MDPI

Article

Impact of Genetic Variations on Thromboembolic Risk in Saudis with Sickle Cell Disease

Mohammad A. Alshabeeb [1,2,*] , Deemah Alwadaani [2,3], Farjah H. Al Qahtani [4], Salah Abohelaika [5,6] ,
Mohsen Alzahrani [2,7], Abdullah Al Zayed [8], Hussain H. Al Saeed [8], Hala Al Ajmi [1,2], Barrak Alsomaie [1,2],
Mamoon Rashid [2,9] and Ann K. Daly [10]

[1]   King Abdullah International Medical Research Center (KAIMRC), Riyadh 11426, Saudi Arabia
[2]   King Saud Bin Abdulaziz University for Health Sciences (KSAU-HS), Ministry of National Guard Health
      Affairs (MNGHA), Riyadh 11426, Saudi Arabia; alzahranimo5@mngha.med.sa (M.A.)
[3]   Medical Genomics Research Department, King Abdullah International Medical Research Center (KAIMRC),
      Riyadh 11481, Saudi Arabia
[4]   Hematology/Oncology Center, King Saud University Medical City (KSUMC), Riyadh 11411, Saudi Arabia;
      falgahtani@ksu.edu.sa
[5]   Research Department, Qatif Central Hospital (QCH), Qatif 32654, Saudi Arabia; sabohelaika@moh.gov.sa
[6]   Pharmacy Department, Qatif Central Hospital (QCH), Qatif 32654, Saudi Arabia
[7]   King Fahad Hospital, Ministry of National Guard Health Affairs (MNGHA), Riyadh 11426, Saudi Arabia
[8]   Hematology Department, Qatif Central Hospital (QCH), Qatif 32654, Saudi Arabia;
      abalzayed@moh.gov.sa (A.A.Z.); alsaeedhha@hotmail.com (H.H.A.S.)
[9]   Department of AI and Bioinformatics, King Abdullah International Medical Research Center (KAIMRC),
      Riyadh 11481, Saudi Arabia
[10] Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University,
      Newcastle Upon Tyne NE1 7RU, UK
*   Correspondence: shabeebmo@mngha.med.sa

Abstract: Background: Sickle cell disease (SCD) is a Mendelian disease characterized by multigenic
phenotypes. Previous reports indicated a higher rate of thromboembolic events (TEEs) in SCD
patients. A number of candidate polymorphisms in certain genes (e.g., FVL, PRT, and MTHFR) were
previously reported as risk factors for TEEs in different clinical conditions. This study aimed to
genotype these genes and other loci predicted to underlie TEEs in SCD patients. Methodology: A
multi-center genome-wide association study (GWAS) involving Saudi SCD adult patients with a
history of TEEs (n = 65) and control patients without TEE history (n = 285) was performed. Genotyping
used the $10\times$ Affymetrix Axiom array, which includes 683,030 markers. Fisher's exact test was used
to generate p-values of TEE associations with each single-nucleotide polymorphism (SNP). The
haplotype analysis software tool version 1.05, designed by the University of Göttingen, Germany, was
used to identify the common inherited haplotypes. Results: No association was identified between
the targeted single-nucleotide polymorphism rs1801133 in MTHFR and TEEs in SCD ($p = 0.79$). The
allele frequency of rs6025 in FVL and rs1799963 in PRT in our cohort was extremely low (<0.01);
thus, both variants were excluded from the analysis as no meaningful comparison was possible. In
contrast, the GWAS analysis showed novel genome-wide associations ($p < 5 \times 10^{-8}$) with seven
signals; five of them were located on Chr 11 (rs35390334, rs331532, rs317777, rs147062602, and
rs372091), one SNP on Chr 20 (rs139341092), and another on Chr 9 (rs76076035). The other 34 SNPs
located on known genes were also detected at a signal threshold of $p < 5 \times 10^{-6}$. Seven of the
identified variants are located in olfactory receptor family 51 genes (OR51B5, OR51V1, OR51A1P,
and OR51E2), and five variants were related to family 52 genes (OR52A5, OR52K1, OR52K2, and
OR52T1P). The previously reported association between rs5006884-A in OR51B5 and fetal hemoglobin
(HbF) levels was confirmed in our study, which showed significantly lower levels of HbF ($p = 0.002$)
and less allele frequency ($p = 0.003$) in the TEE cases than in the controls. The assessment of the
haplotype inheritance pattern involved the top ten significant markers with no LD (rs353988334,
rs317777, rs14788626882, rs49188823, rs139349992, rs76076035, rs73395847, rs1368823, rs8888834548,
and rs1455957). A haplotype analysis revealed significant associations between two haplotypes

(a risk, TT-AA-del-AA-ins-CT-TT-CC-CC-AA, and a reverse protective, CC-GG-ins-GG-del-TT-CC-TT-GG-GG) and TEEs in SCD (p = 0.024, OR = 6.16, CI = 1.34–28.24, and p = 0.019, OR = 0.33, CI = 0.13–0.85, respectively). Conclusions: Seven markers showed novel genome-wide associations; two of them were exonic variants (rs317777 in OLFM5P and rs147062602 in OR51B5), and less significant associations ($p < 5 \times 10^{-6}$) were identified for 34 other variants in known genes with TEEs in SCD. Moreover, two 10-SNP common haplotypes were determined with contradictory effects. Further replication of these findings is needed.

---

1. Introduction

Sickle cell disease (SCD) affects millions of people around the world but focuses more commonly on certain ethnicities in African, Caribbean, and Middle Eastern populations. A large proportion of Saudis, in particular people who live in the Eastern province (up to 24%), are affected by the disease or carry a single copy of the genetic trait [1,2]. It is an autosomal recessive disease caused by a point gene mutation: a transversion change in a single nucleotide base (rs334 A>T) in the codon region of the sixth amino acid of the hemoglobin β gene (HBB), located at chromosome (Chr) 11. This substitution results in an amino acid change from glutamine to valine, which ultimately encodes sickled hemoglobin, which polymerizes when deoxygenated and tends to produce abnormal (crescent-shaped) red blood cells (RBCs) [3]. The sickled reticulocytes (immature RBCs) in SCD have increased adhesive properties, which may trigger the vaso-occlusive process [4]. A clog of blood capillaries may lead to a disabling systemic syndrome, including chronic anemia, which may require frequent blood transfusions. Complications of SCD may also comprise difficult-to-treat leg ulcers, eye damage, stroke, vaso-occlusive (thrombotic) crises accompanied by acute pain, and organ infarction [5]. SCD patients may develop chronic organ malfunctions too, such as splenic sequestration crises, the formation of gallstones, and lung crises (the development of acute chest syndrome), that ultimately result in a poor quality of life, a poor prognosis, and a shorter life expectancy [6].

Several previous reports confirmed the increased risk of thromboembolic events (TEEs), both arterial and venous thromboembolism (VTE), in both SCD trait (HbAS) and SCD patients (HbSS) in comparison to control groups [7–10]. The reported level of risk in African-Americans with sickle cell trait is approximately twofold for VTE in general and fourfold for pulmonary embolism (PE) [11]. These results were confirmed recently in the UK, where sickle cell carriers were found to be at a higher risk of VTE, in particular PE (OR = 2.27), than healthy individuals [12]. In addition, in a study that involved 7000 SCD patients in the United States, the risk of developing PE was found to be fourfold the risk in patients without SCD [13]. Similarly, higher rates of infarctive cerebrovascular accidents and hemorrhagic strokes were noted in individuals with SCD than in normal subjects [14,15]. The prevalence of TEEs in SCD patients reported in African-Americans is about 11.3% to 11.5% [16,17]; this percentage is similar to the findings seen in the Saudi population (11.3%) [18], although a larger study showed a slightly lower prevalence (8.4%) [19]. Rates are extremely low in healthy individuals, approximately 1 per 1000 (0.1%) [20,21]. This might be attributed to several common risk factors, such as the placement of central venous catheters, obesity, pregnancy, and thrombophilias, or related to specific SCD factors, such as a history of splenectomy and the genetic makeup [22].

A number of reported candidate genes demonstrate a potential role in inducing thromboembolism in several clinical conditions. Factor V Leiden (FVL), prothrombin (PRT), and methylenetetrahydrofolate reductase (MTHFR) are the most common genes associated with a hypercoagulability status and are involved in various thrombophilic conditions. The risk of TEEs in women positive for the FVL (G1691A, rs6025) mutation increases by

five times when they are exposed to oral contraceptives [23]. Thrombotic episodes and graft rejection were also noted in patients who underwent kidney transplantation and were heterozygous for the FVL variant [24]. Carriers of this variant are more susceptible to deep venous thrombosis (DVT) [25]. PRT (FII) (G20210A, rs1799963) was also identified as a risk factor for pulmonary embolism [26] and myocardial infarction [27]. Both variants in FVL and PRT were reported as the main contributing factors underlying recurrent pregnancy loss in Saudi females from different regions [28,29]. These two variants were also found to be more prominent in 250 TEE patients from Kashmir [30]. Furthermore, MTHFR variants (C677T (rs1801133) and A1298C (rs1801131)) were reported as risk markers in addition to FVL for developing DVT in two Iranian studies [31,32], although a previous study on a smaller number of patients failed to detect an association between DVT and MTHFR [33]. Different types of thrombotic conditions, such as portal vein thrombosis [34,35], postcardiac surgery thrombosis [36], and arterial thrombosis [37] were significantly associated with MTHFR, in particular the C677T variant. The polymorphisms in MTHFR are common in the healthy Saudi population; 23.9% are positive for C677T and 33.9% are positive for A1298C. In contrast, a lower number of individuals carry the variants G1691A in FVL (an average of 2%) and G20210A in PRT (2%) [38].

