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

November 30, 2016

Author: Jef D Boeke, George Church, Andrew Hessel, Nancy J Kelley (GP-write Leadership Group) and The GP-Write Consortium†
†The authors and their affiliations maybe found at www.gpwrite.org

INTRODUCTION

- The Human Genome Project (HGP-read) was initiated in December 1984 at the DOE Alta meeting and launched on October 1, 1990. Aimed at a nearly complete "reading" of the human and other genomes with annotations of plausible gene functions, as well as the improvement of the technology, cost, and quality of DNA sequencing (*1, 2*), it was the largest life science project ever conducted and one of the great feats of exploration in history.

- The first draft of the human genome was announced in June 2000, and a more complete version in April 2003.

- Since this announcement, interpretation of that data was made possible by sequencing humans from other populations around the world (HapMap and 1KG) and by other "omics" tools yielding insights into the functionality at base pair precision (ENCODE), correlations between natural genome variants and various traits and diseases via genome-wide association studies (GWAS), and identification of genes and variants responsible for rare Mendelian traits by a combination of genomic sequencing and advanced computational methods.

- Incredibly powerful and inexpensive DNA sequencing and genetic analysis technologies, developed, in part, by the National Human Genome Research Institute's (NGHRI) Advanced DNA Sequencing Technology Development program, reduced the cost of sequencing a human genome more than a million fold from $3 billion to less than $1,000 in just over a decade.

- Since then, the world has embarked on a revolution in science and healthcare that is changing the way we live and carries the promise of individualizing clinical delivery to improve health, and prevent and cure human disease. The Precision Medicine Initiative (for example), announced by the White House in 2015, would use the insights gained from "reading" each individual's genome to personalize medical care based on an individual's genes,

environment and lifestyle in order to keep that individual healthier longer and target proper treatments for disease when it does occur.

- Francis Collins, the leader of HGP-read, described the working draft of the human genome as "the first glimpse of our own instruction book." Today, many scientists believe that the way to truly understand that instruction book requires us to "write" DNA and build large Giga-base (Gb) animal and plant genomes, including the human genome, using the four basic building blocks of DNA, the nucleotide bases that encode biological functions. If successful, such an effort would result in and demonstrate an understanding of the fundamental biological processes of the natural world that could be used to solve important global challenges.

- Although sequencing, analyzing, and editing DNA continues to advance at breakneck speed, our capability to "write" DNA sequences in cells is mostly limited to a small number of short segments, restricting our ability to manipulate and understand biological systems.

GP-write: A Grand Challenge

- The Genome Project-write (GP-write) will use synthesis and genome editng technologies to understand, engineer and test living systems. It will build on the knowledge and technological advances of HGP-read, and could be an equally transformative next step.

- Focused on writing and building variations on large Giga-base (Gb) animal and plant genomes, including the human genome, the project would generate a wealth of information connecting the sequence of nucleotides in DNA with their physiological properties and functional behaviors, and would drive the development of tools and methods that facilitate large scale synthesis and editing of genomes. Furthermore, because DNA synthesis, like sequencing and computation, is a foundational technology, GP-write could accelerate research and development (R&D) across a broad spectrum of life sciences, supporting basic R&D of new bio-based therapies, vaccines, materials, energy sources, disease vector control, and nutrition.

- Another proposed benefit could be the commercial development of new genomics analysis, design, synthesis, assembly and phenotyping technologies. Overall the project would develop enabling tools of broad applicability throughout biomedical research, such as:

1) Computational tools, which allow the redesign of any genome, followed by compilation and testing of the redesigned code *in silico* before hitting the print button,

2) Phenotypic screening platforms such as low-cost organoid cultures, which allow characterization of performance of synthetic DNA and variants of unknown significance,

3) Cheaper, more accurate and longer DNA synthesis and assembly for pathways and whole genomes (*3*)

4) Improved genome editing techniques and recombination methods, and

5) Targeted delivery to specific cell types or systemically throughout multiple organ systems.

DEFINITION OF GP-write AND HGP-write

- GP-write will be an international scientific research project, led by a multi-disciplinary group of US-based, global scientific leaders who are highly credible in genetics, genomics and engineering biology.

- The Human Genome Project-write (HGP-write) will be a critical core activity within GP-write focused on synthesizing human genomes in whole or in part. Because of the special challenges surrounding human genomes, this activity will include an expanded ELSI component. It will also be explicitly limited to work in cells, and organoids derived from them only.

- The primary goal of GP-write is to reduce the costs of engineering and testing large (0.1 to 100 billion bp) genomes in cell lines by more than 1,000-fold within 10 years.

- An additional goal is to synthesize variations on complete human (and other) genomes in order to:

  (1) understand the functional properties and phenotypic consequences of the genome sequences provided by HGP-read,

  (2) transform the quality of DNA tools, assembly methods, automated infrastructure, artificial intelligence, standards and data management systems,

(3) massively reduce the cost of writing and editing new genomes and creating DNA at scale,

(4) disseminate broadly information and knowledge generated by GP-write through publicly available databases on the Internet to promote rapid application of research results,

(5) drive development and commercialization of new related technologies, and

(6) address the ethical, legal and social issues that arise from the project.

