Genetic prediction of 33 blood group phenotypes using an existing genotype dataset

Moslemi, Camous; Sækmose, Susanne G.; Larsen, Rune; Bay, Jakob T.; Brodersen, Thorsten; Didriksen, Maria; Hjalgrim, Henrik; Banasik, Karina; Nielsen, Kaspar R.; Bruun, Mie T.; Dowsett, Joseph; Dinh, Khoa M.; Mikkelsen, Susan; Mikkelsen, Christina; Hansen, Thomas F.; Ullum, Henrik; Erikstrup, Christian; Brunak, Søren; Krogfelt, Karen Angeliki; Storry, Jill R.; Ostrowski, Sisse R.; Olsson, Martin L.; Pedersen, Ole B.

Published in:
Transfusion

DOI (link to publication from Publisher):
10.1111/trf.17575

Publication date:
2023

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

Received: 16 March 2023 | Revised: 3 August 2023 | Accepted: 10 August 2023

DOI: 10.1111/trf.17575

BLOOD GROUP GENOMICS

TRANSFUSION

Genetic prediction of 33 blood group phenotypes using an existing genotype dataset

Camous Moslemi[1,2,3] | Susanne G. Sækmose[1] | Rune Larsen[1] | Jakob T. Bay[1] | Thorsten Brodersen[1] | Maria Didriksen[4] | Henrik Hjalgrim[5] | Karina Banasik[6] | Kaspar R. Nielsen[7] | Mie T. Bruun[8] | Joseph Dowsett[4] | Khoa M. Dinh[2] | Susan Mikkelsen[2] | Christina Mikkelsen[4,9] | Thomas F. Hansen[6,10] | Henrik Ullum[11] | Christian Erikstrup[2] | Søren Brunak[6] | Karen Angeliki Krogfelt[3] | Jill R. Storry[12,13] | Sisse R. Ostrowski[4,14] | Martin L. Olsson[12,13] | Ole B. Pedersen[1,14]

[1]Department of Clinical Immunology, Zealand University Hospital, Køge, Denmark

[2]Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark

[3]Department of Science and Environment, Roskilde University, Roskilde, Denmark

[4]Department of Clinical Immunology, Copenhagen University Hospital, Rigshopitalet, Copenhagen, Denmark

[5]Danish Cancer Society Research Center, Copenhagen, Denmark

[6]Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

[7]Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark

[8]Department of Clinical Immunology, Odense University Hospital, Odense, Denmark

[9]Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark

[10]Department of Neurology, Dansk Hovedpine Center and Multiple Sclerosis Center, Rigshospitalet, Glostrup, Denmark

[11]Statens Serum Institut, Copenhagen, Denmark

[12]Department of Laboratory Medicine, Lund University, Lund, Sweden

[13]Department of Clinical Immunology and Transfusion Medicine, Office for Medical Services, Region Skåne, Sweden

[14]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

Correspondence
Camous Moslemi, Department of Clinical Immunology, Zealand University Hospital, Køge, Denmark.
Email: kamiboy@gmail.com

Funding information
A.P Møller Fonden, Grant/Award

Abstract

Background: Accurate blood type data are essential for blood bank management, but due to costs, few of 43 blood group systems are routinely determined in Danish blood banks. However, a more comprehensive dataset of blood types is useful in scenarios such as rare blood type allocation. We aimed to investigate the viability and accuracy of predicting blood types by

Abbreviations: CHB, Copenhagen Hospital Biobank; CI, confidence interval; DBDS, The Danish Blood Donor Study; GSA, global screening array; NPV, negative predictive value: true negatives/(false negatives + true negatives); PCR, polymerase chain reaction; RBC, red blood cell; rsID, reference SNP cluster ID; SNV, single nucleotide variant.

Sisse R. Ostrowski, Martin L. Olsson, and Ole B. Pedersen contributed equally to this study.

Number: 45000 DKK; Bloddonornes forskningsfond, Grant/Award Number: 64931 DKK; Novo Nordisk Foundation, Grant/Award Numbers: nnf14cc0001, nnf17oc0027594

leveraging an existing dataset of imputed genotypes for two cohorts of approximately 90,000 each (Danish Blood Donor Study and Copenhagen Biobank) and present a more comprehensive overview of blood types for our Danish donor cohort.

Study Design and Methods: Blood types were predicted from genome array data using known variant determinants. Prediction accuracy was confirmed by comparing with preexisting serological blood types. The Vel blood group was used to test the viability of using genetic prediction to narrow down the list of candidate donors with rare blood types.

Results: Predicted phenotypes showed a high balanced accuracy >99.5% in most cases: A, B, C/c, Co^a/Co^b, Do^a/Do^b, E/e, Jk^a/Jk^b, Kn^a/Kn^b, Kp^a/Kp^b, M/N, S/s, Sd^a, Se, and Yt^a/Yt^b, while some performed slightly worse: Fy^a/Fy^b, K/k, Lu^a/Lu^b, and Vel ~99%–98% and C^W and P_1 ~96%. Genetic prediction identified 70 potential Vel negatives in our cohort, 64 of whom were confirmed correct using polymerase chain reaction (negative predictive value: 91.5%).

Discussion: High genetic prediction accuracy in most blood groups demonstrated the viability of generating blood types using preexisting genotype data at no cost and successfully narrowed the pool of potential individuals with the rare Vel-negative phenotype from 180,000 to 70.

KEYWORDS

ABO, blood group systems, blood groups, Danish blood type rates, Danish population, Denmark, Diego, Dombrock, donor blood typing, Duffy, erythrocyte antigens, genetic blood typing, Kell, Kidd, Knops, Lewis, Lutheran, MNS, P1PK, Rh, secretor, Vel, Yt

1 | INTRODUCTION

Blood group antigens are epitopes on glycans or (glyco) proteins with antigenic potential found on the surface of red blood cells. Understanding the antigenic nature of blood groups was instrumental in enabling the transfusion of blood between donors and recipients at dramatically decreased risk for potentially lethal complications.

Blood typing is a crucial component in proper management and allocation of donated blood. However, with 44 officially recognized blood group systems containing 354 antigens,[1] testing every donor for all blood groups is currently neither economically nor logistically feasible.

While the exact testing protocol differs slightly among regional Danish blood banks, most routinely test donors for selected antigens in the ABO and Rh blood groups. Additionally, in most places, first-time donors are tested for selected antigens in the following blood group systems: Duffy, MNS, Kell, and Kidd.

In contrast, the blood groups Colton, Dombrock, Knops, Lutheran, P1Pk, Sid, Vel, Yt, and secretor status are not tested at all, or only tested in a limited capacity, or only routinely tested in certain regions.

