1

Competition and the Future of Reading and Writing DNA

Robert Carlson

Biodesic and Bioeconomy Capital, 3417 Evanston Ave N, Ste 329, Seattle, WA 98103, USA

Constructing arbitrary genetic instruction sets is a core technology for biological engineering. Biologists and engineers are pursuing even better methods to assemble these arbitrary sequences from synthetic oligonucleotides (oligos) [1]. These new assembly methods in principle reduce costs, improve access, and result in long sequences of error-free DNA that can be used to construct entire microbial genomes [2]. However, an increasing diversity of assembly methods is not matched by any obvious corresponding innovation in producing oligos. Commercial oligo production employs a very narrow technology base that is many decades old. Consequently, there is only minimal price and product differentiation among corporations that produce oligos. Prices have stagnated, which in turn limits the economic potential of new assembly methods that rely on oligos. Improvements may come via recently demonstrated assembly methods that are capable of using oligos of lower quality and lower cost as feedstocks. However, while these new methods may substantially lower the cost of gene-length double-stranded DNA (dsDNA), they also may be economically viable only when producing many orders of magnitude with more dsDNA than what is now used by the market. The commercial success of these methods, and the broader access to dsDNA they enable, may therefore depend on structural changes in the market that are yet to emerge.

1.1 Productivity Improvements in Biological Technologies

In considering the larger impact of technological monoculture in DNA synthesis, it is useful to contrast DNA synthesis and assembly with DNA sequencing. In particular, it is instructive to compare productivity estimates of commercially available sequencing and synthesis instruments (Figure 1.1). Reading DNA is as crucial as writing DNA to the future of biological engineering. Due to not just commercial competition but also competition between sequencing technologies, both

Synthetic Biology: Parts, Devices and Applications, First Edition. Edited by Christina Smolke.

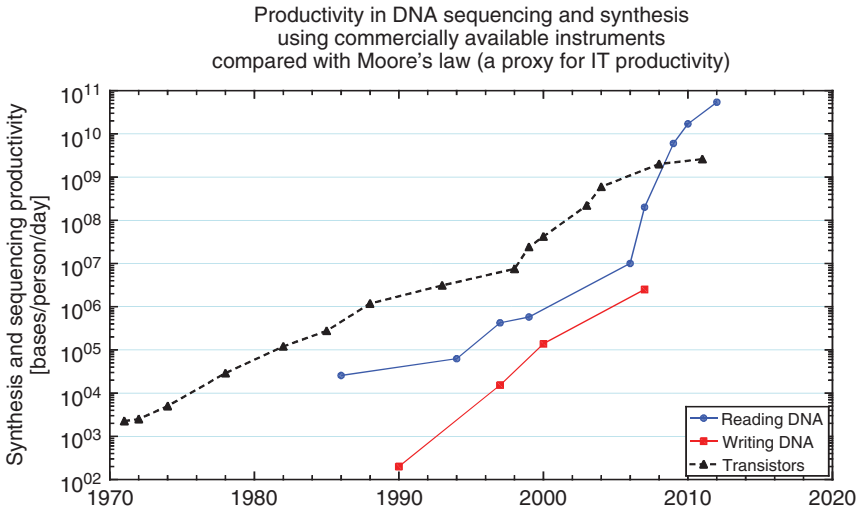

Figure 1.1 Estimates of the maximum productivity of DNA synthesis and sequencing enabled by commercially available instruments. Productivity of DNA synthesis is shown only for column-based synthesis instruments, as data for sDNA fabricated on commercially available DNA arrays is unavailable; exceptions are discussed in the text. Shown for comparison is Moore's law, the number of transistors per chip. (Intel; Carlson, 2010 [3]; Loman *et al*. 2012 [4]; Quail *et al*. 2012 [5]; Liu, 2012 [6].)

prices and instrument capabilities are improving rapidly. The technological diversity responsible for these improvements poses challenges in making quantitative comparisons. As in previous discussions of these trends, in what follows I rely on the metrics of price [$/base] and productivity [bases/person/day].

Figure 1.1 also directly compares the productivity enabled by commercially available sequencing and synthesis instruments to Moore's law, which describes the exponential increase in transistor counts in CPUs over time. Readers new to this discussion are referred to References 3 and 4 for in-depth descriptions of the development of these metrics and the utility of a comparison with Moore's law [3, 7]. Very briefly, Moore's law is a proxy for productivity; more transistors enable greater computational capability, which putatively equates to greater productivity.

