Nucleic Acids Research, Vol. 20, Supplement 2201–2206

The GDB™ Human Genome Data Base Anno 1992

P.L.Pearson, N.W.Matheson[1], D.C.Flescher[1] and R.J.Robbins[1]
Department of Medicine and [1]Welch Medical Library, Johns Hopkins University School of Medicine, Baltimore, MD 21205-2100, USA

INTRODUCTION

The GDB™ Human Genome Data Base, developed under funding from the Howard Hughes Medical Institute (HHMI), made its debut as an integrated environment for editors to enter new data at HGM 10.5 in Oxford, England in September, 1990. At that meeting the GDB was declared the official database to support the development, maintenance, and distribution of a consensus view of human gene mapping information. GDB Release 1 became a public database available online for use by the scientific community on September 15, 1990. (1)

The overall goals of the GDB have remained constant since its inception. The first goal is to support the international research effort of the Human Genome Initiative by creating a system and an environment that supports timely, collaborative data entry and evaluation for chromosome editors worldwide. The second goal is to provide an interactive environment for simple and easy accessibility to the information by the scientific community internationally through communication networks and strategically located service nodes. Towards this goal, full versions of GDB are available at the British Medical Research Council in the U.K. and the German Cancer Research Center in Heidelberg under the auspices of the EEC and the EMBL laboratories. Additional access and service nodes are under development in Australia and Japan. As the network technology matures the goal of interactive distributed editing of GDB will become a reality.

With the establishment of the U.S. Human Genome Project in 1991, and the subsequent withdrawal of HHMI involvement, the GDB successfully competed for funding through September, 1994 from this source. While the GDB is jointly funded by two U.S. agencies, the Department of Energy and the National Institutes of Health, the international orientation of the GDB has not altered. To ensure an international perspective and eventual international funding support, GDB is guided by a well-defined group of external advisory committees.

The Inter-Agency Coordinating Committee (IACC) is comprised of representatives from all agencies who may contribute to the funding of the GDB. At present, its membership includes representatives from agencies in the following countries (in alphabetic order): Canada, the EEC, France, Germany, Italy, Japan, the Netherlands, Sweden, the U.K., and the U.S. A second body is the International Scientific Advisory Council (ISAC), the composition of which is given in Table 1. The Council will advise the GDB Scientific Director and the IACC on the scientific priorities to be established. Its composition will reflect the scientific interests and aims of the GDB user community in particular and of the Human Genome Project in general. In addition, the formation of a number of ad hoc advisory groups, including an Informatics Working Group and a Mapping Working Group, are anticipated.

PRESENT SCOPE AND FUNCTION OF GDB

GDB is used as a collective term for two databases, the GDB Human Genome Data Base and OMIM (Online Mendelian Inheritance in Man). GDB is a relational database incorporating four key categories of data: map objects, map locations, genetic disease and locus descriptions, and bibliographic sources. Map objects include genes, DNA segments and fragile sites, chromosome breakpoints, restriction sites, meiotic cross overs, and partial maps. OMIM is Victor A. McKusick's comprehensive catalog of inherited disorders and includes genetic disease descriptions, bibliographic sources, clinical signs and symptoms. Table 2 gives a breakdown of the information stored in GDB as of March 1, 1992 compared to the information available a year earlier. Table 3 shows the number and activity of registered users as of February, 1992.

MECHANISMS OF DATA ACQUISITION AND EDITING

Data is entered into GDB either from the literature, by direct submission from researchers and centers, or by individual editoGDB employs a central support group that is charged with assisting in the data-entry process. Researchers may submit data electronically using standard formats or manually using standard paper forms.

Nearly 100 editors assist in the management of GDB by providing editorial oversight on information regarding specific chromosomes or on special topics such as nomenclature, DNA, comparative mapping, neoplasia and clinical disorders. The editors are distributed world-wide, with approximately 50% coming from the U.S.. Table 4 lists the editors and their duties. Researchers are encouraged to communicate concerns regarding editorial topics directly to the relevant editors. A messaging system built into the GDB allows users logged on to the system to send messages directly to editors.

