Automated typing of red blood cell and platelet antigens from whole exome sequences

William J. Lane,[1,2] *Sunitha Vege*,[3] *Helen H. Mah*,[1] *Christine Lomas-Francis*,[3] *Maria Aguad*,[1] *Robin Smeland-Wagman*,[1] *Christopher Koch*,[4] *Jacqueline M. Killian*,[5] *Cubby L. Gardner*,[5] *Mauricio De Castro*,[6] *Matthew S. Lebo*,[1,2,4,7] *Richard M. Kaufman*,[1,2] *Robert C. Green*,[2,4,8,9] *and Connie M. Westhoff*,[3] *for the MilSeq Project*

BACKGROUND: Genotyping has expanded the number red blood cell (RBC) and platelet (PLT) antigens that can readily be typed, but often represents an additional testing cost. The analysis of existing genomic data offers a cost-effective approach. We recently developed automated software (bloodTyper) for determination of RBC and PLT antigens from whole genome sequencing. Here we extend the algorithm to whole exome sequencing (WES).

STUDY DESIGN AND METHODS: Whole exome sequencing was performed on samples from 75 individuals. WES-based bloodTyper RBC and PLT typing was compared to conventional polymerase chain reaction (PCR) *RHD* zygosity testing and serologic and single-nucleotide polymorphism (SNP) typing for 38 RBC antigens in 12 systems (17 serologic and 35 SNPs) and 22 PLT antigens (22 SNPs). Samples from the first 20 individuals were used to modify bloodTyper to interpret WES followed by blinded typing of 55 samples.

RESULTS: Over the first 20 samples, discordances were noted for C, M, and N antigens, which were due to WES-specific biases. After modification, bloodTyper was 100% accurate on blinded evaluation of the last 55 samples and outperformed both serologic (99.67% accurate) and SNP typing (99.97% accurate) reflected by two Fy^b and one N serologic typing errors and one undetected SNP encoding a Jk_{null} phenotype. *RHD* zygosity testing by bloodTyper was 100% concordant with a combination of hybrid Rhesus box PCR and PCR–restriction fragment length polymorphism for all samples.

CONCLUSION: The automated bloodTyper software was modified for WES biases to allow for accurate RBC and PLT antigen typing. Such analysis could become a routing part of future WES efforts.

In recent years genotyping has greatly expanded the number of red blood cell (RBC) and human platelet (PLT) antigens that can readily be typed for both donors and recipients. RBC and PLT genotyping has

ABBREVIATIONS: NGS = next generation sequencing; SNP = single-nucleotide polymorphism; WES = whole exome sequencing; WGS = whole genome sequencing.

From the [1]Department of Pathology and [9]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, the [2]Harvard Medical School, the [4]Partners Personalized Medicine, the [7]Laboratory for Molecular Medicine, the [8]Broad Institute of MIT and Harvard, Boston, Massachusetts; the [3]New York Blood Center, New York, New York; the [5]United States Air Force, San Antonio, Texas; and the [6]United States Air Force, Biloxi, Mississippi.

Address reprint requests to: William J. Lane, MD, PhD, Pathology Department, Brigham and Women's Hospital, Hale Building for Transformative Medicine, Room 8002L, 60 Fenwood Road, Boston, MA 02115; e-mail: wlane@bwh.harvard.edu.

Additional members of the MilSeq Project are listed in the acknowledgments.

RCG is supported by grant funding from NIH and the Department of Defense. This material is based on research sponsored by Air Force Medical Support Agency, Research and Acquisition Directorate AFMSA/SG5 under cooperative agreement number FA8650-17-2-6704. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Medical Support Agency, Research and Acquisition Directorate AFMSA/SG5 or the U.S. Government. The voluntary, fully informed consent of the subjects used in this research was obtained as required by 32 CFR 219 and AFI 40-402, Protection of Human Subjects in Biomedical and Behavioral Research.

Received for publication May 9, 2019; revision received July 5, 2019, and accepted July 8, 2019.

doi:10.1111/trf.15473

TRANSFUSION 2019;9999;1–11

improved transfusion medicine practice by facilitating complex reference laboratory workups, identifying antigen-negative RBC and PLT donors, and expediting selection of reagent RBCs. RBC genotyping has also increased the availability of RBC units for routine extended antigen matching in chronically transfused patients and has enabled selection of antigen-matched units when compatibility cannot be demonstrated due to warm autoantibodies or drug interference in pretransfusion testing. Although single-nucleotide polymorphism (SNP)-based assays currently dominate the RBC and PLT genotyping market, next generation sequencing (NGS) approaches including targeted NGS, whole exome sequencing (WES), and whole genome sequencing (WGS) are being actively pursued by several groups.[1–23]

Conventional genotyping assays typically target a limited number of SNPs, which allows for the development of predefined rules for automated interpretation. Complex samples, especially those with changes not present on SNP assays, require further analysis by Sanger sequencing or an alternative approach. Many NGS blood typing efforts have simply expanded on this paradigm and only target a limited number of genetic regions with analysis requiring subject matter experts. However, the potential of NGS is much broader, since it can be used to detect essentially any nucleotide change or structural variation. Nevertheless, the volume and complexity of the NGS data make it difficult to evaluate the more than 2000 allelic variants across 46 RBC and six PLT antigen–associated genes. Although manual analysis of all antigen encoding genes is possible,[6] rapid and accurate automated interpretation of NGS data is desirable and necessary to scale for routine use. Toward this goal, the development of automated interpretative software has been described to analyze whole genomes for all molecularly understood RBC and PLT antigens.[5,10,17] We recently demonstrated the feasibility of using software (bloodTyper) to automatically type for all genetically defined RBC and PLT antigens from 330 whole genomes (WGS) with more than 99% accuracy when compared to conventional serologic and SNP typing for 38 RBC and 22 PLT antigens.[17] Compared to WGS, WES is less expensive and faster,[24] but we reasoned that differences in method might require modifications to bloodTyper. For example, with WES, only exons and intronic regions near the exon boundaries are sequenced; thus relevant intronic regions might be missing. WES enrichment can also introduce capture biases leading to unequal sequence read depth of coverage that could affect copy number calculations and thereby impact the accuracy of detecting some antigens (i.e., C antigen).

Here, we present evidence that bloodTyper software can be used for automated RBC and PLT antigen typing from WES data. We describe specific modifications made to bloodTyper for WES data analysis including correcting for capture biases to more accurately calculate gene copy number. WES outperformed conventional hybrid Rhesus box polymerase chain reaction (PCR) and PCR–restriction fragment length polymorphism (RFLP) assays for *RHD* zygosity determination. WES-based typing was more accurate than serology for extended antigen typing and was able to detect nucleotide changes not targeted by SNP assay.