Previous comparative genetic studies between SCD patients and the normal population revealed variable allele frequencies of variants in the selected thrombotic genes. Two Indian studies that involved collectively 391 patients versus 447 controls [39,40] indicated a higher frequency of risk alleles in both the FVL and MTHFR (C677T) genes in patients compared to controls. However, the result failed to be replicated in 180 patients from the western region of India [41]. Two other studies conducted on Brazilian SCD patients showed a higher variant allele frequency in MTHFR (C677T) but not FVL in patients than in controls [42,43]. A recent study conducted on Tunisians showed more frequent polymorphisms in PRT (G20210A) and MTHFR (C677T) in patients with SCD compared with healthy subjects [44]. On the other hand, investigating the three polymorphisms in these genes (FVL, PRT, and MTHFR) in the Saudi population from the eastern province showed nonsignificant differences between 87 patients and 105 healthy individuals [45]. Due to the known role of these genes in the induction of thrombosis, we considered that we might see a higher rate of the selected mutations in SCD patients with a history of TEEs than in patients without a history of thrombosis. Thus, this study assessed the association of specific variants in FVL, PRT, and MTHFR with TEEs in Saudis with SCD and also performed a genome-wide association study (GWAS) to identify novel risk variants.

2. Methods

2.1. Study Design

This is a genetic association, case–control, multicenter study conducted to screen genes predicted to be associated with TEEs in SCD patients. The study was approved by the Institutional Review Boards of KAIMRC (Ref: IRBC/1414/19) and Qatif Central Hospital (QCH-SREC0216/2020). The recruited subjects were genotyped for selected variants in known thrombotic genes such as FVL (G1691A, rs6025), PRT (FII) (G20210A, rs1799963), and MTHFR (C677T, rs1801133). Moreover, a GWAS analysis comparing SCD patients with and without a TEE history was performed.

2.2. Sample Recruitment

Unrelated adult patients with SCD confirmed by a positive sickling test and homozygosity for the sickle mutation (rs334) (n = 350; 65 SCD patients with a history of TEEs plus 285 controls (SCD patients without a TEE history)) were recruited through their regular follow up with hematology units in three different settings: (i) King Fahad Hospital (KFH), Ministry of National Guard Health Affairs (MNGHA), Riyadh (n = 106 SCD patients; 27 cases with a TEE history plus 79 controls), (ii) King Khalid University Hospital (KKUH), King Saud University Medical City, Riyadh, (n = 82 SCD patients; 25 cases with a TEE history plus 57 controls), and (iii) Qatif central hospital (QCH), Eastern province, Qatif

(162 SCD patients; 13 SCD with a TEE history plus 149 controls). Most of the patients who attended KKUH and KFH hospitals in the central region (Riyadh) were referred from southwestern or northern areas of Saudi Arabia. The average age of participants at the time of recruitment was 32.7 ± 10.2 years. Demographic and phenotypic data were collected for all participants. The focus in this study was on HbSS patients; thus, others with different types of SCD, such as HbSC, HbSbeta-thalassemia, HbSD, and HbSO, were identified via hemoglobin electrophoresis and were excluded. Exclusion criteria for TEE cases involved obese patients (BMI ≥ 30), smokers, females on contraceptive pills, pregnant women, history of postpartum, recurrent miscarriages, diabetes mellitus, Behçet's disease, varicose veins, thrombophilia, chronic renal disease, ST elevation myocardial infarction (STEMI), cancer, immobilization, recent history of surgery, central venous catheter placement, admission at the intensive care units, trauma, fracture, recent long-distance travel, elevated levels of homocysteine, protein C, protein S, D-dimer, thrombin-antithrombin, and those with abnormal prothrombin time (PT, APTT). Only two cases fit these criteria and were excluded from our study cohort: a female with a postoperative DVT and another with a PE during pregnancy. Informed consent was obtained from the participants, and blood samples were collected from each patient for a DNA analysis.

2.3. TEE Diagnosis

Patients with TEEs were diagnosed through physician clinical judgment in conjunction with confirmatory imaging such as ultrasonography or computed tomography (CT). Stroke and transient ischemic attack (TIA) were confirmed via physical and neurological examination, e.g., electroencephalogram (EEG), laboratory (blood) tests, and imaging tests such as a doppler ultrasound, CT, or magnetic resonance imaging (MRI) scan.

2.4. Genomic Analysis

Total genomic DNA was extracted from whole blood using Puregene Blood Kits (Qiagen, Hilden, Germany, Catalog Number # 158389) according to the supplier's instructions. An automated DNA extractor machine (KingFisher™ magnetic system, Thermo Fisher Scientific, Fresno, CA, USA) was used. A nanodrop (2000/2000c), Thermo Fisher Scientific, Fresno, CA, USA, was used to measure the absorbance of DNA at 260 nm. Working DNA stocks were aliquoted at a concentration of 50 ng/ul and stored at 4 °C. Cases and control samples were genotyped using the 10× Affymetrix Axiom array (Axiom 2.0 reagent kit designed by Applied Biosystems™, Waltham, MA, USA, catalog number 901758), which includes 683,030 markers for the GWAS. The GWAS association analysis involved both common (minor allele frequency (MAF) ≥ 5%) and rare (MAF < 0.5%) variants. To ensure the validity and reliability of the used platform, a control sample with known variant calls was added to each test run. Axiom Analysis Suite software version 5.1 was used to cluster the genomic data; the average quality control call rate (CR) for the passing samples was 99.896%, and samples with >93% CR were retested. Genotype imputation was not performed here, as no reference panel is currently available for the Saudi population. The main five HBB haplotypes (Benin, Arab/Indian, Cameroon, CAR/Bantu, and SEN) can be ascertained through the genotyping of four SNPs (rs3834466, rs28440105, rs10128556, and rs968857) [46]. However, these SNPs are not included among the used GWAS panel. Thus, the association between HBB haplotypes and TEEs was not assessed.

2.5. Sample Size Calculation

Minitab version 16 was used to calculate the sample size needed as representative samples of the Saudi population to provide a statistical power of 80% at a 0.05 *p*-value cutoff significance. Our calculations based on MAF showed that a genotyping of 337 cases was needed for the FVL variant and 222 cases for the MTHFR variant to provide the suggested study power (Table 1). Therefore, we decided to recruit 350 cases in total.

Table 1. Sample size calculation based on MAF of FVL (rs6025) and MTHFR (rs1801133) at cut-off significance *p*-value of 0.05.

MAF Reported by Fawaz et al. [45]	FVL (rs6025)	MTHFR (rs1801133)
Controls	0.01	0.14
Cases	0.04	0.25
Sample size with 80% power	337	222

2.6. Statistical Analysis

Plink software version 1.9 was used for the analysis of genomic data and generating linkage disequilibrium (LD) between SNPs. Samples with genotypes that were not in Hardy–Weinberg equilibrium ($p < 0.05$) were removed from the study. The *p*-values were calculated for categorical covariates such as genotype differences between different groups using Fisher's exact test (Graphpad PRISM version 5.0). For GWAS, a *p*-value $< 5 \times 10^{-8}$ was set as a strict significance point to identify risk loci [47]. However, markers with $p > 5 \times 10^{-6}$ were also considered. Calculations of means and standard deviations were performed to assess age matching, and the differences were assessed using the 2-independent-sample *t*-test. Manhattan, quantile–quantile (Q-Q), and principal component analysis (PCA) plots were generated using the R statistical package (qqman) version R-4.2.2. The permutation test (T1) was used to assess sample heterogeneity based on pairwise identity-by-state (IBS) distance. The haplotype analysis software tool version 1.05, prepared by Eliades N-G. and Eliades D. G. [48], was used to determine the common significant haplotypes.

3. Results

As shown in Table 2, matching was confirmed for age (35.7 ± 9.8 years vs. 34.4 ± 10.3 years, *p* = 0.35) and sex (49.2% vs. 56.8% females, *p* = 0.27) between cases and controls, respectively. The majority of TEE cases had stroke (41.5%), PE (38.5%), and DVT (29.2%). The recruited patients with TEEs were mainly from the southwestern province (78.5%), with few cases from the eastern province (21.5%). A Q-Q plot of the observed versus expected *p*-values showed a late departure of the observed *p*-values from the null (Supplementary Figure S1), indicating that the obtained results were not affected by genotyping quality, sample relatedness, or population stratification. Furthermore, the PCA plot indicated homogenous clusters of genetic variations that characterize cases and control cohorts (Figure 1). The genotype data of participants shown on the scatter plot revealed a matched background distribution. This was confirmed by the permutation test (T1) (*p* = 0.93), which ruled out the stratification effects between cases and controls. No significant difference was seen in the allele frequency of the MTHFR (rs1801133) polymorphism between the cases and control group (*p* = 0.74; Table 3). Moreover, the other two candidate genes, FVL (rs6025) and PRT (rs1799963), were not included in the analysis as the allele frequency of their variants was extremely low (<0.01). No meaningful comparison could be achieved in such a condition.

Table 2. Demographic data of SCD patients who developed TEEs versus patients with no TEE history.

