

DNAコンピュータによるSNPsおよび遺伝子発現解析法の開発

●陶山 明

東京大学大学院総合文化研究科

＜研究の目的と進め方＞

DNAコンピュータはDNAの分子反応を利用して計算処理を行う分子コンピュータである。その超並列処理能力に加えて、DNAやRNA分子を入力データとして直接受理してその情報を処理できる特徴を利用すると、DNAコンピュータにより生体分子の多様な情報解析を正確かつ高速に行うことができる。我々はDNAコンピュータを用いて遺伝子の発現解析を行う方法を小規模な実験により世界ではじめて示し、従来のDNAチップやマイクロアレイを用いる方法と比較して、汎用性、定量性、演算性の点で優れていることを明らかにした。本研究では、その成果を踏まえて、DNAコンピュータを用いて大規模なSNP解析および遺伝子発現解析を行う方法（図1）の基盤を確立する。

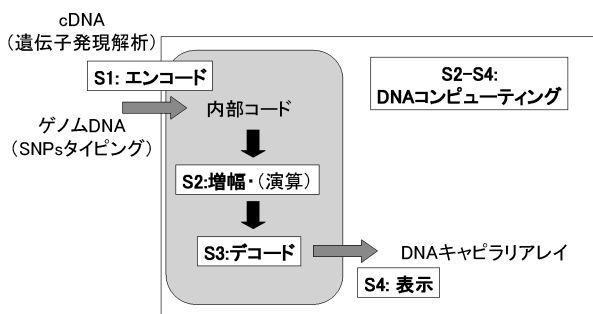


図1 DNAコンピュータによるSNP及び遺伝子発現解析法

＜研究開始時の研究計画＞

以下の項目の研究を実施することにより、DNAコンピュータを利用した大規模SNPタイピング法と遺伝子発現解析法の基盤を確立する。

- 1) DNAコンピュータを用いた大規模なSNPおよび遺伝子発現解析で必要とされる要素技術の開発を行う。そのために、
 - ・大規模な正規直交配列の設計と評価
 - ・DNAキャピラリアレイによる検出の定量性の評価
 - ・DNAコード化数（DCN）の増幅の定量性の評価
 - ・大規模な解析のためのエンコード反応条件の確立を行う。
- 2) DNAコンピュータを用いた並列性の高いSNPタイピング法を開発する。そのために、
 - ・SNPタイピングのためのエンコード反応の開発
 - ・エンコード反応の特異性の検討
 - ・マルチプレックスSNPタイピング実験を行う。
- 3) 遺伝子発現解析における定量性を向上させるための技術を開発する。そのために、
 - ・エンコードステップの定量性の評価
 - ・エンコードの定量性を向上させる方法の開発
 - ・合成DNAを用いた定量性の評価
 - ・他の発現解析法との定量性の比較を行う。

4) SNPのアリルパターン及び遺伝子発現パターンの判定をDNAコンピュータにより行う技術を開発する。

＜研究期間の成果＞

1. 大規模解析のための要素技術の開発[1,5,6]

1.1 大規模な正規直交配列の設計と評価

DNAコンピュータを用いた大規模なSNP解析や遺伝子発現解析において必要とされる多数の正規直交配列の設計とその特性の実験による評価を行った。これまでのDNAコンピュータの研究において開発された設計方法を改良することにより、ハイブリダイゼーション反応において正規直交性が保たれる条件（融解温度幅2℃、ハミング距離8塩基、最大連続一致長7塩基）の下で、500個を超える正規直交配列を生成することに成功した。500個の正規直交配列があると、たとえば、2桁のDNAコード化数(DCN)を用いて、最大で約6万個のSNPや標的遺伝子を同時に解析することができる。

設計された配列の中からランダムにいくつかの配列を選び、実験により正規直交性を調べた。融解温度の測定、融解曲線を利用したハイブリダイゼーション効率の測定、DNAチップによるミスハイブリダイゼーションの測定を行った結果、必要十分な正規直交性が確認された。

1.2 DNAキャピラリアレイによる検出の定量性の評価

DNAコンピュータを用いた大規模なSNP解析や遺伝子発現解析では、正規直交配列をもつプローブを固定したDNAキャピラリアレイあるいはDNAチップに、蛍光ラベルされたデコード産物をハイブリダイゼーションすることにより検出と定量が行われる。解析の内容が異なっても同じ正規直交配列をもつユニバーサルなDNAキャピラリアレイを使用することができるため、有効性の評価の手間が省けるだけでなく、大量製造によるコストの削減と信頼性の向上を図ることができる。デコード産物も正規直交配列しか含まないため、検出の際のエラーは小さく、しかも定量性の高い検出が可能である。

