



**HAL**  
open science

## Association of ABO haplotypes with the risk of venous thrombosis: impact on disease risks estimation

Louisa Goumidi, Florian Thibord, Kerri Wiggins, Ruifang Li-Gao, Michael Brown, Astrid van Hylekama Vlieg, Juan Carlos Souto, Jose Manuel Soria, Manal Ibrahim-Kosta, Noémie Saut, et al.

### ► To cite this version:

Louisa Goumidi, Florian Thibord, Kerri Wiggins, Ruifang Li-Gao, Michael Brown, et al.. Association of ABO haplotypes with the risk of venous thrombosis: impact on disease risks estimation. *Blood*, 2021, 137 (17), pp.2394-2402. 10.1182/blood.2020008997 . hal-03135635

**HAL Id: hal-03135635**

**<https://hal.science/hal-03135635>**

Submitted on 9 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Association of ABO haplotypes with the risk of venous thrombosis: impact on disease risks estimation

Louisa Goumidi<sup>1</sup>, Florian Thibord<sup>2,3,4,5</sup>, Kerri L Wiggins<sup>6</sup>, Ruifang Li-Gao<sup>7</sup>, Mickael R Brown<sup>8</sup>, Astrid van Hylckama Vlieg<sup>7</sup>, Joan-Carles Souto<sup>9,10</sup>, José-Manuel Soria<sup>11</sup>, Manal Ibrahim-Kosta<sup>1,12</sup>, Noémie Saut<sup>1,12</sup>, Delphine Daian<sup>3,13</sup>, Robert Olaso<sup>3,13</sup>, Philippe Amouyel<sup>14,15</sup>, Stéphanie Debette<sup>2,16</sup>, Anne Boland<sup>3,13</sup>, Pascal Bailly<sup>17,18</sup>, Alanna C Morrison<sup>8</sup>, Denis O Mook-Kanamori<sup>7</sup>, Jean-François Deleuze<sup>3,13,19</sup>, Andrew Johnson<sup>4,5</sup>, Paul S de Vries<sup>8</sup>, Maria Sabater-Lleal<sup>20,21</sup>, Jacques Chiaroni<sup>17,18</sup>, Nicholas L Smith<sup>22,23,24</sup>, Frits R Rosendaal<sup>7</sup>, Daniel I Chasman<sup>25,26</sup>, David-Alexandre Tréguët<sup>2,3\*</sup>, Pierre-Emmanuel Morange<sup>1,3,12\*</sup>

<sup>1</sup> Aix Marseille Univ, INSERM, INRAE, C2VN, Marseille, France

<sup>2</sup> INSERM U1219, Bordeaux Population Health Research Center, University of Bordeaux, France

<sup>3</sup> Laboratory of Excellence GENMED (Medical Genomics)

<sup>4</sup> Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, MA, USA

<sup>5</sup> The Framingham Heart Study, Framingham, MA, USA.

<sup>6</sup> Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA

<sup>7</sup> Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands

<sup>8</sup> Human Genetics Center, Department of Epidemiology, Human Genetics & Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA.

<sup>9</sup> Thrombosis and Hemostasis Research Group, Institute of Biomedical Research (IIB-Sant Pau), 08025 Barcelona, Spain.

<sup>10</sup> Unit of Hemostasis and Thrombosis, Hospital de la Santa Creu i Sant Pau, 08025 Barcelona, Spain.

<sup>11</sup> Unit of Genomic of Complex Disease. Institut de Recerca Hospital de la Santa Creu i Sant Pau. IIB Sant Pau. 08041 Barcelona, Spain.

<sup>12</sup> Hematology Laboratory, La Timone University Hospital of Marseille, Marseille, France

<sup>13</sup> Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057, Evry, France

<sup>14</sup> Lille University, Inserm, Institut Pasteur de Lille, RID-AGE - Risk Factors and Molecular Determinants of Aging-Related Diseases and Labex Distalz, France

<sup>15</sup> Lille University, Inserm, CHU Lille, Institut Pasteur de Lille, RID-AGE, France

<sup>16</sup> Department of Neurology, CHU de Bordeaux (University Hospital), Bordeaux, France

<sup>17</sup> Etablissement Français du Sang PACA-Corse "Biologie des Groupes Sanguins", Marseille, France

<sup>18</sup> Aix Marseille Univ, EFS, CNRS, ADES, "Biologie des Groupes Sanguins", Marseille, France

<sup>19</sup> Centre d'Etude du Polymorphisme Humain, Fondation Jean Dausset, Paris, France

<sup>20</sup> Genomics of Complex Diseases. Research Institute of Hospital de la Santa Creu i Sant Pau (IIB Sant Pau), Barcelona, Spain

<sup>21</sup> Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden

<sup>22</sup> Department of Epidemiology, University of Washington, Seattle, WA, USA

<sup>23</sup> Kaiser Permanente Washington Health Research Unit, Kaiser Permanente Washington, Seattle, WA USA

<sup>24</sup> Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Office of Research & Development, Seattle, WA, USA

<sup>25</sup> Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA

<sup>26</sup> Harvard Medical School, Boston, MA, USA

\* These two authors equally contributed to the work.

