

- ① The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74 (2012).
- ② Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T. *et al.* The accessible chromatin landscape of the human genome. *Nature*, **489**, 75–82 (2012).
- ③ Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, **489**, 83–90 (2012).
- ④ Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature*, **489**, 91–100 (2012).
- ⑤ Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T. *et al.* Landscape of transcription in human cells. *Nature*, **489**, 101–108 (2012).

Nature 誌の2012年9月6日号に、ENCODE 計画の論文が5編掲載された^{1)~5)}。本稿ではこれらの論文のうち、特に第1の論文についての問題点を指摘し、真核生物のゲノムにおける転写と、その進化的意義について論じる。

英語で“genetic code”を「遺伝暗号」と訳するが、暗号を解読することを“decode”，暗号を作成することを“encode”という。自然科学の本来の目的は自然現象の解明であるから、「暗号を解読する」という表現のほうが科学研究の計画名称にはふさわしいのだが、残念ながらアイスランドの全国民のゲノムを解読しようとしたdeCODE genetics (www.decode.com) という会社がすでに存在していた。おそらくそのために、米国を中心に進められたヒトゲノム配列の機能解明をするための巨大プロジェクトにENCODEという名前がつけられたのだと思われる。なお、ENCODEは、一応は“ENCyclopedia of DNA Elements”の略称ということになっている〔ENCODE計画ホームページ(www.genome.gov/10005107)を参照〕。

5論文の筆頭は、ENCODE計画連合が著者となっている旗艦論文である¹⁾。

ゲノム計画が、「ある生物のゲノム塩基配列を決定する」という目標をもっていたのに対して、ENCODE計画では、「ゲノム中で転写される領域、転写因子が結合する領域、およびヒストンが修飾を受けてクロマチン構造が変化する領域の特定」をめざした。この論文で用いられた生物はヒトだが、他の論文ではマウスも使われている。第2論文²⁾はおよそ300万個のDNase過敏サイトを、第3論文³⁾はDNaseI フットプリント法を用いて転写因子が結合する840万カ所の領域を、第4論文⁴⁾は転写に関連した119種類のタンパク質相互の結合を調べ、第5論文⁵⁾は転写産物(RNA)の細胞内局在を調べている(以下、それぞれの論文を「論文①」~「論文⑤」と表記する)。これらの膨大な実験結果は、培養細胞を使っている場合にはやや問題点があるものの、ヒトゲノムの機能解明にむけた貴重な礎石となるだろう。

ENCODE計画は2007年に、ヒトゲノムの1%だけを調べた結果をパイロット計画として発表している⁶⁾。さらにその2年前には、日本の理化学研究所を中心とする研究チームFANTOMが、マウスゲノムの転写産物を大規模に調べた論文⁷⁾を発表している。これらの結果は、宇宙論研究の「暗黒物質」に

なぞらえられたり⁸⁾、「RNA大陸」と提唱されたりしている⁹⁾。ただ、当時は“Chip-chip”と呼ばれるDNA雑種形成法を用いた技術を使っており、これについては以前からその定量性に疑問が出ていた。ゲノムのあちこちが転写されているといっても、以前からごく少量のRNAが出ることは知られており、それを過大に評価してしまったのではないか、という可能性がある。実際に、2010年にはカナダの研究グループが、ほとんどの「暗黒物質」転写産物は既知遺伝子と関連しているという実験結果を発表している¹⁰⁾。彼らは塩基配列決定を用いるChip-seq法を用いた高い定量的な結果がChip-chip法による結果と大きく異なっていることを示したのである。その後はENCODE計画でもChip-seq法を導入したようだ。

論文①の抄録中には、ヒトゲノムの80%に「生化学的機能」を見出したとした部分がある。「生化学的機能」とわざわざ断っているのは、通常われわれが考える生物やタンパク質、あるいはDNAやRNAの「機能」とは異なっているからのようだ。進化学では、ゲノム中に機能をもつ領域であれば、有害突然変異が除去されるので、進化速度が純粋中立進化(機能がなく、自然淘汰を受けない)の

