Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

RHD/RHCE paralogous misalignment at exon 4 causes false-negative whole-exome sequencing result in a serologically confirmed DFR partial- D individual: Case report

Sara Shaikh

sara_shaikh@ymail.com

Baqai Medical University

Munazza Rashid

Baqai Medical University

Muhammad Usama

Southern University of Science and Technology

Maeesa Sajeel

Baqai Medical University

Aisha Shaikh

Baqai Medical University

Muhammad Rizwan

Baqai Medical University

Amna Shaikh

Baqai Medical University

Case Report

Keywords: RHD, RHCE, partial D, DFR, RHD–RHCE micro-conversion, Case report

Posted Date: October 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10687930/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background.

The Rh blood group is encoded by the highly homologous paralogues RHD and RHCE (~ 97% identity). Partial-D phenotypes arise from RHD variants, including localised RHD–RHCE gene conversions. DFR (RHD*DFR) is a clinically important partial-D allele defined by a discrete RHCE-derived micro-conversion within RHD exon 4 (RHD-CE(4:505–514)-D), introducing the missense changes p.Met169Leu, p.Met170Arg and p.Ile172Phe. As whole-exome sequencing (WES) is increasingly used opportunistically for blood-group genotyping, its behavior at this locus warrants explicit documentation.

Case/Methods.

We analyzed WES data from an individual with a serologically confirmed DFR partial-D phenotype (BioRad ID-Partial RHD typing). RHD was interrogated by three independent approaches: (i) GATK4 best-practice germline calling with functional annotation; (ii) paralogue-aware targeted genotyping at diagnostic weak-D/partial-D/DEL positions with dual mapping-quality thresholds; and (iii) RHtyper, a published RH-genotyping algorithm, applied with the tool's WES mode.

Results.

Standard variant calling returned a normal RHD coding sequence and did not detect DFR; the common weak-D types (1, 2, 3) and the Asian-type DEL allele were correctly excluded. Critically, RHD exon 4, which harbors the DFR-defining c.505–514 micro-conversion had zero read depth at any mapping quality, while the paralogous RHCE exon 4 was covered at 161×. Reads originating from RHD exon 4 were therefore mis-assigned to RHCE owing to the near-identity of the two genes in this region, leaving no reads at the RHD coordinates from which the DFR variants could be called. A true deletion was excluded both by this reciprocal coverage and by the D-expressing DFR phenotype, which requires exon 4 to be present. RHtyper reproduced the false-negative, defaulting to a deletion/reference call.

Conclusions.

In this case, RHD/RHCE paralogous misalignment at exon 4 rendered short-read WES structurally blind to the DFR-defining variants, producing a clinically consequential false-negative that a purpose-built RH genotyper could not rescue. Because the reads are mis-assigned to the paralogue rather than absent, the failure is intrinsic to short-read alignment at this locus and is not remediable by deeper sequencing. WES should not be relied upon to exclude partial-D alleles at the RH locus; definitive characterization requires RHD-specific serological or molecular methods, or long-read sequencing.

1. Introduction

The Rh blood group system is the most polymorphic of the human blood groups and is second only to ABO in transfusion importance. It is encoded by two closely linked, highly homologous genes, *RHD* and *RHCE*, which share approximately 97% nucleotide identity and lie in opposite orientation on chromosome 1p36.11 [1,2]. This paralogy is the central analytical challenge of the locus: short sequencing reads originating from one gene frequently map ambiguously to the other, collapsing mapping quality and confounding variant calling [2,3].

Partial-D phenotypes in which some D epitopes are absent, predisposing carriers to anti-D alloimmunisation are commonly caused by hybrid alleles in which one or more *RHD* exons are replaced by their *RHCE* counterparts through gene conversion [1,4]. DFR (*RHD*DFR*) is one such partial-D allele. Rather than a whole-exon substitution, DFR is defined by a discrete RHCE-derived micro-conversion within *RHD* exon 4 (RHD-CE(4:505–514)-D), introducing the missense substitutions p.Met169Leu, p.Met170Arg and p.Ile172Phe and altering the third extracellular loop of the RhD protein, thereby producing a partial-D phenotype capable of anti-D alloimmunisation [4,5].

