GP-write: A Grand Challenge Using DNA Synthesis, Gene Editing and Other Technologies to Understand, Engineer and Test Living Systems

Meeting Summary

Jef Boeke, George Church, Andrew Hessel, and Nancy J Kelley

MAY 9-10, 2017
NEW YORK, NY

PREPARED BY TRACY LESSOR, PHD, MBA, ELS

NYU Langone MEDICAL CENTER

WYSS INSTITUTE

AUTODESK.

NANCY J KELLEY + ASSOCIATES

Table of Contents

Executive Summary

"What I cannot create I cannot understand" – Richard Feynman

In May, 2016, more than 130 scientists, industry leaders, ethicists and policy makers steeped in human biology, health and synthetic biology came together in Boston for an organizing meeting for GP-write. This meeting attracted a great deal of thought and interest, especially after the online publication of a Commentary about the project in *Science* on June 2, 2016. More than 195 news outlets worldwide covered the meeting and Commentary.

Since then, the GP-write Leadership Group has been hard at work responding to comments from the press and scientists around the world. In large part due to these conversations, the project's name was changed from HGP-write to GP-write and its scope expanded. The project is focused on using synthesis and genome editing technologies to understand, engineer and test living systems of model organisms, including the human genome, and plants in cell lines. The goal of GP-write is to not only deepen our understanding of life but to develop pragmatic technologies of general use in biology, improving the cost and quality of DNA synthesis, DNA assembly in cells, and testing of many DNA variations on tissue characteristics. The goals and expanded scope of this project is described in a revised White Paper, which can be found on the GP-write website at http://engineeringbiologycenter.org/.

The second GP-write annual meeting was held on May 9th and 10th, 2017, at the New York Genome Center in New York City. This two-day meeting explored the concrete steps that GP-write can take to solve some of the most important problems facing humanity, including how to move towards a sustainable, biological-based economy? How to further advance cures for disease? How to better understand biology processes? How to responsibly communicate the risks and benefits of this project to the world? The Agenda also included the introduction of new pilot projects, as well as the commencement of GP-write working groups to discuss project roadmaps, including scientific direction, technology development, ethical, social and legal engagement, standards and infrastructure development, amongst others. Although not every perspective on every topic discussed throughout the two days could be captured, the most common themes discussed are represented in this meeting summary.

There were more than 250 attendees from 10 countries at this oversubscribed meeting. The meeting participants were academic and industry scientists, ethicists, lawyers, educators, citizen scientists, artists, policymakers, technologists and lay people. For those unable to attend the meeting in person, the entire first session of the program, which provided a project overview and update, was live streamed and all sessions have been made available on the GP-write website. This meeting also attracted a great deal of interest from the press, many of whom were on-site throughout the meeting to attend a press briefing as well as a tour of the genome foundry at the Institute for Systems Genetics at NYU Langone Health. Representative coverage of the 2017 meeting can be found at http://engineeringbiologycenter.org/media/.

GP-write will build on the knowledge and technological advances of the Human Genome Project, and could be an equally transformative next step. Focused on writing, editing and building large genomes, the project will generate a wealth of information connecting the sequence of nucleotide bases in DNA with their physiological properties and functional

behaviors, enabling the development of safer, less costly and more effective therapeutics and a broad range of applications in other areas such as energy, agriculture, healthcare, chemicals and bioremediation. Another proposed benefit would be the commercial development of new genomics analysis, design, synthesis, assembly and testing technologies, with the goal of making these technologies affordable and widely available to everyone.

To date, nearly 200 scientists from over 100 institutions/companies in 14 countries have expressed interest in participating in GP-write, and several countries have expressed a willingness to provide financial support for the project. At the beginning of the year, Nature cited HGP-write as a "project to watch" in 2017.

We invite you to get involved in this project and to take part in the conversation.

Jef Boeke, Ph.D.
Director, Institute for Systems Genetics
Professor, Department of Biochemistry and Molecular Pharmacology
NYU Langone Medical Center

George Church, Ph.D.
Robert Winthrop Professor of Genetics,
Harvard Medical School
Core Faculty Member,
Wyss Institute for Biologically Inspired Engineering at Harvard University
Professor of Health Sciences and Technology,
Harvard and the Massachusetts Institute of Technology (MIT)
Associate Faculty Member,
Broad Institute

Andrew Hessel
Distinguished Researcher,
Bio/Nano Research Group
Autodesk

Nancy J Kelley, J.D., M.P.P.
President & CEO
Nancy J Kelley & Associates
Founding Executive Director,
New York Genome Center

Overview of Project: From Reading to Writing Genomes

The opening session provided an overview of Genome Project (GP)-write with a focus on the evolution of the project to date. GP-write is a grand challenge designed to support technology development in genome writing (i.e., genome design and synthesis) and evaluation of the impact of changes in genome design. The aim of this project is to reduce the cost of designing, synthesizing, assembling and testing genomes by 1,000-fold over the next 10 years. Such technology will revolutionize how we learn about the biological world – and how we engineer it.

GP-write is the natural evolution of GP-read, or what is commonly referred to as the Human Genome Project (HGP). GP-write will build on the knowledge and technological advances of GP-read by utilizing genome-engineering technologies to responsibly synthesize large genomes in cell lines, including human (HGP-write is an initiative within GP-write that focuses on the human genome). Responsible innovation has been at the forefront in the consideration of the design of GP-write, and will continue to remain at the forefront as the project plan advances. For example, GP-write will be limited to cell lines only, and will not involve germ line modification in any way.

