



HAL
open science

Missense RHD SNVs induce weakened D antigen expression by altering splicing and/or protein expression

Loann Raud, Marlène Le Tertre, Léonie Vigneron, Chandran Ka, Gaëlle Richard, Isabelle Callebaut, Jian-Min Chen, Claude Férec, Gérald Le Gac, Yann Fichou

► To cite this version:

Loann Raud, Marlène Le Tertre, Léonie Vigneron, Chandran Ka, Gaëlle Richard, et al.. Missense RHD SNVs induce weakened D antigen expression by altering splicing and/or protein expression. *Transfusion*, 2021, 61 (8), pp.2468-2476. 10.1111/trf.16538 . hal-03329916

HAL Id: hal-03329916

<https://hal.sorbonne-universite.fr/hal-03329916>

Submitted on 31 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Missense *RHD* SNVs induce weakened D antigen expression by altering splicing and/or protein expression

Loann Raud^{1,2}, Marlène Le Tertre¹⁻³, Léonie Vigneron¹, Chandran Ka¹⁻³, Gaëlle Richard^{1,2},
Isabelle Callebaut⁴, Jian-Min Chen^{1,2}, Claude Férec¹⁻³, Gérald Le Gac¹⁻³, Yann Fichou^{1,2}

¹Univ Brest, Inserm, EFS, UMR1078, GGB, Brest, France

²Laboratory of Excellence GR-Ex, Paris, France

³Service de Génétique Médicale, CHRU Brest, Brest, France

⁴Sorbonne Université, Muséum National d'Histoire Naturelle, UMR CNRS 7590, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie (IMPMC), Paris, France

Running title: Missense *RHD* SNVs and D antigen expression.

Correspondence

Dr. Yann Fichou, UMR1078 Génétique, Génomique fonctionnelle et Biotechnologies (Etablissement Français du Sang – Bretagne/Inserm/UBO), Faculté de Médecine, 22 avenue Camille Desmoulins, 29238 Brest Cedex 3, France.

E-mail: Yann.Fichou@efs.sante.fr.

Sources of support

This work was supported by the Région Bretagne, the Association Gaétan Saleün – Don du Sang et Recherche en Génétique, the Etablissement Français du Sang (EFS) – Bretagne, and the Institut National de la Santé et de la Recherche Médicale (Inserm).

Word count: 3,076

Number of tables: 0

Number of figures: 4

References: 50

Conflict of interest

The authors declare no potential conflict of interest.

ABSTRACT

Background: Although D variant phenotype is known to be due to genetic defects, including rare missense single nucleotide variations (SNVs), within the *RHD* gene, few studies have addressed the molecular and cellular mechanisms driving this altered expression. We and others showed previously that splicing is commonly disrupted by SNVs in constitutive splice sites and their vicinity. We thus sought to investigate whether rare missense SNVs located in ‘deep’ exonic regions could also impair this mechanism.

Study Design and Methods: Forty-six missense SNVs reported within exons 6 and 7 were first selected from the Human RhesusBase. Their respective effect on splicing was assessed by using an *in vitro* assay. An RhD-negative cell model was further generated by using the CRISPR-Cas9 approach. RhD mutated proteins were overexpressed in the newly created model and cell membrane expression of the D antigen was measured by flow cytometry.

Results: Minigene splicing assay showed that 14/46 (30.4%) missense SNVs alter splicing. Very interestingly, further investigation of two missense SNVs, which both affect codon 338 and confer a weak D phenotype, showed various mechanisms: c.1012C>G (p.Leu338Val) disrupts splicing only, while c.1013T>C (p.Leu338Pro) alters only the protein structure, in agreement with *in silico* prediction tools and 3D protein structure visualization.

Conclusion: Our functional dataset suggests that missense SNVs damage quantitatively D antigen expression by, at least, two different mechanisms (splicing alteration and protein destabilization) that may act independently. These data thereby contribute to extend the current knowledge of the molecular mechanisms governing weakened D expression.

1. INTRODUCTION

In the Rh blood group system, D antigen is carried by the transmembrane RhD protein encoded by the *RHD* gene.¹⁻³ Rh D typing is routinely performed by serological testing with monoclonal antibodies, resulting typically in the identification of either a D-positive (D+) or D-negative (D-) phenotype, and indicating respectively the presence or absence of the antigen at the surface of red blood cells (RBCs).⁴ Beside these 'common' phenotypes, 'D variant phenotypes', characterized by a discrepancy or ambiguity in the standard typing procedure, are observed in ~1% of the general population,⁴ and are due to a quantitative and/or qualitative defect in the expression of the RhD protein. In clinics, a patient presenting with a quantitative defect expresses a normal D antigen, albeit at a lower density, and can be safely transfused with D+ RBC units. Conversely, an individual presenting with a qualitative defect exhibits an altered D antigen characterized by the absence of one or more D epitopes, and must be transfused with D- RBC units to prevent from alloimmunization. Although the definitions are not satisfactory, these individuals are typically referred to as 'weak D' and 'partial D', respectively.⁴⁻⁷

D variant phenotypes result from genetic alterations within the *RHD* gene. At the molecular level, it is considered that a missense single nucleotide variant (SNV) changing the nature of an amino acid predicted to be located in an extracellular domain of the protein confers a partial D phenotype by altering the nature of the exposed D epitope(s) to which it contributes. Conversely, changing an amino acid located in either a transmembrane helix or an intracellular loop of the RhD protein causes a weak D phenotype. In the latter situation, the missense change is thought to disrupt protein folding/stability, intra-/intermolecular interaction and/or integration within the RBC plasma membrane, inducing a significant decrease in Rh D antigen expression.⁴⁻⁹ Although valuable models depicting the three-dimensional (3D) structure of the protein within the RBC plasma membrane have been built on the basis of information from bacterial homologs,⁹⁻¹³ predicting the impact of a missense variant on the 3D structure, and thus

the potential alteration of antigen expression, remains complicated (and unusable in diagnostics and clinics in practice), especially as information about the folding process and the dynamical behavior of the protein is lacking.

Splicing is a key mechanism of the regulation of eukaryotic gene expression. In physiological (normal) conditions, alternative splicing, i.e. the balance between exon inclusion and exclusion resulting in various messenger RNA (mRNA) molecules transcribed from a single gene, is pivotal for guiding essential biological process as a function of the tissue and/or developmental stage and/or environment. Over the past years, additionally to the constitutive donor and acceptor splice sites that are both known to be critical for cellular splicing, several reports have noted the common disruption of *cis*-regulatory elements within exons, namely the exonic splicing regulatory elements (ESRs), in various genes.¹⁴⁻¹⁸ ESRs are typically 6- to 8-mers in size, and have either an activating or a repressive effect on the inclusion of exons due to their interaction with *trans*-regulatory factors.¹⁹ Indeed, a variation within an ESR can modify RNA splicing, thereby altering quantitatively and/or qualitatively the biosynthesis of the protein that may be potentially a key actor towards pathophysiological conditions.^{20,21}

In the context of the *RHD* gene, beyond the typical example of the Asian *DEL* allele,²²⁻²⁵ it has been demonstrated that synonymous SNVs can impair splicing both *in vivo* and *in vitro*.²⁶⁻²⁸ Additionally, evidence that the mechanism governing D phenotype alteration due to a missense SNV is not only position-specific, but also variant-specific, resulting in different phenotypes was provided.²⁸ On the basis of these observations, we wondered whether missense SNVs could alter Rh D antigen expression by impairing splicing only. Therefore, we 1/ investigated the effect of reported missense SNVs in a subset of *RHD* exons on splicing, and 2/ analyzed the relative expression of recombinant RhD variants in a D- cell model, in order to get insights into the respective effect of splicing and protein expression destabilization independently.

2. MATERIALS AND METHODS

2.1. Missense SNVs selected for analysis

All missense SNVs reported in *RHD* exons 6 and 7 were extracted from The Human RhesusBase.²⁹ All those located >10 nucleotides downstream and upstream the constitutive splice sites, respectively, were included for analysis (Table S1).

2.2. Minigene Splicing Assay

Minigene splicing assay in the pSplice*POLR2G.3* vector was carried out as previously described (supplemental Materials and Methods).²⁸ Splicing patterns were analyzed using Peak Scanner Software v1.0 (Thermo Fisher Scientific, Illkirch-Graffenstaden, France). The Percentage of exon Inclusion π was calculated for each sample. A ratio ρ ($\pi_{\text{SNV}}/\pi_{\text{wild-type}}$) was then calculated to compare directly the impact of all SNVs on exon inclusion. All experiments were performed independently at least three times.

2.3. Prediction of ESR disruption

A method based on the quantitative measure of the splicing impact of individual hexamer sequences obtained by deep sequencing functional analysis, i.e. ESRseq,³⁰ was used to predict the potential effect of SNVs on splicing, as previously described.¹⁶ For each studied SNV, ESRseq score was calculated and compared to the wild-type situation to generate the $\Delta\text{tESRseq}$ score. An SNV is predicted to promote exon skipping if the $\Delta\text{tESRseq}$ score is < -0.5 .³¹

2.4. Generation and characterization of the K562^{RHD}^{-/-} cells by the CRISPR-Cas9 approach

The K562^{RHD^{-/-}} cell line was generated by inactivating both *RHD* alleles in the K562 cell line following a protocol previously described (supplemental Materials and Methods; Tables S1 & S2; Figure S1).³²

Genomic disruption of the *RHD* gene structure and inactivation of *RHD* gene expression were confirmed by PCR, sequencing, semi-quantitative RT-PCR and flow cytometry (supplemental Materials and Methods).

2.5. Production of the recombinant expression vectors, transfection and flow cytometry

The whole *RHD* ORF (NM_016124) contained within a commercial plasmid (pCMV6, Clinisciences, Nanterre, France) was subcloned into a linearized Vector (pIRES2-DsRed2, Ozyme, Saint-Cyr-l'Ecole, France) to generate the newly created p*RHD*-IRES2-DsRed2 vector (supplemental Materials and Methods). Variants of interest were generated by site-directed mutagenesis using the latter vector as a template.

