The Microsoft Cloud empowers organizations of all sizes to re-envision the way they bring together people, data, and processes that better engage patients and customers, empower care teams and employees, optimize clinical and operational effectiveness, and digitally transform health.

Partnering to Advance Clinical Genomics

Table of contents

"We've done medicine the way we've done medicine because we haven't had the benefit of genomics," said Dr. Simon Kos, Chief Medical Officer at Microsoft.

"We prescribe a therapy based on how the average population will respond, but for individual patients, that might not be the right therapy—which is cold comfort for you as an individual patient when you get a therapy that doesn't work. Over time, we've managed to elevate medicine to a science, but right now that science is driven by aggregate evidence rather than personalized evidence."

Advancing genomic understanding

The availability of more powerful tools is driving advances in genomics research and applications that are bringing tailored, personalized care plans for patients into greater focus. "When you've got genomic information about a patient, with all of its implications understood, and add the patient's phenotypic and demographic information, then you start to get the real picture," said Kos.

A more in-depth understanding of how individuals who have certain genes might respond to certain therapies will open the doors to precision medicine, a tantalizing prospect to clinicians. Technology and medical research advances are combining to dramatically improve the timeliness and cost effectiveness of genomics, making it relevant in clinical scenarios, even for medical emergencies.

Of course, the ability to tailor treatments much more specifically based on an individual's genomic profile requires more than their genetic data. "It's not enough to have the genome," said Kos. "We've got to know what the genome means, and that requires collaborative research." Since volumes of genomes must be compared to yield insights, the more genomes that are available, the more accurate and informative the analysis will be.

"Any one organization may have petabytes of genomic information, which is huge, but it's just a fraction of all of the genomes that have been sequenced in the world," said Kos. "Because these are large file sizes, because historically they've been held on-premises, and because the data is highly sensitive, there's no way to aggregate the world's genomic information to do de-identified ethical studies into what these genomes mean. We need large amounts of data, aggregated in a single place, secure but accessible. That's what our cloud platforms afford."

Infinitely salable, readily available resources in the cloud

As they gather information from more and more patients, researchers and clinicians may find that the volume of data outstrips their in-house resources. A paper published in the journal ***PLOS Biology*** reports that "as much as 2–40 exabytes of storage capacity will be needed by 2025 just for the human genomes."* Attempting to manage that level of information on-premises can quickly become a perfect storm of problems: massive storage requirements, massive compute requirements, and strict requirements around security compliance, privacy and data sovereignty.

Fortunately, the cloud, with its practically unlimited ability to scale, presents a viable alternative. The cloud can automatically provide additional storage in minutes —in other words, as soon as it's needed. In contrast, the process of requisitioning, purchasing, and setting up additional disk space on-premises can take months in many organizations.

Computing resources are similarly scalable on demand. Translating raw genomic data into useful information requires extensive processing that only supercomputers or cloud computing resources can handle. While only a small number of elite research centers, major medical centers, or large pharmaceutical companies have access to super computers, any organization—from the smallest startup to the largest enterprise—can access the cloud.

Cloud services can be resources, like utilities, that genomics solutions draw on only as required, keeping costs down. Industry experts who run datacenters for a living keep the underlying software current with the latest features and security patches as soon as they become available, ensuring that clinicians and researchers always have immediate access to the latest technology. Moreover, cloud providers take physical, technical, and operational measures to achieve and maintain security and privacy standards that meet compliance requirements.

