



Dr. Sanger, meet Mr. Moore

Next-generation sequencing is driving new questions and new modes of research

Hallam Stevens

Introduction

Since 2005, so-called next-generation sequencing machines have given biologists the ability to sequence ever-faster and ever-cheaper [1, 2]. In the medium term, as these machines are sold to hospitals and companies selling personalized genomic tests, this phenomenon is likely to have a profound effect on medical care [3]. However, these machines are finding wide use in fundamental biology too. Here too, they are likely to have effects on the production of biological knowledge. This Commentary is an attempt to begin a discussion about what these effects might be. Next-generation machines have depended on advances in laser optics, solid-state electronics, and chip-engineering that have drawn biology into a race for more nucleotides per dollar. If – as is often claimed – this is a “Moore’s Law for biology”, we might be able to understand more about the effects of next-generation by understanding some of the origins and history of Moore’s Law. The massive drops in cost and increases in computing power since the mid-1960s have had profound consequences for what a computer is and what we can do with one. Similar changes may be in store for biology.

A brief history of Moore’s Law

The rapid miniaturization of semiconductor components was in large part made possible by the unique context of Silicon Valley. William Shockley came west to commercialize the transistor that he (with John Bardeen and Walter Brattain) had invented in 1947 [4]. Shockley Semiconductor, founded in Palo Alto in 1956, aimed to capture a rapidly growing market. The competition was already fierce: the US Air Force had begun to digitize its avionics in 1956, replacing unreliable and slow vacuum tube switches and computers with transistors; long-range missiles, especially, required small, reliable computers for on-board navigation [5]. Bell Telephone and Texas Instruments were already manufacturing large quantities of transistors. Some of Shockley’s young engineers thought that Shockley was moving too slowly – in 1957, the so-called “traitorous eight” split from Shockley to form Fairchild Semiconductor. Fairchild’s founders set up the company to take advantage of the group’s expertise in order to enter the market quickly and outstrip the competition [6].

What gave Fairchild the advantage was the invention of the integrated

circuit (IC). It was at Fairchild that this experimental object was transformed into a manufacturable product [6].

The IC was the crucial step toward the miniaturization of circuitry, reducing the number of assembly steps and interconnections and, thereby, increasing reliability. Fairchild was the first company to announce the production of an IC in March 1961, just 26 months after their breakthrough. They shipped the first ICs by the end of that year [6]. In the early 1960s, the market for semiconductor devices became even more crowded as Fairchild employees left to found their own companies. It was in this context that Gordon Moore wrote “Cramming more circuits onto integrated circuits” (1965), making his famous prediction that the “complexity” of components on an IC would double every two years [7]. This should be read not as a prediction but as an advertisement: “The future of integrated electronics is the future of electronics itself. The advantages of integration will bring about a proliferation of electronics, pushing this science into many new areas. ICs will lead to such wonders as home computers – or at least terminals connected to a central computer – automatic controls for automobiles, and personal portable communications equipment” [7]. This is a powerful vision for where an industry should go, a road-map that gave a clear direction and clear targets. Moore was selling the potential and power of ICs for the future. Moore’s idea was powerful enough that it oriented the whole industry toward scale in the long term – cramming more transistors on a chip

DOI 10.1002/bies.201100146

History Group, Nanyang Technological University, Singapore, Singapore

Corresponding author:
Hallam Stevens
E-mail: hstevens@ntu.edu.sg

was pursued at the expense of other possible innovations. Innovation meant getting smaller. Intel, the company that Gordon Moore founded in 1968, marketed scale: its 4,004 microprocessor (or “computer on a chip”) was much smaller than a mainframe but remarkably powerful [8]. This re-scaling changed the idea of what a computer was – it was not just faster and smaller, but an object that could be used in all sorts of new ways, just as Moore predicted. Increasing the market for semiconductor devices meant making computers not just for military or big business machines, but for a variety of small-scale uses at home and in small businesses [9].

Applying Moore’s Law to sequencing

Next-generation sequencing is promoted as “Moore’s law for biology”. Countless advertisements, journal papers, reviews, magazines, and newspapers make the comparison (often with an accompanying figure showing the comparison of Moore’s Law and DNA sequencing cost) [2, 10, 11]. One of the main proponents of this view is Jonathan Rothberg. Rothberg sees himself not only as a scientist and inventor, but also as an entrepreneur – significantly, his greatest hero is Steve Jobs. Rothberg founded the biotech startup CuraGen while still in graduate school and then the first next-gen sequencing company, 454 Life Sciences, as its subsidiary in 1999. 454 produced the first next-gen machine (the GS20) in 2005. Rothberg’s newest invention, the Ion Torrent Personal Genome Machine, makes the analogy between Moore’s Law and next-gen literal. Rothberg often recounts the story of when one of his children was rushed to hospital: “In the hospital waiting room, [I] saw a picture of Intel’s Pentium microprocessor, with its millions of transistors, on the cover of a computer magazine. That gave [me] the inspiration to speed up sequencing by working on numerous DNA snippets in parallel” [12]. The Ion Torrent’s thousands of tiny pH meters on the surface of a CMOS chip directly links semiconductor electronics to sequencing: Moore’s Law advances in chip technology now translate directly to greater sequencing speed.

