

## FPGA を用いた高速プライマー配列探索システムの開発

千葉 啓和<sup>1</sup>, 中川 草<sup>2</sup>, 谷口 丈晃<sup>3</sup>, 藤渕 航<sup>1</sup>  
<sup>1</sup>産業技術総合研究所 生命情報工学研究センター,  
<sup>2</sup>国立遺伝学研究所, <sup>3</sup>三菱総合研究所

近年ではパイロシーケンシングという技術を用いて 1 細胞内の mRNA を絶対定量することが可能になっている。しかし、プライマーの特異性を保証するためには mRNA データベースとの配列アラインメントが必要であり、膨大な計算時間が必要となる。我々は FPGA(field programmable gate array)上に、オリゴマーのアラインメントを並列に実行する回路を実装し、計算の高速化を行った。その結果、本システムにより general CPU に比べて約 1,000 倍のアラインメント高速化が可能であることが分かった。本研究では、アラインメントの他にも、クロスハイブリダイゼーション、プライマーダイマーの形成、プライマー内の 2 次構造形成の割合も評価して、最良のプライマー配列を探索する。

## Development of a fast primer search system using FPGA

Hirokazu Chiba<sup>1</sup>, So Nakagawa<sup>2</sup>, Takeaki Taniguchi<sup>3</sup>, Wataru Fujibuchi<sup>1</sup>  
<sup>1</sup>Computational Biology Research Center, Advanced Industrial Science and Technology,  
<sup>2</sup>National Institute of Genetics, <sup>3</sup>Mitsubishi Research Institute Inc.

Quantitation of mRNA levels in a single cell can be realized by the recent pyro-sequencing technology. However, to assure primer specificity, we need to perform a huge number of sequence alignments of the primer candidates against mRNA database, which requires extremely large computational costs. Thus, we adopt field programmable gate array (FPGA), on which logic circuits for hundreds of parallel computation of oligomer comparison are implemented. This system can execute sequence alignments about 1,000 times faster than a general CPU. To select the best primers, we also check the possibilities of cross-hybridization, primer dimer formation and secondary structure formation.

## 1. はじめに

パイロシーケンシング技術を用いることにより、1細胞内の mRNA 量を絶対測定することが可能である。しかしながら、プライマーの特異性を保証するためには、プライマーがクロスハイブリダイゼーションを起こすターゲット（オフターゲット）を探索し、特異性を評価することが必要である。オフターゲットの探索の際には、mRNA データベースとのアラインメントを正確に行うため、膨大な計算時間が必要となる。そこで我々は、FPGA(field programmable gate array)[1]上にプライマーのアラインメント専用の回路を実装し、計算の高速化を行った。また、アラインメントの後、熱力学的計算[2]を行いプライマーの特異性の評価を行った。特異的プライマー設計の流れを図1に示す。以下、FPGA によるアラインメントの高速化と、熱力学的計算による評価に分けて説明する。

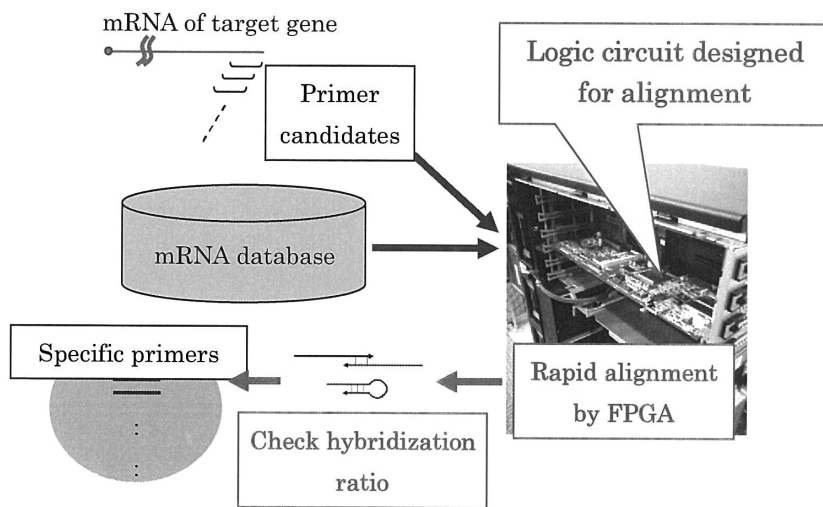


図 1. 特異的プライマー設計の流れ

## 2. FPGA によるアラインメントの高速化

FPGA とは、書き換え可能な回路を納めることのできるハードウェアである。我々はアラインメントを行う回路を Verilog HDL を用いて記述し、Xilinx ISE Design Suite 10.1 によって回路の合成、インプリメンテーションを行い、Xilinx Virtex-5 FPGA 上で実行した。このアラインメント回路によって、25nt のプライマー配列と任意の長さの塩基配列のギャップなしアラインメントを行うことができる。25nt の配列 2 つを比較する回路をユニットとして、それらを並列につなぎ合わせている。配列比較ユニットは 20ns で処理を完了するため、これを 100 個並列につなぐことにより、25nt と 100nt のアラインメントを 20ns で実行することができる。アラインメント回路の概念図を図 2 に示す。同じアラインメントアルゴリズムを C 言語を用いて実装し、Intel Xeon CPU 3.8GHz で実行時間を計測したところ、25nt どちらの比較に 200ns、25nt と 100nt のアラインメントに 20 $\mu$ s を要した。従って、アラインメントを FPGA 上で実装することで 1000 倍の高速化が可能であることが分かった。

