Supplemental Material

Supplemental Methods
Supplemental Figure 1. Pathway enrichment analysis of SCT-associated proteins
Supplemental Figure 2: SCT-protein association p-values with and without adjustment for genetic ancestry (i.e. PC1 or % African ancestry).
Supplemental Figure 3. Correlation between 35 SCT-associated proteins and baseline eGFR or hemoglobin in WHI.
Supplemental Table 1: Fully summary statistics for protein associations with SCT in discovery WHI cohort.
Supplemental Table 2: Demographics table for replication cohorts
Supplemental Table 3: Replication results using SomaScan platform in ARIC and JHS, including information on estimated correlation between Olink and SomaScan from prior publications.
Supplemental Table 4: Additional proteins that were assay-wide significant for association with SCT in JHS (with replication in ARIC).
Supplemental Table 5: Protein risk score (PrRS) weights

Supplemental Methods

The Women's Health Initiative (WHI) discovery data set and data collection

WHI is a long-term, prospective national health study focused on identifying optimal strategies for preventing chronic diseases in older women of all race/ethnicity and socioeconomic backgrounds[1,2]. WHI initially recruited 161,808 post-menopausal women aged 50-79 years between 1993 and 1998 from 40 US clinical centers, including 14,618 self-identified African American women. For the current study, there were 14,118 African American eligible participants who had previously undergone genotyping for the rs334 sickle cell mutation. We randomly selected 592 participants with SCT and 592 age-matched individuals with the major allele rs334 genotype, who underwent proteomic profiling on stored baseline EDTA plasma samples using the Olink® Explore 1536 panel[3]. Data were quality controlled and normalized using an internal extension control and a plate control and presented as relative protein expression using Normalized Protein eXpression (NPX) values, which is an arbitrary unit on a log2-scale.

Baseline socio-demographic and lifestyle characteristics and medical history were collected using standardized questionnaires at WHI recruitment.[1] Covariate adjustment for diabetes and hypertension status was based on self-reported treated diabetes and treated hypertension. Hemoglobin concentration (in grams per deciliter) at the baseline exam was measured using an automated electronic cell counter. White blood cell count was also measured at baseline using an automated electronic cell counter. Serum creatinine was measured at baseline using an enzymatic method that was traceable to an isotope dilution mass spectrometry reference creatinine standard. Glomerular filtration rate was estimated (eGFR) from serum creatinine using the CKD Epidemiology Collaboration (CKD-EPI) equation[4]. For ascertainment of incident end-stage kidney disease (ESKD) during follow-up, WHI participants were linked to the US Renal Data System (USRDS)[5]. ESKD events as of 2010 are included in these analyses.

Description of the Jackson Heart Study (JHS) and ARIC study validation data sets and data collection

The Jackson Heart Study (JHS) is a community-based cohort study begun in 2000 of 5,306 self-identified African American individuals from the Jackson, Mississippi, metropolitan area[6]. At the JHS baseline exam (2000-2004), EDTA plasma from 2,054 participants were assayed using the aptamer-based proteomic platform (SomaScan 1.3k) as previously described[7].

The Atherosclerosis Risk in Communities (ARIC) study is a prospective community-based cohort of 15,792 individuals (4,314 self-identified African American and 11,478 White participants) who were recruited and enrolled between 1987 and 1989 from four US communities (Forsyth County, NC; Jackson, MS; Minneapolis suburbs, MN; Washington County, MD).[8,9] In ARIC, nearly 5,000 unique plasma proteins were quantified at exam 2 (1990-1992) in n=2,463 African American participants using the SomaScan® platform v4[10]. We excluded from the ARIC analysis 415 overlapping African American participants from the Jackson MS field center who were included in the JHS analysis, leaving n=2,048 ARIC participants for the replication analysis.

Genotyping for the rs334 sickle cell variant was performed in JHS and ARIC using whole genome sequence data[11] and target genotyping[12], respectively. Institutional review boards approved of each study and written informed consent was obtained from study participants.