100種類の正規直交配列を固定したユニバーサルDNAキャピラリアレイを用いて検出の定量性を評価したところ、一色法では3桁、別の蛍光分子でラベルした一定濃度のD1相補鎖DNAを競合ハイブリダイズさせた二色法では5桁弱のダイナミックレンジをもつ高い定量性が得られた。二色法を用いると、固定化されたプローブ量の影響を補正することができるため、より高い濃度まで定量性が得られた。

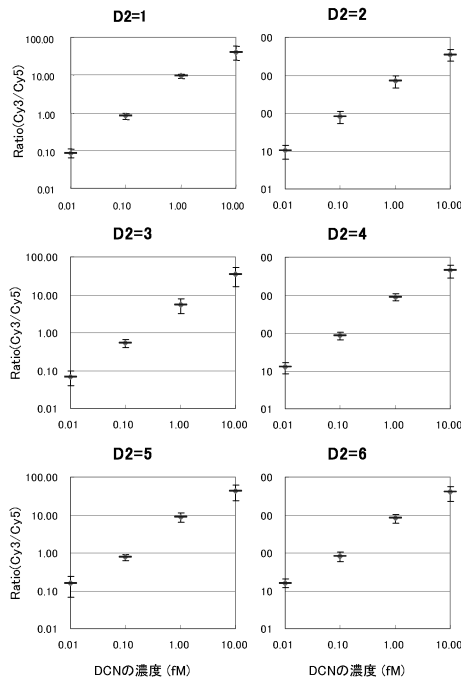


図2 DCNの定量的増幅

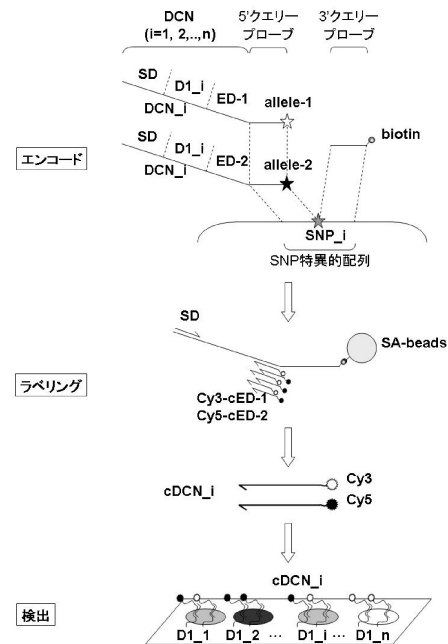


図3 DNAコンピュータによるSNPタイピング法

1.3 DCNの増幅の定量的性の評価

多数のSNP部位におけるアリの同時解析及び多数の遺伝子の発現量の同時解析といった大規模同時解析では、解析のための反応に利用できる個々のSNP部位を含むDNA断片及び個々の遺伝子の転写産物量は少なくならざるを得ない。したがって、このような解析において、検出感度を上げるための増幅は重要である。

DNAコンピューティング技術を用いた解析では、シグナルの増幅は内部コードであるDCNをPCRで増幅することにより行われる。DCNは正規直交配列を連結して作られた、DNA配列で表現された数字である。同じ集合から選択した正規直交配列を使用しても、DCNの形式は解析の内容によって異なる。本研究で用いたDCNの形式は、SNPタイピングではSD-D1-EDであり、遺伝子発現解析ではSD-D1-D2-EDである。ここで、SD、D1、D2、EDは正規直交配列で、D1及びD2は個々のDCNごとに異なるが、SD及びEDはすべてのDCNで共通である。多種類のDCNの混合物はSDとEDを共通のプライマー対の配列とするPCRにより増幅される。

同じ性質を有する正規直交配列を連結して作られたDCNであっても、最初の濃度の相対分布を崩さずに増幅される保証はない。そこで、D1=0~99、D2=0~5の600種類のDCNを等量ずつ加えた混合物をPCR増幅し、ほぼ同じ増幅産物量を示した300種類のDCNを選択した。これらのDCNは比較的一様な増幅効率を有していると考えられる。選択された300種類のDCNの混合物が最初の相対濃度分布を崩さずに増幅されるか否かを調べるために、様々な相対濃度比の混合物を作り、増幅実験を行った。その結果、図2に示したように、共通のプライマー対を用いて相対濃度分布を崩さずに増幅されることが確認された。

1.4 大規模解析のためのエンコード反応条件の確立

DNAコンピュータを用いた解析の最初の処理はエンコード反応である。この処理ではTaq DNAリガーゼ酵素によるライゲーション反応が用いられる。多変換テーブル分子を用いた大規模なSNPや遺伝子発現解析のエンコード反応を酵素の量を変えないで行うためには、変換テーブル分子の濃度を下げなくてはならない。

同一反応チューブ内で使用する変換テーブル分子の種類は、解析対象にも依るが、100から3,000種類程度である。遺伝子発現解析の場合、解析対象の種類はそれよりも多いが、エンコード反応の効率を一律に保つために標的的同定配列の融解温度を一律にすると、同一反応チューブの中で同時に処理できる解析対象の種類はその程度になる。

