



The 65th ASH Annual Meeting Abstracts

POSTER ABSTRACTS

401. BLOOD TRANSFUSION

Optimization of RH Genotyping from Whole Exon Sequencing Data By Machine Learning

Ti-Cheng Chang¹, Jing Yu, MSc², Zhaoming Wang, PhD³, Jane S Hankins, MD MS⁴, Mitchell Weiss, MD PhD¹, Gang Wu, PhD¹, Connie M. Westhoff, PhD⁵, Sunitha Vege⁶, Stella Chou, MD⁷, Yan Zheng, MD PhD⁸

¹ St. Jude Children's Research Hospital, Memphis, TN

² Pathology, St. Jude Children's Research Hospital, Memphis

³ Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN

⁴ Departments of Hematology and Global Pediatric Medicine, St. Jude Children's Research Hospital, Memphis, TN

⁵ New York Blood, New York, NY

⁶ New York Blood Center, Long Island City, NY

⁷ The Children's Hospital of Philadelphia, Philadelphia

⁸ Pathology, St. Jude Children's Research Hospital, Memphis, TN

Introduction

Repetitive exposures to donor's red blood cell (RBC) antigens result in alloimmunization. This problem occurs at especially high rates in patients with sickle cell disease (SCD), in part because Black individuals frequently harbor numerous genetic variations in *RHD* and *RHCE* genes. The *RH* genetic variations can result in loss of common epitopes or expression of neo-epitopes, predisposing patients with Rh variants to Rh alloimmunization. Since these variants are not distinguishable by standard serological typing, *RH* genotyping to facilitate genotype-matched transfusion has become necessary. *RH* genotyping by DNA sequencing-based approaches is complicated by highly homologous sequences shared by *RHD* and *RHCE*. We previously developed the RHtyper, an automated system to ascertain complex *RH* genotypes of Black individuals from standard whole-genome sequencing (WGS) data (Chang, T.C., et al., *Blood Advances*, 2020). In a validation cohort of 57 SCD patients, RHtyper achieved 100% accuracy for *RHD* and 98.2% accuracy for *RHCE* genotypes compared to the genotypes obtained from single nucleotide variant (SNV)-based BeadChip and targeted molecular assays, which represent the current standard. Since whole-exome sequencing (WES) is more cost-effective and widely available than WGS, we optimized RHtyper for analyzing WES data by incorporating a machine learning approach that minimized errors in genotyping caused by non-uniform sequencing coverage and misalignment of sequencing reads in WES data.

Methods

WES and WGS data from 396 SCD patients enrolled in the Sickle Cell Clinical Research and Intervention Program (SCCRIP) study and 3030 childhood cancer survivors enrolled in St. Jude Lifetime Cohort Study (SJLIFE, with 15.2% Blacks) were included. RHtyper was optimized for WES data by using machine learning to improve prediction accuracy for *RHD* zygosity/hybrid alleles, *RHCE**C/*RHCE**c alleles, zygosity of *RHD* c.1136C>T and *RHCE* c.48G>C. Specifically, hundreds to thousands of informative features specific for each of the allele/SNVs were selected by Boruta algorithm and incorporated into a prediction model using XGboost. The model was trained by 75% of the SCCRIP data, followed by validation using the remaining 25% of SCCRIP data. WGS-based genotypes served as references. The optimized RHtyper was further validated in the SJLIFE cohort.

Results

Genotyping *RH* using WES data with the original RHtyper was less accurate. For 396 patients from the SCCRIP study, the concordance between WGS and WES data was 90.2% for *RHD* and 96.3% for *RHCE*. It was particularly problematic in determining 1) *RHD* zygosity and hybrid alleles, 2) *RHCE**C vs. *RHCE**c alleles, 3) *RHD* c.1136C>T zygosity, 4) *RHCE* c.48G>C zygosity. We optimized RHtyper by incorporating machine learning specific for those affected alleles/SNVs, and substantially improved the concordance between WGS- and WES-based genotypes to 97.2% for *RHD* and 98.2% for *RHCE*. We further validated the optimized RHtyper using 3030 patients from the SJLIFE cohort and achieved concordance of 96.3% for *RHD*, 94.6% for *RHCE*. The predicted C antigen frequency per WES data was 59.0% for Whites and 24.2% for Blacks for the SJLIFE cohort, similar to previously reported racial distributions. In addition, for 1036 patients with blood type records, the predicted D serologic types using WES data were 99.8% consistent with clinical serology results.

Conclusion

We improved RHtyper for WES by integrating machine learning, which allowed for incorporation of information from a large number of diverse informative features, enabling more accurate predication.

Disclosures Weiss: *Vertex Pharmaceuticals:* Consultancy; *Cellarity:* Current equity holder in private company; *Novartis Inc.:* Consultancy; *GlaxoSmithKline:* Consultancy; *bluebird bio:* Consultancy.

<https://doi.org/10.1182/blood-2023-182089>