

Large-scale *de novo* DNA synthesis: technologies and applications

Sriram Kosuri¹ & George M Church^{2,3}

For over 60 years, the synthetic production of new DNA sequences has helped researchers understand and engineer biology. Here we summarize methods and caveats for the *de novo* synthesis of DNA, with particular emphasis on recent technologies that allow for large-scale and low-cost production. In addition, we discuss emerging applications enabled by large-scale *de novo* DNA constructs, as well as the challenges and opportunities that lie ahead.

DNA is the predominant information carrier for life. The development of synthetic techniques to construct DNA has led to marked improvements in our ability to understand and engineer biology. For example, despite extensive efforts to unravel the genetic code using molecular genetics, modest capabilities to synthesize nucleic acids ultimately led to the code's unraveling¹. Today, reconstructions of complete viral and bacterial genomes are testaments of how far our synthetic capabilities have come.

Despite the improvements, our ability to read DNA is better than our ability to write it. Over the last decade, high-throughput sequencing technologies, here referred to as next-generation sequencing (NGS), have revolutionized the discovery and understanding of natural DNA sequence, with current installed capacity estimated at ~15 petabases per year². Large-scale data-sharing initiatives, such as GenBank, and continued improvements in bioinformatics software have made computational analyses on these data easier than ever. Such analyses help generate powerful statistical hypotheses for how genome sequence controls cellular functions across organisms and populations. In addition, NGS-based measurement tools allow for the analysis of many genetic and biochemical processes at unprecedented scale and low cost³. However, even though our ability to both generate hypotheses and measure outcomes has increased in scale owing to NGS, our ability to test such hypotheses experimentally still lags and is among the most limiting steps in the study of natural and engineered biology.

Specifically, a designed DNA construct is a physical instance of a hypothesis to be tested, whether it be a simple plasmid-based reporter or a whole-genome synthesis of an organism. Progress in large-scale, low-cost construction of desired DNA sequences could rapidly engender progress in both fundamental and applied biological research.

Although rapid modification of natural DNA sequence both *in vitro* and *in vivo* is useful for a variety of purposes, methodologies for *de novo* synthesis of DNA from nucleosides confer a number of unique advantages. First, engineering new functions often requires vastly modified or wholly new genetic sequences that are most easily accessed by *de novo* synthesis methodologies. Second, synthesized constructs are often superior to natural sequence for the study of genetic mechanisms because they can be designed to specifically test hypotheses for how sequence affects function. Finally, sequences that are targeted to be amplified or modified from natural sequences can be difficult to access (for example, from metagenomic data sets); thus, synthesis is the only practical way to experimentally study them.

Here we review technological innovations and applications for *de novo* DNA synthesis as distinct from assembly and modification of natural DNA sequence. We cover large-scale single-stranded DNA oligonucleotide (oligo) synthesis, assembly of these oligos into longer double-stranded DNA constructs, and emerging applications (Fig. 1).

¹Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California, USA. ²Wyss Institute for Biologically Inspired Engineering, Boston, Massachusetts, USA. ³Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. Correspondence should be addressed to S.K. (sri@ucla.edu).

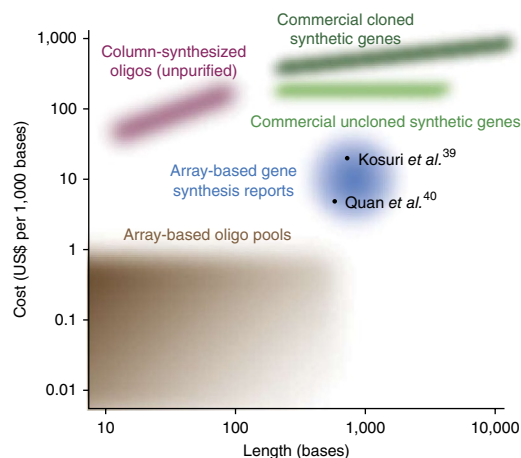


Figure 1 | Lengths and costs of different oligo and gene synthesis technologies. Commercial oligo synthesis from traditional vendors (pink) and array-based technologies (brown) are plotted according to commonly available length scales and price points. Costs of gene synthesis from commercial providers for cloned, sequence-verified genes (dark green) and unpurified DNA assemblies (light green) are shown, as are costs of gene synthesis from oligo pools (blue) derived from academic reports^{39,40}.

Oligo synthesis

Oligo synthesis has a long history, beginning in academic labs in the 1950s, followed by automation and commercialization in the 1980s and progressing into high-throughput array-based methods in the 1990s. This history has been extensively reviewed⁴, and we will mostly cover current approaches here to understand their advantages and trade-offs and their effect on downstream gene synthesis processes.

Column-based oligo synthesis. The first synthetic oligos were reported in the 1950s by Todd, Khorana and their coworkers, who used phosphodiester⁵, *H*-phosphonate⁶ and phosphotriester⁷ approaches. Today, the dominant chemistry for oligo synthesis occurs in automated instruments employing solid-phase phosphoramidite chemistry first developed by Marvin Caruthers in the 1980s⁸ (Fig. 2). Phosphoramidite-based oligo synthesis most commonly consists of a four-step cycle that adds bases one at a time to a growing oligo chain attached to a solid support. First, a dimethoxytrityl (DMT)-protected nucleoside phosphoramidite that is attached to a solid support is deprotected by removal of the DMT using trichloroacetic acid. Second, a new DMT-protected phosphoramidite is coupled to the 5' hydroxyl group of the growing oligo chain to form a phosphite triester. Third, a capping step acetylates any remaining unreacted 5' hydroxyl groups, making the unreacted oligo chains inert to further nucleoside additions, helping to alleviate deletion errors. Fourth, an iodine oxidation converts the phosphite to a phosphate, producing a cyanoethyl-protected phosphate backbone. The DMT protecting group is removed to allow the cycle to continue. This detritylation step is usually monitored to track coupling efficiencies as individual bases are added. After all nucleosides are added in series from 3' to 5', the completed oligo is removed from the solid support, and protecting groups on bases and the phosphate backbone are removed.

