# Targeted exome sequencing defines novel and rare variants in complex blood group serology cases for a red blood cell reference laboratory setting

*Elizna M. Schoeman*,[1] *Eileen V. Roulis*,[1] *Yew-Wah Liew*,[2] *Jacqueline R. Martin*,[2] *Tanya Powley*,[2] *Brett Wilson*,[2] *Glenda M. Millard*,[1] *Eunike C. McGowan*,[1] *Genghis H. Lopez*,[1] *Helen O'Brien*,[1] *Jennifer A. Condon*,[3] *Robert L. Flower*,[1] *and Catherine A. Hyland*[1]

**BACKGROUND:** We previously demonstrated that targeted exome sequencing accurately defined blood group genotypes for reference panel samples characterized by serology and single-nucleotide polymorphism (SNP) genotyping. Here we investigate the application of this approach to resolve problematic serology and SNP-typing cases.

**STUDY DESIGN AND METHODS:** The TruSight One sequencing panel and MiSeq platform was used for sequencing. CLC Genomics Workbench software was used for data analysis of the blood group genes implicated in the serology and SNP-typing problem. Sequence variants were compared to public databases listing blood group alleles. The effect of predicted amino acid changes on protein function for novel alleles was assessed using SIFT and PolyPhen-2.

**RESULTS:** Among 29 unresolved samples, sequencing defined SNPs in blood group genes consistent with serologic observation: 22 samples exhibited SNPs associated with varied but known blood group alleles and one sample exhibited a chimeric RH genotype. Three samples showed novel variants in the CROM, LAN, and RH systems, respectively, predicting respective amino acid changes with possible deleterious impact. Two samples harbored rare variants in the RH and FY systems, respectively, not previously associated with a blood group allele or phenotype. A final sample comprised a rare variant within the KLF1 transcription factor gene that may modulate DNA-binding activity.

**CONCLUSION:** Targeted exome sequencing resolved complex serology problems and defined both novel blood group alleles (*CD55*:c.203G>A, *ABCB6*:c.1118_1124del CGGATCG, *ABCB6*:c.1656-1G>A, and *RHD*:c.452G>A) and rare variants on blood group alleles associated with al tered phenotypes. This study illustrates the utility of exome sequencing, in conjunction with serology, as an alternative approach to resolve complex cases.

The key role of a modern red blood cell (RBC) reference laboratory in transfusion medicine is to employ serology to solve complex problems. Often, the lack of appropriate RBCs, sera, or reagents makes confident resolution of a complex case difficult. Resources can be costly and scarce, if available at all, and often require collaboration and generosity of colleagues from other international laboratories.[1] In addition, reference laboratories commonly employ single-nucleotide polymorphism (SNP) typing microarrays for RBC genotyping investigations; however, these arrays are not comprehensive in their coverage of blood groups or variants.[2-5] When the combination of traditional serologic methods and typing by SNP microarray fails to resolve,

**ABBREVIATIONS:** ExAC = Exome Aggregation Consortium; MPS = massively parallel sequencing; RCRL = Red Cell Reference Laboratory; SNP(s) = single-nucleotide polymorphism(s); SNV(s) = single-nucleotide variant(s); TSO = TruSight One.

From [1]Clinical Services and Research; and the [2]Red Cell Reference Laboratory, Australian Red Cross Blood Service, Kelvin Grove, Queensland, Australia; and the [3]Red Cell Reference Laboratory, Australian Red Cross Blood Service, Melbourne, Victoria, Australia.

*Address reprint requests to*: Catherine A. Hyland, PhD, MSc, Clinical Services and Research, Australian Red Cross Blood Service, 44 Musk Avenue, Kelvin Grove, Queensland 4059, Australia; e-mail: CHyland@redcrossblood.org.au.

The Australian government funds the Australian Red Cross Blood Service for the provision of blood, blood products, and services to the Australian community.

Received for publication February 21, 2017; revision received August 21, 2017; and accepted September 5, 2017.

doi:10.1111/trf.14393



**TRANSFUSION** 2018;58;284–293

further avenues of investigation are required. For some reference laboratories the test algorithm at this stage will include Sanger sequencing of targeted genes.

As an alternative approach, massively parallel sequencing (MPS—also known as next-generation sequencing) technologies are becoming more widely used in blood banking and hematologic investigation, due to the capability of identifying all variants in target blood group genes. A recent proof-of-principle study demonstrated the capability of using whole genome sequencing data for comprehensive antigen prediction of RBCs and platelets.[6] A targeted exome sequencing approach has been recognized for its utility in blood group genotyping and has been predicted to facilitate investigation of novel blood group alleles.[7,8] While studies have shown proof of principle for blood group genotyping by MPS, no studies have investigated to date its clinical application to resolve complex cases in a reference laboratory setting.[6,8-12]

Previously, we validated the use of a commercially available targeted exome sequencing panel for investigation of genetic variation within 28 protein-based blood group systems on a collection of well-characterized reference panel samples.[13] Our analyses confirmed the phenotypes obtained using SNP microarray and serologic methods, as well as identifying a number of variants considered "noninformative"—that is, they may be common population variants or variants that have not previously been associated with established blood group alleles or phenotypes.

Here we investigate the potential application for blood group genotyping using targeted exome sequencing to resolve complex cases, where sequencing was requested by the Australian Red Cross Blood Service Red Cell Reference Laboratory (RCRL). These requests were made either where serology typing was discordant to SNP genotyping or where confirmation was needed to support antigen typing.

## MATERIALS AND METHODS

### RCRL serology investigations

Samples referred to investigate the D status were first tested with two anti-D antisera: one, Epiclone-2 anti-D (IgM) (Seqirus) that does not detect most partial D and weak D types, and another, anti-D (monoclonal blend) Gamma-clone (Immucor) that does detect most partial D and weak D types. Samples were further tested with ALBAclone advanced partial Rh(D) typing kit (Alba Bioscience) as per the manufacturer's instruction.

