2012 Curt Stern Award Address[1]

Jay Shendure[2,*]

I would like to start by thanking my nominators and the Society for this recognition of our work. I would also like to thank Evan for the very kind introduction. Since I started my lab at the University of Washington in 2007, Evan has been a close collaborator, a good friend, and, moreover, my faculty mentor, so it's a particular pleasure to be introduced by him today.

When I learned that I would be receiving this award, it seemed natural to educate myself a little better about the man after whom the award is named, Curt Stern.

Google searches were followed by an actual trip to the library. It's telling of how much research has changed that this was the first time that I'd ever set foot in the library in my 5 years at the University of Washington. I was so excited to be there that I felt moved to take a picture (Figure 1).

Next to Stern's seminal textbook, *Principles of Human Genetics*,[1] I found *Genetic Mosaics and Other Essays*,[2] a series of lectures by Stern in the 1960s. At the conclusion of this volume is a short "sermon," as he called it, entitled "Thoughts on Research." It's really a very nice set of reflections by Stern relatively late in his career. I will quote one passage that I found particularly compelling:

> "…the course of science resembles that of evolution. It may be pictured as an exploration of an unending series of mountain chains. When you enter a new valley you cannot know whether it will end blindly or lead to a pass through which one may reach a vast new area. There are few passes and many dead ends."

I wouldn't say that this describes all of science, but as I was reflecting on my own research, Stern's metaphor resonated. Rather than review my lab's work of the past few years, which is recent history and, moreover, is already quite well-represented at this meeting, I thought that I would instead use my graduate work as an example of what I mean.

In 2002—10 years ago—I was at the beginning of my third year of graduate school in George Church's lab at Harvard Medical School. We were within the Department of Genetics, but my research wasn't so much the actual practice of genetics or genomics as it was early-stage technology development for genetics and genomics.

George's lab was—and I imagine still is—an amazing and somewhat unstructured environment where graduate students had almost complete scientific freedom and what seemed to be unlimited resources. In my first 2 years of graduate school, I took advantage of this to a fault by meandering through a hodgepodge of ideas with little to no coherence, because although I had some modest successes, they had very little to do with one another and were vastly outnumbered by half-baked ideas that led nowhere.

Fortunately, Rob Mitra, who was a postdoc in the lab and who is now a professor at Washington University in St. Louis, took me under his wing to help him with polony sequencing, a technology that he was developing. This is one of maybe a dozen strands of research (in both academia and industry) that collectively gave rise to what we refer to today as massively parallel or next-generation DNA sequencing.

The pseudocolored polonies that I'm showing are representative of where we were in 2003; here, we are visualizing 20 or so millimeter-scale PCR colonies (or polonies) on the surface of a microscope slide (Figure 2). Through serial single-base extensions, we achieved 6–8 bp reads and in total sequenced a few hundred base pairs—less than a single Sanger sequencing read but an early proof-of-concept result for massively parallel sequencing.[3]

Hindsight is usually 20/20, but in this case, foresight was also 20/20. One of the bolder aspects of the paper—given the state of the field in 2003—was a few discussion sentences in which we argued that it would be straightforward to improve on this result by a factor of one billion, and if we could just hurry up and do that, then the cost of

[1]This article is based on the address given by the author at the meeting of the American Society of Human Genetics (ASHG) on November 10, 2012, in San Francisco, CA, USA. The audio of the original address can be found at the ASHG website.

[2]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
*Correspondence: shendure@uw.edu

http://dx.doi.org/10.1016/j.ajhg.2012.11.020.

Figure 1. Photograph of a Bookshelf at the University of Washington's Health Sciences Library
The circled volumes on the top shelf are *Genetic Mosaics and Other Essays* and *Principles of Human Genetics* by Curt Stern. The circled volume on the bottom shelf is *Mouse Genetics: Concepts and Applications* by Lee M. Silver.

sequencing a human genome would be a mere $6,000, which is of course only slightly off from where technologies implementing related schemes are today.

At the time, I was not remotely enthused about this prediction because Rob was moving on to a faculty position and George had implied that I could probably knock off those nine orders of magnitude solo before I graduated. The main thing that I had going for me was a technology-development framework that I had learned from Rob and George.

It's the nitty-gritty of technology development that is a great match to Stern's metaphor of "exploring an unending series of mountain chains." Rob had four rules for how to go about this process of exploration without getting too lost. First, anything that works on paper will almost always work in the lab but orders of magnitude less efficiently than you need it to work. Second, latch on to that barely working experiment and optimize by changing one, and only one, thing at a time. Third, quantify everything that you possibly can in every experiment. Fourth, design your experiments such that you learn something from every failure.

