

ゲノム構造情報研究グループ

Human Genome Research Group

プロジェクトディレクター 榊 佳之
SAKAKI, Yoshiyuki

当研究グループは、ゲノム科学の最も基本となるゲノム塩基配列（シーケンス）を大規模、迅速かつ高精度に決定し、それをもとにゲノム上に書かれた生物の遺伝情報の設計原理や法則性を見いだすことを目標としている。具体的には現在ヒト 21 番染色体を対象として、慶大、ドイツグループと協力してその全シーケンスの決定を進め、2000 年 5 月 18 日号の「Nature」にその全解析を報告した。その後、マウスの相同染色体との対比解析によって遺伝情報の詳細な解析を進めている。また国際ヒトゲノムシーケンスコンソーシアムの主要メンバーとして、11 番、18 番染色体を中心にドラフトシーケンス 203 Mb を決定し、2000 年 6 月に国際チームとしてドラフト完了を宣言、2001 年 2 月 15 日号の「Nature」にその全解析を報告した。このシーケンス決定の過程を通して世界のトップレベルのゲノムシーケンス決定能力を持つセンターを確立した。またこれに伴うマッピングおよび情報処理システムも確立した。その一環としてヒトドラフトシーケンスデータベース HGREP を構築し公開した。

1. ゲノム地図開発研究（藤山、朴）

ゲノム DNA、cDNA のいずれを対象にする場合でも、良質なライブラリ（ゲノムリソース）が利用可能な状態に整備されていることが必須である。ヒトゲノム構造研究が染色体を単位に進められていることに鑑み、本研究課題は、ヒトなどについて染色体特異的なライブラリの作成と、それに基づく高精度かつ効率的なゲノム地図作成と構造解析の技術開発を中心に進められている。本年度の特筆すべき成果は、ヒト 21 番染色体全構造決定のためマッピングとシーケンシングを並行的に進行させる新しい戦略を塩基配列解析研究チームと共同で策定し、最終的にドイツ 3 センター、慶大と共同して 21 番染色体長腕部分のセントロメア領域からテロメアに至るコンティグを完成し、配列決定の成功に導いたことである。これには、従来から作成を続けていた染色体特異的なライブラリの使用が極めて有効であった。さらに、11 番、18 番染色体については、ドラフト配列決定のためのシードバッククローンと伸長クローンを出発点に配列決定と並行的にコンティグ作成を進める戦略を採った。特に 11 番染色体長腕については、推定で 99% をカバーする地図が完成しており、その一部は国際コンソーシアムによるヒトの概要配列論文に含まれている。また、当研究チームが作成した 14 番染色体ライブラリはフランス GENOSCOPE 研究所による同染色体の地図作成に大きな貢献を果たし、共同著者として論文発表されている。また、新たに 9～12 番染色体について 15 万クローンから成る Fosmid ライブラリを作成し、スクリーニングシステムを完成させた。

2. ゲノム塩基配列解析研究（服部、豊田）

ヒト染色体 21 番について当研究チームは、約 18 Mb の配列の決定を完了した。国際協力プロジェクトとして慶大、ドイツチームと合わせて 21 番染色体の全配列を決定したことになる。これらのデータは遺伝子の同定などの解析も終了し、平成 12 年 5 月に「Nature」に発表した。ヒト 21

番以外では、11q（約 80 Mb）および 18p（約 20 Mb）のドラフトシーケンスが完了し、これは染色体全体の 99% をカバーしている。さらにこのドラフトデータの内約 70 個の BAC クローン（約 15 Mb）については完成データを得た。これらのデータは既に公的データベース（DDBJ）に登録している。11q、18p のドラフトシーケンスの解析結果は、平成 13 年 2 月に「Nature」に発表された。このほかマウス 16 番（ヒト 21 番のシンテニー）マウス 7 番（ヒト 11 番のシンテニー）のシーケンシングを進めている。現在までに、約 5 Mb を完了している。

当研究チームでは我々自身が開発したシーケンス戦略であるネステッドデレクション（ND）法を用い、それをベースとした配列決定システムを確立してきた。さらに本年度はドラフトシーケンスのための高速システムを確立した。大腸菌コロニーからの PCR による鋳型 DNA 調製を行う超高速プロトコルの確立、キャピラリー型シーケンサー対応シーケンス反応プロトコルの確立などを進めた結果、本システムにより、約 2 万リード/日のスループットが実現されている。このほか、チンパンジーゲノムの BAC クローン約 2 万の両端シーケンスを実行し、これら BAC クローンをヒト染色体上にマップした。

