Deletions in the *MAL* gene result in loss of Mal protein, defining the rare inherited AnWj-negative blood group phenotype

Short title: Mal protein carries the AnWj blood group antigen

Louise A. Tilley[1†], Vanja Karamatic Crew[1†], Tosti J. Mankelow,[1,2†], Samah A. AlSubhi[1,3,4†], Benjamin Jones[1], Abigail Borowski[1], Vered Yahalom[5,6], Lilach Finkel[7], Belinda K. Singleton[8], Piers J. Walser[9], Ashley M. Toye[3,10], Timothy J. Satchwell[3,10], and Nicole M. Thornton[1‡]

[1]International Blood Group Reference Laboratory, NHS Blood and Transplant, Bristol, UK. [2]Component Development Laboratory, NHS Blood and Transplant, Cambridge, UK. [3]School of Biochemistry, University of Bristol, Bristol, UK. [4]Laboratory Medicine Department, Faculty of Applied Medical Sciences, Umm Al-Qura University, Makkah, Saudi Arabia. [5]Blood Services & Apheresis Institute, Rabin Medical Center, Petah Tiqva, Israel. [6]Faculty of Medical and Health Science, Tel Aviv University, Tel Aviv, Israel. [7]National Blood Group Reference Laboratory, Magen David Adom National Blood Services, Ramla, Israel. [8]NIHR Blood and Transplant Research Unit in Genomics to Enhance Microbiology Screening, NHS Blood and Transplant, Bristol, UK. [9]Clinical Biotechnology Centre, NHS Blood and Transplant, Bristol, UK. [10]National Institute for Health Research (NIHR) Blood and Transplant Research Unit in Red Blood Cell Products, University of Bristol, Bristol, UK.

[†]shared first authors, [‡] corresponding author

Corresponding author: Nicole M. Thornton
International Blood Group Reference Laboratory
NHS Blood and Transplant
500 North Bristol Park
Northway
Filton
Bristol
BS34 7QH

Tel: +44 (0)117 9217469
Fax: +44 (0)117 9217329
Email: Nicole.thornton@nhsbt.nhs.uk

Data availability: Nucleotide sequence data are available in the DDBJ/EMBL/GenBank databases under the accession number PP694609. All other data supporting the findings of this study are available from the corresponding author upon reasonable request.

Abstract word count: 243

Text word count: 4278

Figures: 5

Tables: 1

References: 64

Key points

- The inherited AnWj-negative blood group phenotype is caused by homozygosity for a deletion in *MAL*, encoding Myelin and lymphocyte protein
- Mal protein is expressed on red blood cell membranes of AnWj-positive, but not AnWj-negative, individuals

Abstract

The genetic background of the high prevalence red blood cell antigen AnWj has remained unresolved since its identification in 1972, despite reported associations with both CD44 and Smyd1 histone methyltransferase. Development of anti-AnWj, which may be clinically significant, is usually due to transient suppression of antigen expression, but a small number of individuals with persistent, autosomally-recessive inherited AnWj-negative phenotype have been reported. Whole exome sequencing of individuals with the rare inherited AnWj-negative phenotype revealed no shared mutations in *CD44H* or *SMYD1*, but instead we discovered homozygosity for the same large exonic deletion in *MAL*, which was confirmed in additional unrelated AnWj-negative individuals. *MAL* encodes an integral multi-pass membrane proteolipid, Myelin and Lymphocyte protein (Mal), which has been reported to have essential roles in cell transport and membrane stability. AnWj-positive individuals were shown to express full-length Mal on their red cell membranes, which was not present on the membranes of AnWj-negative individuals, whether of an inherited or suppression background. Furthermore, binding of anti-AnWj was able to inhibit binding of anti-Mal to AnWj-positive red cells, demonstrating the antibodies bind to the same molecule. Over-expression of Mal in an erythroid cell-line resulted in expression of AnWj antigen, regardless of the presence or absence of CD44, demonstrating that Mal is both necessary and sufficient for AnWj expression. Our data resolve the genetic background of the inherited AnWj-negative phenotype, forming the basis of a new blood group system, further reducing the number of remaining unsolved blood group antigens.

Introduction

Knowledge of the complexities of blood groups is crucial, not only in transfusion and pregnancy, but also in understanding disease susceptibility, population genetics, and evolutionary studies. The molecular bases of nearly all known blood group antigens have now been elucidated, with 45 blood group systems, comprising 362 antigens, currently recognised by the International Society of Blood Transfusion (ISBT). A small number of remaining antigens have yet to be genetically resolved; 16 of low prevalence (700 series) and only 3 of high prevalence (901 series).[1]

Amongst these, the AnWj or Anton (ISBT 901009) high prevalence antigen, previously also known as Wj,[2] was first identified in 1972.[3] AnWj is present on epithelial tissues and red cells of more than 99.9% of individuals but is absent on cord cells.[2] Transition of red cells from AnWj-negative to AnWj-positive normally occurs between 3-46 days after birth and is completed within 24-48 hours[4] although the mechanism remains unknown. AnWj expression is also markedly reduced on In(Lu) phenotype cells,[3-6] caused by mutation in the *KLF1* gene encoding the erythroid transcription factor Krueppel-like factor 1 (KLF1/EKLF).[7] In(Lu) cells show reduced expression of several blood group antigens in addition to AnWj, including those of the Lutheran and Indian (CD44) systems.[8,9]

Antibodies directed against AnWj are rare and usually associated with transient suppression of AnWj expression on an individual's red cells, associated with various disorders, including lymphoid malignancies and autoimmune haemolytic anaemia.[10-13] In addition, an autosomal recessive inherited, persistent AnWj-negative phenotype was first reported in a consanguineous family of Middle Eastern origin.[14] Alloanti-AnWj in the AnWj-negative propositus and her sibling was presumably produced as a result of pregnancy, but was not associated with hemolytic disease of the fetus and newborn (HDFN). However, anti-AnWj may be clinically significant and has been associated with acute hemolytic transfusion reactions.[12,13,15]

The AnWj antigen has been reported to be the erythrocyte receptor for *Haemophilus influenzae,*[4] however the carrier molecule for AnWj on the red cell surface has remained elusive. It has been reported that AnWj may be carried on CD44, or dependent on CD44 for its expression.[16,17] More recently, a missense mutation in the *SMYD1* gene, encoding a histone methyltransferase, has been reported to be the cause of the inherited AnWj-negative phenotype, but the mechanism for this has not been elucidated.[18]

Myelin and lymphocyte protein (Mal) is an integral multi-pass membrane proteolipid of 17KDa, which resides in glycosphingolipid and cholesterol-enriched microdomains (GEMs). It appears to shuttle between the trans-Golgi network, plasma membrane and endosomes.[19] The *MAL* gene, located on chromosome 2, was first identified in human T-lymphocytes[20] and is organised into 4 exons, each encoding a membrane-associated segment and an adjacent hydrophilic sequence (Figure 1).[21] Mal is also expressed in myelin-forming cells and in polarised epithelial cells,[22-24] where it plays important roles in assembly and stabilisation of GEMs and in protein targeting. Mal is required for apical sorting and transport of influenza virus haemagglutinin in

polarised cells[25] and has also been identified as the receptor for binding of *Clostridium perfringens* epsilon toxin on erythrocyte membranes.[26]

Mal has been reported to have conflicting roles in cancer, apparently acting both as an oncogene and a tumour suppressor.[27] Oesophageal cancer cell-lines demonstrated a lack of Mal expression, caused by promoter methylation[28] and Mal downregulation has also been associated with a number of epithelial malignancies, including colon cancer,[29,30] breast cancer[31] and bladder cancer.[32] Contrastingly, over-expression of Mal appears to be associated with poor prognosis in several lymphocyte-derived cancers such as acute adult T-cell leukaemia,[33] primary mediastinal large B-cell lymphoma[34] and Hodgkin's lymphoma,[35] as well as in ovarian cancer.[36,37] In addition to these opposing cancer associations, a rare, missense, mutation in *MAL* has been reported to cause leukodystrophy, with mutated Mal retained in the endoplasmic reticulum.[38]

In this study, we identify homozygous deletions in the *MAL* gene associated with the inherited AnWj-negative phenotype, finally elucidating the genetic basis of this long-established blood group antigen. We demonstrate lack of red blood cell Mal expression in AnWj-negative individuals, both transient and inherited, and show that over-expression of Mal in an erythroid cell-line results in expression of AnWj antigen, thereby establishing a new blood group system.

Methods

Subjects

Blood samples were procured, and the study conducted according to NHS Blood and Transplant (NHSBT) Research and Development governance requirements and ethical standards, in accordance with the Declaration of Helsinki. Ethics approval was granted by NHS Health Research Authority, Bristol Research Ethics Committee reference 12/SW/0199. Samples from AnWj-negative individuals (n=7; P1-P7) and family members (n=4; P2_F1-P2_F4) were obtained from cryopreserved rare reference material International Blood Group Reference Laboratory (IBGRL) collections. Fresh blood samples were obtained from a further AnWj-negative individual and her sibling (P8, P8_F1) with informed consent. Samples from In(Lu) phenotype individuals (n=3) were obtained from cryopreserved IBGRL reference collections. Samples are detailed in Supplemental Table 1.

Ethylenediaminetetraacetic acid peripheral blood samples from voluntary NHSBT donors, consented to the use of their blood for research purposes, were used as AnWj-positive controls.

Serological testing

Standard agglutination techniques were used for assessment of anti-AnWj reactivity and red blood cell (RBC) phenotyping.[39] For the indirect antiglobulin test, a low-ionic strength saline (LISS) tube method was used, where the secondary antibody was anti-human globulin (AHG) (Millipore) polyspecific for human primary antibodies (in-house IBGRL reference collection), or polyclonal rabbit anti-mouse immunoglobulin G (IgG; Dako) for mouse primary monoclonal antibodies (IBGRL Research Products). Anti-AnWj eluates were prepared using Gamma ELU-KIT™ II rapid acid elution kit (Immucor). Direct antiglobulin test (DAT) was carried out using DC-Screening I cards (BioRad) and LISS tube method using AHG. For papain treatment, one volume of washed packed RBCs was incubated with two volumes of papain (NHSBT) for 3 minutes. Following incubations, RBCs were washed a minimum of four times with PBS until supernatant was clear.

For agglutination tests using anti-Mal, 15 µL of a 2% solution of RBCs in PBS 1% BSA was incubated for 20 minutes at room temperature (RT) with either anti-GPA (BRIC256 and BRIC163; IBGRL Research Products), polyclonal mouse anti-Mal (ab167374; Abcam) or monoclonal anti-AnWj (H86).[40] Cells were washed twice with PBS 1% BSA before incubation with secondary antibody (goat anti-mouse; Jackson) and viewing under a light microscope at x100 magnification.

For blocking, 15 µL of a 2% solution of RBCs in PBS 1% BSA was incubated at RT for 60 minutes with plasma containing human anti-AnWj or a 1:1 mix of plasma containing human anti-Lu^a and human anti-Lu^b. After two washes in PBS 1% BSA, cells were incubated for 20 minutes at RT with either anti-GPA, anti-Mal or anti-AnWj. Following incubation, cells were washed twice with PBS 1% BSA before a 1 in

100 dilution of goat anti-mouse was added and the plate immediately spun prior to reading.

NGS library preparation

Genomic DNA was extracted using a commercial isolation kit (QIAamp DNA blood mini kit; Qiagen). Library preparation, indexing and target enrichment for exome or targeted blood group panel sequencing was carried out using Nextera Rapid Capture Exome or DNA prep with enrichment (Illumina). The exome kit targets 45 Mb of coding sequence, whilst the custom panel targets 0.5 Mb coding sequence of known blood group genes (recognised by ISBT as of 2022), plus several other genes of interest including *MAL*. Paired-end (2x87 cycle) single-plex sequencing was carried out on a MiSeq (Illumina). Secondary data analysis, including alignment of reads against human reference genome hg19, was performed using MiSeq Reporter v2.5.1 (Illumina) and variants called and annotated using Illumina Variant Studio v2.2. Selected alignments were further visualised using Integrative Genomics Viewer (IGV v2.3).[41]

Genetic analysis of *SMYD1* and *CD44H*

SMYD1 exon 7 and *CD44H* exons present in erythroid isoform (1-5 and 15-17) were amplified by PCR (primers and conditions shown in Supplemental Table 2). PCR products were prepared for sequencing using ExoSAP-IT (Thermo Fisher Scientific) and Sanger sequenced with forward and reverse PCR primers using a capillary automated DNA sequencer (3130xL Genetic Analyzer; Applied Biosystems). Sequences were aligned to *Homo sapiens SMYD1* (NM_001330364.2) or *CD44* (NG_008937.1) reference sequence using SeqScape (Applied Biosystems).