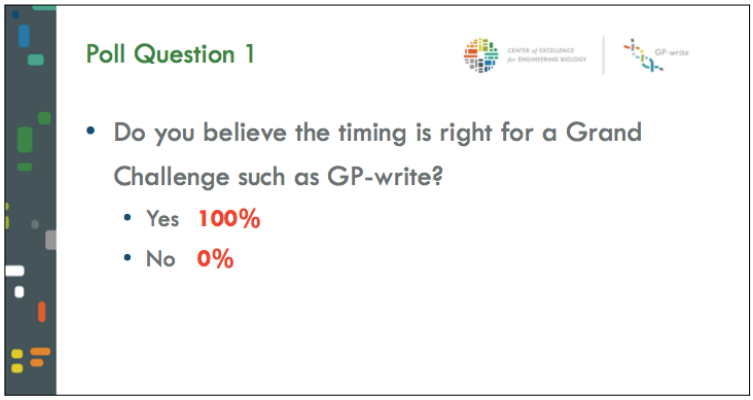
Over the past year, the scope of GP-write has been expanded beyond the human genome to understand, engineer and test living systems of model organisms and plants in cell lines. This session highlighted the benefits to studying other model systems to understand human biology and disease. For example, highly complex mouse models of human disease (Mouse Models 2.0) are being developed using the tools of CRISPR to create genetic diversity in the model systems, and studies of cephalopod genomes are providing fundamental insights into the nervous system of vertebrates. Designing plants for environmentally sustainable systems was also discussed. For example, plants have already been designed to detect explosive TNT at 23 parts per trillion. In the future, plants may also be designed to purify and filter seawater, providing unlimited water for sustainable life on earth.

The ability to efficiently write DNA could pull us forward into a better future, particularly with respect to human health challenges. Some potential applications include growing transplantable human organs, engineering universal T cells, and engineering virus, cancer and aging resistance into cell lines used for manufacturing and therapy. Additionally, DNA synthesis and assembly are foundational technologies, which could accelerate research and development across a broad spectrum of areas that threaten humanity today, including environmental destruction, invasive species, emerging pathogens, food security and climate change. All of these threats have potential biological solutions for which GP-write may play a role.

The first five years of the project will focus heavily on technology development as well as advancing

pilot projects that may have immediate benefits to society. During this time period, intense debate about the HGP-write component is expected, which will involve the engagement of a diverse group of stakeholders. A worthy design for HGP-write will also be identified. After this time period, additional worthy genome projects will be pursued together with ongoing technology development. As of May 2017, approximately $200 million in GP-write related funding has been made available across multiple institutions.

There are a few challenges that must be overcome in order for this project to be successful. First, there must be a reduction in the cost of DNA synthesis. The cost today is approximately 10 cents per base pair – this needs to be reduced by 1000x fold. Second, the DNA molecules need to be efficiently and accurately assembled into larger structures. Third, these larger DNA structures then need to be delivered into a cell, which is a significant bottleneck, particularly in mammalian systems. And finally, the cost of testing the phenotypes of these cells needs to be reduced.



Ethical, Legal, and Social Implications (ELSI)

A "March for Science" in America would have been unthinkable 50 years ago. Unfortunately, today there is a great deal of public mistrust in science and scientific institutions that needs to be mitigated. In the *Science* Commentary, various ELSI concepts were proposed, including inclusive decision-making, equitable distribution of benefits (e.g., distributive justice), appropriate regulation, global harmonization (e.g., stem cell research guidelines), and a better future (e.g., transplantable human organs, revolutionizing gene therapy, etc.). In this session, it was proposed that GP-write also address additional critical questions, including:

- What is a "human genome"? There is disagreement among scientists about whether a reference human genome exists because the published reference sequence has hundreds of gaps.
- What would be the unique benefits of synthesizing a human genome?
- Should priority be given to synthesizing less controversial genomes first?
- Does the cell line only policy apply only to the human genome?

It was proposed that the GP-write ELSI working group engage in a comprehensive review of all regulations, from the Genetic Information Nondiscrimination Act (GINA) to the Biological and Toxic Weapons Convention (BTWC), and examine the downstream threats from a lack of regulatory structure. Moreover, the working group should pledge to never put publications on ethical topics for the project behind a pay wall.

GP-write will advance the foundational tools and technologies for genome design, synthesis and testing that will have many potential future uses, many of which cannot even be imagined today. The ethical issues associated with this project should be addressed as they arise instead of debating whether scientific progress should be stopped because of what could potentially happen in 100 years. Transparency, responsible communication, and clear goals are needed in order to respond to ethical issues in a timely manner.

Everyone has a voice in setting bioethics, particularly in this era of citizen science, which heralds a new citizen-driven bioethics. If scientists want to gain public trust, then scientists must also trust the public with the assumption that the public can make some reasonable decisions about science when faced with the facts about the risks and benefits to society. Thus, public engagement in GP-write will require a new, 21st-century bioethics framework that fosters participation by the public.

Discussion of New Pilot Projects

Similar to other large-scale genome projects, including HGP-read, Encyclopedia of DNA Elements (ENCODE), and the Synthetic Yeast Project (Sc2.0), GP-write will be conducted in phases with explicit milestones, metrics, and assessments. Each of these earlier projects began with a pilot project that focused on a fraction of the genome, typically about 1%. For GP-write, the projects chosen will provide resources valuable for advanced biomedical research and/or biotechnology development.