This automated process usually synthesizes 96–384 oligos simultaneously at scales from 10 to 100 nmol. Over the years, improvements in raw materials, automation, processing and purification have enabled routine synthesis of up to

~100 nt at costs of ~\$0.05–0.15 per nucleotide with error rates of ~1 in 200 nt or better. The limits on length and error rates of this process are due to a few major reasons. First, the yield for each step in the synthetic cycle must be very high, especially for the production of long oligos. For example, even 99% yield from each turn of the cycle will result in 13% final yield for a 200-nt oligo synthesis. In addition, depurination, particularly of adenosine, can occur during acidic detritylation and becomes particularly problematic in the production of long oligos^{9–11}. During the final removal of protecting groups from the bases and phosphate backbone, these abasic sites lead to cleavages that reduce the yield of long-length oligos. Finally, even successfully synthesized oligos contain appreciable errors^{12,13}. The dominant errors in purified oligos are single-base deletions that result from either failure to remove the DMT or combined inefficiencies in the coupling and capping steps. Newer chemistries and improved processes continue to arise and will further augment oligo length and quality⁴.

Array-based oligo synthesis. Starting in the early 1990s, Affymetrix developed methods for spatially localized polymer synthesis on surfaces using light-activated chemistries, which paved the way for the development of DNA microarrays^{14,15}. They used standard mask-based photolithographic techniques to selectively deprotect photolabile nucleoside phosphoramidites. Today, several technologies coexist to make spatially decoupled DNA microarrays. Maskless procedures (used, for example, by NimbleGen and LC Sciences) greatly simplified photolithographic techniques using programmable micromirror devices—similar to those found in modern-day digital projectors—to direct the light-based chemistries^{16,17}. Ink-jet-based printing of nucleotides on an arrayed surface (as Agilent uses) allowed for oligo synthesis using standard phosphoramidite chemistries^{18–20}. In addition, CombiMatrix (now CustomArray) developed semiconductor-based electrochemical acid production to selectively deprotect nucleosides²¹. Many other promising extensions and variations in microfluidic and microarray syntheses have been reported

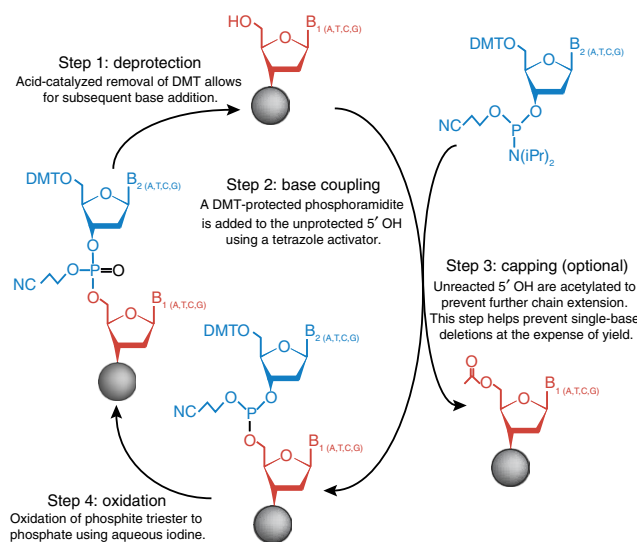


Figure 2 | Phosphoramidite chemistry. The four-step synthetic oligo synthesis is the most commonly used chemistry for the production of DNA oligos.

but are yet to become widely available or commercialized²². The use of microarray-derived oligos, whereby all the oligos synthesized from an array are cleaved and harvested as one ‘oligo pool’, has become increasingly popular as a cheap source of designed oligos. The scales, lengths and error rates vary greatly between vendors, but, to date, Agilent Technologies and CustomArray have provided oligos used in most recent publications (see “Emerging applications for large *de novo* DNA synthesis” below). Oligos produced from microarrays are 2–4 orders of magnitude cheaper than column-based oligos, with costs ranging from \$0.00001–0.001 per nucleotide, depending on length, scale and platform.

Gene synthesis

Small sets of oligos (usually 5–50 oligos) provide the raw substrate for constructing larger synthetic fragments (usually 200–3,000 bp) via a variety of methods collectively termed gene synthesis (‘gene’ refers to ‘gene length’ rather than the classic genetic definition). The first synthetic genes were short (80–200 bp); Gobind Khorana’s group used T4 DNA ligase to seal chemically synthesized oligos together^{23,24}. In ligation-based approaches, complementary overlapping strands are enzymatically joined to produce larger fragments. Initially this was done sequentially, but higher-quality oligos, use of thermostable ligases to improve hybridization stringency²⁵, and methods to produce and select for circular DNA²⁶ have allowed for ‘one-pot’ production of gene-length fragments. Polymerase cycling assembly (PCA)-based techniques use polymerase to extend overlapping oligos into a double-stranded fragment by cycling in a non-exponential process²⁷. Both ligation and PCA approaches usually rely on PCR to isolate and amplify full-length from partially assembled fragments and are often used in combination. More recently, Gibson and colleagues developed both *in vivo*^{28,29} and *in vitro*^{30,31} one-step protocols for assembling and cloning oligos directly into plasmid backbones. All of these protocols have been iteratively improved and underlie most academic and commercial gene synthesis efforts and have been reviewed elsewhere^{22,32–34}. Finally, because oligo synthesis and assembly techniques are prone to errors, gene-length fragments are often cloned and sequence verified, which can substantially add to the final cost.

There are advantages and disadvantages of each approach. High-stringency ligation-based syntheses reduce error rates because sequences with errors are less likely to hybridize and ligate. However, as both top and bottom strands need to be synthesized and oligos require phosphorylated 5’ ends, oligo costs are higher. PCA-based approaches can rely on overlapping regions of only 15–25 nt, thus allowing for fewer oligos per gene synthesized, but suffer from higher error rates owing to lack of hybridization-based error filtration. Also, because PCA-based approaches can contain regions encoded by only a single oligo, targeted diversity can be introduced into these locations. For both methods, sequences containing high GC content and secondary structure can inhibit assembly owing to misannealing and loss

Figure 3 | Different strategies for dealing with microarray oligo complexities. Top, Kosuri *et al.*³⁹ use amplification of barcoded subpools by PCR (thus eliminating background complexity), remove the barcode sequences and then assemble the genes. Bottom, Quan *et al.*⁴⁰ use a custom synthesizer to print oligos needed for any assembly into separate micropatterned wells. Leveraging the spatial separation that enables microarray synthesis, they then amplify and assemble these genes within the microwells themselves.

in ligation stringency. Also, because PCR is used as a final step to amplify constructs from partial assemblies, the lengths of synthetic genes generated from these methods are usually kept <5 kb for reliability; better assembly techniques exist for larger assemblies (see below).