Sequencing and deletion analysis of *MAL*

Confirmation of deletions involving *MAL* exons 3 and 4 was carried out by PCR amplification of individual coding exons (1-4), and of products spanning deleted regions (Supplemental Table 2). PCR products were prepared for sequencing with ExoSAP-IT (Thermo Fisher Scientific) and sequenced with forward and reverse PCR primers, or internal sequencing primers as shown in Supplemental Table 2.

Flow cytometry

Erythrocytes (0.25 µL packed cells per test) and BEL-A cells ($3x10^5$ cells per test) were analysed by flow cytometry after labelling with; anti-Mal (ab167374), anti-AnWj (H86), human anti-AnWj eluate, anti-CD44 (BRIC222; IBGRL Research Products), anti-Lutheran (BRIC221 and BRIC224; IBGRL Research Products). Isotype controls (BioRad) were tested in parallel to ascertain background fluorescence. Cells were incubated with primary antibody at RT for 30 minutes and washed once. Bound antibody was detected by adding R-phycoerythrin anti–mouse globulins (Dako) or Alexa Fluor® 647-AffiniPure F(ab')2 fragment goat anti-human IgG (Stratech

Scientific) and incubated at 4°C for 30 minutes. Cells were washed once, and fluorescence geometric mean was recorded for each test. Flow cytometry was performed on a Navios (Beckman Coulter).

RBC membrane preparation and immunoblotting

RBC membranes were prepared as previously described[42] and protein content was determined using a commercial assay kit (Pierce™ BCA; Pierce Biotechnology). Membranes (50 µg) were separated by SDS PAGE on a 4-20% gradient precast gel (BioRad) under reducing conditions, transferred to PVDF membrane (Thermo Fisher Scientific), blocked overnight in 5% milk and immunoblotted using mouse monoclonal anti-Mal antibodies E1:sc-390687 (Santa Cruz) or 6D9.[43]

Generation of CRISPR/Cas9 DNA constructs and *MAL* editing of BEL-A cells

Knockout of *MAL* in BEL-A cells was achieved by nucleofection of RNPs using an Amaxa 4D nucleofector (Lonza) – programme DZ100. RNPs were prepared by addition of 22.5 pmol of each of two chemically modified sgRNAs (Synthego) targeting *MAL* to 18 pmol Cas9 and incubating at 25˚C for 15 minutes. Guide sequences were as follows (*MAL* exon 1- ATGGCCCCCGCAGCGGCGAC and exon 2- TCCTCCCCCGCAGATCTTCG). Transfected cells were FACS sorted to obtain single clones and the presence of biallelic disruptive mutations assessed by Sanger sequencing. Knockout *CD44* BEL-A cell-line has been previously described.[44]

Overexpression of Mal in BEL-A cells

Preparation of lentiviral particles and transduction of BEL-A cells was performed as previously described[45] using XLG3 vector for expression of *MAL* constructs. Open reading frames of full-length and truncated (exons 1 and 2 only) *MAL* were synthesised, with and without N-terminal Green Fluorescent Protein (GFP) tag, and cloned into XLG3 by Genscript.

Confocal microscopy of GFP-expressing BEL-A cell-lines

BEL-A cells (8×10^4) were centrifuged at 400 g for 5 minutes and pellets were resuspended in 200 µL of BEL-A expansion medium. Cell suspensions were placed into an uncoated 8-well chamber slide (Ibidi) and incubated at RT for 10 minutes. Samples were imaged at 22°C using 40x oil immersion lenses on a TCS SP8 confocal imaging system (Leica). Images were obtained using Leica LAS AF software.

Results

Serological identification of AnWj-negative individuals

Individuals with AnWj-negative phenotype were originally identified through clinical investigations of unresolved antibodies to a high prevalence antigen. The antibody in each case was identified as anti-AnWj and AnWj-negative phenotype determined by lack of reactivity with at least two examples of anti-AnWj. Reactivity with autologous control cells and Direct Antiglobulin Test (DAT) status were scrutinised to predict if AnWj-negative phenotype was due to inheritance or transient suppression of antigen expression. Suppression was suspected if a positive DAT was observed, and comparable weak reactivity was seen with the RBCs when typed with anti-AnWj. Weak reactivity detected with monoclonal anti-AnWj under papain enhancement conditions, or the ability to adsorb and elute anti-AnWj from the cells also supported suppression. Conversely, inheritance was inferred from family information where available, a negative DAT and/or inability to adsorb/elute anti-AnWj from RBCs, demonstrating complete lack of the antigen. Additional AnWj-negative individuals were identified through family studies (Supplemental Table 1).

Determination of genetic basis of inherited AnWj-negative phenotype

To uncover the genetic basis of the AnWj antigen, DNA from two unrelated individuals, with persistent, inherited AnWj-negative phenotype (P1, P2) was subjected to whole exome sequencing. The *CD44* and *SMYD1* genes, previously reported to be associated with AnWj expression, were analysed in both individuals. No mutations were identified in *CD44H* (encoding the erythroid CD44 isoform) in either individual. The *SMYD1* variant 959G>A (NM_001330364.2, R320Q, rs114851602) was homozygous in P2, but P1 was wild-type at this position, suggesting an alternative cause for the AnWj-negative phenotype. Thus, whole exome sequences were analysed, filtering for genes carrying rare (frequency <0.01), homozygous missense or nonsense mutations, consistent with lack of the high prevalence AnWj antigen.

No variant fitting these criteria was shared by both individuals, so data was further analysed to search for possible deletions shared by P1 and P2, but not present in control sequences. A homozygous deletion of exons 3 and 4 of *MAL* gene, encoding Mal protein, was observed in both P1 and P2, but was not present in AnWj-positive controls (n=5). To determine if this deletion was responsible for lack of AnWj, further samples with serologically determined AnWj-negative phenotypes were subjected to targeted NGS and/or Sanger sequencing. AnWj-negative family members of P2 (P2_F1, P2_F2), two siblings from the same region (apparently unrelated to the original family; P8, P8_F1), and a further two unrelated AnWj-negative individuals (P3, P4) were all shown to be homozygous for the same 6646 nt deletion in *MAL*, encompassing coding exons 3 and 4 (NC_00002.12, NM_002371.4; 262-423_462+2348del) as shown in Figure 2. AnWj-positive family members of P2 (P2_F3, P2_F4) were either homozygous wild-type or heterozygous for the *MAL* deletion allele. Three serologically AnWj-negative individuals (P5 to P7) had wild-type *MAL*, all of which have evidence suggesting the negative phenotype is due to suppression of AnWj expression (such as positive DAT, clinical condition; Supplemental Table 1). No coding mutations in *CD44H* were identified in any AnWj-

negative individuals, except P6 who carried a heterozygous rare coding mutation in *CD44* exon 17; 2018G>A, Arg673Gln (rs61752932). Only P2 and family members, plus the unrelated AnWj-negative siblings from the same region, were homozygous for *SMYD1* variant 959G>A, with all other AnWj-negative individuals showing wild-type sequence at this position. Sequencing results are detailed in Table 1.

Mal expression in AnWj-positive and AnWj-negative red cells

Previous studies[26,46,47] suggest that Mal is expressed in low levels on human red cell membranes. To confirm this, AnWj-positive red cells were tested with polyclonal mouse anti-Mal, both serologically and by flow cytometry. Anti-Mal was able to bind and agglutinate AnWj-positive red cells, confirming its expression on normal erythrocyte membranes. Anti-Mal did not, however, bind to AnWj-negative red cells, even those with wild-type *MAL* (Figure 3A-B). Interestingly, flow cytometry of AnWj-positive control erythrocytes shows a bimodal expression pattern with both anti-Mal and anti-AnWj (Figure 3B), indicative of dual populations of antigen-positive and antigen-negative cells. A band consistent in size with full-length Mal was observed by Western blotting of red cell membranes from AnWj-positive individuals, which was not present with AnWj-negative membranes, regardless of presence or absence of the *MAL* deletion (Figure 3C; Supplemental Figure 1).

Furthermore, it was shown that binding of anti-Mal to AnWj-positive red cells was inhibited by binding of human anti-AnWj, consistent with both antibodies binding to the same molecule (Supplemental Figure 2). Together, these results demonstrate that Mal protein is required for expression of AnWj antigen on erythrocytes.

Mal expression in BEL-A erythroid cell-line

Although transcriptomics data (not shown) suggest a low level of Mal expression in the erythroid cell line BEL-A, no expression of either Mal or AnWj was detectable on unmodified BEL-A cells (Figure 4A). However, to ensure that no endogenous Mal expression was possible in these cells, a BEL-A erythroid *MAL* KO cell-line was established by CRISPR-mediated gene editing. Sequencing of a selected BEL-A clone confirmed presence of a biallelic disruptive *MAL* mutation; Ile32AlafsTer21, confirming the clone to be genetically *MAL* KO. As expected, no expression of Mal or AnWj was detectable on BEL-A *MAL* KO cells (Figure 4A). To determine specificity of anti-AnWj antibodies for Mal protein, a stable Mal over-expression BEL-A cell-line was established. Both Mal and AnWj expression were detectable by flow cytometry in these cells (Figure 4A), confirming that Mal is required for AnWj expression. However, over-expression of the truncated form of *MAL* observed in AnWj-negative patients (exons 1 and 2 only) did not result in detectable surface expression of either Mal or AnWj (Figure 4A). Confocal imaging of GFP-tagged Mal showed that the majority of full-length Mal was located on the plasma membrane of the BEL-A cells, whilst truncated Mal was restricted to the cytoplasm (Figure 4B), consistent with the lack of surface expression detected by flow cytometry.

Investigation of CD44 association with AnWj expression

Since previous studies[16,17] have suggested that CD44 is required for expression of AnWj antigen, expression of Mal and AnWj were investigated in *CD44* KO BEL-A cells. These cells showed no detectable levels of AnWj or Mal expression (as observed in unmodified BEL-A cells), but over-expression of Mal in *CD44* KO cells resulted in expression of both Mal and AnWj, whereas over-expression of CD44 alone had no effect (Figure 5A). These data demonstrate that CD44 is neither sufficient, nor required for, expression of AnWj antigen, whilst expression of Mal alone results in AnWj expression, even in the absence of CD44.

Furthermore, normal expression of CD44 is observed on AnWj-negative cells[14] (Supplemental Figure 3) and on Mal KO cells (Figure 5B) showing that CD44 expression does not require the presence of Mal/AnWj antigen. In(Lu) cells, with mutations in erythroid-specific transcription factor KLF1, show reduced expression of several red cell antigens, including CD44.[7-9] AnWj and Mal expression are extremely weak or absent on these cells, even in the presence of (albeit reduced) CD44 expression (Supplemental Figure 3). These data suggest that, rather than lack of AnWj antigen arising as an indirect result of reduced CD44 expression, Mal protein may be a direct target for KLF1 transcriptional regulation. Analysis of the *MAL* proximal promoter region using JASPAR[48] shows a KLF1 consensus binding site sequence (GGGGCGGGG) starting at position –179 relative to the start of coding sequence.

Discussion

Despite reported associations with CD44[16,17] and *SMYD1*,[18] the genetic basis of the AnWj antigen has remained unresolved for over 50 years. Here we demonstrate that the AnWj-negative phenotype is associated with loss of expression of Mal proteolipid on erythrocytes, and that individuals with inherited AnWj-negative phenotype are homozygous for the same large deletion in the *MAL* gene. We also show that over-expression of Mal in an erythroid cell-line results in expression of AnWj antigen, independently of CD44 expression.

