LUND UNIVERSITY

Elucidating Genetic and Biochemical Aspects of the P1 and Sd^a Carbohydrate Histo-Blood Group Antigens

Stenfelt, Linn

2020

Document Version:
Publisher's PDF, also known as Version of record

Elucidating Genetic and Biochemical Aspects of the P1 and Sd^{a} Carbohydrate Histo-Blood Group Antigens

LINN STENFELT
FACULTY OF MEDICINE | LUND UNIVERSITY

At the beginning of the 20th century Karl Landsteiner discovered the ABO histo-blood group system. This discovery became the cornerstone for prediction of safe human-to-human blood transfusions in modern medicine. Today, close to 40 blood group systems have been acknowledged by the International Society of Blood Transfusion (ISBT). Donor-patient matching can be achieved by traditional serological hemagglutination methods as well as newly developed blood group genotyping techniques. Genotyping requires an understanding of the underlying locus and genetic variations that determine blood group status.

The aim of this thesis was to establish the molecular mechanisms and genetic bases of the carbohydrate histo-blood group antigens P1 and Sd^a. In addition, the biochemical carriers of the antigenic structures were investigated. The findings presented here open up for genotypic prediction of these blood groups and also allowed ISBT to establish the SID histo-blood group system.

Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2020:90
ISBN 978-91-7619-952-7
ISSN 1652-8220

Elucidating Genetic and Biochemical Aspects of the P1 and Sd^a Carbohydrate Histo-Blood Group Antigens

Elucidating Genetic and Biochemical Aspects of the P1 and Sd^a Carbohydrate Histo-Blood Group Antigens

Linn Stenfelt

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended September 18 at 13.15 in Belfragesalen, BMC, Lund

Faculty opponent
Professor Steven L. Spitalnik, M.D.

Department of Pathology & Cell Biology,
Columbia University, New York City, NY, USA

Organization LUND UNIVERSITY Department of Laboratory Medicine Division of Hematology and Transfusion Medicine	**Document name** DOCTORAL DISSERTATION **Date of issue** September 18th 2020
Author Linn Stenfelt	Sponsoring organization

Title and subtitle

Elucidating Genetic and Biochemical Aspects of the P1 and Sd^a Carbohydrate Histo-Blood Group Antigens

Abstract

Human histo-blood groups are inherited polymorphic variants that occur in the molecular structures on the human red blood cell (RBC) surface. Introducing foreign RBCs into a recipient lacking an antigen may activate the humoral defence leading to a hemolytic transfusion reaction. Antigenic differences can also cause hemolytic disease of the fetus and newborn (HDFN). Blood group antigens are implicated as receptors in pathogen invasion and their expression are often altered in cancerous tissues. Blood group antigens are carried by protein or carbohydrate structures. Carbohydrate antigens are synthesized stepwise by glycosyltransferases and are carried on glycosphingolipids or glycoproteins anchored into the RBC membrane. The aim of this work was to elucidate the molecular genetic mechanisms behind the P1 and Sd^a antigens, as well as to study their glycan structures. The P1 antigen belongs to the P1PK blood group system. Silencing of *A4GALT* causes the null phenotype (P^k–, P1–) of this system. However, the consequence of the genetic differences between the P_1 (P^k+, P1+) and P_2 (P^k+, P1–) phenotypes, i.e. the molecular mechanism underlying how P1 antigen is expressed, has remained unknown. Additionally, there have been divided views regarding the molecular carriers of the P1 antigen, Galα1-4Galβ1-4GlcNAc-R. The Sd^a antigen GalNAcβ1-4(NeuAcα2-3)Gal-R was associated with the *B4GALNT2* gene already in 2003. However, the genetic basis of the Sd(a–) phenotype was never revealed.

Through EMSA experiments the Runt-related transcription factor 1 (RUNX1) was identified to bind *P1* alleles specifically, dependent on rs5751348 in *A4GALT*. Knock-down of RUNX1 decreased the *A4GALT* mRNA levels, establishing its effect as a P_1/P_2-discriminating factor. Based on these findings a genotyping assay was implemented at the Nordic Reference Laboratory for Genomic Blood Group Typing in Lund, Sweden. P1 was also established to be carried on glycoproteins in N-glycan conjugates, in addition to glycosphingolipids.

Sequencing of *B4GALNT2* in nine Sd(a–) individuals identified the missense mutation rs7224888 as highly associated with the phenotype. Additionally, the splice-site polymorphism rs72835417, and the rare missense variants rs148441237 and rs61743617 were encountered in the Sd(a–) cohort. *In silico* studies identified a close correlation between expression of *B4GALNT2* and the cancer-associated lncRNA RP11-708H21.4 locus, located directly downstream of the gene. Finally, the Sd(a–) associated SNP rs7224888 was shown to abolish Sd^a synthase activity in over-expression experiments. The epitope was evaluated with DBA lectin binding, fluorescence microscopy, enzyme immunoblots and mass spectrometry. The latter confirmed that the glycotransferase utilizes substrates on both on N- and O-glycan elongation.

Understanding the molecular mechanism underlying the P1 antigen as well as defining the genetic background of the Sd(a–) phenotype has enabled genotyping approaches for clinical practice. Additionally, the confirmation of *B4GALNT2* expressing the Sd^a synthase, has allowed the International Society of Blood Transfusion (ISBT) to move the Sd^a antigen from the series of high-frequency antigens to its own, new blood group system designated SID, no. 038.

Key words Blood group, P1 antigen, Sd^a antigen, *A4GALT, B3GALNT2,* glycosyltransferase, red blood cell, transfusion medicine

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language** English
ISSN 1652-8220		**ISBN** 978-91-7619-952-7
Recipient's notes	**Number of pages** 76	Price
	Security classification	

Signature  Date 2020-08-27

Elucidating Genetic and Biochemical Aspects of the P1 and Sd^a Carbohydrate Histo-Blood Group Antigens

Linn Stenfelt

LUND
UNIVERSITY

Cover by Linn Stenfelt

Faculty of Medicine
Department of Laboratory Medicine

ISBN 978-91-7619-952-7
ISSN 1652-8220
Lund University, Faculty of Medicine Doctoral Dissertation Series 2020:90

Printed in Sweden by Media-Tryck, Lund University
Lund 2020

Table of Contents

Abbreviations

α1,4Gal-T	4-α-galactosyltransferase
ACKR1	Atypical chemokine receptor 1
β1,4GalNAc-T1	4-β-*N*-acetylgalactosaminyltransferase 1
β1,4GalNAc-T2	4-β-*N*-acetylgalactosaminyltransferase 2
CAZy	Carbohydrate Active Enzymes database
DBA	*Dolichus biflorus* agglutinin
EGR1	Early growth response factor 1
GLUT1	Glucose transporter type 1
GSL	glycosphingolipid
GT	glycosyltransferase
HDFN	hemolytic disease of the fetus and newborn
HSC	hematopoetic stem cell
HTR	hemolytic transfusion reaction
ISBT	International Society Blood of Transfusion
kb	kilo base pairs
KLF1	Krüppel-like factor 1
PBMC	peripheral blood mononuclear cell
RBC	red blood cell
RUNX1	Runt-related transcription factor 1
SNP	single nucleotide polymorphism
TF	transcription factor
TTID	transfusion-transmitted infectious diseases

List of publications

I. Westman JS, **Stenfelt L**, Vidovic K, Möller M, Hellberg Å, Kjellström S, Olsson ML. Allele-selective RUNX1 binding regulates P1 status by transcriptional control of *A4GALT*. *Blood*, 2018;131(14):1611-1616

II. **Stenfelt L**, Westman JS, Hellberg Å, Olsson ML. The P1 histo-blood group antigen is present on human red blood cell glycoproteins. *Transfusion*, 2019;59(3):1108-1117

III. **Stenfelt L**, Hellberg Å, Möller M, Thornton N, Halverson GR, Larson G, Olsson ML. Missense mutations in the C-terminal portion of the *B4GALNT2*-encoded glycosyltransferase underlying the Sd(a–) red blood cell phenotype. *Biochemical and Biophysical Reports*, 2019;19:100659

IV. **Stenfelt L**, Nilsson J, Hellberg Å, Larson G, Olsson ML. Glycoproteomic and phenotypic elucidation of *B4GALNT2* expression variants in the newly established SID histo-blood group system. *Manuscript*

The following paper, written during the Ph.D. studies, is not included in the thesis:

Stenfelt L, Hellberg Å, Westman JS, Olsson ML. The P1PK blood group system: Revisited and resolved.
Review accepted for publication in *Immunohematology*

Abstract

Human histo-blood groups are inherited polymorphic variants that occur in the molecular structures on the human red blood cell (RBC) surface. Introducing foreign RBCs into a recipient lacking an antigen may activate the humoral defence leading to a hemolytic transfusion reaction. Antigenic differences can also cause hemolytic disease of the fetus and newborn (HDFN). Blood group antigens are implicated as receptors in pathogen invasion and their expression are often altered in cancerous tissues. Blood group antigens are carried by protein or carbohydrate structures. Carbohydrate antigens are synthesized stepwise by glycosyltransferases and are carried on glycosphingolipids or glycoproteins anchored into the RBC membrane. The aim of this work was to elucidate the molecular genetic mechanisms behind the P1 and Sd^a antigens, as well as to study their glycan structures. The P1 antigen belongs to the P1PK blood group system. Silencing of *A4GALT* causes the null phenotype (P^k–, P1–) of this system. However, the consequence of the genetic differences between the P_1 (P^k+, P1+) and P_2 (P^k+, P1–) phenotypes, i.e. the molecular mechanism underlying how P1 antigen is expressed, has remained unknown. Additionally, there have been divided views regarding the molecular carriers of the P1 antigen, Galα1-4Galβ1-4GlcNAc-R. The Sd^a antigen GalNAcβ1-4(NeuAcα2-3)Gal-R was associated with the *B4GALNT2* gene already in 2003. However, the genetic basis of the Sd(a–) phenotype was never revealed.

Through EMSA experiments the Runt-related transcription factor 1 (RUNX1) was identified to bind P^1 alleles specifically, dependent on rs5751348 in *A4GALT*. Knock-down of RUNX1 decreased the *A4GALT* mRNA levels, establishing its effect as a P_1/P_2-discriminating factor. Based on these findings a genotyping assay was implemented at the Nordic Reference Laboratory for Genomic Blood Group Typing in Lund, Sweden. P1 was also established to be carried on glycoproteins in N-glycan conjugates, in addition to glycosphingolipids.

Sequencing of *B4GALNT2* in nine Sd(a–) individuals identified the missense mutation rs7224888 as highly associated with the phenotype. Additionally, the splice-site polymorphism rs72835417, and the rare missense variants rs148441237 and rs61743617 were encountered in the Sd(a–) cohort. *In silico* studies identified a close correlation between expression of *B4GALNT2* and the cancer-associated lncRNA RP11-708H21.4 locus, located directly downstream of the gene. Finally, the Sd(a–) associated SNP rs7224888 was shown to abolish Sd^a synthase activity in

over-expression experiments. The epitope was evaluated with DBA lectin binding, fluorescence microscopy, enzyme immunoblots and mass spectrometry. The latter confirmed that the glycotransferase utilizes substrates on both on N- and O-glycan elongation.

Understanding the molecular mechanism underlying the P1 antigen as well as defining the genetic background of the Sd(a−) phenotype has enabled genotyping approaches for clinical practice. Additionally, the confirmation of *B4GALNT2* expressing the Sd^a synthase, has allowed the International Society of Blood Transfusion (ISBT) to move the Sd^a antigen from the series of high-frequency antigens to its own, new blood group system designated SID, no. 038.

Introduction

Transfusion medicine

Blood transfusion as a treatment has been a part of healthcare for more than 200 years. It was James Blundell who, in 1818 recommenced the controversial idea to transfuse blood from one individual to another.[1,2] The first successful human-to-human blood transfusion was given to a woman suffering from a major bleeding after giving birth.[3] Inspired by a colleague, John Leacock, Blundell had started his studies on dogs, concluding that transfusions between individuals of the same species were more prone to succeed while transfusion between species were dangerous.[2] The idea to transfuse blood had appeared already in the 17th century, after William Harvey established the concept that blood circulates in the body. The first transfusions of blood to humans came from animals and the procedure was disapproved both in France and England. In fact, the practise was abandoned altogether in 1679 when the Pope banned it.[4] Unfortunately, the revival of transfusion in the 19th century was not successful for all and several attempts had a lethal outcomes.[5] In the year 1900, Karl Landsteiner observed that human serum could have an agglutinating effect on blood cells from other humans.[6] It was at the time known that agglutinates occurred as a result of mixing serum and red blood cells (RBCs) from different species, while it, between humans only was recognized in pathological conditions. The underlying reason for this phenomenon was described to be due to so-called agglutinins in the serum.[7] Landsteiner investigated the matter in healthy humans, who indeed had agglutinins, and concluded that there were certain regularities in how serum could agglutinate blood cells of different human individuals. Based on this, three blood groups were defined, A, B and C.[8] The latter was later renamed O, probably originating from the German word “ohne”, translating to “without”.[7] Colleagues of Landsteiner soon added a fourth, less common blood group, AB,[9] and thereby the major blood group phenotypes of the first blood group system were defined. Although Landsteiner remarked that his finding could explain the variable outcomes of blood transfusions, practitioners in the field did not embrace the idea immediately. A blood cross-matching test, made to establish donor-patient compatibility prior to transfusion, was initiated and started

to be used in 1907 by Reuben Ottenberg.[10,11] Another challenge in order to give blood transfusions was the clotting of blood when it was drawn. At the time, transfusions therefore had to be conducted through a direct connection between the blood vessels of the donor and the patient. Insights into how citrate and other chemicals could be used as anticoagulants paved the way for the ability to store blood and subsequently give indirect transfusions.[12,13] These major steps made it possible for the modern concepts of transfusion medicine to take form. Further advances were made as new blood group antigens were discovered, as well as insights in how to avoid transfusion-transmitted infectious diseases (TTID), a complication that came along with blood transfusion as a medical treatment. Today, the vast majority of blood transfusions to patients are conducted without complications.[14] However, even though the knowledge of the human RBC has expanded tremendously during the past century, incompatibilities still occur and more research on this intriguing subject is still needed.

The need of blood in modern medicine

Today transfusion medicine is a well-established practice required for modern healthcare. In Sweden, approximately 400 000 units of whole blood were donated during 2019.[15] Blood is given to patients of various diagnostic categories. Studies conducted in the Netherlands, UK and USA,[16-18] conclude that in these regions most of the blood is distributed to medical patients, mainly with blood diseases or cancers. Other examples of conditions for which blood is needed are surgical and, as was obvious already during Blundell´s time, obstetrics. It can also be noted that neonatal patients in general make up 1-2% of the transfused,[18] for causes that are mentioned in a later section. These data represent high-income countries where the most frequently transfused patient group is above 60 years old. In low-income countries the distribution is very different, up to 54% of units are transfused to children under 5 years of age. Donated blood is usually fractionated into RBCs, platelets and plasma, and stored separately in conditions optimized for each product. The components may then be transfused separately depending on the patient´s need. Again this does not describe a majority of the low-income countries.[19] Lately, a need for whole blood transfusion has become apparent in cases with severe hemorrhage. In the setting of a component-based blood bank the trauma care has adopted a standard combination of RBC, plasma and platelet units to mimic whole blood. However, some investigators claim that there are benefits to gain by using real whole blood products.[20,21] The decision to transfuse blood to a patient must always be balanced against the risks of transfusion-related complications, which occur in approximately 1% of the transfused patients.[22] A relatively new but very successful concept, Patient Blood Management, has resulted in a sharp decrease in the numbers of transfusions given in many countries, and the trend goes towards stricter indications and more stringent requirements for doctors to prescribe blood transfusion as a treatment.[23,24]

Histo-blood groups and the immune response

Human histo-blood groups are polymorphic variants that occur in molecular structures on the human RBC surface. If this is introduced into another human, these variants can provoke an immune response with antibody production, given that the individuals do not carry the antigen themselves. The antibodies that recognize and bind to these antigens can initiate downstream immunological consequences.[25] The cell surface is covered by molecules anchored into the lipid bilayer that constitutes the cell membrane. These molecules are proteins, glycolipids or glycoproteins that carry out functions such as ion transportation across the membrane, maintaining the cell shape, assembly of other surface molecules, conveying surface charge, adhesion, enzymatic activity, complement regulation or act as receptors for chemokines.[26] Proteins are made up of amino acids in a linear sequence which obtains secondary and tertiary structures. Glycans (the glycoconjugates carried on lipids or proteins) and lipids on the other hand are constructed by serial action by different enzymes, which are proteins with catalytic ability.[27,28] Glycans are constructed of carbohydrate moieties which are added in a sequential manner, and often branched.[28] This process is described closer in the Glycosylation section below. Though the proteins and glycans of the cell membrane are built to exert certain functions, a specific type of protein or glycan can display small differences between individuals. It can be variations as small as the substitution of a single amino acid or lack of a terminal carbohydrate moiety and this will be enough to make up the structural basis of a histo-blood group antigen, such as the A or B antigens.[29] By definition, antigens dependent on the same (or closely related) gene locus belong to the same system. Blood group phenotypes within each system are defined based on which combination of antigens that is present on the RBC surface. To lack all antigens in a blood group system constitutes the null phenotype of the system.[25] Even if the antigens of some blood group systems, like Rh, appear to be mainly expressed on erythroid cells,[30] blood group antigens are typically not restricted to RBC expression only, and may consequently be found on cells in other tissues. Hence, in some cases histo-blood group antigen is a more appropriate term.[31]

Antibody production against specific antigens may be initiated following alloimmunization, i.e. when the antigen is presented to the immune system of an individual lacking it, e.g. during pregnancy or following a blood transfusion. Some antibodies are said to be “naturally occurring” and often do not undergo a complete class switch from IgM to IgG, which is otherwise observed for many blood group antibodies. An example of this is the antibodies against the A or B antigens, which persist throughout life as IgM antibodies. This antibody production begins at the age of 3 months in humans not carrying the antigen in question.[25] This was noted already in the early experiments by Landsteiner[8] and is probably due to immunologic cross-reaction against similar antigens occurring on non-pathogenic gut bacteria.[32] The antibodies mainly implicated in hemolytic transfusion complications are immunoglobulin (Ig) classes IgG1-4, IgM or (rarely) IgA.

Antibodies can bind to a specific blood group antigen, which may then start a reaction chain leading to the destruction, neutralization and/or elimination of the antigen-carrying cell, here the RBC. This takes form through the initiation of the complement cascade or binding to Fc receptors (FcRs). Briefly described, the antibody-initiated complement cascade constitutes the classical pathway although other pathways may also be active in a transfusion reaction. This will lead to activation of key components starting the cascade and leads to activation of e.g. mast cells, endothelial cells and phagocytes (e.g. inducing phagocytosis through opsonization). Furthermore, the cascade may cause pore formation in the membrane of the target cell and finally osmotic lysis of the same, i.e. intravascular hemolysis through the membrane attack complex. Engagement of the first complement component requires at least two binding sites, an IgG antibody carries one such site while an IgM antibody carries five. Hence, one IgM molecule binding to its antigen can initiate the complement cascade on its own while at least two IgGs binding antigens in proximity are needed.[33] Whilst IgM is more likely to act by direct lysis, IgG-mediated activation mainly acts through the previously mentioned FcRs. There are eight different human FcγRs known, with varying characteristics regarding binding affinity, IgG type preference, structure, pathway of signalling and effects. The latter may be activation which induces phagocytosis through opsonization and thereby clearing the blood stream of the foreign cells.[34] Opsonization of the foreign cell, either by complement or antibodies (IgG), leads to extravascular haemolysis by phagocytosis in the liver or spleen. Destruction of foreign RBCs increases the levels of free hemoglobin, particularly following direct intravascular lysis by complement attack. There are different mechanisms in the body to clear the plasma of these substances however these may be saturated and thereby the more severe complications appear.[35] A primary alloimmunization, i.e. the first time a specific foreign antigen is encountered, is initially dominated by the production of IgM antibodies while approximately ten days later this has switched to IgG.[34] A secondary immune response causes high levels of IgG within three to seven days.[35]

Since Landsteiner's discovery of the ABO blood groups it has become evident that the RBC surface displays far more antigens than A and/or B. To date there are 39 blood group systems acknowledged[36] containing 1-55 antigens each. (Table 1) Additionally, at least another two candidate systems have recently been reported.[37,38] The antigens in focus of this thesis are found in blood group systems no. 003 and 038. There are also several so-called orphan antigens not yet sorted into a system due to lack of genetic evidence, and still other blood group antigens are emerging. So even though the compatibility is well evaluated before a transfusion, virtually all patients receive blood containing blood group antigens unknown to them in every transfusion. In some cases, this also leads to a real incompatibility situation that results in one of the above-mentioned RBC destruction mechanisms. This can be due to human error, suboptimal sensitivity in the testing procedure,[33,35] or emergency transfusions but can also be the natural consequence of a routinely unmatched transfusion.[22]

Table 1. Blood group systems recognized by the ISBT.[36]

No.	System name	System symbol	Gene name(s)	Number of antigens	Chromosomal location
001	ABO	ABO	*ABO*	4	9q34.2
002	MNS	MNS	*GYPA, GYPB (GYPE)*	49	4q31.21
003	P1PK	P1PK	*A4GALT*	3	22q13.2
004	Rh	RH	*RHD, RHC*	55	1p36.11
005	Lutheran	LU	*BCAM*	25	19q13.2
006	Kell	KEL	*KEL*	36	7q33
007	Lewis	LE	*FUT3*	6	19p13.3
008	Duffy	FY	*ACKR1*	5	1q21—q22
009	Kidd	JK	*SLC14A1*	3	18q11-q12
010	Diego	DI	*SLC4A1*	22	17q21.31
011	Yt	YT	*ACHE*	5	7q22
012	Xg	XG	*XG, MIC2*	2	Xp22.32
013	Scianna	SC	*ERMAP*	7	1p34.2
014	Dombrock	DO	*ART4*	10	12p13-p12
015	Colton	CO	*AQP1*	4	714p
016	Landsteiner-Wlener	LW	*ICAM4*	3	19p13.2
017	Chido/Rogers	CH/RG	*C4A, C4B*	9	6p21.3
018	H	H	*FUT1*	1	19q13.33
019	Kx	XK	*XK*	1	Xp21.2
020	Gerbich	GE	*GYPC*	11	2q14-q21
021	Cromer	CROM	*CD55*	20	1q32
022	Knops	KN	*CR1*	9	1q32.2
023	Indian	IN	*CD44*	6	11p13
024	Ok	OK	*BSG*	3	19p13.3
025	Raph	RAPH	*CD151*	1	11p15.5
026	John Milton Hagen	JMH	*SEMA7A*	6	15q22.3-q3
027	I	I	*GCNT2*	1	6p24.2
028	Globoside	GLOB	*B3GALNT1*	2	3q25
029	Gill	GIL	*AQP3*	1	9p13
030	Rh-associated glycoprotein	RHAG	*RHAG*	3	6p12.3
031	FORS	FORS	*GBGT1*	1	9q34.13-q34.3
032	JR	JR	*ABCG2*	1	4q22.1
033	LAN	LAN	*ABCB6*	1	2q36
034	Vel	VEL	*SMIM1*	1	1p36.32
035	CD59	CD59	*CD59*	1	11p13
036	Augustine	AUG	*SLC29A1*	4	6p21.1
037	KANNO	KANNO	*PRNP*	1	20p13
038	Sid	SID	*B4GALNT2*	1-2	17q21.32
039	CTL2	CTL2	*SLC44A2*	2	19p13.2

Color codes: Red text indicates systems with carbohydrate antigens, Grey text indicates provisional system status.