Whole-genome sequencing (WGS) coupled with purpose-built algorithms notably RHtyper [3] and bloodTyper and, long-read sequencing [6] can resolve many RH variants, including hybrids. Whole-exome sequencing (WES) is far more widely generated in clinical and research settings and has been applied to RH genotyping and alloimmunisation-risk prediction [7]. However, WES captures only exons and omits the intronic paralogous sequence variants on which hybrid phasing and breakpoint resolution depend. Indeed, when RHtyper was subsequently adapted for exome data, the authors identified *RHD* zygosity and hybrid alleles as among the specific predictions rendered unreliable by the uneven coverage and exacerbated read misalignment of WES, requiring dedicated machine-learning correction [8]. To our knowledge, no study has documented how a standard exome pipeline behaves when it encounters a defined partial-D hybrid in an individual serologically confirmed case.

Here we report a serologically confirmed DFR partial-D individual analysed by WES using three independent bioinformatic strategies: standard best-practice variant calling, paralogue-aware targeted genotyping, and RHtyper. We show that the DFR-defining micro-conversion in *RHD* exon 4 fell within a region of complete capture dropout, rendering both variant calling and the purpose-built genotyper structurally unable to detect DFR consistent with the known WES limitations for this locus [8]. We characterize the extent and cause of this dropout and discuss its implications for opportunistic RH genotyping from exome data.

2. Materials and Methods

2.1 Serological typing and Sequencing

Peripheral blood was collected from a 34 years old male with a serologically weak/partial D phenotype. No relevant personal, family or psychosocial history was reported, and no physical examination findings

were relevant to this laboratory based case. Partial-D status was determined to be DFR using the BioRad ID-Partial RHD Typing kit (monoclonal anti-D panel), which served as the ground-truth reference for this study. Whole-exome sequencing was performed on the DNBSEQ platform (paired-end 150 bp). Post-filtering (SOAPnuke) sequencing yield and quality metrics are summarized in Table 1.

Table 1
Sequencing and quality metrics.

Metric	Value
Platform	DNBSEQ, PE150
Clean read pairs	44,279,399
Q20 / Q30	98.31% / 94.30%
GC content	52.73%
Mapping rate	99.99%
Properly paired	98.36%

2.2 Alignment and standard variant calling

Raw reads were quality-filtered with SOAPnuke [9] and aligned to the GRCh38 reference (Homo_sapiens_assembly38) using BWA-MEM [10], with the read group platform set to DNBSEQ. Duplicates were marked with GATK4 MarkDuplicates [11]; optical-duplicate detection was disabled for the patterned DNBSEQ flow cell (READ_NAME_REGEX null). Base quality scores were recalibrated against dbSNP, Mills, and known-indel resources. Germline variants were called with HaplotypeCaller in GVCF mode, hard-filtered according to GATK best practices, and annotated with snpEff [12] (GRCh38.105) and, independently, wANNOVAR [13]. SAMtools and BCFtools [14] were used for alignment and variant-file processing.

2.3 Paralogue-aware targeted RH genotyping

To mitigate RHD/RHCE cross-mapping, the RH locus was genotyped directly from the pileup at two mapping-quality thresholds (MAPQ ≥ 0 and MAPQ ≥ 10; base quality ≥ 15) using bcftools mpileup/call, and read depth was compared reciprocally between matched RHD and RHCE exons. Diagnostic weak-D, DEL and partial-D positions were interrogated explicitly, with each coordinate self-verified against the reference base before genotyping.

2.4 RHtyper

RHtyper (v1.1) [3] was applied in WES mode (-wes) to the RHD and paired RHCE genes with variant calling enabled. Because the tool is designed for WGS with numeric contig names, execution on the GRCh38 (chr-prefixed) alignment required (i) generation of a contig-renamed reference and alignment and (ii) local correction of several hardcoded/un-normalised contig references in the coverage and large-

deletion modules (see Supplementary Note S1). Coverage was supplied as an estimate; large-deletion, exon-CNV and hybrid-detection stages were run with default parameters.

