

事例 1

目的： GWAS の SNP 情報より、B 型慢性肝炎(chronic hepatitis B)；CHB と肝細胞癌(HCC)を深層学習により分類するパイロット研究

背景： B 型慢性肝炎(CHB)治療の最終目標は HBs 抗原消失ですが、その一方で肝発癌発症の有無が予後を規定します。またウイルス療法の進歩により、B 型慢性肝疾患からの肝細胞癌(HCC)発症は抑制傾向が見られますが、まだ肝発癌する症例も散見されます。ゲノムワイド関連解析(GWAS)の登場以降、肝疾患領域においてもさまざまな疾患病態についての遺伝子多型解析が急速に拡大しました。その結果 IL28B のように予測マーカーとして臨床応用が急速に進展した SNP がある一方で、進捗のないものも多い。また発見された全 SNP でもその遺伝性のごく一部分しか説明し得ない事例が多数報告されており、GWAS の改良のみならず rare SNP や構造多型、遺伝子間作用、遺伝子環境間作用、epigenetics などへのアプローチの必要性が提唱されています。そこで、最新の深層学習でどの程度分析、推定が可能であるかを調査しました。

入力： 40万座位、80万塩基 ACGT SNP 情報 -(欠損),A,C,G,T: 計5種類

染色体番号	1	1	1	1	1	1	1	1	1	1	1	1
SNP No	SNP1		SNP2		SNP3		SNP4		SNP5		SNP6	
物理距離	565433		1036959		1120431		1733219		4263677		4284919	
ID1	C	C	G	G	-	-	C	T	A	G	A	A
ID2	C	C	G	G	C	C	-	-	A	G	G	G
ID3	C	C	-	-	G	G	G	G	A	A	A	G
ID4	C	C	A	A	C	T	C	T	G	G	G	G
ID5	C	C	G	T	A	G	A	A	A	A	G	G
ID6	C	C	A	A	C	T	C	T	G	G	A	G

出力：CHB/HCC の 2 値予測

平均 5,000 座位/人の欠損含む 40 万座位/80 万塩基情報をそのまま深層学習して、かなり精度よい CHB/HCC 推定モデルを構築することができました。SNP などの遺伝子情報は全ゲノムのデータを用いると Outcome と関連を持たない座位が多く、また一定割合の欠損データが避けられないため、事前に選別して(関連するエクソン領域など)解析するのが普通ですが、深層学習モデルを用いることによりほぼすべての SNP 情報、しかも欠損をかなり含んだままの SNP 情報を深層学習して、正解率、PPV、感度ともかなり良い値を得ることが出来ました。多くの関連性のないデータ、欠損データを含む多量のデータでも有効な推定が出来ることを示すことができ、全ゲノム解析でも深層学習モデルを使い、従来手法よりより良い結果を出す可能性を示すことができました。