

ORIGINAL ARTICLE

Genotyping complex structural variation at the malaria-associated human glycoporphin locus using a PCR-based strategy

Walid Algady¹ | Eleanor Weyell¹ | Daria Mateja¹ | André Garcia² |
David Courtin² | Edward J. Hollox¹ 

¹ Department of Genetics and Genome Biology, University of Leicester, Leicester, UK

² UMR 261 MERIT, Institut de Recherche pour le Développement (IRD), Université de Paris, Paris, France

Correspondence

Edward J. Hollox, Department of Genetics and Genome Biology, University of Leicester, University Rd, Leicester LE1 7RH, UK.
Email: ejh33@le.ac.uk

Funding information

Saudi Arabia Cultural Bureau in London

Abstract

Structural variation in the human genome can affect risk of disease. An example is a complex structural variant of the human glycoporphin gene cluster, called DUP4, which is associated with a clinically significant level of protection against severe malaria. The human glycoporphin gene cluster harbours at least 23 distinct structural variants, and accurate genotyping of this complex structural variation remains a challenge. Here, we use a polymerase chain reaction-based strategy to genotype structural variation at the human glycoporphin gene cluster, including the alleles responsible for the U- blood group. We validate our approach, based on a triplex paralogue ratio test, on publically available samples from the 1000 Genomes project. We then genotype 574 individuals from a longitudinal birth cohort (Tori-Bossito cohort) using small amounts of DNA at low cost. Our approach readily identifies known deletions and duplications, and can potentially identify novel variants for further analysis. It will allow exploration of genetic variation at the glycoporphin locus, and investigation of its relationship with malaria, in large sample sets at minimal cost, using standard molecular biology equipment.

KEYWORDS

copy number variation, genotyping, glycoporphin, malaria, structural variation

1 | INTRODUCTION

The malaria parasite *Plasmodium falciparum* is a major cause of childhood mortality in Africa (Liu et al., 2016; Snow et al., 2017; Whitty & Ansah, 2019). Genome-wide association studies have identified multiple alleles that increase protection from severe malaria symptoms; understanding the mechanistic basis of this protection, the pre-

cise phenotype affected by the allelic variation and the evolutionary history of the protective alleles is an important area of current research (Malaria Genomic Epidemiology Network, 2015; Ravenhall et al. 2018). One of the association signals was shown to be due to a complex structural variant, called DUP4, involving the human glycoporphin gene cluster on chromosome 4 (Algady et al., 2018; Leffler et al., 2017).

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Annals of Human Genetics* published by University College London (UCL) and John Wiley & Sons Ltd.

The human glycoporphin genes *GYPE*, *GYPB* and *GYP A* are located on three paralogous repeat units approximately 120 kb in size sharing 97% identity at chromosome 4. *GYP A* and *GYP B* encode glycoporphin A and glycoporphin B, respectively, both of which are major components of the erythrocyte membrane and are receptors for ligands on the surface of the *P. falciparum* merozoite (Baldwin, Li, Hanada, Liu, & Chishti, 2015; Mayer et al., 2009; Orlandi, Klotz, & Haynes, 1992; Pasvol et al., 1982; Sim, Chitnis, Wasniowska, Hadley, & Miller, 1994). No protein product of *GYPE* has yet been detected, and it is thought therefore to be a pseudogene. There is extensive structural variation at this locus resulting in at least eight distinct deletions of one or two repeat units, 13 distinct duplications of one or two repeat units, and two more complex variants (Leffler et al., 2017; Louzada et al., 2020). Some of these variants have been shown to be responsible for particular blood group phenotypes – for example the DUP4 variant codes for the Dantu blood group (Leffler et al., 2017), and homozygosity for either DEL1 and DEL2 variants results in the U– blood group as part of the S-s-U– phenotype observed in Africans (Gassner et al., 2020). These structural variants can be identified and genotyped using sequence read depth of mapped high-throughput sequencing reads (Algady et al., 2018; Leffler et al., 2017). However, the ability to rapidly genotype the different structural variants at this locus using simple polymerase chain reaction (PCR) approaches would have a number of benefits. Firstly, it would allow access to large cohorts with limited DNA, where whole genome amplification is both prohibitively costly and known to introduce bias in copy number measurement (Pugh et al., 2008; Veal et al., 2012). Secondly, it would allow laboratories in resource-limited situations to genotype glycoporphin structural variation, as only standard molecular genetic laboratory equipment and an instrument suitable for DNA fragment analysis are needed. Previously we published a PCR-based approach to genotype the malaria-protective DUP4 variant using a paralogue-specific PCR targeting a junction unique to that variant. We used this PCR approach to genotype a Tanzanian cohort and demonstrate association of the DUP4 variant with haemoglobin levels in peripheral blood (Algady et al., 2018). In this report, we describe a simple triplex paralogue ratio test (PRT) system based on a single PCR of 10-ng DNA that can distinguish other structural variants of the glycoporphin gene cluster.

PRT is a well-established form of quantitative PCR that is particularly robust in copy number detection in genomic DNA. It relies on co-amplification of a test and reference locus using the same PCR primer pair, and distinguishing the two products either by single nucleotide difference (Saldanha et al., 2011) or, more commonly, by a small size difference between the products (Armour et al., 2007).

Because the same primer pair is used to amplify test and reference regions that are similar in sequence, segmental duplications or diverged, dispersed and repeats, are usually targeted. The similarity in sequence means that the amplification of both the test and reference loci follows similar kinetics, and quantification of endpoint products reflects the relative amounts of the starting molecules. PRT has been shown empirically to be more robust than more common forms of quantitative PCR that use different primer pairs to amplify test and reference (Aldhous et al., 2010; Cantsilieris, Western, Baird, & White, 2014; Fode et al., 2011).

