ARTICLE

Association between ABO and Duffy blood types and circulating chemokines and cytokines

Sarah C. Van Alsten [1] · John G. Aversa[2] · Loredana Santo[3] · M. Constanza Camargo[4] · Troy Kemp[5] · Jia Liu[6] ·
Wen-Yi Huang[4] · Joshua Sampson[7] · Charles S. Rabkin [8]

Abstract
Blood group antigens are inherited traits that may play a role in immune and inflammatory processes. We investigated associations between blood groups and circulating inflammation-related molecules in 3537 non-Hispanic white participants selected from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. Whole-genome scans were used to infer blood types for 12 common antigen systems based on well-characterized single-nucleotide polymorphisms. Serum levels of 96 biomarkers were measured on multiplex fluorescent bead-based panels. We estimated marker associations with blood type using weighted linear or logistic regression models adjusted for age, sex, smoking status, and principal components of population substructure. Bonferroni correction was used to control for multiple comparisons, with two-sided p values < 0.05 considered statistically significant. Among the 1152 associations tested, 10 were statistically significant. Duffy blood type was associated with levels of CXCL6/GCP2, CXCL5/ENA78, CCL11/EOTAXIN, CXCL1/GRO, CCL2/MCP1, CCL13/MCP4, and CCL17/TARC, whereas ABO blood type was associated with levels of sVEGFR2, sVEGFR3, and sGP130. Post hoc pairwise t-tests showed that individuals with type Fy(a+b−) had the lowest mean levels of all Duffy-associated markers, while individuals with type A blood had the lowest mean levels of all ABO-associated markers. Additional work is warranted to explore potential clinical implications of these differences.

1. Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

2. Surgical Oncology Program, Center for Cancer Research, NCI, Bethesda, MD, USA

3. Occupational and Environmental Epidemiology Branch, DCEG, NCI, Bethesda, MD, USA

4. Metabolic Epidemiology Branch, DCEG, NCI, Bethesda, MD, USA

5. HPV Immunology Laboratory, Frederick National Laboratory for Cancer Research, NCI, Frederick, MD, USA

6. Cancer Genomics Research Laboratory, DCEG, NCI, Bethesda, MD, USA

7. Biostatistics Branch, DCEG, NCI, Bethesda, MD, USA

8. Infections and Immunoepidemiology Branch, DCEG, NCI, Bethesda, MD, USA

Introduction

Blood group systems are collections of polymorphic proteins and oligosaccharides expressed on the surface of red blood cells and other tissues. Within each blood group are the antigenic variants that comprise the system, commonly referred to as "blood type," which correspond to allo-antibody reactions among blood group phenotypic variants [1, 2]. Although some of these antigenic families are defined at the protein level, most are determined in a monogenic fashion, including some of the more commonly studied variants of ABO, H, Rhesus (Rh), Duffy, Lewis, MNS, and Lutheran blood groups [3].

While our understanding of the clinical significance of many less-studied blood group antigens continues to evolve, several blood group systems have well-established clinical associations. ABO's roles in mediating acute hemolytic reactions and modulating hyperacute solid organ rejection have been extensively described and now comprise much of the scientific basis for blood product administration and allotransplantation, respectively [4]. Similarly, Rh antigenicity remains crucial in the management and prevention

of fetal hemolytic disease. In addition, more recent data have elucidated associations between various blood group systems and susceptibility to infectious pathogens such as *Plasmodium vivax* malaria, *Helicobacter pylori*, norovirus, and the severe acute respiratory syndrome coronavirus (SARS-CoV; [5–8]). In findings, more germane to the present pandemic, individuals with non-O blood type, and particularly carriers of the A antigen, have been consistently shown to have higher risk of SARS-CoV2 infection [9–11], and severe complications as a result of SARS-CoV2 [12, 13], than individuals with O blood type, although the mechanisms underlying these associations are unclear.

Beyond the role of the blood group antigens in immunologic reactions and initial susceptibility to infections, the functions of these antigens in host immunity and inflammation are poorly understood. Histo-blood group antigens (i.e., ABO and Lewis) have been found to affect the stability or organization of other glycans on the cell surface [14], possibly modulating binding of other antigens or marking cells for degradation [15]. For instance, it is well known that individuals expressing the A or B ABO antigens have lower rates of clearance of von-Willebrand factor and consequently higher risk of venous thromboembolism than individuals expressing O antigen [16, 17]. Similar investigations measuring inflammatory markers directly, however, have not yielded such unequivocal results. Specifically, while genome-wide association studies (GWAS) have identified associations of the ABO locus and ABO methylation with levels of circulating inflammatory markers [16, 18–21], the relationships failed to replicate in validation samples [16, 20]. Furthermore, the Duffy antigen, a marker with less-defined clinical significance, has been shown to act as an atypical chemokine receptor and assists in regulating circulating CXC and CC-motif chemokine levels [22, 23].

These findings have raised interest for detecting and describing associations between circulating marker levels and blood group antigens. While the existing data justify further investigation, the evaluated markers were all selected a priori for investigation and represent only a small subset of key immune and inflammatory processes that may involve blood groups. Characterizing the additional associations between blood types and immune markers may clarify the broader role of blood group antigens in host immunity and disease susceptibility, and contribute to fundamental description of an individual's inflammatory state. We have therefore cataloged associations between blood group genotypes and a broad spectrum of circulating immune and inflammation markers in an exploratory reanalysis of existing data. To our knowledge, this is the first study to examine associations between blood groups and inflammation using marker data measured on multiplex panels. Ultimately, our aim is to improve understanding of how blood group antigens relate to systemic inflammation.

Results

The final analytic sample size was 3537. No participants were missing information on age, sex, or smoking status. As shown in Table 1, the weighted percentage of cancer cases in the study sample was 3.7%, while 9.7% were current smokers, 41% were former smokers, and 56.5% were male, corresponding to the respective prevalence values in the underlying Prostate, Lung, Colorectal, and Ovarian (PLCO) cohort.

Of the 1152 associations tested (12 blood types × 96 markers), we found ten markers to be associated at the Bonferroni-adjusted significance level of $p < 4.3 \times 10^{-5}$ (Fig. 1). The Duffy antigen group was associated with seven inflammatory markers (CXCL6/GCP2, CXCL5/ENA78, CCL11/EOTAXIN, CXCL1/GRO, CCL2/MCP1, CCL13/MCP4, and CCL17/TARC), and ABO was associated with three markers (soluble vascular endothethial growth factor receptor 2 (sVEGFR2), sVEGFR3, and soluble glycoprotein 130 (sGP130). All associations remained statistically significant in additive models (Supplementary Table 1). No other blood types appeared to be strongly associated with other inflammatory markers (Supplementary Fig. 1). Complete lists of associations from categorical and additive models are listed in the Supplementary Tables 2 and 3. Findings were similar when associations were evaluated with Tobit models (Supplementary Tables 4 and 5).