Phenotyping was performed, where possible, with commercially available antisera. When commercial antisera were not available, in-house antisera was used with the optimum technique for that particular antiserum. Where weak antigen expression was suspected, adsorption-elution techniques were used to confirm presence of the antigen. Cell separation of a dual population (Sample 7) was performed by agglutinating the D+ cells with anti-D and separating from the unagglutinated cells using a 20% albumin gradient.

Antibody identification was performed using standard serologic techniques. Routinely, a commercial 11-cell panel (Seqirus) and autologous cells are tested by three tube methods: saline room temperature (24°C), polyethylene glycol indirect antiglobulin test (IAT), and papain IAT. Results from this routine testing directed further serologic investigation with various enzymes or chemical modification of cells, rare phenotype cells, cord cells, antibody adsorptions-elutions, and peptide inhibitions.

### SNP genotyping in the reference laboratory

Genomic DNA was extracted from fresh EDTA whole blood samples using the DNA blood mini kit (QIAamp DSP, QIAGEN) for use with a robotic workstation for automated purification of DNA, RNA, or proteins (QIACube, QIAGEN) or manually as per the manufacturer's instructions if required. The DNA was tested using a high-throughput molecular assay (BioArray BeadChip technology, Immucor) for the simultaneous detection of a wide range of SNPs affecting RBC antigen expression, using a multiplexed molecular assay (either Immucor PreciseType HEA molecular BeadChip test, herein referred to as "HEA" or the RHD Molecular BeadChip test).

The HEA test predicts antigen phenotype for RH (C, c, E, e, V, VS), KEL (K, k, $Kp^a$, $Kp^b$, $Js^a$, $Js^b$), FY ($Fy^a$, $Fy^b$, GATA, $Fy^x$), JK ($Jk^a$, $Jk^b$), MNS (M, N, S, s, Uvar, Uneg), LU ($Lu^a$, $Lu^b$), DO ($Do^a$, $Do^b$, Hy, $Jo^a$), LW ($LW^a$, $LW^b$), DI ($Di^a$, $Di^b$), CO ($Co^a$, $Co^b$), and SC (Sc1, Sc2).[14] The RHD test utilizes 35 genetic markers associated with D phenotypes to make *RHD* variant calls.[15]

### MPS library preparation and sequencing

We prepared DNA libraries for sequencing using a sequencing panel (TruSight One [TSO], Illumina) as per manufacturer's instructions. The TSO panel amplifies exons and untranslated regions for 4813 clinically significant genes, including the 41 genes related to 33 blood group systems and genes for two transcription factors, KLF1 and GATA1 (Table S1, available as supporting information in the online version of this paper). The genes for three (JMH, VEL, and XG) of the 36 blood group systems were excluded by the manufacturer of the kit. MPS was performed on the MiSeq sequencing platform (Illumina) to generate paired-end 150-bp reads as described previously.[13]

### Data analysis

FASTQ files were generated by the MiSeq system on-instrument software and exported to computer software

(CLC Genomics Workbench software, Version 8.5, QIAGEN) for secondary data analysis (Fig. 1). Reads were trimmed for size and quality using default settings before alignment to the human genome reference (Hg19/GRCh37). Variant detection was subsequently performed using an adjusted Illumina manifest file (targeting only 33 blood group systems, KLF1 and GATA1 genes) to mask data from non–blood group–related genes. Data filtering excluded variants with quality score of less than 30 and/or heterozygote variant calls where the proportion of sequence reads that contain the variant allele is less than 15.9%.[13] Finally, the CLC Genomics Workbench was used to annotate all gene variants detected, for example, with information from NCBI SNP database (dbSNP) build 142, where available. We considered a variant to be of low confidence, but still of possible value, where the coverage was less than 20× and/or strand bias occurred. For sequence data analysis for Sample 7 where serology suggested a mixed Rh phenotype, we adjusted the minimum frequency variable during variant calling to 1% from 15.9%.

For tertiary data analysis, only blood group variants related to the request were investigated. Variants listed on the ISBT database were used to correlate predicted phenotype with observed phenotypes where provided.[16] Additional sources were queried where required. These sources included Erythrogene, The Blood Group Antigen Gene Mutation Database (NCBI dbRBC), RhesusBase, version 2.0, Human Blood Groups and The Blood Group Antigen Facts Book.[17-21] For genotypic and phenotypic prediction for the homologous *RH* gene system, we also considered results from CNV analysis of MPS data as described previously.[13,22,23] We have previously shown that CNV allows accurate determination of the *RHCE*C* allele and *RHD* zygosity.[13] To determine whether novel nonsynonymous SNPs had a deleterious effect on protein function, amino acid sequences for genes encoding novel alleles were compared to reference alleles using SIFT[24] and PolyPhen-2.[25] These online tools assess protein phenotypic tolerance to nonsynonymous mutations using in silico comparative homology predictions.