そこで、通常使用する酵素量でこの程度の種類の変換テーブル分子を用いたエンコード反応を行うことができるか否かを調べた。同時に解析する対象が1,000種類の場合、通常使用する酵素量で変換テーブル分子の量は50 pmol/1000 = 50 fmol程度になるが、10 fmolの変換テーブル分子を用いた15分のエンコード反応でも、変換後のPCRの条件を調整することにより十分な検出感度が得られることがわかった。

2. SNPタイピング法の確立[7,9]

2.1 SNPタイピングのためのエンコード反応の開発

DNAコンピュータによるSNPタイピング法は、エンコードとDNAコンピューティングの大きく2つのステップから構成される(図1)。SNPのアリのタイプはエンコードのステップで対応するDCNに変換される。DCNの形式としては、3つの正規直交配列、SD、D1、EDを連結したSD-D1-EDを用いた。SDはすべてのDCNで共通である。

D1は各DCNにごとに異なり、SNPの位置を識別するために使用した。EDは2種類（ED1及びED2）で、各SNP位置でのアリのルのタイプを表すために用いた。開発した方法では、Taq DNAリガーゼ酵素を用いたライゲーション反応により複数箇所のSNP情報をDCNへ一括して変換した後、SDとED1/ED2を用いた非対称PCRによりDCNを増幅し、D1配列をもつプローブを固定したDNAキャピラリアレイにハイブリダイズして各箇所でのSNPのアイルタイプを決定する（図3）。ED1とED2をそれぞれCy5とCy3でラベルすることにより、DNAキャピラリアレイの各スポットの蛍光色から、アリのルのタイプ、ヘテロカホモの判定を直接行うことができる。

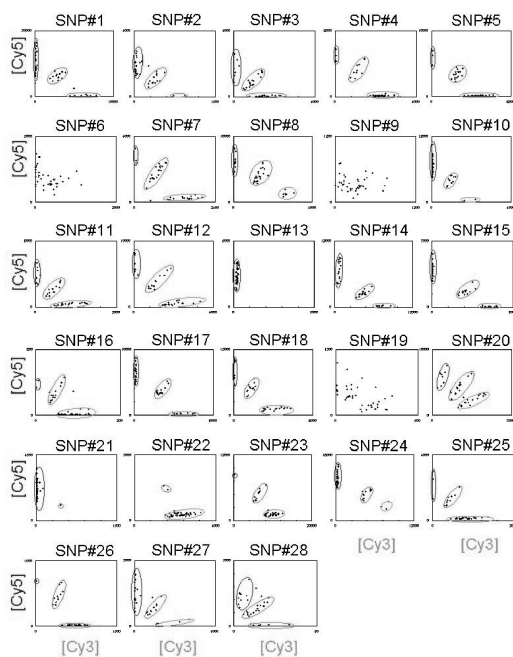


図4 28マルチプレックスSNPタイピング実験の結果

2.2 エンコード反応の特異性の検討

遺伝子発現解析の場合とは異なり、SNPタイピングにおいては、エンコード反応の特異性が最も重要である。そこで、エンコードステップの特異性の検討を行なった。エンコードのステップでSNPのアリルタイプを対応するDCNに変換する際、変換テーブル分子を構成するクエリープローブの5'末端にSNP部位を設定する方法と3'末端に設定する方法が考えられる。どちらの方法の特異性が高いか、SNPタイプとクエリープローブの可能なすべての組み合わせについて、エンコード反応の温度及び変換テーブル分子を構成するコモンプローブの長さを変えながら、エンコードの特異性を調べた。その結果、3'末端にSNP部位を設定した場合に高い特異性が得られることがわかった。最も識別が困難なG/Tミスマッチの場合、5'末端ではミスエンコードにより正確なタイピングを行うことができなかったが、3'末端ではミスエンコードされたシグナルは5%以下になり、正確なSNPタイピングが可能であることがわかった。

2.2 マルチプレックスSNPタイピング実験

ヒト一番染色体上に存在するPLOD遺伝子上の1箇所のSNP及びヒト五番染色体上のIL4、IL13を含む約500kbの領域に存在する28箇所のSNPを対象としたマルチプレックスSNPタイピングを40検体で行った。その結果、多型が見られなかったSNP#13とクラスター分離の悪かった3箇所のSNP（#6、#9、#19）を除く24箇所においてコール率は99.2%となることがわかった（図4）。また、シーケンシング結果との一致率は100%となり、2回の独立した実験から再現性は、2箇所のSNPを除いて、R二乗値で0.99以上となることがわかった。多型が見られた27箇所のうち、24箇所のSNPでは3つのクラスターに分離されている様子が観察され、SNPタイピングに成功していることがわかった（タイピング成功率88.8%）。しかし、SNP#6、#9、#19ではクラスター分離が悪く、SNPタイピングはできなかった。そこで、5'クエリープローブの3'末端から4塩基目の位置にミスマッチを導入して再度マルチプレックスSNPタイピングを行ったところ、SNP#6とSNP#9においてクラスター分離が改善されSNPタイピングに成功した。この結果、ミスマッチ導入プローブを用いた際のタイピング成功率は96.3%となり、既存の方法と比べて高い成功率が得られた。SNP#19だけがタイピングに失敗した原因は、SNP部位を含む入力DNAの量が他よりも少なく、シグナル強度が弱かったことにあることがわかった。

