

多型マイクロサテライトの収集とヒトゲノム多様性に関する研究

I 試験研究の全体計画

1. 研究の趣旨

ヒトゲノムの全塩基配列決定の完了が間近にせまった現在、今後の「ポストゲノムシーケンシング」としての大きな課題の一つは、同定された遺伝子の機能解析であることは間違いない。そのための有力なアプローチである、ある遺伝的多型とヒトでみられる個性的な生物現象との相関を追及することは、遺伝子とヒトがしめす表現型の機能的因果関係、例えば疾患についてはその遺伝要因を解明することが可能となる。このような多様性検索こそが、実験動物で行われる遺伝子のノックアウト操作が許されないヒトについての、ゲノム機能の解析の突破口を開く、と期待される。

そこで本研究では、世界的に進行しようとしているヒトゲノム多様性プロジェクトの標的である SNP (single nucleotide polymorphism: 単一ヌクレオチド多型; 遺伝子内の一塩基置換、欠失、挿入による差異にもとづく多型) に比べ、多型に富み、したがってより精度の高いマッピングが可能なマイクロサテライトに注目した。すなわち、本研究はまず大量のマイクロサテライトについて、その繰り返し多型を質量分析法により検索しうる DNA チップ技術 (MS チップ) を開発し、続いてこの技術を用いてゲノムワイドにヒトの多型マイクロサテライト 30,000 個を収集し、さらにマイクロサテライトを多型遺伝マーカーとする相関解析により、複合遺伝疾患の原因遺伝子など、ヒト表現型を規定している遺伝要因を同定することを目的とする。このような多型マイクロサテライトを用いたゲノム多様性に関する大規模な研究は、疾患に関係する SNP の同定を効率的に推進し、疾患の治療と予防、創薬に役立つのみならず、学習、記憶などの高次行動を始め、ヒトの遺伝子やゲノム機能など、生物種としてのヒトの生命現象の分子レベルでの理解へと展開するであろう。

第 I 期の目標は、質量分析法によるマイクロサテライト多型の検索基盤技術の確立と 30,000 個のマイクロサテライトの設定で、第 II 期の目標は、多型マイクロサテライトのデータベースの構築とマイクロサテライト多型検索技術の効率化である。

2. 研究概要

1. 質量分析法によるマイクロサテライト多型検索のための DNA チップ (MS チップ) 技術の開発に関する研究 (島津製作所分析機器部)

ゲノムシーケンシングなどの解析でえられたヒトゲノム塩基配列情報を利用して、多型マイクロサテライトを 30,000 個収集・設定し、それらの多型マイクロサテライトを遺伝マーカーとして、相関解析により複合遺伝疾患の原因遺伝子を同定することが、本研究の最終目的である。そこでこの研究サブテーマでは、大量のマイクロサテライトの多型の検索に資するため、PCR 増幅したマイクロサテライトをチップにスポッティングし、MALDI-TOF 型質量分析計にて、マイクロサテライトの分子量を測定することにより、チップ上のマイクロサテライトの繰り返し多型を検索する技術の開発と DNA 抽出からマイクロサテライト多型検出までの自動化を目指す。具体的には下記の 3 つの目標を設定する。

(1) 高分子量検体用の精度の高い質量分析法を用いた DNA チップ (MS チップ) の開発に関する研究

まずヒト第 6 染色体 HLA 領域の SNP 部位を含む約 60 bp の領域、並びに 17 個の対立遺伝子がみられる 2 塩基 GA 繰り返しマイクロサテライトの約 60 bp の領域を PCR 増幅する。えられた PCR 産物を熱変性により一本鎖にして、ポリ L-リシンで表面処理したスライドガラスに、1 cm 画のチップに低い密度 (100 スポット/1 チップ) でスタンピングする。このチップに、MALDI-TOF 型質量分析計を用いてレーザーを照射して、PCR 産物をイオン化、放出させて分子量を測定する。このとき、十分な距離に分離して、多型検索が明瞭になるような至適条件、すなわち試料調製時の多価イオンやスライドガラスでコートするポリ陽イオンの検討を行い、高い精度で分子量を決定するための条件を検討する。さらに、分子量が 20,000 以上の高い 3~5 塩基繰り返しマイクロサテライトについても、チップ上で質量分析法を用いた分子量測定で、多型性検索が可能になるように、高い分子量で高分解能をしめす期待される炭酸ガス赤外線レーザーによる照射、高加速電圧の適用やイオン化条件の検討を行う。