Besides its potential in research-related applications, more comprehensive blood type data in these lesser tested blood groups have utility in improving the matching of donors and recipients, for example, to locate donors with an exceedingly rare blood type or a rare combination of blood types or to avoid alloimmunization in patients with chronic transfusion needs.[2,3]

Fortunately, the vast majority of blood group antigens have single nucleotide determinants, making them ideal candidates for genetic blood typing.[4–6] While custom sequencing solutions exist for the purpose of genetic blood group typing,[7,8] sequencing a large portion of Danish donors from scratch using these solutions would come at considerable cost and time.

The Danish Blood Donor Study (DBDS)[9,10] already has an existing dataset of over 90,000 donors genotyped on an Illumina Global Screening Array (GSA) chip, with subsequent imputation by deCODE genetics.[11,12]

The main aim of this study was to test the viability of using our existing genetic dataset to enrich DBDS donor blood types with a more comprehensive coverage of certain lesser tested blood group antigens. This was considered a worthwhile pursuit primarily because the genotypes were already available and using them would come at only a

limited additional cost. Furthermore, an extended dataset of blood types could be used to improve blood bank management and aid in future blood-type-related research.

While there is interest in and benefits[6] to replacing traditional serological methods completely with genetic methods, such an approach would require techniques, such as whole genome sequencing, that offer more precision than GSA typing alone can provide.[13] As such, the aim of the project from the onset was not to attempt to replace traditional serological blood typing methods, but to complement them.

The secondary goals of the project were to compare our genetic prediction accuracy to recent studies using alternate genetic methods[14,15] and using the available data to provide an overview of the blood antigen prevalence in the Danish population.

2 | MATERIALS AND METHODS

2.1 | Study cohorts

This study is based on two cohorts (Table 1). The main cohort consisted of genotyped donors ($N = 92{,}841$) participating in DBDS, a nationwide prospective epidemiological study of Danish blood donors[12] who were included during the period 2010–2016. The secondary cohort consists of genotyped patients ($N = 96{,}022$) from the Copenhagen Hospital Biobank (CHB)[16] genetic study on transfusion and transfusion outcome. Both cohorts have been genotyped by deCODE Genetics, Iceland, using the same Infinium GSA from Illumina. This array includes more than 600,000 genetic variants across the genome. Using additional data from the deCODE North European reference panel, the Illumina array data were used to impute genetic variants across the genome with reasonable accuracy down to a minor allele frequency (MAF) of 0.01. This dataset can therefore be used to detect genetic blood type variants directly located on the array as well as blood type variants imputed based on adjacent markers. Each imputed locus has an associated imputation information score[17] (info-score) between 0 and 1 which can be seen as a measure of its probabilistic certainty. The info-score threshold where one considers an imputation too unreliable to use is a matter of preference according to need or desire for accuracy. At deCODE, imputations with an info-score below 0.75 were considered too unreliable and were discarded.

2.2 | Phenotype cohorts

The serological blood types for the CHB cohort were sourced from a larger historical dataset of blood types, not confined to DBDS alone; however, the dataset covered only ABO and RhD blood types. All serologically determined blood types for the DBDS cohort were sourced from the five Danish regional blood banks. In addition to the serological blood types, in the Capital Region, 19,142 randomly selected donors had been tested using TaqMan[18] polymerase chain reaction (PCR) probes specific for the rs566629828[19] variant, identifying 15 Vel-negative and 19,127 Vel-positive donors.

Blood types determined by each region are stored in local databases, from which they are manually exported on a regular basis and shared with researchers as part of the DBDS study. Differences in the number and types of tests available from each region are due to different local testing policies, practices, and equipment. Since the dataset covers millions of tests across dozens of types made over the span of decades, during which local regional test policies, equipment, and protocols were subject to change, providing an overview of the history behind their creation is beyond the scope of this study.

TABLE 1 Cohort demographic data.

Cohort	*N* (%)	Median age [min:max]
DBDS	92,841	49 [23:70]
DBDS (M)	47,314 (51%)	
DBDS (F)	45,527 (49%)	
CHB	96,022	79 [23:115]
CHB (M)	52,622 (56%)	
CHB (F)	43,400 (44%)	

Note: Demographic statistics for the two cohorts, Danish Blood Donor Study (DBDS) and Copenhagen Hospital Biobank (CHB). Tally (*N*), median, maximum and minimum age, broken down for total cohort, males (M), and females (F) separately.

2.3 | Blood group genetics

Previous research has placed antigens that are products of single or closely linked genes in the same blood group system and traced the origin of many antigen types to single nucleotide variants (SNVs) within these genes.[4] This simplified our genotype-based blood type prediction by focusing the process on examination of SNVs at specific locations in the genome.

Although many causative polymorphisms can exist within one blood group gene, in most blood groups, all but a few of these variants are typically rare (<1%). In such cases, we ignored the rare variants and instead focused our predictions around the few variants that confer the blood group antigen status in a majority of the population.[20]

TABLE 2 Blood group single nucleotide variants.

Blood group	Antigen: allele 1/2	Gene	rsID	Chromosome	Position (GRCh38)	Allele 1	Allele 2	MAF DK	MAF Europe	Info-score
ABO	A+/A−	*ABO*	rs550057	chr9	133,271,182	T	C	0.2738	0.2553	0.9994
ABO	A−/A+	*ABO*	rs60937319	chr9	133,264,504	TTGCCAAA	–	0.2737	0.2639	0.9996
ABO	B−/B+	*ABO*	rs8176743	chr9	133,256,028	C	T	0.0785	0.0806	0.9985
ABO	B−/B+	*ABO*	rs8176741	chr9	133,256,074	G	A	0.0798	0.0876	0.9982
ABO	B−/B+	*ABO*	rs8176747	chr9	133,255,928	C	G	0.0788	0.0809	0.9991
ABO	B−/B+	*ABO*	rs8176749	chr9	133,255,801	C	T	0.0788	0.0796	0.9991
Colton	Coa/Cob	*AQP1*	rs28362692	chr7	30,912,043	C	T	0.0457	0.0422	0.8743
Dombrock	Doa/Dob	*ART4*	rs11276	chr12	14,840,505	C	T	0.3891	0.3864	0.9992
Duffy	Fya/Fyb	*ACKR1*	rs12075	chr1	159,205,564	G	A	0.4291	0.4212	0.9830
Kell	K/k	*KEL*	rs8176058	chr7	142,957,921	A	G	0.0396	0.0415	0.9183
Kell	Kpb/Kpa	*KEL*	rs8176059	chr7	142,954,267	G	A	0.0106	0.0102	0.9654
Kidd	Jka/Jkb	*SLC14A1*	rs1058396	chr18	45,739,554	G	A	0.4808	0.4843	0.9995
Knops	Kna/Knb	*CR1*	rs41274768	chr1	207,609,424	G	A	0.0364	0.0285	0.9997
Lutheran	Lub/Lua	*BCAM*	rs28399653	chr19	44,812,188	G	A	0.0424	0.0314	0.9695
MNS	N/M	*GYPA*	rs7682260	chr4	144,120,567	A	G	0.4632	0.4809	0.9966
MNS	S/s	*GYPB*	rs7683365	chr4	143,999,443	A	G	0.3009	0.3142	0.9971
P1PK	P1+/P1−	*A4GALT*	rs5751348	chr22	42,717,787	C	A	0.4653	0.4875	0.9985
Rh	C/c	*RHCE*	rs586178	chr1	25,430,739	G	C	0.4420	0.4616	0.9927
Rh	Cw−/Cw+	*RHCE*	rs138268848	chr1	25,420,665	T	C	0.0169	0.0145	0.9161
Rh	E/e	*RHCE*	rs609320	chr1	25,390,874	G	C	0.1595	0.1537	0.9909
Secretor[a]	Se+/Se−	*FUT2*	rs601338	chr19	48,703,417	G	A	0.4496	0.4776	0.9993
Sid	Sda+/Sda−	*B4GALNT2*	rs7224888	chr17	49,168,801	T	C	0.1062	0.0998	0.9997
Vel	Vel+/Vel−	*SMIM1*	rs566629828	chr1	3,775,433	AGCCTAGGGGCTGTGTC	–	0.0221	0.0155	0.8719
Yt	Yta/Ytb	*ACHE*	rs1799805	chr7	100,893,176	G	T	0.0398	0.0455	0.9995
Diego	Wra+/Wra−	*SLC4A1*	rs75731670	chr17	44,254,581	C	T	Absent	0.0005	Absent