3. 遺伝子発現解析の定量性を向上する技術の開発 [2,3,8,10]

3.1 エンコードステップの定量性の評価

DNAコンピュータによる遺伝子発現解析は、エンコードとDNAコンピューティングの大きく2つのステップに分けられる（図1）。エンコードのステップでは、遺伝子発現の情報であるcDNAの量の分布がDNAコンピュータの内部コードであるDCNの量の分布に変換される。その後、DNAコンピューティングのステップで、DNAコンピュータの命令を用いて、DCNの増幅と解析のための演算処理などが行なわれる。DNAコンピュータによる遺伝子発現解析の定量性は、これら2つのステップの定量性に左右される。DNAコンピューティングのステップの定量性は1.2及び1.3において確認されたので、さらに、エンコードのステップの定量性を調べた。

エンコードのステップの定量性を左右する大きな要因は、ライゲーション反応の効率の遺伝子同定配列依存性である。そこで、その依存性を酵母遺伝子の同定配列をもつ100種類のDNAオリゴ（30塩基長）を用いて詳細に調べた。その結果、一様な融解温度をもつ遺伝子同定配列であっても、配列により有意に反応効率が異なることがわかった。配列に関する性質、たとえば、ライゲーション部位の塩基対スタックの安定性、その部位に隣接する塩基の種類、両側の配列部分の安定性などとライゲーション反応効率との関係を調べたが、顕著な関係は見出されなかった。また、エンコードのための変換テーブル分子を調製する際に生じる活性の非一様性にも原因があることがわかった。変換テーブル分子のリン酸化の割合、ライゲーション反応に寄与できる正しい構造を取っている変換テーブル分子の割合などがすべての遺伝子同定配列に対して同一であるかどうかは不明確である。これらのことから、解析対象の遺伝子の数が少ない場合は可能であっても、遺伝子の数が多い場合は、遺伝子同定配列の設計を工夫してエンコード反応効率を一様にすることは難しいことがわかった。

3.2 エンコードの定量性を向上させる方法の開発

エンコードステップのライゲーション反応の効率は遺伝子同定配列に依存したが、いずれの遺伝子についても同定配列をもつDNAオリゴの濃度に比例することがわかった。そこで、この性質を利用してエンコード反応効率の遺伝子同定配列依存性を補正する、規格化エンコード法を開発した。この方法を用いると、変換テーブル分子の活性の不揃いに起因する場合であっても補正が可能である。また、仮にDCNの増幅効率やデコード効率にばらつきもあっても、それによりシグナル強度が極端に小さくならない限り、それらの影響も補正できる。

規格化エンコード法では、発現解析を行うcDNAサンプルだけでなく、同じ変換テーブル分子を含む反応液を用いて参照溶液のエンコードも同時に行う(図5)。参照溶液中には解析対象遺伝子の同定配列をもつDNAオリゴが一定の濃度で含まれている。その後、両者を増幅・デコードし、ユニバーサルDNAキャピラリーアレイにハイブリダイズして定量を行う。cDNAサンプルの定量値を参照溶液の対応する配列の定量値で補正することにより、cDNAを正確に定量することができる。補正は数値的に行うことも可能であるが、デコードの際にcDNAサンプルと参照溶液の反応産物を異なる蛍光色素でラベルすると、競合ハイブリにより直接補正することが可能である。

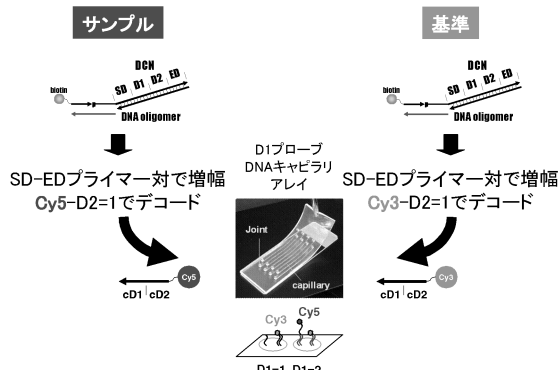


図5 規格化エンコード法 (NE法)