Transfusion-related complications

Blood group incompatibility between blood donor and patient is the most common cause of a clinically significant hemolytic transfusion reaction (HTR).[22] HTR is defined as an increased haemolysis through mechanisms described in the above section. It can be both an acute event, appearing within 24 hours after transfusion, or delayed, up to a month post transfusion. The symptoms in acute HTR vary and may include fever, pain in the chest, abdomen, flank or back, reddish to black urine, chills, rigors, respiratory distress, anxiety, pain at the infusion site, hypotension or oliguria. The delayed reaction normally displays much milder symptoms in general, if any, but also others including tiredness and jaundice due to the increased RBC destruction by phagocytosis.[22,35] The reaction of the immune system is complex, and many factors affect the outcome. Incompatibility does not always cause HTR, but if it does, particularly the acute intravascular HTR may lead to such severe complications as renal dysfunction, shock, disseminated intravascular coagulation (DIC) or death.[34,35] Unfortunately, there is no specific treatment for the worst forms of such transfusion reactions. Symptomatic treatment aiming to reduce the immune response is often tried, as is various kidney-protective measures including forced diuresis. Another principle is to remove free hemoglobin and the implicated antibodies by plasmapheresis as soon as possible in order to protect the kidneys to avoid organ failure and a subsequent need for transplantation.[33,35,39] Increasing the degree of compatibility and minimizing the risk of handling errors are two examples of ways to avoid RBC incompatibility and thereby decrease the risk of HTRs. [34]

Other complications important to mention are transfusion-associated circulatory overload (TACO) and transfusion-related acute lung injury (TRALI). Although the pathophysiological mechanisms causing the latter syndrome is poorly or incompletely understood, it is often described as a two-hit model. The first hit depends on recipient-risk-factor while the second hit is defined by the transfusion product characteristics. The latter may be antibodies against leukocytes or other factors from the donor.[40] This is why some countries only provide male plasma for transfusion to avoid the passive transfer of pregnancy-induced alloantibodies, mainly antibodies against human leukocyte antigens (HLA) and human neutrophil antigens (HNA).[41] These syndromes are not believed to involve antibodies directed towards RBC antigens and are thereby not the main scope of this thesis. Transfusion-associated graft-versus-host disease (TA-GVHD), a very rare but severe transfusion complication, is dependent on donor lymphocytes that may induce an immune response against the recipient tissue, particularly in certain immunodeficient patient groups or if the donor is homozygous for an HLA haplotype present in the recipient. Again, the insights regarding key factors underlying this syndrome are not fully understood, however severe immunodeficiency in the recipient is the most obvious risk factor. Irradiation of the blood components (to eliminate the lymphocytes' ability to divide and expand) prior to transfusion must be done to prevent TA-GVDH.[42]

As previously mentioned, when the practice of blood transfusion expanded in the 20th century it became apparent that certain TTID was a risk in the procedure. In the 1940s the first donor screening was introduced to target syphilis.[43] Posttransfusion hepatitis was well recognized in the 1970s, however, it was substantially decreased as screening for HBsAg, detecting the hepatitis B virus (HBV), along with donor deferral were initiated.[44] In the beginning of the 1980s, acquired immunodeficiency syndrome (AIDS) was emerging but prevention could not take place until the discovery of human immunodeficiency virus (HIV) was made in 1983-84.[14] In addition to HIV it became of interest to include screening for related retroviral risks like human T-cell lymphotropic viruses I and II as well. In the late 1980s the hepatitis C (HCV) was discovered and a majority of remaining incidences of posttransfusion hepatitis (previously known as non-A, non-B hepatitis) were explained and screening also for this agent could be started.[44] To maintain a safe transfusion service, the World Health Organization (WHO) recommends universal mandatory screening of HIV-1 and -2, HBV, HCV and syphilis. In case of areas with certain endemic infections it is recommended to include screening for such agents specific to a certain region as well.[45] Currently, bacterial contamination, especially of platelet concentrates (as these are stored at room temperature) can still cause clinical problems. Aims to advert this are taken e.g. by screening with culturing techniques or to perform pathogen inactivation (also called pathogen reduction). In general, precautions must be taken as new pathogens continuously emerge.[44]

Other medical implications where histo-blood groups play a role

Hemolytic disease of the fetus and newborn (HDFN) is caused by maternal antibodies of IgG type, which can transfer over the placenta. The antibodies bind to paternally derived blood group antigens on the fetal RBCs, foreign to the maternal immune system. The fetal RBCs are phagocytized in the fetal spleen, which can generate anemia, enlargement of liver and spleen, jaundice and eventually the edamatous swelling known as *hydrops fetalis*, a potentially lethal condition. Antibodies against certain blood group antigens expressed early during erythropoiesis may cause a non-hemolytic anemia but also very severe effect by targeting erythroid progenitor cells. Antibodies within the ABO blood group system usually do not cause any severe HDFN but the IgG component, particularly in group O maternal plasmas, may pass the placenta and cause fetal jaundice. The vast majority of HDFN cases are caused by antibodies targeting antigens within the Rh blood group system, especially anti-D followed by anti-c and -E, or the Kell system, e.g. anti-K. Treatments of varying degrees of HDFN include phototherapy, exchange or top-up transfusions of the newborn, or if symptoms are serious already during the pregnancy, even intrauterine blood transfusion to the fetus.[46-48] The frequency of HDFN is very low today in high-income countries, much thanks to the advent of Rh prophylaxis, which involves antenatal and postnatal injections of anti-

D of IgG type in RhD-negative women carrying or having delivered RhD-positive fetuses/babies.[48] Despite this efficient treatment of the RhD-induced HDFN, the progress has not reached low-income parts of the world. Currently, about 50% of the pregnant women who need Rh prophylaxis globally will not receive it. This is why commitments such as the Worldwide Initiative for Rh Disease Eradication (WIRhE) are needed.[49]

Another complication involving blood group antigens during pregnancy involves antibodies targeting the placenta. Though rare the consequences can be severe, potentially leading to reoccurring spontaneous abortions early during pregnancy. This has been noted in women with the rare phenotypes P_1^k, P_2^k (of the Globoside blood group system) or p (of the P1PK blood group system). All these phenotypes lack the P antigen which has been demonstrated to be expressed in placental tissue, and therefore prone to attack by the naturally-occurring anti-P.[50-52]

In solid organ transplantation the histo-blood group antigens may interfere and lead to antibody mediated rejection (AMR). Usually this is caused by antibodies against HLA, antigens of the ABO system or other incompatible endothelial antigens. Therefore, matching of ABO blood group antigens is of importance in solid organ transplantation. In later years, alternative strategies to prevent AMR have evolved: plasmapheresis or immunoadsorption that targets the antibodies against HLA or ABO antigens, IVIg, anti-CD20 directed to members of the B-cell lineage, proteasome inhibitors or complement inhibitors, all with the aim to interfere with the humoral reaction.[53-55] The advances of strategies to prevent humoral reactions have made it possible to transplant across the ABO incompatibility barrier, e.g. kidneys.[56]

Xenotransplantation is another option for broadening organ availability, however it carries along multiple challenges regarding compatibility (and ethics, etc.). Like the antibodies against human A and B antigen the xenoreactive antibodies are naturally occurring.[53] One example is the human reactive antibodies against at least three known pig carbohydrate structures; the Galα1-3Gal epitope, Neu5Gc and the non-Gal antigen, where the genetic basis of the latter is a homolog to the human *B4GALNT2*.[57] In a study the underlying genes of the above-mentioned antigens, *GGTA1, CMAH,* and *B4GALNT2,* where knocked out in pigs. The reactivity against peripheral blood mononuclear cells (PBMCs) from the pigs, with antibodies in human sera, was reduced compared to cells from pigs having a functional *B4GALNT2*. This displays the relevance of these xenoantigens.[58]

Finally, there are examples of where the expression of blood group antigens is either up- or down-regulated in cancerous tissue. This often appears to involve glycan blood groups.[59] Cell signaling, recognition and adhesion may be conducted through the blood group molecules and therefore involved in cell functions affected in the transformation to malignant cells.[60] A very recent example is the discovery of the

molecular basis of the MAM blood group antigen, the carrier molecule (EMP3) of which has been implicated both as a tumor suppressor and an oncogene product.[38]

Blood groups and infectious disease

Many blood groups are associated with disease susceptibility. In several cases the explanation is that pathogens utilize these antigens as receptors. Bacteria, toxins, viruses and parasites bind to cell surfaces and blood group variation is thought to have evolved under this pressure. These epitopes are utilized by the microorganisms to adhere to and often be able to invade the cell.[61] E.g., antigens of at least 12 blood group systems have been shown to be involved in the *Plasmodium* parasites' invasion of RBCs, causing the potentially fatal malaria disease.[62] An interesting case concerns the antigens of the Duffy blood group system, situated on the Atypical chemokine receptor 1 (ACKR1) protein, which is a receptor for *Plasmodium vivax*. The frequency of the Duffy null phenotype Fy(a–b–) is markedly increased among individuals of African origin, especially in areas like western Africa where the parasite is endemic. This exemplifies how pathogens can be a selective force in human evolution.[61,63] Another example of how blood groups are natural resistance factors against severe malaria disease involves the ABO system. Group O has been shown to protect against the most severe forms of *Plasmodium falciparum* malaria, particularly for children in whom cerebral malaria due to rosetting of RBCs is a common reason for mortality.[64] Whilst invasion of RBCs occurs independent of ABO phenotype, group O RBCs are more easily cleared by phagocytosis and also less prone to form rosettes and adhere to endothelial cells since malarial RIFIN proteins bind the A antigen to stabilize cell-to-cell interactions.[65,66]

Furthermore, susceptibility to infection varies due to blood group status in many gastrointestinal illnesses, e.g. those caused by some highly infectious viruses in the *Calici* group, particularly certain strains of *Norovirus*. Individuals with non-secretor status, mainly due to the c.428G>A polymorphism in the *FUT2* gene, are protected against the commonly known winter-vomiting disease. Additionally, higher susceptibility has been seen with the H and A antigens, compared to the B antigen with some strains.[61,67] The naturally-occurring antibodies against carbohydrate blood group antigens may be an advantage when it comes to neutralizing pathogenic agents. The severe acute respiratory syndrome (SARS) caused by the SARS coronavirus (SARS-CoV) has been associated with the A, B and AB blood group phenotypes, while individuals of the O blood group phenotype may be relatively resistant. High titers of anti-A have showed effective blocking of the virus.[61] The current pandemic, initiated during the winter 2019/2020 and still active, is caused by the related virus SARS-CoV-2. Early studies on patients suggest that A blood group phenotype incidence is increased among the infected, while O phenotype is decreased.[68-70] Like the related SARS-CoV infection, this might be due to presence of anti-A (of IgG type), rather than the A antigen being a receptor for the virus[71] but

other possible ABO-related effects have also been discussed. Similar effects of neutralization of viruses have been shown previously for HIV and measles,[72,73] for instance. Pathogen interaction of antigens within the P1PK and SID blood group system will be discussed in the section of respective system.

Blood group characterization

Blood group laboratories routinely use antibody-based hemagglutination methods in clinical practice to avoid transfusion of incompatible blood units, prevent hyperacute rejection following solid organ transplantation and also to monitor incompatibility between the pregnant woman and her fetus. Some examples where traditional immunohematology is very useful include standard blood group (ABO and RhD) typing, cross-matching, screening for and identification of the specificity of blood group antibodies, phenotyping of minor blood groups, antibody titration, adsorption/elution etc. Prior to the 1940s, blood group serology relied solely on direct agglutination, which limited the scope of detection to blood group antibodies of IgM class. In 1945, the direct antiglobulin test (DAT) was reported, a method where sensitized (IgG covered) RBCs are detected adding anti-human globulin (AHG). If the RBCs then form agglutinates, antibodies are apparently bound to the cell surface.[74] Based on this idea, indirect antiglobulin test (IAT) is a method used to detect or identify antibodies in the patient's plasma. These methods may also involve antibodies detecting complement fragments binding to the RBC.[75] For blood group phenotyping commercial antibody reagents are available for the most common antigens to use with RBCs.

Unfortunately, these serological methods cannot always be applied, e.g. a suitable sample of RBCs may not always be accessed, or serological reagents may not be available. There are also conditions, such as cold agglutinins in autoimmune hemolytic anemia, and medical treatments like monoclonal antibodies against CD38 (present also on the RBCs) that challenge the typing.[22,23] A rare but troublesome phenomenon is polyagglutination, a condition where the RBCs are agglutinated by most adult human sera. The cause is exposure of cryptic antigens due to bacterial or viral infections, mutations in hematopoietic progenitor cells or inherited rare antigens.[76,77] There may also be a problem to assess whose RBCs are actually typed, e.g. following massive or chronic transfusion.

An alternative in all these situations is blood group genotyping, a concept made possible as the genetic origin of many blood groups has been mapped over the past 30 or so years. Genetic analysis may thereby predict the blood group phenotype. Early adopters set up in-house methods based on polymerase chain reaction (PCR)-based methods, however as increased through-put was needed, other applications were developed, often in collaboration with corporate partners.[78] These genotyping platforms made it possible to determine many antigens in one panel. Applying serological techniques in combination with genomic predictions constitute a

powerful combination of tools in the efforts towards secure blood matching between donor and patient.[25] Genomic data covering several blood groups especially benefit sickle cell anemia patients who have a particularly high risk of forming multiple, complicated blood group antibodies, e.g. against antigens within the Rh system.[79] In cases of immunized pregnant women, blood group genotyping of the fetus can now be made using maternal plasma as a source of fetal DNA. Thereby is the fetal RhD status and need of Rh prophylaxis determined without risk of pregnancy complications.[48] However, genotyping requires knowledge of the genetic basis of blood group, what genomic region to target. To date there are more than 360 formally acknowledged blood group antigens, and it is currently not possible to target all in a single assay. To be able to do so next-generation-sequencing (NGS) of the whole genome or targeting of known blood group loci can be utilized.[25,80]

To keep track of all the variations known, antigens are categorized into systems by the International Society of Blood Transfusion (ISBT) Working Party for Red Cell Immunogenetics and Blood Group Terminology. The antigens encoded, or in other ways determined, by the same gene or related genes belong to the same system. To assign an antigen to a system (Table 1) it must fulfil the following requirements; 1) the antigen is defined by a human alloantibody, 2) the antigen is inherited, and 3) the genetic background is identified, and chromosomal location known so that it can be proven independent from all other systems but the one it is included in. There are also so-called orphan blood group antigens defined by alloantibodies but not meeting the requirements due to unknown molecular and/or genetic origin. These are either placed in collections, typically when there is a serologically implicated relation between antigens, or defined as solitude high- or low-prevalence antigens, dependent on the frequency. Information about and current status of blood groups are accessible on ISBT's webpage.[36] Additionally, there are several databases that provide information on the blood group genes known today and the vast variations of alleles. One such example is the *Erythrogene* platform.[81]

Erythropoiesis

The RBC, also called the erythrocyte, is the most abundant cell in the whole body, constituting about 84% of all our own cells.[82] It is formed in a biological process called erythropoiesis which is a part of the hematopoiesis, the formation of mature blood and immune cells from hematopoetic stem cells (HSCs), see Fig. 1A. During embryonic development, blood cells are emerging in three waves. First primitive erythroblasts, diploid platelet progenitor cells and macrophages are produced in the yolk sac. In the second wave, erythroid-myeloid progenitors develop and create colonies. Finally, HSCs emerge from specific endothelial cells in the dorsal aorta of the aorta-gonad-mesonephros region.[83,84] The HSCs migrate to the fetal liver and later the bone marrow, where hematopoiesis takes place in the adult human. Burst

forming unit erythroid (BFU-E) and colony-forming unit erythroid (CFU-E) cells are the first progenitors committed to the erythroid lineage, as seen in Fig. 1B. These form colonies and as the erythroblasts differentiate to RBCs the cell nucleus reduces in size and an increase of hemoglobin occurs.[84] The maturating cells gather around a macrophage in the centre of the colonies. These cell formations, located in anatomic niches, are called erythroblastic islands. The macrophage assists in the erythroid differentiation and proliferation. In the final steps of erythrocyte maturation, the nucleus is removed, generating the reticulocyte.[85] Reticulocytes are released into the circulation and are translationally active until the cells finally lose ribosomes, organelles and mRNA in the maturation and reshaping to the characteristic biconcave RBC.[86] The RBC life spans over 120 days, after which it is removed by macrophages in the liver and spleen.[85]

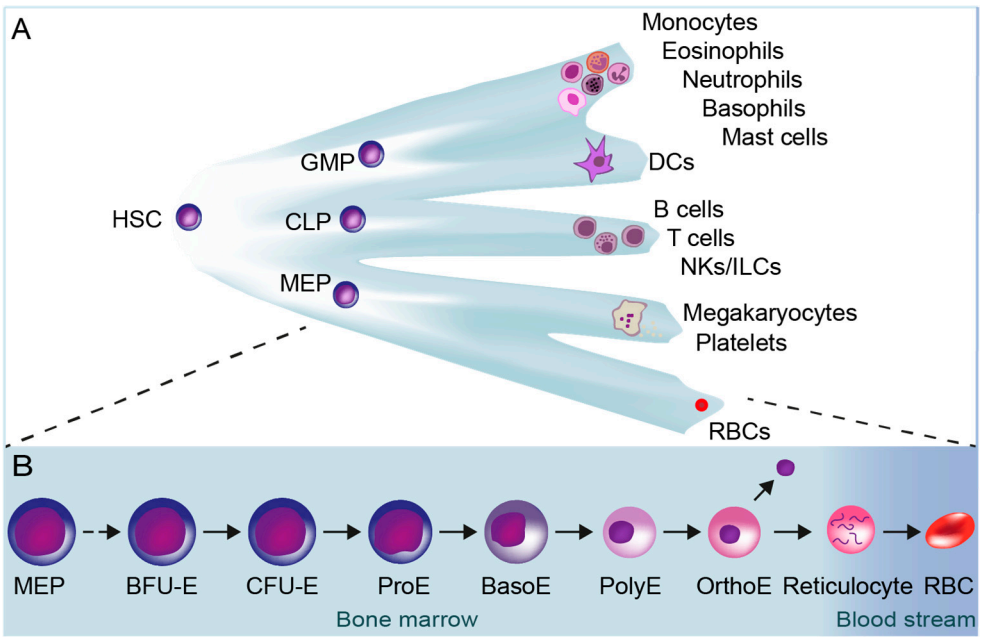


Figure 1. The formation of the blood cells, with a focus on the erythroid lineage.
A) Hematopoiesis is the process during which the maturation of blood cells takes place. HCSs are the progenitors for all blood cells and situated in the bone marrow of the adult human. They give rise to the GMPs, CLPs and MEPs. The MEPs differentiate to the megakaryocytes and platelets or **B)** RBCs, through the erythropoiesis. Abbreviations: HSC, hematopoietic stem cell; GMP, granulocyte monocyte progenitor; CLP, common lymphoid progenitor; MEP, megakaryocyte erythroid progenitor; DCs, dendritic cells; NKs, natural killer cells; ILCs, innate lymphoid cells, BFU-E, blast colony forming unit – erythroid; CFU-E, colony forming unit – erythroid; ProE, proerythroblast; BasoE, basophilic erythroblast; PolyE polychromatic erythroblast; OrthoE, orthochromatic erythroblast; RBC, red blood cell.

