The document that follows is an initial draft of the **2020 National Human Genome Research Institute (NHGRI) Strategic Plan**, which will be published in October 2020. This document is being openly distributed in an effort to solicit final community input about the 2020 NHGRI Strategic Plan; this will be the final major step of the ~2-year Genomics2020 Strategic Planning Process (for additional details, go to genome.gov/about-nhgri/strategic-plan/overview).

Because of the intentionally incomplete and draft nature of this document – and the anticipated substantial changes that will occur prior to its submission for final publication – NHGRI is respectfully requesting adherence to the following etiquette:

Please do not:

- Write or report about this document in any press coverage
- Cite this document in any publication or news stories
- Post or share any part of this document (e.g., excerpts, screenshots, figures, etc.) on social media
- Post this document to a preprint server or other distribution network

Adhering to this requested etiquette would match the spirit in which this draft document has been made broadly available – that is, encouraging discussion and soliciting edits that will ultimately yield the highest quality and most impactful 2020 NHGRI Strategic Plan. *Press coverage and social media postings about the final strategic plan will be welcome upon its publication in October 2020!*

Such input is being collected by various means, including feedback that can be given about any aspect of the draft document at genome.gov/2020draft. The deadline for providing that feedback is **5:00 pm ET on April 26, 2020**.

For questions, please email genome2020@nih.gov.

Thanks for your engagement and contributions for finalizing the 2020 NHGRI Strategic Plan!

Version April 13, 2020

Strategic research priorities and opportunities for improving human health at *The Forefront of Genomics*

Abstract

To be drafted following receipt of feedback about draft document in April.

Introduction

Three decades ago this month, a pioneering group of international researchers began an audacious journey to generate the first map and sequence of the human genome, marking the start of a 13-year odyssey called the Human Genome Project. The project's successful and early completion in 2003 catalyzed enormous progress in genomics research. Leading the signature advances has been a greater than one million-fold reduction in the cost of DNA sequencing. This has allowed the generation of hundreds of thousands of human genome sequences, both in research and clinical settings. With this, researchers have been enabled to create rich catalogs of genomic variants, to gain an ever-deepening understanding of the functional complexities of the human genome, and to elucidate the genomic bases of thousands of human diseases. In turn, the last half-decade has brought the initial realization of genomic medicine, as research successes have been converted into powerful tools for use in healthcare, including somatic genomic testing for cancer (enabling the development and use of targeted therapeutics) and genome-sequencing tests for autism, intellectual disability, and birth defects, among others. These advances were unimaginable when the Human Genome Project began and continually lead to scientific and clinical opportunities beyond our expectations.

Embracing its leadership role in genomics, the National Human Genome Research Institute (NHGRI) has developed strategic visions for the field at key inflection points, in particular at the end of the Human Genome Project in 2003 and then again at the beginning of the last decade in 2011. These visions outlined the most compelling opportunities for human genomics research, in each case informed by a multi-year

engagement process that captured input from numerous stakeholders with diverse expertise.

Catalyzed by the implementation of these strategic visions, the human genomics enterprise has expanded considerably since the end of the Human Genome Project, becoming widely disseminated across biomedicine. Historical funding trends at the U.S. National Institutes of Health (NIH) offer one means for illustrating this phenomenon (Figure 1). During the Human Genome Project, NHGRI was NIH's primary funder of human genomics research, but the past two decades have brought a greater than ten-fold increase in the relative fraction of funding coming from other parts of NIH as genomics became integrated in their research portfolios.