We also present a strategy to confirm the most common variants by junction fragment PCRs. Junction fragment PCR relies on knowledge of the copy number breakpoint to design paralogue-specific allele-specific primers flanking the breakpoint, so that if the copy number variant is present, the two paralogue-specific primers are brought into close proximity, enabling a short PCR product to be successfully amplified. Using these approaches, we present a cost-effective strategy to genotype structural variation in the human glycoporphin region to facilitate large-scale genotyping studies of valuable DNA cohorts.

2 | METHODS

2.1 | DNA samples

Control DNA samples from the 1000 Genomes project were chosen because DNA samples are publically available, and high throughput Illumina sequencing has been used to genotype structural variation at this locus (Leffler et al., 2017; Sudmant et al., 2015). A subset of genotypes of this sample set have been validated by fibre-FISH and breakpoints identified (Louzada et al., 2020). DNA samples for CEPH Europeans from Utah (CEU), Chinese from Beijing, Japanese from Tokyo and Yoruba from Ibadan Nigeria (YRI) were previously purchased from Coriell Cell Repositories. DNA from the Beninese infants came from a longitudinal birth cohort established in the district of Tori Bossito (Le Port et al., 2012). The protocol was approved by both the Ethical Committee of the Faculté des Sciences de la Santé (FSS) in Benin and the IRD Consultative Committee on Professional Conduct and Ethics (CCDE). All genomic DNA samples were diluted to a concentration of 10 ng/ μ l in water.

2.2 | Paralogue ratio test (PRT)

Triplex PRT was conducted essentially as previously published (Armour et al., 2007; Hollox, 2017; Walker,

Janyakhantikul, & Armour, 2009). Briefly, three primer pairs targeting different regions of the glycoprotein repeat were combined in a single PCR reaction as follows: In a final volume of 10 μ l, 10 ng of genomic DNA was amplified using 0.5 μ l each of 10 μ M forward and reverse primers for the three PRT assays (Table 1), 1 μ l of 10 \times low dNTP buffer (Hollox, 2017), 1 μ l of 10 \times *Taq* DNA polymerase buffer (KAPA Biosystems, with 15 mM MgCl₂) and 0.5 U *Taq* DNA polymerase. The forward primer of each PRT assay was 5' labelled with a fluorescent tag, either 6-FAM or HEX, to allow visualisation of products by capillary electrophoresis. Cycling conditions were an initial denaturation at 94°C for 2 min, followed by 25 cycles of 94°C for 30 s, 65°C for 30 s, 70°C for 30 s and a final extension of 70°C for 5 min. PCR products were denatured and run on an ABI3130xl capillary electrophoresis machine using POP-4 polymer with an injection time of 30 s under standard fragment analysis parameters and ROX-labelled Mapmarker 1000XL (Eurogentec) and areas under the peak of each PCR product subsequently analysed using the GeneMapper software. Samples with PCR product peak areas between 300 and 60,000 were accepted.

For each sample, PRT area ratios were calculated using test and reference peaks as shown in Table 1. These values were then normalised using DNA samples of known copy number (Handsaker et al., 2015; Leffler et al., 2017; Louzada et al., 2020) that were included in every experiment: NA19190 (CN = 1; DEL2), NA19818 (CN = 1; DEL2), NA19085 (CN = 2; reference), NA19777 (CN = 2; reference) and NA19084 (CN = 3; DUP2). Following normalisation, data were further normalised to account for small batch effects by dividing against the median value for that experiment.

2.3 | Junction fragment duplex PCR

Junction fragment duplex PCR was conducted using 10-ng genomic DNA, 1 μ l each of 10 μ M breakpoint-specific primers with a 3' terminal-linked nucleic acid to enhance specificity (Table 2), 0.1 μ l each of 10 μ M control primers (5'GAGTACGTCGCTTACC-3' and 5'CTTCCACACTTTTGGCATGA-3'), 0.5 U *Taq* DNA polymerase and 2 μ mol each of dATP, dCTP, dGTP and dTTP in a final volume of 10 μ l 1 \times Kapa Buffer A (ammonium sulphate-based buffer, final concentration MgCl₂ 1.5 mM). PCR cycling was 95°C for 2 min, followed by x cycles of 95°C for 30 s, y°C for 30 s, 72°C for 30 s and a final extension of 72°C for 5 min. The number of cycles (x) and annealing temperature (y) were assay specific and shown in Table 2. PCR products were analysed using standard 2% agarose gel electrophoresis, staining by ethidium bromide

TABLE 1 PRT primers and product details

	PRT1	PRT2	PRT4
Primer F	AAACAAGCCACTTCTGCTCAC	TCACCCGAGTTGTGTACATTG	GAAAGGTTTTCCAGAGCAAGC
Primer R	CAAAAATTTCCACC AAGTGCTC	TGTCATTTGCCAACCCAAATTC A	TCAGACCATTGCTCATGCTG
Fluorescent label	6-FAM	6-FAM	HEX
Amplicon size (GYPE repeat) (bp)	418	382	105
Amplicon size (GYPB repeat)	396	363	102
Amplicon size (GYPA repeat)	399	366	105
Test amplicon	PRT1B (396 bp)	PRT2A (366 bp)	PRT4B (102 bp)
Reference amplicon	PRT1A (399 bp)	PRT2B + PRT2E (363 + 382 bp)	PRT4A + PRT4E (105 bp)

Note. The PRT1E amplicon has a size of 418 bp but was not included in subsequent analysis as it was not present in all reference controls.