- Related goals would be training the next generation of scientific leaders, public education and engagement, technology development and transfer.

The Basic Primer

- To understand how this would be accomplished, consider that the genetic instructions in DNA are found in the nucleotides: adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T), whose sequence holds the biological information for functions within an organism. In humans, this sequence consists of two near-identical sets of 3 billion nucleotide pairs.

- Just as software programmers use infinite permutations of zeroes and ones to design and write software code that drive all functions of technology, so scientists can write computer programs using these nucleotides to design, re-write the genetic code for and build entirely new biological systems from the ground up.

- The actual DNA is produced using DNA synthesis, a technology improved during the past few decades, which makes it possible to assemble a synthetic DNA molecule *de novo* with no limits on sequence or size. Using these techniques of engineering biology, scientists are redesigning genes to improve gene expression, produce new regulatory structures, metabolic pathways and even chromosomes (*4*).

Activity to Date

- Small viral and bacterial genomes synthesized from scratch and organisms with recoded genomes derived from large-scale genome editing have demonstrated the feasibility and utility of synthetic genomes.

- The first whole genomes to be created using the techniques of engineering biology were those of viruses, beginning with the poliovirus in 2002 (*5*). Since 2003, only a few other virus and transposon genomes have been synthesized in their entirety. In 2010, a bacterial genome (1.05 Mbp, Mycoplasma) was synthesized as a nearly identical copy of the natural genome (*6*). The most radical departure from a natural bacterial genome was the 2013 alteration of the genetic code throughout the ~4.7 Mbp *E. coli* genome (*7, 8*). This broadly enabled virus resistance (*9*) and in 2015 was extended to biocontainment strains, also known as rE.coli 2.0, fully dependent on a 21st amino acid (*10, 11*).

- The largest genome synthesis project currently underway is an international effort to write the 12Mb genome of *Saccharomyces cerevisiae* (baker's yeast) led by Dr. Jef Boeke (NYU) (Synthetic Yeast Project) (*12, 13*). Dubbed Sc2.0, the project has participation from groups in the US, the UK, Australia, France, Germany, Singapore and China. The consortium has successfully assembled a synthetic chromosome (chromosome 3) and shown it to be functionally equivalent to the natural chromosome, but it has new features that native chromosomes lack, such as genome scrambling, a genome evolution and analysis method that generates many variants of a base genome. Extensive progress is being made to complete the remaining chromosomes.

What Comes Next?

- With completion of the synthetic yeast genome coming into sight, possibly as early as 2017, the choice of what multicellular genome to synthesize and design next has been discussed. Just as HGP-read started with sequencing simple genomes through progressively more complicated ones, several candidates have been advanced, including *Drosophila melanogaster* (fruit fly) and *Caenorhabditis elegans* (nematode worm). Neither alone is likely to drive the transformations needed to make true progress in understanding the genetic blueprint defined in HGP-read.

- Among genomes worthy of synthesis, the human genome would be truly transformative, with maximal implications for advancing the biomedical field, deciphering functional consequences of human genome variation and offering related benefits. By focusing on building the 3Gb of human DNA, HGP-write would push current conceptual and technical limits by orders of

magnitude and deliver significant scientific advances. Recent technological advances in whole genome synthesis (*14*) and genome editing (*15, 16*) technology are currently revolutionizing the field of genome engineering (*4*). The newfound availability of haploid stem cells (*17*) could similarly revolutionize human genome writing by significantly reducing the complexity of the genome. The other mammalian genome that is the most obvious choice for synthesis is that of the mouse, the best understood experimental mammal.

- In order to complete large-scale design and synthesis of new genomes, many scientific and technical issues must be overcome. While not insurmountable, biological systems, compared to computers, have less predictable processes, evolve over time and are intrinsically very complex. A "Grand Challenge" project will help deepen our understanding and improve our design capabilities.

- In addition, continued breakthroughs in basic science and the development of innumerable tools, assembly methods, and standards will be needed as the project progresses. DNA synthesis and recombination methods, while improving in quality and cost effectiveness, are still too limited, slow and costly and produce DNA fragments that are too small to allow the creation of the large scale projects necessary to create the breakthroughs in applications desired.

- A new "Grand Challenge" like GP-write, will galvanize the scientific community, engage a wide network of educational institutions, attract the required resources to successfully conclude the project, and accelerate research and development in order to make the advances engineering biology requires to solve many of the global problems we face.