Visual inspection of Figure 1.1 reveals several interesting features. First, general synthesis productivity has not improved for several years because no new instruments have been released publicly since about 2008. Productivity estimates for instruments developed and run by oligo and gene synthesis service providers are not publicly available.[1]

1 It is likely that array-based DNA synthesis used to supply gene assembly operates at a much higher productivity than column-based synthesis. For example, Agilent reportedly produces and ships in excess of 30 billion bases of ssDNA a day, the equivalent of more than 10 human genomes, on an undisclosed number of arrays (Darlene Solomon, Personal Communication).

Second, it is clear that DNA sequencing platforms are improving very rapidly, now much faster than Moore's law.

Moore's law and its economic and social consequences are often used to benchmark our expectations of other technologies. Therefore, developing an understanding of this "law" provides a means to compare and contrast it with other technological trends.

1.2 The Origin of Moore's Law and Its Implications for Biological Technologies

Moore's law is often mistakenly described as a technological inevitability or is assumed to be some sort of physical phenomenon. It is neither; Moore's law is a business plan, and as such it is based on economics and planning. Gordon Moore's somewhat opaque original statement of what became the "law" was a prediction concerning economically viable transistor yields [8]. Over time, Moore's economic observation became an operational model based on monopoly pricing, and it eventually enabled Intel to outcompete all other manufacturers of general CPUs. Two important features distinguish CPUs from other technologies and provide insight into the future of trends in biological technologies: the first is the cost of production, and the second is the monopoly pricing structure.

Early on Intel recognized the utility of exploiting Moore's law as a business plan. A simple scaling argument reveals the details of the plan. While transistor counts increased exponentially, Intel correspondingly reduced the price per transistor at a similar rate. In order to maintain revenues, the company needed to ship proportionally more transistors every quarter; in fact, the company increased its shipping numbers faster than prices fell, enabling consistent revenue to grow for several decades. This explains why Intel former CEO Andy Grove reportedly constantly pushed for an even greater scale [9].

In this sense, Moore's law was always about economics and planning in a multibillion-dollar industry. In the year 2000, a new chip fab cost about $1 billion; in 2009, it cost about $3 billion. Now, according to *The Economist*, Intel estimates that a new chip fab costs about $10 billion [9]. This apparent exponential increase in the cost of semiconductor processing is known as Rock's law. It is often argued that Moore's law will eventually expire due to the physical constraints of fabricating transistors at small length scales, but it is more likely to become difficult to economically justify constructing fabrication facilities at the cost of tens to hundreds of billions of dollars. Even through the next several iterations, these construction costs will dictate careful planning that spans many years. No business spends $10 billion without a great deal of planning, and, more directly, no business finances a manufacturing plant that expensive without demonstrating a long-term plan to repay the financiers. Moreover, Intel must coordinate the manufacturing and delivery of very expensive, very complex semiconductor processing instruments made by other companies. Thus Intel's planning and finance cycles explicitly extend many years into the future. New technology has certainly been required to achieve each planning goal, but this is part of the ongoing research, development, and planning process for Intel.

Moore's law served a second purpose for Intel and one that is less well recognized but arguably more important; it was a pace selected to enable Intel to win. Intel successfully organized an entire industry to move at a pace only it could survive. And only Intel did survive. While Intel still has competitors in products such as memory or GPUs, companies that produced high volume, general CPUs have all succumbed to the pace of Moore's law. The final component of this argument is that, according to Gordon Moore, Intel could have increased transistor counts faster than the historical rate.[2] In fact, Intel ran on a faster internal innovation clock than it admitted publicly, which means that Moore's law was, as one Intel executive put it, a "marketing head fake" [10]. The inescapable conclusion of this argument is that the management of Intel made a very careful calculation; they evaluated product rollouts to consumers – the rate of new product adoption, the rate of semiconductor processing improvements, and the financial requirements for building the next chip fab line – and then set a pace that nobody else could match but that left Intel plenty of headroom for future products. In effect, if not intent, Intel executed a strategy that enabled it to set CPU prices and then to reduce those prices at a rate no other company could match.

This long-term planning, pricing structure, and the resulting lack of competition contrasts quite strongly with the commercial landscape for biological technologies. Whereas the exponential pace of doubling of transistor counts was controlled by just one company, productivity in DNA sequencing has recently improved faster than Moore's law due to competition not just among companies but also among technologies. Conversely, the lack of improvement in synthesis productivity suggests that the narrow technology base for writing DNA has reached technical and, therefore, economic limits. Nonetheless, while Figure 1.1 may suggest a temporary slowdown in the rate of improvement for sequencing, and in effect shows zero recent improvement for synthesis, new technologies will inevitably facilitate continued competition and, therefore, continued productivity improvement.