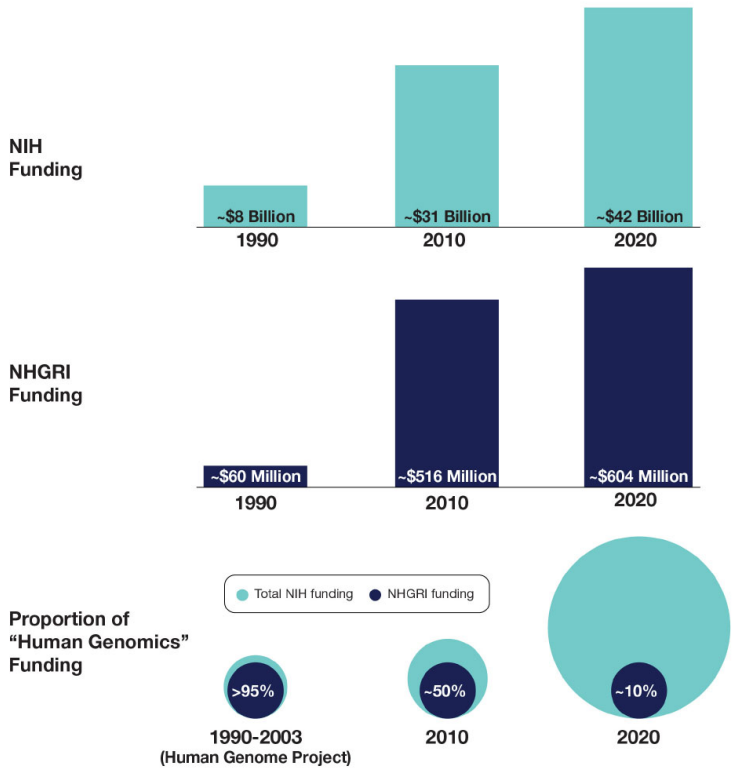

Figure 1: Funding trends of NIH and NHGRI over the past 30 years. The total funding levels for NIH (top panel) and NHGRI (middle panel) are indicated for 1990, 2010, and 2020, respectively. Also shown is the relative proportion of funds supporting "human genomics" research provided by NHGRI versus all of NIH (as derived from NIH funding databases) for three corresponding time intervals (bottom panel). Note that while the NHGRI budget increased roughly ten-fold over the last 30 years (middle panel), the Institute's proportion of total NIH funding for human genomics research decreased from >95% during the Human Genome Project to ~10% at the beginning of the current decade (bottom panel), reflecting a greater than ten-fold increase in non-NHGRI funding.

Noting the accelerating pace of genomic advances since 2011, NHGRI endeavored to start the new decade with an updated strategic vision for human genomics research. Through an engagement and outreach process that involved over 50 events (e.g., dedicated workshops, conference sessions, and webinars) over the last two-plus years (genome.gov/genomics2020), the Institute collected input from an even-more diverse set of stakeholders than was done during previous rounds of strategic planning. The process consistently revealed the highly disseminated nature of human genomics research, the impracticality of being comprehensive with respect to planning across all areas of genomics, and thus the need to focus on the most cutting-edge opportunities in human genomics research. We ultimately linked NHGRI's recently conceived organizational mantra *The Forefront of Genomics* directly to this process and asked the participating stakeholders to reflect on this theme in formulating their strategic input. From the ensuing discussions, it became apparent that *responsibility* is a central aspect of being at *The Forefront of Genomics*, specifically in the following four areas:

I. Providing a socially responsible and highly ethical framework for conducting human genomics research by establishing and adhering to guiding **principles and values**;
II. Maintaining and improving a robust **foundation** for genomics research;
III. **Breaking down barriers** in genomics through advances that create new research opportunities; and
IV. Defining and pursuing audacious genomics **research projects** focused on elucidating genome function, understanding human disease, and improving human health.

What follows are summaries of each of these four areas, which together highlight NHGRI's view of the most compelling biomedical and clinical research opportunities in human genomics for the immediate future. This strategic vision is intentionally limited to applications of genomics in biomedicine, which is provided with a deep appreciation of the growing impact of genomics in numerous others research areas. In articulating a forward-thinking vision for human genomics, we faced the inherent tension between describing a view of the broader field and representing an identity for NHGRI as a research funder; we acknowledge the challenge of consistently separating these two aims, but nonetheless endeavored to emphasize broad strategic goals for the field as opposed to implementation tactics for actually funding the research.

I. Establish and Adhere to Guiding Principles and Values for Conducting Human Genomics Research

As genomics has matured as a discipline, the field has embraced a growing set of fundamental principles and values that together serve as a guiding compass for the research efforts. Anticipating the growing complexities of genomics and its many applications (especially into medicine), it is important to reaffirm and even sharpen these tenets, such as those highlighted in Box 1. This commitment to exploring the real and potential impact of science on society reflects the long-standing attentiveness of the genomics research community to the broader implications of scientific progress; this is exemplified by a subdiscipline of the field that focuses on ethical, legal, and social implications (ELSI) research – an area of continued importance and relevance at *The Forefront of Genomics* (see Box 2).

