Future Opportunities for Genome Sequencing… and Beyond

Welcome, Charge, and Context

Eric Green, M.D., Ph.D.
Director, NHGRI

National Human Genome Research Institute
Charting a course for genomic medicine from base pairs to bedside

Eric D. Green*, Mark S. Guyer* | National Human Genome Research Institute*

There has been much progress in genomics in the ten years since a draft sequence of the human genome was published. Opportunities for understanding health and disease are now unprecedented, as advances in genomics are harmonized to obtain clearer functional knowledge about the structure and function of the human genome and about the genetic contributions to human health and disease. Here we articulate a 2011 vision for the future of genomics research and describe the path forwards for an era of genomic medicine.

Since the launch of the Human Genome Project (HGP) in 2001 and the publication of a reference human genome sequence,3 genomic research has become a mainstay of biomedical research. The research community has driven through many issues and is evident in the broad range of scientific advances that the HGP has enabled, as shown in Fig. 1 (see online). Genomic research is central to understanding the contributions of genetics to human health and disease. For example, many genes have been identified that contribute to heart disease, Alzheimer's disease, cancer, and nearly every other major health disorder.

Although the human genome project has already begun to improve diagnostics and treatments in a few instances, profound improvements in the effectiveness of health care are not yet clinically apparent for many (Fig. 2). Achieving such dramatic improvements will depend not only on research, but also on new policies, practices, and other developments. We believe that we can see some signs of progress towards achieving these goals.

Understanding the biology of genomes

Substantial progress in understanding the human genome has revealed much about the complexity of genomic biology. Continued acquisition of basic knowledge about genome structure and function will be needed to illuminate further these complexities (Fig. 2). The contributions of genomic research will include comprehensive catalogs of human genetic variation and novel research tools, which will enable the capabilities of all researchers to reveal fundamental principles of biology.

Comprehensive catalogs of genomic data

Comprehensive genomic catalogs have been uniquely valuable and widely used. There is a compelling need to improve existing catalogs with: 1) genome variation, such as complete collections of genetic variation, functional genomic elements, microRNAs, proteins, and other biological molecules, for both human and model organisms; 2) comprehensive catalogs of genomic expression; and 3) comprehensive catalogs of chromatin structure.

Germologic studies of the genome and pathways associated with disturbed function require comprehensive catalogs of genetic variation, which provide critical markers for association studies and means for identifying candidate genes. Developing a detailed catalog of variation in the human genome has been an intensive effort that began with the SNP Consortium3 and the International HapMap Project (http://hapmap.ncbi.nlm.nih.gov), and is ongoing with the 1000 Genomes Project (http://www.1000genomes.org).

Over the past decade, these catalogs have been critical to the discovery of the specific genes for many major human diseases,

Figure 1. Genomic achievements since the Human Genome Project (see accompanying color).

*National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. E-mail: mark.guyer@nih.gov.

†Support for Eric D. Green and 50% of the work of Mark S. Guyer was provided by the National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 50% of the work of Mark S. Guyer was supported by the National Library of Medicine, National Institutes of Health, Bethesda, MD, USA.
NHGRI’s ‘Flagship’: The Genome Sequencing Program

1. History

2. Characteristics
NHGRI’s ‘Flagship’: The Genome Sequencing Program

1. History

2. Characteristics
Human Genome Project
1990-2003

“Large-Scale Genome Sequencing Centers”
2003-2006

Post-HGP Planning
~2002

Program Review Workshop
2005

Comparative Genomics
HapMap Project
Workshop on the Future of the Large-Scale Sequencing Program

June 13, 2005

Executive Summary

The National Human Genome Research Institute convened a workshop to obtain opinions from the scientific community on the current status and potential future directions of the NHGRI large-scale sequencing program. Participants were asked to consider the scientific, technological, and strategic opportunities in evaluating NHGRI's future investment in sequencing, and to specifically address several general questions and challenges:

- Given what has already been accomplished - very high quality assembled genome sequences of the human and major model organisms, draft sequence assemblies of genomes representing many of the nodes of the metazoan lineage, concerted application of comparative sequencing to annotate mammalian genomes - what are the best future opportunities for large-scale sequencing? What is the proper balance of these types of projects going forward? Should other kinds of large-scale sequencing projects be considered? What is the continuing priority of large-scale sequencing as a source of genomic data compared with other types of genomic data?

- Disruptive technologies appear to be promising enough that a significant reduction in the cost of DNA sequencing could occur within the next three years. What are the realistic prospects for the introduction of such a disruptive technology? How should it be anticipated and encouraged? How would it affect sequencing costs and capacity? How would it affect the types of scientific questions that can be addressed? How should the possibility of future significant cost reductions affect the decisions about the types of sequencing projects that should be initiated in the next two to three years?

- How should NHGRI evaluate the ongoing value of its investment in a large-scale sequencing program? How should it assess the contribution that continued sequencing will make to scientific research overall and genomic research in...
Human Genome Project
1990-2003

Post-HGP Planning
~2002

Program Review Workshop
2005

Program Review Workshop
2009

“Large-Scale Genome Sequencing Centers”
2003-2006

“Large-Scale Genome Sequencing Centers”
2006-2011

Comparative Genomics
HapMap Project

Microbiome
1000 Genomes

Cancer Genomics (TSP, TCGA)

Medical Sequencing
Pathogens & Vectors
Workshop Report
The Future of DNA Sequencing at the National Human Genome Research Institute
March 23-24, 2009

What are the most important biomedical questions that can be addressed with large-scale sequence data? What are the most compelling sequence-based community resources that should be generated? What are the consequences of the rapid increase in sequencing capacity, and the rapid decrease in cost, afforded by the new technology platforms? In order to answer these questions, the National Human Genome Research Institute (NHGRI) convened a workshop to discuss the future of large-scale sequencing as one component of the Institute’s current two-year planning process for all of its scientific programs.

