

## 論文の内容の要旨

ヒト mRNA のスプライシングによる多様性の解析とその遺伝子機能に及ぼす影響に関する研究

若松 愛

### [背景]

ヒトゲノム配列情報より、ヒト遺伝子が 2~2.5 万しかないことがわかってきた。Alternative splicing(AS)という機構は、1 つの遺伝子から異なる複数種の mRNA を産生させるという重要な役割を果たすが、疾患の起因となるような mRNA が AS によって産み出される場合があることもわかってきた。さらに、ガンなどの疾患に関与しているといわれる環境要因によるエピジェネティックな遺伝子発現制御は、AS といった RNA の発現調節にも及んでいると考えられている。これらのことから、ヒト mRNA の AS による多様性の解析を行ない、その遺伝子機能に及ぼす影響を検討することは、遺伝子のもつ未知なる機能の解明や遺伝子発現制御機構の解明につながると考えた。

### 第 1 章 ヒト全長 cDNA の構造解析

mRNA の多様性の解析には、AS によって産み出される protein-coding transcript の同定が必要であると考えた。しかし、多くの遺伝子は、複数のエクソンから構成されているため、それをゲノム配列から予測するのは困難であった。そこで、実際に生体内で産生される mRNA に由来する全長 cDNA を網羅的に取得して構造解析を行なうことにした。

完全長 cDNA ライブラリー作製法の一つであるオリゴキャップ法の反応条件を最適化し、mRNA 約 150ng からでも、5'末端の完全長率が 92%のヒト cDNA ライブラリー作製が可能という微量化と高効率化に成功した。次いで、約 100 種類のヒト組織・ヒト細胞から完全長 cDNA ライブラリーを作製し、そこから取得した 146 万の全長 cDNA(FLJ cDNA)の 5'-EST の配列解析(平均長 約 500bp)を行なった。それより、約 3.5 万個を選別し、全長を配列解析後、Coding sequence(CDS)と機能の予測を行った。その結果とリソース等をもとに、様々な実験手法に応用できる Gateway

cloning system を導入し、33,275 個の Gateway エントリークローンを作製した。13,364 個については、小麦胚芽を用いた無細胞系のタンパク合成技術によってタンパク合成し、SDS-PAGE で解析した。さらに、他の機関の公共データベースの配列と合わせてヒトゲノム配列へのマッピングを行った。これらの一連の解析結果を搭載した Human Gene and Protein Database (HGPD, <http://www.HGPD.jp>)を構築した。

改良オリゴキャップ法による高効率化、微量化の成功は、細胞や組織を細分化しての解析を可能にすることから、より限定した条件下での遺伝子機能の解明につながると考えられる。取得した 146 万の FLJ cDNA は、10 万にも及ぶと言われる protein-coding transcript の多くをカバーしていることから、mRNA の多様性の解析に大きく貢献すると予測される。取得した 5'-EST は、全長率の高い方法で作製され、平均長が 500bp あることから、転写開始点(TSS)を含む領域の情報を得ることが可能であり、転写調節の解析に特に有用であると考えられる。

## **第 2 章 ヒト mRNA の多様性により産生される転写産物の探索・解析**

遺伝子の機能解析には、AS によって産み出される mRNA の中でも特にこれまで同定されていない CDS をもつ protein-coding transcript の探索が重要であると考えた。そこで、FLJ cDNA の 5'-EST を利用した探索を行い、選別した約 2 万の全長 cDNA の配列解析から、既知の全長 cDNA と異なる CDS をもつ protein-coding transcript を 11,769 個同定することに成功した。

mRNA の多様性解析には、その標的となる遺伝子を特定することが重要であると考え、ヒトゲノムではなく、ヒト cDNA による遺伝子数の予測を行なった。FLJ cDNA より全長配列解析を行なった 5.5 万の全長配列、146 万 cDNA の EST 配列、公共データベースより集めた他の機関による配列 (ヒト全長配列、RefSeq 配列、Ensembl 配列、EST 配列) とヒトゲノム配列との比較を行なった。そこから、染色体領域ごとにマニュアルで protein-coding transcript が発現している領域であるかどうかの評価を行ない、遺伝子数を 23,241 個と予測した。また、FLJ Human cDNA Database ver. 3.0 (<http://flj.lifesciencedb.jp>) を構築し、全長 cDNA の配列情報と機能予測の結果を搭載した。

同定した約 1.2 万の全長 cDNA のスプライスパターンの解析から、約 30% は、既知の全長 cDNA の TSS と異なる TSS から転写される mRNA 由来の全長 cDNA であることがわかった。その内 1,962 個は、First exon variation (FEV) と言われる TSS が既知の全長 cDNA の TSS と異なるエクソン上に存在する全長 cDNA だった。遺伝子の機能と mRNA の多様性の関係を調べるためには、protein-coding transcript ごとの発現プロファイルの取得が必要である。そこで、FEV によって産生される全長 cDNA に注目し、その TSS 領域の発現頻度を FLJ cDNA の 5'-EST を用いて解析した。その結果、261 個の全長 cDNA (155 遺伝子に相当) の TSS が、組織特異的な発現を示すことを見出した。さらに、そのうち 13 遺伝子については、TSS 領域の発現頻度を Real-time PCR によって解析した。その結果、例えば FGD4 遺伝子では、既知の全長 cDNA の TSS は胎児脳、精巣で、同定した FLJ55905 cDNA の TSS は免疫系組織(骨髄、脾臓)において発現上昇することがわかった。これらの結果から、組織ごとに TSS を使い分けて多様な mRNA を産生している遺伝子を明らかにした。

5'-EST を用いた mRNA の多様性解析により、TSS や N 末側のアミノ酸配列に起こる多様性に

ついて、かなりの知見を得た。しかし、5'-EST で配列がカバーされていない TSS から 500 bp 以上下流の領域で起こる mRNA の多様性については、今後解析を行っていく必要がある。TSS を選択的に使うことにより複数の異なる CDS をもつ protein-coding transcript を産み出す遺伝子の存在を明らかにしたが、それらの発現調節と転写のメカニズムには関係性があることが予想されることから、その究明についても今後行っていく必要がある。

### **第 3 章 NT2 細胞のレチノイン酸誘導によって発現レベルが変化する遺伝子の選択的スプライシングの解析**

疾患解明には、環境要因によって起こる mRNA の多様性の解析が重要であると考えた。そこで、遺伝要因が同一であるヒト培養細胞株 NT2 がレチノイン酸(RA)によって神経に分化する系をモデルとして、RA 誘導によって発現レベルが変化する遺伝子の AS の解析を行なった。RA 誘導における 4 点のサンプル(0-day, 1-day, 2-day, 7-day)を用いて DNA マイクロアレイによる網羅的な発現頻度解析を行なった。その結果、コントロール(0-day)と比較して、各点で経時順に 40, 106, 340 個のプローブの発現レベルが変動していた。プローブの染色体位置の同定から、それが 358 遺伝子に相当することがわかった。それらの GO 分類による機能予測の結果から、3 つの特徴的な機能カテゴリーが見つかったが、その一つである“Transcription regulator activity”というカテゴリーに分類された 18 の遺伝子は、転写因子であった。また、FLJ Human cDNA Database ver. 3.0 を用いた mRNA の多様性の解析より、274 遺伝子は AS による多様性をもつ遺伝子であることがわかった。さらに、136 遺伝子は、N 末端側がオルタナティブに変化して(Alt. N-term)、59 遺伝子は、C 末端側がオルタナティブに変化して(Alt. C-term)多様性をもつ遺伝子であることがわかった。

転写因子は、様々な遺伝子の発現調節に関与している可能性が高い。そこで、GO 分類で同定した 18 の転写因子について AS により産生される CDS の異なる protein-coding transcript ごとの発現プロファイルとそれの神経分化に対する影響を調べた。Protein-coding transcript ごとの発現レベルは、それぞれの特異的な領域に設計したプライマーを用いて Real-time PCR で解析した。各遺伝子から産生される 2 種の protein-coding transcript の発現レベルの変動が異なる遺伝子(PEG3, HOXA2, RARB)も、それらの発現レベルの変動が同じである遺伝子(HNF1B, PAX6, ETV4, RFX2)も存在していた。さらに、各遺伝子から産生される 2 種の protein-coding transcript が、発現レベルの変動の大きさだけでなく、発現レベルを変化させる時間も異なっている遺伝子(POU5F1, HOXA3, ETV5, ETV1, ZNF483)も見出した。神経分化において、選択的な発現を行なうシステムが存在し、それが遺伝子の機能に影響を及ぼす可能性が高いことが示唆される。

次に、Alt. C-term によって多様性をもつ 26 の遺伝子も同様の解析を行なった。その結果、12 遺伝子は、各遺伝子から産生される protein-coding transcript ごとの発現プロファイルが異なっていた。また、RA 誘導 14 日後、35 日後のサンプルを用いて同様の解析を行なって発現プロファイルを比較した。

RA 応答遺伝子の mRNA の多様性に注目した解析から、産生する protein-coding transcript ごとの RA 誘導による発現プロファイルが異なっている遺伝子を見出した。それらの機能の予測から、その多様性が神経分化に対して影響を及ぼしている可能性も示唆された。遺伝子の機能に影響

を及ぼす mRNA の多様性が環境要因によってもおこることもわかった。このことから神経分化に関係する遺伝子のさらなる機能解明には、**protein-coding transcript** ごとの機能や多様性を調節するメカニズムについてさらなる解析を行い検証していく必要がある。

#### [まとめ]

以上の結果より、遺伝子がASによって産み出す **protein-coding transcript** についての知識の蓄積に大きく貢献できた。また、環境要因に応じて変化する mRNA の多様性の解析から、遺伝子が状況に応じて産生する mRNA の多様性とその機能との関係についての知見を得た。これらの結果を利用してさらなる解析を行い、産生される **protein-coding transcript** ごとの機能や、多様性を制御するメカニズムを解明していく必要がある。それは、遺伝子の未知なる機能の解明だけでなく、新規医薬品の標的候補遺伝子の探索や、副作用の少ない医薬品の開発へとつながる可能性があると考えられる。