TABLE 2 Junction fragment PCR primers and product details

Variant	Positive control sample	Forward primer 5'3'	Reverse primer 5'3'	Annealing temp	Cycles	Product size (bp)	Breakpoint location (hg19 GRCh37 chr4)
DEL1	NA19223	AGGATATCAAAATTCATTCTGCG[A]	CACACGCTGCTTTGTCTGT[T]	58	35	331	144835144–144945518
DEL2	NA19144, NAI8858	GGACTCCATGTCTCACATCCAG[T]	CATTGCGGTGGTTTCCCCT[A]	55	34	478	144912871–145016257
DEL7	HG02716	GAAAACTTTTCTCCCT[G]	CCCTTATTATACTGGGTATG[G]	54	34	428	144780450–144901335

Note. The 3' nucleotide of each primer is bracketed, indicating that this nucleotide is a linked nucleic acid (LNA).

and visualisation by UV light. The control primers amplify a 216 bp product, and the breakpoint-specific primers amplify product sizes shown in Table 2.

3 | RESULTS

3.1 | Development of triplex PRT system

A triplex PRT consists of three distinct PRTs that are amplified together in a single PCR reaction, which generates three independent measures of copy number in one single tube and electrophoresis capillary (Walker et al., 2009). If the PRTs have a test and reference on the same chromosome, they are called *cis*-PRTs, in contrast to *trans*-PRTs where the reference is located on a different chromosome from the test amplicon (Hollox, 2017). In this study, we designed three *cis*-PRTs, each one targeting a different region of the glycophorin repeat, with test and reference distinguished on the basis of product size by electrophoresis (Figure 1). Because the most frequent deletion affects the *GYPB* gene, the test amplicon was the *GYPB* amplicon with the reference amplicons being either *GYPB* or *GYPE* amplicons. Because the extent of the most frequent structural variants are known, and the three different PRTs measure copy number at different regions of the glycophorin repeat, it is possible to predict the relative test/reference ratios of these different structural variants (Figure 2).

We validated our triplex PCR by analysing a cohort of 177 1000 Genomes samples with matching short read sequence data, where structural variants of the glycophorin region had previously been identified (Supporting Information). We could then directly compare our PRT results with estimates of copy number from previous sequence read depth analysis and fibre-FISH (Louzada et al., 2020). Analysis of PRT1 shows complete concordance with copy number measured by short-read sequencing depth, detecting the deletions and duplications identified previously (Figure 3a). Analysis of PRT2 shows a more complex situation; although the duplications are identified correctly, some of the deletion samples are identified by a gain of signal by PRT2 rather than a loss of signal (Figure 3b; Table 3). This discrepancy is expected, as PRT2 measures DEL2 as an increase in signal, because the amplicon corresponding to *GYPB* (reference) is lost not *GYPB* (test) (Figure 2). Three outliers were not previously identified as known structural variants and are likely to represent a variant with a gene conversion event affecting one of the PRT amplicons (Figure 3b). PRT4 yields similar results to PRT1 for this cohort, with a few individuals with no structural variant identified showing a higher than expected value, again probably due to gene conversion events.

TABLE 3 Predicted normalised PRT ratios for glycoporphin variants observed in 1000 Genomes project cohort

Variant	PRT1	PRT2	PRT4	Glycoporphin gene haplotype (proximal to distal)
Reference	1	1	1	E-B-A
DEL1	0	2	0	E-A
DEL2	0	0	0	E-A
DEL6	No reference	0	0	A
DEL7	1	2	1	B-A
DEL13	No reference	0	0	E-B/A
DUP2	2	2	2	E-B-A/B-A
DUP3	2	1.33	1.33	E-B-E-B-A
DUP4	2	3	0.5	E-E-A/B-A/B-A
DUP5	2	0.4	0.5	E-B-B-E-E-A
DUP7	1	0.66	0.66	E-E-B-A
DUP8	1	1.5	1.5	E-B-E-B-A
DUP14	1	0.66	0.66	E-E-B-A
DUP22	2	0.66	1	E-A
DUP26	1	0.66	0.66	E-B-A-A
DUP27	2	2	2	E-B-A/B-A
DUP29	0.5	0.66	0.66	E-B-B-A

Note. B/A indicates a fusion gene, orientated proximal to distal along the chromosome. Because the glycoporphin genes are transcribed distal to proximally, B/A indicates a gene where the 5' portion of the transcript is GYP4 and the 3' portion of the fusion transcript is GYPB.