規格化エンコード法により、エンコード反応効率の非一様性などが補正できる理由を以下に示す。サンプル溶液中の遺伝子*i*のcDNA濃度を $C_S(i)$ とする。エンコード反応での遺伝子*i*の変換効率を $\gamma(i)$ とすると、遺伝子*i*のcDNAがエンコードされて生成されるDCNの量は $\gamma(i)C_S(i)$ となる。ただし、ライゲーション反応がMichaelis-Menten型で、基質の濃度がMichaelis係数よりも低いとする。エンコードされたDCNを*n*サイクルのPCRで増幅すると、 $\beta_{PCR}(i)\gamma(i)C_S(i)\exp(\alpha(i)n)$ のPCR産物が得られる。ただし、 $\alpha(i)$ は遺伝子*i*に対応するDCNの増幅率(1.3で述べたように、この*i*依存性は小さい)、 $\beta_{PCR}(i)$ は適当な比例係数である。その後のデコードの効率を $\beta_{DCD}(i)$ (この*i*依存性も小さい)とすると、チップへのハイブリダイゼーション溶液中の標的産物量は

$$\beta_{DCD}(i)\beta_{PCR}(i)\gamma(i)C_S(i)\exp(\alpha(i)n)$$

となる。同様に、参照用溶液については、すべての遺伝子*i*に対して同じ濃度 C_R の合成DNAオリゴを使用するので、チップへのハイブリダイゼーション溶液中の標的産物量は

$$\beta_{DCD}(i)\beta_{PCR}(i)\gamma(i)C_R\exp(\alpha(i)n)$$

となる。サンプルと参照用溶液からのデコード産物溶液を適当な量比で混合し、競合ハイブリさせると、測定される蛍光強度比は

$$\frac{\beta_{DCD}(i)\beta_{PCR}(i)\gamma(i)C_S(i)\exp(\alpha(i)n)}{\beta_{DCD}(i)\beta_{PCR}(i)\gamma(i)C_R\exp(\alpha(i)n)} = \frac{C_S(i)}{C_R}$$

に比例することになる。したがって、エンコード反応効率の非一様性などは相殺される。比例係数の値は不確定なので、蛍光強度の比から直接に $C_S(i)$ の値を決定することは難しい。しかし、サンプル溶液中に既知濃度のサンプル(合成DNAあるいは可能ならcDNAでも良い)を入れておき、それらに対する蛍光強度比から作成した検量線を用いれば、 $C_S(i)$ の絶対定量を行うことができる。

3.3 合成DNAを用いた定量性の評価

規格化エンコード法の定量性を評価するために、遺伝子同定配列をもつ濃度が既知の30塩基長DNAオリゴの混合溶液を定量する実験を行った。100種類のDNAオリゴを1 amolから100 amolの範囲の量で混合したサンプルの定量を行ったところ、図6aに示すように、精度よく定量できることがわかった。

3.4 他の発現解析法との定量性の比較

DNAコンピュータによる遺伝子発現解析法の有効性を調べるために、規格化エンコード法によりcDNAの定量を行い、他の発現解析法による結果と比較した。実験には、

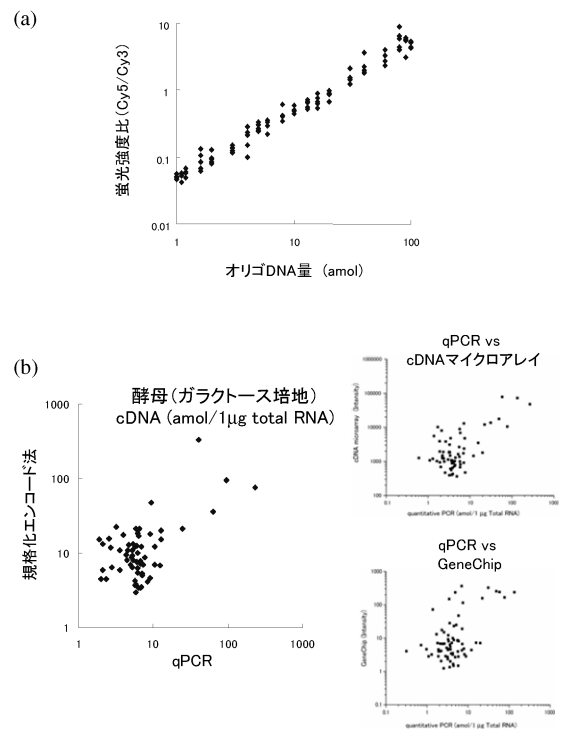


図6 (a) NE法による合成DNAオリゴの定量。(b) NE法、市販DNAチップ及びDNAマイクロアレイによるcDNAの定量結果と定量的PCRによる定量結果との比較。

ヒト培養細胞 (MOLT-4及びRAJI) とガラクトース培地で培養した出芽酵母から調製したcDNAサンプルを使用した。これらのサンプル中のcDNA量は、市販されているDNAチップやDNAマイクロアレイ、定量的PCRにより別途、調べた。