K562^{RHD^{-/-}} cells were transfected with wild-type and variant constructs mixed with the transfection reagent (Lipofectamine 2000, Invitrogen). After 48 h in culture, D antigen expression was assessed by flow cytometry using monoclonal antibodies (anti-D LHM76/55 and LHM169/80, Bio-Rad, Marnes-la-Coquette, France) (supplemental Materials and Methods).

2.6. Analysis of the amino acid substitution at the structure level

A human RhAG-RhD-RhD heterotrimer 3D structure model¹² was used to visualize the position of amino acids and to evaluate the possible impact of variations with UCSF Chimera software,³³ as previously described.²⁸

3. RESULTS

3.1. A significant number of ‘deep’ SNVs reduce exon inclusion

We showed previously that some synonymous SNVs in *RHD* exon 7 alter splicing by minigene splicing assay.^{25,28} Based on our experience, our study focused on missense SNVs located in exon 7 first, as well as those in exon 6 to extend the scale of the work. Also, in order to prevent from disturbing potentially the local environment of the constitutive splice sites, the selected missense SNVs were located >10 nucleotides from those sites. Thus, the functional effect of 46 natural SNVs (26 and 20 in exons 6 and 7, respectively) was assessed by minigene splicing assay. A variant exhibiting a decrease of >20% in the inclusion of the exon due to the tested SNV (i.e. $\rho < 0.80$), as previously reported,³¹ was considered as a variant altering splicing.

In our conditions, exon skipping was the sole mechanism observed in the presence of SNVs. As many as 14/46 (30.4%) missense SNVs were shown to be deleterious to the inclusion of exon within the mature transcript: 5/26 (19.2%) and 9/20 (45.0%) in exons 6 and 7, respectively (Figure 1A,B; Table S1). In particular, four SNVs located in exon 7 (c.1015G>A, c.1054G>A, c.1057G>A, and c.1058G>T) displayed a strong decrease (>60%) in exon inclusion. More interestingly, c.919G>A and c.919G>C in exon 6, which involve the same nucleotide position, were shown to behave differentially. Exon inclusion was impaired in the former ($\rho_{c.919G>A} = 0.543 \pm 0.064$, $p = 0.0060$), but not in the latter ($\rho_{c.919G>C} = 0.980 \pm 0.091$, $p = 0.7334$) (Table S1). Taken together, our data show that missense SNVs in ‘deep’ exonic regions of *RHD* can significantly impair exon inclusion in our *in vitro* model. This effect appears to be exon-dependent, because the deleterious impact observed in exon 6 was found to be milder than in exon 7 at both the qualitative and quantitative levels. It is worth mentioning that the apparent correlation between π and ρ values is purely coincidental, and does not reflect any bias in the assay conditions (personal unpublished data).

Our functional dataset was next compared to an *in silico* prediction tool.^{16,31} Overall, prediction of SNVs resulting in a significant alteration of splicing within *RHD* exons 6 and 7

by calculating the Δ tESRseq score correlates well with the functional data (Figure 1C,D: accuracy = 0.89; sensitivity = 0.86; specificity = 0.91).

3.2. Splicing defect due to ‘deep’ exonic SNVs can solely contribute to weak D phenotype

While the deleterious effect of several SNVs on splicing was clearly demonstrated, it remained unknown to what extent this mechanism is involved in the decrease of antigen expression at the surface of the cells. Therefore, we investigated the expression of the protein independently from the transcriptional process. To address this question, a cell model deficient in RhD, but expressing RhAG, which is required for cell surface expression of RhD,^{34,35} was engineered. The K562 cell model was selected and modified by inactivating both wild-type *RHD* alleles using the CRISPR-Cas9 approach (Figure 2A; Figure S1). Genomic characterization of the novel K562^{*RHD*^{-/-}} cell model showed that both alleles were altered: one by inclusion of the CMV-Zeocin sequence, and the other by a 2-bp deletion (Figure 2B,C). Complete absence in *RHD* gene expression was further confirmed by real-time RT-PCR and flow cytometry at the transcript and protein levels, respectively (Figure 2D,E).

We then sought to validate the cell model by overexpressing the wild-type and variant *RHD* ORFs subcloned individually into the pIRES2-DsRed2 vector. The c.1154G>A (p.Gly385Asp) and c.1154G>C (p.Gly385Ala) missense SNVs, which were both shown to alter splicing at a comparable level,²⁸ but are respectively associated with a D– and weak D phenotype, were the preferred option. Flow cytometry analysis with monoclonal antibodies exhibited a significant decrease in the fluorescence intensity of both variants at the plasma membrane when compared to the wild-type protein, albeit at various levels (Figure 3): residual expression for p.Gly385Asp; marked decrease for p.Gly385Ala. Our data suggest a cooperative effect of ‘splicing disruption’ previously shown and ‘protein alteration’ for both variants. These results show that assessment of D antigen expression by flow cytometry mimics the expression

of proteins reported *in vivo*, thus validating our model for studying variant RhD protein expression.

To gain insights into the respective involvement of both mechanisms on D antigen expression, we paid attention to a single codon position affected by two missense SNVs: c.1012C>G (p.Leu338Val) (GenBank Accession Number: FM212558) and c.1013T>C (p.Leu338Pro),³⁶ which are both associated with a weak D phenotype. While the former was previously shown to alter significantly inclusion of exon 7, the latter did not (Figure 1B). Both variant ORFs were then overexpressed in the K562^{RHD-/-} cell model to analyze experimentally the direct effect of amino acid change on fluorescence intensity. Flow cytometry analysis with monoclonal antibodies clearly showed that only p.Leu338Pro alters RhD protein expression at the membrane, while p.Leu338Val is expressed at a level similar to the wild-type protein (Figure 4A,B). Therefore, we conclude that although both missense SNVs affecting Leu338 confer a weak D phenotype, the molecular mechanisms are different: c.1012C>G (p.Leu338Val) only alters cellular splicing, while the amino acid change due to c.1013T>C (p.Leu338Pro) directly impairs protein expression. In the 3D structure model, RhD p.Leu338 is located in the middle of the transmembrane helix H10, oriented towards the lipid bilayer, in contact with helix H9. It interacts with the aliphatic chains of hydrophobic amino acids included in helices H9 (Val300 and Ile304) and H10 (Leu334 and Ile342). The p.Leu338Val mutant does not affect the hydrophobic character of the position and does not cause steric hindrance (Table S4). Interestingly, a valine is also observed at this position in *E. coli* AmtB, which has been used to model the RhD protein, in a region where the two sequences can be aligned with high accuracy.¹¹ This model thus suggests that p.Leu338Val is not deleterious to protein folding and/or intramolecular interaction, in agreement with a ‘wild-type-like’ expression of the mutant protein observed by functional analysis. In contrast, introduction of a proline in the helix at the

same position is predicted to disrupt its H-bond network and to destabilize the whole membrane domain architecture, in accordance with the observed alteration of protein expression.

4. DISCUSSION

4.1. Evolving the paradigm: missense SNVs alter quantitatively D antigen expression by different mechanisms independently

Disruption of constitutive splice sites is a common mechanism causing human genetic disorders. In the Human Gene Mutation Database (www.hgmd.cf.ac.uk/; accessed on October 20, 2020), ~9% of the mutations reported are splicing variants. Splicing is a complex process finely regulated – thus altered – by *cis*-acting elements and *trans*-regulatory factors. Although SNVs in the *RHD* gene do not cause disease *stricto sensu*, they can disrupt existing, create novel or activate cryptic splice sites, thereby altering quantitatively and/or qualitatively the expression of the D antigen.^{23-28,37-42} Also there are increasing experimental evidence showing that missense SNVs cause various diseases by altering the recognition of splicing regulatory elements by *trans*-regulatory factors.¹⁵⁻¹⁸ In this work, we investigated systematically the potential effect of deep missense SNVs in exons 6 and 7 of the *RHD* gene by minigene splicing assay and showed that these variants commonly impair exon inclusion. SNVs are thought to disrupt ESRs by altering the primary sequence, thereby decreasing the amount of full-length mRNA template available for subsequent translation of the RhD protein, and resulting in a weakened D antigen expression at the surface of RBCs. Therefore, our findings confirm our previous postulate that splicing disruption is a mechanism commonly involved in D variant phenotype.²⁸

Next we investigated the direct effect of amino acid substitution on D antigen expression independently from splicing alteration. Because fresh biological materials were not available, we thought to overexpress mutated *RHD* ORFs in a relevant cell model. Previously, Stegmann

and collaborators developed and implemented a consistent model based on the transduction of human D- erythroblasts by a lentiviral vector expressing the *RHD* gene.⁴³ Another strategy was chosen here. Both *RHD* alleles in the K562 cell line were disrupted by the CRISPR-Cas9 technology to generate the K562^{*RHD*^{-/-}} cell model. The newly created model can be transfected by the construct of interest, which is basically easier and faster to implement than Stegmann and colleagues' strategy, and thus constitutes a convenient alternative to study the expression of rare D variants.

Missense SNVs associated with both weak and partial D phenotypes were investigated in the study. It is important to make a distinction between the two situations. On the one hand, as indicated above, weak D variants typically involve changes in amino acids that are thought to impair protein folding.⁶ As illustrated by the c.1012C>G (p.Leu338Val) variant, it does not seem to be an exclusive mechanism. Indeed, our data suggest that the nucleotide substitution impairs exon inclusion, while amino acid change does not alter D antigen expression. Therefore, for the first time in the field, we conclude that missense SNVs may alter splicing only, resulting in a weak D phenotype. This mechanism may be used by other SNVs, which remain to be identified and characterized by functional analysis in both fresh materials and *in vitro* models.