### Sanger sequencing

Sanger sequencing was employed for Sample 1 to confirm the presence of a putative novel SNP defined by MPS in the CROM blood group system. A primer pair (CD55_Ex2F 5′-TCCTTCAGTTCTGCTTTTGT-3′ and CD55_Ex2R 5′-CTCAGGGTGGTAAATGTCTTC-3′) was designed to amplify Exon 2 for *CD55* and primers were manufactured by Sigma-Aldrich (St Louis, MO). Polymerase chain reaction (PCR) conditions were 95°C for 15 minutes, followed by 35 cycles of 94°C for 30 seconds, 54°C for 30 seconds, 72°C for 1 minute, and a final extension of 72°C for 10 minutes using a ready-to-use 2× reaction mix (HotStart Taq Master Mix kit, QIAGEN). Amplicons (for Sample 1 and a control sample) with an expected product size of approximately 386-bp were purified using a PCR purification kit (MinElute, QIAGEN) and were sequenced at Australian Genome Research Facility (St Lucia, Queensland, Australia)

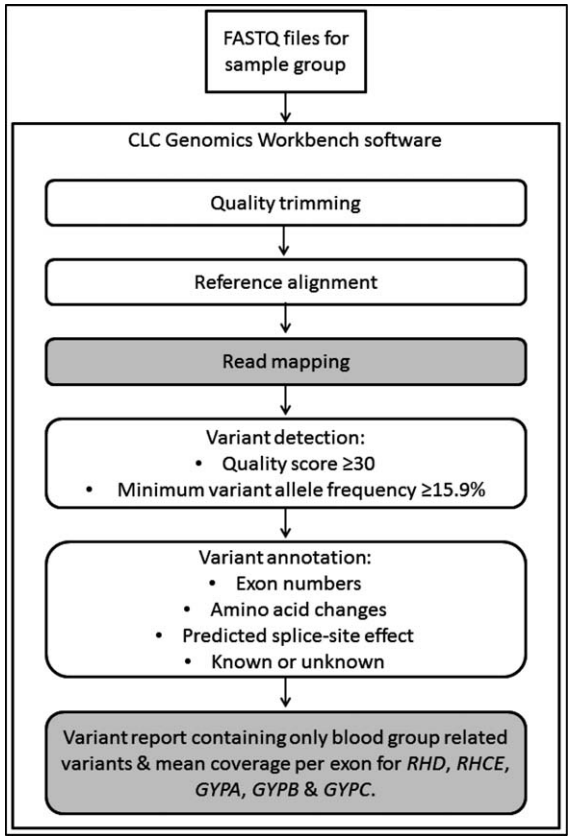


**Fig. 1. Overview of secondary data analysis using the CLC Genomic Workbench software. White rounded rectangles = processes; gray rounded rectangles = data output files.**

## RESULTS

### Overview of MPS typing and serology

Thirty samples were received with a request for sequencing over 2015. For 22 of the 30 samples, MPS defined previously reported SNPs (details outlined in Table S2, available as supporting information in the online version of this paper). For 12 of the 22 samples, SNPs were defined that are not targeted on the SNP microarray platform but are known to associate with blood group antigen polymorphisms listed on at least one blood group database. These 12 samples include eight cases where individual SNPs are not targeted and four cases where the blood group system is not targeted by the SNP microarray platform. In an additional 10 of the 22 samples MPS confirmed serology and/or SNP genotyping results. We report here on the remaining

eight samples, which include six with novel or rare blood group alleles, and a final sample, which we believe provides the first report on the possible application of a targeted exome sequencing approach to investigate chimerism.

Of these eight samples, one presented with an antibody of unknown specificity and another two presented with antibodies against antigens in the CROM and LAN blood group systems. Four samples showed weak antigen reactivity against antigens in the RH (two samples), LU, and FY blood groups systems. One sample presented with a mixed-field reaction in Rh typing.

### MPS performance and overview of variant detection

For the eight samples reported here, the average sequencing depth per target region (n = 429) ranged from 9× to 110×. Target regions included exons and untranslated regions. Details on the average sequencing read depth of coverage per target region are provided in Table S1.

Variant reports for all eight samples contained a total of 71 ± 13 variants on average across 33 blood group genes, *KLF1* and *GATA1*. VEL, JMH, and XG are not included in this panel. Of the total number of variants, an average of 23 ± 4 were synonymous variants, 42 ± 9 led to nonsynonymous amino acid changes, and 5 ± 3 were situated in intronic positions.

For all eight samples, the minimum coverage for variants ranged from 3× to 57× and the maximum from 21× to 399×. Of the total number of single-nucleotide variants (SNVs) contained in the variant reports, an average of 2 ± 2 SNVs were used to provide genotype calls for the requested target blood group system(s).

### Overview of MPS genotype calls for problem cases

For seven of the eight samples, MPS detected nucleotide sequence variations on the blood group gene of interest, which predicted a phenotype that could possibly explain the serology observations (Table 1). The one exception was a sample that presented with an antibody(-ies) of unknown specificity suspected to be either high-titer, low-avidity, or anti-Vel. No variants were detected for the 33 blood group systems on the panel that could explain the observed serology; however, the TSO panel does not include targets for *SMIM1* (VEL blood group) and therefore a causative VEL variant could not be excluded. In this case, MPS could not resolve our investigation.

For the remaining samples, MPS defined SNVs suggesting novel alleles in Samples 1 to 3 and rare SNVs not previously associated with blood group alleles in Samples 4 to 6. MPS variant calling and CNV analysis supported the Rh mixed-field serology for Sample 7.

### Novel blood group alleles predicting changed serology profiles

For three samples (Samples 1 to 3 in Table 1) a novel SNV was observed. These were samples that showed antibodies to the CROM and LAN blood group system antigens and weak D antigen, respectively.

Sample 1 was from a transfused patient with an antibody appearing to recognize a high-frequency antigen by reacting with all panel cells. Soluble CD55 completely inhibited binding of the patient's antibody to RBCs, indicating that the antibody recognized an epitope within the CROM blood group. Sequencing was requested since the CROM system is not included in the BeadChip test. MPS analysis of *CD55* identified only one SNV: a homozygous c.203A>G predicting p.Ser68Asn in the first extracellular complement control repeat domain of the GPI-anchored CROM protein, in close proximity to positions of known low-frequency antigens Tc$^{c}$ and WES$^{a}$ (Fig. 2). There are no previous reports of c.203A>G in dbSNP or Exome Aggregation Consortium (ExAC).[26,27] Sanger sequencing confirmed the presence of the novel *CD55*:c.203A>G variant at a homozygous level in Exon 2 (Fig. 3). We propose that the proband carries a novel homozygous mutation that gives rise to a new CROM allele resulting in a lack of a hitherto unrecognized high-frequency antigen. It is possible that the novel allele gives rise to a yet unrecognized low-frequency antigen but there is as yet no evidence for this.