Corresponding authors:

Tréguët David-Alexandre: david-alexandre.tregouet@u-bordeaux.fr; INSERM U1219, VINTAGE , case 11, Bordeaux Population Health research center, 146 rue Léo Saignat, 33076 Bordeaux, France. Tel: +33 5 47 30 42 54

Morange Pierre-Emmanuel: pierre.morange@ap-hm.fr; Service d'hématologie biologique, Hôpital de la Timone, 264 rue Saint Pierre, 13005 Marseille, France. Tel.: +33 4 91 38 60 49.

## Key points

1. 4 *ABO* SNPs, rs2519093, rs1053878 , rs8176743, rs41302905, must be studied in any work assessing the risk of *ABO* locus on venous thrombosisdetermined (A/B/O) groups for estimating risk of venous thrombosis associated to *ABO* blood types
2. Molecularly defined *ABO* blood groups are more accurate than serologically determined (A/B/O) groups for estimating risk of VT.

## Abstract

Genetic risk score (GRS) analysis is an increasingly popular approach to derive individual risk prediction models for complex diseases. In the context of venous thrombosis (VT), any GRS shall integrate information at the *ABO* blood group locus, the latter being one of the major susceptibility locus for this disease. However, there is yet no consensus about which single nucleotide polymorphisms (SNPs) must be investigated when one is interested in properly assessing the association of *ABO* locus with VT risk. Using comprehensive haplotype analyses of *ABO* blood group tagging SNPs in up to 5,425 cases and 8,445 controls from 6 studies, we demonstrated that using only rs8176719 (tagging O1) to correctly assess the impact of *ABO* locus on VT risk is suboptimal as 5% of rs8176719-delG carriers are not exposed at higher VT risk. Instead, we recommend to use 4 SNPs, rs2519093 (tagging A1), rs1053878 (A2), rs8176743 (B) and rs41302905 (O2) in any analysis aimed at assessing the impact of *ABO* locus on VT risk to avoid any risk misestimation. Compared to O1 haplotype that can be inferred from these 4 SNPs, the A2 haplotype is associated with a modest increase in VT risk (odds ratio ~1.2), A1 and B haplotypes are associated with a ~1.8 fold increased risk while O2 tend to be slightly protective (odds ratio ~0.80). In addition, our analyses clearly showed that while the A1 an B blood group are associated with increased vWF and FVIII plasma levels only the A1 blood group is associated wih ICAM plasma levels but in an opposite direction, leaving additional avenues to be explored in order to fully understand the whole spectrum of biological effect of *ABO* locus on cardiovascular traits.

## Introduction

The *ABO* blood groups are determined by the *ABO* locus located on chromosome 9 (9q34.1-q34.2) which codes for two glycosyltransferases (GT), A and B, whose roles are to transfer a saccharide unit to polypeptides and membrane glycolipids (*ABO*(H) carbohydrate structures). *ABO* blood group is one of the first identified risk factor for venous thrombosis (VT) <sup>1</sup> and its thrombotic effect is considered to be mediated by the *ABO*(H) structures, which are

expressed on the glycans of both Factor VIII (FVIII) and von Willebrand Factor (VWF) <sup>2</sup> but also by modulation of soluble Intercellular Adhesion Molecule 1 (sICAM-1), sP-selectin, and sE-selectin levels <sup>3</sup>.

Despite its robust association with VT risk, there is currently a lack of consensus about the proper way to accurately estimate its impact on disease risk at the population or individual level. While some consider that individuals with A1 or B ABO blood groups are at ~1.5 fold increased VT risk compared with individuals of O and A2 blood groups who have little or no A glycosyltransferase (AGT) activity <sup>4-6</sup>, others <sup>7-12</sup> propose to contrast O and non-O blood groups without further granularity. Some have even proposed that, compared to O, the A2 blood group is associated with increased VT risk but to a lesser extent than A1 and B <sup>13</sup>. 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 pertains to the fact that 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<sup>14</sup>. The A2 blood group is resulting 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 while the rs41302905-T allele (p. Arg268) corresponds to the O2 group <sup>15</sup>. Finally, the B blood type results from a Glycine to Serine substitution at amino acid 235 and is defined by the rs8176743-T allele. These 4 SNPs then define 5 main common haplotypes in European ancestry populations (Figure 1) and, 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) <sup>6</sup>. 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 <sup>8</sup> and rs687621 <sup>9</sup> have been used as proxies for rs8176719 ( $r^2 \sim 0.84$  for both).

This work was initially designed 1 – to fine map the association of *ABO* haplotypes with VT risk, 2 – to identify the optimal set of SNPs needed to tag the VT-associated *ABO* haplotypes and 3 – to address the clinical utility 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 were deployed in relation to AGT activity, VT risk and also to quantitative cardiovascular traits whose inter-individual variability have been demonstrated to be strongly influenced by the *ABO* locus.

## Methods

### Selection of *ABO* SNPs.

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

### Studies contributing to the association with VT risk

Association of *ABO* haplotypes with VT risk was investigated in six case-control /prospective studies for VT risk that 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, and studies that have been previously described in Germain et al<sup>17</sup>(Germain *et al.*, 2015) and the RETROVE study<sup>18</sup>. Briefly, CHS is a US population based cohort of adults aged more than 65yrs and recruited in 4 different centers, HVH is a population based case-control study for cardiovascular events conducted in Washington State, EOVT and MARTHA two French case-control studies for VT, and MEGA and RETROVE are case-control studies from Netherlands and Spain, respectively. Detailed description of each study is given in Supplementary Materials. In total, 5,425 VT cases and 8,445 controls were assembled through these six studies

### Studies contributing to cardiovascular quantitative traits analysis

*ABO* haplotypes were tested for association with 4 quantitative traits that are well known to be modulated by *ABO* blood groups:

- a) AGT activity was available in 213 VT patients of the MARTHA study<sup>19</sup>;
- b) plasma levels of VWF were assessed in 13,527 individuals from 4 studies including the Atherosclerosis Risk in Communities (ARIC) study<sup>20</sup> the Framingham Heart study<sup>21</sup>, MARTHA and RETROVE;
- c) plasma levels of Factor VIII (FVIII) were assessed in 19,085 individuals from 5 studies including ARIC, CHS, MARTHA, The Netherlands Epidemiology of Obesity (NEO) study<sup>22,23</sup> and RETROVE;

d) plasma levels of ICAM-1 were available in 22,639 participants of the Women's Genome Health Study <sup>24</sup>.

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. Detailed description of the study samples and the biological measurements are given in the Supplemental data.

### **Haplotype/diplotype association analyses**

In all these studies, haplotype association analyses were conducted using either the haplo.stats R package <sup>25</sup> or the THESIAS software <sup>26</sup>. From best guessed genotypes obtained from imputed data, these two programs implement linear (for quantitative trait) and logistic (for disease status) regression models to estimate haplotype effects by comparison to a reference haplotype (taken here as the O1 haplotype) under the assumption of additive effects. Analyses were adjusted for cohort specific covariates (see Supplemental data).

Because of the very strong negative LD between studied *ABO* SNPs ( $D' \sim -1$ ) and of the resulting haplotype structure, it is possible to determine without nearly any ambiguity the pair of haplotypes (called diplotypes) carried by each individual. The percentages of individuals with no 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 hold off the assumption of additivity in haplotype effects, we then 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 heterogeneity of associations across studies was assessed with the Cochran-Mantel-Haenszel test statistic <sup>27</sup>.

### **Genetic risk score and risk prediction analyses**

Based on the haplotypic and diplotypic odds ratios (ORs) for VT derived from the meta-analyzed results of the 6 studies, two genetic risk scores (GRS) were computed. A first GRS assuming additive effects of haplotypes (GRSh) 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) that does not assume additivity of haplotype effects and 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 GRS by comparison to a simple GRS defined

by the number of O1 alleles carried by an individual. GRS were calculated in an independent sample of 18,595 VT cases and 92,983 randomly selected controls, frequently matched to cases on age at the baseline exam, from the UK Biobank (UKBB) cohort (see Supplement Materials). We employed this independent sample to minimize the risk of overfitting when GRS are evaluated in the cohorts from which they are 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. GRS were compared in terms of area under the curve (AUC), continuous net Reclassification Index (NRI) and integrated discrimination improvement (IDI). These metrics were computed using the PredictABEL R program (<https://cran.r-project.org/web/packages/PredictABEL/PredictABEL.pdf>).

## Results

As a first step, we assessed the association of *ABO* haplotypes with AGT activity, the main glycosyltransferase encoded by the A1 or A2 alleles, keeping in mind that, by definition, AGT shall be quantified only in A1 or A2 carriers. As shown in Table 1, the six studied *ABO* SNPs generated 5 common haplotypes. The rs2519093-T, rs8176719 delG, rs1053878-A, rs8176743-T and rs41302905-T alleles were each carried out by a unique haplotype, while the rare C allele at the rs579459 was carried by two distinct haplotypes. Two haplotypes demonstrated significant increased AGT activity compared with the O1 tagging haplotype: the haplotype carrying the rs1053878-A allele (i.e., A2) with modest AGT increase ( $\beta = 1655.6 \pm 708.36$ ,  $p = 0.019$ ) and the haplotype carrying the rs2519093-T allele associated with a strong AGT modulation ( $\beta = 7983.2 \pm 304.9$ ,  $p = 4.5 \cdot 10^{-150}$ ). By contrast, the haplotype carrying both the rs579459-C allele and the rs41302905-T allele was not associated with AGT ( $p = 0.864$ ). 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 as it also captures O2 carriers and in that case is not associated with AGT; 2) the rs2519093 variant is a better marker of the A1 blood group.

We then assessed the association of *ABO* haplotypes with VT risk. Detailed haplotype association results obtained in each participating study for VT are provided in Supplementary Tables 1-6. A Forest plot reporting the haplotypic ORs in each individual study is shown in Figure 2. Results indicate that, compared to the O1 haplotype, the rs2519093-T and rs8176743-T alleles were associated with an increased risk for VT of OR [95% Confidence Interval -CI] = 1.78 [1.67 – 1.92] ( $p = 3.15 \cdot 10^{-60}$ ) and 1.76 [1.58 – 1.96] ( $p = 1.88 \cdot 10^{-24}$ ), respectively, with little heterogeneity across studies ( $p = 0.136$  and  $p = 0.018$ , respectively). The rs1053878-A allele (A2 allele) also showed a trend for moderate increased VT risk with OR of 1.16 [1.03 – 1.31],  $p = 0.013$  ( $p$  for heterogeneity = 0.738) (Figure 2, Table

2). Interestingly, the A2-associated OR for VT was significantly ( $p = 3.25 \cdot 10^{-10}$ ) lower than the common OR (OR = 1.78 [1.68 – 1.88]) associated with the rs2519093-T and rs8176743-T alleles. Of note, of the two haplotypes carrying the rs579459-C allele, only the one also carrying the rs2519093-T (A1) allele associated with increased VT risk. By contrast, the second, which is also the sole haplotype carrying the rs41302905-T allele coding for O2, was associated with decreased risk of VT (OR = 0.70 [0.57 – 0.86],  $p = 7.75 \cdot 10^{-4}$ ) but with some evidence of statistical ( $p = 0.001$ ) heterogeneity across studies.

The association of ABO diplotypes with VT is illustrated in Figure 3 where meta-analyzed results from the 6 contributing studies are shown. Diplotype association results per study are given in Supplementary Table 7. Compared to O1O1 diplotype, diplotype ORs ranged from 0.81 [0.60 – 1.09] ( $p = 0.166$ ) for the O1O2 diplotype to 3.45 [1.88 – 6.32] ( $p = 5.91 \cdot 10^{-5}$ ) 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 ( $p = 0.002$ ) across individuals from the same A serological group, A1A1 (OR = 2.98 [2.50 – 3.57]), A1A2 (OR = 2.65 [2.10 – 3.33], A1O1 (OR = 2.12 [1.92 – 2.35]) and A1O2 (OR = 1.68 [1.17 – 2.41]).

From Table 2, it can be deduced that about 19% ( $= 0.07/0.37$ ) of non O1 individuals are A2 individuals at moderate risk of VT (OR  $\sim 1.2$ ),  $\sim 76\%$  are individuals with A1 or B blood type at significant stronger risk (OR  $\sim 1.8$ ) of VT while the remaining 5% with O2 blood type are not exposed to a higher risk of 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 UKBB to evaluate the AUC of three GRS: GRSh defined from haplotype effect estimates reported in Table 2, GRSd defined from diplotype effect estimates reported in Figure 3 and GRS<sub>O1</sub> defined by the number of O1 alleles carried by individuals. Corresponding AUCs [95% CI] were 0.561 [0.557 – 0.566], 0.559 [0.555 – 0.564] and 0.553 [0.548 – 0.558], respectively. GRSh- and GRSd- associated AUCs were significantly higher ( $p = 1.01 \cdot 10^{-9}$  and  $p = 1.31 \cdot 10^{-5}$ , respectively) than the AUC obtained using only O1 polymorphism, but the improvement remains modest. Similarly, the IDI associated with GRSh and with GRSd compared to GRS<sub>O1</sub> were significantly different from zero ( $p < 10^{-10}$ ) but weak (IDI = 0.2% for both). The NRI associated with GRSh was also modest (0.023 [0.011 - 0.035],  $p = 1.7 \cdot 10^{-4}$ ).

As a final step, we investigated how these observations extend to VWF (Supplementary Tables 8-11), FVIII (Supplementary Tables 12-16) and ICAM (Supplementary Table 17) plasma levels. First, in 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) alleles were homogeneously associated with strong increased VWF and FVIII levels while the rs1053878-A (A2) allele was also associated, but more modestly, with increased VWF and FVIII levels. These results are completely in line with those observed for VT risk. Interestingly, only the haplotypes tagging for A1 or A2 blood groups, not B, were associated with ICAM-1 levels but, unlike VWF and FVIII, with decreased levels (Supplementary Table 17, Table 2).

## Discussion

This work started by 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 association between *ABO* haplotypes and AGT activity. *ABO* haplotypes explained ~80% of the variability in AGT activity, and this was mainly due to A1 haplotype associated with extremely increased AGT, and to a lesser extent to A2 haplotype associated with a modest increase in AGT activity. Very interestingly, this haplotype analysis demonstrates that the rs579459 often used to tag the A1 blood group is an imperfect marker of the phenotypic A1 blood group. Indeed, about 10% (~0.02/0.22) of rs579459-C allele carriers are in fact O2 carriers and have AGT activity similar to O1 blood group. By contrast, our analyses demonstrated that the rs2519093, initially thought to be an independent *ABO* blood group polymorphism contributing to VT risk<sup>16</sup>, appears to be an optimal tag for the A1 group as it uniquely defines the haplotype associated with the highest AGT levels.

Capitalizing on 6 independent studies for VT risk totaling 5,425 VT cases and 8,445 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 similar risk of VT (OR~1.8) and that the A2 haplotype is also associated, but more modestly (OR ~1.2), to increased VT risk, the latter observation corroborating the recent findings of Vasan et al.,<sup>13</sup>. The diplotype analysis suggested that there is a substantial variability in disease risk among individuals that 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 to address the *ABO* locus contribution to VT risk, as it has been proposed, is suboptimal. Indeed, ~76% of non O1 carriers are at higher risk (OR ~1.8) of VT, ~19% at moderate risk (OR~1.2) and ~5% at no risk. However, we showed in an independent sample of 18,595 cases and 92,983 non cases that a GRS based on *ABO* haplotypes adds marginal information compared to a GRS defined from the sole rs8176719. This work is not aimed at encouraging the use of *ABO* information in clinical practice, especially as the use of others VT-associated SNPs with stronger effects (such as FV Leiden) is still debated<sup>28,29</sup>, but rather at emphasizing the need to study

haplotypes rather than only one SNP when assessing the impact of ABO locus in relation to 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 expressions, ABO participates in controlling the degree of VWF glycosylation, and consequently in its clearance and/or cleavage by ADAMTS13<sup>30,31</sup>. This would explain the robust association found between the ABO blood group and circulating concentrations of VWF<sup>32</sup> and FVIII of which VWF is the transport protein. However, the effect of the ABO locus on VTE risk cannot be explained only 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<sup>12</sup>. While the pattern of associations of ABO haplotypes with VWF/FVIII parallels that observed with VT risk, our work provides strong evidence that only A1- and A2-tagging haplotypes, and not B, are associated with plasma levels of ICAM-1. Additionally, the directions of haplotype effects are opposite to those observed for VWF and FVIII as 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 analysis 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 (e.g., P-selectin, E-selectin) and cardiovascular diseases (e.g., 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 has introduced some extra variability in the estimated haplotype frequencies and the corresponding haplotype/diplotype effects. Besides, built upon GWAS data, 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 more efficiently investigated using sequencing strategies. All the analyses reported in this work were performed in European ancestry individuals where ABO blood group tagging SNPs are in complete negative LD and where the rs8176719-delG allele tagging O1 is carried by only one haplotype. It would be very interesting to perform similar investigations in other population where LD may not be so complete. For instance, the rs8176719-delG allele is found on 4 different haplotypes in African populations (<https://ldlink.nci.nih.gov/?tab=ldhap>). In such case, the impact of using haplotypes rather than rs8176719 alone in estimating disease risk may be even greater than that observed here in Europeans. The GRS proposed in this work was evaluated in an independent sample of the UKBB population whose specific design (incl inclusion/exclusion criteria) differs from those of the derivation cohorts and may have

contributed to impact on GRS properties. In addition, the assessment of the proposed haplotype/diplotypes GRS in high-risk populations (including cancer patients or women under oral contraceptives) warrants further investigations.

In conclusion, 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.

### **Authors ' contribution**

Genotype and phenotype data were collected/produced by M.R.B, A.vH-K, J-C.S, J-M.S,M.I-K, N.S, D.B, R.O, P.A, S.D, A.B, P.B, A.M, D.O-K, J-F.D, A.J, P.deV, J.C, N.S, F.R, D.C, P-E.M. L.G, F.T, KL.W, R.L-G, P.deV, M.S-L, D.C and D-A.T performed statistical analyses. The general scientific project was designed by D-A.T and P-E.M. The manuscript was drafted by L.G,D-A.T and P-E.M and further completed by P.deV, M.S-L, D-C, N.S, A.J, J.C and P.B.

### **Acknowledgments**

Maria Sabater-Lleal is supported by a *Miguel Servet* contract from the ISCIII Spanish Health Institute (CP17/00142) and co-financed by the European Social Fund. David-Alexandre Trégouët is partially supported by the *EPIDEMIO-M-VT* Senior Chair from the University of Bordeaux initiative of excellence IdEX. Florian Thibord was partially supported by the GENMED Laboratory of Excellence on Medical Genomics, Agence Nationale de la Recherche (ANR-10-LABX-0013). Paul S. de Vries was additionally supported by American Heart Association grant number 18CDA34110116. The International Network Against Venous Thrombosis (INVENT) Consortium and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Working Group were partially funded by NIH HL134894.

MARTHA genetic study was supported by the GENMED Laboratory of Excellence on Medical Genomics, Agence Nationale de la Recherche (ANR-10-LABX-0013). Measurement of Glycosyltransferase A activity was supported by Etablissement Français du Sang, France (APR-2014-17). The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL059367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health

contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions.

The Cardiovascular Health was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, and U01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health

This Framingham Heart Study was supported by the National Heart, Lung, and Blood Institute contracts NO1-HL 25195 and HHSN268201500001I. Florian Thibord and Andrew Johnson were supported by NHLBI Intramural funds. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

The research of the HVH study has been funded in part by National Institute of Health grants HL40628, HL43201, HL53375, HL60739, HL68639, HL68986, HL73410, HL74745, HL85251, HL95080, and HL134894.

The RETROVE study was supported by grants PI 12/612, PI 15/269 and PI 18/434 from the ISCIII Spanish Health Institute.

UK Biobank has received funding from the UK Medical Research Council, Wellcome Trust, UK Department of Health, British Heart Foundation, Cancer Research UK, Diabetes UK, Northwest Regional Development Agency, Scottish Government, and Welsh Assembly Government. This research has been conducted using the UK Biobank Resource under Application number 40713.

### **Conflicts of Interest**

Dennis O. Mook-Kanamori is a part-time clinical research consultant for Metabolon, Inc.

## Data sharing

For original data, please email the corresponding author.

**Figure 1:** Linkage disequilibrium matrix and haplotype frequencies at the ABO locus.

Displayed SNPs are those discussed in the manuscript (and in supplemental materials). Numerical values shown in the upper triangle of the matrix are pairwise linkage disequilibrium  $r^2$  values. Shown values for  $r^2$  and allele/haplotype frequencies were estimated from European populations available at <https://ldlink.nci.nih.gov/?tab=ldmatrix>.

**Figure 2:** Forest plot showing the association of ABO haplotypes with VT risk

The figure shows the haplotypic odds ratio for VT associated with the O2, A1, A2 and B haplotypes compared to the O1 haplotype taken as reference. Odds ratio were computed under the assumption of haplotype effect additivity.

**Figure 3:** Forest plot showing the association of ABO diplotypes with VT risk

In this figure, the O1O1 diplotype was used as reference. Diploypic Odds Ratio were those obtained from the meta-analysis of results observed in each of the 6 contributing study (Supplementary Table 7).

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

ABO blood group	rs2519093	rs8176719	rs579459	rs1053878	rs8176743	rs41302905	Haplotype Frequencies	Haplotype Additive effects
O1	C	delG	T	G	C	C	0.528	reference
O2	C	G	C	G	C	<b>T</b>	0.014	$\beta = -249.54$ [-3115.8 - 2616.8] p=0.864
A1	<b>T</b>	G	C	G	C	C	0.357	$\beta = 7983.2$ [7385.6 - 8580.8] p = 4.5 10 <sup>-150</sup>
A2	C	G	T	<b>A</b>	C	C	0.046	$\beta = 1655.6$ [267.2 - 3044.0] p=0.019
B	C	G	T	G	<b>T</b>	C	0.033	$\beta = -239.2$ [-1475.0 - 996.6] p=0.704

The analysis was conducted using THESIAS, and adjusted for age, sex and the 4 first principal components

**Table 2** Summary information about the association of ABO haplotypes with studied traits.

ABO blood group	rs2519093	rs8176719	rs579459	rs1053878	rs8176743	rs41302905	Frequency <sup>1</sup>	OR for VT <sup>2</sup>	Impact on VWF/FVIII levels <sup>4</sup>	Impact on ICAM1 levels <sup>4</sup>
O1	C	delG	T	G	C	C	0.63	reference	reference	reference
O2	C	G	C	G	C	<b>T</b>	0.02	0.70 <sup>3</sup> [0.57 – 0.86]	-	-
A1	<b>T</b>	G	C	G	C	C	0.20	1.78 [1.67 – 1.92]	↑↑↑	↓↓↓
A2	C	G	T	<b>A</b>	C	C	0.07	1.16 [1.03 – 1.31]	↑	↓
B	C	G	T	G	<b>T</b>	C	0.08	1.76 [1.58 – 1.96]	↑↑↑	-

<sup>1</sup> Haplotype frequencies in European ancestry populations

<sup>2</sup> Haplotypic risk estimate [95% Confidence Interval] for VT obtained from the meta-analysis of results listed in Supplementary Tables 1-6

<sup>3</sup> This risk estimate exhibited statistical heterogeneity ( $p = 1.210^{-3}$ ) across the 6 investigated studies

<sup>4</sup> Direction (↑= increase; ↓= decrease) of haplotype effects on plasma levels. Strong compared to modest effects are illustrated with three arrows instead of one.

## 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, Barbalić 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. Vasani 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. 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.

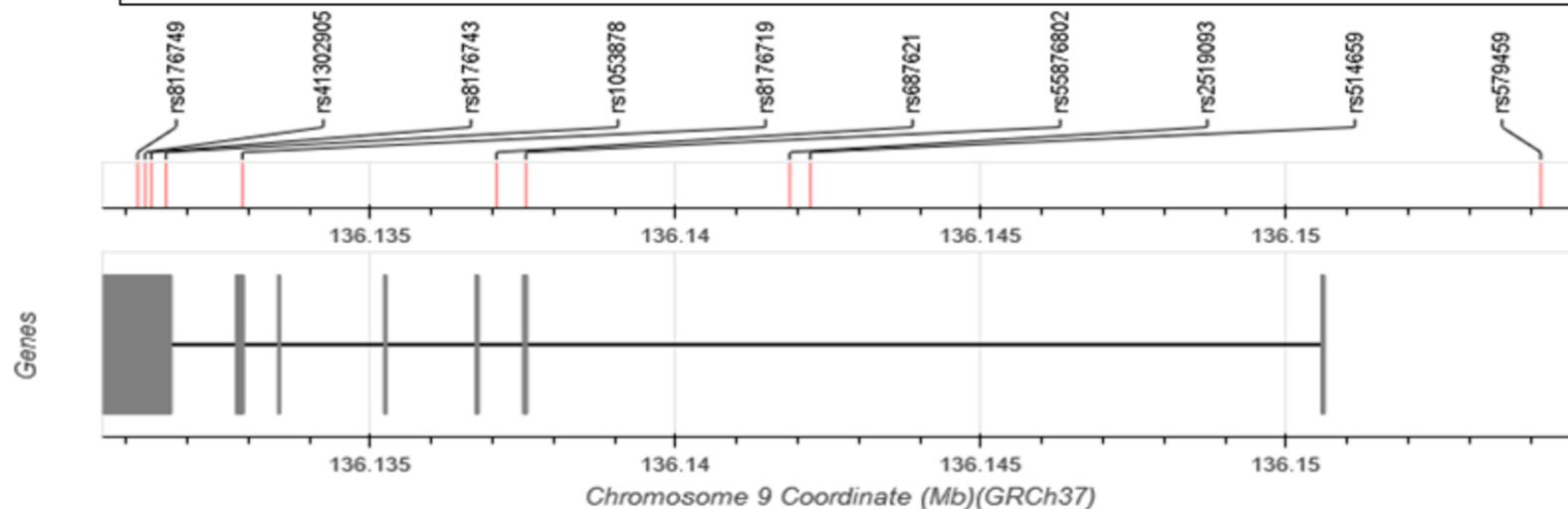
21. Smith NL, Chen M-H, Dehghan A, et al. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE Consortium. *Circulation*. 2010;121(12):1382–1392.
22. de Mutsert R, den Heijer M, Rabelink TJ, et al. The Netherlands Epidemiology 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 NFKB1K, 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.

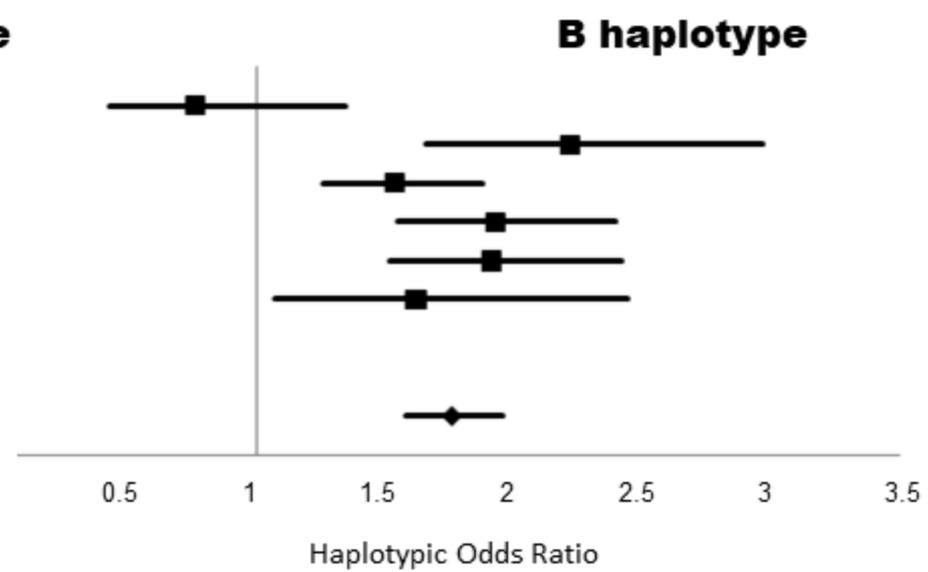
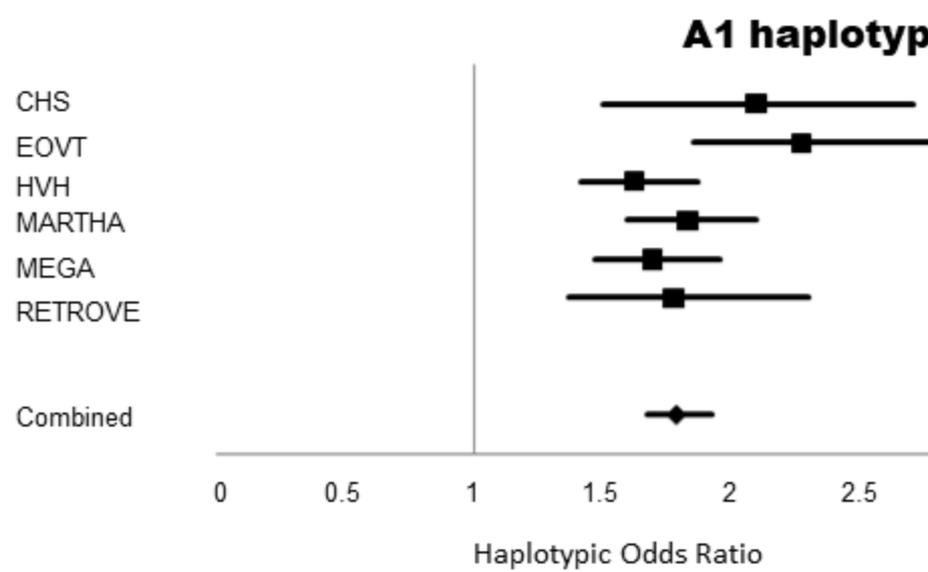
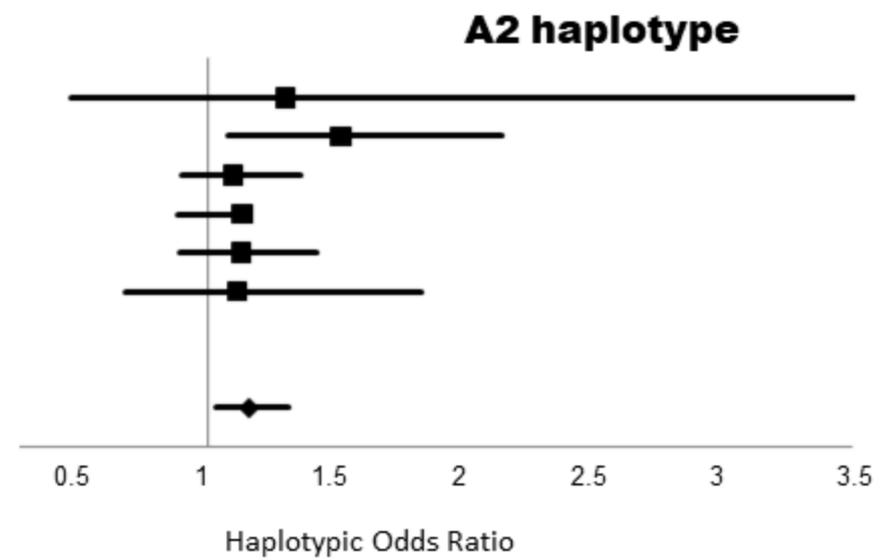
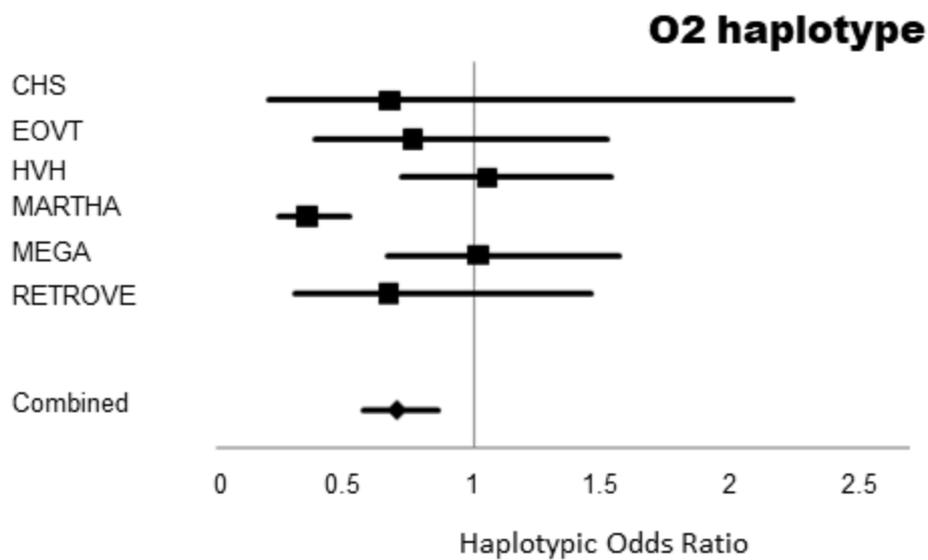
Figure 1