On the other hand, partial D variants involve amino acids expressed at the surface of RBCs. Expression of some partial *RHD* alleles alters quantitatively D antigen density.⁴⁴ Intuitively, changing an amino acid in an extracellular loop is thought to damage the protein structure and/or its interaction with partners to a lesser extent than if substituting an amino acid in a transmembrane or an intracellular domain, suggesting an alternative mechanism resulting in a decrease in D antigen density in the former change. We showed that c.1057G>A and c.1063G>A, defining the partial *DNU* and *DNB* alleles respectively,⁴⁴⁻⁴⁶ significantly impair exon inclusion in our model, although they affect amino acids located in extracellular loops (Table S1: $\rho_{c.1057G>A (DNU)} = 0.3093 \pm 0.1776$; $\rho_{c.1063G>A (DNB)} = 0.5780 \pm 0.0289$). Interestingly,

D antigen density per RBC presenting with these alleles was reported previously to be 8073 and 5908, respectively.⁴⁶ Therefore, it appears that there is no clear correlation between ρ values due to missense SNVs in our model and antigen density. This observation suggests that 1/ mechanisms other than splicing, which remain to be precisely elucidated, contribute to the quantitative defect in D antigen expression, as expected; and 2/ the respective effect of the different mechanisms involved in antigen reduction remains difficult to evaluate by the current *in vitro* functional assays. It is also worthwhile to note, as a potential limitation of our system, that the use of monoclonal antibodies may not reflect alteration in protein expression but alteration of specific epitopes. It will be valuable to test polyclonal antisera in future experiments.

Taken together, our data generated in *in vitro* models demonstrate that the quantitative expression of the D antigen due to missense *RHD* SNVs may result from splicing alteration and/or protein destabilization, the former being likely secondary to the latter. In combination with clinical/phenotype information, our findings pave the way towards a larger study in order to help classifying rare missense SNVs according to their quantitative and/or qualitative impact on D antigen expression by functional analysis.

4.2. Relevance of *in silico* prediction and modeling tools for interpreting the role of rare variants in genetics

With the advent of next-generation sequencing, interpreting the pathogenicity of rare variants with unknown significance has become a critical issue in clinical genetics. In research, *in silico* prediction can be used as an informative filter to prioritize SNVs preliminary for functional tests. Many prediction tools have been developed for assessing the potential effect of SNVs on splicing.⁴⁷ As they are by definition ‘predictive’, their use for diagnostics in clinical genetics is still not recommended. Interestingly, our attempt to assess the potential deleterious effect of

SNVs on ESR disruption by the calculation of the Δ ESRseq score displayed correlation with the experimental data, further confirming experimentally the relevance of the approach.⁴⁸

Generating such novel datasets and increasing their number will undoubtedly contribute to refine algorithms and further improve the robustness as well as the accuracy of those tools.^{49,50}

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

FUNDING AND RESOURCES

This work was supported by the Région Bretagne, the Association Gaétan Saleün – Don du Sang et Recherche en Génétique, the Etablissement Français du Sang (EFS) – Bretagne, and the Institut National de la Santé et de la Recherche Médicale (Inserm).

The authors thank Dr. Laurent Corcos (UMR1078, GGB, Brest) for providing access to the flow cytometry facility.

ORCID

Loann Raud	https://orcid.org/0000-0002-7219-9886
Isabelle Callebaut	https://orcid.org/0000-0003-3124-887X
Jian-Min Chen	https://orcid.org/0000-0002-2424-3969
Claude Férec	https://orcid.org/0000-0002-2325-0710
Gérald Le Gac	https://orcid.org/0000-0003-3236-7280
Yann Fichou	https://orcid.org/0000-0002-5104-9125

REFERENCES

1. Avent ND, Ridgwell K, Tanner MJ, Anstee DJ. cDNA cloning of a 30 kDa erythrocyte membrane protein associated with Rh (Rhesus)-blood-group-antigen expression. *Biochem J.* 1990;271:821–825.

2. Chérif-Zahar B, Bloy C, Le Van Kim C, et al. Molecular cloning and protein structure of a human blood group Rh polypeptide. *Proc Natl Acad Sci U S A*. 1990;87:6243–6247.
3. Le van Kim C, Mouro I, Chérif-Zahar B, et al. Molecular cloning and primary structure of the human blood group RhD polypeptide. *Proc Natl Acad Sci U S A*. 1992;89:10925–10929.
4. Daniels G. Rh and RHAG blood group systems. In: Daniels G, editor. *Human Blood Groups*. 3rd ed. Chichester: Wiley-Blackwell; 2013. p. 182–258.
5. Flegel WA. The genetics of the Rhesus blood group system. *Blood Transfus*. 2007;5:50–57.
6. Flegel WA. Molecular genetics and clinical applications for RH. *Transfus Apheres Sci*. 2011;44:81–91.
7. Sandler SG, Chen LN, Flegel WA. Serological weak D phenotypes: a review and guidance for interpreting the RhD blood type using the *RHD* genotype. *Br J Haematol*. 2017;179:10–19.
8. Wagner FF, Gassner C, Müller TH, Schönitzer D, Schunter F, Flegel WA. Molecular basis of weak D phenotypes. *Blood*. 1999;93:385–393.
9. Srivastava K, Polin H, Sheldon SL, et al. The *DAU* cluster: a comparative analysis of 18 *RHD* alleles, some forming partial D antigens. *Transfusion*. 2016;56:2520–2531.
10. Conroy MJ, Bullough PA, Merrick M, Avent ND. Modelling the human rhesus proteins: implications for structure and function. *Br J Haematol*. 2005;131:543–551.
11. Callebaut I, Dulin F, Bertrand O, et al. Hydrophobic cluster analysis and modelling of the human Rh protein three-dimensional structures. *Transfus Clin Biol*. 2006;13:70–84.
12. Silvy, M., Chapel-Fernandes, S., Callebaut, I., et al. Characterization of novel *RHD* alleles: relationship between phenotype, genotype, and trimeric architecture. *Transfusion*. 2012;52:2020–2029.
13. de Brevern AG, Floch A, Barraud A, et al. Alloimmunization risk associated with amino acid 223 substitution in the RhD protein: analysis in the light of molecular modeling. *Transfusion*. 2018;58:2683–2692.
14. Pagani F, Buratti E, Stuaní C, Baralle FE. Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in Cystic Fibrosis Transmembrane Regulator exon 9. *J Biol Chem*. 2003;278:26580–26588.

15. Gaildrat P, Krieger S, Giacomo DD, et al. Multiple sequence variants of *BRCA2* exon 7 alter splicing regulation. *J Med Genet.* 2012;49:609–617.
16. Di Giacomo D, Gaildrat P, Abuli A, et al. Functional analysis of a large set of *BRCA2* exon 7 variants highlights the predictive value of hexamer scores in detecting alterations of exonic splicing regulatory elements. *Hum Mutat.* 2013;34:1547–1557.
17. Greenwald SH, Kuchenbecker JA, Rowlan JS, Neitz J, Neitz M. Role of a dual splicing and amino acid code in myopia, cone dysfunction and cone dystrophy associated with L/M opsin interchange mutations. *Transl Vis Sci Technol.* 2017;6:2.
18. Donadon I, McVey JH, Garagiola I, et al. Clustered *F8* missense mutations cause hemophilia A by combined alteration of splicing and protein biosynthesis and activity. *Haematologica.* 2018;103:344–350.
19. Matlin AJ, Clark F, Smith CWJ. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol.* 2005;6:386–398.
20. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA.* 2008;14:802–813.
21. Daguene E, Dujardin G, Valcarcel J. The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. *EMBO Rep.* 2015;16:1640–1655.
22. Wagner FF, Frohmajer A, Flegel WA. RHD positive haplotypes in D negative Europeans. *BMC Genet.* 2001;2:10.
23. Luettringhaus TA, Cho D, Ryang DW, Flegel WA. An easy *RHD* genotyping strategy for D– East Asian persons applied to Korean blood donors. *Transfusion.* 2006;46:2128–2137.
24. Shao CP, Xiong W, Zhou YY. Multiple isoforms excluding normal RhD mRNA detected in Rh blood group D_{e1} phenotype with *RHD* 1227A allele. *Transfus Apheres Sci.* 2006;34:145–152.
25. Chen DP, Sun CF, Ning HC, Wang WT, Tseng CP. Comprehensive analysis of *RHD* splicing transcripts reveals the molecular basis for the weak anti-D reactivity of D_{e1}-red blood cells. *Transfus Med.* 2016;26:123–129.

26. Fichou Y, Gehannin P, Corre M, et al. Extensive functional analyses of *RHD* splice site variants: insights into the potential role of splicing in the physiology of Rh. *Transfusion*. 2015;55:1432–1443.
27. Ogasawara, K., Sasaki, K., Isa, K., et al. Weak D alleles in Japanese: a c.960G>A silent mutation in exon 7 of the *RHD* gene that affects D expression. *Vox Sang*. 2016;110:179–184.
28. Raud L, Ka C, Gourlaouen I, et al. Functional analysis of novel *RHD* variants: splicing disruption is likely to be a common mechanism of variant D phenotype. *Transfusion*. 2019;59:1367–1375.
29. Wagner FF, Flegel WA. The Rhesus Site. *Transfus Med Hemother*. 2014;41:357–363.
30. Ke S, Shang S, Kalachikov SM, et al. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res*. 2011;21:1360–1374.
31. Soukarieh O, Gaildrat P, Hamieh M, et al. Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using *in silico* tools. *PLoS Genet*. 2016;12:e1005756.
32. Zhou Y, Zhang H, Wei W. Simultaneous generation of multi-gene knockouts in human cells. *FEBS Lett*. 2016;590:4343–4353.
33. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605–1612.
34. Cherif-Zahar B, Raynal V, Gane P, et al. Candidate gene acting as a suppressor of the *RH* locus in most cases of Rh-deficiency. *Nat Genet*. 1996;12:168–173.
35. Mouro-Chanteloup, I., D'Ambrosio, A.M., Gane, P., et al. Cell-surface expression of RhD blood group polypeptide is posttranscriptionally regulated by the RhAG glycoprotein. *Blood*. 2002;100:1038–1047.
36. Kamesaki T, Kumada M, Omi T, et al. A novel mutation in the *RHD* gene in Japanese individuals with weak D, encoding an amino acid change in the 11th transmembranous domain of the RhD protein. *Vox Sang*. 2003;84:141.
37. Ye LY, Guo ZH, Li Q., Zhu ZY. Molecular and family analyses revealed two novel *RHD* alleles in a survey of a Chinese RhD-negative population. *Vox Sang*. 2007;92:242–246.
38. Liu HC, Eng HL, Yang YF, et al. Aberrant RNA splicing in *RHD* 7-9 exons of DEL individuals in Taiwan: a mechanism study. *Biochim Biophys Acta*. 2010;1800:565–573.