Note: List of each single nucleotide variant used for genetic blood type prediction. Column 1 denotes blood group system. Column 2 denotes the antigen status associated with each allele of the variant. Column 3 denotes the gene where the variant is located. Column 4 denotes the rsID of the variant. Columns 5 and 6 denote the chromosome and position of the variant. Columns 7 and 8 denote the variant alleles. Column 9 denotes the minor allele frequency (MAF) of variant within our cohort. Column 10 denotes the European MAF of the variant according to genomeAD. The last column denotes the information score (info-score) of the variant, a value between 0 and 1, with higher scores indicating higher quality imputation. In some cases, such as ABO, several variants can be used to determine antigen status, where the consensus of all the used variants determines the final status. Please observe that the +/− signs in A+/A− and B+/B− do not refer to RhD status.

[a]Not formally a blood group but included here because of its relationship to the ABH antigens and great interest for disease susceptibility studies.

TABLE 3 Prediction accuracy.

System	Antigens	Predicted	Confirmed	Prediction accuracy	Prediction sensitivity	Prediction specificity	TP	TN	FP	FN	Balanced accuracy
ABO	A	187,974	99,701	99.7%	0.9990	0.9977	45,194	54,162	157	106	0.9974
ABO	B	187,933	99,691	99.996%	0.9990	0.9997	14,357	85,281	27	15	0.9993
Colton	Coa/Cob	182,290	2424	99.6%	0.9907	1.0000	960	1455	0	9	0.9954
Dombrock	Doa/Dob	188,142	1358	100%	1.0000	1.0000	992	366	0	0	1.0000
Duffy	Fya/Fyb	185,875	119,264	99.4%	0.9978	0.9826	87,460	31,061	549	194	0.9902
Kell	K/k	181,904	80,640	99.5%	0.9771	0.9979	11,330	68,898	147	265	0.9875
Kell	Kpb/Kpa	186,284	16,713	99.8%	0.9930	0.9996	3257	13,427	6	23	0.9963
Kidd	Jka/Jkb	188,270	122,471	99.8%	0.9985	0.9957	91,006	31,189	135	141	0.9971
Knops	Kna/Knb	188,053	6034	99.97%	1.0000	0.9993	3217	2815	2	0	0.9996
Lutheran	Lub/Lua	183,506	12,166	99.4%	0.9842	0.9963	2683	9405	35	43	0.9903
MNS	N/M	186,076	65,438	99.7%	0.9978	0.9930	49,626	15,591	110	111	0.9954
MNS	S/s	186,843	81,624	99.8%	0.9992	0.9972	55,691	25,813	73	47	0.9982
P1PK	P1	187,119	9075	98.1%	0.9971	0.9287	6965	1941	149	20	0.9629
Rh	C/c	186,400	141,903	99.89%	0.9995	0.9975	102,577	39,180	98	48	0.9985
Rh	Cw	186,197	31,097	99.78%	0.9390	0.9995	878	30,147	15	57	0.9693
Rh	E/e	186,804	139,616	99.8%	0.9984	0.9986	84,992	54,416	74	134	0.9985
Secretor	Se	187,672	8872	100%	1.0000	1.0000	2732	686	0	0	1.0000
Sid	Sda	187,499	0								
Vel	Vel	179,855	19,142	99.96%	0.9996	0.9846	16,282	64	1	6	0.9921
Yt	Yta/Ytb	188,392	1646	100%	1.0000	1.0000	989	657	0	0	1.0000

Note: Accuracy of genetic blood types. Columns 1 and 2 denote the blood group system and antigen. Column 3 denotes the number of genetic predictions generated. Column 4 denotes the number of confirmatory serological blood types that existed. Column 5 is the calculated prediction accuracy, while columns 6 and 7 denote prediction sensitivity TP/(TP + FN) and specificity TN/(TN + FP). Columns 8–11 count the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) among the predicted blood types that could be confirmed. The last column denotes balanced accuracy, calculated as (sensitivity + specificity)/2.

TABLE 4 Positive rate of genetic and serological blood types.