3. Results

3.1 Variant-level analysis returns a normal RHD coding sequence

Standard GATK best-practice calling, independently confirmed by wANNOVAR, identified no protein-altering variant across the covered RHD coding sequence. The common weak-D alleles (types 1, 2, 3), additional single-nucleotide weak-D types, and the Asian-type DEL allele were all absent (Table 2). The only RHD variant identified, c.149-29G > C (rs2301153), is a common intronic polymorphism of no clinical significance. Critically, the three DFR-defining missense variants (RHD c.505–514; p.Met169Leu, p.Met170Arg, p.Ile172Phe) lie within RHD exon 4, which received zero coverage across its entire length (Section 3.3; Table 3). The variants that define DFR were therefore not merely uncalled but structurally unobservable no reads spanned the positions at which they occur. The normal variant-level result is thus a false-negative produced by capture dropout, not by true absence of the allele.

Table 2
Diagnostic weak-D / DEL / partial-D screen (representative positions).

Allele screened	Defining change	Result
Weak D type 1	RHD c.809T > G	Not detected
Weak D type 2	RHD c.1154G > C	Not detected
Weak D type 3	RHD c.8C > G	Not detected
DEL (Asian-type)	RHD c.1227G > A	Not detected
All RHD coding exons	any protein-altering	None detected
DFR (RHCE exon-4 conversion)	*hybrid, not SNV*	Not detectable at variant level

3.2 RHtyper cannot name the DFR allele from WES

RHtyper completed all stages but failed to assign the DFR allele. Its variant-calling stage recovered no quality-passing RHD variants (all candidate positions had zero quality-passing reads). Allele matching returned a poor fit to reference RHD (RHD*01; log-likelihood − 17.66) with no database match, and the top-ranked call defaulted to a deletion state. At the allele-naming level, therefore, RHtyper reproduced the variant-level false-negative.

3.3 RHD/RHCE paralogous misalignment at exon 4 explains the false-negative

Per-exon coverage analysis identified the direct cause of the false-negative. RHD exon 4 which contains the DFR-defining c.505–514 micro-conversion had a mean depth of 0.0×, with no reads at any mapping quality across the exon or its 500-bp flanking introns (chr1:25,297,085–25,298,233, GRCh38). In striking contrast, the paralogous RHCE exon 4 region (chr1:25,402,823–25,402,971) was covered at 161×. This reciprocal pattern demonstrates that reads originating from the RHD exon 4 region were not absent from the library but were mis-assigned to RHCE, reflecting the near-sequence-identity of the two genes across this segment: the aligner cannot resolve the true gene of origin and places the reads on RHCE, leaving the RHD coordinates empty. Consequently the DFR-defining variants, though present in the genome, are represented by no reads at the RHD locus and cannot be called by any variant-caller or genotyping algorithm. The per-exon RHD/RHCE coverage profile is shown in Fig. 1, and the RHD-only coverage, showing the exon-4 gap, in Fig. 2.

A true genomic deletion of RHD exon 4 was excluded on two independent grounds. First, a deletion would remove the underlying DNA and could not give rise to the 161× coverage observed at the paralogous RHCE exon 4; the presence of those reads shows the sequence is captured and sequenced, merely mis-assigned. Second, and decisively, the individual expresses a DFR partial-D phenotype: DFR arises from an alteration of exon 4 (the c.505–514 micro-conversion), and D-antigen expression requires a translated exon 4. A homozygous deletion of exon 4 would yield a D-negative phenotype, not partial D. The serological result therefore constitutes independent proof that exon 4 is physically present, confirming that the zero RHD coverage is an alignment artefact rather than a structural deletion.

Table 3

Per-exon RHD read depth at a 47× mean on-target depth. Exon 4 (bold) carries the DFR-defining c.505–514 micro-conversion and had zero coverage at any mapping quality, whereas the paralogous RHCE exon 4 was covered at 161× (main text), indicating reads were mis-assigned to RHCE rather than absent.

