

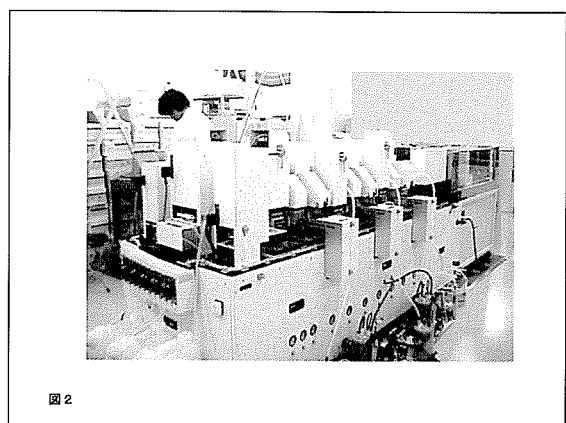
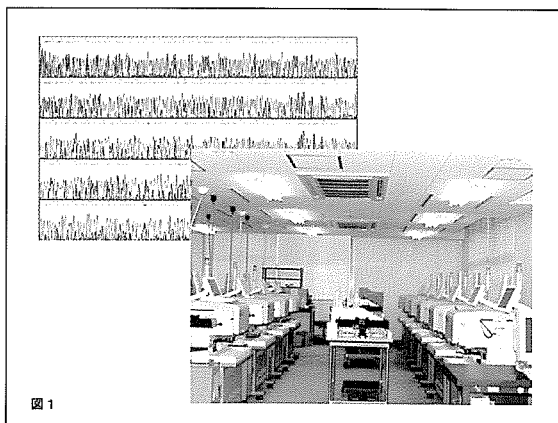
# 高速ゲノム解析システムの開発と疾病遺伝子の単離 新生児一過性糖尿病とヒトインプリント遺伝子

林 崎 良 英

(理化学研究所 ゲノム科学総合研究センター プロジェクトディレクター)

