〔シンポジウム 1 : ゲノミクス・プロテオミクスを中心に -omics への電気泳動法の応用〕

ゲノム科学新時代に向けて

林 崎 良 英

SUMMARY

The FANTOM Consortium analyzed over 2 million sequences of RNAs produced from the mouse genome, obtaining more than 100,000 full length copies of these RNAs, and used several new technologies to generate more than 20 million tags representing their starts and ends in a technical tour-de-force that provides a amazing new view of our genome. These studies clearly indicate that while proteins comprise the essential components of our cells, the development of multicellular organisms like mammals is controlled by vast amounts of regulatory noncoding RNAs that until recently have not been suspected to exist or be relevant to biology. Moreover, since most proteins are similar among mammals it also suggests that many of the differences between species may be embedded in differences in the RNA regulatory control systems, which are evolving much faster than the protein components. If correct, these findings will radically alter our understanding of genetics and how information is stored in our genome, and how this information is transacted to control the incredibly complex process of mammalian development, with major implications for the future of biological research, medicine and biotechnology.

Key words: transcriptome, RNA continent, FANTOM.

近年の技術革新は目覚しく、中でも、最近米国で進められている 1,000 ドルゲノムプロジェクトに代表されるような、ごく短時間にゲノム配列を解読する技術の発展には目を見張るものがある。ヒトやマウスをはじめとして様々な生物のゲノム解読が進んでいる現在、大量のデータを基盤とした、いわゆるデータドリブン型の研究が中心となりつつある。いまや、ライフサイエンスは新たな局面をむかえている。

我々は、1995 年、マウスの遺伝子を網羅的に解析する大規模な計画をスタートした。当時、米国ではヒトゲノム計画が進められており、日本はこの面では遅れをとっていた。そこで、我々は、一歩進んで、成熟 mRNA を完全長 cDNA として収集することにより、ゲノム配列の中でどの部分が最終的に機能しているかを明らかにする先駆的な戦略をとった。これが、マウスエンサイクロペディア計画である。この解析においては、384 キャピラリー・シーケンサーを独

自に開発し活用した。

このようにして得られた完全長 cDNA の配列情報は、それだけでは意味を成さない。そこで、2000 年には、筆者が中心となって国際コンソーシアム FANTOM (Functional Annotation of Mouse cDNA) を組織し、マウスエンサイクロペディアプロジェクトで収集された cDNA のアノテーションを行った。この結果、noncoding RNA (ncRNA) が少なからず存在することが明らかになった^{1,2)}。また、このころから、ncRNA が生体内で重要な役割を持つことを示唆するデータが発表され始めていた。これを受けて、我々は、完全長 cDNA がクローニングされていない RNA も含めたより大規模な解析を行った。本稿では、この解析により得られた知見を紹介する^{3,4)}。

完全長 cDNA が得られていない RNA を含めた包括的解析を実施するため、我々は、CAGE (Cap Analysis Gene Expression) および GSC (Gene Signature Cloning) とよば

The new phase of transcriptome analysis.

Yoshihide Hayashizaki; 独立行政法人理化学研究所ゲノム科学総合研究センター遺伝子構造・機能研究グループ

Correspondence address: Yoshihide Hayashizaki; RIKEN Genomic Sciences Center, Genome Exploration Research Group, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.

略語一覧: FANTOM, Functional Annotation of Mouse cDNA; CAGE, Cap Analysis of Gene Expression; GSC, Gene Signature Cloning.

第 56 回日本電気泳動学会総会・シンポジウム

(受付 2005 年 11 月 21 日, 受理 2005 年 12 月 5 日, 刊行 2006 年 6 月 15 日)

Fig. 1. CAGE and GSC.

Fig. 2. Map of CAGE and GSC tags.