Array-based gene synthesis. Even though microarray-based oligo pools are cheap, there are several challenges in using them for gene synthesis. First, although the numbers of oligos that can be produced in a pool are large, their individual concentrations are quite low for most existing gene synthesis protocols. Second, the error rates for oligo pools are usually higher than those for column-synthesized oligos. Finally, the sheer number of oligos produced leads to interference between gene assemblies, making it difficult to scale up.

Tian *et al.*³⁵ were the first to show how these problems could be overcome. They used PCR amplification to increase the concentration of the oligos before assembly, error-corrected them by hybridization to reverse complementary oligos (also constructed on the chip), and designed protein sequences computationally to avoid potential mishybridizations of the sequences. However, this work and contemporaneous reports^{36,37} used only dozens to hundreds of oligos at once. Scaling these methods proved difficult beyond pool sizes of 1,000 oligos³⁸. At greater pool complexities, where the advantages in cost would come to play, constructing any individual gene became difficult, presumably owing to spurious cross-hybridization during the assembly process. In addition, the methods required sufficient sequence orthogonality between synthesized genes, which limited potential applications. To alleviate these and other issues, two approaches were used that first isolated subpools of oligos required for any single assembly, thereby overcoming the concerns about both pool complexity and sequence orthogonality (Fig. 3). Kosuri *et al.*³⁹ used predefined barcodes that allowed PCR amplification of oligos participating in only a particular assembly and then removed the barcodes by digestions, which was followed by standard assembly of the genes. Quan *et al.*⁴⁰ used a custom ink-jet synthesizer²⁰ that synthesized subsets of oligos in physically separated microwells, where amplification and assembly were then done *in situ*. Both methods used much larger oligo pools (>10,000 oligos) and enzymatic error correction, which paved the way for commercialization in recent years (Gen9). Finally, two reports for using one-pot assembly of libraries of genes directly

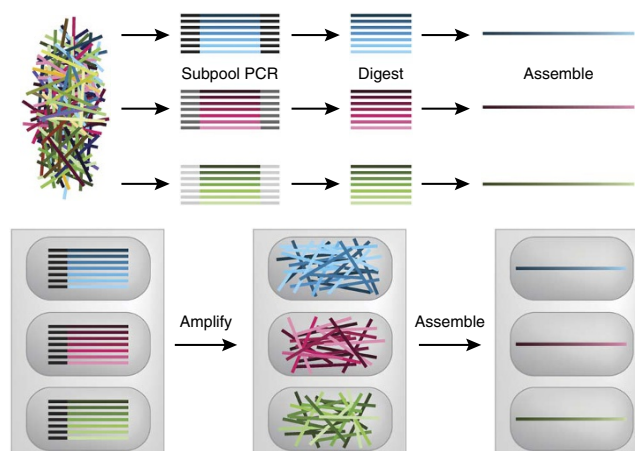


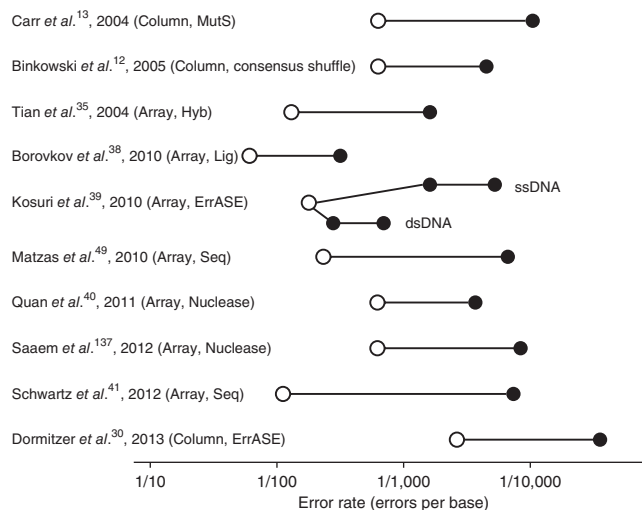
Figure 4 | Comparison of reported error rates from error-correction techniques. The error rates are included along with the indicated oligo source and error-correction methodology. When starting error rates were unreported, we estimated the error rates on the basis of the oligo sources and assembly method. Open circles denote starting error rates; filled circles denote error rates of assembled genes (two filled circles denote error rates before and after error correction). ssDNA, single-stranded DNA; dsDNA, double-stranded DNA; Column, column-synthesized oligos; Array, microarray-based oligo pools; Hyb, oligo hybridization-based error correction; Seq, NGS-based error correction; Lig, high-temperature ligation/hybridization-based error correction; Nuclease, nuclease-based error correction.

from large pools have been attempted, but these have been limited to joining one or two oligos simultaneously and suffer from large differences in dynamic range and the inability to make sequences that are similar to one another^{41,42}.

Cloning, error correction and verification. Once synthetic genes have been assembled, they contain a mixture of perfect and imperfect sequences resulting from both oligo synthesis and assembly errors (Fig. 4). Usually, synthesized genes are cloned into plasmids in *Escherichia coli* or yeast and then sequence verified by Sanger sequencing. These steps are expensive, time consuming and difficult to automate. Thus, reducing the number of clones required to get a perfect sequence is paramount. To help alleviate synthesis errors, early approaches fused synthetic protein-coding sequences in frame with a selectable marker encoding antibiotic resistance or a fluorescence marker^{43,44}. Because single-base deletions will most likely result in a frameshift, and thus loss of activity, it serves as a useful error-correction method but is limited to only protein-coding sequences.

More general methods of error correction usually depend upon a number of enzymatic techniques to reduce errors. All of these techniques rely on the fact that at any given position, most molecules possess the correct base. Heating and reannealing can force the formation of heteroduplexes that will contain disruptions to the canonical helical DNA structure. Such disruptions can be recognized and acted upon by several proteins. MutS binds heteroduplexes and can be used to filter errors by reverse purification¹³. Certain polymerases with exonuclease activity, endonucleases and resolvases can all cut or nick at such heteroduplex sites and, upon reamplification, can help filter errors^{12,26,45–47}. Commercial enzymatic cocktails such as ErrASE have been commonly used to help reduce errors in synthetic genes as well^{30,39}. Such error reduction can greatly reduce the cost and time of gene synthesis by bringing error rates low enough that the genes can be directly used for functional assays without *in vivo* cloning and sequence verification^{30,48}.

