Sequencing drives innovation

The Stanley Institute for Cognitive Genomics is home to 10 Next-Gen sequencing workhorses—HiSeq 2000 machines—with the capacity to sequence 3000 genomes a month.

It took 10 years and $2.7 billion to sequence the first human genome. Today, that same sequence can be read in a single week for one-millionth the cost, as little as $3000. Genome science has arrived.

What does this mean for biology? Because of astonishing advancements in sequencing technology, an entirely new field called genomics has come to life. It has enabled scientists to spell out the genomes of organisms across all branches of the tree of life, from yeast to plants to animals. The broad goal of this young field is to understand how changes in genetic information affect life. It is applied to single cells, individual organisms, even entire species. Genomics enables researchers to understand how breast tumors develop from an initial set of genetic errors. It also enables scientists to reach back through eons of time to understand how species evolve.

In labs around the world, sequencing technology has become a staple of basic and applied genomics research. At Cold Spring Harbor Laboratory, the same technology is also being used in creative and unconventional ways that reflect this institution's unique collaborative culture. The Laboratory has had the foresight to plan and build a cutting-edge genome-sequencing "core" facility that is available to every member of the faculty and, unlike most sequencing facilities elsewhere, is fully integrated into every facet of research on the campus.

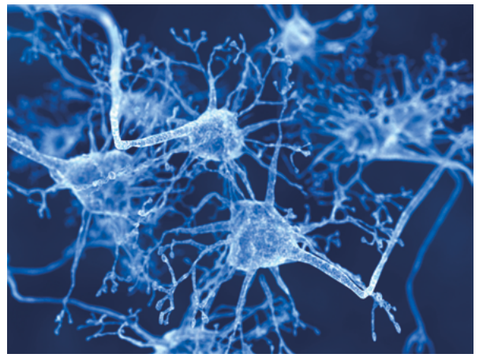
Moving beyond the straightforward compilation of genomic sequences—"read-outs" of the myriad individual DNA "letters" that make up individual genomes—CSHL scientists are developing new sequencing applications that have already generated impressive scientific results. They have found a gene that substantially increases tomato yields. They have identified a DNA element that pushes leukemia cells to keep growing. They are exploiting sequencing expertise to draw a complete circuit diagram of the mouse brain. And they are developing new diagnostics to improve cancer treatments that will cost as little as $10.

"Our core facility, the genome sequencing facility at Woodbury, has changed the way we collectively think about science," says Professor Richard McCombie, a pioneer of genomics and genome sequencing and director of the facility. "Because of our small size, we are flexible, cost-effective, and constantly at the forefront of technology. If somebody has an idea, we can jump on it—which provides a huge advantage to our researchers in this rapidly evolving scientific landscape."

The dawn of Next-Gen sequencing

Our genetic information is stored within our DNA, a long molecule that looks like a twisting ladder, whose rungs are made up of four chemicals called bases. The full human genome is composed of 3 billion bases, whose sequence, among other things, encodes some 20,000 genes. Roughly one out of every 1,000 bases varies from person to person. This means we differ from one another by about 3 million bases over the length of the genome. Inevitably, these single-base variations affect the way our genes are used (or "expressed," in the language of biologists). These differences are also what make each of us unique.

The first genomes were sequenced with technology that now seems quaint. In the early 1990s, researchers at CSHL worked in shifts around the clock to load DNA samples onto machines. At full capacity, they might be able to read 1–2 million bases a day, amounting to just 60 million bases a month. At that rate, it would have taken 50 years to read a full human genome.



The mouse brain contains more than 10 million neurons, and each makes nearly 1000 connections with its neighbors. CSHL Professor Tony Zador is developing a novel sequencing-based approach to build a full circuit map of the more than 10 billion neural connections in the brain.

3 billion
number of letters in the human genome

1.2 million
equivalent number of pages in a book

"Millions of bases may sound impressive," says McCombie, who also heads CSHL's Stanley Institute for Cognitive Genomics, "but today, the technology has advanced so that we are reading just shy of 10 *trillion* bases a month." That is the equivalent of more than 3,000 genomes.