39. Silvy M, Simon S, Gouvitsos J, et al. *Weak D* and *DEL* alleles detected by routine SNaPshot genotyping: identification of four novel *RHD* alleles. *Transfusion*. 2011;51:401–411.
40. Fichou Y, Le Maréchal C, Jamet D, et al. Establishment of a medium-throughput approach for the genotyping of *RHD* variants and report of nine novel rare alleles. *Transfusion*. 2013;53:1821–1828.
41. Ye L, He Y, Gao H, Guo Z, Zhu Z. Weak D phenotypes caused by intronic mutations in the *RHD* gene: four novel weak D alleles identified in the Chinese population. *Transfusion*. 2013;53:1829–1833.
42. Chun S, Yun JW, Park G, Cho D. The synonymous nucleotide substitution *RHD* 1056C>G alters mRNA splicing associated with serologically weak D phenotype. *J Clin Lab Anal*. 2018;32:e22330.
43. Stegmann TC, Veldhuisen B, Bijman R, et al. Frequency and characterization of known and novel *RHD* variant alleles in 37 782 Dutch D-negative pregnant women. *Br J Haematol*. 2016;173:469–479.
44. Wagner FF, Frohmajer A, Ladewig B, et al. Weak D alleles express distinct phenotypes. *Blood*. 2000;95:2699–2708.
45. Avent ND, Jones JW, Liu W, et al. Molecular basis of the D variant phenotypes DNU and DII allows localization of critical amino acids required for expression of Rh D epitopes epD3, 4 and 9 to the sixth external domain of the Rh D protein. *Br J Haematol*. 1997;97:366–371.
46. Wagner FF, Eicher NI, Jorgensen JR, Lonicer CB, Flegel WA. DNB: a partial D with anti-D frequent in Central Europe. *Blood*. 2002;100:2253–2256.
47. Abramowicz A, Gos M. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet*. 2018;59:253–268.
48. Grodecká P, Buratti E, Freiburger T. Mutations of pre-mRNA splicing regulatory elements: are predictions moving forward to clinical diagnostics? *Int J Mol Sci*. 2018;18:1668.
49. Leman R, Gaildrat P, Le Gac G, et al. Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined *in silico/in vitro* studies: an international collaborative effort. *Nucleic Acids Res*. 2018;46:7913–7923.

50. Leman R, Tubeuf H, Raad S, et al. Assessment of branch point prediction tools to predict physiological branch points and their alteration by variants. *BMC Genomics*. 2020;21:86.

FIGURE LEGENDS

FIGURE 1 Experimental and computational assessment of the inclusion of *RHD* exons 6 and 7 in the presence of missense SNVs. Functional analysis of the effect of SNVs by minigene splicing assay in exons 6 (A) and 7 (B) (Table S3: $N \geq 3$; black: wild-type; grey: $\rho < 0.8$; white: $\rho > 0.8$; error bars: standard deviation (SD)). The exonic splicing regulator sequence (Δt ESRseq) score calculated for each SNV in exons 6 (C) and 7 (D) was plotted vs the percentage of exon inclusion experimentally obtained by minigene splicing assay.

FIGURE 2 Generation and characterization of the K562^{RHD^{-/-}} cell model by the CRISPR-Cas9 strategy. (A) Schematic representation of the strategy used to generate the K562^{RHD^{-/-}} cell model (see supplemental Materials and Methods for detailed protocol). sgRNA: single guide RNA; LDT: linear donor template (Figure S1); DSB: double strand break; NHEJ: non-homologous end joining; HRi-KI: homologous recombination-independent knock-in. (B) Genomic DNA characterization by PCR using either primer pair 1-3 or 1-2 (right panel; arrows: relative position of the primers). N: native K562; C: CRISPR-Cas9-engineered K562; *, †, and ‡: PCR products visualized on a 2% agarose gel (ladder in base pairs (bp)). (C) Sequences of the PCR products obtained in (B). Underlined nucleotides: sgRNA binding sequence; nucleotides in bold: PAM sequence; -: deleted nucleotides. (D) Relative expression of *RHD* mRNA in the native K562 and K562^{RHD^{-/-}} cells by qRT-PCR ($N = 3$; internal control: *ACTB*). (E) Flow cytometry analysis of Rh D antigen expression with LHM169/80 monoclonal anti-D. Unstained = stained with secondary antibody reagent only.

FIGURE 3 Experimental analysis of RhD cell surface intensity by flow cytometry with two monoclonal antibodies. (A) Overlay plots of the fluorescence intensity (light grey: mock; dark grey: construct of interest indicated above the figure). (B) Normalized relative fluorescence intensity (RFI) of the D antigen (unpaired *t*-test: $N = 3$; error bars: \pm SD; *: $p < 0.01$; **: $p < 0.001$).

FIGURE 4 Experimental analysis and predictive assessment of the effect of amino acid change on RhD protein expression. Flow cytometry analysis with two monoclonal antibodies showing (A) overlay plots of the fluorescence intensity (light grey: mock; dark grey: construct of interest indicated above the figure) and (B) normalized relative fluorescence intensity (RFI) of the D antigen (unpaired *t*-test: N = 3; error bars: \pm SD; *: $p < 0.01$). (C) Visualization of the respective amino acid lateral chain of at position 338 (Leu (left) vs Val (right)) in the RhAG-RhD-RhD heterotrimer model¹² with UCSF Chimera (<https://www.cgl.ucsf.edu/chimera/>).