Genetic regulation of RBC formation

The mature RBCs lack nuclei and organelles, and thereby no active transcription or translation is thought to occur in these cells. Instead, the cellular composition takes its form during erythropoiesis. The expression pattern of genes to push a cell into a certain fate is a well-orchestrated process where the transcription factors (TFs) are key components. However, a first requirement for gene transcription is that the gene's promoter is accessible. The genomic DNA is organized around histone octamers, forming nucleosomes and only when the chromatin is in open formation a gene can be transcribed. A gene is constituted by sets of exons, which are the coding sequence, divided by intron sequences, the noncoding sequences of the gene. Transcription of the gene to RNA is performed by the RNA Polymerase II (Pol II), which binds to the transcription starting site (TSS) at the 5´end of a gene. The area surrounding the TSS is considered as the core promoter. The Pol II binds the area in complex with general TFs, e.g. the TATA-box-binding protein is a part of such a complex and it binds to the TATA-box, situated in the area. Other characteristics of a promoter may be CpG islands (regions of elevated GC content) overlapping the promoter or AT areas downstream of the TSS. The regulatory motifs are usually more distally located. TFs bind and recruit transcriptional cofactors generating an enhancement of transcription.[87] To generate the definitive hematopoietic progenitors and HSCs, the TFs Runt related TF 1 (RUNX1) and GATA2 have shown to be vital.[83] In further commitment of the HSCs towards the erythroid lineage GATA1, T-cell acute lymphocytic leukemia protein 1 (TAL1) and Krüppel-like factor 1 (KLF1), LIM domain binding protein 1 (LDB1) and LIM domain only protein 2 (LMO2) are required, as all have been associated with DNA regulatory elements in erythroblasts affecting erythroid genes. A very recent finding is also that KLF1 is stabilized by the G-coupled protein pathway suppressor 2 (GPS2), which promotes erythroid differentiation.[88,89] The TFs bind to the DNA in complexes and each TF association to the DNA is often dependent on other TFs. Abolished binding or lack of a single TF may thereby impair the whole complex.[84] Long noncoding RNAs (lncRNAs) is another family of regulatory factors. These may act through recruiting and regulating TF activity and chromatin modifiers, organize chromosomal domains and control downstream events. In this manner, several lncRNAs have been identified to modulate gene expression required for erythrocyte maturation.[90]

The expressed transcript is processed further; one important step is the removal of the noncoding intron sequences which is performed by the spliceosome. The remaining mRNA carrying the sequence of the joined exons is transferred from the nucleus (Fig. 2-I) and the ribosome translates it into the amino acid sequence producing the protein. (Fig. 2-II) Further posttranslational modifications of the protein take place mainly in the endoplasmic reticulum (ER) and Golgi apparatus.[91]

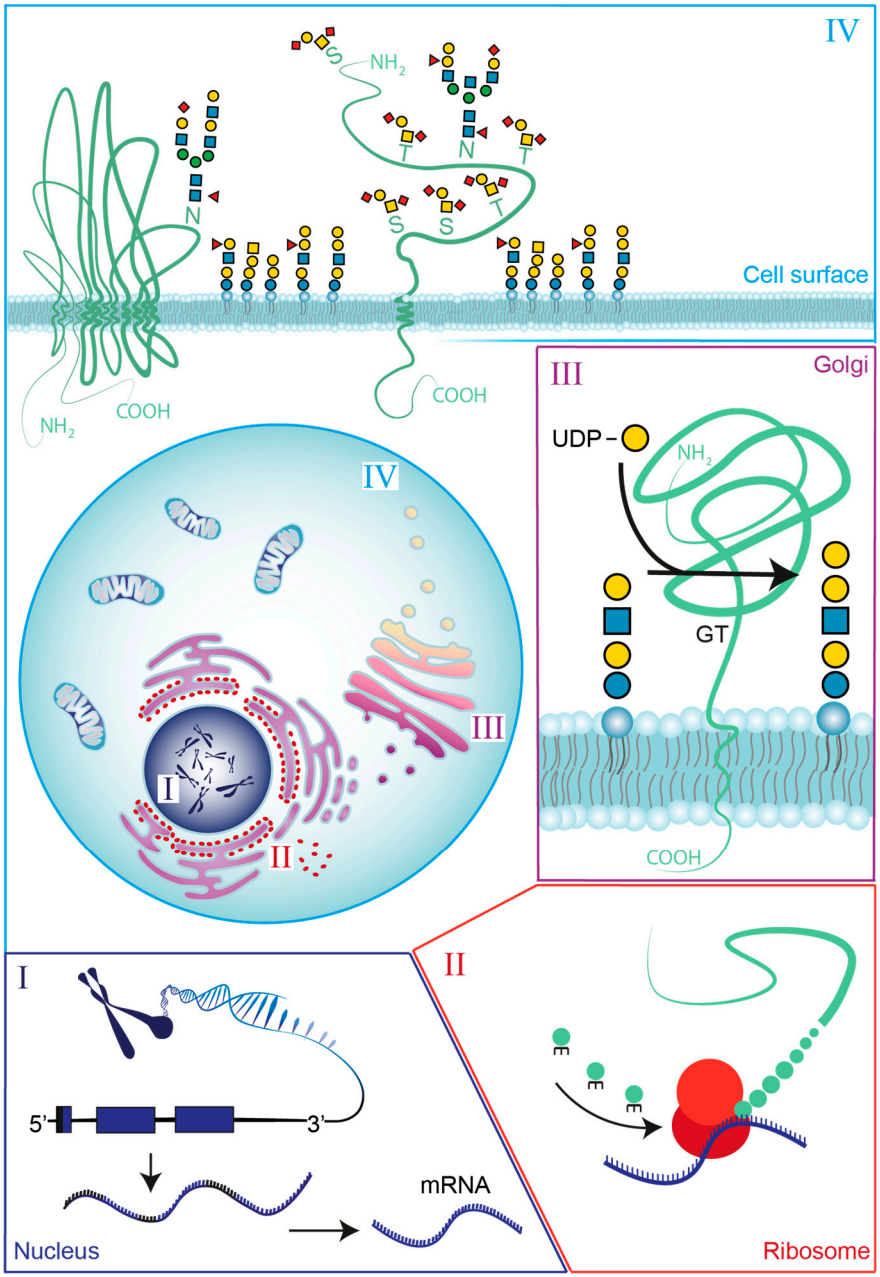


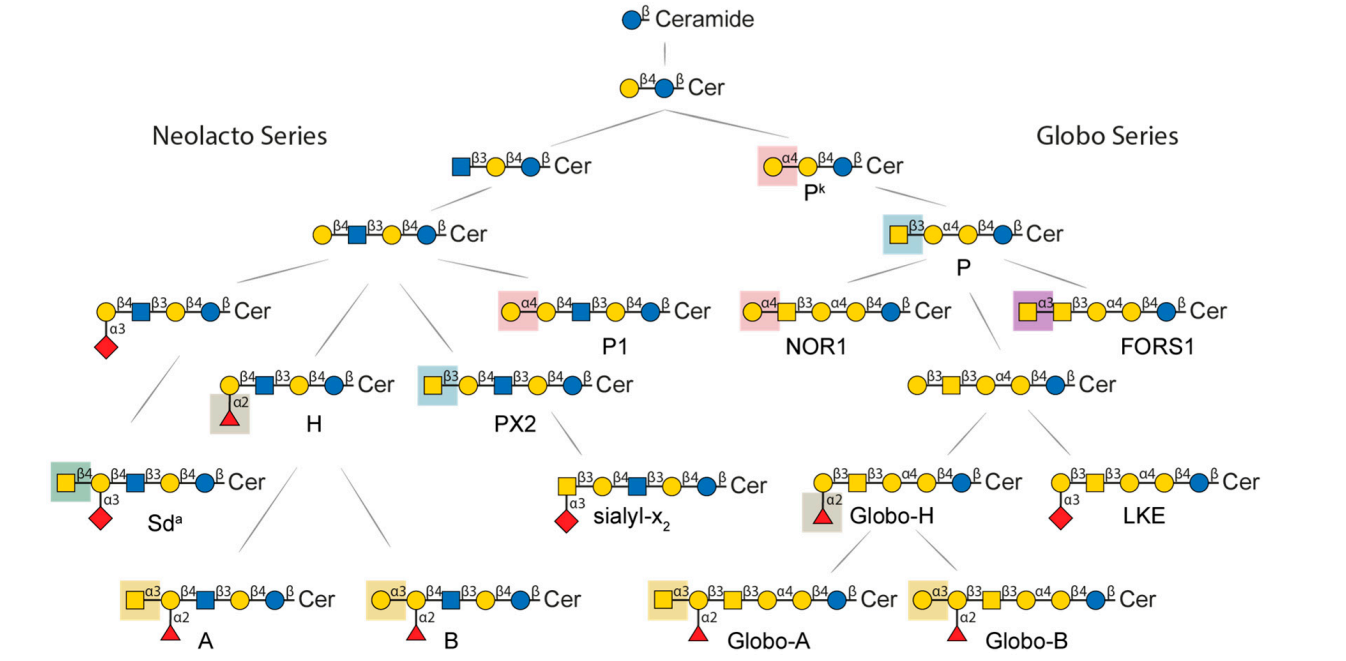
Figure 2. Gene expression, protein synthesis and glycosylation in the cell.
In the cell nucleus **(I)** the transcription of the DNA generates a pre-mRNA, which is further spliced and modified before export from the nucleus. The ER-membrane-bound or free ribosomes **(II)** translate the amino acid sequence of the RNA, and thereby synthesize the protein. After folding and other modifications in the ER-Golgi apparatus, the membrane-bound GT proteins are maintained in Golgi **(III)** where the glycosylation take place. The GTs transfer carbohydrate moieties from activated nucleotide sugars to their acceptor substrates. The glycosylated lipids and proteins may then be transferred to the cell surface **(IV)** where they are a part of the glycocalyx. Symbol nomenclature for glycans according to Varki et al. 2015.[92]

Glycosylation

The cell surface encompasses a diverse dense layer of glycan structures, which make up the glycocalyx.[93] As other posttranslational modifications, the glycosylation of cell surface structures takes place intracellularly, more specifically in the cell's ER and most importantly in the Golgi apparatus. The glycans are synthesized stepwise by glycosyltransferases (GTs). The GTs are expressed as any other protein, described in the previous section. However, after its final modifications the GT is retained in the ER or Golgi compartment. There, it utilizes activated donor sugar substrates, usually nucleoside diphosphate sugars (such as UDP-Gal) to catalyze the elongation of a specific acceptor substrate, e.g. glycosphingolipids (GSLs) Fig. 2-III or glycoproteins. The glycosylated structure is then transported in vesicles to the plasma membrane and the cell surface to become a part of the glycocalyx, Fig. 2-IV.[28]

The GTs are type II transmembrane proteins with a short N-terminal cytosolic tail, a transmembrane domain (TMD), a stem and a globular domain with the catalytic site. The correct localization is maintained depending on the GTs cytoplasmic tail, TMD and/or stem.[94] Among GTs there are two major types of structural folds, GT-A and GT-B. The GT-A enzymes display a Rossman structure (interacting with the nucleotide) and a majority encompass a DXD motif and are metal-ion dependent. The GT-B fold enzymes have two domains with Rossman folds where the C-terminal domain interacts with the nucleotide sugar donor and the N-terminal domain recognizes the acceptor molecule.[28,95,96] Furthermore, the transferases may form homodimers, often interacting through their catalytic domain. The GTs that add the terminal carbohydrate in the A and B antigen can form dimers but do so by interacting mainly through the stem regions.[97] There are even data to suggest that many other GTs not only form homodimers but possibly also heterodimers or heterooligomers built up as functional units to optimize the glycosylation process.[97] However, this is still considered controversial by many investigators. GTs catalyze a retaining or inverting reaction of the anomeric carbon of the sugar to be added, connecting it to the precursor.[28] In the carbohydrate active enzymes (CAZy) database (www.cazy.org/), GTs are classified based on sequence similarities, which indicate a similarity in 3D structure and specificity.[98]

GSLs are usually initiated with a glucose covalently attached to the ceramide and there are seven neutral core sequences (the ganglio, lacto, neolacto, globo, isoglobo, mollu and arthro series) based on the four initial moieties in the carbohydrate chain. The ganglio, globo and neolacto series are the major series in vertebrates.[28] The cores and selected elongation of the two latter series are depicted in Fig. 3, where several blood group antigens are also recognized. The most common glycans of the glycoproteins are divided into O-GalNAc- and N-glycans. The sequence of the O-GalNAc-glycans may comprise of 1-20 or more carbohydrate moieties and have



Blood group system	Glycosyltransferase	Gene
ABO	α1,3GalNAc-T/α1,3Gal-T	*ABO*
P1PK	α1,4Gal-T	*A4GALT*
H	α1,2Fuc-T1	*FUT1*
Globoside	β1,3GalNAc-T1	*B3GALNT1*
FORS	Gb α1,3GalNAc-T	*GBGT1*
SID	β1,4GalNAc-T2	*B4GALNT2*

Symbolic representation
- Glucose (Glc)
- Galactose (Gal)
- *N*-acetylglucosamine (GlcNAc)
- *N*-acetylgalactosamine (GalNAc)
- Fucose (Fuc)
- *N*-acetylneuraminic acid (NeuAc)

Figure 3. Biosynthesis flow sheet of selected structures in the Globo and Neolacto pathways.
Many carbohydrate blood group antigens are found on GSLs and are synthezied stepwise by GTs adding each carbohydrate moiety to a growing acceptor chain. The table states the names, active GTs and implicated genes of six blood group systems. Color codes indicate which carbohydrate(s) each GTs adds. The depicted Sd[a]/Cad epitope has been caracterized in Cad RBCs.[99] Further extentions of ABO and NOR1 antigens have been described but are not included here. Symbol nomenclature for glycans is according to Varki et al. 2015.[92]
Abbreviations: α1,3GalNAc-T, 3-α-*N*-acetylgalactosaminyltransferase; α1,3Gal-T, 3-α-galactosyltransferase; α1,4Gal-T, 4-α-galactosyltransferase; α1,2Fuc-T1, 2-α-fucosyltransferase 1; β1,3GalNAc-T1, 3-β-*N*-acetylgalactosaminyltransferase Gb α1,3GalNAc-T, globoside 3-α-*N*-acetylgalactosaminyltransferase; β1,4GalNAc-T2, 4-β-*N*-acetylgalactosaminyltransferase 2.

four major core structures all initiated with GalNAc covalently bound to serine, threonine[28] or rarely to tyrosine[100] in the peptide chains, Fig. 4A. N-glycans have a core of Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1-4β1-4GlcNAcβ1-Asn, which is further branched into oligo-mannose, complex or hybrid structures, Fig. 4B. The structure resides on asparagine in the sequence of Asn-X-Ser/Thr (X may be any other amino acid except proline).[28] In the RBCs several major surface proteins are glycosylated. The previously mentioned protein ACKR1, carrier of the Duffy blood group system antigens, is N-glycosylated.[101] The glucose transporter type 1 (GLUT1) has no blood group antigen associated to its amino acid sequence, but carries an N-glycan which may be terminated with carbohydrate blood group antigens like ABH. [26] The same goes for the band 3 anion transporter which is associated with the cytoskeleton of the RBC. In fact, band 3 is a major carrier of the ABH antigens on the RBC surface. Concurrently, variations in its amino acid sequence make up the antigens of the Diego histo-blood group.[30] The glycophorins (GP) A and B, carriers of the protein blood group antigens of the MNS system, are highly glycosylated. The GPA carries 12 possible O-GalNAc sites and one N-glycan, the latter of which may carry the A and/or B antigen.[102]

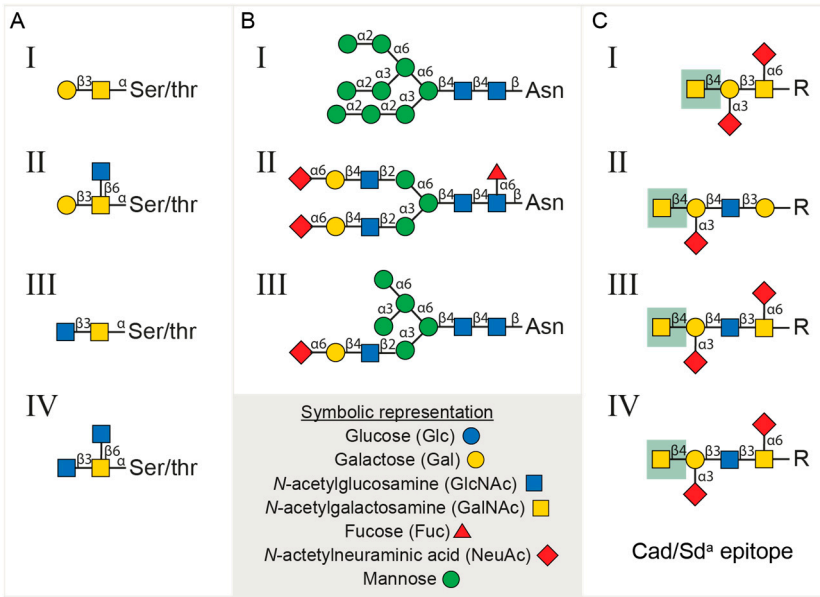


Figure 4. Glycoprotein carried glycan structures. A) The most common core structures of O-GalNAc-glycans in humans, I-IV. **B)** Three general types of elongations on the eukaryote N-glycans, I: Oligomannose, II: Complex and III: Hybrid, all sharing the N-glycan core structure Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1-4β1-4GlcNAcβ1-Asn, **C)** The anti-Sd[a] reactive glycans on glycoproteins, (the green box follow colorcode in Fig. 3) I: GPA-carried O-GalNAc glycan from Cad RBCs, II: terminal saccharide residing on the N-glycans of a Sd(a+) uromodulin glycoprotein, III and IV: the Sd[a] epitope found on core 3 O-GalNAc-glycans on the mucins in human descending colon. Symbol nomenclature for glycans according to Varki et al. 2015.[92]

Carbohydrate blood group systems

Among the 39 known blood group systems there are eight containing carbohydrate antigens: ABO, P1PK, Lewis, H, I, Globoside, FORS and SID. These are highlighted in red in Table 1. The antigen(s) of each system is/are dependent on the same GT-encoding gene. Alleles coding for different phenotypes differ in the gene sequence. The polymorphisms underlying expression of different blood groups may occur in the exons, causing amino acid changes or creating premature stop codons or frame shifts due to insertions or deletions. The alterations may also be found in the introns, e.g. affecting the splice sites, or in regulatory areas flanking the gene such as the promoter region or enhancer motifs, which determine how much of the transferase is expressed. Different alleles of the same gene may thereby express transferases with different characteristics in donor or acceptor substrate affinity. The null phenotype of a blood group system is generally due to completely silenced alleles, e.g. homozygosity for a premature stop codon. However, within the carbohydrate blood groups, a null phenotype may also be due to alleles expressing inactive but full-length forms of the GTs.[30]

The GTs synthesizing the antigens of the ABO blood group system are encoded by the *ABO* gene.[103,104] It has a GalNAc or Gal transferase activity, building the A or B antigen, respectively. This is due to so called single nucleotide polymorphisms (SNPs) in the DNA sequence. In total there is as little as seven SNPs that differ between the A and B transferase coding alleles, but only four will alter residues in the amino acid sequence, three are so-called silent or synonymous polymorphisms. The changed amino acids have varied impact on the donor substrate specificity.[105] These GTs, or ABO synthases, may be active on several different core structure precursors, utilizing glycoproteins (as previously mentioned) or GSLs of both neolacto and globo series, (Fig. 3) forming type 2 and 4 chain antigens, respectively.[28,106] The ABO synthases belong to the GT6 family and so does the GT building the Forssman (FORS1) antigen of the FORS system, which highlights the similarities between the transferases. Forssman synthase encoded by *GBGT1* is active in the globo GSL synthetic pathway, transferring a GalNAc moiety and uses globoside as its acceptor substrate. It was previously believed that this gene was not expressed in humans, until the ABO subgroup A_{pae} was identified to be due to a SNP in the *GBGT1* gene sequence. The SNP in this case activates the FORS synthase's ability to make the Forssman antigen.[107] Antigens of both the Lewis and the H blood group systems are synthesized by the addition of a terminal fucose moiety.[108,109] The GTs synthesizing the A and B antigens utilize the H antigen as their precursor. Thereby the ABO antigens are dependent of a functional H synthase in addition to functional A or B synthase.[108] The Globoside blood group system contains the P (also known as globoside or Gb4) and PX2 antigens. The P antigen is the most abundant neutral GSL on the RBC surface[110,111] and PX2 was characterized as a blood group only five years ago.[111] The I antigen of the I blood group system, is synthesized in the branching of poly-*N*-acetyllactosaminoglycans, by the *GCNT2*-

encoded GT.[112] The remaining two carbohydrate blood group system P1PK and SID will be described in detail in following sections since they constitute the lion's share of the studies presented in this thesis.

The P1PK histo-blood group system

ISBT blood group system no. 003 has a long history. Both the name and the antigens belonging to it have changed over the years. It was established even though the requirement of the genetic background was lacking. The reason for this was the early discovery of the antigens and their obvious relation. Clarification through the years and several studies on the subject have resulted in that system no. 003 is now dedicated to the antigens P1, P^k and the very rare NOR1. However, the antigens of another system, the Globoside blood group system (no. 028), must be considered when addressing certain of the phenotypes, as the antigens of this system are in close relation to the antigens in the P1PK system.[113]

P1PK phenotypes

There are two common blood group phenotypes in this system, the P_1 phenotype in which individuals express both P1 and P^k on the RBC surface while individuals of the P_2 phenotype express P^k at low amounts and only cryptic levels of the P1 antigen, just detectable following treatment of proteases, if at all.[114] (Table 2) Initially, P^k was not believed to be expressed on P_1 and P_2 erythrocytes,[115] for reason explained in the next section. The prevalence of the P_1 vs. P_2 phenotypes varies in different populations. The frequency of the P_1 phenotype is highest in populations of African origin. It seems as a gradient stretches over Afro-Eurasia, where P_1 dominates in Africa/Europe and P_2 in Asia. (Table 3) The reason for this diverse ethnic distribution is unknown. During pregnancy, P1 appears to be easier to detect on fetal RBCs early in pregnancy than later, suggesting that expression is decreasing as the fetus grows.[116]

Table 2. A summary of the phenotype distributions of the P1PK and Globoside blood group systems.

Phenotype	Frequency (%)	Antigens present on RBCs
P_1	19-94*	P1, P^k, P, PX2
P_2	6-81*	P^k, P, PX2
p	Rare	PX2
P_1^k	Rare	P1, P^k
P_2^k	Rare	P^k

*Regional variation, see table 3.

Table 3. P1 antigen distribution in different populations.