Exon	RHD depth (×)	RHD CN status	RHCE CN status	Assessable
1	27.7	Gain	normal	yes
2	20.9	Normal	normal	yes
3	0.0	Loss	elevated	conversion
4	0.0	Loss	gain	conversion
5	18.7	Reduced	gain	moderate
6	11.3	Reduced	gain	moderate
7	0.6	Reduced	elevated	no
8	0.0	Loss	elevated	no
9	24.8	Reduced	normal	yes
10	5.4	Reduced	normal	no

3.4 The WES deletion call is an artefact of uneven capture

RHtyper's machine-learning module classified RHD as a heterozygous deletion (probability 0.91). This is an artefact rather than a true genomic deletion: as shown in Section 3.3, the zero RHD-exon coverage arises because reads are mis-assigned to RHCE, not because the DNA is absent, and the D-expressing DFR phenotype independently proves exon 4 is present. A deletion model that infers copy number from RHD read depth alone interprets these mis-assigned exons as deleted. RHD zygosity and large-deletion status therefore cannot be reliably determined from WES at this locus. Because the effect is driven by paralogous alignment rather than by insufficient sequencing, it is not remediable by additional depth—a limitation intrinsic to short-read exome data at this locus.

4. Discussion

This case documents, against a serologically confirmed ground truth, three distinct behaviours of short-read WES at the RHD locus for a partial-D hybrid allele. First, and most importantly for patient safety, variant-level analysis the standard output of essentially all exome pipelines returned a normal RHD coding sequence and thereby a false-negative for DFR. A laboratory relying on exome SNV calling to exclude clinically significant RHD variation would have missed a partial-D allele capable of driving anti-D alloimmunisation.

Second, a purpose-built RH genotyper (RHtyper [3]) could not rescue the allele call from exome data: with no quality-passing variants in the converted region and without the intronic paralogous sequence variants that WGS provides, its allele-matching defaulted to a deletion/reference state. This is the expected consequence of applying a genotyper whose exome adaptation explicitly required machine-learning correction for hybrid alleles [8], and it underscores that tool availability does not by itself make an unmodified WES pipeline adequate for this locus.

Third, the failure was one of read assignment, not of variant interpretation. The reciprocal coverage zero at RHD exon 4, 161× at RHCE exon 4 shows that reads from the region were captured and sequenced but placed on the paralogue, because RHD and RHCE are near-identical across this segment. Because DFR is a discrete ~ 10-bp micro-conversion rather than a whole-exon substitution, it produces no distinctive read-depth signature that could be used to detect it; the exon-4 zero-coverage is a property of the locus and would occur in a non-DFR sample with the same alignment behaviour. We therefore make no claim that coverage analysis can detect or localise DFR. The practical consequence is simple and general: a clinically significant partial-D allele was invisible to WES because the reads carrying its defining variants were assigned to RHCE, and no variant-caller or genotyping algorithm can recover a variant from reads placed on the wrong gene.

Our findings are consistent with, and extend, the established literature. Paralogy, hybrid rearrangements and phasing are recognised as the core analytical challenges of the Rh locus [2,3], and the adaptation of RHtyper to exome data specifically flagged *RHD* zygosity and hybrid alleles as the predictions most

degraded by WES [8]. Our single confirmed case provides a concrete, worked illustration of that degradation against a monoclonal serological ground truth, the DFR-defining micro-conversion fell within a completely uncaptured exon, so the false-negative was structural and unavoidable at the variant level. Consistent with the field's movement toward long-read sequencing and *RHD*-specific long-range PCR for definitive hybrid resolution [6], the reference-laboratory monoclonal partial-D panel provided the answer that exome sequencing could not.

4.1 Limitations

- Single case (n = 1); the extent of RHD/RHCE mis-alignment may vary with read length, aligner and reference build, and should be assessed per platform.
- Exome capture of the RH locus was uneven and incomplete, precluding assessment of a subset of rarer alleles and confounding copy-number/zygosity inference.
- WES cannot resolve intronic breakpoints; DFR subtype could not be determined and was not attempted.
- RHtyper required source-level patching to run on GRCh38 exome data (Supplementary Note S1); results should be interpreted with that caveat.

