POLICY FORUM

SYNTHETIC BIOLOGY

Technological challenges and milestones for writing genomes

Synthetic genomics requires improved technologies

By Nili Ostrov[1], Jacob Beal[2], Tom Ellis[3], D. Benjamin Gordon[4], Bogumil J. Karas[5], Henry H. Lee[1], Scott C. Lenaghan[6], Jeffery A. Schloss[7], Giovanni Stracquadanio[8], Axel Trefzer[9], Joel S. Bader[10], George M. Church[1,11], Cintia M. Coelho[12], J. William Efcavitch[13], Marc Güell[14], Leslie A. Mitchell[15], Alec A. K. Nielsen[16], Bill Peck[17], Alexander C. Smith[18], C. Neal Stewart Jr.[19], Hille Tekotte[20]

Engineering biology with recombinant DNA, broadly called synthetic biology, has progressed tremendously in the last decade, owing to continued industrialization of DNA synthesis, discovery and development of molecular tools and organisms, and increasingly sophisticated modeling and analytic tools. However, we have yet to understand the full potential of engineering biology because of our inability to write and test whole genomes, which we call synthetic genomics. Substantial improvements are needed to reduce the cost and increase the speed and reliability of genetic tools. Here, we identify emerging technologies and improvements to existing methods that will be needed in four major areas to advance synthetic genomics within the next 10 years: genome design, DNA synthesis, genome editing, and chromosome construction (see table). Similar to other large-scale projects for responsible advancement of innovative technologies, such as the Human Genome Project, an international, cross-disciplinary effort consisting of public and private entities will likely yield maximal return on investment and open new avenues of research and biotechnology.

The ability to design and write genomes of living cells provides distinctive opportunities to tackle problems that are intractable with current technologies. These transformative technologies could have widespread scientific, social, and economic impacts; thus, their development and adoption require proactive identification of potential pitfalls through ongoing public discussion. Accordingly, the outlook presented here is part of a broader effort by the international Genome Project-write (GP-write) consortium (*1*) to encourage inclusive conversation among scientists, lawyers, ethicists, educators, environmentalists, other experts, and stakeholders, as well as the general public, and ensure responsible, safe, and coordinated implementation of these new technologies.

Synthetic genomics is a relatively new field, and the majority of writing technologies (with the exception of commercial DNA synthesis) are still developed by academic research laboratories. Thus, unlike the reasonably predictable progress of engineering in more established industries, such as semiconductors, prediction of timelines and costs in these nascent fields remains highly speculative (see table).

GENOME DESIGN

Genome design aims to encode higher-level design criteria into DNA sequences at the chromosome scale. This will require computer-aided design (CAD) technologies to (i) reliably produce desired phenotypes; (ii) maximize the impact of the design, both in terms of experimental feedback and technical feasibility; and (iii) facilitate collaboration by employing standards for handling and exchanging design information. Current synthetic genomic CAD software, such as used for Sc2.0 (*2*), uses automation and collaboration tools for scaling DNA design from plasmids to entire chromosomes. Currently, estimation of functional effects of edits (e.g., gene deletions) is left to human experts. Looking forward, these tools will need to accurately predict the viability and phenotype of a cell from its genome design.

Although simple models are sufficient to handle silent edits (e.g., synonymous editing of coding sequences for watermarking), models of increasing complexity and precision will be necessary to predict how changes to genome sequence affect gene regulation and protein function. Although comprehensive mechanistic models of higher eukaryotes are likely decades away, machine learning approaches could improve phenotype prediction by leveraging the wealth of high-dimensional, high-throughput systems biology data in public databases and generated from genome writing projects, similar to how such techniques have been used to predict protein structures.

Leveraging these models, there is a need for experimental design tools to minimize the number of expensive iterations required to obtain successful genome designs. For instance, redesign of the relatively small *Mycoplasma* genome required four iterations, as well as genomic screens to identify essential genes (*3*). Larger projects will require many more intermediate stages of construction and supporting experiments. To provide the most valuable feedback and leverage the resulting data in subsequent designs, new algorithms are required to automate design of experiments and selection of appropriate engineering technologies for their implementation (e.g., "write versus edit").