(2) 高密度 DNA チップの開発に関する研究

レーザーの照射位置を電場により移動操作して、1 cm 画のチップに最大限 1,000 個スポットされた PCR 産物のそれぞれについて分子量の測定を行い、1 チップで多量検体のマイクロサテライト多型の検索を可能にする機器を開発する。それによって、1 cm 画のチップに 1,000 個のマイクロサテライト多型の検索が可能となり、迅速、高精度かつ大量検体処理用のゲノムワイドなマッピングや分子量測定を原理とする DNA 診断法を確立する。

(3) マイクロサテライト多型検索の自動化に関する研究

DNA 抽出, PCR 増幅, ゲルろ過スピンカラムの各行程は BIOMEK 型全自動ワークステーションで自動化されているので, スポットング以下の質量分析法による多型性検索までの行程を機能的に組み合わせて, 自動化を試みる。以上の研究により, DNA 抽出, PCR 増幅, マイクロサテライト多型検出までの一連の過程が自動化され, 迅速, 高精度かつ大量検体処理用の DNA 診断機器が開発されるであろう。

2. ゲノムワイドな多型マイクロサテライト 30,000 個の設定と各対立遺伝子の頻度の検索に関する研究 (東海大学医学部)

ゲノムシーケンシングなどの解析でえられたヒトゲノム塩基配列情報を利用して, 多型マイクロサテライトを 30,000 個収集・設定し, それらの多型マイクロサテライトを遺伝マーカーとして, 相関解析により複合遺伝疾患の原因遺伝子を同定することが, 本研究の最終目的である。そこでこの研究サブテーマでは, このような繰り返し数に多型をしめすマイクロサテライト 30,000 個をゲノムワイドに設定し, それぞれについて対立遺伝子の頻度を調べる。具体的には下記の 2 つの目標を設定する。

(1) 多型マイクロサテライト 30,000 個の設定に関する研究

Genethon の Weissenbach グループが白人での多型性マーカーとして DNA データバンクに登録している, ヒトマイクロサテライト 6,000 個について日本人で多型をしめすか否かを検討するため, 200 名の健常日本人の PCR 産物の混合物を, DNA シーケンサーを用いた電気泳動法により検討する。また, DNA データバンクに登録されているヒトゲノムシーケンスデータより, 2~5 塩基繰り返しマイクロサテライトをまず抽出し, それらのなかで 5 回以上の繰り返しをしめすマイクロサテライト (5 回以上の繰り返しをしめすマイクロサテライトは, 我々の経験上約 54% がヘテロ接合度 70% 以上の高度な多型性をしめす) を 50,000 個抽出して, PCR プライマーを設定する。これらについて同様の方法で, 日本人で多型をしめすマイクロサテライトを選別する。

(2) 多型マイクロサテライト 30,000 個の各対立遺伝子の頻度の検索に関する研究

前項(1)で多型が認められたマイクロサテライト 30,000 個について, 健常日本人 200 名由来 DNA を用いて, DNA シーケンサーや質量分析法により検索する DNA チップ技術により, 各マイクロサテライトの各対立遺伝子 (allele) 頻度を明らかにする。以上の研究により, 各対立遺伝子 (allele) 頻度が既知の 30,000 個の多型マイクロサテライトが収集・設定される。

3. 多型マイクロサテライトを用いた複合遺伝疾患の原因遺伝子の相関解析によるマッピングに関する研究 (東海大学医学部)

ゲノムシーケンシングなどの解析でえられたヒトゲノム塩基配列情報を利用して, 多型マイクロサテライトを 30,000 個収集・設定し, それらの多型マイクロサテライトを遺伝マーカーとして, 相関解析により複合遺伝疾患の原因遺伝子を同定することが, 本研究の最終目的である。そこでこの研究サブテーマでは, 前項目 2. で設定した多型マイクロサテライト 30,000 個を遺伝マーカーとして,