With these rules in hand, I went to work on those nine orders of magnitude, but it turned out to be not so straightforward at all, and after a year of minimal progress, I decided to throw in the towel. I had managed to accumulate a few publications, so I went before my thesis committee—which, as it happens, included last year's Curt Stern Award winner, David Altshuler—and asked for permission to graduate.

To my surprise and dismay, they said no.

In retrospect, I'm incredibly grateful that my committee had more faith in me than I had in myself—I'm fairly certain that I wouldn't be standing up here today if they had let me graduate. However, at the time, I was just bummed out. My experiments weren't working. I was living in a shabby apartment in Central Square. My girlfriend, whom I was convinced was "the one," had broken up with me. And I was in the middle of an MD/PhD program that seemed as if it would go on forever.

Eventually, I snapped out of it. I teamed up with another graduate student, Greg Porreca, and we kept at it nights and weekends and eventually navigated our way onto flatter terrain.

Moving entirely out of our comfort zones and the scope of our training, Greg and I started building sequencing instruments, first this one, the central component of which is a bicycle wheel (Figure 3A), and then the considerably more sophisticated instrument shown here (Figure 3B). I also didn't give up on the girlfriend—I put inordinate amounts of time into making the perfect mix tape—even though, like polony sequencing, it was a long shot.

If I fast forward again to 2005, somehow it all worked out. We succeeded in advancing the sequencing technology to the point where it was 10-fold less expensive than Sanger sequencing—roughly five of those nine

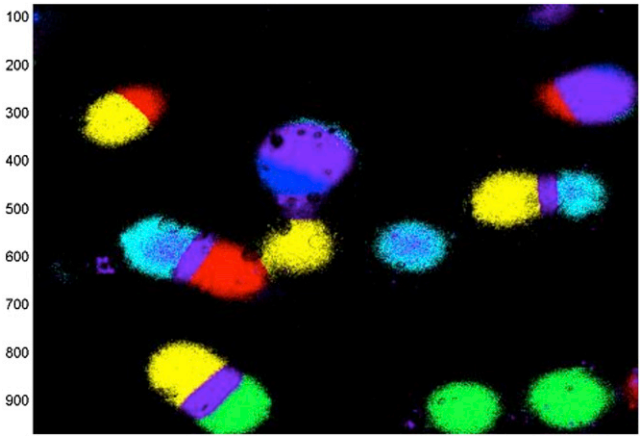

Figure 2. Fluorescent In Situ Sequencing on Polymerase Colonies in 2003
Reprinted from the supplemental data of Mitra et al.[3]