	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 groups		O2	B	A2	O1					

	Allele Frequencies
rs8176749	C=0.915, T=0.085
rs41302905	C=0.971, T=0.029
rs8176743	C=0.915, T=0.085
rs1053878	G=0.901, A=0.099
rs8176719	delG=0.605, G=0.395
rs687621	A=0.629, G=0.371
rs55876802	C=0.971, A=0.029
rs2519093	C=0.815, T=0.185
rs514659	A=0.629, C=0.371
rs579459	T=0.786, C=0.214

Linkage Disequilibrium ( $r^2$ ) matrix										Haplotype Frequencies					
rs8176749	1.0	0.003	1.0	0.01	0.142	0.157	0.003	0.21	0.157	0.025	C	C	C	T	C
rs41302905		1.0	0.003	0.003	0.046	0.017	1.0	0.007	0.017	0.109	C	C	C	C	T
rs8176743			1.0	0.010	0.142	0.157	0.003	0.021	0.157	0.025	C	C	C	T	C
rs1053878				1.0	0.148	0.187	0.003	0.020	0.187	0.022	G	G	A	G	G
rs8176719					1.0	0.864	0.046	0.342	0.857	0.411	delG	G	G	G	G
rs687621						1.0	0.017	0.387	0.991	0.285	A	G	G	G	A
rs55876802							1.0	0.007	0.017	0.109	C	C	C	C	A
rs2519093								1.0	0.385	0.823	C	T	C	C	C
rs514659									1.0	0.285	A	C	C	C	A
rs579459										1.0	T	C	T	T	C
											0.5994	0.1809	0.0924	0.0845	0.0288
											O1	A1	A2	B	O2





Serological  
Blood group

Diplotype

AB

A2B

A1B

B

BB

BO2

BO1

A

A1A1

A1A2

A2A2

O1A1

O1A2

O2A1

O2A2

O

O1O2

0.5

1

1.5

2

2.5

3

3.5

4

4.5

5

Diploypic Odds Ratio

3.35

1.77

2.53

3.45

2.42

2.98

2.65

1.91

2.12

1.07

1.68

1.08

0.81

