報道発表資料

2001 年 2 月 12 日
独立行政法人 理化学研究所

ヒトゲノムのドラフトシーケンス解析結果を公表

日、米、英、仏、独、中の 6 カ国 20 研究センターから構成される「国際ヒトゲノムシーケンス決定コンソーシアム」は、ヒトゲノムドラフトシーケンスの解析を終え、その成果をまとめました。理化学研究所（小林俊一理事長）は、横浜研究所ゲノム科学総合研究センター（GSC）のゲノム構造情報研究グループ（榊佳之プロジェクトディレクター）が中心となり、ヒトゲノムのシーケンス決定に大きく貢献、約 203Mb のデータを取得し、100Mb 以上のデータ解析を行った 6 研究機関の一つに数えられました。さらに同グループは、東京大学と共同で、ゲノム情報を効果的に利用するための「ヒトゲノムドラフト配列データーベース」を世界に先駆けて開発し、ホームページ（http://hgrep.ims.u-tokyo.ac.jp）で公開しています。また、遺伝子予測においては、GSC 遺伝子構造・機能研究グループ（林崎良英プロジェクトディレクター）が生産したマウス cDNA 情報が、レファレンスデータとして大きな役割を果たしています。

国際チームは、ヒトゲノム全体の約 90%をカバーする 27 億 2，500 万塩基の配列を決定しました。解析の結果、ヒトゲノムの 45～50%は、直接生体機能に関与しない繰り返し配列が占め、残りの 50～55%に様々な遺伝子が存在することが分かりました。さらに、ヒト遺伝子の総数は約 31，000 種と推定され、そのうち動物固有のものが約 5 割、そのうちの約半分が脊椎動物に固有の遺伝子であるなど、ヒトゲノムの特色が明らかになりました。ヒトゲノムの全体像が判明したことにより、今後、生活習慣病の遺伝子の同定や、ヒトの発生・分化に関する詳細なメカニズムの解析が加速度的に進行すると期待されます。

本研究成果は、英国科学雑誌「nature」2 月 15 日号に掲載されます。

1.背　景

ヒトの遺伝設計図であるヒトゲノムの全解読は、人類科学の最も輝かしい成果のひとつとして歴史に残るものです。このヒトゲノムの全解読を目指す“ヒトゲノム計画”は国際協調のもと、1991 年より進められてきました。そして、日・米・英・仏・独・中の 6 カ国 20 センターから構成される国際チームは、昨年 6 月 26 日にヒトゲノムのドラフトシークエンス完了を宣言しました。今回、その全データと、それに基づくヒトゲノム全体像の解析結果を、論文として公式に発表することになりました。

2. 方法及び戦略

国際チームは、ヒトゲノム DNA を大きく断片化し、それをクローン化した後各々の断片をショットガン法で配列決定する階層的ショットガン法という戦略を取りました。全ゲノムショットガン法に比べ、この方法は多くのセンターが協力して進めるのに特に有効でした。また配列の位置情報が分かり、解析途中でもデータの有効利用が可能なことから、国際チームでは即時公開の方針を打ち出しました。

3. 研究成果

1）シークエンス決定・編集

ヒト血液または、精子由来 DNA から作成された 11 種の BAC[*1]、PAC[*1] ヒトゲノム DNA ライブラリーに由来する 29，298 クローンを用いて、ヒトゲノム全体の約 90%をカバーする 27 億 2，500 万塩基の配列を決定しました。そのうち約 30%は完成データであり、残りの部分についても、その 9 割は 99．99%の信頼性を持っています。

シーケンス決定においては、日本はヒトゲノム計画の当初から参画、データー生産量では、米、英に次ぐ位置を占めており、理研ゲノム科学総合研究センター（203Mb）と、科学技術振興事業団（JST）[*6]などの支援を受けた慶應義塾大学医学部のグループ（17Mb）がシーケンス決定に大きく貢献しています。さらに全解読した 21 番、22 番染色体は日本の研究チームが中心的、指導的な役割を果たしました。