Variables	TEE Cases, n = 65 (%)	No TEE Controls; n = 285 (%)	*p*-Value
Males/Females	32/33	162/123	0.27
Age in years, Mean ± SD	35.7 ± 9.8	34.4 ± 10.3	0.35
Fetal hemoglobin (HbF), Mean ± SD	10.7 ± 8.0	14.1 ± 7.6	0.002
Southwestern province	51 (78.5)	133 (46.7)	
Eastern province	14 (21.5)	152 (53.3)	
DVT only	11 (16.9)		

Table 2. *Cont.*

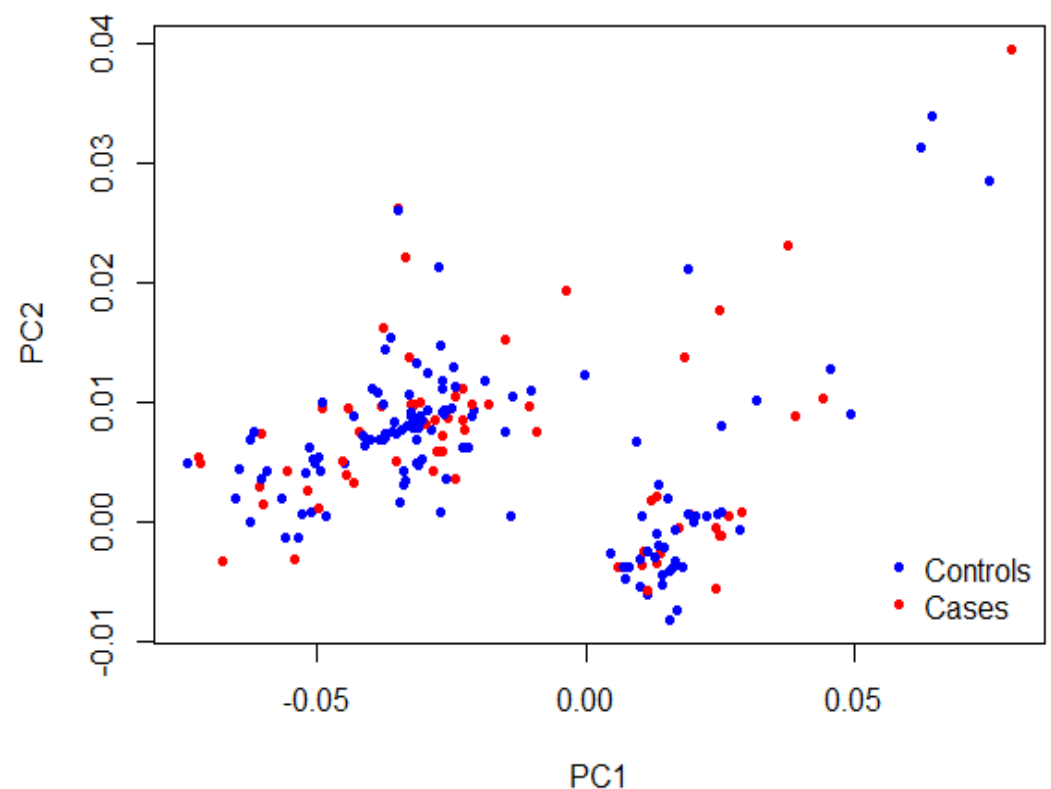| Variables	TEE Cases, n = 65 (%)	No TEE Controls; n = 285 (%)	*p*-Value
PE only	18 (27.7)		
Stroke only	22 (33.8)		
DVT and PE	5 (7.7)		
DVT and stroke	3 (4.6)		
PE and stroke	2 (3.1)		
Other TEEs affected different organs (eye, ear, and liver)	4 (6.2)		

Figure 1. PCA plot showing sample stratification. The first two principal components (PC1 and PC2) of the cases in red and the controls in blue were plotted. The permutation test (T1) confirmed the samples' heterogeneity (*p* = 0.93).

Table 3. Association results of selected SNPs (reported previously as risk factors for thrombosis) with TEE cases in our SCD cohort.

Gene, Variant (Alleles)	SCD Cases with TEEs (n = 65)	SCD Controls without TEEs (n = 285)	*p*-Value
	Allele Frequencies (MAF%)		
Known thrombotic gene			
MTHFR, rs1801133 (A/G)	21/109 (0.16)	85/481 (0.15)	0.74
Previously reported with TEEs			
ITPR3, rs2229637 (A/G)	24/106 (0.18)	166/404 (0.29)	0.014
LINC02651-RPL5P26, rs10998957 (C/T)	18/112 (0.14)	131/435 (0.23)	0.020
H6PD-SPSB1, rs10746487 (T/C)	40/88 (0.31)	123/441 (0.22)	0.023
Intergenic, rs1985317 (T/C)	46/84 (0.35)	252/310 (0.45)	0.050
LINC00877, rs6771316 (A/G)	1/129 (0.01)	28/542 (0.05)	0.032

Five SNPs (rs2229637 in ITPR3, rs10998957 in LINC02651-RPL5P26, rs10746487 in H6PD-SPSB1, rs1985317, and rs6771316 in LINC00877) were previously tested in GWASs among French [49], African-Americans [50], and other mixed populations [51] that showed associations at $p < 5 \times 10^{-6}$ between the tested markers and venous thromboembolism (VTE). These results were replicated among our Saudi cohort but at lower significant levels ($p = 0.016, 0.024, 0.028, 0.0496,$ and $0.044,$ respectively). Other variants in various genes were previously reported among the German population [52] (SLCO1B1 (rs4149056), PRIM1 (rs2277339), APOB (rs676210), TYK2 (rs12720356), TSEN15 (rs1046934), CYP4F2 (rs2108622), and MST1 (rs3197999)), and the Brazilians (ADAMTS (rs1364044)) [53] as risk factors for stroke in SCD patients with p-values $< 1.0 \times 10^{-5}$ in GWASs. These SNPs were tested in our stroke patients (n = 27), and no significant associations were detected except for rs1364044 in ADAMTS12 ($p = 0.75, 0.74, 0.55, 1.0, 0.53, 0.55, 0.38,$ and $0.036,$ respectively).

Additional stroke markers (rs2084898 in TRIM29, rs469568 in ADAMTS2 [51], rs17696736 in SH2B3, rs4792143 in FLJ45455, rs16851055 in SPSB4, rs12646447 in PITX2, rs225132 in ERRFI1, rs161802 in PARK7 [54], rs12425791 in NINJ2 [55], rs2200733 in NR [56], rs556621 in SUPT3H [57], rs173686 in CSPG2 [58], and rs11853426 in ANXA2 [59,60]) were further tested here, but no associations with stroke cases in SCD were found except for the TRIM29 marker ($p = 0.02, 0.65, 0.70, 0.81, 0.12, 0.07, 0.91, 0.88, 0.76, 0.08, 0.17, 0.32,$ and $0.06,$ respectively). The GWAS analysis of our data identified seven markers that surpassed genome-wide significance; five of them were located on Chr 11, (rs35390334 and rs331532 are in strong LD, rs317777, rs147062602, and rs372091), one SNP located on Chr 20 (rs139341092), and another on Chr 9 (rs76076035) (Figure 2). The SNPs rs317777 (an upstream variant in OLFM5P) and rs147062602 (a frame shift variant located in OR51B5) are exonic polymorphisms. In addition, we found 34 additional SNPs located on different genes with an association threshold of $p < 5 \times 10^{-6}$ (Table 4). Most of them were located at Chr 11, and some resided in potential sites such as rs73395847 (a splice acceptor variant in the C11orf40 gene) and rs80034548 (a noncoding transcript exon variant in the OR51A1P pseudogene). In addition, rs6554634 is a transcription factor binding site variant in the SLC6A19 gene at Chr 5, rs7933549 in OR51V1 is a missense variant, and rs73402629 in HBG1 and rs2071348 in HBBP are transcript enhancers located at the human β-globin locus. Two of the detected intronic variants (rs1455957 and rs11035718), which were in a perfect LD ($r^2 = 0.9$, Supplementary Excel Sheet S1), were both found to be in a strong LD ($r^2 = 0.77$ and 0.85, respectively) with a missense variant (rs2472530) in the OR52A5 gene. These SNPs were all located at Chr 11. An SNP (rs5006884-A) in OR51B5 was less frequent in the cases than in the controls (MAF = 0.15 vs. 0.27, $p = 0.003$, OR = 0.46) and interestingly, the cases showed significantly lower levels of HbF compared to the controls (mean $\pm$ SD = 10.7 $\pm$ 8.0 vs. 14.1 $\pm$ 7.6, $p = 0.002$), Table 2.

The subgrouping analysis, which included stroke patients only, showed significant associations beyond the GWAS threshold of $p < 5 \times 10^{-8}$ with 84 variants; 26 of them were among the identified 41 loci detected for all TEE patients, though the associations seen in stroke cases were drastically stronger; see Table 4 and the Supplementary Excel Sheet S2. A haplotype analysis was conducted, focusing on the top 10 SNPs with independent significance (no or weak LD) at a GWAS threshold of $< 1.0 \times 10^{-6}$. The haplotype inheritance pattern of the selected markers (rs35390334, rs317777, rs147062602, rs4910823, rs139341092, rs76076035, rs73395847, rs1368823, rs80034548, and rs1455957) emphasized 131 haplotypes. The common haplotypes (observed in ≥ 5 individuals) were 14 (Table 5), and 2 of them indicated significant associations: a risk haplotype (haplo-91: TT-AA-del-AA-ins-CT-TT-CC-CC-AA) and its protective inverse (haplo-22: CC-GG-ins-GG-del-TT-CC-TT-GG-GG) with TEEs in SCD ($p = 0.024$, OR = 6.16, CI = 1.34–28.24, and $p = 0.019$, OR = 0.33, CI = 0.13–0.85, respectively).

Table 4. Markers with the highest association significance suggested to impact TEEs in SCD.

