help guide industry and support government efforts to prevent similar occurrences in the future.

PulseNet Laboratories Transition to Whole-genome Sequencing

In 2019, the PulseNet national laboratory network completed the six-year transition from pulsed-field gel electrophoresis (PFGE) to whole-genome sequencing (WGS) to combat foodborne diseases more effectively. The shift to this AMD technology provides public health scientists with a highly detailed characterization of foodborne pathogens and helps investigators find outbreaks that they couldn't using the older technology. Scientists from public health laboratories across the United States are now generating, analyzing, and collaborating with the other PulseNet participants.

Illuminating HIV outbreaks with AMD

A community of people experiencing homelessness, many of whom injected drugs, lived beneath an overpass north of Boston. Beginning in August 2016, they sought care at a clinic where they were diagnosed with HIV. By June 2018, the outbreak involved more than 120 people, predominantly in the cities of Lawrence and Lowell in northeastern Massachusetts. The Massachusetts Department of Public Health (MDPH) asked for CDC's help in investigating and responding to the outbreak. A key tool in the investigation was MicrobeTrace.

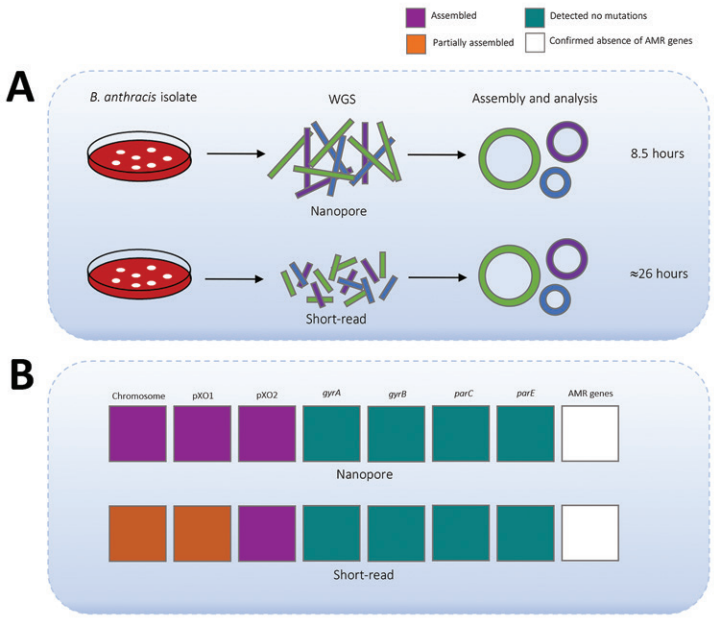
Created with support from the AMD program, MicrobeTrace is a secure, browser-based platform that allows public health workers to create network diagrams of outbreaks. In this investigation, MicrobeTrace helped MDPH define the extent of the outbreak, leading to the discovery that 55 percent more people were affected than the health department had initially realized. This allowed local health officials to ensure that needed public health services were delivered appropriately.

CDC and health departments use molecular analysis to identify transmission clusters that might not be recognized with traditional public health tools and to better understand networks experiencing rapid transmission. These HIV sequences are analyzed with other epidemiologic data, giving investigators a more complete picture of an outbreak. This helps public health departments target interventions more specifically, bring people into treatment and care, and prevent new transmissions.

AMD project provides bioinformatic support for scientists studying *H. pylori*

Infection with *Helicobacter pylori*, a digestive bacterium, has been linked to gastric cancer and ulcers and is common among Alaska Native people and people living in rural communities in Alaska. *H. pylori* can colonize and damage the protective lining of the stomach and establish infections that last for years or decades. To better understand this pathogen, scientists want to study its genetic data to look for markers related to *H. pylori's* virulence and learn why it causes disproportionately high rates of gastric cancer in Alaska Native people.

CDC's Arctic Investigations Program (AIP) leads epidemiology and laboratory efforts focused on *H. pylori*. With AMD support, CDC's Division of Scientific Resources deployed a bioinformatician to work with AIP in Alaska for 6 months. By being on site and working directly with the AIP scientists, the bioinformatician was better able to create a *Helicobacter pylori* pipeline (HPype) to process, clean, and analyze *H. pylori* sequencing data. The bioinformatician modified existing resources to meet specific needs rather than creating something from scratch and established a computer system to track and manage the data. Once all of the HPype analysis modules are complete, AIP will install them on the AMD portal to provide an interface for laboratorians and epidemiologists outside of Alaska so that they, too, can analyze the sequence data and learn more about this infectious disease.



