

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

I 試験研究の全体計画

1. 研究の趣旨

偶発的な事故による負傷を除き、糖尿病・高血圧などの生活習慣病、がん、アレルギーなどすべての疾患の発症や増悪には、その程度に違いがあるものの、遺伝的素因、すなわち、遺伝子産物の質や量の差が関与する。ヒトゲノム解析研究の最も重要な目的は、このような遺伝的要因を解明して、それを疾患の予防・診断に生かしたり、画期的な治療の開発につなげる点にある。遺伝子の発現量の差（遺伝子発現プロフィール）や遺伝子機能産物の性質の差（翻訳領域における遺伝子多型や遺伝子のスプライシングの差）を体系的に解析し、これを疾患、あるいは、特異な病態との関連を調べて遺伝子機能を類推していく研究は、わが国の将来の医療を見据えた観点からきわめて重要であることは明白である。21世紀は情報と医療が重要な産業の柱となると予測されているが、医学・生物学研究の基盤整備の遅れは、国益の点からみてもきわめて危機的な状況を招くことは必至であり、国家財産の確保という観点からも提案の研究計画の強力な推進が必要である。

本研究は、ヒト遺伝子機能研究、疾患関連遺伝子研究を効率的に進めていく観点から計画されたものであり、(1)研究の基盤材料となる完全長 cDNA クローンバンクの構築、(2)疾患遺伝子研究のための基盤情報となる遺伝子多型情報の収集とデータベース化、(3)遺伝子発現プロフィールを効率よく検索する方法の確立の3つの柱から構成される。

(1)の完全長 cDNA バンクの構築では、オリゴキャップ法というわが国で開発された手法を利用して、多数の臓器と細胞株から抽出した mRNA を材料に効率よく完全長 cDNA を取得する。これらの cDNA クローンを大量にシーケンスすることによってクローンを選別化し、5年間で7万種類からなるヒト完全長 cDNA の単離とバンク化を行うことを目的とする。このバンクの構築は、遺伝子の機能解析を速やかに進めていく観点から非常に重要であり、東京大学医科学研究所のヒトゲノム解析センターを通して国内外の研究者に提供する予定である。(2)の遺伝子多型解析は、遺伝子産物の質・量の差と疾患遺伝子との関連を効率よく検索していくために必要不可欠な基盤情報を提供するものであり、疾患遺伝子の解明やそれに基づく画期的治療法や診断法の開発につながることが期待される。多型を見つけるために、SSCP 法によるスクリーニングや DNA シーケンス法を利用し、5年間に1000 遺伝子のエクソン領域での多型と2000 のエクソン以外の領域での多型の同定と

それらのデータベースの構築を目指す。(3)の遺伝子発現プロフィールを効率よく検索する方法の確立は、さまざまな病態や環境の変化に応じて発現量の変化する遺伝子を効率的に解析し遺伝子機能解明に応用するために重要な基盤技術であり、5年間で数万種類の cDNA クローンからなる DNA チップ等のマイクロアレイ技術の確立を目標とする。

2. 研究内容及び目標

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癌ウイルス研究部）

5年間で約7万種類の遺伝子（全遺伝子の7割）に相当する完全長 cDNA クローンの単離を最終目標とする。

(1) 上記の目標を達成するために、多数のヒト組織（皮下脂肪・内臓脂肪・子宮・骨格筋・大腸・胃・肝臓・脾臓・脾臓・肺・脳の各領域）および、多数の培養細胞（大動脈・冠動脈・胎盤等の血管内皮細胞や種々のがん細胞株）から mRNA を抽出し、これらを材料に、オリゴキャップ法による完全長 cDNA ライブラリーの作製を行う（年間10種類程度の組織もしくは細胞株からのライブラリーの作製を目標とする）。

(2) 各完全長 cDNA ライブラリーより20,000 クローンをマイクロタイタープレートに拾い上げ、これらのクローンより発現量の多い遺伝子を高密度フィルターを利用したハイブリダイゼーションで除去する。それぞれのライブラリーの発現量の低い遺伝子について10,000~15,000 クローンずつ部分配列の決定を行う。年間100,000 クローン（平成10年度）~200,000 クローン（平成11年度以降）の塩基配列を決定する。

(3) これら cDNA 部分配列から得られた膨大な情報を解析して、完全長 cDNA クローンの分離同定とバンク化を行う。

2. DNA シーケンス法によるヒト遺伝子多型の同定及び新規 cDNA の全塩基配列の決定（助癌研究会癌化学療法センター及び東京大学医科学研究所附属ヒトゲノム解析センター）

(1) 新規 cDNA の全塩基配列の決定と DNA シーケンス法によるエクソン領域でのヒト遺伝子多型開発研究（東京大学医科学研究所附属ヒトゲノム解析センター）

単離同定した cDNA クローンを解析した結果、新規遺伝子と考えられたもの3,000種類について、その全塩基配列の決定を行う。

(2) DNA シーケンス法によるエクソン領域でのヒト遺伝子多型開発研究（1800 遺伝子については助癌研究会癌

化学療法センター、1200 遺伝子については東京大学医科学研究所附属ヒトゲノム解析センターで実施。)