場合よりも低くなるということが知られている^{11)~14)}。この論理を用いて、真核生物ゲノムの非コード領域の中から、進化的に高度に保存された塩基配列が発見されている。私の研究室でも数年前からそれを調べており^{15)~17)}、昨年(2012年)には高橋真保子¹⁶⁾が霊長類と齧歯類でそれぞれ独自に進化的な保存を生じた非コード領域を多数発見し、今年(2013年)になって松波雅俊¹⁷⁾は脊椎動物の祖先で2回生じたゲノム重複の後に進化的な保存を生じた傍系相同な非コード領域の解析結果を報告している。

図1はヒトゲノムと新世界猿の一種、マーモセットのゲノムを比較した結果である。縦軸は両ゲノム間の塩基置換数を、横軸は発見された霊長類特異的に高度に保存された非コード塩基配列(HCNS)の中心を起点として、上流と下流にプロットしたものである。ゲノムの大部分はほぼ10%近い違いを示しているが、HCNSの近傍だけは、あたかも平原に深く掘られた井戸のように、進化的に高度に保存されていることがわかる。

図2では、こうして発見されたHCNSの多くが転写因子遺伝子の近くに位置していることを示している。UCEとは、ultra-conserved elementの略であり、ヒト、マウス、ラットのゲノム配列を比較して480個ほど発見されたものである¹⁸⁾。図2でわかるように、同じ転写因子の近くにあっても、少しずつゲノム上の位置が異なっており、系統特異的に生じてきたことがわかる。

図3には、異なる染色体上で見出された保存配列が示してある。2回のゲノム重複で生じてきた4本の染色体すべてで保存されている配列は発見できなかったが、3本の染色体の、しかも

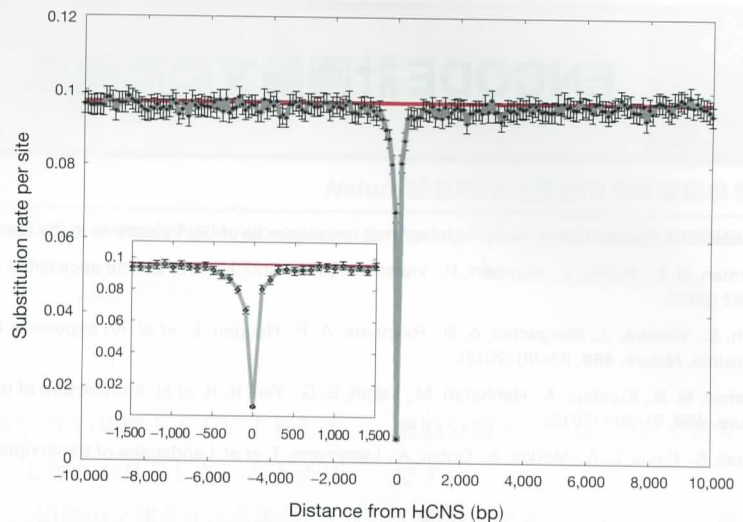


図1 HCNSを中心として、その前後1万塩基についてヒトとマーモセットのゲノム配列の違いをプロットしたもの

HCNS = highly conserved noncoding sequence
くわしくは本文を参照

(文献16より)

傍系相同なタンパク質コード遺伝子のイントロン、あるいはその近傍にある非コード領域保存配列(CNS)を3組見出した(図3A~図3C)。2本の染色体で保存されているCNSは全部で150組あり、それらの中で同一傍系相同遺伝子の近傍にあるものを図3D~図3Hに示した。CNSが6個ともっとも多かったのは、FOXP1とFOXP2の組である(図3D)。位置関係も大部分が保存されているのが興味深い。

なお、文献17では、ENCODE計画の一環ではあるが、論文①~⑤より一足先に発表されたマウス遺伝子の転写解析結果¹⁹⁾を用いて、今回見出したCNSの多くが脳で発現する遺伝子の近傍にあることを示した。