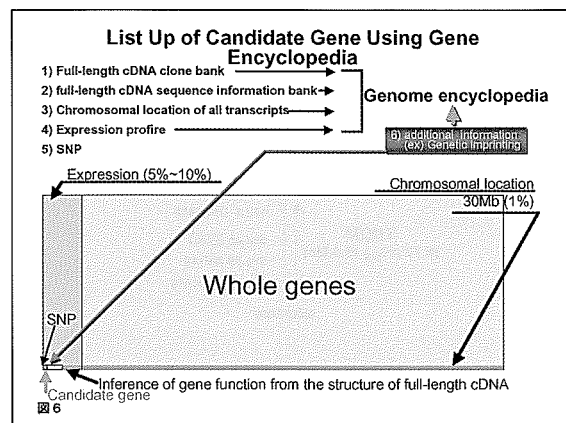
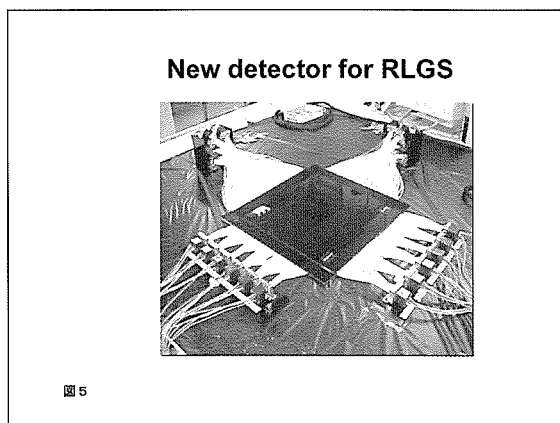
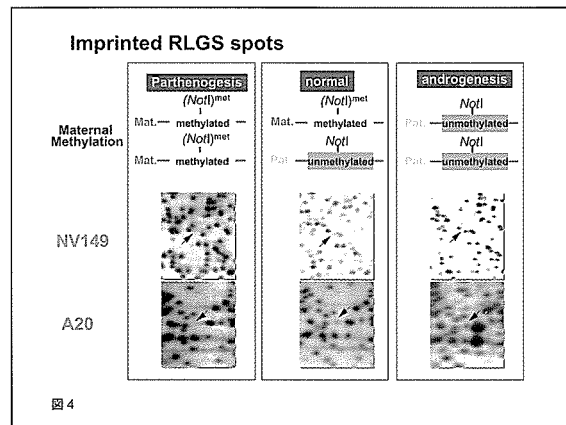
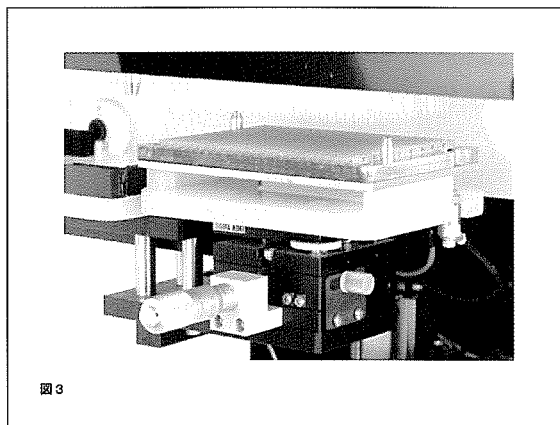
ゲノムDNA上に書きこまれている情報は、塩基配列と修飾（メチル化）が、現在知られている限り、ゲノムDNA上にある情報の全てである。本プロジェクトでは、高速に引き出すシステムを開発し、遺伝子単離に応用するため、その各々につき最高速のシステムとして、RISA sequencer (RISA:Riken Integrated Sequence Analysis System) (ref.1, 2, 3, 4, 5, 6, 7, 8) とRLGS (Restriction Landmark Genomic Scanning) の総合システムの開発に成功した (ref.9)。RISAシステムは塩基配列を高速に決定するシステムであり、それを用いて、マウスの転写単位を70%以上カバーすると推定されるマウス完全長cDNAの収集を可能にした。近年、高度の機能注釈を付けたデータベースと完全長cDNAクローンバンクを完成し、そのうち21,000個に関して、論文発表した (ref.10, 11)。一方、ゲノム上のメチル化をスクリーニングするためには、RLGS法が、最高速の手法である (ref.12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22)。メチル化がGenomic Imprintingの主要なメカニズムであることが予想されることから、これを指標として、ヒトゲノム上においてGenomic Imprintingという遺伝形式に従う伝達様式をしめすメチル化を見つけ出し、ヒト新生児一過性糖尿病遺伝子、偽副甲状腺機能亢進症の遺伝子のpositional candidate cloningに成功した (ref.23, 24)。

RISAシステムは、ゲノムDNA上の塩基配列情報解読を高速化する為の、高速シーケンシングシステムである。シーケンシングの速度を上げるためには、検体の多本並列化をもって高速化を図った (ref.1, 2, 3, 4, 5, 6, 7, 8)。そのため、本プロジェクトでは、384 Capillary Sequencer (RISAシーケンサー) を開発した (図1)。このマシンは、384プレートから一度に384キャピラリーアレイを通してサンプルがInjectionでき、独自のゲル泳動系で平均600から650bpを解読できる。一台により、90000bp/台/時間が実現できるようになった。また、理研で開発した完全長cDNA技術 (ref.25, 26, 27, 28, 29, 30) やプラスミド調製機 (図2 ; ref.31, 32) などその他一連のシステムと合わせ、50000検体/日/16台の速度で、生体組織からデータまでの384穴プレートによる (図3) 一貫システムが完成した。これにより、マウスの全完全長cDNAを収集する理研マウスcDNAプロジェクトが急速に加速された。このRISAシーケンサー



は、現在商品化されるに至った。また、メチル化をスクリーニングするシステムとして、NotIのようなメチル化感受性（図4）GC-rich制限酵素サイトランドマークを二次元パターン上に視覚化するRLGSを開発した（ref.12, 33, 9）。このシステムにより、ゲノム上の多くのCpG-islandをスクリーニングできる。この二次元パターンを迅速かつ精度良く行うデバイスの一つとして、シンチレーションファイバーを用いたβ線二次元画像検出器（図5；ref.9）を開発し、そのパターンを自動比較するコンピュータプログラムも新規開発した（ref.19）。これらを用いて、ヒトのImprint遺伝子のスクリーニングを行い、それを用いてヒト新生児一過性糖尿病遺伝子、偽副甲状腺機能亢進症の遺伝子を単離した。