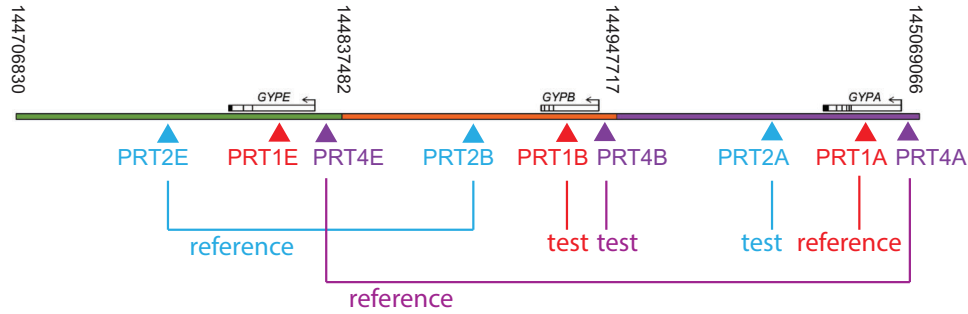
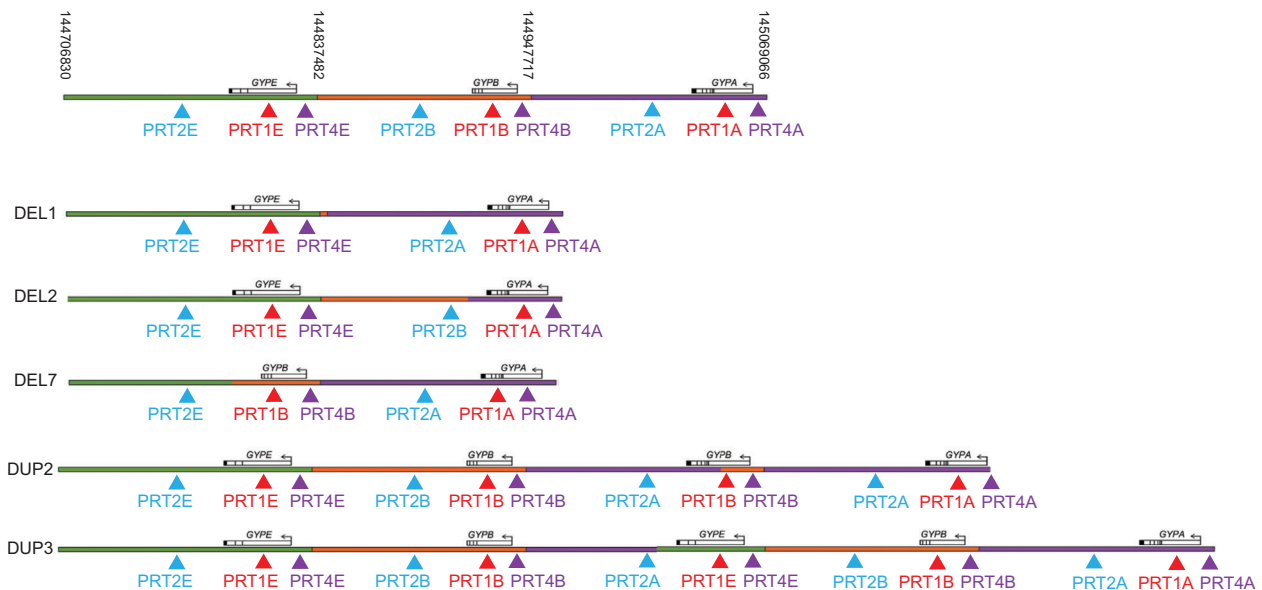


FIGURE 1 Multiplex paralogue ratio test approach in measuring structural variation at the human glycoprotein gene cluster. The hg19 GRCh37 reference sequence is shown, with numbers indicating boundaries among the three glycoprotein repeats coloured green, orange and purple carrying *GYPE*, *GYPB* and *GYPA*, respectively. Amplicons for PRT1, PRT2 and PRT4 are indicated by triangles. Test and reference amplicons for each PRT are also shown. The PRT1E amplicon was not used as a test or a reference [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.com)]



Assay	Test	Reference	WT	DEL1	DEL2	DUP2	DUP3	DEL7
PRT1	B	A	1	0	0	2	2	1
PRT2	A	B+E	1	2	0	2	1.33	2
PRT4	B	A+E	1	0	0	2	1.33	1

FIGURE 2 Distinguishing different structural variants using the multiplex paralogue ratio test. (a) The hg19 GRCh37 reference genome, together with locations of PRT amplicons. (b) Structures of five variants (DEL1, DEL2, DEL7, DUP2 and DUP3) that have been determined previously by fibre-FISH are shown, together with PRT amplicon sites. (c) Expected dosage values for each PRT assay for the reference genome (WT) and the five variants shown. Note that the PRT amplicons used as test and reference are shown, and the reference variant is defined with a value of 1. For example, the ratio of PRT1A to PRT1B is 1 for the WT allele, and therefore predicted to be 2 for DUP3 (two PRT1B sites and one PRT1A site) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.com)]

By comparing across all three PRTs, the four different structural variants can be distinguished from individuals not carrying a structural variant. The exceptions are DUP2 and DUP27, which cannot be distinguished and are either

the same variant or share very close breakpoints (Louzada et al., 2020), the deletions DEL6 and DEL13 and the duplications DUP26, DUP14 and DUP7 (Table 3). These variants are rare and would need to be distinguished

