

NHGRI Informatics and Analysis Planning Meeting

April 21-22, 2010

Draft Agenda

January 22, 2010

Goals: As part of the NHGRI planning process on the future of genomics, NHGRI is considering the informatics and analysis needs for producing, distributing, and using large genomic datasets. The goal of this meeting is to define the key issues and paths to address them.

Wednesday, April 21

8:30 – 9:00 Breakfast

9:00 – 9:20 **The NHGRI planning process on the future of genomics** Eric Green
(including the goals and outputs of this meeting, and that this meeting will not focus on databases or training)

The future of genomic data: (each 10 min talk, 10 min disc)
The questions that will be addressed, the types of data that will be produced, and the size of the datasets (including basic and clinical research)

9:20 – 9:40 **Sequence and functional elements**

9:40 – 10:00 **Comparative genomics**

10:00 – 10:20 **Cancer genomics**

10:20 – 10:40 **Common-disease genomics**

10:40 – 11:00 Break

Infrastructure for computing hardware and software

Ever-increasing volumes of data will need to be managed, processed, and integrated. These vast data volumes exceed the capacity of the computing systems of many users.

11:00 – 11:15 **Summary of the cloud computing meeting** (10 talk, 5 disc)

11:15 – 12:30 **The current situation and future needs for genomic computing:** (15 talk, 60 disc)

What is already occurring and what will be needed for distributing and computing on large amounts of data?

12:30 – 1:00 **Security:** (10 talk, 20 disc)

Although many genomic data types should be released rapidly and publicly, some data types, such as those on human subjects, require data access restrictions and high security. How do human subjects concerns, limited access, and security affect computing infrastructure needs?

1:00 – 2:00 Lunch

2:00 – 2:45 **Path:** What obstacles are hindering progress, how can they be overcome, and what is needed to support progress?

Data processing and scientific analysis

Most data types require some processing to produce, by taking raw data from machines, extracting derived data such as assemblies, alignments, expression levels, and variants, and assessing data quality. These processed data are then analyzed to address scientific questions.

- 2:45 – 3:45 **Central vs. distributed:** (15 talk, 45 disc)
What conditions are best for central, regional, or distributed data storage, processing of raw data, analysis of processed data, and high-level scientific analyses? When are project Data Analysis Centers needed, and what roles should they play?
- 3:45 – 4:00 Break
- 4:00 – 5:00 **Integration:** (15 talk, 45 disc)
What is needed to support integration and analysis across various types of data?
- 5:00 – 5:45 **Path:** What obstacles are hindering progress, how can they be overcome, and what is needed to support progress?
- 5:45 – 7:00 Dinner

Thursday, April 22

- 8:30 – 9:00 Breakfast

Data tools

Many tools are needed to deal with the various data types and accommodate different levels of user expertise. Tools are developed by data-production centers, by Data Analysis Centers, by central databases, and by small groups.

- 9:00 – 9:45 **Development** (15 talk, 30 disc)
How can tool development be facilitated?
- 9:45 – 10:30 **Maintenance** (15 talk, 30 disc)
What is needed to support the maintenance and improvement of tools? What criteria should be used to decide which tools merit continued support?
- 10:30 – 10:45 Break
- 10:45 – 11:15 **Path:** What obstacles are hindering progress, how can they be overcome, and what is needed to support progress?
- 11:15 – 12:00 Another topic?
- 12:00 – 1:00 Lunch

Summary and next steps

- 1:00 – 3:00 **What are the major issues and steps to take?**