blood advances

Ancestry-independent risk of venous thromboembolism in individuals with sickle cell trait vs factor V Leiden

Keng-Han Lin,[1] Julie M. Granka,[1] Anjali J. Shastri,[1] Vence L. Bonham,[2,*] and Rakhi P. Naik,[3,*] the 23andMe Research Team

[1]23andMe, Sunnyvale, CA; [2]National Human Genome Research Institute, Social and Behavioral Research Branch, National Institutes of Health, Bethesda, MD; and [3]Division of Hematology, Department of Medicine, Johns Hopkins University, Baltimore, MD

Sickle cell trait (SCT) is a risk factor for venous thromboembolism (VTE). Prior studies investigating the association between SCT and VTE have been performed nearly exclusively in Black populations. However, race-based research can contribute to systemic racism in medicine. We leveraged data from the 23andMe research cohort (4 184 082 participants) to calculate the ancestry-independent risk of VTE associated with SCT as well as comparative risk estimates for heterozygous factor V Leiden (FVL). Odds ratios (ORs) were calculated using a meta-analysis of 3 genetic ancestry groups (European [n = 3 183 142], Latine [n = 597 539], and African [n = 202 281]) and a secondary full-cohort analysis including 2 additional groups (East Asian [n = 159 863] and South Asian [n = 41 257]). Among the full cohort, 94 323 participants (2.25%) reported a history of VTE. On meta-analysis, individuals with SCT had a 1.45-fold (confidence interval [CI], 1.32-1.60) increased risk of VTE compared with SCT noncarriers, which was similar to the full-cohort estimate. The risk of pulmonary embolism (PE) in SCT (OR, 1.95; CI, 1.72-2.20) was higher than that of isolated deep venous thrombosis (DVT; OR, 1.04; CI, 0.90-1.21). FVL carriers had 3.30-fold (CI, 3.24-3.37) increased risk of VTE compared with FVL noncarriers, with a higher risk of isolated DVT (OR, 3.59; CI, 3.51-3.68) than PE (OR, 2.72; CI, 2.64-2.81). In this large, diverse cohort, the risk of VTE was increased among individuals with SCT compared with those without, independent of race or genetic ancestry. The risk of VTE with SCT was lower than that observed in FVL; however, the pattern of VTE in SCT was PE predominant, which is the opposite to that observed in FVL.

Key Points

- The risk of VTE was increased among individuals with SCT, independent of genetic ancestry, and this risk was lower than heterozygous FVL.

- The risk of PE in SCT is significantly higher than the risk of isolated DVT; this pattern suggests a unique mechanism of thrombosis in SCT.

Introduction

Three million individuals in the United States and an estimated 300 million people worldwide have sickle cell trait (SCT). Due to its protective effects against severe malarial infection, SCT prevalence is highest in global regions with high malaria endemicity, including West/Central Africa, Mediterranean Europe, India, and the Middle East.[1] In the Americas, the highest prevalence of SCT is found in individuals with ancestry from West and Central Africa.[2] As a result, despite its widespread geographic and population distribution, SCT in the United States is often incorrectly assumed to affect only those who identify as Black or African American.[3]

SCT is generally an asymptomatic carrier state but has been found to be a risk factor for select clinical outcomes, including venous thromboembolism (VTE), conferring a 1.6- to 1.8-fold increased risk of VTE compared with individuals without SCT.[4] However, prior large studies investigating the association between SCT and VTE have been performed nearly exclusively in Black individuals.[5,6] Although prior studies have often had to limit their analyses to high prevalence populations due to finite funding, the recent availability of cheaper and more readily available genetic testing technology promises to allow for affordable analysis of more inclusive populations. In addition, there has been increasing recognition that race-based research can contribute to bias and inequities.[7] In light of these concerns, the National Academies of Science, Engineering, and Medicine recently issued a consensus committee report recommending against the use of race as a proxy for human genetic variation or assigning genetic ancestry group labels to individuals based on their race.[8,9]

In the context of SCT, racialization has led to discrimination and stigmatization of Black people in the United States. Although the prevalence of SCT in the United States is currently highest among individuals who identify as Black, the perpetuation of SCT as a "Black trait" has resulted in and continues to cause considerable harm. Historically, Black individuals with SCT have faced employment and insurance discrimination due to misinformation about the health risks of SCT.[10-12] More recently, when coupled with systemic racism, biased interpretation of epidemiologic studies has proved particularly harmful and has led to overattribution of SCT to systemic disease. For example, a 2021 *New York Times* investigative report uncovered the inappropriate use of SCT as a rationale for deaths of Black Americans who died in police custody.[13]

The aim of this study was to provide appropriate contextualization of the association between SCT and VTE by evaluating the risk of VTE among individuals with SCT, irrespective of race or ethnicity, and by calculating direct comparative risk estimates with heterozygous factor V Leiden (FVL). Leveraging the large, diverse cohort of participants enrolled in the 23andMe research program, we evaluated the association between SCT and VTE across multiple genetic ancestry groups, defined by genetic similarity, to establish ancestry-independent risk estimates for VTE, pulmonary embolism (PE), and isolated deep venous thrombosis (DVT). We also evaluated the association between FVL and VTE to provide comparative risk estimates to a well-established inherited thrombophilia.

Methods

Study population

Individuals included in this study are research participants of 23andMe Inc, a direct-to-consumer genetics company, who were genotyped as part of the 23andMe personal genome service. Participants provided informed consent and volunteered to participate in the research online, under a protocol approved by the external Association for Accreditation of Human Research Protection Programs–accredited Institutional Review Board, Ethical & Independent Review Services. Most participants of 23andMe are from the United States (~90%), with the remainder from Canada, Great Britain, and other countries where 23andMe conducts research. Individuals of the ages 18 to 100 years (inclusive) were included in this analysis.