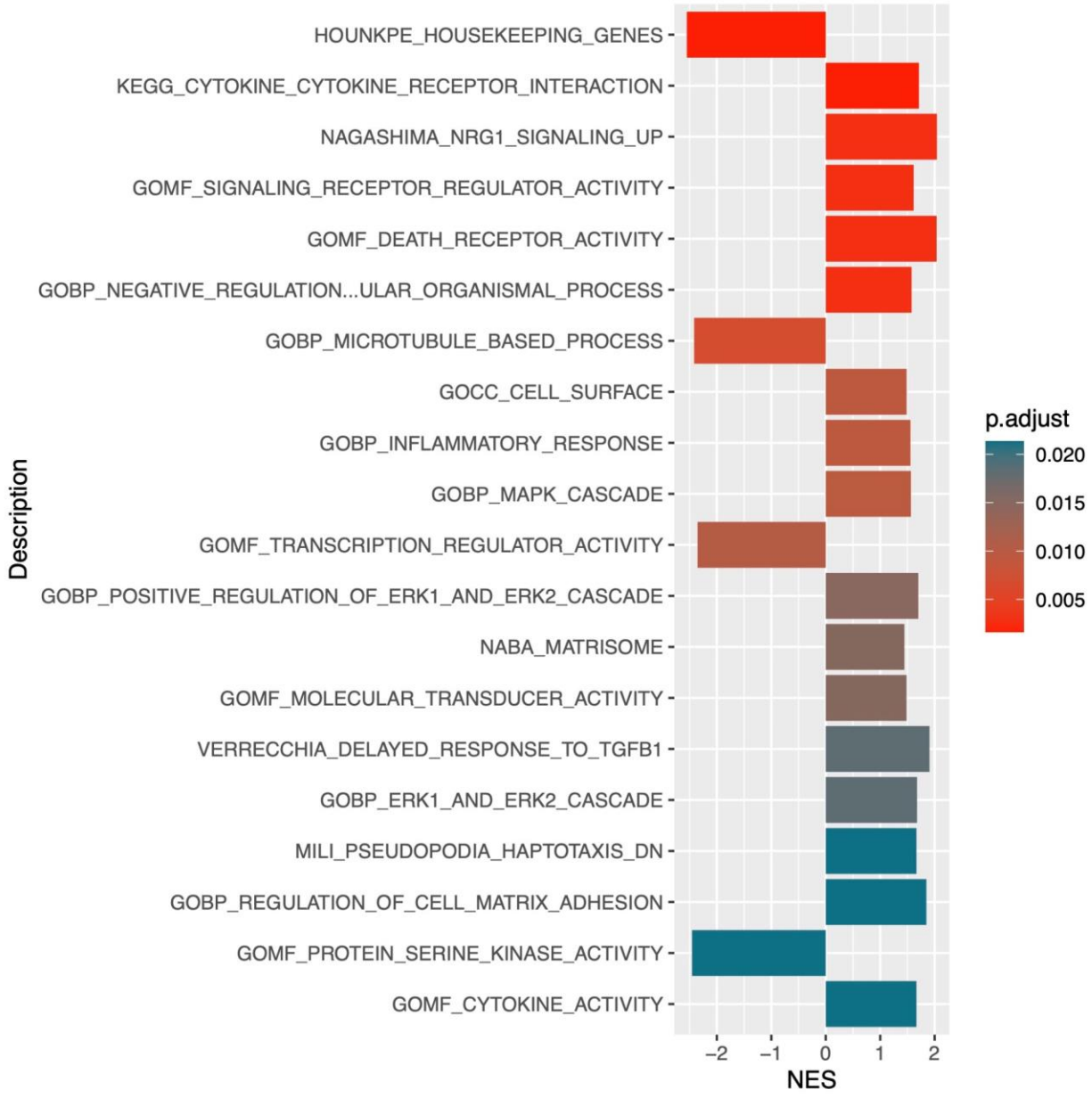
Estimation of genetic ancestry in WHI

For estimation of % African ancestry in WHI, for most participants, genotyped on the MEGA array, we used previous estimates of global ancestry from ADMIXTURE. Global ancestry proportions were estimated for 43,753 unrelated individuals (African American, N=6,280; Asian, N=4,193, Hispanic/Latino, N=19,400; Native American, N=630; Native Hawaiian, N= 2,267; and 983 individuals who identified themselves as “Other”) as part of the PAGE Study using ADMIXTURE v1.3[13] assuming five clusters (k=5). In short, genotype data (cleaned as previously described in[14]), was subset to variants with minor allele frequency (MAF)>1% and then pruned for independent sites (r^2<0.20) using Plink v1.90. Global ancestry proportions were estimated for unrelated individuals using ADMIXTURE v1.3 assuming five clusters (k=5). These five clusters represent African, European, Indigenous to the Americas, East Asian, and Pacific Islander ancestry. A total of ten runs were completed with 99.99% similarity as determined by pong[15]. WHI African American participants were then subset for inclusion in models detailed below. For participants genotyped on other arrays, we used PC1 to impute ADMIXTURE % African ancestry values, which was possible given the high correlation of ADMIXTURE results with PC1 in an admixed African American population (r^2>0.99). Participants without genome-wide genotyping data (with SCT status derived from targeted genotyping) were not included in the sensitivity analysis adjusting for % African ancestry.