The AnWj-negative phenotype is usually caused by transient suppression of antigen expression, associated with certain haematological disorders and malignancies.[10-13] Such suppression of blood group antigens, with associated antibody production against the suppressed antigen, is an established phenomenon, seen for example in Kell[49], Kidd[50] and LW[51,52] systems. In such cases, it has been hypothesised that acquired reduction of antigen expression on the patient's red cells leads to development of the antibody, typically during pregnancy or following transfusion[13,51,52]. Establishing whether an AnWj-negative phenotype is due to suppression, or the rare inherited phenotype, is not always clearcut. Demonstration of a long-standing AnWj-negative phenotype (as in P1), or a clear pattern of inheritance (P2, P8) provides strong evidence for the autosomal recessive genetic phenotype. Conversely, serological indicators such as a positive DAT, or very weakly detectable antigen expression, combined with clinical information, may suggest suppression of antigen expression. Transient suppression would ideally be proven by follow-up samples, demonstrating the return of antigen expression and resulting loss of antibody (as has been shown in previous studies[10,13]), however this is often not practicable. The three AnWj-negative individuals in this study without *MAL* mutations all had indications suggestive of a suppression cause of their negative phenotype (Supplemental Table 1), although unfortunately re-testing to confirm return of AnWj expression was not possible. Genotyping for *MAL* mutation can now be used to determine whether an AnWj-negative phenotype is inherited, or due to suppression, which may aid clinical decision-making processes. It is interesting that all the inherited AnWj-negative individuals tested in this study share the same *MAL* deletion. It remains to be seen whether other *MAL* variants affecting AnWj expression may be identified in future, but no homozygous nonsense variants, or missense mutations encoding external amino acids, are currently listed in the GnomAD database.

All AnWj-negative individuals tested lacked Mal expression on their erythrocytes, regardless of presence or absence of the *MAL* deletion. The mechanism of loss of Mal expression in those with no *MAL* deletion remains unclear. Mal is known to be down-regulated in certain cancers,[28-32] likely due to promoter hypermethylation, which may explain the transient loss of AnWj antigen in these individuals. Our results support the theory that Mal may be down-regulated at the transcriptional or translational level, as no Mal protein is detectable on red cell membranes of the suppressed AnWj-negative phenotype by Western blotting, serology or flow cytometry. No pathological phenotype appears to be directly associated with lack of Mal in either transient or inherited AnWj-negative individuals. The original Anton

propositus (P1), together with P2 and several of her AnWj-negative family members, have been well-studied over the course of many years, with no disease phenotype or red cell abnormalities found, despite their *MAL* deletions. Mal function may be non-essential, or compensated, perhaps by another member of the MAL protein family.

AnWj antigen has previously been reported to be carried on CD44,[16,17] or at least require CD44 for its expression. In(Lu) phenotype cells, with reduced levels of CD44, have a significant reduction in AnWj expression,[3-6] and a CD44-deficient individual was reported to be AnWj-negative.[53] These observations have been used to support the theory that AnWj expression is dependent on CD44. However, our results show that both Mal and AnWj can be expressed on erythroid cells *in vitro* in the absence of CD44, so reduced CD44 levels do not explain aberrant AnWj antigen expression in In(Lu) or CD44-deficient phenotypes. Both phenotypes are associated with mutations in erythroid-specific transcription factor, KLF1,[7, 54, 55] which has known roles in transcriptional regulation of multiple blood group genes. There is at least one KLF1 consensus binding site in the GC box region of the *MAL* proximal promoter,[48,56] so erythrocyte Mal expression may be regulated by KLF1, explaining reduced AnWj expression in KLF1 mutant phenotypes, although this remains to be proven.

More recently, inherited cases of AnWj-negative phenotype were reported to be associated with homozygosity for a rare, missense mutation in *SMYD1*,[18] encoding a histone methyltranferase. Although members of two AnWj-negative families in our study do indeed show segregation of this mutation with the AnWj-negative phenotype, we also show other unrelated AnWj-negative individuals (including the original Anton propositus, P1) to lack the *SMYD1* mutation. Both *SMYD1* and *MAL* are located on chromosome 2, on either side of the centromere, so this association may be due to linkage between the *SMYD1* variant and the *MAL* deletion, which appears to be restricted to a specific (Middle Eastern) population. Strong linkage disequilibrium has been demonstrated across centromeric gaps.[57] In our study, all *SMYD1* variant homozygotes were also homozygous for the *MAL* deletion, whilst heterozygous carriers of the *SMYD1* mutation were concomitantly heterozygous for the *MAL* deletion allele. However, the observation of *MAL* deletion homozygotes with no *SMYD1* mutation show that this linkage does not exist in all populations.

Many blood group antigens have known roles in infection and disease susceptibility, with several acting as pathogen receptors on the red cell surface. Perhaps the most well-known example is the Duffy blood group antigen, carried on atypical chemokine receptor 1 protein, acting as a receptor for binding of *Plasmodium* v*ivax* to reticulocytes.[58,59] AnWj antigen has been reported to act as the erythrocyte receptor for *Haemophilus* i*nfluenzae*[6] and erythrocytes from individuals with the inherited AnWj-negative phenotype (including P2 from this current study) are not agglutinated by *H*. *influenzae*.[60] Mal has been identified as the erythrocyte receptor for binding of *Clostridium perfringens* epsilon toxin (ETX).[26] ETX has been shown to cause hemolysis in humans, but not in other mammals[61] and ETX does not bind to Mal KO mouse cells.[62] Exposure to ETX has been proposed as a potential environmental trigger in the development of Multiple Sclerosis.[63] More work is needed to elucidate the role of Mal/AnWj in red cell infection.

Levels of Mal membrane expression on mature red cells are low, and detection of Mal in the red cell proteome is variable,[47] perhaps due to different sample preparation protocols and the nature of the Mal proteolipid. We were unable to demonstrate Mal membrane expression at any stage of primary cell culture (data not shown), so the pattern of Mal expression during erythropoiesis remains unclear. It is interesting to note that normal adult red cells show bimodal expression of AnWj/Mal by flow cytometry and serologically, anti-AnWj exhibits characteristic mixed field agglutination. This is indicative of only a proportion of cells expressing AnWj antigen, whilst others remain negative. A similar pattern of reactivity is seen with Lutheran antibodies,[64] the reason for which has not, as yet, been explained.

Both cord cells and red cells of newborn babies lack AnWj antigen expression, and rapid transition of circulating red cells from AnWj-negative to AnWj-positive occurs shortly after birth.[4] It is interesting therefore, that production of anti-AnWj in inherited AnWj-negative individuals appears to have been stimulated by pregnancy, albeit with no evidence of HDFN. Antibody production may have been stimulated by presence of AnWj antigen on other fetal cell types. Although anti-AnWj is not implicated in HDFN, it is significant in transfusion,[12,13] so identification of AnWj-negative individuals, whether inherited or due to suppression, remains clinically important. Our findings allow development of new tests for detecting such rare individuals, and the identification of natural human Mal knockouts allows further study of the role of Mal in erythroid cells and wider biological processes.

Acknowledgements: The authors would like to thank the following individuals: Jan Frayne and Kongtana Trakarnsanga for provision of the BEL-A cell line; IBGRL Molecular Diagnostics Department for Sanger sequencing support; Andrew Herman from the University of Bristol Faculty of Biomedical Sciences Flow Cytometry Facility for cell sorting support; Cyril Levene and Michael Bennett for their work in describing the first inherited AnWj family; Miguel A Alonso (Centro de Biologia Molecular, CSIC-UAM), for providing the anti-Mal 6D9 monoclonal antibody utilized in this study; Joyce Poole and Geoff Daniels for initiating this project; finally to the late Dave Anstee, for his insights and many fruitful discussions and ideas about this long study, we know he would have been proud of the outcome.

This study was supported by NHSBT R&D funding, the Cultural Bureau at the Royal Saudi Embassy in the UK and the Saudi Ministry of Education (SAA PhD scholarship), Medical Research Council (MR/V010506/1) and by infrastructure funding from the National Institute for Health Research Blood and Transplant Research Unit (NIHR BTRU) in Red Cell Products (IS-BTU-1214-10032).The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR, or the Department of Health and Social Care.

Authorship contributions: N.M.T, V.K.C, L.A.T, T.J.M, A.M.T and T.J.S conceived, designed and coordinated the study. N.M.T, T.J.M, A.B, B.J performed serology experiments. L.A.T, V.K.C, S.A.A performed whole exome/targeted sequencing, Sanger sequencing and associated data analysis. T.J.S, T.J.M, S.A.A performed expression, CRISPR/Cas9, confocal and flow cytometry experiments. T.J.S and S.A.A performed Western blotting. V.Y, L.F performed antibody identification on patients P8 and P8_F1. B.K.S made CD44 KO cell-line and performed transcriptomics. P.J.W carried out protein modelling. L.A.T, V.K.C, T.J.M, N.M.T wrote the manuscript and all authors contributed to review of the final manuscript.

Disclosure of conflicts of interest: AMT is a co-founder, a Director and consultant to Scarlet Therapeutics Ltd. TJS is a co-founder and scientific consultant to Scarlet Therapeutics Ltd.

References

1. International Society of Blood Transfusion (ISBT). Red cell immunogenetics and blood group terminology. Accessed 22 February 2024. https://www.isbtweb.org/isbt-working-parties/rcibgt.html
2. Marsh WL, Brown PJ, DiNapoli J, et al. Anti-Wj: an autoantibody that defines a high-incidence antigen modified by the In(Lu) gene. *Transfusion.* 1983;23(2):128-130.
3. Boorman KE, Tippett P. Unpublished observations, 1972; cited in: Blood groups in man. Oxford: Blackwell; 1975:274-275.
4. Poole J, Van Alphen L. *Haemophilus influenzae* receptor and the AnWj antigen. *Transfusion.* 1988;28(3):289.
5. Poole J, Giles CM. Observations on the Anton antigen and antibody. *Vox Sang.* 1982;43(4):220-222.
6. Van Alphen L, Poole J, Overbeeke M. The Anton blood group antigen is the erythrocyte receptor for Haemophilus influenzae. *FEMS Microbiol Lett.* 1986;37:69-71.
7. Singleton BK, Burton NM, Green C, et al. Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype. *Blood.* 2008;112(5):2081-2088.
8. Taliano V, Guévin R-M, Tippett P. The genetics of a dominant inhibitor of the Lutheran antigens. *Vox Sang.* 1973;24(1):42-47.
9. Spring FA , Dalchau R , Daniels GL , et al. The In^a and In^b blood group antigens are located on a glycoprotein of 80 000 MW (the CDw44 glycoprotein) whose expression is influenced by the In(Lu) gene. *Immunology.* 1988;64(1):37-43.
10. Mannessier L, Rouger P, Johnson CL, Mueller KA, Marsh WL. Acquired loss of red-cell Wj antigen in a patient with Hodgkin's disease. *Vox Sang*. 1986;50(4):240-244.
11. Harris T, Steiert S, Marsh WL, Berman LB. A Wj-negative patient with anti-Wj. *Transfusion*. 1986;26(1):117.
12. De Man AJ, van Dijk BA, Daniels GL. An example of anti-AnWj causing haemolytic transfusion reaction. *Vox Sang*. 1992;63(3):238.
13. Xu Z, Duffett L, Tokessy M, Cote J, Goldman M, Saidenberg E. Anti-AnWj causing acute hemolytic transfusion reactions in a patient with aplastic anemia. *Transfusion*. 2012;52(7):1476-1481.
14. Poole J, Levene C, Bennett M, Sela R, van Alphen L, Spruell PJ. A family showing inheritance of the Anton blood group antigen AnWj and independence of AnWj from Lutheran. *Transfus Med*. 1991;1(4):245-251.

15. Yong J, Johnstone M, Callaghan T, Chandrasekar A, Pervaiz MS. Anti-AnWj antibody induced haemolytic transfusion reaction in a patient with primary acquired pure red cell aplasia. *Transfus Med*. 2022;32(4):346-348.

16. Telen M, Rao N, Udani M, Liao H-X, Haynes BF. Relationship of the AnWj blood group antigen to expression of CD44. *Transfusion.* 1993;33(suppl):48. Abstract S182.

17. Rao N, Udani M, Telen MJ. Demonstration by monoclonal antibody immobilization of erythrocyte antigens and dot blot that both the In and AnWj blood group antigens reside on CD44. Transfusion 1994;34(suppl):25. Abstract S99.

18. Yahalom V, Pillar N, Zhao Y, et al. SMYD1 is the underlying gene for the AnWj-negative blood group phenotype. *Eur J Haematol*. 2018;101(4):496-501.

19. Puertollano R, Alonso MA. Targeting of MAL, a putative element of the apical sorting machinery, to glycolipid-enriched membranes requires a pre-golgi sorting event. *Biochem Biophys Res Commun*. 1999;254(3):689-92.

20. Alonso MA, Weissman SM. cDNA cloning and sequence of MAL, a hydrophobic protein associated with human T-cell differentiation. *Proc Natl Acad Sci USA*. 1987;84(7):1997-2001.

