Internet-Based Support for a Collaborative Genome Center to Map Human Chromosome 12

Perry Miller, M.D., Ph.D.[1], Prakash Nadkarni, M.D.[1], Amy Banks[1], Patricia Bray-Ward, Ph.D.[1]
Kei Cheung[1], Lisa Cupelli, Ph.D.[2], Valerie Herdman[2], Kenneth Kidd, Ph.D.[1]
Kenneth Krauter, Ph.D.[2], Ivonne Marondel, Ph.D.[2], Kate Montgomery, Ph.D.[2]
Beatrice Renault, Ph.D.[2], David C. Ward, Ph.D.[1], Sung-Joo Yoon, Ph.D.[2], Raju Kucherlapati, Ph.D.[2]
[1]Yale University School of Medicine, New Haven, CT
[2]Albert Einstein College of Medicine, Bronx, NY

We will demonstrate a shared database (DB/12) and a set of software tools developed for a genome center to map human chromosome 12. The approach involves an internet-based client-server architecture to support collaborating laboratories at the Albert Einstein College of Medicine and at Yale.

INTRODUCTION

We will demonstrate an approach developed to provide internet-based informatics support for a genome center[1] to map human chromosome 12, in a collaboration between the Albert Einstein College of Medicine (AECOM) and Yale.

Our chromosome mapping strategy uses yeast artificial chromosomes (YACs), which contain segments of human DNA growing in yeast. The YACs were produced at the Centre d'Etude du Polymorphisme Humain (CEPH) in Paris, France. Each YAC contains roughly 400 Kbp to 1.2 Mbp of human DNA. The CEPH YAC library contains thousands of overlapping YACs.

The main mapping activity involves taking known chromosomal markers from human chromosome 12 (as well as newly created markers) and experimentally determining which markers are found on each of a set of overlapping YACs. Since the various overlapping YACs begin and end at different places on chromosome 12, each will contain a different set of markers. Once enough data is produced for a set of overlapping YACs, one can start to infer the order of the markers along the chromosome, and the pattern of overlap of the YACs. (Such a set of overlapping YACs is called a "contig.")

A SHARED DATABASE

DB/12, a shared database containing the experimental mapping data, is implemented using Sybase on a Sun SPARCstation-2. It is accessed from Yale and AECOM by Macintosh computers running the 4th Dimension (4D) database as a user-friendly front-end. The client-server interaction is mediated by SQLGEN, which allows SQL to be constructed on-the-fly from a tabular description of DB/12's structure. Users interact with DB/12 either through 4D or through a World Wide Web (WWW) interface, e.g., MOSAIC.

MAP MANIPULATION

4D routines for map manipulation allow the display of contig maps (which can also be imported into a drawing program for editing). Routines also let the user view alternative contig maps consistent with the data, ranked by "goodness of fit." This contig analysis routine allows considerable flexibility. For example, the user may specify a fixed order for certain markers, and request that the relative positions of other markers be automatically fit to the data.

EXTERNAL RESOURCES

We are also in the process of integrating external software tools and resources into the overall framework. For example, the 4D front-end can use SQLGEN to access the national Genome Data Base (GDB) in Baltimore, in addition to DB/12. This capability allows a single user query to retrieve information from both DB/12 and GDB in an integrated fashion.

Acknowledgements
This work was supported in part by NIH Grants R01 HG00175 and P01 HG00965 from the National Center for Human Genome Research, by NIH Grant G08 LM05583 from the National Library of Medicine, and by an equipment grant from the Sun Microsystems Corporation.

Reference

1. Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years FY 1991-1995. DOE/ER-0452P. Springfield, VA: National Technical Information Service, U.S. Department of Commerce, 1990.

0195-4210/95/$5.00 © 1995 AMIA, Inc.