Box 1: Guiding Principles and Values for Human Genomics Research

- **Maintain an overarching focus on using genomics to understand biology, to enhance knowledge about disease, and to improve human health** – *genomics is now foundational across the entire continuum of biomedical research, from deciphering fundamental principles of biology to translating that knowledge into medical advances.*

- **Embrace the interdisciplinary and team-oriented nature of genomics research** – *starting with the Human Genome Project, some of the most challenging genomics endeavors have benefited from the creation and management of large, interdisciplinary research collaborations.*

- **Pursue advances in genomics as part of a vibrant global community of genomics researchers and funders** – *the challenges in genomics require the collective energies and creativity of an international ecosystem of researchers, funders, and other stakeholders.*

- **Champion open science, data sharing, and attention to rigor and reproducibility in genomics research** – *the genomics enterprise has a well-respected history of leading in these areas, and that commitment must be built upon and continually reaffirmed.*

- **Promote robust and consistently applied standards in genomics research** – *adherence to carefully defined standards (e.g., those for generating, analyzing, storing, and sharing data) has benefited genomics in significant ways, and this must include appropriate privacy and data-security protections for those participating in genomics research.*

- **Embrace equity and global diversity in all aspects of genomics research, committing to the systematic inclusion of ancestrally diverse and underrepresented individuals in major genomic studies** – *attention to diversity in genomics research is both socially just and scientifically essential, including meaningful, sustained partnerships with communities in the design and implementation of research studies, the propagation of research findings, and the development and use of new technologies.*

- **Maximize the utility of genomics for all members of the public, including the ability to access genomics in healthcare** – *engagement, inclusion, and understanding the needs of diverse and underrepresented groups are required to ensure that all members of society benefit equally from genomic advances, with particular attention paid to the use of genomics in healthcare so as to avoid exacerbating healthcare disparities.*

- **Provide a conceptual framing that consistently conveys the role of both genomic and non-genomic contributors to health and disease** – *routinely considering the importance of social and environmental contributions to human health (and the interactions among those components and genomics) will be important for the comprehensive understanding of most human diseases.*

Box 2: ELSI Research at *The Forefront of Genomics*

Starting with the Human Genome Project, the field of genomics adopted a proactive approach to studying the ethical, legal, and social implications (ELSI) of genomics research. Rather quickly, the ELSI name became a well-recognized area of research associated with genomics. What began as a series of important questions related to the initial sequencing of the human genome has broadened in scope and scale within genomics and even extended to other fields (e.g., neuroscience and data science). ELSI research efforts have directly informed policies to address genetic discrimination, approaches for appropriately consenting participants in genomics research, and strategies for returning genomic information to individuals. ELSI research often attempts to anticipate and prepare for future research developments – for example, the initial policies governing CRISPR-Cas germline editing were informed by ELSI research conducted a decade before this specific genome-editing method was actually developed.

Looking ahead, a broad and robust ELSI research portfolio will be essential to ensure that the potential benefits of genomics are available to all members of society. This will be particularly important as genomics expands beyond the core group of investigators familiar with ELSI research. Specifically, researchers new to genomics are likely to re-discover ELSI research questions that have already been addressed and generate new questions as the use of genomics expands into new areas. For example, the rapidly expanding use of genomic and other multi-omic data coupled with health records and non-medical data raises challenging questions about data control, privacy, consent, and participation of under-represented populations. Similarly, new types of data collection and uses call for research examining downstream implications ranging from law-enforcement applications to considerations of race, ancestry, community, and family. ELSI research could also seek to understand the implications of studying genetic associations with non-medical traits (e.g., intelligence, sexual behavior, and social status), to address the challenges of health equity that accompany the implementation of genomic medicine (including the communication of genomic risk and therapeutic interventions), and to assess how genomics influences fundamental concepts of privacy and social identity. Finally, to support the core values of equity and diversity, ELSI research should explore how priorities and values concerning the use of genomics differ among communities. The successful use of genomics requires engaging and developing partnerships with diverse communities during all phases of research. Descriptions of some emerging and future ELSI research efforts that address issues at the interface between genomics and society are provided below.