遺伝子多型情報を有効に疾患解析に活用するために、エクソン部分を中心として多型の有無をスクリーニングする。これまでに単離された遺伝子の内、その機能から考えて生活習慣病・がん・アレルギー疾患に関与する可能性のあるもので遺伝子構造の決定されているものについて、エクソンを挟むかたちで PCR 反応により DNA を増幅し、これを T4DNA エンドヌクレアーゼ VII で処理して、遺伝子多型をスクリーニングする。この際、5 人ずつの DNA を混合してシーケンスを行う予定である。5 人を 1 サンプルとして 20 サンプル、計 200 染色体分に相当する試料を解析する。多型の認められると考えられるサンプルについては DNA シーケンスで塩基配列を決定する。この方法により、アレル頻度が 0.5 % の遺伝子多型を検出することが可能であると考えている。また、本研究において新規に単離された cDNA など cDNA の配列しか得られていないものについては、白血球 mRNA、もしくは、がん細胞から精製した mRNA の 5~10 人分をプールして計 50 人分について、RT-PCR (reverse transcriptase-PCR) を行い、上記と同様の方法で多型の有無をスクリーニングする。5 年間の目標は 3000 遺伝子についての多型解析を行い、1000 遺伝子についてなんらかの多型を同定することにある。

3. マイクロアレー解析技術の開発及びそれを応用した病態と遺伝子発現変化との関連性の研究 (武田薬品工業株開拓第 2 研究所)

マイクロアレー技術についてマイクロアレーの作製法・定量性・感度などを検討し、定量的かつ高感度の検出法や微量な検体でも検出可能な方法を開発する。これらを見きわめた上で、ヒト主要組織における発現パターンを解析し、データベースを作成する。中間目標としては、まず、技術の確立を目指し、1 万種類の cDNA に関する発現パターンデータベースを作成する。これと並行して、マイクロアレー技術を用いて、さまざまな病態の変化や薬剤・物理的刺激により発現が変化する遺伝子の単離を目指す。最終的には、cDNA シーケンス解析チームによって単離された 7 万種類の cDNA クロンをマイクロアレー技術により解析するシステムの確立を目標とする。

4. 一塩基多型の体系的解析・収集のための研究 (九州大学遺伝情報実験施設)

ヘテロ体出現頻度の高い、すなわち、連鎖解析等に有効な一塩基多型 (single nucleotide polymorphism, 以下 SNP) を見つけるために、SSCP 法に基づく多型解析を利用して体系的な解析を行う。また、個々の試料に対する効率的な遺伝子型決定法を確立する。

(1) 集団中のヘテロ体出現頻度を効率よく推定するためには、多数の PCR 産物を SSCP 分析する必要がある。そこで、初年度は PLACE-SSCP 分析の多重化と、キャピラリー

電気泳動の多チャンネル化をはかり、解析の大容量化を確立する。

(2) 多チャンネルキャピラリー電気泳動装置による大規模多重 PLACE-SSCP 法を確立させるとともに、これを用いて現在マップされている約 20,000 種類の全 STS を対象とする日本人集団における多型検索を行う。これには、まず 50~100 人のゲノム DNA を混合して特定の STS について SSCP 分析を行い、多数のゲノム中でのアレルを同時に検出する。これにより、二つのアレルの存在比の低い STS をヘテロ体出現頻度の高いものとして選別する。次いで、その塩基配列を決定することによりヘテロ体が高頻度出現する SNP を得ることができ (一般的に STS の約 10 % に多型があると推定されているので、本研究によって約 2,000 種類の SNP を得ることが期待される。これは、平均して 1.5Mb に 1 個の分布で SNP を獲得することに相当する)。

(3) 得られた高頻度 SNP を用いて個々の試料について効率的かつ安価に遺伝子型を決定できる方法を確立する。得られた高頻度 SNP を利用した、個人のゲノムワイドな遺伝子型決定法を確立する。これには、SNP 遺伝子型決定のための DNA チップのデザインを行い、遺伝子型決定法としての有効性を判定する。一方、DNA チップによる決定はその判定精度に問題がある (疑陽性、疑陰性の頻度が高い) ので、これに代わる多数の SNP を塩基配列レベルで同時に決定する方法の開発を試みる。

