

審査の結果の要旨

氏名 内藤 龍彦

本研究では、深層学習を応用した新規の HLA imputation 法、DEEP*HLA を開発し、頑健な性能評価により、その有用性を評価した。さらに DEEP*HLA を Parkinson 病 (Parkinson's disease, PD) と (type I diabetes, T1D) のゲノムデータに適用し主要組織適合性複合体 (major histocompatibility complex, MHC) 領域の trans-ethnic fine-mapping を行うことで、両疾患それぞれにおいて発症リスクに関わるヒト白血球抗原 (human leukocyte antigen, HLA) 遺伝子変異を新規に同定した。

本研究で開発した DEEP*HLA は、一塩基変異 (single nucleotide variant, SNV) ジェノタイプを入力データとし、複数の HLA 遺伝子のアレルの予測値を同時に出力する、マルチタスク畳み込みニューラルネットワークを利用したモデルである。従って、SNV ジェノタイプと HLA アレルの組み合わせである参照パネルを予め学習することで、SNV ジェノタイプのみから HLA アレルを予測することができる。精度評価は、日本人集団用の参照パネル、欧米人集団用の参照パネルを用いて交差検証法により行った。また、日本人集団用の参照パネルを用いて独立サンプルの imputation を行った場合の精度検証も行った。結果、いずれの検証においても DEEP*HLA は既存手法である SNP2HLA や HIBAG と比較して、特に頻度が低い希少アレルにおいて精度の上昇を認めた。計算負荷の評価も行ったところ、総実行時間の点ではサンプル数が増えるに従い既存手法と比べて有利であり、またメモリ負荷の点からもバイオバンクレベルのサンプル数に対しても実行可能であった。

次に DEEP*HLA を用いて T1D の MHC 領域における trans-ethnic fine-mapping を行った。BioBank Japan、UK Biobank の両コホートの SNV ジェノタイプデータにそれぞれ日本人集団用参照パネル、欧米人集団用参照パネルで学習した DEEP*HLA を適用して HLA imputation を行った。両集団の imputation 結果を統合の上、trans-ethnic fine-mapping を行った。結果、HLA クラス II 遺伝子の HLA-DRβ1 pos. 71 に最も強い関連を認めた。さらに、同領域内では、HLA-DRβ1 pos. 71 に加えて、HLA-DRβ1 pos. 74、HLA-DQβ1 pos. 30, 70, 185 にも独立した関連を認めた。また、HLA クラス I 遺伝子として、HLA-A pos. 62, HLA-B*54:01 にも独立した関連を認めた。それらは集団間で共通したリスク効果を持っていた。検出されたリスク関連アミノ酸位置の多くは、HLA 分子の構造においてペプチド結合溝に位置しており、抗原提示能への機能的影響を介して T1D の発症に関わると考えられた。

さらに本研究では、Parkinson 病 (Parkinson's disease, PD) についても MHC 領域における trans-ethnic fine-mapping を行った。UK Biobank の SNV ジェノタイプデータに対して DEEP*HLA を適用し fine-mapping を行い、さらに他の複数の先行研究の要約統計量に対し

て Z 値 imputation を適用した結果と合わせてメタアナリシスを行い trans-ethnic fine-mapping を行った。結果、HLA クラス II 遺伝子の HLA-DRβ1 pos. 13 に最も強い関連を認め、それ以外に HLA-B pos. 69 にも独立した関連を認めた。HLA 分子が PD の発症にどのように関わるかにおいては先行研究では、HLA 分子が PD の脳組織内蓄積物質である α-シヌクレインの断片をエピトープとして認識し、抗原提示・免疫応答を引き起こす可能性が報告されていた。従って、本研究では *in silico* の HLA-ペプチド結合親和性予測ツールである NetMHCpan II を用いて、*HLA-DRB1* アレルと α-シヌクレイン・エピトープの結合親和性を網羅的に評価した。結果、HLA-DRβ1 pos. 13 においてヒスチジン（PD の発症に保護的な効果をもつアレル）を有する *HLA-DRB1* アレルはその他のアレルに比べて弱い結合親和性を持ち、逆にアルギニン（PD の発症にリスク効果をもつアレル）を有する *HLA-DRB1* アレルは、他のアレルに対して強い結合親和性を呈した。このことから、HLA-DRβ1 pos. 13 は、α-シヌクレインエピトープへの結合親和性の変化を介して PD の発症に関わっている可能性が考えられた。

本研究で開発した DEEP*HLA は他の疾患にも適用可能であり、今後様々な疾患において発症リスクに関わる HLA 遺伝子変異を同定することに寄与するだろう。1 型糖尿病においては、異なった集団間でリスク HLA 遺伝子変異が共有されていないことが遺伝学上の問題であったが、本研究により集団間で共通したリスクを持つ HLA 遺伝子変異のセットを初めて同定した。Parkinson 病においては、近年その病態における神経炎症が注目を浴びており、遺伝的リスクと HLA 分子の関わりに仮説を与えることができた。さらに、本研究は集団ゲノムデータに対する深層学習の応用という課題に対する成功例であるという広い意義をもつとともに、深層学習の新たな可能性を示すものである。

よって本論文は博士（医学）の学位請求論文として合格と認められる。