II. Maintain and Improve a Robust Foundation for Genomics Research

Over the last three decades, genomics has grown from an emerging field to a well-established discipline. The tools of genomics are now routinely and expansively used throughout biomedical research. As a consequence, there is widespread reliance on a robust foundation for using genomic approaches. A key responsibility at *The Forefront of Genomics* is the maintenance and improvement of that foundation. Together, the various parts of this foundation are important for the full range of anticipated genomic advances, from increasing our basic understanding of genome biology to unraveling the genomic bases of human disease to the use of genomic information in medicine. But to be clear, this foundation reflects more than just infrastructure and resources; rather it also includes a number of dynamic areas of technology development and research, including those involved in generating an ever-advancing understanding of genome structure and function, providing key capabilities in data science focused on genomic analyses, addressing myriad issues at the interface of genomics and society, and ensuring appropriate training and genomic workforce development.

Genome structure and function

Enable the routine generation of genome sequences, transcriptome sequences, and epigenomic datasets. No truly complete and contiguous mammalian genome sequence has been generated to date, yet the comprehensive analyses of genomes requires the ability to generate high-quality and complete (telomere-to-telomere and phased) genome sequences on a routine basis. Such efforts will rely on the continued refinement and enhanced utility of the human genome reference sequence(s). Similarly, understanding how genomes choreograph the establishment of different cell states during development and how this choreography is influenced by environmental exposures requires the ability to generate high-quality transcriptomic and epigenomic data, including at single-cell resolution and for multiple data types simultaneously. Removing the obstacles for generating all of these datasets in a comprehensive, scalable, and inexpensive fashion requires sustained efforts in the development of technologies and analytic tools, with success in this area being critical for understanding the genomic bases of health and disease in the research setting and, ultimately, for using genomics as a clinical tool for prevention, diagnostics, and management.

Establish robust approaches for analyzing genome-sequence, transcriptome, epitranscriptome, and epigenome data. The ability to analyze increasingly heterogeneous genomic data types in a comprehensive and inexpensive fashion will enable characterizing the genomic contributions to health and disease in both research and clinical practice. However, the sheer size of the datasets and the various dimensions under study – including biological context (e.g., life stage and cell state), environment, and genomic variation – present major computational and inferential barriers. The underlying biology also presents inherent complexities, such as linking individual regulatory elements to their functionally associated gene(s) and establishing the complete repertoire of potential regulatory influences on a given gene. To address these challenges, scalable and inexpensive methods for analyzing genomic data that

keep pace with the generation of large and multidimensional datasets are needed; such methods must be both robust in their performance and readily accessible for use by a broadening community of researchers. Particularly in this area, partnerships involving academic and industrial/commercial groups have proven to be effective for improving data quality and reducing the costs of data generation.

Facilitate the use of evolutionary and comparative genomic data to advance understanding of genome function. Understanding natural genomic variation within and between species, conservation of genomic elements across species, and rapid evolutionary changes in genomic regions that can be correlated with specific traits is critical for attaining a comprehensive view of genome function. The study of a wide range of organisms continues to be instrumental for understanding the structure and function of genomes and for experimentally determining how genomic variation influences biology and how evolution shapes genomes. To capitalize on the multi-species genomic data being generated, community accepted standards for data, metadata, and data interoperability are needed, along with new methods for integrating functional data from diverse species and for visualizing and exploring increasingly complex comparative genomic datasets.

Genomic data science

Ensure facile storing, sharing, and computing on large genomic datasets. Managing large genomic datasets requires robust hardware, algorithms, software, standards, and platforms. Barriers to the efficient and effective utilization of such datasets include the lack of interoperable genomic data resources (which limits access, integration, and analyses, thereby inhibiting innovation) and the absence of controlled and consistently adopted data and metadata vocabularies and ontologies. To enable the effective management of genomic datasets requires a broad and cooperative effort aimed at developing the necessary skills, creating resources, and building consensus across the community about various data-related issues. For example, appropriate incentives are needed for biologists to learn data-management skills and for data scientists to learn biology (see below). Also needed are user-friendly systems that capture scalable, intelligent, and cost-effective metadata. Success will depend on improving access to all types of genomic data and the associated phenotypic and other metadata. Meanwhile, all relevant elements must function in a future data ecosystem that will be highly distributed.