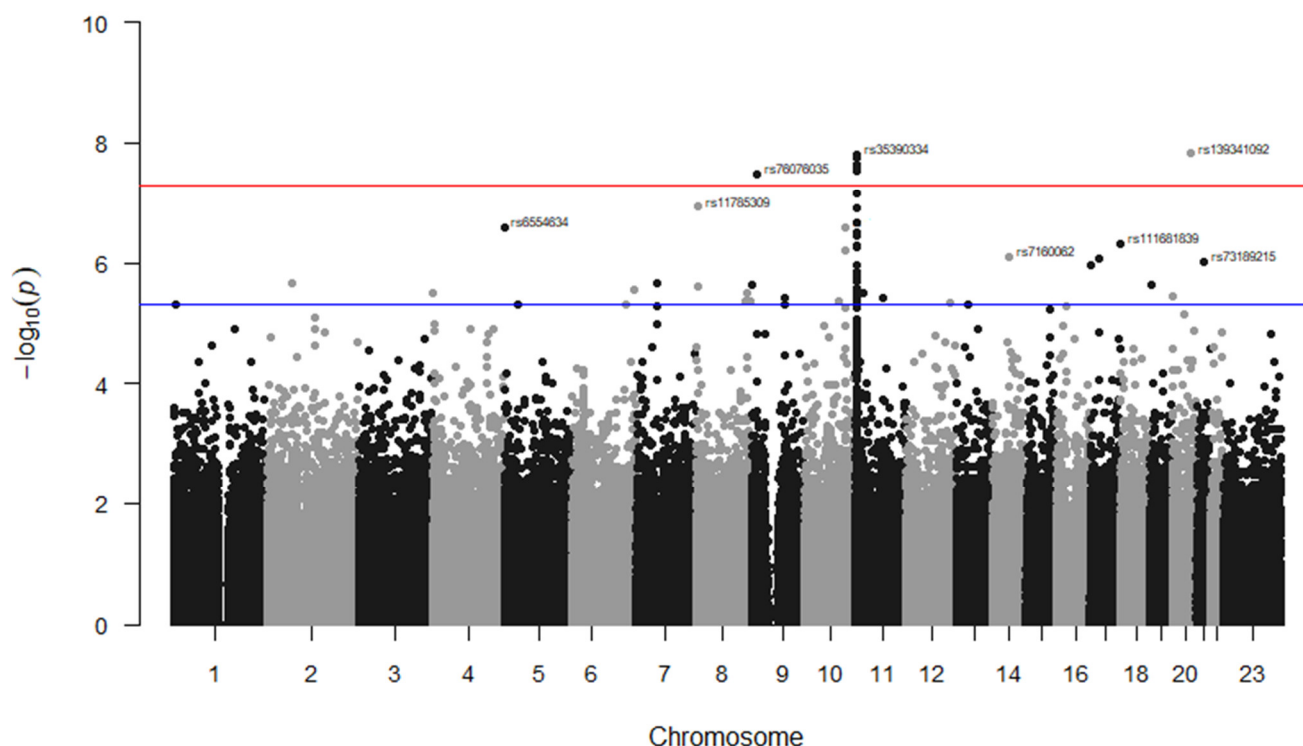| Variant	Chr	Variant Type (Gene)	Alleles	*p*-Value	
				All TEEs	*Stroke only*
Signals at a threshold of p < 5 × 10⁻⁸ (n = 7)					
rs139341092	20	Intergenic	CTCA/-	1.50×10^{-8}	
rs35390334	11	Intronic	C > T	2.40×10^{-8}	3.21×10^{-9}
rs331532	11	Intronic	G > A	2.80×10^{-8}	2.63×10^{-9}
rs317777	11	Exonic (OLFM5P)	G > A	2.29×10^{-8}	3.52×10^{-11}
rs147062602	11	Exonic (OR51B5)	CCAGGTCTGTGGCAGCC/-	2.47×10^{-8}	8.70×10^{-11}
rs372091	11	Intronic (OR51B5)	G > A	2.95×10^{-8}	4.43×10^{-11}
rs76076035	9	Intronic	T > C	3.23×10^{-8}	
Signals located on known genes with an association threshold of p < 5 × 10⁻⁶ (n = 34)					
rs4910823	11	Intronic (TRIM5, TRIM6-TRIM34)	A > G	7.14×10^{-8}	3.71×10^{-10}
rs11785309	8	Intronic (CSMD1)	T > C	1.18×10^{-7}	2.67×10^{-11}
rs73395847	11	Splice acceptor (C11orf40)	C > T	1.25×10^{-7}	7.08×10^{-10}
rs12412726	10	Intronic (LINC01435)	G > A	2.57×10^{-7}	
rs6554634	5	Transcription factor binding site (SLC6A19)	G > A	2.64×10^{-7}	
rs80034548	11	Exonic (OR51A1P)	G > C	3.08×10^{-7}	1.95×10^{-9}
rs1368823	11	Intronic (MMP26)	C > T	5.15×10^{-7}	9.32×10^{-9}
rs1455957	11	Linked with rs2472530 in OR52A5	A > G	5.37×10^{-7}	4.70×10^{-9}
rs73388885	11	3 prime UTR variant (OR51E2)	G > A	5.46×10^{-7}	1.05×10^{-9}
rs7933549	11	Missense (OR51V1)	G > A	1.08×10^{-6}	3.17×10^{-8}
rs73405065	11	Intronic (MMP26)	G > T	1.10×10^{-6}	5.57×10^{-9}
rs62071691	11	Intronic (CAMKK1)	A > G	1.10×10^{-6}	
rs9667878	11	Upstream variant (OR51V1)	C > T	1.11×10^{-6}	3.25×10^{-8}
rs73402629	11	Upstream transcript enhancer (HBG1)	C > A	1.41×10^{-6}	
rs11035718	11	Linked with rs2472530 in OR52A5	G > A	1.63×10^{-6}	2.33×10^{-8}
rs4525262	11	Intronic (TRIM22)	A > C	1.94×10^{-6}	1.41×10^{-10}
rs73405021	11	Intronic (MMP26)	A > G	2.06×10^{-6}	2.24×10^{-8}
rs317781	11	Intronic (UBQLN3)	A > C	2.06×10^{-6}	9.84×10^{-10}
rs35504021	9	Intronic (DOCK8)	G > A	2.36×10^{-6}	
rs1566860	8	Intronic (CSMD1)	G > A	2.53×10^{-6}	3.49×10^{-8}
rs141682404	6	Intronic (PACRG)	C > T	2.74×10^{-6}	
rs73402631	11	Upstream variant (HBG1)	C > G	2.98×10^{-6}	
rs10742697	11	Upstream variant (UBQLN3)	C > A	3.16×10^{-6}	1.63×10^{-9}
rs2213170	11	Intronic (HBE1)	A > G	3.16×10^{-6}	
rs2071348	11	Enhancer (HBBP1)	T > G	3.25×10^{-6}	4.19×10^{-9}
rs2599587	11	Intronic (TRIM21)	C > A	3.68×10^{-6}	
rs73427713	11	Downstream variant (OR52K1)	A > G	3.80×10^{-6}	4.43×10^{-8}
rs7026337	9	Intronic (NTRK2)	A > G	3.93×10^{-6}	
rs116926964	11	Intronic (ATG16L2)	G > T	3.93×10^{-6}	
rs36030093	8	Intronic (LOC105375776)	C > T	4.21×10^{-6}	
rs11033350	11	Intronic (OR51E2)	A > G	4.34×10^{-6}	2.77×10^{-8}
rs12286769	11	Upstream variant (OR52K2)	C > T	4.62×10^{-6}	2.33×10^{-8}
rs72981995	6	Intronic (LOC105378031)	A > C	4.85×10^{-6}	
rs967664	11	Downstream variant (OR52T1P)	T > G	4.92×10^{-6}	1.48×10^{-8}

Figure 2. Manhattan plot for SNP associations with TEEs in the Saudi SCD cohort. The GWAS association *p*-values of the detected SNPs were plotted across the 23 chromosomes. Only the top marker on each chromosome was annotated. The SNPs which met the genome-wide significance threshold ($p < 5 \times 10^{-8}$) are shown just above the red line. The rs ID numbers of some SNPs with $p < 5 \times 10^{-6}$ (above the blue line) are demonstrated. The SNPs were plotted in two different colures (black and grey dots) to show a distinction between the chromosomes.

Table 5. The common 10-SNP * haplotypes (observed in ≥5 individuals) seen in the tested SCD cohort of 65 cases and 285 controls.

Number of Observations	Haplotype	Haplotype Code	Inherited Cases (%)	Inherited Controls (%)	*p*-Value
11	CC-GG-ins-AG-del-TT-CC-TT-GG-GG	haplo-12	0 (0.0)	11 (16.92)	0.14
20	CC-GG-ins-GG-del-CT-CC-TT-GG-GG	haplo-17	2 (3.08)	18 (27.69)	0.39
8	CC-GG-ins-GG-del-TT-CC-TC-GG-GG	haplo-20	1 (1.54)	7 (10.77)	1.0
63	CC-GG-ins-GG-del-TT-CC-TT-GG-GG	haplo-22	5 (7.69)	58 (89.23)	**0.019**
5	CC-GG-ins-GG-ins-TT-CC-TT-GG-GG	haplo-25	1 (1.54)	4 (6.15)	1.0
5	TC-AA-del-AA-del-TT-TC-CC-CC-AA	haplo-40	0 (0.0)	5 (7.69)	0.59
5	TC-AG-W-AA-del-TT-TC-CC-CC-AA	haplo-50	0 (0.0)	5 (7.69)	0.59
5	TC-GG-ins-GG-del-TT-CC-TC-GG-GG	haplo-77	0 (0.0)	5 (7.69)	0.59
5	TC-GG-ins-GG-del-TT-CC-TT-GG-GG	haplo-78	0 (0.0)	5 (7.69)	0.59
15	TT-AA-del-AA-del-CT-TT-CC-CC-AA	haplo-84	5 (7.69)	10 (15.38)	0.17
23	TT-AA-del-AA-del-TT-TT-CC-CC-AA	haplo-89	8 (12.31)	15 (23.08)	0.051
7	TT-AA-del-AA-ins-CT-TT-CC-CC-AA	haplo-91	4 (6.15)	3 (4.62)	**0.024**
9	TT-AG-W-AA-del-TT-TT-CC-CC-AA	haplo-102	0 (0.0)	9 (13.85)	0.21
5	TT-AG-W-AG-del-TT-TT-CC-CC-AA	haplo-108	0 (0.0)	5 (7.69)	0.59

ins = insert, del = delete, W = wild-type copy (CAGCCCCAGGTCTGTGG), *p*-values in bold are significant. * The selected 10 SNPs are rs35390334, rs317777, rs147062602, rs4910823, rs139341092, rs76076035, rs73395847, rs1368823, rs80034548, and rs1455957.

