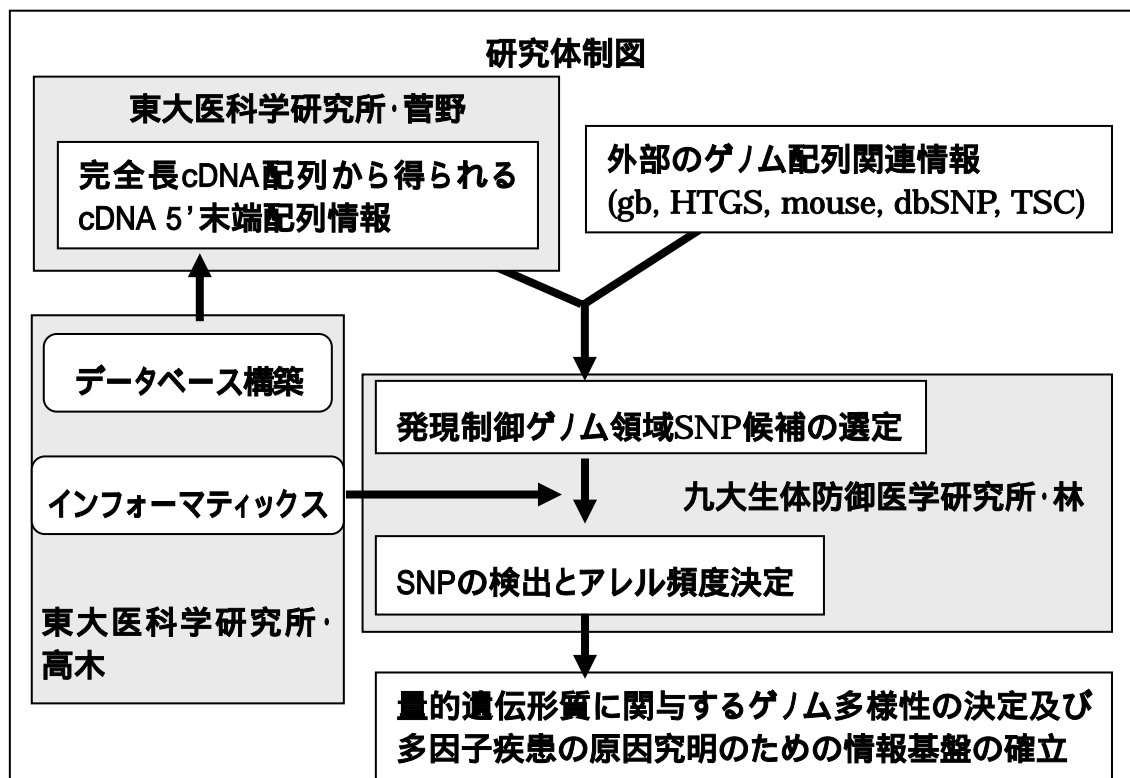


ヒト完全長cDNAクローンの単離とそのバンク化及びそれらを用いた多型マーカーの開発並びに発現情報解析方法の確立のための研究

< 研究の概要・目標 >	< 諸外国の現状 >	< 研究進展によるメリット >
<p>1 何をを目指しているのか 遺伝子発現量を決定するゲノム配列の個人間差異を究明し、多因子疾患の遺伝学的解析に役立つ遺伝情報基盤を確立する。</p> <p>2 何を研究しているのか ・完全長cDNA配列のカatalog化とこれを用いた多数遺伝子の発現制御ゲノム領域の同定を行っている。(菅野) ・キャピラリー電気泳動式SSCP法による高効率SNPアレル頻度解析法を確立し、これを用いた大規模SNP定量解析を行っている。(林) ・DNA配列の情報学的解析アルゴリズムを多数開発している。特に多種配列データベースを総合し、発現支配配列等の意味付けを解析している。(高木)</p> <p>3 何が新しいのか 完全長cDNA配列情報から厳密な情報学的手法により遺伝子発現ゲノム領域を同定し、そこにあるSNPを大規模に同定し、かつそれらのアレル頻度を正確に決定するシステムティックな研究は全く新しい。</p>	<p>1 現状及び我国の水準 ・完全長cDNAライブラリー作成とその配列決定は菅野等が世界に先駆けて開発、完成させた手法である。既に20万クローンの配列を決定しており、さらにその数を拡大しつつある。諸外国でも近年、完全長cDNA研究の重要性が認識され、急速に配列情報が蓄積されつつあるので、我国の当該研究は一層の拡充が望まれる。 ・現在、諸外国では200万余のSNP候補が収集されているが、これらの多くはアレル頻度情報が欠如している。SNPは遺伝解析マーカーとして使用することを目的として収集されているので、その正確なアレル頻度情報は不可欠であるが、これには多数個体の検索を要し、その方法論の選択が世界的に議論されている。林等の開発したPLACE-SSCP法はプールしたDNAを用いてSNPアレル頻度を正確に定量する画期的なもので、我が国から発信する数少ないゲノム研究分野での新技術である。</p>	<p>1 世界の水準との関係 完全長cDNA単離技術は我が国が独自技術を持ち世界レベルで競争し得る分野の一つ。またPLACE-SSCP法によるSNPアレル頻度の正確な定量も我が国の独自開発技術である。これらを結集して、他国に先駆けた大規模完全長cDNAライブラリーの構築、これを基にした遺伝子構造の効率的解析、インフォーマティックスの動員による遺伝子発現支配ゲノム配列の決定、そこにおける遺伝的多様性情報基盤の確立、を行うことが本研究の使命である。これによって特許取得競争における欧米との差を縮められることが大いに期待される。</p> <p>2 波及効果 多因子性疾患の多くは遺伝子発現の量的異常に起因すると想定されるので、本研究によってこれら疾患の予防、早期発見、早期治療、新治療法の開発、個人別医療の確立等、医学分野への幅広い波及効果が期待される。また新規遺伝子の同定、発現調節機構の解明等、基礎生物学分野への波及効果が期待される。</p>