Editors work on the database by logging on to the system in Baltimore, which is available for use 23 hours per day, every day of the year. At present, editorial access (i.e., write-only access to the database) is only available at the Baltimore site. Plans are being made to extend editorial access to other sites (see below), perhaps by late 1993.

DISTRIBUTION MECHANISMS

To date, a number of strategies have been employed to make GDB data readily available to the international scientific community. Online access to the interactive database is available via either network or telephone connections. Surveys indicate that the majority of registered GDB users connect to the GDB via modem and use only simple terminal emulation software to

interact with the database. To meet the needs of these users, interactive access to the GDB was originally developed with only a text-based, VT100 interface. The next release of the GDB (scheduled for late summer, 1992) will extend the GDB interface to include full interactive support (including for the use of a mouse) for networked computers running X-windows.

Currently there are three basic ways that users may access the information in the GDB: 1) accessing the GDB information present in the relational database in Baltimore; 2) accessing the GDB information in one of a growing number of GDB clones (known as 'official nodes') at other locations around the world; 3) downloading GDB data from an anonymous ftp server maintained in Baltimore. The scientific data contained in the GDB and OMIM databases are available for ftp access on this server.

1. Accessing GDB in Baltimore

Online access to the GDB in Baltimore is available using either via a modem-based dialup connection through Sprintnet or via the Internet. Recently, Sprintnet has added 9600-baud service in a number of cities in North America and elsewhere. Users should contact both Sprintnet and the GDB to determine what level of connectivity is potentially available between their site and the GDB in Baltimore. In practice, 9600-baud modem connections may be as fast as a direct network connection, particularly when the networks are extensively used.

A login name and password, supplied by the GDB Product Service group, are necessary to gain access to the data. These are available without charge upon request. Effective use can be made of GDB with any computer system capable of connecting with GDB via modem and of providing VT100 terminal emulation. However, superior ease of use and functionality are available to users with networked Sun workstations that connect to GDB via the Internet. To gain this increased functionality, users must obtain a copy of the GDB front-end software and mount it on their workstation. Because this front-end software contains some licensed Sybase products, appropriate licenses must be obtained in order to mount this software on a workstation (see below). With the availability of X-Windows support in the next release of GDB, equivalent functionality (mouse point-and-click interface) will be available to all users with networked computers capable of supporting an X-Windows session. No Sybase licenses are required to access the system using X-Windows.

2. Accessing GDB at other sites

The limited penetration of network technologies worldwide is a serious problem to many scientists. In some cases, network connections may be availaare expensive, difficult to use, and unreliable. To overcome these barriers, GDB nodes are being established at other institutions in other countries.

The same level of read-only accessibility as that accorded to users of GDB Baltimore is also maintained at GDB 'official nodes', which currently include sites in London, Heidelberg, Sydney, and in the near future Tokyo followed by Uppsala. These sites have been established to ease the burden of accessing GDB information over the international networks. The information at the official nodes is updated weekly from the GDB database in Baltimore. Investigators in countries where GDB nodes are established should contact their local organization to become registered as users of the node and to obtain help on either the data or use of the system. A contact address list for current nodes is provided in the section on user information.

A successful network of distributed nodes will depend on uniformity of quality and standards of service. The scientists of the world should never be in a position of having to shop for the 'best' node. For these reasons potential nodes must meet certain criteria which have been established jointly by the GDB funding agencies and GDB scientific advisors. A potential node must have the support and endorsement of its own scientific community for, except for the data, the cost of the operation must be borne entirely by the node's institution or community. A potential node must exhibit demonstrated technical computational competence. It must be willing and able to provide educational and service support at a level equal to that available from GDB in Baltimore. All nodes operate under a formal agreement with JHU, which defines mutual responsibilities and obligations.