MATERIALS AND METHODS

Study overview

Between October 2017 and August 2018, samples for RBC and DNA isolation were collected from 75 individuals undergoing WES as part of the MilSeq Project: Enabling Personalized Medicine through Exome Sequencing in the US Air Force (Clinical Trials Gov Number NCT03276637; see supplemental tables for self-identified ethnicities). DNA and RBC samples were frozen for follow-up testing if needed. In this substudy, we evaluated the performance of our automated RBC and PLT typing algorithm (bloodTyper) on WES data compared to serologic and SNP based typing (Fig. 1). The first 20 samples were analyzed by bloodTyper and compared to serologic testing, SNP typing, and PCR-based *RHD* zygosity and the algorithm was modified to address WES-specific data challenges. The last 55 participant samples were then analyzed as blinded samples by bloodTyper.

Serologic typing

Blood samples were collected in EDTA, and conventional RBC serologic antigen typing was performed according to standard tube typing methods.[25] Commercially available serologic typing reagents were used to type for the A, B, D, C, E, c, e, K, k, $Fy^{a/b}$, $Jk^{a/b}$, M, N, S, s (Samples 1-47—Bio-Rad for all except Ortho Clinical Diagnostics for Fy^{b} and Immucor for $Jk^{a/b}$; Samples 48-75—Immucor for all except Ortho Clinical Diagnostics for D and Quotient for Fy^{b}).

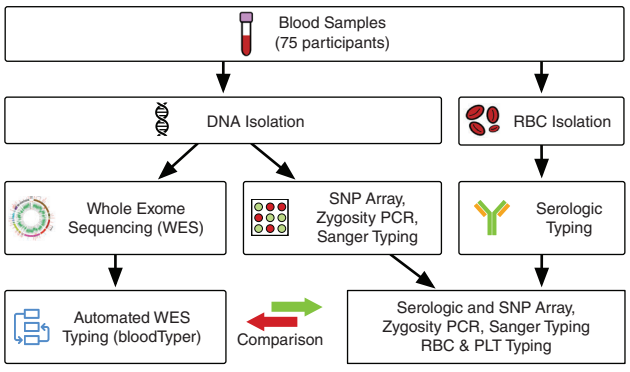

Fig. 1. WES-based RBC and PLT antigen typing study overview. DNA and RBC samples were collected from 75 individuals and used for conventional serologic and DNA-based SNP array typing of RBC and PLT antigens. The serologic, SNP array, and WES-based typing results were compared and bloodTyper modified for WES. [Color figure can be viewed at wileyonlinelibrary.com]

SNP typing

DNA was isolated from WBCs by standard methods (Qiagen). The PreciseType HEA (human erythrocyte antigen) SNP-typing BeadChip array (Immucor) was performed according to the manufacturer's instructions to type for M, N, S, s, U, C, c, E, e, V, VS, $Lu^{a/b}$, K, k, $Kp^{a/b}$, $Js^{a/b}$, $Fy^{a/b}$, $Jk^{a/b}$, $Di^{a/b}$, Sc1, Sc2, $Do^{a/b}$, Hy, Jo^{a}, $Co^{a/b}$, and $LW^{a/b}$. The HPA (PLT antigen) BeadChip RUO array was used to type for the following PLT antigens: HPA-1a/b, 2a/b, 3a/b, 4a/b, 5a/b, 6a/b, 7a/b, 8a/b, 9a/b, 11a/b, and 15a/b (Immucor). To confirm results of samples found to be *RHD*DAU0* by WES, an *RHD* Exon 8 PCR-RFLP NlaIII was performed for c.1136C>T and visualized by agarose gel electrophoresis and ethidium bromide staining. *RHCE* BeadChip (Immucor) assay was performed to confirm a C^{W}+ sample found by WES.

Sanger sequencing

Sanger sequencing of genomic DNA was performed according standard procedures. Sequencing was initiated for samples in which WES indicated nucleotide changes not present on the SNP-typing platform. *RHD* Exons 5 (c.733C for *RHD*DUC2*) and 9 (c.1195A for *RHD*weak D Type 45*), *SLC14A1* Exon 5 (c.342-1A for JK_{null}), and *ITGA2B* Exon 26 (c.2614A for HPA-27bw+) were amplified using published primers,[26–28] sequenced by Genewiz (Warren, NJ), and analyzed with ClustalX.[29]

PCR-based *RHD* zygosity testing

RHD zygosity testing was performed using two different PCR assays; detection of the hybrid Rhesus box by allele-specific PCR and by PCR-RFLP with PstI according to published methods.[30,31] Briefly, allele-specific PCR targets a product of 1507 bp within the hybrid box sequence with the presence indicating at least one deleted *RHD* allele and absence of a product indicating *RHD* homozygote. The PCR-RFLP pattern differentiates *RHD* homozygote, hemizygote, and homozygote *RHD* deletion.

WES-based bloodTyper analysis

The custom interpretive blood typing software (bloodTyper) utilizes a curated antigen allele database (http://bloodantigens.com) to analyze NGS data.[17] Variant calls for RBC and PLT antigen–associated genes and promoter regions were made using Genomic Analysis Tool Kit (GATK) v3.7-0-gcfedb67 (UnifiedGenotyper and EMIT_ALL_SITES output mode) and saved as a variant calling format file (.vcf).[32] Sequencing coverage was extracted from the alignment file using BEDTools v2.17.0.[33] The Integrative Genomics Viewer[34] was used as needed to verify coverage and sequence identity. bloodTyper used a 4× nucleotide calling cutoff when detecting antigen alleles. Toward the end of this study, the KANNO blood group antigen was identified to be located on the *PRNP* gene.[35] After adding the KANNO genetic change to our allele database all 75 samples were retrospectively typed for KANNO. bloodTyper also evaluated for the most common HbS sickle cell disease *HBB* gene variant c.20A>T[36] (see supplemental information for more details about WES sequencing workflow and copy number analysis method).

RESULTS

WES sequencing coverage

One limitation of WES is that nonexonic regions (e.g., upstream promoter and introns) are only sequenced if they are near exon boundaries. We estimated the effect of this limitation on RBC and PLT antigen typing by reviewing the antigen allele tables[37–39] for nonexonic nucleotide changes that impact antigen phenotype predictions. Other than the recently published *transcription factor binding regions deep within A4GALT1 Intron 1 and upstream of XG that control the expression the P1 and* Xg^{a} *antigens,*[40–44] all other nonexonic changes associated with antigenic variation are located within 13 bp of an exon boundary. In fact, analysis of WES data from all 75 participants showed a similar depth of coverage pattern to previous WGS efforts,[6] with overall adequate sequencing coverage of all exons, but with some low-sequence coverage regions in nonantigen allele nucleotide positions within *CR1*, *C4A*, *C4B*, and several other genes (Fig. 2A). Other than the nucleotide positions that control the expression of the P1 and Xg^{a} antigens, there was adequate sequencing coverage (>4×) for all other allele nucleotide positions with an average sequence depth of coverage of 159× for exons and 125× for nonexon nucleotide changes (Fig. 2B). Of note, the very common silencing *FY*B* GATA mutation, c.–67T>C, found in those of African ancestry is within the exonic untranslated region of *ACKR1* and this position is present in the WES data with an average 121× depth of coverage.