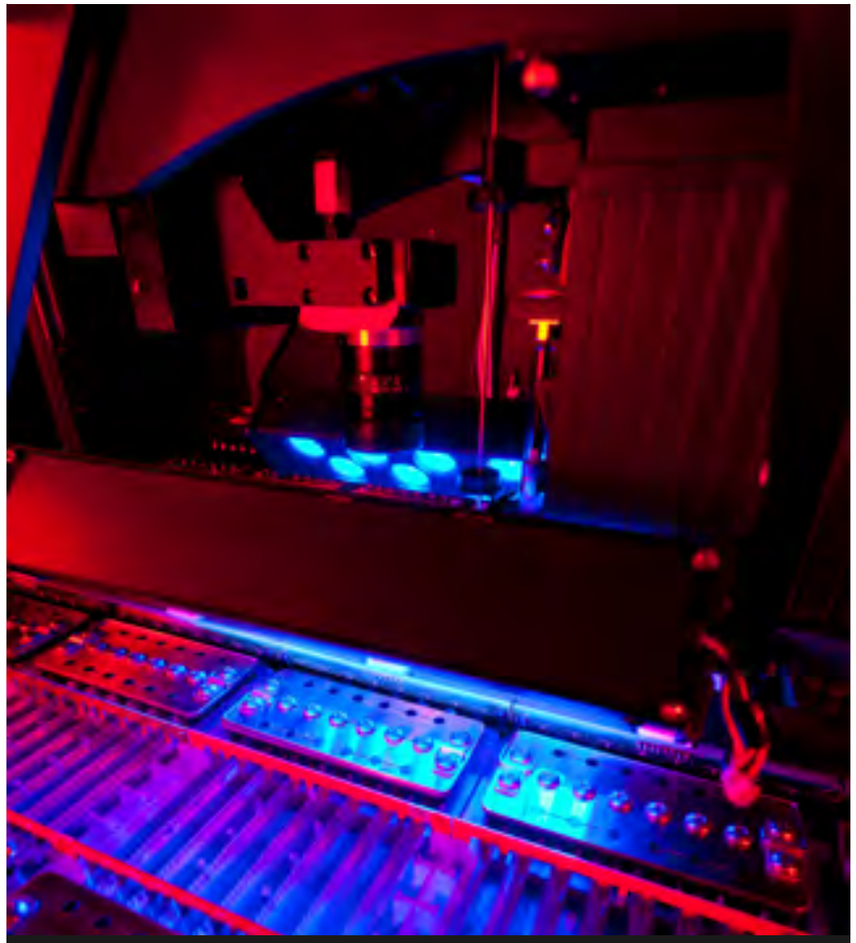
Learn more about privacy, security, and compliance in the Microsoft cloud

"Someday, with the added speeds and feeds from cloud infrastructure, we'll be able to sequence and analyze a full genome for less than a hundred dollars," said Kos. "As more sequenced genomes become available, genomics will move into the mainstream. Your primary doctor will be able to use your genomic information for practical purposes."

Microsoft's investment in genomics

In 2007, Microsoft established a team focused on genomics in its Artificial Intelligence and Research division. This team of engineers, data scientists, and scientists, led by Geralyn Miller, works hand in hand with Microsoft's cloud computing team to create genomics-specific platforms, consistently running heavy, long-running workloads that push Microsoft's cloud platform, Azure, to its limits. "We're the only cloud provider that has a research team dedicated to genomics," said Miller. "We understand the needs and concerns of researchers because we're researchers ourselves." To advance genomics, the team takes an open-source approach, sharing much of its work with the research community.

In addition to the work her team has done to improve secondary and tertiary analysis, they've helped integrate genomics-specific analytical tools and capabilities into the Microsoft platform. "Combining these tools with the research we've done makes our results significantly more powerful," Miller explained. "In essence, we're blending data science and artificial intelligence with core bioinformatics."



Microsoft Research offers Factor Spectrally Transformed Linear Mixed Models (FaST-LMM), a set of tools for genome-wide association studies (GWAS). The tools help researchers handle huge volumes of data by cataloging and comparing nucleotide variations at specific positions.

Learn more here.

Project Premonition seeks to detect pathogens in the wild before they cause outbreaks of disease by capturing mosquitoes and sequencing the genomes of the mosquito, the blood they have ingested, and any viruses they are carrying. The project is using drones to find mosquito hotspots and breeding sites and catches specimens with robotic traps.

