

国際的ヒト遺伝子地図データベース としてのGDB

鈴木 政彦^{*1}、平川 美夏^{*1}、諏訪 秀策^{*1}、宇津野 宏二^{*1}、清水 信義^{*2}

^{*1} 日本科学技術情報センター 技術開発部

^{*2} 慶應義塾大学 医学部 分子生物学教室

Abstract

ゲノム情報の発生に関し、各種のデータベース化の試みがなされている。中でもヒト遺伝子の地図に関するデータベースとしては、GDBが世界的な取りまとめを行っている。ヒト遺伝子地図情報をデータベースシステムとして見たとき、従来のデータベースで扱われてきたやり方とは異なるアプローチが必要となることがある。本稿では、GDBの概要と現状を述べ、ヒトゲノム解析プロジェクトとしてヒト遺伝子地図データベースの今後の在り方について考察した。

THE GENOME DATA BASE (GDB) : INTERNATIONAL HUMAN GENE MAPPING DATA BASE

Masahiko Suzuki^{*1}, Mika Hirakawa^{*1}, Shusaku Suwa^{*1}, Kouji Utsuno^{*1},
Nobuyoshi Shimizu^{*2}

^{*1} The Japan Information Center of Science and Technology(JICST)
5-2, Nagata-cho 2 chome, Chiyoda-ku, Tokyo 100, Japan

^{*2} Keio University School of Medicine
35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan

Abstract

The Genome Data Base(GDB), which stores all information on human gene mapping such as location of genes and genome segments, is produced under an international collaboration and is now available all over the world. An outline of GDB and its present status are described in this paper. The future development of such a human gene mapping data base as a human genome project is also discussed.

1. はじめに

遺伝子組み換えを始めとするバイオテクノロジーのめざましい発展により、動物・植物を問わず各種生物の遺伝情報の解析が近年可能になってきた。この遺伝情報はゲノム (genome: 生物を形作る最小限度の遺伝子のセットを持った染色体の一群) とよばれ、マクロには染色体のレベルからDNA、タンパク質の分子構造といったミクロのレベルまで、含んだものである。いくつかの生物を対象として、全てのゲノムを解析しようとする多くのゲノム解析プロジェクトが発足し、既に欧米諸国では成果を上げつつある。

ヒト (人間) に関しても、数多くの国際プロジェクトが進められており、発生したデータは、以下のようなデータベースに登録されている。

表 1.1

1)	GDB (Genome Data Base) (ヒト遺伝子マッピング)
2)	OMIM (Online Mendelian Inheritance in Man) (ヒトの遺伝病と遺伝形質)
3)	GenBank と EMBL (European Molecular Biology Laboratory) (DNA シーケンス)
4)	PIR (Protein Identification Resource) (アミノ酸シーケンス)
5)	ATCC (American Type Culture Collection) (DNA プローブ)

これらの中で、近年ヒトゲノムプロジェクトとして注目を集めているものがGDBである。GDBは、米国エル大学のHGM L (Human Gene mapping Library) に端を発し、ハーワード・ヒューズ医学財団の支援の下にジョンズ・ホプキンス大学で開発運営された。その後、米国DOE (エネルギー省)、NIH (国立衛生研究所) が支援する形に変わったが、本年5月、国際プロジェクトとして日米欧3種の共同事業化の方向が打出された。本稿では、以下GDBの概要について述べる。

2. ヒト遺伝子解析法の一例

ヒトのゲノムは22種類の常染色体と2種類の性染色体から成る (図2.1)。1番~22番染色体は長さにより番号が付けられたもので、XとYは性差を表現することから区別されている。染色体の上に存在する遺伝子の位置を特定するには様々な方法があり、それらの結果をまとめて共通遺伝子地図が作成される (図2.2)。

具体的な方法としては、家系分析による連鎖地図の様な遺伝学的方法と、染色体やDNAに直接マッピングする物理学的方法がある。物理学的方法には、蛍光で染めた染色体のバンド上に目印を付け、目的遺伝子の位置を決める等のトップダウンの方式と、染色体から抽出したDNAを断片化し、酵母や大腸菌中で増幅したクローンを、既知の遺伝子や位置情報に基づいて染色体へと再構成する等のボトムアップ方式がある。これらの方法は、精度やスケールの違いから、互いに補い合いながら、より正確な地図情報を提供する。