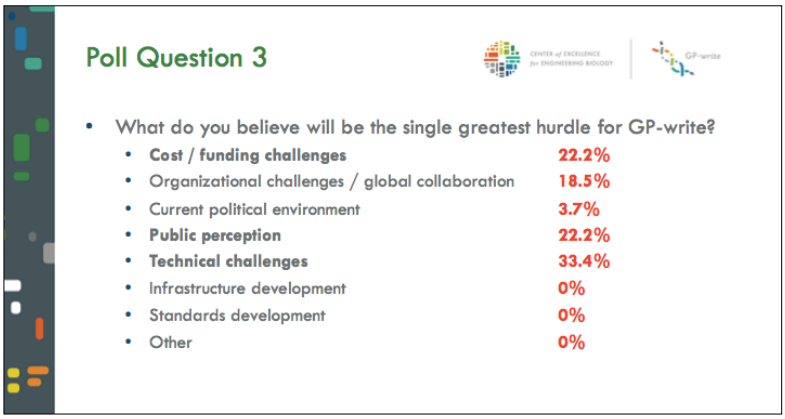
For the pilot projects, scientists – including citizen scientists – have been sending in their abstracts to be evaluated by the scientific leadership of GP-write. Those proposals that are deemed to have merit and have already been approved are briefly described below. Until the project is fully funded, GP-write is supporting the scientists submitting pilot projects with letters of recommendation and support. In the future, the goal is to financially support the pilot projects directly. There are two ways to apply for a pilot project: 1) Scientists can apply for funding through their university and GP-write will support the project, or 2) scientists can partner with the Center of Excellence to apply for funding.

Synthetic Screening for Essential Introns and Retroelements

In the pre-genome-synthesis era, due to their large size and repetitiveness, the physiological significance of introns and retroelements, respectively, could not have been systematically evaluated in the context of chromosomes within living systems. This pilot project proposes to use a synthetic genome approach to pursue the significance of introns and retroelements in three different models: human, mouse, and fly. This project could help to improve the sophistication of genome design by understanding the patterns of dispensable introns and retroelements, which could ultimately save money, time, and labor by downsizing animal genomes.

Precision Human Genome Engineering of Disease-Associated Noncoding Variants

Advances in genome sequencing have enabled rapid discovery of disease-associated gene variants, but the ability to perturb and interrogate genetic variants has lagged behind sequencing. New CRISPR/Cas-based genome engineering technologies have enabled complete gene knock-out in a rapid and efficient manner by taking advantage of the efficient non-homologous end-joining repair pathway to create frame-shift mutations that abolish protein production. However, it remains a challenge to make precise single nucleotide variants using homology-directed repair, which tends to occur at ~100-fold lower efficiencies. This pilot project proposes the development of novel selection techniques to enrich for homologous recombination repair in disease-relevant cell types, such as human pluripotent stem cells (hPSCs). Given the growing set of relevant CRISPR systems, bioinformatics tools will also be developed to help identify the best CRISPR system for introducing a particular mutation/edit. The goal is to create a complete pipeline for rapid engineering of disease-specific variants into human cells with significantly higher efficiency than previously possible.



Framing and Addressing the Ethical Issues Raised by GP-write

This session explored three real-world research scenarios – cerebral organoids, synthetic human entities with embryo-like features, and recoding – to highlight the critical need for a model that can deepen ethical analysis while keeping pace with the dynamics of science. This pilot project seeks to address, in real-time, the conceptual and normative issues at the cutting edge of synthetic biology. Additional goals of this pilot project include expanding collaborations in the humanities and sciences to ultimately advance both, as well as providing education and training.

Isothermal Amplification Array to Extend Synthetic Gene Sequences

At today's prices, a fully synthetic human genome would cost hundreds of millions of dollars. DNA synthesis is limited by traditional phosphoramidite chemistry, with base-by-base addition of nucleotides as the only real option for a synthetic gene sequence. While the price has dropped marginally to ~$0.10 per base pair, simply refining the process will not bring about the

exponential reduction in costs necessary to write an entire genome. Furthermore, bridging the gap between short synthetic fragments and entire chromosomes is an essential step in the GP-write project. This pilot project exploits isothermal amplification array and recombinase-mediated assembly as an alternative approach to DNA synthesis as it allows for extremely long sequences to be generated in an automated fashion.

Ultrasafe Cell Line

There is an unmet need for an "ultrasafe" human cell line designed to serve as a platform for many biomedical applications, from production of biologics, to modeling cell and tissue behaviors, to *ex vivo* and ultimately *in vivo* therapeutic applications. Such a cell line will be engineered to be ultrasafe from many distinct perspectives, including viral, prion and cancer resistance, as well as being engineered to minimize immune rejection, and to prevent germ line transmission. Moreover, these cells will have additional advantages such as biocontainment by adding new chemistry through the introduction of non-standard amino acids. This cell line will potentially be of great value to the pharmaceutical, vaccine, and biotechnology industries.

This session provided an update on a radically recoded *E.coli*, in which 62,000 positions of 7 codons are being replaced throughout the genome. The result of this extensive recoding is the creation of a much more virus-resistant bacterium than a previous version that is missing only a single stop codon. However, scaling up current synthesis, assembly and delivery technologies to create safe mammalian and human cell lines will be a much greater challenge.

Understanding and Anticipating Governance Systems

Governance systems, which seek to expose and address gaps in safety, are constantly challenged by new technology developments. Examples of successful governance in the biotechnology realm include the establishment of biosafety norms (BL3) with the emergence of DNA recombination technologies in the 1970's, and the Biological Weapons Convention, which is an international treaty against the development, production and stockpiling of bioweapons. This pilot project seeks to bring the GP-write community together to expose and address potential gaps in safety, and to ensure that this project remains in the public interest.

Synthetic Regulatory Genomics

An expansive atlas of the regulatory DNA landscape is now available. Some variants have assay-defined functions (promoters, enhancers, silencers, etc.), but most have complex activities that are still being sorted out. Mapping human disease- and trait-associated variation by genome-wide association studies (GWAS) has demonstrated that the majority of variants (95%) lie in non-coding regions. Thus, there is a need to study regulatory variation in context. This pilot project proposes the use of a synthetic approach to study regulatory genomics. The key advantages of a synthetic approach include multi-edited "synthetic haplotypes", cross-species

function, gene fusions, chromosomal rearrangements, position effect variegation, and gene therapy. A synthetic approach also allows flexible incorporation of alternate functional modules.