論文①にもどろう。彼らの論理を使うと、これまで「がらくたDNA」だと考えられてきたゲノムの大部分はなんらかの「生化学的機能」があることになる。実際に、*Science*誌のNews &

Viewsには、「ENCODE計画はがらくたDNAへの弔辞を書いた」と題した紹介記事²⁰⁾が掲載された。しかしこの見方には多くの反論があるようだ。がらくたDNAとENCODE計画について論評した論文²¹⁾が昨年発表されたが、最後は以下の文章で終わっている。

(ENCODEが発表した)データは、ヒトゲノムの80~95%を占める、進化的に保存されていない領域は大部分が機能を持たず、くずれつつあるトランスポゾンだという見解を変更するものではない。彼らは「がらくた」なのだ。

また、今年の2月には、米国ヒューストン大学のDan Graurらが、「テレビ機器の不死性について：進化を否定したENCODEの福音にもとづくヒトゲノムの機能」という一風変わったタイトルのコメンタリー²²⁾を発表し、ENCODE計画の論文①をきびしく批判した。

ENCODE計画の論文②~⑤については、進化的保存性とゲノム領域の機能

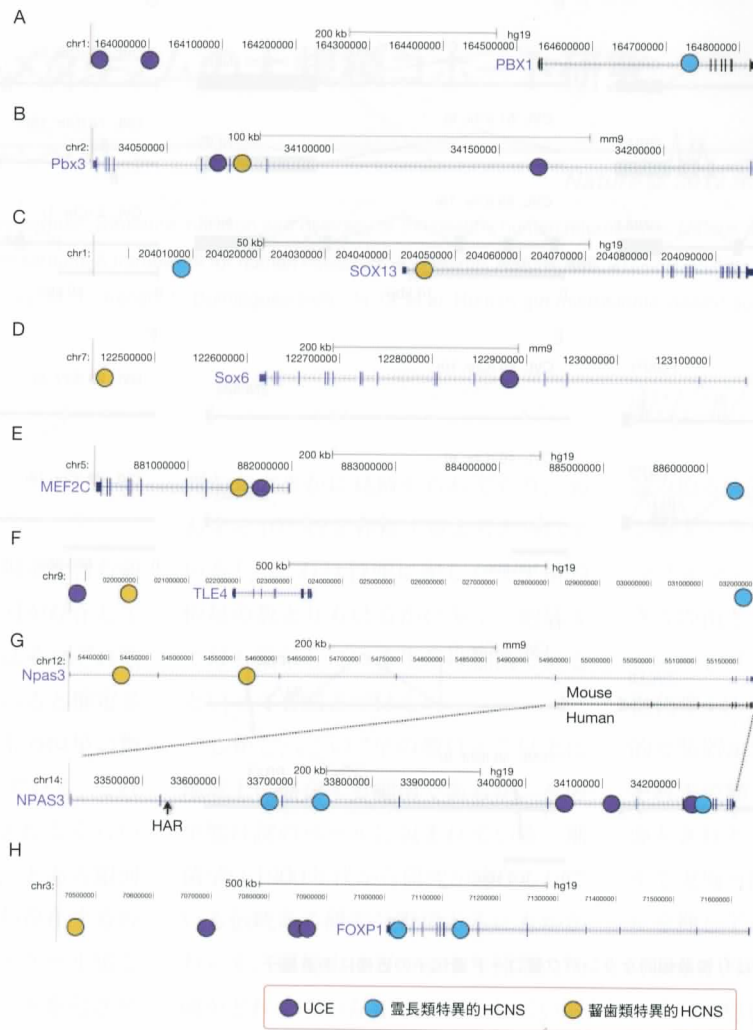


図2 霊長類および齧歯類特異的HCNSとUCEがともに近傍に存在するタンパク質コード遺伝子

UCE = ultra-conserved element
くわしくは本文を参照

(文献16より)