3. ゲノム情報比較解析研究

ゲノム情報比較解析研究チームでは、ゲノム配列データの産出からその構造情報・機能情報のアノテーションに至る一連の解析作業を効率化することを目指し、本年度は以下のコンピュータシステムを設計・開発・運用した。

（1）シーケンシング戦略立案支援システム（矢田、高木^{*1}）

昨年度は、ヒト 11 番、18 番、21 番染色体に関する完成およびドラフト配列データを収集・整理し、さらに、インターネット上に散在する既知のマーカーや遺伝子に関する情報を併せて収集・整理することによって、シーケンシングの戦略立案を効果的に支援するコンピュータシステムを開発した。本年度は、このシステムを大幅に改良し、ヒトゲノ

ムの90%に相当する全ドラフト配列データを収集・整理・統合することで、ヒトゲノム全体を再構築するコンピュータシステムを開発した。このシステムによる解析結果は、インターネット上に公開され、内外の研究者によって広く利用されている。

(2) ゲノムプロジェクト進捗状況管理システムの開発と活用(渡邊, 榎)

世界各地で進められている、ヒトを含むさまざまな生物のゲノムプロジェクトの進捗状況についての情報収集システムを開発した。本システムにおいては、各センター、各ゲノム、各染色体毎に、データ産出量とその経時変化などが示される。本システムを用いて、何時、どのセンターがどのゲノムについてどれだけのデータを産出したかを見ることにより、当研究グループにおける配列決定・解析の計画立案等に用いられている。

(3) ドラフトシーケンスデータの高速解析システム(矢田, 高木^{*1})

昨年度は、ホモロジー検索に基づいてドラフト配列データ中の遺伝子領域を高速に検出するコンピュータシステムを開発した。本年度は、上記システムに加え、ホモロジー情報に頼らずに遺伝子領域を検出するコンピュータシステムを開発した。ここで開発したホモロジー情報に頼らずに遺伝子領域を検出するコンピュータプログラムは、既存のプログラムを大きく上回る予測精度で遺伝子領域を検出することが実証された。さらに、本システムを利用して、上述の3(1)で再構築されたヒトゲノム全体を解析し、包括的で体系的な遺伝子領域の抽出を行った。

(4) ゲノム構造情報アノテーションシステムの構築と応用(Taylor, 榎)

ゲノム塩基配列を短時間のうちに大量に決定できるようになった今日、配列内に存在する遺伝子やその他の構造についての情報を取得し、配列データに意味付けする自動システムが必要不可欠になってきている。そこで、複数の遺伝子領域推定プログラム利用することでもっともらしい遺伝子領域情報を取得し、CpG島、GC量、繰り返し配列の位置等の情報を与えるシステムを開発した。このシステムを21番染色体の完成配列に応用した。その結果新たに発見された遺伝子については一部(例えばDSCR5)実験による確認を行うことで、データの質を高めている。

(5) 非経験的アプローチによるゲノム構造情報アノテーションシステム(十時, 矢田, 高木^{*1})

ここ10年間、たくさんのヒト遺伝子発見プログラムが開発されてきたが、未だにヒトゲノム配列の遺伝子同定に利用できる信頼性の高いプログラムはない。一方、これらのプログラムは各々が異なるエクソンを正しく予測する傾向があることが分かってきた。このことは、複数のプログラムを組み合わせると、相補的な効果により、予測精度が向上する可能性があることを示している。実際に21番染色体のゲノム配列が決定されたときに行われた遺伝子同定では、2, 3個のヒト遺伝子発見プログラムが使われ、複数のプログラムが予測した領域がエクソンであると同定する方法が用いられた。しかしこの方法では、予測されたエクソンを遺伝子としてつなげた場合に、エクソン間のフレームの整合性が保たれる保証はない。さらに、複数のプログラムが予測したエクソンでも、すべてのプログラムのスコ

アが低い場合には、そのエクソンが本当に正しいエクソンかどうかは疑わしい。これらの問題を解決するために、当研究グループでは、複数のヒト遺伝子発見プログラムを統合する新しい手法を開発した。エクソン間のフレームの整合性は、隠れマルコフモデルを用いて保証した。異なるプログラムによって予測されたエクソンの配置とそれらのスコアを勘案し、考えられるエクソン領域についての正解確率を導出する手法をベイズ推定に基づいて定式化した。最近集められたデータセットを用いて我々の手法を評価すると、注目に値する予測精度の向上がみられた。特に、遺伝子全体のエクソンを完全に予測したときのみ正解とする基準では、単独の最も予測精度の高いプログラムの予測精度は45%であったが、我々の手法を用いることで予測精度を54%に向上することができた。