1 μ gのトータルRNAを用いて、約100種類の遺伝子の発現を同時に解析した結果、他の発現解析法と同等か、それ以上に高い再現性のある結果が得られた。しかし、最も多くの人に信頼されている定量法である定量的PCRによる結果と比較したところ、十分に高い相関は得られなかった。相関の低さは、市販されているDNAチップやDNAマイクロアレイによる結果を定量的PCRによる結果と比較したときも同様であった。DNAコンピュータ、DNAチップ、DNAマイクロアレイの中では、DNAコンピュータによるcDNAの定量結果が定量的PCRによる結果と最も高い相関を示した (図6b)。

それぞれの解析方法の再現性は高いが、互いの結果の一致は良くないという状況において、どの解析方法の結果が最も正しいのか、結論はまだ得られていない。サンプル中に存在する各遺伝子のcDNAの正しい量は誰も知らないで、どの方法の結果が最も正しいかを判定することは容易でない。しかし、少なくとも、DNAコンピュータによる方法は、正解がわかっている合成オリゴDNAのサンプルでは正しい結果を与えることが実証された。cDNAは合成オリゴDNAより長いため、二次構造がエンコード反応の効率に及ぼす影響による誤差が心配される。高温にしてその影響を回避するようにエンコード反応を行っていること、DNAコンピュータによる結果と定量的PCRによる結果が最も高い相関を示していることを考えると、DNAコンピュータによる方法は、高い並列性を維持しながら精度の良い定量が可能であると考えられる。

4. SNP及び遺伝子発現のパターン判定法の開発[2,4]

DNAコンピュータによる方法が備えている演算性を利用して、SNPのアリルタイプや遺伝子の発現パターンを判定する方法、パターンの探索を行う方法を考案した。方法は、人手やロボットを用いて解析のための反応操作を実行する他律的方法と、外部からの操作無しに一定の温度の下で解析のための一連の反応を実行する自律的方法に分けられる。

他律的方法について、遺伝子の発現パターンを判定する場合を例にして、方法の概要を示す。パターン判定のためのDNAコンピューティングは、発現遺伝子に対応するDCNを増幅するステップ、増幅されたDCNから発現状態を二値化した別の内部コードを生成する二値化データ生成ステップ、それが発現パターンを表す論理式を満たすか否かを判定する論理式判定ステップから構成されている (図7)。二値化データ生成ステップでは、2つに分けられたDCNは、それぞれget+命令とget-命令により処理される。発現量が一定量以上の場合、get+命令の実行により発現されていることを表す内部コードが出力される。そのとき、get-命令の実行は何も出力しない。発現量が一定量以下の場合はその逆で、get+命令の実行は何も出力しないのに対して、get-命令の実行は発現されていないことを表す内部コードを出力する。それらの内部コードを一緒にしたのち、論理式の内容を表す引数を用いたappend命令を実行し、発現パターンが指定した論理関係を満たしているか否かを判定する。論理関係を満たしていると、蛍光ラベルが取り込まれ、磁気ビーズから蛍光が検出される。DCN配列をもつDNAを用いて、各ステッ

プの演算がきちんと進むか否か調べたところ、正確に演算が実行されることが確認された。

一方、自律的方法については、レトロウイルスのゲノムの複製の仕組みを模して構築した自律型DNAコンピュータを用いた方法を考案した。遺伝子の発現パターンを判定する場合、発現されて遺伝子のmRNAが存在すると、それに対応した内部コードが生成され、AND関数に入力される。2つのmRNAが同時に存在するとAND関数から1の状態を表す内部コードが出力され、分子ビーコンでそれを検出すれば判定結果が得られる。すべての処理は、単一の試験管内で一定温度の下、自律的に進行する。DNAチップスキャナーや質量分析器などの高価な装置を必要としない、簡便な診断法のための基盤技術として優れている。

<国内外での成果の位置づけ>

DNAコンピュータを用いたSNP解析および遺伝子発現解析は、国内外を問わず他に類似研究のない、新規性の高い研究である。

大規模なSNPタイピングや遺伝子発現解析を行うための方法として、既にDNAチップやDNAマイクロアレイを用いた方法が開発され、広く利用されている。これらの方法は、疾患に関連する候補SNPや機能に関連する候補遺伝子を絞り込むための大規模な第一次スクリーニングには適している。しかし、解析の精度が十分でないため、疾患との関連や遺伝子機能に関する詳細な解析を進めるためには適していない。一方、SNPのアリルタイプや遺伝子転写産物を個別に決定、定量する既存の技術は、解析の精度は高いが並列処理には適さないため、多数のSNPや遺伝子を同時に効率よく解析することができない。本研究で開発した方法は、両者の良いところを有する解析技術である。すなわち、実用的に十分に高い並列性をもち、しかも個別に解析する技術と同等の高い精度の解析が可能である。DNAマイクロアレイによる方法とは異なり、定量的PCRのように個々の遺伝子の転写産物量を定量することができるので、異なる研究機関で実施された発現解析データを集積して規格化データベースを構築する用途にも適している。