なお、今後は解析精度を高める一方、完全解読にむけて今回の未解読部分（約 3 億塩基）の解読を行っていきます。

2）データ解析・注釈付け

配列決定したデータをもとに、遺伝子の予測や繰り返し配列の分布など、ヒトゲノム全体の特色を明らかにしました。その結果、ヒトゲノムの 45～50%は直接生体機能に関与しないと思われる分散型繰り返し配列[*2]（LINE21%、SINE13%、レトロウィルス種エレメント 8%）や単純な繰り返し配列が占め、残りの領域に様々な遺伝子が存在することが分かりました。ヒトのタンパク質をコードする遺伝子の総数は、既知遺伝子、EST 配列から編集された Unigene データ[*3]、マウス完全長 cDNA データといくつかの遺伝子予測プログラム[*4]から、約 31，000 種と推定されます。これは、ハエや線虫の総遺伝子数の 2 倍程度でしかありません。それらの機能を推定したところ、物質代謝や転写・翻訳、シグナル伝達に関わるものが比較的多く、また動物固有のものが約 5 割、そのうちの約半分が脊椎動物に固有のタイプの遺伝子であると推定されました。tRNA 遺伝子[*5]は 497 種。このほか、ヒトゲノムが進化の過程で遺伝子の重複やゲノムの一部の重複を繰り返して遺伝子の数や種類を増加させてきたこと、染色体の動原体には特色的な繰り返し構造があること、マウスとの相同性の関係など、ヒトゲノムの様々な特色が明らかとなりました。また、配列解析から 140 万以上の SNP（1 塩基多型）も発見されました。

なお、解析・注釈付けに当たっては、理研ゲノム科学総合研究センターの遺伝子構造・機能研究グループが生産した約 21，000 種の完全長マウス cDNA の情報が大きく貢献[*7]しており、遺伝子予測、データの精度解析の重要なレファレンスデータとして活用されました。

3）データの公開

ドラフトシーケンスデータは、人類共通の財産として公開されています。わが国では、理研ゲノム科学総合研究センターと、東京大学医科学研究所ヒトゲノム解析センターが、世界に先駆けてヒトゲノムの配列情報をゲノム研究に利用しやすい形でデーターベース化（HGREP：Human Genome Reconstruction Project）し、ホームページ上で公開しています。このデーターベースは、様々な高度解析

手法を用いて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りの様子、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多数の関連情報を参照､検索できるようになっています。現在、世界各国から月に数千件のアクセスがあり、わが国が誇るヒトゲノム分野の国際貢献の一つとして、評価されています。

4. 今回の成果の意義

ヒトゲノムシーケンスの完全解読は、2003 年春までには終了する予定です。現在、7 番、14 番、19 番、20 番、Y 染色体は完全解読完了間近です。今後も生命科学の最も基盤となるヒトゲノム全体の高精度配列決定を進めることが、科学の発展に必要ですが、今回、ヒトゲノムの全体像がほぼ判明したことによって、今後生活習慣病の遺伝子の同定や、ヒトの発生・分化に関する詳細なメカニズムの解析が加速度的に進行すると期待されます。

（問い合わせ先）
独立行政法人理化学研究所　横浜研究所
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　榊　佳之
Tel ： 045-503-9171 / Fax ： 045-503-9170
Tel ： 03-5449-5622 / Fax ： 03-5449-5445（東大医科研）
横浜研究所　研究推進部　堤　精史
Tel ： 045-503-9117 / Fax ： 045-503-9112

（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

<補足説明>

＊1「BAC／PAC」

ヒト DNA などを大腸菌にクローン化する時に用いるベクターの一種。BAC はバクテリアの人工染色体、PAC は P1 人工染色体の略。数十万塩基の長さの DNA をクローン化できるのが特色である。

＊2「分配型繰り返し配列（トランスポゾン）」

ヒトゲノムは、主に単純配列の直列型繰り返しと、転移性の因子が分散した分配型繰り返しがある。そのほかに、偽遺伝子や部分的な重複から生じた繰り返しが存在する。転移性の繰り返しの主なものは、短い単位の「SINE」と長い単位の「LINE」である。

＊3「Unigene データ」

世界中で生産された大量の cDNA の断片データをコンピューター上で編集し、分類したもので、コンピューター上で作成された遺伝子が多数リストされている。

＊4「遺伝子予測プログラム」

遺伝子のエキソンやプロモーターの特色などをもとに新規配列の中から遺伝子領域を予測するプログラム。

＊5「tRNA 遺伝子」

タンパク合成に使われる転移 RNA の遺伝子のこと。タンパク質を合成するときに使われる 20 種のアミノ酸に対応して多数の tRNA が存在している。

＊6「JST の取り組み」

科学技術振興事業団による JST シークエンシングプロジェクトは、1995 年より開始され、北里大学、慶應義塾大学、東海大学、（財）癌研究会が参加し、詳細塩基配列を行ってきた。東海大学及び、癌研究会による解読結果の一部も、ドラフトシーケンスの解析データとして取り込まれており、日本全体ではヒトゲノムの約 5%を担った。