Genotyping and phenotyping

DNA samples were genotyped using several customized Illumina single nucleotide polymorphism arrays, as has been previously described in detail.[14] For hemoglobin variants, available genotyping included hemoglobin S (HbS; rs334), HbC (rs33930165), and 9 β-thalassemia mutations (see supplemental Table 1); α-thalassemia genotyping was not available. We also genotyped 2 common and widely known thrombophilias, FVL mutation (rs6025) and the prothrombin G2021e gene mutation (rs1799963). Additional known thrombophilias were not included in this study. Participants with HbSS as well as carriers of HbC and genotyped β-thalassemia mutations were excluded. We also excluded individuals with homozygous FVL and all carriers of the prothrombin G20210A gene mutation.

The outcome of VTE, as well as the subsets of any DVT and PE, was assessed by self-report on a baseline health questionnaire (see supplemental Table 2). Individuals were asked, "have you ever been diagnosed with, or treated for, a blood clot or condition causing repeated blood clots?" If respondents answered yes, they were asked to select the type(s) of blood clot(s) with which they were diagnosed from a list of options.

Provoking factors were also assessed by self-report on a baseline questionnaire (see supplemental Table 2); only a single provoking factor could be chosen. For this study, we defined provoked VTE as those associated with any of the categorically assessed factors of "another condition, such as cancer," "medication," "pregnancy," "prolonged immobility," "overweight," or "surgery or injury."

We performed association analyses separately for the following phenotypes: VTE, isolated DVT, and PE (±DVT).

Genetic ancestry and relatedness calculations

To reduce the effects of population structure on downstream association analyses, we performed analyses stratified by genetic similarity. We inferred local genetic ancestry using 23andMe's ancestry composition algorithm,[15] then grouped individuals into genetic ancestry groups by their proportion of inferred genetic ancestry, as well as the lengths and numbers of segments of inferred ancestry (see supplemental Methods). Genetic ancestry groups evaluated via this method include African, East Asian, European, Latine, and South Asian, following the precedent of prior 23andMe studies.[14,16-18] We acknowledge that this is 1 of many possible ways to group individuals by genetic similarity for the purpose of minimizing the effects of population structure on downstream analyses. Per the recommendations of the National Academies for Science, Engineering, and Medicine,[8] the population descriptors we use in this study are solely for the purpose of reflecting genetic similarity; see supplemental Methods and supplemental Table 4 for additional details.

For the association analyses within each genetic ancestry group, a maximal set of unrelated individuals was filtered using a segmental identity-by-descent (IBD) estimation algorithm.[19] Following prior 23andMe association studies,[14,18] participants are defined as related if they share >700 cM IBD, including the regions where the 2 individuals share either 1 or both genomic segments IBD. This level of relatedness (~20% of the genome) corresponds approximately to the minimal expected sharing between first cousins in a nonendogamous population.

Association analyses

Logistic regression was used to calculate odds ratios (ORs) and 95% confidence intervals (CIs) for its association with HbS carrier status and FVL carrier status within the European, African, and Latine genetic ancestry groups. East Asian and South Asian genetic ancestry groups were excluded from the within-genetic ancestry group analyses because very few (<5) VTE events among SCT carriers were observed in those groups (note that, to respect 23andMe research participant privacy, we are unable to provide counts of results with n <5). We included covariates for age, sex, genotyping platform, and 9 principal components (PCs; see supplemental Methods) to account for residual population structure. A primary crossancestry meta-analysis was performed across the 3 studied genetic ancestry groups (European, African, and Latine), using a fixed effects model (inverse variance method[20]). As a comparison with the meta-analysis, a secondary mega-analysis (including individuals across all 5 studied genetic ancestry groups but in a single association analysis) was performed, including covariates for age, sex, genotyping platform, and top 20 PCs to account for residual population structure. Significance was conservatively defined as P value < .017 (0.05/3) to account for multiple comparisons across the 3 studied phenotypes (VTE, isolated DVT, and PE).

Results

Baseline characteristics

Starting with all participants who were classified into 1 of the 5 genetic ancestry groups, we excluded those with missing covariate or outcome data (n = 2 153 141), individuals with HbSS disease, HbC, or β-thalassemia mutations (n = 25 033), and those with homozygous FVL or prothrombin gene mutation (n = 221 907). In total, 4 184 082 participants were included in the final full cohort. The baseline characteristics of the studied cohort and of each genetic ancestry group are described in Table 1. The prevalence of SCT in the overall cohort was 0.46% (n ≥ 19 055). Among genetic ancestry groups, SCT prevalence differed significantly, with the highest prevalence among individuals of African ancestry (7.02%) and the lowest prevalence among individuals of European (0.02%) and East Asian ancestry (~0%). In contrast, the prevalence of FVL in the overall cohort was 4.45% (n = 186 277), with the highest prevalence among individuals of European ancestry (5.19%) and the lowest prevalence among individuals of East Asian ancestry

(0.06%; note again that, to respect 23andMe research participant privacy, we are unable to provide specific statistics or counts for n <5).

A history of VTE was reported by ~490 of 19 055 individuals (2.57%) with SCT compared with ~93 833 of 4 165 027 participants (2.25%) without SCT. A history of PE was found in ~301 of 19 055 individuals (1.58%) with SCT compared with ~37 552 of 4 165 027 (0.90%) without SCT. The prevalence of isolated DVT was 1.01% in SCT carriers and 1.36% in noncarriers (Table 2). In addition, 219 of 490 individuals (44.69%) with SCT reported a provoked VTE compared with 42 544 of 93 833 (45.34%) without SCT. Accounting for genetic ancestry group, we did not find a significant difference in the proportion of individuals with provoked VTE between individuals with and without SCT (see supplemental Methods).