Like the semiconductor industry in the 1960s, the market for next-gen sequencing was (and remains) a highly competitive field. 454 Life Sciences was quickly in competition with Applied Biosystems, Illumina, and Helicos Biosciences. Tony Smith, the chief science officer at Illumina, reported, “It was a race. When you have as a competitor ABI, you better not only have the very best technology but also better commercialize it with ruthless efficiency. If we had been two years later to market, we had have been head to head with Helicos, as opposed to a year ahead of ABI. That made an enormous difference” [13]. This is creating a rapid acceleration of sequencing and a great deal of enthusiasm for the new technologies: if size was driving Moore’s law, then speed is driving Moore’s law for biology. Most of the literature on next-gen technologies celebrates its enabling of a bigger and faster biology [2]. The historian Michael Fortun has written about how the Human Genome Project engendered cultures of speed and acceleration [14]. But next-gen is not merely a speeding-up: I argue here that it is creating three kinds of qualitative changes in biological work.

First, it is contributing to the reorientation of biology toward data-greedy questions. Many studies using next-gen technology are directed at highly general questions that can only be answered with massive data volumes. For instance, Pan et al.’s [15] study of alternative splicing examines this phenomenon not in one or a few genes, but generally, across all the genes in the genome. Others have collected massive data-sets of single nucleotide polymorphisms (SNPs) [16] or exome sequence [17] in order to search for genetic variation as markers of diseases. Such studies, and the questions that motivate them, are only possible because of the data that next-gen tools provide. This kind of work often goes under the label of “hypothesis free” or “data driven” or “exploratory science”, where Baconian induction replaces Popperian deduction [18, 19]. Next-gen technology drives more general and wide-scale questions.

Second, sequencing will be used in a variety of new ways, not just for collecting more and more genomes. Just as the expanding market for computers in the 1970s required a radical diversification

of use, next-gen will result in a variety of new uses for sequencing. This is already beginning to take place. For instance, next-gen has begun to be deployed to investigate the transcriptome. Here sequencing is not used to collect genome data, but to begin to examine the complexity of expression. Cloonan et al.’s study [20] examined SNPs, transcription in repeat elements, and signaling pathways. Other studies have examined non-coding RNAs [21] and alternative splicing [15] on a genome-wide scale. Likewise, CHIP-seq techniques use next-gen to give snap-shots of the dynamic environment within the cell at a level of high-resolution and specificity. This has not only been used to identify protein binding sites genome-wide [22], but also to analyze epigenomic elements such as histone modifications [23] and the DNA methylome [24]. A recent article series in *Nature Reviews Genetics* devoted to the topic of “Applications of next-generation sequencing” has included articles devoted to exome sequencing, transcriptome assembly, human population history, genome structural variation, DNA replication, cancer genomics, and RNA processing. Next-gen also offers novel possibilities for metagenomics [10], tracking SNPs, tracking alleles, and tracking somatic mutations. These examples suggest that next-gen is contributing to a “post-genomic” biology in which the genome yields to a multiplicity of other -omes. In the long term, this may result in a radical displacement of the genome-centered view of biology.

Third, next-gen opens up the possibility of expanding biological work to different sites and different individuals. Just as Moore’s Law drove the personalization of computing, next-gen may drive a “personalization” of biology. The Ion Torrent is already attempting to open up the sequencing market in this way: its \$50,000 price tag, short run times, its name (the *Personal Genome Machine*), and user-friendly design (you can plug in your iPod and its controls make it look like a Playstation) suggest that it is a device for bringing sequencing to the people. Next-gen, like semiconductors, is driven by market forces, and this is changing the way we think about where sequencing belongs and what it might be used for. The techniques of synthetic biology

are creating many possibilities for personal or “Do It Yourself” (DIY) biology in people’s garages or small businesses [25–28]. As sequencing becomes ubiquitous, DIY biology will include DIY sequencing. *Who* does sequencing is likely to change.