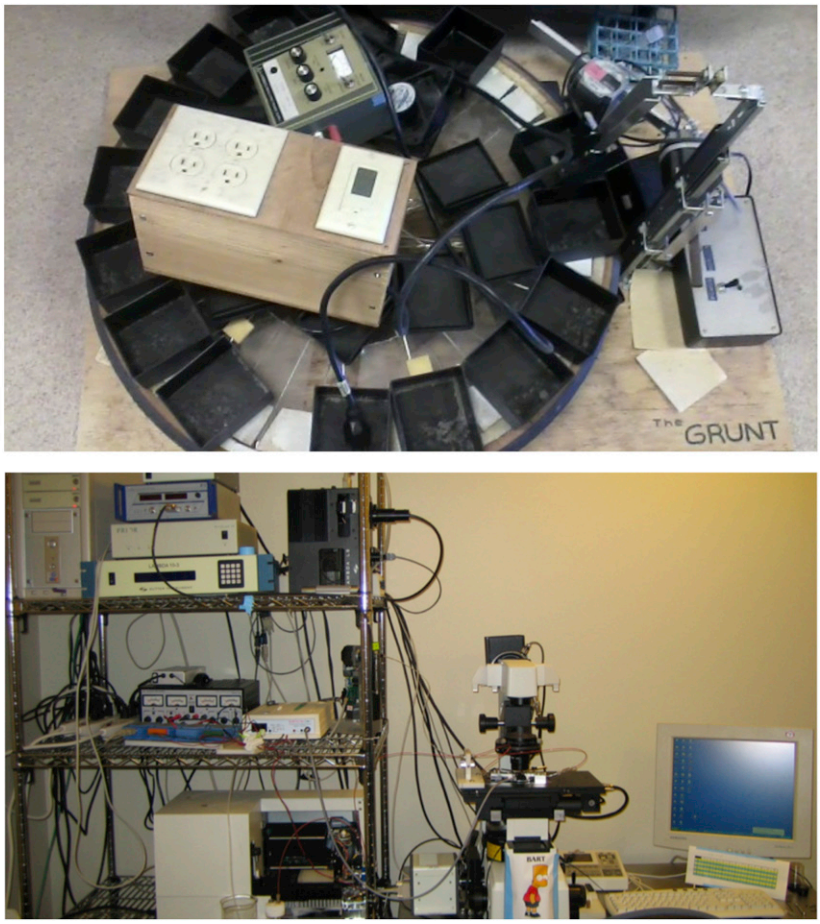

Figure 3. Early Instrumentation for Massively Parallel Sequencing
(A) Prototype automation for massively parallel sequencing in 2003.
(B) The sequencing instrument used for the work described in Shendure et al.[4]

orders of magnitude—and published a description of the technology in *Science*[4] (Figure 4). My committee finally agreed to let me graduate. Lastly, the mix tape worked, and the girlfriend (now my wife) got back together with me.

To come back to Stern's metaphor, I would estimate that I spent more than 90% of my 5 years of graduate school in valleys that proved to be dead ends and less than 10% in passes that eventually proved to be fruitful.

For many practical scientific problems, including technology development, this endurance race, rather than any single brilliant insight, is simply the nature of the terrain. You can learn to become a better navigator, but that learning is entirely dependent on experiencing those repeated failures and surviving to tell about it.

Stern's metaphor also holds if you zoom out from one lab to the sequencing technology field as a whole. The specific technology that we developed, although at the leading edge in 2005, would soon be overtaken by related but different technologies. However, that doesn't mean it wasn't worthwhile, if not necessary.

To put this another way, here is the curve that I'm sure you've all seen illustrating the cost of sequencing over the past decade (Figure 5). However, when you take a moment to think about it, this is really the superimposition of two curves.

The first simply reflects electrophoretic or Sanger-sequencing technologies, whereas the second reflects the disruptive introduction of massively parallel sequencing.

My point is simply that an incredibly rich period of exploration and natural selection—a period of intense innovation within and between both academic and commercial groups—predated the moment when these curves crossed one another.

Moreover, I would argue that the eventual breakthrough and survival of a few species is predicated on a rich ecosystem of ultimately noncompetitive approaches—an evolutionary process including recombination and natural selection and more blind alleys than successes.

Here is a quotation from the remarkable Sydney Brenner, who, among other things, is a prolific inventor of technologies, including some of the earliest manifestations of massively parallel sequencing: "Progress in science depends on new techniques, new discoveries, and new ideas, *probably in that order*" (emphasis added).

Of course I love this quotation, but I also believe that human genetics is a field that recognizes the central value of advancing technologies, and this is in part because it is a field whose progress for the past half century has been largely rate limited by the state of available technologies for ascertaining genetic variation in individual human genomes.

However, we are also clearly at the tail end of that technological evolution. It is now plausible to consider sequencing every genome that needs to be sequenced in order to get at the underlying genetics of both rare and common human disease.

At the same time, I don't believe that the field of human genetics will simply analyze these data sets and call it a day. Rather, the challenges will shift, and probably they already are shifting. I'll name just two of these that are already becoming quite clear.

First, there is of course tremendous interest in bringing genomics to the clinic to inform patient care. But the challenge posed by "variants of uncertain significance" is likely to profoundly deepen as the clinical sequencing of human genomes accelerates and as the

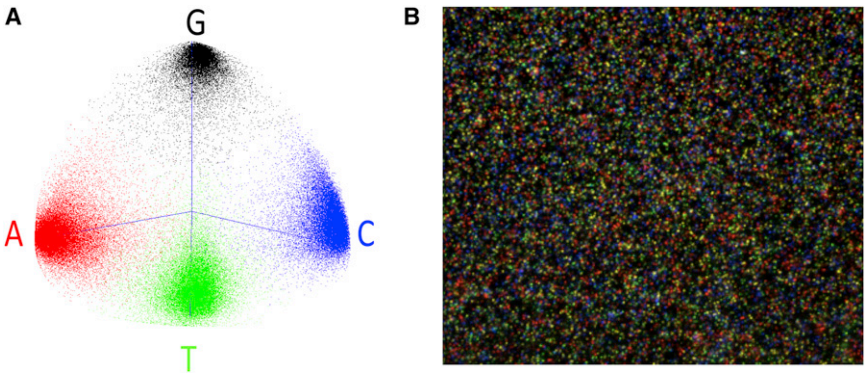

Figure 4. An Early Proof of Concept for Massively Parallel Sequencing
Base calling (A) and massively parallel sequencing (B) in 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. Reprinted from Shendure et al.[4]

list of genes that are clinically actionable grows. Is this a problem that we simply have to live with, or are there solutions?