Operation of GP-write and HGP-write

- GP-write and HGP-write would proceed in phases according to a roadmap (Appendix 1) to evaluate feasibility and value, just as HGP-read, ENCODE and the Synthetic Yeast Project also began with pilot projects (Appendix 2) focusing on a fraction of the genome: typically ~ 1% but not limited to this amount. Intense technology development would also occur during this 3-5 year period to facilitate the other 99%. During this 3-5 year period, there would be ample time to debate whether a specific full HGP-write project (i.e. whole genome) design was scientifically, economically and ethically justified, or whether whole genome activities would be better slanted toward a model system such as the mouse.

- In earlier projects, the pilots were arbitrary with respect to prioritization of functional value or practical applications. But for HGP-write, the pilot projects would be chosen to provide early stage resources valuable for biomedical research and/or biotech production, and/or safer stem cell therapeutic platforms. For example, using induced pluripotent stem cells (iPSCs) (*18*) to create a more stable biomanufacturing human cell resistant to any virus could help reduce costly incidents of virus contamination of vaccines seen in the past, including SV40 contamination of the Sabin oral polio vaccine, and Vesivirus in the Genzyme orphan drug production fermenters in both the EU and USA (*19*). This particular project could also allow the powerful scrambling system developed for the synthetic yeast genome to be incorporated at the same time. Other desirable features could be added to this genome; a specific proposal for an "ultrasafe" human cell engineering platform, together with other suggested proposals for possible pilot projects, are appended to this document.

- Redesigning genes in such a human cell line offers a way to understand how the HGP-write genome might be made more resilient without impact on function, while providing new valuable features such as resistance to viruses, cancer and other diseases, and aging processes.

- These sub-scale HGP-write projects would also need to address the social, ethical, public perception etc. issues likely to arise as a reaction to this proposal (*20, 21*).

- The use of human organoids derived from iPSCs would help reduce animal testing and in many cases increase accuracy of assessment of function and/or therapy testing. The project would help researchers develop skills in testing and developing innovative technologies. Moreover, it would serve as an excellent readout for whether partially synthetic genomes retained all natural functions, such as "epigenomic marks" which are heritable changes to DNA that do not involve a change in DNA sequence such as cytosine methylation and modifications of the structure of nucleosomes in which DNA is packaged. The epigenomic tools to systematically and faithfully program organoids could be as transformative as the genomic tools.

- The goals of GP-write and HGP-write are necessarily ambitious, since building a human genome at today's prices would cost more than HGP-read. The costs of the project lie not only in the raw DNA, but in the design infrastructure, assembly, integration, functional assays and biological interpretation required to understand the resulting phenomes. However, an expectation of GP-write will be to catalyze a sharp price drop as new

technology development occurs apace with advancement of the project as with the cost of DNA sequencing in HGP-read. Indeed, recent and continued improvements in enabling technologies are not only improving the cost, quality and time period necessary to complete the project, but also, most other fields of biomedicine; thus the return on investment will be swift and substantial.

- The estimated cost of GP-write and HGP-write will be worked out over the next few months, but the goal would be to launch the project with $100 million in committed support.

- Roughly $40 million has been spent on microbial and mammalian genome engineering pilot projects represented by co-authors on this paper to date and Autodesk, has offered a leadership gift of $250,000 to seed the planning and launch of GP-write and HGP-write. The project will require four matching gifts to cover the planning and launch effort.

Responsible Innovation

- Genome synthesis is a natural extension of the genetic engineering tools that have been used safely within the biotechnology industry for the past 40 years and have provided significant benefits to society. However, recent technological advancements, such as chip-synthesis, standardized libraries of genomic parts and genome editing technologies are revolutionizing the field and creating uncertainty in how these technologies will be applied.

- As human genome-scale synthesis appears increasingly feasible, a coordinated scientific effort to understand, discuss, and apply large genome editing technologies is timely; and public discourse is both expected and encouraged. HGP-write will require public involvement and consideration of ethical, legal, and social implications (ELSI) from the start. However, responsible innovation requires more than ELSI; it also involves identifying common goals important to scientists and the wider public through timely and detailed consultation among diverse stakeholders.

- HGP-write will enable public discourse; having these conversations well in advance of project implementation will guide emerging capabilities in science and contribute to societal decision-making. Through open and ongoing dialog, 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.

- To assure responsible innovation and ongoing consideration of ELSI, a percentage of all research funds will be dedicated to these issues, enabling inclusive decision-making. 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 calls 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 of setting standards rests with the scientists and their community.

Organization/Administration of GP-write and HGP-write

- Dr. Jef Boeke from New York University School of Medicine, a geneticist and the lead designer/investigator of the Sc2.0 project, and Dr. George Church, from Harvard Medical School, geneticist/technologist behind the rE.coli project volunteered to assemble a key group of scientists, ethicists and policy makers steeped in human biology, health and synthetic biology. This group met on Saturday, October 31, 2015, to conceptualize and discuss the feasibility of HGP-write. The participants in this organizing meeting were invited to co-author this White Paper, to be sent to interested funders.