GP-write welcomes new proposals for pilot projects. Please download the Pilot Proposal form from the GP-write website at http://engineeringbiologycenter.org/wp-content/uploads/2017/10/GP-write-Pilot-Project-Proposal-Form.pdf.

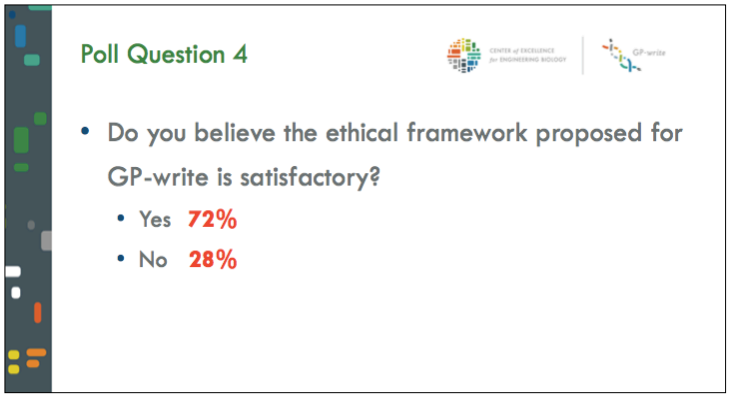
Required Elements of an Ethical, Social, and Legal Roadmap for GP-write

GP-write requires significant consideration of the social, ethical, and legal implications of the project. Specifically, responsible innovation in the form of deep conversations with diverse stakeholders – including the public – will be key to identifying common goals of importance to scientists and the larger community. During this panel discussion, the concern was raised that bioethics is frequently presented as a serious, philosophical, and weighty topic, which could deter the general public from engaging in these important discussions. Therefore, new approaches may be needed to engage the public in the issues that could ultimately impact them.

The question was raised about how to establish ethical principles for the project and how best to make this process transparent. It is important for the GP-write community to be clear about the purpose of the project, the benefits and risks to society, and the defined boundaries (e.g., no embryos will be created). Although these conversations should begin now, there is lead-time to address ethical issues both during and after the project's completion. Moreover, it is important to not generalize bioethics; instead, specific focus should be placed on the issues associated with each pilot project separately.

GP-write Related Research Areas

A series of topics with far-ranging fields of relevance to GP-write were covered, including microorganism genomes, microbiome engineering, engineering intracellular communication and genome engineering in mammalian systems.



Microorganism Genomes

This session covered scientific talks related to the design and construction of synthetic yeast and bacteria genomes. The goal of the Synthetic Yeast Genome Project (Sc2.0) is to design, build, assemble, and test the function of an entirely synthetic designer yeast genome. The Sc2.0 International Consortium hopes to have the entire genome – all 16 chromosomes – completed before the end of 2017. Presentations covered the design of tRNA neochromosomes, the power of synthetic chromosome rearrangement and modification of loxP-mediated evolution (SCRaMbLE) to accelerate genome evolution, key issues in the application of designer cells, and cell-free cloning of large circular DNA.

Microbiome Engineering

The human body is shared with trillions of commensal microbes that are involved in development, immunity, digestion, and mood/behavior. An improved ability to engineer the microbiome could open many opportunities to advance human health and disease treatment. This session highlighted recent developments in microbiome engineering, including efforts to recode the Salmonella genome, engineer the microbiome and mammalian genome with enhanced metabolic functions, and manipulate the skin microbiome for therapeutic development.

Engineering Intercellular Communication and Other Complex Systems

Engineering biology is rapidly progressing the development of synthetic intercellular communication, aiding the formation of higher-order, complex regulatory networks that will greatly advance the field. This session highlighted current efforts to design and build a scalable yeast intercellular communication system, engineer programmable cell-based therapies, and human to yeast pathway transplantation.

Genome Engineering in Mammalian Systems

Numerous genome engineering developments are being made in mammalian systems that will greatly aid the efforts of GP-write. This session covered some of these developments, including efforts to assemble human centromeres *in silico*, developing a human artificial chromosome (HAC) that efficiently shuttles from yeast/bacteria and establishes itself in human cells, overcoming challenges associated with rewriting pig genomes for xenotransplantation, and creating haploid human embryonic stem cells to explore genome functionality in virtually all cell types.

Center of Excellence for Engineering Biology

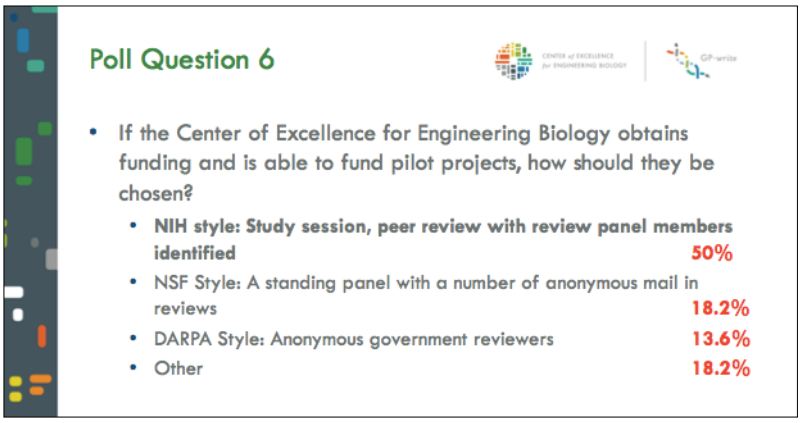
Over the past year, the GP-write Leadership Team has been working on the general governance structure for the Center of Excellence for Engineering Biology (CEEB). It was decided that five governance committees will be created, each with a Charter:

- Science Executive Committee
  - Co-chaired by Jef Boeke and George Church
- International Committee
- Clinical Advisory Board
- Industry Advisory Board
  - Chaired by Andrew Hessel
- Social, Ethical, and Legal Advisory Board
  - Co-chaired by Jonathan Moreno, and Barbara Evans

The scientific executive committee will be responsible for creating the scientific roadmap for the project, for evaluating, selecting and overseeing the pilot projects, and addressing any issues of safety and compliance as they arise. Supporting the work of these five governance committees are seven working groups:

- Technology Development
- Infrastructure Development
- Safety Engineering
- Standards, Quality Control, and Reporting
- Intellectual Property
- Public Communications, Outreach and Education
- Policy Development

Each working group has co-chairs and 6-10 members who will meet monthly to discuss roadmaps and provide progress updates. The working groups kicked off at the GP-write meeting in May 2017.



