

理研ゲノム科学総合研究センターと東大医科研ヒトゲノム解析センター が共同でヒトゲノムドラフト配列統合データベースを作成・公開

理化学研究所ゲノム科学総合研究センター（GSC）・ゲノム構造情報研究グループ（榑佳之プロジェクトディレクター、矢田哲士研究員ほか）と、東京大学医科学研究所ヒトゲノム解析センター・ゲノムデータベース分野（高木利久教授）は共同で、ドラフトシーケンスが終了した「ヒトゲノム」の配列情報をゲノム研究に利用しやすい形式で公開することを目的に、新しいデータベースを作成し、両センターのホームページ上で18日から公開します（<http://hgp.gsc.riken.go.jp>, <http://www.hgc.ims.u-tokyo.ac.jp>）。このような一般に使いやすい形でヒトゲノムドラフト配列データが提供されたのは、世界で初めてです。

1. 背景

「国際ヒトゲノムシーケンス決定コンソーシアム」は平成12年6月26日、ヒトゲノムドラフト配列の終了を宣言しましたが、その内容はDNA断片ごとの配列情報しかなく、一般に分かりやすい形では公開されていませんでした。このため、理研GSCゲノム構造情報研究グループと、東大医科研ヒトゲノム解析センター・ゲノムデータベース分野では、ヒトゲノム情報を有効に活用するため、配列情報を利用しやすい形でまとめる作業を共同で進めていました。

2. 成果及び今後の展開

両グループは、DNA断片ごとの配列情報をコンピュータ上で独自の手法で編集し、ゲノム全体に整列化させた統合データベースの作成に成功しました。HGREP（Human Genome Reconstruction Project）と名付けられたこのデータベースは、染色体の領域ごとにドラフトシーケンスのもとになったクローン情報、クローン同士の重なる様子、クローンの配列データ、さらにそのデータに基づく遺伝子予測などの関連情報を参照・検索できるようになっています。

全体では、大腸菌を用いてクローン化したヒトDNA断片23,032のドラフト配列を解析し、そのうち17,277クローン（DNA断片）について整列化を行いました。これらのクローンは全体でヒトゲノムの約60%を占めていると予測されています。残りの5,755クローンの情報については、現段階では染色体上の位置などが正確に特定できていませんが、それらの重なる様子は既に解析済みで、併せてデータベースに格納されています。

研究チームでは、今後これらの未帰属クローンの位置の特定や、さらなるデータの注釈付けを続け、より完成度の高いものを作成し、多くの研究者の利用に供する予定です。

<参 考>

理化学研究所ゲノム科学総合研究センター・ゲノム構造情報研究グループは、先に 21 番染色体解析の中心的役割を果たすなど、英・米の 5 大シーケンシングセンターに並ぶデータ生産・処理能力を持つ我が国最大のシーケンシングセンターで、今回の国際ヒトゲノム計画によるドラフトシーケンス全体の約 6% のデータを生産しました。一方、東京大学医科学研究所ヒトゲノム解析センター・ゲノムデータベース分野は、これまでヒト SNPs データベース、微生物ゲノム比較データベースなど、数多くのデータベースを構築した我が国を代表するデータベース開発部門であります。今回の成果は、この両者の特色を活かして生まれたものです。

(問い合わせ先)

独立行政法人理化学研究所 ゲノム科学総合研究センター
ゲノム構造情報研究グループ
プロジェクトディレクター 榎 佳之
Tel : 042-778-9923 / Fax : 042-778-9924
Tel : 03-5449-5622 / Fax : 03-5449-5445
(東大医科研)

東京大学 医科学研究所 ヒトゲノム解析センター
ゲノムデータベース分野 教授 高木 利久
Tel : 03-5449-5613 / Fax : 03-5449-5434

(報道担当)

独立行政法人理化学研究所 広報室 嶋田 庸嗣
Tel : 048-467-9272 / Fax : 048-462-4715

HGREP Stats.

Mapped :
17277 Clones
to
5797 Contigs

UnMapped :
5755 Clones
to
4392 Contigs

Mapping :
2741 Clones

2000/6/27

Site navigator

[Top](#)

[Chromosome](#)

[Contig](#)

[Clone](#)

[Fragmnet](#)

[Gene](#)

Keyword search

Keyword

Clone name

Human Genome REconstruction Project **HGREP**

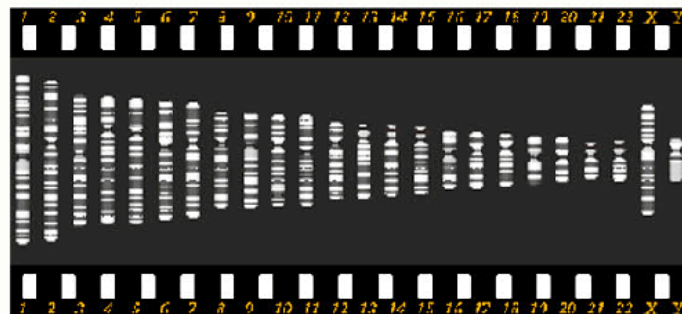
[Top](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [X](#) [Y](#) [Unmapped](#)

What's HGREP

We have developed a database which enables us to overview the human genome structure. The database contains working draft and finished sequences which cover more than 85% of the human genomic sequences. Moreover, sequence entries are aligned along the chromosomes based on sequence similarities to STS markers, BAC-end and other entry sequences. Further, biological features, such as genes, gene functions, repeats and GpC islands, are fully annotated on these entries.

HGREP is a joint project between [Human Genome Research Group](#) (Genomic Sciences Center, RIKEN) and [Laboratory of Genome Database](#) (University of Tokyo, Institute of Medical Science, Human Genome Center).

Access to Contig Map from Chromosome



[click >> Unmapped](#)

Direct Access to Contig Map

Input GenBank Accession number,
and access analysis result of it.

Blast Server

You can perform homology searches with your query against Human genome sequences (Finished and Unfinished).

[click >> Blast Server](#)

What's New...

26 Jun. 2000. Human Genome Reconstruction Project HomePage is opened.

Future Works

- Integration of information about SNPs (positions, substitution patters) into "Contig map page".
- Representation of gene positions (annotated by informatic analysis) in "Chromosome page" and "Contig map page".
- Functional extension of "Keyword search".
- Homology search with Human genome sequences against SWISS-PROT for gene function annotation.
- Integration of analysis result of Fgenes, Fgenesh and HMMgene into annotation.
- Construction of consensus sequences within contigs, and annotate (mapping of genes, calculation of G+C contents and prediction of CpG Islands) the sequences.