YALE JOURNAL OF BIOLOGY AND MEDICINE 80 (2007), pp.191-194.


PROFILE

# 454 Life Sciences: Illuminating the future of genome sequencing and personalized medicine

Kristin L. Patrick

*Yale University School of Medicine, New Haven, Connecticut*

Imagine going to the doctor with a migraine headache and knowing the next day whether your DNA is to blame. Or better yet, imagine knowing whether your DNA encodes a risk for migraines even before a headache strikes. If 454 Life Sciences' dream is realized, doctors will be able to quickly and inexpensively unlock the mysteries of patients' genomes, revealing indicators of disease susceptibility and shedding light on the potential efficacy of particular drug regimens. Due to advances in technology made by companies like 454 Life Sciences in Branford, Connecticut, personalized genomic medicine is becoming a reality and is sure to revolutionize not only your doctor's visits, but also the way society views human health and disease.

Founded in 2000 with the mission of making individual human genome sequencing a reality, 454 Life Sciences has proven itself a front-runner in the race toward personalizing medicine. As the story goes, 454's founder Jonathan Rothberg, PhD, was faced with a family health emergency. His newborn son was ill, and as he waited to hear news from doctors, he pondered the idea of pinpointing his child's condition by reading his genome — and, thus, his quest to provide fast, affordable genome sequencing began.

The first gene sequencing technology was developed by Frederick Sanger in 1977. The innovation behind the Sanger method involved the use of dideoxynucleotide triphosphates to terminate elongation of a DNA primer (a short fragment complementary to a single-stranded DNA template of interest). Variations of the Sanger method have contributed to almost all of our current knowledge of gene and genome sequences. However, the technique has certain limitations. In particular, large-scale projects like the sequencing of full organism genomes by the Sanger method involve the rather laborious task of creating a library of genomic fragments in bacterial plasmids, which then are sequenced and assembled. Costly reagents and time-consuming sample preparation make this approach unfeasible for use in a clinical setting, where cheap, fast, and accurate sequencing is demanded.

The Sanger method was used for the Human Genome Project and was the gold standard for decades — until 1998, when scientist Mostafa Ronaghi at Stanford University [1,2] worked out the chemistry behind the technology of pyrosequencing, the method eventually adapted by 454 Life Sciences. Pyrosequencing refers to a series of

enzymatic reactions during which incorporation of DNA bases is measured by the release of visible light. Briefly, a DNA template is incubated with several enzymes, including DNA polymerase and ATP sulfurylase. The subsequent incorporation of one of the four deoxynucleotide triphosphates, or dNTPs, complementary to the DNA template strand, releases pyrophosphate, a compound which then is converted to ATP by ATP sulfurylase and fuels the luciferase enzyme to release protons. The light released by the addition of each new nucleotide is detected and recorded, and the process is repeated to extend the chain of sequenced DNA.

The 454 sequencing technology takes pyrosequencing one step further, allowing for bulk sequencing of an entire genome [3]. First, genomic DNA is fragmented and ligated to specific adapter molecules, which are used as templates for primers in a polymerase chain reaction (PCR). The PCR reaction, carried out on synthetic beads in an oil emulsion, takes a single piece of DNA and replicates it until 10 million copies of a discrete DNA molecule are bound to each bead. Next, each bead is deposited into an individual well of a fiber-optic slide, where it meets a cocktail of the enzymes required for the pyrophosphate reaction. These steps take the place of the laborious task of cloning individual DNA molecules and eliminate biases that can be introduced by cloning a population of fragments (transformed bacteria will preferentially replicate certain sequence stretches over others). A loaded slide is then fed into the sequencing instrument, which washes deoxynucleotides over the plate, extending the chains of DNA templates in each well and promoting photon release. A computer records the release of light, logs the sequence of the DNA in each well, and eventually interprets these data to align smaller bits of sequence into a full genome sequence. In 2005, 454 Life Sciences released the genome of *Mycoplasma genitalium*, the first organism sequenced by this technology [3].

Currently, a machine called the Genome Sequencer FLX does the dirty work at 454 Life Sciences. The strength of this instrument is best expressed not by measures of length, but of depth: While it only reads short DNA molecules of 200 to 300 bases, it can read more than 400,000 of these short sequences in a single run, providing a truly ultra-high throughput solution to genome sequencing. This is equivalent to more than 100 million total bases per 7.5 hour run, meaning the machine can sequence the 4.5 million base pair genome of a bacterium such as *E. coli* in about two days. Not only is the newest instrument at 454 Life Sciences fast, it is also incredibly accurate — consensus reads, meaning the alignments of many identical sequenced molecules, are more than 99.99 percent correct.

The Genome Sequencer FLX is not without shortcomings, however. Because the technology relies on short reads of only a couple of hundred base pairs, genomes with highly repetitive regions cannot be assembled accurately. More broadly, the sequencing of truly novel organisms presents a challenge. Without a related species' genome to use as a template, short base reads may not contain enough information to accurately piece together a new genome. In these cases, several additional sequencing runs may need to be performed in order to create a great enough number of reads for the computer software to overlap separate sequences with high confidence.