GP-write will have an organizational document (similar to the Statement of Ethics and Governance used in Sc2.0) that will set forth a Statement of Principles and Governance, set forth an ELSI white paper that will delineate safety/standards and address public outreach, and set forth an approach/agreement for sharing IP, data, and innovation. All participants in GP-write must sign and adhere to the principles and terms of this Memorandum of Understanding. (MOU).

With respect to funding, discussions are underway with philanthropic and public funders for large-scale project funding. In a Center-based approach, CEEB will help scientists to apply for grants in partnership with the Center (and will be entitled to a small percentage of funds). Alternatively, CEEB can support grant applications generated through a University/Institution.

The National Institutes of Health (NIH) invested $4 billion in the Human Genome Project (GP-read), which unleashed approximately $1 trillion in economic activity. A Grand Challenge such as GP-write is also expected to have a significant positive economic impact.

Pilot Project Pitches

Five-minute "elevator pitch" speaking opportunities were given by scientists to present new ideas for GP-write pilot projects selected from abstracts submitted prior to the meeting. These proposed projects are currently under consideration by the Science Executive Committee.

Long And Precise Genomic DNA Construction Using Bacillus Subtilis

Saccharomyces cerevisiae is almost always used as an assembly host for bacterial genome-sized (re)construction due to ease of assembly for long DNA strands. Despite its popularity, the yeast assembly system has several constraints related to GC-content and/or replication origin, unexpected mutations associated with direct use of PCR-amplified DNA as a material, etc. These constraints may be overcome with Ordered Gene Assembly in *B. subtilis* (OGAB). OGAB is an efficient DNA assembly method that can assemble more than 50 DNA fragments, including long fragments (~100 kb), in one-step using plasmid transformation. This method requires high skill in precise control of molar concentration of material DNAs. However, this pilot project proposes to overcome this limitation through the development of an automated, liquid handling robot to make the system user-friendly. The system developed in this pilot project could enable the construction of long (~100 kb) and precise genomic DNA for GP-write.

Rearrangement and Reduction at Human Genomic Loci

The human genome contains several regions of importance where groups of genes with similar functions are clustered together along with sequences and motifs that are structurally unstable. As the chromosomes replicate and divide in mitosis and meiosis these areas are hotspots for genotypic diversity, where genes and other DNA sequences can be deleted, duplicated or rearranged. Previous work on Sc2.0 and its inducible SCRaMbLE system for genome rearrangement has demonstrated the power of genomic rearrangement in eukaryotes. Dramatic new phenotypes can be generated in hours, and new information about gene content and the roles of genomic sequence can be combinatorially determined, especially now with routine long-read DNA sequencing. As the Sc2.0 project is scaled to the human genome, the SCRaMbLE tools can be applied to human chromosomal regions that naturally recombine and rearrange in order to further understand the role of specific sequences within these loci. This pilot project proposes to test megabase-scale construction and design in the human genome and investigate the utility of SCRaMbLE and genome refactoring in medically relevant target loci.

Rewriting Regulation? A Comparative Policy Study of Natural vs. "Synthetic Cells", Organoids and Human Genomes

Around the world, policies governing diverse scientific fields extend across a continuum that differentiates between degrees of regulatory control (i.e. restrictive vs. liberal) and

harmonization. This pilot project proposes to conduct an international comparative policy study of the core overarching scientific human applications GP-write will focus on. The goal is to refine the current knowledge and understanding of the policy and regulatory frameworks applicable to GP-write. In particular, this project will identify, describe and analyze national policy approaches to the regulation of 'natural' versus engineered/synthetic human cells and organoids, which is a fertile area of study. By mapping the policy landscapes, it will explore their strengths and challenges, their ability to adapt to scientific advances, evolving technological uptake, as well as social interests. It will further assist in determining whether the issues arising in GP-write fall within, or outside, the remit of existing policy dealing with interconnected scientific fields. This work would help elucidate the consistency of the ethical principles, social values and scientific rationale underlying the policy choices, and would set the groundwork for a more comprehensive international policy 'roadmap' and analysis. This project would also complement the activities conducted by the ELSI working group.

Stable Haploid Human Pluripotent Stem Cells

In yeast, and more recently in mammalian cells, haploid cells have provided an effective tool to evaluate gene function because the phenotype provides a functional readout of the single allele present in the cells. By contrast, in diploid cells the change or deletion of one copy often has no functional consequences because of allelic complementation. Haploid human pluripotent stem cells have two key advantages relevant to GP-write: they are capable of differentiation into any cell type, and they are human, providing a tool to functionally interrogate the human genome. However, current haploid cell cultures have a significant disadvantage: they diploidize and require repeated sorting for DNA content. This pilot project proposes to determine the mechanism of diploidization in haploid human pluripotent stem cells in order to understand how to stabilize these cells. Stable haploid human pluripotent stem cells would be exceedingly useful for GP-write, as they could reduce the workload by half.