ヒトゲノムプロジェクトは、当初5年間で遺伝子地図を完成し、その後地図上の位置を参照してゲノムDNA全クローンの塩基配列を決定する予定で進められている。

3. GDBの構築と利用

GDBは、ヒトの遺伝子に関する地図情報をHUGO (Human Genome Organisation; 国際的なヒトゲノム研究者の組織) の協力の下、国際的に統一されたデータベースとして構築が進められている。GDBの構築は、HGM (Human Gene Mapping; 国際ヒト遺伝子地図ワークショップ) の際に集積されたデータを評価して大量に追加するものが主で、この他に世界各地の研究機関にいるHUGO委員によるデータベースへの直接入力が行われている (図3.1)。

GDBは、データベース本体を米国メリーランド州ボルチモア市のジョンズ・ホプキンス大学に置き、米国内の各種ネットワーク (NSFNET; 米国科学財団の運営するアカデミック・ネットワーク、一般商用ネットワーク、電話回線等) により米国内の各研究機関へ提供されている。またInternetを経由することにより、世界各地のネットワークから接続できる。ゲノムデータベースの利用料金は無料で、基本的には、利用者が最寄りのネットワーク接続ポイント迄の料金を負担するだけで済む。ネットワークに加入できない利用者には、一般公衆電話回線からの接続も準備されており、全ての利用者に対しオープンなシステムとなっている。

4. GDBデータベースの概要

GDBシステムは、ワークステーションとネットワークによるクライアントサーバーシステムで、遺伝子地図情報としてのGDBと文献情報としてのOMIMの2種の含む。特にGDBの検索結果からOMIMの検索を実行できる等、両者を連携して利用できるようなデータベースシステムを構成している。

GDBシステムのハードウェアは、クライアントサーバーシステムで、サーバーワークステーション (SUN4/490: サンマイクロシステムズ社) の構成は図4.1の様のものである。このワークステーションが、ジョンズ・ホプキンス大学内LANに接続され、NSFNETならびに電話回線に接続され、全世界からの利用を受けている (図4.1)。

GDBシステムのソフトウェアは、ワークステーションのオペレーティングシステム（UNIX）とリレーショナルデータベース（Sybase）から構成され、システムの記述はSybaseのアプリケーション・プログラムで記述されている。このため、GDBシステムは移植性が高く、同一の環境を準備することで、容易に移植できるようになっている（図4.2）。

GDBデータベースはいくつかのデータマネージャ（個別検索システム）から成り、概略は以下のようなものである（図4.3、表4.1）。

表4.1

LOCUS マネージャ	: 遺伝子座の名称、シンボル、位置、多型の有無、実験方法、等
MAP マネージャ	: CEPH（フランスで作成されたヒト遺伝子地図）と互換の遺伝子地図、及び各種物理地図とその作成方法
PROBE マネージャ	: クローン化DNAプローブとPCRプライマー、ASO等のオリゴヌクレオチド情報
CONTACT マネージャ	: GDBのエディタや委員、ユーザーの情報
SOURCE マネージャ	: 遺伝子座、遺伝子地図の情報源となった文献情報
OMIM	: NIHのIRXを検索システムとして利用した、フルテキストデータベース。メンデル遺伝に関する文献より遺伝子座ごとに抄録を蓄積している。GDBとは、MIM番号を通じて、LOCUS、SOURCEマネージャと接続している。

GDBに格納されているデータの種類の量としては、遺伝子座（約1万）、プローブ（約2万）、多型情報（約3,500）、出典としての研究論文（約2万文献）と研究者自身の直接登録（約6,000）、等のヒトに関する研究情報と関連のマウスゲノム情報、更にDNAシーケンスデータベースであるGenBankの登録情報、ゲノム研究者情報、等を含む。

データベースのデータスキーム、検索出力のフロントエンドであるマネージャ、及び検索画面については以下の通りである。

①データスキーム

GDBは、3,000種余りのデータを約400テーブルに保持し、データベース管理システムSybase上に構築された、リレーショナルデータベースの形式を採用している。テーブルは関連データの集合体として構成され、最少1個のユニークな内部キーを共有することによりテーブル間のデータの相関を保っている（図4.4）。テーブル間の関係はデータスキームによって図示される。データスキームは、保持するデータの種類や特徴によって図4.5の様なダイアグラムに分けて表示されている。