Ethnicity/Origin	P1 positive (%)*	Total no. of samples tested	Study
African and Afro-American	94	33	Lai *et al.* 2014[117]
Ogoni, Nigeria	81	101	Christian *et al.* 2020[118]
Swedes, South Sweden	73	208	Thuresson *et al.* 2011[119]
Caucasian	70	46	Lai *et al.* 2014[117]
Iranian, Northeast Iran	66	522	Keramati *et al.* 2011[120]
Indian, South Gujarat, India	64	115	Kahar *et al.* 2014[121]
Indian	63	32	Lai *et al.* 2014[117]
Indian, Singapore	52	12.5**	Mohan *et al.* 1989[122]
Chinese Han, mainland China	38	1412	Yu *et al.* 2016[123]
Japanese, Japan	36	478965	Kawai *et al.* 2017[124]
Unspecific origin, Singapore	30	4.3**	Mohan *et al.* 1989[122]
Tawanese, Taiwan	28	227	Lai *et al.* 2014[117]
Malays, Singapore	27	13.7**	Mohan *et al.* 1989[122]
Chinese, Singapore	19	65.9**	Mohan *et al.* 1989[122]

*All studies but one analyzed the serology using commercial reagents, Kawai *et al.* 2017 used an in-house monoclonal reagent.
**Estimated number based on ethnic distribution in the donor population at Singapore general hospital.

The expression on RBCs is believed to reach adult levels by age 7.[125] The anti-P1 as found in the plasma of some P_2 individuals, is typically of IgM type and optimally reactive at lower temperatures. Although rare, anti-P1 can induce HTR. [126-128] The null phenotype in the P1PK system, denoted as p, is very rare. It is distinguished by the absence of both P1 and P^k antigens along with the P antigen of the Globoside system. (Table 2) Initially, p was called Tj(a–) based on patient J who had a tumor, but the name was changed as a relation was established to the other antigens.[129] The anti-P1PP^k occurring in plasma of p phenotype individuals is associated with recurrent miscarriages,[30] due to antibodies attacking the placenta.[50,52] The third antigen of the system, the NOR1 antigen is very rare and has so far only been detected in the members of two families.[130,131] When it was discovered in 1982 some unusual characteristics were reported. The inheritable RBC phenotype displayed polyagglutination, which is often an acquired phenomenon. However, this could be inhibited by hydatid cyst fluid, known to inhibit anti-P^k and anti-P1, and avian P1 blood group substance.[130] The inhibition pattern suggested a relation between all three antigens and family studies showed that it was likely to be inherited.

Both P1 and P^k have been suggested to be expressed on the surface of other blood cells, such as lymphocytes, granulocytes and are also found on skin fibroblasts.[132,133] The P^k antigen has been shown to be widely expressed in other human tissues, such as kidney, gastric tissue and capillary endothelial cells.[134] Anti-P1PP^k serum has shown reactivity with leukocytes, which suggests that one or several of its targets are present on these cells.[135] As previously mentioned, the same antibodies may target the placenta and cause spontaneous abortions, both the P and P^k antigens are found in large amount in placental tissue.[50]

The antigen structures

The P1 antigen was encountered already in 1927, initially only denoted as P.[136] The neolacto-series core sequence, paragloboside, is the precursor of P1, Galα1-4Galβ1-4GlcNAc-R, on GSLs.[137-140] The P^k antigen, discovered in 1959,[115] has the structure Galα1-4Galβ1-4Glcβ-R and is the initial carbohydrate sequence in the biosynthetic pathway of the globo-series.[28,141] (Fig. 3) In the study from 1959, the antigen today called P (i.e. globoside) was discovered as well.[115] This P antigen was established to be the elongation of P^k in the globo-series and now belongs to the Globoside blood group system because it is synthesized by another glycosyltransferase than P1 and P^k.[110] The P antigen thereby masks the P^k antigen, which is why anti-P^k may not be reactive with P_1 and P_2 RBCs.[141] In 1999, the NOR1 antigen was suggested to be a neutral glycolipid, terminated with a galactose moiety.[131] The whole structure was later deduced as Galα1-4GalNAcβ1-3Galα1-4Galβ1-4Glc-ceramide, which is an elongation of the P antigen.[142]

The synthase and genetics behind the antigens

The *A4GALT* on chromosome 22, encodes the 4-α-galactosyltransferase (α1,4Gal-T) that synthesizes the antigens of the P1PK system. α1,4Gal-T is categorized into the GT32 family as it transfers Gal from a UDP-Gal to the precursor/acceptor by retention. The fold of the enzyme has not yet been elucidated, only predictions have been made so far.[96] The gene *A4GALT* was cloned in 2000 and silencing mutations were identified in p individuals. Thereby a connection between the P^k antigen and *A4GALT* was established.[143-145] No polymorphism correlating to the P_2 phenotype (individuals lacking the P1 antigen) was identified at the time, but in 2011 such a SNP (rs8138197) was found in an area of the gene previously established as non-coding. The P1 antigen was thereby also declared to be dependent of the gene.[119] Although the concordance between phenotype and rs8138197 genotype was very high (correct in 207 of 208 samples), the molecular mechanism behind the P_1/P_2 phenotypes remained unknown. Transcriptional regulation was suggested already in 2003[114] as the *A4GALT* mRNA levels are higher in cells from P_1 individuals compared to P_2.[114,119] In addition, the In(Lu) phenotype of RBCs has been shown to be due to low levels of the TF KLF1 which coincides with decreased levels of both P1 and some other blood group antigens on the cell surface.[146,147] Additional SNPs rs5751348 and rs2143918, with an apparent 100% concordance with the phenotypes P_1/P_2, and located in the same region as the previously found rs8138197 (5'-end of the large intron 1) were identified in 2014.[117] The NOR1 antigen was associated with a SNP in the coding region, rs397514502, causing an amino acid change, p.211Q>E, and thereby broadening the GT's acceptor substrate specificity to include globoside GSL, the P antigen. Moreover, NOR1 may be further elongated to NORint and NOR2,[148,149] but this is not further discussed here.

Disease associations

The antigens of the P1PK system have been shown to affect susceptibility for several pathogens. PBMCs from individuals with the P_1^k phenotype express P^k highly and were shown to be resistant to HIV-1 infection *in vitro*, compared to common phenotypes, whilst cells of the P^k-deficient p phenotype are hypersusceptible.[150] The terminal carbohydrate sequence (Galα1-4Gal-R) is shared by the two antigens P1 and P^k (Gb3) and constitute receptors for uropathogenic P-fimbriated *Escherichia coli* (*E. coli*).[151] The P1 phenotype has been associated with susceptibility for urinary tract infection.[152] The Galα1-4Gal determinant is also recognized as a binding motif for Shiga toxin (Stx) produced by enterohemorrhagic *E. Coli* (EHEC).[153] The binding of Stx to P^k is described as crucial for the EHEC-caused renal and cerebral vascular injury[154] and important for the pathology of hemolytic uremic syndrome.[155] It has even been suggested to utilized this as a drug delivery system, i.e. use the non-toxic subunit B of the Stx, which binds to P^k after which the complex is endocytosed and transferred to Golgi. This is especially interesting as the epitope is overexpressed in several primary human cancers.[156] As an example, the P1 epitope is detected in the pleural mesothelioma cell line NCI-H226.[157] It has been suggested that *A4GALT* is involved in ovarian and peritoneal cancers due to its high expression in ovarian cancer cell line IGROV1, and the P1 antigen may assist in migration.[158] The P^k structure is also involved in Fabry disease, a lysosomal storage disorder. Affected individuals have a decreased lysosomal α-galactosidase A activity and thereby an accumulation of α-galactosyl moieties. This especially applies to the P^k, however, P1 and B (from the ABO blood group system) antigens may also accumulate. The gene *GLA*, coding for the enzyme is located on the X chromosome which is why males are more severely affected.[159] Finally, it can also be mentioned that PBMCs from Fabry patients are resistant to HIV infection *in vitro* because of the increased P^k levels, thus resembling the null phenotypes of the Globoside system, P_1^k and P_2^k.[160]

The P1PK antigens in other species

As already indicated, the antigens of the P1PK blood group system are present in hydatid cyst fluid.[130] The cysts are caused by parasites, i.e. tapeworms of the genus *Echinococcus* which may utilize humans as an intermediate host. The tapeworm egg or larvae enter through the wall of the small intestine to the circulatory system to transfer, most commonly, to the liver or lung where it develops to the hydatid cyst.[161]

The P1 antigen has also been detected in birds, e.g. it decorates the major proteins in egg white from pigeons. The carrier structures are N-glycans of complex type.[162,163] Interestingly, anti-P1 has a higher incidence in plasma among pigeon breeders with the P_2 phenotype as compared to a control group.[164]

The SID histo-blood group system

The SID histo-blood group system was only recently acknowledged as system no. 038 by the ISBT Working Party for Red Cell Immunogenetics and Blood Group Terminology. The Sd^a antigen was described more than 50 years earlier[165,166] and the gene sequence, suggested to express the GT synthesizing the antigen, was cloned in 2003 by two independent groups.[167,168] However, the reason for leaving the antigen in the high-frequency series of orphan blood groups was the last missing piece, namely to identify the genetic alteration that causes the Sd(a–) phenotype. Data presented in this thesis, **Paper III**, contributed to solve the question and consequently led to announcement of a new system.

SID phenotypes

The high-frequency histo-blood group antigen Sd^a was described for the first time in 1967. Anti-Sd^a had been detected in incompatible cross-matching tests for years, however the characterization was challenging. The strength of the antibody reaction with Sd(a+) RBCs varies greatly between individuals and the agglutination pattern is characterized by well-defined but small agglutinates in a sea of free cells.[165,166] This pattern seems not to be explained by RBC age.[166] The frequency of the different SID blood group phenotypes in the European population can be seen in Table 4. Sd^a is detected in various tissues and a discrepancy in the frequency of apparently Sd(a–) individuals was discovered early on. That is, some individuals have a phenotype where the Sd^a antigen is lacking on the RBCs but present in other tissues, see Table 4. This makes the true incidence of Sd(a–) individuals to be in the range of 2-4%.[169,170] Furthermore, it is more common to find what looks like a Sd(a–) blood group phenotype among pregnant women compared to normal blood donors, and possibly the number increases closer to term.[165,169,171] This is a feature that the Sd^a antigen shares with the Lewis system antigens Le^a and Le^b and may at least partially be attributed to increased blood volume during pregnancy.[172]

Table 4. A summary of RBC SID blood group phenotype frequencies in the European population.

RBC phenotype	Frequency (%)*
Cad+	0,07**
Sd(a+)	90,6
Sd(a–)	9,5 (2-4 in other tissue)

*Compiled data from the following studies: Macvie *et al.* 1967[165], Renton *et al.* 1967[166], Morton *et al.* 1970[169], Gerbal *et al.* 1976[173], Morton *et al.* 1988[170] and Conte *et al.* 1991[174]
**RBCs of B and O blood group phenotype

The Sd^a antigen is currently the only antigen of this system, however, in addition to the regular Sd(a+) and Sd(a–) phenotypes, there is another variant in which RBCs are highly reactive with anti-Sd^a. Whether the epitope is structurally altered from that on the common, weaker reactive RBCs or if there is a quantitative difference, or indeed both, is not known. This Cad blood group phenotype, a.k.a. Sd(a++) or super-Sid, was first described in three individuals of a Mauritian family of Indian origin. Their RBCs were agglutinated when mixed with sera of 83 individuals and appeared to be polyagglutinable. Like Sd(a+) RBCs only a portion of the cell population is agglutinated, see Fig. 5. In addition, *Dolichus biflorus* agglutinin (DBA), a plant lectin generally used for blood group A phenotyping, bound to the RBCs. This was unexpected as the three Mauritians were of blood type B and O. In a French population, 250 000 subjects were tested for this highly reactive antigen but all were negative.[175] The characteristics described above is named Cad 1 as two other Cad subgroups since then have been defined. Cad 2, found on RBCs in individuals of Asian origin (Vietnamese, Thai and Chinese) occurs relatively frequent, 0,26% among Thai blood group donors of B and O phenotype.[176-178] Cad 3 was identified in two European families (Polish and Sicilian).[176] The Cad frequency among Europeans (subgroup not defined) is very rare, see Table 4. The association between the Sd^a antigen and the Cad+ phenotype was made in 1971, as the RBCs of four unrelated Cad individuals were strongly agglutinated by anti-Sd^a. The reaction was much greater than that of the original Sd(a+) RBCs, which had been the strongest reactor to that point.[179] Interestingly, the Cad phenotype is inherited in a dominant way.[178] This is in contrast to the Sd(a–) phenotype which appears to be a recessive feature.

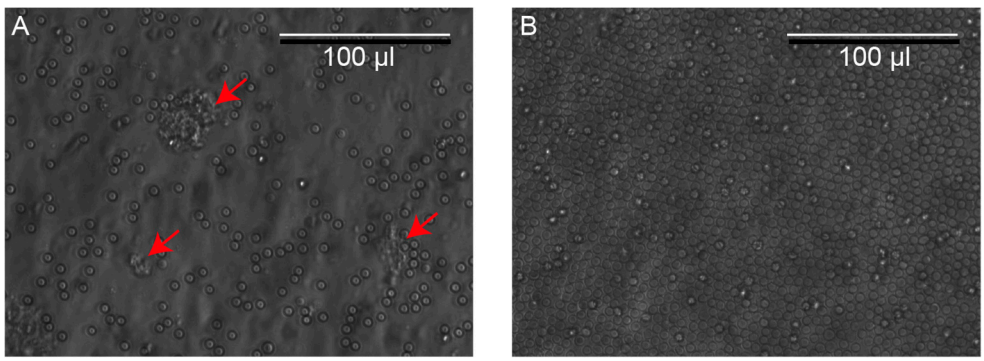


Figure 5. Hemagglutination pattern with anti-Sd^a. Cad RBCs mixed with **A)** plasma from a Sd(a–) individual with anti-Sda or **B)** with an ABO-matched plasma control without anti-Sd^a. Cells were observed under Evos FL microscope (Thermo Fisher Scientific) at x40 magnification.

The antibodies against Sd^a are naturally occurring. Anti-Sd^a is generally considered to be of IgM type, however IgG has been detected and a rise in IgG titre followed transfusion has been observed.[165,166,171] The antibodies have generally been regarded

as clinically insignificant, however HTR has occurred.[180,181] Especially blood transfusion from a donor with the Cad phenotype should not be distributed to a patient of the Sd(a–) phenotype, particularly not if anti-Sd^a is present.

In adults, the Sd^a antigen has been detected in tissues of the colon, kidney as well as soluble in urine and saliva. The blood cells and colon of the human fetus have been shown to be low or negative for the antigen. However, it has been detected in the urine and at high levels in the saliva of the newborn infant and the antigen appears on the RBCs in humans at age 10 weeks to 7 months.[165,166,169,170,182]

The antigen structures

The initial studies of the Cad phenotype indicated that a terminal GalNac moiety was involved in the antigen structure as GalNAc binding lectins such as the previously mentioned DBA and those from *Helix pomatia*, *Helix aspersa* and *Wisteria sensis* all agglutinated the Cad RBCs.[175,176,183] The antigenic structures of the Sd^a and Cad have later been assessed in different types of human tissue, with the common terminal trisaccharide being identifed as GalNAcβ1-4(NeuAcα2-3)Gal-R. In Cad RBCs the epitope was defined on the glycoproteins GPA and GPB as O-GalNAc core 1 glycans, see Fig. 4C-I, as well as on GSLs, as the elongated paragloboside in the neolacto series, Fig. 3.[99,184] However, in RBC membranes with Sd(a+) phenotype this kind of characterization has not been achieved, although attempts have been made.[185] In the urine of Sd(a+) individuals the abundant uromodulin, also known as Tamm-Horsfall glycoprotein, carries the antigen on N-glycans, Fig. 4C-II.[186-188] Sd^a was found to be a major structural feature in the descending colon, Fig. 4C-III. The carriers are the highly glycosylated mucins, and the epitope was found here on core 3 carriers.[189]

The synthase and genetics behind the antigen

The 4-β-*N*-acetylgalactosaminyltransferase 2 (β1,4GalNAc-T2), has been detected in human healthy kidney, plasma, urine and colon. The GT has the capacity to synthesize the Sd^a/Cad antigen by using the donor substrate UDP-GalNAc and the acceptor group NeuAcα2-3Gal-R. β1,4GalNAc-T2 requires sialylated acceptor molecules and can utilize acceptor substrate on both O-GalNAc- and N-glycans, as well as the sialylparagloboside in the GSL neolacto synthetic pathway.[190-193] It belongs to the GT12 family, as does its homolog β1,4GalNAc-T1, which synthesizes the GM2 epitope. GM2, in the ganglioside series is composed of an identical terminal trisaccharide as the Sd^a epitope. However, the β1,4GalNAc-T2 cannot use the precursor GM3 and build GM2.[57]

In 1996, a transcript corresponding to *B4GALNT2* was partially cloned[194] and seven years later two groups cloned the whole gene situated on chromosome 17. [167,168] It

was suggested that *B4GALNT2* expressed the GT building the Sd^a epitope based on a number of arguments: 1) sequence similarities with the murine homologue, *B4galnt2* and human *B4GALNT1*, 2) the human tissues where the gene is expressed overlap with Sd^a locations, and 3) the GT had the expected substrate specificities. There is a region of GpC islands upstream of the gene that indicates the promoter area.[195,196] The gene expresses two isoforms utilizing different exon 1 alternatives, where one will carry an extended cytoplasmic tail usually not seen in GTs. This feature alters the GT cellular location, placing it post Golgi but the significance of this is not clear.[197] As expected, there is a DXD motif in the catalytic domain where the UDP part of the donor substrate is coordinated to bind. This enzymatically active domain is highly homologous with the corresponding amino acid sequence in β1,4GalNAc-T1.[167] According to RNAseq data there is no expression of the *B4GALNT2* in erythroid cells differentiating towards RBCs. Additionally, no GATA1 motif has been associated with the gene, a feature otherwise seen in most of the erythroid expressed genes coding for blood group antigen synthesizing GTs.[198] This suggests that Sd^a is acquired from a different source. It might be analogous with the GSLs carrying the Lewis antigens, which are adsorbed from the plasma onto the RBC membrane.[199] More research on the subject is required to establish the cellular origin of the RBC carried Sd^a antigen but the gastrointestinal tract in general, and the descending colon in particular, appear to be the likely source. Finally, the genetic base and molecular mechanism underlying the Sd(a–) and Cad phenotype had not been resolved, prior this study.

Disease associations

B4GALNT2 expression is severely decreased in colorectal cancer, resulting in downregulation of the Sd^a epitope. Instead the sLe^x, an epitope built on the same precursor as the Sd^a epitope can be detected. In cancer cells sLe^x may be used as a ligand for E-selectin on endothelial cells, thereby important in the metastasis process. It has been shown that forced *B4GALNT2* expression reduces metastasis.[190,200,201] The promoter region of the gene, containing a CpG island, is heavily methylated in cancerous tissue which may explain the down regulation.[195,196] Finally, serum glycoprotein carried Sd^a antigen has been suggested to be used as a tumor marker.[202]

When it comes to microbial pathogens, Cad RBCs have been shown highly resistant to *Plasmodium falciparum* invasion. GPA on the cell surface is involved in the attachment of the parasite to the RBC and the sialic acid residues (often terminating glycan structures) may be involved. The addition of a GalNAc to the sialic acid carrying structures, in creating the Cad epitope on GPA, may interfere with parasite-cell interaction.[203] Similarly, *B4GALNT2* expression has shown an inhibitory effect on influenza A virus. Again, the pathogen attaches to α2,3-bound sialic acid, which is the type of precursor the β1,4GalNAc-T2 uses.[204]

Research on muscular dystrophy is often performed using mouse models. In dystrophic mice, overexpression of *B4galnt2* (a homolog of human *B4GALNT2*) reduces the pathology. Unfortunately, the mechanism is poorly understood and to be able to translate this to humans a better understanding of the glycosylation in human muscles is needed.[205]

The Sd^a antigen in other species

Several *B4GALNT2* homologues are of interest in various species. Early on, the Sd^a antigen was recognized in species such as guinea pig, mole and hedgehog.[169] Actually, guinea pig urine has high concentrations of the Sd^a antigen, and are therefore a useful tool in inhibitory assays, identifying antibody specificity.[25] Yet another set of animals has lately been the focus in several research fields.

B4GALNT2 in pigs express a synthase with 76% amino acid sequence identity to the human β1,4GalNAc-T2. It builds similar structures as the Sd^a/Cad epitope, for instance the DBA lectin binds to this structure. Nevertheless, this has been acknowledged as a xenoantigen and most humans have both IgG and IgM antibodies against it. Supposedly, the carrier structure differs from that in humans and the implicated moiety might be Neu5Gc, a type of sialic acid that humans lack.[57]

Mice express *B4galnt2*, as previously mentioned in the section about disease associations. However, the muscles are not the only tissue this homologue is expressed in. Most strains express the gene predominantly in intestinal endothelium. However, some strains have an allele called *Mvwf1*, which carries a *cis*-acting mutation 30 kb upstream of the gene. This motif causes a regulatory switch, which makes the gene express in the vascular endothelium. A downstream consequence of this is decreased half-life, and thereby lower plasma levels of the von Willebrand factor (vWF). The vascular endothelium is the primary site of vWF synthesis, and the altered glycosylation has been hypothesized to cause clearance via the GalNAc-binding hepatic asialoglycoprotein receptor in the liver. This feature has been seen in both inbread and wildtype mouse strains.[206-208] Fertility is another area of research interest since there are indications that *B4galnt2* expression is regulated by progesterone and estrogen, and involved in embryo implantation.[209]

In sheep, the gene is primary expressed in the ovaries and it seems to have an impact on fertility. Several breeds carry a SNP in intron 7 of their *B4GALNT2*, called the FecL^L mutation, which leads to increased expression and glycosylation in ovarian granulosa cells. This alteration has been shown to correlate with ovulation rate and large litter size, however the mechanism of action is yet unknown.[210] Another study identified two missense mutations in a different breed also suggested to affect litter size.[211]

Present investigation

Aims of this work

The overall aim of this work was to bring clarification to the genotypic backgrounds of certain carbohydrate blood group antigens that puzzled the research field for decades. In addition, insights on the antigen carrier structures have been achieved. The specific aims were:

- To define the molecular mechanism underlying the P_1 vs. P_2 blood groups.