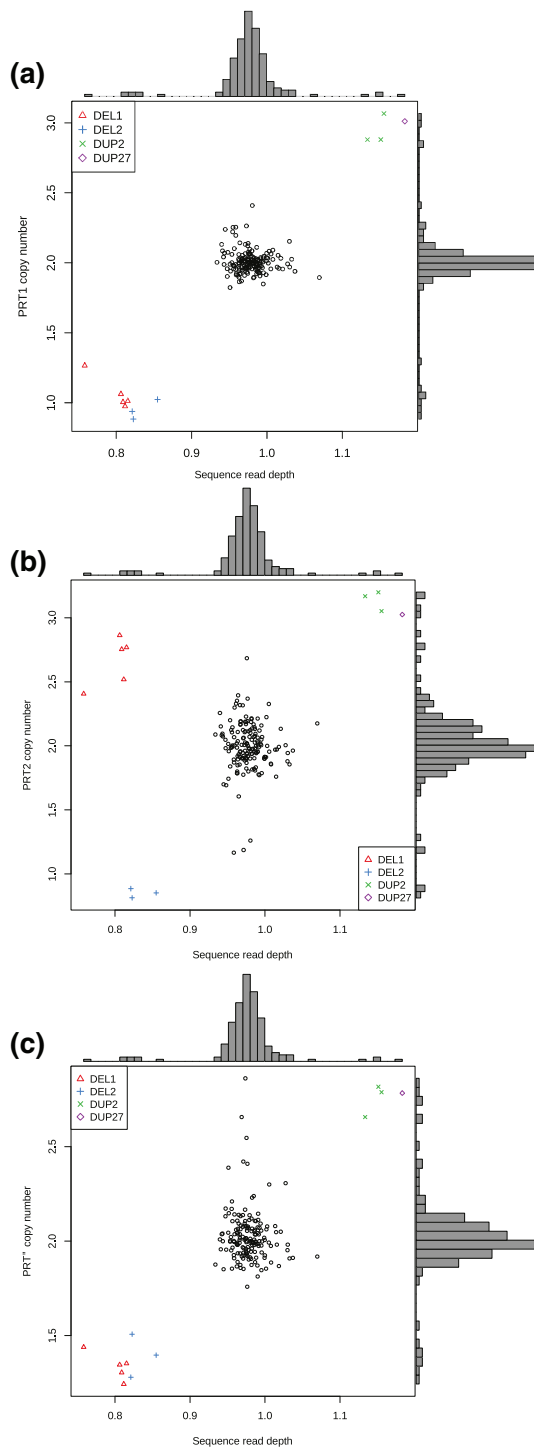


FIGURE 3 Comparison of PRT-based copy numbers to sequence read depth-based copy numbers in 1000 Genomes samples. The scatterplots show comparison of sequence read depth estimation of copy number (with 1.0 being normal read depth) on the *x*-axis and PRT copy number estimates on the *y*-axis for (a) PRT1, (b) PRT2 and (c) PRT3. Each point represents a different individual, with point shape/colour indicating the inferred structural variant carried as a heterozygote by that sample, as shown in the legend. Black circles indicate samples interpreted as homozygous reference variants. A total of 177 samples are shown on each plot [Colour figure can be viewed at wileyonlinelibrary.com]

either by developing a particular junction fragment PCR as we have done for three deletions described below or by genome sequencing.

This analysis shows that by using different loci as test and reference loci, and by combining PRT assays together, different structural variants can be distinguished on the basis of copy number and breakpoint (Table 3). It should also be noted that, for PRT4, the clusters for the deletions are higher than expected, with a median value for heterozygotes around 1.25 compared to a theoretical value of 1 (Table 3), and a median value for homozygotes of 0.5 compared to a theoretical value of 0 (Table 3). Of the two DEL1 control samples, NA19818 shows the expected level of test amplicon amplification, whereas NA19190 shows a much lower level. The reason for this is unclear – it might be due to an undetected variant in the PRT4 test locus primer binding site on the reference allele in NA19190 – but the consequence is that the calibration curve for the positive controls for PRT4 is slightly distorted at lower copy number values. For future studies, we recommend using only NA19818 as the DEL1 positive control.

3.2 | Analysis of Benin population

We used our triplex PRT approach on a cohort of 574 individuals from Benin in west Africa. Individuals from west African populations have been shown to have appreciable frequencies of different glycoprotein structural variants (Leffler et al., 2017). A plot of PRT1 copy number estimates against PRT2 copy number estimates shows clear clustering of individuals (Figure 4a), with genotype inferred using the relative dosages predicted from the structures of the different variants (Figure 2). A plot of PRT1 copy number estimates against PRT4 copy number shows dosages consistent with the assigned genotypes, but with several WT/WT samples as outliers for PRT4. The molecular basis for these outliers is not yet clear; they have not previously been identified as novel structural variants affecting copy number, so it seems likely that they are due to gene conversion events altering the ratio of test/reference amplicon dosage without altering the overall copy number of the locus.

3.3 | Confirmation using junction fragment PCR

Junction fragment PCR uses the fact that, across a duplication or deletion breakpoint, two different sequences come into juxtaposition. By designing one PCR primer to the sequence one side of the breakpoint, and the other PCR

