

論文審査の結果の要旨

氏名 村上 勝彦

ヒトゲノム計画によるヒトゲノム塩基配列の決定や、その前後から盛んに研究されている網羅的なヒト遺伝子の特徴付け研究（オームクス研究）の進展を契機として、さまざまな角度からヒトゲノムにコードされた遺伝情報の理解への努力が試みられている。本論文もその一環として、論文提出者がこれまで約10年にわたり行ってきた研究の集大成として位置づけられる。

本論文の構成は大きく2部に分けられる。前半では、ゲノム配列からの遺伝子構造予測について述べられている。具体的には二つの内容、複数プログラムの統合による遺伝子予測（第2章）と、スプライス部位のクラスタリング（第3章）が扱われている。バイオインフォマティクスの分野では、与えられたDNA塩基配列中に存在するタンパク質コード領域をイントロン・エキソン構造まで含めて予測するという、遺伝子発見問題として知られる有名な問題がある。論文提出者はこの問題に対する一つの先駆的なアプローチとして、複数の既存の予測プログラムを組み合わせることで、全体の予測性能を5%程度向上させることに成功した（第2章）。一方、スプライス部位のクラスタリング研究では、遺伝子予測において5'スプライス部位をゲノム配列から識別するステップがその精度向上の要であるという認識のもとに、スプライス部位内の塩基の位置依存性を考慮した新規クラスタリング方法および識別方法を提案した。この方法によって、スプライス部位は5つのクラスのいずれかに属するとして認識されるが、特定のクラスに分類された場合には、それまでの方法に比べて、擬陽性を減らして認識できることを示した（第3章）。

本論文の後半では、ヒト遺伝子のプロモーターの構造解析について述べられている。こちらにも二つの内容にわかれ、予測された転写因子結合部位（モチーフ）のクラスターの存在とプロモーターの相関解析（第4章）、並びにモチーフペアと転写開始点（TSS）の位置関係の解析（第5章）について論じている。遺伝子の転写制御情報を含むと思われるプロモーター領域がどのような原理で構成されているのかを理解することは、バイオインフォマティクスにおける最重要の課題の一つと言えるが、今のところその実態はほとんどわかっていないと言っても過言ではない。一つの有力な仮説として、脊椎動物一般のプロモーター領域では、複数のモチーフ配列（転写因子結合部位）、特に同一のモチーフ配列が複数集まってモジュールを構成していると言われている。そこで、論文提出者らは、プロモーター領域には同一モチーフクラスターが存在するという仮説が、モチーフによらず一般的に成り立つかどうかを調べるため、同一モチーフのクラスターとプロモーター領域との相関をモチーフ毎に調べた。その結果、約47%についてはモチーフクラスターとプロモーターに相関があり、仮説の正当性はモチーフに依存することを明らかにした（第4章）。さらに、論文提出者らは、プロモーター構造を特徴づけるために、転写因子結合部位（モチーフ）と転写開始点（TSS）との相対距離に着目した。従

来、単独のモチーフで、TSS からの位置特異的に頻出するものが知られていたもので、モチーフのペアが同じプロモーター領域に重ならない位置関係で一緒に出現する場合に、TSS からの相対位置に有意に偏りが存在するかどうかを調べた。位置のバイアスをヒートマップにより可視化すると様々なパターンが見られたが、大きく TSS 付近でペアのピークがあるもの（クラス 1 : 250 ペア）と、ペア間距離が 200bp 以下で、かつ TSS からの距離に関係ない位置でピークがあるもの（クラス 2 : 4,119 ペア）およびその他（クラス 3 : 4,479 ペア）に分類した。クラス 1 のペアはクラス 2 のペアと比べて、それぞれの成分であるモチーフやそこに結合する転写因子ファミリー、そのペアをもつ遺伝子の発現の組織特異性などなど、いろいろな面でその性質が異なっていることがわかり、ヒト遺伝子のプロモーター構造の特徴として、共起する位置に有意なバイアスをもつモチーフペアが数多く存在しているという現象を示し、それらのペアには特徴の異なるクラスが少なくとも二つ存在するという見方を提案した（第 5 章）。

なお、本論文第 2 章および第 3 章は高木利久との共同研究、第 4 章は小島俊男、榊佳之との共同研究、第 5 章は今西規、五條堀孝、中井謙太との共同研究であるが、論文提出者が主体となって分析及び検証を行ったもので、論文提出者の寄与が十分であると判断する。

したがって、博士（生命科学）の学位を授与できると認める。