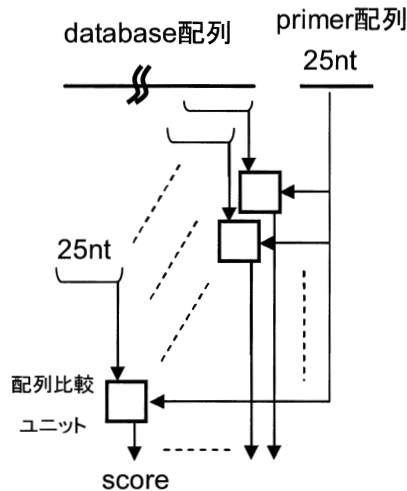


図 2. アラインメント回路の概念図

### 3. プライマー配列の評価

プライマー配列とヒトの mRNA データベース(Ensembl 47)とのアラインメントを行い、オフターゲット候補の上位 100 配列を取り出した。その後、プライマー配列とオフターゲット候補との hybridization の自由エネルギーを hybrid-min[3]によって計算し、さらに hybridization ratio[4]を求めた。この値によって、プライマーがどの程度のクロスハイブリダイゼーションを起こすかを評価することができる。また、プライマーとターゲット配列との Tm 値も求めて評価に用いた。プライマーどうし、あるいはプライマー内での相補性に関しても、以下のようにして評価を行った。プライマーダイマー形成の自由エネルギーを hybrid-min によって計算し、プライマーどうしの hybridization ratio を求めた。プライマー内の 2 次構造形成の自由エネルギーは hybrid-ss[3]によって計算し、2 次構造を形成する割合を求めた。このようにして求めたプライマーの 3 つの特性 (1) プライマーとオフターゲットとの hybridization ratio、(2) プライマーどうしの hybridization ratio、(3) プライマー内 2 次構造形成の割合を、以下に示す関数によって統合し、プライマー候補の評価に用いた。

$$F = -2 \sum_{i=1}^3 \log(x_i)$$

プライマー評価の一例を以下に示す。図 3 は、ヒトの ERBB2 遺伝子に対するプライマー候補を評価した結果である (mRNA の 3'端から 1000nt の部分)。この評価結果に対して適当な条件を設定することによって、実験者のためのプライマーリストを得ることができる。例えば、オフターゲットに対する hybridization ratio が 0.05 未満のもの、かつプライマーとターゲット配列との Tm 値が 60℃～70℃のものに限定することで、プライマー配列 46 候補(図 3 の影の部分)が得られる。これらのプライマー候補を、上記 F の値でソートした上で実験者に提供する。

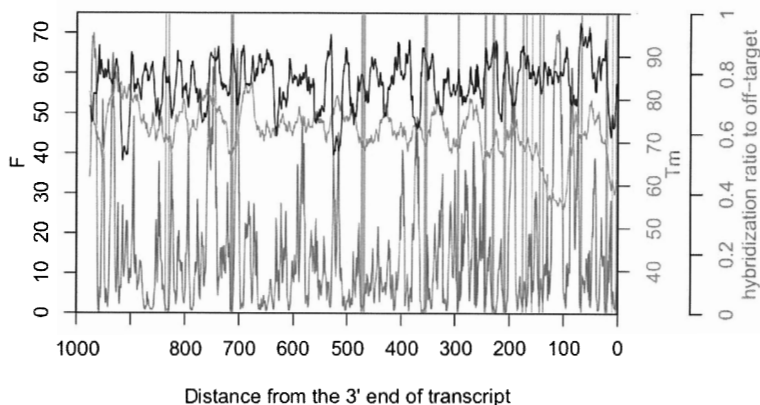


図 3. ヒト ERBB2 遺伝子に対するプライマー候補の評価結果

#### 4. おわりに

本研究で設計された特異的プライマーは、実際にパイロシーケンシングによる実験で使われ始めており、有効性の確認が行われている。我々のシステムは、アラインメント回路の並列度を高めることによってさらに高速化することが可能であり、400 の回路を並列につなぎあわせると、全ヒト遺伝子のプライマー設計が 4 日で可能となる計算である。このように設計したプライマーセットを用いて、細胞内の mRNA 量を網羅的に絶対定量したデータが得られれば、細胞の状態をデジタル化した貴重な情報となり、細胞の解析のためのプラットフォームとして重要な役割を果たすであろう。本研究は、細胞解析のプラットフォームを形成していくための、計算機の基礎を構築するものである。

#### 参考文献

- [1] Li ITS, Shum W, Truong K 160-fold acceleration of the Smith-Waterman algorithm using a field programmable gate array (FPGA). *BMC Bioinformatics* 2007, **8**:185.
- [2] SantaLucia J, A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*. 1998, **95**:1460-1465
- [3] Markham NR, Zuker M, UNAFold: software for nucleic acid folding and hybridization. In Keith JM, editor, *Bioinformatics, Volume II. Structure, Functions and Applications (Methods in Molecular Biology)* 2008, chapter 1, pages 3–31.
- [4] Yamada T, Soma H, Morishita S, PrimerStation: a highly specific multiplex genomic PCR primer design server for the human genome. *Nucleic Acids Res.* 2006, **34**(Web Server issue):W665-W669.