所要経費

(単位：千円)

研究項目	担当機関等	研究担当者	H10-12年度	H13年度	H14年度	所用経費
1. ヒト完全長cDNAクローンの単離とそのバンク化、及び、それらを用いた多型マーカー開発並びに発現情報解析方法の確立のための研究						
(1) 完全長cDNAライブラリーの作製とそのバンク化に関する研究	東京大学医科学研究所	菅野 純夫	313,130	148,700	119,900	581,730
(2) DNA シークエンス法によるヒト遺伝子多型の同定に関する研究	東京大学医科学研究所	中村 祐輔	213,498	0	0	213,498
(3) マイクロアレイ解析技術の開発及びそれを利用した病態の遺伝子発現変化との関連性の検索	武田薬品工業株式会社	藤澤 幸夫	103,983	0	0	103,983
(4) 一塩基多型の体系的解析・収集のための研究	九州大学生体防御医学研究所	林 健志	191,875	107,800	88,300	387,975
(5) ヒト多型DNAデータベースの構築及び有用情報発見・抽出アルゴリズムの研究	東京大学医科学研究所	高木 利久	55,202	10,400	5,100	70,702
合計			877,688	266,900	213,300	1,357,888

研究成果の概要

総括

ヒトゲノム参照配列決定が完了しつつある現在、遺伝子発現制御機構のゲノムワイドな解析と、ゲノム配列の個人間多様性の解析は最も重要なゲノム科学のテーマである。その理由のひとつは、これによって遺伝子発現量異常を原因とすると考えられる種々の多因子性疾患の原因遺伝子探索の研究が飛躍的に進むと期待されているからである。本研究において菅野等の完全長 cDNA グループは、独自開発技術であるオリゴキャップ法を用いた完全長 cDNA クローンの大規模収集を行い、各クローンの配列を決定した。またこの情報を基に多数の遺伝子の転写開始点をゲノム上に同定した。林等の SNP グループは、独自開発の配列変化検出法である SSCP 法を発展させて、一般に普及しているキャピラリーシークエンサーを用いた一塩基多型 (SNP) の大規模検索、アレール頻度高精度決定システムを確立した。またこのシステムを用いて、菅野らが決定した転写開始点を含む遺伝子発現制御領域にある SNP を網羅的に収集し、複数人種集団でのアレール頻度を決定しつつある。高木等の情報グループは上記 2 グループをインフォーマティクス面で支援し、データベース構築と公開に寄与した。これらの成果は、多因子性疾患原因遺伝子同定のための関連解析を多くの研究者にとって現実的なものとするに寄与した。

