POLICY FORUM

GENOME ENGINEERING

The Genome Project–Write

We need technology and an ethical framework for genome-scale engineering

By **Jef D. Boeke,* George Church,* Andrew Hessel,* Nancy J. Kelley,* Adam Arkin, Yizhi Cai, Rob Carlson, Aravinda Chakravarti, Virginia W. Cornish, Liam Holt, Farren J. Isaacs, Todd Kuiken, Marc Lajoie, Tracy Lessor, Jeantine Lunshof, Matthew T. Maurano, Leslie A. Mitchell, Jasper Rine, Susan Rosser, Neville E. Sanjana, Pamela A. Silver, David Valle, Harris Wang, Jeffrey C. Way, Luhan Yang**

The Human Genome Project ("HGP-read"), nominally completed in 2004, aimed to sequence the human genome and to improve the technology, cost, and quality of DNA sequencing (*1*, *2*). It was biology's first genome-scale project and at the time was considered controversial by some. Now, it is recognized as one of the great feats of exploration, one that has revolutionized science and medicine.

Although sequencing, analyzing, and editing DNA continue to advance at a breakneck pace, the capability for constructing DNA sequences in cells is mostly limited to a small number of short segments, which restricts the ability to manipulate and understand biological systems. Further understanding of genetic blueprints could come from construction of large, gigabase (Gb)–sized animal and plant genomes, including the human genome, which would, in turn, drive development of tools and methods to facilitate large-scale synthesis and editing of genomes. To this end, we propose the Human Genome Project–Write (HGP-write), named to honor HGP-read but embracing synthesis of all large genomes.

*The list of author affiliations is available in the supplementary materials. *These authors contributed equally to this work. Email: jef.boeke@nyumc.org*

RESPONSIBLE INNOVATION

Genome synthesis is a logical extension of the genetic engineering tools that have been used safely within the biotech industry for ~40 years and have provided important societal benefits. However, recent technological advancements—e.g., standardized gene parts, whole-genome synthesis, and clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9 genome editing technology (*3*, *4*)—are revolutionizing the field (*5*). Some applications are controversial; human germline editing in particular has raised intense moral debate (*6*). As human genome-scale synthesis appears increasingly feasible, a coordinated scientific effort to understand, discuss, and apply large-genome engineering technologies is timely. HGP-write will require public involvement and consideration of ethical, legal, and social implications (ELSI) from the start. Responsible innovation requires more than ELSI, though, and involves identifying common goals important to scientists and the wider public through timely and detailed consultation among diverse stakeholders.

We will enable broad public discourse on HGP-write; having such conversations well in advance of project implementation will guide emerging capabilities in science and contribute to societal decision-making.

Human chromosomes and nucleus as visualized by scanning electron microscopy.

Through open and ongoing dialogue, common goals can be identified. Informed consent must take local and regional values into account and enable true decision-making on particularly sensitive use of cells and DNA from certain sources. Finally, the highest biosafety standards should guide project work and safety for lab workers and research participants, and ecosystems should pervade the design process. A priority will be cost reduction of both genome engineering and testing tools to aid in equitable distribution of benefits—e.g., enabling research on crop plants and infectious agents and vectors in developing nations.

To ensure responsible innovation and ongoing consideration of ELSI, a percentage of all research funds could be dedicated to these issues, enabling inclusive decision-making on the topics mentioned above (*7*). In addition, there should be equitable distribution of any early and future benefits in view of diverse and pressing needs in different global regions. The broad scope and novelty of the project call for consideration of appropriate regulation alongside development of the science and societal debates. National and international laws and regulations differ, and as in stem cell research, a major burden of responsibility for setting standards rests with the scientists and their community. Existing stem cell research guidelines (*8*) may serve as a useful template.

FROM OBSERVATION TO ACTION

The primary goal of HGP-write is to reduce the costs of engineering and testing large genomes (0.1 billion to 100 billion base pairs) in cell lines to 1/1000th of previous efforts within 10 years. This will include whole-genome engineering of human cell lines and other organisms of agricultural and public health significance or those needed to interpret human biological functions—i.e., gene regulation, genetic diseases, and evolutionary processes.

This goal is necessarily ambitious, because building a human genome at today's prices would cost more than HGP-read (*9*) (see fig. S1). However, an expectation of HGP-write is that a sharp price drop will be catalyzed as new technology development occurs apace with advancement of the project, as with the cost of DNA sequencing in HGP-read. Small viral (*10*) and bacterial (*11*) genomes synthesized from scratch and organisms with recoded genomes (*12*) derived from large-scale genome editing (*13*) have demonstrated the feasibility and utility of synthetic genomes. By focusing on building the 3 Gb of human DNA, HGP-write would

push current conceptual and technical limits by orders of magnitude and deliver important scientific advances.