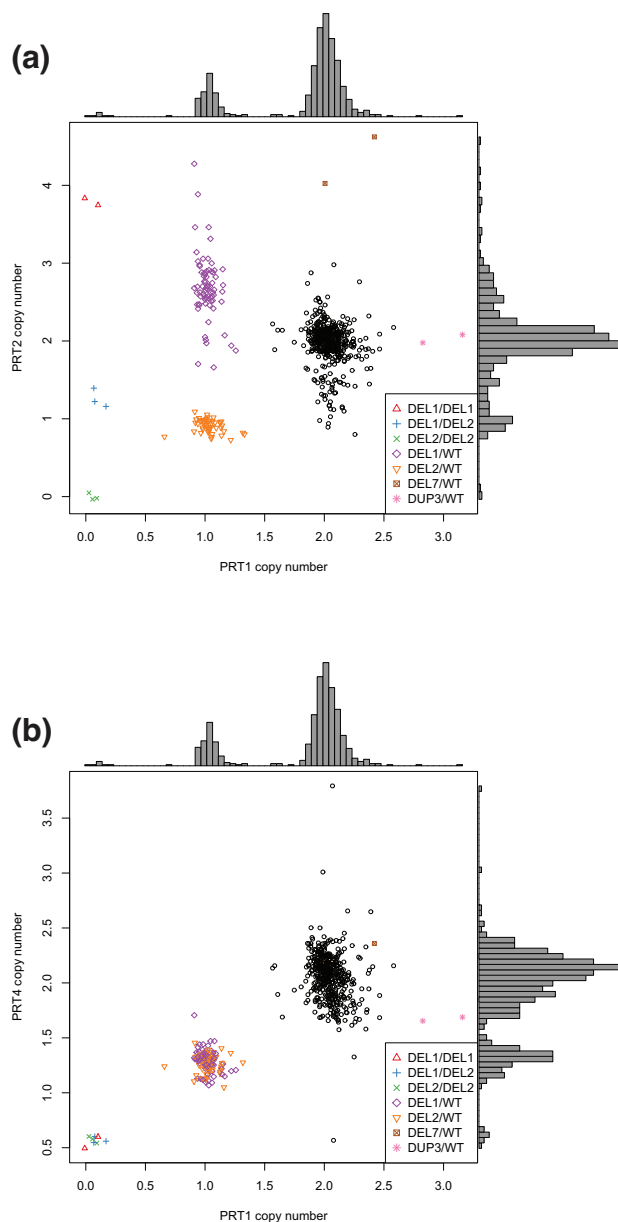


FIGURE 4 Comparison of copy numbers from different PRT tests allows determination of structural variant genotype. The scatterplots show comparison of PRT copy number estimates for (a) PRT1 versus PRT2 and (b) PRT1 versus PRT4 on a cohort of individuals from Benin. Each point represents a different individual, with point shape/colour indicating the inferred structural variant genotype. Black circles indicate samples interpreted as homozygous reference variants. A total of 574 individuals are shown on each plot [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

primer to the other side of the breakpoint, a PCR amplification that is specific to genomic DNA containing this juxtaposition of sequences – that is the deletion or duplication – can be designed (Figure 5). This approach can generate a positive test of the presence of a particular deletion or duplication in a particular genomic DNA. There are two main challenges in applying this approach

to the glycoprotein gene cluster deletions and duplications. Firstly, the location of the breakpoint must be known precisely to enable PCR primers to be designed to flank the breakpoint and still generate a small amplicon that is still likely to be amplified even if the genomic DNA is degraded. Secondly, because the sequences either side of the breakpoint are from glycoprotein paralogous repeats that show >97% identity, each PCR primer must be carefully designed to be specific to that particular paralogue.

To confirm genotypes, we designed junction fragment PCRs to span the known breakpoints of DEL1, DEL2 and DEL7, as determined previously (Louzada et al., 2020). These were designed to generate short amplicons with a co-amplified non-variable control amplicon from an unrelated locus in the genome, to act as a positive control for DNA amplification. To ensure paralogue specificity, the 3' end of each primer was designed over a nucleotide that distinguished the paralogue to be amplified from the other two paralogues, and was synthesised using a linked nucleic acid nucleotide analogue to enhance specificity in base-pairing (Latorra, Campbell, Wolter, & Hurley, 2003). These assays were developed using the samples for each variant as shown in Table 2. The assays for DEL1 and DEL2 were confirmed using the samples HGDP00449 and HGDP00474, whose genotypes had previously been inferred from short read genome sequencing (Louzada et al., 2020).

4 | DISCUSSION

In this study, we present a genotyping strategy for structural variants of the glycoprotein gene cluster in humans. We use a PCR-based approach called the PRT to estimate copy number gain and loss across the region. By combining information from three separate PRT assays, the genotype of particular structural variants can be called. These can then be confirmed by paralogue-specific junction fragment PCR assays.

There are several strengths in the PRT approach. It calls copy number and does not rely on homogeneity of variant breakpoints, unlike junction fragment PCR. This allows a variant to be called that may not share the exact breakpoint with other examples of the same apparent variant, and this is known to be a complication in DUP2, encoding the Sta antigen. Because PRT is based on PCR of small fragments, it can be applied to large sample cohorts containing small quantities of degraded DNA. The two-stage approach of PRT followed by junction fragment PCR allows initial triage of samples homozygous for the reference variant and selection of samples carrying rarer variants for confirmatory junction fragment PCR, or other analyses, such