サブテーマ毎、個別課題毎の概要

< 完全長 cDNA ライブラリーの作製とそのバンク化に関する研究 >

ヒト完全長 cDNA の単離と転写開始点の同定及びプロモーター領域の同定を目指し、完全長ライブラリーの作製と 5' 端配列の決定による cDNA クローンのバンク化を行った。そのために、完全長及び 5' 端濃縮 cDNA ライブラリーの作製をオリゴキャッピング法によって行い、各ライブラリーから 1 万 - 1 万 5 千個の独立の cDNA クローンをマイクロタイタープレートに分離し、5' 端ワンパスシーケンスを行った。得られた部分配列データは既存の Genbank データと照合し、機能既知のヒト遺伝子に対応するもの (Known)、機能未知だが EST データに対応するもの (EST)、全く新規のもの (New) に分類した。cDNA ライブラリー作製にあたっては、ヒト正常組織 (唾液腺、臍帯、副腎、筋肉、心臓、直腸、全脳、視床下部、骨髄、扁桃腺、回腸、胎盤等)、癌細胞由来細胞株 (MKN28、HL60 等)、癌組織 (白血病、消化管腫瘍等) および培養細胞 (臍帯静脈内皮細胞、マクロファージ等) からの mRNA を用い、87 種類の cDNA ライブラリーを作製した。そのうち、55 が完全長 cDNA ライブラリーであり、残り 32 が 5' 端濃縮 cDNA ライブラリーである。総数で約 60 万のクローンを分離し、配列決定により約 40 万の 5' 端部分塩基配列を得た。これらの配列につき、Known、EST、New の割合を解析したところ、機能既知の遺伝子に当たるものが約 70% を占めた。一方、EST にのみ対応するものは全体の 20% を、全く新規のものは全体の 10% を占めた。既知の遺伝子配列に当たるものをクラスター化することにより、約 9000 の既知遺伝子の転写開始部位を同定するに至った。このデータと、ゲノムドラフト配列を比較した結果、約 8000 の遺伝子について、プロモーター領域を含むと考えられる転写開始点上流 1000 塩基長のゲノム配列を得ることが出来た。これらのデータをもとに、転写開始点データベース DBTSS を作製し公開した。

< 一塩基多型の体系的解析・収集のための研究 >

PLACE-SSCP 法は PCR 産物を非変性条件下で電気泳動することにより塩基配列変化を検出する独自開発の手法である。この方法を利用して、個人の DNA を混合して作製した DNA プールを解析することにより、その集団内での SNP アレール頻度を高精度に定量する方法を確立した。更に、本法により大規模に SNP 検出・アレール頻度定量を行うための新しい技術として、蛍光標識ジデオキシヌクレオチドを用いた PCR 産物の増幅後蛍光標識法・短鎖多重標識 SSCP 法の開発、及び自動化キャピラリーアレイ式電気泳動装置を用いた SSCP 解析条件の確立を行った。また、多色蛍光検出自動化キャピラリーアレイ装置から得られた SSCP データの解釈を効率よく行う専用ソフトウェアを

開発した。これらの技術開発により完成したプロトコールおよび新規開発ソフトウェアと既存の有用ソフトウェアを組み込むことにより、PLACE-SSCP法を基幹技術とする大規模 SNP 検出・定量解析データ処理システム「dbQSNP」を構築した。さらに、同システムを用いて完全長 cDNA プロジェクトの成果として同定された遺伝子発現制御ゲノム領域について、SNP の網羅的探索を行い、同時に日本人集団と西欧人集団における SNP アレル頻度解析を行った。その成果はデータベースに格納され、dbQ public として公開されている。