(6) ゲノム塩基配列の比較解析(渡邊, 榎)

ゲノム配列に存在する特徴と機能を明らかにするために、大量のゲノム配列データの比較を可能にするシステムを開発し、データベースに登録され公開されているさまざまな生物種のゲノム配列間の類似関係解析を行った。その結果、例えばヒトゲノムでは、これまでに報告されていない多くの重複領域がヒト染色体のセントロメアやテロメアの近傍領域で発見された。また、各生物種が生じる進化過程で重複や欠失が起こった遺伝子を同定し、高等真核生物においても、遺伝子の誕生と死が高頻度で起こっていることを明らかにした。

(7) 細胞内共生細菌ゲノムの決定と解析(重信^{*2}, 渡邊, 榎)

世界で初めて、細胞内共生細菌(ブフネラ)の完全ゲノム配列を決定し、その特徴を明らかにした。1995年以降、多くの細菌ゲノムが決められてきたが、その多くは自由生活性か寄生性の細菌の物である。我々は、共生生物の特徴を遺伝情報の特徴として記述することを試みた。そのために、ゲノムにコードされている遺伝子セットを自由生活性や寄生性の細菌と比較した。その結果、共生細菌の遺伝情報がさまざまな点で特異であることが明らかになった。その特徴を要約すると、共生細菌は、宿主が必要とする物を作り、宿主が作れる物は自ら作ることを完全に放棄して宿主に全面的に依存するように特殊化し、宿主を攻撃したり自分の身を守る機構の一切を放棄していることが、共生細菌の遺伝子セットの機能解析から明らかになった。これは、寄生細菌のゲノムに見られる特徴と正反対である。今後、遺伝子レベルでの共生細菌の特徴づけが行われる。

^{*1} 共同研究員, ^{*2} ジュニア・リサーチ・アソシエイト

誌上発表 Publications

(原著論文) *印は査読制度がある論文誌

Yatsuki H., Watanabe H., Hattori M., Joh K., Soejima H., Komada H., Xin Z., Zhu X., Higashimoto K., Nishimura M., Kuratomi S., Sasaki H., Sakaki Y., and Mukai T.: "Sequenced-based structural features between *Kvlt1* and *Tapa1* on mouse chromosome 7F4/F5 corresponding to the Beckwith-Wiedemann syndrome region on human 11p15.5: Long-stretches of unusually well conserved intronic sequences of *Kvlt1* between mouse and human",

- DNA Res. **7**, 195–206 (2000). *
- Togashi T., Choi D.-K., Taylor T. D., Suzuki Y., Sugano S., Hattori M., and Sakaki Y.: “A novel gene, *DSCR5*, from the distal down syndrome critical region on chromosome 21q22.2”, DNA Res. **7**, 207–212 (2000). *
- Slavov D., Hattori M., Sakaki Y., Rosenthal A., Shimizu N., Minoshima S., Kudoh J., Yaspo M.-L., Ramser J., Reinhardt R., Reimer C., Clancy K., Rynditch A., and Gardiner K.: “Criteria for gene identification and features of genome organization: Analysis of 6.5 Mb of DNA sequence from human chromosome 21”, Gene **247**, 215–232 (2000). *
- Bono H., Kasukawa T., Okido T., Sakai K., Furuno M., Kohtsuki S., Yoshida K., Okazaki Y., and Hayashizaki Y.: “FANTOM+: The interface for functional annotation of mouse cDNA”, Genome Inf. Ser., No. 11, pp. 219–221 (2000). *
- Okamura K., Hagiwara-Takeuchi Y., Li T., Vu T. H., Hirai M., Hattori M., Sakaki Y., Hoffman A. R., and Ito T.: “Comparative genomic analysis of the mouse imprinted gene *impact* and its nonimprinted human homolog *IMPACT*: Toward the structural basis for species-specific imprinting”, Genome Res. **10**, 1878–1889 (2000). *
- Hida A., Koike N., Hirose M., Hattori M., Sakaki Y., and Tei H.: “The human and Mouse *Period1* genes: Five well-conserved E-boxes additively contribute to the enhancement of m*Per1* transcription”, Genomics **65**, 224–233 (2000). *
- Sakaki Y., Hattori M., Fujiyama A., Taylor T. D., Watanabe H., Yada T., Park H. S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., and the Chromosome 21 Mapping and Sequencing Consortium: “The DNA sequence of human chromosome 21”, Nature **405**, 311–319 (2000). *
- Shigenobu S., Ishikawa H., Hattori M., Sakaki Y., and Watanabe H.: “Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS”, Nature **407**, 81–86 (2000). *
- Sakaki Y., Fujiyama A., Hattori M., Yada T., Toyoda A., Itoh T., Kawagoe C., Watanabe H., Totoki Y., Taylor T. D., and International Human Genome Sequencing Consortium: “Initial sequencing and analysis of the human genome”, Nature **409**, 860–921 (2001). *
- Ohira M., Kageyama H., Mihara M., Furuta S., Machida T., Shishikura T., Takayasu H., Islam A., Nakamura Y., Takahashi M., Tomioka N., Sakiyama S., Kaneko Y., Toyoda A., Hattori M., Sakaki Y., Ohki M., Horii A., Soeda E., Inazawa J., Seki N., Kuma H., Nozawa I., and Nakagawara A.: “Identification and characterization of a 500-kb homozygously deleted region at 1p36.2-p36.3 in a neuroblastoma cell line”, Oncogene **19**, 4302–4307 (2000). *
- Ito T., Tashiro K., Muta S., Ozawa R., Chiba T., Nishizawa M., Yamamoto K., Kuhara S., and Sakaki Y.: “Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins”, Proc. Natl. Acad. Sci. USA **97**, 1143–1147 (2000). *