3. Anonymous FTP (file transfer protocol)

GDB data and other information is available as a number of files for downloading from a general access computer maintained at JHU in Baltimore. These include GDB and OMIM data files as tab-delimited ASCII tables, forms for submitting data to GDB, and GDB schema files and user documentation in PostScript format. These files are updated weekly. In the future, standard report files containing results from a series of predefined searches such as all mapped genes on particular chromosomes or all polymorphic markers fulfilling particular selection criteria will be made available. Users are encouraged to submit suggestions for standard reports to GDB staff. Organizations that wish to may arrange for routine formatted tape distribution on a cost-recovery basis from GDB. Other than providing access to its files and Sybase software, GDB provides neither technical nor user support to organizations who wish to redistribute GDB.

Institutions are free to redistribute GDB and OMIM to their users under whatever software interface they choose. GDB is built on Sybase software. To facilitate the use of GDB files, GDB is an authorized Sybase VAR distributor and can provide front-end Sybase software at a discount. Users who wish to acquire a copy of OMIM must first obtain a no-fee license for IRX software from the National Library of Medicine. GDB can facilitate this process.

As stipulated in the current GDB funding agreement, commercial redistributors of GDB and OMIM must negotiate royalty agreements with GDB. Any royalties that accrue from these agreements will used to fund projects proposed by the scientific community that enhance and improve GDB.

USERS HELP TOOLS

To date, all GDB services (access to the system and to user help, written documentation, diskettes containing access software for PCs and Macintoshes, etc.) have been provided to the user community without charge. Recognizing that these costs will grow as the user community grows, the GDB review committee and the funding agencies have urged the GDB to find ways to limit these expanding costs. We expect, over the next year, to develop online help tools that will reduce, if not eliminate, the need for individual manuals and other written documentation. Users who prefer offline help will still be able to obtain it at a modest charge. This will substantially reduce the overhead cost to the sm, and the majority of users will find their use of the system greatly simplified and improved.

USER INFORMATION

Potential users can obtain information for accessing the database files from:

GDB Human Genome Data Base
William H. Welch Medical Library
1830 E. Monument Street
Baltimore, MD 21205, USA

General Information: (410) 955-9705
User Support: (410) 955-7058
Fax: (410) 955-0054
E-mail: help@welch.jhu.edu

Anonymous Access to FTP Server
FTP Server Address: mendel.welch.jhu.edu (128.220.59.42)

The Human Genome Program Resource Center (HGMP-RC)
Clinical Research Center
Watford Road
Harrow, Middlesex HA1 3UJ
England

Telephone: 44-81-869-3446
Fax: 44-81-869-3807
E-mail: cbates@uk.ac.crc

Deutsches Krebsforschungszentrum (DKFZ)
German Cancer Research Center (DKFZ)
Molecular Biophysics Group
Im Neuenheimer Feld 280
D-6900 Heidelberg 1
FRG

Telephone: 49-6221-42-2372
Fax: 49-6221-40-1271
E-mail: dok261@cvx12.dkfz-heidelberg.de

Australian National Genomic Information Service (ANGIS)
Electrical Engineering Building, J03
The University of Sydney
Sydney, N.S.W. 2006
Australia

Telephone: 61-2692-2948
Fax: 61-2692-3847
E-mail: reisner@ee.su.oz.au

1. Pearson PL The genome data base (GDB) a human gene mapping repository. Nucleic Acid Research 1991:19 Supplement 2237–9.

