



TRANSFUSION MEDICINE

Disruption of a GATA1-binding motif upstream of *XG/PBDX* abolishes Xg^a expression and resolves the Xg blood group system

Mattias Möller,¹ Yan Quan Lee,¹ Karina Vidovic,¹ Sven Kjellström,^{2,3} Linda Björkman,⁴ Jill R. Storry,^{1,4} and Martin L. Olsson^{1,4}

¹Division of Hematology and Transfusion Medicine, Department of Laboratory Medicine, ²Department of Clinical Sciences Lund, and ³Center of Excellence in Biological and Medical Mass Spectrometry, Lund University, Lund, Sweden; and ⁴Department of Clinical Immunology and Transfusion Medicine, Office of Medical Services, Lund, Sweden

KEY POINTS

- Expression of the Xg blood group protein is governed by rs311103, and its minor allele disrupts a GATA motif to cause the $Xg(a-)$ phenotype.
- These data elucidate the genetic basis of the last unresolved blood group system and make genotyping for Xg^a status possible.

The Xg^a blood group is differentially expressed on erythrocytes from men and women. The underlying gene, *PBDX*, was identified in 1994, but the molecular background for Xg^a expression remains undefined. This gene, now designated *XG*, partly resides in pseudoautosomal region 1 and encodes a protein of unknown function from the X chromosome. By comparing calculated Xg^a allele frequencies in different populations with 2612 genetic variants in the *XG* region, rs311103 showed the strongest correlation to the expected distribution. The same single-nucleotide polymorphism (SNP) had the most significant impact on *XG* transcript levels in whole blood ($P = 2.0 \times 10^{-22}$). The minor allele, rs311103C, disrupts a GATA-binding motif 3.7 kb upstream of the transcription start point. This silences erythroid *XG* messenger RNA expression and causes the $Xg(a-)$ phenotype, a finding corroborated by SNP genotyping in 158 blood donors. Binding of GATA1 to biotinylated oligonucleotide probes with rs311103G but not rs311103C was observed by electrophoretic mobility shift assay and proven by mass spectrometry. Finally, a luciferase reporter assay indicated this GATA motif to be active for rs311103G but not rs311103C in HEL cells. By using an integrated bioinformatic and molecular biological approach, we elucidated the underlying

genetic basis for the last unresolved blood group system and made Xg^a genotyping possible. (*Blood*. 2018;132(3):334-338)

Introduction

Unraveling the molecular genetic bases of blood groups has facilitated implementation of novel concepts and diagnostic tools in transfusion medicine and related fields. Recent developments include the elucidation of several important blood group systems like JR,^{1,2} Lan,³ FORS,⁴ Vel,⁵⁻⁷ and AUG.⁸ Furthermore, understanding of the molecular genetic mechanism underlying P1 antigen expression was recently reported.^{9,10} Currently, all blood group systems except 1 have been resolved, and their polymorphic antigens have been predicted by genotyping efforts.¹¹

Anti- Xg^a was first described in 1962 by Mann et al.¹² Approximately 66% of men and 90% of women express the Xg^a antigen on their red blood cells (RBCs). Because of its skewed sex distribution, Xg was the first blood group system to be assigned to a specific chromosome: X. The underlying gene, *PBDX*, was identified by Ellis et al,¹³ but despite these landmark studies, Xg has remained the only system for which the genetic basis of antigen negativity is unknown. *PBDX*, now renamed *XG*, partly resides in pseudoautosomal region 1 (PAR1) on both sex chromosomes; the

first 3 exons lie in PAR1, whereas the remaining 7 exist only on the X chromosome (Figure 1A). Thus, *XG* is disrupted on the Y chromosome and results in no protein product.¹⁴ Consistent with its location across PAR1, *XG* is one of few X-borne genes not inactivated. Despite >50 years of investigations into this enigmatic blood group, the presence or absence of Xg^a on RBCs cannot yet be predicted by genotyping, and the function of the Xg protein remains unknown.

Using an integrated bioinformatic and molecular biological approach, we aimed to establish the genetic basis underlying the $Xg(a+)$ vs $Xg(a-)$ phenotype. We hypothesized that Xg^a expression is transcriptionally regulated by a single SNP within the *XG* region, potentially disrupting an erythroid transcription factor binding site.