As pointed out by Rothberg and Leamon [11] next-gen may also affect the organization of professional biological work. The hallmarks of genomics have been centralization, big funding, and big labs. Sequencing the human genome required an international collaboration stretching over 15 years and costing \$3 billion. Sanger methods required large-scale work and large-scale money. However, next-gen sequencing on a bench-top shifts the center of gravity back toward individual investigators, reversing trends toward Big Biology. An individual lab will be able to afford a machine that can sequence whole genomes in a few days or generate large volumes of other kinds of data. This could lead to less collaborative work, less top-down organization of biological research, more small-scale projects, less sharing of sequence data, and an even wider proliferation in the variety of uses being found for next-gen technologies.

Conclusions

I am not suggesting here that next-gen technologies are completely responsible for recent trends in biology (e.g. toward data-driven biology). However, I am arguing that there is a synergy or relationship of mutual reinforcement between the technologies and the trends. Next-gen has the potential not only to lead us toward more data more quickly, but also toward more general and data greedy questions and problems. In particular, the analogy with

Moore’s Law suggests (somewhat paradoxically) that next-gen may actually move biology’s focus away from static genome sequences and toward multiple, dynamic, interacting “-omes”. Taking seriously the analogy with semiconductor electronics suggests that market forces may be seeding changes in who, how, and where biology is done: individual, small-scale work may become the norm. Between 1960 and 1980, Moore’s Law transformed the computer into a new kind of technology; Moore’s Law for biology could have a similar effect.

References

- Kircher M, Kelso J. 2010. High-throughput DNA sequencing – concepts and limitations. *BioEssays* **32**: 524–36.
- Mardis ER. 2007. The impact of next generation sequencing technology on genetics. *Trends Genet* **24**: 133–41.
- Davies K. 2010. *The \$1000 Genome: The Revolution in DNA Sequencing and the New Era of Personalized Medicine*. New York: Free Press.
- Riordan M, Hodgeson L. 1997. *Crystal Fire: The Birth of the Information Age*. New York: Norton.
- Lécuyer C. 2006. *Revolution in Silicon. In Making Silicon Valley: Innovation and the Growth of High Tech*. Cambridge, MA: MIT Press. p. 129–67.
- Lécuyer C, Brock DC. 2010. *Makers of the Microchip: A Documentary History of Fairchild Semiconductor*. Cambridge, MA: MIT Press.
- Moore GE. 1965. Cramming more components onto integrated circuits. *Electronics* **38**: 114–7.
- Ceruzzi P. 2003. *A History of Modern Computing*. 2nd ed. Cambridge, MA: MIT Press.
- Intel Corporation. 1974. “From CPU to software, the 8080 microprocessor is here”. *Electron Des* **12**: 4–5.
- Editorial. 2009. Metagenomics versus Moore’s Law. *Nat Methods* **6**: 623.
- Rothberg JM, Leamon JH. 2008. The development and impact of 454 sequencing. *Nat Biotechnol* **26**: 1117–24.
- Pollack A. 2011. *Taking DNA Sequencing to the Masses*. New York, NY: New York Times. 4th January.
- Davies K. 2010. The Solexa Story. *BiolT World* 28th September. Available at: <http://www.bio-itworld.com/2010/issues/sept-oct/solexa.html>.
- Fortun M. 1999. Projecting speed genomics. In Fortun M, Mendelsohn E, eds; *The Practices of Human Genetics*. Dordrecht: Kluwer. p. 25–48.
- Pan Q, Shai O, Lee LJ, Frey BJ, et al. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**: 1413–5.
- The 1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467**: 1061–73.
- Yingrui L, Vinckenbosch N, Tian G, Huerta-Sanchez E, et al. 2010. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* **42**: 969–72.
- Franklin L. 2005. Exploratory experiments. *Philos Sci* **72**: 888–99.
- Kell DB, Oliver SG. 2003. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays* **26**: 99–105.
- Cloonan N, Forrest ARR, Kolle G, Gardiner BBA, et al. 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* **5**: 613–9.
- Lu C, Meyers BC, Green PJ. 2007. Construction of small RNA cDNA libraries for deep sequencing. *Nat Methods* **43**: 110–7.
- Robertson G, Hirst M, Bainbridge M, Bilenky M, et al. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **4**: 651–7.
- Zhou VW, Goren A, Bernstein BE. 2011. Charting histone modifications and the function organization of mammalian genomes. *Nat Rev Genet* **12**: 7–18.
- Chung CAB, Boyd VL, McKernam KJ, Fu Y. 2010. Whole methylome analysis by ultra-deep sequencing using two-base encoding. *PLoS ONE* **5**: e9320.
- Biba E. 2011. Genome at home: biohackers build their own labs. *Wired* 19.09.
- Dyson F. 2007. Our biotech future. *N Y Rev Books* **54**: 4–8.
- Ledford H. 2010. Garage biotech: life hackers. *Nature* **467**: 650–2.
- Lipinski J. 2010. *Turning Geek into Chic*. New York Times. New York, NY: The New York Times Company. December 17.