- On May 10, 2016, more than 130 international scientists, industry leaders, ethicists and policy makers steeped in human biology, health and synthetic biology came together in Boston for an expanded organizing meeting of HGP-write. This meeting focused on the design and technical issues, ethical and social issues, and industry involvement of HGP-write. Based on discussions at this meeting, the GP-write Leadership Group decided to broaden the scope of the project to include all large animal and plant genomes, including the human genome, and to make HGP-write one component of GP-write.

- The October and May meetings are part of a series of scientific discussions amongst the community that have been unfolding over the past several years. Initially, these meetings were focused on the synthesis and testing of yeast and bacterial genomes, and the future of synthetic biology. More recently, the focus has shifted to synthesis of variations within large genomes as a pathway to addressing some of the many global challenges facing humanity, including healthcare and the environment.

- The launch of GP-write is scheduled for 2016, a significant milestone since it falls during the 25th anniversary year of the original HGP-read, and marks another big leap forward. This is a chance to inspire a new generation of scientists and engage the public in sustained discourse.

- GP-write will be implemented through a new, independent nonprofit organization, the Center of Excellence for Engineering Biology. The Center, which will operate virtually at first, will manage initial planning and coordination efforts. These efforts include supporting the formation and work of multi-institutional and interdisciplinary research teams working in a highly integrated fashion, responsive to and engaged with a broach public outreach. Additional Centers could be included in the future.

- The Center will create a neutral environment for international participants and will accept funding from the public, private, philanthropic and academic sectors, including international funding agencies. However, this in no way precludes direct grants and sponsored research agreements by government agencies and others to academic and industrial laboratories through more traditional mechanisms.

- A Scientific Executive Committee of the Center, supported by scientific working groups, will set the scientific priorities for GP-write and supervise peer-reviewed research projects. The Center will enter into affiliation agreements with the major universities and philanthropies participating in GP-write. Acting as a coordinating center, it will support and coordinate the work of multi-institutional and interdisplinary research teams working in a highly integrated fashion responsive to and engaged with a broad public outreach. This is similar to the way HGP-read was administered within the federal government.

- Key governance and operational issues for the Center will be worked out during the initial organizing and planning process for GP-write. This process will be led by Nancy J Kelley, JD/MPP, previously the Founding Executive Director of the New York Genome Center, who has been leading a strategic planning process with the synthetic biology community since 2014. Wilmer Hale will assist in these efforts.

- The IT infrastructure for GP-write will be designed to foster large-scale collaborative science across international boundaries, institutions and scientific disciplines. Using the power of open networks to solve complex scientific problems, an open sourced, fully automated design, test, build platform would support the sharing of knowledge and expertise and large-scale pooling of data. Essentially, this platform would create a

community hub designed to enhance communication and grow communities of scientists with diverse skill sets that can work together on the complex and diverse scientific problems that would need to be tackled in GP-write.

- Leading experts in this area can offer important insights in order to design and build this infrastructure as well as provide tools and the environment to conduct dynamic, large-scale collaborative biomedical research. These include, among others, Global Alliance for Genomics and Health, UCSC Genome Browser, Sanger Institute, Broad Institute, Beijing Genomics Institute, OpenHumans.org, Sage Bionetworks, Google Genomics, the Wyss Institute for Biologically Inspired Engineering and Cloudera.

- Intellectual property developed in GP-write will encourage broad access and use through the use of patent pooling and common Licensing Agreements. The open source free-sharing of engineered cells policies of the Sc2.0 project will also be considered for at least some components of the project.

- A diverse array of perspectives—including leading experts, but more broadly harnessing the collective wisdom of ordinary members of the public--will be engaged to consider the ethical, legal/policy, social, religious, and other impacts as well as the public education needs of GP-write in collaboration with the Center of Excellence in Engineering Biology (15). As examples, the Human Practices group within SynBERC, the JCVI, and the Woodrow Wilson Center for International Scholars have been working with the Synthetic Biology Community on these issues for several years, supported by the Sloan Foundation.

- We strongly encourage and value public discourse surrounding HGP-write. Having these conversations well in advance of deliverables will help society better prepare for and guide emerging capabilities. Through a transparent and inclusive dialog, common goals can be identified and concerns, based on cultural values, can inform the direction and output of the project. Finally, the highest biosafety standards should guide project work, and safety for lab workers, research participants, and ecosystems should pervade the design process.

- A dedicated percentage of all research dollars for the overall project for HGP-write will be specifically targeted toward the issues above. The project infrastructure will include active engagement of voices from the humanities and social sciences, including but not limited to bioethicists and legal scholars, and it will establish robust mechanisms for community engagement to enable inclusive decision-making.

- Funding for GP-write will come from public, private, philanthropic, industry and academic sources around the world as well as from emerging sources, such as crowdfunding, that can amplify the voices of ordinary citizens in defining worthy projects to pursue. Fundraising efforts have begun.