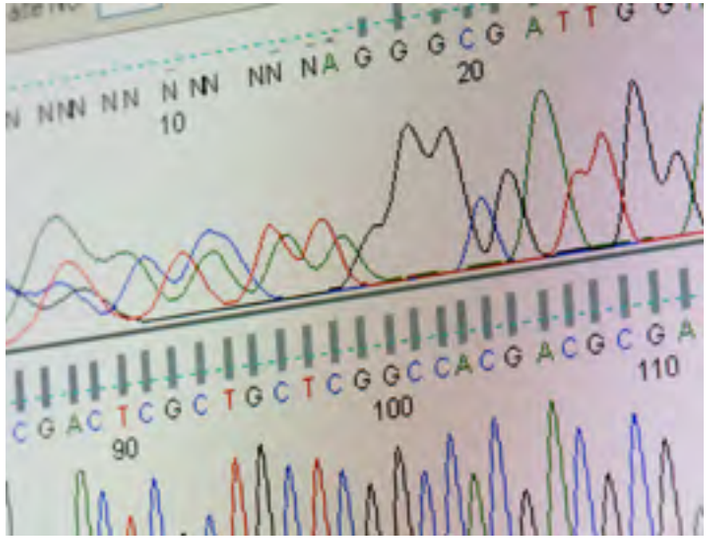
A core focus of Microsoft's genomics team is to develop analytics, machine learning, and artificial intelligence algorithms to aid clinicians who want to make more accurate diagnoses more quickly, and reduce trial-and-error when prescribing treatments.

The goal is not only to inform a patient's current condition, but also predict, based on data collected from the patient and similar subpopulations, how their condition may evolve, and help suggest the optimal course of action.

"For example, just because you have lung cancer doesn't mean all lung cancers are the same," Miller explained. "Tumors don't always identify with the tissue of origin. Today, clinicians generally start with the treatment that's most commonly used. With genomics, you can get more specific about the type of cancer and the best type of treatment. If you can tell upfront whether a person has an inherited tendency for an adverse reaction to a drug, you can make a better choice as the doctor. It's all about taking the raw data and getting to an insight that's actionable in the context of a care situation."

Partnering for success

Building on its collaboration with researchers, Microsoft has developed a vibrant ecosystem of solution partners who focus solely on genomics. As these partners build solutions based on input from researchers and scientists, they help Microsoft further develop and accelerate commercialization of their cloud-based platforms, optimizing services for storing and analyzing vast amounts of genomics data.



Curoverse: pooling genomic data while protecting privacy

"The number of genomes being sequenced today is growing at three times the rate of Moore's Law," said Keith Elliston, Chief Commercial Officer for Curoverse, which provides open source genomics enabling technology. "You're not done once you've sequenced a single genome. The germ line genome is just the base map. Things change. When you develop tumors, that's a change. Then there's the whole microbiome, which is all the bacteria living on your skin, in your gut and in your hair, in your nose. These have a big impact on what's happening. You look at terminally differentiated cells. Your immune system has rearrangements of the whole genome in every cell. You have all these genomes that become very important, so we're not talking about one genome per person. We're talking about literally thousands of genomes per person."

Learn more about genomic and biomedical computing solutions from Curoverse

Since single genomes don't yield insight in isolation, the ability to amass genomic data, and then share it broadly and easily, is particularly important. But moving data into a single repository—even within the cloud—isn't practical.

"If you have a hundred thousand genome centers in the world, and they each have a hundred thousand genomes, the model of bringing all of them into one database will never work," said Elliston. "It's all going to live in the hundred thousand different centers out there. Your genome sequence is the most private and personal data you have. It's really the fundamental blueprint that makes you you, and from that perspective, there are lots of laws about how that data must be protected. If you put data up on the cloud securely and privately, you're not going to mix it with anybody else's."

The cloud makes it possible to assemble genome data from multiple organizations into a data lake, while retaining the appropriate privacy and security controls. Cloud-based algorithms can run on individual datasets regardless of whether they're cloud-based or stored in localized datacenters.