Rapid nanopore whole-genome sequencing for anthrax

Anthrax is a rare but serious disease, requiring rapid identification to reduce illness and death. *Bacillus anthracis*—the bacterium that causes anthrax—is found naturally worldwide but has also been used as a biological weapon. Genetic characterization of *B. anthracis* can help health officials detect whether the samples are resistant to antibiotics or whether they have been genetically modified, information that can inform treatment and be useful in tracing the origin of the infection.

On August 2, 2019, CDC received a *B. anthracis* isolate (Ba0914) from an infected person in Texas who had been exposed to an animal that had died of anthrax. Conventional DNA sequencing takes 26 hours. Conventional antimicrobial susceptibility testing takes over 16 hours. But scientists in CDC's Division of Preparedness and Emerging Infections used a new method and a portable, pocket-sized nanapore sequencer. They were able to sequence the whole genome and confirm that the bacteria did not carry genetic markers for antibiotic resistance or evidence of genetic engineering in only 8.5 hours.

This study demonstrates the speed and usefulness of a portable sequencer. Although the same-day DNA sequencing results are not as precise as second-generation short-read sequencing, CDC had actionable data in approximately 1/3 the time. This can give public health decision-makers the information they need to make rapid decisions during an emergency, when every minute counts.

INFLUENZA

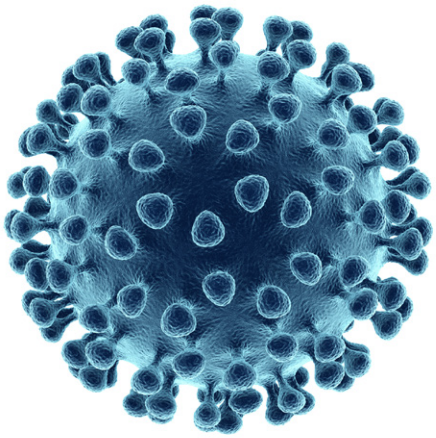
MIA: Mobile Influenza Analysis

CDC's Influenza Division developed a rapid and reliable viral genomic sequencing kit called the Mobile Influenza Analysis (MIA). MIA is a portable influenza (flu) laboratory that can determine which flu strain is circulating on-site. The kit produces real-time data that can be used to evaluate whether a virus is new, what the associated public health risk is, and whether influenza vaccines against the circulating strain are available. These are key components to detecting and responding to new influenza strains and containing outbreaks before they start.

CDC scientists used nanopore sequencing to conduct field surveillance and inform real-time, actionable public health countermeasures for Influenza A at a swine exhibition.

Influenza A Virus Surveillance at a Swine Exhibition Event

In a recent field test at a large swine exhibition, MIA was used to sequence and analyze swine influenza A virus (IAV) samples from pigs. Traditional sequencing methods normally take over a week, given the need to ship virus samples to the laboratory for sequencing. The deployment of MIA for the field test resulted in the on-site identification of three genetically distinct swine IAV lineages in less than 15 hours. The team electronically transferred the sequence data to CDC, where a synthetic candidate vaccine virus (CVV) was developed before the first human case of the outbreak strain was identified. CVVs are flu viruses that have been prepared for vaccine manufacturers to aide them in producing a flu vaccine.



Investigating an influenza outbreak in Louisiana

The 2019–2020 influenza season began early in Louisiana with a large outbreak in the New Orleans area. CDC studied samples from New Orleans and determined they belonged to a specific group of influenza B Victoria viruses. While this group of viruses has been seen before, the amount of flu activity in Louisiana was concerning to CDC's Influenza Division.

A team of CDC staff traveled to the outbreak zone in November to work with Louisiana state public health officials. The team brought MIA , the portable flu laboratory, to investigate the cause of the outbreak. In less than 24 hours, the team tested 89 samples and confirmed that 87 of them were B Victoria viruses. The team also produced full genome sequences of some of these viruses on-site and within hours after sample preparation.

CDC staff then compared the outbreak sequences with recently circulating viruses. They determined that the outbreak viruses were very similar to other viruses seen months earlier in Washington state and were not a newly emerged group of viruses with pandemic potential. This was a reassuring finding to public health officials.