Appendix 1 -- Draft Roadmap for GP/HGP-write

- Years 1-5: Massive technology development effort on writing, delivering and testing aspects on *E. coli*, yeast, and mammalian cells, with global coordination.
- Year 1: Select Genome Pilot Projects to move forward toward as international consortia.
- Years 2-5: Human and other whole genome synthesis Pilot Projects selected with a well defined therapeutically, agricultural or diagnostically valuable "product" (early return on investment).
- Years 1-4: Major effort to engage with representative members of the public as well as with the scientific community, humanities, focus groups etc., to identify common outcomes of various possible whole genome synthesis projects, to identify potential benefits that resonate best with the public.
- Years 1-5: Safety and Surveillance Technology Development.
- Years 5-10: Ongoing Technology Development; additional pilots, as costs drop and resources increase, tackle scales from epigenomic to ecosystem.
- Years 5-10: Shift into high gear on whole genome projects. Human cells and/or organismal (mouse), plant and others, depending on input from the public, the funding agencies and the scientific and industrial communities
- Transition to environmental, farm, factory, clinical tests

Appendix 2 -- Themes for Pilot Projects for HGP-write

A. Ultrasafe Cell Line, Jef Boeke, Farren Isaacs, Marc Lajoie, Nili Ostrov

HGP-write proposes to engineer a human cell line for use as a basic and potentially universal platform for human biotechnology. This can be done by altering roughly 1% of the genome, including the exons of all of the genes, and nearby sequences, leaving the vast majority of the noncoding regions, which are at this point far less well understood, untouched.

It is anticipated that many of the details of this cell line remain to be extensively discussed and the plans for exactly how to construct it will be jointly worked out by a group of human genetics and genomics experts as well as synthetic genomicists.

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 actual *ex vivo* and ultimately *in vivo* therapeutic applications. As described below, this cell line will be engineered to be ultrasafe from many distinct perspectives. Because this cell line would potentially be of great value to the pharmaceutical, vaccine, and biotechnology industries, this theme as a pilot project should be very attractive to philanthropic and industry, potentially as a public/private partnership designed to benefit the entire biomedical community.

Safety and other properties of an "ultrasafe" cell line

Property	Explanation
Virus resistant	Cell line resists viruses through codon/tRNA recoding
Prion resistant	Endogenous prion gene deleted or recoded
Retroelement/transposon free	All mobile DNAs inactivated
Triplet repeat resistant	Triplet repeat regions made safe by recoding
Germ line negative	Engineered to prevent germ line transmission
Radiation resistant	Very active DNA Repair systems
Multiple self destruct circuits	Orthogonal mechanisms to control cell growth
Cancer resistant	P53 and other tumor suppressors re-engineered to minimize chance of deleterious mutations
Immuno-negative	Engineered to minimize immune rejection
Multiple safely targetable sites	Safe insertion of future engineered circuits and pathways
Major allele for every coding region SNP and indel	"Average" human cells maximizes compatibility with diverse human populations
Scramble-able	Allows rapid evolutionary optimization for desirable traits

B. Synthesizing Prototrophic Mammalian Genomes, Pam Silver, Harris H. Wang

Chronic malnutrition is a global health challenge leading to a variety of systemic diseases, mortality and negative life outcomes, afflicting especially children and adults in underdeveloped and developing countries. In contrast to bacterial cells, which can produce all metabolites needed for growth from simple sugars and elemental building blocks, humans are metabolically incapable of biosynthesis of 9 of 20 amino acids needed for life. Instead, these 9 essential amino acids are derived from diet. In addition to amino acids, at least 10 essential vitamins needed to be consumed as well since they cannot be produced by the body. While certain vitamins can be produced by humans, they are often generated at insufficient levels and require significant dietary supplementation. These deficiencies result in chronic malnutrition and food shortage and can potentially be addressed using synthetic genomic approaches.

In general, the simpler amino acids are easily derived from core intermediates of metabolism. In comparison, more complex amino acids, requiring up to 15 steps, are essential. Due to unknown evolutionary pressures, humans (and related mammals) have yet to evolve capabilities to synthesize these essential compounds. The biosynthesis of essential amino acids and vitamins are generally well-characterized and the genes and pathways responsible for them are known.

We propose to heterologously introduce these biosynthetic pathways for amino acids and vitamins (as well as the regulatory factors) into the human genome as to enable human cells to be less dependent on (or completely independent of) exogenous supplementation of these metabolites. It is anticipated that the proof-of-concept studies will be done in cell lines including stem cells, which can be grown in defined media. In the future, one could imagine extending other pathways such as photosynthesis machinery from algae or plants to generate autotrophic human cells.

Utility: In addition to the eventual goal of generating prototrophic human cells to combat malnutrition, one could imagine this study to be of utility for understanding the biochemical milieu needed for mammalian organismic development, cell differentiation, and nutrition-associated aging processes. Since current mammalian production cell lines (e.g. CHO cells) require expensive growth medium, one could imagine using these prototrophic cells with cheaper growth medium formulations at larger bio-production scales, leading to more economical biosynthesis of various drugs and biologics that require mammalian cell lines.