Table 1. International Scientific Advisory Council

Table 2. GDB Data Statistics

	Mar 1, 1991	Mar 6, 1992	% change
Locus			
Total Genes:	2,217	3,029	37
Total D-segments	na	9,435	
Mapped Genes	1,883	2,332	24
Mapped D-segments	5,369	7,195	34
Mapped fragile sites	113	113	
Mapped Breakpoints	na	53	
Map sets	na	254	
Total Mapped loci	7,365	9,947	35
Total loci		12,884	
Disease loci & mendelian phenotypes (OMIM)			
Total	5,248		
Probes			
PCR	519	1,426	175
ASO	432	434	
Clones	14,032	20,399	45
Total probes	14,983	22,259	49
Polymorphisms			
Polymorphic genes	521	653	25
Polymorphic D-segments	2,145	2,929	37
Total Polymorphisms	4,435	5,898	33
Sources (References)			
Journal articles	15,467	19,831	28
Personal communications	5,508	5,910	7
Books	12	24	100
Theses	1	1	
Total references	20,988	25,766	23
People:			
Total probe contacts	1,922	2,562	33

Table 3. GDB User Statistics February 1992

Registered Users	
Johns Hopkins University (USA)	3800
Medical Research Council (UK)	689
Deutsches Krebsforschungszentrum (FRG)	76
Total Users	4565
Average Number of GDB Sessions Per Month	
Johns Hopkins University (USA)	2400
Medical Research Council (UK)	400
Deutsches Krebsforschungszentrum (FRG)	80
Total Average Number of GDB Sessions Per Month	2880
Average Session Connect Time	50 minutes

Table 4. Editors

Chromosome 1

Nicholas C. Dracopoli
Massachusetts Institute of Technology
Center for Cancer Research
E17-541
77 Massachusetts Ave.
Cambridge, MA 02139
USA

Gail A. Bruns (Senior Editor)
Children's Hospital Medical Center
Department of Clinical Genetics
300 Longwood Ave.
Boston, MA 02115
USA

Chromosome 2

Andrew J. Pakstis
Yale University School of Medicine
Department of Human Genetics
1-309 SHM
333 Cedar St.
New Haven, CT 06510
USA

Nigel K. Spurr (Senior Editor)
Imperial Cancer Research Fund
Clare Hall Laboratories
Blanche Lane
South Mimms
Potters Bar, Hertfordshire EN6 3LD
ENG

Chromosome 3

Yusuke Nakamura
Japanese Foundation for Cancer Research
Division of Biochemistry
1-37-1 Kami-ikebukro
Toshima-ku
Tokyo 170
JPN

Susan Naylor
University of Texas Health Science
Center at San Antonio
Dept. of Cellular and Structural Biology
7703 Floyd Curl Rd.
San Antonio, TX 78284-7762
USA

Ben Carritt (Senior Editor)
University College of London
MRC Human Biochemical Genetics Unit
4 Stephenson Way
London NW1 2HE
ENG

Chromosome 4

Ken H. Buetow
Fox Chase Cancer Center
Division of Population Science
7701 Burholme Ave.
Philadelphia, PA 19111
USA

Jeffrey C. Murray
University of Iowa
Department of Pediatrics
Division of Medical Genetics
JCP 2613
Iowa City, IA 52242
USA

Gert J.B. van Ommen (Senior Editor)
University of Leiden
Department of Human Genetics
Sylvius Laboratory
P.O. Box 9503
2300 Leiden
NL

Chromosome 5

John D. McPherson
University of California at Irvine
Department of Biological Chemistry
Medical Sciences I, Room D420
Irvine, CA 92717
USA

Larry L. Deaven (Senior Editor)
Los Alamos National Laboratories
Life Sciences Division
P.O. Box 1663
Los Alamos, NM 87545
USA

Chromosome 6

R. Duncan Campbell
University of Oxford
Department of Biochemistry
MRC Immunochemistry Unit
South Parks Rd.
Oxford OX1 3QU
ENG

Howard Cann
Centre d'Etude du Polymorphisme Humain
27 rue Juliette Dodu
F-75010 Paris
FRA