5. Conclusion

In a serologically confirmed DFR partial-D individual, short-read whole-exome sequencing produced a clinically consequential false-negative because reads from RHD exon 4 which carries the DFR-defining c.505–514 micro-conversion were mis-assigned to the near-identical RHCE paralogue, leaving zero coverage at the RHD coordinates. Reciprocal RHCE coverage and the D-expressing DFR phenotype together exclude a true deletion and confirm the effect as an alignment artefact. A purpose-built RH genotyper could not overcome it, as no variant can be recovered from reads placed on the wrong gene. Because the effect is intrinsic to short-read alignment at this locus rather than a matter of depth, it cannot be remedied by deeper sequencing. WES should not be used to exclude partial-D alleles at the RH locus; clinically important characterisation should rely on RHD-specific serological and molecular methods, or long-read sequencing.

Declarations

Ethics and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. Ethical approval was granted by the Baqai Medical University review board. Written informed consent to participate was obtained from the individual.

Consent to publish

Written informed consent was obtained from the patient for publication of this case report and any accompanying images. A copy of the written consent is available for review by the Editor-in-Chief of this journal.

Availability of data and materials

The data supporting the findings of this study are available from the corresponding author upon request. Raw sequence data are subject to participant-consent and regulatory constraints and cannot be shared publicly.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgments

Not applicable

Authors' contributions

S.S. Conceptualization, investigation, serological evaluation, data collection, interpretation of findings, literature review, and preparation of the original manuscript. M.R. Investigation, data collection, clinical and laboratory interpretation, and critical review of the manuscript. M.U. Computational analysis, data interpretation, conceptualization, manuscript writing and critical revision of the manuscript. M.S. Conceptualization, supervision, clinical and transfusion-medicine interpretation, validation of findings, and critical revision of the manuscript for important intellectual content. A.S. Investigation, laboratory support, data collection, and review of the manuscript. M.R. Methodology, laboratory supervision, validation of laboratory findings, and critical review of the manuscript. A.S. Data collection, literature review, and assistance in manuscript preparation. All authors reviewed the manuscript.

References

1. Chang T-C, Haupfear KM, Yu J, Rampersaud E, Sheehan VA, Flanagan JM, Hankins JS, Weiss MJ, Wu G, Vege S, Westhoff CM, Chou ST, Zheng Y. A novel algorithm comprehensively characterizes human RH genes using whole-genome sequencing data. Blood Adv. 2020;4(18):4347–4357. doi:10.1182/bloodadvances.2020002148.
2. Wagner FF, Flegel WA. RHD gene deletion occurred in the Rhesus box. Blood. 2000;95(12):3662–3668. doi:10.1182/blood.V95.12.3662.