RISAシステムを用いて、マウスゲノムエンサイクロペディアを完成させつつある。完全長cDNAを200種類の組織から単離し（ref.25, 26, 27, 28, 29, 30）100万クローンをピックアップした。それらの3'端のシーケンスを決定し、分類分けしたところ、128000種類に分けられた。その内、21000に付き、全長に渡り、シーケンスを決定したところ、16000種類の独立したシーケンスに分類された。マウス遺伝子エンサイクロペディアは、完全長cDNAクローンバンク（ref.10, 11）、完全長cDNAシーケンスデータベース（ref.34, 35, 36, 37, 10, 11）、全遺伝子の染色体の位置情報（ref.10, 11）、全遺伝子の発現データベース（ref.38, 39）を完備した物である。このようにマウスの完全長cDNAをベースとするマウス遺伝子エンサイクロペディアが完成しつつあるので、このゲノムリソースを使い、表現形質と遺伝子を結びつける高速遺伝子単離システムを作った。表現形質とそれを支配する遺伝子を結びつけるためには、全遺伝子の中から、目的の物を選ぶ指標が重要である。まずひとつは、染色体上の位置情報である。これは遺伝学により、病気の原因遺伝子をマッピングできる。さらに、発現情報はさらに候補遺伝子を絞り込むのに有用である。特に、Genetic Imprintingのような、発現情報などが加わると特定の遺伝子を

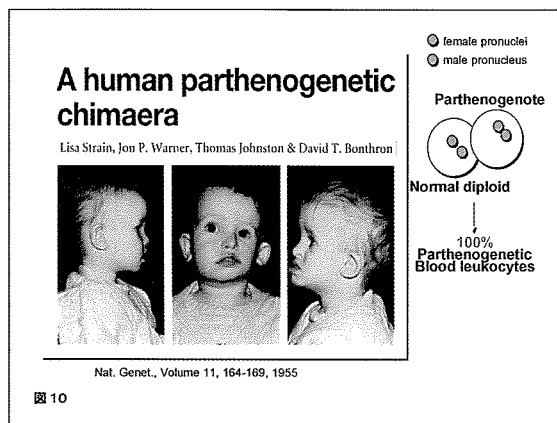
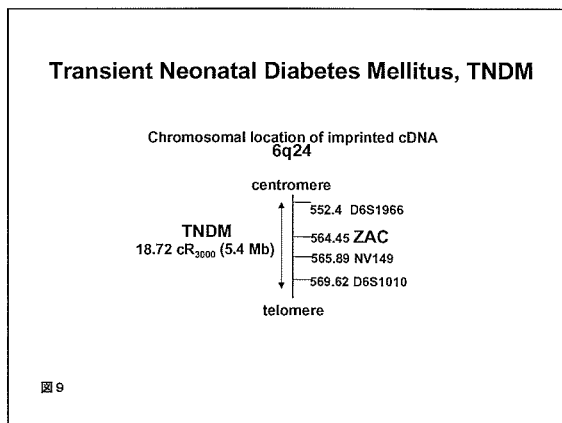
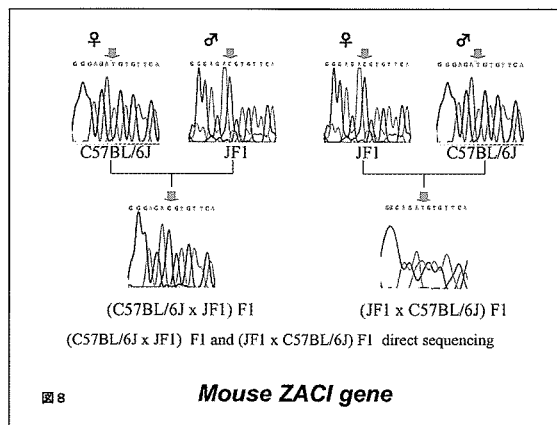
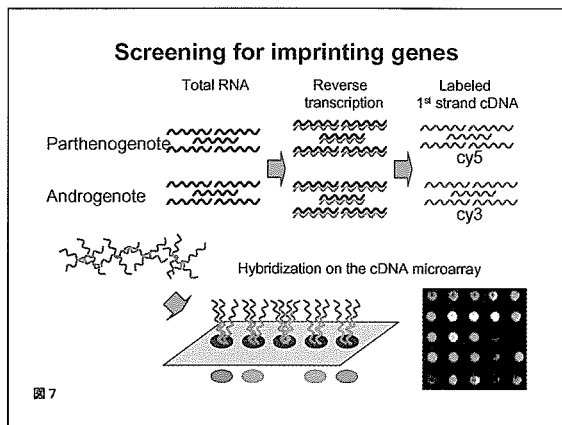