21. Rancaño C, Rubio T, Alonso MA. Alternative splicing of human T-cell-specific MAL mRNA and its correlation with the exon/intron organization of the gene. *Genomics*. 1994;21(2):447-50.

22. Zacchetti D, Peränen J, Murata M, Fiedler K, Simons K. VIP17/MAL, a proteolipid in apical transport vesicles. *FEBS Lett*. 1995;377(3):465-9.

23. Kim T, Fiedler K, Madison DL, Krueger WH, Pfeiffer SE. Cloning and characterization of MVP17: a developmentally regulated myelin protein in oligodendrocytes. *J Neurosci Res.* 1995;42(3):413-22.

24. Schaeren-Wiemers N, Valenzuela DM, Frank M, Schwab ME. Characterization of a rat gene, rMAL, encoding a protein with four hydrophobic domains in central and peripheral myelin. *J Neurosci.* 1995;15(8):5753-64.

25. Puertollano R, Martín-Belmonte F, Millán J, et al. The MAL proteolipid is necessary for normal apical transport and accurate sorting of the influenza virus hemagglutinin in Madin-Darby canine kidney cells. *J Cell Biol*. 1999;145(1):141-51.

26. Geng Z, Huang J, Kang L, et al. Clostridium perfringens epsilon toxin binds to erythrocyte MAL receptors and triggers phosphatidylserine exposure. *J Cell Mol Med*. 2020;24(13):7341-7352.

27. Lara-Lemus R. On the role of Myelin and Lymphocyte protein (MAL) in cancer: A puzzle with two faces. *J Cancer*. 2019;10(10):2312-2318.

28. Mimori K, Shiraishi T, Mashino K, et al. MAL gene expression in esophageal cancer suppresses motility, invasion and tumorigenicity and enhances apoptosis through the Fas pathway. *Oncogene*. 2003;22(22):3463-71.

29. Lind GE, Ahlquist T, Kolberg M, et al. Hypermethylated *MAL* gene - a silent marker of early colon tumorigenesis. *J Transl Med*. 2008;6:13.

30. Kalmár A, Péterfia B, Hollósi P, et al. DNA hypermethylation and decreased mRNA expression of MAL, PRIMA1, PTGDR and SFRP1 in colorectal adenoma and cancer. *BMC Cancer*. 2015;15:736.

31. Horne HN, Lee PS, Murphy SK, Alonso MA, Olson JA Jr, Marks JR. Inactivation of the *MAL* gene in breast cancer is a common event that predicts benefit from adjuvant chemotherapy. *Mol Cancer Res*. 2009;7(2):199-209.

32. Blaveri E, Simko JP, Korkola JE, et al. Bladder cancer outcome and subtype classification by gene expression. *Clin Cancer Res*. 2005;11(11):4044-55.

33. Kohno T, Moriuchi R, Katamine S, Yamada Y, Tomonaga M, Matsuyama T. Identification of genes associated with the progression of adult T cell leukemia (ATL). *Jpn J Cancer Res*. 2000;91(11):1103-10.

34. Copie-Bergman C, Gaulard P, Maouche-Chrétien L, et al. The *MAL* gene is expressed in primary mediastinal large B-cell lymphoma. *Blood*. 1999;94(10):3567-75.

35. Hsi ED, Sup SJ, Alemany C, et al. MAL is expressed in a subset of Hodgkin lymphoma and identifies a population of patients with poor prognosis. *Am J Clin Pathol*. 2006;125(5):776-82.

36. Zanotti L, Romani C, Tassone L, et al. MAL gene overexpression as a marker of high-grade serous ovarian carcinoma stem-like cells that predicts chemoresistance and poor prognosis. *BMC Cancer*. 2017;17(1):366.

37. Lee PS, Teaberry VS, Bland AE, et al. Elevated MAL expression is accompanied by promoter hypomethylation and platinum resistance in epithelial ovarian cancer. *Int J Cancer*. 2010;126(6):1378-1389.

38. Elpidorou M, Poulter JA, Szymanska K, et al. Missense mutation of MAL causes a rare leukodystrophy similar to Pelizaeus-Merzbacher disease. *Eur J Hum Genet*. 2022;30(7):860-864.

39. Judd JW, Johnson ST, Storry J. Judd's methods in immunohematology, 3rd edition. Bethesda: AABB Press; 2008

40. Knowles RW, Bai Y, Lomas C, Green C, Tippett P. Two monoclonal antibodies detecting high frequency antigens absent from red cells of the dominant type of Lu(a-b-) Lu:-3. 1982. *J. Immunogenet.* 9: 353-357.

41. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. 2013. *Briefings in Bioinformatics* 14: 178-192.

42. Dodge JT, Mitchell C, Hanahan DJ. The preparation and chemical characteristics of hemoglobin-free ghosts of human erythrocytes. *Arch Biochem Biophys.*1963;100:119-130.

43. Millán J, Alonso MA. MAL, a novel integral membrane protein of human T lymphocytes, associates with glycosylphosphatidylinositol-anchored proteins and Src-like tyrosine kinases. Eur J Immunol. 1998;28(11):3675-3684.

44. Thornton N, Karamatic Crew V, Tilley L, et al. Disruption of the tumour-associated EMP3 enhances erythroid proliferation and causes the MAM-negative phenotype. *Nat Commun*. 2020;11(1):3569.

45. Karamatic Crew V, Tilley LA, Satchwell TJ, et al. Missense mutations in PIEZO1, which encodes the Piezo1 mechanosensor protein, define Er red blood cell antigens. Blood. 2023;141(2):135-146.

46. Bryk AH, Wiśniewski JR. Quantitative Analysis of Human Red Blood Cell Proteome. *J Proteome Res*. 2017;16(8):2752-2761.

47. Ravenhill BJ, Kanjee U, Ahouidi A, et al. Quantitative comparative analysis of human erythrocyte surface proteins between individuals from two genetically distinct populations. *Commun Biol*. 2019;2:350.

48. Rauluseviciute I, Riudavets-Puig R, Blanc-Mathieu R, et al. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2024;52(D1):D174-D182.

49. Lee E, Burgess G, Win N. Autoimmune hemolytic anemia and a further example of autoanti-Kpb. *Immunohematology*. 2005;21(3):119-121.

50. Issitt PD, Obarski G, Hartnett PL, Wren MR, Prewitt PL. Temporary suppression of Kidd system antigen expression accompanied by transient production of anti-Jk3. *Transfusion*. 1990;30(1):46-50.

51. Giles CM, Lundsgaard AA. A complex serological investigation involving LW. *Vox Sang*. 1967;13:206-416.

52. Chown B, Kaita H, Lowen B, Lewis M. Transient production of anti-LW by LW-positive people. *Transfusion.* 1971;11:220-222.

53. Parsons SF, Jones J, Anstee DJ, et al. A novel form of congenital dyserythropoietic anemia associated with deficiency of erythroid CD44 and a unique blood group phenotype [In(a-b-), Co(a-b-)]. *Blood*. 1994;83(3):860-868.

54. Singleton BK, Fairweather VSS, Lau W, et al. A novel EKLF mutation in a patient with dyserythropoietic anemia: the first association of EKLF with disease in man. *Blood*. 2009;114:162 abstract

55. Arnaud L, Saison C, Helias V, et al. A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. *Am J Hum Genet*. 2010;87(5):721-727.

56. Tugores A, Rubio T, Rancaño C, Alonso MA. A tandem array of Sp-1 sites and a reverse initiator element are both required for synergistic transcriptional activation of the T-cell-specific MAL gene. *DNA Cell Biol*. 1997;16(3):245-255.

57. Langley SA, Miga KH, Karpen GH, Langley CH. Haplotypes spanning centromeric regions reveal persistence of large blocks of archaic DNA. *Elife*. 2019;8:e42989. Published 2019 Jun 25.

58. Miller LH, Mason SJ, Clyde DF, McGinniss MH. The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med*. 1976;295(6):302-304.

59. Miller LH, Mason SJ, Dvorak JA, McGinniss MH, Rothman IK. Erythrocyte receptors for (Plasmodium knowlesi) malaria: Duffy blood group determinants. *Science*. 1975;189(4202):561-563.

60. Van Alphen L, Levene C, Geelen-van den Broek L, Poole J, Bennett M, Dankert J. Combined inheritance of Epithelial and Erythrocyte Receptors for Haemophilus influenzae. *Infection and Immunity*. 1990;58(11):3807-3809.

61. Gao J, Xin W, Huang J, et al. Hemolysis in human erythrocytes by Clostridium perfringens epsilon toxin requires activation of P2 receptors. *Virulence*. 2018;9(1):1601-1614.

62. Rumah KR, Ma Y, Linden JR, et al. The Myelin and Lymphocyte Protein MAL Is Required for Binding and Activity of Clostridium perfringens ε-Toxin. *PLoS Pathog*. 2015;11(5):e1004896. Published 2015 May 20.

63. Rumah KR, Linden J, Fischetti VA, Vartanian T. Isolation of Clostridium perfringens type B in an individual at first clinical presentation of multiple sclerosis provides clues for environmental triggers of the disease. *PLoS One*. 2013;8(10):e76359. Published 2013 Oct 16.

64. Novitzky-Basso I, Spring F, Anstee D, Tripathi D, Chen F. Erythrocytes from patients with myeloproliferative neoplasms and splanchnic venous thrombosis show greater expression of Lu/BCAM. *Int J Lab Hematol*. 2018;40(4):473-477.

Table 1. Sequencing results for AnWj-negative individuals and family members

Sample	Phenotype	Genetic investigations	*CD44H* (Exons 1-5, 15-18) NM_00610.4	*SMYD1* NM_001330364.2	*MAL* NC_0002.12, NM_002371.4
P1	AnWj-negative (inherited)	WES, Sanger *SMYD1* & *MAL*, breakpoint analysis	No mutations	No mutation	Homozygous *MAL* 6646 nt deletion encompassing exons 3 and 4; (NC_00002.12, NM_002371.4; 262-423_462+2348del)
P2	AnWj-negative (inherited)	WES, Sanger *SMYD1* & *MAL*	No mutations	Homozygous 959G>A, R320Q (rs114851602; gnomAD v4.0.0 freq 0.0006)	Homozygous *MAL* 6646 nt deletion as in P1
P2_F1	AnWj-negative (inherited)	Targeted NGS, Sanger *SMYD1* & *MAL*, breakpoint analysis	No mutations	Homozygous 959G>A, R320Q	Homozygous *MAL* 6646 nt deletion as in P1
P2_F2	AnWj-negative (inherited)	Targeted NGS, Sanger *SMYD1*	No mutations	Homozygous 959G>A, R320Q	Homozygous *MAL* 6646 nt deletion as in P1
P2_F3	AnWj-positive	Targeted NGS, Sanger *SMYD1*	No mutations	No mutation	Wild-type *MAL*, no deletion
P2_F4	AnWj-positive	Targeted NGS, Sanger *SMYD1*	No mutations	Heterozygous 959G>A, R320Q	Heterozygous *MAL* 6646 nt deletion
P3	AnWj-negative (inherited)	Targeted NGS, Sanger *SMYD1* & *MAL,* breakpoint analysis	Homozygous synonymous mutation in exon 3; 255C>T, H85H (rs1071695; gnomAD v4.0.0 freq. 0.16)	No mutation	Homozygous *MAL* 6646 nt deletion as in P1
P4	AnWj-negative (inherited)	Sanger *SMYD1* & *MAL,* breakpoint analysis	n.t	No mutation	Homozygous *MAL* 6646 nt deletion as in P1
P5	AnWj-negative (suppression)	Targeted NGS, Sanger *SMYD1* & *MAL*	No mutations	No mutation	Wild-type *MAL*, no deletion
P6	AnWj-negative (suppression)	Targeted NGS, Sanger *SMYD1* & *MAL*	Heterozygous mutation in exon 17 2018G>A; R673Q (rs61752932; gnomAD v4.0.0 freq. 0.003)	No mutation	Wild-type *MAL*, no deletion
P7	AnWj-negative (suppression)	Sanger *SMYD1* & *MAL*	n.t	No mutation	Wild-type *MAL*, no deletion
P8	AnWj-negative (inherited)	Targeted NGS, Sanger *SMYD1*	No mutations	Homozygous 959G>A, R320Q	Homozygous *MAL* 6646 nt deletion as in P1
P8_F1	AnWj-negative (inherited)	Targeted NGS, Sanger *SMYD1*	No mutations	Homozygous 959G>A, R320Q	Homozygous *MAL* 6646 nt deletion as in P1

WES: Whole Exome Sequencing

Figure Legends

Figure 1. *MAL* gene and Mal protein organisation. (A) Representation of *MAL* gene showing the four coding exons represented by different coloured cylinders. (B) Schematic showing Mal protein organisation in the membrane, with two predicted external loops. Amino acids are colour-coded according to the exon which encodes them (exon 1; green, exon 2; blue, exon 3; orange, exon 4; purple). (C) All-atom explicit solvent molecular dynamics calculation of modelled Mal protein, depicting membrane insertion. Mal protein is depicted in green cartoon representation while the lipid components are shown as wire representation coloured by atom type. The solvent has been omitted for clarity. Phosphate atoms are represented as gold balls for orientation of the membrane leaflet.