WES RBC and PLT antigen typing accuracy

Although all antigens with a known genetic basis and WES sequencing were interpreted by bloodTyper, it was only possible to evaluate the actual performance for antigens typed by confirmatory serologic and/or SNP-based typing (see supplemental tables for full typing results). As shown in Fig. 3, in the first 20 WES samples there were eight discordances between WES-based bloodTyper analysis when compared with serologic and SNP typing: C (n = 5), M (n = 1), N (n = 1), and Fy^{b} (n = 1). Repeat serologic typing showed the Fy^{b} discordance was due to serologic typing error. The cause for the remaining discordances were due to WES-specific issues and modifications were made to bloodTyper to address them (see following sections for more details).

The remaining 55 samples were analyzed with the modified WES-based algorithm, blinded to the information

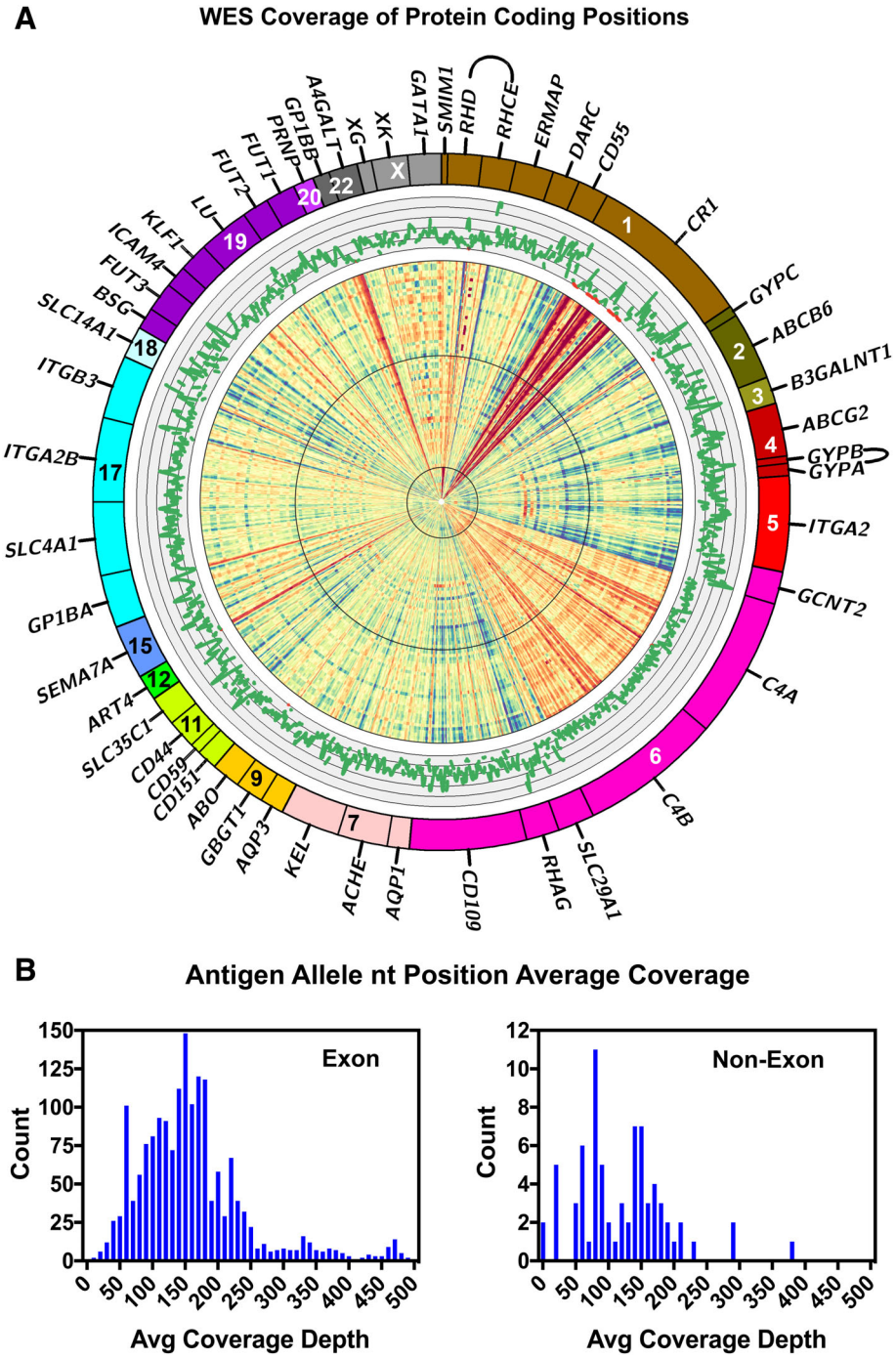

Fig. 2. WES sequence read depth of coverage. (A) Circos plot[53] showing the read depth of coverage for protein coding DNA positions. The mean coverage for all 75 samples shown as a scatter plot with the grid lines represent intervals of 100× depth of coverage and dots colored green at least 10× and red less than 10×. The central area contains stacked heat maps showing the depth of coverage for each of the 75 samples ordered by *RHD* zygosity with 2× outer ring, 1× middle ring, and 0× center ring. (B) Frequency plot of the mean sequence read depth of coverage for exon (protein coding positions and untranslated regions) and nonexonic (intronic and upstream promoter) positions found in the antigen allele lists. Note: For *RHD* positions the mean only included samples with 2× *RHD* zygosity. nt = nucleotide. [Color figure can be viewed at wileyonlinelibrary.com]

generated by the other antigen typing techniques for 38 RBC and 22 PLT antigens. There were four discordances, all of which underwent repeat serologic typing from frozen stored RBC samples. Two Fy^b and one N typing discordance were shown to be serologic typing error. Additionally, a discordance was noted between a serologic Jk(b−) phenotype and