Among SCT carriers with VTE, 41.22% reported a history of PE alone, 20.20% had a history of both PE and DVT, and 38.57% reported a history of DVT only (see also supplemental Table 3 for a breakdown across the full cohort).

Among carriers of FVL, a history of VTE was reported by ~12 912 of 186 271 (6.93%) compared with ~81 410 of 3 997 811 individuals (2.04%) without FVL. In participants with FVL, the prevalence of PE was 2.45% and isolated DVT was 4.59%, compared with 0.83% and 1.21%, respectively, in FVL noncarriers (Table 3). In addition, 3977 of 12 913 individuals (30.8%) with FVL reported a provoked VTE compared with 38 786 of 81 410 (47.6%) without FVL. Accounting for genetic ancestry group, we found a significant difference in the proportion of individuals with provoked VTE between FVL carriers and noncarriers, with a lower proportion of FVL carriers reporting provoked VTE (see supplemental Methods).

Meta-analysis results for SCT carriers compared with SCT noncarriers

Table 4 and Figure 1A show the ORs for VTE, PE, and isolated DVT within each genetic ancestry group (European, African, and Latine) and on meta-analysis for individuals with SCT compared with those without. The ORs for VTE ranged from 1.38 (CI, 1.24-1.54; P = 3.1e−09) in individuals of African ancestry to 1.97 (CI, 1.39-2.79; P = .00014) in Europeans. The point estimates (ORs) were higher for PE than VTE in all genetic ancestry groups, ranging from 1.73 (CI, 1.51-1.99; P = 7.7e−15) in participants of African ancestry to 3.19 (CI, 2.08-4.89; P = 1.1e−07) in Europeans.

Table 1. Baseline cohort characteristics, by genetic ancestry group

	Full cohort	European	Latine	African	East Asian	South Asian
Sample size, n (%)	4 184 082	3 183 142 (76.08%)	597 539 (14.28%)	202 281 (4.83%)	159 863 (3.82%)	41 257 (0.99%)
Age, mean (SD), y	49.37 (17.21)	51.67 (17.19)	42.04 (15.11)	43.8 (15.75)	39.97 (14.17)	41.64 (13.57)
Female, n (%)	2 384 323 (56.99%)	1 801 651 (56.60%)	347 700 (58.19%)	120 312 (59.48%)	97 600 (61.05%)	17 121 (41.50%)
SCT, n (%)	≥19 055 (0.46%)	795 (0.02%)	3 990 (0.67%)	14 205 (7.02%)	<5*	65 (0.16%)
FVL, n (%)	186 271 (4.45%)	165 345 (5.19%)	16 909 (2.83%)	2 869 (1.42%)	97 (0.06%)	1 051 (2.55%)
VTE, n (%)	94 323 (2.25%)	82 445 (2.59%)	7 229 (1.21%)	4 005 (1.98%)	433 (0.27%)	211 (0.51%)
PE, n (%)	37 853 (0.90%)	32 830 (1.03%)	2 833 (0.47%)	1 952 (0.97%)	160 (0.10%)	78 (0.19%)
Isolated DVT, n (%)	56 470 (1.36%)	49 615 (1.58%)	4 396 (0.74%)	2 053 (1.03%)	273 (0.17%)	133 (0.32%)

SD, standard deviation.
*Note that in order to respect 23andMe research participant privacy, we are unable to provide specific counts where n < 5.

Table 2. VTE prevalence, by SCT status

	Full cohort		European		Latine		African		East Asian		South Asian	
	Noncarriers	SCT carriers	Noncarriers	SCT carriers	Noncarriers	SCT carriers	Noncarriers	SCT carriers	Noncarriers	SCT carriers	Noncarriers	SCT carriers
VTE, n (%)	93 833 (2.25%)	≥490 (2.57%)	82 411 (2.59%)	34 (4.28%)	7162 (1.21%)	67 (1.68%)	3616 (1.92%)	389 (2.74%)	433 (0.27%)	<5*	211 (0.51%)	<5
PE, n (%)	37 552 (0.90%)	≥301 (1.58%)	32 808 (1.03%)	22 (2.77%)	2789 (0.47%)	44 (1.10%)	1717 (0.91%)	235 (1.65%)	160 (0.10%)	<5	78 (0.19%)	<5
Isolated DVT, n (%)	56 281 (1.36%)	≥189 (1.01%)	49 603 (1.58%)	12 (1.55%)	4373 (0.74%)	23 (0.58%)	1899 (1.02%)	154 (1.10%)	273 (0.17%)	<5	133 (0.32%)	<5

*Note that, to respect 23andMe research participant privacy, we are unable to provide specific counts in which n was <5.

Table 3. VTE prevalence, by FVL status

	Full cohort		European		Latine		African		East Asian		South Asian	
	Noncarriers	FVL carriers	Noncarriers	FVL carriers	Noncarriers	FVL carriers	Noncarriers	FVL carriers	Noncarriers	FVL carriers	Noncarriers	FVL carriers
VTE, n (%)	81 410 (2.04%)	≥12 912 (6.93%)	70 332 (2.33%)	12113 (7.33%)	6582 (1.13%)	647 (3.83%)	3865 (1.94%)	140 (4.88%)	432 (0.27%)	<5*	199 (0.49%)	12 (1.14%)
PE, n (%)	33 286 (0.83%)	≥4 565 (2.45%)	28 552 (0.95%)	4278 (2.59%)	2604 (0.45%)	229 (1.35%)	1894 (0.95%)	58 (2.02%)	159 (0.10%)	<5*	77 (0.19%)	<5*
Isolated DVT, n (%)	48 124 (1.21%)	≥8 346 (4.59%)	41 780 (1.40%)	7835 (4.87%)	3978 (0.69%)	418 (2.51%)	1971 (1.00%)	82 (2.92%)	273 (0.17%)	<5*	122 (0.30%)	11 (1.05%)

*Note that, to respect 23andMe research participant privacy, we are unable to provide specific counts in which n was <5.