＊7「GSC 遺伝子構造・機能研究グループの貢献」

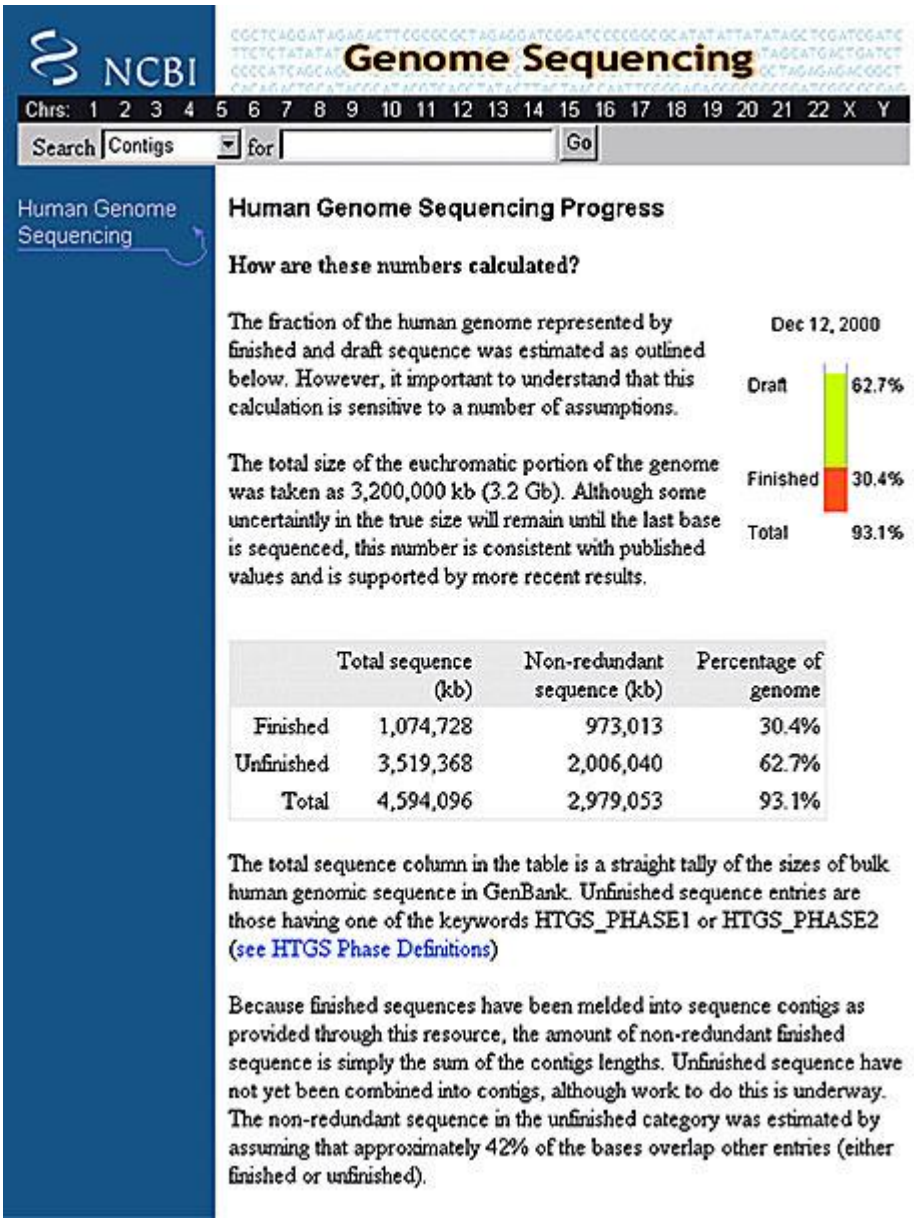
未探索なヒトの遺伝子予測に関して、ヒト既知遺伝子の塩基配列をもとにした解析では予測出来ない部分を、ヒトモデル動物であるマウスにおいて採取された約 21,000 種ものマウス完全長 cDNA の塩基配列をもとにした解析を行い、補完することによって、大きな役割を果たした。

例えば、ヒト既知遺伝子は、線虫やショウジョウバエの既知遺伝子と 8 割近くオーバーラップしているが、理研で得られたマウス遺伝子と、線虫やショウジョウバエの既知遺伝子とは、4 割程度しかオーバーラップしていない。このことから、ヒト既知遺伝子の中には、ほ乳類特有な遺伝子がまだまだ知られていないと考えられるため、 また、国際チームが蓄積したヒト候補遺伝子及び、タンパク質データセットの評価にも理研のマウス完全長 cDNA の塩基配列情報が用いられている。