1.3 Lessons from Other Technologies

Compared with that in other industries, the financial barrier to entry in biological technologies is quite low. Unlike chip manufacturing, there is nothing in biology with a commercial development price tag of $10 billion. The Boeing 787 reportedly cost $32 billion to develop as of 2011 and is on top of a century of multibillion-dollar aviation projects that preceded it [11]. Better Place, an electric car company, declared bankruptcy after receiving $850 million in investment [12]. Tesla Motors has reported only one profitable quarter since 2003 and continues to operate in the red while working to achieve manufacturing scale-up [13, 14].

There are two kinds of costs that are important to distinguish here. The first is the cost of developing and commercializing a particular product. Based on the

2 Gordon Moore to Danny Hillis, as related by Danny Hillis, Personal Communication.

money reportedly raised and spent by Illumina, Pacific Biosciences, Oxford Nanopore, Life, Ion Torrent, and Complete Genomics (the latter three before acquisition), it appears that developing and marketing a second-generation sequencing technology can cost more than $100 million. Substantially more money gets spent, and lost, in operations before any of these product lines is revenue positive. Nonetheless, relatively low development costs have enabled a number of companies to enter the market for DNA sequencing, resulting in a healthy competition in a market that is presently modest in size but that is expected to grow rapidly over the coming decades.

1.4 Pricing Improvements in Biological Technologies

The second kind of cost to keep in mind is the use of new technologies to produce an object or produce data. Figure 1.2 is a plot of commercial prices for column-synthesized oligos, gene-length synthetic DNA (sDNA), and DNA sequencing. Prior to 2006, the sequencing market was dominated by Sanger-based capillary instruments produced by Applied Biosystems, in effect another pricing monopoly. After 2006, the market saw a rapid proliferation of not just commercial but also technological competition with the launch of next-generation systems from 454, Illumina, Ion Torrent, Pacific Biosciences, and Oxford Nanopore based on a diversity of chemical and physical detection methodologies [15]. Illumina presently dominates the market for sequencing instruments but is facing competition from Oxford Nanopore and various Chinese insurgents. There also remains technological diversity between companies, which contributes to competitive

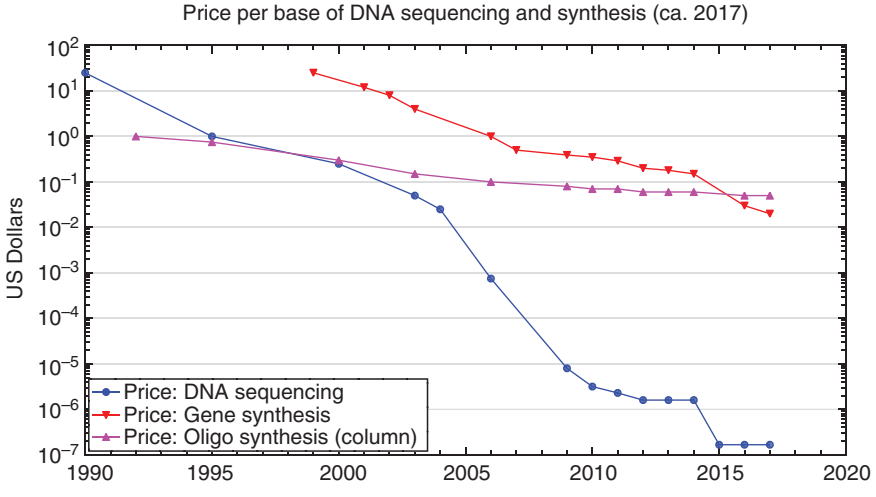

Figure 1.2 Commercial prices per base for DNA sequencing, column-synthesized oligonucleotides, and gene-length sDNA. Reported prices for array-synthesized oligos vary widely, and no time series is available. Market pricing for genes can vary by up to an order of magnitude, depending on sequencing composition and complexity. (Carlson (2010), Commercial price quotes.)

pressures. An important consequence of the emergence of technological competition in the DNA sequencing market is a rapid price decrease. The NIH maintains a version of this plot that compares sequencing prices with cost per megabyte for memory, another form of Moore's law [16]. Both Figure 1.2 and the NIH plot show that sequencing costs kept pace with Moore's law while a pricing monopoly was in effect. The emergence of technological competition produced both productivity improvements and price changes that outpaced Moore's law.