Figure 2. All individuals with inherited AnWj-negative phenotype are homozygous for the same deletion in *MAL*, encompassing exons 3 and 4. (A) Portion of chromosome 2 sequence, spanning *MAL* exons 2-4, from NGS targeted-panel sequencing of three representative individuals. Wild-type sequence is shown in the upper panel, whilst individual P2 (middle) shows no sequencing reads mapping to exons 3 and 4, with reads spanning the deleted area mapping to intron 2 (boxed in purple) and in the region downstream from exon 4 (boxed in green). The bottom panel shows a deletion heterozygote (P2_F4), with only approximately 50% of expected reads mapping to *MAL* exons 3 and 4, and the end of sequencing reads spanning the deleted area clearly visible, mapping in the same region as with the homozygous deletion sample above. (B) Gene schematic showing deletion breakpoints. Wild-type *MAL* (top) consists of four coding exons (blue cylinders), whilst the AnWj-negative individuals lack exons 3 and 4 and parts of the adjacent introns (grey cylinders). Deletion (6646 bp) represented by dashed red line. (C) Details of portions of *MAL* intron 2 (upper left panel) and 3' region downstream from *MAL* (lower right panel) sequence alignment in representative AnWj-negative individual (P2) as compared to wild-type control. Sanger sequencing of P2 across the deletion breakpoints is shown in bottom panel; sequence boxed in purple derives from intron 2, whilst sequence boxed in green derives from the 3' region downstream from exon 4. Position of deletion is indicated by red triangle. Breakpoints confirmed by Sanger and/or NGS sequencing to be identical in all AnWj-negative samples (Table 1).

Figure 3. Anti-Mal binds AnWj-positive red cells, but not AnWj-negative cells. (A) Red cells sequentially incubated with either mouse polyclonal anti-Mal (ab167376, AbCam; 1 in 10) or mouse monoclonal anti-AnWj (H86; 1 in 5) followed by goat anti-mouse IgG, show agglutination in AnWj-positive red cells (top row) but not in AnWj-negative red cells (lower three rows). Two representative examples of inherited AnWj-negative phenotype (P1 and P2_F1), both carrying homozygous *MAL* deletions, and one example of suspected AnWj antigen suppression, without *MAL* mutation, (P5) cells are shown. Negative (secondary antibody only) and positive (anti-GPA; BRIC256) controls were included (data not shown). Scale bars are 40μm. Cells were imaged using a Leica DM750 microscope (Leica Microsystems) at 100x magnification and imaged using a Pixera Penguin 600CL camera (Digital Imaging

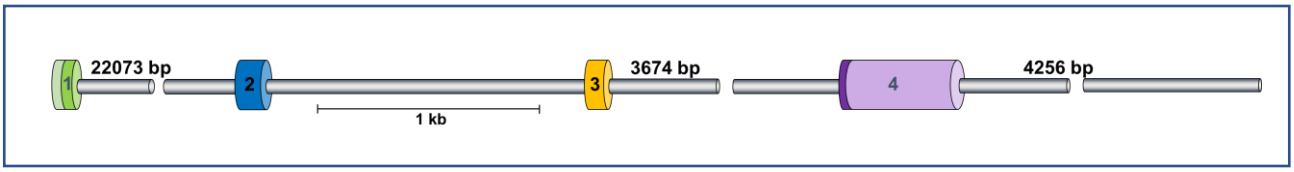
Systems). (B) Flow cytometry with anti-Mal (ab167376) and anti-AnWj (H86) on control RBC (top row) and four representative examples of AnWj-negative cells, two with *MAL* deletion (P1 and P4) and two lacking the *MAL* deletion (P5 and P6). No detectable expression of Mal or AnWj is observed in any AnWj-negative samples. (C) Red cell membranes prepared from one AnWj-positive control and three AnWj-negative samples (P1, inherited; P5 and P6, suppression) were immunoblotted using mouse monoclonal anti-Mal antibody E1 (sc-390687; Santa Cruz). A band consistent with full-size Mal (15 kDa) was present in the control sample and absent in all AnWj-negative samples tested. The anti-protein 4.2 (BRIC273) control demonstrates consistent protein loading. A further two AnWj-negative samples showed the same results (Supplemental Figure 1A). Multiple commercially available anti-Mal antibodies were tested (Supplemental Table 3), but only one worked in our hands by red cell serology/flow cytometry (ab167376) and two, E1 (shown here and Supplemental Figure 1A) and 6D9[47] (Supplemental Figure 1A-B), by immunoblotting.

Figure 4. Over-expression of full-length, but not truncated, Mal protein in BEL-A cells results in expression of AnWj antigen at the cell surface. (A) Flow cytometry with mouse polyclonal anti-Mal (ab167374), human anti-AnWj eluate and monoclonal anti-AnWj (H86) wild-type BEL-A cells (top row), Mal KO BEL-A cells (second row) and BEL-A cells over-expressing either full length Mal or truncated Mal (exons 1 and 2 only). Only BEL-A cells over-expressing full-length Mal (row three) show any expression of Mal protein and AnWj antigen detectable by flow cytometry. (B) Full-length Mal (i), N-terminal GFP tagged full-length Mal (ii) or GFP truncated Mal (iii; exons 1 and 2 only, as potentially expressed in the AnWj-negative patients) were expressed in KO Mal BEL-A cells and observed by confocal microscopy. Full length Mal (ii) was located throughout the cell (but not in the nucleus) but was predominantly observed on the cell surface, whilst truncated Mal (iii) is localised inside the cell (but again not in the nucleus). Yellow scale bars are 10 µm and white are 5µm. Samples were imaged at 22°C using 40x oil immersion lenses (magnification 101.97µm at zoom 3.8, 1.25 NA) on a TCS SP8 confocal imaging system (Leica). Images were obtained using Leica LAS AF software.

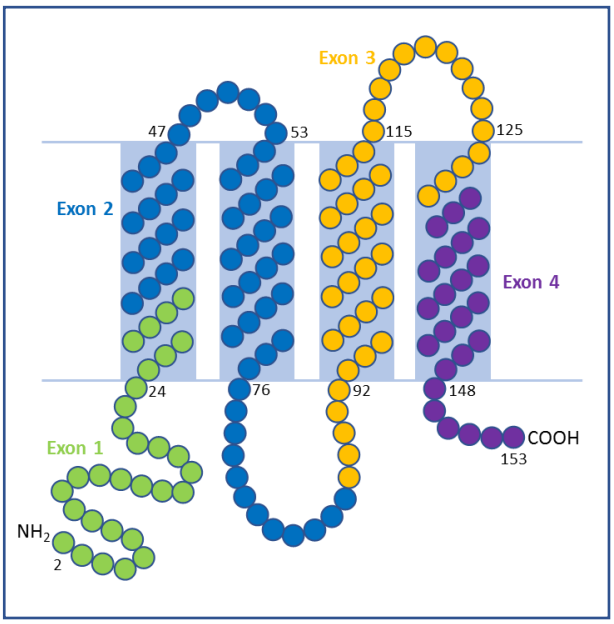
Figure 5. CD44 is neither sufficient, nor required for, expression of AnWj antigen, and CD44 expression is not affected by lack of Mal protein. (A) Flow cytometry with anti-Mal (ab167374), anti-CD44 (BRIC222) and anti-AnWj (H86) on wild-type BEL-A cells (top row), CD44 KO BEL-A cells (second row), and CD44 KO BEL-A cells over-expressing either Mal or CD44. Wild-type BEL-A cells express CD44, which is abolished in the CD44 KO BEL-A cells. Over-expression of Mal in CD44 KO BEL-A cells results in detectable expression of Mal and AnWj antigen, which is not observed in either wild-type, CD44 KO, or CD44 KO BEL-A cells rescued by over-expression of CD44. (B) Flow cytometry with anti-Mal (ab167374), anti-CD44 (BRIC222) and anti-AnWj (H86) on wild-type BEL-A cells (top row), Mal KO BEL-A cells (second row) and BEL-A cells over-expressing Mal (row 3). CD44 expression is not altered by either KO or OE of Mal protein.

Figure 1

A


1
22073 bp
2
1 kb
3
3674 bp
4
4256 bp

B


Exon 3
47
53
115
125
Exon 2
Exon 4
24
76
92
148
COOH
153
Exon 1
NH2
2

C

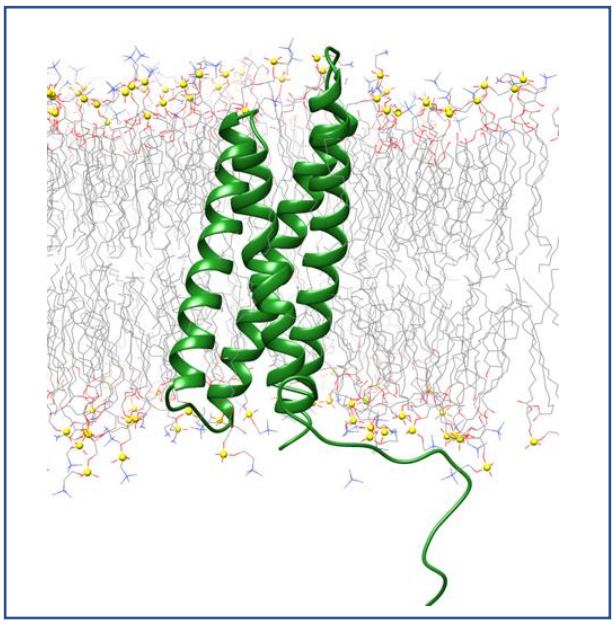


Figure 2

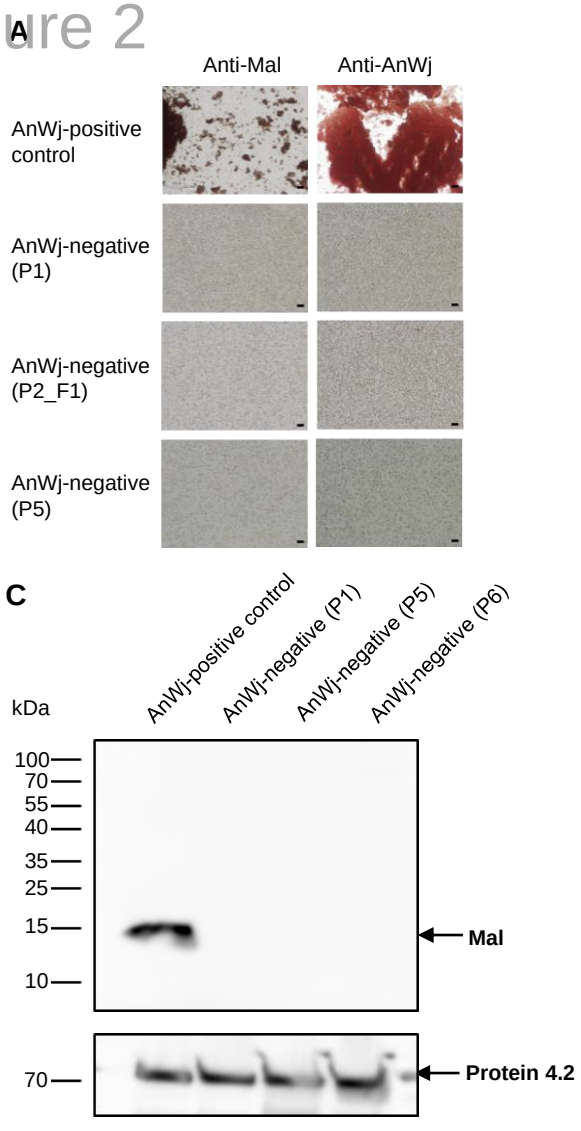

A
Anti-Mal
Anti-AnWj
AnWj-positive control
AnWj-negative (P1)
AnWj-negative (P2_F1)
AnWj-negative (P5)
B
Anti-Mal
Anti-AnWj
AnWj-positive control
AnWj-negative (P1)
AnWj-negative (P4)
AnWj-negative (P5)
AnWj-negative (P6)
C
AnWj-positive control
AnWj-negative (P1)
AnWj-negative (P5)
AnWj-negative (P6)
kDa
100
70
55
40
35
25
15
10
70
Mal
Protein 4.2