Second, human genetics has implicated thousands of genes in Mendelian disorders and thousands of genomic regions in common diseases. How do we go about the broader task of exploiting those findings to understand the underlying biology? Moreover, given that there are now thousands of genes and thousands of regions to be followed up on, how do we pull this off at the requisite scale?

These are big problems, but if we take the very long view, the ultimate impact of the field of human genetics depends not only on comprehensively cataloging the genes underlying human disease but also on whether we can successfully tackle these ensuing challenges.

At the same time, I don't think there is any obvious way forward, and facing down these challenges is most likely going to require entirely new technologies and new experimental paradigms.

I can't honestly say that I know what those new experimental paradigms are going to be, but I do think that it will require an evolutionary process similar to what I described for the early days of new sequencing technologies.

In other words, this evolutionary process might involve the encouragement of a rich ecosystem that includes the messy, uncoordinated exploration of lots of half-baked ideas and a tolerance for the fact that nearly all of these will end in failure while also keeping a sharp eye out for the eventual passageway that will break us through to the other side.

I thought that I would close on a memory of how I actually got interested in genetics in the first place. This memory was sparked on that same library trip, in which I also chanced upon a mouse genetics textbook[5] written by my undergraduate research advisor, Lee Silver (Figure 1).

I hardly knew anything about genetics before enrolling in a course with Lee in my sophomore year at Princeton. Although the course itself was phenomenal, it was really

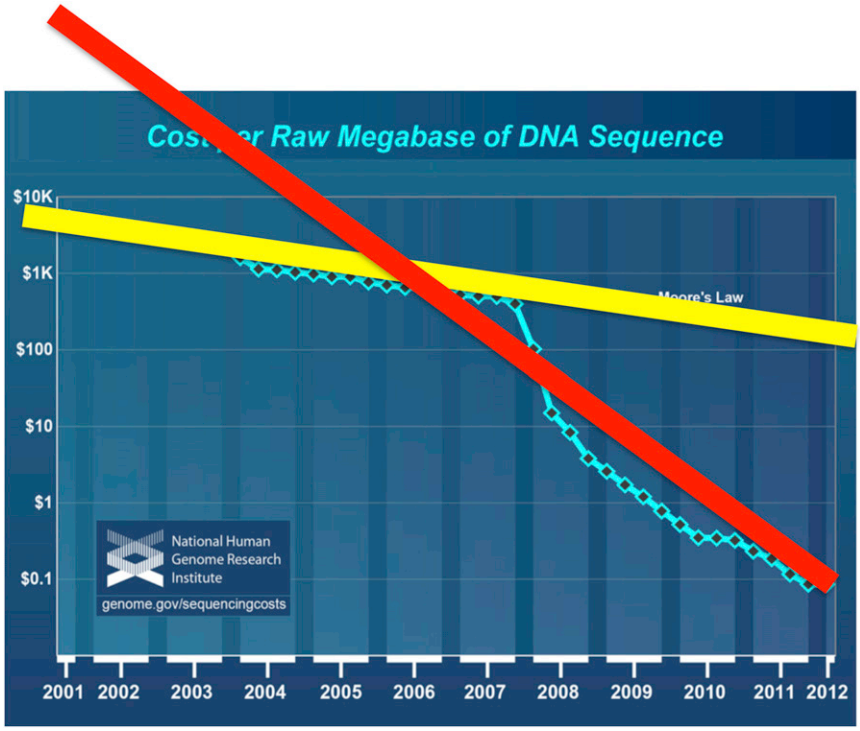

Figure 5. Disruptive Technology Development and the Cost of DNA Sequencing
The log-scale cost curve for declining sequencing costs from the National Human Genome Research Institute (see Web Resources) is superimposed with lines showing the separate trajectories of Sanger sequencing (yellow) and massively parallel sequencing (red).