The remarkable technology that makes this feat possible is known prosaically as Next Generation Sequencing (NGS, or "Next-Gen"). On average, it takes a little less than two weeks to sequence a single human genome, and each machine can process six genomes at once.

"We have 10 'Next-Gen' machines in our Woodbury core facility," says McCombie. Each one costs about $750,000, a steep price for individual researchers to pay on a lab-by-lab basis. "We pooled our resources—both in terms of finances and brain-power—so that our researchers have tremendous access to sequencing power, better than some of the largest institutions in the world."

"Our genomics investigators benefit from the large number of machines at their disposal, while scientists with smaller projects—even a single experiment—also have direct access to this technology, which is unusual." This, in turn, encourages creative applications of the technology, two of which are described below.

Using sequencing to map the brain

In one example, CSHL Professor Tony Zador, a neuroscientist, is using Next-Gen sequencing to determine how all of the neurons in the brain are connected. A map of the "connectome" will provide us with a better understanding of how the healthy brain works. But also, notes Zador, "we are beginning to understand that autism, schizophrenia and many other mental disorders are actually wiring problems in the brain. Projects like this one will help us pinpoint what goes wrong, and develop treatments for these illnesses."

The project, though simple in concept, is actually daunting in technological terms. There are more than 10 million

neurons in the mouse cortex and about 1000 synapses, or communications bridges, between *each* neuron. This means that scientists must map 10 billion connections to get a good idea of how the mouse brain is wired. "The conventional approach relies on microscopy to create this map," says Zador. But these projects are highly labor intensive and come at the tremendous cost of more than $10 million per year.

Zador has devised a method to genetically tag individual neurons with short sequences of DNA. This is akin to stamping a barcode on each one. In a trick of genetic engineering, Zador's team has found a way to drag the barcodes to the synapses where neighboring cells exchange messages. There, the barcodes are glued together. The fused DNA barcodes then can be isolated and sequenced with Next-Gen sequencing technology, just like any other DNA sample. Mathematical programs called algorithms enable Zador's team to make sense of the sequences in order to map connections not just between two cells but throughout the brain. If successful—it remains in the proof-of-concept stage—barcoding and Next-Gen sequencing may enable the team to generate a full-brain map for as little as $10,000 per brain, rather than tens or hundreds of millions.

$2.7 billion/10 years

to sequence the first human genome

$3000/2 weeks

to sequence six human genomes today

The CSHL sequencing facility is what turned this project into a reality. "I absolutely would not have thought of this project or been able to pursue it anywhere else," says Zador. "With all of the open discussion here at the Lab between people in different fields, I was able to see the power of sequencing and all it can do."

Sequencing single cells to diagnose cancer

CSHL scientists are also using Next-Gen sequencing to revolutionize cancer research. "Our goal has been to develop new kinds of diagnostics—to inform clinicians about the type of cancer cells they are treating so they can choose the best therapeutics," says Research Professor Jim Hicks, one of the lead scientists on the project.

A single tumor is made up of many different types of cells. Some may be susceptible to specific cancer treatments while others may be resistant to these same drugs. Doctors currently must biopsy a tumor in order to identify the aberrant cell types it harbors. Even then, pathology reports only provide a limited overview of the cancer.

Hicks and other members of Professor Mike Wigler's lab made a breakthrough when they developed methods to sequence cells one at a time. "We can extract DNA from blood samples or urine, so a minimally invasive blood test can replace a biopsy."

The next challenge has been to distinguish one type of cancer cell from another. Hicks and colleagues found that, in individual cancer cells, regions of the genome are duplicated or deleted. "These changes, called copy number variations, can be used to identify different types of cancer cells." This discovery means that it is no longer necessary to sequence the entire genome. Rather, Next-Gen sequencing is employed as a simple counting tool. "You can think of it as a small survey of the genome that lets us see where regions are deleted or duplicated," says Hicks. The advantage is that you can identify populations of cancer cells with a fraction of the sequencing data.

"The cheapest full genome sequence is at least a few thousand dollars, but we have devised a way to determine the origins of a cell for just $10," says Hicks, who hopes the work will lead to a marketable tumor diagnostic in the next few years.

Jaclyn Jansen