について、論文①のような指摘はないようである。たとえば、論文③の抄録には、「高精度のDNaseI切断パターンは、塩基配列レベルの進化的保存性にびたりと対応している」という部分がある。論文⑤の抄録も、「ヒトゲノムの3/4は転写されることができる」というおだやかな表現である。結論として、ENCODE計画の5論文は、どれも貴重で膨大なデータを提供してくれている。これらは今後の哺乳類ゲノムの研究にとって重要な基盤になるだろう。

しかし、哺乳類ゲノムの大部分ががらくたDNAから構成されているという見方がくずれることはないようだ。

ところで、人間を含む多くの生物のゲノムの大部分を「がらくたDNA」だと1972年に名付けたのは、大野乾^{オノノ}である²³⁾。21世紀のゲノムの時代に、ますます輝いているこの言葉を創り上げた故大野博士を顕彰することもふくめて、今年の6月22日(土)と23日(日)に、Dan Graur教授と、日本で唯一ENCODE計

画に関係している理化学研究所の林崎良英博士、および国内の多数の分子進化学研究者を招いて、静岡県三島市の国立遺伝学研究所で「がらくたDNAの進化」と題した研究集会を開催する予定である。どなたでも聞きにきていただけるし、入場は無料である。くわしくは、国立遺伝学研究所のホームページ(www.nig.ac.jp)にいずれ掲載されるので、そちらをご覧ください。