Table 4. Association of SCT and VTE outcomes within genetic ancestry groups and in a crosspopulation meta-analysis

	European	Latine	African	Meta-analysis
VTE				
OR (95% CI)	1.97 (1.39-2.79)	1.62 (1.27-2.08)	1.38 (1.24-1.54)	1.45 (1.32-1.60)
P value	.00014	.00013	3.1e−09	1.1e−14
PE				
OR (95% CI)	3.19 (2.08-4.89)	2.67 (1.97-3.64)	1.73 (1.51-1.99)	1.95 (1.72-2.20)
P value	1.1e−07	3.6e−10	7.7e−15	4.9e−27
Isolated DVT				
OR (95% CI)	1.15 (0.65-2.04)	0.93 (0.61-1.41)	1.05 (0.89-1.25)	1.04 (0.90-1.21)
P value	.64	.73	.53	.58

Isolated DVT risk estimates were close to 1 for all genetic ancestry groups. Using a test for heterogeneity (see supplemental Methods), the risk estimates were not significantly different across genetic ancestry groups for either VTE ($P = .19$) or isolated DVT ($P = .6$); however, the PE estimates were significantly different ($P = .009$). We investigated the potential reasons for this difference; 1 hypothesis is that because the African genetic ancestry group has a higher baseline risk for PE, the additional effect of having SCT is attenuated. Supplemental Figure 2C demonstrates that individuals without SCT or FVL in the African Ancestry cohort demonstrate a higher prevalence of PE, especially across younger age groups, than the European cohort. Causes for this increased baseline risk in the African genetic ancestry group were not further investigated.

On meta-analysis, individuals with SCT had a 1.45-fold (CI, 1.32-1.60; $P = 1.1e−14$) higher risk of VTE than those without. This risk was predominantly explained by an increased risk of PE (OR, 1.95; CI, 1.72-2.20; $P = 4.9e−27$) compared with isolated DVT (OR, 1.04; CI 0.90-1.21; $P = .58$). In the meta-analysis, CIs for PE did not overlap 1, whereas CIs for isolated DVT did. Additionally, the meta-analyzed CIs for PE and isolated DVT did not overlap (Figure 1B).

Meta-analysis results for FVL carriers compared with FVL noncarriers

ORs for VTE, PE, and isolated DVT within each ancestry group and on meta-analysis for individuals with FVL compared with those without are shown in Table 5 and Figure 2A. The risk estimates for
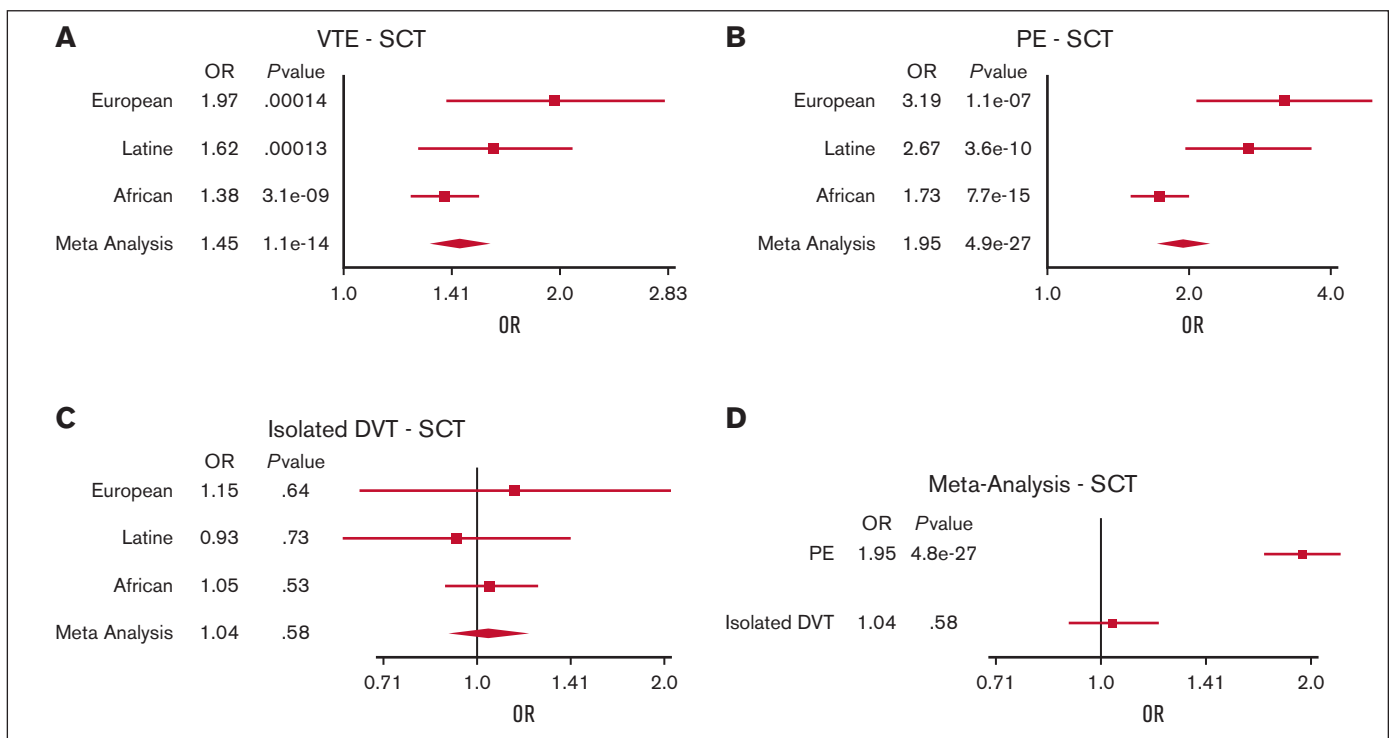
Figure 1. Risk estimates for VTE in sickle cell trait. (A-C) Meta-analysis of ORs for VTE, PE, and isolated DVT, respectively, comparing SCT carriers with noncarriers. (D) Meta-analysis estimate for PE vs isolated DVT comparing SCT carriers with noncarriers.