同定するのに非常に有用である (図6)。

ゲノミックインプリンティングとは、ある遺伝子が父親由来または母親由来のアリールより選択的に発現する現象をいう。多くのインプリント遺伝子の近傍から、メチル化が一方の親由来のアリールに特異的な領域が見つかっており、インプリントの成立と維持に必要な要素であると考えられている。インプリント遺伝子をスクリーニングするために、マウス遺伝子エンサイクロペディアから、約20000種類のcDNAをcDNAマイクロアレイ上に並べ、それをもちいて、マウスの雄性単為発生体、雌性単為発生体からmRNAを得、両者で発現の差のある遺伝子をインプリント遺伝子の候補とした (図7, 8)。それらの候補遺伝子を染色体上にマップすると、新生児一過性糖尿病がマップされているヒト6p24にマップされた (図9) マウスインプリント候補遺伝子があった。ZAC/PLAGL1であった (ref.23, 24)。

一方、ヒトで直接ゲノム上のメチル化を指標に、インプリント遺伝子のスクリーニングも行った。Bonthronらの報告したヒト単為発生キメラ (図10; ref.40) は、そのリンパ球が全て単為発生由来で、インプリント遺伝子の母親アリール特異的なメチレーションを保持している事が確認されている。一方、胞状奇胎は雄核発生を原因として発症し、メチレーションの状態も父親アリールの状態を反映している。そこで、これらのサンプルのメチル化状態を上記のRLGS法で比較し、ヒトインプリント遺伝子の系統的な探索を行った。全ゲノム領域にわたってホモ接合体の細胞を使用するので、多型を示さないスポットであっても、ゲルパターン上の2000から3000の全てのスポットが検出できる (ref.12)。

この戦略を用い、約4,500の座位のメチル化状態をスクリーニングする事に成功し、インプリントの挙動を示す二つの座位、A20とNV149を発見した。A20はG-protein- $\alpha$ -subunit 1 (GNAS1) 遺伝子のAlternative Spliceされた転写単位中にあることがわかり、この遺伝子がインプリント