4. Discussion

Major progress has been made in the genetics field since the advent of GWASs, which allows genetic testing across the whole genome with improved resolution to identify variations with the highest level of association [61]. A GWAS usually requires thousands

of cases and controls to detect modest to strong associations with sufficient statistical power [62]. However, for very strong associations, small (approximately 50 to 100) numbers of cases may suffice [63,64]. Various GWASs conducted on SCD patients have succeeded in identifying several distinct loci predicted to be additional phenotypic modifiers. For example, rs3115229 at Chr 4 showed a significant association with acute, severe vaso-occlusive pain in children with SCD [65]. Furthermore, GWAS data from two African cohorts showed significant associations between variable HbF levels and variants in GLP2R and near BCL11A and HBS1L-MYB [66,67].

A head-to-head single SNP comparison of variants located in selected thrombotic genes (FVL, PRT, and MTHFR) between our cases and controls showed no association with TEEs in SCD patients. The genotyping results related to MTHFR (rs1801133) were not consistent with TEE associations reported previously in other medical conditions among Iranians [31,32], Japanese [34], Italians [35], Americans [36], and Georgians [37]. This may indicate that the MTHFR variant has no general role in TEE susceptibility. In contrast, the genotyping comparisons for FVL and PRT variants in our study cohort were considered unreliable due to the extremely low frequency of their alleles (<0.01). In such a condition, inductive reasoning is not possible, and therefore the analysis of both variants was removed from the Section 3. These variants in FVL and PRT are known thrombotic factors impacting TEEs but not in SCD across various ethnic groups, including the Saudi population, as supported by previous studies [23–30]. A recent meta-analysis that included 18 studies from mixed populations comparing 30,234 VTE cases and 172,122 controls detected FVL (rs6025) as the highest signal marker (1.4×10^{-188}) [51]. The variant rs1799963 (in LD with rs191945075) in PRT was also reported with a very strong association (9.5×10^{-32}). Thus, our data do not rule out the possibility of the involvement of both markers in TEE development. Multiple variants in the SLCO1B1, PRIM1, APOB, TYK2, TSEN15, CYP4F2, and MST1 genes were previously suggested as risk factors for stroke with *p*-values $< 1.0 \times 10^{-5}$ in a GWAS on German pediatrics [52]. Our study on a larger number of adult stroke patients (n = 27) revealed no associations between the suggested markers and stroke. In contrast, the association between rs1364044 in ADAMTS and stroke in SCD, reported in pediatric patients from Brazil [53], was detected among our adult Saudis with stroke and SCD ($p = 3 \times 10^{-6}$ vs. 0.036, respectively). This finding may imply a role for the variant in stroke induction, specifically in SCD. An association suggested by Arning et al. for another stroke marker (rs2084898 in TRIM29) [52] was also confirmed in our study. Furthermore, GWAS signals at the $p < 5 \times 10^{-6}$ threshold for five SNPs (rs2229637 in ITPR3, rs10998957 in LINC02651-RPL5P26, rs10746487 in H6PD-SPSB1, rs1985317 in an intergenic region, and rs6771316 in LINC00877) were reported previously with TEEs [49–51]. These signals were detected in our study as well, but at lower association levels. This replication provides evidence that the reported associations are not chance findings.

Genome-wide significant or close to significant associations with 41 markers were identified in the current study, with many of them (73.2%) located on Chr 11. SCD is a monogenic disorder involving Chr 11 [3], with various complications also predicted to be partially influenced by parental genetics. Thus, it is not surprising that the majority of genetic modifiers were seen in Chr 11. Four of the top detected signals that met the GWAS significant threshold cut-off are intronic variants with no previous phenotypic association reports. Introns are noncoding loci; however, some intronic variants may play a role in manipulating gene function via disrupting the RNA splicing process that takes place at exon–intron boundaries [68].

Our data also showed a possible association with rs317777 in a pseudogene (OLFM5P). Pseudogenes are DNA segments that have lost their coding ability; hence, they have no direct impact on phenotype occurrence but have the potential to influence the expression and activity of other coding genes possibly related to phenotypic diversity [69]. The associations seen with variants in olfactory receptor genes, family 51 (OR51B5, OR51V1, OR51A1P, and OR51E2) and family 52 (OR52A5, OR52K1, OR52K2, and OR52T1P), represented 29.3% of the suggested risk variants. The results highlighted frameshift and missense variants in

two olfactory receptor genes (rs147062602 in OR51B5 and rs7933549 in OR51V1). These genes are homologous (paralog) and have closely similar structures and functions [70]. A frameshift variant is an insertion or deletion of a number of base pairs that generates a stop codon and usually causes a premature termination of a DNA sequence [71]. Previous reports have identified two variants in the olfactory receptor region associated with various phenotypes: rs7948471-A in OR51B5 associated with a higher degree of hemolysis in SCD ($p = 3 \times 10^{-10}$), rs5006884-A in OR51B5 associated with an increased HbF level ($p = 3 \times 10^{-8}$) [72], and rs7950726-A in OR51V1 associated with variable HbA2 levels in healthy adults ($p = 1 \times 10^{-11}$) [73]. In line with these findings, our results confirmed the association of rs5006884-A in OR51B5, where it was significantly less frequent in the cases with lower HbF levels than in the controls. This may suggest a linkage between HbF levels and TEE development. More findings in our study signify the role of olfactory receptors on TEEs. For instance, two of the detected markers were found to have a strong LD with a missense variant (rs2472530) in another olfactory receptor gene (OR52A5). Similarly, another suggested risk variant (rs12286769) is located in the upstream transcriptional region near OR52K2. A different variant (rs569117290) in this gene was previously reported with variable mean corpuscular hemoglobin concentrations [74]. The mutations in the olfactory receptor region (OR genes) may modulate the chromatin structure of the CTCF binding site within the β-globin gene and interfere with the gene-receptor interaction [75]. Moreover, selected variants within OR genes may regulate the expression of HBG [76]. A few olfactory receptor genes, such as OR2L13, OR4D6, and OR1N1, were recently suggested as risk factors for a number of TEE incidents [77]. A functional analysis indicated an upregulation of the transduction pathway through which the signaling of olfactory receptors modulates platelet activation underlying TEE [77,78].

A splice acceptor variant (rs73395847) in the C11orf40 gene was among the top-detected markers in our study that could possibly impact TEE development. Several mutations in C11orf40 were reported previously in patients with fibromyalgia syndrome, but the gene function is not clearly known [79]. This study also suggested that two variants (rs73402629 in HBG1 and rs2071348 in the human β-globin locus (HBBP1)) might be relevant to TEE risk ($p = 1.41 \times 10^{-6}$ and 3.25×10^{-6}, respectively). The variant rs2071348 was previously reported as a predictor for disease severity in anemic patients with β^0-thalassemia/hemoglobin E ($p = 3 \times 10^{-15}$, OR = 4.05) [80], whereas a strong association between HBG1 (rs998870472) and low hemoglobin levels was recently reported ($p = 1 \times 10^{-273}$) [81]. HBBP1 (a pseudogene) and HBG1 are both involved in the interaction of the locus control region with globin genes, which is a critical step for γ-globin regulation in adults [82]. Furthermore, rs6554634 is a predicted transcription factor binding site located approximately 12 kilobases (kb) upstream of SLC6A19, denoted in our study as a possible marker for TEE susceptibility. Certain cis-regulatory elements can be found hundreds of kb away from the actual transcriptional site [83]. The variant rs6554634 is not currently reported in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/, accessed on 1 January 2023), but was suggested as a biomarker affecting patients' response to cetuximab in the treatment of colorectal cancer ($p = 1.76 \times 10^{-6}$) [84]. Two haplotypes of independent markers showed significant associations: a haplo-91 with a sixfold risk of TEEs in the carriers and a haplo-22 with a protective effect. The haplotype analysis involved only the top ten independent risk loci. In addition, two-thirds of the suggested variants (n = 26) showed genome-wide significance when the GWAS analysis was restricted to stroke cases only. This may imply a role for these variants in the induction of thromboembolism. The recruitment of further cases may confirm this. Four common SNPs were previously suggested to assess HBB haplotype frequencies [46], but these SNPs were missing in our GWAS markers, so it was not possible to assess the haplotypes frequencies in our study cohort. A recent study on the Saudi population with SCD identified the common HBB haplotypes, where the Arab/Indian haplotype was predominant in the patients from the eastern province, whereas the Benin haplotype was most common in the southwestern individuals [85]. Moreover, the study showed higher incidents of stroke in 318 Southwesterners with SCD than in the

159 examined patients from the eastern region. A previous study on Egyptians indicated a higher risk of stroke in homozygous Benin/Benin than other haplotypes [86]. Consistent with this finding, we noticed a higher rate of TEE cases in the southwestern participants (34.4%), who are known to inherit Benin haplotype more frequently, in comparison to the eastern patients (8.4%).