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附属ヒトゲノム解析センター)

cDNA マイクロアレー等を利用した実験結果から得られる大量のヒト遺伝子発現情報やヒト多型に関するデータを効率よく格納する方式、および、それらのデータと DNA 配列データ等との統合化に関する研究を行うことによって、それらから導き出される知識発見支援システムの開発研究を行う。また、並行して、そのようにして作られたデータベースから有用な医学的・生物学的知識、例えばヒト多型と疾患との関係など、の発見を支援する計算技術の研究を行う。

3. 年次計画

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

年間 10~15 種類ずつの cDNA ライブラリーを作製し、それぞれから 10,000~20,000 クロンについて、計年間 100,000 クロン (平成 10 年度)~200,000 クロン (平成 11 年度以降) の塩基配列を決定し、これら cDNA 部分配列から得られた膨大な情報を解析して、

第 I 期では約 5 万種類 (全遺伝子の 5 割) の cDNA バンクの作製を目標とする。

第Ⅱ期ではさらに約2万種類の新規の完全長 cDNA クロウンを単離し、7万種類の完全長鎖 cDNA バンクの構築を最終目標とする。

2. DNA シークェンス法によるヒト遺伝子多型の同定及び新規 cDNA の全塩基配列の決定

(1) 新規 cDNA の全塩基配列の決定と DNA シークェンス法によるエクソン領域でのヒト遺伝子多型開発研究

単離同定した cDNA クロウンを解析した結果、新規遺伝子と考えられたものについて、その全塩基配列の決定を行う。

第Ⅰ期においては2,000種類の cDNA についてプライマーウォーキング法あるいはショットガンクローニング法によって全塩基配列を決定する。

第Ⅱ期についてはさらに1,000種類 cDNA についてプライマーウォーキング法あるいはショットガンクローニング法によって全塩基配列を決定する。

(2) DNA シークェンス法によるエクソン領域でのヒト遺伝子多型開発研究

エクソン領域でのヒト遺伝子多型については、第Ⅰ期は1800 遺伝子のスクリーニングと600 遺伝子における多型の同定を目標とする。

第Ⅱ期については1200 遺伝子についての多型解析を行い、400 遺伝子についての多型の同定を目標とする。

3. マイクロアレー解析技術の開発及びそれを応用した病態と遺伝子発現変化との関連性の研究

第Ⅰ期の目標はマイクロアレーの作製法・定量性・感度などを検討し、定量的かつ高感度の検出法や微量な検体でも検出可能な方法を開発し、ヒト主要組織における1万種類の cDNA に関する発現パターンデータベースを作成することにある。

第Ⅱ期については、cDNA シークェンス解析チームによって単離された7万種類の cDNA クロウンをマイクロアレー技術により解析するシステムの確立とそれを用いて、さまざまな病態の変化や薬剤・物理的刺激により発現が変化する遺伝子の単離を目指す。

4. 一塩基多型の体系的解析・収集のための研究

第Ⅰ期の目標は PLACE-SSCP 分析の多重化・キャピラリー電気泳動の多チャンネル化・解析の大容量化の確立および約20,000 種類の全 STS を対象とする日本人集団における多型検索による約2,000 種類の SNP の獲得にある。

第Ⅱ期については高頻度 SNP を個々の試料について効率的かつ安価に遺伝子型を決定できる方法の確立である。

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

第Ⅰ期の目標は、ヒト多型 DNA データベースの構築を主な目標とする。

この多型データベースはゲノムの配列データのうち人種や個人によって異なる部分のデータを網羅的に納めたもの

である(本プロジェクトでは日本人に特有多型データベースの構築を目指している)。具体的には、それらの情報を実験装置からでてくるデータから効率よく作製する方法を開発したり、それらの情報を有効にかつ効率よく活用するには、データベースにどのような形で表現し、どのようなフォーマットで格納すべきかなどの研究を行う。なお、これらの多型情報や発現情報は、ゲノム配列をはじめとする他の生物学医学情報と関連づけてはじめて意味を持つ。そこで、多様なデータベースとの統合化の研究を並行して行う。また、関連づけるべき情報をインターネットから自動的に収集してくるエージェントシステムについても併せて開発する。

第Ⅱ期については cDNA マイクロアレー等を利用した実験結果から得られる大量のヒト遺伝子発現情報を効率よく格納する方式、および、それらのデータと DNA 配列データ等との統合化に関する研究を行うことによって、それらから導き出される知識発見支援システムの開発研究を行う。具体的には、膨大な発現情報データベースに格納されている発現パターンを分類したり、それをもとにデータベースから特徴的な発現パターンを抽出したりするような計算機ソフトウェアを開発する。また、多型データと薬剤の効果との関連や遺伝子発現パターンと疾患との関係などの医学生物学的知識情報をデータベースから自動的に発見するためのシステムを開発する。