3. Fichou Y, Chen J-M, Férec C. Hybrids and microconversions in RH genes: investigation and implication in transfusion therapy. Ann Blood. 2022;7:6. doi:10.21037/aob-21-73.
4. International Society of Blood Transfusion (ISBT) Working Party on Red Cell Immunogenetics and Blood Group Terminology. RHD blood group alleles (RHD*DFR = RHD*17). Available at: isbtweb.org [or rheference.org, allele RHef00041].
5. Rouillac C, Colin Y, Hughes-Jones NC, Beolet M, D'Ambrosio A-M, Cartron J-P, Le Van Kim C. Transcript analysis of D category phenotypes predicts hybrid Rh D-CE-D proteins associated with alteration of D epitopes. Blood. 1995;85(10):2937–2944. doi:10.1182/blood.V85.10.2937.
6. Gueuning M, Thun GA, Koller S, Sigurdardottir S, Trost N, Wagner L, Mayer B, Engström C, Mattle-Greminger MP, Meyer S. Proof-of-principle: nanopore adaptive sampling enables full blood group genome analysis and resolution of hybrid alleles. Blood Adv. 2026 Feb 10;10(3):864-875. doi: 10.1182/bloodadvances.2025017463. PMID: 41213005; PMCID: PMC12887796.
7. Chou ST, Flanagan JM, Vege S, Luban NLC, Brown RC, Ware RE, Westhoff CM. Whole-exome sequencing for RH genotyping and alloimmunization risk in children with sickle cell anemia. Blood Adv. 2017;1(18):1414–1422. doi:10.1182/bloodadvances.2017007898.
8. Chang T-C, Yu J, Wang Z, Hankins JS, Weiss MJ, Wu G, Westhoff CM, Chou ST, Zheng Y. Machine learning to optimize automated RH genotyping using whole-exome sequencing data. Blood Adv. 2024;8(11):2651–2659. doi:10.1182/bloodadvances.2023011660.
9. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 2018;7(1):1–6. doi:10.1093/gigascience/gix120.
10. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–1760. doi:10.1093/bioinformatics/btp324.
11. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–1303. doi:10.1101/gr.107524.110.
12. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 2012;6(2):80–92. doi:10.4161/fly.19695.
13. Chang X, Wang K. wANNOVAR: annotating genetic variants for personal genomes via the web. J Med Genet. 2012;49(7):433–436. doi:10.1136/jmedgenet-2012-100918.
14. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008. doi:10.1093/gigascience/giab008.