		Serology[a]		Genetic		Serology[a]		Genetic		
System	Antigen	Positives	Negatives	Positives	Negatives	Ratio	95% CI	Ratio	95% CI	χ^2
ABO	A	353,876	403,272	88,151	99,772	0.4674	[0.4663–0.4685]	0.4691	[0.4668–0.4713]	0.187
ABO	B	122,687	634,505	27,409	160,523	0.162	[0.1612–0.1629]	0.1458	[0.1443–0.1474]	<0.001
Colton	Coa	2397	7	181,938	352	0.9971	[0.994–0.9988]	0.9981	[0.9979–0.9983]	
Colton	Cob	439	4017	15,478	166,812	0.0985	[0.0899–0.1076]	0.0849	[0.0836–0.0862]	
Diego	Wra	237	145,777			0.0016	[0.0014–0.0018]			
Dombrock	Doa	1250	756	118,045	70,097	0.6231	[0.6015–0.6444]	0.6274	[0.6252–0.6296]	
Dombrock	Dob	1694	312	159,708	28,434	0.8445	[0.8279–0.8601]	0.8489	[0.8472–0.8505]	
Duffy	Fya	204,455	99,208	124,852	61,023	0.6733	[0.6716–0.675]	0.6717	[0.6696–0.6738]	0.2491
Duffy	Fyb	201,639	50,173	152,061	33,814	0.8008	[0.7992–0.8023]	0.8181	[0.8163–0.8198]	<0.001
Kell	K	29,179	358,938	12,258	169,646	0.0752	[0.0744–0.076]	0.0674	[0.0662–0.0685]	<0.001
Kell	k	26,723	512	181,670	234	0.9812	[0.9795–0.9828]	0.9987	[0.9985–0.9989]	
Kell	Kpa	2475	107,122	3644	182,640	0.0226	[0.0217–0.0235]	0.0196	[0.0189–0.0202]	<0.001
Kell	Kpb	11,088	38	186,263	21	0.9966	[0.9953–0.9976]	0.9999	[0.9998–0.9999]	<0.001
Kidd	Jka	230,140	66,923	144,505	43,765	0.7747	[0.7732–0.7762]	0.7675	[0.7656–0.7694]	<0.001
Kidd	Jkb	178,251	68,329	137,734	50,536	0.7229	[0.7211–0.7247]	0.7316	[0.7296–0.7336]	<0.001
Knops	Kna	8938	12	187,804	249	0.9987	[0.9977–0.9993]	0.9987	[0.9985–0.9988]	1
Knops	Knb	8321	625	13,659	174,394	0.0699	[0.0647–0.0753]	0.0726	[0.0715–0.0738]	0.0339
Lewis	Lea	23,376	98,990			0.191	[0.1888–0.1932]			
Lewis	Leb	62,419	22,861			0.7319	[0.7289–0.7349]			
Lutheran	Lua	7773	81,278	14,643	168,863	0.0873	[0.0854–0.0892]	0.0798	[0.0786–0.081]	<0.001
Lutheran	Lub	8870	181	183,201	305	0.98	[0.9769–0.9828]	0.9983	[0.9981–0.9985]	
MNS	N	108,137	42,474	133,225	52,851	0.718	[0.7157–0.7203]	0.716	[0.7139–0.718]	0.1976
MNS	s	150,091	15,150	170,061	16,782	0.9083	[0.9069–0.9097]	0.9102	[0.9089–0.9115]	0.0551
MNS	S	106,856	105,701	95,307	91,536	0.5027	[0.5006–0.5048]	0.5101	[0.5078–0.5124]	<0.001
MSN	M	191,364	53,683	145,543	40,533	0.7809	[0.7793–0.7826]	0.7822	[0.7803–0.784]	0.3302
P1PK	P1	70,829	20,966	146,582	40,537	0.7716	[0.7689–0.7743]	0.7834	[0.7815–0.7852]	<0.001
Rh	c	282,786	63,425	152,709	33,691	0.8168	[0.8155–0.8181]	0.8193	[0.8175–0.821]	0.0273
Rh	C	237,634	135,420	123,179	63,221	0.637	[0.6355–0.6385]	0.6608	[0.6587–0.663]	<0.001
Rh	Cw	5169	158,909	5336	180,861	0.0315	[0.0307–0.0324]	0.0287	[0.0279–0.0294]	<0.001

TABLE 4 (Continued)

System	Antigen	Serology[a]		Genetic		Serology[a]		Genetic		
		Positives	Negatives	Positives	Negatives	Ratio	95% CI	Ratio	95% CI	χ^2
Rh	D	619,764	140,503			0.8152	[0.8143–0.8161]			
Rh	D-weak	1808	41,586			0.0417	[0.0398–0.0436]			
Rh	e	304,605	8893	182,112	4692	0.9716	[0.971–0.9722]	0.9749	[0.9742–0.9756]	<0.001
Rh	E	107,201	269,042	53,611	133,193	0.2849	[0.2835–0.2864]	0.287	[0.2849–0.289]	0.1068
Sid	Sda			169,235	4868			0.9749	[0.9742– 0.9756]	
Secretor	Se	7136	1736	149,942	37,730	0.8043	[0.7959–0.8125]	0.799	[0.7971–0.8008]	0.2223
Vel	Vel	19,127	15	179,780	75	0.9959	[0.9987–0.9996]	0.9996	[0.9995–0.9997]	0.0367
Yt	Yta	2790	32	188,091	301	0.9992	[0.984–0.9922]	0.9984	[0.9982–0.9986]	
Yt	Ytb	136	1876	14,175	174,217	0.0676	[0.057–0.0795]	0.0752	[0.0741–0.0764]	

Note: Comparison of positive antigen rates calculated based on genetic or serological blood types. Columns 1 and 2 denote blood group system and antigen, respectively. Columns 3 and 4 tally existing positive and negative serological blood type results (serology). Columns 5 and 6 tally positive and negative predicted blood types (genetic). Columns 7–10 denote the calculated positive blood rate and 95% confidence interval for serology and genetic blood types. The last column denotes Pearson's Chi-squared test comparing the blood type rates of each source. Differences in rates are to be expected in blood groups without comprehensive serological testing as they might only tested under certain circumstances, which could introduce a bias. The ABO serology cohort is sourced from the entire donor dataset, instead of just the Danish Blood Donor Study (DBDS) subset; therefore, it is larger. In the case of the large deviance in the ABO blood group, which is comprehensively tested, the difference reveals a bias for donors with the blood antigen A in Danish blood banks. The other deviances are rather small and likely due to chance. The large size of the ABO cohort is due to inclusion of all available donor ABO blood types, not restricted to only DBDS donors.
Abbreviation: CI, confidence interval.

[a]Serological tests in all cases except for Vel blood group, which is based on polymerase chain reaction tests.

TABLE 5 Estimated Danish blood group antigen rates.