The need for this workshop was particularly underscored by the recent and ongoing rapid changes in sequencing technology, propelled by the “next generation” sequencing platforms. Introduced into production activities less than two years ago, the new sequencing platforms have already afforded an increase in throughput of two orders of magnitude over the previous platforms, and this is likely to increase by nearly another order of magnitude in the next year or two. Furthermore, yet newer technologies are being developed and are expected to be available in the next three to five years. These rapid changes offer incredible new opportunities as well as major new challenges for the use of sequencing technology in general and to NHGRI’s sequencing program specifically. As the technology continues to improve, new applications of genomic sequencing are constantly being developed, for example the sequencing of genomes from large numbers of individuals for disease and population studies, quantitative transcriptional analysis and epigenomics.

The ‘disruptive’ technological change has many other consequences. Most obviously, the ability to apply large-scale sequencing efficiently towards a larger number of problems will result in unprecedented demands on scientists’ ability to find enough samples that are appropriate to addressing an expanded range of questions. To date, the most difficult problem has been obtaining samples for human disease or population studies that are properly consented for the work. One can also foresee
Human Genome Project
1990-2003

“Large-Scale Genome Sequencing Centers”
2003-2006

“Large-Scale Genome Sequencing Centers”
2006-2011

4-Component “Genome Sequencing Program”
2012-2015

Post-HGP Planning
~2002

Program Review Workshop
2005

Program Review Workshop
2009

2003

2006

2012

2016

Comparative Genomics
HapMap Project

Microbiome
1000 Genomes

Cancer Genomics (TSP, TCGA)
Medical Sequencing
Pathogens & Vectors
NHGRI Genome Sequencing Program
Circa 2012-2015

Large-Scale Genome Sequencing and Analysis Centers

Centers for Mendelian Genomics

Clinical Sequencing Exploratory Research

iSEQTOOLS
Plummeting Cost of Genome Sequencing

Cost per Genome

Moore's Law

NIH National Human Genome Research Institute

gene.gov/sequencingcosts
Human Genome Project 1990-2003

“Large-Scale Genome Sequencing Centers” 2003-2006

“Large-Scale Genome Sequencing Centers” 2006-2011

4-Component “Genome Sequencing Program” 2012-2015

To Be Determined 2016-???

Comparative Genomics

HapMap Project

Microbiome

1000 Genomes

Cancer Genomics (TSP, TCGA)

Medical Sequencing

Pathogens & Vectors

Common Diseases ???

Mendelian Diseases ???

Clinical Genomics ???

Computational Tools ???
NHGRI’s ‘Flagship’: The Genome Sequencing Program

1. History

2. Characteristics
NHGRI Genome Sequencing Program: Characteristics (To Date)

Large (i.e., Scale)
Consortia-oriented
Highly managed
Resource-generating
‘Technology’-advancing
Scientifically/medically relevant
Nimble
Going Forward: What Does NHGRI Want?

- Continue being ‘genomics trailblazers’
- Alignment with strategic vision/plan
- Impact that correlates with program size
- If continuation of a major program, then retain 7 characteristics (previous slide)
- Importance of ‘moving on’ past initial catalytic role (e.g., organism sequencing, microbes, microbiome, and cancer)
- Increased ‘cost-sharing’ to broaden impact
NHGRI and Cost-Sharing

- Genomics = Huge; NHGRI = Small
- NHGRI cannot support ‘everything genomics’
- Partnerships are key (past and future)
- Consider formalizing an approach for cost-sharing in large-scale NHGRI-funded genomics projects
Purpose of Workshop

1. For Starters: It’s what we do...
2. General: Natural time for strategic input (e.g., 3 years since 2011 strategic plan)
3. Critical: Synchronize strategic thinking in light of rapidly changing (and complicated) landscape
4. Practical: Fiscal Year 2016 and ~$100M (~25% of extramural funds)
Highest Priority: Discussion about Areas Associated with:

Large-Scale Genome Sequencing and Analysis Centers

Also Important: Discussion about Areas Associated with:

Centers for Mendelian Genomics

Clinical Sequencing Exploratory Research

iSEQTOOLS
Questions to Address (Among Many)

1. What are the ‘grand opportunities’ appropriate for a ‘flagship’ NHGRI program(s)?

2. What is NHGRI not doing that it should be doing?

3. How to balance ‘democratization’ of genome sequencing and benefits of consortia-based, large-scale pursuits?

4. How to properly tune a ‘flagship’ program’s(s’) funding level with its impact?

5. Should NHGRI develop a formal cost-sharing approach for large non-generic (e.g., disease-specific) projects?

6. How should NHGRI more efficiently obtain ‘commodity’ genome sequencing to meet programmatic needs?
Going Forward (Quickly)…

TOUGH DECISION AHEAD
There are Likely Too Many Good Options
Advancing human health through genomics research