Study design

Calculated Xg^a allele frequencies in different populations were compiled from historical data based on Xg^a phenotyping (supplemental Table 1, available on the *Blood* Web site). Comparisons were made with frequencies for multiple *XG*

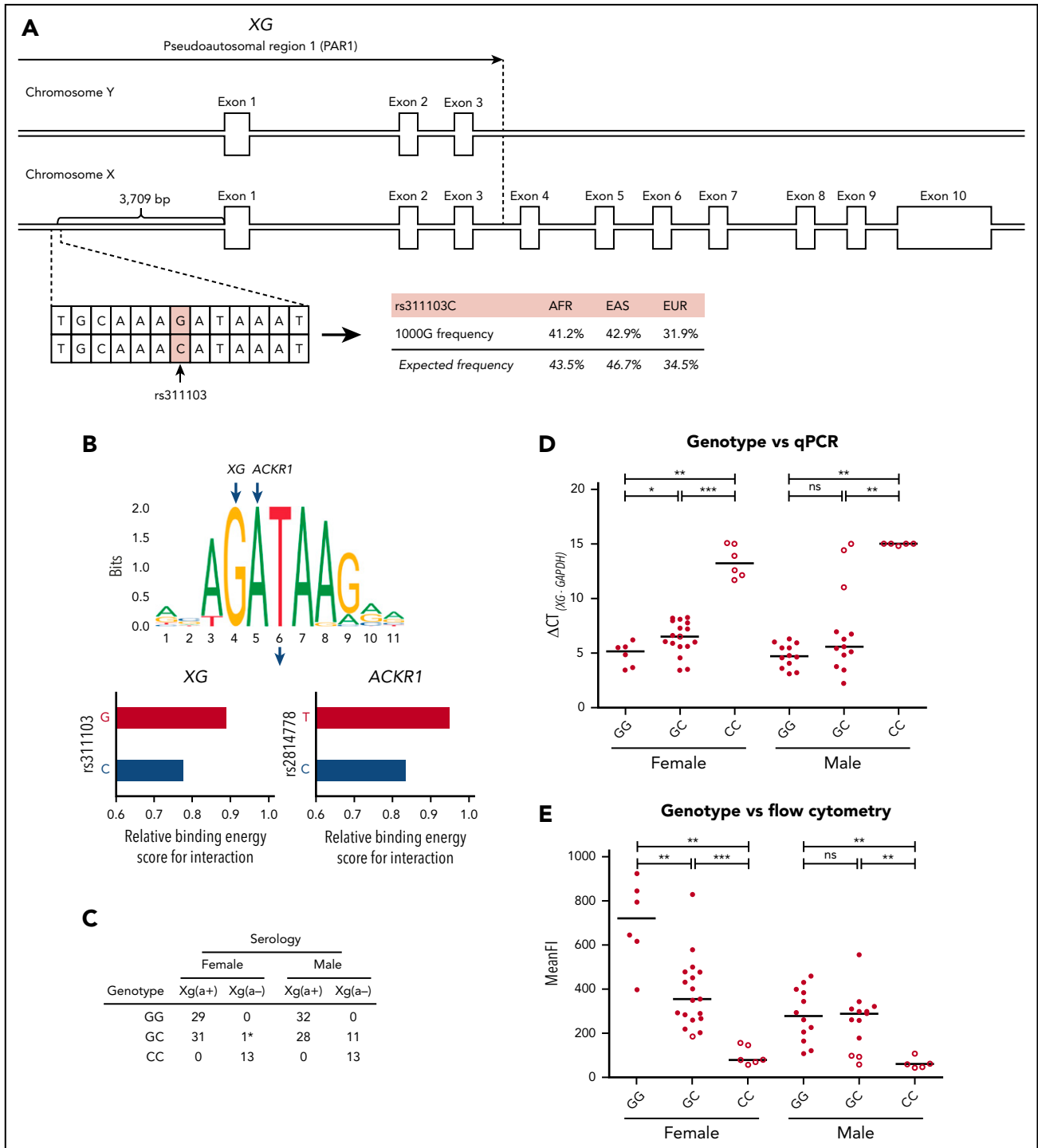


Figure 1. The rs311103C allele correlates with the Xg(a⁻) phenotype. (A) The XG gene partly resides in pseudoautosomal region 1 (PAR1) on both sex chromosomes; the first 3 exons lie in PAR1, whereas the remaining 7 exist only on the X chromosome. Thus, XG is truncated to the first 3 exons on the Y chromosome, where it does not produce a functional transcript. The frequencies of rs311103C, a single-nucleotide polymorphism (SNP) located 3709 bp upstream of the transcription start site, as derived from the 1000 Genomes Project¹⁵ closely match the expected allele distributions for Xg^a negativity. (B) Transcription factor binding analyses of the GATA motif identified at rs311103 show decreased binding preferences with nucleotide substitutions, as adapted from the JASPAR database.¹⁷ The XG GATA motif with rs311103C converts GATA to CATA and reduces the relative binding energy score for GATA1 from 0.888 to 0.775, thus bringing it below the default threshold at 0.8. Similarly, the ACKR1 upstream GATA site may carry the rs2814778C SNP on the complementary strand and results in a similarly lowered score for GATA1 and the Fy(b⁻) phenotype.¹⁸ (C) The 158 donors were serologically typed as Xg(a⁺) or Xg(a⁻) and genotyped for the rs311103 allele. The asterisk indicates 1 serologically Xg(a⁻) heterozygous female donor who was weakly positive on flow cytometric analysis; see also panel E. The XG complementary DNA (cDNA) sequence from this donor was found to be identical to the reference sequence (NM_175569.2). (D) For a subset of donors (n = 59; 29 female and 30 male donors), messenger RNA (mRNA) was isolated and converted to cDNA for XG transcript analysis. (E) We also performed flow cytometry on this subset. Genotypes of the donors correlated with XG transcript levels by reverse transcription quantitative polymerase chain reaction (RT-qPCR) and Xg^a antigen levels by flow cytometry (mean fluorescence intensity [MeanFI]); solid and open circles represent serologically Xg(a⁺) and Xg(a⁻) samples, respectively, on 29 females and 30 males. Note the overlap between 2 dots on the median line for the GC males. Bars represent the median. *P < .05, **P < .01, ***P < .001. ns, not significant.