口頭発表 Oral Presentations

(国際会議等)

Maekawa K., Nishimura Y., Sakaki Y., and Soeda E.: “Contig assembly from 21q21.1-22.1 by STS-content mapping”, 13th Ann. Cold Spring Harbor Meet. on Genome Sequencing & Biology, New York, USA, May (2000).

Sakaki Y.: “Complete sequence of chromosome 21 and its comparative genomics”, 3rd Int. Workshop on Advanced Genomics, (164th Committee on Genome Technology, Japan Society for the Promotion of Science), Yokohama, Nov. (2000).

Ohira M., Kageyama H., Mihara M., Furuta S., Machida T., Takayasu H., Islam A., Nakamura Y., Takahashi M., Shishikura T., Sakiyama S., Kaneko Y., Toyoda A., Hattori M., Sakaki Y., Ohki M., Horii A., Chen Y.-Z., Soeda E., Hayashi Y., Inazawa J., Seki N., Kuma H., Nozawa I., and Nakagawara A.: “Identification and characterization of a 500-kb homozygously deleted region at chromosome 1p36.2 in neuroblastoma”, 3rd Int. Workshop on Advanced Genomics, (164th Committee on Genome Technology, Japan Society for the Promotion of Science), Yokohama, Nov. (2000).

Chikayama E., Kuroda Y., Tani K., Matsuo Y., Kawai J., Hayashizaki Y., and Yokoyama S.: “Automated protein domain prediction system for the selection of structure determination targets”, Int. Conf. on Structural Genomics 2000 (ICSG2000), Yokohama, Nov. (2000).

Sakaki Y., Watanabe H., Fujiyama A., Yada T., Toyoda A., Taylor T. D., Totoki Y., Murakami K., Kawagoe C., Itoh T., and Hattori M.: “Why we sequence the chimpanzee genome”, Genes and Minds Initiative Workshop on Ape Genomics (GEMINI), Tokyo, Mar. (2001).

(国内会議)

荒井康仁, 久保崇, 細田文恵, 豊田敦, 服部正平, 榊佳之, 添田栄一, 中川原章, 大木操: “ヒト 11 番染色体 q23 が ん抑制遺伝子領域のゲノム解析”, 第 59 回日本癌学会総会, 横浜, 10 月 (2000).

大平美紀, 影山肇, 三原基弘, 古田繁行, 町田泰一, 服部正平, 榊佳之, 添田栄一, 隈秀和, 中川原章: “神経芽腫細胞株における 1p36 ホモ欠失領域の同定とそのゲノム解析”, 第 23 回日本分子生物学会年会, 神戸, 12 月 (2000).