HGP-write will aim to address a number of human health challenges. Potential applications include growing transplantable human organs; engineering immunity to viruses in cell lines via genome-wide recoding (*12*); engineering cancer resistance into new therapeutic cell lines; and accelerating high-productivity, cost-efficient vaccine and pharmaceutical development by using human cells and organoids. The project could encourage broad intellectual property access via patent pooling. Extreme cost-reduction is feasible, as demonstrated by the $1000 genome grant program (*2*), as well as sharing of CRISPR tools from over 80 labs through www.addgene.org. Furthermore, because DNA synthesis, like sequencing and computation, is foundational technology, HGP-write could also facilitate biological engineering of many organisms, accelerating research and development (R&D) across a broad spectrum of life sciences and supporting basic R&D of new bio-based therapies, vaccines, materials, energy sources, disease vector control, and nutrition.

Ultrasafe cell lines

Some properties of "ultrasafe" cells with a pervasively reengineered genome.

Virus resistance
Can be conferred by systematically recoding certain codons across all genes. Subsequent deletion of tRNA genes would generate a cell line resistant to viruses.

Improved cancer resistance
Tumor suppressor genes could be made multicopy; genes like p53 could be recoded to eliminate CpG dinucleotides that give rise to "hotspot" mutations.

Other useful traits
Delete potentially deleterious genes such as prion genes.

Improved genome stability
Comprehensively eliminate endogenous repetitive "selfish DNA" elements.

Fail-safe security
Prevent formation of germ cells, e.g., by removing transcriptional regulators.

Applications
"Go to" cell line(s) for stem cell therapies; robust production of biologics.

PILOT PROJECTS

Similarly to other Gb-scale genomic projects, including HGP-read, ENCODE (which aims to map genome functional elements), and Sc2.0 (which is synthesizing a heavily edited yeast genome) (*14*, *15*), HGP-write would be conducted in phases with explicit goals and metrics. Each of the earlier large-scale projects began with pilot projects focused on a fraction of the genome, typically ~1%. For HGP-write, the pilots should provide resources of immediate value for advanced biomedical research and/or biotech development. Technology development will likely also occur early in the project to propel large-scale genome design and engineering.

A series of pilot projects making use of very long DNA sequences that are nonetheless short of a full genome are anticipated: (i) synthesizing "full" gene loci with accompanying noncoding DNA to help explain still-enigmatic roles of noncoding DNA variants in regulating gene expression and leading to more comprehensive models for the role of noncoding genetic variation in common human diseases and traits; (ii) constructing specific chromosomes—e.g., chromosome 21—or complex cancer genotypes to more comprehensively model human disease; (iii) producing specialized chromosomes encoding one or several pathways—e.g., all genes needed to make a prototrophic human cell or pathways to transform the pig genome to make it more amenable as a source for human organ transplantation; (iv) a potential transformation of gene therapy, with freedom to deliver many genes and control circuits to improve safety and efficacy, provided that delivery challenges can be met. Indeed, many substantial and useful innovations may be realized in such "stepping-stone" projects that are short of whole-genome reengineering but require substantial improvement in synthesis capacity of Mb- to Gb-sized DNA. Both genome-wide and more modest changes could be tested for their impact on, e.g., organoid development and function in vitro, facilitated by ongoing progress in stem cell differentiation and "organ-on-a-chip" technologies. Novel cell culture technologies may, in some cases, be many times more cost-effective and accurate than current whole-organism testing.

Additional pilot projects being considered include (v) using induced pluripotent stem cells (*16*) to construct an "ultrasafe" human cell line via comprehensive recoding of protein-coding regions and deletion of corresponding genome features to increase safety of such a cell line (see the box); and (vi) developing a homozygous reference genome bearing the most common pan-human allele (or allele ancestral to a given human population) at each position to develop cells powered by "baseline" human genomes. Comparison with this baseline will aid in dissecting complex phenotypes, such as disease susceptibility. The pervasive nature of the required changes makes whole- or partial-genome synthesis an efficient route to these goals.

PROJECT LAUNCH AND ADMINISTRATION

The goal is to launch HGP-write in 2016 with $100 million in committed support, from public, private, philanthropic, industry, and academic sources from around the world. The costs of the project lie not only in obtaining de novo synthesized DNA but in the assembly, integration, and functional assays required to evaluate and understand the modified genomes. Total project costs are difficult to estimate but would likely be less than the $3 billion cost of HGP-read.