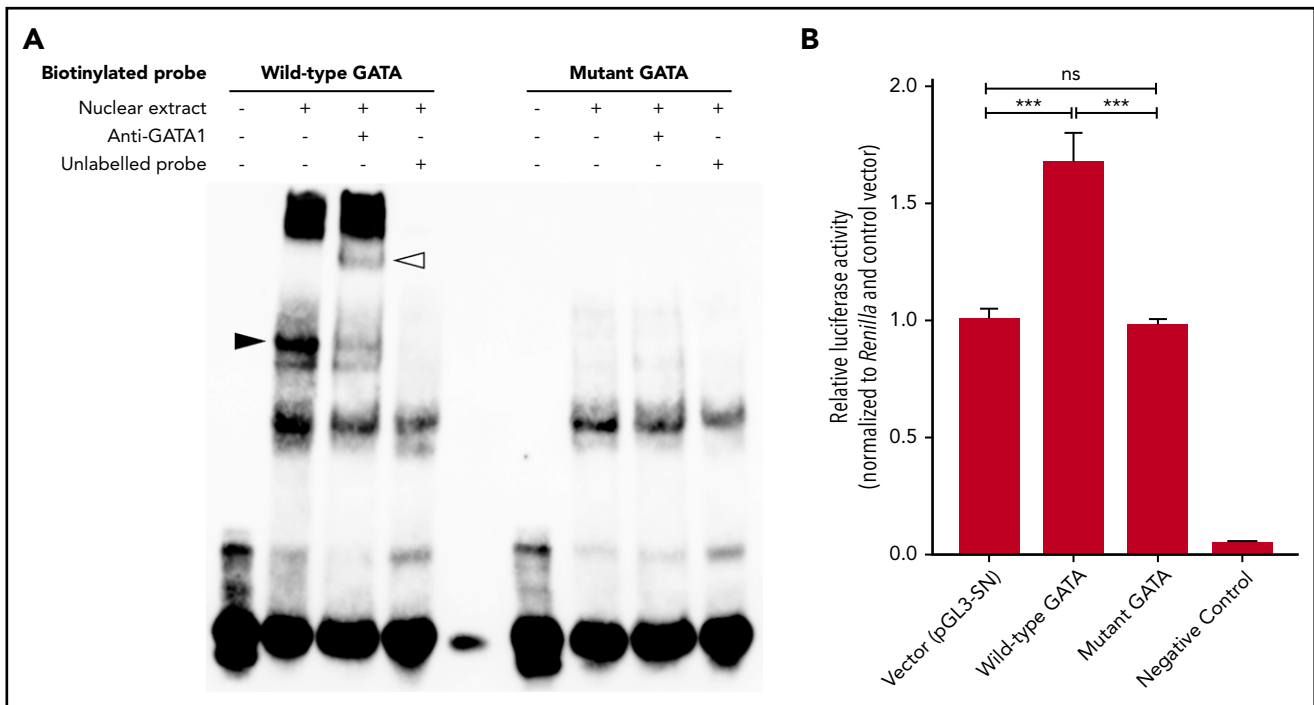


Figure 2. Functional assays indicate that the XG upstream GATA motif binds GATA1 and can enhance transcription. (A) Electrophoretic mobility shift assays were performed with 35-bp biotinylated probes spanning the rs311103 SNP. Only the wild-type probe exhibited a shift (black arrowhead) upon incubation with nuclear extract from K562 cells and a further supershift (white arrowhead) with addition of anti-GATA1, highlighting its affinity for GATA1. Preincubation with 200-fold unlabeled probes abolished the mobility shifts, indicating specificity. The figure is representative of 3 independent experiments. (B) HEL cells were transfected with plasmids carrying the wild-type or mutant XG GATA sequence and a luciferase gene driven by the ABO promoter ($n = 9$). The relative luciferase activity obtained with the pGL3-SN vector was used as reference and normalized to 1. Data represent mean values; error bars indicate standard errors of the mean. *** $P < .001$.

variants as found in the 1000 Genomes Project.¹⁵ Expression quantitative trait loci were analyzed in the GTEx Portal.¹⁶ Transcription factor binding site analysis was performed in JASPAR.¹⁷

Phenotype/genotype correlation was performed on blood samples from 158 anonymized blood donors. An electrophoretic mobility shift assay with biotinylated oligonucleotide probes and mass spectrometric analysis of protein pulldowns were performed as described previously,⁹ and a luciferase reporter assay was run to assess function. Details on experiments are provided in supplemental Methods.

Results and discussion

Among 2612 investigated variants in the XG region, rs311103G/C located 3709 bp upstream of the erythroid transcription start site (Figure 1A) was identified as the SNP with the strongest correlation to the expected distribution (supplemental Table 2). Furthermore, rs311103 not only showed the best fit to the 1000 Genomes Project super populations¹⁵ (Figure 1A) but was also identified as the eQTL with the most significant impact on XG transcript levels in whole blood in the GTEx Portal (normalized effect size, -0.59 for C, the minor allele; $P = 2.0 \times 10^{-22}$; supplemental Table 3).