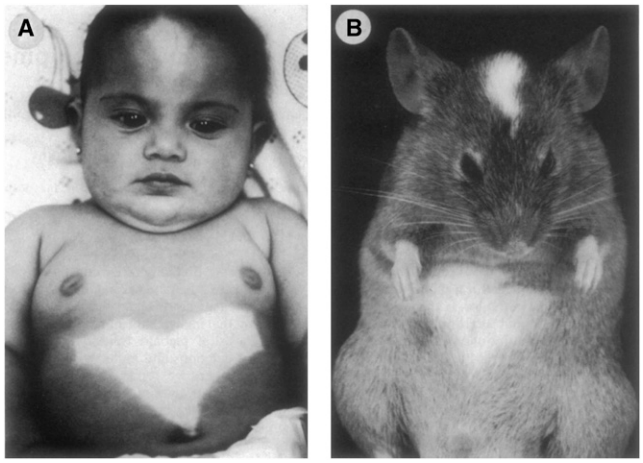

Figure 6. Mammalian Genetics
Cutaneous malformations resulting from heterozygous mutation of the *c-kit* proto-oncogene in a child (A) and a mouse (B). Reprinted from Fleischman et al.[6]

just one figure that I came across in the assigned reading that did it for me (Figure 6).[6] The figure, which I am showing here, juxtaposes the highly similar cutaneous phenotype of a human child and a knockout mouse, both with heterozygous *c-kit* deletions.

Even if you entirely accept the worldview and basic facts that allow for this to be true, which I did, seeing those facts so succinctly and powerfully summarized just knocked me over. I tacked the figure above my desk, and not long after the course ended, I approached Lee and pleaded with him to let me join his lab. I spent the last 2 years of college receiving a very practical education in mouse—and by extension, human—genetics.

I now have small children, and it's of course never too early to start teaching them some basic genetics—really simple things, such as that all living things are descended from a common ancestor and that you are a blend of your father and your mother.

What amaze me are the questions that I get back from my 4-year old daughter. For example, "If we had the same great-great grandmother as a gorilla, did she look like a gorilla or a person?" —which I hear as "What was the phenotype of the common ancestor of primates?" Or "Why am I half brown all over instead of this half brown and this half yellow?"—which I hear as "Galton or Mendel?"

And now that I'm buried in the minutia of running a lab, it's been extraordinarily refreshing to be reminded by my daughter of how natural and compelling the central questions of our field are once one is confronted with a few basic facts. Along the same lines, attending this meeting each year is a great reminder of what a privilege it is to be a part of this amazing community of individuals who share a burning curiosity for these same questions.

There are a number of people who I would like to thank.

First, I thank my scientific mentors, Lee Silver and George Church: Lee, for introducing me to research and for teaching me genetics; George, for providing me with a framework for thinking about technology and genetics, as well as for being a kind and inspiring mentor.

Second, I would like to thank Rob Mitra and Greg Porreca, who were my scientific partners for the key periods of my training.

Third, I thank my lab. From the inception of my lab in 2007, I have been blessed with trainees and staff who inspire me every day with their creativity, enthusiasm, boldness, rigor, and support of one another. They refer to themselves as a "hive mind," and in a certain way, I think of myself here today as accepting this award on behalf of the hive.

And of course, no lab is an island. I have simply amazing colleagues in genome sciences, medical genetics, and pediatric genetics at the University of Washington. I don't have the words to express how thankful I am for this personally warm, intellectually engaging, and intensely collaborative community within which my lab resides and on which the work being recognized here has been completely dependent.

Lastly, I would like to thank my family, including my parents, my brother, my wife, Alex, and my children, Ariya and Daniel, for their ongoing support and for making my life what it is today.

Thank you.

Web Resources

The URL for data presented herein is as follows:

NHGRI Genome Sequencing Program, http://www.genome.gov/sequencingcosts/

References

1. Stern, C. (1973). Principles of human genetics (San Francisco: W. H. Freeman).
2. Stern, C. (1968). Genetic mosaics, and other essays (Cambridge: Harvard University Press).
3. Mitra, R.D., Shendure, J., Olejnik, J., Edyta-Krzymanska-Olejnik, and Church, G.M. (2003). Fluorescent in situ sequencing on polymerase colonies. Anal. Biochem. *320*, 55–65.
4. Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science *309*, 1728–1732.
5. Silver, L.M. (1995). Mouse genetics: Concepts and applications (New York: Oxford University Press).
6. Fleischman, R.A., Saltman, D.L., Stastny, V., and Zneimer, S. (1991). Deletion of the c-kit protooncogene in the human developmental defect piebald trait. Proc. Natl. Acad. Sci. USA *88*, 10885–10889.