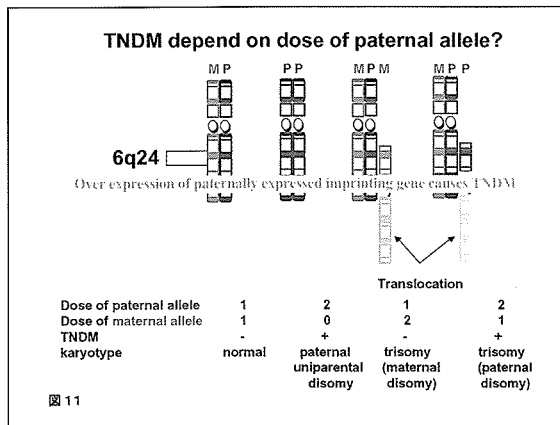


図 11

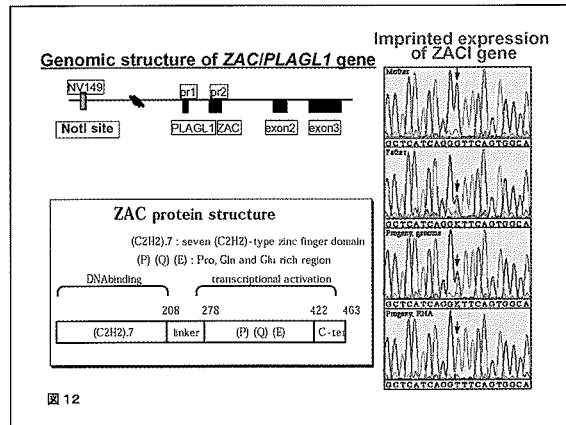


図 12

していることが判明した。この転写単位は従来のインプリントしていない転写単位の上流から始まる新しい転写単位でこの新しいG-protein-subunit 1 (GNAS1) 遺伝子の転写単位こそが、インプリント疾患、偽副甲状腺機能亢進症type Iaの原因遺伝子であることをつきとめた (ref.24)。

驚いたことに、もう一方のNV149は、その症状がインプリントしている新生児一過性糖尿病 (TNDM) の領域に位置している事を発見した。

このNV149は、cDNAマイクロアレイで単離されてきた、ZAC/PLAGL1その物であった。そもそも、TNDMはUPD (UniparentalDisomy) の患者や6q24領域の染色体転座の研究から、父方対立遺伝子特異的に発現している遺伝子の発現量が増加することが原因と考えられていた (図11)。NV149近傍の転写単位を調べた結果、ZAC/PLAGL1が父親由来のアレルから特異的に発現していることを明らかにした。ZAC/PLAGL1はZinc-Finger (図12) 蛋白質をコードしており、転写因子である事、癌抑制遺伝子である事、アポトーシスを起こす事などが既に示されていた。さらに最も興味深い点として、(図13) PACAP (pituitary adenylate cyclase activating polypeptide) -type1 receptorを誘導することが報告されている。PACAPは、膵臓では、グルコース存在下でインスリンの分泌を促進する事が報告されている。このことから、ZAC/PLAGL1が新生児一過性糖尿病の原因遺伝子であると考えられる (ref.23)。

このように、我々は、DNA塩基配列を高速決定する手法である、RISAシステムと、DNAのメチル化を決定するRLGSシステムを完成した。これらは、高速に個体における表現形質から遺伝子を単離同定するために極めて有用かつ強力なツールとなる。

#### References:

1. Ito, M. et al. *Nucleic Acides Res.*, **25**, 1315-1316 (1997)
2. Sasaki, N. et al. *Proc. Natl. Acad. Sci. USA.*, **95**, 3455-3460 (1998)
3. Izawa, M. et al. *J. Biol. Chem.* **273**, 14242-14246 (1998)
4. Sasaki, N. et al. *GENE*, **222**, 17-24 (1998)
5. Itoh, M. et al. *Genome Res.*, **9**, 463-470 (1999)
6. Shibata, K. et al. *Genome Res.*, **10**, 1757-1771 (2000)
7. Sasaki, N. et al. *DNA Res.*, **4**, 387-391 (1997)
8. Mizuno, Y. et al. *Nucleic Acides Res.*, **27**, 1345-1349 (1999)