<ヒト多型 DNA データベースの構築及び有用情報発見・抽出アルゴリズムの研究>

ヒト多型 DNA データベースに入力すべきデータを効率よく産生するための実験支援システムの開発を行った。また、並行してヒト多型データベースの設計を行った。さらに、cDNA データ、SNP データとヒトゲノム配列データを統合して転写開始点のデータベースを構築した。また、プロモータ領域の解析を支援するために、共通配列検出プログラムを開発した。これらの研究成果により得られた DBTSS データベース、JSNP データベースは世界的に広く使われている。

波及効果、発展方向、改善点等

ヒトゲノム参照配列決定が完了し、配列の意味づけ、アノテーションが重要な課題となっている現時点で、本研究の成果は重要な意義を持つ。即ち、完全長 cDNA クローンの大量分離とこれらの配列決定、これに基づいた各遺伝子の転写開始点のゲノム上での同定と公開によって、その情報が世界各地で利用され、転写開始配列シグナルの検索、プロモーター選択機構の究明、翻訳開始機構の解析、オルターナティブスプライシング機構の解析等に多大し寄与しつつある。また、PLACE-SSCP法を用いた SNP アレル頻度の正確な決定は、今後 DNA プールを用いた関連解析で、広く多くの研究者にとって利用可能な方法として普及することが期待される。これら二つの研究課題は、我が国において独自開発された実験技術をコアとしたものであり、今後も世界をリードする方法論として発展し続けるために、解析量の更なる拡大を図る。また、この成果を利用して遺伝子発現量の異常に起因する可能性の高い遺伝性疾患である多くの多因子性疾患の遺伝学的解析と発症機構の解明を積極的に推進する予定である。このために本研究でのグループ間の一層の協力体制を推し進めるつもりである。

研究成果公表等の状況

(1) 研究発表件数

	原著論文による発表	左記以外の誌上発表	口頭発表	合計
国内	第 期 0 件	第 期 11 件	第 期 18 件	第 期 29 件
	第 期 0 件	第 期 2 件	第 期 10 件	第 期 12 件
国際	第 期 45 件	第 期 0 件	第 期 18 件	第 期 63 件
	第 期 60 件	第 期 3 件	第 期 15 件	第 期 78 件
合計	第 期 45 件	第 期 11 件	第 期 36 件	第 期 92 件
	第 期 60 件	第 期 5 件	第 期 15 件	第 期 80 件

(2) 特許等出願件数

第 期 0 件 (うち国内 0 件、国外 0 件)

第 期 3 件 (うち国内 3 件、国外 0 件)

合計 3 件 (うち国内 3 件、国外 0 件)