This study included the largest genome-wide scan of Saudi SCD patients, and by considering markers that just failed to meet the normal GWAS threshold, it identified some interesting novel variants and a haplotype as possible risk factors; however, further work is needed to replicate the findings in an independent cohort to confirm that the detected associations are true signals and not related to technical or methodological bias [87]. Some of the detected signals in GWAS are located in noncoding regions, but these may influence the binding between enhancer elements and transcription factors, which ultimately modulate genes' expression [88]. Functional studies are further needed to interpret GWAS findings and provide possible mechanisms through which the detected variants impact TEE development [89].

5. Conclusions

This study showed no impact of the known thrombotic gene, MTHFR, on TEEs induced in Saudi patients with SCD. However, significant GWAS associations, at levels of $p < 5 \times 10^{-6}$, were identified between 41 variants (30 of them on Chr 11) and TEEs in SCD. Seven of the markers showed novel and stronger associations ($p < 5 \times 10^{-8}$), two of them were exonic variants (rs317777 in OLFM5P and rs147062602 in OR51B5), but these findings need further replication studies to be confirmed.

References

1. Jastaniah, W. Epidemiology of sickle cell disease in Saudi Arabia. *Ann. Saudi Med.* **2011**, *31*, 289–293. [CrossRef]
2. Al-Suliman, A.; Elsarraf, N.A.; Baqishi, M.; Homrany, H.; Bousbiah, J.; Farouk, E. Patterns of mortality in adult sickle cell disease in the Al-Hasa region of Saudi Arabia. *Ann. Saudi Med.* **2006**, *26*, 487–488. [CrossRef]
3. Wu, D.Y.; Ugozzoli, L.; Pal, B.K.; Wallace, R.B. Allele-specific enzymatic amplification of β-globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 2757–2760. [CrossRef]
4. Carden, M.A.; Fasano, R.M.; Meier, E.R. Not all red cells sickle the same: Contributions of the reticulocyte to disease pathology in sickle cell anemia. *Blood Rev.* **2020**, *40*, 100637. [CrossRef]
5. Vichinsky, E. Chronic organ failure in adult sickle cell disease. *Hematol. Am. Soc. Hematol. Educ. Program* **2017**, *2017*, 435–439. [CrossRef]
6. Janjua, T.K.; Haider, S.A.; Raza, N. Multiple complications in sickle cell anaemia. *J. Pak. Med. Assoc.* **2018**, *68*, 154–156.
7. van Hamel Parsons, V.; Gardner, K.; Patel, R.; Thein, S.L. Venous thromboembolism in adults with sickle cell disease: Experience of a single centre in the UK. *Ann. Hematol.* **2016**, *95*, 227–232. [CrossRef]
8. Wun, T.; Brunson, A. Sickle cell disease: An inherited thrombophilia. *Hematol. Am. Soc. Hematol. Educ. Program* **2016**, *2016*, 640–647. [CrossRef]
9. Noubouossie, D.; Key, N.S.; Ataga, K.I. Coagulation abnormalities of sickle cell disease: Relationship with clinical outcomes and the effect of disease modifying therapies. *Blood Rev.* **2016**, *30*, 245–256. [CrossRef]
10. Shet, A.S.; Lizarralde-Iragorri, M.A.; Naik, R.P. The molecular basis for the prothrombotic state in sickle cell disease. *Haematologica* **2020**, *105*, 2368. [CrossRef]
11. Austin, H.; Key, N.S.; Benson, J.M.; Lally, C.; Dowling, N.F.; Whitsett, C.; Hooper, W.C. Sickle cell trait and the risk of venous thromboembolism among blacks. *Blood* **2007**, *110*, 908–912. [CrossRef] [PubMed]
12. Little, I.; Vinogradova, Y.; Orton, E.; Kai, J.; Qureshi, N. Venous thromboembolism in adults screened for sickle cell trait: A population-based cohort study with nested case-control analysis. *BMJ Open* **2017**, *7*, e012665. [CrossRef] [PubMed]
13. Stein, P.D.; Beemath, A.; Meyers, F.A.; Skaf, E.; Olson, R.E. Deep venous thrombosis and pulmonary embolism in hospitalized patients with sickle cell disease. *Am. J. Med.* **2006**, *119*, 897.e7–897.e11. [CrossRef] [PubMed]
14. Ohene-Frempong, K.; Weiner, S.J.; Sleeper, L.A.; Miller, S.T.; Embury, S.; Moohr, J.W.; Wethers, D.L.; Pegelow, C.H.; Gill, F.M.; The Cooperative Study of Sickle Cell Disease. Cerebrovascular accidents in sickle cell disease: Rates and risk factors. *Blood* **1998**, *91*, 288–294. [PubMed]
15. Kirkham, F.J.; Lagunju, I.A. Epidemiology of stroke in sickle cell disease. *J. Clin. Med.* **2021**, *10*, 4232. [CrossRef] [PubMed]
16. Naik, R.P.; Streiff, M.B.; Haywood, C., Jr.; Segal, J.B.; Lanzkron, S. Venous thromboembolism incidence in the Cooperative Study of Sickle Cell Disease. *J. Thromb. Haemost.* **2014**, *12*, 2010–2016. [CrossRef]
17. Brunson, A.; Lei, A.; Rosenberg, A.S.; White, R.H.; Keegan, T.; Wun, T. Increased incidence of VTE in sickle cell disease patients: Risk factors, recurrence and impact on mortality. *Br. J. Haematol.* **2017**, *178*, 319–326. [CrossRef]
18. Alsultan, A.; Aleem, A.; Ghabbour, H.; AlGahtani, F.H.; Al-Shehri, A.; Osman, M.E.; Kurban, K.; Alsultan, M.S.; Bahakim, H.; Al-Momen, A.M. Sickle cell disease subphenotypes in patients from Southwestern Province of Saudi Arabia. *J. Pediatr. Hematol. Oncol.* **2012**, *34*, 79–84. [CrossRef]
19. Al-Gahtani, F.H. Thromboembolic Events among Patients with Sickle Cell Disease: Risk Factors and Prevalence in a Tertiary Hospital in Saudi Arabia. *J. Hematol. Thrombo. Dis.* **2016**, *4*, 1000254. [CrossRef]
20. Cushman, M. Epidemiology and risk factors for venous thrombosis. *Semin. Hematol.* **2007**, *44*, 62–69. [CrossRef]
21. Porter, B.; Key, N.S.; Jauk, V.C.; Adam, S.; Biggio, J.; Tita, A. Impact of sickle hemoglobinopathies on pregnancy-related venous thromboembolism. *Am. J. Perinatol.* **2014**, *31*, 805–809. [CrossRef]
22. Naik, R.P.; Streiff, M.B.; Lanzkron, S. Sickle cell disease and venous thromboembolism: What the anticoagulation expert needs to know. *J. Thromb. Thrombolysis* **2013**, *35*, 352–358. [CrossRef]
23. Wu, O.; Robertson, L.; Twaddle, S.; Lowe, G.D.; Clark, P.; Greaves, M.; Walker, I.D.; Langhorne, P.; Brenkel, I.; Regan, L.; et al. Screening for thrombophilia in high-risk situations: Systematic review and cost-effectiveness analysis. The Thrombosis: Risk and Economic Assessment of Thrombophilia Screening (TREATS) study. *Health Technol. Assess.* **2006**, *10*, 1–110. [CrossRef]
24. Wuthrich, R.P.; Cicvara-Muzar, S.; Booy, C.; Maly, F.E. Heterozygosity for the factor V Leiden (G1691A) mutation predisposes renal transplant recipients to thrombotic complications and graft loss. *Transplantation* **2001**, *72*, 549–550. [CrossRef]
25. Al-Allawi, N.A.; Badi, A.I.; Goran, M.A.; Nerweyi, F.F.; Ballo, H.M.; Al-Mzury, N.T. The Contributions of Thrombophilic Mutations to Genetic Susceptibility to Deep Venous Thrombosis in Iraqi Patients. *Genet. Test. Mol. Biomark.* **2015**, *19*, 500–504. [CrossRef] [PubMed]
26. Konecny, F. Inherited trombophilic states and pulmonary embolism. *J. Res. Med. Sci.* **2009**, *14*, 43–56. [PubMed]

27. Li, C.; Ren, H.; Chen, H.; Song, J.; Li, S.; Lee, C.; Liu, J.; Cui, Y. Prothrombin G20210A (rs1799963) polymorphism increases myocardial infarction risk in an age-related manner: A systematic review and meta-analysis. *Sci. Rep.* **2017**, *7*, 13550. [CrossRef] [PubMed]

28. Gawish, G. The Prevalence of Inherited Thrombophilic Polymorphisms in Saudi Females with Recurrent Pregnancy Loss Confirmed using Different Screening Protocols of PCR. *J. Mol. Genet. Med.* **2015**, *9*, 1747.