各センターの解析状況

機関名	全データ (kb)	そのうち完成データ分 (kb)
MIT ホワイトヘッド ゲノム研究所（米国）	1,196,888	46,560
サンガーセンター（英国）	970,789	284,353
ワシントン大学 ゲノムシーケンスセンター（米国）	765,898	175,279
エネルギー省合同ゲノム研究所（米国）	377,998	78,486
ベーラー医科大（米国）	345,125	53,418
理化学研究所 ゲノム科学総合研究センター（日本）	203,166	16,971
ジェノスコープ（仏国）	85,995	48,808
ゲノム治療コーポレーション（米国）	71,357	7,014
分子ゲノムテクノロジー研究所（独国）	49,865	17,788
中国科学院 ヒトゲノムセンター（中国）	42,865	6,297
システム生物学研究所 シーケンスセンター（米国）	31,241	9,676
スタンフォード ゲノムテクノロジーセンター（米国）	29,728	3,530
スタンフォード大学 ヒトゲノムセンター（米国）	28,162	9,121
ワシントン州立大 ゲノムセンター（米国）	24,115	14,692
慶應義塾大学 医学部（日本）	17,364	13,058
テキサスウェスターン大学 医科学センター（米国）	11,670	7,028
オクラホマ州立大学（米国）	10,071	9,155
マックスプランク研究所（独国）	7,650	2,940
ゲノムテクノロジーセンター（独国）	4,639	2,338
コールドスプリングハーバー研究所（米国）	4,338	2,104
その他※	59,574	35,911
合計	4,338,224	842,027

※JST の支援による東海大　癌研究会の 11783kb を含む



「ヒトゲノム計画」解析状況を示す HP

（http://www.ncbi.nlm.nih.gov/genome/seq）

HGREP Stats.

Mapped :
25065 Clones / 4908 Contigs
UnMapped :
4359 Clones / 2593 Contigs
Mapping :
0 Clones
2001/2/2

Related sites

- Ensembl
- Map viewer (NCBI)
- Human Genomics Studio (DDBJ/CIB)

Site navigator

Top
Chromosome
Contig
Clone
Fragment
Gene

Keyword search

Keyword
Clone name

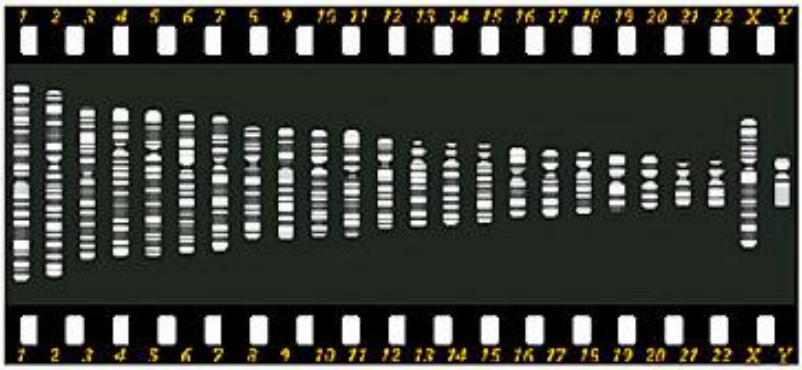
Human Genome REconstruction Project HGREP

What's HGREP

The international consortium for human genome sequencing finished the working draft of the human genome on Jun 26, 2000. This working draft is expected to reveal various features of the human genome. For this reason, we have developed a database which gives an overview of the entire 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 CpG islands, are fully annotated on these entries.

