

統合的 sequence motif解析システムの開発とその応用

明石浩史¹⁾ 丸山玲緒²⁾ 青木文夫¹⁾ 豊田 実³⁾ 佐々木泰史²⁾ 戸倉 一¹⁾ 西城一翼¹⁾ 山口徳蔵¹⁾
見田裕章²⁾ 苗代康可²⁾ 辰巳治之⁴⁾ 今井浩三³⁾ 時野隆至²⁾

札幌医科大学附属情報センター¹⁾ 札幌医科大学医学部附属がん研究所分子生物学分野²⁾
札幌医科大学医学部第一内科³⁾ 札幌医科大学医学部第一解剖⁴⁾

Development and appriciation of comprehensive system for analyzing sequence motif

Hirofumi Akashi¹⁾ Reo Maruyama²⁾ Fumio Aoki¹⁾ Minoru Toyota³⁾ Yasushi Sasaki¹⁾ Hajime Tokura¹⁾
Kazuyoku Saijo²⁾ Yasuyoshi Naishiro²⁾ Haruyuki Tatsumi⁴⁾ Kohzoh Imai³⁾ Takashi Tokino²⁾

Information Center of Computer Communication, Sapporo Medical University¹⁾
Department of Molecular Medicine, Cancer Research Institute, Sapporo Medical University²⁾
First Department of Internal Medicine, Sapporo Medical University³⁾
First Department of Anatomy, Sapporo Medical University⁴⁾

Abstract: We have developed a web-based genome-wide analyzing system on DNA motifs. The system was predicted too many p53 binding sites to analyze. In this time, we have looked for ways to narrow the list down to adequate number. In results, we demonstrated that the functional p53 binding sites are within 10kb around TSS(Transcription start site) and detected in similar sites to mouse ortholog.

Keywords: bioinformatics, genome, sequence motif, ortholog

1. 背景と目的

現在、ゲノムプロジェクトの進展により、ヒトはもとよりマウスをはじめとして多くの生物種のゲノム情報を解析に利用できるようになってきている。われわれは、これまでネットワークパラレルコンピューティング技術により全ゲノム塩基配列を対象に癌抑制遺伝子p53の結合配列を高速に検索できるWEBベースアプリケーションを開発し報告してきた¹⁾。開発したシステムによりゲノム上にp53結合部位のコンセンサス配列 [RRRCWWGYYY]+スペーサ+[RRRCWWGYYY]と80%以上一致する部位をヒトで483万箇所、マウスで464万箇所予測し、これらの予測箇所のデータベース化、Image Map上へのマッピングによる可視化を行ってきた。これら予測箇所がゲノム上の全遺伝子数と比較し莫大であり、それら全てを分子生物学的手法により解析するのは不可能であり、またそれらのうち実際に機能している結合部位は一部であると考えられる。そのためp53標的遺伝子検索のためには、これら予測箇所を絞込み有意な結合部位を同定する必要がある。この目的のために、転写開始点からの距離およびマウスOrthologとの比較による絞り込み解析を行った結果につき報告する。

2. 転写開始部位からの距離

ゲノム上に遺伝子情報をマッピングした遺伝子アノテーションデータとしてUCSCのRefFlat (<http://genome.ucsc.edu/goldenPath/hg16/database/refFlat.txt.gz>)を用いて、ヒト19174遺伝子、マウス13406遺伝子の解

析を行なった。具体的には図1に示すように転写開始点(TSS)と、予測したp53結合部位の相対的な位置関係を検討した。その結果、既知のp53標的遺伝子においては、TSSの上流、下流10kb以内にp53結合部位コンセンサス配列と17/20以上一致する部位が半数近くの遺伝子で見出されることが明らかになった。

3. マウス Ortholog との比較

有意な結合配列は種間で保存されうると仮定し、NCBIのGenBankより抽出した7444組のヒト-マウスOrthologにおいて、p53結合部位の位置、配列を比較した(図2)。その結果、既知p53標的遺伝子18個において、ヒト-マウスOrtholog間でp53結合部位の位置、配列(ATGCレベル)が極めて類似していることが分かった。同様の傾向を示す幾つかの遺伝子に関して分子生物学的解析を行い、それらのほとんどがp53標的遺伝子である可能性が示唆されており現在解析中である。

4. 今後の検討

現在、更なる絞り込み方法の開発をめざし、他の生物種のゲノムとの比較、既存の汎用データマイニングシステム(WekaおよびVisual Mining Studio)と開発したシステムの連携により決定木ならびにニューラルネットワークを使用した結合部位周囲の配列解析などを施行している。

参考文献

- [1] Aoki F, Akashi H, Goudge M, Toyota M, Sasaki Y, Guo X, Li S, Tokino T, Tatsumi H. Post-Genome Applications Based on Multi-Parallel Computing over High Performance Network. International Workshop on Next Generation Internet and its Applications -Bio Medical Applications ? Pre-Proceedings, 61-67, 2001.

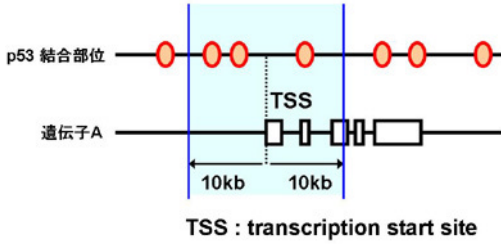


図1 転写開始点からの距離による絞込み

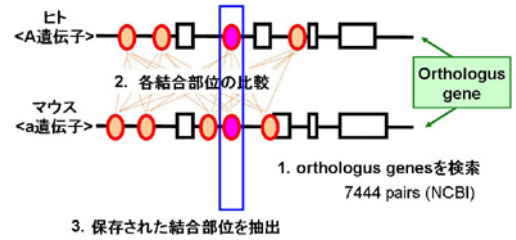


図2 Ortholog 間の比較による絞込み