

機械学習で日本人集団の白血球の血液型を解明

大阪大学 大学院医学系研究科 教授

岡田 随象

(お問い合わせ先) E-MAIL: yokada@sg.med.osaka-u.ac.jp



研究の背景

ヒト白血球の血液型はHLA遺伝子のゲノム配列の個人差で決定され、様々な疾患に対して発症リスクの高い遺伝子型があることが知られています。しかし、HLA遺伝子には複数の種類があるだけでなく、各遺伝子にも数十以上の遺伝子型があるため、白血球の血液型の組み合わせは膨大な数になります。そのため、その全体像は長らく不明でした。HLA遺伝子の構造が複雑で、配列の解読に多額の費用と専門技術が必要なことも、解明を遅らせる原因になっていました。

私たちは以前、7個の主要なHLA遺伝子を対象にして、日本人集団における白血球の血液型の構成を決定し、日本人集団に特異的な白血球の血液型が存在することを報告しました (Okada Y et al. *Nat Genet* 2015)。今回、次世代シーケンス技術や機械学習を活用して、白血球の血液型とHLA遺伝子の構成や病気との関連を詳細に検討しました (Hirata J et al. *Nat Genet* 2019)。

研究の成果

まず、次世代シーケンス技術を駆使し、日本人集団1120名を対象に33のHLA遺伝子における700種類以上のゲノム配列を詳細に決定しました (図1)。さらに、HLA遺伝子ゲノム配列情報の膨大な組み合わせを効率的に分類する目的で、機械学習手法のひとつであるt-SNEを適用しました。その結果、日本人集団の白血球の血液型は11パターンの組み合わせに分類可能なが明らかになりました (図2)。これは、複雑なヒトゲ

ノム情報の解釈を、機械学習により実現した先進的な成功例です。

次に、日本人集団17万人における大規模ゲノムワイド関連解析 (GWAS) のゲノムデータから、白血球の血液型をコンピューター上で高精度に推定することができました。また、病気 (免疫疾患・生活習慣病・悪性腫瘍) や臨床検査値 (身長・肥満、血液検査値、生理検査結果) を含む100を超える多彩な表現型との関連を網羅的に調べるフェノムワイド関連解析 (PheWAS) を実施したところ、半数を超える52の表現型において、白血球の血液型が関与していることが明らかになりました。

今後の展望

本研究により、日本人集団における白血球の血液型の全容が明らかになり、この成果が移植医療や個別化医療へ応用されることが期待されます。また、今回、機械学習による白血球の血液型の分類に成功したことから、今後も複雑なヒトゲノム情報に対する機械学習の応用研究を加速していきたいと考えています。

関連する科研費

2015-2017年度 若手研究(A)「HLA imputation法を用いた自己免疫疾患のバイオマーカーの同定」

2015-2019年度 新学術領域研究(研究領域提案型)「遺伝統計学とビッグデータの邂逅がもたらす新たながんゲノム創薬」

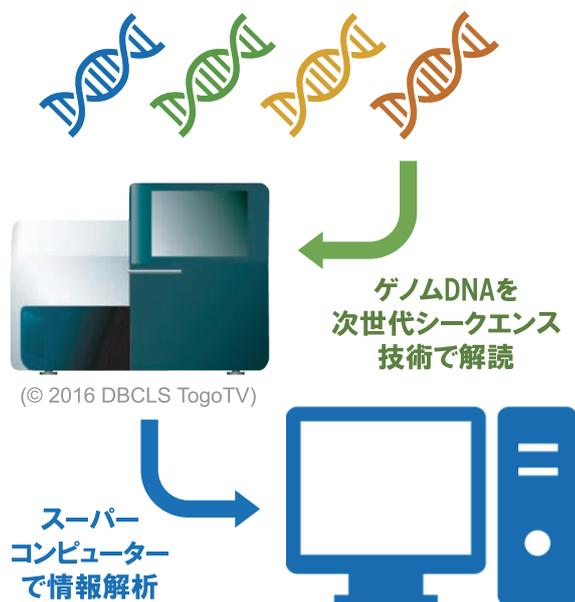


図1 次世代シーケンス技術によるHLA遺伝子配列の解読

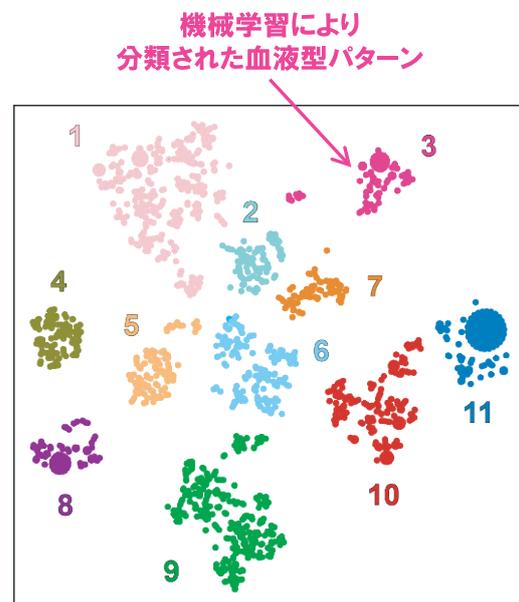


図2 機械学習による日本人集団の白血球の血液型の分類