In contrast, despite modest commercial competition in the DNA synthesis market, the lack of technological competition has limited price decreases in the last 5 years. The industry as it exists today is based on chemistry that is several decades old, in which oligos are synthesized step by step on an immobilized substrate. Using array-synthesized oligos for gene assembly appears to be lowering the market price, though quality and delivery time are reportedly inconsistent across the industry. Improved error correction and removal technologies may further reduce the assembly cost for genes and thereby improve the profit margins [17]. My informal conversations with industry insiders suggest that oligo producers may no longer include the cost of goods in calculating prices; that is, oligo prices are evidently determined largely by the cost of capital rather than the cost of raw materials. This suggests that very little pricing improvement can be expected for genes produced from standard oligo synthesis.

1.5 Prospects for New Assembly Technologies

Array synthesis has the advantage of a low volume production of oligos with high library diversity [18]. Gene assembly based on array synthesis has proved difficult to commercialize. At least three companies in this space, Codon Devices, Gen9, and Cambrian Genomics, have gone bankrupt or been acquired in recent years. Twist, a more recent entrant, now quotes prices in the neighborhood of $10 per base and publicly asserts it will push prices much lower in the coming years.

With prices potentially soon falling by orders of magnitude, one must ask about the subsequent impact on the market for synthetic genes. New firms entering the market are implicitly working on the hypothesis that supply-side price reductions will drive increased demand. The most obvious source of that demand would be forward design of genetic circuits based on rational models. Yet the most sophisticated synthetic genetic circuits being constructed in industrial settings are designed largely using heuristic models rather than quantitative design tools [19]. Moreover, these circuits contain only a handful of components, which stand as a substantial bottleneck for demand. Alternatively, customers may employ less up-front predictive design and instead rely on high-throughput screening of pathway variants; screening libraries of pathways has the potential to create substantial demand for synthetic genes [20].

Considering the interplay between market size and price reveals challenges for companies entering the gene synthesis industry. Recalling the lessons of Moore's law, a relatively simple scaling argument will reveal the performance constraints

of the gene synthesis industry. Intel knew that it could grow financially in the context of exponentially falling transistor costs by shipping exponentially more transistors every quarter – that is, the business model of Moore's law. But that was in the context of an effective pricing monopoly, and Intel's success required a market that grew exponentially for decades. The question for synthetic gene companies is whether the market will grow fast enough to provide adequate revenues when prices fall. For every order of magnitude drop in the price of synthetic genes, the industry will have to ship an order of magnitude of more DNA just to maintain constant revenues. More broadly, in order for the industry to grow, synthesis companies must find a way to expand their market at a rate faster than when prices fall. Unfortunately, as best as I can tell, despite falling prices and putative increases in demand, the global gene synthesis industry generated only about $150 million in 2015 [21]. The total size of the industry appears to have been static, or even to have decreased, over the prior decade.

Ultimately, for a new wave of gene synthesis companies to be successful, they have to provide their customers with something of value. Academic customers are likely to become more plentiful as it becomes even more obvious that ordering genes is cheaper than cloning genes, even with graduate student labor costs. Gene synthesis pioneer John Mulligan used to cite NIH expenditures on cloning – approximately $3 billion annually – as a potential market size for gene synthesis [22]. This is certainly an attractive potential market. However, with the price per base potentially falling dramatically in the near term, the comparison to cloning must focus on the total number of cloned bases replaced by synthesis and at what exact price.

For commercial customers, it is less obvious that lower prices will equate to substantial increases in demand. The cost of sDNA is always going to be a small cost of developing a product, and it is not obvious that making a small cost even smaller will affect the operations of an average corporate lab. In general, research only accounts for 1–10% of the cost of the final product [23]. The vast majority of development costs are in scaling up production and in polishing the product into something customers will actually buy. For the sake of argument, assume that the total metabolic engineering development costs for a new product are in the neighborhood of $50–100 million, a reasonable estimate given the amounts that companies such as Gevo and Amyris have reportedly spent. In that context, reducing the cost of sDNA from $50 000 to $500 may be useful, but the corporate scientist-customer will be more concerned about reducing the $50 million overall costs by a factor of two, or even an order of magnitude, a decrease that would drive the cost of sDNA into the noise. Thus, in order to increase demand adequately, the production of radically cheaper sDNA must be coupled with innovations that reduce the overall the product development costs. As suggested above, forward design of complex circuits is unlikely to provide adequate innovation anytime soon. An alternative may be high-throughput screening operations that enable testing many variant pathways simultaneously. But note that this is not just another hypothesis about how the immediate future of engineering biology will change but also another generally unacknowledged hypothesis. It might turn out to be wrong, and elucidating one final difference between transistors and DNA may explain why.