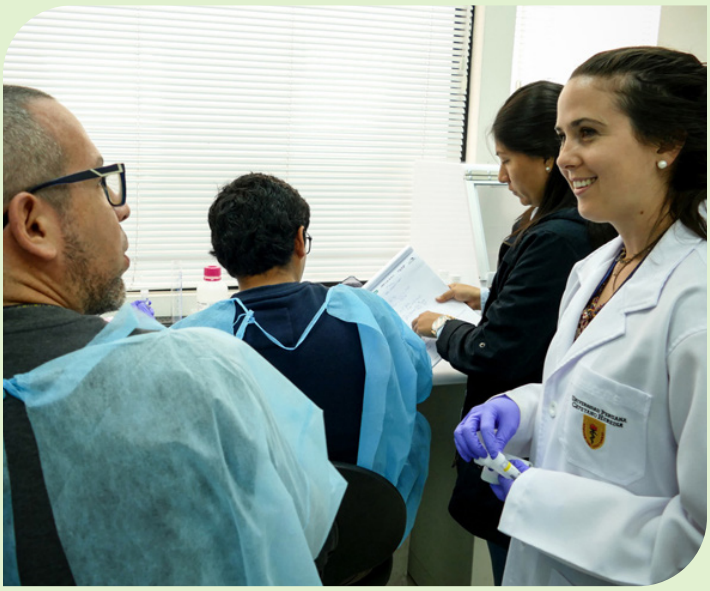
The ability to use sensitive diagnostic and genomic equipment at the site of an outbreak is a major advancement for public health because it greatly decreases response times and allows rapid risk assessments of viral outbreaks. In addition to being useful during outbreak responses, this portable technology can also widen surveillance networks in a more affordable way than conventional genomic approaches. This new technology is a breakthrough for pandemic preparedness and global health security.

GLOBAL HEALTH

Using AMD to fight drug-resistant malaria in Peru

In the summer of 2019, members of CDC's Division of Parasitic Diseases and Malaria (DPDM) led a program to teach researchers in Peru to identify drug-resistant malaria using AMD technologies. In three weeks of sessions, university students, professors, naval researchers, and members of Peru's National Institute of Health worked with trainers from Atlanta to run samples through sequencers in search of antimalarial drug-resistance mutations.

The training was part of CDC's Malaria Resistance Surveillance project, or MaRS, which provides tools for regional laboratories to use. The goal is to standardize and speed up the data flow that tracks drug resistance, making it easier for decision makers to shape their anti-malaria strategies. Identifying the genes that allow some forms of the parasite to survive current therapies can tell doctors which specific drugs to prescribe and improve efforts to battle this disease.



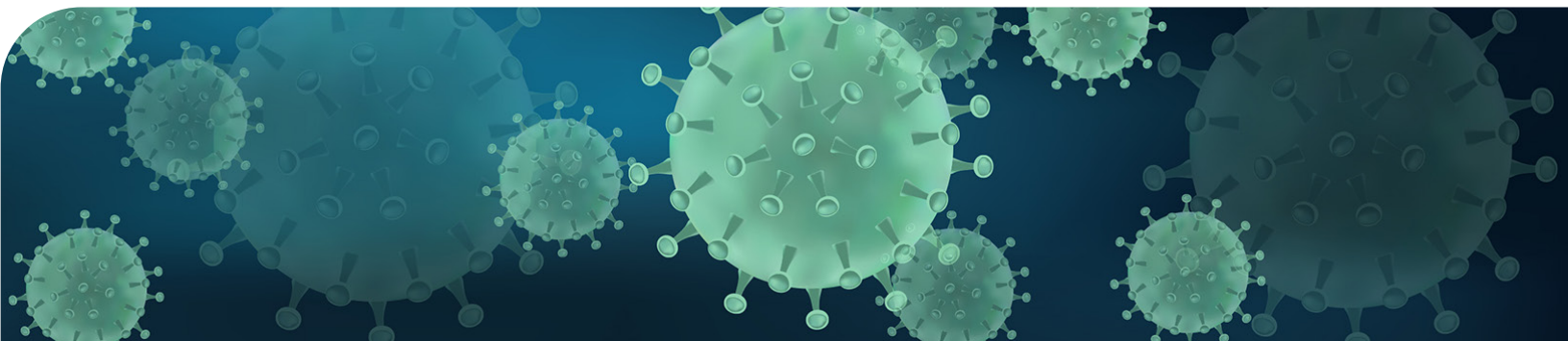
CDC's Julia Kelley, M.S., works with trainees to run samples through sequencers in the Research and Development Laboratories at Peruvian University Cayetano Heredia in Lima, Peru.