Andreas Ziegler (Senior Editor)
Institute for Experimental Oncology and Transplantation Medicine
Universitats-Klinik Rudolf Virchow
Standort-Charlottenberg
Spandauer Damm 130
D-1000 Berlin 19
FRG

Chromosome 7

Karl-Heinz Grzeschik
Institut fur Humangenetik
Bahnhofstrasse 7A
D-3550 Marburg
FRG

Lap-Chee Tsui (Senior Editor)
Hospital for Sick Children
Department of Genetics
555 University Ave.
Toronto, ON M5G 1X8
CAN

Chromosome 8

Stephen Wood
University of British Columbia
Department of Medical Genetics
216 Wesbrook Building
6174 University Boulevard
Vancouver, BC V6T 1Z3
CAN

Dennis T. Drayna (Senior Editor)
Genentech, Inc.
460 Point San Bruno Blvd.
Room 1277
South San Francisco, CA 94080
USA

Chromosome 9

Sue Povey
University College of London
MRC Human Biochemical Genetics Unit
Wolfson House
4 Stephenson Way
London NW1 2HE
ENG

Moyra Smith (Senior Editor)
University of California at Irvine
Department of Pediatrics/ Human Genetics
Medical Sciences I, C-234
Irvine, CA 92717
USA

Chromosome 10

Bruce A.J. Ponder
CRC Human Cancer Genetics Group
Department of Pathology
Tennis Court Road
Cambridge,
Cambridgeshire CB2 1QP
ENG

Nancy E. Simpson
Queens University
Department of Pediatrics
Kingston, ON K7L 3N6
CAN

Kenneth K. Kidd (Senior Editor)
Yale University School of Medicine
Department of Human Genetics
333 Cedar St.
New Haven, CT 06510
USA

Chromosome 11

Glen A. Evans
The Salk Institute
Molecular Genetics Laboratory
P.O. Box 85800
10010 N. Torrey Pines Road
San Diego, CA 92138-9216
USA

Carol Jones
Eleanor Roosevelt Institute for Cancer Research
1899 Gaylord St.
Denver, CO 80206
USA

Veronica van Heyningen (Senior Editor)
Western General Hospital
MRC Medical Genetics Unit
Crewe Rd.
Edinburgh EH4 2XU
SCOT

Chromosome 12

Tobias Gedde-Dahl, Jr.
University of Oslo
Institute of Forensic Medicine
State Hospital
0027 Oslo 1
NOR

Raju Kucherlapati
Albert Einstein College of Medicine
Department of Molecular Genetics
1300 Morris Park Avenue
New York, NY 10461
USA

Ian W. Craig (Senior Editor)
University of Oxford
Genetics Laboratory
South Parks Road
Oxford OX1 3QU
ENG

Chromosome 13

Aravinda Chakravarti
University of Pittsburgh
Department of Human Genetics
A310 Crabtree Hall
130 De Soto St.
Pittsburgh, PA 15261
USA

Dorothy Warburton (Senior Editor)
Columbia-Presbyterian Medical Center
Babies Hospital
Room BHS-B7
3959 Broadway
New York, NY 10032
USA

Chromosome 14

Tobias Gedde-Dahl, Jr.
University of Oslo
Institute for Forensic Medicine
State Hospital
0027 Oslo 1
NOR

Diane W. Cox (Senior Editor)
The Hospital for Sick Children
555 University Ave.
Toronto, ON M5G 1X8
CAN

Chromosome 15

Jacques S. Beckman
Centre d'Etude du Polymorphisme Humain
27 Rue Juliette Dodu
Paris 75010
FRA

Timothy A. Donlon (Senior Editor)
Stanford University Hospital
Molecular and Clinical Cytogenetics Unit
H1517
Stanford, CA 94305
USA

Chromosome 16

David Frederick Callen
The Adelaide Children's Hospital
Department of Histopathology
Cytogenetics Unit
72 King William Rd.
North Adelaide, South Australia
5006 AUSL