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

Access to Contig Map from Chromosome



click >> Unmapped

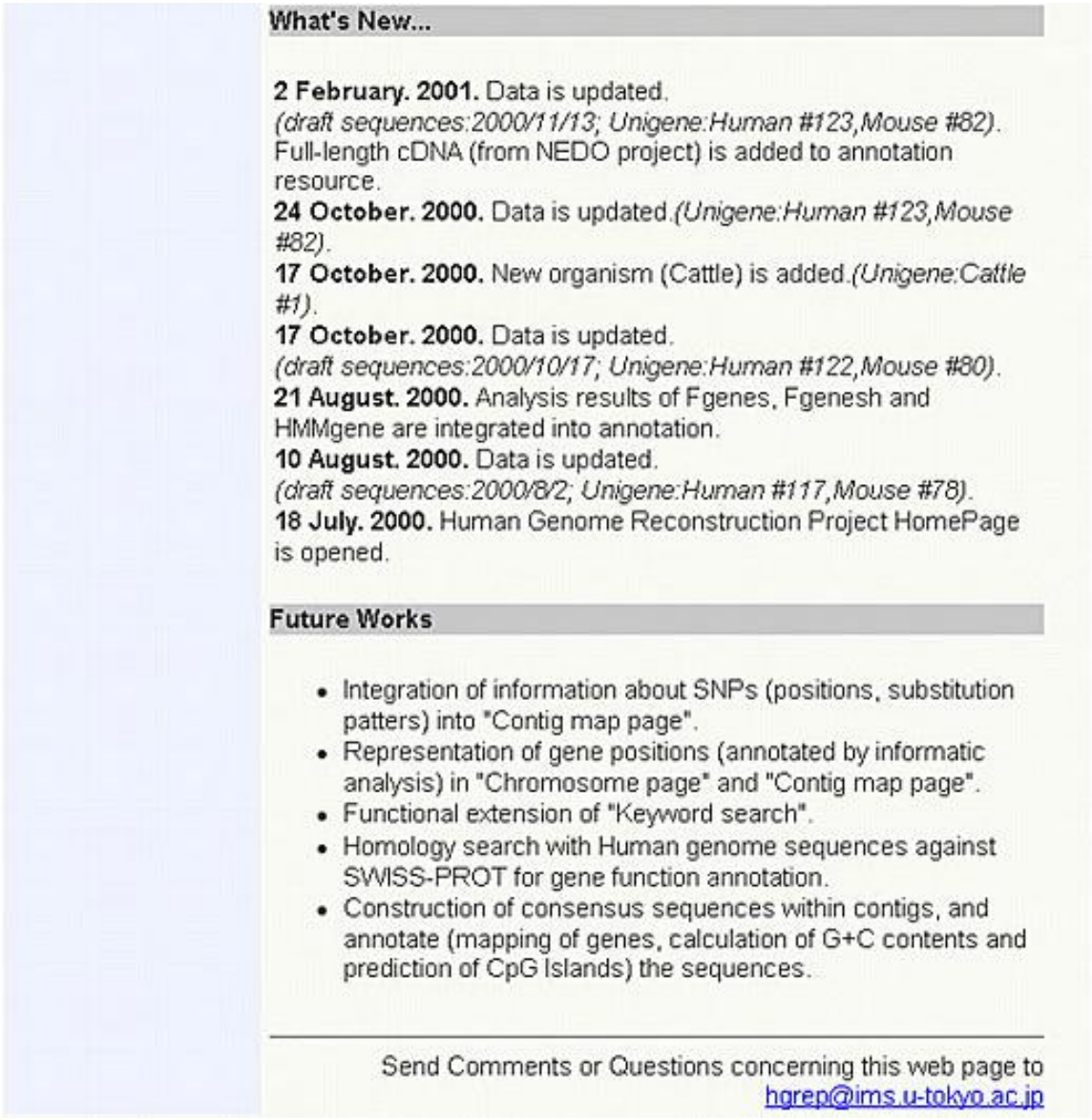
Direct Access to Contig Map

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

Blast Server

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

click >> Blast Server



What's New...

2 February. 2001. Data is updated.
(draft sequences:2000/11/13; Unigene:Human #123,Mouse #82).
Full-length cDNA (from NEDO project) is added to annotation resource.
24 October. 2000. Data is updated.*(Unigene:Human #123,Mouse #82).*
17 October. 2000. New organism (Cattle) is added.*(Unigene:Cattle #1).*
17 October. 2000. Data is updated.
(draft sequences:2000/10/17; Unigene:Human #122,Mouse #80).
21 August. 2000. Analysis results of Fgenes, Fgenesh and HMMgene are integrated into annotation.
10 August. 2000. Data is updated.
(draft sequences:2000/8/2; Unigene:Human #117,Mouse #78).
18 July. 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.
- Construction of consensus sequences within contigs, and annotate (mapping of genes, calculation of G+C contents and prediction of CpG Islands) the sequences.

Send Comments or Questions concerning this web page to
hgrep@ims.u-tokyo.ac.jp

ヒトゲノムドラフト配列結合データベースの HP（http://hgrep.ims.u-tokyo.ac.jp）

ヒトゲノムの解析に用いられたシーケンサー（理研横浜研究所）

ヒトゲノム解析に参加した理化学研究所の研究者ら