Figure 3

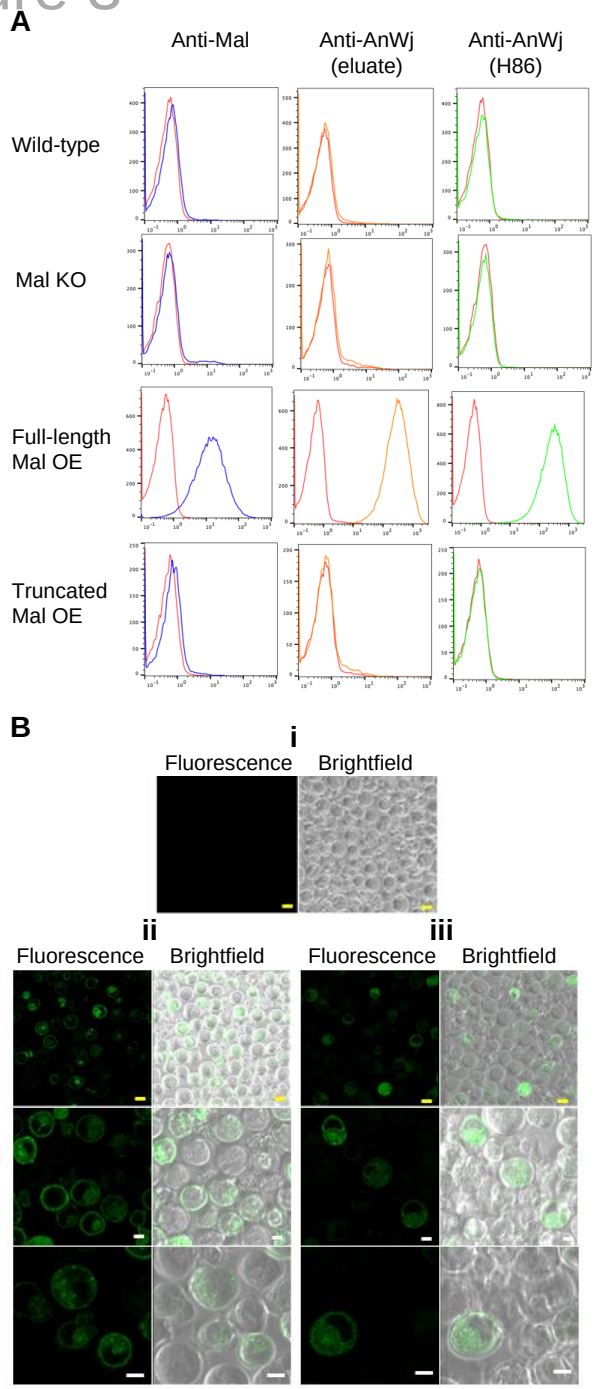

A
Anti-Mal
Anti-AnWj (eluate)
Anti-AnWj (H86)
Wild-type
Mal KO
Full-length Mal OE
Truncated Mal OE
B
i
Fluorescence
Brightfield
ii
Fluorescence
Brightfield
iii
Fluorescence
Brightfield

Figure 4

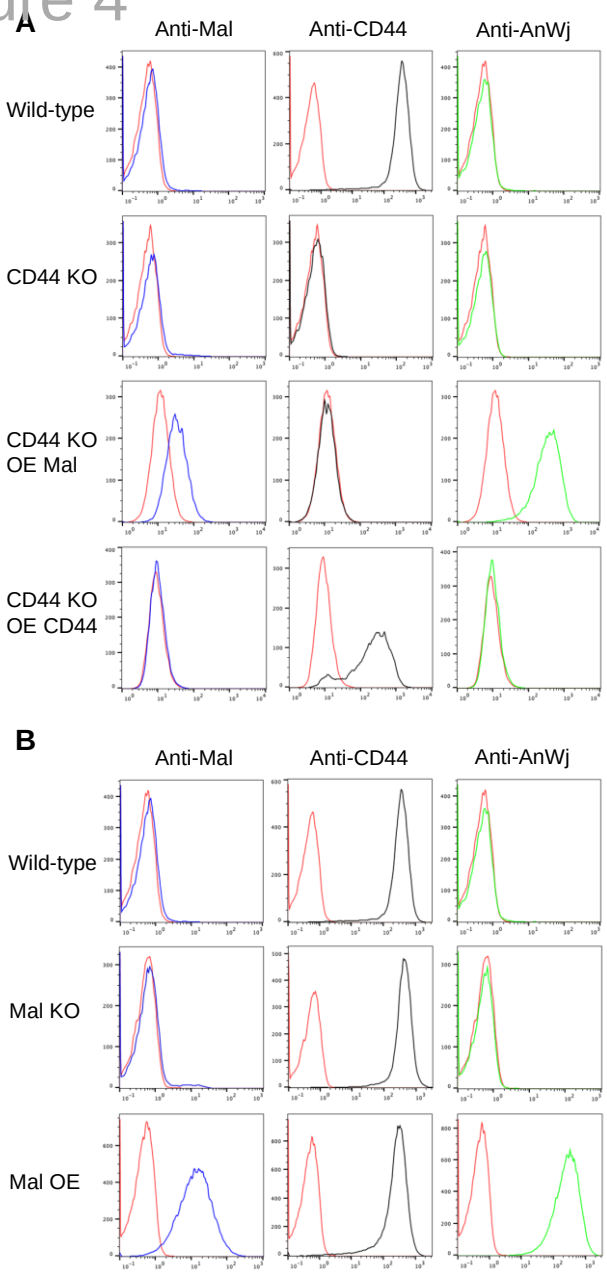
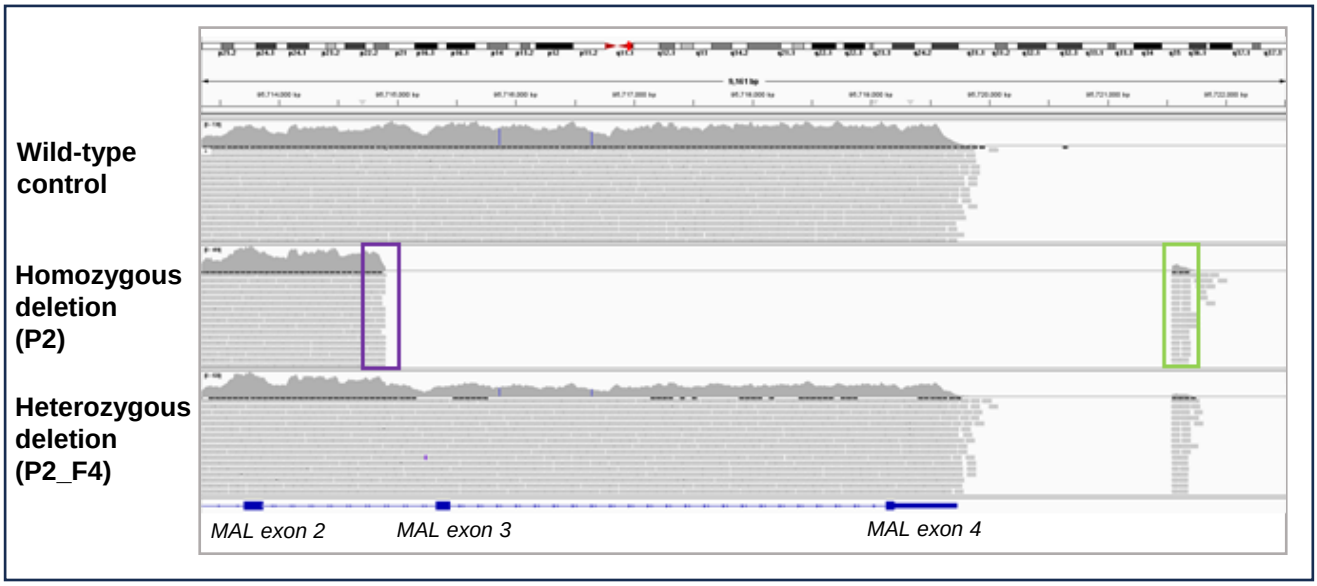
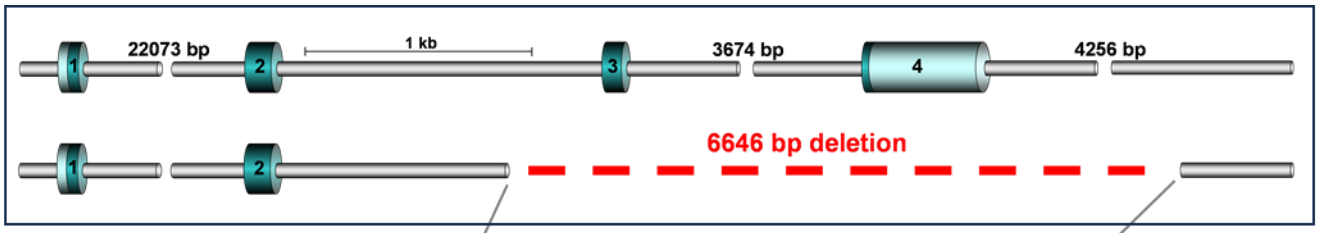

A
Anti-Mal
Anti-CD44
Anti-AnWj
Wild-type
CD44 KO
CD44 KO OE Mal
CD44 KO OE CD44
B
Anti-Mal
Anti-CD44
Anti-AnWj
Wild-type
Mal KO
Mal OE

Figure 5

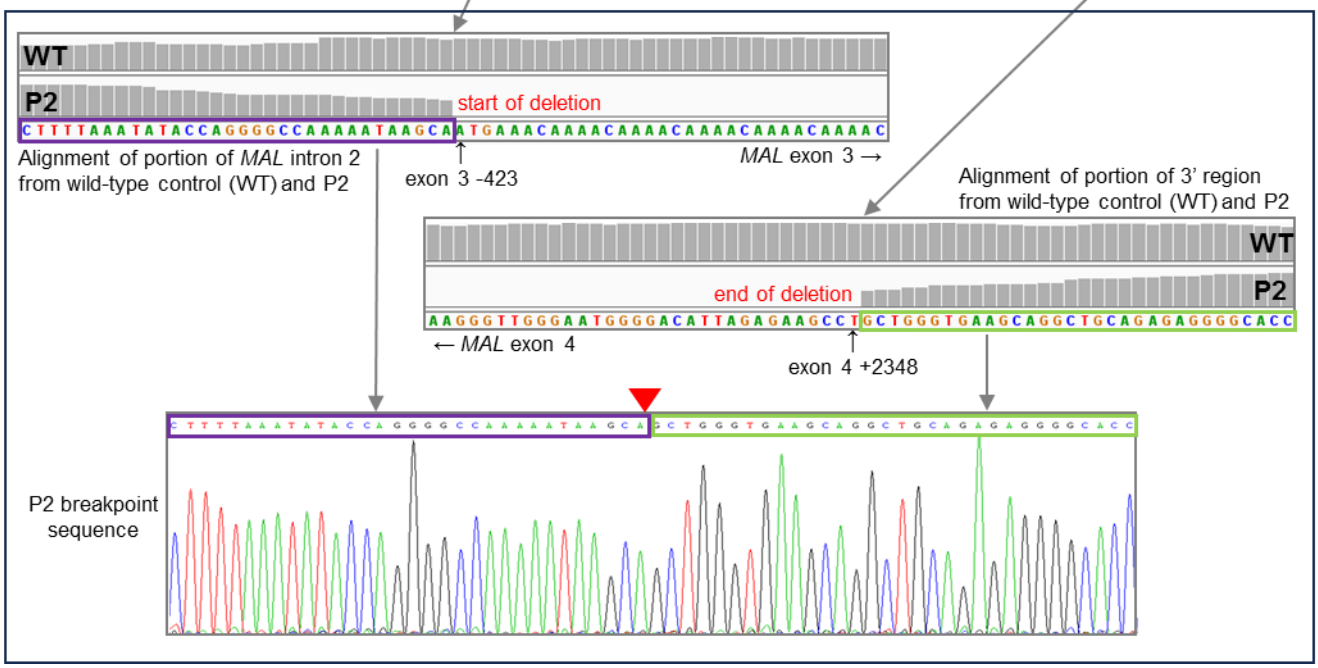
A


Wild-type control
Homozygous deletion (P2)
Heterozygous deletion (P2_F4)
MAL exon 2
MAL exon 3
MAL exon 4

