MANAGING DATA FOR GENETIC LINKAGE ANALYSIS

To the Editor: Genetic linkage analysis is one of the oldest and most powerful tools for gene mapping. Recombinant DNA technology has heralded an explosion in activity in this area. Whereas the number of polymorphic genetic markers detected in red cell antigens and serum proteins is less than 40, every 2 years since 1979 there has been approximately a sixfold increase in the total number of DNA restriction fragment length polymorphisms [1]. The rapid discovery of new DNA markers for linkage and development of alternative strategies for analysis of linkage data has exacerbated the already tedious task of storing and managing data. Because such data are usually manipulated by hand or with the assistance of a computer text editor, data coding, recoding, and formatting are not trivial tasks. In fact, data manipulation has become one of the most frustrating and time-consuming aspects of modern linkage studies.

At the recent Genetic Analysis IV Pre-Workshop Conference held at Snowbird, Utah, on October 7–9, 1985, participants presented and compared results from genetic analyses of very large data sets. Many members of the conference recognized data coding and transcription errors as critical obstacles during the course of their analyses. In addition, several investigators mentioned that the proportion of time devoted to data management impeded them from tackling more than one data set.

A data management system that is useful for linkage analysis must be flexible enough to allow for multiple hypothesis testing, subsetting of data, and conversion of data between linkage analysis programs. User friendliness and some degree of data error checking are other desirable features. IMAGENARI (Interactive MAnagement of data for GENetic linkage Analysis and Related Investigations) is a series of three standard Pascal data management programs designed to interface two of the most widely used computer programs for linkage analysis, LIPED [2] and LINKAGE [3]. These programs feature interactive formatted entry of pedigree data (INTERACT), extraction of selected pedigree records and phenotype fields from a large pedigree data file (EXTRACT), and interconversion of pedigree data between LIPED and LINKAGE (CONVERT).

Although a well-designed database management system may more efficiently meet the needs of some genetics software users [4], very few computer programming skills are required for IMAGENARI use, other than familiarity with the data input schemes for LIPED and LINKAGE. Each of the IMAGENARI programs is menu driven and can be operated in many mainframe, minicomputer, and 16-bit microcomputer environments. For a free copy of programs and documentation, send a blank 5.25-inch diskette (compatible with MS or PC DOS, version 2.1) to Dr. Lindsay A. Farrer, Department of Human Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510. Please specify whether Turbo, Microsoft, or DEC VMS Pascal is desired.

LINDSAY A. FARRER[1]

Received February 7, 1986.

[1] Department of Human Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510.

REFERENCES

1. WILLARD HF, SKOLNICK MH, PEARSON PL, MANDEL J-L: Eighth International Workshop on Human Gene Mapping. *Cytogenet Cell Genet* 40:360–489, 1985
2. OTT J; Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am J Hum Genet* 26:773–775, 1974
3. LATHROP GM, LALOUEL J-M, JULIER C, OTT J: Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81:3443–3446, 1984
4. FARRER LA, HAINES JL, YOUNT EA: Automating data manipulation for genetic analysis using a data base management system. *Hum Hered* 35:296–301, 1985

LIPIN: AN INTERACTIVE DATA ENTRY AND MANAGEMENT PROGRAM FOR LIPED

To the Editor: In 1974, Ott [1] introduced the LIPED computer program utilizing the maximum likelihood method of linkage analysis for the calculation of the recombination fractions between two loci in extended and/or complex pedigrees. LIPED has been widely used and allows the use of large pedigrees in linkage analysis. However, recent advances in molecular genetic technology have complicated the linkage analysis process by creating hundreds of useful marker loci. Problems now arise in simply managing the large amounts of data available.

A further problem arises in that LIPED, although requiring a simple pedigree structure and coding scheme for phenotype data, requires strict adherence to FORTRAN formatting requirements. A data file containing all the pedigree, locus, and FORTRAN formatting information must be created from scratch. These files also must be meticulous in the column spacing of all variables involved. A single value out of place can cause the program to function improperly. To correctly format large amounts of information into this data file is a tedious venture and concern about setting up correct formatting cards slows the process even more. Additionally, once set, the data file is a static entity and addition or deletion of information may involve hours of tedious editing using the generally cumbersome line editors provided by most computer operating systems.

We have written an interactive data entry and management program called LIPIN, which will produce a functional, nonstatic data file for use with LIPED. LIPIN is a program written in PASCAL that manages all data formatting, thus freeing the LIPED user from such worry and hassle. A data file created by LIPIN also can be reread by the program and then modified. These modifications can entail the deletion or addition of loci and/or pedigrees or simply the modification of existing data. Also, subsets of marker loci and/or pedigrees can be extracted and new files written. These new files are readily usable by widely available data conversion programs to create files compatible with the multipoint linkage analysis package LINKAGE [2].

This program contains many internal data checks that will reduce errors of data entry. We have found this program efficient and easy to use and believe it can streamline linkage data management and analysis. LIPIN is currently avail-