Using AMD to quickly predict protective immunity after influenza vaccination

CDC has developed a process that allows scientists to sequence the antibodies from the human B cells that make up part of the immune response to influenza viruses. Sequencing these antibodies provides a high-resolution view of how the human immune system responds to an influenza vaccine, how each specific antibody functions, and what part of the influenza virus they target.

This information can help scientists create more effective tests, treatments, and vaccines for seasonal and novel influenza (flu) viruses, which present a serious threat to global public health. Vaccination is the best way to prevent flu illnesses, hospitalizations, and deaths. However, flu viruses are constantly changing, which means vaccines need to be updated each year to match the cocirculating seasonal influenza virus types and maximize the vaccines' effectiveness. Understandably, there is great interest in developing broadly cross-protective "universal" flu vaccines, or vaccines that would protect against many different flu viruses. CDC's Influenza Division is using AMD tools to analyze how the human immune system responds to the vaccines and whether they create a protective effect.

AMD tools may also help determine whether a person is immune to certain flu viruses. Using a combination of next-generation sequencing, bioinformatics, and molecular cloning, CDC can identify the functions of the individual antibodies and what part of the influenza virus they target. This information can then be used to create flu tests, treatments, and vaccines–possibly even a universal flu vaccine.



WORKFORCE DEVELOPMENT

OAMD joins CSTE and APHL to host AMD Academy

In 2019, the Office of Advanced Molecular Detection (OAMD) welcomed epidemiologists and microbiologists from state and local health departments across the nation to the first year of the AMD Academy, hosted with the Association of Public Health Laboratories (APHL) and the Council of State and Territorial Epidemiologists (CSTE). AMD Academy is a new addition to the AMD portfolio and supports the program's goal to expand workforce development opportunities to its partners.

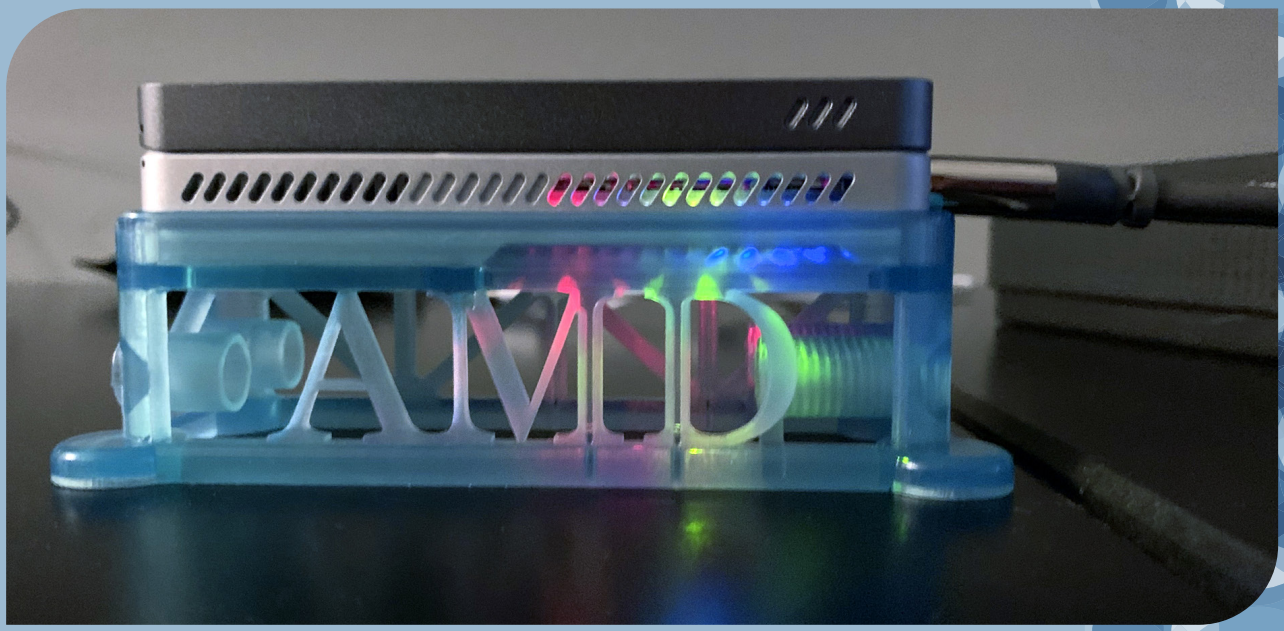
OAMD developed the AMD Academy to meet two emerging needs: a demand for more advanced training among microbiologists and a growing interest in genomics among epidemiologists. Epidemiologists in the program receive an introduction to AMD technologies and learn how to integrate it into their work. The success of the inaugural courses highlights the growing AMD capacity in state public health labs and health departments and the need for dynamic and innovative training options. The third component of AMD Academy—continuing education for state-based bioinformaticians—is scheduled to launch in 2020.