System	Antigen	Positive rate	95% CI	Source
Colton	Coa	0.9981	[0.9979–0.9983]	SNV
Colton	Cob	0.0849	[0.0836–0.0862]	SNV
Diego	Wra	0.0016	[0.0014–0.0018]	Serology
Dombrock	Doa	0.6274	[0.6252–0.6296]	SNV
Dombrock	Dob	0.8489	[0.8472–0.8505]	SNV
Duffy	Fya	0.6717	[0.6696–0.6738]	SNV
Duffy	Fyb	0.8181	[0.8163–0.8198]	SNV
Kell	K	0.0752	[0.0744–0.076]	Serology
Kell	k	0.9812	[0.9795–0.9828]	Serology
Kell	Kpa	0.0196	[0.0189–0.0202]	SNV
Kell	Kpb	0.9999	[0.9998–0.9999]	SNV
Kidd	Jka	0.7675	[0.7656–0.7694]	SNV
Kidd	Jkb	0.7316	[0.7296–0.7336]	SNV
Knops	Kna	0.9987	[0.9985–0.9988]	SNV
Knops	Knb	0.0726	[0.0715–0.0738]	SNV
Lewis	Lea	0.1910	[0.1888–0.1932]	Serology
Lewis	Leb	0.7319	[0.7289–0.7349]	Serology
Lutheran	Lua	0.0873	[0.0854–0.0892]	Serology
Lutheran	Lub	0.9800	[0.9769–0.9828]	Serology
MNS	N	0.7160	[0.7139–0.718]	SNV
MNS	s	0.9102	[0.9089–0.9115]	SNV
MNS	S	0.5101	[0.5078–0.5124]	SNV
MSN	M	0.7822	[0.7803–0.784]	SNV
P1PK	P1	0.7716	[0.7689–0.7743]	Serology
Rh	c	0.8193	[0.8175–0.821]	SNV
Rh	C	0.6608	[0.6587–0.663]	SNV
Rh	Cw	0.0315	[0.0307–0.0324]	Serology
Rh	D	0.8152	[0.8143–0.8161]	Serology
Rh	D-weak	0.0417	[0.0398–0.0436]	Serology
Rh	e	0.9749	[0.9742–0.9756]	SNV
Rh	E	0.2870	[0.2849–0.289]	SNV
Sid	Sda	0.2013	[0.1995–0.2031]	SNV
Secretor	Se	0.7990	[0.7971–0.8008]	SNV
Vel	Vel	0.9992	[0.9987–0.9996]	PCR
Yt	Yta	0.9984	[0.9982–0.9986]	SNV
Yt	Ytb	0.0752	[0.0741–0.0764]	SNV

Note: Estimated rate of positive antigen status for the Danish population in several blood groups. Columns 1 and 2 denote blood group system and antigen, respectively. Columns 3 and 4 denote the rate of positive phenotype and 95% confidence interval (CI) in the Danish population. The last column denotes the source of the estimated rate, either imputed single nucleotide variants (SNVs), or serological testing (serology), or polymerase chain reaction (PCR).

2.4 | Validation of predicted Vel genotypes

A selection of potential Vel-negative cases within the DBDS ($N = 38$) and CHB ($N = 37$) cohorts was identified based on the imputed rs566629828[19] variant. Validation was performed by assessing DNA extracted from archived frozen blood samples if available, and subsequent genotype confirmation was performed using a *SMIM1* exon 3-specific PCR spanning

the 17-bp deletion that confers the Vel-negative blood type.[19]

3 | STATISTICS AND PROCESSING

3.1 | Genotype processing

The unphased genetic data were stored in plink format and were read using custom Python (v3.5) scripts. Blood group antigen types were generated either using interpretation of single SNVs or a consensus of several SNVs when multiple associated SNVs are known to confer a certain blood type (Table 2). Associated SNVs were identified using relevant textbooks[4] and online resources.[21]

3.2 | Statistics

Rstudio (R v4.1.0) was used for data processing and statistical analysis. Serological blood type data were manually inspected for any invalid entries, and cohort identification was performed using unique donor IDs.

Pearson's Chi-squared testing[22] was used to compare blood group antigen rates derived from genetic and serological data. Fisher's exact test[23] was used to obtain 95% confidence intervals for each blood type rate. Accuracy of genetic blood typing was calculated using the balanced accuracy F-measure (F1-score).[24–26] Balanced accuracy was selected since it weighs both precision and recall equally, thereby making it a better metric for judging prediction accuracy in low- or high-prevalent phenotypes where class distribution is uneven.

4 | RESULTS

4.1 | Accuracy estimation

The accuracy of genetically determined blood types was estimated using preexisting serologically determined blood types as the golden standard (Table 3). Most of the genetic blood types had a high balanced accuracy of over 99.5%, with some achieving a good balanced accuracy of ~99% (Fy^a/Fy^b, Lu^a/Lu^b, and Vel) and ~98% (K/k). However, two (C^W and P1) only demonstrated a lower balanced accuracy of ~96% (Table 3).

4.2 | Antigen rate estimation

We estimated the rate of antigen types based on serological tests and genetic predictions separately. Comparison of calculated blood type rates for the serological and genetic sources revealed slight differences (Table 4). We selected a mix of predicted and serological sources to estimate final representative rates for the Danish population (Table 5).

The use of a mix was necessary since not all blood types were consistently serologically tested, in which case they could not be expected to be representative of the entire cohort, nor the Danish population.

4.3 | Vel antigen

Out of 179.855 individuals with an available imputed genotype for the rs566629828 variant, 75 were predicted to be homozygous for the 17 bp *SMIM1* deletion that results in the rare Vel-negative phenotype. Frozen blood samples were available for 70 of these, of which 64 were confirmed to be true Vel negatives (negative predictive value [NPV]: 91.5%). Upon including the 19,142 donors whose Vel status had previously been verified by Capitol Region using PCR tests, the balanced accuracy for genetic prediction of Vel status rises above 99%.

5 | DISCUSSION

The imputed genotypes from the Infinium GSA were used to predict phenotypes with an achieved balanced accuracy of over 99.5% for a majority of the attempted blood groups: A, B, C/c, Co^a/Co^b, Do^a/Do^b, E/e, Fy^a/Fy^b, Jk^a/Jk^b, Kn^a/Kn^b, Kp^a/Kp^b, M/N, S/s, Se, Vel, and Yt^a/Yt^b.

A small subset of antigens did not reach such a high accuracy but came close by achieving a balanced accuracy of 99% (Fy^a/Fy^b, Lu^a/Lu^b, and Vel) and 98% (K/k). While no definitive cause for any discordant predictions can be offered, it should be noted that in case of Lu^a/Lu^b, K/k, and Vel, the MAF and info-scores are lower than ideal, which could well explain the larger-than-ideal number of false positives and false negatives.

Two antigens (C^W and P1) fell somewhat shorter than the rest with a balanced accuracy of only 96%. The C^W variant (rs138268848) is located in the highly polymorphic region of the homologous *RHD* and *RHCE* genes and is further impacted by a relatively low MAF of 0.0169. For the above reasons, imputation of this variant is challenging, which explains its lower info-score and resulting lower prediction accuracy. As such, the only way to improve the genetic prediction of C^W, Lu^a/Lu^b, K/k, and Vel is by way of future improvements to the imputation process or possibly by using a denser genotyping array that directly includes these and other SNVs of interest with low MAFs.

The C^w antithetical antigen MAR and the alternative C^x antigen were never included in this study because of their high and low prevalence, respectively, which means that their respective variants were not imputed. However, even if they had been imputed, they may well have presented a similar challenge. They are of even greater importance in some populations, particularly in Finland, and all three will likely need strategies for improvement.

The lower observed prediction accuracy for the P1 antigen cannot be explained by info-score or low MAF. The cause could be unknown factors beyond the variant (rs5751348) controlling P1 antigen status. This is a likely explanation as said SNV is a potential binding site for transcription factors that influence the expression levels of the involved *A4GALT* gene.[27] This manner of influencing antigen status is different and less direct than most other causative SNVs covered in this study. In most other cases, the causative SNV either directly causes a protein residue change or is a null mutation.