Among the ten blood type-marker associations identified among non-Hispanic whites, only the associations of Duffy with CXCL6/GCP2, CXCL1/GRO, and CXCL5/ENA78 were evident in other racial and ethnic groups (Supplementary Table 6 and Supplementary Fig. 2).

Post hoc evaluation of individual terms showed that individuals with Duffy type Fy(a−b+) and Fy(a+b+) had higher levels of all associated circulating chemokines than individuals with Duffy type Fy(a+b−), although the pairwise comparison between Fy(a+b+) and FY(a+b−) for CXCL5/ENA78 was not significant after Bonferroni correction (Fig. 2). Individuals with ABO A blood type had significantly lower levels of sVEGFR2, sVEGFR3, and sGP130 than individuals with O or B blood types; no significant differences were noted between type A and AB individuals (Fig. 3). Most correlations among the ten significantly associated markers were weak with $|r| < 0.4$, although moderate associations were observed between sVEGFR2 and sGP130 ($r = 0.55$), CCL2/MCP1 and CCL11/EOTAXIN ($r = 0.55$), and CXCL5/ENA78 and CXCL1/GRO ($r = 0.48$).

Fig. 1 Observed versus expected distribution of *p* values for associations between circulating inflammatory markers and blood antigen groups. Each dot represents the joint Wald test of a linear or logistic model predicting marker concentration from blood type, adjusting for sex, age (5-year category), smoking status (never/former/current), study, and the first-five principal components of population substructure. Labeled points denote associations which were significant after Bonferroni correction. Shaded bands represent a bootstrapped 95% confidence interval generated from 10,000 replicates, and dotted lines represent predicted empirical distribution. Note different *y*-axis scales for top and bottom figures.
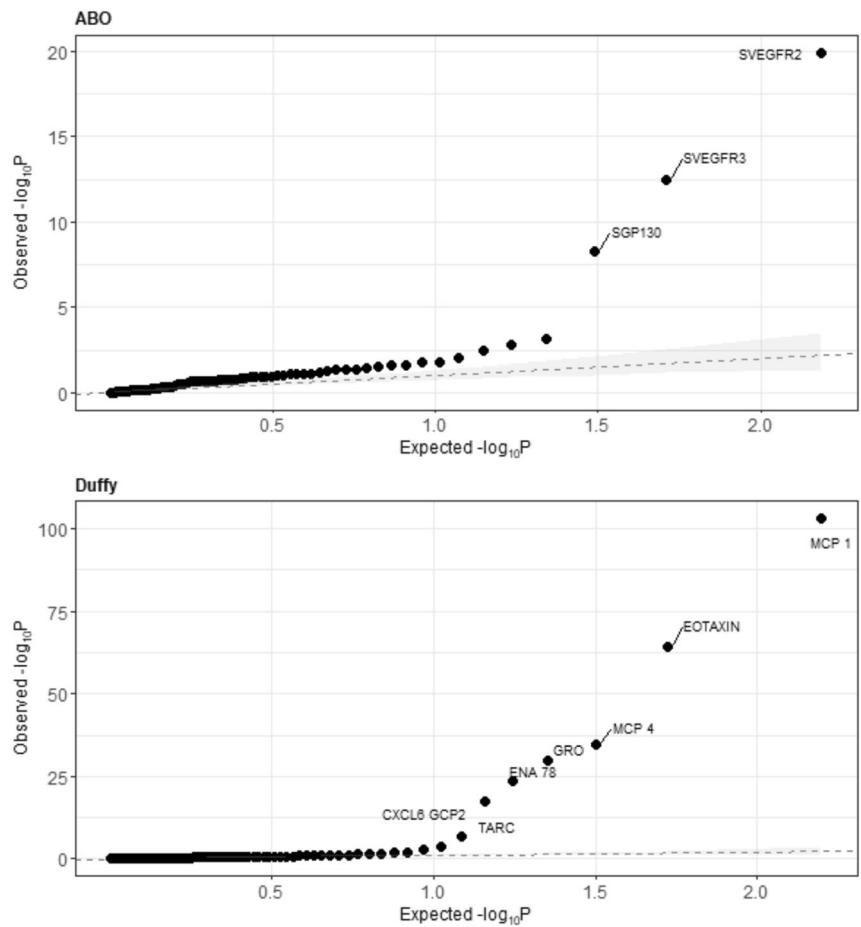
Fig. 2 Estimated marginal mean (EMM) concentrations of circulating biomarkers among individuals with different Duffy phenotypes, adjusting for age (5-year categories), sex, study, smoking status (never/former/current), and the first-five principal components of population substructure. Bars represent 95% confidence intervals of the EMM. Overlapping black lines between bars represent pairwise comparisons that are not significantly different after Bonferroni correction.
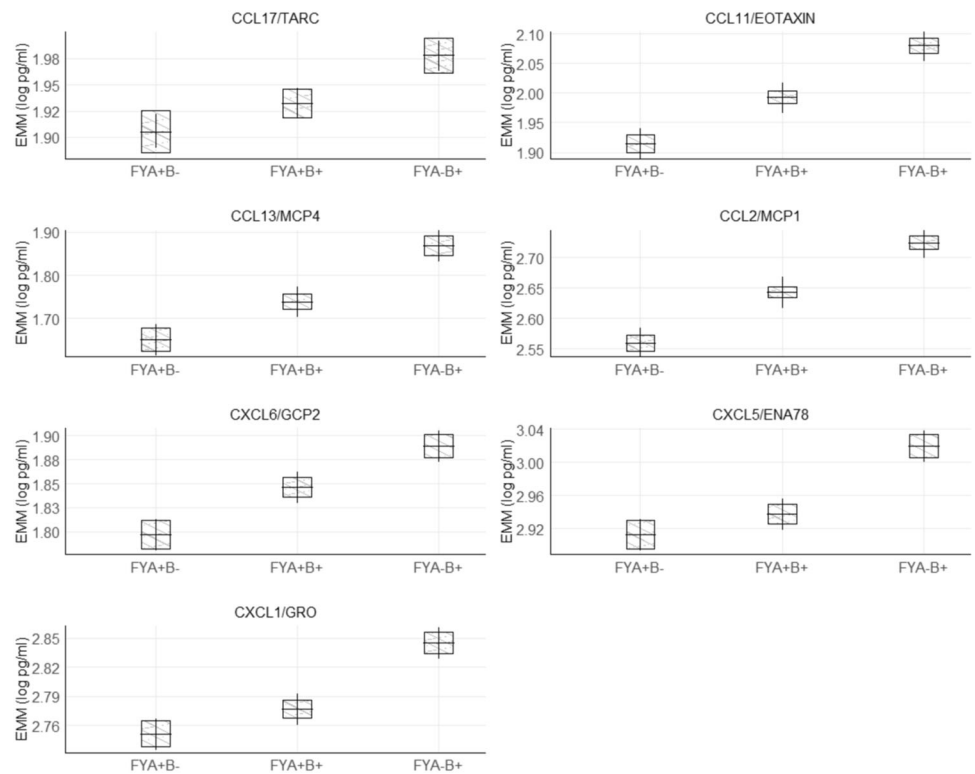
Fig. 3 Estimated marginal mean (EMM) concentrations of circulating biomarkers among individuals with different ABO phenotypes, adjusting for age (5-year categories), sex, study, smoking status (never/former/current), and the first-five principal components of population substructure. Bars represent 95% confidence intervals of the EMM. Overlapping black lines between bars represent pairwise comparisons that are not significantly different after Bonferroni correction.
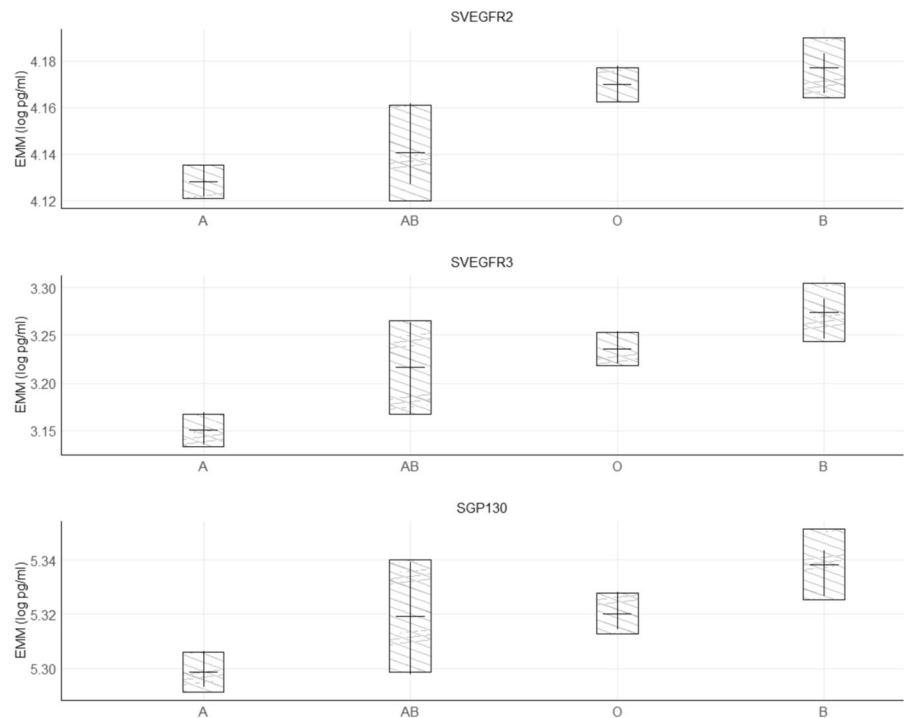
Discussion