More recent approaches have leveraged NGS technologies to screen and then select for perfect sequences at either the oligo or gene level. Matzas *et al.*⁴⁹ used 454 sequencing (Roche) combined with a robotic pick-in-place pipette to mechanically pull reads with perfect sequence off the sequencing array and used these oligos for synthetic genes. Kim *et al.* also used 454 but marked individual molecules with random tags and then amplified constructs with perfect sequence⁴². Both approaches help reduce error rates, although they still suffer from sequencing errors and require more expensive long-read platforms. Schwartz *et al.*⁴¹ used Illumina sequencing of barcodes to pull out perfect sequences but



overcame limitations in length and sequencing errors through a tag-based consensus-sequencing approach. All three approaches substantially cut error rates and will continue to improve with the rapid progress in DNA sequencing technologies. These NGS-based error-correction approaches are also especially exciting for library-based constructions and synthesis methods because they allow correction without having to separate each gene assembly into individual reactions.

Larger DNA assemblies. Methods to produce larger assemblies from combinations of sequence-verified *de novo*-synthesized or amplified gene-length fragments have seen rapid advances⁵⁰. Both commercial and academic systems now allow combinations and libraries to be formed at high efficiency, fidelity and reliability for reasonably low reaction costs. Today, seamless assembly methods that do not leave behind scars at assembly junctions—including ligase cycling reaction⁵¹, Gibson assembly⁵², seamless ligation cloning extract⁵³, yeast assembly^{54,55}, circular polymerase extension cloning⁵⁶, sequence- and ligation-independent cloning⁵⁷, Golden Gate⁵⁸ and others—are routinely used and automated both in academic and industrial settings. Most of these methods can also be used for the generation of large, multicomponent libraries with little extra effort. Thus, for *de novo* synthesis applications, the cost and errors associated with generating gene-sized fragments dominate over those required for constructing larger DNA assemblies.

Emerging applications for large *de novo* DNA synthesis

Molecular tools. In 2004, one of the first and largest uses of synthetic DNA was in the development of human and mouse short hairpin RNA libraries⁵⁹. Using Agilent oligo pools, researchers synthesized 447,410 short hairpin RNAs targeting all human and mouse genes, which in total comprised ~44 Mb of *de novo* sequence. Improved lengths and NGS-based screening approaches have greatly expanded both the usage and applicability of such libraries. Now oligo pools are also being used routinely for targeted capture and resequencing of exons and other genomic regions of interest^{60–64} as well as to study genetic regulatory mechanisms such as genome-wide CpG methylation^{65,66}, RNA editing⁶⁷ and allele-specific expression⁶⁸. Another interesting use was the creation of a human peptidome phage-display library by

Larman *et al.*^{69,70} (413,611 peptides using ~58 Mb of DNA) for identifying autoimmune targets from patient samples. This same group also constructed a rationally designed human antibody library using oligo pools that were optimized for NGS analysis⁷¹. Similar phage-display methods were used to profile the interaction of PDZ domains against all known human and viral proteins' C termini⁷². Warner *et al.*⁷³ used oligo pools to construct genome-wide barcoded knockout and overexpression libraries in *E. coli* to facilitate selections for traits of interest. Recently, two groups have leveraged clustered, regularly interspaced, short palindromic repeats (CRISPR)-mediated gene targeting combined with large oligo pools to construct comprehensive, pooled and bar-coded knockout libraries in human cell lines^{74,75}. These molecular tools have all been directly enabled by the availability of microarray-based oligo pools, and we can only expect more in the coming years, with improvements in length, quality and the scale of such libraries.

Understanding and engineering regulatory elements.

Microarray-based oligo libraries have also been used to help uncover the structure and quantitative effects of regulatory elements that drive expression. In one of the earliest examples, Patwardhan *et al.*⁷⁶ used an oligo pool of 18,492 synthetic promoter mutants for bacteriophage and mammalian Pol II core promoters with corresponding barcodes for NGS readout to quantify expression differences and map important bases for core activity in these promoters. Later, Schlabach *et al.*⁷⁷ used 52,429 oligos designed to contain arrays of transcription factor binding sites to screen for synthetic strong promoters that work in a variety of human cell lines. Recent efforts focused on understanding the structural and functional characteristics of thousands of *cis*-regulatory sequences governing transcriptional, translational and other regulatory processes in mammalian, yeast and bacterial systems^{78–87}. Over the coming years, NGS-based methodologies that are developed to measure transcription, translation, epigenetics, splicing and other gene regulatory phenomena will also be used to analyze synthetic libraries. The goal is to understand which sequences are responsible for causal changes to these processes and how we can use them to engineer new functionalities.

Protein engineering. Protein engineering has always benefited from improvements in synthetic capabilities such as DNA shuffling, site-directed mutagenesis and low-cost gene synthesis. *De novo* synthesis, however, provides a more powerful tool to engineer new protein functions by taking advantage of computational design and metagenomic information. For example, Bayer *et al.*⁸⁸ synthesized 89 methyl halide transferase enzymes found in metagenomic sequences from diverse organisms and showed large improvements in enzymatic activities. As another example, Kudla *et al.*⁸⁹ and Quan *et al.*⁴⁰ constructed and characterized libraries of reporter genes (154 and 1,468 genes, respectively) to study codon usage. More recently, our group has used oligo pools combined with multiplexed reporter assays to construct >14,000 reporter constructs and measured their transcriptional and translational rates to understand how N-terminal codon bias affects protein expression⁷⁸. Finally, the development of deep mutational scanning techniques to measure structure-function relationships in multiplex will enable rapid characterization of large designed synthetic gene libraries as synthesis methods improve^{90–95}.

Genetic refactoring. To better understand and engineer particular genetic systems, researchers have begun to redesign and *de novo* synthesize these systems with orthogonal, well-defined gene sequences and control elements. Through refactoring, researchers hope to define and include known elements in a pathway while simultaneously disrupting any unknown control elements; this may serve as a better starting point for improvement or transplantation of these genetic systems. Early work resynthesized the first ~11 kb of bacteriophage T7 with a refactored surrogate that separated and defined individual genes and control elements and showed that the resultant phage was viable⁹⁶. More recent bacteriophage genome refactorings have helped improve biological understanding and usefulness^{97,98}. Temme *et al.*⁹⁹ extended these approaches to refactor the *Klebsiella oxytoca* nitrogen-fixation 20-gene cluster in *E. coli*. They removed noncoding sequences, eliminated non-essential genes, removed transcription factors, randomized codons and placed all the genes into seven operons with synthetic regulatory elements governing transcription and translation. The refactored system reconstituted functionality, albeit at reduced production levels. Improvements in the design and automated assembly of these refactored segments allowed reconstitution to wild-type production levels. Finally, Lajoie *et al.*⁴⁸ synthesized sequence-orthogonal variants of 42 *E. coli* essential genes using DNA microarrays and selected for function in order to explore the limits of genetic recoding. Again, such studies can powerfully explore regulatory requirements of genetic sequences but require currently expensive *de novo* synthesis methodologies and would greatly benefit from lower-cost gene synthesis.