②データマネージャ

GDBからユーザが得られるデータは、8個に分類されたデータマネージャのそれぞれの1画面として表示される。日付やデータ評価などは、各データに記載される（図4.3、図4.6）。

③検索画面

Welcome画面から直接メニュー選択により検索画面に入る、LOCUS、SOURCE、PROBE、CONTACTの各画面がある。検索方法はフィールドにより、キーボードから直接文字を入力、表示項目から選択、フィールドバリューの一覧表中から選択の3種の検索方法が存在する（図4.7、図4.8）。

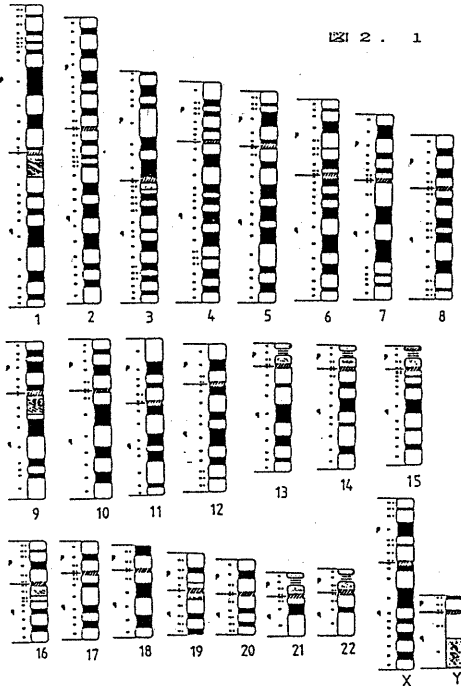
5. 今後のヒト遺伝子地図データベース

以上のように国際的ヒト遺伝子地図データベースGDBについて紹介した。日本国内でも科学技術庁科学技術振興調整費によるヒトゲノムプロジェクト、文部省科学研究費重点領域研究によるゲノム情報、といったゲノム解析プロジェクトが進められているが、今後はこれら研究を支援すべくGDBの国内利用体制の整備、国内利用に馴染むシステム（日本語表記、日本の研究機関に独特なデータの処理等の追加）への改良、遺伝子地図のグラフィック表示等、データ蓄積型からデータ利用解析型データベースに向けての利便性の向上、等を進めていく必要がある。

参考文献

- 1) ニュートンspecial issue; ヒトゲノム解析計画 遺伝情報を解説する巨大プロジェクトの全容: 教育社
- 2) 人類遺伝学の基礎; 近藤、葛巻、吉田: 南江堂
- 3) GDB/OMIMマニュアル