While the algorithms generate insight, the genome data stays where it is. "When our platform, Arvados, does genetic variation analysis," Elliston explained, "it federates the data in a way that encapsulates the analysis in a workflow, which we can distribute across the datacenters where those data live. We do the analysis, aggregate the results, and only export metadata. We don't move the private individual data. Even on the same cloud infrastructure, it's stored in distinct places. With Arvados, it can still be analyzed with everyone else's data, while respecting all that privacy and ownership."

Microsoft has a unique ability to do hybrid cloud installations using local datacenters in combination with Azure. Microsoft can deploy the **Azure Stack** on-premises at the customer site or at the facility of a hosting partner. The clinical data remains securely in the local data warehouse to comply with data sovereignty requirements, while the processing and analysis takes place in the Azure cloud.

BC Platforms: turning raw genomic data into analytical insights

It took 13 years and USD $2.7 billion to sequence the first genome. Steady improvements in algorithms reduced sequencing time to 28 hours. Algorithms Microsoft has developed with the Broad Institute can yield results in just four hours.

Even with faster sequencing, however, the resources required to process the millions of genomes that will be necessary to yield a broad range of useful clinical insights are gargantuan. "When you go from ten thousand genotype patients to one hundred thousand to one million patients," said Nino Da Silva, Executive Vice President of BC Platforms, a provider of bioinformatics and genome data management software headquartered in Switzerland, "it will take 2.4 million CPU core hours just to impute the raw data. And at this point we haven't even started analyzing."

Fortunately, the cloud makes such intensive computing requirements manageable without a large, upfront investment in infrastructure. The calculations required to sequence genomes, and then analyze them, can take place in the cloud, even if sensitive patient data remains on-premises. The role of companies like BC Platforms is to offer end-to-end solutions that provide ready access to genomic information, even for clinicians who aren't geneticists, or don't have a high level of training in genetics.

"We take the data and make it understandable," Da Silva explained. "When a patient comes in with a referral for a specific question, such as whether they have the BRCA breast cancer gene, our system will take the raw data, align and create the usable file (gVCF) and from there create the variant core report. We then inject the report into the electronic medical record or lab system without any manual intervention, so doctors can see the results where they need it, with the rest of the patient data." The genomic information becomes one of many tools that clinicians use in treating their patients.

Microsoft's method of running the Burrows-Wheeler Aligner (BWA) and the Broad Institute's Genome Analysis Toolkit (GATK) on its Azure cloud computing system is seven times faster than the previous version, allowing researchers and medical professionals to get results in just four hours instead of 28.

Learn more here.

Learn more about genomic data management and analysis solutions from BC Platforms

"Once we've optimized the whole flow from the diagnostics devices in the laboratory to the patient records, we'll need to form a standard of what's going to be reported and how it will be used in terms of the actual patient care," said Da Silva. "It's also important to understand that clinical decision makers, university researchers and potentially commercial partners can all benefit from different pieces of the same knowledge base. We must ensure that doctors and patients always have the most up-to-date, personalized, valid interpretation available."

Da Silva believes that at some point, clinicians will rely on genetic data about patients as a matter of course. "Genomics is like any other medical system," he said, "and I think we should perceive it that way. Nobody would today consider replacing a hip without doing an X-ray or an MRI first and interpreting it. I think genetics will become the same, that nobody will treat a patient for cancer without having specific genetic information. Not doing so will be considered as uninformed as doing hip surgery without an X-ray."

DNAnexus: employing genomics in clinical settings

Dr. David Shaywitz, Chief Medical Officer at DNAnexus, a Mountain View, California company that offers a cloud-based genome informatics and data management platform, agrees that genomics will become a standard component of clinical evaluation. "There's an increased recognition of the relevance of genetics as people have a better understanding of it and as the tools have become basically faster, cheaper, and better," he said.

Yet many clinicians remain skeptical about genomics. "For over a generation, physicians, even in medical school, have been hearing about the promise of genetics, and have caught glimpses of ways genetic diagnosis and genetically informed treatments could really be helpful," Shaywitz explained. "But there's

always been a disconnect between the promise of genetic medicine and the reality of precision or personalized medicine." The successful use of genomics to improve diagnoses, drug selection, and outcomes for patients with specific cancers, such as breast cancer, is encouraging skeptics to reevaluate.