A related, unaddressed need is to ensure compatibility of designed DNA with constraints from downstream synthesis, assembly, delivery, and analytical stages (e.g., software-guided parsing of chromosome sequence into synthesis-compatible fragments, or introduction of designated sequences to facilitate assembly and delivery). Tools will need to be adaptable to anticipate advances in writing technologies.

All of these efforts rely on and generate large datasets, which require stewardship to streamline model definition and facilitate sharing of results. Two major barriers for integrating biological information are data incompatibility and a lack of sufficiently descriptive metadata. We encourage researchers to use widely adopted data-exchange formats, e.g., GenBank and general feature format, and to continue to establish and adhere to standards for experimental metadata, e.g., the Synthetic Biology Open Language (SBOL).

Funding agencies and industrial stakeholders should prioritize support for long-term development of software and standard data formats, including collaboration, visualization, and quality control capabilities, similar to initiatives already in place for genomics, such as the Wellcome Trust Open Research Fund and the Chan Zuckerberg BioHub. Synthetic genomics software will not only enhance our ability to plan and execute large-scale genetic projects, but will likely also lead to fundamental methodological advances to move from correlation to causation of genotype-phenotype relationships.

DNA SYNTHESIS

Genome writing projects depend on large numbers of long [>5000 base pairs (bp)] and precise synthetic DNA constructs (*4*, *5*). How-

Author affiliations are listed in the supplementary materials.
Email: schlossjeff500@gmail.com

Key challenges and milestones for synthetic genomes

KEY TECHNOLOGY DEVELOPMENT TARGET	EXAMPLE OF DESIRED MILESTONES	ESTIMATED TIME (YEARS)
Genome design		
Develop tools for genome-scale design, visualization, and quality control.	Design a virus-proof mammalian chromosome.	3
Integrate structural information (2D and 3D) into genome design software	Predict the conformation of a synthetic yeast chromosome.	5
Develop sequence-to-phenotype whole-cell modeling.	Optimize metabolic profile, accurate to within twofold, for 100 key gene products of a synthetic virus-proof chromosome.	10
DNA synthesis		
Increase coupling efficiency for oligonucleotide synthesis.	Synthesize high-fidelity oligonucleotides longer than 500 nucleotides.	3
Increase efficiency of in vitro DNA assembly for fragments >20 kb.	Assemble 20 kb with >50% yield.	4
Develop methods for synthesis of difficult sequences, including homopolymers, high-GC content, and secondary structure.	Synthesize a centromere.	5
Develop enzymatic methods for direct synthesis of multikilobase DNA fragments.	Synthesize a 10-kb fragment (without assembly).	7
Decrease cost of DNA synthesis by 1000-fold.	Synthesize and assemble DNA for one haploid human genome (i.e., 3.2×10^9 bases) for $1000.	10
Genome editing		
Expand multiplexity and precision of DNA editing.	Simultaneously edit 1000 different targets in a single bacterial, mammalian, or plant cell with 1 off-target hit per 10,000 genomes.	2
Increase efficiency of homologous-directed repair (HDR)–mediated editing in mammalian and plant cells.	Perform HDR-mediated editing in nondividing mammalian cells at >90% efficiency.	3
Develop editing enzymes for precise substitution of any nucleotide at any desired genomic locus, with increased efficiency.	Perform allele editing of human cells at sites lacking PAM sequence, with >95% efficiency.	5
Chromosome construction		
Develop methods for temporal and spatial control of single chromosomes, such as chromatin state.	Engineer segregating, stable human artificial chromosomes (HAC).	2
Develop specialized host cells with high efficiency for DNA assembly, particularly for difficult-to-assemble sequences.	Establish in vivo chromosome assembly methods in the host *Streptomyces coelicolor* (72% GC content).	5
Develop efficient, inexpensive methods for routine and automated delivery of entire chromosomes into cells.	Demonstrate routine, device-based chromosome delivery in mammalian cells by cell fusion.	3
Develop methods for assembly and testing of Mb-size chromosomes.	Assemble a synthetic recoded human chromosome 21 from DNA fragments.	10

ever, chemical synthesis of DNA remains limited to production of short oligonucleotides (oligos), commonly 200 bp long. Although oligos have driven major advances in recombinant DNA technologies, larger DNA constructs require assembly of multiple oligos, a process that is laborious and lossy. Routine production of long, precise fragments of synthetic DNA would be desirable.