Engineered genetic networks and metabolic pathways. Many researchers in synthetic biology are focused on building and optimizing genetic networks to control cellular behavior and metabolic pathways for chemical production¹⁰⁰. Although many of these efforts are focused on assembling already existing DNA in myriad combinations, *de novo* synthesis is still an important mainstay and will become increasingly so as we improve our ability to design and measure the effects of such assembled pathways. For instance, when building large, multicomponent systems, the number of orthogonal components becomes limiting. Large-scale studies of hundreds to tens of thousands of regulatory elements such as promoters, ribosome-binding sites and transcriptional terminators in *E. coli* usually use *de novo* synthesis of designs culled from both natural and designed sequences^{79,101–104}. The Voigt lab has also applied synthetic metagenomics for part mining to find libraries of orthogonal repressors (73 synthetic genes)¹⁰⁵ and transcription factors (62 synthetic genes)¹⁰⁶. Thus, as the genetic networks and pathways of engineered systems in synthetic biology get larger and studies move to new organisms, there will be increasing reliance on *de novo* DNA synthesis to generate requisite system components.

Whole-genome syntheses. *De novo* synthesis of genomes offers the promise of complete control of an organism's genetic code. Owing to the compact size of viruses and their importance in health and biotechnology, there has been tremendous progress in viral genomic reconstructions. Most synthetic reconstructions have been of RNA viruses by chemical synthesis of the required cDNA. In 2002, Eckard Wimmer's group first generated

infectious poliovirus from synthetic reconstruction of its full cDNA¹⁰⁷. Since then, dozens of RNA viruses have been chemically reconstructed—including the 1918 Spanish influenza¹⁰⁸, the likely coronavirus progenitors to severe acute respiratory syndrome¹⁰⁹ and many others^{110–120}—for purposes of viral attenuation, historical reconstructions, vaccine development and viral genomic studies. Several DNA-based bacteriophages have been synthesized *de novo* as well^{110,121}. Beyond viral genomes, over a series of studies, the Venter Institute designed, built, assembled and transplanted a fully synthetic bacterial genome to encode a viable organism⁵⁰. Such efforts are only increasing. For example, the design, synthesis and viability of synthetically designed yeast chromosomal arms was shown by Dymond *et al.*¹²², and work on a fully synthetic yeast genome is ongoing.

DNA nanotechnology. As a chemical polymer, DNA has several unique properties that make it intriguing. First, the compact helical form and simple base-pairing rules of double-stranded DNA allow us to consider DNA as a technology to reliably position atoms in three-dimensional (3D) space at nanometer resolutions. The emergence of DNA origami¹²³ and single-stranded tiles to form complex 2D and 3D shapes^{124–126} has been used by researchers to tackle problems from materials¹²⁷ to therapeutics¹²⁸. Base-pairing and strand-invasion properties of DNA have also allowed researchers to explore interesting information processing and computational capabilities using small libraries of oligos^{129–132}. Finally, direct encoding of digital information into DNA sequence has recently been shown to outpace most other technologies for data density in three dimensions^{133,134}. We are still in the early stages of this field, but harnessing advances in oligo-pool synthesis for such applications^{20,135} will allow researchers to test orders of magnitude more designs and hypotheses.

Future developments

Given the requisite investments, what is the cost of gene synthesis that we might expect to attain? Today, the cost of gene synthesis is on the same order as the cost of column-synthesized oligos used in their assembly. If gene synthesis transitioned to array-based oligos, there are no *prima facie* reasons why costs could not fall 3–5 orders of magnitude to be on par with the cost of oligo pools (\$1 per 10³–10⁵ bp). The benefits would likely be as dramatic as productivity gains due to NGS, because testing genetic hypotheses would become as simple as the design and analyses allow them to be. However, the large private investments that drove massive drops in the costs of integrated circuits and DNA sequencing were largely motivated by the reasonable expectation for their broad-based consumer-level uses: a processor in every pocket and a genome sequence for every person¹³⁶. Whereas potentially larger markets stand to benefit from cheap gene synthesis, including those of agriculture, chemicals, enzymes, materials and medicine, synthetic DNA serves only as a research tool for the ultimate product (with the possible exception of DNA nanotechnologies).

Can larger-scale synthetic biology efforts help increase demand sufficiently to spur investments? Even in academic research labs, the downstream cost of testing individual biological constructs for function is often far more expensive than the costs of the synthetic constructs themselves. Thus, reduction in gene synthesis costs will not tremendously affect the throughput and scale of current experimental workflows. However, the types of

