

# **Lessons from Other Genomics Projects**

**Inaugural Meeting of the  
H3Africa Consortium**

**Mark Guyer**

**October 9, 2012**

# **Bottom Line**

## **(according to me)**

- 1. Rapid data and materials sharing is a high-level “good” and a primary characteristic of genomics**
- 2. Policies for implementing effective data release and data sharing are situational**
  - a. Dependent on the nature of the data to be released**
  - b. Has to accommodate other high-level “goods”**
    - 1) Human subjects concerns**
    - 2) Investigator needs**

# 1988

**"... access to all sequences and material generated by these publicly funded projects should and even must be made freely available**

**..."** -- National Research Council Report on *Mapping and Sequencing the Human Genome*

**1993**

**NIH-DOE Guidelines for Access to Mapping and Sequencing Data and Material Resources**

**“...to promote and encourage the rapid sharing of material and data...especially information that has not yet been published or may never be published in its entirety.”**

# 1993

## NIH-DOE Guidelines for Access to Mapping and Sequencing Data and Material Resources

**“...to promote and encourage the rapid sharing of material and data...especially information that has not yet been published or may never be published in its entirety.”**

# 1993

## NIH-DOE Guidelines for Access to Mapping and Sequencing Data and Material Resources

**“...to promote and encourage the rapid sharing of material and data...especially information that has not yet been published or may never be published in its entirety.”**

**“...maximize outreach to the scientific community [but] also necessary to give investigators time to verify the accuracy of their data and to gain some scientific advantage from the effort they have invested.”**

# 1993

## NIH-DOE Guidelines for Access to Mapping and Sequencing Data and Material Resources

**“...to promote and encourage the rapid sharing of material and data...especially information that has not yet been published or may never be published in its entirety.”**

**“...maximize outreach to the scientific community [but] also necessary to give investigators time to verify the accuracy of their data and to gain some scientific advantage from the effort they have invested.”**

# 1996

## International Large-Scale Sequencing Meeting Hamilton, Bermuda

- “All human genome sequence data generated by centers funded for large-scale human sequencing should be freely available and in the public domain to encourage research and development and to maximize the benefit to society.”

# 1996

## International Large-Scale Sequencing Meeting Hamilton, Bermuda

- “All human genome sequence data generated by centers funded for large-scale human sequencing should be freely available and in the public domain to encourage research and development and to maximize the benefit to society.”

# 1996

## International Large-Scale Sequencing Meeting Hamilton, Bermuda

- “All human genome sequence data generated by centers funded for large-scale human sequencing should be freely available and in the public domain to encourage research and development and to maximize the benefit to society.”

# 1996

## International Large-Scale Sequencing Meeting Hamilton, Bermuda

- “All human genome sequence data generated by centers funded for large-scale human sequencing should be freely available and in the public domain to encourage research and development and to maximize the benefit to society.”
- “These principles should apply to all human genomic sequences generated by public large-scale sequencing centers to avoid having such centers establish a privileged position in exploitation and control of human sequence information.”

# 1996

## International Large-Scale Sequencing Meeting Hamilton, Bermuda

- **“All human genome sequence data generated by centers funded for large-scale human sequencing should be freely available and in the public domain to encourage research and development and to maximize the benefit to society.”**
- **“These principles should apply to all human genomic sequences generated by public large-scale sequencing centers to avoid having such centers establish a privileged position in exploitation and control of human sequence information.”**

# 1996

## International Large-Scale Sequencing Meeting Hamilton, Bermuda

- “Sequence assemblies should be released as soon as possible; in some centers, assemblies larger than 1 kb would be released automatically on a daily basis.”
- “Finished annotated sequence should be submitted immediately to public databases.”

**1997**

**NHGRI Policy on Release  
of Human Genome Sequence Data**

**No...[award]... will be funded until an  
acceptable plan has been approved.**

# Scientific Developments necessitate changes in data release policies

Late 1990's –large clone sequences (e.g. BACs) and whole genome shotgun data sets typically not assembled until late in a project → tying data release to assembly would have **delayed** rapid access to the underlying data set

1998 SNPs. Data needed to be verified to avoid release of false positives

Requirement to release sequence traces, prior to assembly, weekly

90 day period allowed for quality control prior to release

# Scientific Developments necessitate changes in data release policies

**Early 2000's – large whole genome assemblies**

**Request users to voluntarily refrain from publication of “whole genome” analyses until after first publication by data producers**

**Mid-2000's – GWAS; sequencing human genomes from phenotyped donors**

**Release to controlled access databases (dbGaP)**

**Mid 2000's – generation of functional genomics data**

**Voluntary publication moratorium to allow data producers to publish**

# Continuous evolution/refinement of data release policies

- **Public meetings**
  - 2003 Fort Lauderdale
  - 2009 Toronto
- **Project-specific policies**
  - HapMap/1000 Genomes
  - ENCODE
  - The Cancer Genome Atlas (TCGA)
  - eMERGE



# H3Africa Data Release Policies

## Balance a number of important considerations

- **“Pro” rapid data release**
  - H3Africa seen as operating by internationally recognized norms of genomics research
  - Production and release of high-quality, important data will contribute to the global recognition of African researchers
- **“Con” rapid data release**
  - H3Africa investigators need time to extract value from their own data; concerns about being scooped.