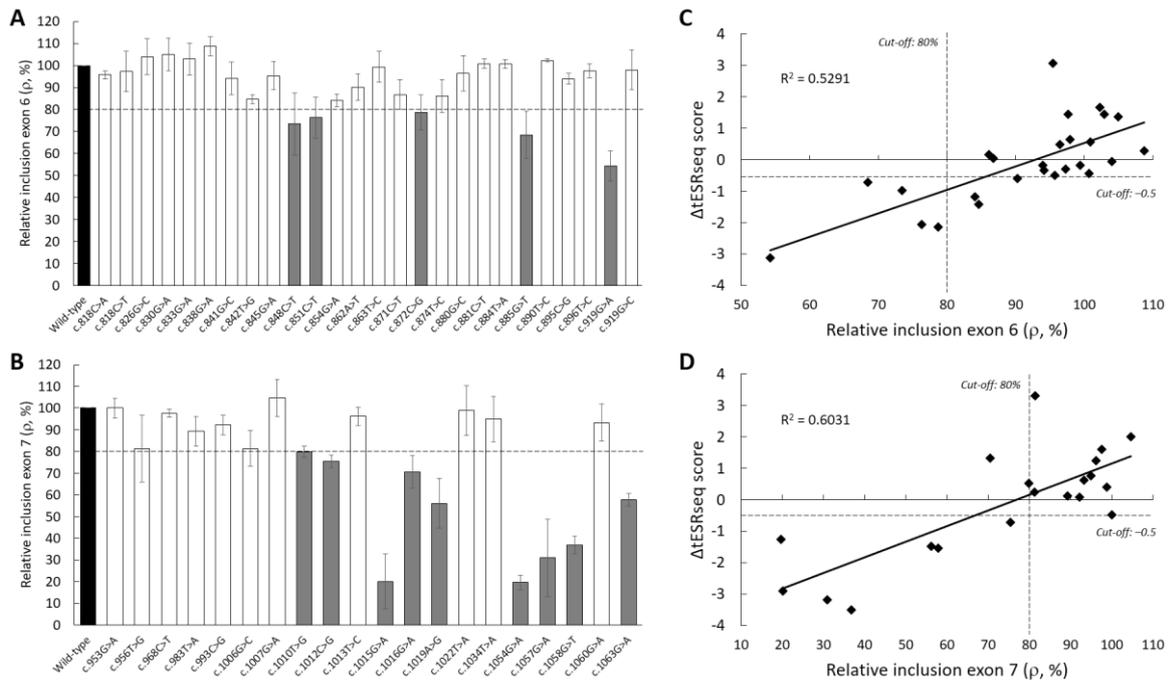


Figure 1

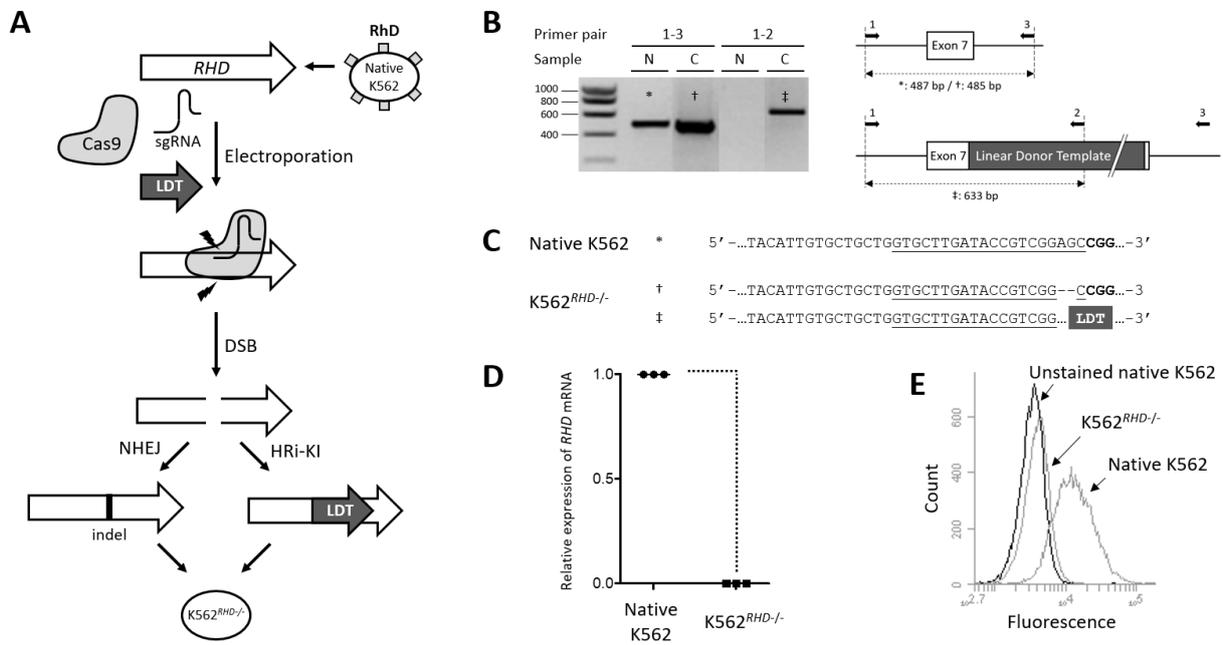


Figure 2

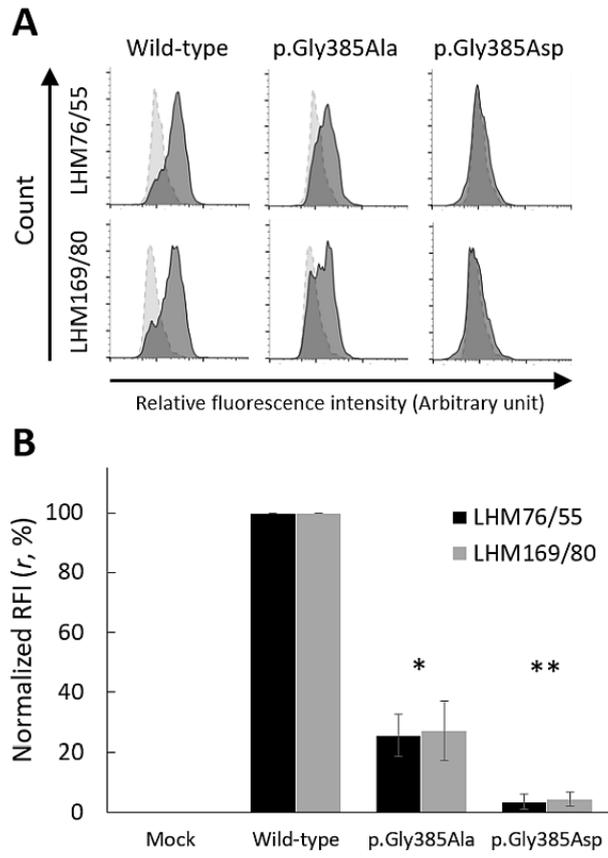


Figure 3

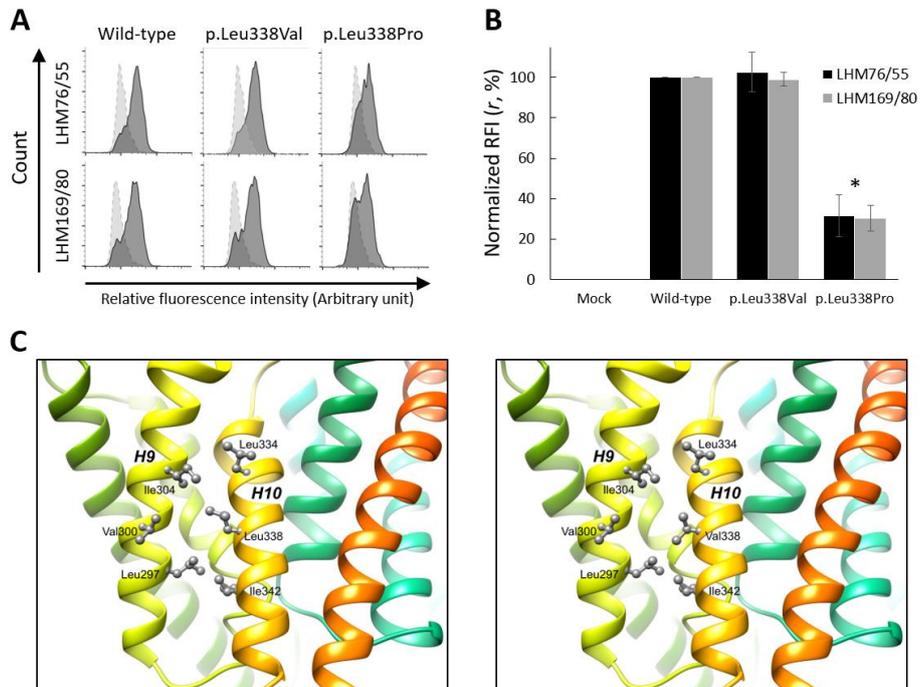


Figure 4

SUPPLEMENTAL DATA

1. SUPPLEMENTAL MATERIALS AND METHODS

1.1. Minigene Splicing Assay

Minigene splicing assay was carried out to analyze the effect of 46 natural missense variants in the *RHD* gene: 26 and 20 variants in exons 6 and 7, respectively. All minigene constructs and site-directed mutagenesis procedures were performed by using the pSplice*POLR2G.3* vector as a template, as previously described.¹ Wild-type and variant minigene constructs were transfected (1 µg/well) separately in K562 cells with a transfection reagent (Lipofectamine 2000, Invitrogen, purchased from Thermo Fisher Scientific, Illkirch-Graffenstaden, France). Total RNA extraction and fluorescent RT-PCR conditions are as previously described,¹ except 30 rounds of PCR cycles. Two microliters of the fluorescent RT-PCR products were mixed with 0.25 µL of an internal size standard (GeneScan 500 ROX, Applied Biosystems, purchased from Thermo Fisher Scientific) and 9.25 µL deionized formamide (Hi-Di Formamide Applied Biosystems), denatured at 95°C for 5 min, and size-separated by capillary electrophoresis (ABI3130xl, Applied Biosystems).

Splicing patterns were analyzed individually using Peak Scanner Software v1.0 (Applied Biosystems). The Percentage of exon Inclusion π was calculated as follow:

$$\pi = [\text{Area of exon inclusion} / (\text{Area of exon exclusion} + \text{Area of exon inclusion}) \times 100],$$

‘Area’ referring to the area under individual peaks in the fragment analysis profile.

A ratio ρ (expressed in percentage, %) was calculated to compare directly the impact of all SNVs on exon inclusion: $\rho = \pi_{\text{SNV}}/\pi_{\text{wild-type}}$. All experiments were carried out independently a least three times. A paired *t*-test was carried out to assess the statistical significance by calculating the *p*-value ($p < 0.05$ was considered as statistically significant).

1.2. Prediction of exonic splicing regulatory element (ESR) disruption

ESRseq prediction tool was used to predict the potential effect of SNVs on splicing, as previously described.² The method is based on the quantitative evaluation of all RNA hexamers as exonic splicing regulatory elements (ESRs) by a high-throughput minigene splicing assay approach.³ For each SNV in both *RHD* exons 6 and 7, ESRseq score was calculated and compared to the wild-type situation to generate the Δ tESRseq score.² The SNV of interest is predicted to promote exon skipping if the Δ tESRseq score < -0.5 .⁴ Accuracy, sensitivity and specificity were subsequently calculated by considering the following criteria:

- True positive (TP): Δ tESRseq score < -0.5 ; $\rho < 0.8$;
- True negative (TN): Δ tESRseq score > -0.5 ; $\rho > 0.8$;
- False positive (FP): Δ tESRseq score < -0.5 ; $\rho > 0.8$;
- False negative (FN): Δ tESRseq score < -0.5 ; $\rho < 0.8$.

1.3. sgRNA design and synthesis

An RNA guide design software (Guide Picker, Deskgen)⁵ was used to identify optimal single guide RNAs (sgRNA) (Table S2).^{6,7} The selected sgRNA (D7A) specifically targets exon 7 of the *RHD* gene and was designed to include a 5'-T7 promoter and a 3'-overlap/RNA scaffold (5'-TTCTAATACGACTCACTATA-(N)₂₀-GTTTTAGAGCTAGA-3'). sgRNA was synthesized by using a commercial kit (EnGen sgRNA Synthesis kit, New England Biolabs, Evry, France), following the manufacturer recommendations. Two transcription products were pooled together and purified (RNA Clean & Concentrator-25 kit, Zymo Research, purchased from Ozyme, Saint-Cyr-l'Ecole, France) according to the manufacturer's instructions.

1.4. Linear donor construction

The procedure followed for linear donor construction is illustrated in Figure S1. The pcDNA3.1(+) backbone vector (Invitrogen) was digested by the PmeI and NheI restriction endonucleases (New England Biolabs) at 37°C, overnight, and gel-purified (NucleoSpin Gel and PCR Clean-up, Macherey-Nagel, Hoerd, France). The Zeocin-resistant gene was PCR-amplified (HotStarTaq Master Mix kit, Qiagen, Courtaboeuf, France) with primers Zeocin_NheI_F and Zeocin_PmeI_R (Table S3) in a 25 µL final volume using 10 ng of a commercial plasmid (pcDNA4/TO, Invitrogen) as a template. PCR conditions were as follow: PCR enzyme activation step at 95°C for 15 min, followed by 40 cycles of denaturation at 94°C for 30 sec, annealing at 59°C for 30 sec, extension at 72°C for 1 min and a final extension at 72°C for 10 min. PCR product was gel-purified (NucleoSpin Gel and PCR Clean-up, Macherey-Nagel) and mixed with the linearized vector (molar ratio 1:1) for subcloning (In-Fusion HD Cloning Kit, Clontech, purchased from Ozyme) in the manufacturer's conditions to generate the donor vector (pDoZ). The plasmid was extracted, purified (NucleoSpin Plasmid Miniprep kit, Macherey-Nagel) and characterized by direct sequencing (BigDye Terminator v1.1 Cycle Sequencing Kit, Applied Biosystems).