experiments conducted might also significantly change. One data point to consider occurred a decade ago when microarrays were first leveraged for cheap oligo pools. Although initial reports used these pools as plug-in replacements for column-synthesized oligos, researchers quickly adapted to this increased synthetic capacity, using powerful bioinformatics tools to design large libraries of synthetic oligos and NGS-based multiplexed assays to measure their functional consequences simultaneously. This has recently led to many fruitful experiments at scales that only a few years ago would have been unimaginable for an individual investigator. Likewise, cheap gene synthesis will likely change how we use synthetic genes through the development of powerful design tools for libraries of genes, pathways and genomes as well as cheap, multiplexed assays to measure or select for their function. Such new experimental paradigms could engender far greater use of synthetic genes than is currently imagined today. The initial progress described in this Review warrants optimism and hopefully enough demand and investment to bring about large advances in our ability to design, build, test and analyze biological hypotheses and designs.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Nirenberg, M.W. & Matthaei, J.H. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA* **47**, 1588–1602 (1961).
 - Schatz, M.C. & Phillippy, A.M. The rise of a digital immune system. *Gigascience* **1**, 4 (2012).
 - Shendure, J. & Lieberman Aiden, E. The expanding scope of DNA sequencing. *Nat. Biotechnol.* **30**, 1084–1094 (2012).
 - Roy, S. & Caruthers, M. Synthesis of DNA/RNA and their analogs via phosphoramidite and *H*-phosphonate chemistries. *Molecules* **18**, 14268–14284 (2013).
 - Michelson, A.M. & Todd, A.R. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3': 5'-internucleotidic linkage. *J. Chem. Soc.* **1955**, 2632–2638 (1955).
 - Hall, R.H., Todd, A. & Webb, R.F. 644. Nucleotides. Part XLI. Mixed anhydrides as intermediates in the synthesis of dinucleoside phosphates. *J. Chem. Soc.* **1957**, 3291–3296 (1957).
 - Khorana, H.G., Razzell, W.E., Gilham, P.T., Tener, G.M. & Pol, E.H. Syntheses of dideoxyribonucleotides. *J. Am. Chem. Soc.* **79**, 1002–1003 (1957).
 - Beaucage, S.L. & Caruthers, M.H. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedr. Lett.* **22**, 1859–1862 (1981).
 - Efcavitch, J.W. & Heiner, C. Depurination as a yield decreasing mechanism in oligodeoxynucleotide synthesis. *Nucleosides Nucleotides* **4**, 267 (1985).
 - LeProust, E.M. *et al.* Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res.* **38**, 2522–2540 (2010).
- Iterative improvements to chemistries and processes for array-based oligo synthesis allow the production of long-length and low-error-rate oligo pools.**
- Septak, M. Kinetic studies on depurination and detriylation of CPG-bound intermediates during oligonucleotide synthesis. *Nucleic Acids Res.* **24**, 3053–3058 (1996).
 - Binkowski, B.F., Richmond, K.E., Kaysen, J., Sussman, M.R. & Belshaw, P.J. Correcting errors in synthetic DNA through consensus shuffling. *Nucleic Acids Res.* **33**, e55 (2005).
 - Carr, P.A. *et al.* Protein-mediated error correction for *de novo* DNA synthesis. *Nucleic Acids Res.* **32**, e1622 (2004).
 - Fodor, S.P. *et al.* Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773 (1991).

15. Pease, A.C. *et al.* Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. USA* **91**, 5022–5026 (1994).
16. Singh-Gasson, S. *et al.* Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.* **17**, 974–978 (1999).
17. Gao, X. *et al.* A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Res.* **29**, 4744 (2001).
18. Blanchard, A.P., Kaiser, R.J. & Hood, L.E. High-density oligonucleotide arrays. *Biosens. Bioelectron.* **11**, 687–690 (1996).
19. Hughes, T.R. *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **19**, 342–347 (2001).
20. Saaem, I., Ma, K.S., Marchi, A.N., LaBean, T.H. & Tian, J. *In situ* synthesis of DNA microarray on functionalized cyclic olefin copolymer substrate. *ACS Appl. Mater. Interfaces* **2**, 491–497 (2010).
21. Ghindilis, A.L. *et al.* CombiMatrix oligonucleotide arrays: genotyping and gene expression assays employing electrochemical detection. *Biosens. Bioelectron.* **22**, 1853–1860 (2007).
22. Tang, N., Ma, S. & Tian, J. in *Synthetic Biology* (ed. Zhao, H.) Ch. 1, 3–21 (Academic Press, 2013).
23. Agarwal, K.L. *et al.* Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* **227**, 27–34 (1970).
24. Sekiya, T. *et al.* Total synthesis of a tyrosine suppressor transfer RNA gene. XVI. Enzymatic joinings to form the total 207-base pair-long DNA. *J. Biol. Chem.* **254**, 5787–5801 (1979).
25. Au, L.C., Yang, F.Y., Yang, W.J., Lo, S.H. & Kao, C.F. Gene synthesis by a LCR-based approach: high-level production of leptin-L54 using synthetic gene in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **248**, 200–203 (1998).
26. Bang, D. & Church, G.M. Gene synthesis by circular assembly amplification. *Nat. Methods* **5**, 37–39 (2008).
27. Stemmer, W.P., Cramer, A., Ha, K.D., Brennan, T.M. & Heyneker, H.L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).
28. Gibson, D.G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.* **37**, 6984–6990 (2009).
29. Gibson, D.G. Oligonucleotide assembly in yeast to produce synthetic DNA fragments. *Methods Mol. Biol.* **852**, 11–21 (2012).
30. Dormitzer, P.R. *et al.* Synthetic generation of influenza vaccine viruses for rapid response to pandemics. *Sci. Transl. Med.* **5**, 185ra168 (2013).
31. Gibson, D.G., Smith, H.O., Hutchison, C.A. III, Venter, J.C. & Merryman, C. Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* **7**, 901–903 (2010).
32. Carr, P.A. & Church, G.M. Genome engineering. *Nat. Biotechnol.* **27**, 1151–1162 (2009).
33. Czar, M.J., Anderson, J.C., Bader, J.S. & Peccoud, J. Gene synthesis demystified. *Trends Biotechnol.* **27**, 63–72 (2009).
34. Xiong, A.S. *et al.* Chemical gene synthesis: strategies, softwares, error corrections, and applications. *FEMS Microbiol. Rev.* **32**, 522–540 (2008).
35. Tian, J. *et al.* Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).
- The first report to show that array-based oligo pools can be used to construct synthetic genes.**
36. Zhou, X. *et al.* Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Nucleic Acids Res.* **32**, 5409–5417 (2004).
37. Richmond, K.E. *et al.* Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Res.* **32**, 5011–5018 (2004).
38. Borovkov, A.Y. *et al.* High-quality gene assembly directly from unpurified mixtures of microarray-synthesized oligonucleotides. *Nucleic Acids Res.* **38**, e180 (2010).
39. Kosuri, S. *et al.* Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.* **28**, 1295–1299 (2010).
- Amplification of subpools of oligos prior to array-based gene assemblies helped solve issues related to gene assemblies from large oligo pools.**
40. Quan, J. *et al.* Parallel on-chip gene synthesis and application to optimization of protein expression. *Nat. Biotechnol.* **29**, 449–452 (2011).
- The development of a custom array-based oligo synthesizer that prints oligos in microwells allowed for cheap downstream gene assemblies.**
41. Schwartz, J.J., Lee, C. & Shendure, J. Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules. *Nat. Methods* **9**, 913–915 (2012).
- By combining oligo pools, molecular barcodes and NGS-based sequence verification and retrieval, this work enables vast reduction in gene synthesis errors.**
42. Kim, H. *et al.* ‘Shotgun DNA synthesis’ for the high-throughput construction of large DNA molecules. *Nucleic Acids Res.* **40**, e140 (2012).
43. Kim, H., Han, H., Shin, D. & Bang, D. A fluorescence selection method for accurate large-gene synthesis. *ChemBioChem* **11**, 2448–2452 (2010).
44. Allert, M., Cox, J.C. & Hellinga, H.W. Multifactorial determinants of protein expression in prokaryotic open reading frames. *J. Mol. Biol.* **402**, 905–918 (2010).
45. Smith, J. & Modrich, P. Removal of polymerase-produced mutant sequences from PCR products. *Proc. Natl. Acad. Sci. USA* **94**, 6847–6850 (1997).
46. Young, L. & Dong, Q. Two-step total gene synthesis method. *Nucleic Acids Res.* **32**, e59 (2004).
47. Fuhrmann, M., Oertel, W., Berthold, P. & Hegemann, P. Removal of mismatched bases from synthetic genes by enzymatic mismatch cleavage. *Nucleic Acids Res.* **33**, e58 (2005).
48. Lajoie, M.J. *et al.* Probing the limits of genetic recoding in essential genes. *Science* **342**, 361–363 (2013).
49. Matzas, M. *et al.* High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing. *Nat. Biotechnol.* **28**, 1291–1294 (2010).
50. Gibson, D.G. Programming biological operating systems: genome design, assembly and activation. *Nat. Methods* **11**, 521–526 (2014).
51. de Kok, S. *et al.* Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth. Biol.* **3**, 97–106 (2014).
52. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
53. Zhang, Y., Werling, U. & Edelmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40**, e55 (2012).
54. Gibson, D.G. *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. USA* **105**, 20404–20409 (2008).
55. Muller, H. *et al.* Assembling large DNA segments in yeast. *Methods Mol. Biol.* **852**, 133–150 (2012).
56. Quan, J. & Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS ONE* **4**, e6441 (2009).
57. Li, M.Z. & Elledge, S.J. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256 (2007).
58. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* **6**, e16765 (2011).
59. Cleary, M.A. *et al.* Production of complex nucleic acid libraries using highly parallel *in situ* oligonucleotide synthesis. *Nat. Methods* **1**, 241–248 (2004).
60. Tewhey, R. *et al.* Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biol.* **10**, R116 (2009).
61. Porreca, G.J. *et al.* Multiplex amplification of large sets of human exons. *Nat. Methods* **4**, 931–936 (2007).
62. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).
63. Depledge, D.P. *et al.* Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS ONE* **6**, e27805 (2011).
64. Geniez, S. *et al.* Targeted genome enrichment for efficient purification of endosymbiont DNA from host DNA. *Symbiosis* **58**, 201–207 (2012).
65. Li, J.B. *et al.* Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. *Genome Res.* **19**, 1606–1615 (2009).
66. Deng, J. *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.* **27**, 353–360 (2009).
67. Li, J.B. *et al.* Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213 (2009).
68. Zhang, K. *et al.* Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nat. Methods* **6**, 613–618 (2009).
69. Larman, H.B. *et al.* Autoantigen discovery with a synthetic human peptidome. *Nat. Biotechnol.* **29**, 535–541 (2011).
- The entire human peptidome is encoded in synthetic oligo pools and used to construct a phage-display library for autoantigen discovery.**