The global market for transistors has grown consistently for decades, driven by an insatiable demand for more computational power and digital storage. Every new product must contain more transistors than the model it replaces. In contrast, while the demand for biological products is also growing, every new biological product is made using, in principle, just one DNA sequence. In practice, while many different DNA sequences may be constructed and tested in developing a new product, these many sequences are still winnowed down to only one sequence that defines a microbial, plant, or mammalian production strain. Nevertheless, this fundamental difference in use between transistors and DNA reveals the gene synthesis industry as the provider of engineering prototypes rather than as a large volume manufacturer of consumer goods. Consequently, while high-throughput synthetic biology companies such as Amyris, Ginkgo Bioworks, and Zymergen may place relatively large orders for sDNA, the price and volume of that sDNA will never have much impact on the final products produced by those companies.

1.6 Beyond Programming Genetic Instruction Sets

At present, the cost of purifying oligos and short dsDNA can exceed the cost of the DNA itself by as much a factor of three. The availability of lower cost, high quality dsDNA may therefore enable applications that are presently not economically viable at large scale. Beyond its utility in programming biological systems, dsDNA can be used as nanoscale structural or functional components [24]. The future of these applications is difficult to predict but could include circuitry assembled from DNA that is modified using proteins and chemistry to create conductive and semiconductive regions useful for computation [25]. It is unclear what sDNA market size these applications may support. Recent progress suggests that new demand might emerge from the use of DNA as a digital information storage medium [26]. Even a single, modestly size data center would consume many orders of magnitude of more sDNA than any prospective use of sDNA in biological contexts [27].

1.7 Future Prospects

Regardless of the particular course of companies entering the gene synthesis market, it appears that prices are likely to fall, potentially fueling an increase in demand. That demand may come in part from customers who fall outside the usual academic and corporate classifications; start-up companies, community labs, and individual, independent entrepreneurs and scientists are likely to use sDNA in new and interesting ways. The standing biosecurity strategy of the United States is to explicitly engage and encourage this innovation, including in contexts such as "garages and basements" [28]. This strategy recognizes the important role of entrepreneurs in innovation and job creation and also recognizes the difficulty of preventing access to biological technologies through regulations or restrictions.

Complementing the engagement strategy is an effort to prevent accidentally synthesizing and shipping potentially hazardous sequences. Most gene synthesis companies have voluntarily signed onto international agreements to screen orders against lists of pathogens and toxins such as the Harmonized Screening Protocol of the International Gene Synthesis Consortium (IGSC) [29].

The technical potential of new sDNA production methods may provide an opportunity to build and test far more genetic circuit designs than what is now feasible. The economic demand for biological production is enormous and is growing rapidly [30, 31]. Whether newly emerging sDNA companies survive economically depends in large part on their ability to increase total market demand sufficiently to offset falling prices. The size of that market, in turn, largely depends on whether less expensive dsDNA enables customers to reduce research and development costs and to create more products. The fundamental problem for the synthesis industry is that, however valuable sDNA is substantively to biological engineering in practice, the monetary value of that DNA is small compared with total development costs and has been falling, at times very rapidly, for decades. Falling prices limit both the maximum profit margin and the incentive to invest in new technology. Any new technology that does enter the market will inevitably drive competition, further depressing prices and margins. Going forward, productivity and prices are likely to display step changes resulting from the emergence of new technology and competition rather than display smooth long-term changes. Finally, given the relatively low barriers to entry for biological technologies and the consequent inevitable competition, it is worth asking whether centralized production is the future of the industry. As with printing documents, it may be that the economics of printing and using DNA favor distributed production, perhaps even a desktop model. There is no fundamental barrier to integrating any demonstrated synthesis and assembly technologies into a desktop gene printer. Ultimately, over the long term, a globally expanding customer base will ultimately determine how sDNA is produced and used. Regardless of how current technology specifically impacts supply and prices, that customer base is increasing, and it is likely that the trends displayed in Figures 1.1 and 1.2 will continue for many years to come.