It is likely that a variety of pathway configurations need to be explored to optimize biosynthetic levels, an effort well-suited for a DNA synthesis project of this size. The

chromosomal integration sites will need to be stable. Thus this effort will likely leverage technical resources, advances, and activities generated by other HGP-write projects (e.g. building additional chromosomes or discovering safe chromosomal insertion sites). In order to generate all essential amino acids, vitamins, and other metabolite pathways, we anticipate that >200 kb of DNA will need to be synthesized, possibly more to test various expression/pathway configurations, which is actually closer to 0.1% of the human genome.

C. Probing Chromosome Structure, Function and Building Better Containment,
Farren Isaacs, Jasper Rine, Ting Wu

I have taken the 1% Pilot to mean projects involving the synthesis of genomes or genome fragments that are less than 30 megabases, focusing attention on microbes and portions of single chromosomes in other species. In reducing the grand vision to practical and achievable goals, two issues provide initial foci. 1.) Safety and containment: Can we build reliable tools that would prevent a synthetic genome from ever emerging following hybridization with a natural genome? 2) What can we learn about the contribution of DNA sequence to chromatin and chromosome structure at a more macro way than previously possible?

The synthetic yeast and *E.coli* genomes can play vital roles. There will likely be substantial pushback on the goals of the GP-write. But demonstration of a class of questions that can only be answered by use of synthetic genomes may help the community to see the goals and opportunities that the GP-write could bring.

Safety and Containment. If synthetic genomes are to play a role in agriculture, we will face a daunting scale issue given that millions of acres are under cultivation for the few major crops that have been engineered (corn, soybean, cotton, and canola in the US). While the probability of an adverse event is low, the opportunities could be enormous. The common occurrence of hybrids in nature suggests that any synthetic organism could find a way to hybridize with a natural organism at some frequency. So, I think methods should be developed that would ensure that no recombinants from such a hybrid would survive. The classic Drosophila balancer chromosomes provide one such inspiration. The synthetic yeast genome provides opportunity to observe how much containment can be achieved in a well-understood context. The project would consist of using the genome scrambling capability being put into Sc2.0 to first scramble one chromosome and then multiple chromosomes to learn the efficiencies by which scrambled balancers could prevent meiotic survivors, and analyze failure modes allowing recombinant gamete survival. Ancillary activities could include development of toxin-antidote combinations, e.g. restriction-modification gene pairs that were inserted into centromeres such that, if the pair were ever separated, the cell would die.

Chromatin and Chromosome Structure. We have been relatively blind to a middle range of chromatin structures larger than nucleosomes, and beyond the limits of light microscopy. The various Chromosome Conformation Capture technologies, especially Hi-C as well as super-resolution Oligopaints (*21*) are making a contribution, but to date they have been largely data collecting exercises rather than experiments. The Sc2.0 and scrambled derivatives provide a logical opportunity to determine the functional consequences of sequences, shapes and contact maps of these chromosomes. Drosophila polytene chromosomes are nature's own magnifier of chromatin structure and provided the classic proof that chromatin structure was determined by underlying DNA sequence. If we could introduce substantial synthetic sequences into Drosophila, we could use polytene chromosomes to uncover principles behind how particular sequence creates particular structure.

D1. Engineering Human Organoids with the Human Variome as a Substrate,
Adam Arkin, George Church, Jennifer Lewis, Elaine Lim

Numerous GWAS studies and population genetic studies have identified variations in the human genome that are correlated with key disease, aging, and susceptibility phenotypes. Further, and either less or more interesting there is an increasing number of hominid genomes that present possible evolutionary variation that may be important in understanding current human function. These variations are likely important for different responses to medical treatments and environmental challenges that have been difficult to resolve with complex clinical studies or the small number of cell lines used for clinical and drug research. A number of key innovations have lead to the ability to make structured tissues with properties very similar to an individual's organ tissue for a number of organ systems. A step beyond this would be to take a set of exemplar tissues we can make in this way and introduce into them with genome editing technology as many of the known important allelic and genome structural variations identified above as possible (*20*). Combinations could be those known from individuals that identify human and hominid subpopulations, or introduced somewhat randomly. The resultant synthetic tissue bank then becomes an excellent substrate for drug and environmental response screening. Validation that these tissues respond differentially in ways similar to different human subpopulations could breed a great deal of interest from the pharmaceutical industry. The existence of differential response of these tissues to chemical, microbial/viral, or other environmental challenge then provides and excellent starting point for mechanistically unravelling the molecular basis for these differences and ways to control these interactions and outcomes.

