

本論文は乳癌に統合的バイオインフォマティクスによる解析を適用し乳癌の悪性度に関連するシス制御モチーフの存在を示すものである。

様々な種類の癌の中でも、特に乳癌についてはマイクロアレイ技術による解析が広汎に行われている。乳癌は組織学的にも、予後、治療に対する反応性においても多様で、マイクロアレイを用いた発現解析は、このような表現型の多様性の下に潜む、トランスクリプトームの多様性を明らかにしてきた。しかしながらマイクロアレイにより得られた膨大なトランスクリプトームに関する知識と比べると、トランスクリプトームの多様性、更には表現型の多様性を生み出す遺伝子制御の機構については、得られている知識はごくわずかである。またこれまで、悪性度に関連する転写プログラムについてはまだほとんど解析はされておらず未知のままである。転写制御プログラムを解明するためには、制御配列情報と網羅的発現プロファイルを統合する計算的アプローチが必須である。これまでに多くのアプローチが開発され、酵母のような下等生物の系に対しての適用はかなりの成功を収めている。しかしヒトのような高等生物のもつ複雑な遺伝子制御システムに対する適用はまだ萌芽的段階にある。

癌細胞における転写制御プログラムを解明するため、本論文は Bayesian network をもちいて制御配列情報と網羅的発現プロファイルデータを統合し、発現プロファイルデータに付随する表現型情報に相関するシス制御モチーフを探索するための方法を確立した。まず始めに個々の遺伝子の発現量と表現型の相関を計算しさらにその値を"メタ発現値 (meta expression value)"として相関するようなシス制御モチーフを Bayesian network により探索した。

はじめに、この方法の実用性を確かめるためにいくつかのヒト細胞の発現プロファイルデータの解析を行った。肝細胞特異的発現においては HNF1 及び HNF4 の結合モチーフ、骨格筋細胞特異的発現においては MEF2 の結合モチーフ、HUVEC における TNF α による発現誘導には NF κ B の結合モチーフが相関していることが示され、既知の報告との一致が見られることからこの方法の実用性が示された。

次に乳癌の組織学的多様性を生み出す転写プログラムに注目し、histological grade に関連付けられるシス制御モチーフの探索を行った。Bayesian Network を用いて発現プロファイルデータ中の全ての遺伝子について G1 (高分化型, 67 サンプル) と G3 (低分化型, 高増殖性, 54 サンプル) の間の発現量の違いに相関するシス制御モチーフを探索し、ELK1、E2F1、NRF1 および NFY の結合モチーフを有意に histological grade に相関する sequence feature として同定した。また実際の依存関係の解析により、これらの sequence feature の存在が G3 サンプル群での遺伝子の upregulation に正に相関していることがわかった。

最後に、より直接的に癌の悪性度を反映している指標としての予後に注目して解析を行った。それぞれの遺伝子に対して生存時間との相関を Cox 回帰モデルを用いて計算し、その値に相関しているようなシス制御モチーフを探索した。解析の結果 histological grade と同様、ELK1、E2F1、NRF1 および NFY の結合モチーフが予後と相関していることが示された。これらの結果を全て考慮に入れると ELK1、E2F1、NRF1 および NFY の結合モチーフが主要な乳癌の悪性度と関連するシス制御モチーフであると考えられる。

本論文は、Bayesian Network によるシス制御配列と発現値データの統合解析方法を確立し、この方法を用いてはじめて乳癌細胞の悪性化に関連するシス制御モチーフの存在を示した。本論文の一部は広子貴俊、笠井真菜、古川洋一、中村祐輔、鈴木稔、菅野純夫、秋山徹との共同研究であるが、論文提出者が主体となって分析及び検証を行ったもので、論文提出者の寄与が十分であると判断する。

従って、博士（理学）の学位を授与できると認める。