Sample 2 presented with an antibody to a high-frequency antigen, which appeared to be anti-Lan upon investigation. In this case sequencing was requested to confirm the LAN phenotype as the patient was Group AB and there was only one suitable anti-Lan antiserum available and the LAN system is not included in the BeadChip test. MPS analysis of *ABCB6* (LAN system) identified only two novel heterozygous variants: a splice site variant (c.1656-1G>A) in Intron 9 and a 7-bp deletion in Exon 5 resulting in a frameshift mutation (c.1118_1124delCGATCCG/p.Ala373Glyfs*47), which terminates at the 3′ end of Exon 6. Neither of these two variants have previously been reported in dbSNP or ExAC. Our data suggest a null compound heterozygote for the two variants. We therefore propose these as new alleles in the LAN blood group system.

Sample 3 was an antenatal patient referred to determine D status. The RBCs phenotyped as C+E–c+e+ and exhibited an unknown reaction profile (no reaction with six of 12 anti-D lines and weak reactions with LHM76/58, LHM76/59ALBA, LHM50/2B, LHM169/81, ESD1, and LHM169/80) with an ALBAclone advanced partial RhD typing kit (Table S3, available as supporting information in the online version of this paper). RHD molecular BeadChip Test predicted the phenotype result as "possible D." MPS analysis for the RH and RHAG systems showed this

**TABLE 1. Summary of MPS results for seven cases with unresolved complex serology**

| Case | Blood group system (gene) analyzed | Variant(s) identified and depth of coverage. | Conclusion and predicted phenotype |
|---|---|---|---|
| 1 | CROM (*CD55*) | Novel homozygous c.203G>A (p.Ser68Asn) at 148× coverage: ClinVar accession number SCV000266303.1 | RBCs lack high-frequency antigen, not previously reported. |
| 2 | LAN (*ABCB6*) | Novel heterozygous c.1118_1124delCGGATCG (p.Ala373Glyfs*47) at 60/141× coverage: ClinVar accession number SCV000297765.1.<br>Novel heterozygous c.1656-1G>A at 31/81× coverage: ClinVar accession number SCV000297764.1. | Neither mutation was reported before. Predict nonfunctional protein. Probable null phenotype. |
| 3 | RH (*RHD* and *RHCE*)<br>RHAG (*RHAG*) | CNV analysis: hemizygous *RHD*.<br>Novel c.452G>A on *RHD* (p.Gly151Asn) at 55× coverage: ClinVar accession number SCV000297766.1.<br>CNV analysis: heterozygous *RHCE*C*. | Altered D expression.<br><br>C+E–c+ e+ |
| 4 | RH (*RHD* and *RHCE*)<br>RHAG (*RHAG*) | CNV analysis: hemizygous *RHD*.<br>*RHD*weak D Type 33* associated c.520G>A (p.Val174Met) at 58× coverage.<br>Homozygous rare *RHD*: c.353T>A (p.Met118Lys) with no current phenotypic associations at 35× coverage.<br>CNV analysis: heterozygous *RHCE*C* | Novel combination of variants leading to either weak D or weak partial D expression.<br><br>C+E–c+e+ |
| 5 | KLF1 transcription factor (*KLF1*) | Heterozygous rare c.964C>A (p.Arg322Ser) with no current phenotypic associations at 20/49× coverage | Possible In(Lu). |
| 6 | FY (*ACKR1*) | Heterozygous c.125G>A (p.Gly42Asp) associated with the $Fy^a/Fy^b$ polymorphism at 71/133× coverage.<br>Heterozygous rare c.977C>T (p.Ser326Phe) with no current phenotypic associations in 17/56X coverage | Fy(a+b+). Serology leads us to conclude p.S326F is implicated in reduced $Fy^a$ expression. |
| 7 | RH (*RHD* and *RHCE*) | CNV analysis: homozygous *RHD* gene deletion with minor population of cells containing *RHD*01*.<br>CNV analysis: homozygous *RHCE**c with minor population of cells containing c.307C>T associated with *RHCE*C*.<br>*RHCE*E* associated c.676G>C (p.Ala226Pro) at 53/60× coverage (88% variant allele frequency). | Major haplotype: cdE.<br>Minor haplotype: CDe. |

sample to be hemizygous for *RHD* with one *RHD* SNV:c.452G>A predicting a p.Gly151Asn in the fifth helical transmembrane domain for the RhD protein. There are no previous reports of c.452G>A in either dbSNP or ExAC. MPS data analysis (including CNV analysis) predicted an RhCE phenotype as C+E–c+e+. The impact of p.Gly151Asn was predicted in silico to be "tolerated" by SIFT[24] and "possibly damaging" by PolyPhen-2.[25] In comparison, the proximate p.Ala149Asn defining the *RHD*weak D Type 5* allele reportedly has a weak partial D reactivity profile.[16,28] The p.Gly151Asn mutation is consistent with a novel allele within the Rh blood group system.

### Rare SNPs associated with serology phenotype

Samples 4, 5, and 6 each presented with a rare SNP reported on polymorphism databases that have not previously been associated with any blood group phenotype (Table 1). For these three samples no other sequence variations were present that could explain the serologic phenotype. These variants occurred on the genes for the RH and FY blood group systems and for the KLF1 transcription factor and are described as follows.