Develop novel methods and build sustainable data resources for genomic sciences. All major breakthroughs in genomics have been accompanied by the groundbreaking development of computational and statistical methods for analyzing the data. Accordingly, a continued focus is needed for developing advanced computational and statistical approaches (including machine learning) that generalize across genomic applications. Work in this area should include enhancing strategies for integrating genomic and non-genomic data, visualizing multi-dimensional information at scale, and assessing analytical performance and accuracy. Further, with increasing amounts and uses of generated genomic data, the biomedical research community requires accurate,

curated, accessible, secure, and interoperable genomic data repositories and informatics platforms. Support for crucial data and informatics resources requires carefully considered funding strategies, including a global, multi-funder model that ensures their long-term sustainability. Approaches for improving the efficiency of such resources include the use of shared storage and computing infrastructure, the adoption of common data-management processes, and the development of automated data-curation methods. Innovations in technology and policy should be integrated to develop data-stewardship models that ensure open science and reduce data-access burdens, all while honoring study consents and participant preferences.

Develop integrated knowledgebases and informatics methods for genomic medicine. Integrating genomics into routine medical practice requires informatics and data-science advances that effectively connect the growing genomic knowledge to clinical decision-making at the point of care. To make genomic information readily accessible and broadly useful to clinicians, user-friendly electronic health record (EHR)-based clinical decision support (CDS) tools that interact with a variety of EHR and data systems (laboratory, pharmacy, and radiology, among others) are needed. Such CDS tools will play a major role in using genomic information about individual patients to stratify risk – be that of disease onset or progression, adverse or salutary response to medications, or rare or undiagnosed diseases. All of this will require well-curated, highly integrated, and up-to-date knowledgebases that connect genomic information to clinical characteristics, other phenotypic data, and information on family health history. Another foundational element in this area will be robust risk-stratification algorithms that incorporate both common and rare genomic variants, phenotypic data, and environmental information into the risk modeling; such algorithms will need to be evaluated in a variety of study designs that can assess the validity and impact of using genomic information with respect to patient outcomes and healthcare utilization. Finally, it will be important to evaluate these genomics-oriented CDS tools through rigorous usability studies to ensure that they are acceptable to practitioners from all clinical disciplines and that their limitations are understood and addressed (if possible) over time.

Genomics and society

Understand the appropriate role of genomics in broader social and cultural contexts. Genomic risk factors interact with environmental risk factors in the manifestation of multi-factorial health conditions. At present, such interactions are, in general, poorly understood. Studies of common, complex diseases must increasingly account for the social and environmental determinants of health in addition to genomic risks. Moreover, most genomic medicine implementation studies to date have occurred in a limited number of social and cultural contexts, limiting generalizability; determining the optimal implementation strategies in a broader set of contexts should be a priority.

Empower people to make well-informed decisions about genomic data and develop data-stewardship systems that reinforce their choices. Genomic data combined with phenotypic data have become a highly valued commodity, especially as

calls for open science increase. Decisions about access to and use of genomic and phenotypic data should involve the individuals and communities from whom those data were generated. Better insights are needed to identify the key issues influencing peoples' choices about the provision of personal data for research, including the values and goals that guide those decisions. Such information should be incorporated into the design and study of tools that acknowledge peoples' concerns, aid decision-making, and promote individuals' access to and use of their own data. Data-stewardship infrastructures that include appropriate policies, robust technologies, and human governance must be developed and assessed to ensure alignment between individuals' decisions about their data and the practices of researchers and clinicians.