References

1 Ellis, T. *et al.* (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol.*, **3** (2), 109–118.

2 Baker, M. (2012) De novo genome assembly: what every biologist should know. *Nat. Methods*, **9**, 333–337.

3 Carlson, R. (2010) *Biology is Technology: The Promise, Peril, and New Business of Engineering Life*, Harvard University Press.

4 Loman, N. *et al.* (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.*, **30**, 434–439.

5 Quail, M. *et al* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.

6 Liu, L. (2012) Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.*, Article ID 251364.
7 Carlson, R. (2004) The pace and proliferation of biological technologies. *Biosecur. Bioterror.*, **1** (3), 203–214.
8 Moore, G.E. (1965) Cramming more components onto integrated circuits. *Electronics*, 114–117.
9 Under new management, The Economist, 2 April 2009.
10 Hutcheson, G.D. (2009) The economic implications of Moore's Law, in *Into the Nano Era: Moore's Law Beyond Planar Silicon CMOS* (ed. H. Huff), Springer.
11 Wikipedia Boeing 787 Dreamliner, http://en.wikipedia.org/wiki/Boeing_787_Dreamliner (accessed 28 May 2013).
12 Wired Better Place Runs Out of Juice, Reportedly Plans Bankruptcy, http://www.wired.com/autopia/2013/05/better-place-bankruptcy-report/ (accessed 10 January 2018).
13 Higgins, T. (2016) Tesla Generates a Profit. The Wall Street Journal (Oct. 26).
14 Reuters (2016) Tesla Just Raised $1.5 Billion to Finance Model 3 Production (May 20).
15 See Carlson, R. (2013) How Competition Improves DNA Sequencing, http://www.synthesis.cc/synthesis/2013/04/how_competition_improves_reading_dna (accessed 3 January 2018).
16 National Institutes of Health DNA Sequencing Costs, http://www.genome.gov/sequencingcosts/ (accessed 10 January 2018).
17 Carlson, R. (2009) The changing economics of DNA synthesis. *Nat. Biotechnol.*, **27** (12), 1091–1094.
18 Tian, J. *et al.* (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*, **432** (7020), 23.
19 Carlson (2009, 2010); Paddon, C.J. *et al.* (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, **496** (7446).
20 See, for example, Integrated DNA Technologies Building Biological Factories for Renewable and Sustainable Products, IDT Decoded Newsletter, http://www.idtdna.com/pages/decoded/decoded-articles/your-research/decoded/2013/01/18/building-biological-factories-for-renewable-and-sustainable-products (accessed 10 January 2018).
21 Carlson, R. (2016) On DNA and Transistors http://www.synthesis.cc/synthesis/2016/03/on_dna_and_transistors (accessed 3 January 2018).
22 See, for example, Carlson, R. (2009) On the Demise of Codon Devices, http://www.synthesis.cc/synthesis/2009/04/on_the_demise_of_condon_devices (accessed 3 January 2018).
23 See, for example, National Research Council (1999) *Funding a Revolution: Government Support for Computing Research*, National Academies Press, Washington, DC.
24 (a) Tørring, T. *et al.* (2011) DNA origami: a quantum leap for self-assembly of complex structures. *Chem. Soc. Rev.*, **12**, 5636; (b) Venkataraman, S. *et al.* (2007) An autonomous polymerization motor powered by DNA hybridization. *Nat. Nanotechnol.*, **2**, 490; (c) Gu, H. (2010) A proximity-based programmable DNA nanoscale assembly line. *Nature*, **465**, 202–205.

25 (a) Kershner, R. *et al.* (2009) Placement and orientation of individual DNA shapes on lithographically patterned surfaces. *Nat. Nanotechnol.*, **4**, 55; (b) Keren, K. *et al.* (2003) DNA-templated carbon nanotube field-effect transistor. *Science*, **302** (5649), 1380.

26 Organick, L., *et al*. Random access in large-scale DNA data storage, *Nat. Biotechnol.*, In Press.

27 Carlson, R. (2017) Guesstimating the Size of the Global Array Synthesis Market, http://www.synthesis.cc/synthesis/2017/8/guesstimating-the-size-of-the-global-array-synthesis-market (accessed 3 January 2018).

28 National Security Council (2009) The National Strategy for Countering Biological Threats.

29 Goldberg, M. (2013) BioFab: applying Moore's Law to DNA synthesis. *Ind. Biotechnol.*, **9** (1), 10–12.

30 Carlson (2010).

31 Carlson, R. (2016) Estimating the biotech sector's contribution to the US economy. *Nat. Biotechnol.*, **34**, 247–255.