"What's so exciting now are the specific tools available for patients," said Shaywitz. "There are more and more clear use cases—clear opportunities where you can demonstrate that patients' lives were dramatically improved through the use of genetic information."

For example, it's estimated that about one-third of the children under the age of one who are hospitalized suffer from a genetic condition, and that 25 to 40 percent of mortality in high acuity neonatal intensive care units (NICUs) may be attributable to a genetic disorder. "An accurate genetic diagnosis—which in some cases can now be achieved within about 24 hours —can immediately influence the direction of care, and also avoid a prolonged and often excruciating diagnostic odyssey," said Shaywitz. "While the primary driver is, as always, humanistic, that is, it's the 'right thing to do,' there are also immediate economic benefits, given the high cost of keeping children in the NICU. The result is that appropriate, early use of genetics in these cases can be a win for everyone."

Learn more about sharing and management of genomic data and tools to accelerate genomics from DNAnexus

DNAnexus has developed a platform for local and distributed sequencing and analysis, built on the Microsoft cloud, that enables an end-to-end, rapid genomics solution to support both researchers and clinicians. "Oncologists seeking to provide the best care for their cancer patients are going to be performing molecular evaluations," Shaywitz predicted, "and cancer centers that want to provide the best care for their patients will want to collect and organize this information—both the genetic information and the patient journey information." Over time, as more medical centers and research facilities around the world gather genetic data, it will become more useful for identifying the best treatments options.

DNAnexus, in partnership with Microsoft, supports **The Stanford Center for Genomics and Personalized Medicine** (SCGPM), which is creating a new paradigm of patient-centered medicine that encompasses the entire genome of individuals "to improve disease prediction, prevention, and treatment of conditions such as cancer, diabetes, heart disease, asthma, schizophrenia, and many others." Efforts include reducing the time to achieve effective treatment, integrating bioinformatics with the patient's electronic medical record, eliminating trial and error in prescribing drugs, identifying disease risks, offering patients genetic counseling and engaging them to enhance their compliance with prescribed treatments.

"Ideally, we'll have a learning healthcare system," said Shaywitz. "We'd learn from the clinical journey of the first patient integrating their clinical and genetic information, and apply it intelligently so that we treat the next patient in a fashion that reflects the knowledge gained from each previous interaction. It will be a virtuous cycle."