AMD ACADEMY

Building a consensus on genomic software needs in public health

In fall 2018, the AMD program partnered with the CDC Foundation and the Bill and Melinda Gates Foundation (BMGF) on a one-year project to define what software infrastructure is needed to support the complex information workflows required by AMD technologies. This led to the creation of the Public Health Alliance for Genomic Epidemiology (PHA4GE) in the fall of 2019. The project, headquartered at the South African National Bioinformatics Institute, launched at the 2019 Grand Challenges Annual Meeting in Addis Ababa, Ethiopia.

At the meeting, PHA4GE presenters discussed the consortium's goals to establish bioinformatic standards and advocate for greater openness, interoperability, accessibility, and reproducibility in the application of pathogen genomics. The Office of AMD joins with other founding members from over two dozen public health, academic, and private sector institutions that will advise on the critical need for public health infrastructure in this area.



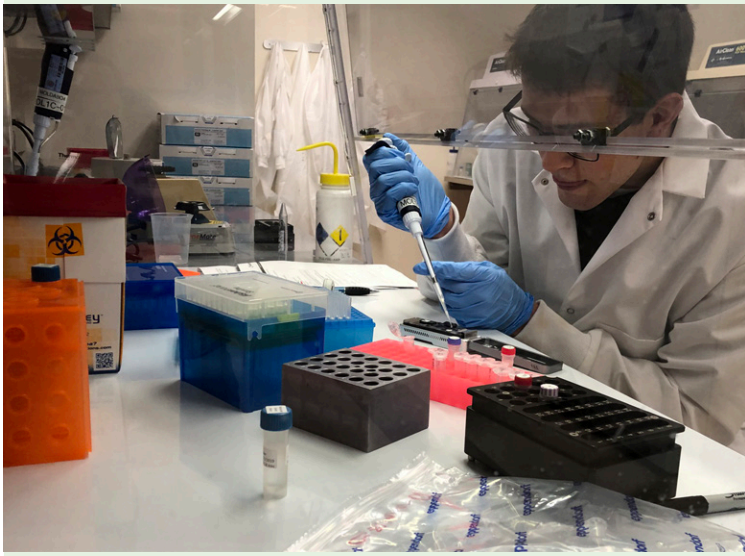
Nanapore sequencing device sits in a customized base during a training session in Utah. Photo taken by Kelly Oakeson

Expanding bioinformatics capacity in state public health labs

The APHL-CDC Bioinformatics Fellowship aims to train and prepare bioinformaticians to apply their expertise in public health settings where they can design tools to help public health personnel use bioinformatic data. This fellowship gives fellows at the post-masters and post-doctoral levels the opportunity to apply their skills to emerging public health issues.

Since the program's inception in 2013, a total of 38 bioinformatics fellows have been placed in host labs at CDC or in state public health laboratories. Starting in 2016, some fellows were also placed in state public health laboratories where they started having an immediate impact. For example, the first fellow at a state public health laboratory helped increase their sequencing capacity from five different types of pathogens to over 20 different types in one year.

Each fellow brings a unique educational background and experience. In several cases, the fellowship introduced the fellow to careers in public health. To date, almost 75% of the fellows still work in public health. They also make an impact beyond their labs, serving as AMD bioinformatics resources for their regions. Collectively the fellows provide technical assistance, analysis, and training to public health laboratory scientists around the country.



Bioinformatics scientist Curtis Kapsak prepares a flow cell at the Colorado Department of Public Health and Environment during his APHL-CDC Bioinformatics Fellowship.