In case of Fy^a/Fy^b, a likely culprit for the number of false positives could be rare causative variants other than the one used in this study (rs12075). The *Plasmodium vivax* parasite has known associations with the Duffy blood group.[28] Since the parasite uses Duffy antigens as binding sites, Duffy-negative individuals that are homozygous for $FY*B^{ES}$ (rs2814778) are protected against infection. As such the mutation that gives rise to this phenotype is common in sub-Saharan Africa.[29] Said variant is very rare in Europe (MAF 0.00344) and not imputed in our genotype dataset due to its MAF falling below 0.01. Therefore, any individuals with African ancestry who is a carrier of the given variant could result in an erroneous prediction in the Duffy blood group system.

Genetic blood type prediction using any genetic variants with lower than 0.01 MAF was not possible due to our genotype data lacking such variants. Other blood groups could not be predicted with acceptable accuracy due to higher genetic complexity (A-subtypes, Le^a/Le^b, RhD, and FORS1).

The accuracy of our genetic predictions for the ABO blood group compared favorably to the transfusion medicine array (TM-array) previously designed and used by Guo et al.[14] While close, our accuracy of ~99% was slightly higher than their ~98% accuracy in the ABO blood group. Like ours, their array can predict several other blood types; however, their prediction accuracy for these blood groups was not published and therefore could not be compared to.

Another study achieved an even higher final accuracy of 99.9% by developing 35 competitive allele-specific PCR assays in 1034 Danish blood donors.[15] This is an impressive figure even if the study involved a limited number of donors compared to ours. The high accuracy of their study is to be expected since PCR is a highly accurate genotyping technique and preferable when absolute accuracy is a high priority. However, to make use of their technique in our cohort would have required retesting over 90,000 DBDS donors, while our predictions used already existing genotypes at no cost and can be applied to all future genotyped DBDS donors and potentially other donors globally where the GSA and similar arrays have been used.

Although the main aim of this study was to predict blood types for GSA-genotyped DBDS donors, we opted to include the CHB cohort as well since they offered access to a larger dataset of genotypes and phenotypes in certain blood groups, thereby improving assessment of genetic blood typing accuracy and blood type rate calculations.

5.1 | Antigen rate estimation

While serological data from the DBDS cohort were initially planned to form the majority of blood group antigen rate estimation for the entire Danish population, the genetic cohort of non-donors unveiled a slight difference between the two. This bias revealed that some rates derived from the DBDS cohort deviated from the CHB cohort by a few percentages in some cases (Table 4). This was to be expected since blood banks have a bias for certain ABO blood types,[30,31] resulting in an overrepresentation of donors with the desired blood type compared to the background population. In such cases, we could default to using the CHB genetic cohort for blood type rate calculations to represent the Danish population more accurately. However, the ABO blood group has known associations to the risk of arterial thrombosis.[32] The magnitude of this potential bias in the CHB cohort is unknown due to a lack of an unbiased cohort for comparison. As such, we chose to omit ABO blood type rate estimations for the Danish population (Table 5). It should be noted that the existence of other types of unknown biases cannot be ruled out in other blood groups, in either the DBDS or CHB cohorts.

5.2 | Serological confirmation of discrepancies

Attempts were made to redo the serological blood tests of 40 individuals whose serologically determined and genetically predicted blood types were in conflict. Certain blood types such as ABO and RhD are serologically confirmed with each donation; in such cases, the serologically determined blood type can be considered correct,

and any conflicting predicted blood type should be erroneous, if not representing a rare weak variant not observed despite repeated clinical typings. However, in the case of blood types where routine retesting does not take place, a serological reconfirmation could be performed, provided that the donor in question was still active. In 38 of 40 selected cases, retesting confirmed the serologically obtained type to be correct, but in the two remaining cases, the genetically predicted type turned out to be the correct one. Neither serological nor genetic testing results are 100% reliable, but the above results lead us to conclude that serological testing is likely to be more reliable than the imputed genotypes in our dataset, which is not too surprising after all.

The sources of inaccuracies in the genotypes could be due to a variety of reasons, such as array typing errors and human factors such as misplaced samples and imputation errors since it is a statistical inference technique. The low balanced accuracy of 96% achieved for C^W is likely due to the genomic region being difficult to impute accurately. This is backed up by the info-score of the C^W-causative variant being lower (0.91). Consequently, there is a high correlation between a lower info-score (Table 2) of the variant used and lower accuracy of the resulting genetic blood typing (Table 3). Another factor could be cases where blood type status is conferred by rare alternative genetic variants that due to their rarity are not part of our genotypes.

The source of inaccuracies in serological results could be experimental or clerical human error in the laboratory or inherent inaccuracies in the antibody-based hemagglutination test being used at each individual blood bank.

Given the above uncertainty and the general complexity of the human genome, the desired accuracy of 100% in any given method for all donors is probably not achievable.

5.3 | Genetic confirmation of inferred Vel blood types

The Vel blood group was an ideal candidate for testing the feasibility of finding rare donors using our existing genotypes. The Vel-negative blood type is rare, found only in approximately 1 in 1276 Danish donors, according to random PCR testing performed by the Capital Region and close to the frequency reported from Southern Sweden.[19] Since the small deletion that confers Vel status[19] was not on the GSA chip, it had to be imputed. The low info-score of the imputed variant (Table 2) can be attributed to the low MAF of the rs566629828 variant (0.0221) rendering accurate imputation difficult.

Due to lack of commercially available reagents approved by regulatory authorities, serological typing of Vel status has remained largely infeasible outside of using in-house reagents based on patient-derived antisera. Besides, recently developed monoclonal antibodies are not yet available for diagnostic use.[33,34] As such, apart from genetic screening efforts, Vel typing is not normally performed in blood banks.

Currently, only one of five regions in Denmark (the Capital Region) performs blood typing tests for Vel status by using PCR probes specific for the SMIM1 deletion. At the time of writing, 19,142 persons had been randomly tested for Vel status using this method, resulting in the identification of 15 individuals with the Vel-negative blood type.

However, by utilizing our genotyped cohort, we managed to identify 75 potential Vel negatives, five of whom had already been confirmed via the aforementioned PCR probes by Capitol Region. We tested the viability of using genetically predicted blood types to narrow down the search for rare blood type donors. We used PCR techniques to confirm the inferred genotype of the remaining potential Vel negatives with available frozen blood samples. Despite the relatively low info-score of 0.87 for rs566629828, PCR confirmed 64 of 70 (NPV: 91.4%) potential Vel negatives to be true negatives (Table 3). In order to identify 64 Vel-negative donors, the Capital Region would have had to genotype over 80,000 persons at random, when we achieved this based on already existing genotypes while performing only 70 confirmations by PCR. This remarkable increase in efficiency proves the viability of using this method to locate donors with rare blood types such as Vel negatives.