(3) 受賞等

なし

(4) 主な原著論文による発表の内訳

* 発表者氏名:「発表題目」,文献名,巻(号),頁,(掲載年)の順

国外誌

- Suzuki Y, Tsunoda T, Sese J, Taira H, Mizushima-Sugano J, Hata H, Ota T, Isogai T, Tanaka T, Nakamura Y, Suyama A, Sakaki Y, Morishita S, Okubo K, Sugano S.: Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res.* 11: 677-684, (2001)
- Suzuki Y, Taira H, Tsunoda T, Mizushima-Sugano J, Sese J, Hata H, Ota T, Isogai T, Tanaka T, Morishita S, Okubo K, Sakaki Y, Nakamura Y, Suyama A, Sugano S.: Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites. *EMBO Rep.* 2: 388-393, (2001)
- Suzuki Y, Yamashita R, Nakai K, Sugano S.: DBTSS: DataBase of human Transcriptional Start Sites and full-length cDNAs. *Nucleic Acids Res.* 30: 328-331, (2002)
- Sakate R, Osada N, Hida M, Sugano S, Hayasaka I, Shimohira N, Yanagi S, Suto Y, Hashimoto K, Hirai M.: Analysis of 5'-End Sequences of Chimpanzee cDNAs. *Genome Res.*, 13: 1022-1026, (2003)
- Hayashi K, Wenz H-M, Inazuka M, Tahira T, Sasaki T, Atha DH: SSCP analysis of point mutations by multi-color capillary electrophoresis, *Methods in Molecular Biology*, 163, 109-126, (2001)
- Sasaki T, Tahira T, Suzuki A, Higasa K, Kukita Y, Baba S, Hayashi K: Precise estimation of allele frequencies of single-nucleotide polymorphisms by a quantitative SSCP analysis of pooled DNA, *Am. J. Hum. Genet.*, 68, 214-218, (2001)
- Higasa K, Hayashi K: Ordered catenation of sequence-tagged sites and multiplexed SNP-genotyping by sequencing, *Nucleic Acids Research*, 30(e11), 1-14, (2002)
- Kukita Y, Higasa K, Baba S, Nakamura M, Manago S, Suzuki A, Tahira T, Hayashi K: A single-strand conformation polymorphism method for the large-scale analysis of mutations/polymorphisms using capillary-array electrophoresis, *Electrophoresis*, 23, 2259-2266, (2002)

9. Kukita Y, Hayashi K: Multicolor post-PCR labeling of DNA fragments with fluorescent dideoxynucleotides, *BioTechniques*, 33, 502-506, (2002)
10. Higasa K, Kukita Y, Baba S, Hayashi K: A software for machine-independent quantitative interpretation of SSCP in capillary array electrophoresis (QUISCA), *BioTechniques*, 33, 1342-1348, (2002)
11. Baba S, Kukita Y, Higasa K, Tahira T, Hayashi K: Single-stranded conformational polymorphism analysis using automated capillary array electrophoresis apparatus, *BioTechniques*, 34, 746-750, (2003)
12. Poluliakh, N., Takagi, T., Nakai, K.: Melina: a web server for motif extraction from promoter regions of potentially co-regulated genes, *Bioinformatics*, 19(3), 423-424, (2003)
13. Hirakawa, M., Tanaka, T., Hashimoto, Y., Kuroda, M., Takagi, T., Nakamura, Y.: JSNP: a database of common gene variations in the Japanese population, *Nucleic Acids Res.*, 30, 158-162, (2002)

(5) 主要雑誌への研究成果発表

Journal	Impact Factor	サブテーマ1	サブテーマ2	サブテーマ3	合計
Nature	28.0	1	0	0	1
Nature Genetics	29.6	1	0	0	1
EMBO Journal	12.5	1	0	0	1
American Journal of Human Genetics.	10.5	0	1	0	1
Blood	9.3	0	1	0	1
Genome Research	8.6	2	0	0	2
Journal of Biological Chemistry	7.3	1	0	0	1
Journal of Immunology	7.1	0	1	0	1
Oncogene	6.7	1	0	0	1
Nucleic Acids Research	6.4	5	1	1	7
EMBO Reports	6.0	1	0	0	1
Journal of Molecular Biology	5.8	1	0	0	1
Electrophoresis	4.2	0	1	0	1
Invest. Ophth. Vis. Sci.	4.2	0	1	0	1
Journal of Molecular Evolution	4.0	1	0	0	1
Gene to Cells	3.8	2	0	0	2
FEBS Letter	3.6	1	0	0	1
Bioinformatics	3.4	0	0	1	1
Genomics	3.4	2	0	0	2
Human Genetics	3.2	0	1	0	1
Gene	3.0	3	0	0	3
Biochem. Biophys. Res. Commun.	2.9	2	1	0	3
Mammalian Genome	2.3	1	0	0	1
BioTechniques	2.0	0	4	0	4
Immunology Letters	2.0	1	0	0	1
Japanese Journal of Cancer Research	2.0	1	1	0	2
Journal of Biochemistry	2.0	0	1	0	1