Development of a Pipeline for Precision Cloning and Effective Storage of Large Synthetic or Natural Human DNA Fragments

One of the main objectives of GP-Write is to develop economical methods to synthesize and store large DNA fragments. It is evident that yeast is an excellent host for cloning and storing DNA fragments up to few megabase pairs. However, the process requires further optimization as it is still very labor-intensive and requires personnel with significant experience. This pilot project proposes the development of a pipeline for the efficient cloning and manipulation of large DNA fragments in yeast, *Saccharomyces cerevisiae*. Specifically, this project would create novel molecular tools to hook and capture large targeted DNA fragments from a vast undifferentiated pool of DNA. The maximum size of DNA that can be cloned and stably propagated on a centromeric plasmid vehicle in *S. cerevisiae* will also be established. The resulting platform for capturing natural and synthetic DNA fragments could aid companies and initiatives that are developing novel methods for synthesizing large DNA fragments.

Genome-scale Design Representation with SBOL

The Synthetic Biology Open Language (SBOL) is a key technology for supporting emerging biological engineering workflows. SBOL is a free and open community standard for the description and exchange of biological designs, supported by a diverse international community of researchers. The SBOL community has already developed many of the representational and software capabilities necessary to support a large-scale construction effort such as GP-write.

This pilot will study the application of these capabilities to key representation, coordination, and integration problems inherent in the GP-write vision. The products of this study will be: 1) a collection of representational examples, workflows, and best practice recommendations addressing key representation, coordination, and integration problems in GP-write; and 2) extension of the existing standards as needed in support.

Winning the Fight Against Diabetes: Singapore to GP-write

There is a global urgent need to better understand Type 2 diabetes (T2D) in a bid to improve treatments for this disorder. In Singapore alone, the prevalence of T2D is forecasted to double in by 2050 and to inflict 15% of the population. Moreover, the lifetime risk of the population is expected to reach approximately 50% by 2050. Being ethnically diverse, Singapore's public health trends are representative of Asia, thus the disconcerting growth in prevalence of T2D may be imminent for other Asian countries. The goal of this pilot project is to accelerate our understanding of T2D through GP-write by constructing cell and animal models to investigate T2D-associated genes. These models would enable the study of genetic factors that render individuals predisposed to T2D and investigate variations at transcriptional and translational levels that cause anomalies, such as impaired glucose transport and suppressed insulin response, leading to T2D.

An In-Situ Digital Annotation System to Document and Safeguard GP-write Applications

GP-write aims for the widespread and routine application of synthetic DNA applications in disparate fields such as clinical medicine, agriculture, biotech, and data storage. The future genome engineer will work similar to a computer programmer, where the 'commenting' of code and 'code signage' are standard quality assuring procedures. It is now standard to keep developing computer code in GIT repositories (sometimes referred to as 'global information trackers'). Further, software based products often have 'watermarks' that allow for the tracking of copies, authentication of originality, and enforcement of access and copy rights. Will DNA engineering adopt similar procedures? Is there a role for DNA watermarks in GP-write products? This pilot project proposes to outline an innovative platform that will allow for a digital global annotation system to document and safeguard GP-write applications. In addition, it will contribute to identifying the regulatory issues and challenges that would need to be resolved in order to implement such a global annotation system. Developing early an open standard to track deeply annotated GP-write products with unique 'in situ' signature barcodes will secure public trust, improve quality, and ensure long-term traceability of engineered DNA independent of the originating laboratory.

Installing "Mini-Symplastomes" in Crops

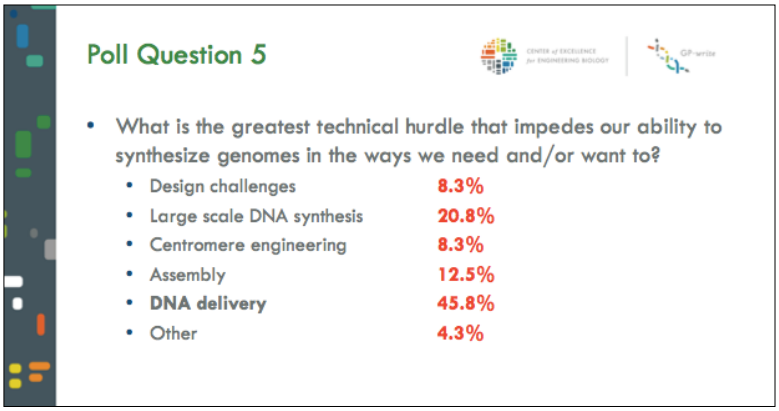
There is currently an effort to design, synthesize, and install the first synthetic chloroplast genome "synplastome^{TM}" in tobacco (funded by ARPA-E) and potato (funded by DARPA). These

species have plastomes of approximately 150 kb in multiple copies that reside in each of the dozens of chloroplasts and other plastids within each plant cell. The synplastome concept is a compelling entry point into synthetic genome design and implementation in plants. The plastome is relatively compact and plastomes have conserved structures among individuals and even plant species. However, the community is facing challenges with the *in vitro* assembly of the synplastome. This pilot project proposes to resynthesize and introduce "mini-symplastomes" as a workaround to the assembly and insertion of a single 150 kb molecule into the chloroplast. The mini-synplastome concept would use fluorescent proteins and selectable markers on each of the three-to-four mini-symplastomes that would be introduced together into chloroplasts in attempt to replace the endogenous plastome via *in vitro* antibiotic selection. Additional funding for DNA synthesis and potential collaboration with other eukaryotic synthetic genomics experts could make a significant difference in the success of introducing symplastomes into crops.

Scientific Talks on Technology Development and Technical Challenges

The development and use of large-scale DNA synthesis and genome engineering technologies will be integral to the success of GP-write. This session highlighted industrial and academic efforts to advance genome-engineering technologies and overcome their many challenges.