本研究は、科学技術庁の平成3年度科学技術振興調整費による「ヒト遺伝子地図作成技術の開発に関する研究」の一環として行った調査・研究の成果の一部である。



ヒト染色体24種の標準分染パターン模式図 (ISCN, 1985を改変)
 p=短腕, q=長腕, 各染色体の左側の数字はそれぞれのバンド領域を示す

図 2 . 1

染色体地図作製と遺伝子解析法

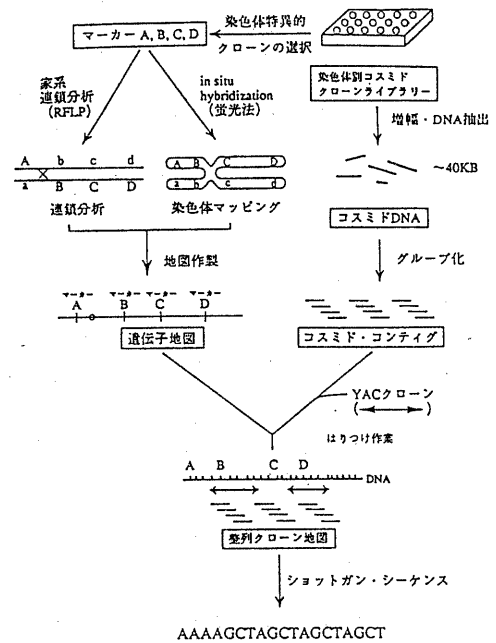


図 2 . 2

Data Flow, Validation, and Distribution Pattern

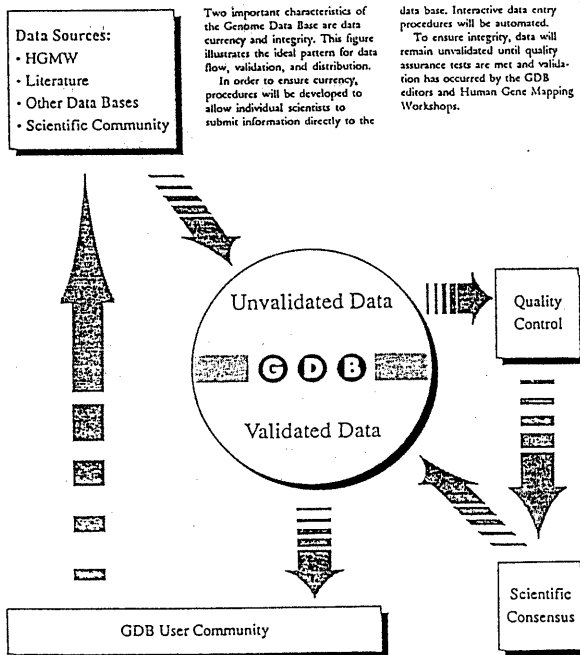
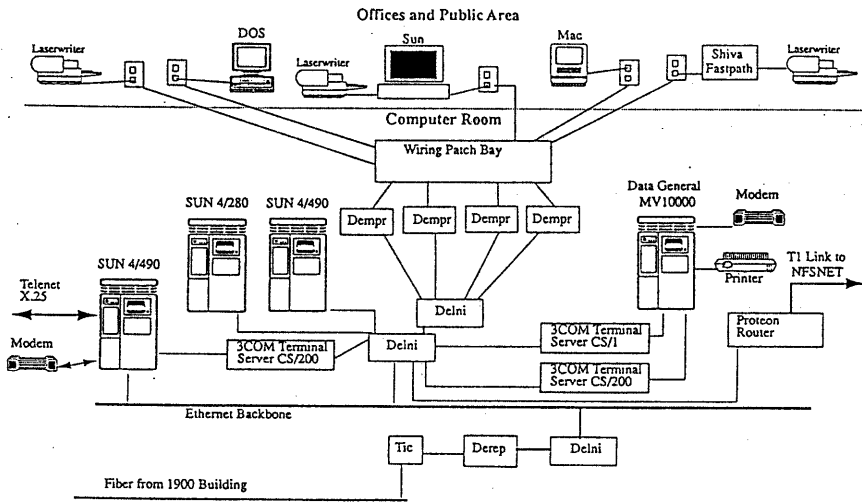
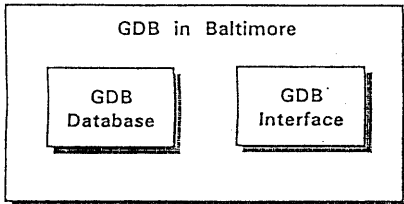


図 3 . 1



Wall Jack

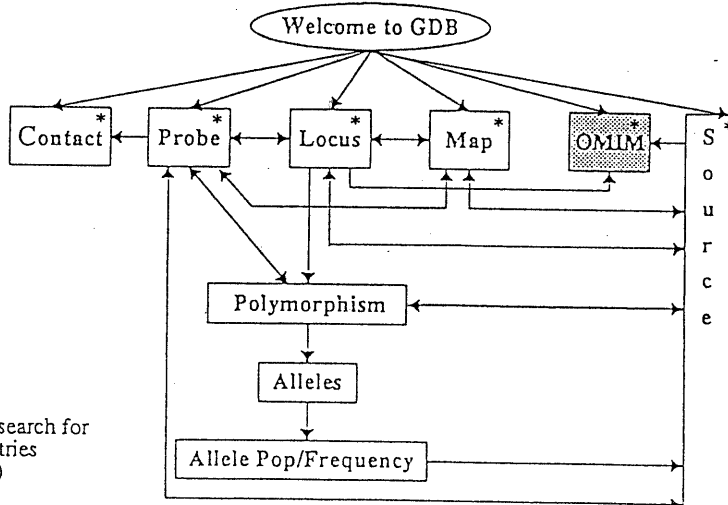
GDB Structure



Both the back-end database and the front-end interface were built using Sybase tools.

GDB was developed to run on Sun workstations running the Unix operating system.

GDB was constructed using a commercial database system (Sybase), with a client-server architecture.



