

6R-1

ニューラルネットワークによる
遺伝子のスプライシング部位の同定

古谷博史、小倉久和、山本皓二、北添康弘
(高知医科大学)

1. はじめに

真核生物では、多くの遺伝子においてアミノ酸をコードする領域(エクソン)の間にアミノ酸として翻訳されない領域(イントロン)が存在し、核内で直接の転写産物(mRNA前駆体)から除去される。この現象(スプライシング)は、生物の進化や遺伝情報の発現などにおいて重要な意味を持つと考えられているが、その内容については、まだ不明な点が多い。エクソンとイントロンの接合部位には一定の規則性(コンセンサス配列)があるが、それだけでは接合部位の特定は難しい。我々は、エクソンとイントロンの接合部位のまわりの配列を教師信号として与えることにより、ニューラルネットワークによる学習機構を用いて、接合部位を認識するネットワークを生成するプログラムを作成したので、その内容と各種の遺伝子に適用した結果を報告する。また、β-サラセミア(遺伝性の貧血性疾患)患者のβ-グロビン遺伝子に我々の方法を適用し、点突然変異によってもたらされるスプライシングの異常について解析した。

2. 方法

入力層・中間層・出力層の三層から構成されるニューラルネットワークを採用し、入力層から中間層・中間層から出力層の一方のみ信号が伝達するものとした。学習アルゴリズムとしてはバックプロパゲーションを用いた。遺伝子DNAは、A・G・T・Cの四種類の塩基から構成され、イントロンの両端の塩基配列は、GTとAGであることが明らかになっている。このためGTとAGを含む配列を遺伝子データベースから抽出し、それを入力信号とした。入力する塩基配列の長さは、これまでの実験から50塩基対とし、各々の塩基対ごとに四種類の塩基が対応するため入力層全体の長さは200とした。出力層は一素子のみとし、入力の配列がイントロン・エクソン接合部の場合は1を、それ以外の場合は0を出力するものとした。GT側の端とAG側の端で各々のネットワークを作成した。教師用信号として、GT側は接合部の配列 159 とそれ以外の配列 283、AG側は接合部配列 120 とそれ以外の配列 272 を用いた。

3. 結果

遺伝子データベースから教師用配列として用いたもの以外の遺伝子 15 種類について上で求めたニューラルネットワークを適用した。出力信号が 0.5以上のものを接合部配列、0.5未満のものを非接合部配列とみなした。その結果は、GT側について 3187配列(接合部72,非接合部3115)を調べ接合部配列を非接合部とみなした誤りが35%(25/72)、非接合部配列を接合部とみなした誤りが5.5%(171/3115)であった。AG側は、4400配列(接合部72,非接合部4328)について調べ、接合部配列の内誤ったものが 28%(20/72)、非接合部配列の内誤って接合部としたもの 5.7%(248/4328)であった。

次にβ-サラセミア患者のβ-グロビン遺伝子における点突然変異が、スプライシングに及ぼす影響をニューラルネットワークにより計算した結果を示す。サラセミアは、ヘモグロビンを構成するα鎖やβ鎖の合成障害にもとづく遺伝性疾患であり、特にβ-グロブリン遺伝子の異常によるβ-サラセミアが最もよく調べられている。図1にβ-グロビン遺伝子の構造と突然変異の位置をしめした。E1, E2, E3がエクソン、I1, I2がイントロンである。表1に、各々の突然変異によるネットワークの出力値の変化をしめした。これは、実験の解析をよく再現している。

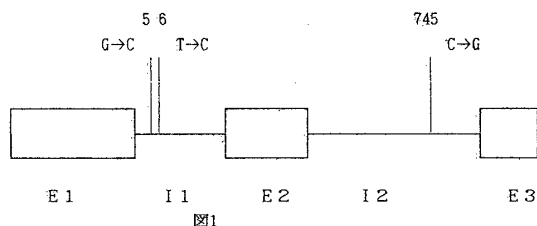


表1

正常	I1-5	I1-6	正常	I2-745
0.83	0.12	0.36	/	0.94