Table 5. Association of FVL and VTE outcomes within genetic ancestry groups and in a crosspopulation meta-analysis

	European	Latine	African	Meta-analysis
VTE				
OR (95% CI)	3.31 (3.25-3.38)	3.24 (2.98-3.52)	2.83 (2.37-3.38)	3.30 (3.24-3.37)
P value	<2.3e−308	5.5e−164	1.5e−30	<2.3e−308
PE				
OR (95% CI)	2.72 (2.64-2.82)	2.74 (2.39-3.15)	2.40 (1.83-3.13)	2.72 (2.64-2.81)
P value	<2.3e−308	1.1e−46	1.5e−10	<2.3e−308
Isolated DVT				
OR (95% CI)	3.61 (3.52-3.70)	3.49 (3.14-3.87)	3.12 (2.48-3.93)	3.59 (3.51-3.68)
P value	<2.3e−308	4.6e−122	2.5e−22	<2.3e−308

the included outcomes were similar within each genetic ancestry group. Using a test for heterogeneity (see supplemental Methods), the risk estimates for all phenotypes were not significantly different across genetic ancestry groups (any DVT, $P = .10$; PE, $P = .56$; isolated DVT, $P = .25$). On meta-analysis, individuals with FVL had a 3.30-fold (CI, 3.24-3.37; $P < 2.3e−308$) higher risk of VTE than those without. The risk was explained both by an increased risk of PE and isolated DVT, with the risk of isolated DVT (OR, 3.59; CI, 3.51-3.68; $P < 2.3e−308$) being higher than that of PE (OR, 2.72; CI, 2.64-2.81; $P < 2.3e−308$). The meta-analyzed CIs for isolated DVT and PE did not overlap (Figure 2B).

Full cohort (mega-analysis) results

As a secondary analysis, we also assessed whether these associations hold in a single association analysis including individuals across all 5 studied genetic ancestry groups. After adjusting for age, sex, and PCs, participants with SCT had an increased risk of prevalent VTE compared with those without SCT (OR, 1.44; CI, 1.31-1.58; $P = 7.9e−14$). Similar to the meta-analysis, this risk was predominantly explained by risk of PE (OR, 1.90; CI, 1.68-2.15; $P = 6.0e−25$) compared with isolated DVT (OR, 1.03; CI, 0.89-1.20; $P = .67$), with nonoverlapping CIs (supplemental Table 5A).
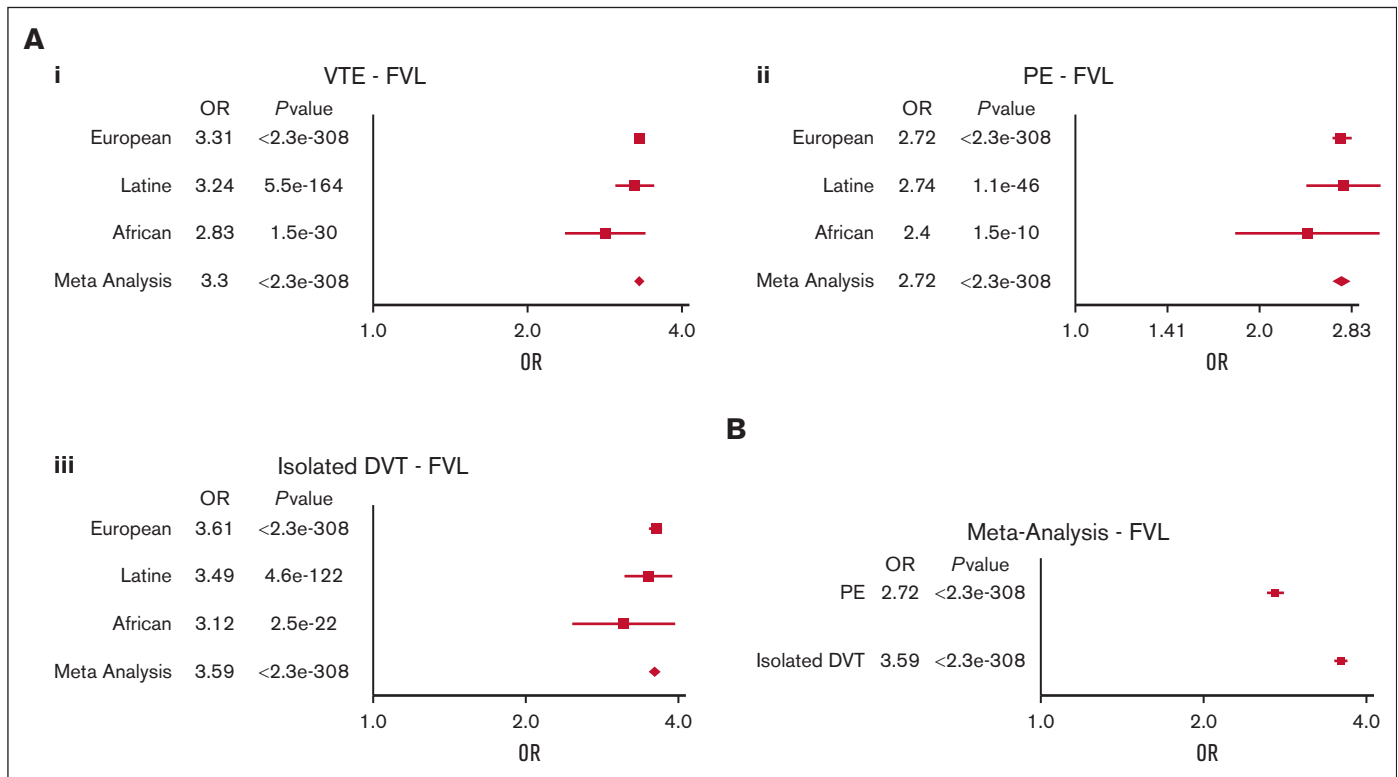
Figure 2. Risk estimates for VTE in factor V leiden. (A) Meta-analysis of ORs for VTE (i), PE (ii), and isolated DVT (iii) comparing FVL carriers with noncarriers. (B) Meta-analysis estimate for PE vs isolated DVT comparing FVL carriers with noncarriers.