HGP-write could be implemented through one or more centers [similar to Centers of Excellence in Genomic Science (CEGS) and the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) initiative centers] that will coordinate and support formation and work of multi-institutional and interdisciplinary research teams working in a highly integrated fashion responsive to and engaged with a broad public outreach.

We celebrate 2016—the 25th anniversary of HGP-read—as a major step forward for human knowledge and health. In this spirit, we look forward to the launch of HGP-write.

REFERENCES AND NOTES

1. International Human Genome Sequencing Consortium, *Nature* **431**, 931 (2004).
2. C. W. Fuller *et al.*, *Nat. Biotechnol.* **27**,1013 (2009).
3. L. Cong *et al.*, *Science* **339**, 819 (2013).
4. P. Mali *et al.*, *Science* **339**, 823 (2013).
5. A. D. Haimovich, P. Muir, F. J. Isaacs, *Nat. Rev. Genet.* **16**, 501 (2015).
6. D. Baltimore *et al.*, *Science* **348**, 36 (2015).
7. Presidential Commission for the Study of Bioethical Issues, *New Directions: The Ethics of Synthetic Biology and Emerging Technologies* (Washington, DC, 2010).
8. J. Kimmelman *et al.*, *Nature* **533**, 311 (2016).
9. P. A. Carr, G. M. Church, *Nat. Biotechnol.* **27**, 1151 (2009).
10. J. Cello, A. V. Paul, E. Wimmer, *Science* **297**, 1016 (2002).
11. D. G. Gibson *et al.*, *Science* **329**, 52 (2010).
12. M. J. Lajoie *et al.*, *Science* **342**, 357 (2013).
13. F. J. Isaacs *et al.*, *Science* **333**, 348 (2011).
14. N. Annaluru *et al.*, *Science* **344**, 55 (2014).
15. J. S. Dymond *et al.*, *Nature* **477**, 471 (2011).
16. K. Takahashi, S. Yamanaka, *Cell* **126**, 663 (2006).

ACKNOWLEDGMENTS

This paper is the result of meetings held at NYU Langone Medical Center 31 October 2015 and Harvard Medical School on 10 May 2016. F.J.I. is a cofounder of enEvolv, Inc. G.C. has financial relationships with Gen9, Editas, enEvolv, and Egenesis (companies directly related to this article; for a full list of G.C.'s financial relationships, see arep.med.harvard.edu/gmc/tech.html). J.D.B. is on the board of directors of Neochromosome, Inc., and owns stock in Recombinetics, Inc., and Sample6, Inc. A.H. has investments in Autodesk, Inc. G.C. is an inventor on patents and patent applications filed by Harvard Medical School that cover synthesis, assembly, and testing of large DNAs. This Policy Forum is the opinion of the authors and not that of their employers or institutions. The authors gratefully acknowledge the financial support of Autodesk, sponsor of the meetings.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/353/6295/126/suppl/DC1

Published online 2 June 2016

10.1126/science.aaf6850

The Genome Project-Write

Jef D. Boeke, George Church, Andrew Hessel, Nancy J. Kelley, Adam Arkin, Yizhi Cai, Rob Carlson, Aravinda Chakravarti, Virginia W. Cornish, Liam Holt, Farren J. Isaacs, Todd Kuiken, Marc Lajoie, Tracy Lessor, Jeantine Lunshof, Matthew T. Maurano, Leslie A. Mitchell, Jasper Rine, Susan Rosser, Neville E. Sanjana, Pamela A. Silver, David Valle, Harris Wang, Jeffrey C. Way and Luhan Yang (June 2, 2016)

Science **353** (6295), 126-127. [doi: 10.1126/science.aaf6850]
originally published online June 2, 2016

Editor's Summary

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

]www.sciencemag.org/cgi/content/full/science.aaf6850/DC1

Supplementary Materials for

The Genome Project–Write

Jef D. Boeke,[1]*† George Church,[2]* Andrew Hessel,[3]* Nancy J. Kelley,[4]* Adam Arkin,[5] Yizhi Cai,[6] Rob Carlson,[7] Aravinda Chakravarti,[8] Virginia W. Cornish,[9] Liam Holt,[10] Farren J. Isaacs,[11] Todd Kuiken,[12] Marc Lajoie,[13] Tracy Lessor,[14] Jeantine Lunshof,[15] Matthew T. Maurano,[16] Leslie A. Mitchell,[17] Jasper Rine,[18] Susan Rosser,[19] Neville E. Sanjana,[20] Pamela A. Silver,[21] David Valle,[22] Harris Wang,[23] Jeffrey C. Way,[24] Luhan Yang[25]