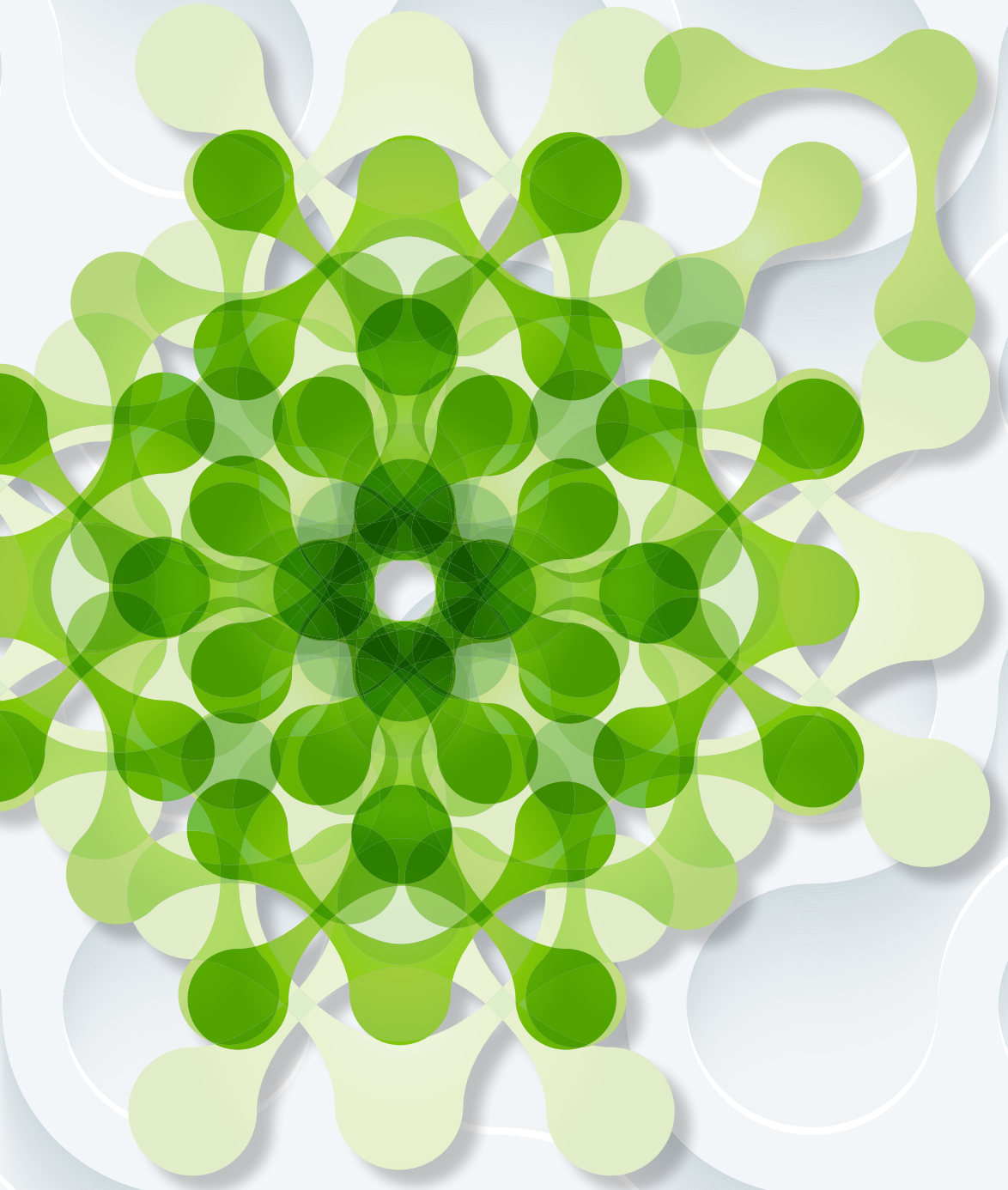
NGS Quality Initiative website goes live

Next-generation sequencing (NGS) is evolving quickly, and even though CDC, APHL, and others have published guidance documents, a need persists for additional laboratory resources.

The NGS Quality Initiative is developing an NGS-focused quality management system to address emerging challenges that public health and clinical laboratories are encountering as they develop and implement NGS-based workflows. The project is funded by the Office of AMD and is co-led by the CDC Division of Laboratory Systems, the CDC Office of the Deputy Director for Infectious Diseases, and APHL.

NGS Quality Initiative

www.cdc.gov/amd



CS314202O