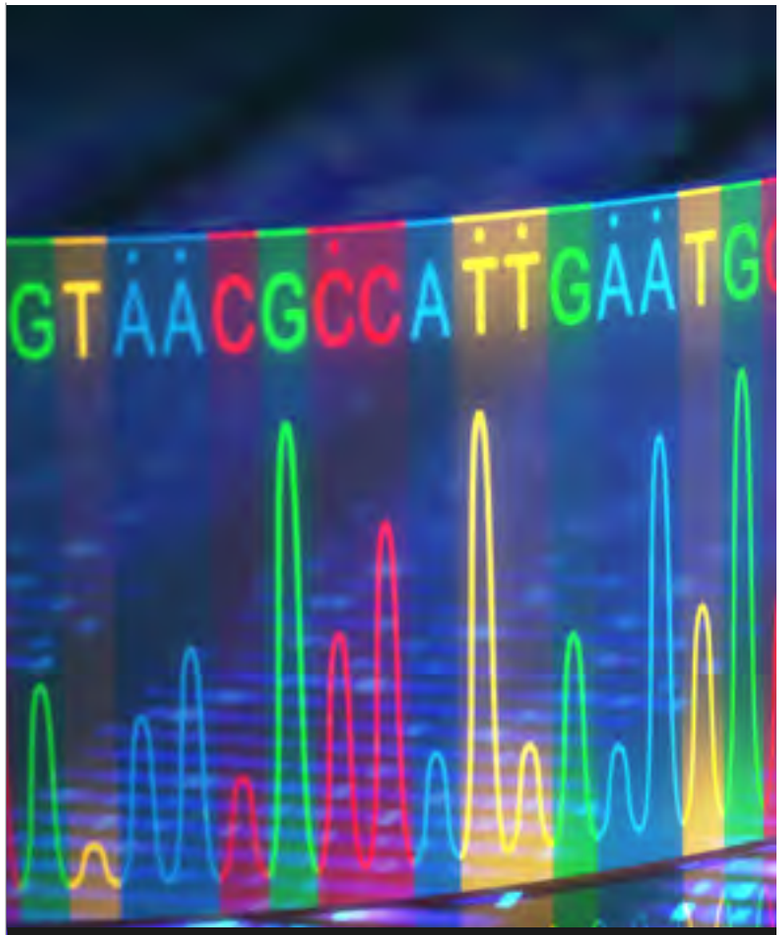
Democratizing genomics

The successful use of genomics to improve diagnoses and outcomes for neonatal infants and patients with specific cancers, such as breast cancer, is encouraging efforts to develop more genomic information.

Microsoft and its partners are hard at work making this information easier to sequence, store, analyze, and digest, thus transforming the promise of genomics into real-world evidence that clinicians can use to improve care.

"We're taking our research investment and getting the IP into the hands of our customers and partners through a series of Azure-based offerings," said Miller. "Our goal with our partners is to offer a turnkey way to do the bioinformatics pipeline—to reduce complexity. When you run analyses across large sample sets through the same pipeline, you end up with more consistent data. In essence, we're democratizing genomics."

Learn more about Microsoft's investments in genomics.



Project Hanover, is a Microsoft Research effort that seeks to empower biomedicine with cutting-edge computer science, focusing on three areas: machine reading of the millions of bio-medical articles to form queryable knowledge bases, cancer decision support using AI with a current focus on Acute Myeloid Leukemia (AML), and the use of AI in chronic disease management.

Benefits of cloud computing

Trust. Security is a big concern when it comes to cloud adoption. Health data needs to be protected. The industry experts who manage cloud data centers take physical, technical, and operational measures to achieve and maintain security and privacy so your organization can meet the compliance requirements of regulators. They deploy the latest security patches as soon as they're available, so you get them automatically.

Speed. Traditionally, getting access to new or better tools, more storage space, or more computing power has involved a lot of waiting. This happens when IT has to get budget, purchase hardware and/or software, and set it up before they can give you what you need. With the cloud, it takes minutes instead of days, weeks, or months, to get more storage space, compute cycles, or updated software.

Scale. The cloud provides an almost infinite set of computing resources, which is necessary for storing or processing massive amounts of data. When you need more resources, you can scale up instantly and automatically. When you need fewer resources, you scale down instantly and automatically. In other words, you get exactly what you need, when you need it—no more, no less.

Economics. Few organizations have the resources to set up or manage very-largescale data centers. The cloud model, like a utility, is pay-as-you-go. This way you never have to worry about having fewer resources than you need, and your organization never has to worry about paying for more than you're using.

Choosing a cloud offering

Health organizations need cloud services that are not only powerful, but also secure and compliant. Here is a short checklist to consider when evaluating cloud offerings:

- Is it Enterprise-ready? Do the services it offers scale to large organizations with many users or customers?
- Does it offer more than infrastructure, such as platforms that support core functions for productivity, communications, relationship management, and analytics, so you can easily tailor solutions to meet your needs?
- Does it work with on-premises environments and with technologies from multiple cloud providers and device manufacturers?
- Does it have certifications to show it complies with health security and privacy regulations in the countries where you operate, so you can safely store and process personal health information?

Microsoft

microsoft.com/health

* Zachary D. Stephens, Skylar Y. Lee, Faraz Faghri, Roy H. Campbell, Chengziang Zhai, Miles J. Efron, Ravishankar Iyer, Michael C. Schatz, Saurabh Sinha, Gene E. Robinson. (2015, July 7). Big Data: Astronomical or Genomical? Retrieved from PLOS Biology: http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002195