¹⁶ This was in stark contrast to the other 47 tissues tested, where no effect of this SNP was noted (supplemental Figure 1). In addition, transcription factor binding site analysis identified disruption of a GATA family-binding motif by the minor allele rs311103C (supplemental Table 4), which lowered the relative binding energy score (Figure 1B). This drop is comparable to the decrease in binding score observed for c.-67T>C in the

GATA1-binding site of the *ACKR1* promoter, known to cause erythroid silencing of *Fy^b* (*FY*02N.01*) and resistance to *Plasmodium vivax* invasion in individuals of African descent.¹⁸

To test if rs311103 determines Xg^a phenotype, 158 blood donors anonymized other than for sex were serologically typed for Xg^a by hemagglutination and flow cytometry. Initially, Sanger sequencing and, subsequently, a TaqMan SNP genotyping assay were used to determine rs311103 genotype, and mRNA analysis was correlated with Xg^a genotype and antigen expression (Figures 1C-E; supplemental Figure 2).

All female Xg(a⁻) samples identified ($n = 13$ [17.6%] of 74) were homozygous for the minor allele (C), whereas all clearly Xg(a⁺) samples ($n = 120$) regardless of sex carried at least 1 copy of the major allele (G). A sample that phenotyped Xg(a⁻) with 2 anti-Xg^a reagents (but that demonstrated weak reactivity by flow cytometry) was heterozygous, indicating that XG genotyping may overcome serological challenges with low sensitivity. Of the male Xg(a⁻) samples identified ($n = 24$ [28.6%] of 84), all carried at least 1 C allele, but in 11 (45.8%) of 24 of the samples, this was accompanied by G (Figure 1C). Because the X and Y chromosomes are assumed to be homologous in this region, the G allele in these samples is likely Y chromosome derived. However, attempts to obtain Y chromosome-specific amplicons by long-range polymerase chain reaction (~40 kb) were unsuccessful. Real-time polymerase chain reaction was used to quantify mRNA transcripts from 59 samples (Figure 1D). Xg(a⁻) individuals regardless of sex had low to undetectable XG mRNA, suggesting that rs311103C prevents transcription of XG.