Genome Engineering Technologies Utilized for Large-Scale Editing



Similar to editing a Word document, sometimes only small modifications to a genome are required; at other times, entire sections need to be re-written. This session highlighted recent developments in the technologies utilized for large-scale genome editing. A particular focus was placed on DNA target recognition systems, making programmable organoids a clinical reality, and overcoming the challenges of delivering large DNA fragments to mammalian cells, including intact transfer, nuclear delivery, and site-specific integration.

Engineering from Synthesized DNA

Synthetic DNA holds the potential to solve many of humanity's grand challenges, including environmental sustainability, hunger, and disease. Phosphoramidite chemistry has reached a plateau in technology development and the next generation of DNA synthesis technologies is rapidly underway in both academic and industry laboratories. This session provided updates on DNA synthesis using error-free oligonucleotides retrieved from next-generation sequencing flow cells, terminal deoxynucleotidyl transferase to make very long single-stranded DNA molecules, a 3D microchip DNA-writing platform, the use of silicon to deliver game-changing throughput, speed, cost and quality, as well as a platform for the rapid assembly and transfer of large DNA constructs (SynGen™).

Software Development Tools

This session covered the latest in software development tools that are advancing the field of engineering biology. Focus was placed on making digital biology a reality, overcoming the challenges in genome design through lessons learned from Sc2.0, and an update on the Parametric Human Project, which aims to be the core for all human data to support organization and correlation efforts enabling eScience.

Infrastructure Development

Infrastructure development is a critical component of GP-write. A series of presentations were given on genome engineering foundries and strain development technologies in industry.

Genome Engineering Foundries

Genome engineering foundries seek to streamline and standardize engineering biology in order to make the process easier, faster and scalable. The latest developments in software, automation and analytics for the rapid design, build and testing of engineered organisms were presented by four foundries, including the Edinburgh Genome Foundry, Genome Foundry at the Institute for Systems Genetics at NYU, MIT-Broad Foundry, and the London DNA Foundry.

Strain Engineering Technologies in Industry

Various topics were covered in this session, including the latest trends in writing DNA, such as DNA printing, single cell writing with trackable multiplex editing, overcoming the challenge of engineering microbes using multiplex assisted genome engineering (MAGE), which enables the building and screening of billions of combinatorial genomic designs in weeks, and a discussion of the many applications of liquid handling technology for engineering biology.

High Performance Computing and Computational Biology

An exponential drop in the cost of DNA sequencing is accompanied with an exponential increase in the amount of data available for interpretation. This session provided an update on the latest advancements in high performance computing with a focus on scaling computation to keep pace with data generation, technological approaches that enable communities of researchers to work across institutional boundaries, and participatory models of biomedical research and innovation.

Working Group Roadmaps

A series of working groups have been formed, the goal of which is to create roadmaps for a number of important topics for GP-write. Each working group has co-chairs and 6-10 members who will meet monthly, beginning in May 2017. Each working group met at the May meeting and provided their first status update.

Social, Legal and Ethical Issues

The ELSI working group saluted the efforts of GP-write for incorporating ethics at the very earliest stages of project development. One of the complexities of early involvement, however, is concretely defining the ethical issues when the scope of the science is still being set.

One of the needs identified by members of the working group is to have greater representation by members of the public as well as greater international representation because every nation has their own distinct ethics community. The ELSI working group will also need to closely interface with the other groups working in parallel because it is important that the group not be siloed.

In terms of next steps, the ELSI working group will develop a whitepaper that will identify the ethics work done to date for other relevant projects such as GP-read, what gaps need to be addressed, and what the ethical agenda should be for GP-write moving forward.

Technology and Infrastructure Development

There was agreement among the working group that GP-write will need a database/central repository (oligonucleotides, constructs up to very large size, engineered DNA in cells, methods, designs) to avoid costly repetition of synthesis and to promote sharing. Early establishment of a universally established coordinate system will be required (once the target is agreed upon) along with extensive annotation. This comes with a need to develop technology to support the archiving of large constructs, as well as for standards for representing the contents of the collection.

In terms of next steps, a whitepaper will be developed that surveys the community for speculative DNA synthesis technologies for long-term development projects. In addition, a call for proposals/pilot projects meeting certain DNA synthesis goals will be submitted to the scientific community.

High Performance Computing and Bioinformatics

High performance computing and bioinformatics are extremely important for GP-write. There was agreement among members of the working group that focus should be placed on five major areas: identity and access management; privacy and information security; permission and consent management; data storage and architecture; and, workflows, automation, and reproducibility.

Identity and access management: Any global collaboration will require significant thought on how researchers, stakeholders, and research participants identify themselves to the project. It is expected that a system of federated identity will be used, and a layered security model will be

supported so that the appropriate level of protection can be applied to each portion of the project.

Privacy and Information Security: Information security will be addressed in terms of both policy and practice. From a policy perspective, strong standards will be defined around the regulatory frameworks to be followed. In terms of practice, a modular, layered approach will be taken to security, building protections in at an architectural rather than at some imaginary system perimeter.

Permission and consent management: The effort will certainly include personally identifiable information on research participants. In order to create globally useful datasets, local laws and standards must be respected around data privacy, while at the same time integrating ever-larger sets of information. This will require a system of machine-readable consent, as well as the ability to re-contact and update participants while still respecting privacy and anonymity.

Data storage and architecture: In order to be useful, data must be both accessible and also well organized. A significant effort is anticipated to organize and present the project's metadata in a comprehensive and scientifically correct framework. This effort will involve data modeling by domain experts, including agreement on ontologies to allow data to be compared and integrated between researcher groups.

Workflows, automation, and reproducibility: A large part of the consortium's focus is on scaling academic or research pipelines to industrial capacities. This will require automation and process engineering. As part of this, modern practices will be leveraged around containerized workflows (Docker) and automated, data driven analysis (Lambda architectures).