Figures

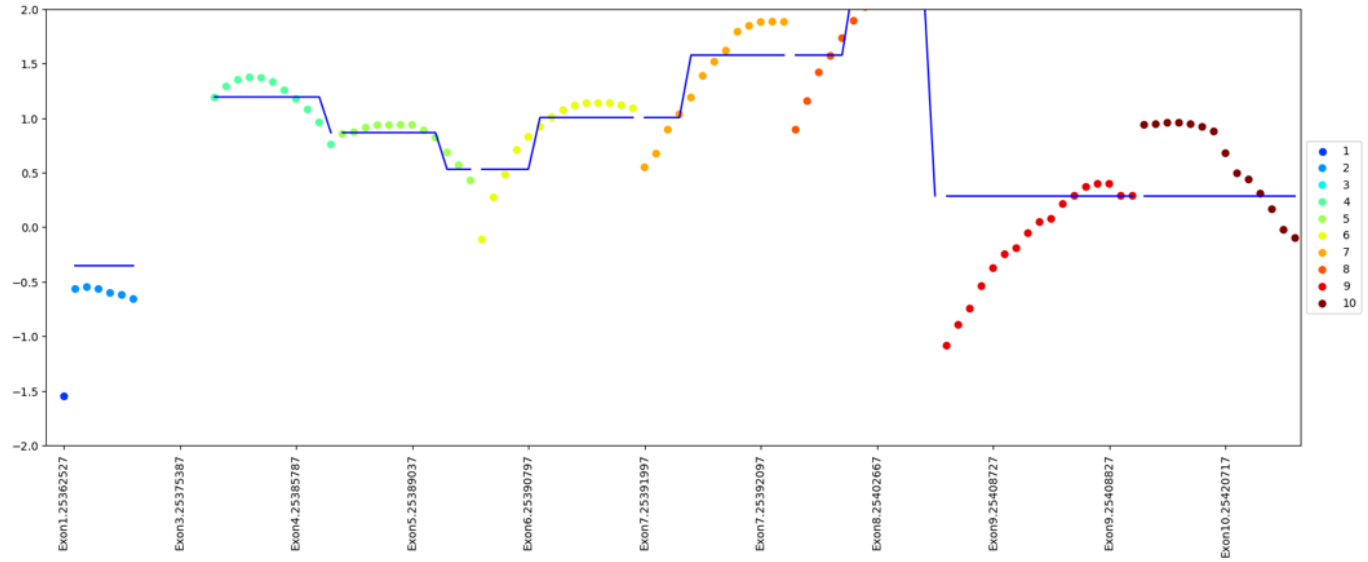


Figure 1

Reciprocal RHD/RHCE per-exon coverage at the RH locus. Binned read depth (RHtyper, $\log_2$ normalised) across the ten exons of RHD (upper track) and the paralogous RHCE (lower track). RHD exon 4 which carries the DFR-defining c.505–514 micro-conversion shows a complete loss of coverage, together with exons 3, 7 and 8, whereas the corresponding RHCE exons, including RHCE exon 4 (161×), retain or gain coverage. The reciprocal RHD-loss/RHCE-retention pattern indicates that reads originating from the RHD exon 4 region are mis-assigned to RHCE owing to the near-identity of the two genes, rather than reflecting a genomic deletion. Exon positions (GRCh38) are shown on the horizontal axis; colours denote individual exons.

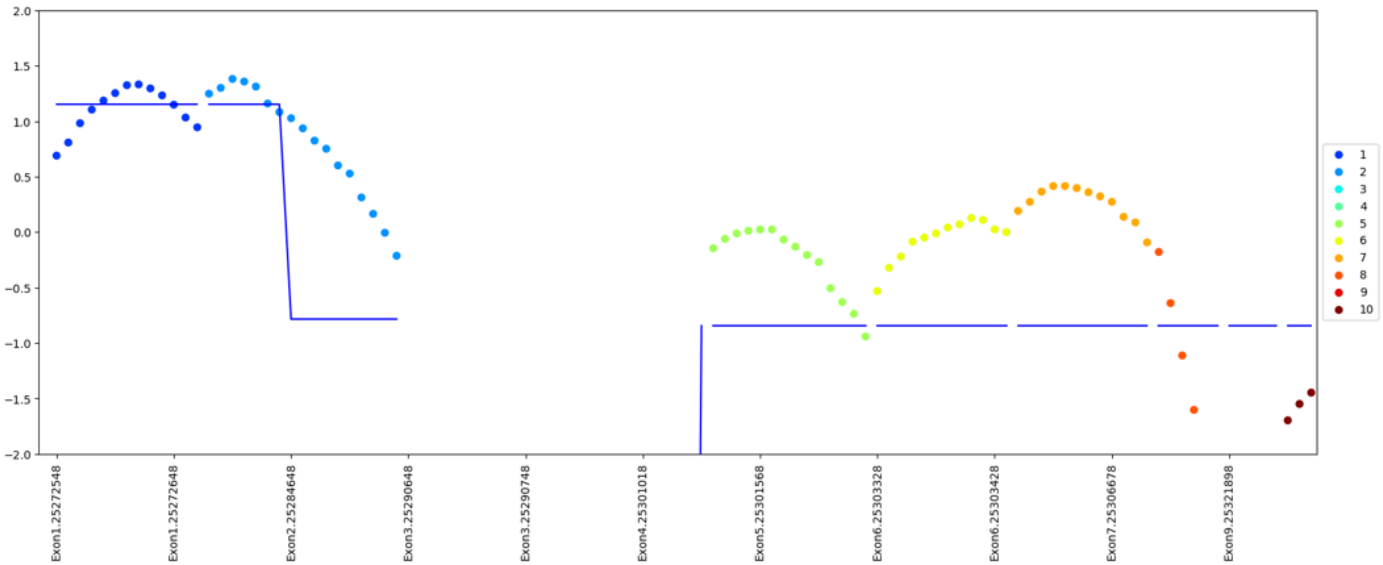


Figure 2

RHD-only per-exon coverage, showing the exon-4 gap. Binned read depth (RHtyper, $\log_2$ normalised) across the ten RHD exons for sample Sal1. Coverage is present at exons 1, 2, 5, 6 and 9 but drops to zero

at exon 4 (and at exons 3, 7 and 8), producing a discrete gap at the region containing the DFR-defining c.505–514 micro-conversion. As shown in Figure 1 and the main text, this gap reflects paralogous mis-assignment of reads to RHCE rather than absence of the underlying sequence; the individual's D-expressing DFR phenotype independently confirms that RHD exon 4 is physically present. Exon positions (GRCh38) are shown on the horizontal axis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx