

# ヒト完全長 cDNA の大量末端塩基配列データ公開について

平成 17 年 10 月 17 日

バイオテクノロジー開発技術研究組合

## 1 . 背景

国際的な共同研究によるヒトゲノム DNA の全塩基配列の解析により、ヒトの遺伝子数は 2 - 2 . 5 万個と推定されています。しかし、ゲノム DNA から転写される mRNA の種類は、mRNA のスプライシングや転写開始点の多様性によって、遺伝子数よりもっと多く、約 10 万と推定されています。

最近では、個の医療に向けて、個々人のゲノム DNA の塩基配列から直接得られる多様性 (SNP など) からは予測できない mRNA のスプライシングや転写開始点の違いによる多様性と疾患の関係を解明する研究が加速されてきています。即ち、ゲノム DNA の塩基配列は同一だが年齢・環境因子・疾患・ストレス・刺激等により選択的に転写される mRNA の多様性やネットワークと疾患との関係が解析される、本格的なポストゲノムシーケンス研究の時代になってきました。

この度、バイオテクノロジー開発技術研究組合は、このような研究を進めるための基盤であり、ヒト mRNA から約 3 万の完全長 cDNA を取得する過程で得られた約 140 万にも及ぶ大量末端塩基配列データを、日本 DNA データバンク (DDBJ) に登録し、公開します。

## 2 . cDNA プロジェクト

バイオテクノロジー開発技術研究組合は、経済産業省の「完全長 cDNA 構造解析」プロジェクトを新エネルギー・産業技術総合開発機構 (NEDO) から受託し、3 年間 (平成 11 年 3 月から平成 14 年 3 月) で約 3 万個のヒト cDNA の全長塩基配列を解析する目標を達成しました。

この全ての全長配列は、既に国際データベースである日本 DNA データバンク (DDBJ) に登録し、公開しています。

上記プロジェクトにおいて、全長の塩基配列を解析する cDNA を選別するために、約 130 万個の cDNA について末端の塩基配列 (大部分が 5'-末端配列) を解析しました。このデータは、選択的スプライシング、転写開始点上流のプロモータ領域等の解析にとり、貴重な研究資源であると考えられています。

このデータについては、プロジェクト終了後、「ヒト cDNA 末端配列情報利用コンソーシアム」を組織し、日本に研究施設のある企業、大学、研究機関における研究に利用して参りました。また、NEDO の研究開発プロジェクトでも、「タンパク質機能解析・活用プロジェクト」のサブテーマ「スプライシング・バリエーション cDNA クローンの取得」において利用してきていますが、情報整理

が完了しましたので、これを機に、全長配列と同様にDDBJに登録し公開します。

なお、「完全長 cDNA 構造解析」プロジェクトは、菅野純夫東京大学教授をリーダーとし、(旧)株式会社ヘリックス研究所、東京大学医科学研究所、製品評価技術基盤機構ゲノム解析部門との共同研究により、実施したものです。

### 3 . 公開日と新規公開配列の DDBJ 登録番号

公開日：平成 17 年 10 月 17 日から 21 日にかけて順次公開

・国際データベース登録機関である DDBJ への登録配列は、日米欧の DDBJ/GenBank/EMBL から公開されます。

新規公開配列の DDBJ 登録番号：

・ヒト完全長 cDNA の 5'-末端配列 129 万個 DA000001 - DA999999,  
DB000001 - DB294747

・ヒト完全長 cDNA の 3'-末端配列 9 万個 DB294748-DB384947

### 4 . 公開するヒト cDNA 末端配列の特徴

全長率の高いオリゴキャップ法で作成したヒト cDNA ライブラリー由来であり、cDNA の大部分のインサートサイズが 2Kb 以上で、5'-末端の全長率が 90%であるため、以下の特徴がある (米国の GenBank で公開されている dbEST では約 18%である)。

ゲノムマッピングにより、転写開始点情報がそのバリエーションも含めて得られる。

タンパク質コード領域を含んだ cDNA が多く得られる。

ヒト組織・細胞由来の cDNA ライブラリーより、1 ライブラリーあたり平均 1 万以上の cDNA の 5'-末端配列を取得しているため、以下の特徴がある。

ヒトゲノムの遺伝子の約 80%を網羅している。

各遺伝子クラスターでの 5'-末端配列での発現プロファイル情報が得られる。

### 5 . 公開処理への謝辞

大量配列の公開処理のために、次の 2 機関に多大な負荷をおかけしましたが、全面的なご支援を頂きましたことに感謝します。

国際データベース登録・公開：日本 DNA データバンク (DDBJ)

DDBJ への登録ファイルの作成：株式会社リバース・プロテオミクス研究所

## 6 . 問い合わせ先

バイオテクノロジー開発技術研究組合

三好 泰克 TEL 03-3595-0371 E-mail y-miyoshi@ra-bio.or.jp

株式会社リバース・プロテオミクス研究所

磯貝 隆夫 TEL 0438-52-3975 E-mail isogai-t@reprori.jp

東京大学大学院新領域創成科学研究科

菅野 純夫教授 TEL 03-5449-5286 E-mail sugano@ims.u-tokyo.ac.jp

日本 DNA データバンク (DDBJ)

五條堀 孝教授 (国立遺伝学研究所 生命情報・DDBJ 研究センター長)

TEL 055-981-6847 E-mail tgojobor@genes.nig.ac.jp