5.4 | Strengths and limitations

The main strength of this study is the high accuracy of genetic blood types derived from our large cohort dataset at very limited cost. However, some shortcomings were identified during the course of the project. Our genotypes proved insufficient for blood types that are determined by more than one SNV or are determined by more complex genetics. As such, attempts to genetically predict the blood types Le^a/Le^b in the Lewis system, RhD in the Rh system, and A-subtypes (A_1 vs. A_2) in the ABO system failed. One of the possible reasons for this failure was our genotype datasets lacking phasing information. When two heterozygous causative SNVs are in partial linkage disequilibrium, phased genotype data that identify the DNA strand placement of each genotype are needed to resolve ambiguity in interpreting the resulting phenotype. In addition, the RhD blood group could not be typed due to its complicated genetic underpinning[35] and the absence of several important variants in our dataset.

The other shortcoming of our genetic dataset is the lack of variants with lower than 0.01 MAF. This prevents prediction of low- or high-prevalence antigens with known SNVs (Cr^a, Dh^a, Di^a, Erik, Hy, Jo^a, Js^a/Js^b, LW^a/LW^b, Mit, Mur, Sc1, Ul^a, U, Wr^a/Wr^b, Wu, etc.). Due to the statistical nature of the imputation process, even if present, many of these variants would likely have had low info-scores. The low info-scores would likely have resulted in unreliable genotypes, with an unknown exact degree of accuracy due to lack of serological results to compare them to. However, as demonstrated by Vel, even relatively unreliable predictions of rare antigen types can prove useful for greatly narrowing down the pool of potential blood donors with certain rare, desired blood types. Antigens with a very skewed prevalence can in rare cases cause hemolytic disease[36–40] and can therefore be of interest in rare blood type allocation.

It is worth mentioning that the antigen rates in our DBDS cohort might differ slightly from the general population as blood donors might not perfectly reflect the whole population. This is demonstrated in the differences seen in the ABO blood group (Table 4) due to specific blood bank preferences for group O blood.[30]

It should also be noted that the genetic predictions used in this project are somewhat specific to a European genetic cohort. Other genetic cohorts such as Asian and African have other blood type variants specific for or more common in their population, which must be accounted for to achieve highly accurate genetic blood typing. This could also mean that blood donors of other ethnic backgrounds than the majority of a cohort like ours may suffer lower accuracy due to the reasons discussed above.

6 | CONCLUSION

The very high accuracy of genetic prediction in selected blood types supports the feasibility of applying this technique to the existing DBDS genetic cohort and potentially blood donor cohorts in other countries with minor methodological modifications. Despite the high accuracy of genetic typing in the clinically important blood group systems ABO and Rh (C/c and E/e), they may fall short of the rigorous requirements needed to supplant traditional serological tests for the purpose of blood transfusion, even if serology is not fool-proof either. However, the accuracy is nevertheless adequate to generate blood type data for other uses such as narrowing down the number of candidate donors with a specific or rare blood type and performing blood type-related correlation studies.

Since genetic data are available for over 180,000 persons, which represent about 3% of the Danish population,[41] antigen ratios derived from this cohort can be regarded as a good approximation for the entirety of the Danish population. Calculated Danish blood group antigen rates are made available here, while the generated antigen status for the 90,000 genotyped DBDS donors is being made available for use by DBDS researchers and regional blood bank staff to aid in future research and blood bank strategies.

ACKNOWLEDGMENTS

The authors thank the Danish Blood Donor Study for making data available for the research performed in this publication and Novo Nordisk Foundation for their grants (nnf17oc0027594 and nnf14cc0001) that funded major parts of the infrastructure for DBDS and CHB. We also thank Bloddonorernes forskningsfond and A.P. Møller Fonden for awarding this project research grants of 64931 DKK and 45000 DKK, respectively.

CONFLICT OF INTEREST STATEMENT

MLO and JS are inventors on patents about Vel blood group genotyping and own 50% each of the shares in BLUsang AB, an incorporated consulting firm that receives royalties for said patents. They are both coauthors of AABB books and members of the Transfusion editorial board. MD received consultant fee for helping with a study tracking COVID-19 infection among Falck Health Care Workers, which has no connection to the present study. CM received grants of 64,931kr and 45,000kr from Bloddonornes forskningsfond and A.-P. Møller Fonden, respectively. All other authors have disclosed no conflicts of interest.

DATA AVAILABILITY STATEMENT

The Python code for the blood typing program (VladTyper) used to generate predicted blood types is available on Github.[42]

ETHICS STATEMENT

DBDS cohort: Scientific ethics approval number: NVK-1700407. Data protection approval number: P-2019-99. *CHB cohort*: Scientific ethics approval number: NVK-1808571. Data protection approval number: P-2019-243.

ORCID

Camous Moslemi https://orcid.org/0000-0001-7905-9774
Thorsten Brodersen https://orcid.org/0000-0003-4431-9972
Maria Didriksen https://orcid.org/0000-0002-4856-496X
Joseph Dowsett https://orcid.org/0000-0001-5381-2633
Jill R. Storry https://orcid.org/0000-0003-2940-2604
Martin L. Olsson https://orcid.org/0000-0003-1647-9610