The pDoZ vector was digested by MluI restriction endonuclease (New England Biolabs) at 37°C, overnight, and gel-purified (NucleoSpin Gel and PCR Clean-up, Macherey-Nagel). To generate the extending cutting site, twenty-five nanograms of genomic DNA extracted from the K562 cell line served as a template for PCR amplification (HotStarTaq Master Mix Kit, Qiagen) with 10 µM of the specific primers containing the extended sgRNA cutting site (eCS) (Table S2) by the in a 25 µL final volume. PCR conditions were as follow: PCR enzyme activation step at 95°C for 15 min, followed by 40 cycles of denaturation at 94°C for 30 sec, annealing at 56°C for 30 sec, extension at 72°C for 1 min 30 sec; and a final extension at 72°C for 10 min. PCR product was pre-treated enzymatically (Cloning Enhancer reagent, Clontech) and mixed with the linearized vector (molar ratio 3:1) (In-Fusion HD Cloning Kit, Clontech) following the

manufacturer's recommendations to generate the pDoZ-eCS vector. Next, plasmid was extracted, purified (NucleoSpin Plasmid Miniprep kit, Macherey-Nagel) and characterized by direct sequencing (BigDye Terminator v1.1 Cycle Sequencing Kit, Applied Biosystems).

Ten nanograms of the pDoZ-eCS plasmid served as a template for PCR amplification (HotStarTaq Master Mix Kit, Qiagen) of the linear donor template in a 50 μ L final volume with the respective specific forward and reverse primers containing protective sequences (Table S3), in the conditions described above. Five reactions were pooled together and purified (DNA Clean & Concentrator-25 kit, Zymo Research) according to the manufacturer's instructions.

1.5. Generation of K562^{RHD-/-} cells and clonal selection

The selected D7A sgRNA (Table S1) and the RNA-guided endonuclease (EnGen® Cas9 NLS, *S. pyogenes*, New England Biolabs) were combined at an equimolar ratio (120 pmol each) and incubate for 15 min at room temperature. The sgRNA/Cas9 complex was mixed with 1 μ g of the linear donor template. The mixture was incubated for 5 min at room temperature, mixed with K562 cells (1×10^6) suspended in 100 μ l of supplemented solution (Nucleofector Solution V, Lonza, Basel, Switzerland), electroporated (Amaxa Nucleofector II Device using program T-016, Lonza) in the manufacturer's conditions and directly seeded into 6-well plate in 2 mL of IMDM (Lonza) supplemented with 10% FBS (Lonza). After three days in culture, K562 cells were treated with a selective medium consisting of 0.5 mg/ml Zeocin (Thermo Scientific) in IMDM supplemented with 10% FBS. After other ten days in culture with the selective medium, 400 cells were plated into semi-solid media (MethoCult H4100, Stemcell Technologies, Grenoble, France) in 6-well plates to isolate single clones. After two weeks in culture, each clone was individually selected and cultured in the selective medium.

1.6. Clonal characterization

Genomic DNA was extracted (QIAamp DNA Investigator Kit, Qiagen). Fifty nanograms of genomic DNA was PCR-amplified (HotStarTaq Master Mix Kit, Qiagen) in a 10 μ L final volume with primer pairs 1-2 and 1-3 (Table S2, Fig. 2B). PCR conditions were as follow: PCR enzyme activation step at 95°C for 15 min, followed by 40 cycles of denaturation at 94°C for 30 sec, annealing at 60°C for 30 sec, extension at 72°C for 1 min 30 sec, and a final extension at 72°C for 10 min. PCR products were loaded on a 2% gel agarose, gel-purified (NucleoSpin Gel and PCR Clean-up, Macherey-Nagel), subcloned into the vector (pCR4-TOPO vector with the TOPO-TA Cloning Kit for Sequencing, Invitrogen) and characterized by direct sequencing (BigDye Terminator v1.1 Cycle Sequencing Kit, Applied Biosystems).

Total RNA was extracted (RNeasy Mini Kit, Qiagen). One microgram of total RNA was treated by DNase I amplification grade (Thermo Scientific) and served as a template for reverse transcription (SuperScript III Reverse Transcriptase, Invitrogen) according to the manufacturer's instructions. Next, the neosynthesized cDNA (1 μ L of a 1:10 dilution) was mixed with 1X SYBR Green Master Mix (Qiagen) and 10 μ M of either the specific *RHD* primers (forward: 5'-GAGCCAAGTACCTGCCGGG-3', reverse: 5'-ATCATGCCATTGCCGGCT-3') or the *ACTB* primers (forward: 5'-TCTCCATGTCGTCCCAGTTG-3', reverse: 5'-AGTCTTCCCCTCCATCGTTG-3') in a 10 μ L final volume. Real-time PCR was performed in the LightCycler II 480 (Roche). Real-time PCR cycling conditions were: PCR enzyme activation step at 95°C for 15 minutes; followed by 45 cycles of denaturation at 94°C for 15 sec, annealing at 60°C for 30 sec, and extension at 72°C for 30 sec. Fluorescence was collected at the extension step. Standard curves were generated and analyzed by the LightCycler 480 SW1.5 software (Roche).

Expression of the D antigen was assessed by flow cytometry also (see below).

1.7. Production of the recombinant expression vector, site-directed mutagenesis and transfection

The whole *RHD* ORF (NM_016124) subcloned within a commercial vector (pCMV6, Origene, purchased from Clinisciences, Nanterre, France) was PCR-amplified in a 25 μ L final volume with 1 U Platinum™ *Taq* DNA Polymerase High Fidelity, 1X High Fidelity Buffer, 2 mM $MgSO_4$, 0.2 mM dNTPmix, 0.4 μ M of both forward (5'-CTCAAGCTTCGAATTCATGAGCTCTAAGTACCCGCGG-3') and reverse (5'-GTCGACTCAGAATTCTTAAAATCCAACAGCCAAATGAGG-3') primers (complementary sequences for subcloning by homologous recombination are underlined), and 20 ng of the plasmid DNA as a template. PCR conditions were: PCR enzyme activation step at 94°C for 1 min; followed by 30 cycles of denaturation at 94°C for 15 sec, annealing at 55°C for 30 sec, extension at 68°C for 2 min.

The pIRES2-DsRed2 Vector (Clontech) was linearized by digesting the single *EcoRI* restriction site. *RHD* ORF was subcloned (In-Fusion HD Cloning Kit, Clontech) following the manufacturer's protocol, thereby generating the newly created p*RHD*-IRES2-DsRed2 vector. In this construct, the expression of both the RhD and DsRed2 fluorescent proteins is driven by a single promoter, i.e. the human cytomegalovirus (CMV) immediate early promoter $P_{CMV IE}$. Variants of interest were generated (QuikChange Lightning Site-Directed Mutagenesis Kit, Agilent Technologies, Les Ulis, France) following the supplier instructions. Direct sequencing was performed and visualized with Sequencher 5.0 for sequence confirmation of the variant p*RHD*-IRES2-DsRed2 vectors.

K562^{*RHD*^{-/-}} cells (5×10^5) were transiently transfected with 2.5 μ g wild-type or variant p*RHD*-IRES2-DsRed2 vector mixed with 7.5 μ l of the transfection reagent (Lipofectamine 2000, Invitrogen from Thermo Fisher Scientific) in 6-well plates. After 48 h in culture, D antigen expression was assessed by flow cytometry.

1.8. Flow cytometry

Rh D antigen expression at the surface of the cells was assessed by flow cytometry by following the recommendations of the 4th International Workshop on Monoclonal Antibodies against Human Red Blood Cells and Related Antigens protocol.⁸ Briefly, native or K562^{RHD-/-} cells (5×10^5) were resuspended in 75 μ l of PBS-EDTA/10% FBS and incubated with 25 μ l of human monoclonal anti-D (LHM76/55 or LHM169/80, Bio-Rad, Marnes-la-Coquette, France) for 30 min at 37°C. After extensive washes with PBS-EDTA, cells were incubated with diluted (1:200) Goat anti-Human IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 (Thermo Scientific) for 1 h at 4°C. After several washes, cells were resuspended in PBS-EDTA and fluorescence in 20,000 events was measured by flow cytometry on a BD Accuri C6 flow cytometer (BD Biosciences, Le-Pont-de-Claix, France). Data analysis was performed with BD Accuri C6 software (BD Biosciences) or FlowJo™ v10 software (BD Biosciences).

For clonal characterization of the K562^{RHD-/-} cells, cells were gated by using unstained (= stained with secondary antibody reagent only), native K562 cells as a RhD-negative (RhD–) control and stained, native K562 cells as a RhD-positive (RhD+) control.

For assessment of D antigen expression in K562^{RHD-/-} cells overexpressing variants, events were gated to select RhD+ cells (by excluding RhD– cells using mock-transfected K562^{RHD-/-} cells as a RhD– control) in DsRed2-positive (DsRed2+) cells (by excluding DsRed2-negative (DsRed2–) cells using unstained K562^{RHD-/-} cells as the DsRed2– control). The relative expression r of RhD variants is presented as the percentage of RhD+ cells over DsRed2+ cells, normalized with the wild-type construct:

$$r = \frac{(\text{RhD}+ / \text{DsRed}+)_{\text{SNV}} - (\text{RhD}+ / \text{DsRed}2+)_{\text{mock}}}{(\text{RhD}+ / \text{DsRed}+)_{\text{WT}} - (\text{RhD}+ / \text{DsRed}2+)_{\text{mock}}}$$

2. SUPPLEMENTAL TABLES

TABLE S1 Missense SNVs in *RHD* exons 6 and 7, and functional analysis by minigene splicing assay.