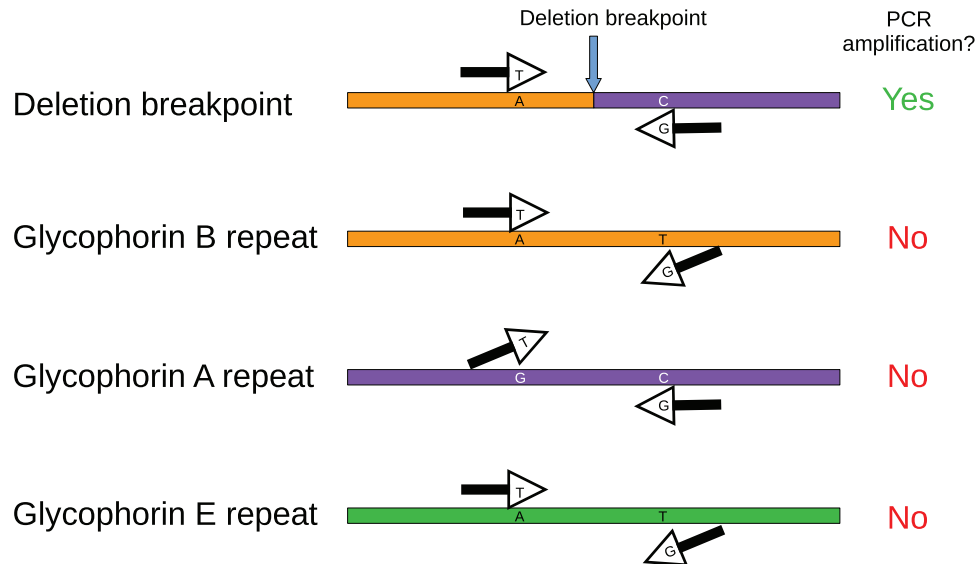


FIGURE 5 Detecting deletions using junction fragment PCR. Design of paralogue-specific PCR primers that span a deletion breakpoint allows detection of particular deletions by junction fragment PCR. In this example, paralogous sequences with a high level of identity from three glycophorin repeats are shown as purple (glycophorin A), orange (glycophorin B) and purple (glycophorin E). At the top is an example of a deletion formed between the glycophorin A and glycophorin B repeats. PCR primers (shown as arrows) are designed with the 3' nucleotide at a position that distinguishes the repeat involved in the deletion from the other two repeats, and the combination of primers amplifies only across the deletion breakpoint. In this example, the primers are designed to distinguish B and E from A (left hand primer) and A from B and E (right hand primer). In combination, these two primers only amplify from the repeat unit at the top, spanning the deletion breakpoint [Colour figure can be viewed at wileyonlinelibrary.com]

as targeted high-throughput sequencing. Because in most cases the breakpoints have previously been defined for these variants (Louzada et al., 2020), junction fragment PCR could also be designed for the very rare variants that currently cannot be distinguished by PRT, for example DUP7 and DUP14.

Our approach uses widely available low-cost molecular biology equipment, with the exception of capillary electrophoresis equipment. Our chosen positive controls are from the 1000 Genomes project, which have been validated extensively by short-read sequence read depth analysis, fibre-FISH and breakpoint Sanger sequencing (Louzada et al., 2020). It is also adaptable, for example using PRT1 alone would be the most straightforward to interpret for common deletions and duplications, and could be used to select for particular genotypes of interest for sequencing.

We demonstrated the usefulness of our approach by genotyping 574 individuals from a Beninese birth cohort. We found two deletions also observed in other west African populations (DEL1 allele frequency 7% and DEL2 allele frequency 5%) that, when homozygous, are responsible for the U- blood group (Gassner et al., 2020). Genotype frequencies are in Hardy-Weinberg proportions, and there is no previous epidemiological evidence that these affect the risk of malarial clinical phenotypes (Leffler et al., 2017). We found two examples of the DEL7 variant and two DUP3

variants, confirming that these are present, but rare, in west African populations. We found no examples of the DUP4 variant that is protective against malaria.

As with any approach, there are weaknesses. As mentioned above, the junction fragment PCR can be challenging to design and assumes homogeneity of breakpoints across all copies of the same variant. Although PRT can determine copy number robustly (Adewoye et al., 2018; Cantsilieris et al., 2014), care must be taken to include internal positive controls to allow for normalization within an experiment and across experiments (Hollox, 2017).

In conclusion, we hope our genotyping approach will be useful to other investigators in tackling this fascinating, complex region where variation has profound consequences on the susceptibility of individuals to malarial disease.

ACKNOWLEDGEMENTS

We would like to thank Rachael Madison for technical support and Mark Jobling for access to the ABI 3100xl capillary electrophoresis machine. We would also like to thank the children and families from Tori Bossito in Benin for their contribution to this study. WA was funded by a PhD studentship from the Saudi Arabia Cultural Bureau in London.

AUTHOR CONTRIBUTIONS

WA and EJH conceived and designed the work. AG and DC contributed samples. WA, EW, DM and EJH collected and analysed data. EJH wrote the paper with contributions from all authors.

DATA AVAILABILITY STATEMENT

Data are available in the Supporting Information.