B


22073 bp
1 kb
3674 bp
4256 bp
1
2
3
4
6646 bp deletion

C


WT
P2
start of deletion
CTTTTAAATATACCAGGGGCCAAAAATAAGCAATGAAACAAAACAAAACAAAACAAAACAAAAC
Alignment of portion of MAL intron 2 from wild-type control (WT) and P2
exon 3 -423
MAL exon 3 →
Alignment of portion of 3' region from wild-type control (WT) and P2
WT
end of deletion
P2
AAGGGTTGGGAATGGGGACATTAGAGAAGCCTGCTGGGTGAAGCAGGCTGCAGAGAGGGGCACC
← MAL exon 4
exon 4 +2348
P2 breakpoint sequence

Supplemental Information

Deletions in the *MAL* gene result in loss of Mal protein, defining the rare inherited AnWj-negative blood group phenotype

Short title: Mal protein carries the AnWj blood group antigen

Louise A. Tilley[1†], Vanja Karamatic Crew[1†], Tosti J. Mankelow,[1,2†], Samah A. AlSubhi[1,3,4†], Benjamin Jones[1], Abigail Borowski[1], Vered Yahalom[5,6], Lilach Finkel[7], Belinda K. Singleton[8], Piers J. Walser[9], Ashley M. Toye[3,10], Timothy J. Satchwell[3,10], and Nicole M. Thornton[1‡]

[1]International Blood Group Reference Laboratory, NHS Blood and Transplant, Bristol, UK. [2]Component Development Laboratory, NHS Blood and Transplant, Cambridge, UK. [3]School of Biochemistry, University of Bristol, Bristol, UK. [4]Laboratory Medicine Department, Faculty of Applied Medical Sciences, Umm Al-Qura University, Makkah, Saudi Arabia. [5]Blood Services & Apheresis Institute, Rabin Medical Center, Petah Tiqva, Israel. [6]Faculty of Medical and Health Science, Tel Aviv University, Tel Aviv, Israel. [7]National Blood Group Reference Laboratory, Magen David Adom National Blood Services, Ramla, Israel. [8]NIHR Blood and Transplant Research Unit in Genomics to Enhance Microbiology Screening, NHS Blood and Transplant, Bristol, UK. [9]Clinical Biotechnology Centre, NHS Blood and Transplant, Bristol, UK. [10]National Institute for Health Research (NIHR) Blood and Transplant Research Unit in Red Blood Cell Products, University of Bristol, Bristol, UK.

[†]shared first authors, [‡] corresponding author

Contents

Supplemental Table 1. Sample information for AnWj-negative individuals and family members

Sample	Sex	Phenotype and Serological Information	Antibody	Clinical history
P1	F	**AnWj negative** Persistent negative phenotype demonstrated.	Yes (the original anti-Anton)[3]	Two children. Antibody first identified during second pregnancy. No HDFN.
P2	F	**AnWj negative** Demonstrated inherited phenotype in large family.[14]	Yes	Multiple pregnancies, antibody detected after 12th pregnancy. No HDFN.
P2_F1	M	**AnWj negative** Family member of P2.	No	
P2_F2	M	**AnWj negative** Family member of P2.	Yes	
P2_F3	M	**AnWj positive** Family member of P2.	N/A	
P2_F4	M	**AnWj positive** Family member of P2.	N/A	
P3	F	**AnWj negative** Negative DAT.	Yes	Multiple pregnancies.
P4	F	**AnWj negative** Negative DAT.	Yes	Not available.
P5	F	**AnWj negative** Query suppression. Negative DAT. Very weakly positive with H86 anti-AnWj (papain technique).	Yes	Acute kidney injury and deep anemia. History of previous transfusion.
P6	F	**AnWj negative** Query suppression**.** Positive DAT. Anti-AnWj eluted from patient's cells, following adsorption. Extensive family study was carried out (three generations, including 15 siblings) all were found to be AnWj-positive.	Yes	Not available.
P7	M	**AnWj negative** Query suppression. History of positive DAT. Very weakly positive with H86 anti-AnWj (papain technique).	Yes	Non-Hodgkins lymphoma.
P8	F	**AnWj negative** Apparent inherited phenotype. Sibling (P8_F1) also AnWj-negative.	Yes	Antibody identified during first known pregnancy.
P8_F1	F	**AnWj negative** Compatible sibling of P8.	No	No known history of pregnancy.

Supplemental Table 2. PCR amplification primers, PCR conditions, and Sanger sequencing primers for *MAL*, *SMYD1* and *CD44* genes used in this study.

Amplified Region	Primer*	Primer Sequence (5' to 3')	Primer Position†	PCR Amplification Enzyme and Annealing Temperature (°C)	Product Size (bp)
MAL exon 1	MAL-1F MAL-1R	CGCTCGTCCCGTCCCAAG CCCCGCTGCCTACATCTGA	Exon 1 (-38 to -20) Intron 1 (+249 to +268)	AmpliTaq Gold, 60	399
MAL exon 2	MAL-2F MAL-2R	AAATGCCTGCCCTGTTCTCTT TGGCACTGGAAAAGCCCATC	Intron 1 (-136 to -115) Intron 2 (+68 to +88)	AmpliTaq Gold, 68	392
MAL exon 3	MAL-3F MAL-3R	GGCATGGGACCTCCGTGAC GGCACAACATGACACTGCCC	Intron 2 (-149 to -130) Intron 3 (+159 to +179)	Expand, 60	454
MAL exon 4	MAL-4F MAL-4R	GAGAAGCAATGACAGCCCAAG GGGAGAGTAAACACAGCACCC	Intron 3 (-129 to -108) Exon 4 (+159 to +180)	AmpliTaq Gold, 65	384
MAL Intron 2 – 3'UTR	MAL-Intr2F MAL-3'UTR	TGCCAGGTGAGACTTCTCCG ACCACATAAGGTCGTCACAGG	Intron 2 (-1371 to -1351) 3' UTR (+4698 to +4720)	AmpliTaq Gold, 68	9965
MAL sequencing primers for breakpoint analysis	MAL-3aF MAL-4aR MAL-3F MAL-4R	GACCTCAGCTCTGCATCTGGG CCATGGACCTCTGGAAAGATCTG GGCATGGGACCTCCGTGAC GGGAGAGTAAACACAGCACCC	Intron 2 (-68 to -47) Exon 4 (+464 to +487) Intron 2 (-149 to -130) Exon 4 (+159 to +180)	N/A‡	N/A
SMYD1 exon 7	SMYD1-7aF SMYD1-7aR	CACCTTAGCCAACAGGACCA ATTGGTCATTAACACTGGAGCAC	Intron 6 (-181 to -161) Intron 7 (+160 to +183)	AmpliTaq Gold, 60	527
CD44 exon 1	CD44_1F CD44_1R	ACTTCCGAGGCAGCCTCATTG ACACAATTCTCCAACGGTTTAGCG	exon 1 (-147 to -127) intron 1 (+156 to +178)	AmpliTaq Gold, 65	392
CD44 exon 2	CD44_2F CD44_2R	TGTTAACCAGGCTGGTCTTGAG AGTTCTAAGCCCAGCTGCCTG	intron 1 (-196 to -175) intron 2 (+48 to +68)	AmpliTaq Gold, 65	430
CD44 exon 3	CD44_3F CD44_3R	TAACTCGGTTGTTGAAACCTCCG AGCTGAGCTCCAAAGACCAGG	intron 2 (-141 to -119) intron 3 (+46 to +66)	AmpliTaq Gold, 60	341
CD44 exon 4	CD44_4F CD44_4R	GCTTCCACAGTGCCTGATATAG TACCACAGAGAACACACCTGAG	intron 3 (-286 to -265) intron 4 (+51 to +72)	AmpliTaq Gold, 60	427
CD44 exon 5	CD44_5F CD44_5R	TCTCCCACCACTGGATAGATAGG GTGCAATGCTCAGGAAGGTCAG	intron 4 (-132 to -110) intron 5 (+55 to +76)	AmpliTaq Gold, 60	439
CD44 exon 15	CD44_15F CD44_15R	GATAGGCTGTATAAGAATGCAAAGG GTGTCAGTATTACCAGGGAACTG	intron 14 (-92 to -68) intron 15 (+223 to +245)	AmpliTaq Gold, 65	400
CD44 exon 16	CD44_16F CD44_16R	GGAGAGCTGCCCTTTATGCAG AGCTAGTTTGCAGAACCCAGG	intron 15 (-86 to -66) intron 16 (+200 to +220)	AmpliTaq Gold, 65	378
CD44 exon 17	CD44_17F CD44_17R	CTGTGGTGCTTGTTTCAACTAGG AGGGACTACGCTCTGAGCAG	intron 16 (-101 to -79) intron 17 (+69 to +88)	AmpliTaq Gold, 60	268

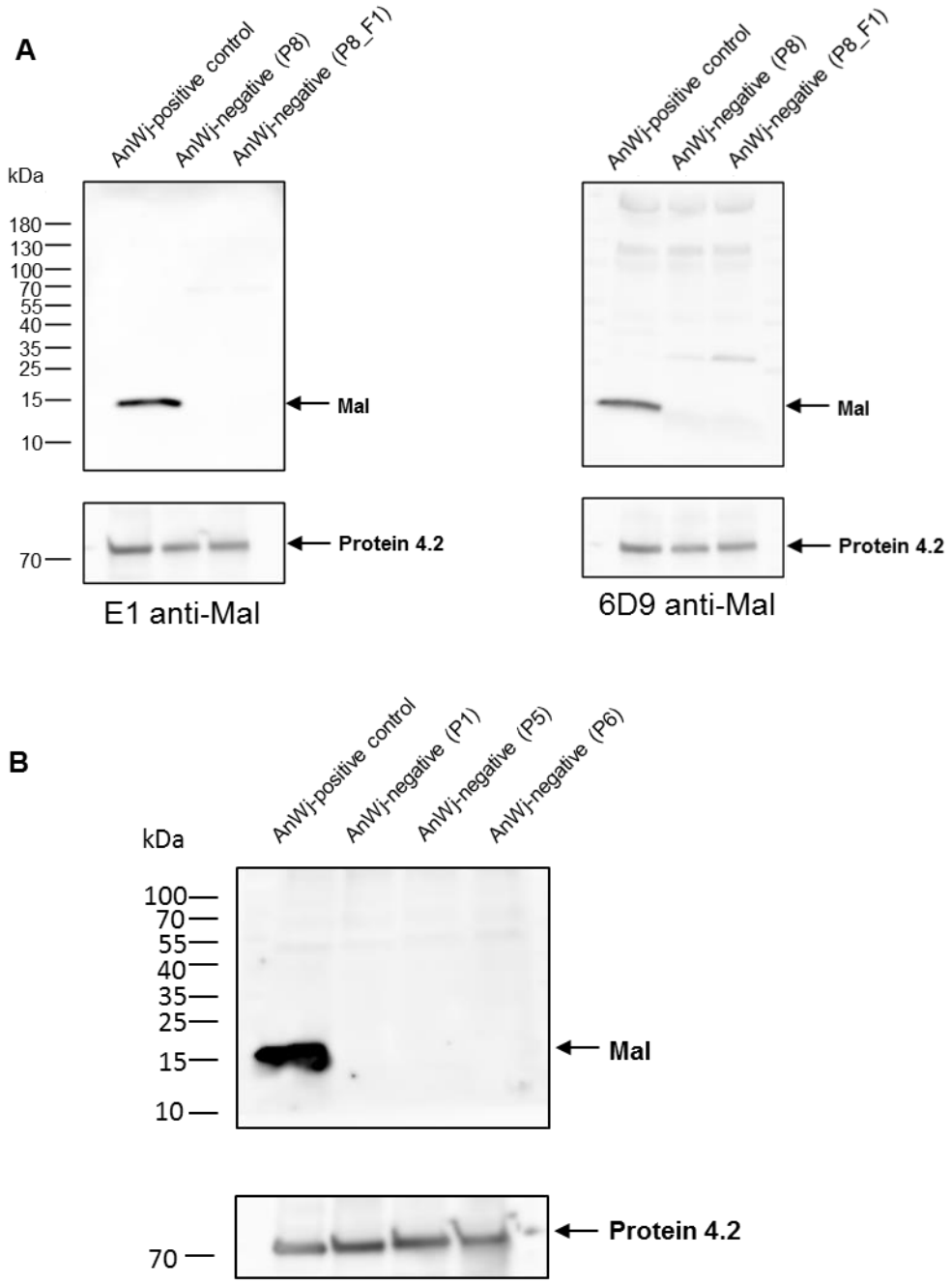
* F and R denote forward (sense) and reverse (antisense) direction, respectively.
†Primer positions relative to the respective exon/intron boundary.
‡Not applicable

Supplemental Table 3. Anti-Mal antibodies tested during this study by indicated techniques.