REFERENCES

1. ISBT. Red cell immunogenetics and blood group terminology, ISBT Working Party. 2023. Available from: https://www.isbtweb.org/isbt-working-parties/rcibgt.html
2. Noizat-Pirenne F. Relevance of alloimmunization in haemolytic transfusion reaction in sickle cell disease. Transfus Clin Biol. 2012;19:132–8.
3. Muniz JG, Arnoni C, Medeiros R, Vendrame T, Cortez A, S Afonso J, et al. Antigen matching for transfusion support in Brazilian female patients with sickle cell disease to reduce RBC alloimmunization. Transfusion. 2021;61:2458–67.
4. Reid ME, Lomas-Francis C, Olsson ML. The blood group antigen FactsBook. Amsterdam: Elsevier Ltd; 2012.
5. Daniels G. The molecular genetics of blood group polymorphism. Transpl Immunol. 2005;14:143–53.
6. Westhoff CM. Blood group genotyping. Blood. 2019;133:1814–20.
7. Finning K, Bhandari R, Sellers F, Revelli N, Villa MA, Muñiz-Díaz E, et al. Evaluation of red blood cell and platelet antigen genotyping platforms (ID CORE XT/ID HPA XT) in routine clinical practice. Blood Transfus. 2016;14:160–7.
8. Paris S, Rigal D, Barlet V, Verdier M, Coudurier N, Bailly P, et al. Flexible automated platform for blood group genotyping on DNA microarrays. J Mol Diagn. 2014;16:335–42.
9. Pedersen OB, Erikstrup C, Kotzé SR, Sørensen E, Petersen MS, Grau K, et al. The Danish Blood Donor Study: a large, prospective cohort and biobank for medical research. Vox Sang. 2012;102:271.
10. Erikstrup C, Sørensen E, Nielsen KR, Bruun MT, Petersen MS, Rostgaard K, et al. Cohort profile: the Danish blood donor study. Int J Epidemiol. 2022;dyac194:e162–71. https://doi.org/10.1093/ije/dyac194
11. COMPANY. deCODE genetics. 2012. Available from: https://www.decode.com/company/
12. Hansen TF, Banasik K, Erikstrup C, Pedersen OB, Westergaard D, Chmura PJ, et al. DBDS Genomic Cohort, a prospective and comprehensive resource for integrative and temporal analysis of genetic, environmental and lifestyle factors affecting health of blood donors. BMJ Open. 2019;9:e028401.
13. Lane WJ, Westhoff CM, Gleadall NS, Aguad M, Smeland-Wagman R, Vege S, et al. Automated typing of red blood cell and platelet antigens from whole genome sequencing. Lancet Haematol. 2018;5:e241–51.
14. Guo Y, Busch MP, Seielstad M, Endres-Dighe S, Westhoff CM, Keating B, et al. Development and evaluation of a transfusion medicine genome wide genotyping array. Transfusion. 2019;59:101–11.
15. Krog GR, Rieneck K, Clausen FB, Steffensen R, Dziegiel MH. Blood group genotyping of blood donors: validation of a highly accurate routine method. Transfusion. 2019;59:3264–74.
16. Sørensen E, Christiansen L, Wilkowski B, Larsen MH, Burgdorf KS, Thørner LW, et al. Data resource profile: the Copenhagen Hospital Biobank (CHB). Int J Epidemiol. 2021;50:719–720e.
17. Lin P, Hartz SM, Zhang Z, Saccone SF, Wang J, Tischfield JA, et al. A new statistic to evaluate imputation reliability. PloS One. 2010;5:e9697.
18. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci U S A. 1991;88:7276–80.
19. Storry JR, Jöud M, Christophersen MK, Thuresson B, Åkerström B, Sojka BN, et al. Homozygosity for a null allele of SMIM1 defines the Vel-negative blood group phenotype. Nat Genet. 2013;45:537–41.
20. Reid ME, Denomme GA. DNA-based methods in the immunohematology reference laboratory. Transfus Apher Sci. 2011;44:65–72.
21. Home–SNP–NCBI. Available from: https://www.ncbi.nlm.nih.gov/snp/
22. Pearson KX. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Lond Edinb Dublin Philos Mag J Sci. 1900;50:157–75.
23. Newman JR. The world of mathematics. North Chelmsford, Massachusetts: Courier Corporation; 2000.
24. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. In: *2010 20th International Conference on Pattern Recognition*. IEEE; 2010. p. 3121–4.
25. Blair DC. Information retrieval, 2nd ed. C.J. Van Rijsbergen. London: Butterworths; 1979: 208 pp. Price: $32.50. J Am Soc Inf Sci. 1979;30:374–5.
26. Sasaki Y. The truth of the F-measure. Manchester: Teach Tutor Mater; 2007.
27. Yeh C-C, Chang CJ, Twu YC, Hung ST, Tsai YJ, Liao JC, et al. The differential expression of the blood group P1-A4GALT and P2-A4GALT alleles is stimulated by the transcription factor early growth response 1. Transfusion. 2018;58:1054–64.
28. Dayananda KK, Achur RN, Gowda DC. Epidemiology, drug resistance, and pathophysiology of *Plasmodium vivax* malaria. J Vector Borne Dis. 2018;55:1–8.
29. Popovici J, Roesch C, Rougeron V. The enigmatic mechanisms by which *Plasmodium vivax* infects Duffy-negative individuals. PLoS Pathog. 2020;16:e1008258.
30. Golding J, Northstone K, Miller LL, Davey Smith G, Pembrey M. Differences between blood donors and a population sample: implications for case-control studies. Int J Epidemiol. 2013;42:1145–56.
31. Zhang H, Mooney CJ, Reilly MP. ABO blood groups and cardiovascular diseases. Int J Vasc Med. 2012;2012:641917.
32. Capuzzo E, Bonfanti C, Frattini F, Montorsi P, Turdo R, Previdi MG, et al. The relationship between ABO blood group and cardiovascular disease: results from the Cardiorisk program. Ann Transl Med. 2016;4:189.
33. van der Rijst MVE, Lissenberg-Thunnissen SN, Ligthart PC, Visser R, Jongerius JM, Voorn L, et al. Development of a recombinant anti-Vel immunoglobulin M to identify Vel-negative donors. Transfusion. 2019;59:1359–66.
34. Danger Y, Danard S, Gringoire V, Peyrard T, Riou P, Semana G, et al. Characterization of a new human monoclonal antibody directed against the Vel antigen. Vox Sang. 2016;110:172–8.
35. Ying Y, Zhang J, Hong X, Xu X, He J, Zhu F. The significance of RHD genotyping and characteristic analysis in Chinese RhD variant individuals. Front Immunol. 2021;12:755661.

36. Anderson C, Hunter J, Zipursky A, Lewis M, Chown B. An antibody defining a new blood group antigen, Bu-a. Transfusion. 1963;3:30–3.
37. Seyfried H, Frankowska K, Giles CM. Further examples of anti-bu-a found in immunized donors. Vox Sang. 1966;11:512–6.
38. Squires A, Nasef N, Lin Y, Callum J, Khadawardi EM, Drolet C, et al. Hemolytic disease of the newborn caused by anti-Wright (anti-Wra): case report and review of literature. Neonatal Netw. 2012;31:69–80.
39. Moise KJ, Morales Y, Bertholf MF, Rossmann SN, Bai Y. Anti-Vel alloimmunization and severe hemolytic disease of the fetus and newborn. Immunohematology. 2019;33:152–4.
40. Scharberg EA, Wieckhusen C, Luz B, Rothenberger S, Stürzel A, Rink G, et al. Fatal hemolytic disease of the newborn caused by an antibody to KEAL, a new low-prevalence Kell blood group antigen. Transfusion. 2017;57:217–8.
41. Danmarks Statistik Statistics Denmark. Danmarks Statistik. Available from: https://www.dst.dk/en/Statistik/emner/befolkning-og-valg/befolkning-og-befolkningsfremskrivning
42. kamiboy/VladTyper: A program for predicting blood types using bed/bim/fam genotypes. Available from: https://github.com/kamiboy/VladTyper

How to cite this article: Moslemi C, Sækmose SG, Larsen R, Bay JT, Brodersen T, Didriksen M, et al. Genetic prediction of 33 blood group phenotypes using an existing genotype dataset. Transfusion. 2023;63(12):2297–310. https://doi.org/10.1111/trf.17575