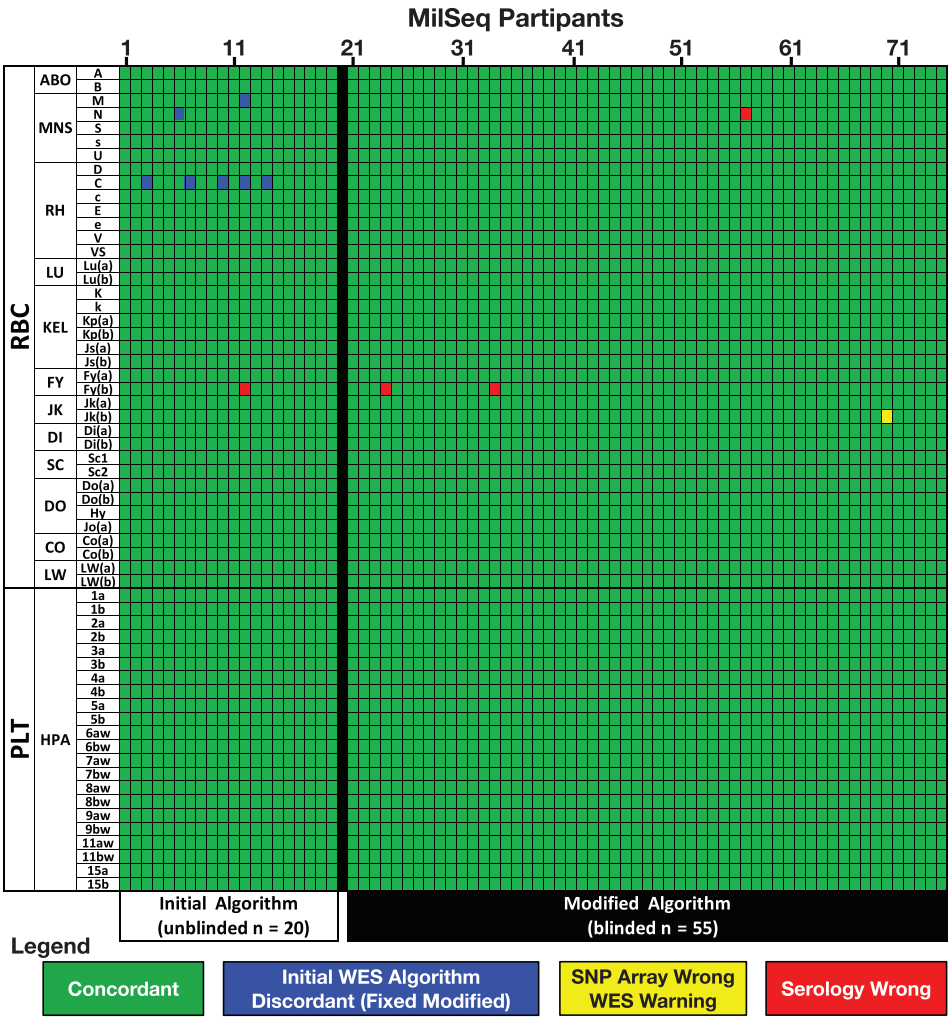

Fig. 3. WES, serology, and SNP typing concordance. Results of automated bloodTyper WES-based RBC and PLT antigen typing compared to conventional serologic and DNA-based SNP typing. Concordance for 75 MilSeq samples for 59 (37 RBC and 22 PLT) antigens. [Color figure can be viewed at wileyonlinelibrary.com]

SNP-predicted Jk(b+) in one sample; WES detected a heterozygous *SLC14A1* nucleotide change c.342−1G>A, which encodes a Jk_{null} phenotype (confirmed by Sanger sequencing). The c.342−1A change has been reported on both *JK*A* and *JK*B* backgrounds and even if intronic sequence reads were present the approximately 5300-bp distance between the relevant nucleotide positions would likely preclude short read NGS-based cis or trans phasing of the heterozygous c.342−1A with the *JK*A* and *JK*B* nucleotide changes. Thus, bloodTyper gave a warning that a Jk_{null} allele was present and that the sample would require further investigation to determine Jk^a and Jk^b antigens.

The modified bloodTyper algorithm for WES typing accuracy was 100% (2088 of 2088 individual antigen typings from 55 participants). Serologic typing accuracy was 99.67% (932 of 935 individual antigen typings from 55 participants), and SNP typing was 99.97% (3499 of 3500 individual antigen typings from 55 participants).

WES *RHD* zygosity modifications

Whole exome sequencing depth of coverage for *RHD* was more variable than WGS-based analysis,[17] especially for *RHD* Exons 4, 7, 9, and 10 (Fig. S1, available as supporting information in the online version of this paper). However, it was possible to calculate copy number correction factors for these exons using homozygous *RHD* samples as a baseline (Fig. S2B). As shown in Fig. 4A, the correction factors improved the copy number agreement across the *RHD* exons for all *RHD* zygosity states: 2× (homozygous), 1× (hemizygous), and 0× (null). Sequence read depth–based WES zygosity determination using both uncorrected and corrected exon copy numbers, agreed with a combination of conventional hybrid box PCR and PCR-RFLP based *RHD* zygosity testing with 29 *RHD* homozygous, 36 hemizygous, and 10 gene deletion samples (Fig. 4A). One sample from a Hispanic individual (#41) was consistent with a homozygote by hybrid box assay, but hemizygote by PCR-RFLP and