Mega-analysis results for VTE, PE, and isolated DVT for FVL carriers were also consistent with the meta-analyzed results (supplemental Table 5B).

Discussion

The use of race in medical research has contributed to systemic bias and inequities. When coupled with systemic racism, the racialization of SCT has resulted in the mis-contextualization of risks and sustained harm to Black communities in both social and patient care contexts.[3,10,11] In this largest study to date, to our knowledge, using crowd-sourced data from the 23andMe research cohort, we found that the risk of VTE was increased among individuals with SCT compared with those without, independent of race or genetic ancestry. We also found that the risk of VTE with SCT was lower than that observed in FVL, a well-known inherited thrombophilia, which provides important data to allow for contextualization of VTE risks.[21]

Throughout medical literature, SCT is widely labeled as a "Black" trait, whereas FVL is labeled as a "White" thrombophilia.[22,23] These notions are partly based on studies that have evaluated statistical differences in prevalence rates among racial and genetic ancestry groups. Unfortunately, these race-based associations coupled with unconscious bias and systemic racism can lead to either overestimation or underestimation of attributable risk. For example, SCT has been erroneously and harmfully implicated as the cause of death in police custody cases, in part due to the overattribution of SCT to systemic disease as opposed to a modest increase in risk.[13,24] Conversely, providers often omit FVL testing in Black patients with VTE, despite its higher prevalence rate than other commonly tested thrombophilias.[25,26] In addition, incorrect assumptions about SCT's prevalence only in Black populations can also inadvertently harm non-Black individuals by limiting testing in those populations. Our findings demonstrate both that SCT is prevalent among diverse ancestry groups and that it confers a similar and consistent pattern of VTE risk independent of genetic ancestry. Our study, therefore, provides ancestry-independent and comparative risk estimates with FVL to allow for providers to appropriately contextualize VTE risk among SCT carriers and inform clinical practice guidelines without unintended bias. Because FVL is the most well-recognized thrombophilia in practice, direct comparative risk estimates of SCT and FVL using the same database are particularly important for contextualization of risk and for genetic counseling.

Although the direction of risk and pattern of VTE was similar for all genetic ancestry groups, we did observe a significantly higher point estimate for the risk of PE associated with SCT in the European cohort than the African cohort. We hypothesize that this observation may be due to a higher baseline risk of PE among all individuals of African ancestry. To investigate this, we explored the prevalence of PE among individuals without thrombophilia (no SCT or FVL), and we found that the baseline prevalence of PE, especially in younger age groups, was highest in the African cohort. A recent national database study similarly demonstrated a higher incidence of PE hospitalization among self-identified African Americans than other race groups.[27] This may suggest that unmeasured confounders, such as socioeconomic factors, may contribute to a higher baseline risk of PE in the African population. For example, we could hypothesize that differences in access to care may

contribute to delay in VTE diagnosis and higher progression to embolization to PE, even among individuals without thrombophilia. Because we were unable to adjust for these potential unmeasured confounders, a higher baseline risk of PE would lead to attenuation of the effect size of a moderate-risk genetic thrombophilia, such as SCT.

Our study also provides important insights about the pattern of VTE in SCT. Venous thrombi most commonly form in the lower extremities, with a 30% to 40% rate of embolization to PE.[28,29] We, therefore, expect that among individuals with VTE, 30% to 40% will experience PE, and 60% to 70% will have isolated DVT alone. Our study did verify this pattern both in our full cohort and among FVL carriers. However, among individuals with SCT and VTE, we observed the opposite pattern, with 61.4% of SCT carriers reporting a history of PE and 38.6% reporting a history of isolated DVT alone. Although previous studies have suggested that the VTE pattern among SCT carriers is PE predominant,[5,6] our study is, to our knowledge, the first to have sufficient sample size and power to demonstrate a statistically significant difference between the risk of PE and isolated DVT, with CIs that do not overlap. This is particularly important to demonstrate clear epidemiologic evidence for a unique mechanism of thrombosis in SCT compared with other thrombophilic states.

Whether the PE predominance of VTE observed in SCT is due to in situ pulmonary thrombosis or due to high risk of embolization from DVT has been a topic of debate. SCT is not associated with overt sickling, but erythrocyte rheologic changes may occur in local areas of extreme hypoxia such as the venous valves.[30] Mouse model studies have demonstrated that the composition of thrombi in sickle cell anemia and SCT mice differs from that of HbAA mice due to the entrapment of sickled erythrocytes and increased fibrin deposition within venous clots.[31] This differential composition has been hypothesized to lead to decreased stability and increased embolization tendency of SCT clots. In line with this, our study demonstrates that nearly one-third of SCT carriers with PE also reported a history of DVT, which is consistent with patterns observed in prior real-world general population studies in which DVT embolization is known to be the primary mechanism of PE.[32] Our study, therefore, suggests that embolization may also underlie the PE predominance among SCT carriers.

The major strengths of our study are its design and its largest-to-date sample size. In this respect, our study highlights the value of opt-in community engagement as a unique avenue for research that allows for a more large-scale and diverse data set than that available in other studies. Our study does have limitations. We did not have genotypic data on α-thalassemia, which may modify phenotype in SCT.[33] In addition, we used self-reported DVT and PE phenotypes. We believe our self-reported DVT and PE phenotypes to be accurate based on our previously published data of replication of genome-wide association signals using self-reported phenotypes,[14,34-36] and because our calculated VTE risk estimates are similar to those reported in previous studies of SCT.[4-6] In addition, a recent study in the UK Biobank found strong agreement of genetic effects for PE (±DVT) between cases identified using hospital records and those identified using verbal questionnaire data, further supporting the potential validity of survey self-report data.[37] However, our reliance on self-reported data includes certain limitations. For example, provoking factors were assessed

by self-report only and were not specific to a particular VTE event, and for individuals who reported both DVT and PE, we were unable to assess the relative timing of clots. Additionally, socioeconomic status may differ between our cohort and certain populations. There is a need for more large-scale studies of SCT to understand clinical risks.

Our study provides important ancestry-independent risk estimates for VTE in individuals with SCT and provides comparative risk estimates to FVL, which is a well-known inherited thrombophilia. Our study also provides high-powered data to confirm the PE-predominant pattern of VTE in SCT carriers. These data may inform clinical practice guidelines, future research, and public health initiatives in SCT.

Authorship

Contribution: K.-H.L. and J.M.G. designed the research, analyzed the data, and wrote the manuscript; A.J.S. provided critical edits to the manuscript; and V.L.B. and R.P.N. designed the research, interpreted the results, and wrote the manuscript.

Conflict-of-interest disclosure: K.-H.L., J.M.G., and A.J.S. are employed by and hold stock or stock options in 23andMe Inc. The remaining authors declare no competing financial interests.

A complete list of the members of the 23andMe Research Team appears in "Appendix."

ORCID profiles: V.L.B., 0000-0002-3649-5442; R.P.N., 0000-0001-5562-1283.

Correspondence: Rakhi P. Naik, Division of Hematology, Department of Medicine, Johns Hopkins University, 1830 E Monument St, Suite 7300, Baltimore, MD 21205; email: rakhi@jhmi.edu.

Appendix

The members of the 23andMe research team who contributed to this study are Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg, Jonathan Bowes, Katarzyna Bryc, Ninad S. Chaudhary, Daniella Coker, Sayantan Das, Emily DelloRusso, Sarah L. Elson, Nicholas Eriksson, Teresa Filshtein, Pierre Fontanillas, Will Freyman, Zach Fuller, Chris German, Julie M. Granka, Karl Heilbron, Alejandro Hernandez, Barry Hicks, David A. Hinds, Ethan M. Jewett, Yunxuan Jiang, Katelyn Kukar, Alan Kwong, Yanyu Liang, Keng-Han Lin, Bianca A. Llamas, Matthew H. McIntyre, Steven J. Micheletti, Meghan E. Moreno, Priyanka Nandakumar, Dominique T. Nguyen, Jared O'Connell, Aaron A. Petrakovitz, G. David Poznik, Alexandra Reynoso, Shubham Saini, Morgan Schumacher, Leah Selcer, Anjali J. Shastri, Janie F. Shelton, Jingchunzi Shi, Suyash Shringarpure, Qiaojuan Jane Su, Susana A. Tat, Vinh Tran, Joyce Y. Tung, Xin Wang, Wei Wang, Catherine H. Weldon, Peter Wilton, and Corinna D. Wong.

References

1. Grant AM, Parker CS, Jordan LB, et al. Public health implications of sickle cell trait: a report of the CDC meeting. *Am J Prev Med*. 2011;41(6 suppl 4): 435-S439.

2. Lindenau JD, Wagner SC, Castro SMd, Hutz MH. The effects of old and recent migration waves in the distribution of HBB*S globin gene haplotypes. *Genet Mol Biol*. 2016;39(4):515-523.

3. Power-Hays A, McGann PT. When actions speak louder than words - racism and sickle cell disease. *N Engl J Med*. 2020;383(20):1902-1903.

4. Naik RP, Smith-Whitley K, Hassell KL, et al. Clinical outcomes associated with sickle cell trait: a systematic review. *Ann Intern Med*. 2018;169(9): 619-627.

5. Austin H, Key NS, Benson JM, et al. Sickle cell trait and the risk of venous thromboembolism among blacks. *Blood*. 2007;110(3):908-912.

6. Folsom AR, Tang W, Roetker NS, et al. Prospective study of sickle cell trait and venous thromboembolism incidence. *J Thromb Haemost*. 2015;13(1): 2-9.

7. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight - reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. 2020;383(9): 874-882.

8. *Using Population Descriptors in Genetics and Genomics Research*. National Academies of Sciences, Engineering, and Medicine; 2023.