Sample 4 was an antenatal patient referred to determine D status. The RBCs phenotyped as C+E–c+e+ and were positive with eight of the 12 anti-D cell lines with the ALBAclone advanced partial RhD typing kit, fitting the reaction profile for the variant DFR (Table S3). RHD molecular BeadChip test predicted phenotype result was "possible D". MPS analysis of the RH and RHAG systems showed this sample to be hemizygous for *RHD* with two *RHD* SNPs in cis: c.520G>A (Exon 4) predicting p.Val174-Met and defining the *RHD*weak D Type 33* allele and c.353T>A (Exon 3) predicting p.Met118Lys (fourth transmembrane helical domain). The p.Met118Lys substitution has only been reported once before in ExAC in a South Asian sample (MAF = 0.0009%; SNP ID rs773065365) and has no current known phenotypic associations. Although mutations in transmembrane helical domains are usually associated with weak D or DEL phenotypes, such changes may also weaken, destroy, or create epitopes.[29] Interestingly, RBCs expressing the weak D Type 33 phenotype showed a complete reactivity profile,[30] suggesting that the partial D reactivity profile for Sample 4 is likely caused by p.Met118Lys (Table S3). Whether the partial reaction profile is caused by p.Met118Lys alone or in combination with the *RHD*weak D type 33* variant is uncertain.

Sample 5 showed an apparent Lu(a–b–) phenotype while performing routine screening for Lu(b–)-negative

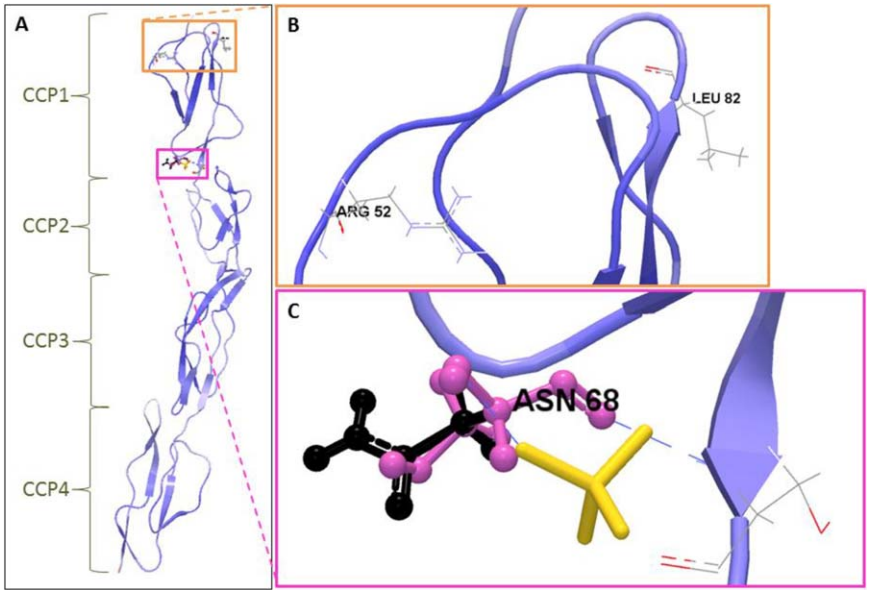


**Fig. 2. Three-dimensional model of CD55 protein showing p.Ser68Asn identified in Sample 1 (Table 1). (A) Three-dimensional backbone model of the complement decay accelerating factor protein (CD55) in purple showing the amino acid change p.S68N in the pink outlined block and the amino acid position for known high-frequency antigens (HFA) in the orange outline block. CCP = complement control protein motif. (B) Zoomed-in view of amino acid positions for HFA (gray wireframe figures) where the amino acid changes p.Arg52Pro and p.Leu82Arg occur that give rise to the low-frequency antigens Tc$^b$, Tc$^c$, and WES$^a$, respectively. (C) Zoomed-in view of amino acid change p.Ser68Asn observed in Sample 1. Pink ball and stick figure = serine at Position 68; black ball and stick figure = asparagine at Position 68; gray wireframe figure = serine at Position 97; yellow stick figure = sulfate cofactor; blue dash line = hydrogen bond. It is possible that the hydrogen bonds are impacted by the change to asparagine observed in Sample 1. Images created using CLC Genomics Workbench version 8.5.**

units using in-house anti-Lu$^b$ by column agglutination technique and with commercial anti-sera. Anti-Lu$^b$ was subsequently adsorbed and eluted from the donor's RBCs, suggesting the In(Lu), or "inhibitor of LU," phenotype. The donor's cells were then typed for the high-frequency LU antigens AnWj, Lu8, Lu17, Lu3, and Lu6 and found to be negative. MPS defined a heterozygous c.964C>A (rs376711350; reported in one out of 96 South East Asian individuals) variant in *KLF1*, predicting a p.Arg322Ser in the second highly-conserved[31] zinc finger domain of the KLF1 protein. The three zinc finger domains of KLF1 are involved in sequence-specific DNA binding that modulates its hematopoietic-specific transcription factor activity. The p.Arg322Ser mutation is likely to impair DNA binding or block it completely and such rare heterozygous loss of function *KLF1* mutations can result in dominant suppression of LU antigens and the In(Lu) phenotype.

Sample 6 exhibited a weak Fy$^a$ antigen and phenotyped as Fy(a+$^w$b+) after observing reduced Fy$^a$ expression by flow cytometry. HEA BeadChip predicted the phenotype Fy(a+b+). Genotyping by MPS defined a rare variant, c.977C>T (rs17851570; MAF = 0.19% in mostly European populations in ExAC) at a heterozygous level in the FY gene *ACKR1*. This nucleotide substitution predicts a p.Ser326Phe mutation in the cytoplasmic C-terminal domain of the FY protein. It is known that a change (p.Arg89Cys) in the first cytoplasmic loop results in protein instability, causing a weakened expression in Fy$^a$ and Fy$^b$.[16,32-34] The p.Ser326Phe substitution, located at the other end of the protein, appears to have the same effect on the Fy$^a$ antigen.