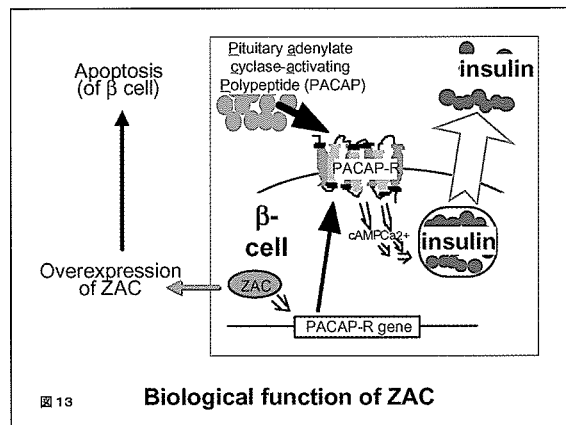


図 13

9. Morimoto, K. et al. *IEEE Transaction on Nuclear Science*, **47**, 2033-2038 (2000)
10. The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium, *Nature*, **409**, 685-690 (2001)
11. International Human Genome Sequencing Consortium, *Nature*, **409**, 860-921 (2001)
12. Hatada, I. et al. *Proc. Natl. Acad. Sci. USA.*, **88**, 9523-9527 (1991)
13. Kawai, J. et al. *Nucl. Acid. Res.*, **21**, 5604-5608 (1993)
14. Hayashizaki, Y. et al. *Nature Genetics*, **6**, 33-40 (1994)
15. Kawai, J. et al. *Mol. Cell Biol.*, **14**, 7421-7427 (1994)
16. Shibata, H. et al. *Electrophoresis*, **16**, 210-217 (1995)
17. Plass, C. et al. *Nature Genetics*, **14**, 106-109 (1996)
18. Akama, T. O. et al. *Cancer Res.*, **57**, 3294-3299 (1997)
19. Sugahara, Y., et al. *Mammal. Genome*, **10**, 611-616 (1999)
20. Takada, S. *Genomics*, **61**, 92-100 (1999)
21. Komatsu, S., et al. *BBRC*, **267**, 109-117 (2000)
22. Tateno, M., et al. *Cancer Research*, **61**, 1144-1153 (2001)
23. Kamiya, M., et al. *Hum. Mol. Genet.*, **9**, 453-460 (2000)
24. Hayward, B.E., et al. *Proc. Natl. Acad. Sci. USA.*, **95**, 10038-10043 (1998)
25. Carninci, P., et al. *Genomics*, **37**, 327-336 (1996)
26. Carninci, P., et al. *Genome Res.*, **10**, 1617-1630 (2000)
27. Carninci, P., et al. *DNA Res.*, **4**, 61-66 (1997)
28. Shibata, Y., et al. *Biotechnology*, in press
29. Carninci, P., et al. *Proc. Natl. Acad. Sci. USA.* **95**, 520-524 (1998)
30. Carninci, P., et al. "Methods in Enzymology" Academic Press, Inc. San Diego, **303**, 19-44 (1999)
31. Itoh, M., et al. *Genome Res.*, **9**, 463-470 (1999)
32. Shibata, K., et al. *Genome Res.*, **10**, 1757-1771 (2000)
33. Sugahara, Y., et al. *Mammal. Genome*, **9**, 643-651 (1998)
34. Fukunishi, Y., et al. *FEBS Lett.*, **464**, 129-132 (1999)
35. Sugahara, Y., et al. *GENE*, **263**, 93-102 (2001)
36. Konno, H., et al. *Genome Res.*, **11**, 281-289 (2001)
37. Fukunishi Y., et al. *Physiological Genomics*, **5**, 81-87 (2001)
38. Miki R., et al. *Proc. Natl. Acad. Sci. USA.*, **98**, 2199-2204 (2001)
39. Kadota, K., et al. *Physiological Genomics*, **4**, 183-188 (2001)
40. Strain L., et al. *Nat. Genet.*, **11**, 164-169 (1995)

#### News:

1. *Science*, **291**, 963-964, Feb. (2000)
2. *Science*, **278**, 1677-1848, Dec. (1997)
3. *Nature*, **392**, 219, Mar. (1998)
4. *Science*, **283**, 455, Jan. (1999)
5. *Nature*, **397**, 98-99, Jan. (1999)
6. *Science*, **286**, 445, Oct. (1999)
7. *Nature Genetics*, **25**, 371, Aug. (2000)
8. *Nature*, **407**, 279, Sep. (2000)
9. *Genome Research*, **10**, 1431-1432, Oct. (2000)
10. *Nature Genetics*, **26**, 255-256, Nov. (2000)
11. *日経サイエンス (Scientific American Japan Edition)*, **350**, 118-120, Dec. (2000)
12. *Science*, **288**, 257, Apr. (2000)
13. <http://www.nature.com/nsu/010208/010208-11.html>  
*Nature Science Update*, Lifelines:Genes of mice and men, Feb. (2000)
14. [http://dailynews.yahoo.com/h/nm/20010207/sc/genome\\_japan\\_dc\\_1.html](http://dailynews.yahoo.com/h/nm/20010207/sc/genome_japan_dc_1.html)  
*Science News*, Japan Firm Finds Potential Cancer-Fighting Genes, Feb. (2001)