9. Bonham VL, Callier SL, Royal CD. Will precision medicine move us beyond race? *N Engl J Med*. 2016;374(21):2003-2005.

10. Naik RP, Haywood C Jr. Sickle cell trait diagnosis: clinical and social implications. *Hematology Am Soc Hematol Educ Program*. 2015;2015(1): 160-167.

11. Bonham VL, Dover GJ, Brody LC. Screening student athletes for sickle cell trait–a social and clinical experiment. *N Engl J Med*. 2010;363(11):997-999.

12. US Court of Appeals Ninth Circuit, Ninth Circuit. Norman-Bloodsaw v. Lawrence Berkeley Laboratory. *Fed Report*. 1998;135:1260-1276.

13. LaForgia M, Valentino-DeVries J. How a genetic trait in Black people can give the police cover. *The New York Times*. 2021. Accessed 15 May 2021. https://www.nytimes.com/2021/05/15/us/african-americans-sickle-cell-police.html

14. Shelton JF, Shastri AJ, Ye C, et al. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat Genet*. 2021;53(6):801-808.

15. Durand EY, Do CB, Mountain JL, Macpherson JM. Ancestry composition: a novel, efficient pipeline for ancestry deconvolution. *bioRxiv*. Preprint posted online 18 October 2014. https://doi.org/10.1101/010512

16. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet*. 2015;96(1):37-53.

17. Shi Z, Zhan J, Wei J, et al. Reliability of ancestry-specific prostate cancer genetic risk score in four racial and ethnic populations. *Eur Urol Open Sci*. 2022;45:23-30.

18. Shelton JF, Shastri AJ, Fletez-Brant K, , 23andMe COVID-19 Team, Aslibekyan S, Auton A. The UGT2A1/UGT2A2 locus is associated with COVID-19-related loss of smell or taste. *Nat Genet*. 2022;54(2):121-124.

19. Henn BM, Hon L, Macpherson JM, et al. Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One*. 2012;7(4):e34267.

20. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191.

21. Koster T, Rosendaal FR, de Ronde H, Briet E, Vandenbroucke JP, Bertina RM. Venous thrombosis due to poor anticoagulant response to activated protein C: Leiden Thrombophilia Study. *Lancet*. 1993;342(8886-8887):1503-1506.

22. Heit JA, Sobell JL, Li H, Sommer SS. The incidence of venous thromboembolism among factor V Leiden carriers: a community-based cohort study. *J Thromb Haemost*. 2005;3(2):305-311.

23. Dowling NF, Austin H, Dilley A, Whitsett C, Evatt BL, Hooper WC. The epidemiology of venous thromboembolism in Caucasians and African-Americans: the GATE Study. *J Thromb Haemost*. 2003;1:80-87.

24. Lichtsinn HS, Weyand AC, McKinney ZJ, Wilson AM. Sickle cell trait: an unsound cause of death. *Lancet*. 2021;398(10306):1128-1129.

25. Jerrard-Dunne P, Evans A, McGovern R, et al. Ethnic differences in markers of thrombophilia: implications for the investigation of ischemic stroke in multiethnic populations: the South London Ethnicity and Stroke Study. *Stroke*. 2003;34(8):1821-1826.

26. Lijfering WM, Brouwer JL, Veeger NJ, et al. Selective testing for thrombophilia in patients with first venous thrombosis: results from a retrospective family cohort study on absolute thrombotic risk for currently known thrombophilic defects in 2479 relatives. *Blood*. 2009;113(21):5314-5322.

27. Afifi AM, Leverich M, Tadrousse K, Ren G, Nazzal M. Racial, biological sex, and geographic disparities of venous thromboembolism in the United States, 2016-2019. *J Vasc Surg Venous Lymphat Disord*. 2024;12(5):101908.

28. Cushman M, Tsai AW, White RH, et al. Deep vein thrombosis and pulmonary embolism in two cohorts: the longitudinal investigation of thromboembolism etiology. *Am J Med*. 2004;117(1):19-25.

29. Heit JA, Spencer FA, White RH. The epidemiology of venous thromboembolism. *J Thromb Thrombolysis*. 2016;41(1):3-14.

30. Hamer JD, Malone PC, Silver IA. The PO_2 in venous valve pockets: its possible bearing on thrombogenesis. *Br J Surg*. 1981;68(3):166-170.

31. Faes C, Ilich A, Sotiaux A, et al. Red blood cells modulate structure and dynamics of venous clot formation in sickle cell disease. *Blood*. 2019;133(23):2529-2541.

32. Spencer FA, Emery C, Lessard D, et al. The Worcester Venous Thromboembolism study: a population-based study of the clinical epidemiology of venous thromboembolism. *J Gen Intern Med*. 2006;21(7):722-727.

33. Steinberg MH, Embury SH. Alpha-thalassemia in Blacks: genetic and clinical aspects and interactions with the sickle hemoglobin gene. *Blood*. 1986;68(5):985-990.

34. Tung JY, Do CB, Hinds DA, et al. Efficient replication of over 180 genetic associations with self-reported medical data. *PLoS One*. 2011;6(8):e23473.

35. Simcoe M, Valdes A, Liu F, et al. Genome-wide association study in almost 195,000 individuals identifies 50 previously unidentified genetic loci for eye color. *Sci Adv*. 2021;7(11):eabd1239.

36. Levey DF, Stein MB, Wendt FR, et al. Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat Neurosci*. 2021;24(7):954-963.

37. DeBoever C, Tanigawa Y, Aguirre M, McInnes G, Lavertu A, Rivas MA. Assessing digital phenotyping to enhance genetic studies of human diseases. *Am J Hum Genet*. 2020;106(5):611-622.