  Paper I

- To investigate if the human P1 antigen is carried on erythrocyte membrane glycoproteins.

  Paper II

- To explain the genetic background of the Sd^a negative phenotype.

  Paper III and IV

- To study the glycoprotein carrier structures of the Sd^a antigen.

  Paper IV

- To investigate the genetic characteristics of the Cad phenotype.

  Paper IV

Methods

In the table below the methods used in the papers included in this thesis are listed, more comprehensive descriptions and manufacturer details are stated in Papers I-IV.

Method	Paper	Description
Hypotonic lysing of RBCs (membrane preparation)	II, IV	RBCs are lysed 1:10 in cold lysis buffer with protease inhibitor and pelleted. The membranes are washed until the hemoglobin has been removed.
Chloroform-methanol lipid extraction of RBC membranes	II	Chloroform-methanol 2:1 solution is mixed with RBC membranes 9:1, followed by phase separation.
Enzymatic digestion of RBC surface antigens	II	RBCs are incubated in enzyme solution in appropriate buffer at 37°C, for 5 min-16 h. Reaction is stopped by washing in PBS.
SDS-PAGE and western blot	I, II, IV	Denatured proteins are separated applying Bio-Rads V3 western workflow. Immunostaining and visualization by HRP-conjugated antibodies, developed signal is detected in the chemidoc imager from Bio-Rad.
Flow cytometry	IV	Cell structures are detected by fluorochrome conjugated antibodies or vector expressed flurochromes, as a single cell passes a laser beam. Light scatter and emitted fluoroscence are recorded and the data anlysed.
Fluorescence microscopy	IV	Cells are mounted upon slides and stained with biotinylated lectin followed by fluorochrome conjugated streptavidin. A cover glass is added and fluorescence detected in a microscope.
Transfections of siRNA into suspension cells by electroporation	I	Cells are mixed in a solution of siRNA and electroporated. After 48 hours culture at 37°C, 5% CO_2 the cells are harvested.
Transfections of plasmid construct into adherent cell line	IV	At 80% confluency cells are transfected by the adding of premixed plasmids in transfection reagents. The cells are further cultured at 37°C, 5% CO_2 and harvested after 24 hours.
Real-time quantitative PCR	I	RNA is extracted and cDNA synthesized, the amplification cycles of targeted transcripts during PCR are recorded and the $\Delta\Delta C_T$ analyzed.
Amplification and Sanger sequencing of DNA	III, IV	DNA is prepared from whole blood and targeted region is amplified. The amplicons are purified and Sanger sequencing applied.
Allelic discrimination assay	II, III	Taqman SNP genotyping. A probe binds allele specific and a fluorescent signal will be detected as the polymerase transcribe the gene.
Statistical analysis	I, II, IV	Nonparametric Mann-Whitney U test is used to evaluate differences between groups. In case of more than two comparrisons in the same experiment, the Bonferroni correction is applied.
EMSA (Julia S. Westman, PhD)	I	Double stranded biotinaylated probes are mixed with nuclear extracts, (for supershift an antibody is added). The samples are separated and processed, according to protocol, the result is recorded using the chemidoc imager from Bio-Rad.
LC-MS/MS and proteomic analysis (Sven Kjellström, Ph.D.)	I	Probe bound proteins immobilised on beads are trypsinated and analyzed on a mass spectrometer.
LC-MS/MS and glycoproteomic analysis (Jonas Nilsson, Ph.D.)	IV	Cell lysates are trypsinated and glycopeptides enriched with HILIC to be analyzed on a mass spectrometer.
Bioinformatic analysis (Mattias Möller, M.D.)	I,III	Applying software tools to analyze large biological data sets. This may be used to predict outcome of variants, protein structure modelling, protein structure conservation evaluation, linkage disequilibrium, gene expression patterns in cell lines, etc.

Abbreviations: SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; HRP, horseradish peroxidase; PCR, polymerase chain reaction; EMSA, electrophoretic mobility-shift assay; LC-MS/MS, liquid chomatography-tandem mass spectrometry HILIC, hydrophilic interaction liquid chromatography.

Summary of results

Paper I

This project was initiated with an evaluation of the nucleotide sequences surrounding three candidate SNPs (rs8138197, rs2143918 and rs5751348), previously proposed to be implicated as P^1/P^2-differentiating. One of them soon became the prime suspect, namely rs5751348. Using bioinformatics tools and visual examination, it was concluded that KLF1 had a binding sequence consistent with this site, but so did RUNX1 and early growth response factor 1 (EGR1). Haploinsufficiency of KLF1 in the In(Lu) phenotype is known to be associated with weakening of several blood group antigens on RBCs,[146,147] including a decreased P1 expression. KLF1 was therefore considered the primary candidate. Encouragingly, all three candidates were predicted bioinformatically to decrease their binding energy score significantly for this sequence in the presence of the P^2-related rs5751348:T, as compared to the P^1-associated rs5751348:G. By designing oligonucleotide probes of different lengths and variants, corresponding to the sequence around rs5751348, it was possible to show by electrophoretic mobility shift assay (EMSA) that a nuclear extract protein bound the P^1 allele specifically, however several results implied it was not KLF1. Importantly, anti-KLF1 did not give a supershift and KLF1 could not be immunoblotted to show its presence on the probes. Instead, liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis identified the binding transcription factor eluted from the probes to be RUNX1, a finding which could also be confirmed by immunoblotting and a competition assay with anti-RUNX1. The functional effect of RUNX1 on *A4GALT* expression was shown by knocking down *RUNX1* in two different cell lines, using siRNA, which in turn decreased the levels of *A4GALT* mRNA. In contrast, *KLF1* knockdown did not change the levels of *A4GALT* mRNA.

In addition to our data, a parallel study was published almost simultaneously in *Transfusion*, identifying EGR1's involvement in the phenotype as well.[212] The genetic basis for the P1 antigen expression, first detected in 1927, has thereby finally been unraveled and the underlying molecular mechanism begun to be understood.

Paper II

According to textbooks, GSLs are the sole carriers of P1 antigen on human RBCs[25,30] and this has remained the dominating opinion, even though there are contradicting studies published on the subject.[213,214] In this study, P1 expression on glycoproteins were evaluated by SDS-PAGE and immunoblots using RBC membranes from P1 positive and negative individuals. The monoclonal IgM anti-P1 (clone P3NIL100) is known to be specific for P1 and not to crossreact with the

related P^k antigen or other RBC surface structures. Here it was also evaluated in experiments of galactosidase-digestion and used to blot the P1 epitope on RBCs. The antibody stained a smear in the lanes of the homo- and heterozygous P_1 samples, while nothing was detected in samples from individuals of P_2 or p blood type. The P^1/P^1 samples showed a significantly higher expression compared to P^1/P^2 samples. In addition, there were some differences within the genotype groups, which is similar to previously reported findings regarding the total expression of P1.[119,215] Papain digestion did not seem to influence the protein carrying P1, however, after PNGase F treatment of the RBCs we could conclude that the epitope is, at least partially, expressed on N-glycans. Experiments were conducted to propose, but not prove, that band 3 and GLUT1 may be major P1 carriers in the RBC membrane. This would nicely mirror their roles as the main ABH antigen carriers on glycoproteins.[30] Additionally, the findings in Paper I were followed up by developing and implementing a P^1/P^2 genotyping assay, based on the rs5751348, which is now in clinical routine use at the Nordic Reference Laboratory for Genomic Blood Group typing in Lund, Sweden.

Paper III

The genetic basis that distinguishes the Sd(a+) and Sd(a–) phenotypes has remained unknown for decades. With this study we set out to elucidate if any correlating mutation could be found. In a cohort of nine true Sd(a–) individuals, known to have anti-Sd^a in their serum, we sequenced the coding region of *B4GALNT2*, and its promoter. A missense mutation, rs7224888:T>C was found in homozygous form in six individuals, and heterozygously in one. This freqence, 13 alleles out of 18, is much higher than in the general population. Screening of 211 Swedish blood donors gave an occurrence of 10%, a representative number for Europeans when comparing to the superpopulations, which varies from 2.2% to 20.2% according to 1000 Genomes.[216] The lower incidence represents east Asia, which agrees well with the 3.3%, that were obtained by screening 183 Thai donors for the mutation. The rs7224888 heterozygous individual also carried a splice-site mutation in intron 8, rs72835417. *In silico* analysis indicated that this variation may have an effect on splicing. Screening the Swedish and Thai donors for this SNP gave incidences of 11% and 0.55%, respectively. Furthermore, one Sd(a–) individual was compound heterozygous for two rare missense mutations, rs148441237 and rs61743617, and one individual had no variation to be found. All the three detected missense mutations cause amino acid changes localised in the GT's catalytic domain. The rs7224888 causes the exchange of an evolutionary conserved cysteine to an arginine. The allele of rs7224888:T>C also carries a synonymous SNP in exon 11, rs16946912:A>G. Linkage disequilibrium between rs7224888:T>C and this SNP along with surrounding SNPs revealed a haplotype that spans over 32 kb, starting in *B4GALNT2* intron 9 and continuing downstream of the gene. This proposed haplotype also includes a lncRNA, RP11-708H21.4, which also came up when

scanning the cancer cell line encyclopedia (CCLE) for *B4GALNT2* co-expressing transcripts. Interestingly, this lncRNA has been associated with poor prognosis of colorectal cancer. Like the expression of *B4GALNT2*, RP11-708H21.4 is downregulated in malignant tissue.[217]

Simultaneously with this publication an abstract was presented by a Dutch group of investigators at the ISBT regional congress 2019 in Basel, Switzerland.[218] The data corroborated our findings regarding rs7224888 and rs72835417 in an independent cohort of Sd(a-) individuals, which emphasizes the correlation identified. Following our formal proposal and a presentation of our data at the ISBT congress in Basel on 22 June 2019, the ISBT Working Party for Red Cell Immunogenetics and Blood Group Terminology, acknowledged SID as a new blood group system.

Paper IV

In the previous study of **Paper III**, rs7224888:T>C, situated in exon 10 of *B4GALNT2,* was identified as the predominant Sd(a–)-associated mutation. In this study we set out to determine if it is causing the null phenotype of the newly-appointed system. The human embryonic kidney cell line, HEK293, was transfected with constructs of the *B4GALNT2* coding region (Genbank accession no. AJ517771). Initial experiments were analysed by fluorescence microscopy and followed-up with experiments where data were collected by flow cytometry. Cells transfected with consensus *B4GALNT2* were positive following DBA lectin staining, while mock-transfected cells (with an empty vector) were negative. Cells transfected with *B4GALNT2* carrying rs7224888:T>C did not bind DBA. Constructs of the gene with either of the two rare SNPs rs148441237 and rs61743617, found in one Sd(a–) individual, did not alter staining compared to consensus. In addition, glycopeptide analysis of the transfected cells was performed in collaboration with glycoproteomics experts at the University of Gothenburg. To achieve this, glycoprotein-carried glycans were analysed in LC-MS/MS. The results confirmed the previous DBA staining: the Sd^a structure was found in all samples but the mock and *B4GALNT2*-rs7224888:T>C transfected cells. The Sd^a antigen terminated glycans on nine different proteins, both on O-GalNAc- and N-glycans. The β1,4GalNAc-T1 was detected in SDS/PAGE gels and western blot in all four samples of construct-transfected cells, while no band was seen in the mock-transfected sample. This showed that both consensus and mutants were readily expressed and detectable.

A remaining unsolved question related to this system is the genetic background of the Cad phenotype. We sequenced the coding region and approximately 2000 base pairs upstream of the gene in five individuals with the Cad phenotype, subtype not defined. Three were of Asian origin, hence it would be reasonable to suspect the Cad2 phenotype. There were no common or critical alterations detected compared to the consensus sequence. To conclude, the results showed that rs7224888:T>C

indeed defines and causes the Sd(a–) phenotype in most cases. However, there are still unexplained genetic causes for a minority of all investigated donors with the Sd(a–) phenotype. The Cad phenotype shows no deviation from consensus in the regions tested here and therefore remains unexplained. Thus, more work is needed to understand both null and exceptionally strong phenotypes of this system.

Discussion

The molecular mechanism and biochemical carrier of the P1 antigen

The hypothesis of regulatory differences causing higher levels of *A4GALT*-encoded GTs in P_1 individuals, and thereby synthesis of readily detectable amounts of P1 antigen, has support in previously published data. Namely, the higher transcript levels in P_1 RBCs compared to P_2,[114,119] and the decreased P1 antigen expression in In(Lu) phenotype RBCs, caused by shortage of the TF KLF1.[146,147] The findings of P_1/P_2-discriminating SNPs in intron 1 of *A4GALT*, indeed suggested that the inherited, functional difference between the phenotypes was to be found in association with this gene. In **Paper I** we were able to distinguish potential TF motifs in the region surrounding the rs5751348:G. However, it was not KLF1 that was binding to P^1 (or for that matter P^2) probes, nor were *A4GALT* transcript levels affected when KLF1 expression was knocked down in erythroid cells. Instead, RUNX1 was identified by mass spectrometric methods to bind the P^1 allele specifically and by knocking its expression down, the levels of *A4GALT* were significantly reduced. EGR1, the third TF with a potential binding motif in the area of the implicated SNP was not detected in our study. However, another group showed that this TF also binds the P^1 allele specifically and induces expression of *A4GALT*.[212] Altogether, this reveals that rs5751348:G and the surrounding sequence constitute an enhancer motif utilized by more than one TF in case of the P^1 allele, but not P^2. It would not be remarkable if several TFs were involved in the enhanced expression of *A4GALT* since TFs often bind in complexes and work together to drive gene expression. TFs may even be dependent on the presence of other enhancer elements to be able to bind.[84] However, it is not apparent that RUNX1 and EGR1 would be associated in the same complex. Instead, the TFs have been implicated at different stages and lineages of blood formation. RUNX1 is not directly associated with the erythroid lineage, but it is active in the early hematopoiesis and the megakaryocytic differentiation. In the embryo RUNX1 is of great importance for definitive hematopoietic progenitors and HSCs, and is one out of several TFs that are needed for the establishment of the hematopoietic program.[83] Further ahead, RUNX1 is associated with silencing the erythroid lineage in favor of the megakaryocytic lineage.[219] *In vitro* cultures, of human bone marrow derived CD34+ cells, driven toward erythroid lineage, display appearance of *A4GALT* mRNA,

before the cells are readily committed.[220] Taken together the *A4GALT* expression may very well be upregulated in the early stages of hematopoiesis and thereby the surface antigens on the RBCs linger from an early stage. EGR1 is involved in macrophage differentiation and it is activated upon erythropoietin stimuli, which supports the erythroid lineage.[221,222] Moreover, one should not entirely rule out the possible influence of KLF1. This TF has a great impact in the erythroid commitment and differentiation of RBCs. In this role it may have a large impact of the overall composition of the RBCs, as seen in the In(Lu) phenotype, where P1 levels also are affected. Thereby, it is possible that even though the KLF1 does not bind directly to the particular site to enhance this gene it may regulate other pathways leading to increased *A4GALT* expression. Studies on primary human bone marrow cells could be tested as a complement to the two published studies.

With **Paper II** the detection of the P1 antigen on glycoproteins rewrites a common perception of the structure. In blots of the P1 antigen, a smear was visible in the lanes of P1-positive RBC membranes, indicating that proteins of various sizes carry the epitope. P1 expression in RBC membranes from individuals of the P^1/P^1 and P^1/P^2 genotypes differ, showing higher levels of P1 among homozygotes. This mimics what has been known from serology for a long time, namely that the P1 expression varies a lot from one P_1 individual to another. That this is dependent on dosage had been proposed earlier but was shown clearly in 2011.[119] The amount of P1-staining glycoprotein from P_2 erythrocytes was in level with the negative control, RBCs of the p phenotype (without functional *A4GALT*-encoded GT). Earlier findings in a study where intact RBCs from P_2 individuals were treated with three different proteases suggested the cells carry cryptic P1. Anti-P1 staining with flow cytometry analysis showed low levels of, presumably GSL carried, P1 on the cell surface of the protease-treated P_2 cells, compared to untreated cells, which stained P1 negative.[114] This could be a sensitivity issue of the different methods, but if not, this may distinguish the GTs preference of substrate. Low levels of GTs will have a high access to its preferred acceptor substrate, compared to if higher levels of GT are produced. In the latter scenario, the first choice of substrate may not be available in large amounts and it may be possible that the GT starts working on acceptor substrates for which it has lower affinity, and so forth. Based on this, the suggested preference would be lactosylceramide (to produce P^k-GSL) > paragloboside (to produce P1-GSL) > glycoproteins (to produce P1 on N-glycans).

The blots of PNGase F-treated membranes in **Paper II**, indicated the epitope being present on N-glycans. Bands around 50 and 100 kDa were observed in the smear-like pattern of all immunoblots probed with anti-P1. These data gave hints of what proteins may be the carriers of P1. Several proteins on the RBC surface are glycosylated and carries N-glycans. The sizes where P1 seemed to accumulate were of interest as these coincide with two such abundant RBC proteins, GLUT1 and band 3. Both are known carriers of the ABH antigens,[30] which have structural similarities with P1. The GTs behind these structures can be active in the neolacto

GSL synthetic pathway, Fig. 3. In fact, both the H antigen and P1 (and PX2) share the same GSL precursor/acceptor, namely paragloboside. Thereby we know they have the capacity to work on similar structures. However, this is only circumstantial evidence and further detailed work is needed before any conclusions about the identity of the protein carrier(s) can be drawn.

Understanding the carrier structures in the P1PK system are of interest not only in transfusion medicine. Recently, a study conducted in a modified cell line, showed that P1 on glycoproteins is a less efficient receptor for Stx, the toxin produced by EHEC, when compared to GSL-carried P^k. The internalisation of Stx bound to P1-glycoprotein stays in the endosomes and does not seem to reach the Golgi. Additionally, when both GSL-P^k and glycoprotein-P1 were expressed, sensitivity for the toxins remained as the P1 epitope acted as a decoy.[223]

In our study western blot was utilised to detect the P1 epitope on glycoprotein. We evaluated and excluded any contamination of GSL-carried P1. GSLs are also known to separate on a gel such as this, however GSLs would generally be seen at much lower sizes than those observed here. Additionally, in blots of the protein fraction of RBC membranes the P1 signal remains, although weakly due to concentration reduction in the process. Additionally, the PNGase treatment would not have affected the GSL fraction, however in our experiment this reduces presence of P1. Even so it would be interesting to confirm and further evaluate the variety of P1 carriers. After PNGase F treatment a fraction of P1 remains, this may be due to O-GalNAc-carried P1, or possibly inefficient glycosidase action. Unfortunately, no deglycosylation approach such as how PNGase F works on N-glycans, was available for O-GalNAc. A possible approach to deduce both the glycan structure as well as the carrier proteins would be LC-MS/MS. However, then the low amount of P1 carried proteins may be a challenging hurdle to overcome and may require some kind of enrichment of targets.

The genetic background of the Sd^a antigen

Among Sd^a-negative individuals one SNP dominated the cohort studied, namely rs7224888:T>C (13 of 18 alleles). This was presented in **Paper III** and further supported as the same conclusion was drawn by another research group in an independent cohort.[218] The study from the Netherlands also presented data on a group of nine Sd(a–) individuals and identified rs7224888 (in 17 of 18 alleles). The SNP causes an amino acid change, where a small cysteine is substituted, an amino acid often known to be structurally important in proteins as it may form disulphide bonds involving other cysteines. In this case, the cysteine is replaced with an arginine, an amino acid with a large and charged side chain. Here, a comparison with the homolog β1,4GalNAc-T1 that has very similar activity as the β1,4GalNAc-T2, and with a high sequence resemblance in the region surrounding the SNP,[167,168] becomes interesting. The β1,4GalNAc-T1's structural conformation in soluble form

has been evaluated with special focus on its cysteine moieties. It was concluded that a disulfide bond is formed between Cys^{429} and Cys^{476}, both residues are conserved in β1,4GalNAc-T2, and the former corresponds to the amino acid exchanged due to rs7224888. In **Paper IV** the rs7224888:T>C substitution was further evaluated with focus on causality, i.e. is it this change that leads to the enzyme's inability to make Sd^a antigen and thereby causes the Sd(a–) phenotype in most cases. Overexpression of *B4GALNT2* resulted in clear Sd^a expression on the cell surface as measured by several methods, while *B4GALNT2* carrying the SNP encoded a GT unable to produce the antigen. This indeed proves the necessity of the cysteine residue for functional Sd^a synthesis.

The two rare SNPs, rs148441237:A>G and rs61743617:C>T, detected as compound heterozygous in one Sd(a–) individual in **Paper III** also cause amino acid changes in the catalytic domain. In the overexpression experiments in **Paper IV**, these two SNPs did not cause any apparent alteration in Sd^a expression compared to consensus as measured by the methods used. However, we know there must be other genetic explanations for lacking the Sd^a antigen than missense mutations in *B4GALNT2*, as one Sd(a–) individual in the cohort of **Paper III** did not reveal any discrepancies from the consensus coding region. The same, or a different cause, might be found in the individual being compound heterozygous for rs148441237:A>G and rs61743617:C>T. The extended, non-coding region of *B4GALNT2* may be of interest, for further investigation of variations in the surrounding area or within the introns, where regulatory motifs may be situated. A regulatory site may be located far away and this has been seen for the *B4galnt2* locus in mice, where such a site is located 30 kb upstream of the gene. It should however also be considered if the experimental set-up might not have been optimal for distinguishing alterations caused by these SNPs. E.g., the alleles should be tested in co-transfection experiments to verify if the products of each allele affect one another, leading to a joint impairment of the catalytic effect only seen when these two rare variants are combined *in trans*. In the LC-MS/MS analysis of glycoproteins from cell line overexpressing *B4GALNT2* in **Paper IV** we also determined the carrier structures of Sd^a. As previously known the GT is able to use carriers of both O-GalNAc- and N-glycan type. No difference of carriers was seen comparing the glycoproteins of consensus-transfected cells comparing to cells transfected with either of the two rare SNPs, previously discussed. Hence, no qualitative differences were able to be established either.