When asked to elaborate on 454's role in shaping the DNA sequencing industry, Manager of Global Marketing Brendon Hill speaks of the "democratization of sequencing." As described by Hill, sequencing organisms' full genomes used to be considered "big science" and only could be tackled by special genome centers. Now, with the technology and service provided by companies such as 454 Life Sciences, the privilege of large-scale sequencing projects has been stripped from these centers and other elite academic institutions, opening up a wide range of opportunities for basic researchers. While, according to Hill, the eventual goal of 454 Life Sciences is to bring genome sequencing technology into the public health sector, in the meantime there is a "lot of great science to happen as progressively faster, better versions of our instrument come out."

Indeed, many researchers have begun taking advantage of 454's service, and the list of peer-reviewed publications made possible by 454's technology continues to grow. To date, 454 has been involved in the *de novo* sequencing of bacterial, fungal, and viral genomes, the sequencing of small RNA populations, and the elucidation of genomes from ancient organisms such as the wooly mammoth and Neanderthal man.

New ways of thinking about pressing biological questions and even entire new fields of biology are springing up in light of the availability of fast, cheap sequencing technology. One such novel discipline is metagenomics, or the study of genetic material collected *en masse* from an environmental source. Because machines like the Genome Sequencer FLX can sequence hundreds of thousands of bases per run, this depth of coverage allows the sequencing of a complex slew of genomes without introducing base changes or biases. 454's technology so far has enabled the sequencing of organisms from places as diverse as hot thermal vents in the depths of the Atlantic Ocean [4] and soil samples collected from across the planet [5]. In a rather unique application, 454 sequencing technology was able to pinpoint the probable cause for the collapse in United States honeybee populations [6]. By comparing the genomes of microflora found in hives experiencing colony collapse disorder to those collected from normal hives, researchers discovered a virus, the presence of which strongly correlated to the suffering hives, presenting a possible explanation for the loss of worker bees from those colonies.

Not surprisingly, 454 Life Sciences is not the only company working to increase the speed and decrease the cost of genome sequencing. Currently, 454 Life Sciences is competing for the Archon X Prize in Genomics alongside Reveo Inc., VisiGen Biotechnologies, base4 Innovation, Personal Genome X-team, and the Foundation for Applied Molecular Evolution. Subsidized by private investors, the X Prize will be awarded to the first team to sequence 100 full human genomes in 10 days at a cost of less than $10,000 per genome. Seeing that it took two months and just under $1 million to sequence DNA double-helix decoder Dr. James Watson's genome, the Archon contest appears quite a daunting challenge. However, technologies at 454 continue to improve rapidly: In 2005, 454's machines could sequence about 20 million bases per run. This number increased to about 100 million bases per run in 2007, and, according to Hill, the next version of the 454 Genome Sequencer will be able to sequence approximately a billion bases in a day. With advances like these, the X Prize may be well within reach of any of its competitors.

For a time, it appeared genomics would be dethroned by even more complex ways of analyzing whole organisms such as transcriptomics or proteomics, whose technologies look downstream of DNA sequences at actual messenger RNAs or proteins in a given cell. When asked to comment on his company's "rivals" for scientific knowledge, Hill at 454 remained undaunted, stating that "DNA sequencing is a fairly universal tool, not a proxy for anything else … DNA information is the most basic you want to have."

Beyond simply creating faster, more affordable technologies, 454 continues to push the limits of genome sequencing. Researchers there are working to develop new methods for highly parallel DNA sequencing using microarray chips, which are collections of microscopic spots, each corresponding to a single human gene. Their so-called "sequence capture" technology enables researchers to enrich portions of the human genome via hybridization to a chip [7]. Such enrichment eliminates the need for PCR-based amplification strategies and will allow for the selective analysis of specific portions of complex genomes. For example, one could use sequence capture to look exclusively at human exons, or the expressed portions of the human genome. Alternatively, the technology could pull out specific regions of the genome encoding genes relating to heart disease or cancer, thereby facilitating the diagnosis of at risk-patients, while at the same time adding to our knowledge of specific mutations or polymorphisms that lend themselves to the development of these complex diseases.

Truly, the future for 454 Life Sciences and their applications of pyrosequencing technology is bright. And so may be the future of health care and medicine — with improvements in genome sequencing, we can look forward to significant advances in understanding human disease, developing effective treatments, and leading longer and healthier lives.

## REFERENCES

1. Ronaghi M, Karamohamed S, Pettersson B, et al. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem. 1996;242:84-9.
2. Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Res. 2001;11:3-11.
3. Marguiles M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437:376-80.
4. Sogin ML, Morrison HG, Huber JA, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc Natl Acad Sci USA. 2006;103:12115-20.
5. Leininger S, Urich T, Schloter M, et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature. 2006;442:806-9.
6. Cox-Foster DL, Conlan S, Holmes EC, et al. A metagenomics survey of microbes in honeybee colony collapse disorder. Science. 2007;318:283-7.
7. Albert TJ, Molla MN, Muzny DM, et al. Direct selection of human genomic loci by microarray hybridization. Nat Methods. 2007;4:903-5.