論 文 の 内 容 の 要 旨

論文題目　無脊椎脊索動物カタユウレイボヤにおける転写開始点の網羅的同定とプロモーター領域の網羅的解析

氏　名　　横森　類

1. 背景

遺伝子がどのように発現調節されているかを理解するには、遺伝子の転写開始点およびプロモーター領域の同定が必要不可欠である。近年、オリゴキャッピング法やキャップトラッパー法と次世代シークエンサーを組み合わせることにより、転写開始点（TSS）をゲノムワイドに同定することが可能となっており、様々な種において転写開始点およびプロモーター領域の解析が行われている。

カタユウレイボヤは、我々ヒトに最も近縁の無脊椎動物であり、その幼生の体制は脊椎動物を含む脊索動物の基本型を表している。ホヤゲノムは約 160Mb で、遺伝子数はおよそ 15,000 個である。このゲノムのコンパクトさは、ショウジョウバエに匹敵する。また、ホヤの胚発生は非常に速く（約 18 時間で幼生になる）、電気穿孔法によってレポーターコンストラクトを大量の卵に一度に導入することが可能であることから、遺伝子上流の転写制御領域の解析が多く行われている。しかしながら、カタユウレイボヤにおいてゲノムワイドな正確な転写開始点の同定およびプロモーター領域解析はほとんど行われていない。

ホヤの遺伝子の約半分は spliced leader（SL）*trans*-splicing が起こると言われている。SL *trans*-splicing を受けた mRNA の 5′ 末端は、16 nt の SL 配列に取って代わる。したがって、SL *trans*-spliced mRNA の 5′ 末端は元の配列を失い、転写開始点の同定が困難となる。重要なことに、*trans*-splicing を受ける遺伝子の中には、*trans*-splicing されやすい遺伝子とされにくい遺伝子があると報告されている。したがって、*trans*-splicing されにくい遺伝子もしくは *trans*-splicing されやすいが高発現している遺伝子は、転写開始点が同定できる可能性がある。

本研究では、脊索動物門尾索動物に属するカタユウレイボヤの５つ（卵巣、心臓、体壁筋、神経複合体、幼生）のサンプルに対してオリゴキャッピング法と次世代シークエンサーを組み合わせた TSS-seq 法を適用し、カタユウレイボヤにおける転写開始点の網羅的同定およびプロモーター領域の解析を目的とする。また、ヒトとの比較解析を行い、類似点･相違点を見いだす。さらに、組織特異的な選択的プロモーターの探索も行う。

2. 方法

2.1 TSS と *trans*-splice acceptor site の同定

5 つのサンプルに対して TSS-seq 法を適用し、mRNA の 5′ 末端配列に由来する 34 bp のリード配列を得た。リードのフィルタリング後、SL 配列の有無に従って、SL 配列をもつ SL(+)リードともたない SL(-)リードに分類した。SL(+)と SL(-)リードをそれぞれリファレンスゲノム上にマッピングすることで、*trans*-splice acceptor site （TAS）と TSS を網羅的に同定した。

2.2 TSS クラスターと TAS クラスターの同定

遺伝子の転写開始点はある一点から始まるのではなく、分布していることが知られている。そこで、同定した TSS をクラスタリングすることで、高密度の TSS 領域（TSS cluster, 以下 TSC）を同定した。同様に、TAS もわずかではあるが分布することが知られており、クラスタリングすることで TAS cluster（以下、TAC）を同定した。この操作によって得られた TSC および TAC は、それぞれプロモーター領域と *trans*-splice acceptor site を示すと考えられる。

2.3 SL *trans*-spliced 遺伝子の TSS の予測

SL *trans*-spliced 遺伝子から生じた mRNA は、5′ 末端が SL 配列に置き換わり、元々の 5′ 末端を持っていないため TSS の同定ができない。しかし、SL *trans*-spliced 遺伝子の中には、*trans*-splicing されにくい遺伝子がある。すなわち、それらの遺伝子からは、5′ 末端が SL 配列に置き換わった（*trans*-splicing を受けた）mRNA だけでなく、元々の 5’ 末端を維持した（*trans*-splicing を受けなかった）mRNA の 2 種類が生じると考えられる。したがって、*trans*-splicing されにくい遺伝子に対しては、*trans*-spliced mRNAs の 5′ 末端由来の TAC だけでなく、non-*trans*-spliced mRNAs の 5′ 末端由来の TSC が得られるはずである。また、この TAC と TSC のペアは同じ遺伝子に由来するので、発現パターンは類似していると期待される。そこで、発現パターンが同じである TAC とその上流に存在する TSC のペアを探索することで、SL *trans*-spliced 遺伝子の TSS を予測した。