### Mixed-field agglutination in Rh serology testing

Sample 7 RBCs from a patient with myelofibrosis phenotyped as Rh:D+C+c+E+e+, with mixed-field reactions observed with anti-D, anti-C, and anti-e typing reagents. Cell separation, using anti-D, left unagglutinated cells that typed as D–C–c+E+e– only, suggesting that the most likely genotype is CDe/cdE with a subset of cells not expressing the CDe haplotype. MPS variant detection analysis identified a homozygous (variant allele frequency = 88%) nucleotide substitution (c.676G>C) in *RHCE* that defines the *RHCE*E* allele.[16] In comparison, we have found the variant allele frequency for other homozygote c.676G>C samples (n = 3) to be equal to 100% and the mean variant allele frequency for heterozygote c.676G>C samples (n = 29) to be equal to 49% (95%

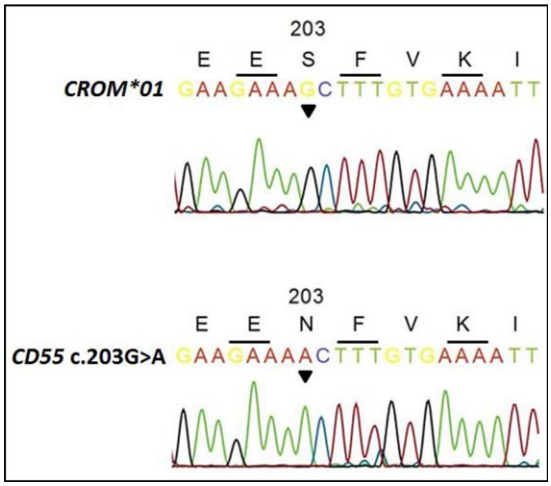


**Fig. 3. Sequencing chromatograms comparing nucleotide sequences between the reference allele (*CROM*01*) and the variant allele, c.203G>A identified in Sample 1 (Table 1). Amino acid position in question is indicated with the number 203. (▼) Nucleotide Position c.203.**

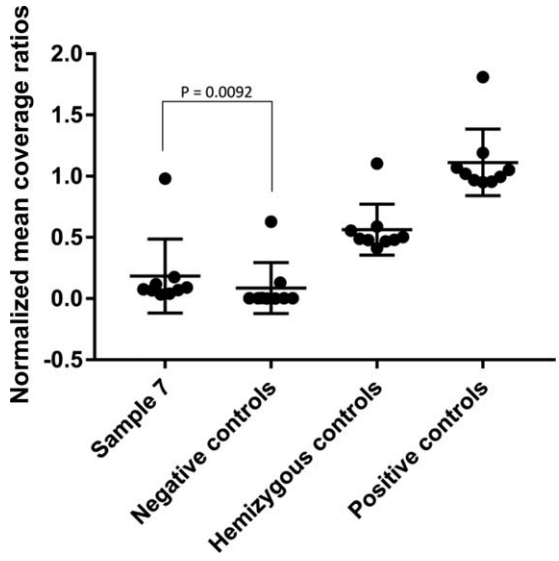


**Fig. 4. Scatter dot plot of the normalized mean coverage ratios of *RHD* exons for Sample 7 compared to that of control samples. Normalized mean coverage ratios for all exons (except Exon 8) were determined during CNV analysis as described previously.[13] The average normalized mean coverage ratios were determined for negative (n = 6), hemizygous (n = 6), and positive (n = 12) control samples for comparison. The ratios for Sample 7 appear to be similar to that of the negative controls on first glance but were calculated to be significantly different (p = 0.0092).**

confidence interval, 46-51). The 88% variant allele frequency for c.676G>C in Sample 7 therefore did not clearly correspond to either homo- or heterozygosity, but suggested a major population of cells containing the *RHCE*E* allele and a minor population (at ±12%) containing the *RHCE*e* allele.

In addition, CNV analysis performed as described previously[13] indicated a homozygous *RHD* gene deletion with some background signals that appeared to be above the levels seen in negative controls. As reported previously, CNV analysis of Exon 2 for *RHD* and *RHCE* can be used to identify the presence of the *RHCE*C* allele.[13] For Sample 7, CNV analysis appeared to indicate homozygosity for *RHCE*c*. However, subsequent analysis showed that the normalized mean coverage ratios of *RHD* exons for Sample 7 were significantly increased (p = 0.0092) compared to the average normalized mean coverage ratios for negative control samples (n = 6; see Fig. 4). These observations could be indicative of a minor population of cells containing the *RHD*01* allele. In addition, the normalized mean coverage ratio for *RHD* Exon 2 appeared increased relative to that observed for other *RHD* exons. This would be consistent with a minor population of cells containing c.307C>T associated with the *RHCE*C* allele. In summary, the major haplotype (±88%) was defined as cdE and the minor (±12%) as CDe. This is consistent with serology results.

## DISCUSSION

We previously showed that MPS using a targeted exome approach provided an extended and accurate blood group genotype profile for samples characterized by extended blood group phenotyping and/or by SNP microarray genotyping.[13] Here, we demonstrate the utility of combining reference laboratory typing methods with the TSO sequencing panel to resolve problematic serologic cases. In the process, we defined three novel blood group alleles in the CROM, LAN, and RH systems and three rare variants in RH, KLF1, and FY not previously associated with blood group phenotypes. In all cases, the variants identified by genotyping predicted amino acid changes that had potential to alter the blood group antigen presentation and were consistent with serology findings.

It has previously been shown that MPS can be applied for accurate blood group genotyping and RBC antigen prediction.[6,8] Whole exome sequencing has proven to be an efficient way to define the genetic basis for "orphan" antigens and to provide a precise *RH* genotype in patients with sickle cell anemia.[35-37] Whole genome sequencing has been proposed as an alternative to targeted sequencing to provide a full blood group genotype profile but this approach generates an enormous amount of data, much of which is inconsequential to blood group genotyping.[6] Fichou and colleagues[7] demonstrated that targeted MPS could be applied for routine blood group genotyping in specific populations such as sickle cell disease patients.