Although DNA has become more available through commercial vendors in recent years as a result of industrialization by parallelization and miniaturization, there has been little improvement to the underlying phosphoramidite chemistry, which limits DNA length, production speed, and cost. Accordingly, construction of whole chromosomes remains expensive and time-consuming. For example, array synthesis of oligos costs approximately $0.0005 per nucleotide, yielding an estimate of $1.5 million for synthesizing 3 gigabases of DNA—the size of a human genome. Radical new approaches to DNA assembly, purification, and synthesis processes are thus required to achieve substantial advancement on cost and ease.

Innovations to minimize or eliminate the need for assembly, error correction, and cloning of DNA fragments assembled from oligos could boost productivity of current DNA synthesis infrastructure. To increase the yield of perfect sequences, currently ranging from 5 to 60% (*6*), hybridization and error correction can benefit from the engineering of high-fidelity polymerases and ligases. These advancements, largely driven by industry, will decrease operating costs and production time. Cloning efficiency can also be enhanced by harnessing hosts with rapid division and/or high recombination rates, or obviated by using cell-free cloning and artificial cells. These technologies require fundamental research before they reach commercial readiness.

New technologies capable of synthesizing high-quality long DNA fragments would fundamentally alter chromosome-scale engineering. Recently, rapid production of short sequence-defined single-stranded DNA (ssDNA) has been reported by using the template-independent DNA polymerase TdT (terminal deoxynucleotidyl transferase) (*7*, *8*). TdT offers the potential for directly synthesizing multikilobase sequences with increased polymerization rate and higher coupling efficiencies. To compete with existing phosphoramidite chemistry, enzymatic synthesis approaches should be further developed to address complex sequences and achieve precise, high-quality DNA in an automated and cost-effective fashion. These efforts, ripe for startup company attention, will benefit from continued investment in fundamental research to elucidate molecular mechanisms of enzymatic terminal transferase reactions.

To support the scale and quality of DNA required for genome-scale projects, enhancement of electromechanical systems as well as innovative biological tools are needed. Continued increase in throughput may be achieved by further parallelization and miniaturization—for example, by semiconductor fabrication or droplet-based techniques. Increases in DNA quality and production speed will be influenced by use of biological tools such as enzymes and organisms.

GENOME EDITING

Powerful new DNA editing tools have lowered the technical barriers for performing highly precise genetic and epigenetic modifications. Multiplexed editing of an intact genome could sharply decrease the time and labor required to generate a large number of modifications and, in some instances, circumvent the need for de novo synthesis and chromosome assembly.

Yet, despite considerable success with programmable nucleases such as Cas9, TALEN, and ZFN in multiple cell types with exquisite temporal and tissue-specific control, genome-scale editing remains limited. A localized, nuclease-induced double-strand break can be used to increase editing efficiency at each locus, but multiple simultaneous breaks often cause cellular toxicity. To avert toxicity, "base editor" enzymes were engineered in which the nuclease is replaced by enzymatic base modification (*9*), achieving simultaneous editing of over 13,000 Alu

repetitive elements in human cells by using a small number of guides (*10*). Other engineered Cas9 tools are used for repression, activation, or targeted insertion of DNA. However, a major bottleneck for multiplex genome-wide editing remains the delivery of guide RNAs (gRNAs), as multiple unique genome changes necessitate the presence of multiple unique gRNAs in the same cell. We anticipate that this barrier, as well as off-target mutagenesis and constraints caused by sequence specificity of editing enzymes [e.g., protospacer adjacent motif (PAM) sequence requirement], will be overcome to enable routine multiplex editing.