29. Badawy, A.; AlSel, B.A.; Fawzy, M. Factor V Leiden G1691A and Prothrombin G20210A mutations are associated with repeated spontaneous miscarriage in Northern area of Saudi Arabia. *Gen. Mol. Res.* **2017**, *16*, gmr16039810. [CrossRef]

30. Shafia, S.; Zargar, M.H.; Khan, N.; Ahmad, R.; Shah, Z.A.; Asimi, R. High prevalence of factor V Leiden and prothrombin G20101A mutations in Kashmiri patients with venous thromboembolism. *Gene* **2018**, *654*, 1–9. [CrossRef]

31. Ehsani, M.; Imani, A.; Moravveji, A. Prevalence of factor V leiden, MTHFR C677T and MTHFR A1298C polymorphisms in patients with deep vein thrombosis in Central Iran. *Mol. Biol. Rep.* **2018**, *45*, 621–624. [CrossRef]

32. Hosseini, S.; Kalantar, E.; Hosseini, M.S.; Tabibian, S.; Shamsizadeh, M.; Dorgalaleh, A. Genetic risk factors in patients with deep venous thrombosis, a retrospective case control study on Iranian population. *Thromb. J.* **2015**, *13*, 35. [CrossRef] [PubMed]

33. Ghaznavi, H.; Soheili, Z.; Samiei, S.; Soltanpour, M.S. Association of Methylenetetrahydrofolate Reductase C677T Polymorphism with Hyperhomocysteinemia and Deep Vein Thrombosis in the Iranian Population. *Vasc. Spec. Int.* **2015**, *31*, 109–114. [CrossRef] [PubMed]

34. Kamei, H.; Nakamura, T.; Nagai, S.; Ishigami, M.; Hamajima, N. Possible Association Between the Methylenetetrahydrofolate Reductase Gene C677T Polymorphism and Preexisting Portal Vein Thrombosis in Liver Transplant Recipients. *Exp. Clin. Transplant.* **2016**, *14*, 313–316. [PubMed]

35. Ventura, P.; Venturelli, G.; Marcacci, M.; Fiorini, M.; Marchini, S.; Cuoghi, C.; Pietrangelo, A. Hyperhomocysteinemia and MTHFR C677T polymorphism in patients with portal vein thrombosis complicating liver cirrhosis. *Thromb. Res.* **2016**, *141*, 189–195. [CrossRef] [PubMed]

36. Spronk, K.J.; Olivero, A.D.; Haw, M.P.; Vettukattil, J.J. Methylenetetrahydrofolate Reductase C677T: Hypoplastic Left Heart and Thrombosis. *World J. Pediatr. Congenit. Heart Surg.* **2015**, *6*, 643–645. [CrossRef]

37. Garakanidze, S.; Costa, E.; Bronze-Rocha, E.; Santos-Silva, A.; Nikolaishvili, G.; Nakashidze, I.; Kakauridze, N.; Glonti, S.; Khukhunaishvili, R.; Koridze, M.; et al. Methylenetetrahydrofolate Reductase Gene Polymorphism (C677T) as a Risk Factor for Arterial Thrombosis in Georgian Patients. *Clin. Appl. Thromb. Hemost.* **2018**, *24*, 1061–1066. [CrossRef]

38. Settin, A.A.; Alghasham, A.; Ali, A.; Dowaidar, M.; Ismail, H. Frequency of thrombophilic genetic polymorphisms among Saudi subjects compared with other populations. *Hematology* **2012**, *17*, 176–182. [CrossRef]

39. Pandey, S.K.; Meena, A.; Kishor, K.; Mishra, R.M.; Pandey, S.; Saxena, R. Prevalence of factor V Leiden G1691A, MTHFR C677T, and prothrombin G20210A among Asian Indian sickle cell patients. *Clin. Appl. Thromb. Hemost.* **2012**, *18*, 320–323. [CrossRef]

40. Nishank, S.S.; Singh, M.P.; Yadav, R. Clinical impact of factor V Leiden, prothrombin G20210A, and MTHFR C677T mutations among sickle cell disease patients of Central India. *Eur. J. Haematol.* **2013**, *91*, 462–466. [CrossRef]

41. Kangne, H.K.; Jijina, F.F.; Italia, Y.M.; Jain, D.L.; Nadkarni, A.H.; Ghosh, K.K.; Colah, R.B. The Prevalence of Factor V Leiden (G1691A) and Methylenetetrahydrofolate Reductase C677T Mutations in Sickle Cell Disease in Western India. *Clin. Appl. Thromb. Hemost.* **2015**, *21*, 186–189. [CrossRef] [PubMed]

42. Moreira Neto, F.; Lourenco, D.M.; Noguti, M.A.; Morelli, V.M.; Gil, I.C.; Beltrao, A.C.; Figueiredo, M.S. The clinical impact of MTHFR polymorphism on the vascular complications of sickle cell disease. *Braz. J. Med. Biol. Res.* **2006**, *39*, 1291–1295. [CrossRef] [PubMed]

43. Hatzlhofer, B.L.; Bezerra, M.A.; Santos, M.N.; Albuquerque, D.M.; Freitas, E.M.; Costa, F.F.; Araujo, A.S.; Muniz, M.T. MTHFR polymorphic variant C677T is associated to vascular complications in sickle-cell disease. *Genet. Test. Mol. Biomark.* **2012**, *16*, 1038–1043. [CrossRef] [PubMed]

44. Nefissi, R.B.; Ouali, F.; Massaoud, T.; Gritli, N. Thrombophilic Mutations among Patients with Sickle Cell Disease. *Clin. Lab.* **2017**, *63*, 1815–1818. [CrossRef]

45. Fawaz, N.A.; Bashawery, L.; Al-Sheikh, I.; Qatari, A.; Al-Othman, S.S.; Almawi, W.Y. Factor V-Leiden, prothrombin G20210A, and MTHFR C677T mutations among patients with sickle cell disease in Eastern Saudi Arabia. *Am. J. Hematol.* **2004**, *76*, 307–309. [CrossRef]

46. Shaikho, E.M.; Farrell, J.J.; Alsultan, A.; Qutub, H.; Al-Ali, A.K.; Figueiredo, M.S.; Chui, D.H.K.; Farrer, L.A.; Murphy, G.J.; Mostoslavsky, G.; et al. A phased SNP-based classification of sickle cell anemia HBB haplotypes. *BMC Genom.* **2017**, *18*, 608. [CrossRef]

47. Witte, J.S.; Elston, R.C.; Schork, N.J. Genetic dissection of complex traits. *Nat. Genet.* **1996**, *12*, 355–356. [CrossRef]

48. Eliades, N.-G.; Eliades, D.G. HAPLOTYPE ANALYSIS: Software for analysis of haplotype data. In *Forest Genetics and Forest Tree Breeding*; Georg-August University: Goettingen, Germany, 2009.

49. Greliche, N.; Germain, M.; Lambert, J.-C.; Cohen, W.; Bertrand, M.; Dupuis, A.-M.; Letenneur, L.; Lathrop, M.; Amouyel, P.; Morange, P.-E. A genome-wide search for common SNP x SNP interactions on the risk of venous thrombosis. *BMC Med. Genet.* **2013**, *14*, 36. [CrossRef]

50. Heit, J.A.; Armasu, S.M.; McCauley, B.M.; Kullo, I.J.; Sicotte, H.; Pathak, J.; Chute, C.G.; Gottesman, O.; Bottinger, E.P.; Denny, J.C. Identification of unique venous thromboembolism-susceptibility variants in African-Americans. *Thromb. Haemost.* **2017**, *117*, 758–768. [CrossRef]

51. Lindström, S.; Wang, L.; Smith, E.N.; Gordon, W.; van Hylckama Vlieg, A.; De Andrade, M.; Brody, J.A.; Pattee, J.W.; Haessler, J.; Brumpton, B.M. Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. *Blood* **2019**, *134*, 1645–1657. [CrossRef]

52. Arning, A.; Hiersche, M.; Witten, A.; Kurlemann, G.; Kurnik, K.; Manner, D.; Stoll, M.; Nowak-Göttl, U. A genome-wide association study identifies a gene network of ADAMTS genes in the predisposition to pediatric stroke. *Blood* **2012**, *120*, 5231–5236. [CrossRef] [PubMed]

53. Santiago, R.P.; Vieira, C.; Adanho, C.S.A.; Guarda, C.C.; Santana, S.S.; Figueiredo, C.V.B.; Aleluia, M.M.; Pitanga, T.N.; Fiuza, L.M.; Maffili, V.V. Genome Wide Association Study of Sickle Cell Disease Individuals with Stroke Risk. *Blood* **2016**, *128*, 4852. [CrossRef]

54. Dichgans, M.; Malik, R.; König, I.R.; Rosand, J.; Clarke, R.; Gretarsdottir, S.; Thorleifsson, G.; Mitchell, B.D.; Assimes, T.L.; Levi, C. Shared genetic susceptibility to ischemic stroke and coronary artery disease: A genome-wide analysis of common variants. *Stroke* **2014**, *45*, 24–36. [CrossRef] [PubMed]

55. Ikram, M.A.; Seshadri, S.; Bis, J.C.; Fornage, M.; DeStefano, A.L.; Aulchenko, Y.S.; Debette, S.; Lumley, T.; Folsom, A.R.; Van Den Herik, E.G. Genomewide association studies of stroke. *N. Engl. J. Med.* **2009**, *360*, 1718–1728. [CrossRef]

56. Gudbjartsson, D.F.; Holm, H.; Gretarsdottir, S.; Thorleifsson, G.; Walters, G.B.; Thorgeirsson, G.; Gulcher, J.; Mathiesen, E.B.; Njølstad, I.; Nyrnes, A. A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat. Genet.* **2009**, *41*, 876–878. [CrossRef]

57. Holliday, E.G.; Maguire, J.M.; Evans, T.-J.; Koblar, S.A.; Jannes, J.; Sturm, J.W.; Hankey, G.J.; Baker, R.; Golledge, J.; Parsons, M.W. Common variants at 6p21. 1 are associated with large artery atherosclerotic stroke. *Nat. Genet.* **2012**, *44*, 1147–1151. [CrossRef]

58. Zholdybayeva, E.V.; Medetov, Y.Z.; Aitkulova, A.M.; Makhambetov, Y.T.; Akshulakov, S.K.; Kaliyev, A.B.; Talzhanov, Y.A.; Kulmambetova, G.N.; Iskakova, A.N.; Ramankulov, Y.M. Genetic risk factors for intracranial aneurysm in the Kazakh population. *J. Mol. Neurosci.* **2018**, *66*, 135–145. [CrossRef]

59. Sebastiani, P.; Ramoni, M.F.; Nolan, V.; Baldwin, C.T.; Steinberg, M.H. Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. *Nat. Genet.* **2005**, *37*, 435–440. [CrossRef]

60. Flanagan, J.M.; Frohlich, D.M.; Howard, T.A.; Schultz, W.H.; Driscoll, C.; Nagasubramanian, R.; Mortier, N.A.; Kimble, A.C.; Aygun, B.; Adams, R.J. Genetic predictors for stroke in children with sickle cell anemia. *Blood* **2011**, *117*, 6681–6684. [CrossRef]

61. Daly, A. Genetic Basis of Drug-Induced Liver Injury Linked to Commonly Prescribed Drugs. In Proceedings of the Abstracts of the XXIX International Congress of the European Association of Poison Centres and Clinical Toxicologists, Stockholm, Sweden, 12–15 May 2009; Volume 47, p. 438.

