

論文内容の要旨

論文題目：

転写因子 STAT6 に直接結合するクロマチン部分の網羅的解析

(Analysis of Chromatin DNA Directly Bound by Transcription Factor STAT6)

氏名： 金 井 昭 教

<背景と目的>

転写因子標的遺伝子の同定には、転写因子の結合の有無、下流遺伝子の発現の変化の情報は必須である。しかし、転写因子結合の有無を調べるために従来のクロマチン免疫沈降法 (ChIP) とマイクロアレイを組み合わせた方法 (ChIP-CHIP 法) では、ゲノムの全領域を解析するためのタイリングアレイの煩雑は処理が必要である等、技術面およびコスト面での困難が伴う。またマイクロアレイを用いた遺伝子発現解析においては、高い定量性をもって、あるいは正確な転写開始点(TSS)情報を加味した形で解析を進めることは困難である。一方、cDNA ライブラリーのランダムシーケンスを基盤とした方法では、十分なカバレッジを得ることは事実上、不可能である。

近年の新型シーケンサーの開発により、短時間で大量に短い塩基配列の決定をすることが可能となった。これらの技術を応用すれば、得られた配列数を指標に従来の検出系を置換することが可能である。すなわち、ChIP-CHIP 法におけるハイブリダイゼーションをシーケンサーで得られる大量の短い塩基配列データに置換し、その配列をリファレンスとなるゲノム配列にマッピングすることでそのゲノム上の位置情報を

得ることが出来る（以下、ChIP-Seq 法と呼ぶ）。また、転写開始点情報、あるいは遺伝子発現情報についても cDNA の 5'末端を大量にシークエンスすることで (TSS-Seq) 正確な転写開始点情報が得られ、さらにその配列数をデジタルにカウントは、その転写開始点からの発現量を示す発現情報として計測することが可能である。筆者はイルミナ社の新型シークエンサーGenome Analyzer (GA) を用いて 1) ChIP-Seq による転写因子の結合 2) TSS-Seq による転写開始点、発現頻度情報 3) それらを組み合わせたプロモーター解析を用いることによって転写因子の細胞内での標的遺伝子を同定することが可能であると考え、その方法論的開発を行った。

本研究では上記の手法を転写因子 STAT6 の標的遺伝子の探索に応用することを試みた。STAT6 は IL-4・IL-13 のようなサイトカインにより転写活性化能が誘導される転写活性化因子であり、喘息をはじめとする様々なアレルギー性疾患との関連が示唆されている。STAT6 は核内では DNA の TTCNNNGAA という 4 つの N によって隔てられたコンセンサス配列を認識して DNA に結合し、イムノグロブリンやサイトカイン・ケモカインのような遺伝子発現を活性化することにより、アレルゲンに対する生体防御反応である Th2 細胞への分化や IgE へのクラススイッチを促す。しかし、その標的遺伝子の全体像について、また標的遺伝子の細胞種特異性について、現在の得られている知見は非常に限られている。

<実験方法と解析手法>

1) ChIP-seq

ヒト B リンパ球 Ramos 細胞・ヒト気道上皮 BEAS2B 細胞に対して、IL-4 刺激を 30 分行った。その後、常法に従って抗 STAT6 抗体を用いた ChIP を行った。回収された DNA に対して、イルミナ GA シークエンサーを用いて、36 bp の塩基配列を各実験条件につき、約 500 万本ずつ決定した。決定された 36 塩基の配列をリファレンスとなるヒトゲノム配列 (UCSC Genome Browser; hg18) に ELAND を用いてアラインメントし、ミスマッチなしでユニークにアラインした領域を同定した。得られたゲノム座標に対して 120 bp でのクラスタリングを行い、各クラスターの塩基配列数を算出した。刺激依存的に 10 倍以上の濃縮が見られるクラスターを STAT6 の結合領域とした。また、クロマチンの状態を調べるために H3K4me3、H3Ac、RNA ポリメラーゼ II (Pol II) の抗体を用いて同様に ChIP-Seq を行った。

2) TSS-seq

Ramos 細胞と BEAS2B 細胞を IL-4 存在、非存在下で 24 時間培養した。Total RNA を抽出し、オリゴキャップ法を用いてイルミナ GA アダプタープライマーをキャップ構造に置換した後、ランダムプライマーを用いて cDNA を合成した。15 サイクルの PCR

の後に 150-250 塩基の領域を PAGE によりサイズ分画した。得られた鋳型 cDNA を用いて、各実験条件につき、約 3000 万の転写開始点近傍塩基配列を決定した。配列をヒトゲノム配列に対して上記と同様の手法でアラインメントし、クラスターを精製した。TSS-Seq については 500 bp でのクラスタリングを行った。転写開始点クラスター (TSS Cluster; TSCs) が約 18000 種類の Ref-seq 遺伝子の転写開始点上流 50 kb から 3'exon の末端までに位置する場合、Ref-seq 遺伝子への対応付けを行った。1 ppm (100 万タグあたり 1 配列=1 タグ) 以上で IL-4 刺激依存的に 2 倍以上のタグ数の変化が見られた TSC を、IL-4 応答転写開始点とした。