D2. The Seven Signals and 1576 Transcription Factor Toolboxes: Reprogramming Cells to Enable Therapies, Liam Holt and Alex Ng

The HGP-write ultrasafe cell-line will likely be generated in an induced pluripotent state. This approach will derive maximum value from the HGP-write project in terms of scientific understanding, commercial application and therapeutic potential because the pluripotent cell can be differentiated into any human cell-type or tissue. To fully capitalize on this flexibility we will develop a synthetic approach to control the differentiation of HGP-write cells.

The astonishing diversity of cell and organ structures in the human body are largely generated from the action of just seven conserved developmental signals: the Hedgehog, Wnt, TGF-β, receptor tyrosine kinase, Notch, JAK/STAT and nuclear hormone pathways. These pathways are activated in many combinations and temporal patterns to drive the differentiation of stem cells into hundreds of cell types during embryogenesis. These pathways all have a tree-like structure, with many 'branches' accepting input information that is funneled to a core 'trunk' that transduces the signal, which is then relayed into many 'roots' leading to hundreds of transcriptional and cell biological changes. Taking Wnt signaling as an example: There are 29 Wnt-family ligands that bind to 8 Frizzled receptors (the branches). These branches feed into a single GSK3- core transducer (the trunk), which then drives activation of many transcription factors to activate or repress hundreds of genes (the roots). Most current efforts in cell reprogramming either modulate growth factors or various combinations of transcription factors. These approaches are searching through the vast combinatorial space in the branches and roots. In contrast, we will develop synthetic methods to perturb the seven major developmental signaling pathways at their conserved core: the trunk of the tree. We will develop a genetic toolbox for the precise temporal modulation of the seven pathways at their 'trunk' in any combination using optogenetic and chemical controls. In addition, we build a complimentary toolbox of reporters for the rapid phenotypic profiling of these cells thereby enabling the mapping of signaling to phenotype in high-throughput.

This approach vastly reduces the search space necessary to find the appropriate reprogramming signals that will generate a wide range of differentiated human cell-types. We recognize that the effect of activating a signaling pathway can be very different in distinct cell types due to the presence or absence of co-activators, repressors and other signaling modulators. However, we believe that this context is generated by the history of activation of other members of the seven core pathways.

Therefore, we believe we can recapitulate most cellular reprogramming by driving the seven pathways at their core in the appropriate temporal combinations. Indeed, cell fate determination is hierarchical, supporting this idea and increasing the feasibility of this project by allowing the problem to be parsed into discreet steps. This project will both create a map for the artificial control of development and generate tools that facilitate the control cell of differentiation *in vitro*. Ultimately, a precise control of cell differentiation is a crucial step toward cell therapies, tissue replacement or even organ transplants either with patient-derived iPS cells or ultrasafe HGP-write cells.

E. Through The Looking Glass: Anticipating And Understanding Governance Systems And The Public's Views On HGP-Write,

Barbara Evans, Todd Kuiken, Jeantine Lunshof,

Understanding and anticipating the governance issues around HGP-write will be a critical component for the project's success which may include local, state and national oversight regimes both at the research stage and for the eventual approval of any product/applications. In addition, establishing a public dialogue and outreach program to not just inform, but incorporate the public's views into the research program, will better enable HGP-write to acquire and sustain funding which will span multiple administrations over time and enable public support to continue such funding.

F. A New Generation of Human Artificial Chromosomes, Mark Budde, Lacramioara Bintu, Alina Chan, Iain Cheeseman, Michael Elowitz, John Glass, Vladimir Larionov, Nicholas Lee, Leslie Mitchell, Pamela Silver, Jeffrey Way

Since the 1990s, researchers have strived to create human artificial chromosomes (HACs) to address the limitations of viral-based mammalian vectors. HACs are anticipated to allow large cloning capacities (megabase-scale), copy number control, long-term gene expression, preclusion of anti-viral responses, and reduced mutagenesis and insertion into host chromosomes. An ideal HAC should be easy to transfer between different mammalian cell lines as well as yeast, have options for deactivation, possess multiple orthogonal integration sites, and permit (epigenetic) regulation by transcription factors and chromatin modifiers.

The two most salient HACs today can persist in cell lines for months and be manipulated in Chinese hamster ovary (CHO) cells using site-specific recombination; genes of interest and genomic regions can be loaded into a single loxP site in these HACs. However, these HACs cannot be manipulated in bacteria or yeast. Gene loading and moving the modified HACs into human cells for functional

studies via microcell-mediated chromosome transfer (MMCT) takes months and requires an expertise rarely found outside the labs that invented the HACs. These steps can be greatly accelerated if HACs can be assembled and propagated in yeast cells as yeast artificial chromosomes (YACs). YACs containing well-defined basic functional units have been available since the 1980s and are convenient to manipulate using conventional cloning protocols. YACs have been used in hundreds of studies going so far as to study human genetic elements in YAC transgenic mice and build exogenous biosynthetic pathways to produce diverse compounds in yeast.

Generating a HAC for easy manipulation and universal transfection holds the potential to transform current practices and widening the breadth of possible experiments and synthetic biology designs in mammalian cell lines.