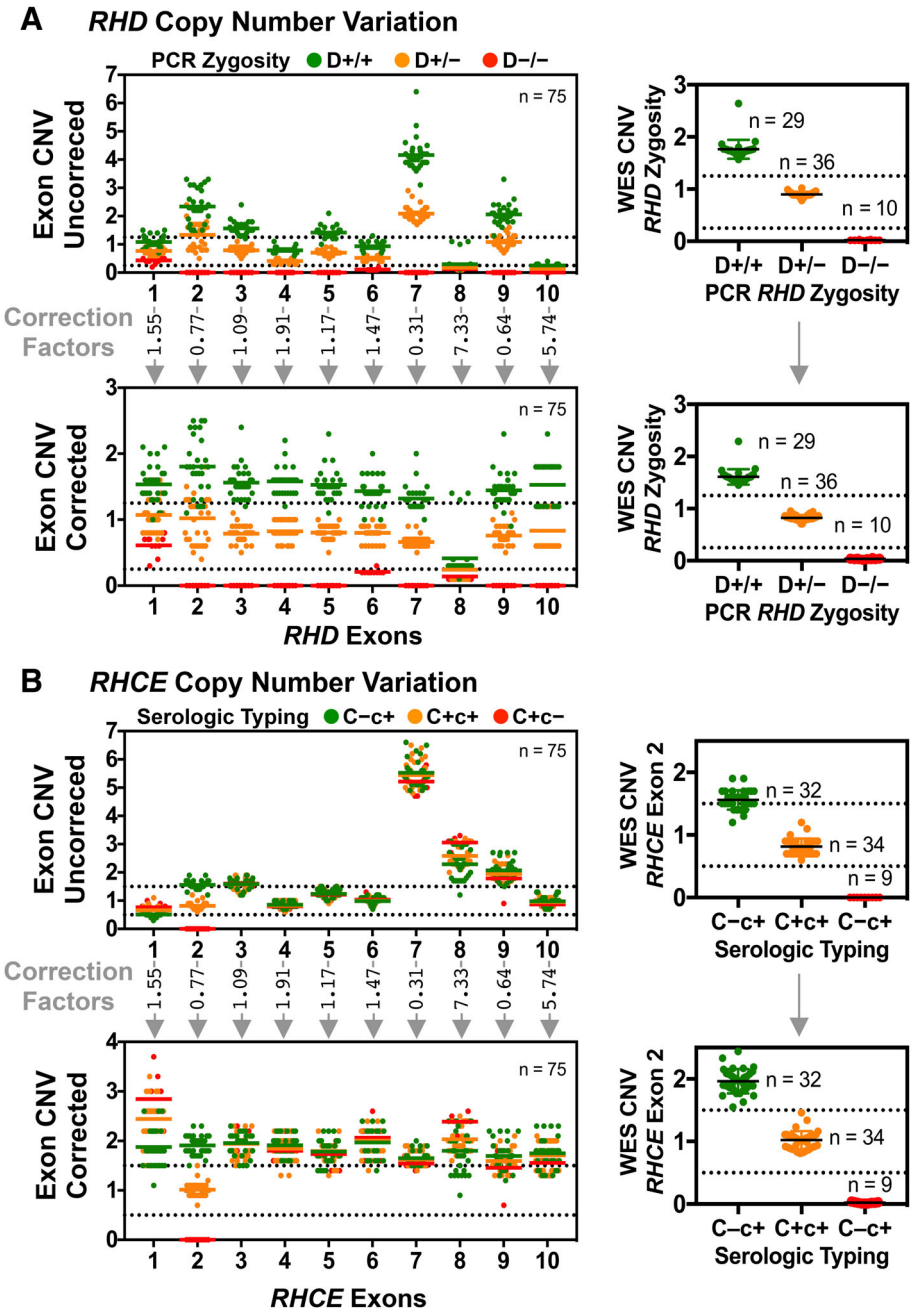

Fig. 4. bloodTyper Rh modifications for WES. Data shown for all 75 participants before and after copy number correction. (A, left) Sequence read depth–based copy numbers for *RHD* Exons 1 to 10. (A, right) Performance of copy number *RHD* zygosity, with dotted line showing cutoff between *RHD* homozygous (D+/+), hemizygous (D+/−), and null (D−/−). Cutoffs were determined using the first 20 WES samples compared to PCR-based zygosity. (B, left) Sequence read depth–based copy numbers for *RHCE* Exons 1 to 10. (B, right) Performance of copy number–based C antigen calling, with dotted line showing cutoff between C− and C+ calls. Cutoffs were determined using the first 20 WES samples compared to C/c serologic typing. [Color figure can be viewed at wileyonlinelibrary.com]

WES. Two samples from individuals of African ancestry (#11 and #40) indicated hemizygote by PCR-RFLP but homozygote by hybrid box assay and WES.

Analysis of Exon 8 is difficult, since the human reference genome Exon 8 contains c.1136T found in *RHD*DAU0*, which causes Exon 8 sequencing from non-DAU0 individuals to incorrectly align.[17] In the first 20 samples, there was one (#11) that without correction appeared to have 2× aligned reads over *RHD* Exon 8, indicative of proper Exon 8 alignment suggesting a homozygous *RHD*DAU0* individual (confirmed by PCR-RFLP). In the last 55, there were two additional samples (#29 and #40) with similar levels of 2× Exon 8 alignment,

but these were heterozygous *RHD*DAU0* individuals by PCR-RFLP (c.1136C/T). As such, it appears that WES can detect the presence of *RHD*DAU0*, but not its zygosity.

WES C antigen modifications

Determination of the C antigen requires special consideration when typing from NGS data because the C antigen is the result of an ancestral 4-kb gene conversion of *RHD* Exon 2 and surrounding intronic regions into *RHCE*.[45] As such *RHCE*C* Exon 2 sequence reads misalign to *RHD* Exon 2.[6,11,13,17,18] For WGS analysis, a copy number approach using sequence depth of coverage was useful to detect misalignment;[17] however, for WES data this approach did not immediately work (Figs. 3 and 4B).

In the first 20 samples, serologic typing identified 10 samples with the C−c+ phenotype, which based on the above pattern should have 2× coverage across *RHCE* including Exon 2. However, over these 10 samples the *RHCE* depth of coverage was more variable than our previous WGS-based analysis[17] especially Exons 4, 6, 7, and 10 (Fig. S1), which affects the baseline read depth used in the Exon 2 copy number calculation. To overcome this, *RHCE* correction factors were calculated to normalize the copy number for each exon (Fig. S2). Without exon copy number correction bloodTyper correctly called the C/c status in 70% of the first 20 samples (14/20), but with correction the C/c antigens were correctly typed in 100% of the last 55 samples (55/55).

WES M and N antigen modifications

M and N antigens require special consideration in NGS-based typing since the sequence alignment efficiency is low. This was shown to be true especially for the M antigen in a previous WGS study with 15× coverage that required a lower 2× calling cutoff.[17] However, despite the 91× average WES depth of coverage for the M/N antigen nucleotide positions (c.59, c.71, and c.72), in the first 20 samples the alignment efficiency of M antigen reads were still very low (averages of 12× for M+N− and 4× for M+N+), but N reads were more than adequate (averages of 195× for M−N+ and 85× for M+N+). One M+N+ serologic and SNP-typed sample (#12) had a false M− antigen typing by WES (M reads of 0× c.59, 0× c.71, and 2× c.72). Some of this low level of alignment could potentially be explained by the fact that the reference genome encodes the N antigen.

However, as the MNS blood group system consists of three highly similar genes (*GYPA*, *GYPB*, and *GYPE*), misalignment to homologous genes might also account for the low M sequence read alignment. Analysis of *GYPB* and *GYPE* for misaligned M-specific *GYPA* sequences showed that as the M dosage (copy number) increased, the *GYPA* Exon 2 coverage decreased, and *GYPE* Exon 2 coverage increased (Fig. 5A). This suggested that Exon 2 M allele reads were misaligning to *GYPE* Exon 2. Examination of *GYPA*, *GYPB*, and *GYPE* Exon 2 sequences are shown in Fig. 4A. M+ *GYPA* reads are homologous to *GYPE* (c.59C, c.71G, and c.72T) and homologous to N+ *GYPA* at c.38C. As