3) プロモーター配列の計算機解析

TSC 上流の塩基配列中における STAT6 結合コンセンサス配列を探索した。各計算機ツールは perl を用いて独自に作成したものを用いた。

<結果と考察>

ChIP-Seq 法による STAT6 結合領域の同定

ChIP-Seq 法において、IL-4 刺激依存的に 10 倍以上のタグの濃縮が見られる領域を STAT6 結合領域としたところ、ゲノム中に Ramos 細胞では 556 ヶ所、BEAS2B 細胞で 467 ヶ所の STAT6 結合領域を同定することができた。これら STAT6 結合領域それぞれのゲノム上での位置について解析を行うと、BEAS2B 細胞では 77 %のサイトが RefSeq 遺伝子のプロキシマルプロモーター領域 (- 10 kbp ~ + 1 kbp; 転写開始点を 0 とする) に存在していたが、Ramos 細胞では同じ領域に 27 %しか存在しておらず、34 %がイントロン領域に存在していた。他にも両細胞の STAT6 結合サイトのコンセンサス配列を知るために MATCH にて解析を行うと Ramos 細胞では既知コンセンサスである TTCNNNGAA が最も多く見られたが、BEAS2B 細胞では既知コンセンサス配列よりも TCTCGCG というコンセンサス配列の方が多く見つかった。

TSS-Seq 法による TSS 情報および遺伝子発現情報と ChIP-Seq 情報の統合的解析

TSS-Seq 法を用いて Ramos 細胞、BEAS2B 細胞の転写開始点の同定と mRNA の発現解析を行った。TSS-Seq のデータと STAT6 ChIP-Seq のデータと組み合わせたところ、RefSeq 遺伝子のプロキシマルプロモーター領域において STAT6 が結合しており、TSCs が発現上昇している遺伝子は Ramos 細胞で 132 サイト、BEAS2B 細胞で 44 サイトであった。これら以外の領域として、これまでの RefSeq 遺伝子モデルでは同定されておらず選択的プロモーターと思われる領域に 45 サイトと 9 サイト、遺伝子間領域に 34 サイトと 6 サイトの STAT6 結合領域が Ramos 細胞、BEAS2B 細胞それぞれについて同定された。これらの STAT6 標的遺伝子について比較を行ったところ、両細胞での標

的遺伝子は全く重なっておらず、GO 解析によってそれら遺伝子の機能についても比較したが異なる働きをする遺伝子群がそれぞれ多く見られていた。さらに近年報告された Th2 細胞での STAT6 標的遺伝子の結果も同様に解析を行い比較してみたところコンセンサス配列は Ramos 細胞と同様に既知配列であったが、標的遺伝子は大きく異なっており、GO 解析での遺伝子機能も異なっていた。このように Ramos 細胞と BEAS2B 細胞では STAT6 標的遺伝子が大きく異なっているにもかかわらず、標的遺伝子群の発現レベルや発現変化量はほぼ同じであった。

次に細胞によって異なる転写誘導が各細胞系でのクロマチン状態が異なるためであると推測し、H3K4me3、H3Ac、Pol II の ChIP-Seq を両細胞にて行った。クロマチン活性化状態のマーカーである H3K4me3 と H3Ac は STAT6 標的サイトの転写開始点周辺に濃縮されており、またこれらの修飾が IL-4 による刺激以前に起きていることが明らかとなった。さらに Ramos 細胞において Ramos 細胞では標的ではないが BEAS2B 細胞では標的となっているような領域（サイレント領域）について解析を行ったところ STAT6 の結合や TSS タグの存在は確認できなかったが、Pol II は結合しておりヒストンは活性化型であった。一方、BEAS2B 細胞では Ramos 細胞同様に STAT6 標的領域のヒストン修飾は活性化の割合が少なく、Pol II の結合は IL-4 依存的に起きていた。これらの結果から STAT6 による転写活性化は活性化状態のクロマチンが形成されている領域に IL-4 依存的に STAT6 が結合し、さらなる細胞依存的な因子がその領域に結合することによって転写誘導が行われると考えた。このような細胞特異的因子を同定するために TRANSFC の転写因子とそのコンセンサス配列を用いて、解析を行ったところ BEAS2B 細胞では FAC1 遺伝子、Ramos 細胞では CEBPG 遺伝子のコンセンサス配列が STAT6 標的領域に多くみられ、それぞれの遺伝子発現は各細胞のみで上昇していた。これらの遺伝子群は転写因子 STAT6 と細胞特異的に STAT6 と協調して働く遺伝子群候補として初めて同定されたものである。

本論文においては超並列型シーケンサを用いる事によって多段階からなる転写制御のゲノムワイドなデータを一度に得て解析する事が可能となり、STAT6 の転写制御には全ての細胞で画一的な活性化クロマチン形成と転写因子結合のみでは表せない複雑な制御機構が細胞ごとにそれぞれ存在している可能性が示唆された。今回に同定された各 STAT6 標的候補遺伝子それぞれについての細胞レベルや組織レベルでの生化学的な解析はゲノムワイドな解析と比べてもさらに多くの詳細な検証を必要とする。しかし、このような細胞ごとの STAT6 標的遺伝子のカタログ化とその分類、特徴付けは IL-4 の免疫系における重要な役割を明らかにしていく上で重要なリソースとなるであろう。