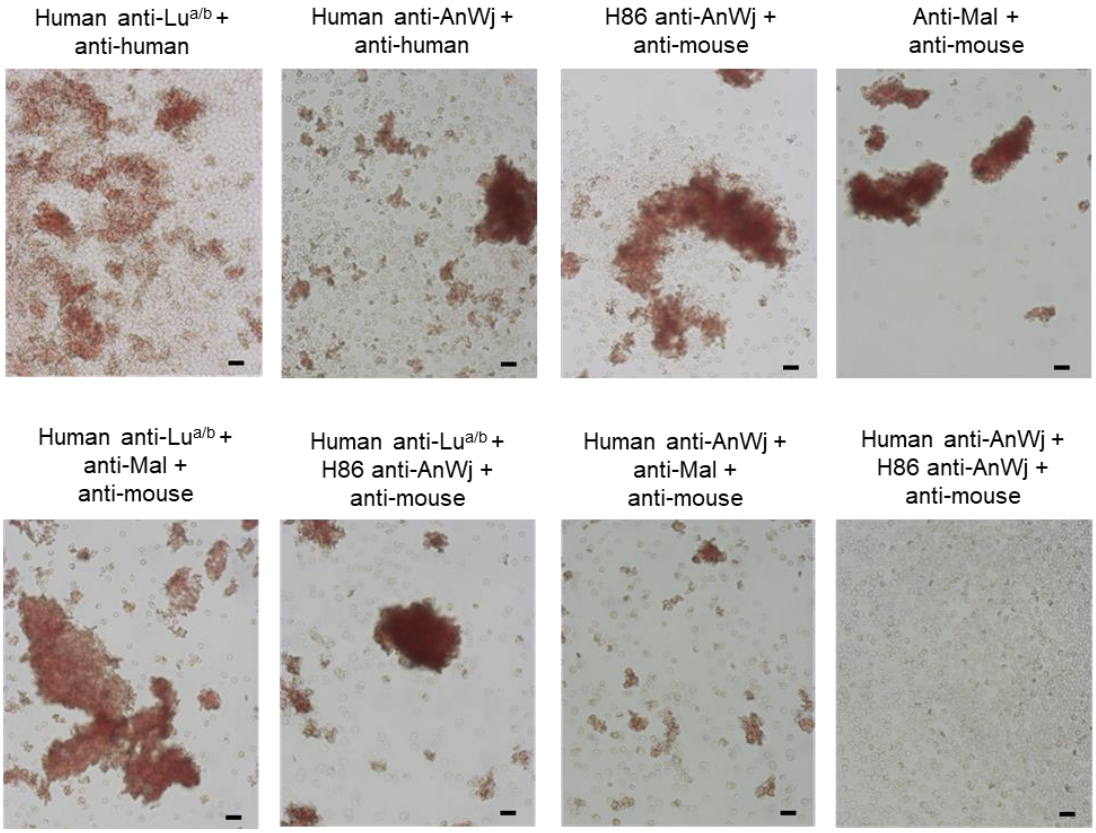
Antibody clone	Clonality	Immunogen	Source	Techniques			
				Serology	Flow Cytometry	Immunobloting	Confocal Microscopy
ab167374	Mouse polyclonal	Recombinant full-length human Mal (aa 1-153)	AbCam	yes	yes	no	NT
E-1, sc-390687	Mouse monoclonal	Internal region of human Mal (aa 61-130)	Santa Cruz	NT	no	yes	NT
6D9	Mouse monoclonal	Peptide himan Mal (aa 114-123)	M. Alonso, gift	NT	no	yes	NT
ab15418	Rabbit polyclonal	Synthetic peptide human Mal (aa 114-126)	AbCam	NT	no	no	no
orb45962	Rabbit polyclonal	Recombinant human Mal (aa 1-84)	Biorbyt	NT	no	no	no
B5-G3, MA5-32924	Mouse monoclonal	Synthetic peptide mouse Mal (aa 97-146)	Invitrogen	NT	no	no	NT
SAB1411741	Mouse polyclonal	Recombinant full-length human Mal (aa 1-153)	Sigma-Aldrich	no	no	NT	NT

NT = not tested

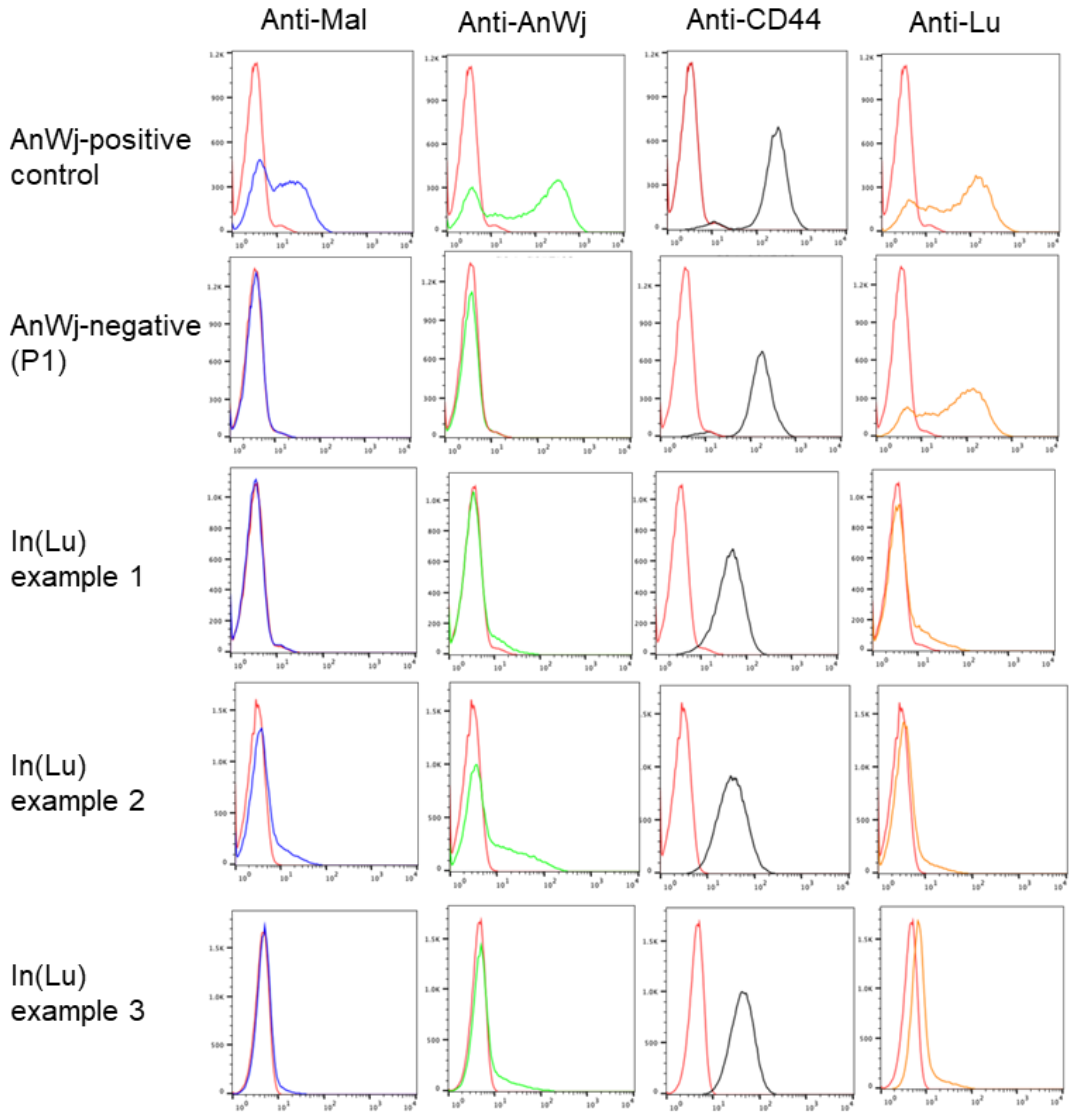
Only antibodies highlighted in grey worked in our hands by the techniques indicated (one by red cell serology/flow cytometry and two by immunoblotting)



Supplemental Figure 1. Western blotting shows full-length Mal is present on AnWj-positive red cell membranes but absent on AnWj-negative membranes. (A) Red cell membranes were prepared from one AnWj-positive control and two AnWj-negative samples (P8, P8_F1). Immunoblotting was performed using mouse monoclonal anti-Mal antibodies E1 (sc-390687; Santa Cruz, left image) and 6D9[47] (right image), with 50 µg of protein loaded per lane. A band consistent with full-size Mal (15 kDa) was present in the control sample and absent in P8 and P8_F1, as seen with other AnWj-negative samples tested (Figure 3B). The anti-protein 4.2 (BRIC273) control demonstrates consistent protein loading. (B) Red cell membranes prepared from one AnWj-positive control and three AnWj-negative samples (P1, P5, P6) immunoblotted with 6D9, replicating the results seen in Figure 3B with E1 anti-Mal.



Supplemental figure 2. Binding of anti-Mal is inhibited by binding of anti-AnWj to AnWj-positive red cells. Red cells sequentially incubated with either a 50/50 mix of human plasma containing anti-Lu^{a} and anti-Lu^{b} (anti-Lu$^{a/b}$), human plasma containing anti-AnWj, mouse polyclonal anti-Mal (ab167374) or H86 (mouse monoclonal anti-AnWj) followed by either goat anti-human IgG or goat anti-mouse IgG (with intervening wash steps), show agglutination (upper row). Incubation of red cells with anti-Lu$^{a/b}$ prior to either anti-Mal or H86 in the agglutination assay has no effect on the level of agglutination caused by these two antibodies (bottom row panels 1 and 2). However, incubation with human plasma containing anti-AnWj severely reduces the agglutination observed with anti-Mal and almost completely blocks agglutination with H86 (bottom row panels 3 and 4). Incubation of cells with either anti-Lu$^{a/b}$, human plasma containing anti-AnWj or BRIC163 (mouse monoclonal to intracellular glycophorin A) followed by goat anti-mouse IgG caused no agglutination at all (data not shown). Incubation with BRIC256 (mouse monoclonal to extracellular glycophorin A) followed by goat anti-mouse IgG caused maximal agglutination (as observed in Figure 2) which was not reduced if the cells had been incubated with either anti-Lu$^{a/b}$ or human plasma containing anti-AnWj beforehand (data not shown). Scale bars are 20 μm. Cells were imaged using a Leica DM750 microscope (Leica Microsystems) at 100x magnification and imaged using a Pixera Penguin 600CL camera (Digital Imaging Systems).



Supplemental Figure 3. CD44 is expressed on AnWj-negative red cells, and weakly expressed on In(Lu) cells. Flow cytometry with mouse polyclonal anti-Mal (ab167374), monoclonal anti-AnWj (H86), anti-CD44 (BRIC222) and anti-Lu (BRIC221 and BRIC224) on AnWj-positive control RBC (top row), AnWj-negative RBC (P1; second row) and three examples of In(Lu) phenotype cells (bottom three rows). Expression of CD44 is weaker in In(Lu) cells than in AnWj-positive control cells, or the AnWj-negative patient cells. Expression of Lutheran is extremely weak or negative in the In(Lu) cells. Expression of Mal and AnWj is almost completely absent in the In(Lu) cells, although very weak expression is observed in In(Lu) example 2. This does not appear to be related to changes in CD44 expression, which appear reasonably consistent between the three In(Lu) examples