* indicates can search for new set of entries (i.e. Retrieve)

TABLE: locus_alias

Column Name	Type	Length
locus_alias_date	datetime	8
locus_alias_status_flag	tinyint	1
locus_alias_symbol	char	12
locus_alias_type_key	tinyint	1
locus_dna_type_key	tinyint	1
locus_id	int	4

TABLE: locus_alias_type_dict

Column Name	Type	Length
locus_alias_type_desc	varchar	255
locus_alias_type_key	tinyint	1

TABLE: locus_brkpt

Column Name	Type	Length
locus_brkpt_chrom_num	char	2
locus_brkpt_id	int	4
brkpt_orient_key	tinyint	1
locus_brkpt_set_id	int	4
locus_id	int	4

DIAGRAM名	表示内容
enzys	制限酵素データベースの情報とリンクキー
gdb.admin	GDB全体の管理、保持に関わるテーブル データサーバや配布先のホスト情報、画面の表示形式、 GDBから各種メッセージ、メール、アウトプット、 プリントの制御などに関する各テーブルの相関
gdb.hist	データの作成・更新の記録に関するテーブル
locus	遺伝子座に関する情報のテーブルと各種リンクキー
moreloc	locusに関わる情報の未公開データ
moreprobe	probeに関わる情報の未公開データ
mouse	Jackson研究所のデータベースGBaseのマウスゲノム データとヒト遺伝子座との関連情報
order.set	マップの作成法、マップの表示方法などの地図情報
people	エディタ、委員などのデータベース作成者やユーザーなど 研究者の連絡先情報とデータベース利用記録情報
polys	多型、対立遺伝子関係の各種データと関連情報
probe	プローブに関する情報と他のデータとのリンク情報
proposals	データタイプやオブジェクトなど入力サポート情報
source	データの出典に関する情報。文献、研究者自身の直接登録等の データ源情報やOMIM、MEDLINEとの関連情報
unique.ids	共有して検索、表示され、また記述内容の多い項目について のリンクを示す場合のキーのテーブル
views	入力サポートのオンライン画面を制御するテーブル

LOCUS MANAGER - GENERAL USER

Go To View **Call** Retrieve! Output!

Source Manager
 Polymorphism Manager
 Probe Manager
 Map Manager

Symbol: HBE1
 Name: hemo
 EC Num:
 MIM Num: 142108

Location: 11p15.5
 Mode: A,R,S
 Status: C
 Het: .497
 Annotat:

Prev Symbol: Locus 5 of 7

Cloned: Yes
 Polymorphic: +
 PIC: .373

Created: 81 Jan 86 08:08 Fully Approved: 22 Aug 98 20:58

GENERAL USER - LOCUS VIEW

Go To **View** Call Retrieve! Output!

Table/Full

Location	History	P Name	Mode	Loci	1 - 15 of 159
7q31.1-q31.3	Messages	+ DNA segment, single copy, pr+	A,L,S		C
7q31.3-qter	Aliases	DNA segment, single copy, pr+	S		D
7q31	D7S8	+ DNA segment, single copy, pr+	A,L,S		C
7q31	D7S14	DNA segment, single copy, pr+	A,S		C
7q31	D7S399	+ DNA segment, single copy, pr+	L,N,S		C
7q31	D7S482	+ DNA segment, single copy, pr+	S		P
7q31	D7S411	+ DNA segment, single copy, pr+	N,S		P
7q31	INT1L1	+ int-1-like protein 1	A,N,R,S		C
7q22 or 7q31	LAMB1	+ laminin, B1 polypeptide	A,R,S		C
7q31	MET	+ met proto-oncogene	A,CH,L,R,S		C
7q31-q32	CF	+ cystic fibrosis	L		C
7q31-q32	D7S20	+ DNA segment, single copy, pr+	N		P
7q31-q32	D7S23	+ DNA segment, single copy, pr+	L,N,S		C
7q31-q32	D7S24	DNA segment, single copy, pr+	N,S		C
7q31-q32	D7S25	+ DNA segment, single copy, pr+	N,S		P

RETRIEVE LOCI - GENERAL USER

Retrieve & Sort Last Query! Clear!

Symbol: Symbol Type: [Human]
 Type: **Gene** Fragile-Site DNA-Segment Breakpoint Map-Set
 Name: *hemoglobin*
 EC Num:
 MIM Num:
 Location: **11** Location Match: [Inclusive]
 Polymorphic: + C .none Variation Type:
 Probe Name: Enzyme:
 Probe Type: Cloned PCR ASO Variation Inherited?: []
 Mode: Status:
 Het: PIC:
 Annotat:
 Created: Fully Approved: