Association between ABO haplotypes and the risk of venous thrombosis: impact on disease risk estimation

Goumidi, L.; Thibord, F.; Wiggins, K.L.; Li-Gao, R.F.; Brown, M.R.; Vlieg, A.V.; ... ; Morange, P.E.

Note: To cite this publication please use the final published version (if applicable).

Regular Article

THROMBOSIS AND HEMOSTASIS

Association between ABO haplotypes and the risk of venous thrombosis: impact on disease risk estimation

Louisa Goumidi,[1] Florian Thibord,[2-5] Kerri L. Wiggins,[6] Ruifang Li-Gao,[7] Mickael R. Brown,[8] Astrid van Hylckama Vlieg,[7] Joan-Carles Souto,[9,10] José-Manuel Soria,[11] Manal Ibrahim-Kosta,[1,12] Noémie Saut,[1,12] Delphine Daian,[3,13] Robert Olaso,[3,13] Philippe Amouyel,[14,15] Stéphanie Debette,[2,16] Anne Boland,[3,13] Pascal Bailly,[17,18] Alanna C. Morrison,[8] Denis O. Mook-Kanamori,[7] Jean-François Deleuze,[3,13,19] Andrew Johnson,[4,5] Paul S. de Vries,[8] Maria Sabater-Lleal,[20,21] Jacques Chiaroni,[17,18] Nicholas L. Smith,[22-24] Frits R. Rosendaal,[7] Daniel I. Chasman,[25,26] David-Alexandre Trégouët,[2,3,*] and Pierre-Emmanuel Morange[1,3,12,*]

[1]Aix Marseille University, INSERM, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), Centre de Recherche en CardioVasculaire et Nutrition, Marseille, France; [2]INSERM U1219, Bordeaux Population Health Research Center, University of Bordeaux, Bordeaux, France; [3]Laboratory of Excellence (LabEx) Genomique Médicale, Evry, France; [4]Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Framingham, MA; [5]The Framingham Heart Study, Framingham, MA; [6]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA; [7]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; [8]Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; [9]Thrombosis and Hemostasis Research Group, Sant Pau Institute of Biomedical Research (IIB Sant Pau), Barcelona, Spain; [10]Unit of Hemostasis and Thrombosis, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; [11]Unit of Genomic of Complex Disease, Institut de Recerca Hospital de la Sant Creu i Sant Pau, IIB Sant Pau, Barcelona, Spain; [12]Hematology Laboratory, La Timone University Hospital of Marseille, Marseille, France; [13]Université Paris-Saclay, Commissariat à l'Energie Atomique, Centre National de Recherche en Génomique Humaine, Evry, France; [14]Lille University, INSERM, Institut Pasteur de Lille, Facteurs de Risque et Déterminants Moléculaires des Maladies Liées au Vieillissement (RID-AGE), LabEx Development of Innovative Strategies for a Trans-disciplinary Approach to Alzheimer's Disease (DISTALZ), Lille, France; [15]Lille University, INSERM, Centre Hospitalier Universitaire (CHU) Lille, Institut Pasteur de Lille, RID-AGE, Lille, France; [16]Department of Neurology, CHU de Bordeaux, Bordeaux, France; [17]Etablissement Français du Sang Provence-Alpes-Côte d'Azur–Corse "Biologie des Groupes Sanguins," Marseille, France; [18]Aix Marseille University, Etablissement Français du Sang, Centre National pour la Recherche Scientifique, Anthropologie Bio-Culturelle, Droit, Ethique et Santé, "Biologie des Groupes Sanguins," Marseille, France; [19]Centre d'Etude du Polymorphisme Humain, Fondation Jean Dausset, Paris, France; [20]Genomics of Complex Diseases, Research Institute of Hospital de la Santa Creu i Sant Pau, IIB Sant Pau, Barcelona, Spain; [21]Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden; [22]Department of Epidemiology, University of Washington, Seattle, WA; [23]Kaiser Permanente Washington Health Research Unit, Kaiser Permanente Washington, Seattle, WA; [24]Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Office of Research and Development, Seattle, WA; [25]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA; and [26]Department of Medicine, Harvard Medical School, Boston, MA

KEY POINTS

- **Four ABO SNPs (rs2519093, rs1053878, rs8176743, rs41302905) must be studied in any work assessing the risk of ABO locus on VT.**

- **Molecularly defined ABO blood groups are more accurate than serologically determined (A/B/O) groups for estimating risk of VT.**

Genetic risk score (GRS) analysis is a popular approach to derive individual risk prediction models for complex diseases. In venous thrombosis (VT), such type of analysis shall integrate information at the *ABO* blood group locus, which is one of the major susceptibility loci. However, there is no consensus about which single nucleotide polymorphisms (SNPs) must be investigated when properly assessing association between *ABO* locus and VT risk. Using comprehensive haplotype analyses of ABO blood group tagging SNPs in 5425 cases and 8445 controls from 6 studies, we demonstrate that using only rs8176719 (tagging O1) to correctly assess the impact of *ABO* locus on VT risk is suboptimal, because 5% of rs8176719-delG carriers do not have an increased risk of developing VT. Instead, we recommend the use of 4 SNPs, rs2519093 (tagging A1), rs1053878 (A2), rs8176743 (B), and rs41302905 (O2), when assessing the impact of *ABO* locus on VT risk to avoid any risk misestimation. Compared with the O1 haplotype, the A2 haplotype is associated with a modest increase in VT risk (odds ratio, ~1.2), the A1 and B haplotypes are associated with an ~1.8-fold increased risk, whereas the O2 haplotype tends to be slightly protective (odds ratio, ~0.80). In addition, although the A1 and B blood groups are associated with increased von Willebrand factor and factor VIII plasma levels, only the A1 blood group is associated with ICAM levels, but in an opposite direction, leaving additional avenues to be explored to fully understand the spectrum of biological effects mediated by *ABO* locus on cardiovascular traits. (*Blood*. 2021;137(17):2394-2402)

Introduction

The ABO blood groups are determined by the *ABO* locus located on chromosome 9 (9q34.1-q34.2) that codes for 2 gly-cosyltransferases (GTs), A and B; their roles are to transfer a saccharide unit to polypeptides and membrane glycolipids [ABO(H) carbohydrate structures]. ABO blood group was one of the first identified risk factor for venous thrombosis (VT),[1] and its thrombotic effect is considered to be mediated by the ABO(H)

structures, which are expressed on the glycans of factor VIII (FVIII) and von Willebrand factor (VWF),[2] as well as by modulation of soluble ICAM-1, soluble P-selectin, and soluble E-selectin levels.[3]

Despite the robust association between the ABO locus and VT risk, there is a lack of consensus about the proper way to accurately estimate its impact on disease risk at the population or individual level. Although some investigators believe that individuals with A1 or B ABO blood groups are at an ~1.5-fold increased risk for VT compared with individuals with O or A2 blood groups who have little or no A glycosyltransferase (AGT) activity,[4-6] others investigators[7-12] propose to contrast O and non-O blood groups without further granularity. Some have even proposed that, compared with O, the A2 blood group is associated with increased VT risk but to a lesser extent than A1 and B.[13] Clarifying this association is required to provide more accurate risk predictions. Several explanations could be proposed to explain the aforementioned inconsistency, including standard epidemiological issues, such as population heterogeneity and limited sample size. However, another important aspect is that the ABO blood group can be phenotypically assessed using serological tests or genetically defined using single nucleotide polymorphisms (SNPs) at the ABO locus. From an evolutionary perspective, the ancestral ABO blood group is A1.[14] The A2 blood group resulted from a proline to leucine substitution at amino acid 156 and is defined by the rs1053878-A allele. The O1 group that accounts for 95% of O group is the consequence of a frameshift deletion (c.del261G, p.88fs118Stop) and is defined by the rs8176719-delG allele, whereas the rs41302905-T allele (p.Arg268) corresponds to the O2 group.[15] Finally, the B blood type resulted from a glycine to serine substitution at amino acid 235 and is defined by the rs8176743-T allele. These 4 SNPs define 5 main common haplotypes in European ancestry populations (Figure 1); as the ancestry allele, the A1 group is generally called when none of the above alleles is present. However, it has been proposed to be tagged by the rs579459-G allele (or any SNP in perfect linkage disequilibrium [LD] with it).[6] The choice of studied SNPs and their groupings strongly influences the downstream association results. Indeed, because of complete negative LD between these SNPs (D′ = −1), the analysis of only a subset of these SNPs may introduce biases in the estimation of genetic association parameters. These biases can be even amplified when these ABO blood group SNPs are substituted by other SNPs expected to serve as good, but imperfect, proxies ($0.8 < r^2 < 1$). For example, rs514659[8] and rs687621[9] have been used as proxies for rs8176719 ($r^2 \sim 0.84$ for both).

This study was initially designed to fine map the association between ABO haplotypes and VT risk, to identify the optimal set of SNPs needed to tag the VT-associated ABO haplotypes, and to address the clinical usefulness of using such optimal set in the context of genetic risk estimation. To achieve these aims, a series of haplotype association analyses, taking into account the LD between ABO SNPs, was undertaken in relation to AGT activity, VT risk, and quantitative cardiovascular traits whose interindividual variability was demonstrated to be strongly influenced by the ABO locus.

Methods

Selection of ABO SNPs

Six polymorphisms located at the ABO locus were selected for our series of haplotype association analyses: the 4 SNPs that characterize main ABO blood groups, as discussed in the Introduction (rs1053878, rs8176719, rs41302905, rs8176743); rs579459, which has been proposed to tag the ABO A1 blood group; and rs2519093, which has been proposed to be a genetic risk factor for VT independently of the ABO blood groups.[16]

Studies contributing to the association with VT risk

The association between ABO haplotypes and VT risk was investigated in 6 case-control/prospective studies of VT risk which have been previously typed for genome-wide genotyped data: the Cardiovascular Health Study (CHS), the Early Onset Venous Thrombosis study (EOVT), the Heart and Vascular Health (HVH), the Marseille Thrombosis Association study (MARTHA), the Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis (MEGA) study, studies that were described by Germain et al,[17] and the RETROVE study.[18] Briefly, CHS is a United States population–based cohort of adults aged older than 65 years that recruited patients in 4 centers. HVH is a population-based case-control study of cardiovascular events that was conducted in Washington state. EOVT and MARTHA are 2 French case-control studies for VT, and MEGA and RETROVE are case-control studies from The Netherlands and Spain, respectively. A detailed description of each study is provided in supplemental Data (available on the Blood Web site). In total, 5425 VT cases and 8445 controls were assembled through these 6 studies.

Studies contributing to analyses of cardiovascular quantitative traits

ABO haplotypes were tested for their association with 4 quantitative traits that are well known to be modulated by ABO blood groups: (1) AGT activity was available for 213 VT patients from the MARTHA study[19]; (2) plasma levels of VWF were assessed in 13 527 individuals from 4 studies, including the Atherosclerosis Risk in Communities (ARIC) study,[20] the Framingham Heart study,[21] MARTHA, and RETROVE; (3) plasma levels of FVIII were assessed in 19 085 individuals from 5 studies, including ARIC, CHS, MARTHA, The Netherlands Epidemiology of Obesity study,[22,23] and RETROVE; and (4) plasma levels of ICAM-1 were available for 22 639 participants in the Women's Genome Health Study.[24]

For these analyses on quantitative traits, only control data from RETROVE participants were used. By contrast, only data from VT patients were available in MARTHA. A detailed description of the study samples and the biological measurements are provided in supplemental Data.

Haplotype/diplotype association analyses

In all of these studies, haplotype association analyses were conducted using the haplo.stats R package[25] or THESIAS software.[26] From best-guessed genotypes obtained from imputed data, these 2 programs implement linear (for quantitative trait) and logistic (for disease status) regression models to estimate haplotype effects by comparison with a reference haplotype (taken here as the O1 haplotype) under the assumption of additive effects. Analyses were adjusted for cohort-specific covariates (supplemental Data).
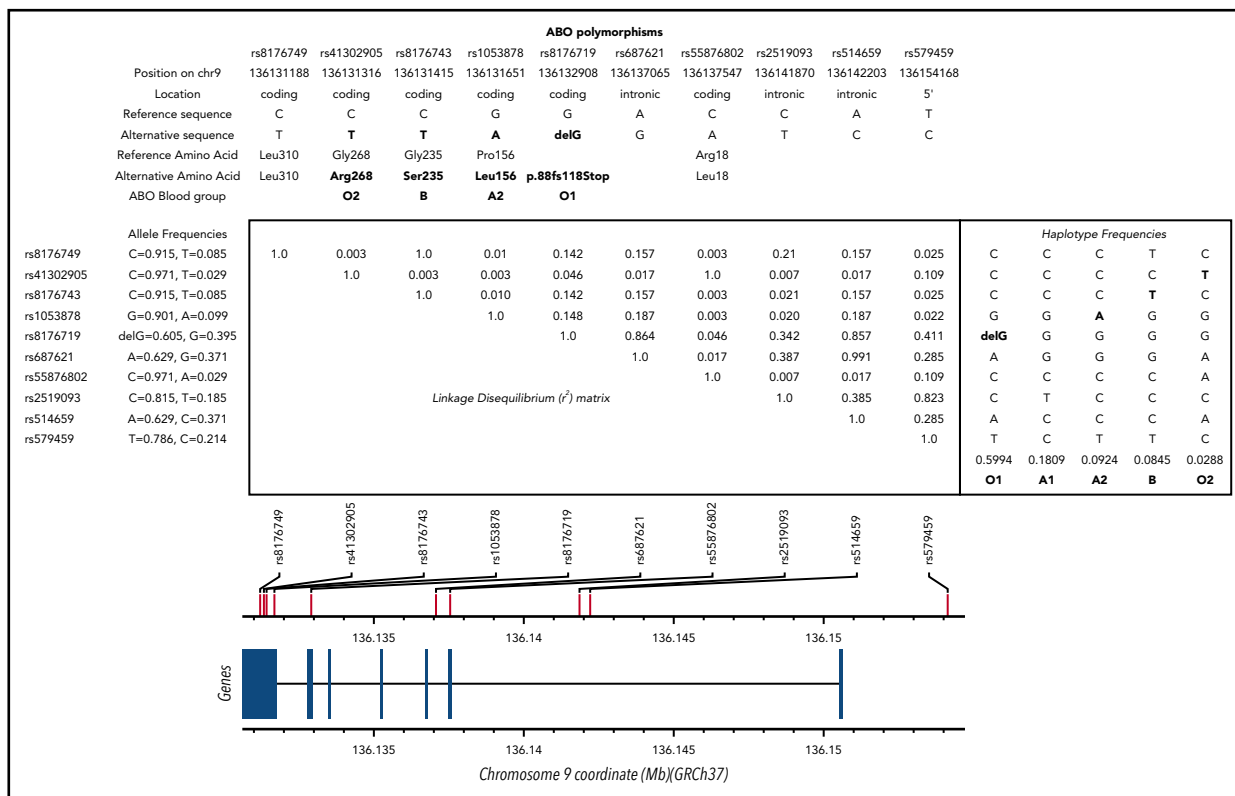
ABO polymorphisms

	rs8176749	rs41302905	rs8176743	rs1053878	rs8176719	rs687621	rs55876802	rs2519093	rs514659	rs579459
Position on chr9	136131188	136131316	136131415	136131651	136132908	136137065	136137547	136141870	136142203	136154168
Location	coding	coding	coding	coding	coding	intronic	coding	intronic	intronic	5'
Reference sequence	C	C	C	G	G	A	C	C	A	T
Alternative sequence	T	**T**	**T**	**A**	**delG**	G	A	T	C	C
Reference Amino Acid	Leu310	Gly268	Gly235	Pro156			Arg18			
Alternative Amino Acid	Leu310	**Arg268**	**Ser235**	**Leu156**	**p.88fs118Stop**		Leu18			
ABO Blood group		**O2**	**B**	**A2**	**O1**					

Linkage Disequilibrium (r^2) matrix

Allele Frequencies	rs8176749	rs41302905	rs8176743	rs1053878	rs8176719	rs687621	rs55876802	rs2519093	rs514659	rs579459
rs8176749 C=0.915, T=0.085	1.0	0.003	1.0	0.01	0.142	0.157	0.003	0.21	0.157	0.025
rs41302905 C=0.971, T=0.029		1.0	0.003	0.003	0.046	0.017	1.0	0.007	0.017	0.109
rs8176743 C=0.915, T=0.085			1.0	0.010	0.142	0.157	0.003	0.021	0.157	0.025
rs1053878 G=0.901, A=0.099				1.0	0.148	0.187	0.003	0.020	0.187	0.022
rs8176719 delG=0.605, G=0.395					1.0	0.864	0.046	0.342	0.857	0.411
rs687621 A=0.629, G=0.371						1.0	0.017	0.387	0.991	0.285
rs55876802 C=0.971, A=0.029							1.0	0.007	0.017	0.109
rs2519093 C=0.815, T=0.185								1.0	0.385	0.823
rs514659 A=0.629, C=0.371									1.0	0.285
rs579459 T=0.786, C=0.214										1.0

Haplotype Frequencies

rs8176749	C	C	C	T	C
rs41302905	C	C	C	C	**T**
rs8176743	C	C	C	**T**	C
rs1053878	G	G	**A**	G	G
rs8176719	**delG**	G	G	G	G
rs687621	A	G	G	G	A
rs55876802	C	C	C	C	A
rs2519093	C	T	C	C	A
rs514659	A	C	C	C	A
rs579459	T	C	T	T	C
	0.5994	0.1809	0.0924	0.0845	0.0288
	O1	**A1**	**A2**	**B**	**O2**

Figure 1. LD matrix and haplotype frequencies at the ABO locus. SNPs shown are discussed in the article (and in supplemental Data). Numerical values in the upper right triangle of the matrix are pairwise LD r^2 values. Values for r^2 and allele/haplotype frequencies were estimated from European populations available at https://ldlink.nci.nih.gov/?tab=ldmatrix.

Because of the very strong negative LD between studied ABO SNPs (D' ~ 1) and the resulting haplotype structure, it is possible to determine the pair of haplotypes (called diplotypes) carried by each individual with little ambiguity. The percentages of individuals without any ambiguous haplotypes/diplotypes were 0.995, 0.997, 0.989, 0.992, 0.999. and 0.996 in CHS, EOVT, HVH, MARTHA, MEGA, and RETROVE, respectively. Therefore, to remove the assumption of additivity in haplotype effects, we tested the association between ABO diplotypes and VT risk using a logistic regression model adjusted for the same covariates as for the haplotype analyses.

Haplotype/diplotype association results were meta-analyzed using a fixed-effects model based on the inverse-variance weighting, and the heterogeneity of associations across studies was assessed with the Cochran-Mantel-Haenszel test statistic.[27]

Genetic risk score and risk prediction analyses

Based on the haplotypic and diplotypic odds ratios (ORs) for VT derived from the meta-analyses of the 6 studies, 2 genetic risk scores (GRSs) were computed. A first GRS assuming additive effects of haplotypes (GRSh) is defined, for each individual i, as $GRSh_i = \log(OR_{hi1}) + \log(OR_{hi2})$, where (h_{i1}, h_{i2}) is the pair of haplotypes carried by individual i, and OR_k is the haplotypic OR associated with haplotype k. A second GRS (GRSd), which does not assume additivity of haplotype effects, is defined from derived diplotypic ORs as $GRSd_i = \log(OR_{hi1,hi2})$, where $OR_{k,l}$ is the OR for VT associated with the diplotype (k,l). We then assessed the discriminative properties of these GRSs by comparison with a simple GRS defined by the number of O1 alleles carried by an

individual. GRSs were calculated in an independent sample of 18 595 VT cases and 92 983 randomly selected controls, frequently matched to cases by age at the baseline examination, from the UK Biobank (UKBB) cohort (supplemental Data). We used this independent sample to minimize the risk of overfitting when GRSs are evaluated in the cohorts from which they were derived. The association between GRS and VT risk was assessed using a logistic regression analysis, adjusting for sex, age at the time of VT event, grouped year at the time of VT event, and the 10 first principal components derived from genome-wide genotype data. Of note, the percentage of UKBB participants with unambiguous haplotypes/diplotypes was 0.992. GRSs were compared in terms of area under the curve (AUC), continuous net Reclassification Index, and integrated discrimination improvement (IDI). These metrics were calculated using the PredictABEL R program (https://cran.r-project.org/web/packages/PredictABEL/PredictABEL.pdf).

Results

As a first step, we assessed the association between ABO haplotypes and AGT activity, the main GT encoded by the A1 or A2 allele, keeping in mind that, by definition, AGT shall be quantified only in A1 or A2 carriers. As shown in Table 1, the 6 studied ABO SNPs generated 5 common haplotypes. The rs2519093-T, rs8176719-delG, rs1053878-A, rs8176743-T, and rs41302905-T alleles were each carried by a unique haplotype, whereas the rare C allele at rs579459 was carried by 2 distinct haplotypes. Two haplotypes demonstrated significantly increased AGT activity compared with the O1 tagging haplotype: the

Table 1. Haplotype association analysis of ABO SNPs with A GT activity in 213 VT patients from the MARTHA cohort

ABO blood group	rs2519093	rs8176719	rs579459	rs1053878	rs8176743	rs41302905	Haplotype frequencies	Haplotype additive effects (95% CI)
O1	C	delG	T	G	C	C	0.528	Reference
O2	C	G	C	G	C	**T**	0.014	$\beta = -249.54$ (−3115.8 to 2616.8); $P = .864$
A1	**T**	G	C	G	C	C	0.357	$\beta = 7983.2$ (7385.6-8580.8); $P = 4.5 \times 10^{-150}$
A2	C	G	T	**A**	C	C	0.046	$\beta = 1655.6$ (267.2-3044.0); $P = .019$
B	C	G	T	G	**T**	C	0.033	$\beta = -239.2$ (−1475.0 to 996.6); $P = .704$

The analysis was conducted using THESIAS and adjusted for age, sex, and the first 4 principal components. Minor alleles are shown in bold.

haplotype carrying the rs1053878-A allele (ie, A2) with a modest AGT increase ($\beta = 1655.6 \pm 708.36$; $P = .019$) and the haplotype carrying the rs2519093-T allele associated with strong AGT modulation ($\beta = 7983.2 \pm 304.9$; $P = 4.5 \times 10^{-150}$). By contrast, the haplotype carrying the rs579459-C allele and the rs41302905-T allele was not associated with AGT ($P = .864$). These observations demonstrate that (1) the rs579459-C allele generally used to tag the A1 blood type is an imperfect genetic marker for the ABO A1 blood group because it also captures O2 carriers and, in that case, is not associated with AGT and (2) the rs2519093 variant is a better marker of the A1 blood group.

We then assessed the association between *ABO* haplotypes and VT risk. Detailed haplotype association results for VT obtained in each participating study are provided in supplemental Tables 1-6. A Forest plot reporting the haplotypic ORs in each individual study is shown in Figure 2. Results indicate that, compared with the O1 haplotype, the rs2519093-T and rs8176743-T alleles were associated with an increased risk for VT (OR, 1.78; 95% confidence interval [CI], 1.67-1.92; $P = 3.15 \times 10^{-60}$ and OR, 1.76; 95% CI, 1.58-1.96; $P = 1.88 \times 10^{-24}$, respectively), with little heterogeneity across studies ($P = .136$ and $P = .018$, respectively). The rs1053878-A allele (*A2* allele) also showed a trend toward a moderate increased risk for VT (OR, 1.16; 95% CI, 1.03-1.31; $P = .013$ ($P = 0.738$ for heterogeneity) (Figure 2; Table 2). Interestingly, the *A2*-associated OR for VT was significantly ($P = 3.25 \times 10^{-10}$) lower than the common OR (OR, 1.78; 95% CI, 1.68-1.88) associated with the rs2519093-T and rs8176743-T alleles. Of note, of the 2 haplotypes carrying the rs579459-C allele, the 1 that also carries the rs2519093-T (A1) allele is associated with increased VT risk. By contrast, the second haplotype, which is also the sole haplotype carrying the rs41302905-T allele coding for O2, was associated with a decreased risk for VT (OR, 0.70; 95% CI, 0.57-0.86; $P = 7.75 \times 10^{-4}$) but with some evidence of statistical ($P = .001$) heterogeneity across studies.

The association between *ABO* diplotypes and VT is illustrated in Figure 3, where meta-analyses from the 6 contributing studies are shown. Diplotype association results per study are given in supplemental Table 7. Compared with the *O1O1* diplotype, diplotype ORs ranged from 0.81 (95% CI, 0.60-1.09; $P = .166$) for the *O1O2* diplotype to 3.45 (95% CI, 1.88-6.32; $P = 5.91 \times 10^{-5}$) for the *BO2* diplotype. This analysis illustrates the variability of risk estimates within a given serological group. For example, diplotype risk estimates were significantly different ($P = .002$) across individuals from the same A serological group (*A1A1*: OR, 2.98; 95% CI, 2.50-3.57; *A1A2*: OR, 2.65; 95% CI, 2.10-3.33; *A1O1*: OR, 2.12; 95% CI, 1.92- 2.35; and *A1O2*: OR, 1.68; 95% CI, 1.17-2.41).

From Table 2, it can be deduced that ~19% (0.07/0.37) of non-O1 individuals are A2 individuals at moderate risk for VT (OR, ~1.2), ~76% are individuals with A1 or B blood type at significantly greater risk for VT (OR, ~1.8), and the remaining 5% with O2 blood type are not exposed to a greater risk for disease. These observations suggest that the sole use of O1-tagging rs8176719 to estimate the individual risk of VT associated with ABO blood groups may lead to risk misclassification. To further assess whether integrating haplotype/diplotype *ABO* information could be more clinically relevant than using the sole information on *O1* polymorphism, we used the independent sample from
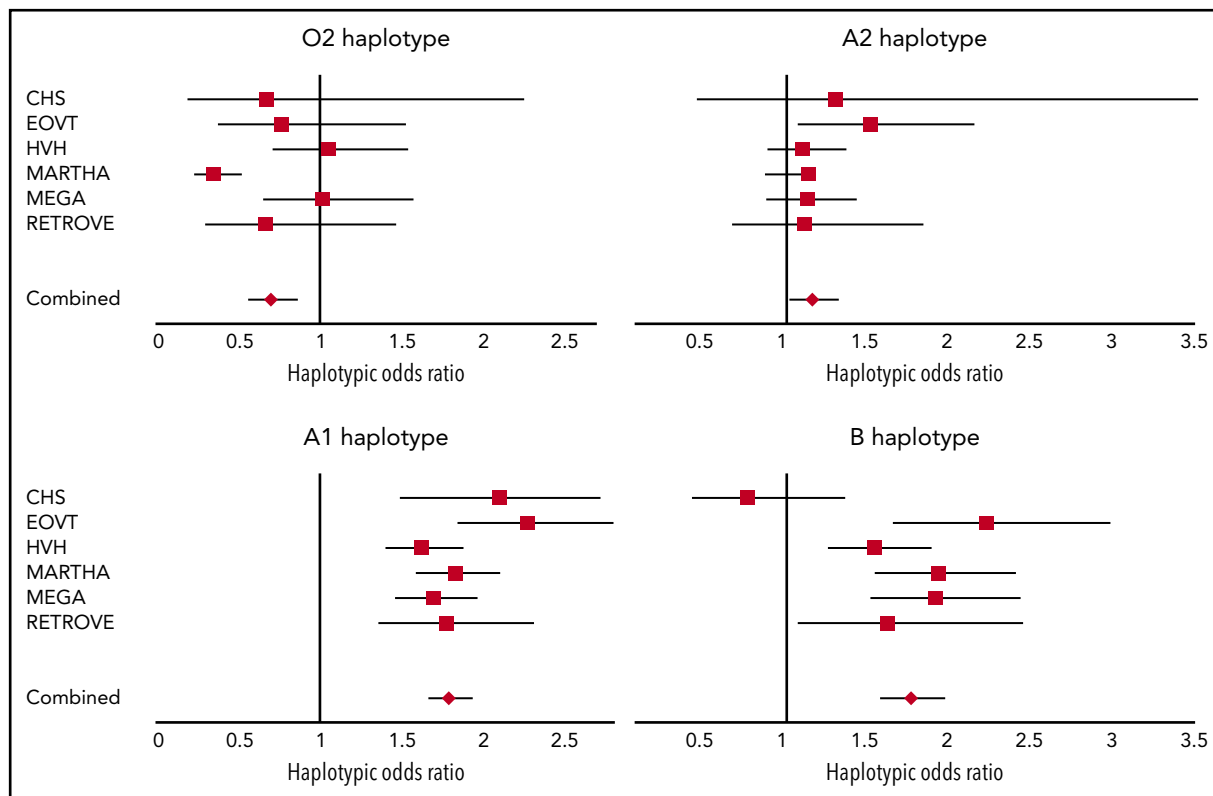
Figure 2. Forest plots showing the association between ABO haplotypes and VT risk. The figure shows the haplotypic OR for VT associated with the O2, A1, A2, and B haplotypes compared with the O1 haplotype taken as reference. ORs were computed under the assumption of haplotype effect additivity.

UKBB to evaluate the AUC of 3 GRSs: GRSh defined from haplotype effect estimates reported in Table 2, GRSd defined from diplotype effect estimates reported in Figure 3, and GRS_{O1} defined by the number of O1 alleles carried by individuals. Corresponding AUCs were 0.561 (95% CI, 0.557-0.566), 0.559 (95% CI, 0.555-0.564), and 0.553 (95% CI, 0.548-0.558), respectively. GRSh- and GRSd-associated AUCs were significantly higher ($P = 1.01 \times 10^{-9}$ and $P = 1.31 \times 10^{-5}$, respectively) than the AUC obtained using only the O1 polymorphism, but the improvement

remains modest. Similarly, the IDIs associated with GRSh and GRSd compared with GRS_{O1} were significantly different from 0 ($P < 10^{-10}$) but weak (IDI = 0.2% for both). The net Reclassification Index associated with GRSh was also modest (0.023; 95% CI, 0.011- 0.035; $P = 1.7 \times 10^{-4}$).

As a final step, we investigated how these observations extend to plasma levels of VWF (supplemental Tables 8-11), FVIII (supplemental Tables 12-16), and ICAM (supplemental Table 17). First, in
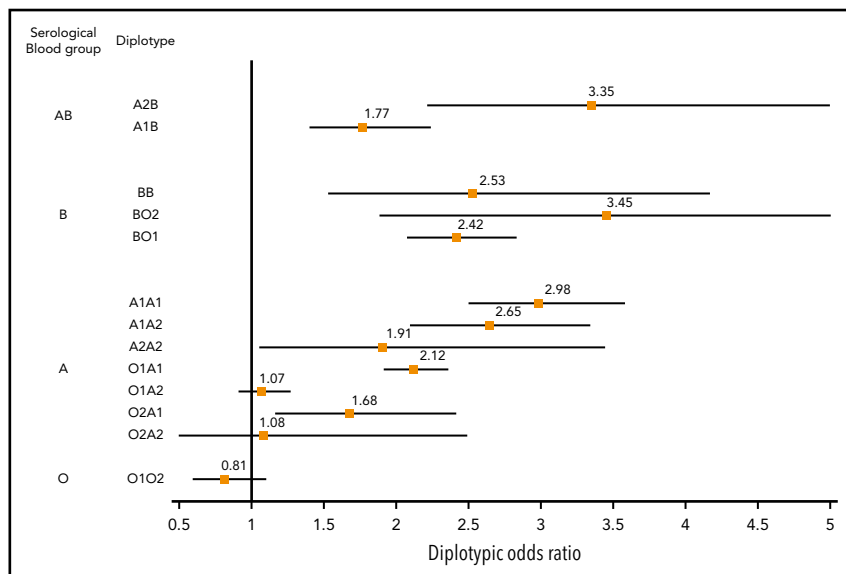
Figure 3. Forest plot showing the association between ABO diplotypes and VT risk. The O1O1 diplotype was used as reference. Diplotypic ORs were obtained from the meta-analysis of results observed in each of the 6 contributing studies (supplemental Table 7).

Table 2. Summary data for the association between ABO haplotypes and studied traits

ABO blood group	rs2519093	rs8176719	rs579459	rs1053878	rs8176743	rs41302905	Frequency*	OR for VT†	Impact on VWF/FVIII levels‡	Impact on ICAM1 levels‡
O1	C	delG	T	G	C	C	0.63	Reference	Reference	Reference
O2	C	G	C	G	C	**T**	0.02	0.70§ (0.57-0.86)	—	—
A1	**T**	G	C	G	C	C	0.20	1.78 (1.67-1.92)	↑↑↑	↓↓↓
A2	C	G	T	**A**	C	C	0.07	1.16 (1.03-1.31)	↑	↓
B	C	G	T	G	**T**	C	0.08	1.76 (1.58-1.96)	↑↑↑	—

Minor alleles are shown in bold.

—, no effect of a given haplotype on the studied trait.

*Haplotype frequencies in European ancestry populations.

†Haplotypic risk estimate (95% CI) for VT obtained from the meta-analysis of results listed in supplemental Tables 1-6.

‡Direction (↑= increase; ↓= decrease) of haplotype effects on plasma levels. Strong vs modest effects are illustrated by 3 arrows vs 1 arrow.

§This risk estimate exhibited statistical heterogeneity ($P = 1.2 \times 10^{-3}$) across the 6 investigated studies.

all studied samples, the same 5 common haplotypes were derived from the 6 studied SNPs. In all investigated studies, the effect of the *O2* allele on the studied phenotypes was not significantly different from that of the most frequent haplotype tagging the O1 blood group. Furthermore, both haplotypes carrying the rs2519093-T (*A1*) or rs8176743-T (*B*) allele were homogeneously associated with substantially increased VWF and FVIII levels; the rs1053878-A (*A2*) allele was also associated with increased VWF and FVIII levels but to a more modest degree. These results are completely in line with those observed for VT risk. Interestingly, the haplotypes tagging for the A1 or A2 blood group, but not the B group, were associated with ICAM-1 levels but in an opposite direction from that observed for their associations with VWF and FVIII levels (Table 2; supplemental Table 17).

Discussion

This work begins with the first haplotype analysis of *ABO* SNPs in relation to AGT activity. The modest sample size of this analysis was counterbalanced by the nature of the studied trait, AGT activity being the most proximal biological trait influenced by the *ABO* locus, and by the strength of the association between *ABO* haplotypes and AGT activity. *ABO* haplotypes explained ~80% of the variability in AGT activity, which was primarily due to the *A1* haplotype being associated with extremely increased AGT and, to a lesser extent, the *A2* haplotype being associated with a modest increase in AGT activity. Very interestingly, this haplotype analysis demonstrates that the rs579459 allele, which is often used to tag the A1 blood group, is an imperfect marker of the phenotypic A1 blood group. Indeed, ~10% (~0.02/0.22) of rs579459-C allele carriers are, in fact, O2 carriers and have AGT activity similar to the O1 blood group. By contrast, our analyses demonstrated that rs2519093, which was initially believed to be an independent *ABO* blood group polymorphism contributing to VT risk,[16] appears to be an optimal tag for the A1 group because it uniquely defines the haplotype associated with the highest AGT levels.

Capitalizing on 6 independent studies for VT risk including 5425 VT cases and 8445 controls, we then performed the first haplotype and diplotype analyses of *ABO* SNPs with VT. The haplotype analysis demonstrated that A1- and B-tagging haplotypes are at a similar risk for VT (OR ~ 1.8) and that the *A2* haplotype is also associated, but more modestly (OR ~ 1.2), with increased VT risk; the latter observation corroborated recent findings by Vasan et al.[13] The diplotype analysis suggested that there is substantial variability in disease risk among individuals who would have been assigned the same serological ABO phenotype (O/A/B/AB), providing support for defining the ABO blood group using genetic markers. This work also demonstrates that using only the *ABO* O1-tagging rs8176719 allele to address the contribution of the *ABO* locus to VT risk, as it has been proposed, is suboptimal. Indeed, ~76% of non-*O1* carriers are at higher risk (OR ~ 1.8) for VT, ~19% are at moderate risk (OR ~ 1.2), and ~5% have no risk. However, we showed in an independent sample of 18 595 cases and 92 983 noncases that a GRS based on *ABO* haplotypes adds marginal information compared with a GRS defined from the sole rs8176719 allele. This work is not aimed at encouraging the use of *ABO* information in clinical practice, especially because the use of

other VT-associated SNPs with stronger effects (such as FV Leiden) is still debated[28,29]; rather, it emphasizes the need to study haplotypes instead of only 1 SNP when assessing the impact of *ABO* locus on VT risk.

Finally, beyond these results pertaining to VT risk, we performed additional haplotype analyses on endothelial biomarkers that revealed some intriguing findings. The most likely hypothesis that is generally put forward to explain the relationship between VT and *ABO* is that, by modifying GT expression, *ABO* participates in controlling the degree of VWF glycosylation and, consequently, its clearance and/or cleavage by ADAMTS13.[30,31] This would explain the robust association found between the ABO blood group and circulating concentrations of VWF[32] and FVIII, of which VWF is the transport protein. However, the effect of the *ABO* locus on VT risk cannot be explained solely by a VWF-dependent mechanism. Indeed, it has been shown that the association between the ABO blood group and VT persists after adjusting for FVIII or VWF levels.[12] Although the pattern of the association between *ABO* haplotypes and VWF/FVIII parallels that observed with VT risk, our work provides strong evidence that *A1*- and *A2*-tagging haplotypes, but not the *B*-tagging haplotype, are associated with plasma levels of ICAM-1. Additionally, the directions of haplotype effects are opposite from those observed for VWF and FVIII, because A carriers displayed lower ICAM-1 levels. Unfortunately, VWF and FVIII levels were not available in participants measured for ICAM-1, which prevented us from conducting in-depth joint analyses of these biomarkers. Further work is needed to clarify this point, as well as to assess whether this phenomenon also holds for other endothelial biomarkers (eg, P-selectin, E-selectin) and cardiovascular diseases (eg, coronary artery disease, stroke) with an ABO-dependent component.

Additional limitations must be discussed. Our association analyses were conducted using imputed genotypes. Even if all studied SNPs were imputed with high accuracy (imputation $r^2 > 0.93$ in all studies for all SNPs), we cannot exclude that this introduced some extra variability in the estimated haplotype frequencies and the corresponding haplotype/diplotype effects. In addition, built upon data from genome-wide association studies, this work focused on common *ABO* SNPs and common ABO blood groups. As a consequence, very rare ABO blood groups, such as A3 and O3, were not addressed here. These rare groups are molecularly defined by extremely rare coding variations (https://www.presentica.com/doc/10230663/names-for-abo-isbt-001-blood-group-alleles-pdf-document) that would be investigated more efficiently using sequencing strategies. All of the analyses reported in this work were performed in individuals of European ancestry, in whom ABO blood group–tagging SNPs are in complete negative LD and the rs8176719-delG allele tagging *O1* is only carried by 1 haplotype. It would be very interesting to perform similar investigations in other populations in which LD may not be so complete. For instance, the rs8176719-delG allele is found on 4 haplotypes in African populations (https://ldlink.nci.nih.gov/?tab=ldhap). In such a case, the impact of using haplotypes, rather than rs8176719 alone, in estimating disease risk may be even greater than that observed in this study in Europeans. The GRS proposed in this work was evaluated in an independent sample of the UKBB population whose specific design (including inclusion/exclusion criteria) differs from that of the derivation cohorts and may have had an impact on GRS properties. In addition,

assessment of the proposed haplotype/diplotype GRS in high-risk populations (including cancer patients or women taking oral contraceptives) warrants further investigation.

In summary, common *ABO* haplotypes and diplotypes are associated with different VT risks. Four common SNPs, rs2519093, rs1053878, rs8176743, and rs41302905, should be analyzed in any study aimed at accurately estimating the risk of VT attributable to the *ABO* locus.

Acknowledgments

Authorship

Contribution: Genotype and phenotype data were collected/produced by M.R.B., A.v.H.V., J.-C.S., J.-M.S., M.I.-K., N.S., D.D., R.O., P.A., S.D., A.B., P.B., A.C.M., D.O.M.-K., J.-F.D., A.J., P.S.d.V., J.C., N.L.S., F.R.R., D.I.C., P.-E.M.; L.G., F.T., K.L.W., R.L.-G., P.S.d.V., M.S.-L., D.I.C., and D.-A.T. performed statistical analyses; the general scientific project was designed by D.-A.T. and P.-E.M.; and the manuscript was drafted by L.G., D.-A.T., and P.-E.M. and further completed by P.S.d.V., M.S.-L., D.I.C., N.L.S., A.J., J.C., and P.B.

Conflict-of-interest disclosure: D.O.M.-K. is a part-time clinical research consultant for Metabolon, Inc. The remaining authors declare no competing financial interests.

ORCID profiles: F.T., 0000-0003-2229-8322; K.L.W., 0000-0003-2749-1279; R.L.-G., 0000-0003-0650-1351; J.-C.S., 0000-0003-2092-5142; N.S., 0000-0002-8365-7731; R.O., 0000-0001-7631-9657; P.A., 0000-0001-9088-234X; A.B., 0000-0001-8789-5676; M.S.-L., 0000-0002-0128-379X; F.R.R., 0000-0003-2558-7496; D.I.C., 0000-0003-3357-0862; D.-A.T., 0000-0001-9084-7800.

Correspondence: David-Alexandre Trégouët, INSERM U1219, VINTAGE, Case 11, Bordeaux Population Health Research Center, 146 Rue Léo Saignat, 33076 Bordeaux, France; e-mail: david-alexandre.tregouet@u-bordeaux.fr; and Pierre-Emmanuel Morange: Service d'Hématologie Biologique, Hôpital de la Timone, 264 Rue Saint Pierre, 13005 Marseille, France; e-mail: pierre.morange@ap-hm.fr.

Footnotes

REFERENCES

1. Jick H, Slone D, Westerholm B, et al. Venous thromboembolic disease and ABO blood type. A cooperative study. *Lancet.* 1969;1(7594):539-542.

2. O'Donnell J, Laffan MA. The relationship between ABO histo-blood group, factor VIII and von Willebrand factor. *Transfus Med.* 2001;11(4):343-351.

3. Kiechl S, Paré G, Barbalic M, et al. Association of variation at the ABO locus with circulating levels of soluble intercellular adhesion molecule-1, soluble P-selectin, and soluble E-selectin: a meta-analysis. *Circ Cardiovasc Genet.* 2011;4(6):681-686.

4. Trégouët D-A, Morange P-E. What is currently known about the genetics of venous thromboembolism at the dawn of next generation sequencing technologies. *Br J Haematol.* 2018;180(3):335-345.

5. Wu O, Bayoumi N, Vickers MA, Clark P. ABO(H) blood groups and vascular disease: a systematic review and meta-analysis. *J Thromb Haemost.* 2008;6(1):62-69.

6. Bruzelius M, Strawbridge RJ, Trégouët D-A, et al. Influence of coronary artery disease-associated genetic variants on risk of venous thromboembolism. *Thromb Res.* 2014;134(2):426-432.

7. Gran OV, Brækkan SK, Hansen J-B. Prothrombotic genotypes and risk of venous thromboembolism in cancer. *Thromb Res.* 2018;164(suppl 1):S12-S18.

8. Bruzelius M, Bottai M, Sabater-Lleal M, et al. Predicting venous thrombosis in women using a combination of genetic markers and clinical risk factors. *J Thromb Haemost.* 2015;13(2):219-227.

9. Smith NL, Rice KM, Bovill EG, et al. Genetic variation associated with plasma von Willebrand factor levels and the risk of incident venous thrombosis. *Blood.* 2011;117(22):6007-6011.

10. de Haan HG, Bezemer ID, Doggen CJM, et al. Multiple SNP testing improves risk prediction of first venous thrombosis. *Blood.* 2012;120(3):656-663.

11. Koster T, Blann AD, Briët E, Vandenbroucke JP, Rosendaal FR. Role of clotting factor VIII in effect of von Willebrand factor on occurrence of deep-vein thrombosis. *Lancet.* 1995;345(8943):152-155.

12. Ohira T, Cushman M, Tsai MY, et al. ABO blood group, other risk factors and incidence of venous thromboembolism: the Longitudinal Investigation of Thromboembolism Etiology (LITE). *J Thromb Haemost.* 2007;5(7):1455-1461.

13. Vasan SK, Rostgaard K, Majeed A, et al. ABO blood group and risk of thromboembolic and arterial disease: a study of 1.5 million blood donors. *Circulation.* 2016;133(15):1449-1457, discussion 1457.

14. Kitano T, Blancher A, Saitou N. The functional A allele was resurrected via recombination in the human ABO blood group gene. *Mol Biol Evol.* 2012;29(7):1791-1796.

15. Gassner C, Degenhardt F, Meyer S, et al. Low-frequency blood group antigens in Switzerland. *Transfus Med Hemother.* 2018;45(4):239-250.

16. Heit JA, Armasu SM, Asmann YW, et al. A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *J Thromb Haemost.* 2012;10(8):1521-1531.

17. Germain M, Chasman DI, de Haan H, et al; Cardiogenics Consortium. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *Am J Hum Genet.* 2015;96(4):532-542.

18. Vázquez-Santiago M, Vilalta N, Cuevas B, et al. Short closure time values in PFA-100® are related to venous thrombotic risk. Results from the RETROVE Study. *Thromb Res.* 2018;169:57-63.

19. Ibrahim-Kosta M, Bailly P, Silvy M, et al. ABO blood group, glycosyltransferase activity and risk of venous thromboembolism. *Thromb Res.* 2020;193:31-35.

20. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol.* 1989;129(4):687-702.

21. Smith NL, Chen M-H, Dehghan A, et al. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE Consortium. *Circulation.* 2010;121(12):1382-1392.

22. de Mutsert R, den Heijer M, Rabelink TJ, et al. The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection. *Eur J Epidemiol.* 2013;28(6):513-523.

23. Blauw LL, Li-Gao R, Noordam R, et al. CETP (cholesteryl ester transfer protein) concentration: a genome-wide association study followed by Mendelian randomization on coronary artery disease. *Circ Genom Precis Med.* 2018;11(5):e002034.

24. Paré G, Ridker PM, Rose L, et al. Genome-wide association analysis of soluble ICAM-1 concentration reveals novel associations at the NFKBIK, PNPLA3, RELA, and SH2B3 loci. *PLoS Genet.* 2011;7(4):e1001374.

25. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet.* 2002;70(2):425-434.

26. Tregouet DA, Garelle V. A new JAVA interface implementation of THESIAS: testing haplotype effects in association studies. *Bioinformatics.* 2007;23(8):1038-1039.

27. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959;22(4):719-748.

28. Connors JM. Thrombophilia testing and venous thrombosis. *N Engl J Med.* 2017;377(12):1177-1187.

29. Middeldorp S. Inherited thrombophilia: a double-edged sword. *Hematology (Am Soc Hematol Educ Program)*. 2016;2016(1):1-9.

30. Jenkins PV, O'Donnell JS. ABO blood group determines plasma von Willebrand factor levels: a biologic function after all? *Transfusion*. 2006;46(10):1836-1844.

31. McGrath RT, McKinnon TAJ, Byrne B, et al. Expression of terminal alpha2-6-linked sialic acid on von Willebrand factor specifically enhances proteolysis by ADAMTS13. *Blood*. 2010;115(13):2666-2673.

32. Swystun LL, Lillicrap D. Genetic regulation of plasma von Willebrand factor levels in health and disease. *J Thromb Haemost*. 2018;16(12):2375-2390.