Stephen Reeders
Yale University School of Medicine
Boyer Center for Molecular Medicine
295 Congress Ave., Suite 136
New Haven, CT 06536-0812
USA

C. Edgar Hildebrand (Senior Editor)
Los Alamos National Laboratory
Genetics Group
M886
Los Alamos, NM 87545
USA

Chromosome 17

Pamela Fain
University of Utah Research Park
410 Chipeta Way, Room 103
Salt Lake City, UT 84108
USA

Ellen Solomon (Senior Editor)
Imperial Cancer Research Fund
Somatic Cell Genetics
P.O. Box 123
44 Lincoln's Inn Fields
London WC2A 3PX
ENG

Chromosome 18

Ad Geurts van Kessel
University of Nijmegen
Department of Human Genetics
P.O. Box 9101
6500 HB Nijmegen
NL

Michelle LeBeau (Senior Editor)
University of Chicago
Department of Medicine
Joint Section of Hematology/Oncology
Box 420, 5841 S. Maryland Ave.
Chicago, IL 60637
USA

Chromosome 19

Harvey Mohrenweiser
Lawrence Livermore National Laboratory
Biomedical Sciences Division
L-452
P.O. Box 5507
Livermore, CA 94550
USA

Hans-Hilger Ropers
Katholieke Universiteit
Anthropogenetisch Instituut
Geert Grootplein Zuid 20
NL-6526 Nijmegen
NL

Chromosome 20

Ingo Hansmann
University Gottingen
Institute of Human Genetics
Gosslerstrasse 12D
Gottingen D-3400
FRG

Tim P. Keith
Collaborative Research
Department of Human Genetics
1365 Main Street
Waltham, MA 02154
USA

Cassandra L. Smith
Lawrence Berkeley Laboratory
Human Genome Center
1 Cyclotron Road
Berkeley, CA 94720
USA

Jean Weissenbach
Institut Pasteur
Unite de Recombinaison et Expression Genet
28, rue de Docteur Roux
75724 Paris, Cedex 15
FRA

Chromosome 21

Stylianos E. Antonarakis
The Johns Hopkins University
School of Medicine
CMSC 1003
600 N. Wolfe St.
Baltimore, MD 21205
USA

Nobuyoshi Shimizu
Keio University School of Medicine
35 Shinanomachi
Shinjuku-ku
Tokyo 160
JPN

David R. Cox (Senior Editor)
University of California at San Francisco
Department of Pediatrics/Psych./Biochem.
505 Parnassus Ave.
Box 0106
San Francisco, CA 94143
USA

Chromosome 22

Ken H. Buetow
Fox Chase Cancer Center
Division of Population Science
7701 Burholme Ave.
Philadelphia, PA 19111
USA

Jean-Claude Kaplan
C.H.U. COCHIN
Institut de Pathologie Moleculaire
24, rue du Faubourg Saint-Jacques
Paris 75014
FRA

Melvin I. Simon
California Institute of Technology
Division of Biology, 147-75
1202 East California Blvd.
Pasadena, CA 91125
USA

Beverly S. Emanuel (Senior Editor)
The Children's Hospital of Philadelphia
Department of Pediatrics and Human Genetics
34th St. and Civic Center Blvd.
Philadelphia, PA 19104
USA

Chromosome X

Tony Monaco
Imperial Cancer Research Fund
Institute of Molecular Medicine
John Radcliffe Hospital
Headington
Oxford OX3 9DU
ENG

David L. Nelson
Baylor College of Medicine
Institute for Molecular Genetics
One Baylor Plaza, T821
Houston, TX 77030
USA

David Schlessinger
Washington University School of Medicine
Department of Molecular Microbiology
Box 8230
660 South Euclid Ave.
St. Louis, MO 63110-1093
USA

Huntington F. Willard
Stanford University Medical Center
Department of Genetics
Room S-337
Stanford, CA 94305-5120
USA