We have designed and are testing a synthetic minimal human chromosome with an artificial centromere, telomeres, and insulator-separated gene expression domains based on current state of the art. These efforts are additionally pushing the limits of complex DNA synthesis. To enhance HAC stability and prevent its rearrangement in the recipient cells, we are tethering human centromeric proteins to the synthetic centromere in the HAC. This strategy has been demonstrated to increase the efficiency of de novo functional kinetochore formation in mammalian cells, which is critical to preventing HAC amplification and rearrangement. Our preliminary findings have confirmed that the presence of a centromere in combination with a specific centromeric protein significantly improves the expression and retention rate of a HAC in human cells. We will proceed to evaluate our HAC in terms of stability, segregation, and maintenance of chromatin boundaries in various cell types. We are also building HAC variants to be amenable to multiple cell lines. With a well characterized HAC, we will also engineer gene expression from multiple loci on the HAC to be compatible with both conventional and chromatin-based regulation. The HAC will in parallel be developed so that it can host large mammalian genetic circuits that can control cell development, intercellular interaction, and other synthetic functions. Our ultimate goal is to derive a stable and user-friendly nucleomics tool for distribution, with many potential applications in cell line diversification and generation, animal transgenesis, and human gene therapy.

G. Pig genome humanization to produce tissues and organs for transplanation, Jef Boeke, Scott Fahrenkrug, Luhan Yang.

Transform the pig genome to make it amenable for human organ transplantation altering immune, complement, coagulation and anti-viral strategies (APOBEC and CRISPR)

H. Synthetic genomes in photosynthetic organisms, Jim Haseloff, June Medford

Developing sensors for primary and secondary metabolites, and new synthetic genomic tools and testing including *Arabidopsis, Zea. Marchantia.*

I. Many synthetic fungal, microbial and viral genomes, Jef Boeke and the Yeast Consortium, Nili Ostrov and the RE.coli project team.

Application of lessons from *S. cerevisae* and *E.coli* to other genomes such as *Lactococcus, Lactobacillus, Synechococcus, Agaricus*

Other pilot projects related to GP-write could include:

J. Altering of endogenous viral genomes and other repetitive elements (SINEs, LINEs, centromeres) and testing for impact on organ development and function *in vitro*.

K. Making cancer resistant, pathogen resistant and/or senescence resistant organs *in vitro* or in pigs to improve understanding as well as outcomes in clinical transplanation settings and preventative medicine.

L. A synthetic version of the NIH Encyclopedia of DNA Elements (ENCODE) project to answer questions about which particular cis-regulatory elements or gene products are "just noise" or have functions evident upon removing one or more motifs, family members or putative parallel pathway members.

M. An ancestral human protein coding genome, bearing the homozygous ancestral allele (and/or the major allele) at each position.

References and Notes:

1. C. International Human Genome Sequencing, Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).
2. C. W. Fuller *et al.*, The challenges of sequencing by synthesis. *Nature biotechnology* **27**, 1013-1023 (2009).
3. P. A. Carr, G. M. Church, Genome engineering. *Nature biotechnology* **27**, 1151-1162 (2009).

4. A. D. Haimovich, P. Muir, F. J. Isaacs, Genomes by design. *Nature reviews. Genetics* **16**, 501-516 (2015).
5. J. Cello, A. V. Paul, E. Wimmer, Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297**, 1016-1018 (2002).
6. D. G. Gibson *et al.*, Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52-56 (2010).
7. M. J. Lajoie *et al.*, Genomically recoded organisms expand biological functions. *Science* **342**, 357-360 (2013).
8. F. J. Isaacs *et al.*, Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348-353 (2011).
9. N. J. Ma, F. J. Isaacs, Genomic Recoding Broadly Obstructs the Propagation of Horizontally Transferred Genetic Elements. *Cell Syst* **3**, 199-207 (2016).
10. D. J. Mandell *et al.*, Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55-60 (2015).
11. A. J. Rovner *et al.*, Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518**, 89-93 (2015).
12. N. Annaluru *et al.*, Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55-58 (2014).
13. J. S. Dymond *et al.*, Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471-476 (2011).
14. S. Kosuri, G. M. Church, Large-scale de novo DNA synthesis: technologies and applications. *Nat Methods* **11**, 499-507 (2014).
15. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
16. P. Mali *et al.*, RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).
17. I. Sagi *et al.*, Derivation and differentiation of haploid human embryonic stem cells. *Nature* **532**, 107-111 (2016).
18. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).
19. V. Bethencourt, Virus stalls Genzyme plant. *Nat Biotech* **27**, 681-681 (2009).
20. D. Baltimore *et al.*, Biotechnology. A prudent path forward for genomic engineering and germline gene modification. *Science* **348**, 36-38 (2015).
21. "Presidential Commission for the Study of Bioethical Issues," *New Directions: The Ethics of Synthetic Biology and Emerging Technologies* (Washington, DC, December 2010).