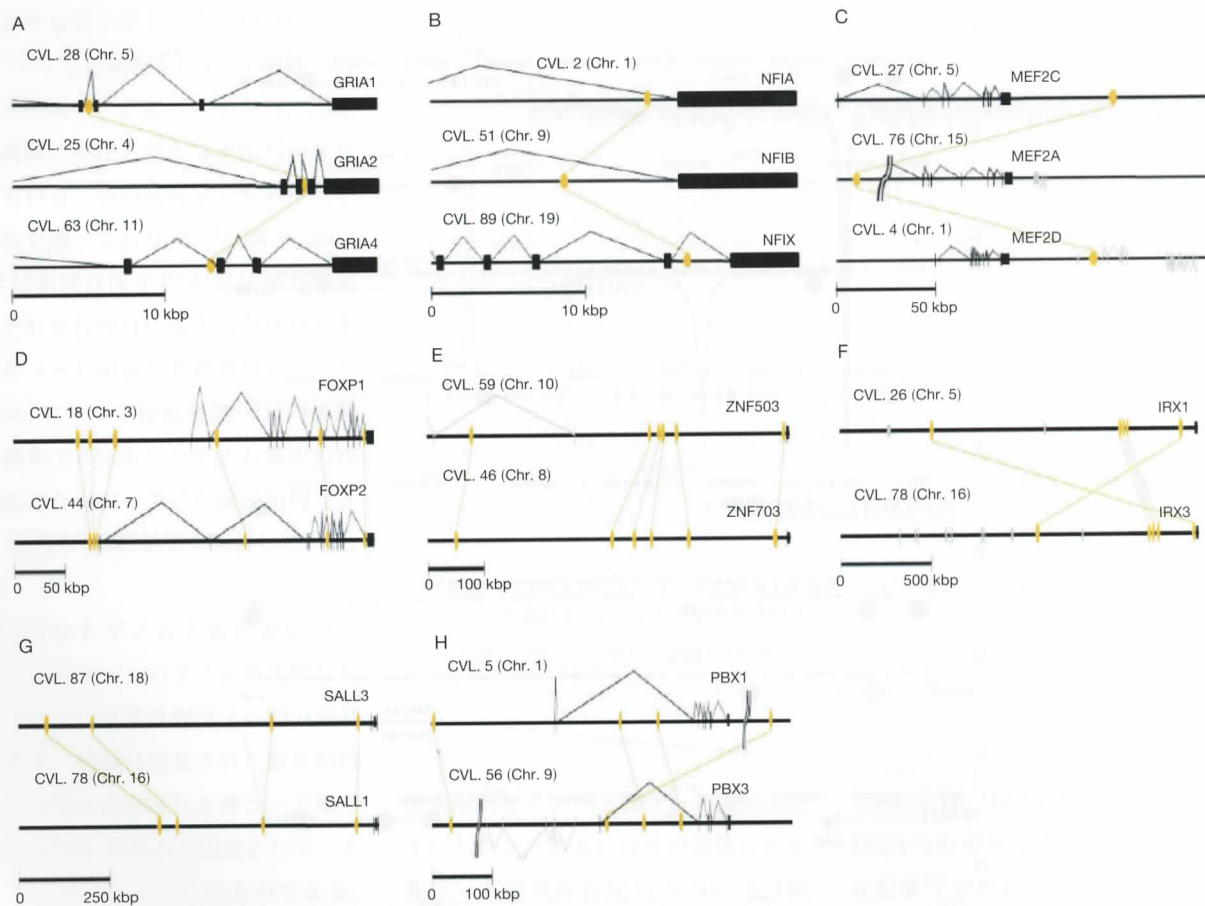


図3 傍系相同なCNSがやはり傍系相同なタンパク質コード遺伝子の近傍にある様子

くわしくは本文を参照

(文献17より)

【文献】

- 1) The ENCODE Project Consortium. (2012) *Nature*, **489**, 57–74.
- 2) Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T. *et al.* (2012) *Nature*, **489**, 75–82.
- 3) Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E. *et al.* (2012) *Nature*, **489**, 83–90.
- 4) Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K. *et al.* (2012) *Nature*, **489**, 91–100.
- 5) Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T. *et al.* (2012) *Nature*, **489**, 101–108.
- 6) The ENCODE Project Consortium. (2007) *Nature*, **447**, 799–816.
- 7) The FANTOM Consortium. (2005) *Science*, **309**, 1559–1563.
- 8) Johnson, J. M., Edwards, S., Shoemaker, D. & Schadt, E. E. (2005) *Trends Genet.*, **21**, 93–102.
- 9) Yasuda, J. & Hayashizaki, Y. (2008) *Adv. Cancer Res.*, **99**, 77–112.
- 10) van Bakel, H., Nislow, C., Blencowe, B. J. & Hughes, T. R. (2010) *PLoS Biol.*, **8**, e1000371.
- 11) 木村資生 (向井輝美/日下部真一・訳). 分子進化の中立説, 紀伊國屋書店, 1986.
- 12) 根井正利 (五條堀孝/齋藤成也・訳). 分子進化遺伝学, 培風館, 1990.
- 13) 齋藤成也. ゲノム進化学入門, 共立出版, 2007.
- 14) Saitou, N. (2013) *Introduction to evolutionary genomics*. Springer-Verlag (出版予定).
- 15) Matsunami, M., Sumiyama, K. & Saitou, N. (2010) *J. Mol. Evol.*, **71**, 427–436.
- 16) Takahashi, M. & Saitou, N. (2012) *Genome Biol. Evol.*, **4**, 641–657.
- 17) Matsunami, M. & Saitou, N. (2013) *Genome Biol. Evol.*, **5**, 140–150.
- 18) Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J. *et al.* (2004) *Science*, **304**, 1321–1325.
- 19) Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L. *et al.* (2012) *Nature*, **488**, 116–120.
- 20) Pennisi, E. (2012) *Science*, **337**, 1159–1161.
- 21) Eddy, S. (2012) *Curr. Biol.*, **22**, R898–R899.
- 22) Graur, D., Zheng, Y., Price, N., Azevedo, R. B. R., Zufall, R. A. & Elhaik, E. (2013) *Genome Biol. Evol.*, (published online on February 20, 2013).
- 23) Ohno, S. (1972) *Brookhaven Symp. Biol.*, **23**, 366–370.

齋藤 成也 *Naruya Saitou*

国立遺伝学研究所 教授