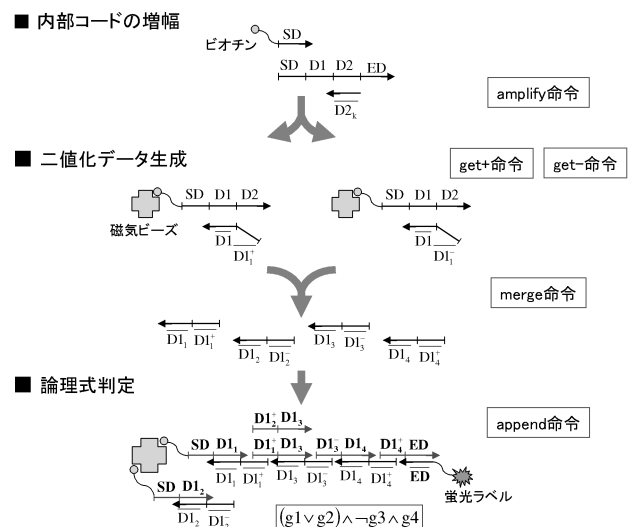


図7 DNAコンピューティングによる遺伝子発現パターンの判定

さらに、本研究の方法は既存のSNPタイピング技術や遺伝子発現解析技術にはない汎用性と演算性を有している。解析の内容や対象によって異なるのは、エンコード反応の部分の変換テーブル分子だけで、その他の部分は共通である。この汎用性により、DNAチップやDNAマイクロアレイによる第一次スクリーニングで候補を選択した後の多様な解析対象に容易に対応できる。また、演算性を有しているため、SNPのアリルタイプのパターンや遺伝子の発現パターンを分子反応だけで直接判定することも可能である。多検体の診断に効果を発する解析法である。

このように、本研究で開発した方法は既存の方法には見られない画期的な新規性を有しているため、以下のよういろいろなメディアでも報道された。Science 295, 951 (2002). DNA-Based Computer Takes Aim at Genes./ 日本工業新聞, 平成14年1月29日号, 朝刊第2面. 世界初の遺伝子解析用./ 日経産業新聞, 平成14年3月7日号, 朝刊第10面./ DNAコンピューター, ナノ部品組み立て工場に, 遺伝子解析に威力./ 日経ビジネス, 平成14年9月9日号, p.155. 世界初の商業用DNAコンピューターを開発./ NHK教育テレビ. サイエンスアイスペシャル, 平成14年1月26日放送./ テレビ朝日, ニュースステーション, 平成14年2月12日放送./ 日経バイオビジネス, 平成15年2月号, pp.96-101. 特別レポート DNAコンピューターはなぜ動く./ 日経PC21, 平成15年2月号, p.17. DNAコンピューター./ 東京大学新聞年鑑2002, pp.67-69. DNAコンピューター./ 東京MXテレビ, ガリレオチャンネル, 平成15年1月26日放送. 『生物学博士・軽部征夫が語るバイオニクス 生物に学ぶ21世紀型テクノロジー』.

達成できなかったこと、予想外の困難、その理由

SNPタイピング法に関しては、タイピングの成功率を100%にすることができなかった。その原因は、ゲノムDNAから調製したSNP部位を含むDNA断片の量がSNP部位によって大きく異なることにあることがわかった。本研究で採用したDCNの形式と増幅法では、アリルタイプによるエンコード効率の違いは補正されて結果に影響を与えない。しかし、SNP部位を含むDNA断片の量の違いは補正されないため、他と比べて非常に量が少なく、検出時のシグナル強度が低くなり、タイピングに失敗する。DCNの形式と増幅方法を修正すると、この問題を回避することができることがわかったため、今後、試みたいと考えている。

遺伝子の発現解析法に関しては、エンコードの効率と遺伝子同定配列の間に何らかの関係が見出されることを期待していたが、結局、明確な関係を見出すことができなかった。そのため、関係を利用して遺伝子同定配列を設計することによりエンコード効率を一様にするという計画は失敗した。しかし、エンコード効率を詳細に調べた結果、規格化エンコード法という、より強力な方法を開発することができた。

また、遺伝子発現解析法に関しては、本研究で開発した方法はかなり良い定量結果を与えると推察されるが、依然として、本当に正しい結果を与えるのか、他の方法と比べてより正しい結果を与えるか、決定することができなかった。トータルRNAから調製したcDNAサンプルでは、その中に含まれる各遺伝子のcDNA量の正しい値は不明である。このようなサンプルを用いて各方法の比較をする限り、どの方法が最も正確なのかを決定することは不可能である。