70. Larman, H.B. *et al.* PhIP-Seq characterization of autoantibodies from patients with multiple sclerosis, type 1 diabetes and rheumatoid arthritis. *J. Autoimmun.* **43**, 1–9 (2013).
71. Larman, H.B., Xu, G.J., Pavlova, N.N. & Elledge, S.J. Construction of a rationally designed antibody platform for sequencing-assisted selection. *Proc. Natl. Acad. Sci. USA* **109**, 18523–18528 (2012).
72. Ivarsson, Y. *et al.* Large-scale interaction profiling of PDZ domains through proteomic peptide-phage display using human and viral phage peptidomes. *Proc. Natl. Acad. Sci. USA* **111**, 2542–2547 (2014).
73. Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B. & Gill, R.T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.* **28**, 856–862 (2010).
74. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
Along with ref. 75, this study used oligo pools encoding Cas9-targeting RNAs to construct genome-wide knockout libraries in human cell lines.
75. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
Along with ref. 74, this study used oligo pools encoding Cas9-targeting RNAs to construct genome-wide knockout libraries in human cell lines.
76. Patwardhan, R.P. *et al.* High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* **27**, 1173–1175 (2009).
77. Schlabach, M.R., Hu, J.K., Li, M. & Elledge, S.J. Synthetic design of strong promoters. *Proc. Natl. Acad. Sci. USA* **107**, 2538–2543 (2010).
78. Goodman, D.B., Church, G.M. & Kosuri, S. Causes and effects of N-terminal codon bias in bacterial genes. *Science* **342**, 475–479 (2013).
79. Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **110**, 14024–14029 (2013).
An analysis of part composability using oligo pools by the multiplexed characterization of DNA, RNA and protein levels.
80. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).
One of the first uses of oligo pools to systematically dissect human enhancers using multiplexed reporter assays.
81. Kheradpour, P. *et al.* Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.* **23**, 800–811 (2013).
82. Kwasnieski, J.C., Mogno, I., Myers, C.A., Corbo, J.C. & Cohen, B.A. Complex effects of nucleotide variants in a mammalian *cis*-regulatory element. *Proc. Natl. Acad. Sci. USA* **109**, 19498–19503 (2012).
83. White, M.A., Myers, C.A., Corbo, J.C. & Cohen, B.A. Massively parallel *in vivo* enhancer assay reveals that highly local features determine the *cis*-regulatory function of ChIP-seq peaks. *Proc. Natl. Acad. Sci. USA* **110**, 11952–11957 (2013).
84. Mogno, I., Kwasnieski, J.C. & Cohen, B.A. Massively parallel synthetic promoter assays reveal the *in vivo* effects of binding site variants. *Genome Res.* **23**, 1908–1915 (2013).
85. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2009).
86. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30**, 521–530 (2012).
87. Smith, R.P. *et al.* Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat. Genet.* **45**, 1021–1028 (2013).
88. Bayer, T.S. *et al.* Synthesis of methyl halides from biomass using engineered microbes. *J. Am. Chem. Soc.* **131**, 6508–6515 (2009).
89. Kudla, G., Murray, A.W., Tollervey, D. & Plotkin, J.B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–258 (2009).
90. Araya, C.L. & Fowler, D.M. Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol.* **29**, 435–442 (2011).
91. Melamed, D., Young, D.L., Gamble, C.E., Miller, C.R. & Fields, S. Deep mutational scanning of an RRM domain of the *Saccharomyces cerevisiae* poly(A)-binding protein. *RNA* **19**, 1537–1551 (2013).
92. Fowler, D.M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat. Methods* **7**, 741–746 (2010).
93. Kim, I., Miller, C.R., Young, D.L. & Fields, S. High-throughput analysis of *in vivo* protein stability. *Mol. Cell. Proteomics* **12**, 3370–3378 (2013).
94. McLaughlin, R.N. Jr., Poelwijk, F.J., Raman, A., Gosal, W.S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491**, 138–142 (2012).
95. Reynolds, K.A., McLaughlin, R.N. & Ranganathan, R. Hot spots for allosteric regulation on protein surfaces. *Cell* **147**, 1564–1575 (2011).
96. Chan, L.Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol. Syst. Biol.* **1**, 2005.0018 (2005).
97. Jaschke, P.R., Lieberman, E.K., Rodriguez, J., Sierra, A. & Endy, D. A fully decompressed synthetic bacteriophage ϕ X174 genome assembled and archived in yeast. *Virology* **434**, 278–284 (2012).
98. Ghosh, D., Kohli, A.G., Moser, F., Endy, D. & Belcher, A.M. Refactored M13 bacteriophage as a platform for tumor cell imaging and drug delivery. *ACS Synth. Biol.* **1**, 576–582 (2012).
99. Temme, K., Zhao, D. & Voigt, C.A. Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc. Natl. Acad. Sci. USA* **109**, 7085–7090 (2012).
100. Brophy, J.A.N. & Voigt, C.A. Principles of genetic circuit design. *Nat. Methods* **11**, 508–520 (2014).
101. Cambray, G. *et al.* Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.* **41**, 5139–5148 (2013).
102. Chen, Y.J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10**, 659–664 (2013).
103. Mutalik, V.K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10**, 354–360 (2013).
104. Mutalik, V.K. *et al.* Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat. Methods* **10**, 347–353 (2013).
105. Stanton, B.C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* **10**, 99–105 (2014).
106. Rhodius, V.A. *et al.* Design of orthogonal genetic switches based on a crosstalk map of σ s, anti- σ s, and promoters. *Mol. Syst. Biol.* **9**, 702 (2013).
107. Cello, J., Paul, A.V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297**, 1016–1018 (2002).
108. Tumpey, T.M. *et al.* Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* **310**, 77–80 (2005).
109. Becker, M.M. *et al.* Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc. Natl. Acad. Sci. USA* **105**, 19944–19949 (2008).
110. Smith, H.O., Hutchison, C.A. III, Pfannkoch, C. & Venter, J.C. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA* **100**, 15440–15445 (2003).
111. Takehisa, J. *et al.* Generation of infectious molecular clones of simian immunodeficiency virus from fecal consensus sequences of wild chimpanzees. *J. Virol.* **81**, 7463–7475 (2007).
112. Burns, C.C. *et al.* Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. *J. Virol.* **83**, 9957–9969 (2009).
113. Dewannieux, M. *et al.* Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res.* **16**, 1548–1556 (2006).
114. Orlinger, K.K. *et al.* An inactivated West Nile virus vaccine derived from a chemically synthesized cDNA system. *Vaccine* **28**, 3318–3324 (2010).
115. Mueller, S. *et al.* Live attenuated influenza virus vaccines by computer-aided rational design. *Nat. Biotechnol.* **28**, 723–726 (2010).
116. Burns, C.C. *et al.* Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *J. Virol.* **80**, 3259–3272 (2006).
117. Lee, Y.N. & Bieniasz, P.D. Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog.* **3**, e10 (2007).
118. Mueller, S., Papamichail, D., Coleman, J.R., Skiena, S. & Wimmer, E. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J. Virol.* **80**, 9687–9696 (2006).
119. Wimmer, E. & Paul, A.V. Synthetic poliovirus and other designer viruses: what have we learned from them? *Annu. Rev. Microbiol.* **65**, 583–609 (2011).
120. Coleman, J.R. *et al.* Virus attenuation by genome-scale changes in codon pair bias. *Science* **320**, 1784–1787 (2008).
121. Liu, Y. *et al.* Whole-genome synthesis and characterization of viable S13-like bacteriophages. *PLoS ONE* **7**, e41124 (2012).