ORCID

Edward J. Hollox  <https://orcid.org/0000-0002-0115-2298>

REFERENCES

- Adewoye, A. B., Shrine, N., Odenthal-Hesse, L., Welsh, S., Malarstig, A., Jelinsky, S., ... Wain, L. V. (2018). Human CCL3L1 copy number variation, gene expression, and the role of the CCL3L1-CCR5 axis in lung function. *Wellcome Open Research*, 3, 13.
- Aldhous, M. C., Bakar, S. A., Prescott, N. J., Palla, R., Soo, K., Mansfield, J. C., ... Armour, J. A. (2010). Measurement methods and accuracy in copy number variation: Failure to replicate associations of beta-defensin copy number with Crohn's disease. *Human Molecular Genetics*, 19, 4930–4938.
- Algady, W., Louzada, S., Carpenter, D., Brajer, P., Farnert, A., Rooth, I., ... Hollox, E. J. (2018). The malaria-protective human glycoprotein structural variant DUP4 shows somatic mosaicism and association with hemoglobin levels. *American Journal of Human Genetics*, 103, 769–776.
- Armour, J. A. L., Palla, R., Zeeuwen, P. L. J. M., den Heijer, M., Schalkwijk, J., & Hollox, E. J. (2007). Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. *Nucleic Acids Research*, 35, e19–e19.
- Baldwin, M. R., Li, X., Hanada, T., Liu, S. C., & Chishty, A. H. (2015). Merozoite surface protein 1 recognition of host glycoprotein A mediates malaria parasite invasion of red blood cells. *Blood*, 125, 2704–2711.
- Cantsilieris, S., Western, P. S., Baird, P. N., & White, S. J. (2014). Technical considerations for genotyping multi-allelic copy number variation (CNV), in regions of segmental duplication. *BMC Genomics*, 15, 329.
- Fode, P., Jaspersgaard, C., Hardwick, R. J., Bogle, H., Theisen, M., Dodoo, D., ... Hollox, E. J. (2011). Determination of beta-defensin genomic copy number in different populations: A comparison of three methods. *PLOS One*, 6, e16768.
- Gassner, C., Denomme, G. A., Portmann, C., Bensing, K. M., Mattle-Greminger, M. P., Meyer, S., ... Frey, B. M. (2020). Two prevalent ~100-kb GYPB deletions causative of the GPB-deficient blood group MNS phenotype S–s–U– in Black Africans. *Transfus Med Hemother*. <https://doi.org/10.1159/000504946>
- Handsaker, R. E., Van Doren, V., Berman, J. R., Genovese, G., Kashin, S., Boettger, L. M., & McCarroll, S. A. (2015). Large multiallelic copy number variations in humans. *Nature Genetics*, 47, 296–303.
- Hollox, E. J. (2017). Analysis of copy number variation using the paralogue ratio test (PRT). *Methods in Molecular Biology*, 1492, 127–146.
- Latorra, D., Campbell, K., Wolter, A., & Hurley, J. M. (2003). Enhanced allele-specific PCR discrimination in SNP genotyping using 3' locked nucleic acid (LNA) primers. *Human Mutation*, 22, 79–85.
- Le Port, A., Cottrell, G., Martin-Prevel, Y., Migot-Nabias, F., Cot, M., & Garcia, A. (2012). First malaria infections in a cohort of infants in Benin: Biological, environmental and genetic determinants. Description of the study site, population methods and preliminary results. *BMJ Open*, 2, e000342.
- Leffler, E. M., Band, G., Busby, G. B. J., Kivinen, K., Le, Q. S., Clarke, G. M., ... Malaria Genomic Epidemiology Network. (2017). Resistance to malaria through structural variation of red blood cell invasion receptors. *Science*, 356, eaam6393
- Liu, L., Oza, S., Hogan, D., Chu, Y., Perin, J., Zhu, J., ... Black, R. E. (2016). Global, regional, and national causes of under-5 mortality in 2000–15: An updated systematic analysis with implications for the Sustainable Development Goals. *The Lancet*, 388, 3027–3035.
- Louzada, S., Algady, W. K., Weyell, E., Zuccherato, L. W., Brajer, P., Almalki, F., ... Hollox, E. J. (2020). Structural variation of the malaria-associated human glycoprotein ABE region. *BMC Genomics*, 21, 446.
- Malaria Genomic Epidemiology Network. (2015). A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature*, 526, 253–257.
- Mayer, D. C. G., Cofie, J., Jiang, L., Hartl, D. L., Tracy, E., Kabat, J., ... Miller, L. H. (2009). Glycophorin B is the erythrocyte receptor of *Plasmodium falciparum* erythrocyte-binding ligand, EBL-1. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 5348–5352.
- Orlandi, P. A., Klotz, F. W., & Haynes, J. D. (1992). A malaria invasion receptor, the 175-kilodalton erythrocyte binding antigen of *Plasmodium falciparum* recognizes the terminal Neu5Ac(alpha 2–3)Gal- sequences of glycoprotein A. *Journal of Cell Biology*, 116, 901–909.
- Pasvol, G., Jungery, M., Weatherall, D. J., Parsons, S. F., Anstee, D. J., & Tanner, M. J. (1982). Glycophorin as a possible receptor for *Plasmodium falciparum*. *Lancet*, 2, 947–950.
- Pugh, T., Delaney, A., Farnoud, N., Flibotte, S., Griffith, M., Li, H., ... Marra, M. (2008). Impact of whole genome amplification on analysis of copy number variants. *Nucleic Acids Research*, 36, e80–e80.
- Ravenhall, M., Campino, S., Sepulveda, N., Manjurano, A., Nadjm, B., Mtove, G. ... Malaria Genomic Epidemiology Network. (2018). Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. *PLoS Genet*, 14, e1007172.
- Saldanha, G., Potter, L., Dyal, L., Bury, D., Hathiari, N., Ehdode, A., ... Pringle, J. H. (2011). Detection of copy number changes in DNA from formalin fixed paraffin embedded tissues using paralogue ratio tests. *Analytical Chemistry*, 83, 3484–3492.
- Sim, B. K., Chitnis, C. E., Wasniowska, K., Hadley, T. J., & Miller, L. H. (1994). Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. *Science*, 264, 1941–1944.
- Snow, R. W., Sartorius, B., Kyalo, D., Maina, J., Amratia, P., Mundia, C. W., ... Noor, A. M. (2017). The prevalence of *Plasmodium*

- falciparum* in sub-Saharan Africa since 1900. *Nature*, 550, 515–518.
- Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., ... Korb, J. O. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, 526, 75–81.
- Veal, C. D., Freeman, P. J., Jacobs, K., Lancaster, O., Jamain, S., Leboyer, M., ... Brookes, A. J. (2012). A mechanistic basis for amplification differences between samples and between genome regions. *BMC Genomics*, 13, 455.
- Walker, S., Janyakhantikul, S., & Armour, J. A. (2009). Multiplex Paralogous Ratio Tests for accurate measurement of multiallelic CNVs. *Genomics*, 93, 98–103.
- Whitty, C. J. M., & Ansah, E. (2019). Malaria control stalls in high incidence areas. *BMJ*, 365, l2216.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Algady W, Weyell E, Mateja D, Garcia A, Courtin D, Hollox EJ. Genotyping complex structural variation at the malaria-associated human glycophorin locus using a PCR-based strategy. *Ann Hum Genet.* 2021;85:7–17. <https://doi.org/10.1111/ahg.12405>