Exon	Nucleotide change	Amino acid change (location*)	N [†]	π (Mean (\pm SD), %) [‡]		ρ [§]	ISBT allele ID	Reference [¶]
				Wild-type	SNV			
6	c.818C>A	p.Ala273Glu (M8)	3	85.8 (\pm 3.1)	82.1 (\pm 4.3)	0.9575	<i>RHD*01W.120</i>	[9]
	c.818C>T	p.Ala273Val (M8)	3	83.4 (\pm 4.8)	81.2 (\pm 8.6)	0.9729	<i>RHD*01W.119</i>	FJ041201
	c.826G>C	p.Ala276Pro (M8)	3	85.8 (\pm 3.1)	89.2 (\pm 5.2)	1.0400	<i>RHD*01W.13</i>	[10]
	c.830G>A	p.Gly277Glu (M8)	3	85.8 (\pm 3.1)	90.0 (\pm 3.4)	1.0498	<i>RHD*01W.12</i>	[10]
	c.833G>A	p.Gly278Asp (M8)	3	85.8 (\pm 3.1)	88.3 (\pm 3.4)	1.0291	<i>RHD*01W.38</i>	[11]
	c.838G>A	p.Ala280Thr (M8)	4	84.0 (\pm 7.6)	91.3 (\pm 5.1)	1.0872	<i>RHD*01EL.24</i>	[12]
	c.841G>C	p.Val281Leu (ECL5)	4	84.0 (\pm 7.6)	79.1 (\pm 10.1)	0.9411	N/A	[13]
	c.842T>G	p.Val281Gly (ECL5)	3	76.8 (\pm 4.6)	65.0 (\pm 4.4)	0.8466	<i>RHD*01W.36</i>	AJ867387
	c.845G>A	p.Gly282Asp (ECL5)	3	76.8 (\pm 4.6)	73.2 (\pm 2.1)	0.9541	<i>RHD*15</i>	[10]
	c.848C>T	p.Thr283Ile (ECL5)	3	76.8 (\pm 4.6)	56.4 (\pm 12.0)	0.7343	<i>RHD*DHMi</i>	[14]
	c.851C>T	p.Ser284Leu (ECL5)	3	76.8 (\pm 4.6)	58.6 (\pm 6.1)	0.7627	<i>RHD*DLO</i>	[11]
	c.854G>A	p.Cys285Tyr (ECL5)	3	76.8 (\pm 4.6)	64.6 (\pm 4.6)	0.8409	<i>RHD*DIM</i>	[15]

	c.862A>T	p.Ile288Phe (ECL5)	3	76.8 (± 4.6)	69.3 (± 3.5)	0.9027	N/A	FR848376
	c.863T>C	p.Ile288Thr (ECL5)	4	84.0 (± 7.6)	83.5 (± 12.5)	0.9937	N/A	[16]
	c.871C>T	p.Pro291Ser (M9)	3	87.2 (± 11.8)	75.7 (± 11.5)	0.8678	<i>RHD*01W.138</i>	[13]
	c.872C>G	p.Pro291Arg (M9)	3	76.8 (± 4.6)	60.4 (± 8.7)	0.7870	<i>RHD*01EL.41</i>	HE999545
	c.874T>C	p.Trp292Arg (M9)	4	91.0 (± 8.4)	78.4 (± 10.4)	0.8612	<i>RHD*01W.113</i>	[17]
	c.880G>C	p.Ala294Pro (M9)	4	90.3 (± 7.7)	87.1 (± 7.4)	0.9642	<i>RHD*01W.9</i>	[10]
	c.881C>T	p.Ala294Val (M9)	3	89.2 (± 8.7)	89.9 (± 10.2)	1.0085	<i>RHD*01W.64</i>	AM902713
	c.884T>A	p.Met295Lys (M9)	3	93.5 (± 1.5)	94.2 (± 2.1)	1.0074	<i>RHD*01W.135</i>	KT819195
	c.885G>T	p.Met295Ile (M9)	3	76.8 (± 4.6)	52.6 (± 10.3)	0.6850	<i>RHD*11</i>	[18]
	c.890T>C	p.Leu297Pro (M9)	3	93.5 (± 1.5)	95.6 (± 0.9)	1.0230	N/A	KX352169
	c.895C>G	p.Leu299Val (M9)	3	93.5 (± 1.5)	87.9 (± 2.1)	0.9399	<i>RHD*01W.55</i>	AM746209
	c.896T>C	p.Leu299Pro (M9)	3	94.7 (± 2.5)	92.4 (± 4.1)	0.9761	<i>RHD*01N.79</i>	[19]
	c.919G>A	p.Gly307Arg (M9)	3	76.8 (± 4.6)	41.6 (± 6.4)	0.5425	<i>RHD*01W.8</i>	[10]
	c.919G>C	p.Gly307Arg (M9)	3	84.1 (± 5.1)	82.4 (± 5.1)	0.9795	N/A	KT894738
7	c.953G>A	p.Arg318Gln (M9)	4	78.4 (± 6.5)	78.5 (± 7.9)	1.0002	<i>RHD*01W.69</i>	FM212559
	c.956T>G	p.Val319Gly (M9)	6	80.5 (± 4.0)	65.4 (± 12.3)	0.8129	N/A	[20]
	c.968C>T	p.Pro323Leu (ICL6)	4	78.2 (± 2.5)	76.4 (± 3.9)	0.9764	<i>RHD*01W.114</i>	[21]
	c.983T>A	p.Met328Lys (M10)	4	78.2 (± 2.5)	69.8 (± 6.5)	0.8929	<i>RHD*01W.115</i>	[21]

c.993C>G	p.Asn331Lys (M10)	4	76.8 (± 4.6)	70.8 (± 5.6)	0.9215	<i>RHD*01W.90</i>	[9]
c.1006G>C	p.Gly336Arg (M10)	8	79.8 (± 5.2)	64.9 (± 7.0)	0.8139	<i>RHD*01W.58</i>	[22]
c.1007G>A	p.Gly336Asp (M10)	7	79.1 (± 5.2)	82.7 (± 8.0)	1.0463	<i>RHD*01N.80</i>	[19]
c.1010T>G	p.Leu337Arg (M10)	4	76.8 (± 4.6)	61.4 (± 4.3)	0.7990	<i>RHD*01EL.38</i>	[9]
c.1012C>G	p.Leu338Val (M10)	5	81.4 (± 3.8)	61.4 (± 0.8)	0.7545	<i>RHD*01W.70</i>	FM212558
c.1013T>C	p.Leu338Pro (M10)	7	79.1 (± 5.2)	74.6 (± 3.2)	0.9621	<i>RHD*01W.24</i>	[23]
c.1015G>A	p.Gly339Arg (M10)	5	79.1 (± 6.1)	15.9 (± 8.2)	0.2011	<i>RHD*01W.39</i>	[11]
c.1016G>A	p.Gly339Glu (M10)	3	76.3 (± 4.9)	53.8 (± 5.1)	0.7051	<i>RHD*01W.7</i>	[10]
c.1019A>G	p.Glu340Gly (M10)	4	76.5 (± 4.1)	42.9 (± 7.8)	0.5611	N/A	HG326208
c.1022T>A	p.Ile341Asn (M10)	3	75.2 (± 4.0)	74.3 (± 12.2)	0.9888	N/A	[24]
c.1034T>A	p.Val345Glu (M10)	4	74.1 (± 3.9)	70.4 (± 4.1)	0.9495	<i>RHD*01W.86</i>	KT377957
c.1054G>A	p.Val352Ile (ECL6)	3	70.8 (± 0.1)	13.9 (± 2.4)	0.1966	N/A	[25]
c.1057G>A	p.Gly353Arg (ECL6)	5	78.1 (± 5.1)	24.2 (± 12.2)	0.3093	<i>RHD*DNU</i>	[26]
c.1058G>T	p.Gly353Val (ECL6)	5	75.4 (± 4.4)	27.7 (± 2.0)	0.3679	N/A	[25]
c.1060G>A	p.Ala354Thr (ECL6)	5	73.4 (± 3.7)	68.5 (± 4.8)	0.9329	<i>RHD*50</i>	[9]
c.1063G>A	p.Gly355Ser (ECL6)	5	78.1 (± 5.1)	45.1 (± 1.8)	0.5780	<i>RHD*DNB</i>	[27]

SNVs (5/26 in exon 6; 9/20 in exon 7) showing a 20% decrease in the inclusion of the respective exon of interest ($\rho < 0.8$) are highlighted in grey.

TM = Transmembrane domain; EC = extracellular loop; IC = intracellular domain; N/A = not applicable.

* In accordance with [28].

† N: number of experiments (= pairs).

‡ Percentage of inclusion (π) of the exon of interest within the transcript.

§ Relative inclusion of the exon of interest with SNV *vs* wild-type (WT): $\rho = \pi_{\text{SNV}}/\pi_{\text{WT}}$.

|| In accordance with the Red Cell Immunogenetics and Blood Group Terminology Working Party (ISBT, URL: <http://www.isbtweb.org/working-parties/red-cell-immunogenetics-and-blood-group-terminology/>).

¶ GenBank Accession Number is provided when no reference is available.

TABLE S2 Best five off-target sgRNAs designed by Guide Picker (Deskgen).

sgRNA ID	Sequence (5'→3')	PAM	Cutting site*	Off-Target†	Activity‡	Location§	Mismatches
D7A	GTGCTTGATACCGTCGGAGC	CGG	25,306,714	100	56	Exon 7	6
D7B	GATACCGTCGGAGCCGGCAA	TGG	25,306,720	100	55	Exon 7	6
D7C	CGGAGCCGGCAATGGCATGT	GGG	25,306,728	99	63	Exon 7	4
D7D	CTGCTGGTGCTTGATACCGT	CGG	25,306,708	97	58	Exon 7	2
D3U	AAAGTCTCCAATGTTCGCGC	AGG	25,329,036	100	57	3'UTR	8

* Cutting site coordinates on chromosome 1 (GRCh38.81).

† Risk of mediating cutting at unintended sites in the genome (range 0-100; 100, lowest risk).⁶

‡ Ability to guide RNA to knock out the target gene (range 0-100; 100, highest predicted guide RNA activity).⁵

§ Location in the *RHD* gene (NM_016124.4).

|| Number of mismatches with *RHCE* (NM_020485.5; *RHCE*ce* allele).

TABLE S3 Primers for linear donor plasmid construct and clonal selection.