They predicted that this approach would be valid in the discovery of novel blood group alleles, and our study supports that prediction.

It is possible that Sanger sequencing would have solved some cases presented here; however, an MPS approach has clear advantages. First, allelic dropout is avoided by multiple overlapping amplicons for target regions. Second, automated assignment of quality scores for each base allows quality trimming and ensures consistency and valid comparison of results. Third, multiple genes can be sequenced simultaneously. Finally, CNV analysis can be performed, for example, to determine *RHD* zygosity. In essence, MPS provides a simplified massive parallel typing approach using a single test format compared to the battery of tests for Sanger sequencing.

It is important to note that our data analysis was restricted to those genes relevant to the requested analysis from the RCRL and that no additional variants were detected in these target blood group genes to confound interpretation. This was consistently applied for the 30 samples referred for testing (the 22 samples with known SNPs are outlined in Table S2). Serology results not only enabled focused analysis of MPS data but in one of the cases (Sample 7) serology led the interpretation of MPS data. This latter was in the case of the mixed Rh serology where MPS was applied to test whether results would be consistent with serology. To the best of our knowledge, this is the first report on the possible application of a targeted exome sequencing approach to investigate chimerism. MPS analysis, however, only supported the phenotype results and no additional tests such as short tandem repeat analysis were performed. Further testing of such samples would be needed to assess the lower limit of detection for this MPS approach.

Samples 3 and 4 were from antenatal patients where correct D phenotype assignment is important to determine whether antenatal and postnatal RhIG prophylaxis will be required to manage the pregnancy appropriately.[38] Serology reagents were unable to provide a clear classification, although admittedly the partial D phenotypes did suggest that RhIG was required in both cases. SNP typing was unable to explain the partial D serological profiles, providing a result of "possible D" in both cases (Table 1). This result is a recognition by the manufacturer that the RHD BeadChip kit does not include all of the *RHD* variants reported in the ISBT database.[15] Subsequent TSO sequencing was able to provide an explanation for the serology findings in both cases. These cases highlight the strengths and weaknesses of both serology and SNP-based genotyping and the importance of sequence-based genotyping to resolve discrepancies.

The targeted exome approach covered 33 of the 36 blood group systems in a single test format. The SNP genotyping array systems, however, are accredited by regulatory bodies whereas the targeted exome sequencing approach described here is currently for research use only. It takes one trained scientist approximately 3 days to prepare up to nine samples using the TSO library preparation kit, and up to 27 hours to sequence a maximum of three pooled samples per run on a MiSeq platform. The time spent on analysis and reporting depends on the complexity of the case involved and the skill set of the bioinformatician. As is the case with a whole genome or exome sequencing approach, the manual review and interpretation of data generated by the targeted exome approach reported here requires expertise in both RBC immunogenetics and bioinformatics. Review and interpreting variants can take approximately 1 hour for a simple case, for example analyzing a single gene, or up to 1 day for a more complex case.

On the other hand, the targeted exome approach is more comprehensive than SNP genotyping array methods even though the TSO sequencing panel does not provide complete blood group genotyping, as genes for JMH, VEL, and XG are excluded, and only Exon 1 of *A4GALT* (P1PK system) is covered.[13] The need for a complete blood group sequencing panel is exemplified by the inability to interrogate the VEL blood group gene system to resolve a sample presenting with putative anti-Vel. This would be overcome by design of a custom panel limited only to blood group genes. A custom panel would also address any ethical concerns regarding the masking of additional non–blood group–associated data obtained from the TSO sequencing panel and potentially actionable clinically significant variants that are therefore not identified. The right to know or not to know incidental findings generated by whole genome or exome sequencing is a heavily debated issue with the American College of Medical Genetics and Genomics recommending informing about incidental findings that are accurate and actionable.[39]

A targeted exome sequencing approach enabled us to expand current knowledge on blood group molecular genetics by defining novel and very rare variants and their first association with serologic phenotypes. Additional RBC functional studies are yet to be performed to validate the phenotypic effect of the novel and rare variants identified in this study.

The current algorithm for the resolution of cases presenting with complex or confounding serology in a RBC reference laboratory involves sequential application of serology, SNP genotyping frequently using broad or Rh-specific modules, and Sanger sequencing on selected blood group genes if required. This process is expensive and time-consuming and has limited utility in identifying novel or rare blood group alleles. In our own study we applied a targeted exome sequencing strategy as part of the algorithm to resolve complex serology cases. As we have shown, the TSO panel covers many blood group genes in a single test. It would even be possible in the future to run the serology and MPS testing concurrently,

with the results of serology guiding the interpretation of the MPS data as exemplified in this study. For now, this study illustrates the utility of targeted exome sequencing, in conjunction with serology, as an alternative approach to resolve complex cases.

### ACKNOWLEDGMENTS

Serology was performed by various staff at the Australian Red Cross Blood Service Red Cell Reference Laboratories in Perth, Melbourne, Brisbane, and Sydney. We acknowledge the two samples that were received from international laboratories for the Red Cell Reference Laboratory.

### CONFLICT OF INTEREST

The authors have disclosed no conflicts of interest.