One of the Sd(a–) individuals in **Paper III** and one in the abstract from the Netherlands cohort were compound heterozygous for the rs7224888:T>C and the splice-site variant rs72835417:G>A. These two individuals are distinguished by the fact that our sample had anti-Sd^a, while the other did not. Rs72835417:G>A may affect the expression of a functional GT, but its exact effect has not been thoroughly investigated yet. The fact that it is not very frequent in the two investigated cohorts and no individuals were found to be homozygous, suggests that it may not abolish

the GT activity completely. This is particularly striking if the frequencies of the predominant polymorphism (rs7224888:T>C) and the splice-site variant (rs72835417:G>A) are compared as was done in **Paper III**. Given how similar they are in Europeans, one would expect to see a more even distribution in the two Sd(a–) cohorts, if both would abolish enzyme activity. It would be interesting to evaluate the variant close to a splice-site further as it did appear in both cohorts. Maybe the SNP indeed causes lowered levels of the antigen, not detectable in serological tests if an individual would be homozygous, but enough for such an individual not to make anti-Sd^{a}. As our cohort only involved individuals with anti-Sd^{a}, this could explain why we only detected the SNP in compound heterozygosity with the now established impairing SNP rs7224888.

The haplotype carrying rs7224888:T>C and the synonymous SNP rs16946912 was also an interesting finding in **Paper III**. This haplotype block connects the *B4GALNT2* and the lncRNA RP11-708H21.4, two loci that are apparently frequently co-expressed. That is, high or low mRNA levels of the two overlap well in cell lines. Both genes are downregulated in cancerous tissue and have been shown to inhibit malignancy.[201,217] Currently, no study has been conducted on colon tissue, healthy or malignant, where both the genetic regions are taken into consideration simultaneously. It would be interesting to establish if and how they affect one another and perhaps even more so if the Sd(a–) phenotype, which is not too uncommon, affects or predicts susceptibility, progression or even mortality of colon cancer.

Lastly, in **Paper IV** the investigated Cad individuals carried no variant in the coding region of *B4GALNT2*, which could explain the phenotype. The genetic basis may very well be found in regulatory motifs in non-coding parts of the gene, but this has not yet been investigated. The enhanced affinity of anti-Sd^{a} to the RBCs from these individuals could in theory be explained by enhanced expression of the gene, increasing the amount of GT and thereby more efficient production of the antigen. However, there are suggestions that there are not only quantitative differences but also qualitative ones as well between Sd^{a} and Cad-related antigens. E.g., this is implied by the polyagglutination effect that can be observed when Cad1 RBCs are mixed with a majority of human plasmas. A weakness of the Cad investigation carried out as part of this study, is that no effort was made to ensure if samples belong to the Cad1, 2 or 3 type. Structural differences between the anti-Sd^{a} reactive structures on Cad RBCs and regular Sd^{a}-positive RBCs would be interesting to evaluate further in the future.

General discussion

Safe blood transfusions from donor to patient must be preceded by reliant blood group matching. If irregular antibodies are detected in a patient's plasma, their specificity must be identified and with that extended blood group typing is

performed with the purpose to be able to give antigen-matched blood. Sometimes serological techniques have limitations and then genetic approaches to predict the phenotype make a great complement. However, knowledge about the blood group locus and SNPs causing the antigen variant in question are required for being able to apply such methods. The same applies if DNA-based blood group typing is needed to identify blood donors of a certain, perhaps rare type and serological reagents are not available or good enough. Genetic markers defined to cause a blood group phenotype, according to scientific investigations like the ones presented here, are then targeted in the analysis. This strategy can be applied based on the findings made as part of this Ph.D. thesis. For instance, the markers discovered or characterized here may be used for prediction of the P_1 vs. P_2 phenotypes as well as Sd(a+) vs. Sd(a–). The techniques utilized for DNA-based blood group typing are constantly evolving and targeted NGS is an approach that can now be applied also to *A4GALT* and *B4GALNT2*. In addition to the polymorphisms discussed here, NGS will also be able to detect variations in these genes, previously not known.[80] However, if restrictions are made in these assays one may miss out on certain blood groups. If NGS is restricted to only the exome, then the regulatory sites might be missed. In such a scenario the P_1/P_2-discriminating SNP would be missed.

Deciphering the antigenic structures may expand our understanding about the strength of expression for a certain antigen, currently often measured as reactivity with a certain serological reagent, which may differ from one individual to the next. Both anti-P1 and anti-Sd^a are generally considered to constitute low risks when it comes to causing HTR, however there are individual cases where the contrary applies.[126,180,181] It may be that the recipient have a strong immunological response. It may also be a donor whose RBCs express the antigen in question at unexpectedly high levels. One such example is Cad, where both qualitative and quantitative aspects of Sd^a may be at play. Thus, it is of importance to avoid Cad RBCs transfusion to Sd(a–) recipients with anti-Sd^a.

Furthermore, investigations of blood group antigens not only benefit the sector of transfusion medicine. Both the P1 and Sd^a are implicated in a number of other medical conditions and play leading or supporting roles in a number of other research fields. One of the most obvious concerns pathogen interaction, which has been described for both antigens. Additionally, the Sd^a antigen has been associated with cancer inhibition.[201] It is of course important to find out more about the detailed mechanisms to understand why that is. In the case of *B4GALNT2* the surrounding genetic region seems to be of interest with the downstream co-expressed lncRNA RP11-708H21.4.

Future perspective

Through the identification of enhancing TFs binding selectively to the P^1 allele in **Paper I** and a study from another group,[212] a new understanding of an old enigma was reached in the P1PK system. Even so, questions remain, such as if and how the involved TFs may interact with each other or at what stage of the hematopoiesis each of them affects P1 expression. As the P_1/P_2 phenotypes differ based on the mRNA levels, there are still antigen dosage variation within the genotype groups which is of interest and has not fully been explained yet. Individuals homozygous for the P^1 allele may be strongly P1-positive but they may also be at the same level of P1 expression as an individual with the P^1/P^2 genotype. In the same way, some in the heterozygous group may have very low levels of P1 on their RBCs. This has been seen by flow cytometry staining the whole P1 surface expression[215] and it is also well displayed in immunoblotting of P1, see **Paper II,** Fig. 1. The intra-genotype differences have been evaluated based on molecular differences as well as altered cell membrane composition,[215] but no conclusions have been drawn. It can be predicted that other regulatory mechanisms, still to be discovered, are at play.

Paper II contributed with evidence of P1 antigen presence on glycoproteins and N-glycans. However, the carrier proteins were never established. LC-MS/MS has proven to be a valuable and useful technique to determine both glycan structures and their protein carriers. The method requires a certain quantity of material to allow detection of the structures of interest with reliability. The proportion of P1 antigen displayed on glycoproteins is likely low and therefore a purification and enrichment step to increase its concentration is probably needed. A suggestion could be to apply an immunoprecipitation step prior to analysis, however the available antibodies directed to P1 are usually of IgM type whilst IgG would serve this application better. Another way could be to overexpress the gene in a cell line and run LC-MS/MS with the aim to optimize a protocol for detecting P1-positive glycopeptides.

Paper III and **IV** identified a relatively common genetic variant, which causes Sd^a negativity. However, there are still individuals lacking the antigen where no cause has been established. A step further would be to evaluate the sequences surrounding the promoter and coding regions of *B4GALNT2* in these individuals but unknown enhancers and suppressor motifs may need to be found first. Furthermore, the consequences of the splice-site mutation rs72835417 need to be experimentally evaluated to finally distinguish the potential effect on the *B4GALNT2*-encoded GT and in turn its effect on Sd^a synthesis.

Like the remaining unexplained Sd(a–) individuals, a genetic explanation of the Cad phenotype, addressed in **Paper IV,** may lie in the surrounding regulatory sequences of *B4GALNT2*. Thus, a project to reveal such motifs may benefit both aspects. A further step would be whole genome sequencing to address if the genetic variation

lies elsewhere in the genome but with few samples available, this may prove challenging.

Another question is the antigenic structure and how or if it differs between the Sd(a+) and Cad phenotypes on RBCs. We have made initial attempts at analyzing glycoproteins from RBC membranes of the different phenotypes by LC-MS/MS in collaboration with our co-authors from **Paper IV**. However, like previous studies on the subject have experienced, the amount of Sd^a is low in RBCs in general and a special problem may be that only a portion of RBCs carry the antigen in Sd(a+), as indicated by the unusual hemagglutination pattern with anti-Sd^a (Fig. 5). We cannot exclude that the amount of Sd^a in the analysed samples is too low to be detected instead of declaring them as negative. If the latter were true, the reactive structures in Sd(a+) RBCs most probably are found on GSLs. To be able to analyse readily the Sd^a structures in the Sd(a+) RBC phenotype a purification step before the glycoproteomics analysis would be appropriate and the addition of analysis for GSLs, e.g. by immune-overlays of thin layer chromatograms (TLC) and mass spectrometry, to be run in parallel.

Finally, there are still carbohydrate antigens of unknown genetic origin, that also are structurally related to the antigens assessed in this thesis. One of them is the LKE antigen, which, like the Sd^a antigen, has been a conundrum since the 1960s. Antibodies against the LKE antigen were found in the plasma in an individual with Hodgkin's disease.[224] It is a high-frequency antigen as approximately 98% carry it, however antibodies have only been detected in a scarce number of individuals. The antigen is found in the globo GSL biosynthesis series (Fig. 3) and LKE-negative RBCs have a high exposure of the P^k antigen.[225] It has been established that the *B3GALT5*-expressed GT elongates the P antigen to the Gb5 structure,[226] which may then be further sialylated, by the GT coded by *ST3GAL2*, to synthesize the LKE epitope.[227] There have been studies where *B3GALT5* has been suggested to have SNPs associated with LKE negativity.[228] However, no final conclusion has been reached to resolve this matter.

Populärvetenskaplig sammanfattning

Under 1800-talet genomfördes de första lyckade blodtransfusionerna mellan människor. Det var en osäker behandling och läkaren James Blundell, som var en av pionjärerna, förespråkade behandlingen endast om inga andra möjligheter fanns. År 1901 publicerade Karl Landsteiner en studie som skulle ändra det transfusionsmedicinska fältet för alltid. Han hade noterat att de röda blodkropparna från en människa klumpade ihop sig (agglutinerades) när de blandades med plasma från vissa personer men inte från andra. När detta testades i olika kombinationer påvisades ett mönster. Detta var upptäckten av det första blodgruppssystemet, ABO. Idag finns 39 olika blodgruppssystem erkända av International Society of Blood Transfusion (ISBT), som är den organisation som samlar och förmedlar kunskap i ämnet, över hela världen.

Blodgrupper är antigen, små ärftliga variationer på den röda blodkroppens yta. Antigener kan utgöras av socker- eller proteinstrukturer som skiljer sig lite åt människor emellan. När en människa transfunderas kan immunförsvaret hos mottagaren reagera på det som är främmande. Immunförsvaret kan då börja producera antikroppar som binder till antigenet. När antikropparna binder in finns det risk att en immunreaktion startar. Transfusion av blod som inte är matchat mellan blodgivare och patient kan, om det finns antikroppar mot en eller flera blodgrupper, orsaka en s.k. hemolytisk transfusionsreaktion (HTR), varvid röda blodkroppar går sönder. För detta tillstånd finns det ingen effektiv behandling utan man måste arbeta förebyggande vilket innebär att säkerställa att blodet från givaren så bra som möjligt passar patientens blod. På grund av blodgruppsvariation finns det även risk för immunreaktioner vid graviditet. Det innebär att den gravida kvinnan kan bilda antikroppar mot antigen som finns på fostrets röda blodkroppar, vilket då kan orsaka hemolytisk sjukdom hos fostret eller den nyfödda. Historiskt har antikroppar mot RhD-antigenet betraktats som de farligaste men i och med en behandling med förebyggande antikroppar, s.k. Rh-profylax, som ges till kvinnor som är RhD-negativa och bär på RhD-positiva foster, så är risken att bilda antikroppar mot fostrets RhD nästintill eliminerad i västvärlden. Detta gäller tyvärr inte för hela världen och fortfarande finns det många gravida som inte får denna hjälp.

Blodgruppering kan göras serologiskt med reagens av antikroppar som detekterar antigen på den röda blodkroppens yta. Under senare decennier har även genetisk typning börjat användas, efter det att man har kunnat kartlägga var och vad i den mänskliga arvsmassan som gör att vi får olika blodgrupper.

Denna avhandling behandlar blodgruppssystem vars antigener liksom ABO-systemets består av kolhydrater, d.v.s. sockermolekyler. Kolhydrater byggs upp i cellen av enzymer av klassen glykosyltransferaser. Ett glykosyltransferas har specificitet dels för en viss bärarstruktur och dels för en viss kolhydratmolekyl, som den lägger till på bärarstrukturen. På så vis bygger olika glykosyltransferaser upp kortare eller längre kolhydratkedjor i eller på celler. Glykosyltransferaser är proteiner som kodas av generna. DNA bär den genetiska koden, själva arvsmassan. DNA består av specifika sekvenser av molekyler som kallas nukleotider. Tre nukleotider i rad kodar för en viss typ av aminosyra som är det som proteinet i sin tur består av. För att en gen ska uttryckas i en viss cell krävs en mängd faktorer. En del i detta är transkriptionsfaktorer (TFs), ett slags proteiner, som reglerar genernas uttryck genom att binda till specifika, korta sekvenser nukleotider i anslutning till en gen och aktivera (eller ibland hämma) den.

DNA-sekvensen hos en människa kan variera på ett stort antal olika positioner. En gens olika varianter av sekvenser kallas alleler. Variationer i en glykosyltransferas-gen kan påverka produktens (alltså glykosyltransferasets) enzymatiska funktion. Kanske byter den specificitet för vilken kolhydrat den lägger till och därmed byggs en annan struktur. Kanske tappar den sin funktion eller kanske genen i sig inte längre uttrycks, i båda fallen leder det till att antigenet helt kommer saknas på cellytan, d.v.s. individen blir negativ för just den blodgruppen.

P1PK-blodgruppssystemet

P1PK-blodgruppssystemet omfattar tre olika antigener P^k, P1 och NOR. Det sistnämnda är mycket ovanligt. Mitt arbete har fokuserat på de två vanligaste blodgruppsfenotyperna i detta system, P_1 och P_2. Den första innebär att individen bär både P^k och P1 antigen på sina röda blodkroppar, medan P_2-cellerna saknar P1 antigen men har P^k på cellytan. P_1-fenotypen är mycket vanlig hos människor av afrikanskt ursprung (över 90%) men även bland den europeiska befolkningen (ca. 80%), medan P_2 dominerar i Asien (ca. 80%). Anledningen till dessa skillnader är inte känd men intressant i sammanhanget är att både P^k, P1 och besläktade antigener har visat sig vara receptorer för sjukdomsalstrande mikroorganismer som t.ex. bakterier och virus. Blodgruppsvariationen skulle alltså kunna förklara att människor är olika känsliga för vissa infektioner. Nollfenotypen i detta system, d.v.s. om man saknar samtliga antigen i systemet, kallas p, och är en mycket sällsynt fenotyp. P1 och P^k byggs båda upp av ett glykosyltransferas som kodas av genen *A4GALT*. I denna gen har man identifierat flera variationer som leder till nollfenotyp, men man har inte kunnat förklara mekanismen som skiljer fenotyperna P_1 och P_2. Associerat till fenotyperna finns tre stycken s.k. singelnukleotid-polymorfismer (SNPs). SNPs är positioner i DNA-sekvensen som kan variera hos olika individer. I detta fall förekommer alltså en viss sorts nukleotider på specifika positioner i DNA hos P_1-individer men är utbytta till andra hos P_2-individer. Vad

som är intressant med dessa är att de återfinns i DNA-regionen för *A4GALT* men inte i den del som kodar för enzymet. Det har noterats att individer med P_1-fenotypen har högre uttryck av *A4GALT* jämfört med P_2-individer. Detta tyder på att det skulle finnas mer glykosyltransferas hos P_1-individer. Sedan tidigare är det känt att människor som har låga mängder av en specifik transkriptionsfaktor, KLF1, också har ett svagt P1-uttryck, varför utgångspunkten för den första studien var att undersöka om KLF1 kunde spela en roll i varför människor blir antingen P_1 eller P_2.

I **Artikel I** undersöktes DNA-sekvensen runt de SNPs som associerats till de båda vanliga fenotyperna. Över en av dessa SNPs, kallad rs5751348, stämmer sekvensen, som finns hos P_1-individer, överens med flera transkriptionsfaktorers inbindningsmotiv. Dessa är KLF1, RUNX1 och EGR1. I experiment kunde vi detektera att något band in specifikt till P^1-allelen men inte till P^2-allelen. Här misstänktes så klart KLF1 som det tidigare funnits indikationer på, men det kunde inte bekräftas experimentellt. Istället identifierades RUNX1 som den inbindande faktorn. I ytterligare undersökningar där uttrycket av RUNX1 tystades i celler som odlades i labbet, så kunde det även noteras hur *A4GALT*-uttrycket gick ned. Detta kunde inte noteras när KLF1 tystades. Därmed kunde det konstateras att mekanismen bakom P1-uttrycket ligger i reglering av genen. När RUNX1 kan binda in ökar uttrycket av *A4GALT* som förekommer i P_1-fenotypen.

Artikel II behandlar samma blodgruppssystem som föregående artikel. Men här är fokus i stället antigen-strukturen P1. I litteraturen är det en allmän uppfattning att kolhydraten P1 endast byggs på lipider, s.k. glykolipider. De få studier som gjorts på ämnet är emellertid inkonsekventa i sina slutsatser. Målet med denna studie var att försöka detektera P1 på glykoproteiner. I röda blodkroppar från P_1-individer, men inte P_2 eller p, kunde P1 detekteras på proteiner. Vi kunde också påvisa att antigenet bars av N-glykaner som är en viss typ av bassekvens för kolhydrater på proteiner. Kolhydraters bärarstrukturer i cellmembranet kan göra skillnad, då lipiderna är små och placerar kolhydraterna nära cellytan. Proteiner är däremot ofta stora strukturer som gör att positionen för kolhydraterna kommer högre upp och därmed längre ifrån cellytan. Sammanfattningsvis anser vi att denna studie en gång för alla visar att P1 inte bara bärs av glykolipider utan också av glykoproteiner.

SID-blodgruppssystemet

Det andra blodgruppssystemet denna avhandling berör baseras på Sd^a-antigenet. Detta antigen saknas på de röda blodkropparna hos ca. 10% av befolkningen i Europa, men egentligen är det bara 2-4% som är helt negativa för det i all vävnad. Dessutom verkar det endast vara en del av de röda blodkropparna som bär antigenet hos individerna med Sd^a-positiv fenotyp, Sd(a+). Det finns även en fenotyp som kallas Cad, Sd(a++) eller super-Sid. Fenotypen är mycket ovanlig och innebär att de röda blodkropparna hos dessa individer reagerar mycket starkt med antikroppar mot Sd^a. Sd^a är inte bara en intressant struktur på blodkropparna utan finns också i

tjocktarmen, där har man noterat att den försvinner i elakartad tumörvävnad. Sd^a-antigenet har varit känt sedan 1960-talet och genen *B4GALNT2,* som man tror kodar för glykosyltransferaset som bygger antigenet, identifierades redan 2003. Det har dock varit okänt om Sd^a-negativitet orsakas av någon variation i *B4GALNT2* eller ej.

I **artikel III** och **IV** behandlas just nollfenotypen och den genetiska orsaken till att Sd^a-antigenet saknas. Genom att kartlägga DNA-sekvensen av den kodande delen av *B4GALNT2*-genen hos nio individer med Sd^a-negativ fenotyp, identifierades en mycket intressant SNP, rs7224888. Den är relativt ovanlig i övrig befolkning men förekom hos sju individer av de nio undersökta. Ytterligare tre SNPs identifierades och en individ konstaterades inte ha några varianter alls. Den SNP (rs7224888) som var starkast associerad med Sd^a-negativ blodgrupp förutsågs ändra i glykosyltransferasets aminosyrasekvens, i en region som troligen är avgörande för proteinets tredimensionella struktur, som i sin tur är viktig för den enzymatiska funktionen. I vidare experiment uttrycktes de olika *B4GALNT2*-allelerna i celler och därefter analyserades Sd^a-uttrycket med olika metoder. Den vanligt förekommande SNP:en, rs7224888, kodade för ett icke funktionellt glykosyltransferas, vilket noterades då Sd^a-antigenet ej kunde detekteras med någon metod på cellerna som utryckte denna allel. Detta innebär att fenotypen orsakas av en variant i *B4GALNT2*, vilket gör att man nu kan prediktera blodgruppen med genomisk analys. Denna kunskap var också avgörande för att ISBT kunde utnämna SID till ett nytt blodgruppssystem med nummer 038. Vidare har de kodande delarna i *B4GALNT2* också sekvenserats hos fem Cad-individer, men inga variationer har hittats. Därmed inte sagt att det inte är denna gen som orsakar fenotypen. Förklaringen kan precis som hos P1-individerna ligga i icke-kodande delar av genen men det krävs mer forskning i framtiden för att klarlägga detta.

Acknowledgements

The work in this thesis has been conducted in collaboration with and supported by many. Without you it would not have been possible.