Training and genomic workforce development

Ensure that the next generation of genomic scientists is sufficiently trained in data science and that there are adequate numbers of data scientists trained in genomics. Appropriate skills and training in data science are now prerequisites for becoming a genomics researcher. Furthermore, given the ever-expanding use of genomics in basic, translational, and clinical research, a greater number of scientists will need fundamental data-science skills appropriate for the specific genomic applications being utilized. Establishing and maintaining data-science competencies for conducting genomics research will require a series of interrelated educational and training efforts, including the recruitment of a cadre of data scientists into genomics and the reciprocal exchange of expertise between genomics researchers and data scientists. While generic training approaches may be suitable in most cases, consideration and support will be needed for individual genomic researchers who require more specialized data-science skills.

Train healthcare providers to integrate genomics into the clinical workflow. Providers need to be equipped with current information to manage questions from patients who receive genomic information, including that from direct-to-consumer testing. The full spectrum of medical professionals should be included in this training effort. Education modules that are components of curricula tailored to specific user groups should be designed to adapt rapidly to advances in genomics and data science technologies; these should be available on demand and, where appropriate, integrated into existing clinical systems. Research is also needed, such as that focused on the educational methodologies for train-the-trainer approaches, the implementation of standards and competency-based education, and the strategies for enhancing genomic literacy among all healthcare providers. The involvement of patients, caregivers, educators, professional organizations, and accreditation boards will be critical to success.

Enhance the diversity of the genomics workforce. Developing the genomics workforce of the future requires intentional efforts to include women, underrepresented racial and ethnic groups, disadvantaged populations, and individuals with disabilities. For example, the U.S. genomics workforce – as with the general biomedical research workforce – does not reflect the diversity of the nation's population, which limits the opportunity for diverse minds to contribute to scientific and clinical innovation. New

strategies and programs to reduce barriers to career opportunities in genomics are required, as are innovative approaches to promote workforce diversity and inclusion practices. Diversity-enhancing efforts should not focus exclusively on early-stage recruitment; rather, they must also include incentives to recruit and retain a diverse workforce at all career stages.

Increase the genomic literacy of public educators to meet the expanding educational and workforce needs. An increase in the understanding of key genomic principles and applications will facilitate the use of genomics for advancing human health, promoting equity and diversity in genomics, and increasing the broad uptake of genomic medicine. The heterogeneous educational needs of the general public must be met by leveraging innovative tools that are shared, assessed, and improved as new strategies are developed. Academic institutions need to build scalable curricula for training the full range of public educators as well as provide basic education about the key ethical and social issues in genomics that will help the public and the genomic workforce navigate complicated genomic questions as they emerge.

III. Break Down Barriers in Genomics to Create New Research Opportunities

Since the inception of the field, genomics has benefited enormously from the proactive identification of major obstacles impeding progress and the subsequent focused efforts to break down those barriers. Prototypic successes include the call for a "$1000 Genome" in 2003 following completion of the Human Genome Project and a proposed set of actions to facilitate the early implementation of genomic medicine in 2011; in these cases, both the risks of failure and the benefits of success were high. Not unexpectedly, some prior barriers remain while some new ones have emerged, with some illustrative examples described below. Once again, breaking down these barriers – in particular those related to fundamental technologies and biological insights – would accelerate progress and create new genomics research opportunities at *The Forefront of Genomics*.

Fundamental technologies

Develop methods for generating and analyzing increasingly heterogeneous data types for studying human health and disease. Data types beyond primary DNA sequence (e.g., readouts of transcription levels, epigenetic markers, proteins, and metabolic markers as well as environmental and lifestyle data) are needed to fully understand the relevance of genomic variants in different biological contexts. Sustained efforts are needed for developing the technologies to generate these various data types in increasingly complex ways (e.g., from different cells in human tissues, at single-cell resolution, and for multiple data types simultaneously) in a comprehensive, scalable, and inexpensive fashion. In parallel, readily available analytic tools, including but not limited to those utilizing machine-learning and artificial-intelligence approaches, will become increasingly vital for assimilating, sharing, integrating, and analyzing these heterogeneous data types, including those in existing and new multi-omic datasets.