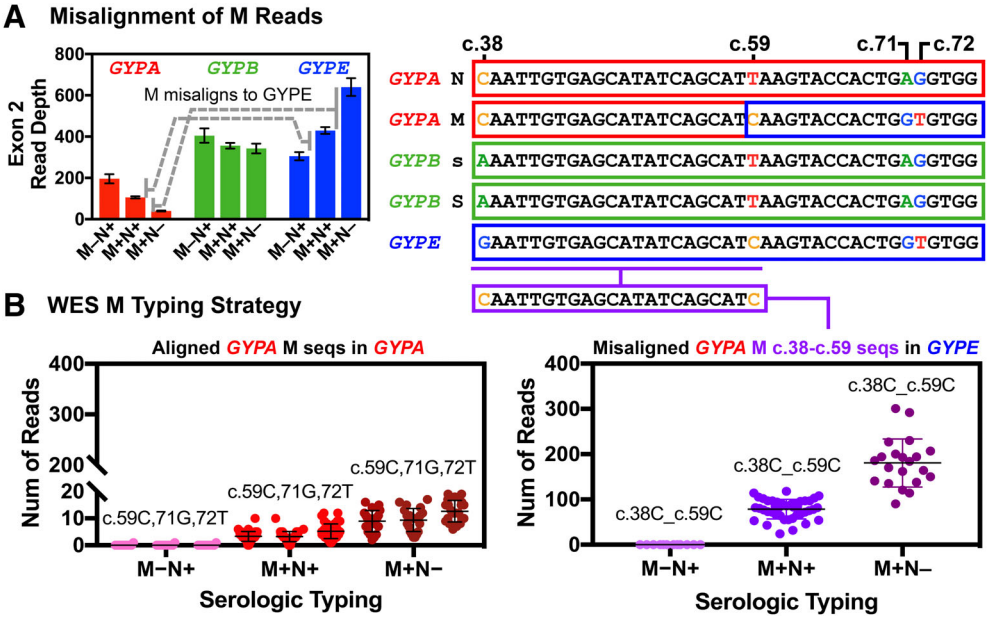

Fig. 5. bloodTyper MNS modifications for WES. Data shown for all 75 participants. (A, left) Mean Exon 2 sequence depth of coverage for *GYPA*, *GYPB*, and *GYPE* based on the M antigen dosage. (A, right) M and N antigen sequence positions in *GYPA*, *GYPB*, and *GYPE*. Sequence similarities between each gene are shown with colored boxes. The sequence pattern to find *GYPA* M antigen reads misaligned in *GYPE* is show in the purple box. (B) Number of normally aligned M sequence reads aligned to *GYPA* (left) and misaligned *GYPA* M sequence reads in *GYPE* (right) compared to M antigen dosage. [Color figure can be viewed at wileyonlinelibrary.com]

shown in Fig. 4B, an understanding of the partial misalignment was used to look for M+ specific reads misaligned to *GYPE* using the following pattern: *GYPA*-specific c.38C along with the M+-specific c.59C (M c.38C_c.59C reads). This approach was able to correctly identify the M antigen status in 100% of the remaining 55 participant samples.

One N antigen discordance in the first 20 samples was due to two N-like sequence reads in an N− individual. This discordance was corrected by increasing the N antigen nucleotide cutoff from 2× to the usual 4× for other nucleotide changes. With this adjustment, bloodTyper was 100% accurate at typing the N antigen over the last 55 WES samples. Poststudy analysis showed that the cutoff adjustment prevented the incorrect calling of another sample with two N-like reads in an N− individual by serologic and SNP typing.

Clinically relevant antigen typings

In the MilSeq Project protocol, under institutional review board approval and with patient consent, extended RBC and PLT antigen profiles were summarized for each participant and provided to the physician as part of the clinical report designed for that study. Among the MilSeq Project participants, the absence of high-prevalence PLT antigens was revealed in two HPA-1a− and one HPA-2a− individuals. None of the participants lacked high-prevalence RBC antigens, but three were e−, which occurs in 2% to 3% of Caucasians and is uncommon less than 0.1% in other ethnic groups. Additional information of relevance for determining risk for alloantibody production for transfusion included three patients of African ancestry whose RBCs were typed as Fy(b−) due to a GATA mutation (*ACKR1* c.−67T>C). Because Fy^b is absent on RBCs but present in tissues, it is presumed that these individuals are not at risk of developing anti-Fy^b alloantibodies. The cohort also included participants positive for uncommon antigens including four V+/VS+, four Co(b+), three Lu(a+), two Kp(a+), one Js(a+), one Di(a+), one weak D Type 45 hemizygote and one heterozygote, one partial D (DUC2) heterozygote, one C^W+, one HPA-9bw+, and one HPA-27bw+. The analysis also identified sickle trait (*HBB* c.20A>T) in one participant of African ancestry, which when present in blood donors can occlude filters during postdonation leukoreduction.[46] All of the above antigen findings were confirmed by serology and/or SNP array testing, except for *RHD***weak D Type 45*, *RHD***DUC2*, HPA-27bw+, and HbS, which were confirmed by Sanger sequencing.

DISCUSSION

In this study, we demonstrated the ability of automated analysis software, bloodTyper, to accurately type RBC and PLT antigens from whole exomes. After testing bloodTyper on the first 20 samples, modifications were made to the typing algorithm to accommodate for WES data, including copy number correction factors for *RHD*/*RHCE* exons and a novel method to detect *GYPA* M antigen-specific sequences misaligned to *GYPE*. With these modifications, WES-based typing was 100% accurate for all antigens tested by serologic and SNP typing methods.

Our study is not the first to evaluate large-scale targeted NGS or WES for RBC antigen typing.[9,11,13–15,20,22] However, ours is the first to automate the analysis and to evaluate its use for extended antigen typing. Similar to our previous experience with WGS[17] the C, M, and N antigens required WES-specific algorithmic modifications. Previous WGS and WES studies have shown that *RHD* Exon 8 does not properly align since the reference sequence is *RHD***DAU0*.[6,13,17] Others have indicated that it was not possible to detec*t RHD***DAU0* in WES data.[13] In this study, we found that it was possible to detect the presence of *RHD***DAU0*, but we were not able to distinguish between homozygous and heterozygous *RHD***DAU0*.

Whole exome sequencing data performed well in determining *RHD* zygosity and was more reliable than conventional PCR methods in samples from minorities which are known to be discordant when altered *RHD* alleles are present.[47–50] However, large *RHD* structural changes such as *RHD***D-CE(2-9)-D*, which were not present in any of the tested samples, would likely complicate WES-based zygosity calculations. Although it might be possible to use read depth-based copy number methods to call for these other large structural changes and then adjust the *RHD* zygosity call accordingly. After correcting for WES coverage biases, we were able to use sequence depth of coverage copy number to accurately type for the C antigen. Such an approach is vital here, since the intronic 109-bp insertion in Intron 2 is often used in SNP-based assays as a marker for the C antigen, but this region is not sequenced by WES.

In contrast to our findings with WGS we were not able to call M antigen simply by lowering the calling cutoff from 4× to 2×, since even at 2× there was a false M− call. However, using a defined sequence read pattern capable of finding *GYPA* M+ specific reads that misaligned to *GYPE*, M antigen typing was 100% accurate over the last 55 samples. This same technique should also be usable with WGS data for more robust M+ antigen typing, especially in analysis of genomes with lower than 15× coverage.