*These authors contributed equally to this work.
†Corresponding author. Email: jef.boeke@nyumc.org

Published 2 June 2016 on *Science* First Release
DOI: 10.1126/science.aaf6850

This PDF file includes:

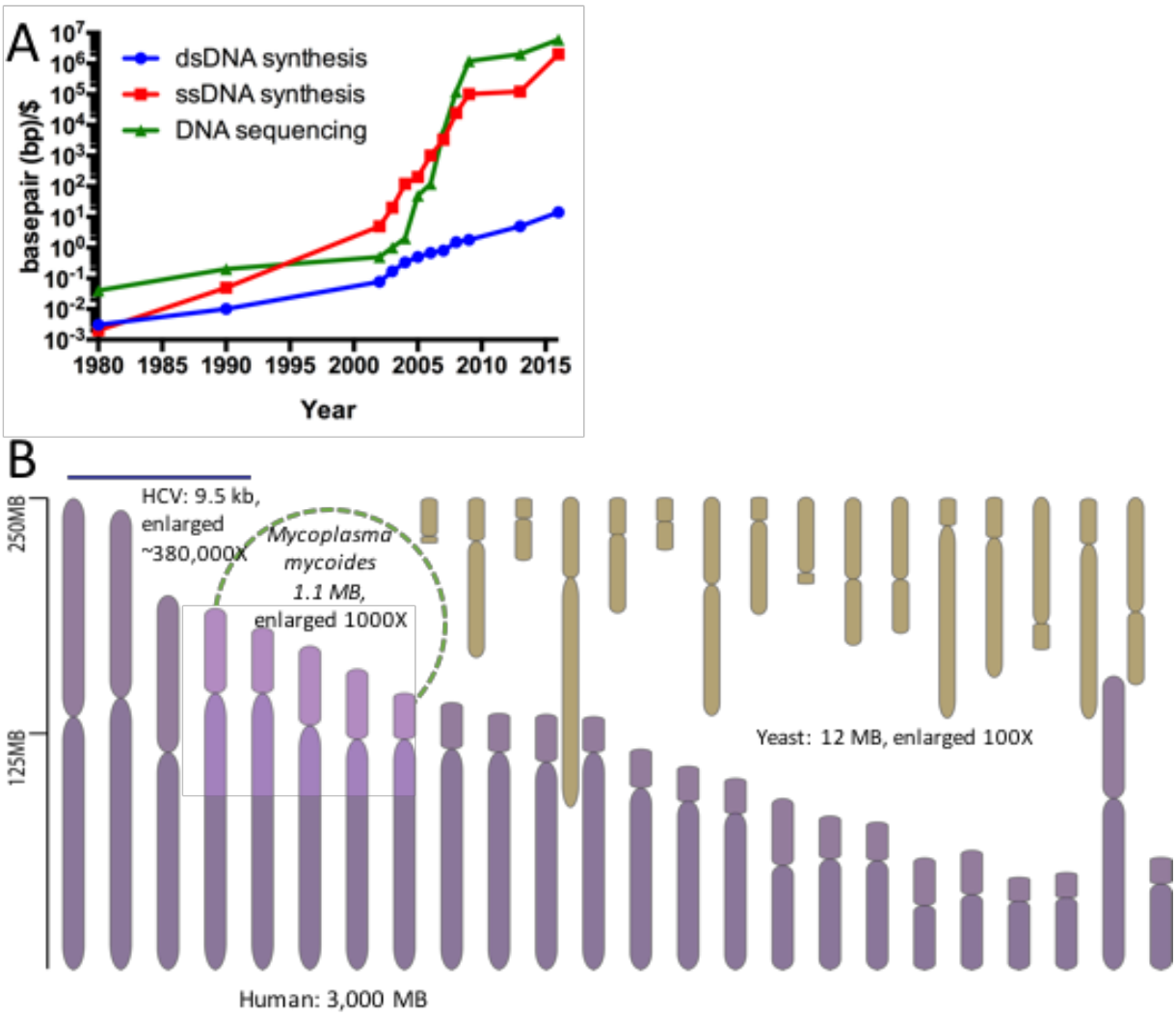

Fig. 1. Synthesizing synthetic or semisynthetic genomes. A. Efficiency trends in DNA sequencing (green) and synthesis of double-stranded DNA (dsDNA, blue) and single-stranded DNA (ssDNA, red) over the past ~35 years. Double-stranded DNA, or gene synthesis, has improved noticeably over the past ~10 years, but still lags behind sequencing and ssDNA synthesis. The disruptive improvement in sequencing and ssDNA (oligonucleotides) synthesis technologies has improved from multiplex and miniaturization technologies in high-throughput DNA sequencing and oligo microarray technologies, respectively. Commercial gene synthesis technologies relies on both oligo synthesis (building blocks) and sequencing (validation of synthesis) technologies. **B.** Graphical representation of four representative genomes benchmarked to the size of the 3,000 MB human chromosomes: 9.5 kb hepatitis C virus (HCV) enlarged ~380,000-fold, 1.1 MB Mycoplasma mycoides enlarged ~1,000-fold, 12 MB yeast enlarged 100-fold.