## REFERENCES

1. Scharberg EA, Richter E, Bugert P. Red cell antigen testing. ISBT Sci Series 2015;10:5-11.
2. Denomme GA. Cost-effectiveness of implementing molecular immunohematology. ISBT Sci Series 2017;12:223-6.
3. McBean RS, Hyland CA, Flower RL. Approaches to determination of a full profile of blood group genotypes: single nucleotide variant mapping and massively parallel sequencing. Comput Struct Biotechnol J 2014;11:147-51.
4. McBean R, Liew YW, Wilson B, et al. Genotyping confirms inheritance of the rare At(a−) type in a case of haemolytic disease of the newborn. J Pathol Clin Res 2016;2:53-5.
5. Storry JR, Castilho L, Chen Q, et al. International Society of Blood Transfusion Working Party on Red Cell Immunogenetics and Terminology: report of the Seoul and London meetings. ISBT Sci Series 2016;11:118-22.
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. Fichou Y, Mariez M, Le Marechal 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.
8. Fichou Y, Audrézet MP, Guéguen P, et al. Next-generation sequencing is a credible strategy for blood group genotyping. Br J Haematol 2014;167:554-62.
9. Avent ND, Madgett TE, Halawani AJ, et al. Next generation sequencing: academic overkill or high-resolution blood group genotyping? [4A-S25-01 abstract]. Vox Sang 2014;107:37.
10. Altayar MA, Halawani AJ, Kiernan M, et al. Extensive genotyping of blood groups Duffy, Kidd and ABO by next generation sequencing [P-407 abstract]. Vox Sang 2014;107:187.
11. Halawani AJ, Altayar MA, Kiernan M, et al. Can next-generation DNA sequencing solve the Rh complexity for genotyping? [P-440 abstract]. Vox Sang 2014; 107:196.
12. Fichou Y, Audrezet MP, Guéguen P, et al. Next generation sequencing (NGS) for blood group genotyping. Vox Sang 2013;105:59.
13. 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.
14. Immucor PreciseType™ HEA molecular beadchip test. Norcross (GA): Immucor; 2014.
15. Immucor. Immucor RHD: Molecular BeadChip test [package insert]. Warren: BioArray Solutions Ltd; 2015.
16. Red cell immunogenetics and blood group terminology [Internet]. Amsterdam: International Society of Blood Transfusion (ISBT); [cited 2017 Feb 21]. Available from: http://www.isbtweb.org/working-parties/red-cell-immunogenetics-and-blood-group-terminology/.
17. 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.
18. Daniels G. Human blood groups. 3rd ed. Oxford: Wiley-Blackwell Publishing; 2013.
19. Patnaik SK, Helmberg W, Blumenfeld OO. BGMUT: NCBI dbRBC database of allelic variations of genes encoding antigens of blood group systems. Nucleic Acids Res 2012;40: D1023-9.
20. Wagner FF, Flegel WA. The human RhesusBase, version 2.0. http://www.rhesusbase.info/. Last updated 2017 Jun 14. The Rhesus Site. Transfus Med Hemother 2014;41:357-63.
21. Reid ME, Lomas-Francis C, Olsson ML. The blood group antigen facts book. 3rd ed. San Diego: Elsevier Academic Press; 2012.
22. Feng Y, Chen D, Wang GL, et al. Improved molecular diagnosis by the detection of exonic deletions with target gene capture and deep sequencing. Genet Med 2015;17:99-107.
23. Baronas J, Westhoff CM, Vege S, et al. RHD zygosity determination from whole genome sequencing data. J Blood Disorders Transfus 2016;7:365.
24. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res 2001;11:863-74.
25. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.
26. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308-11.
27. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536: 285-91.
28. Wagner FF, Frohmajer A, Ladewig B, et al. Weak D alleles express distinct phenotypes. Blood 2000;95:2699-708.
29. Daniels G. Variants of RhD—current testing and clinical consequences. Br J Haematol 2013;161:461-70.
30. McGowan EC, Lopez GH, Knauth CM, et al. Diverse and novel RHD variants in Australian blood donors with a weak

D phenotype: implication for transfusion management. Vox Sang 2017;112:279-87.
31. Bieker JJ. Krüppel-like factors: three fingers in many pies. J Biol Chem 2001;276:34355-8.
32. Arndt PA, Horn T, Keller JA, et al. First example of an FY*01 allele associated with weakened expression of Fya on red blood cells. Immunohematology 2015;31:103-7.
33. Lopez GH, Condon JA, Wilson B, et al. A novel FY*A allele with the 265T and 298A SNPs formerly associated exclusively with the FY*B allele and weak Fy(b) antigen expression: implication for genotyping interpretative algorithms. Vox Sang 2015;108:52-7.
34. Tournamille C, Le Van Kim C, Gane P, et al. Arg89Cys substitution results in very low membrane expression of the Duffy antigen/receptor for chemokines in Fy(x) individuals. Blood 1998;92:2147-56.
35. McBean RS, Hyland CA, Hendry JL, et al. SARA: a "new" low-frequency MNS antigen (MNS47) provides further evidence of the extreme diversity of the MNS blood group system. Transfusion 2015;55:1451-6.
36. Cvejic A, Haer-Wigman L, Stephens JC, et al. SMIM1 underlies the Vel blood group and influences red blood cell traits. Nat Genet 2013;45:542.
37. 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.
38. The Australian & New Zealand Society of Blood Transfusion Ltd, The Royal Australian and New Zealand College of Obstetricians and Gynaecologists. Guidelines for blood grouping & antibody screening in the antenatal & perinatal setting. Sydney: Australian & New Zealand Society of Blood Transfusion Ltd; 2007.
39. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 2013;15:565-74.

## SUPPORTING INFORMATION

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

**Table S1.** The average sequencing read depth of coverage obtained for the eight reported samples across the 41 genes related to 33 blood group systems and two transcription factor genes that are incorporated into the design of the TruSight One Sequencing (TSO) Panel.
**Table S2.** Overview of 22 samples: confirming and/or resolving Red Cell Reference Laboratory investigations by MPS analysis.
**Table S3.** Reaction Profile for Samples 3 and 4 using the ALBAclone Advanced Partial RhD Typing Kit compared to a control weak D type 33.*