1. で開発した質量分析法により検索する DNA チップ技術を用いて, 複合遺伝疾患のゲノムワイドなマッピングを行い, 疾患原因遺伝子候補領域を 100 kb 以内に絞り込むことを目指す。具体的には下記の 1 つの目標を設定する。

(1) 多型マイクロサテライト 30,000 個を用いた複合遺伝疾患の原因遺伝子の相関解析によるマッピングに関する研究

代表的な多因子遺伝子疾患で, 原因遺伝子候補がすでに 15 個見いだされている若年性糖尿病 (IDDM1~IDDM15) をモデル疾患として, 前項目 2. で設定した, 全ゲノムをカバーする 30,000 個のマイクロサテライトマーカー (設定された多型マイクロサテライトより, 順次行う) に関し, 患者 200 人について繰り返し多型検索を DNA シーケンサーを用いた電気泳動法 (可能であるなら, 質量分析法により検索する DNA チップ技術) を行い, 相関解析 (P 検定, ハプロタイプ解析, Hardy-Weinberg 平衡からのずれの解析) による原因遺伝子のマッピングを試み, 本法の精度や感度を評価する。さらに, 同様の方法により多因子遺伝子疾患である, インスリン非依存性糖尿病 (NIDDM), 脳梗塞, 心筋症, リウマチ, グレーブス病, 高血圧, 無精子症の原因遺伝子のマッピングを行う。質量分析法により検索する DNA チップ技術が実行可能になった時点で, マイクロサテライトの繰り返し多型検索は, DNA チップで行い, マッピングの効率化を計る。

4. 絞りこまれた原因遺伝子候補領域のゲノムシーケンシングによる疾患の原因遺伝子の同定に関する研究 (東海大学医学部)

ゲノムシーケンシングなどの解析でえられたヒトゲノム塩基配列情報を利用して, 多型マイクロサテライトを 30,000 個収集・設定し, それらの多型マイクロサテライトを遺伝マーカーとして, 相関解析により複合遺伝疾患の原因遺伝子を同定することが, 本研究の最終目的である。そこでこの研究サブテーマでは, 前項目 3. で行った, 多型マイクロサテライトを用いたゲノムワイドな複合遺伝疾患のマッピングにより, 絞り込まれた原因遺伝子候補領域 100 kb について, 患者のゲノムシーケンシングを行い, 原因遺伝子の特定を目指す。同時に, 最終的に決定された原因遺伝子の位置を参考にして, マッピングに用いる相関解析プログラムを改善を試みる。具体的には次の 2 つの目標を設定する。

(1) 複合遺伝疾患の原因遺伝子候補領域内の発現遺伝子の

探索に関する研究

前項目3.の解析により、100kb前後に絞り込まれた複合遺伝疾患の原因遺伝子候補がまだ、ゲノム塩基配列が決定されていなければ、PACやBACクローニングとショットガンシーケンシング、またはPCR産物の直接塩基配列決定法で、ゲノム塩基配列を決定する。続いて、候補領域中に、BLAST、GENESCAN等のコンピュータプログラムによってエクソン配列やEST配列の存在を予測する。この情報をもとにして、RT-PCRやRACE、cDNAライブラリーのスクリーニングを行う。同時に、候補領域のエクソントラッピングを行い発現遺伝子を同定する。同定された発現遺伝子に関しては、各種ヒト組織由来のRNAを用いたRT-PCRおよびノザン解析により、発現組織を同定し、それぞれ発現量や鎖長などの情報を収集する。

(2) 患者の原因遺伝子候補領域内のPCR産物のゲノムシーケンシングと相関解析に関する研究

患者群で候補領域内の組換え体を含むと考えられる、少なくとも10個体程度の患者群と同数程度の健常者群のDNAプールを用いてlong-PCRにて候補領域を3kbごとに増幅したのち、シーケンシングにより塩基配列を決定し、同領域内における全gSNPs (genomic SNPs)を同定する。同定されたgSNPsで構成されるハプロタイプを推定し、候補領域内の組換えパターンを詳細に追跡して、候補領域をさらに絞り込む。同時に、前項目(1)で同定した候補領域内の発現遺伝子のエクソン内に存在する全cSNPs (coding SNPs)に関して、保有している全ての患者群と健常者群で検索を行い、アミノ酸多型とそれぞれの疾患との関連を相関解析により調べ、疾患の発症を規定するアミノ酸置換あるいは欠失・挿入を同定する。これによって、真の原因遺伝子を排他的に同定することが可能となる。