Research Subjects and Members of Human Genome Research Group

1. Genome Mapping and Resource

2. Genome Sequencing
3. Genome Informatics

Head

Dr. Yoshiyuki SAKAKI

Team Leader

Dr. Asao FUJIYAMA (Genome Mapping and Resources Team)

Dr. Masahira HATTORI (Genome Sequencing Team)

Dr. Yoshiyuki SAKAKI (Genome Informatics Team)

Members

Genome Mapping and Resources Team

Dr. Asao FUJIYAMA

Dr. Hong-Seng PARK

Ms. Mai KONDO

Ms. Yuko ODAMA

Ms. Miwako TOCHIGI

Ms. Kyoko FUKANO

Ms. Ayuko MOTOYAMA

Ms. Rina TSUZUKI

Ms. Miho MAEDA

Ms. Ryoko SAKAI

Ms. Yoriko TERADA

Ms. Yumi TSUKAMOTO

Ms. Fumio EJIMA

Mr. Yoshikazu KOBAYASHI

Ms. Nozomi UYAMA

Ms. Miho YONEZAWA

Ms. Reina OKUMURA

Ms. Mariko IWATATE

Ms. Chikako NAKANO

Genome Sequencing Team

Dr. Masahira HATTORI

Dr. Atsushi TOYODA

Ms. Saori NAKAGAWA

Ms. Noriko ITO

Ms. Minami HORISHIMA

Ms. Maho TANAKA

Ms. Wakako HASHIMOTO

Ms. Rie ARAI

Ms. Hiromi WADA

Ms. Kaoru YAKUSHIJI

Ms. Emi ISOZAKI

Ms. Naomi INAGAKI

Ms. Miho KIYOOKA

Ms. Tomoko ISHII

Ms. Yasue YAMASHITA

Ms. Hinako ISHIZAKI

Ms. Aki NISHIDA

Ms. Noriko YAMAMOTO

Ms. Noriko KODAKA

Mr. Tomoyuki AIZU

Mr. Satoru YOSHIDA

Ms. Yuka MATSUMURA

Ms. Hiroko MORITA

Ms. Konomi NAKAMURA

Ms. Yui BANDO

Ms. Ryoko TANAKA

Ms. Maki UEDA

Ms. Kayo MARUYAMA

Genome Informatics Team

Dr. Yoshiyuki SAKAKI

Dr. Tadayuki TAKEDA

Dr. Todd TAYLOR

Mr. Yasushi TOTOKI

Dr. Hidemi WATANABE

Dr. Tetsushi YADA

Dr. Dong-kug CHOI

Mr. Yutaka SUZUKI

Dr. Akiko FUKUDA

Dr. Mina HIRAOKA

Dr. Hideki NOGUCHI

Ms. Matsumi HIROSE

Ms. Ritsuko OZAWA

Ms. Kayo KOBAYASHI

Mr. Chiho TANAKA

Mr. Shuuji INOUE

Mr. Sven Minoru DEMIYA

Ms. Yoko ONO

Ms. Naoko TOMIOKA

Ms. Yumie OHYAMA

Ms. Yuko MITANI

Visiting Members

Dr. Toshihisa TAKAGI (Human Genome Cen., Univ. Tokyo)

Dr. Satoshi OHMURA (Kitasato Inst.)

Dr. Akira HORII (Sch. Med., Tohoku Univ.)

Dr. Toshimichi IKEMURA (Lab. Evol. Genet., Natl. Inst. Genet.)

Dr. Masayoshi MAI (Cancer Res. Inst., Kanazawa Univ.)

Dr. Tatsuya NISHI (Kyowa Hakko)

Dr. Hajime ISHIKAWA (Dept. Biol. Sci., Fac. Sci., Univ. Tokyo)

Dr. Hiroyuki SASAKI (Div. Human Genet., Dept. Integr. Genet., Natl. Inst. Genet.)

Dr. Katsuhiko SHIRAHIGE (Grad. Sch. Biol. Sci. Res., Nara Inst. Sci. Technol.)

Dr. Takashi ITOH (Cancer Res. Inst., Kanazawa Univ.)

Mr. Katsuhiko MURAKAMI (Life Sci. Group, Hitachi Ltd.)

Mr. Yoshihiro NAKAO (Inst. Fund. Res., Suntory Ltd.)

Mr. Shuji SHIGENOBU (Sci. Univ. Tokyo)

Trainees

Ms. Yuki KATO (Grad. Sch. Biol. Sci. Res. Nara Inst.
Sci. Technol.)

Mr. Kohei MAEKAWA (Grad. Sch. Med., Univ. Tokyo)