We evaluated a wide range of circulating inflammation-related proteins and found that ten inflammatory markers were associated with Duffy and ABO antigenic variants. The associations were robust against alternative model specifications, specifically when controlled for population stratification, and nonrepresentative sampling. These findings may reflect baseline differences in the inflammatory state of individuals with these blood types, and similarly could help to explain immune reactivity and infection susceptibility among corresponding subgroups of individuals.

All of the seven markers associated with Duffy antigen type were CC or CXC motif chemokines, five of which (CCL11/EOTAXIN, CXCL1/GRO, CCL2/MCP1, CCL13/MCP4, and CCL17/TARC) are known to have high affinity for the Duffy antigen receptor [24]. Mounting evidence suggests that the Duffy antigen receptor acts as a homeostatic reservoir for chemokines, buffering intracellular concentrations of these molecules, and sequestering them for release under conditions of immune challenge [22]. Prior GWAS have also noted biomarker associations between rs12075, which distinguishes the Fy(a) and Fy(b) antigens of the Duffy blood group and circulating CCL2/MCP1 [25]. From a clinical standpoint, inverse associations have been noted between Duffy antigen expression and metastatic potential in breast [26], pancreatic ductal [27], and non-small cell lung carcinomas [28], possibly due to lower concentrations of pro-angiogenic chemokines in the tumor microenvironment. Notably, however, these models

of tumor aggressiveness are based on comparisons between Duffy-positive (carriers of Fy(a) and/or Fy(b)) and Duffy-null individuals (carriers of neither Fy(a) nor Fy(b)). Given the heterogeneity within the Duffy-positive subgroup observed in the current study, our results suggest that distinctions between the Fy(a) and Fy(b) alleles may also be etiologically relevant for metastatic potential.

Two of the markers associated with ABO were soluble receptors for molecules in the VEGF superfamily, a class of proteins, which regulate angiogenesis [29]. The soluble forms of VEGF receptors 2 and 3 compete with membrane-bound VEGF receptors for binding of pro-angiogenic VEGF-A and VEGF-C, thereby inhibiting outgrowth of blood and lymphatic vessels [30, 31]. The finding that individuals with type A blood had lower levels of circulating sVEGFR2, and sVEGFR3 is consistent with the higher risks of acute cardiovascular events, such as venous thromboembolism and myocardial infarction, observed in those with type A blood relative to those with type O [32]. Decreased expression of sVEGFR2 also correlates with metastatic potential in laboratory models of neuroblastoma [33] and breast cancer [34], which suggests that ABO-associated differences in angiogenic capacity may have implications for oncologic progression and prognosis. For instance, lower levels of prediagnostic sVEGFR2 and sVEGFR3 have also been associated with lung cancer risk in never smokers [35], while type A blood is associated with increased risks of pancreatic [36, 37], ovarian [38], and gastrointestinal cancers [32]. The extent to which these associations are related to inflammation and angiogenesis

Table 1 Selected characteristics (%) of participants in component case–control studies nested in the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial.

	Laboratory pilot	NHL	Lung (Study 1)	Lung (Study 2)	Ovarian	Endometrial	Colorectal	Upper GI	Overall	PLCO[a]
Case/control										
Cases	–	45.2	40.3	43.5	49.4	44.3	54.7	47.9	46.4	3.7
Controls	100	54.8	59.7	56.5	50.6	55.7	45.3	52.1	53.6	96.3
Sex										
Male	47.4	63.6	64.9	58.2	0	0	62.5	95	50.1	56.5
Female	52.6	36.4	35.1	41.8	100	100	37.5	5	49.9	43.5
Age (years)										
55–59	25.6	25.1	19.1	20.4	32.3	32.6	14.8	29.4	22.6	33.6
60–64	34.6	31.1	28.7	27.5	28.7	35	32	37.8	30.7	32
65–69	20.5	27.7	31.6	30.4	26.3	22.3	32.5	24.4	28.9	20.3
70–74	19.2	16.1	20.6	21.6	12.7	10.1	20.6	8.4	17.8	14
Smoking status										
Never	38.5	46.9	13.3	18.2	60.6	59.8	45.8	35.3	36.3	49.4
Former	39.7	10.1	32.7	37.1	8.4	6	6.3	10.1	19.6	41
Current	21.8	43	54.1	44.7	31.1	34.2	47.8	54.6	44	9.7

[a]Overall percentages re-weighted to the total numbers of participants in the PLCO screening arm (see Supplementary Methods).

versus other probable mechanisms, such as *H. pylori* colonization [39], requires additional study.

Differences in circulating levels of sVEGFR2 and sVEGFR3 among individuals with type A blood may also have prognostic implications for COVID19. For instance, an autopsy study of lungs from COVID19 patients who died of acute respiratory distress found upregulated mRNA expression of both VEGF-A and VEGF-C [40]. Given that individuals with type A blood are also at increased risk of experiencing severe COVID19 (defined as requiring a ventilator, death, or development of respiratory distress; [9, 12, 13]), metabolic pathways involving sVEGFR2/sVEGFR3 may underlie higher COVID19 susceptibility of individuals with type A blood.

Like the other two markers associated with ABO, sGP130 has an anti-inflammatory role, inhibiting the IL6/soluble IL6 receptor (sIL6R) trans-signaling cascade [41, 42]. IL6 trans-signaling induces T-cell recruitment to local sites of inflammation and plays an important role in the development of atherosclerosis when activated in vascular tissue [43]. Interestingly, exogenous administration of sGP130 attenuates this effect in mice [44], yet in humans, elevated levels of sGP130 are correlated with higher blood pressure and increased risk of adverse cardiovascular events [45] in individuals with pre-existing cardiovascular conditions. We found individuals with type A blood to have lower levels of SGP130 than those with type O or B blood, consistent with an anti-atherosclerotic activity of the molecule. However, additional research is needed to understand the complex regulatory loops governing this association and its apparent pleiotropic effects.

Notably, we did not find associations between levels of inflammatory markers and blood groups other than Duffy or ABO. However, mechanisms other than inflammation may mediate some associations between blood types and diseases, such as the Kidd system with bladder cancer [46].

The strengths of our study include its large sample size, use of a well-characterized, population-based cohort, comprehensive assessment of inflammatory biomarkers using a reproducible method, the broad spectrum of blood group phenotypes measured using high-quality genetic data, and statistical control for potential nonrepresentativeness due to reliance on case–control data. Nonetheless, our results should be interpreted amidst several limitations. First, because the PLCO was predominantly comprised of non-Hispanic white participants, we had limited power to detect meaningful differences in inflammation levels by blood type for individuals of non-European ancestry. Thus, the findings may not generalize to racial and ethnic groups other than non-Hispanic whites. Second, our data are based on single measurements of circulating biomarker levels, which may require serial assessment for better characterization. In addition, because assays were conducted separately for each component study, laboratory variation may have limited our power to detect weaker associations. Third, several of the assays used to detect marker levels, including that of sCD30, may have been less sensitive than those used in previous studies since we identified fewer samples above the lower limits of detection. Finally, misclassification of blood group phenotypes may have occurred due to rare variants in genomic regions,

Table 2 Single-nucleotide polymorphisms (SNPs) used to define blood antigen phenotypes of participants in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial.

Blood type	Chromosome:position	SNP [63–66]	Nucleotide	Weighted prevalence (%)
ABO				
A1	Ref	Ref	Ref	11.5
A2	9:133256264	rs1053878	A	7.9
B	9:133256205/133257486	rs7853989/rs8176720	G/G	17.3
O (null)	9:133257521-133257522	rs8176719	Del	63.3
Auburger				
A	Ref	Ref	Ref	66.4
B	19:44819487	rs1135062	G	33.6
Colton				
A	Ref	Ref	Ref	96.4
B	7:30912043	rs28362692	T	3.6
Dombrock				
A	Ref	Ref	Ref	69.8
B	12:14840505	rs11276	T	30.2
Duffy				
A	Ref	Ref	Ref	57.4
B	1:159205564	rs12075	A	40.8
x	1:159205704	rs34599082	T	1.8
Null	1:159204893	rs2814778	C	<0.1
Kell				
k	Ref	Ref	Ref	94.7
K	7:142957921	rs8176058	A	5.3
Kidd				
A	Ref	Ref	Ref	49.8
B	18:45739554	rs1058396	A	50.2
Knops				
A	Ref	Ref	Ref	97.4
B	1:207609424	rs41274768	G	2.6
Lewis				
A	Ref	Ref	Ref	32.1
le1 (Null)	19:5844781/5844332	rs28362459/rs3745635	C/T	6.5
le2 (Null)	19:5844781/5843773	rs28362459/rs3894326	C/T	3.4
B	19:48703417	rs601338	G	58
Lutheran				
A	Ref	Ref	Ref	96.5
B	19:44812188	rs28399653	G	3.5
RhE				
E	Ref	Ref	Ref	86.1
e	1:25390874	rs609320	G	13.9
Secretor				
Se	Ref	Ref	Ref	58
se (null)	19:48703417	rs601338	A	42

which we did not assess, although the degree of misclassification would correspondingly be small and likely nondifferential.

Overall, our study suggests that Duffy and ABO blood types are associated with differences in levels of several circulating inflammatory markers. Further studies are warranted to determine whether differences in circulating levels of chemokines and angiogenic factors are etiologically or clinically relevant for Duffy- and ABO-associated conditions.

Materials and methods

Study population

Data for this study come from the PLCO Cancer Screening Trial. The design of the PLCO has been described in detail elsewhere [47, 48]. Briefly, ~155,000 men and women aged 55–74 years were recruited for the PLCO from the general population between 1992 and 2000 at ten sites across the United States. Participants randomized to the screening arm received periodic cancer screenings in the first 6 years of follow-up. In addition to basic demographic and behavioral information, nonfasting blood plasma samples were collected from screening arm participants at baseline and annually for five follow-up visits [48]. All participants provided informed consent.

We included participants from one laboratory pilot ($n =$ 78) and seven cancer case–control studies nested within PLCO from prior investigations of non-Hodgkin lymphoma (NHL; 358 cases/295 controls), and lung (study 1: 347 cases/513 controls; study 2: 411 cases/533 controls), ovarian (124 cases/127 controls), endometrial (280 cases/352 controls), colorectal (432 cases/358 controls), and upper gastrointestinal (57 cases/62 controls) cancers [49–55]. For participants included in multiple studies ($n = 219$), we used marker data from the study with fewer missing marker values or, given equal numbers of missing values, the first study of inclusion. Due to small numbers with other racial and ethnic backgrounds among the PLCO cohort, we limited our primary analyses to self-identified non-Hispanic white participants, as confirmed by genetically determined ancestry. Additional details regarding individual study inclusion/exclusion and matching criteria are provided in the Supplementary Table 7. Sample size was determined by availability of extant data.

Laboratory analyses

Using prediagnostic serum blood samples from either recruitment (NHL, first lung, endometrial, colorectal, upper gastrointestinal studies) or follow-up (second lung and ovarian studies) visits, serum levels of 102 immune and inflammation-related markers were measured by a single laboratory on Luminex fluorescent bead-based assay panels (Millipore Inc., Billerica, MA). As described previously, assays were run at a centralized laboratory according to manufacturer protocol, with samples from cases and controls included in the same batches. Laboratory staff were blinded to the patient data and clinical outcome. Concentrations were determined with four- or five-parameter standard curves, and samples were run in duplicate to calculate average measurements. In addition, blinded replicates were included within and across batches to assess reproducibility. Ninety-one percent of log-transformed marker intraclass correlation coefficients (ICCs) were >0.80 in the lung and NHL studies, 98% in the lung replication study, 84% in the endometrial study, and 76% in the ovarian cancer study [50–54]. The upper gastrointestinal cancer study was itself comprised of multiple case–control studies and thus did not report overall ICCs. Due to the use of multiple assay lots for markers in the colorectal cancer study, we only included results for an analyte from the lot with the largest number of samples. Depending on the analyte, the number of included participants from the colorectal cancer study ranged from 668 (319 cases/349 controls) to 768 (365 cases/403 controls).

After pooling the data, we excluded six markers (interleukin 3, leukemia inhibitory factor, soluble CD30, soluble interleukin 1 receptor 1, soluble receptor for advanced glycation end products, and soluble vascular endothelial growth factor receptor 1) that were detectable in fewer than 10% of samples. A complete list of markers assessed and the studies in which they were measured is available in the Supplementary Table 8.

Genotyping

Participants were genotyped on either the Global Screening Array, Oncoarray, OmniX, or Omni 2.5 platforms. We evaluated twelve blood group systems based on the commonly encountered single-nucleotide polymorphisms (SNPs) that define their phenotypic variance. After excluding samples with call rates <95% and SNPs with call rates <95%, we pre-phased SNPs using Eagle, version 2.4 [56] and the TOPMed reference panel 5b [57]. We then imputed genotypes against the same reference panel on Minimac, version 4 [58], and removed variants with Rsq values <0.3. We partitioned remaining variants into bins based on minor allele frequencies ([0,0.0005], (0.0005,0.002], (0.002,0.005], (0.005,0.01], (0.01,0.03], (0.03,0.05], (0.05, 0.5)), then iteratively filtered out variants in each bin with the lowest Rsq values until the average bin Rsq was at least 0.9 [59]. Of the SNPs we used to evaluate blood types, only one (rs12075) was genotyped directly on any of the four platforms. The remaining SNPs were imputed against the TOPMed reference panel.

We also generated a set of principal components to accommodate potential population stratification. The Genetic Relationships and Fingerprinting method was used to split participants into genetically determined ancestry groups, with splits made separately by genotyping platform [60, 61]. For groups with at least ten individuals, we removed variants with minor allele frequencies <0.01, <98% call rates, or Hardy Weinberg Equilibrium p values < 0.001. We then pruned remaining SNPs in a stepwise manner and removed heterozygosity outliers ($|F| > 0.2$),

using remaining individuals and SNPs to generate principal components using EIGENSTRAT smartpca [62].

Table 2 depicts the blood antigens assessed in this study, as well as the SNPs used to infer them [63–66]. Except for Lewis null (le1/le2), Secretor null (se), and the O allele of ABO, all blood group alleles were considered codominant.

Statistical analyses

Our objective was to estimate associations between blood types and levels of circulating inflammatory markers. Because we analyzed individuals selected for case–control studies of cancer, as opposed to a random selection of individuals, we first developed a set of propensity score-adjusted sampling weights [67, 68] as previously used in similar studies [69] to estimate each participant's probability of being included in each of the eight component studies. Weights were generated separately by gender and case–control status estimate weights, we fit logistic regressions on (a) all PLCO screening arm cases (for case models) or (b) all PLCO screening arm potential controls (for control models) where the outcomes were inclusion in a given study and the predictors were age (5-year categories), smoking status (never/former/current), and pack years for former and current smokers (<15, 15+). We imposed basic inclusion criteria (having a valid blood draw, providing consent, having questionnaire data, complete smoking history, no history of cancer, non-Hispanic white) on all subjects. In addition, we applied any study inclusion criteria to applicable models, such as limiting controls in the endometrial cancer study to women with an intact uterus. We estimated predicted probabilities from each regression, and constructed weights as the inverse of predicted probabilities. Thus, cancer cases, all of whom were included in our study given the case–control sampling scheme, were given low weights in comparison to controls, of whom only 2064 (3.8%) were selected. Because our analysis was weighted to the entire PLCO screening arm, individual weights summed to the total of number of participants meeting eligibility criteria ($N = 56,709$). In addition, because we performed weighting separately by sex and case–control status, summed weights in each sex and case/non-case stratum of our study equaled the total number of eligible PLCO screening arm participants with the same sex and case/non-case status (male case $N = 1054$, female case $N = 1042$; male non-case $N = 28,244$; female non-case $N = 26,369$).

Virtually all markers were measured in more than one study, so we then generated marker-specific weights by taking the average of study-specific weights for studies including that marker (sixteen different combinations). For instance, interferon gamma was measured in the lung, NHL, and ovarian cancer studies. Weights for this marker were computed as the average of lung, NHL, and ovarian cancer weights for women and the average of lung and NHL weights for men, as men were ineligible for the ovarian cancer subsample.

We next fit-weighted linear or logistic regressions to assess associations between log-transformed marker concentrations (outcome) and blood types (exposure), accounting for the multiple sampling schemes with the survey package in R [70]. The type of regression depended on the percentage of marker samples that fell below a given assay's lower limit of detection (LLOD): for analytes with fewer than 50% of samples below the LLOD (64 markers), we used linear regression, imputing concentrations below the LLOD as one-half the LLOD, and for analytes with 50–90% of samples below the LLOD (31 markers), we used logistic regression, dichotomizing concentrations as detectable or undetectable. One marker (soluble CD40 ligand; CD40L) with 71% of samples above the upper limit of quantitation was also analyzed with a logistic model, although in this instance, quantitated samples were set as the reference group. For all other markers, the percentage of samples with concentrations above the upper limit of detection was small and ignorable (<5%; Supplementary Table 8); we set the values of these samples as the upper limit of quantitation.

In accordance with phenotypic inheritance patterns, we evaluated associations for all blood types using categorical genetic models. Wald tests were used to determine joint statistical significance for all levels of a given blood group combined. Because few participants were homozygous recessive for Colton ($n = 3$), Knops ($n = 5$), and Kell ($n = 4$) phenotypes, we combined heterozygous and homozygous recessive individuals for these blood groups, thus comparing marker levels for carriers of any recessive allele versus carriers of two dominant alleles. We excluded participants with weak (Fyx; $n = 1$) or null ($n = 4$) Duffy antigen types and participants with null Lewis types ($n = 30$) from our models due to small sample size. To examine the robustness of findings to alternate model specifications, we also estimated associations using additive genetic models, where the primary predictors were number of minor alleles for a given blood group. All models were adjusted for age (5-year categories), sex, study, the first-five principal components of population substructure, and smoking status (never/former/current). In addition, to determine whether imputation of marker concentrations below an assay's LLOD affected our findings, we also performed sensitivity analyses by fitting Tobit models for all markers evaluated with linear regression [71]. Each Tobit model was left-censored at an assay's LLOD plus a small constant.

To further explore which blood group antigens contributed to significant results, we conducted post hoc pairwise t-tests to compare the estimated marginal mean marker levels of different groups. Marginal means were calculated

using the beta values from the respective regressions. To explore the relationship between significantly associated markers, we report Pearson's partial correlations, controlling for study, for all markers associated with the same blood group. Finally, to determine whether results were generalizable to other races and ethnicities, we modeled all significant blood-type-marker associations identified among non-Hispanic whites in Black ($n = 322$), Asian ($n = 107$), and Hispanic ($n = 66$) participants using the same covariate adjustments as in main analyses. All statistical tests were two-sided with a Bonferroni threshold of $p < 4.3 \times 10^{-4}$ considered statistically significant (based on 12 blood types $\times$ 96 markers $= 1152$ tested associations).

Code availability

Code used for analyses can be accessed at https://github.com/sarahVanAlsten/blood_type_inflammation.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Nydegger UE, Tevaearai H, Berdat P, Rieben R, Carrel T, Mohacsi P, et al. Histo-blood group antigens as allo- and auto-antigens. Ann N Y Acad Sci. 2005;1050:40–51. https://doi.org/10.1196/annals.1313.006.

2. Poole J, Daniels G. Blood group antibodies and their significance in transfusion medicine. Transfus Med Rev. 2007;21:58–71. https://doi.org/10.1016/j.tmrv.2006.08.003.

3. Cooling L. Blood groups in infection and host susceptibility. Clin Microbiol Rev. 2015;28:801–70. https://doi.org/10.1128/CMR.00109-14.

4. Bohmig GA, Farkas AM, Eskandary F, Wekerle T. Strategies to overcome the abo barrier in kidney transplantation. Nat Rev Nephrol. 2015;11:732–47. https://doi.org/10.1038/nrneph.2015.144.

5. Hendrickson JE, Delaney M. Hemolytic disease of the fetus and newborn: modern practice and future investigations. Transfus Med Rev. 2016;30:159–64. https://doi.org/10.1016/j.tmrv.2016.05.008.

6. King CL, Adams JH, Xianli J, Grimberg BT, McHenry AM, Greenberg LJ, et al. Fy(a)/Fy(b) antigen polymorphism in human erythrocyte Duffy antigen affects susceptibility to plasmodium vivax malaria. Proc Natl Acad Sci U S A. 2011;108:20113–8. https://doi.org/10.1073/pnas.1109621108.

7. Moonens K, Gideonsson P, Subedi S, Bugaytsova J, Romão E, Mendez M, et al. Structural insights into polymorphic abo glycan binding by helicobacter pylori. Cell Host Microbe. 2016;19:55–66. https://doi.org/10.1016/j.chom.2015.12.004.

8. Guillon P, Clement M, Sebille V, Rivain JG, Chou CF, Ruvoen-Clouet N, et al. Inhibition of the interaction between the SARS-CoV spike protein and its cellular receptor by anti-histo-blood group antibodies. Glycobiology. 2008;18:1085–93. https://doi.org/10.1093/glycob/cwn093.

9. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide association study of severe COVID-19 with respiratory failure. N Engl J Med. 2020;383:1522–34. https://doi.org/10.1056/NEJMoa2020283.

10. Kolin DA, Kulm S, Christos PJ, Elemento O. Clinical, regional, and genetic characteristics of COVID-19 patients from UK Biobank. PLOS ONE. 2020;15:e0241264. https://doi.org/10.1371/journal.pone.0241264.

11. Barnkob MB, Pottegård A, Støvring H, Haunstrup TM, Homburg K, Larsen R, et al. Reduced prevalence of SARS-CoV-2 infection in ABO blood group O. Blood Adv. 2020;4:4990–3. https://doi.org/10.1182/bloodadvances.2020002657.

12. Hoiland RL, Fergusson NA, Mitra AR, Griesdale DEG, Devine DV, Stukas S, et al. The association of abo blood group with indices of disease severity and multiorgan dysfunction in COVID-19. Blood Adv. 2020;4:4981–9. https://doi.org/10.1182/bloodadvances.2020002623.

13. Ray JG, Schull MJ, Vermeulen MJ, Park AL. Association between ABO and Rh blood groups and SARS-CoV-2 infection or severe COVID-19 illness. Ann Intern Med. 2021;174:308–15. https://doi.org/10.7326/m20-4511.

14. Cohen M, Hurtado-Ziola N, Varki A. ABO blood group glycans modulate sialic acid recognition on erythrocytes. Blood. 2009;114:3668–76. https://doi.org/10.1182/blood-2009-06-227041.

15. Varki A. Biological roles of glycans. Glycobiology. 2017;27:3–49. https://doi.org/10.1093/glycob/cww086.

16. Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS Genet. 2008;4:e1000072. https://doi.org/10.1371/journal.pgen.1000072.

17. Preston RJ, Rawley O, Gleeson EM, O'Donnell JS. Elucidating the role of carbohydrate determinants in regulating hemostasis: insights and opportunities. Blood. 2013;121:3801–10. https://doi.org/10.1182/blood-2012-10-415000.

18. Polfus LM, Raffield LM, Wheeler MM, Tracy RP, Lange LA, Lettre G, et al. Whole genome sequence association with e-selectin levels reveals loss-of-function variant in African Americans. Hum Mol Genet. 2019;28:515–23. https://doi.org/10.1093/hmg/ddy360.

19. Sliz E, Kalaoja M, Ahola-Olli A, Raitakari O, Perola M, Salomaa V, et al. Genome-wide association study identifies seven novel loci associating with circulating cytokines and cell adhesion molecules in Finns. J Med Genet. 2019;56:607–16. https://doi.org/10.1136/jmedgenet-2018-105965.

20. Kiechl S, Pare G, Barbalic M, Qi L, Dupuis J, Dehghan A, 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;4:681–6. https://doi.org/10.1161/CIRCGENETICS.111.960682.

21. Aslibekyan S, Agha G, Colicino E, Do AN, Lahti J, Ligthart S, et al. Association of methylation signals with incident coronary heart disease in an epigenome-wide assessment of circulating tumor necrosis factor alpha. JAMA Cardiol. 2018;3:463–72. https://doi.org/10.1001/jamacardio.2018.0510.

22. Fukuma N, Akimitsu N, Hamamoto H, Kusuhara H, Sugiyama Y, Sekimizu K. A role of the Duffy antigen for the maintenance of

plasma chemokine concentrations. 2003;303:137–9. https://doi.org/10.1016/s0006-291x(03)00293-6.

23. Pruenster M, Mudde L, Bombosi P, Dimitrova S, Zsak M, Middleton J, et al. The Duffy antigen receptor for chemokines transports chemokines and supports their promigratory activity. Nat Immunol. 2009;10:101–8. https://doi.org/10.1038/ni.1675.

24. Comerford I, Nibbs RJ. Post-translational control of chemokines: a role for decoy receptors? Immunol Lett. 2005;96:163–74. https://doi.org/10.1016/j.imlet.2004.08.018.

25. Comuzzie AG, Cole SA, Laston SL, Voruganti VS, Haack K, Gibbs RA, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the hispanic population. PLoS ONE. 2012;7: e51954. https://doi.org/10.1371/journal.pone.0051954.

26. Wang J, Ou ZL, Hou YF, Luo JM, Shen ZZ, Ding J, et al. Enhanced expression of Duffy antigen receptor for chemokines by breast cancer cells attenuates growth and metastasis potential. Oncogene. 2006;25:7201–11. https://doi.org/10.1038/sj.onc.1209703.

27. Maeda S, Kuboki S, Nojima H, Shimizu H, Yoshitomi H, Furukawa K, et al. Duffy antigen receptor for chemokines (DARC) expressing in cancer cells inhibits tumor progression by suppressing CXCR2 signaling in human pancreatic ductal adenocarcinoma. Cytokine. 2017;95:12–21. https://doi.org/10.1016/j.cyto.2017.02.007.

28. Addison CL, Belperio JA, Burdick MD, Strieter RM. Overexpression of the Duffy antigen receptor for chemokines (DARC) by NSCLC tumor cells results in increased tumor necrosis. BMC Cancer. 2004;4:28. https://doi.org/10.1186/1471-2407-4-28.

29. Pavlakovic H, Becker J, Albuquerque R, Wilting J, Ambati J. Soluble VEGFR-2: an antilymphangiogenic variant of VEGF receptors. Ann N Y Acad Sci. 2010;1207:E7–15. https://doi.org/10.1111/j.1749-6632.2010.05714.x.

30. Albuquerque RJ, Hayashi T, Cho WG, Kleinman ME, Dridi S, Takeda A, et al. Alternatively spliced vascular endothelial growth factor receptor-2 is an essential endogenous inhibitor of lymphatic vessel growth. Nat Med. 2009;15:1023–30. https://doi.org/10.1038/nm.2018.

31. Singh N, Tiem M, Watkins R, Cho YK, Wang Y, Olsen T, et al. Soluble vascular endothelial growth factor receptor 3 is essential for corneal alymphaticity. Blood. 2013;121:4242–9. https://doi.org/10.1182/blood-2012-08-453043.

32. Vasan SK, Hwang J, Rostgaard K, Nyren O, Ullum H, Pedersen OBV, et al. ABO blood group and risk of cancer: a register-based cohort study of 1.6 million blood donors. Cancer Epidemiol. 2016;44:40–3. https://doi.org/10.1016/j.canep.2016.06.005.

33. Becker J, Pavlakovic H, Ludewig F, Wilting F, Weich HA, Albuquerque R, et al. Neuroblastoma progression correlates with downregulation of the lymphangiogenesis inhibitor sVEGFR-2. Clin Cancer Res. 2010;16:1431–41. https://doi.org/10.1158/1078-0432.CCR-09-1936.

34. Shibata MA, Ambati J, Shibata E, Albuquerque RJ, Morimoto J, Ito Y, et al. The endogenous soluble vegf receptor-2 isoform suppresses lymph node metastasis in a mouse immunocompetent mammary cancer model. BMC Med. 2010;8:69 https://doi.org/10.1186/1741-7015-8-69.

35. Shiels MS, Shu XO, Chaturvedi AK, Gao YT, Xiang YB, Cai Q, et al. A prospective study of immune and inflammation markers and risk of lung cancer among female never smokers in shanghai. Carcinogenesis. 2017;38:1004–10. https://doi.org/10.1093/carcin/bgx075.

36. Wolpin BM, Chan AT, Hartge P, Chanock SJ, Kraft P, Hunter DJ, et al. ABO blood group and the risk of pancreatic cancer. J Natl Cancer Inst. 2009;101:424–31. https://doi.org/10.1093/jnci/djp020.

37. Risch HA, Lu L, Wang J, Zhang W, Ni Q, Gao YT, et al. ABO blood group and risk of pancreatic cancer: a study in shanghai and meta-analysis. Am J Epidemiol. 2013;177:1326–37. https://doi.org/10.1093/aje/kws458.

38. Poole EM, Gates MA, High BA, Chanock SJ, Cramer DW, Cunningham JM, et al. ABO blood group and risk of epithelial ovarian cancer within the ovarian cancer association consortium. Cancer Causes Control. 2012;23:1805–10. https://doi.org/10.1007/s10552-012-0059-y.

39. Risch HA, Yu H, Lu L, Kidd MS. ABO blood group, helicobacter pylori seropositivity, and risk of pancreatic cancer: a case-control study. J Natl Cancer Inst. 2010;102:502–5. https://doi.org/10.1093/jnci/djq007.

40. Ackermann M, Verleden SE, Kuehnel M, Haverich A, Welte T, Laenger F, et al. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in covid-19. N Engl J Med. 2020;383:120–8. https://doi.org/10.1056/NEJMoa2015432.

41. Jostock T, Mullberg J, Ozbek S, Atreya R, Blinn G, Voltz N, et al. Soluble gp130 is the natural inhibitor of soluble interleukin-6 receptor transsignaling responses. Eur J Biochem. 2001;268:160–7. https://doi.org/10.1046/j.1432-1327.2001.01867.x.

42. Wolf J, Rose-John S, Garbers C. Interleukin-6 and its receptors: a highly regulated and dynamic system. Cytokine. 2014;70:11–20. https://doi.org/10.1016/j.cyto.2014.05.024.

43. Interleukin-6 Receptor Mendelian Randomisation Analysis Consortium, Swerdlow DI, Holmes MV, Kuchenbaecker KB, Engmann JE, Shah T, et al. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. Lancet. 2012;379:1214–24. https://doi.org/10.1016/S0140-6736(12)60110-X.

44. Schuett H, Oestreich R, Waetzig GH, Annema W, Luchtefeld M, Hillmer A, et al. Transsignaling of interleukin-6 crucially contributes to atherosclerosis in mice. Arterioscler Thromb Vasc Biol. 2012;32:281–90. https://doi.org/10.1161/ATVBAHA.111.229435.

45. Gwechenberger M, Pacher R, Berger R, Zorn G, Moser P, Stanek B, et al. Comparison of soluble glycoprotein 130 and cardiac natriuretic peptides as long-term predictors of heart failure progression. J Heart Lung Transpl. 2005;24:2190–5.

46. Garcia-Closas M, Ye Y, Rothman N, Figueroa JD, Malats N, Dinney C, et al. A genome-wide association study of bladder cancer identifies a new susceptibility locus within SLC14A1, a urea transporter gene on chromosome 18q12. 3. Hum Mol Gen. 2011;20:4282–9. https://doi.org/10.1093/hmg/ddr342.

47. Gohagan JK, Prorok PC, Hayes RB, Kramer BS, Ovarian Cancer Screening Trial Project T, Prostate LC. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the national cancer institute: history, organization, and status. Control Clin Trials. 2000;21:251S–272S. https://doi.org/10.1016/s0197-2456(00)00097-0.

48. Prorok PC, Andriole GL, Bresalier RS, Buys SS, Chia D, Crawford ED, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. Control Clin Trials. 2000;21:273S–309S. https://doi.org/10.1016/s0197-2456(00)00098-2.

49. Purdue MP, Hofmann JN, Kemp TJ, Chaturvedi AK, Lan Q, Park JH, et al. A prospective study of 67 serum immune and inflammation markers and risk of non-hodgkin lymphoma. Blood. 2013;122:951–7. https://doi.org/10.1182/blood-2013-01-481077.

50. Shiels MS, Pfeiffer RM, Hildesheim A, Engels EA, Kemp TJ, Park JH, et al. Circulating inflammation markers and prospective risk for lung cancer. J Natl Cancer Inst. 2013;105:1871–80. https://doi.org/10.1093/jnci/djt309.

51. Trabert B, Pinto L, Hartge P, Kemp T, Black A, Sherman ME, et al. Pre-diagnostic serum levels of inflammation markers and risk of ovarian cancer in the Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial. Gynecol Oncol. 2014;135:297–304. https://doi.org/10.1016/j.ygyno.2014.08.025.

52. Trabert B, Eldridge RC, Pfeiffer RM, Shiels MS, Kemp TJ, Guillemette C, et al. Prediagnostic circulating inflammation markers and endometrial cancer risk in the Prostate, Lung, Colorectal

and Ovarian Cancer (PLCO) Screening Trial. Int J Cancer. 2017;140:600–10. https://doi.org/10.1002/ijc.30478.

53. Huang WY, Berndt SI, Shiels MS, Katki HA, Chaturvedi AK, Wentzensen N, et al. Circulating inflammation markers and colorectal adenoma risk. Carcinogenesis. 2019;40:765–70. https://doi.org/10.1093/carcin/bgz027.

54. Shiels MS, Katki HA, Hildesheim A, Pfeiffer RM, Engels EA, Williams M, et al. Circulating inflammation markers, risk of lung cancer, and utility for risk stratification. J Natl Cancer Inst. 2015;107. https://doi.org/10.1093/jnci/djv199.

55. Cook MB, Barnett MJ, Bock CH, Cross AJ, Goodman PJ, Goodman GE, et al. Prediagnostic circulating markers of inflammation and risk of oesophageal adenocarcinoma: a study within the national cancer institute cohort consortium. Gut. 2019;68:960–8. https://doi.org/10.1136/gutjnl-2018-316678.

56. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, et al. Reference-based phasing using the haplotype reference consortium panel. Nat Genet. 2016;48:1443–8. https://doi.org/10.1038/ng.3679.

57. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 2021;590:290–9. https://doi.org/10.1038/s41586-021-03205-y.

58. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44:955–9. https://doi.org/10.1038/ng.2354.

59. Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, et al. Use of >100,000 nhlbi trans-omics for precision medicine (topmed) consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed african and hispanic/latino populations. PLoS Genet. 2019;15: e1008500. https://doi.org/10.1371/journal.pgen.1008500.

60. Jin Y, Schaffer AA, Sherry ST, Feolo M. Quickly identifying identical and closely related subjects in large databases using genotype data. PLoS ONE. 2017;12:e0179106. https://doi.org/10.1371/journal.pone.0179106.

61. Jin Y, Schaffer AA, Feolo M, Holmes JB, Kattman BL. GRAF-pop: a fast distance-based method to infer subject ancestry from multiple genotype datasets without principal components analysis. G3. 2019;9:2447–61. https://doi.org/10.1534/g3.118.200925.

62. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38:904–9. https://doi.org/10.1038/ng1847.

63. Reid ME, Denomme GA. DNA-based methods in the immuno-hematology reference laboratory. Transfus Apher Sci. 2011;44:65–72. https://doi.org/10.1016/j.transci.2010.12.011.

64. Cakir B, Pankow JS, Salomaa V, Couper D, Morris TL, Brantley KR, et al. Distribution of lewis (FUT3)genotype and allele: Frequencies in a biethnic united states population. Ann Hematol. 2002;81:558–65. https://doi.org/10.1007/s00277-002-0508-x.

65. Parsons SF, Mallinson G, Daniels GL, Green CA, Smythe JS, Anstee DJ. Use of domain-deletion mutants to locate lutheran blood group antigens to each of the five immunoglobulin super-family domains of the lutheran glycoprotein: elucidation of the molecular basis of the Lu(a)/Lu(b) and the Au(a)/Au(b) polymorphisms. Blood. 1997;89:4219–25.

66. Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB. Sequence and expression of a candidate for the human secretor blood group alpha(1,2)fucosyltransferase gene (fut2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. J Biol Chem. 1995;270:4640–9. https://doi.org/10.1074/jbc.270.9.4640.

67. Korn EL, Graubard BI. Analysis of health surveys. 1st ed. New York, NY: John Wiley & Sons; 1999.

68. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivar Behav Res. 2011;46:399–424. https://doi.org/10.1080/00273171.2011.568786.

69. Kitahara CM, Trabert B, Katki HA, Chaturvedi AK, Kemp TJ, Pinto LA, et al. Body mass index, physical activity, and serum markers of inflammation, immunity, and insulin resistance. Cancer Epidemiol Biomark Prev. 2014;23:2840–9. https://doi.org/10.1158/1055-9965.EPI-14-0699-T.

70. Lumley T. Analysis of complex survey samples. J Stat Softw. 2004;9:1–19.

71. Persson T, Rootzen H. Simple and highly efficient estimators for a type i censored normal sample. Biometrika. 1977;64:123–8. https://doi.org/10.1093/biomet/64.1.123.