Primer ID	Sequence (5'→3')	Comments
Zeocin_NheI_F	ACCCAAGCTGGCTAGCATGGCCAAGTTGACCAGTGC	Insertion of Zeocin resistance gene
Zeocin_PmeI_R	ATCAGCGGGTTTAAACTCAGTCCTGCTCCTCGGC	
RHD_MluI_F	GGCCAGATATACGCGTGTGCTTGATACCGTCGGAG	Insertion of extended cutting site
RHD_MluI_R	ATCAATGTCAACGCGTCAGTGACCCACATGCCATT	
Donor_RHD_F	TCCACTGCGACGTCGCGAGTGTGCTTGATACCGTCGGAG	Linear donor template
Donor_R	GGCTTAGGATTGTTACGCCCCATAGAGCCCACCGCAT	
1	TTCATTTCAACAAACTCCCCGA	Clonal selection and sequencing
2	GGCTATGAACTAATGACCCCGTAA	
3	CCAAGGTAGGGGCTGGACAG	

Respective restriction sites are underlined. Protective sequences are in bold.

Primers 1 and 3 correspond to RHD_e7seqF and RHD_e7seqR, respectively.²⁵

TABLE S4 Valine Dunbrack rotamer library in the RhD p.Leu338Val variant.²⁹

Chi 1	Probability	Clashes
171.9	<u>0.939144</u>	0
68.0	0.041442	2
-61.0	0.019414	1

Chi 1 corresponds to the sidechain torsion angle. The highest probability (underlined) was considered to estimate the most likely number of clashes, i.e. 0.

3. SUPPLEMENTAL FIGURE

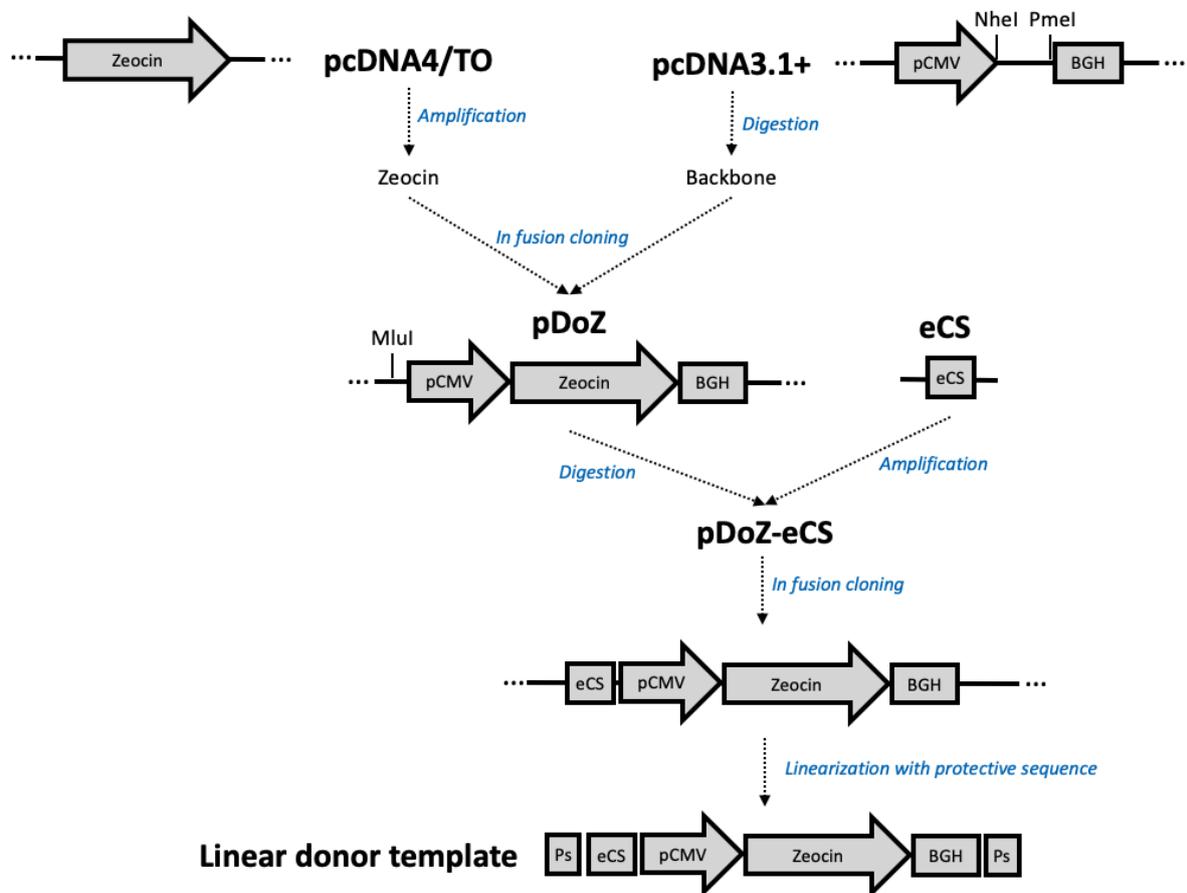


FIGURE S1 General strategy to generate the linear donor template (LDT) containing the Zeocin-resistant gene. eCS: extended sgRNA cutting site. Ps: Protective sequence.

SUPPLEMENTAL REFERENCES

1. Raud L, Ka C, Gourlaouen I, et al. Functional analysis of novel *RHD* variants: splicing disruption is likely to be a common mechanism of variant D phenotype. *Transfusion*. 2019;59:1367–1375.
2. Di Giacomo D, Gaildrat P, Abuli A, et al. Functional analysis of a large set of *BRCA2* exon 7 variants highlights the predictive value of hexamer scores in detecting alterations of exonic splicing regulatory elements. *Hum Mutat*. 2013;34:1547–1557.
3. Ke S, Shang S, Kalachikov SM, et al. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res*. 2011;21:1360–1374.
4. Soukarieh O, Gaildrat P, Hamieh M, et al. Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using *in silico* tools. *PLoS Genet*. 2016;12:e1005756.
5. Hough SH, Kancleris K, Brody L, et al. Guide Picker is a comprehensive design tool for visualizing and selecting guides for CRISPR experiments. *BMC Bioinformatics*. 2017;18:167.
6. Hsu PD, Scott DA, Weinstein JA, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013;31:827–832.
7. Doench JG, Fusi N, Sullender M, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*. 2016;34:184–191.
8. Flegel WA, Curin-Serbec V, Delamaire M, et al. Section 1B: Rh flow cytometry. Coordinator's report. Rhesus index and antigen density: an analysis of the reproducibility of flow cytometric determination. *Transfus Clin Biol*. 2002;9:33–42.
9. Garcia F, Rodriguez MA, Goldman M, et al. New *RHD* variant alleles. *Transfusion*. 2015;55:427–429.
10. Wagner FF, Gassner C, Müller TH, Schönitzer D, Schunter F, Flegel WA. Molecular basis of weak D phenotypes. *Blood*. 1999;93:385–393.
11. Ansart-Pirenne H, Asso-Bonnet M, Le Pennec PY, Roussel M, Patereau C, Noizat-Pirenne F. RhD variants in Caucasians: consequences for checking clinically relevant alleles. *Transfusion*. 2004;44:1282–1286.
12. Gu J, Wang XD, Shao CP, et al. Molecular basis of DEL phenotype in the Chinese population. *BMC Med Genet*. 2014;15:54.
13. Floch A, Barrault A, De Brevern AG, et al. Molecular characterization of 13 new *RHD* alleles. *Transfusion*. 2017;57:1089–1091.
14. Jones J. Identification of two new D variants, DHMi and DHMii using monoclonal anti-D. *Vox Sang*. 1995;69:236–241.

15. Wagner FF, Frohmajer A, Ladewig B, et al. Weak D alleles express distinct phenotypes. *Blood*. 2000;95:2699–2708.
16. Tilley L, Bullock T, Kingdom S, Poole J, Daniels G. A novel RhD variant associated with extremely low serological expression of D-antigen. *Transfus Med*. 2005;15(suppl.1):48–49.
17. Isa K, Sasaki K, Ogasawara K, et al. Prevalence of *RHD* alleles in Japanese individuals with weak D phenotype: identification of 20 new *RHD* alleles. *Vox Sang*. 2016;111:315–319.
18. Wagner FF, Frohmajer A, Flegel WA. RHD positive haplotypes in D negative Europeans. *BMC Genet*. 2001;2:10.
19. Krog GR, Clausen FB, Berkowicz A, et al. Is current serologic RhD typing of blood donors sufficient for avoiding immunization of recipients? *Transfusion*. 2011;51:2278–2285.
20. Vege S, Hong H, Burgos A, et al. D typing discrepancies and anti-D production associated with six new *RHD* alleles. *Transfusion*. 2016;56:S31-020C.
21. Ogasawara K, Sasaki K, Isa K, et al. Weak D alleles in Japanese: a c.960G>A silent mutation in exon 7 of the *RHD* gene that affects D expression. *Vox Sang*. 2016;110:179–184.
22. Le Maréchal C, Guerry C, Benech C, et al. Identification of 12 novel *RHD* alleles in western France by denaturing high-performance liquid chromatography analysis. *Transfusion*. 2007;47:858–863.
23. Kamesaki T, Kumada M, Omi T, et al. A novel mutation in the *RHD* gene in Japanese individuals with weak D, encoding an amino acid change in the 11th transmembranous domain of the RhD protein. *Vox Sang*. 2003;84:141.
24. He Y, Yu L, Xu D, Guo W, Deng G. A case with a novel weak D type. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi*. 2019;36:731–733.
25. Fichou Y, Le Maréchal C, Jamet D, et al. Establishment of a medium-throughput approach for the genotyping of *RHD* variants and report of nine novel rare alleles. *Transfusion*. 2013;53:1821–1828.
26. Avent ND, Jones JW, Liu W, et al. Molecular basis of the D variant phenotypes DNU and DII allows localization of critical amino acids required for expression of Rh D epitopes epD3, 4 and 9 to the sixth external domain of the Rh D protein. *Br J Haematol*. 1997;97:366–371.
27. Wagner FF, Eicher NI, Jorgensen JR, Lonicer CB, Flegel WA. DNB: a partial D with anti-D frequent in Central Europe. *Blood*. 2002;100:2253–2256.
28. Silvy M, Chapel-Fernandes S, Callebaut I, et al. Characterization of novel *RHD* alleles: relationship between phenotype, genotype, and trimeric architecture. *Transfusion*. 2012;52:2020–2029.

29. Dunbrack RL. Rotamer libraries in the 21st century. *Curr Opin Struct Biol* 2002;12:431–440.