Enable the routine generation and use of synthetic nucleic acids in genomics research. The ability to generate nucleic acid molecules with defined sequences (i.e., synthetic genomics) would enable a wide range of genomic applications. Novel and innovative technologies are needed for longer, more accurate, faster, and cheaper nucleic acid synthesis on a large scale. Current limitations in our understanding of genome function hamper the ability to judiciously design synthetic nucleic acids for experimentation, but this situation will improve to create new opportunities. The use of synthetic genomics could also enable significant advances in genome editing, including the ability to make precise *in vivo* genomic changes, to characterize large genomic regions, and to test the effect of changes across large genomic distances. Improved systems for delivering large synthetic nucleic acid fragments into cells would also be important.

Biological insights

Establish the means to determine functional consequences of any genomic variant affecting human health and disease. Understanding the biological relevance of genomic variants, the genes they affect, their interplay with other genomic variants, and how they impact human health and disease is central to the application of genomics to human biology and medicine. The last decade has brought major leaps in the identification of genomic variants that are statistically associated with a given phenotype, but reliably establishing the mechanistic links connecting a specific variant to that phenotype remains challenging. Multiple complementary efforts are needed to overcome this barrier; these range from developing assays that model and test the effects of every possible variant in a genomic region to integrating multiple data types (e.g. association data, gene-expression information, model organism information, and protein localization/interaction data) for deducing influences of variation on genome function.

Identify and characterize somatic mutations and genetic mosaicism in human phenotypes. As progress is made in establishing the general consequences of inherited genomic variants, there is also the need to account for the reality that humans are mosaics with different variants in different cells (both somatic and germline). Despite its prevalence and the increasing familiarity of several well-studied examples, there is a paucity of detailed knowledge of mosaicism, such as how and when it contributes to human phenotypes. New methods for readily detecting genomic mosaicism at high spatial and temporal resolution are needed as are new experimental and clinical approaches for assaying for mosaicism, especially in ways that are relatively non-invasive (e.g., requiring minimal amounts of tissue).

Understand and leverage population structure and admixture to facilitate human genetics studies. Technological advances have enhanced the ability to generate, integrate, and analyze genomic and phenotypic data from highly diverse and recently admixed human cohorts. These large datasets can be leveraged for human genetics studies that address questions relevant to health and disease. In particular, the growing availability of high-quality, well-characterized genome-sequence data can be analyzed with new analytic tools to enable rapid progress in several areas. For example, examining genome-wide association signals in different populations can be used to more precisely map functional genomic elements (e.g., regulatory segments) and to reduce the size of genomic regions defined by genetic-association studies. In the case of the latter, such regions often contain many genomic variants, and identifying the exact one(s) conferring risk for the disease remains difficult. Analyzing the genomes of individuals from diverse ancestral origins can be particularly helpful for delimiting these regions; eventually pinpointing the precise genomic variant(s) responsible for disease risk can lead to a better understanding of the molecular mechanism underlying that disease. More routine and robust abilities to study diverse and recently admixed collections of healthy individuals and those with known diseases would both reveal the range of genomic variation present in our species and lead to a deeper understanding of the variants contributing to both rare and common diseases.

IV. Lead the Pursuit of Compelling Genomics Research Projects

Starting with the Human Genome Project, the field of genomics has routinely benefited from a willingness to articulate ambitious – at times even audacious – research efforts that aim to address questions and acquire knowledge that (at the time) may seem out of reach. Such boldness has often served to stimulate interest in emerging opportunities, recruit new expertise for addressing difficult problems, and propel the field forward. While by no means comprehensive, the areas described below illustrate the broadening range of compelling research projects that should be pursued at *The Forefront of Genomics*; each has meaningful potential for leading to the next set of significant and, at times, unexpected genomic advances.

Establish the roles and relationships of all genes and regulatory elements in pathways, networks, and phenotypes. The current and anticipated future technological and computational capabilities to analyze genomes provide an opportunity to decipher the integrated function of individual genomic components in an increasingly comprehensive way, including how genomic variants influence those functions. Because genes and regulatory elements do not function in isolation, the challenge of assessing their individual and combined biological roles and their resulting contributions to phenotype encompasses a nearly boundless experimental space. Nonetheless, that is the space that must be explored in an increasingly comprehensive fashion. In addition to using new methodologies for data generation, robust computational models that accurately predict cellular and organismal phenotypes from genome-sequence data need to be developed and applied in conjunction with enhanced visualization tools for understanding multidimensional information. A complete understanding of these genomic complexities must also account for functional redundancy and differences in biological activity in different physiological contexts, environmental conditions, and genetic backgrounds.