また、最終的に決定された原因遺伝子の位置を参考に、マッピングに用いる相関解析用プログラムGENEPOPの改善を試み、より正確なマッピング用プログラムの開発を試みる。

5. 多型マイクロサテライトの抽出プログラムの作成とデータベースの構築に関する研究 (国立遺伝学研究所生命情報研究センター)

ゲノムシーケンシングなどの解析でえられたヒトゲノム塩基配列情報を利用して、多型マイクロサテライトを30,000個収集・設定し、それらの多型マイクロサテライトを遺伝マーカーとして、相関解析により複合遺伝疾患の原因遺伝子を同定することが、本研究の最終目的である。そこでこの研究サブテーマでは、効率的な多型マイクロサテライトの30,000個のゲノムワイドな設定に資するため、多型マイクロサテライトを大量のゲノム塩基配列より取り出すプログラムの作成、並びに設定された多型マイクロサテライトの管理と他のゲノム研究者への情報提供に資するため、多型マイクロサテライトのデータベースの構築を行

う。具体的には下記の3つの目標を設定する。

(1) HTG配列の情報の各染色体毎の整理や管理に適したデータベースの構築に関する研究

今年春に発表されるドラフトシーケンス、すなわちギャップを含む90%精度の粗配列(HTG: high throughput generation配列と呼ばれる)配列情報を収集し、各染色体毎に整理する。次にこの情報をデータベース化して管理および更新を行う。これはHTG配列の整列化データを基幹とするデータベース(以下、HTGデータベースと呼ぶ)であり、このデータベースを利用してマイクロサテライトマーカーの抽出作業などを進めて行く。また、ESTや、既知遺伝子などの付加情報を付与する。実際には、HTGデータベースの構築はUNIXサーバーマシン上でORACLE8システムをベースにSQL (Structured Query Language)言語を用いて作成する。つまり、SQL-DDL (Data Definition Language)にてデータベースの仕様や定義を行い、SQL-DML (Data Manipulate Language)によりデータの入力、変更などの操作を行う。

(2) マイクロサテライトマーカーを抽出するためのソフトウェアプログラムの開発に関する研究

その時点までにえられている多型をしめすマイクロサテライト周辺の塩基配列から帰納的な推察をもとに、効率的に多型マイクロサテライトを抽出するため、さらにはプライマー作成を自動化するために、UNIX上でPerl言語およびC/C++言語を使用してソフトウェアプログラムの開発を行う。このプログラムによって生成された情報は、自動でHTGデータベースに付加情報として登録されるものとする。

(3) 多型マイクロサテライトデータベースの構築に関する研究

設定したマイクロサテライトの多型情報、染色体上の位置、PCRプライマー、PCR産物のサイズなどと実際にマイクロサテライトを用いて行ったマッピングデータを基幹とするデータベースを設計し、情報の整理・入力を逐次行っていく。さらに、塩基配列データ、付加情報の整理、更新によって、HTGデータベースの拡充をはかる。次に、より簡便なマッピング解析の為に、HTGおよびマッピングデータベースを包括した統合データベースの構築を行う。さらに、これらのデータを、WWWを利用してインターネット上で公開する。

3. 年次計画

本プロジェクトでは、ゲノムシーケンシングなどの解析でえられたヒトゲノム塩基配列情報を利用して、多型マイクロサテライトを30,000個収集・設定し、それらの多型マイクロサテライトを遺伝マーカーとして、相関解析により複合遺伝疾患の原因遺伝子を同定することを、本研究の最終目標とする。

上記の最終目標を達成するため、
 第Ⅰ期では、1) DNA チップによる、マイクロサテライト繰り返し多型の検索技術の開発 2) ゲノムワイドな 30,000 個の多型マイクロサテライトの設定
 第Ⅱ期では、1) 第Ⅰ期に、設定された 30,000 個の多型マ