Genome-scale editing can also be accomplished by oligo recombineering (*11*), which relies on homologous recombination (HR) and has reduced in vivo toxicity. However, this technique is currently limited to a handful of organisms in which high-efficiency HR can be catalyzed by a recombinase that uses a donor ssDNA for targeting. To enable recombineering to edit plants and mammalian cells, new recombinase enzymes must be discovered or designed. It may also be necessary to map and modulate an organism's repair pathways to improve HR.

Additionally, comprehensive suites of molecular tools should be generated to accelerate testing and optimization of genome editing. For example, programmable TALEN or ZFN nucleases can be generated for targeting all UAG stop codons in human cells. Similarly, CRISPR-Cas9 guide libraries targeting all PAMs can be used to explore multiplexed, allele-specific targeting in plant, human, or fungal cells. Efforts to generate these genome-wide resources will provide experimental evidence of "accessibility maps" that reflect editing efficiency variability across genomic targets. Such data will optimize the choice of target sequences, inform predictive computational models, and deepen our knowledge of chromosome structure, folding, and repair.

CHROMOSOME CONSTRUCTION

The most critical hurdle facing synthetic genomics is the assembly and introduction of synthetic chromosomes into host cells. How does one stitch together all the DNA pieces required to construct a fully functional chromosome? Once constructed, how can we control chromosome localization and architecture to ensure cell viability? How do we replace all chromosome copies in polyploid organisms? As the genomes of most free-living organisms are larger than 2 Mb, methods for routine manipulation of large DNA fragments are critically needed.

Despite recent improvements in DNA synthesis and in vitro cloning, such methods are not efficient for construction of entire chromosomes. Higher-order assembly of chromosomes at least 1 Mb in length can be performed by in vivo HR in the yeast *Saccharomyces cerevisiae*—a robust technique used in all synthetic chromosomes reported to date, including viral, bacterial, yeast, and algal chromosomes, as well as fragments of mice and human genomes (*12*, *13*).

The efficiency of DNA assembly in *S. cerevisiae* has not been found in other genetically tractable organisms. To expand the toolkit for writing specialized synthetic chromosomes that are difficult to assemble in *S. cerevisiae*, new HR-proficient cloning organisms should be developed that tolerate high GC, direct repeats, and desired posttranscriptional modifications. Organisms compatible with extreme environments such as desert, deep ocean, or space travel may provide new routes for DNA assembly, such as found in the polyextremophile *Deinococcus radiodurans*.

Once a chromosome is constructed, delivery and manipulation become the primary engineering bottleneck in most desired hosts. To deliver megabase-scale constructs, robust, high-throughput DNA transformation methods must be developed in a variety of organisms spanning genera. For example, breakthroughs in DNA delivery can revolutionize plant engineering, which is currently hindered by species-specific, labor-intensive transformation methods and limited by traditionally conservative funding. High-risk, high-reward funding to support modernization of plant research, such as development of tissue-culture–independent DNA delivery methods, is pertinent for synthetic biology enabled improvements of agricultural organisms. Automation of specialized methods for chromosome transfer between yeast, bacteria, plants, and mammalian cells (such as cell fusion, genome transplantation, or microinjection) requires multidisciplinary funding opportunities aimed at bridging microfluidics with traditional cell and molecular biology work. It is essential that early proof-of-concept efforts be supported at government and foundation levels.

Many cellular forces that shape genome structure and function remain largely unknown. Fundamental studies are needed to elucidate mechanisms by which sequence and epigenetic regulation guide inter- and intrachromosomal interactions and determine genome architecture. Emerging technologies for programmable modifications of chromatin, such as insulators guiding chromatin remodeling, safe harbor sites for DNA insertion, and orthogonal recombinase enzymes, will be necessary for developing gene therapies (*14*). Better understanding of organelle genomes (plastid, mitochondrion), which remain extremely difficult to engineer, would offer new routes for stable maintenance and incorporation of artificial chromosomes.