- Martin L Olsson, my main supervisor. Thank you for giving me this opportunity. I have learnt a lot during this journey, it has been inspiring and challenging at the same time.
- Åsa Hellberg, my co-supervisor. Thank you for guidance in the experimental work, interesting discussions and always there to help.
- Julia S Westman, co-supervisor in the studies that preceded this. You encouraged me to strive further. Thank you for being a great friend and colleague.
- Annika Hult, Jill R Storry and Karina Vidovic. Your excellence in the profession of research are truly inspiring. I take great pride in having you as my colleagues.
- Fellow Ph.D. students, post-docs and students of the group, past and present. Thank you for making the working place creative and fun, but also for support when pressure was on.
- Co-authors. I've had the honor of working in projects involving highly skilled and professional people, you brought the projects to new levels.
- Colleagues at BMC C14. Thank you for a friendly working environment and generosity.
- The people working at the clinical laboratories of transfusion medicine, who implement the knowledge research comes up with. Speciellt tack till alla er på Klinisk Immunologi och Transfusionsmedicin vid Skånes Universitetssjukhus i Lund, vars lokaler, reagens och råd jag tagit del av. Stort tack också till personalen på Klinisk Immunologi och Transfusionsmedicin vid Sahlgrenska universitetssjukhuset, Göteborg, som var mina kollegor 2011–2014 och introducerade mig i detta ämne.
- Vänner. Tack för all glädje och för att ni finns.
- Familjen. Mamma och pappa, tack för allt ni lärt mig, för stöd i alla lägen och den uppmuntran ni ger. Jennifer och Jeppe, tack för att ert hem alltid står öppet, och tack till Eije och Edith för all glädje. GG, tack för att du är en otrolig förebild och vän.
- Jonas. Tack för kärlek. Tack för stöd. Tack för allt!

References

1. Blundell J. Some account of a case of obstinate vomiting, in which an attempt was made to prolong life by the injection of blood into the veins. Med Chir Trans 1819;10:296-311.
2. Blundell J. Experiments on the transfusion of blood by the syringe. Med Chir Trans 1818;9:56-92.
3. Waller C. Case of uterine hemorrhage, in which the operation of transfusion was successfully performed. Lond Med Phys J 1825;54:273-7.
4. Learoyd P. The history of blood transfusion prior to the 20th century-part 1. Transfus Med 2012;22:308-14.
5. Baskett TF. James Blundell: the first transfusion of human blood. Resuscitation 2002;52:229-33.
6. Landsteiner K. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe (Translation in Camp & Ellis, 1966). Zbl Bakt 1900;27:357-62.
7. Boulton FE. Blood transfusion; additional historical aspects. Part 1. The birth of transfusion immunology. Transfus Med 2013;23:375-81.
8. Landsteiner K. Über Agglutinationserscheinungen normalen Menschlichen Blutes (Translation in Camp & Ellis, 1966). Wien Klin Wochenschr 1901;14:1132-4.
9. Decastello A, Sturli A. Ueber die isoagglutinine im gesunder und kranker menschen (Translation in Camp & Ellis, 1966) Munch Med Wochenschrift 1902;49:1090-5.
10. Ottenberg R. Transfusion and arterial anastamosis. Am J Surg 1908;47:486.
11. Ottenberg R. Studies in Isoagglutination.1 transfusion and the question of intravascular agglutination. J Exp Med 1911;13:425-38.
12. Learoyd P. The history of blood transfusion prior to the 20th century-part 2. Transfus Med 2012;22:372-6.
13. Boulton F, Roberts DJ. Blood transfusion at the time of the First World War--practice and promise at the birth of transfusion medicine. Transfus Med 2014;24:325-34.
14. Alter HJ, Klein HG. The hazards of blood transfusion in historical perspective. Blood 2008;112:2617-26.
15. Auvinen M-K. Blodverksamheten i Sverige 2019 [Internet]: Swedish Blood Alliance (SweBA). Available from: http://www.sweba.se/filedepot?cid=7&fid=1200 [2020-08-14]

16. Karafin MS, Bruhn R, Westlake M*, et al.* Demographic and epidemiologic characterization of transfusion recipients from four US regions: evidence from the REDS-III recipient database. Transfusion 2017;57:2903-13.
17. Borkent-Raven BA, Janssen MP, van der Poel CL, Schaasberg WP, Bonsel GJ, van Hout BA. The PROTON study: profiles of blood product transfusion recipients in the Netherlands. Vox Sang 2010;99:54-64.
18. Tinegate H, Pendry K, Murphy M*, et al.* Where do all the red blood cells (RBCs) go? Results of a survey of RBC use in England and North Wales in 2014. Transfusion 2016;56:139-45.
19. WHO. Blood safety and availability [Internet]: World Health Organization. Available from: https://www.who.int/news-room/fact-sheets/detail/blood-safety-and-availability. [2020-08-14]
20. Haley NR, Hess JR. Whole blood: back to the future. Transfusion 2019;59:3293-4.
21. Huish S, Green L, Curnow E, Wiltshire M, Cardigan R. Effect of storage of plasma in the presence of red blood cells and platelets: re-evaluating the shelf life of whole blood. Transfusion 2019;59:3468-77.
22. Panch SR, Montemayor-Garcia C, Klein HG. Hemolytic Transfusion Reactions. N Engl J Med 2019;381:150-62.
23. Storch EK, Custer BS, Jacobs MR, Menitove JE, Mintz PD. Review of current transfusion therapy and blood banking practices. Blood Rev 2019;38:100593.
24. Murphy MF, Goodnough LT. The scientific basis for patient blood management. Transfus Clin Biol 2015;22:90-6.
25. Daniels G. Human Blood Groups. 3rd ed. Oxford, UK: Blackwell Scientific, 2013.
26. Anstee DJ. The functional importance of blood group-active molecules in human red blood cells. Vox Sang 2011;100:140-9.
27. Harayama T, Riezman H. Understanding the diversity of membrane lipid composition. Nat Rev Mol Cell Biol 2018;19:281-96.
28. Varki A. Essentials of glycobiology. 3rd ed. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press, 2017.
29. Zimring JC, Hudson KE. Cellular immune responses in red blood cell alloimmunization. Hematology Am Soc Hematol Educ Program 2016;2016:452-6.
30. Reid ME, Lomas-Francis C, Olsson ML. The Blood Group Antigen FactsBook. 3rd ed. London, UK: Academic Press, 2012.
31. Clausen H, Hakomori S. ABH and related histo-blood group antigens; immunochemical differences in carrier isotypes and their distribution. Vox Sang 1989;56:1-20.
32. Springer GF, Horton RE. Blood group isoantibody stimulation in man by feeding blood group-active bacteria. J Clin Invest 1969;48:1280-91.
33. Stowell SR, Winkler AM, Maier CL*, et al.* Initiation and regulation of complement during hemolytic transfusion reactions. Clin Dev Immunol 2012:307093.
34. Zimring JC, Spitalnik SL. Pathobiology of transfusion reactions. Annu Rev Pathol 2015;10:83-110.

35. Strobel E. Hemolytic Transfusion Reactions. Transfus Med Hemother 2008;35:346-53.
36. ISBT. Red Cell Immunogenetics and Blood Group Terminology Working Party [Internet]: International Society of Blood Transfusion. Available from: https://www.isbtweb.org/working-parties/red-cell-immunogenetics-and-blood-group-terminology. [2020-08-12]
37. Azouzi S, Mikdar M, Hermand P*, et al.* Lack of the multidrug transporter MRP4/ABCC4 defines the PEL-negative blood group and impairs platelet aggregation. Blood 2020;135:441-8.
38. 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;11:3569.
39. Fiorellino J, Elahie AL, Warkentin TE. Acute haemolysis, DIC and renal failure after transfusion of uncross-matched blood during trauma resuscitation: illustrative case and literature review. Transfus Med 2018;28:319-25.
40. Semple JW, Rebetz J, Kapur R. Transfusion-associated circulatory overload and transfusion-related acute lung injury. Blood 2019;133:1840-53.
41. Roubinian N. TACO and TRALI: biology, risk factors, and prevention strategies. Hematology Am Soc Hematol Educ Program 2018:585-94.
42. Kopolovic I, Ostro J, Tsubota H*, et al.* A systematic review of transfusion-associated graft-versus-host disease. Blood 2015;126:406-14.
43. Busch MP, Bloch EM, Kleinman S. Prevention of transfusion-transmitted infections. Blood 2019;133:1854-64.
44. Perkins HA, Busch MP. Transfusion-associated infections: 50 years of relentless challenges and remarkable progress. Transfusion 2010;50:2080-99.
45. WHO. Screening donated blood for transfusion-transmissible infections: Recommendations Publication. Geneva: World Health Organization, 2009.
46. de Haas M, Thurik FF, Koelewijn JM, van der Schoot CE. Haemolytic disease of the fetus and newborn. Vox Sang 2015;109:99-113.
47. Hendrickson JE, Delaney M. Hemolytic disease of the fetus and newborn: Modern practice and future investigations. Transfus Med Rev 2016;30:159-64.
48. Fasano RM. Hemolytic disease of the fetus and newborn in the molecular era. Semin Fetal Neonatal Med 2016;21:28-34.
49. WIRhe. Worldwide Initiative for Rh Disease Eradication (WIRhE) [Internet]: Worldwide Initiative for Rh Disease Eradication 2020. Available from: https://wirhe.org/. [2020-08-17]
50. Lindström K, Dem Borne AE, Breimer ME*, et al.* Glycosphingolipid expression in spontaneously aborted fetuses and placenta from blood group p women. Evidence for placenta being the primary target for anti-Tj^{a}-antibodies. Glycoconj J 1992;9:325-9.
51. Ricci Hagman J, Hult AK, Westman JS*, et al.* Multiple miscarriages in two sisters of Thai origin with the rare P(k) phenotype caused by a novel nonsense mutation at the B3GALNT1 locus. Transfus Med 2019;29:202-8.

52. Lopez M, Cartron J, Cartron JP*, et al.* Cytotoxicity of anti-PP1Pk antibodies and possible relationship with early abortions of p mothers. Clin Immunol Immunopathol 1983;28:296-303.
53. Montgomery RA, Cozzi E, West LJ, Warren DS. Humoral immunity and antibody-mediated rejection in solid organ transplantation. Semin Immunol 2011;23:224-34.
54. Dieudé M, West LJ, Muruve DA*, et al.* New answers to old conundrums: What antibodies, exosomes and inflammasomes bring to the conversation. Canadian national transplant research program international summit report. Transplantation 2018;102:209-14.
55. McCaughan JA, Tinckam KJ. Donor specific HLA antibodies & allograft injury: mechanisms, methods of detection, manifestations and management. Transpl Int 2018;31:1059-70.
56. Böhmig GA, Farkas AM, Eskandary F, Wekerle T. Strategies to overcome the ABO barrier in kidney transplantation. Nat Rev Nephrol 2015;11:732-47.
57. Zhao C, Cooper DKC, Dai Y, Hara H, Cai Z, Mou L. The Sda and Cad glycan antigens and their glycosyltransferase, beta1,4GalNAcT-II, in xenotransplantation. Xenotransplantation 2018;25:e12386.
58. Estrada JL, Martens G, Li P*, et al.* Evaluation of human and non-human primate antibody binding to pig cells lacking GGTA1/CMAH/β4GalNT2 genes. Xenotransplantation 2015;22:194-202.
59. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer 2015;15:540-55.
60. Ewald DR, Sumner SC. Blood type biochemistry and human disease. Wiley Interdiscip Rev Syst Biol Med 2016;8:517-35.
61. Cooling L. Blood Groups in Infection and Host Susceptibility. Clin Microbiol Rev 2015;28:801-70.
62. Jongruamklang P. Studies of Thai blood group and platelet polymorphism: implications for malaria susceptibility: Faculty of Medicine, Lund University 2020.
63. Anstee DJ. The relationship between blood groups and disease. Blood 2010;115:4635-43.
64. Cserti CM, Dzik WH. The ABO blood group system and plasmodium falciparum malaria. Blood 2007;110:2250-8.
65. Wolofsky KT, Ayi K, Branch DR*, et al.* ABO blood groups influence macrophage-mediated phagocytosis of Plasmodium falciparum-infected erythrocytes. PLoS Pathog 2012;8:e1002942.
66. Goel S, Palmkvist M, Moll K*, et al.* RIFINs are adhesins implicated in severe Plasmodium falciparum malaria. Nat Med 2015;21:314-7.
67. Choi JM, Hutson AM, Estes MK, Prasad BV. Atomic resolution structural characterization of recognition of histo-blood group antigens by Norwalk virus. Proc Natl Acad Sci U S A 2008;105:9175-80.
68. Wu Y, Feng Z, Li P, Yu Q. Relationship between ABO blood group distribution and clinical characteristics in patients with COVID-19. Clin Chim Acta 2020;509:220-3.

69. Göker H, Aladağ Karakulak E, Demiroğlu H*, et al.* The effects of blood group types on the risk of COVID-19 infection and its clinical outcome. Turk J Med Sci 2020;50:679-83.
70. Li J, Wang X, Chen J, Cai Y, Deng A, Yang M. Association between ABO blood groups and risk of SARS-CoV-2 pneumonia. Br J Haematol 2020;190:24-7.
71. Gérard C, Maggipinto G, Minon JM. COVID-19 and ABO blood group: another viewpoint. Br J Haematol 2020;190:e93-e4.
72. Arendrup M, Hansen JE, Clausen H, Nielsen C, Mathiesen LR, Nielsen JO. Antibody to histo-blood group A antigen neutralizes HIV produced by lymphocytes from blood group A donors but not from blood group B or O donors. AIDS 1991;5:441-4.
73. Preece AF, Strahan KM, Devitt J, Yamamoto F, Gustafsson K. Expression of ABO or related antigenic carbohydrates on viral envelopes leads to neutralization in the presence of serum containing specific natural antibodies and complement. Blood 2002;99:2477-82.
74. Coombs RRA, Mourant AE, Race RR. A new test for detection of weak and incomplete Rh agglutinins. Bri J Exp Pathol 1945;26:821-30.
75. Hamilton JR. Saline-indirect antiglobulin test. Immunohematology 2019;35:156-8.
76. Horn KD. The classification, recognition and significance of polyagglutination in transfusion medicine. Blood Rev 1999;13:36-44.
77. Beck ML. Red blood cell polyagglutination: clinical aspects. Semin Hematol 2000;37:186-96.
78. Quill E. Blood-matching goes genetic. Science 2008;319:1478-9.
79. Chou ST, Evans P, Vege S*, et al.* RH genotype matching for transfusion support in sickle cell disease. Blood 2018;132:1198-207.
80. Westhoff CM. Blood group genotyping. Blood 2019;133:1814-20.
81. 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;1:240-9.
82. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol 2016;14:e1002533.
83. Dzierzak E, Bigas A. Blood Development: Hematopoietic Stem Cell Dependence and Independence. Cell Stem Cell 2018;22:639-51.
84. Nandakumar SK, Ulirsch JC, Sankaran VG. Advances in understanding erythropoiesis: evolving perspectives. Br J Haematol 2016;173:206-18.
85. Zivot A, Lipton JM, Narla A, Blanc L. Erythropoiesis: insights into pathophysiology and treatments in 2017. Mol Med 2018;24:11.
86. Vatikioti A, Karkoulia E, Ioannou M, Strouboulis J. Translational regulation and deregulation in erythropoiesis. Exp Hematol 2019;75:11-20.
87. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Biol 2018;19:621-37.
88. Ma WB, Wang XH, Li CY*, et al.* GPS2 promotes erythroid differentiation by control of the stability of EKLF protein. Blood 2020;135:2302-15.

89. Huang LJ, Bieker JJ. A master erythroid regulator gets its own GPS. Blood 2020;135:2209-10.
90. Alvarez-Dominguez JR, Hu W, Gromatzky AA, Lodish HF. Long noncoding RNAs during normal and malignant hematopoiesis. Int J Hematol 2014;99:531-41.
91. Cooper GM, Hausman RE. The cell : a molecular approach. 7th ed: Sinauer Associates, 2016.
92. Varki A, Cummings RD, Aebi M*, et al.* Symbol Nomenclature for Graphical Representations of Glycans. Glycobiology 2015;25:1323-4.
93. Varki A. Biological roles of glycans. Glycobiology 2017;27:3-49.
94. Welch LG, Munro S. A tale of short tails, through thick and thin: investigating the sorting mechanisms of Golgi enzymes. FEBS Lett 2019;593:2452-65.
95. Albesa-Jove D, Giganti D, Jackson M, Alzari PM, Guerin ME. Structure-function relationships of membrane-associated GT-B glycosyltransferases. Glycobiology 2014;24:108-24.
96. Breton C, Fournel-Gigleux S, Palcic MM. Recent structures, evolution and mechanisms of glycosyltransferases. Curr Opin Struct Biol 2012;22:540-9.
97. Harrus D, Kellokumpu S, Glumoff T. Crystal structures of eukaryote glycosyltransferases reveal biologically relevant enzyme homooligomers. Cell Mol Life Sci 2018;75:833-48.
98. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 2014;42:D490-5.
99. Blanchard D, Piller F, Gillard B, Marcus D, Cartron J-P. Identification of a novel ganglioside on erythrocytes with blood group Cad specificity*. J Biol Chem 1985;260:7813-6.
100. Halim A, Brinkmalm G, Ruetschi U*, et al.* Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein/amyloid beta-peptides in human cerebrospinal fluid. Proc Natl Acad Sci U S A 2011;108:11848-53.
101. Tournamille C, Filipe A, Wasniowska K*, et al.* Structure-function analysis of the extracellular domains of the Duffy antigen/receptor for chemokines: characterization of antibody and chemokine binding sites. Br J Haematol 2003;122:1014-23.
102. Fredriksson SA, Podbielska M, Nilsson B, Krotkiewska B, Lisowska E, Krotkiewski H. ABH blood group antigens in N-glycan of human glycophorin A. Arch Biochem Biophys 2010;498:127-35.
103. Yamamoto F, Clausen H, White T, Marken J, Hakomori S. Molecular genetic basis of the histo-blood group ABO system. Nature 1990;345:229-33.
104. Yamamoto F, Marken J, Tsuji T, White T, Clausen H, Hakomori S. Cloning and characterization of DNA complementary to human UDP-GalNAc:Fuc-alpha1-2Gal alpha1-3GalNAc transferase (histo-blood group A transferase) mRNA. J Biol Chem 1990;265:1146-51.
105. Seto NO, Compston CA, Evans SV, Bundle DR, Narang SA, Palcic MM. Donor substrate specificity of recombinant human blood group A, B and hybrid A/B glycosyltransferases expressed in Escherichia coli. Eur J Biochem 1999;259:770-5.

106. Koscielak J. ABH blood group active glycoconjugates from human red cells. Transfus Med 2001;11:267-79.
107. 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;121:1459-68.
108. Kelly RJ, Ernst LK, Larsen RD, Bryant JG, Robinson JS, Lowe JB. Molecular basis for H blood group deficiency in Bombay (Oh) and para- Bombay individuals. Proc Natl Acad Sci U S A 1994;91:5843-7.
109. Kukowska-Latallo JF, Larsen RD, Nair RP, Lowe JB. A cloned human cDNA determines expression of a mouse stage-specific embryonic antigen and the Lewis blood group α(1,3/1,4)fucosyltransferase. Genes Dev 1990;4:1288-303.
110. Hellberg Å, Poole J, Olsson ML. Molecular basis of the globoside-deficient P^k blood group phenotype. Identification of four inactivating mutations in the UDP-*N* -acetylgalactosamine:globotriaosylceramide 3-β-*N*-acetylgalactosaminyltransferase gene. J Biol Chem 2002;277:29455-9.
111. Westman JS, Benktander J, Storry JR*, et al.* Identification of the molecular and genetic basis of PX2, a glycosphingolipid blood group antigen lacking on globoside-deficient erythrocytes. J Biol Chem 2015;290:18505-18.
112. Yu LC, Twu YC, Chang CY, Lin M. Molecular basis of the adult i phenotype and the gene responsible for the expression of the human blood group I antigen. Blood 2001;98:3840-5.
113. Hellberg Å, Westman JS, Thuresson B, Olsson ML. P1PK: the blood group system that changed its name and expanded. Immunohematology 2013;29:25-33.
114. Iwamura K, Furukawa K, Uchikawa M*, et al.* The blood group P1 synthase gene is identical to the Gb3/CD77 synthase gene. A clue to the solution of the P1/P2/p puzzle. J Biol Chem 2003;278:44429-38.
115. Matson GA, Swanson J, Noades J, Sanger R, Race RR. A "new" antigen and antibody belonging to the P blood group system. Am J Hum Genet 1959;11:26-34.
116. Ikin EW, Kay HEM, Playfair JHL, Mourant AE. P1 antigen in the human fetus. Nature 1961;192 883-.
117. Lai YJ, Wu WY, Yang CM*, et al.* A systematic study of single-nucleotide polymorphisms in the *A4GALT* gene suggests a molecular genetic basis for the P_1/P_2 blood groups. Transfusion 2014;54:3222-31.
118. Christian SG, Eze EM, Elekima I. Blood group antigen dominant among Indigenes of Ogoni ethnicity in Rivers state of Nigeria. Asian Hematol Res J 2020;3:30-5.
119. Thuresson B, Westman JS, Olsson ML. Identification of a novel *A4GALT* exon reveals the genetic basis of the P_1/P_2 histo-blood groups. Blood 2011;117:678-87.
120. Keramati MR, Shakibaei H, Kheiyyami MI*, et al.* Blood group antigens frequencies in the northeast of Iran. Transfus Apher Sci 2011;45:133-6.
121. Kahar MA, Patel RD. Phenotype frequencies of blood group systems (Rh, Kell, Kidd, Duffy, MNS, P, Lewis, and Lutheran) in blood donors of south Gujarat, India. Asian J Transfus Sci 2014;8:51-5.