イクロサテライトと、第Ⅰ期に開発された DNA チップによるマイクロサテライト多型の検索技術を用いた、複合疾患の原因遺伝子のマッピングと同定 2) 第Ⅰ期に設定された 30,000 個の多型マイクロサテライトのデータベースの構築をそれぞれ目標とする。

研究項目	12年度	13年度	14年度	15年度	16年度
1. 質量分析法によるマイクロサテライト多型検索のためのDNAチップ技術(MSチップ)の開発に関する研究	低密度MSチップの開発	高密度MSチップの開発		MSチップ自動化の開発	
2. ゲノムワイドな多型マイクロサテライト 30,000 個の設定と各対立遺伝子の頻度の検索に関する研究	多型マイクロサテライトの設定と対立遺伝子頻度の検索				
3. 多型マイクロサテライトを用いた複合遺伝疾患の原因遺伝子の相関解析によるマッピングに関する研究			疾患の原因遺伝子の相関解析によるマッピング		
4. 絞り込まれた原因遺伝子候補領域のゲノムシーケンシングによる疾患の原因遺伝子の同定に関する研究				疾患の原因遺伝子候補領域の発現遺伝子の検索	
5. 多型マイクロサテライトの抽出プログラムの作成とデータベースの構築に関する研究	HTG配列のデータベースの構築およびマイクロサテライトマーカーを抽出するためのソフトウェアの開発		多型マイクロサテライトデータベースの構築		
所要経費(合計)	243百万円				

4. 平成12年度における実施内容と達成目標

1. 質量分析法によるマイクロサテライト多型検索のための低密度DNAチップ(MSチップ)技術の開発に関する研究

SNP部位及び2塩基GA繰り返しマイクロサテライトを含む領域をPCR増幅し、えられたPCR産物をスライドグラスに、1cm画のチップに低い密度(100スポット/1チップ)でスタンピングする。このチップに、MALDI-TOF型質量分析計を用いてレーザーを照射して、分子量を測定する。このとき、十分な距離に分離して、多型検索が明瞭になるような至適条件を検討をする。また、分子量が20,000以上の高い3~5塩基繰り返しマイクロサテライトについても、チップ上で質量分析法を用いた分子量測定で、多型性検索が可能になるように、高い分子量で高分解能をしめす期待される炭酸ガス赤外線レーザーによる照射、高加速電圧の適用やイオン化条件の検討を行う。

2. ゲノムワイドな多型マイクロサテライト 20,000 個の設定と各対立遺伝子の頻度の検索に関する研究

GenethonのWeissenbachグループが白人での多型性マーカーとしてDNAデータバンクに登録している、ヒトマイクロサテライト6,000個について日本人で多型をしめすか否かを検討するため、200名の健常日本人のPCR産物の混合物を、DNAシーケンサーを用いた電気泳動法により検討する。また、DNAデータバンクに登録されているヒトゲノムシーケンスデータより、5回以上の繰り返しをしめすマイクロサテライトを30,000個抽出して、同様の方法で日本人で多型をしめすマイクロサテライトを選別する。予想では、平成12年度中に日本人での多型マイクロサテライト20,000個の収集・設定を完了する。多型が認められたマイクロサテライト20,000個について、健常日本人200名由来DNAを用いて、DNAシーケンサーや質量分析法により検索するDNAチップ技術により、各

マイクロサテライトの各対立遺伝子 (allele) 頻度を順次明らかにする。平成 12 年度中には、約半分の多型マイクロサテライト 10,000 個について、決定することを目標とする。

5. HTG 配列のデータベースの構築と多型マイクロサテライトの抽出プログラムの作成に関する研究

今年春発表されるドラフトシーケンスの HTG 配列情報を収集し、各染色体毎に整理する。次にこの情報を UNIX サーバマシン上で、ORACLE8 システムをベ-

スに SQL (Structured Query Language) 言語を用い、EST や既知遺伝子などの付加情報を付与して、データベースを構築する。また、このデータベースを利用してマイクロサテライトマーカーの抽出作業を行う。多型をしめすマイクロサテライト周辺の塩基配列から帰納的な推察をもとに、効率的に多型マイクロサテライトを抽出するためのプログラムを作成する。さらにはプライマー作成を自動化するために、UNIX 上で Perl 言語および C/C++ 言語を使用してソフトウェアプログラムの開発を行う。