Similar to our previous WGS bloodTyper validation,[17] we also encountered serologic Fy^b errors. Initial Fy^b serologic typing was falsely negative in two single-dose Fy^b [Fy(a+b+)] samples and one double dose [Fy(a−b+)] sample, none of which on SNP or WES showed the presence of a known weak Fy^b, that is, Fy^x phenotype. Based on our prior experience with donor typing discrepancies, we have observed that serologic Fy^b reagents are not good at picking up single-dose expression of Fy^b (data not shown). The double-dose sample was homozygous for c.298A is which is part of a known weak allele (*FY***02W.01*), but c.298A is not

thought to be responsible for weak Fy^b without c.265T,[51] which was not found in this homozygous c.265C sample. It should also be noted that although allelic dropout of c.265T could account for this discordance, this is unlikely given that the 146× depth of coverage for c.265C in this sample was similar to other samples at this same position (mean of 194× with a standard deviation of 38×).

A limitation of the current study is that not every known RBC or PLT antigen could be tested for comparison because some antigens are very rare or only common in specific ethnicities. Similarly, the data set did not include samples with hybrid Rh and MNS changes. The copy number analysis algorithm will likely require future optimizations to address Rh and MNS hybrids. Full validation of bloodTyper for all known antigenic backgrounds will require the testing of additional data sets representing untested phenotypes. Other than for the P1 and Xg^a nucleotide changes, all other nucleotide changes were within 13 bp of exons and captured by WES. There are also rare weak A and B_3 subgroup nucleotide changes, not yet added to official allele tables, in *ABO* Intron 1 and upstream gene promoter regions that would not be detected by WES.

The ability to detect less common null nucleotide changes is an important advantage of both WGS and WES over available SNP typing assays. As shown here, the sequencing of intronic regions near exon boundaries covers the vast majority of allele nucleotide changes, including a Jk_{null} c.342−1A nucleotide change not detected by the SNP assay in a participant of self-declared Native Hawaiian/Pacific Islander/Asian ancestry (#70). Although the c.342−1A nucleotide is very rare in most ethnicities, it is present in 1.1% of Asians (chr18:43314238A).[52] Although bloodTyper detected the c.342−1A nucleotide change, it issued a warning since it could not cis/trans phase c.342 −1A with the *JK*A* and *JK*B* nucleotide changes. While c.342−1A has been reported on both *JK*A* and *JK*B* backgrounds, it is more prevalent on *JK*B*. Therefore, future modifications to bloodTyper will impute this and other allele haplotypes based on prevalence with notation of potential exceptions. As additional genomic data sets are made available, ethnicity-based haplotype frequencies should allow for even better imputation of short read NGS data. In addition, long and linked read sequencing NGS will increase this advantage since it should be able to cis/trans phase many more heterozygous changes like the one found here.

In summary, we have modified an automated WGS-based antigen typing software bloodTyper to accurately type RBC and PLT antigen from WES data. bloodTyper-based analysis could become a routine part of WES analysis. Given that WES is less expensive than WGS, it opens up the possibility of routinely using WES to aid in complex alloantibody workups. In addition, this work demonstrates the flexibility of bloodTyper to adapt to different data types, which in the future could include targeted NGS, high-density SNP arrays, and long-read NGS. This flexibility should also make it possible to extend bloodTyper to analyze for human neutrophil antigens and hemoglobinopathies.

ACKNOWLEDGMENTS

The authors thank the staff and participants of the MilSeq Project, as well as the staff of the New York Blood Center and Brigham and Women's Blood Bank. Members of the MilSeq Project are as follows: Jacqueline M. Killian, Cubby L. Gardner, Mauricio De Castro, Phoebe S. Lee, and Amanda C. Zehring, United States Air Force; Robert C. Green, Joel B. Krier, Kurt D. Christensen, Matthew S. Lebo, Jason L. Vassy, Megan Maxwell, Carrie L. Blout, and Bethany Zettler, Brigham and Women's Hospital; Amy McGuire, Stacey Pereira, Jill Robinson, Mary Majumder, Rebecca Hsu, and Amanda Guiterrez, Baylor College of Medicine; Maxwell Mehlman, Case Western Reserve University; and Efthimios Parasidi, Ohio State University.

CONFLICT OF INTEREST

RCG receives compensation for advising the following companies: AIA, Helix, Ohana, OptraHealth, Prudential, and Veritas; and is cofounder of Genome Medical, Inc., a nationwide telemedicine service providing expert advice in genetics. The other authors have disclosed no conflicts of interest.

REFERENCES

1. Stabentheiner S, Danzer M, Niklas N, et al. Overcoming methodical limits of standard RHD genotyping by next-generation sequencing. Vox Sang 2011;100:381-8.
2. Rieneck K, Bak M, Jønson L, et al. Next-generation sequencing: proof of concept for antenatal prediction of the fetal Kell blood group phenotype from cell-free fetal DNA in maternal plasma. Transfusion 2013;53(Suppl 2):2892-8.
3. Fichou Y, Audrézet M-P, Guéguen P, et al. Next-generation sequencing is a credible strategy for blood group genotyping. Br J Haematol 2014;167:554-62.
4. Wieckhusen C, Bugert P. 454-Sequencing™ for the KEL, JR, and LAN Blood Groups. Methods Mol Biol 2015;1310:123-33.
5. Giollo M, Minervini G, Scalzotto M, et al. BOOGIE: predicting blood groups from high throughput sequencing data. PLoS One 2015;10:e0124579.
6. Lane WJ, Westhoff CM, Uy JM, et al. Comprehensive red blood cell and platelet antigen prediction from whole genome sequencing: proof of principle. Transfusion 2016;56:743-54.
7. Lang K, Wagner I, Schöne B, et al. ABO allele-level frequency estimation based on population-scale genotyping by next generation sequencing. BMC Genomics 2016;17:374.
8. Baronas J, Westhoff CM, Vege S, et al. RHD zygosity determination from whole genome sequencing data. J Blood Disord Transfus 2016;7:1-5.