A final challenge when introducing synthetic constructs is to quickly determine whether they perform as desired in the destination cell. Whole-genome DNA and RNA sequencing will serve as a first-pass verification of chromosome integrity. Tailored cell lines with phenotypic reporters may be developed for assessing the performance of large synthetic constructs. Reliable organoid models and a clear understanding of regulation and expression changes that drive organismal development will be key to extrapolating results from single cells to the design of functional chromosomes for multicellular organisms.

GLOBAL AND MULTIDISCIPLINARY

New technologies may come from efforts in synthetic biochemistry, such as programmable synthetic protocells, from progress at the interface of hardware and wetware such as solid-phase DNA assembly platforms, or from findings in basic bioscience research—for example, by uncovering valuable new enzymes or delivery systems. Innovation will be driven by government grants and genomic and cancer institutes, with a growing role for BioFoundries, emerging hubs for automation of bioengineering. A highly interdisciplinary, multinational effort from government and private sectors will help achieve and disseminate these advances to make an impact in biomedical, pharmaceutical, agricultural, and chemical industries. ■

REFERENCES AND NOTES

1. J. D. Boeke *et al.*, *Science* **353**, 126 (2016).
2. S. M. Richardson *et al.*, *Science* **355**, 1040 (2017).
3. C. A. Hutchison 3rd *et al.*, *Science* **351**, aad6253 (2016).
4. J. Fredens *et al.*, *Nature* **569**, 514 (2019).
5. N. Ostrov *et al.*, *Science* **353**, 819 (2016).
6. N. B. Lubock, D. Zhang, A. M. Sidore, G. M. Church, S. Kosuri, *Nucleic Acids Res.* **45**, 9206 (2017).
7. S. Palluk *et al.*, *Nat. Biotechnol.* **36**, 645 (2018).
8. H. H. Lee *et al.*, *Nat. Commun.* **10**, 2383 (2019).
9. H. A. Rees, D. R. Liu, *Nat. Rev. Genet.* **19**, 770 (2018).
10. C. J. Smith *et al.*, bioRxiv 574020 (2019). https://doi.org/10.1101/574020.
11. M. J. Lajoie *et al.*, *Science* **342**, 357 (2013).
12. B. J. Karas *et al.*, *Chromosome Res.* **23**, 57 (2015).
13. L. A. Mitchell *et al.*, *Science* **355**, eaaf4831 (2017).
14. F. Ceroni, T. Ellis, *Nat. Rev. Mol. Cell Biol.* **19**, 481 (2018).

ACKNOWLEDGMENTS

We thank members of the GP-write technology and infrastructure working group for discussions. We also thank N. J. Kelley and A. Schwartz for administration and coordination of group discussions. B.J.K. is a cofounder, chief executive officer, and shareholder of Designer Microbes Inc. H.H.L. is a cofounder of Kern Systems. J.S.B. is a founder and director of Neochromosome, Inc. G.M.C.'s financial interests are listed at http://arep.med.harvard.edu/gmc/tech.html. J.W.E. is a cofounder and board member of Molecular Assemblies Inc. L.A.M. is a cofounder, president, and shareholder of Neochromosome, Inc. A.A.K.N. is a cofounder, chief executive officer, and shareholder of Asimov Inc. B.P. is a cofounder, chief technology officer, and shareholder of Twist Bioscience

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6463/310/suppl/DC1

10.1126/science.aay0339

Technological challenges and milestones for writing genomes

Nili Ostrov, Jacob Beal, Tom Ellis, D. Benjamin Gordon, Bogumil J. Karas, Henry H. Lee, Scott C. Lenaghan, Jeffery A. Schloss, Giovanni Stracquadanio, Axel Trefzer, Joel S. Bader, George M. Church, Cintia M. Coelho, J. William Efcavitch, Marc Güell, Leslie A. Mitchell, Alec A. K. Nielsen, Bill Peck, Alexander C. Smith, C. Neal Stewart Jr. and Hille Tekotte

Science **366** (6463), 310-312.
DOI: 10.1126/science.aay0339