62. Witte, J.S. Genome-wide association studies and beyond. *Annu. Rev. Public Health* **2010**, *31*, 9–20. [CrossRef]

63. Link, E.; Parish, S.; Armitage, J.; Bowman, L.; Heath, S.; Matsuda, F.; Gut, I.; Lathrop, M.; Collins, R. SLCO1B1 variants and statin-induced myopathy—A genomewide study. *N. Engl. J. Med.* **2008**, *359*, 789–799. [CrossRef] [PubMed]

64. Daly, A.K.; Donaldson, P.T.; Bhatnagar, P.; Shen, Y.; Pe'er, I.; Floratos, A.; Daly, M.J.; Goldstein, D.B.; John, S.; Nelson, M.R. HLA-B* 5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat. Genet.* **2009**, *41*, 816–819. [CrossRef]

65. Chaturvedi, S.; Bhatnagar, P.; Bean, C.J.; Steinberg, M.H.; Milton, J.N.; Casella, J.F.; Barron-Casella, E.; Arking, D.E.; DeBaun, M.R. Genome-wide association study to identify variants associated with acute severe vaso-occlusive pain in sickle cell anemia. *Blood* **2017**, *130*, 686–688. [CrossRef] [PubMed]

66. Bhatnagar, P.; Purvis, S.; Barron-Casella, E.; DeBaun, M.R.; Casella, J.F.; Arking, D.E.; Keefer, J.R. Genome-wide association study identifies genetic variants influencing F-cell levels in sickle-cell patients. *J. Hum. Genet.* **2011**, *56*, 316–323. [CrossRef]

67. Mtatiro, S.N.; Singh, T.; Rooks, H.; Mgaya, J.; Mariki, H.; Soka, D.; Mmbando, B.; Msaki, E.; Kolder, I.; Thein, S.L. Genome wide association study of fetal hemoglobin in sickle cell anemia in Tanzania. *PLoS ONE* **2014**, *9*, e111464. [CrossRef]

68. Anna, A.; Monika, G. Splicing mutations in human genetic disorders: Examples, detection, and confirmation. *J. Appl. Genet.* **2018**, *59*, 253–268. [CrossRef] [PubMed]

69. Lindsay, M.A.; Griffiths-Jones, S.; Pink, R.C.; Carter, D.R. Pseudogenes as regulators of biological function. *Essays Biochem.* **2013**, *54*, 103–112. [CrossRef] [PubMed]

70. Ibn-Salem, J.; Muro, E.M.; Andrade-Navarro, M.A. Co-regulation of paralog genes in the three-dimensional chromatin architecture. *Nucleic Acids Res.* **2017**, *45*, 81–91. [CrossRef]

71. Roth, J.R. Frameshift mutations. *Annu. Rev. Genet.* **1974**, *8*, 319–346. [CrossRef]

72. Solovieff, N.; Milton, J.N.; Hartley, S.W.; Sherva, R.; Sebastiani, P.; Dworkis, D.A.; Klings, E.S.; Farrer, L.A.; Garrett, M.E.; Ashley-Koch, A. Fetal hemoglobin in sickle cell anemia: Genome-wide association studies suggest a regulatory region in the 5′ olfactory receptor gene cluster. *Blood* **2010**, *115*, 1815–1822. [CrossRef]

73. Menzel, S.; Garner, C.; Rooks, H.; Spector, T.D.; Thein, S.L. H b A 2 levels in normal adults are influenced by two distinct genetic mechanisms. *Br. J. Haematol.* **2013**, *160*, 101–105. [CrossRef] [PubMed]

74. Wojcik, G.L.; Graff, M.; Nishimura, K.K.; Tao, R.; Haessler, J.; Gignoux, C.R.; Highland, H.M.; Patel, Y.M.; Sorokin, E.P.; Avery, C.L. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **2019**, *570*, 514–518. [PubMed]

75. Hou, C.; Dale, R.; Dean, A. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 3651–3656. [CrossRef] [PubMed]

76. Steinberg, M.H.; Sebastiani, P. Genetic modifiers of sickle cell disease. *Am. J. Hematol.* **2012**, *87*, 795–803.

77. Morrell, C.N.; Mix, D.; Aggarwal, A.; Bhandari, R.; Godwin, M.; Owens, P.; Lyden, S.P.; Doyle, A.; Krauel, K.; Rondina, M.T. Platelet olfactory receptor activation limits platelet reactivity and growth of aortic aneurysms. *J. Clin. Investig.* **2022**, *132*, e152373. [CrossRef]

78. Bosch, F.; van Dijk, F.; Briedé, S.; Groen, J.; Hanna-Sawires, R.; Klok, F.; Kaasjager, H.; Brosens, L.; Molenaar, I.; Bonsing, B.; et al. Tumor Gene Expression Associates with Venous Thromboembolism in Patients with Pancreatic Cancer. *Res. Pract. Thromb. Haemost.* **2021**, *5* (Suppl. S2). Available online: https://abstracts.isth.org/abstract/tumor-gene-expression-associates-with-venous-thromboembolism-in-patients-with-pancreatic-cancer/ (accessed on 7 February 2023).

79. Feng, J.; Zhang, Z.; Wu, X.; Mao, A.; Chang, F.; Deng, X.; Gao, H.; Ouyang, C.; Dery, K.J.; Le, K. Discovery of potential new gene variants and inflammatory cytokine associations with fibromyalgia syndrome by whole exome sequencing. *PLoS ONE* **2013**, *8*, e65033. [CrossRef]

80. Nuinoon, M.; Makarasara, W.; Mushiroda, T.; Setianingsih, I.; Wahidiyat, P.A.; Sripichai, O.; Kumasaka, N.; Takahashi, A.; Svasti, S.; Munkongdee, T. A genome-wide association identified the common genetic variants influence disease severity in β 0-thalassemia/hemoglobin E. *Hum. Genet.* **2010**, *127*, 303–314. [CrossRef]

81. Lee, C.-J.; Chen, T.-H.; Lim, A.M.W.; Chang, C.-C.; Sie, J.-J.; Chen, P.-L.; Chang, S.-W.; Wu, S.-J.; Hsu, C.-L.; Hsieh, A.-R. Phenome-wide analysis of Taiwan Biobank reveals novel glycemia-related loci and genetic risks for diabetes. *Commun. Biol.* **2022**, *5*, 1175. [CrossRef]

82. Huang, P.; Keller, C.A.; Giardine, B.; Grevet, J.D.; Davies, J.O.; Hughes, J.R.; Kurita, R.; Nakamura, Y.; Hardison, R.C.; Blobel, G.A. Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. *Genes Dev.* **2017**, *31*, 1704–1713. [CrossRef]

83. Palstra, R.-J.T.S. Spatial Organisation of the [β]-globin Locus: The Active Chromatin Hub. Ph.D. Thesis, Erasmus University Rotterdam, Rotterdam, The Netherlands, 2005.

84. Wormald, S.; Milla, L.; O'Connor, L. Association of candidate single nucleotide polymorphisms with somatic mutation of the epidermal growth factor receptor pathway. *BMC Med. Genom.* **2013**, *6*, 43. [CrossRef]

85. Al-Ali, A.K.; Alsulaiman, A.; Alfarhan, M.; Safaya, S.; Vatte, C.B.; Albuali, W.M.; Qutub, H.O.; Alzahrani, A.J.; Milton, J.N.; Steinberg, M.H. Sickle cell disease in the Eastern Province of Saudi Arabia: Clinical and laboratory features. *Am. J. Hematol.* **2021**, *96*, E117–E121. [CrossRef] [PubMed]

86. Abou-Elew, H.H.; Youssry, I.; Hefny, S.; Hashem, R.H.; Fouad, N.; Zayed, R.A. β^S globin gene haplotype and the stroke risk among Egyptian children with sickle cell disease. *Hematology* **2018**, *23*, 362–367. [CrossRef] [PubMed]

87. Marigorta, U.M.; Rodríguez, J.A.; Gibson, G.; Navarro, A. Replicability and prediction: Lessons and challenges from GWAS. *Trends Genet.* **2018**, *34*, 504–517. [CrossRef] [PubMed]

88. Morova, T.; Ding, Y.; Huang, C.-C.F.; Sar, F.; Schwarz, T.; Giambartolomei, C.; Baca, S.C.; Grishin, D.; Hach, F.; Gusev, A. Optimized high-throughput screening of non-coding variants identified from genome-wide association studies. *Nucleic Acids Res.* **2022**, *51*, e18.

89. Cano-Gamez, E.; Trynka, G. From GWAS to function: Using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.* **2020**, *11*, 424. [CrossRef]