II 平成 12 年度における実施体制

研 究 項 目	担 当 機 関	研究担当者
1. 質量分析法によるマイクロサテライト多型検索のための DNA チップ技術 (MS チップ) の開発に関する研究	㈱島津製作所分析機器部	川 畑 慎一郎
2. ゲノムワイドな多型マイクロサテライト 30,000 個の設定と各対立遺伝子頻度の検索に関する研究	東海大学医学部	安 藤 麻 子
3. 多型マイクロサテライトを用いた複合遺伝疾患の原因遺伝子の相関解析によるマッピングに関する研究	東海大学医学部	田 宮 元
4. 絞り込まれた原因遺伝子候補領域のゲノムシーケンシングによる疾患の原因遺伝子の同定に関する研究	東海大学医学部	椎 名 隆
5. HTG 配列のデータベースの構築と多型マイクロサテライトの抽出プログラムの作成に関する研究	文部省国立遺伝学研究所生命情報研究センター	五條掘 孝
6. 研究課題の推進	東海大学医学部	猪 子 英 俊

Ⅲ 研究推進委員会

委 員	所	属
[プロジェクト内委員]		
○猪 子 英 俊	東海大学	医学部分子生命科学系遺伝情報部門教授
安 藤 麻 子	東海大学	医学部分子生命科学系遺伝情報部門講師
川 畑 慎 一 郎	(株)島津製作所	分析機器事業部ライフサイエンス部主任
木 村 稔	東海大学	医学部分子生命科学系遺伝情報部門教授
五 條 堀 孝	文部省	国立遺伝学研究所生命情報研究センター遺伝情報分析教授
椎 名 隆	東海大学	医学部分子生命科学系遺伝情報部門助手
田 宮 元	東海大学	医学部分子生命科学系遺伝情報部門助手
[プロジェクト外委員]		
木 村 彰 方	東京医科歯科大学	難治疾患研究所成人疾患研究部門教授
笹 月 健 彦	九州大学	生体防御医学研究所遺伝部門教授
田 代 弘 行	(株)不二家	バイオサイエンス研究所所長
谷 口 寿 章	藤田保健衛生大学	総合医科学研究所助教授
辻 省 次	新潟大学	脳研究所神経内科学教室教授
徳 永 勝 士	東京大学	大学院医学系医学研究科人類遺伝学教室教授

(注：○は研究推進委員長)

Ⅳ 研究連絡会議

委 員	所	属
安 藤 麻 子	東海大学	医学部分子生命科学系遺伝情報部門講師
安 東 正 徳	(株)島津製作所	分析機器事業部ライフサイエンス部副主任
池 尾 一 穂	文部省	国立遺伝学研究所生命情報研究センター遺伝情報分析助手
猪 子 英 俊	東海大学	医学部分子生命科学系遺伝情報部門教授
岡 晃	東海大学	医学部分子生命科学系遺伝情報部門大学院
川 畑 慎 一 郎	(株)島津製作所	分析機器事業部ライフサイエンス部主任
館 野 義 男	文部省	国立遺伝学研究所生命情報研究センター遺伝情報分析教授
木 村 稔	東海大学	医学部分子生命科学系遺伝情報部門教授
五 條 堀 孝	文部省	国立遺伝学研究所生命情報研究センター遺伝情報分析教授
椎 名 隆	東海大学	医学部分子生命科学系遺伝情報部門助手
田 宮 元	東海大学	医学部分子生命科学系遺伝情報部門助手
野 上 正 弘	東海大学	医学部分子生命科学系遺伝情報部門奨励研究員
バトムンフ ムンフバト	東海大学	医学部分子生命科学系遺伝情報部門特定研究員
林 英 樹	東海大学	医学部分子生命科学系遺伝情報部門特定研究員
牧 野 悟 士	東海大学	医学部分子生命科学系遺伝情報部門大学院
松 坂 恭 成	東海大学	医学部分子生命科学系遺伝情報部門大学院
東 内 健 一	東海大学	医学部分子生命科学系遺伝情報部門特定研究員
山 形 哲 司	東海大学	医学部分子生命科学系遺伝情報部門奨励研究員