今後の課題

SNPタイピング法については、SNP部位を含むDNA断片の量の違いを補正するDNAコンピューティングを試みることで、タイピングの成功率を100%にする。また、同時にタイピングするSNP部位の数を100箇所まで増やすことを行う。

遺伝子の発現解析法については、濃度が既知の長鎖DNAを含むサンプルについて、本研究の方法と定量的PCR法による定量結果を比較することにより、DNAコンピューターを用いた方法の優位性を実証する。また、同時解析できる遺伝子の数を1000程度にまで増やすとともに、解析に必要なトータルRNAの量を0.01 μ g程度までに減らすことを試みる。

SNPのアリルタイプのパターンや遺伝子発現パターンの判定と探索については、これまでの合成DNAオリゴや *in vitro* 転写産物を用いた実験の成果を踏まえて、実サンプルを用いた実験を行い、手法の評価を実施する。

研究期間の全成果公表リスト

1) 論文/プロシーディング

- 304082253
Rose, J. A., Deaton, R. J., Hagiya, M., and Suyama, A., The fidelity of the tag-antitag system, LNCS, 2340, 138-149 (2002).
- 304082303
Suyama, A., Programmable DNA computer with application to mathematical and biological problems, Preliminary Proceedings of The Eighth International Meeting on DNA Based Computers, 91 (2002).
- 304082312
Nishida, N., Wakui, M., Hatta-Ohashi, Y., Tokunaga, K., and Suyama, A., Highly specific and quantitative gene expression profiling based on DNA computing, Proceedings of The Eighth International Meeting on DNA Based Computers, 79 (2002).
- 0404212325
Nitta, N. and Suyama, A., Autonomous biomolecular computer modeled after retroviral replication, Lect. Notes Comput. Sc., 2943, 203-207 (2004).
- 0404220238
Ogura, Y., Kawakami, T., Sumiyama, F., Suyama, A., and Tanida, J., Parallel translation of DNA clusters by VCSEL array trapping and temperature control with laser illumination, Lect. Notes Comput. Sc., 2943, 10-18 (2004).
- 0602241815
Rose, J. A., Deaton, R. J., and Suyama, A., Statistical thermodynamic analysis and design of DNA-based Computers, Natural Computing, 3, 443-459 (2004).
- 0602241636
Nishida, N., Tanabe, T., Hashido, K., Hirayasu, K., Takasu, M., Suyama, A., and Tokunaga, K., DigiTag assay for multiplex single nucleotide polymorphism typing with high success rate, Anal. Biochem., 346, 281-288 (2005).
- 0602241728
Gotoh, O., Sakai, Y., Mawatari, Y., Gunji, W., Murakami, Y., and Suyama, A., Normalized molecular encoding method for quantitative gene expression profiling, Proc. 11th Int. Meeting on DNA computing, 395 (2005).

- 2) データベース/ソフトウェア
- 3) 特許など
9. 2004年10月8日出願. DNAコンピューティングの原理に基づくSNPタイピング法. 発明者: 徳永勝士, 西田奈央, 陶山明, 橋渡賢四郎, 田邊哲也. 出願人: 東京大学, オリンパス.
10. 2005年2月4日出願. DNAコンピュータ技術による核酸の定量検出方法. 発明者: 陶山明, 後藤理, 森本伸彦. 出願人: 東京大学, オリンパス.
- 4) その他顕著なもの
11. 304082322
陶山明, 実用化されはじめたDNAコンピューター, 日経サイエンス, 32(5), 114 (2002).
12. 304082330
陶山明, DNAコンピュータの現状と将来, *Computer Today*, 109, 11-16 (2002).
13. 304082340
陶山明, 分子生物学の教科書に隠されていたDNAコンピュータ, *細胞工学*, 21(11), 1346-1349 (2002).
14. 304082346
陶山明, 超並列DNAコンピュータの誕生: エーデルマンの実験, *細胞工学*, 22(1), 75-79 (2003).
15. 304082350
陶山明, 汎用型DNAコンピュータを目指して, *細胞工学*, 22(3), 348-352 (2003).
16. 304082358
陶山明, 生命体は分子コンピュータである, *細胞工学*, 22(5), 570-572 (2003).
17. 0404220221
陶山明, DNAコンピュータで生命の謎を解く, *細胞工学*, 22(8), 882-886 (2003).
18. 0404220215
陶山明, 自律的に動作するin vivo DNAコンピュータ, *細胞工学*, 22(11), 1244-1249 (2003).
19. 304090002
陶山明, DNAコンピュータとその応用, *ゲノム医学*, 3(2), 223-228 (2003).
20. 0404220201
陶山明, DNAコンピュータとナノテクノロジー, *細胞工学*, 23(2), 241-246 (2004).