Bibliography

As further support for the arguments in our paper, this is a (non-comprehensive) sampling of precedents for projects that could take advantage of radical reduction in cost of genome-scale synthesis and high-throughput cellular/organismal testing of consequences. As with HGP-read, this effort need not be restricted to human but could and should include mouse, pig, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae*, etc.

The bibliography, along with proposals for pilot projects, maybe found online at the project web site www.hgpwrite.org

Author affiliations

[1]Institute for Systems Genetics at New York University Langone Medical Center, New York, NY, USA. Email: jef.boeke@nyumc.org

[2]Wyss Institute for Biologically Inspired Engineering at Harvard Medical School, Massachusetts Institute of Technology, Boston/Cambridge, MA, USA. Email: gmc@harvard.edu

[3]Autodesk Bio/Nano Research Group, San Rafael, CA, USA. Email: andrew.hessel@autodesk.com

[4]Nancy J Kelley & Associates. New York, NY, USA. Email: nancy@nancyjkelley.com

[5]Department of Bioengineering, University of California, Environmental Genomics and Systems Biology Division, E.O. Lawrence Berkeley National Laboratory, Berkeley, CA, USA. Email: aparkin@lbl.gov

[6]Edinburgh Genome Foundry, School of Biological Sciences at University of Edinburgh, Edinburgh, Scotland. Email: yizhi.cai@ed.ac.uk

[7]Bioeconomy Capital, Seattle, WA, USA. Email: rob@synthesis.cc

[8]McKusick-Nathans Institute of Genetic Medicine at Johns Hopkins University School of Medicine, Batimore, MD, USA. Email: achakra1@jhmi.edu

[9]Departments of Chemistry and Systems Biology at Columbia University, New York, NY, USA. Email: vc114@columbia.edu

[10]Institute for Systems Genetics at New York University Langone Medical Center, New York, NY, USA. Email: liam.holt@nyumc.org

[11]Department of Molecular, Cellular and Developmental Biology, Systems Biology Institute at Yale University, New Haven, CT, USA. Email: farren.isaacs@yale.edu

[12]Science and Technology Innovation Program at Woodrow Wilson International Center for Scholars, Washington, DC, USA. Email: todd.kuiken@wilsoncenter.org

[13]Department of Biochemistry at University of Washington, Seattle, WA, USA. Email: mlajoie@uw.edu

[14]Feinstein Kean Healthcare, Cambridge, MA, USA. Email: tracy.lessor@fkhealth.com

[15]Department of Genetics, Harvard Medical School, Boston, MA, USA; Department of Genetics, University of Groningen, Groningen, The Netherlands. Email: j.e.lunshof@umcg.nl

[16]Institute for Systems Genetics, NYU Langone Medical Center, New York, NY, USA. Email: maurano@nyu.edu

[17]Institute for Systems Genetics, NYU Langone Medical Center, New York, NY, USA. leslie. Email: mitchell@nyumc.org

[18]Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA, USA. Email: jrine@berkeley.edu

[19]Edinburgh Genome Foundry, School of Biological Sciences at University of Edinburgh, Edinburgh, Scotland. Email: yizhi.cai@ed.ac.uk

[20]New York Genome Center; Department of Biology, New York University, New York, NY, USA. Email: nsanjana@nygenome.org

[21]Department of Systems Biology, Wyss Institute of Biologically Inspired Engineering at Harvard Medical School, Boston, MA, USA. Email: pamela_silver@hms.harvard.edu.

[22]McKusick-Nathans Institute of Genetic Medicine at Johns Hopkins University School of Medicine, Baltimore, MD, USA. Email: dvalle@jhmi.edu.

[23]Department of Systems Biology, Department of Pathology and Cell Biology at Columbia University, College of Physicians and Surgeons, New York, NY, USA. Email: hw2429@columbia.edu

24Wyss Institute for Biologically Inspired Engineering at Harvard Medical School, Boston, MA, USA. Email: jeff.way@wyss.harvard.edu

[25]eGenesis, Inc., Boston, MA, USA. Email: luhan.yang@egenesisbio.com