To determine if the GATA motif in this enhancer region is functional, electrophoretic mobility shift assay was performed. Strong binding was observed, with biotinylated oligonucleotides corresponding to rs311103G but not C (Figure 2A; supplemental Figure 3). Supershifts were noted after addition of anti-GATA1, and conversely, all binding was inhibited by addition of unlabeled probe. Oligonucleotide probe/nuclear extract complexes were analyzed by liquid chromatography–tandem mass spectrometry, and GATA1 was identified in the complex bound by the wild-type oligonucleotide probe only (supplemental Tables 5 and 6). GATA1 binding at rs311103 was further corroborated by available ChIP-seq data (supplemental Figure 4).¹⁹ Finally, we used luciferase reporter assays to show that the intact GATA1-binding motif could drive transcription of a downstream gene (Figure 2B).

Taken together, the *in vitro* data support the *in silico* prediction that the Xg(a+) blood group phenotype depends on an intact GATA1-binding motif 3.7 kb upstream of the XG transcription start site. The Xg(a−) phenotype is therefore the consequence of markedly decreased erythroid transcript levels, which in turn follow from disruption of the GATA1 site. Of other possible candidate regulatory SNPs identified, most were linked to rs311103, and none of the top candidates disrupted other potential GATA-binding motifs (supplemental Table 3). The regulatory site XGR, between XG and its neighbor gene, *MIC2*, was postulated as early as 1987,²⁰ even before Xg^a expression was linked to *XG/PBDX*.¹³ *MIC2* (now *CD99*) encodes CD99, which is also expressed on RBCs and other tissues.²¹ Interestingly, the expression of CD99 on RBCs correlates with Xg^a antigen status and sex.²² Although the function of Xg protein is completely unknown, it shows moderate homology with CD99, a widely distributed adhesion molecule involved in leukocyte migration and lymphocyte maturation.²³

A vast majority of protein blood groups depend on SNP-based structural changes in the antigen-carrying protein,²¹ which leads to risk of an immune response when individuals are exposed to foreign antigens during pregnancy, transfusion, or transplantation. Anti-Xg^a is a relatively unusual blood group specificity, given how common Xg(a+) transfusion is to Xg(a−) individuals. Even if this were the result of weak antigenicity, our results offer a plausible explanation as to how trace amounts of Xg on RBCs, or Xg expression on nonerythroid cells, may prevent Xg^a immunization. Strikingly, the major antigens of the 2 last blood group systems to be resolved at the genetic level (P1 and Xg^a) both turned out to be quantitatively regulated to cause the antigen-negative phenotypes P₂ and Xg(a−), respectively.^{9,10} In fact, the ErythroGene database reveals only 1 nonsynonymous XG SNP

(c.178G>A; p.Asp60Asn) with a frequency ≥1% among the 2504 individuals in the 1000 Genomes database.²⁴ Importantly, this SNP does not correlate to Xg^a status on RBCs (unpublished data).

We have solved a longstanding conundrum in the field of immunohematology and opened up the possibility of predicting Xg^a status of blood donors and transfusion recipients by rs311103 genotyping. Future studies are required to address the seemingly heterozygous men in whom prediction is hampered by the identical 5′ end of XG on the Y chromosome.

Acknowledgments

The authors acknowledge Marion Darlison for providing technical assistance.

This study was supported by the Knut and Alice Wallenberg Foundation (grant #2014.0312) (M.L.O.), the Swedish Research Council (grant #2014-71X-14251) (M.L.O.), and governmental Avtal om Läkarutbildning och Forskning grants (#ALFSKANE-446521) (M.L.O.) to University Healthcare in Skåne, Sweden.

Authorship

Contribution: M.M. performed bioinformatic analyses and interpreted data; Y.Q.L., K.V., S.K., L.B., and J.R.S. performed experiments and interpreted data; M.M., J.R.S., and M.L.O. designed the study; M.M., Y.Q.L., K.V., J.R.S., and M.L.O. wrote the paper; and all authors read, revised, and approved the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: M.M., 0000-0001-6065-0046; Y.Q.L., 0000-0001-7509-7474; K.V., 0000-0003-4310-3264; S.K., 0000-0003-4646-5279; J.R.S., 0000-0003-2940-2604; M.L.O., 0000-0003-1647-9610.

Correspondence: Martin L. Olsson, Division of Hematology and Transfusion Medicine, Department of Laboratory Medicine, Lund University, BMC C14, SE-22184 Lund, Sweden; e-mail: martin_l.olsson@med.lu.se.