9. Fichou Y, Mariez M, Le Maréchal C, et al. The experience of extended blood group genotyping by next-generation sequencing (NGS): investigation of patients with sickle-cell disease. Vox Sang 2016;111:418-24.
10. Möller M, Jöud M, Storry JR, et al. Erythrogene: a database for in-depth analysis of the extensive variation in 36 blood group systems in the 1000 Genomes Project. Blood Adv 2016; 1:240-9.
11. Schoeman EM, Lopez GH, McGowan EC, et al. Evaluation of targeted exome sequencing for 28 protein-based blood group systems, including the homologous gene systems, for blood group genotyping. Transfusion 2017;57:1078-88.
12. Dezan MR, Ribeiro IH, Oliveira VB, et al. RHD and RHCE genotyping by next-generation sequencing is an effective strategy to identify molecular variants within sickle cell disease patients. Blood Cells Mol Dis 2017;65:8-15.
13. Chou ST, Flanagan JM, Vege S, et al. Whole-exome sequencing for RH genotyping and alloimmunization risk in children with sickle cell anemia. Blood Adv 2017;1:1414-22.
14. Jakobsen MA, Dellgren C, Sheppard C, et al. The use of next-generation sequencing for the determination of rare blood group genotypes. Transfus Med 2017;29:162-8. https://doi.org/10.1111/tme.12496.
15. Schoeman EM, Roulis EV, Liew Y-W, et al. Targeted exome sequencing defines novel and rare variants in complex blood group serology cases for a red blood cell reference laboratory setting. Transfusion 2018;58:284-93.
16. Orzińska A, Guz K, Mikula M, et al. A preliminary evaluation of next-generation sequencing as a screening tool for targeted genotyping of erythrocyte and platelet antigens in blood donors. Blood Transfus 2018;16:285-92.
17. Lane WJ, Westhoff CM, Gleadall NS, et al. Automated typing of red blood cell and platelet antigens: a whole-genome sequencing study. Lancet Haematol 2018;5:e241-51.
18. Wheeler MM, Lannert KW, Huston H, et al. Genomic characterization of the RH locus detects complex and novel structural variation in multi-ethnic cohorts. Genet Med 2019;21:477-86.
19. Wu PC, Lin Y-H, Tsai LF, et al. ABO genotyping with next-generation sequencing to resolve heterogeneity in donors with serology discrepancies. Transfusion 2018;58:2232-42.
20. Montemayor-Garcia C, Karagianni P, Stiles DA, et al. Genomic coordinates and continental distribution of 120 blood group variants reported by the 1000 Genomes Project. Transfusion 2018;58:2693-704.
21. Tounsi WA, Madgett TE, Avent ND. Complete RHD next-generation sequencing: establishment of reference RHD alleles. Blood Adv 2018;2:2713-23.
22. Schoeman EM, Roulis EV, Perry MA, et al. Comprehensive blood group antigen profile predictions for Western Desert Indigenous Australians from whole exome sequence data. Transfusion 2019;59:768-78.
23. Orzińska A, Guz K, Mikula M, et al. Prediction of fetal blood group and platelet antigens from maternal plasma using next-generation sequencing. Transfusion 2019;59:1102-7.
24. Schwarze K, Buchanan J, Taylor JC, et al. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genet Med 2018;20:1122-30.
25. Fung MK, Eder A, Westhoff CM, et al. Technical manual. Bethesda: AABB; 2017 745 p.
26. Westhoff CM, Vege S, Wylie D, et al. The JAL antigen (RH48) is the result of a change in RHCE that encodes Arg114Trp. Transfusion 2009;49:725-32.
27. Horn T, Castilho L, Moulds JM, et al. A novel JKA allele, nt561C>A, associated with silencing of Kidd expression. Transfusion 2012;52:1092-6.
28. Pillitteri D, Pilgrimm A-K, Kirchmaier CM. Novel mutations in the GPIIb and GPIIIa genes in Glanzmann thrombasthenia. Transfus Med Hemother 2010;37:268-77.
29. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947-8.
30. Wagner FF, Flegel WA. RHD gene deletion occurred in the Rhesus box. Blood 2000;95:3662-8.
31. Murdock A, Assip D, Hue-Roye K, et al. RHD deletion in a patient with chronic myeloid leukemia. Immunohematology 2008;24:160-4.
32. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 2013;43:11.10.1-11.10.33.
33. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26: 841-2.
34. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;14:178-92.
35. Omae Y, Ito S, Takeuchi M, et al. Integrative genome analysis identified the KANNO blood group antigen as prion protein. Transfusion 2019;59:2429-35. https://doi.org/10.1111/trf.15319.
36. Patrinos GP, Giardine B, Riemer C, et al. Improvements in the *HbVar* database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. Nucleic Acids Res 2004;32:D537-41.
37. Metcalfe P, Watkins NA, Ouwehand WH, et al. Nomenclature of human platelet antigens. Vox Sang 2003;85:240-5.
38. Reid ME, Lomas-Francis C, Olsson ML. The blood group antigen FactsBook. San Diego: Academic Press; 2012. p. 758.
39. International Society of Blood Transfusion. Red cell immunogenetics and blood group terminology [Internet]. Amsterdam: International Society of Blood Transfusion; [cited 2017 Sep 1]. Available from: https://www.isbtweb.org/working-parties/red-cell-immunogenetics-and-blood-group-terminology/
40. Lai YJ, Wu WY, Yang CM, et al. A systematic study of single-nucleotide polymorphisms in the A4GALT gene suggests a molecular genetic basis for the P1/P2 blood groups. Transfusion 2014;54:3222-31.
41. Westman JS, Stenfelt L, Vidovic K, et al. Allele-selective RUNX1 binding regulates P1 blood group status by transcriptional control of A4GALT. Blood 2018;131:1611-6.

42. Möller M, Lee YQ, Vidovic K, et al. Disruption of a GATA1-binding motif upstream of XG/PBDX abolishes Xga expression and resolves the Xg blood group system. Blood 2018;132:334-8.
43. Yeh CC, Chang CJ, Twu YC, et al. The molecular genetic background leading to the formation of the human erythroid-specific Xga/CD99 blood groups. Blood Adv 2018;2:1854-64.
44. Lane WJ, Aguad M, Smeland-Wagman R, et al. A whole genome approach for discovering the genetic basis of blood group antigens: independent confirmation for P1 and Xga. Transfusion 2019;59:908-15.
45. Carritt B, Kemp TJ, Poulter M. Evolution of the human RH (rhesus) blood group genes: a 50 year old prediction (partially) fulfilled. Hum Mol Genet 1997;6:843-50.
46. Stroncek DF, Rainer T, Sharon V, et al. Sickle Hb polymerization in RBC components from donors with sickle cell trait prevents effective WBC reduction by filtration. Transfusion 2002; 42:1466-72.
47. Wagner FF, Moulds JM, Flegel WA. Genetic mechanisms of Rhesus box variation. Transfusion 2005;45:338-44.
48. Matheson KA, Denomme GA. Novel 3'Rhesus box sequences confound RHD zygosity assignment. Transfusion 2002;42: 645-50.
49. Grootkerk-Tax MG, Maaskant-van Wijk PA, van Drunen J, et al. The highly variable RH locus in nonwhite persons hampers RHD zygosity determination but yields more insight into RH-related evolutionary events. Transfusion 2005;45:327-37.
50. Kacem N, Jemni-Yacoub S, Chiaroni J, et al. Paternal RHD zygosity determination in Tunisians: evaluation of three molecular tests. Blood Transfus 2015;13:59-65.
51. Yazdanbakhsh K, Rios M, Storry JR, et al. Molecular mechanisms that lead to reduced expression of duffy antigens. Transfusion 2000;40:310-20.
52. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. BioRxiv 2019 [cited 2019 March 1]. Available from: https://www.biorxiv.org/content/10.1101/531210v2.
53. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. Genome Res 2009;19: 1639-45.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Supporting Information.
Table S1. Automated typing of red blood cell and platelet antigens from whole exome sequences" by Lane WJ et al.