122. Dymond, J.S. *et al.* Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011).
123. Rothmund, P.W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
124. Dietz, H., Douglas, S.M. & Shih, W.M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725–730 (2009).
125. Douglas, S.M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
126. Ke, Y., Ong, L.L., Shih, W.M. & Yin, P. Three-dimensional structures self-assembled from DNA bricks. *Science* **338**, 1177–1183 (2012).
127. Maune, H.T. *et al.* Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates. *Nat. Nanotechnol.* **5**, 61–66 (2010).
128. Douglas, S.M., Bachelet, I. & Church, G.M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831–834 (2012).
129. Venkataraman, S., Dirks, R.M., Rothmund, P.W., Winfree, E. & Pierce, N.A. An autonomous polymerization motor powered by DNA hybridization. *Nat. Nanotechnol.* **2**, 490–494 (2007).
130. Soloveichik, D., Seelig, G. & Winfree, E. DNA as a universal substrate for chemical kinetics. *Proc. Natl. Acad. Sci. USA* **107**, 5393–5398 (2010).
131. Qian, L. & Winfree, E. Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**, 1196–1201 (2011).
132. Qian, L., Winfree, E. & Bruck, J. Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368–372 (2011).
133. Church, G.M., Gao, Y. & Kosuri, S. Next-generation digital information storage in DNA. *Science* **337**, 1628 (2012).
This report, along with ref. 134, describes the use of large DNA oligo pools to encode digital information at high density.
134. Goldman, N. *et al.* Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **494**, 77–80 (2013).
This report, along with ref. 133, describes the use of large DNA oligo pools to encode digital information at high density.
135. Marchi, A.N., Saaem, I., Tian, J. & LaBean, T.H. One-pot assembly of a hetero-dimeric DNA origami from chip-derived staples and double-stranded scaffold. *ACS Nano* **7**, 903–910 (2013).
136. Kosuri, S. & Sismour, A.M. When it rains, it pours. *ACS Synth. Biol.* **1**, 109–110 (2012).
137. Saaem, I., Ma, S., Quan, J. & Tian, J. Error correction of microchip synthesized genes using Surveyor nuclease. *Nucleic Acids Res.* **40**, e23 (2012).