Footnotes

Submitted 30 March 2018; accepted 7 May 2018. Prepublished online as *Blood* First Edition paper, 10 May 2018; DOI 10.1182/blood-2018-03-842542.

The online version of this article contains a data supplement.

There is a *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

REFERENCES

- Saison C, Helias V, Ballif BA, et al. Null alleles of ABCG2 encoding the breast cancer resistance protein define the new blood group system Junior. *Nat Genet*. 2012;44(2):174-177.
- Zelinski T, Coghlan G, Liu XQ, Reid ME. ABCG2 null alleles define the Jr(a-) blood group phenotype. *Nat Genet*. 2012;44(2):131-132.
- Helias V, Saison C, Ballif BA, et al. ABCB6 is dispensable for erythropoiesis and specifies the new blood group system Langereis. *Nat Genet*. 2012;44(2):170-173.
- Svensson L, Hult AK, Stamps R, et al. Forssman expression on human erythrocytes: biochemical and genetic evidence of a new histo-blood group system. *Blood*. 2013;121(8):1459-1468.
- Cvejic A, Haer-Wigman L, Stephens JC, et al. SMIM1 underlies the Vel blood group and influences red blood cell traits. *Nat Genet*. 2013;45(5):542-545.
- Ballif BA, Helias V, Peyrard T, et al. Disruption of SMIM1 causes the Vel-blood type. *EMBO Mol Med*. 2013;5(5):751-761.
- Storry JR, Jöud M, Christophersen MK, et al. Homozygosity for a null allele of SMIM1 defines the Vel-negative blood group phenotype. *Nat Genet*. 2013;45(5):537-541.
- Daniels G, Ballif BA, Helias V, et al. Lack of the nucleoside transporter ENT1 results in the Augustine-null blood type and ectopic mineralization. *Blood*. 2015;125(23):3651-3654.
- Westman JS, Stenfelt L, Vidovic K, et al. Allele-selective RUNX1 binding regulates P1 blood group status by transcriptional control of A4GALT. *Blood*. 2018;131(14):1611-1616.

10. Yeh CC, Chang CJ, Twu YC, et al. The differential expression of the blood group P¹-A4GALT and P²-A4GALT alleles is stimulated by the transcription factor early growth response 1. *Transfusion*. 2018;58(4):1054-1064.
11. Veldhuisen B, van der Schoot CE, de Haas M. Blood group genotyping: from patient to high-throughput donor screening. *Vox Sang*. 2009;97(3):198-206.
12. Mann JD, Cahan A, Gelb AG, et al. A sex-linked blood group. *Lancet*. 1962;1(7219):8-10.
13. Ellis NA, Tippett P, Petty A, et al. PBDX is the XG blood group gene. *Nat Genet*. 1994;8(3):285-290.
14. Weller PA, Critcher R, Goodfellow PN, German J, Ellis NA. The human Y chromosome homologue of XG: transcription of a naturally truncated gene. *Hum Mol Genet*. 1995;4(5):859-868.
15. Auton A, Brooks LD, Durbin RM, et al; 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
16. Battle A, Brown CD, Engelhardt BE, Montgomery SB; eQTL Manuscript Working Group. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213.
17. Khan A, Fomes O, Stigliani A, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res*. 2018;46(D1):D260-D266.
18. Tournamille C, Colin Y, Cartron JP, Le Van Kim C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet*. 1995;10(2):224-228.
19. Xu J, Shao Z, Glass K, et al. Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. *Dev Cell*. 2012;23(4):796-811.
20. Goodfellow PJ, Pritchard C, Tippett P, Goodfellow PN. Recombination between the X and Y chromosomes: implications for the relationship between MIC2, XG and YG. *Ann Hum Genet*. 1987;51(Pt 2):161-167.
21. Reid ME, Lomas-Francis C, Olsson ML. The Blood Group Antigen FactsBook. 3rd ed. London, United Kingdom: Academic Press; 2012.
22. Fouchet C, Gane P, Cartron JP, Lopez C. Quantitative analysis of XG blood group and CD99 antigens on human red cells. *Immunogenetics*. 2000;51(8-9):688-694.
23. Pasello M, Manara MC, Scotlandi K. CD99 at the crossroads of physiology and pathology. *J Cell Commun Signal*. 2018;12(1):55-68.
24. Möller M, Jöud M, Storry JR, Olsson ML. ErythroGene: a database for in-depth analysis of the extensive variation in 36 blood group systems in the 1000 Genomes Project. *Blood Adv*. 2016;1(3):240-249.