4. 平成11年度における実施内容と達成目標

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

肺、胃、十二指腸、筋肉、脾、脳、心、腎各組織及び大動脈・冠動脈の血管内皮培養細胞の完全長鎖 cDNA ライブラリーを作製し、それぞれから20,000 クロウンについて、計年間200,000 クロウンの塩基配列を決定し、これら cDNA 部分配列から得られた膨大な情報を解析して、30,000 の cDNA バンクを構築する。

2. DNA シークェンス法によるヒト遺伝子多型の同定及び新規 cDNA の全塩基配列の決定

(1) 新規 cDNA の全塩基配列の決定と DNA シークェンス法によるエクソン領域でのヒト遺伝子多型開発研究

単離同定した cDNA クロウンを解析した結果、新規遺伝子と考えられるもの400種類について、その全塩基配列の決定を行う。

(2) DNA シークェンス法によるエクソン領域でのヒト遺伝子多型開発

エクソン領域でのヒト遺伝子多型については、300種類の遺伝子のスクリーニングと100遺伝子における多型の同定を目指す。

3. マイクロアレー解析技術の開発及びそれを応用した病態と遺伝子発現変化との関連性の研究

(1) バンク化された全長ヒト cDNA クローン及びその配列情報をもとに、20,000 種類のヒト cDNA 断片を増幅、精製する。

(2) 上記 cDNA 断片をスライドガラス上にスポットし DNA チップを作製する。

(3) 作製された DNA チップを用い、ヒト癌細胞／組織における遺伝子の発現変動を解析する。

4. 一塩基多型の体系的解析・収集のための研究

現在の PLACE-SSCP 法では、単一チャンネルキャピラリー電気泳動装置を用いているので、本研究での目標達成には処理能力が不足している。そこで今年度は、多チャンネルキャピラリー電気泳動装置を導入し、これを用いた PLACE-SSCP 法の実験系を確立する。さらに PLACE-SSCP 法を用いて、ヘテロ体出現頻度が高い STS を収集し、

SNP 化する作業を開始する。すなわち、ヘテロ体頻度 25 %以上の PCR 産物について、そこに含まれる SNP を塩基配列決定によって同定する。得られた SNP をデータベース化するシステムの構築を開始する。

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

(1) 平成 10 年度に開発した全長鎖 cDNA クローンバンクのシーケンシング及びアノテーションを支援する計算機システムを実際の研究現場で使用し、その経験をもとにシステムの評価、改良を図る。

(2) ヒト多型 DNA に関するデータベースの設計を行い、そのプロトタイプを開発する。

(3) 上記のデータに関連した情報をインターネットから自動的に収集するエージェントシステムを設計する。

II 平成 11 年度における実施体制

研究項目	担当機関	研究担当者
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癌ウイルス研究部	菅野 純 夫
2. DNA シーケンス法によるヒト遺伝子多型の同定及び新規 cDNA の全塩基配列の決定	東京大学医科学研究所附属ヒトゲノム解析センター 東京大学医科学研究所附属ヒトゲノム解析センター	中村 祐 輔 田 中 敏 博
3. マイクロアレー解析技術の開発及びそれを応用した病態と遺伝子発現変化との関連性の研究	武田薬品工業(株)開拓第 2 研究所	白 藤 英 夫
4. 一塩基多型の体系的解析・収集のための研究	九州大学遺伝情報実験施設	林 健 志
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附属ヒトゲノム解析センター	高 木 利 久
○研究管理	(協)癌研究会	向 井 孝 始

III 研究推進委員会

委員	所 属
[プロジェクト内委員]	
○中村 祐 輔	東京大学 医科学研究所附属ヒトゲノム解析センター長
白藤 英 夫	武田薬品工業(株) 開拓第 2 研究所部長
菅野 純 夫	東京大学 医科学研究所癌ウイルス研究部助教授
林 健 志	九州大学 遺伝情報実験施設ゲノム解析分野教授
[プロジェクト外委員]	
勝木 元 也	東京大学 医科学研究所附属ヒト疾患モデル研究センター長
鶴尾 隆	東京大学 分子細胞生物学研究所分子生物活性研究分野教授
寺田 雅 昭	厚生省 国立がんセンター研究所長

(注：○は研究推進委員長)

IV 研究連絡会議

委 員	所 属
中 村 祐 輔	東京大学 医科学研究所附属ヒトゲノム解析センター長
林 健 志	九州大学 遺伝情報実験施設ゲノム解析分野教授
菅 野 純 夫	東京大学 医科学研究所癌ウイルス研究部助教授
白 藤 英 夫	武田薬品工業(株) 開拓第2研究所部長
高 木 利 久	東京大学 医科学研究所附属ヒトゲノム解析センター教授
田 中 敏 博	東京大学 医科学研究所附属ヒトゲノム解析センター助手