Elucidate the genetic architecture of all human diseases and traits. While progress in the last decade has illustrated the promise of genomics for unraveling the genetic underpinnings of human disease (to date most notably for rare diseases), the field is now poised to begin capturing a more complete understanding of the genetic architecture of all human diseases and traits. Systematic approaches will be needed for establishing the phenotypic consequences of all genomic variants – individually and in combinations – and this will undoubtedly require multifaceted experimental modalities (including perturbation assays, large-scale molecular phenotyping in cells and tissues, and studies of animal models). However, myriad complexities can be anticipated. For example, any given genomic variant or combination of variants may affect more than one disease or trait (i.e., pleiotropy). Similarly, variant effects can differ by orders of magnitude; can confer risk or be protective; can act additively, synergistically, and/or through intermediates; and can be mosaic in an individual. Multiple, sometimes thousands, of genes can contribute to a phenotype. The use of new methods for analyzing data from large studies of ancestrally diverse human participants, coupled with a growing knowledge of genotype-phenotype relationships, will enable the discovery of associations and interactions among genomic variants and environmental

factors as well as estimates of penetrance and expressivity. These efforts would be greatly enabled by redesigning the approaches used for assembling samples (e.g., DNA) for genomic studies of human disease. In addition to advancing the general knowledge about pathophysiology, more comprehensive insights about the genetic architecture of human diseases will enhance the clinical utility of genomic information for predicting risk, prognosis, treatment response, and, ultimately, clinical outcomes.

Determine the value of linking genome sequencing to the delivery of clinical care within healthcare systems. The evidence base documenting the value of genomic information in clinical care is currently in a nascent state, but new opportunities for significant growth are quickly being realized. For example, healthcare systems are emerging as effective testbeds for genomic medicine implementation, including the use of genome-sequence data. Providing genome sequencing to all members of a healthcare system, performed in conjunction with research and participant engagement, can establish the clinical utility of genomic information for a broad range of conditions and may allow providers to improve disease diagnosis and management, lead to better informed patient care, and enable more favorable patient and provider outcomes. As an example, understanding the needs for (and utilization of) genomic information by providers is critical, especially for those without extensive expertise in genomics (who actually deliver most clinical care). Opportunities to investigate how genomics-related interventions can reduce health disparities should also be prioritized. Studies within a single healthcare system allow for examination of the interactions among patients, providers, and laboratories, which, in turn, can be used to optimize approaches for effectively returning genomic information to both patients and providers. Extending such studies across multiple healthcare systems should reveal common challenges and solutions, thereby enhancing the learning healthcare model for genomic medicine more broadly. Such clarity would pave the way to conducting research in these learning healthcare environments, yielding more data at lower cost and encouraging the widespread use of genomic information.

Evaluate multi-omic approaches for the diagnosis and management of human disease. Tissue- and cell-specific data from multiple assays (i.e., multi-omics) can provide real-time snapshots of biological and disease processes, yielding insights beyond those that can be captured from currently available datasets. For example, integrating data about genomes, epigenomes, transcriptomes, proteomes, and metabolomes with clinical variables and outcomes should advance understanding of disease onset and progression, leading to improved predictive and prognostic models for a wide range of conditions. Such studies will require the integration of high-dimensional multi-omic data with CDS tools and EHRs to enable adoption and appropriate clinical use. Ultimately, these efforts could yield an understanding of the relationships among genomic, environmental, and behavioral variation and facilitate a transition from treating disease to maintaining health.

Closing Thoughts

To be drafted following receipt of feedback about draft document in April (and will include a reference to Box 3).

Box 3: Bold Predictions for Human Genomics by 2030

To be written following receipt of feedback about draft document in April.

- **Bold prediction 1** – *brief explanatory text.*
- **Bold prediction 2** – *brief explanatory text.*
- **Bold prediction 3** – *brief explanatory text.*
- **Bold prediction 4** – *brief explanatory text.*
- **Bold prediction 5** – *brief explanatory text.*
- **Total Number: To Be Determined**