



The 65th ASH Annual Meeting Abstracts

POSTER ABSTRACTS

114. SICKLE CELL DISEASE, SICKLE CELL TRAIT AND OTHER HEMOGLOBINOPATHIES, EXCLUDING THALASSEMIAS: CLINICAL AND EPIDEMIOLOGICAL
Investigating Thrombosis Risk Factors in Sickle Cell Disease Utilizing Archetypal Analysis

Olufunke Y. Martin, MD^{1,2}, Bindu Parachalil Gopalan, PhD², Jian Sun, PhD³, Ashima Singh, PhD⁴, Neal Jeffries, PhD⁵, Arun S. Shet, MDPhD²

¹Center for Cancer and Blood Disorders, Children's National Hospital, Woodbury, MN

²Sickle Thrombosis and Vascular Biology Lab, Sickle Cell Branch, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD

³Integrated Data Sciences Section, Research Technologies Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

⁴Department of Pediatrics, Division of Hematology and Oncology, Medical College of Wisconsin, Pewaukee, WI

⁵Office of Biostatistics Research, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD

Background: Thrombosis in sickle cell disease (SCD) affects both the arterial and venous vasculature with devastating consequences. Individuals with SCD have an increased risk of venous thromboembolism (VTE). The risks are multifactorial with prior studies identifying both VTE-specific and SCD-specific risk factors. Machine learning (ML) offers the potential to detect non-linear relationships in complex multifactorial diseases and has successfully predicted outcomes in deeply phenotyped patients. We investigated whether archetypal analysis (AA), a method of ML would identify clinical and laboratory factors associated with arteriovenous thrombosis in patients with SCD.

Methods: We retrospectively analyzed a prospective longitudinal cohort of patients with SCD (NCT00011648) studied between 2000-2017. A prior history of VTE and/or stroke at the time of study entry as captured by the electronic medical record was used to define index "cases". SCD patients without a history of thrombosis were "controls". The demographic and clinical characteristics of cases and controls were compared using Wilcoxon rank-sum test for continuous and Fisher's exact test for categorical variables. Unsupervised AA was used to identify patterns of clinical and laboratory test characteristics within the entire cohort of patients including cases and controls. AA identifies patterns are on the convex hull that encloses all patients in the cohort in the high-dimensional feature space. Once specific archetypes were identified, the frequency of each archetype was determined within cases and controls. In addition, we describe the age and laboratory characteristics of the archetypes that were significantly more prevalent in cases as compared to controls.

Results: The cohort consisted of 633 out of 862 patients that had their whole exome sequenced. A combined total of 160 arterial and venous thrombosis events was detected in 148/633 patients (24%) with half (n=79) attributable to arterial stroke and half (n=81) attributable to VTE [Table 1]. A small proportion of patients had both arterial and venous thrombosis (n=12). The median age (31 years, p=0.94) and sex were similar between the cases (n=148) and controls (n=485). The distribution of genotypes was similar between the two groups (p=0.09). A significantly higher proportion of patient with thrombosis had a prior history of blood transfusion (p<0.0001) and history of insertion of a central venous access device (p<0.0001) [Table 1]. History of medical management with hydroxyurea was similar among the cases and controls (p=0.70). Compared to controls, the cases had elevated laboratory parameters including AST levels (p=0.02), and serum ferritin (p<0.0001), while HbS percentages were significantly lower (p<0.0001). Based on the minimum error in matrix factorization, a total of 15 archetypes were identified in the cohort [Figure 1]. Each patient was decomposed as a weighted combination of the 15 archetypes. The contributing archetypes for each patient were those archetypes that have non-zero weights. Among those 15 archetypes, archetypes 9, 14 and 15 were significantly more enriched among the cases compared to controls (p<0.02) [Figure 1]. We observed elevated serum ferritin levels and low levels of HbS across the three archetypes, which are consistent with the previously identified risk factors [Figure 1].

Conclusions: Machine-identified archetypes are objective and reproducible feature representations. Archetype analysis-based machine learning appears to hold promise to identify both traditional VTE and SCD-specific risks for clinical thrombosis. Future studies will utilize archetypal analysis to investigate associations between clinical, laboratory and genetic markers for thrombosis and assess their reproducibility using validation cohorts.

Disclosures No relevant conflicts of interest to declare.

TABLE 1: CLINICAL VARIABLES STRATIFIED BY THROMBOTIC EVENTS

All patients N=633	Thrombosis (Arterial + Venous) "Cases" N= 148, 24%	No Thrombosis "Controls" N=485, 76%	*p value
Age, median (IQR), years	31 (24-44)	31 (24-41)	0.94
Sex (N,%)			0.85
Male	73 (49%)	233 (48%)	
Female	75 (51%)	252 (52%)	
Race			
Black/AA	137 (93%)	456 (94%)	0.53
Sickle Cell Genotype N (%)			0.09
Hgb SS	119 (80%)	382 (79%)	
Hgb SC	21 (14%)	69 (14%)	
Hgb Sβ0 Thal	3 (2%)	13 (3%)	
Hgb Sβ+ Thal	2 (1%)	20 (4%)	
BMI (IQR), N=627	22.9 (20.6 – 27.7)	23 (20.7 – 26.3)	0.74
History of Transfusion, N (%)	142 (96%)	425 (88%)	<0.0001
History of Tobacco Use, N (%)	28 (19%)	63 (13%)	0.08
Current Hydroxyurea Use, N (%)	56 (38%)	174 (36%)	0.70
History of CVAD, N (%)	58 (39%)	73 (15%)	<0.0001

Figure 1: Unsupervised Learning. The contributing machine-identified patterns based on the weight were identified for each of the Archetype patients. Archetypes with weights of zero or close to zero are not considered contributing patterns. The y axis indicated the percentile of the levels in the original cohort. The x axis are the variables used in the archetypal analysis. The red box indicates a significant archetype.

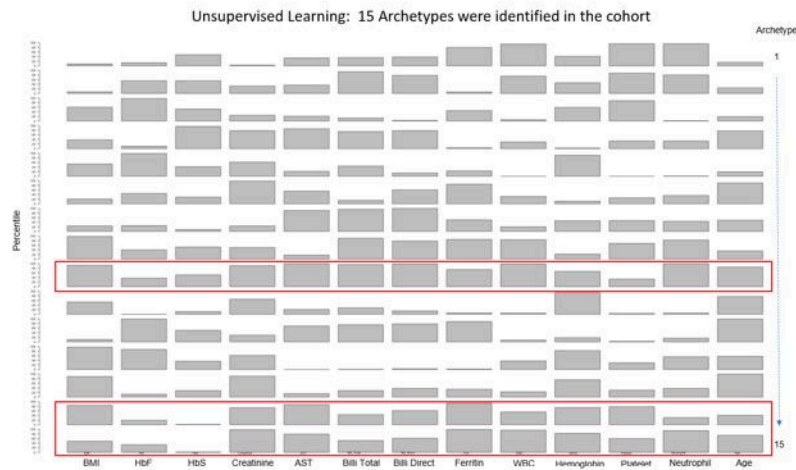


Figure 1

<https://doi.org/10.1182/blood-2023-190569>