Safety Engineering

Safeguarding biology is important for realizing the full potential of GP-write. The case for engineering safeguards is strong and includes, but is not limited to, protecting against the accidental release of genetically-modified organisms (GMOs) into the environment; protecting scientific investigators; preventing industrial espionage and economic threats to national security; and enabling the safe use of GMOs in open systems (e.g., probiotics, bioremediation).

For the past 40 years, society has benefited from the many advancements that have been made in biotechnology. This research has mostly been conducted in closed systems, but now these systems could eventually get into the environment either intentionally or otherwise. Therefore, the early safeguards that have been developing in this area over the past four decades need to be advanced and integrated into the systems being worked on today. Examples include gene drives, kill switches, unnatural amino acids, conjugation resistance, alternative DNA backbones, biocontainment, and genetic isolation.

Standards, Quality Control and Reporting

The development of standards enables data sharing across organizational boundaries, including laboratories, institutions, and companies. Standards also enable comparability across organizations, facilitating reproducibility and interoperability. Moreover, the use of standards supports confidence in experimental results and they play a deep role in commercialization. Given the importance of standards development, there was general agreement among the working group that enduring standards become a valuable spin-off from GP-write. Otherwise,

standards development will not happen nearly as well – or as soon – without the project. In terms of next steps, the working group will develop a framework that will identify the specific problems that standards can solve for GP-write.

Intellectual Property

With respect to intellectual property (IP) versus open source, the community will have to decide on the best model for the project. Open source as a development model may be the best approach at the early stages of development so that everyone can freely innovate, but the community is also responsible for translating these discoveries into commercial products with the freedom to operate. Finding the right balance/approach for GP-write is one of the goals of this working group, which is actively recruiting for open-source experts to serve on the committee. In terms of next steps, the working group will develop a framework for evaluating intellectual property and finding the right balance of IP versus open source.

Public Communications and Outreach

Ongoing communications with all stakeholders, particularly the general public, will shape the perception – and success – of GP-write. The vision of this working group is to get the world as excited about GP-write as they were about getting a man on the moon. The goal is to ensure that all interested constituencies – including scientists, laymen, policymakers, funders, media, et al. – can access accurate, timely, transparent, and understandable information about GP-write and its activities, and can provide input and feedback to the GP-write leadership group on an ongoing basis.

Education

The goal of education is to teach in an accessible and inspiring manner so people want to learn more, and teaching is most effective through open questions. The working group identified a few questions that will need to be addressed as the roadmap is developed:

- What needs to be taught? What is the big message? Is it the idea that a genome can be built, and not just edited?
- How should the community/demographic for education be prioritized (available funding, ease, impact)?
- What can be learned from GP-read?
- How should this working group integrate with others on a coherent message?

This working group would like to recruit additional members of the community to join, particularly STEM educations at college and high school level, informal educators, educational video experts, social media experts, and scientists.

bioRxiv

Launched in 2013 as a non-profit service out of Cold Spring Harbor Laboratory and now funded by the Chan Zuckerberg Initiative, bioRxiv is a preprint server that allows for scientific results to be shared before the papers are sent to a journal for peer-review. Both submission and access are free.

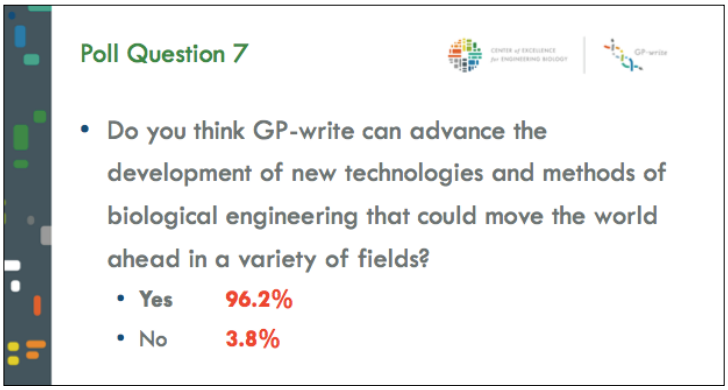
bioRxiv is a hub for communication, which is why a GP-write channel has been proposed. The benefits include rapid transmission of results, pre-publication feedback/discussion before

submission to a journal, visibility for early-career scientists, and immediate availability to grant/hiring committees. To date, more than 11,000 submissions have been approved and more than 400 journals have published papers first posted on bioRxiv. But behavioral, policy and rule changes are also being made. The number of biologists posting and reading preprints is growing daily, most journals are allowing preprint posting, and funding bodies are allowing preprint citations in grants and/or are mandating preprint posting.

Next Steps

In continuing efforts to make GP-write projects more open, accessible and understandable, all communications from the leadership team, working groups and pilot projects will be shared through the GP-write website at http://engineeringbiologycenter.org/. Please check back frequently as updates will be made continuously throughout the year.

Discussions are ongoing regarding holding two meetings per year: a smaller scientific meeting in October to advance pilot projects and working groups, and a larger annual meeting in May (such as this one) that provides an annual update on all activities for the larger community. Please visit the website for all meeting updates.



In the meantime, everyone is invited to become involved in the project and have his or her voice be heard. There are multiple ways to become involved in HGP-write:

- Suggest a new 1% pilot project
- Join the conversation via social media
- Join a working group
- Fund the project

We would like to thank our Partners and Sponsors for their support.

Partners

AUTODESK.

NANCY J KELLEY + ASSOCIATES

Feinstein Kean Healthcare
An Ogilvy Company

WILMERHALE

LABCYTE

Sponsors

HUNTER ROBERTS CONSTRUCTION GROUP

JB&B JAROS BAUM & BOLLES

milrose consultants, inc.

Modern Meadow

Newmark Grubb Knight Frank