122. Mohan TC, Koo WH, Ng HW. A study of the P blood group system in the Singaporean population. Singapore Med J 1989;30:372-5.
123. Yu Y, Ma C, Sun X*, et al.* Frequencies of red blood cell major blood group antigens and phenotypes in the Chinese Han population from Mainland China. Int J Immunogenet 2016;43:226-35.
124. Kawai M, Obara K, Onodera T*, et al.* Mutations of the KLF1 gene detected in Japanese with the In(Lu) phenotype. Transfusion 2017;57:1072-7.
125. Heiken A. Observations on the blood group receptor P1 and its development in children. Hereditas 1966;56:83-98.
126. Arndt PA, Garratty G, Marfoe RA, Zeger GD. An acute hemolytic transfusion reaction caused by an anti-P1 that reacted at 37 degrees C. Transfusion 1998;38:373-7.
127. Bezirgiannidou Z, Christoforidou A, Kontekaki E*, et al.* Hyperhemolytic syndrome complicating a delayed hemolytic transfusion reaction due to anti-P1 alloimmunization, in a pregnant woman with HbO-Arab/beta-Thalassemia. Mediterr J Hematol Infect Dis 2016;8:e2016053.
128. Smith D, Aye T, Er LS, Nester T, Delaney M. Acute hemolytic transfusion reaction due to anti-P1: A case report and review of institutional experience. Transfus Med Hemother 2019;46:380-3.
129. Sanger R. An association between the P and Jay systems of blood groups. Nature. 1955;176:1163-4.
130. Harris PA, Roman GK, Moulds JJ, Bird GW, Shah NG. An inherited RBC characteristic, NOR, resulting in erythrocyte polyagglutination. Vox Sang 1982;42:134-40.
131. Kusnierz-Alejska G, Duk M, Storry JR*, et al.* NOR polyagglutination and St^a glycophorin in one family: relation of NOR polyagglutination to terminal α-galactose residues and abnormal glycolipids. Transfusion 1999;39:32-8.
132. Ingolfsdottir R, Hult A, Olsson ML. Flow cytometric assessment of carbohydrate histo-blood group expression on leucocyte subpopulations: University of Iceland, 2009.
133. Fellous M, Gerbal A, Tessier C, Frezal J, Dausset J, Salmon C. Studies on the biosynthetic pathway of human P erythrocyte antigens using somatic cells in culture. Vox Sanguinis 1974;26:518-36.
134. Oosterwijk E, Kalisiak A, Wakka JC, Scheinberg DA, Old LJ. Monoclonal antibodies against Galα1-4Galβ1-4Glc (P^k, CD77) produced with a synthetic glycoconjugate as immunogen: reactivity with carbohydrates, with fresh frozen human tissues and hematopoietic tumors. Int J Cancer 1991;48:848-54.
135. Gurner BW, Coombs RR. Examination of human leucocytes for the ABO, MN, Rh, Tja, Lutheran and Lewis systems of antigens by means of mixed erythrocyte-leucocyte agglutination. Vox Sang 1958;3:13-22.
136. Landsteiner K, Levine P. Further observations on individual differences of human blood. Proc Soc Exp Biol Med 1927;24:941-2.
137. Morgan WTJ, Watkins WM. Blood-group P1 substance - (I) Chemical properties. Bibl Haematol 1964;19:225-9.

138. Watkins WM, Morgan WT. Blood-group P1 substance - (II) Immunological properties. Bibl Haematol 1964;19:230-4.
139. Marcus DM. Isolation of a substance with blood-group P1 activity for human erythrocyte stroma. Transfusion. 1971;11:16-8.
140. Naiki M, Fong J, Ledeen R, Marcus DM. Structure of the human erythrocyte blood group P1 glycosphingolipid. Biochemistry 1975;14:4831-7.
141. Naiki M, Marcus DM. Human erythrocyte P and P^k blood group antigens: identification as glycosphingolipids. Biochem Biophys Res Commun 1974;60:1105-11.
142. Duk M, Reinhold BB, Reinhold VN, Kusnierz-Alejska G, Lisowska E. Structure of a neutral glycosphingolipid recognized by human antibodies in polyagglutinable erythrocytes from the rare NOR phenotype. J Biol Chem 2001;276:40574-82.
143. Kojima Y, Fukumoto S, Furukawa K*, et al.* Molecular cloning of globotriaosylceramide/CD77 synthase, a glycosyltransferase that initiates the synthesis of globo series glycosphingolipids. J Biol Chem 2000;275:15152-6.
144. Steffensen R, Carlier K, Wiels J*, et al.* Cloning and expression of the histo-blood group P^k UDP-galactose: Gal-β1-4Glc-β1-Cer α1,4-galactosyltransferase. Molecular genetic basis of the p phenotype. J Biol Chem 2000;275:16723-9.
145. Furukawa K, Iwamura K, Uchikawa M*, et al.* Molecular basis for the p phenotype. Identification of distinct and multiple mutations in the α1,4-galactosyltransferase gene in Swedish and Japanese individuals. J Biol Chem 2000;275:37752-6.
146. Singleton BK, Burton NM, Green C, Brady RL, Anstee DJ. Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype. Blood 2008;112:2081-8.
147. Crawford MN, Tippett P, Sanger R. Antigens Aua, i and P1 of cells of the dominant type of Lu(a-b-). Vox Sang 1974;26:283-7.
148. Duk M, Singh S, Reinhold VN, Krotkiewski H, Kurowska E, Lisowska E. Structures of unique globoside elongation products present in erythrocytes with a rare NOR phenotype. Glycobiology 2007;17:304-12.
149. Suchanowska A, Kaczmarek R, Duk M*, et al.* A single point mutation in the gene encoding Gb3/CD77 synthase causes a rare inherited polyagglutination syndrome. J Biol Chem 2012;287:38220-30.
150. Lund N, Olsson ML, Ramkumar S*, et al.* The human P(k) histo-blood group antigen provides protection against HIV-1 infection. Blood 2009;113:4980-91.
151. Källenius G, Möllby R, Svenson SB, Winberg J, Hultberg H. Identification of a carbohydrate receptor recognized by uropathogenic Escherichia coli. Infection 1980;8 Suppl 3:288-93.
152. Lomberg H, Edén CS. Influence of P blood group phenotype on susceptibility to urinary tract infection. FEMS Microbiol Immunol 1989;1:363-70.
153. Lindberg AA, Brown JE, Stromberg N, Westling-Ryd M, Schultz JE, Karlsson KA. Identification of the carbohydrate receptor for Shiga toxin produced by Shigella dysenteriae type 1. J Biol Chem 1987;262:1779-85.

154. Bauwens A, Betz J, Meisen I, Kemper B, Karch H, Müthing J. Facing glycosphingolipid-Shiga toxin interaction: dire straits for endothelial cells of the human vasculature. Cell Mol Life Sci 2013;70:425-57.
155. Karpman D, Loos S, Tati R, Arvidsson I. Haemolytic uraemic syndrome. J Intern Med 2017;281:123-48.
156. Luginbuehl V, Meier N, Kovar K, Rohrer J. Intracellular drug delivery: Potential usefulness of engineered Shiga toxin subunit B for targeted cancer therapy. Biotechnol Adv 2018;36:613-23.
157. Spitalnik SL, Spitalnik PF, Dubois C*, et al.* Glycolipid antigen expression in human lung cancer. Cancer Res 1986;46:4751-5.
158. Jacob F, Anugraham M, Pochechueva T*, et al.* The glycosphingolipid P(1) is an ovarian cancer-associated carbohydrate antigen involved in migration. Br J Cancer 2014;111:1634-45.
159. Schiffmann R. Fabry disease. Pharmacol Ther 2009;122:65-77.
160. Lund N, Branch DR, Sakac D*, et al.* Lack of susceptibility of cells from patients with Fabry disease to productive infection with R5 human immunodeficiency virus. AIDS. 2005;19:1543-6.
161. Salamone G, Licari L, Randisi B*, et al.* Uncommon localizations of hydatid cyst. Review of the literature. G Chir 2016;37:180-5.
162. Takahashi N, Khoo KH, Suzuki N, Johnson JR, Lee YC. N-glycan structures from the major glycoproteins of pigeon egg white: predominance of terminal Galalpha(1)Gal. J Biol Chem 2001;276:23230-9.
163. François-Gérard C, Brocteur J, André A. Turtledove: a new source of P1-like material cross-reacting with the human erythrocyte antigen. Vox Sang 1980;39:141-8.
164. Radermecker M, Bruwier M, François C*, et al.* Anti-P1 activity in pigeon breeders' serum. Clin Exp Immunol 1975;22:546-9.
165. Macvie SI, Morton JA, Pickles MM. The reactions and inheritance of a new blood group antigen, Sd^a. Vox Sang 1967;13:485-92.
166. Renton PH. Anti-Sd^a new blood group antibody. Vox Sang 1967;13:493-501.
167. Lo Presti L, Cabuy E, Chiricolo M, Dall'Olio F. Molecular cloning of the human beta1,4 *N*-Acetylgalactosaminyltransferase responsible for the biosynthesis of the Sd^a histo-blood group antigen: The sequence predicts a very long cytoplasmic domain. J Biochem 2003;134:675-82.
168. Montiel MD, Krzewinski-Recchi MA, Delannoy P, Harduin-Lepers A. Molecular cloning, gene organization and expression of the human UDP-GalNAc:Neu5Acalpha2-3Galbeta-R beta1,4-N-acetylgalactosaminyltransferase responsible for the biosynthesis of the blood group Sd^a/Cad antigen: evidence for an unusual extended cytoplasmic domain. Biochem J 2003;373:369-79.
169. Morton JA, Pickles MM, Terry AM. The Sda blood group antigen in tissues and body fluids. Vox Sang 1970;19:472-82.
170. Morton JA, Pickles MM, Vanhegan RI. The Sda antigen in the human kidney and colon. Immunol Invest 1988;17:217-24.

171. Spitalnik S, Cox MT, Spennacchio J, Guenther R, Blumberg N. The serology of Sd^a effects of transfusion and pregnancy. Vox Sang 1982;42:308-12.
172. Hammar L, Månsson S, Rohr T*, et al.* Lewis phenotype of erythrocytes and Leb-active glycolipid in serum of pregnant women. Vox Sang 1981;40:27-33.
173. Gerbal A, Lopez M, Chassaigne M*, et al.* [Cad antigen in the French population]. Rev Fr Transfus Immunohematol 1976;19:415-29.
174. Conte R, Serafini-Cessi F. Comparison between the erythrocyte and urinary Sd^a antigen distribution in a large number of individuals from Emilia-Romagna, a region of northern Italy. Transfus Med 1991;1:47-9.
175. Cazal P, Monis M, Caubel J, Brives J. [Herediatry dominant polyagglutinability: private antigen (Cad) corresponding to a public antibody and a lectin of Dolichos biflorus]. Rev Fr Transfus 1968;11:209-21.
176. Cazal P, Monis M, Bizot M. [The Cad antigens and their relation to A antigens]. Rev Fr Transfus 1971;14:321-34.
177. Sringarm S, Chupungart C, Giles CM. The use of Ulex europaeus and Dolichos biflorus extracts in routine ABO grouping of blood donors in Thailand. Some unexpected findings. Vox Sang 1972;23:537-45.
178. Sringarm S, Chiewsilp P, Tubrod J. Cad receptor in Thai blood donors. Vox Sang 1974;26:462-6.
179. Sanger R, Gavin J, Tippett P, Teesdale P, K. E. Plant agglutinin for another human blood-group. Lancet 1971;297.
180. Peetermans ME, Cole-Dergent J. Haemolytic transfusion reaction due to anti-Sd^a. Vox Sang 1970;18:67-70.
181. Reznicek MJ, Cordle DG, Strauss RG. A hemolytic reaction implicating Sd^a antibody missed by immediate spin crossmatch. Vox Sang 1992;62:173-5.
182. Robbe-Masselot C, Maes E, Rousset M, Michalski JC, Capon C. Glycosylation of human fetal mucins: a similar repertoire of O-glycans along the intestinal tract. Glycoconj J 2009;26:397-413.
183. Bird GW, Wingham J. Some serological properties of the Cad receptor. Vox Sang 1971;20:55-61.
184. Blanchard D, Cartron JP, Fournet B, Montreuil J, van Halbeek H, Vliegenthart JF. Primary structure of the oligosaccharide determinant of blood group Cad specificity. J Biol Chem 1983;258:7691-5.
185. Blanchard D, Capon C, Leroy Y, Cartron J-P, Fournet B. Comparative study of glycophorin A derived O-glycans from human Cad, Sd(a+) and Sd(a -) erythrocytes. Biochem J 1985;232:813-8.
186. Donald AS, Soh CP, Watkins WM, Morgan WT. N-Acetyl-D-galactosaminyl-beta-(1 goes to 4)-d-galactose: a terminal non-reducing structure in human blood group Sd^a-active Tamm-Horsfall urinary glycoprotein. Biochem Biophys Res Commun 1982;104:58-65.
187. Donald AS, Yates AD, Soh CP, Morgan WT, Watkins WM. A blood group Sda-active pentasaccharide isolated from Tamm-Horsfall urinary glycoprotein. Biochem Biophys Res Commun 1983;115:625-31.

188. van Rooijen JJ, Kamerling JP, Vliegenthart JF. The abundance of additional N-acetyllactosamine units in N-linked tetraantennary oligosaccharides of human Tamm-Horsfall glycoprotein is a donor-specific feature. Glycobiology 1998;8:1065-75.
189. Capon C, Maes E, Michalski JC, Leffler H, Kim YS. Sd(a)-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. Biochem J 2001;358:657-64.
190. Malagolini N, Dall'Olio F, Di Stefano G, Minni F, Marrano D, Serafini-Cessi F. Expression of UDP-GalNAc:NeuAc alpha 2,3Gal beta-R beta 1,4(GalNAc to Gal) N-acetylgalactosaminyltransferase involved in the synthesis of Sda antigen in human large intestine and colorectal carcinomas. Cancer Res 1989;49:6466-70.
191. Piller F, Blanchard D, Huet M, Cartron JP. Identification of a alpha-NeuAc-(2-3)-beta-D-galactopyranosyl N-acetyl-beta-D- galactosaminyltransferase in human kidney. Carbohydr Res 1986;149:171-84.
192. Takeya A, Hosomi O, Kogure T. Identification and characterization of UDP-GalNAc: NeuAc alpha 2-3Gal beta 1-4Glc(NAc) beta 1-4(GalNAc to Gal)N-acetylgalactosaminyltransferase in human blood plasma. J Biochem 1987;101:251-9.
193. Serafini-Cessi F, Malagolini N, Dall'Olio F. Characterization and partial purification of beta-N-acetylgalactosaminyltransferase from urine of Sd(a+) individuals. Arch Biochem Biophys 1988;266:573-82.
194. Dohi T, Yuyama Y, Natori Y, Smith PL, Lowe JB, Oshima M. Detection of N-acetylgalactosaminyltransferase mRNA which determines expression of Sd^a blood group carbohydrate structure in human gastrointestinal mucosa and cancer. Int J Cancer 1996;67:626-31.
195. Wang HR, Hsieh CY, Twu YC, Yu LC. Expression of the human Sd(a) beta-1,4-N-acetylgalactosaminyltransferase II gene is dependent on the promoter methylation status. Glycobiology 2008;18:104-13.
196. Kawamura YI, Toyota M, Kawashima R*, et al.* DNA hypermethylation contributes to incomplete synthesis of carbohydrate determinants in gastrointestinal cancer. Gastroenterology 2008;135:142-51.e3.
197. Groux-Degroote S, Schulz C, Cogez V*, et al.* The extended cytoplasmic tail of the human *B4GALNT2* is critical for its Golgi targeting and post-Golgi sorting. FEBS J 2018;285:3442-63.
198. Jöud M, Möller M, Olsson ML. Identification of human glycosyltransferase genes expressed in erythroid cells predicts potential carbohydrate blood group loci. Sci Rep 2018;8:6040.
199. Marcus DM, Cass LE. Glycosphingolipids with Lewis blood group activity: uptake by human erythrocytes. Science 1969;164:553-5.
200. Groux-Degroote S, Wavelet C, Krzewinski-Recchi MA*, et al.* B4GALNT2 gene expression controls the biosynthesis of Sda and sialyl Lewis X antigens in healthy and cancer human gastrointestinal tract. Int J Biochem Cell Biol 2014;53:442-9.
201. Kawamura YI, Kawashima R, Fukunaga R*, et al.* Introduction of Sd(a) carbohydrate antigen in gastrointestinal cancer cells eliminates selectin ligands and inhibits metastasis. Cancer Res 2005;65:6220-7.

202. Tanaka-Okamoto M, Hanzawa K, Mukai M, Takahashi H, Ohue M, Miyamoto Y. Identification of internally sialylated carbohydrate tumor marker candidates, including Sda/CAD antigens, by focused glycomic analyses utilizing the substrate specificity of neuraminidase. Glycobiology 2018;28:247-60.
203. Cartron JP, Prou O, Luilier M, Soulier JP. Susceptibility to invasion by Plasmodium falciparum of some human erythrocytes carrying rare blood group antigens. Br J Haematol 1983;55:639-47.
204. Heaton BE, Kennedy EM, Dumm RE*, et al.* A CRISPR Activation Screen Identifies a Pan-avian Influenza Virus Inhibitory Host Factor. Cell Rep 2017;20:1503-12.
205. McMorran BJ, Miceli MC, Baum LG. Lectin-binding characterizes the healthy human skeletal muscle glycophenotype and identifies disease-specific changes in dystrophic muscle. Glycobiology 2017;27:1134-43.
206. Mohlke KL, Purkayastha AA, Westrick RJ*, et al.* Mvwf, a dominant modifier of murine von Willebrand factor, results from altered lineage-specific expression of a glycosyltransferase. Cell 1999;96:111-20.
207. Johnsen JM, Levy GG, Westrick RJ, Tucker PK, Ginsburg D. The endothelial-specific regulatory mutation, Mvwf1, is a common mouse founder allele. Mamm Genome 2008;19:32-40.
208. Millar CM, Brown SA. Oligosaccharide structures of von Willebrand factor and their potential role in von Willebrand disease. Blood Rev 2006;20:83-92.
209. Li PT, Liao CJ, Wu WG, Yu LC, Chu ST. Progesterone-regulated B4galnt2 expression is a requirement for embryo implantation in mice. Fertil Steril 2011;95:2404-9, 9.e1-3.
210. Ben Jemaa S, Ruesche J, Sarry J, Woloszyn F, Lassoued N, Fabre S. The high prolificacy of D'man sheep is associated with the segregation of the FecL(L) mutation in the *B4GALNT2* gene. Reprod Domest Anim 2019;54:531-7.
211. Guo X, Wang X, Liang B*, et al.* Molecular cloning of the B4GALNT2 gene and its single nucleotide polymorphisms association with litter size in small tail han sheep. Animals 2018;8:160.
212. Yeh CC, Chang CJ, Twu YC*, et al.* The differential expression of the blood group P(1) -A4GALT and P(2) -A4GALT alleles is stimulated by the transcription factor early growth response 1. Transfusion 2018;58:1054-64.
213. Haselberger CG, Schenkel-Brunner H. Evidence for erythrocyte membrane glycoproteins being carriers of blood-group P1 determinants. FEBS Lett 1982;149:126-8.
214. Yang Z, Bergström J, Karlsson K-A. Glycoproteins with Galalpha4Gal are absent from human erythrocyte membranes, indicating that glycolipids are the sole carriers of blood group P activities. J Biol Chem 1994;269:14620-4.
215. Kaczmarek R, Szymczak-Kulus K, Bereznicka A*, et al.* Single nucleotide polymorphisms in *A4GALT* spur extra products of the human Gb3/CD77 synthase and underlie the P1PK blood group system. PLoS One 2018;13:e0196627.
216. The 1000 Genomes Project Consortium, Auton A, Brooks LD*, et al.* A global reference for human genetic variation. Nature 2015;526:68-74.

217. Sun L, Jiang C, Xu C*, et al.* Down-regulation of long non-coding RNA RP11-708H21.4 is associated with poor prognosis for colorectal cancer and promotes tumorigenesis through regulating AKT/mTOR pathway. Oncotarget 2017;8:27929-42.
218. Veldhuisen B, Ligthart P, van der Mark-Zoet J*, et al.* Identification of a single homozygous mutation in the B4GALNT2 gene in individuals lacking the Sd(a) (SID) antigen on red blood cells. Vox Sang 2019;114:193 (P-401).
219. Kuvardina ON, Herglotz J, Kolodziej S*, et al.* RUNX1 represses the erythroid gene expression program during megakaryocytic differentiation. Blood 2015;125:3570-9.
220. Thuresson B. Studies on polymorphism and transcriptional regulation of the ABO and P1PK histo-blood group genes.: Faculty of Medicine, Lund University, 2011.
221. Nguyen HQ, Hoffman-Liebermann B, Liebermann DA. The zinc finger transcription factor Egr-1 is essential for and restricts differentiation along the macrophage lineage. Cell 1993;72:197-209.
222. Schulze C, Büchse T, Mikkat S, Bittorf T. Erythropoietin receptor-mediated Egr-1 activation: structural requirements and functional implications. Cell Signal 2008;20:1848-54.
223. Morimoto K, Suzuki N, Tanida I*, et al.* Blood group P1 antigen-bearing glycoproteins are functional but less efficient receptors of Shiga toxin than conventional glycolipid-based receptors. J Biol Chem 2020;295:9490-501.
224. Tippett P, Sanger R, Race RR, Swanson J, Busch S. An agglutinin associated with the P and the ABO blood group systems. Vox Sanguinis 1965;10:269-80.
225. Cooling LL, Kelly K. Inverse expression of P^k and Luke blood group antigens on human RBCs. Transfusion. 2001;41:898-907.
226. Zhou D, Henion TR, Jungalwala FB, Berger EG, Hennet T. The β1,3-galactosyltransferase β3GalT-V is a stage-specific embryonic antigen-3 (SSEA-3) synthase. J Biol Chem 2000;275:22631-4.
227. Saito S, Aoki H, Ito A*, et al.* Human α2,3-sialyltransferase (ST3Gal II) is a stage-specific embryonic antigen-4 synthase. J Biol Chem 2003;278:26474-9.
228. Cooling LL, Gu Y, Judd WJ. A Missense mutation in B3GalT5, the glycosyltransferase responsible for galatosylgloboside and Lewis c synthesis, may be associated with the LKE-weak phenotype in African Americans. Transfusion 2002;42:9S.