

[別紙 1]

論文の内容の要旨

論文題目 データベースを使用した遺伝子機能解析支援システム

指導教官 金澤一郎教授

東京大学大学院医学系研究科

平成10年4月入学

医学博士課程

脳神経医学専攻

氏名 村上 大勇

ディファレンシャル・ディスプレイ (Differential Display、以下 DD) は、mRNA より作成した cDNA (群) に、より多くの cDNA と hybridize すると思われる塩基配列を持ったプライマーと、3'-末端の Poly-A 配列に対応するプライマーとで PCR 法を行う実験法であり、多数の遺伝子についてその発現を同時かつ系統的に探索可能な優れた手法である。

これまでの DD におけるデータベースの使用は、主に得られた遺伝子の塩基配列決定後にデータベースから遺伝子の種類を推定するだけであった。また、これまで単一の DNA 塩基配列より PCR 法のプライマーを選定するプログラムはあったが、より多くの塩基配列をカバーするプライマーを選択するシステムはなく、現在のプログラムを使って実現するのが困難であった。DD 法で使用するプライマーの選択については、これまでは経験則に頼るしかなかった。また、どのプライマーを選べば重複が少なく新規遺伝子を選べるかも、知る手段が無かった。

また、DD の実験産物をゲル電気泳動してパターンを見ても、ある特定の遺伝子が、どのような位置に、バンドとして現れるかを予測するのは、既存のデータベース、バイオインフォマティクス用ソフトウェアでは困難であった。あるバンドについて、それが具体的に何の遺伝子であるかを調べる為には、実際にそのバンドをゲルより取得して DNA 塩基配列を決定するしか方法はなかった。

DD は RT-PCR により検出精度が高く、安価で、比較的簡便な実験法である。DNA アレイのように膨大な量の PCR 産物とプライマーを予め用意する必要がないので、多数の遺伝子の発現量を測定

するのにコンピューター化された DD は有用であると考えられる。

そこで我々は DD 実験の結果を予測するコンピュータープログラムを作成した。実際の DD 電気泳動バンドパターンと比較し、データの解析を行うことで、DD の実効性を高めることが出来る。その遺伝子データベースを使用した選択された DD プライマーを使用した DD 結果予測を行う新規コンピューターシステムを *in silico* Differential Display (*isDD*) と名付けた。

isDD システムは JAVA スクリプトを含むダイナミック HTML (ハイパーテキスト・マークアップ・ランゲージ) で作成され Windows® 2000™ ベースの IBM PC/AT 互換機の上で開発、クライアントプログラム及び IIS (インターネット・インフォメーション・サーバ) 上で実行される。

まず、Genbank または UniGene DNA 塩基配列データベースの DD に使用可能なエントリーを取得し、すべての cDNA 塩基配列から、DD の実際の実験で検出出来る cDNA 3'-端 1000 塩基または 2000 塩基以内により多く現れる配列をコンピュータープログラムにより総当りで調べ列挙した。その際、パンドローム禁止、3つ以上続いた T 配列を含まない、より多い T 配列を含んだものは除外、という条件付けをした。

より多くの cDNA の発現を確かめられるプライマーの組み合わせを、遺伝的アルゴリズム (GA) によって選択した。GA による選択では 99.70% の *isDD* データベース中のエントリーをカバーできたのに対し、上位 20 個のプライマーからランダムに選択する場合 98.0% のカバー率に留まることが示された。

さらに、その選択されたバンドがどのような電気泳動パターンを示すかを予測するプログラムを作成した。*isDD* バンドビューは Windows (R) ベースのプログラムで、フルスクリーンの DirectDraw 方式を使い高速描画を行う。プログラムは実際の電気泳動ゲルの画像を白黒 256 階調の BMP (ウィンドウズ・ビットマップ)、フルカラー (24bit) のビットマップに変換したものを読み込み、シミュレーションのゲルイメージと重ね合わせられる。

加えて、我々は WWW ベースのバンドビューワーシステムも用意した。*isDD* コモン・ゲートウェイ・インターフェース (CGI) は WWW サーバー上で実行される。クライアントの WWW ブラウザーでの表示は、ダイナミック HTML (DHTML) を使用したインタラクティブなアプローチを取っており、ウィンドウズ版と同様に *isDD* エントリー名をマウスカーソルが横切ると、対応するバンド位置が赤色で示される。エントリーをクリックするとそのエントリーの塩基配列が別ウィンドウで表示され、カット&ペーストすることで次のプログラムまたは Web サイトを使用して解析可能となる。

次に、市販酵母 RNA (Stratagene®) を使用して実際の DD 実験を行った。合計 20 個のバンドをゲルより取得し、TA ベクターにサブクローニング後、それぞれのバンドについて 20 個の cDNA クロオンを蛍光シーケンシング法で塩基配列を読んだところ、*isDD* データベースに含まれる ORF と EST のエントリー上で 5 塩基以上の精度で実際に存在したものが相当するクローン中の 4 個あり、さらに 2 クロオンは EST のエントリー上でのみ一致した。それ以外の *isDD* の結果に出てこない 12 のクローンのうち少なくとも 6 個は Poly-A 配列を持った新規遺伝子で、酵母染色体上の推定 ORF 配列と推定 ORF 配列の 1000 塩基以上離れた場所に存在した。その 12 クローンのうち 9 はオープン・リーディング・フレームを欠いており、リボソーム RNA のような翻訳されないで使用される RNA であると考えられる。以上の結果から *isDD* でのシミュレーションの効率率は *isDD* で予測できるものにおいては 5 割、ORF 全体では約 20% 程度であるが、*isDD* に使用可能な 3'-端を含む遺伝子データ

ベースが充実するか、推定 ORF の精度が上がる、EST の数が増える等の手法により改善可能であると考えられる。

実際の DD のゲルイメージで変化が観察されたバンドは取得して解析する価値があり、*isDD* でその遺伝子の名称を絞り込んで予測することができる。*isDD* で見られないバンドは新規遺伝子である可能性が高い。多数の遺伝子発現の有無を、統一的に確認することが出来る。数本のプライマーで例えば酵母の場合、6 塩基配列のプライマー 8 種類の組み合わせで、全 Genbank と EST の *isDD* エントリー中の、99%以上の発現を調べられると考えられる。これは、従来の莫大な量（数千）のプライマーを用いて PCR を行い、DNA アレイ等を使用して調べる方法よりも、非常に安価で簡便に、さまざまな条件下での遺伝子発現についての知見が得られる。将来的には、シグナル伝達系、代謝系等のカテゴリー別のエントリー群が、どのようなバンドパターンを描くか予測して示すことが考えられる。

現在、さまざまな種のゲノムプロジェクトが進行中である。特にヒトとマウスに関しては、データベース中の発現遺伝子プロファイリングが連日のように更新されている。*isDD* システムは新規に作成されたデータベースに対して実行できるので、将来的にさらに効率が高くなることを期待出来、DD 法による遺伝子の機能解析と新規遺伝子取得の一助になることが期待される。