3. 結果と考察

3.1 TSC と TAC の同定

ノイズ TSC の除去後、計 3,206 個の TSC と 5,373 個の TAC を同定した（表 1）。60%の TSC は既知の TSS 上に存在した。残りの 40%は、既知の TSS 以外の場所に存在しており、これらの TSC は、未知の転写単位の存在を示していると考えられる。カタユウレイボヤにおいては、マイクロ RNA を含む多くのノンコーディング RNA 候補が同定されており、これらのアノテーションされていない転写産物のいくつかはノンコーディング RNA をコードするかもしれない。

また、TSC は 5′ UTR やイントロンにも多く発見された。これらの TSC は選択的プロモーターを表していると考えられる。面白いことに、幾つかの TSC は TAS と重なっていた。TAS 付近の領域が SL *trans*-spliced 遺伝子の選択的プロモーターとして機能するのかもしれない。

3.2 プロモーター解析

3.2.1 リボソームタンパク質（RP）遺伝子のプロモーターの性質

本研究によって、初めて正確な RP 遺伝子の主要 TSS 候補を明らかにした。その結果、RP 遺伝子のプロモーターはヒトやショウジョウバエと同じく、pyrimidine rich な Inr motif を持ち、シャープな TSS 分布を示すことが分かった。しかし、ヒトと異なり、TATA box は持っていないことが示唆された。RP 遺伝子プロモーターの構造の一般的特徴はヒト、ニワトリ、両生類、魚においてよく保存されているので、カタユウレイボヤにおいて TATA box がなかったことは、RP 遺伝子プロモーターが TATA box を獲得したのは脊椎動物の進化の初期段階であると考えられる。そのため、ヤツメウナギなど原始的な脊椎動物において TATA box があるのかどうかを調べるのは興味深いと思われる。

表 1: 同定した TSC と TAC の数

Location	TSCs	TACs
TSS	1923 (60.0%)	1 (0.0%)
TAS	32 (1.0%)	4748 (88.4%)
5UTR	143 (4.5%)	70 (1.3%)
CDS	0 (0.0%)	40 (0.7%)
3UTR	0 (0.0%)	8 (0.1%)
Intron	239 (7.5%)	104 (1.9%)
Intergenic	869 (27.1%)	402 (7.5%)
Total	3206 (100%)	5373 (100%)

3.2.2 Non-RP 遺伝子のプロモーター

RP 遺伝子以外のプロモーターの TSS を調べたところ、ヒトでは多くの場合 CA の A から転写が始まるのに対し、ホヤにおいては CA だけでなく TA の A から始まることが非常に多いことが分かった。このことは、ホヤゲノムが比較的 AT-rich であることに起因するものだと考えられる。実際、カタユウレイボヤにおいては、TSS 付近において T 含量がその他の塩基の含量より高かった。この高 T 含量がカタユウレイボヤにおける TA の高い使用率の原因だと思われる。また、ヒトでは TATA box を持たないプロモーターは CpG islands と関連があり、発現特異性の低い house-keeping 遺伝子と結びつきが強いことが知られているが、ホヤでは CpG islands がないにもかかわらず発現特異性の低い遺伝子との関連が強かった。果たしてホヤプロモーターにおいて CpG islands に代わるものがあるのかというのは非常に興味深いことだが、現在のところ不明である。

3.3　SL *trans*-spliced 遺伝子の TSS の予測

発現パターンが同じ TAC と TSC のペアを探索した結果、約 230 個の既知 SL *trans*-spliced 遺伝子の転写開始点候補が得られた。これらの情報は、今後の SL *trans*-spliced 遺伝子の転写制御領域解析に役立つと考えられる。また、予測された *trans*-spliced 遺伝子のプロモーターは、TSS 下流 10 から 20 bp 付近で non-*trans*-spliced 遺伝子プロモーターよりも高い G+T 含量を示すことが分かった。しかしながら、この高 G+T 含量がどのような役割を果たすのかはまだ明らかではない。

3.4　組織特異的な選択的プロモーターの探索

表 1 で示した TSC を用いて、組織特異的に使われている選択的プロモーターの探索を行った。その結果、いくつかの興味深いプロモーターをもつ遺伝子が見つかった。図 1 に Myosin-light chain 遺伝子の例を示す。この遺伝子は、体壁筋と心臓で異なるプロモーターが使われている。この転写制御がどのように行われているのかは興味深く、今後の解析が待たれる。

図 1 ：Myosin-light chain 遺伝子の組織特異的プロモーター

4.　結論

本研究では、脊策動物モデルであるカタユウレイボヤにおいて、正確な転写開始点を網羅的に同定した。プロモーター構造に関して、ホヤとヒト間での類似点および相違点を示した。また、本研究で初めて、多数の SL *trans*-spliced 遺伝子の転写開始点候補を示した。これらの結果およびデータは、今後の脊索動物の転写制御領域の解析に役立つものである。