れるゲノム上の転写開始点および転写終止点を同定する方法を開発した (Fig. 1). CAGE では RNA の 5' 端を 20 塩基程度のタグとして取り出す。また、GSC は、転写された RNA の両端をそれぞれタグとして取り出し、この 2 つをライゲーションしたタグを作成する。いずれも、複数のタグをコンカテマー化して一度に効率的にシーケンシングする。このようにして得られた配列をゲノム上にマップすることにより、ゲノム上における RNA の転写開始点、および転写領域を示した (Fig. 2)。我々は、この CAGE タグをマウスで 1,156 万個、ヒトで約 600 万個、また GSC については約 250 万個収集した。この結果、3 つの事実が明らかになった。

第一に、ゲノムの全体像が一変した。従来、ゲノムの大半は生体における機能を持たないいわゆるジャンク DNA であるといわれてきた。2004 年のヒトゲノムコンソーシアムの報告でも、RNA に転写される領域は、ゲノム全体の約 2% に過ぎないと見積もられていた⁵⁾。この推測は、完全長 cDNA、EST などの既存データおよび計算によって見積もられたものであった。しかし、今回の大規模解析の結果、ゲノム領域の実に 70% が RNA に転写されていることが明らかになった。例えて言うならば、従来は、広大な砂漠の中にポツポツとオアシス (遺伝子) があるようなイメージがあった。しかし、いまやこのイメージは一転し、ゲノムは遺伝子がひしめく大都会のイメージとなったと感じている。

第二の点は、大量の ncRNA の存在が明らかになったことである。これまで ncRNA は数百個といわれていたのに対し、今回の研究では 23,218 個の ncRNA の存在が示された。驚くべきことに、これは、全 RNA のおよそ 53% が ncRNA

Fig. 3. Discovery of the RNA continent.

であることを意味する (Fig. 3)。クリックがセントラルドグマを提唱してから約 50 年、我々はタンパク質にばかり注目してきたが、今回の解析で、RNA がこれに匹敵する広大な研究領域として開けたといえる。この意味で、我々はこの大量の RNA の発見を「RNA 大陸」の発見といっている。

最後のポイントは、全 RNA の 72% 以上で、二重鎖 DNA の同一領域において両鎖が転写されていることが明らかになったことである。これは、多量のセンス/アンチセンス対の存在を示唆する。マクロファージから単球への分化誘導を経時的に観察すると、センス及びアンチセンス RNA 量の変動に相互関連性が見られた。加えて、ノックアウト実験を通して、センス/アンチセンスは何らかの形で転写・発現の制御に関わっていることが示された。siRNA を用いてある遺伝子のアンチセンス RNA をノックアウトすると、該当するセンス鎖 RNA の発現量に顕著な変化が見られた。そこで、センス/アンチセンスをゲノム上にマップし、データベースとしてまとめたところ、多くの重要疾患の関連遺伝子においてアンチセンス RNA が転写されていることがわかった。これは、センス/アンチセンスが何らかの機構により、疾患関連遺伝子の発現および転写の制御に関わっていることを示唆するといえるであろう。この知見は、将来、様々な疾患に対する治療薬および治療法の開発につながると期待される。

このように、大規模解析により、多量の ncRNA の存在のみならず、ncRNA が生体内において機能を有することが明確に示された。しかし、我々は、まだ RNA 大陸の全貌を捉えるには至っていないと考えている。というのは、今回の解析では、Ploy A をもつ RNA を解析したが、これをもたない RNA も存在することがわかってきたからである。また、RNA による制御の事実は見えてきたが、その機構は RNAi だけでは説明しきれない。RNA による制御機構の解明も今後の重要な研究課題である。

文 献

- 1) Kawai J, Shinagawa A, Shibata K, et al. Functional annotation of a full-length mouse cDNA collection. Nature 2001 Feb 8;409(6821):685-690.
- 2) Okazaki Y, Furuno M, Kasukawa T, et al. Analysis of the mouse transcriptome based on functional annotation of

- 60,770 full-length cDNAs. Nature 2002 Dec 5;420(6915): 563–573.
- 3) Carninci P, Kasukawa T, Katayama S, et al. The transcriptional landscape of the mammalian genome. Science 2005 Sep 2;309(5740):1559–1563.
- 4) Katayama S, Tomaru Y, Kasukawa T, et al. Antisense transcription in the mammalian transcriptome. Science 2005 Sep 2;309(5740):1564–1566.
- 5) International Humane Genome Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931–945.