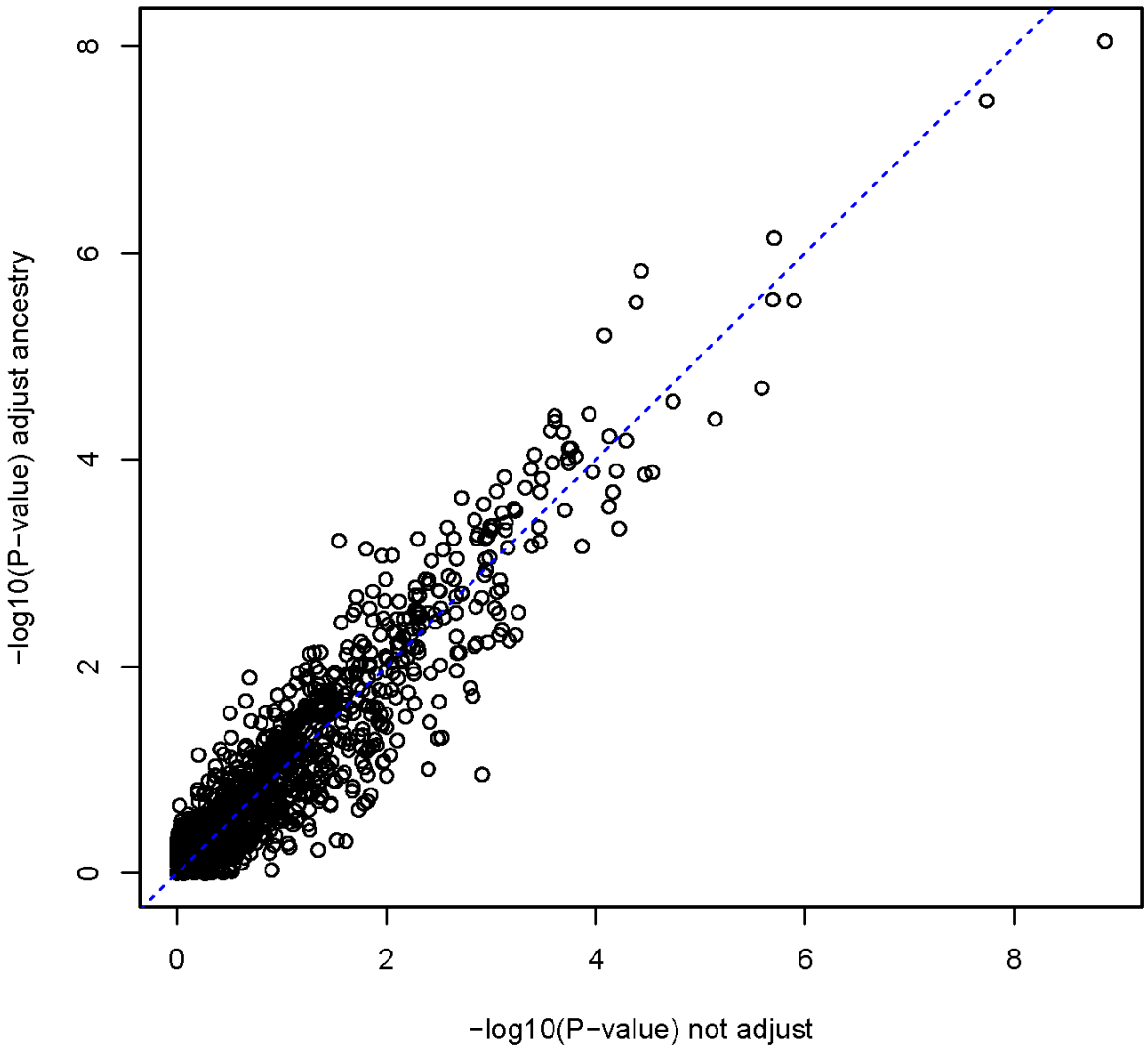
Sensitivity analyses in WHI

To assess whether the association between a particular protein and SCT was influenced by genetic ancestry or other known SCT-related phenotypes, in sensitivity analyses we included as covariates either % African genetic ancestry, baseline hemoglobin concentration, or baseline eGFR. Due to missing covariate data (15 missing hemoglobin, 201 missing eGFR, and 254 missing genetic ancestry from array-based data), inclusion of these covariates led to variable reductions in sample size. Therefore, we present pre- and post- conditioning results for each covariate using identical sample sizes, to ensure sample exclusion due to missing covariate data does not lead to any of the observed changes in effect estimates and p-values before and after covariate adjustment.

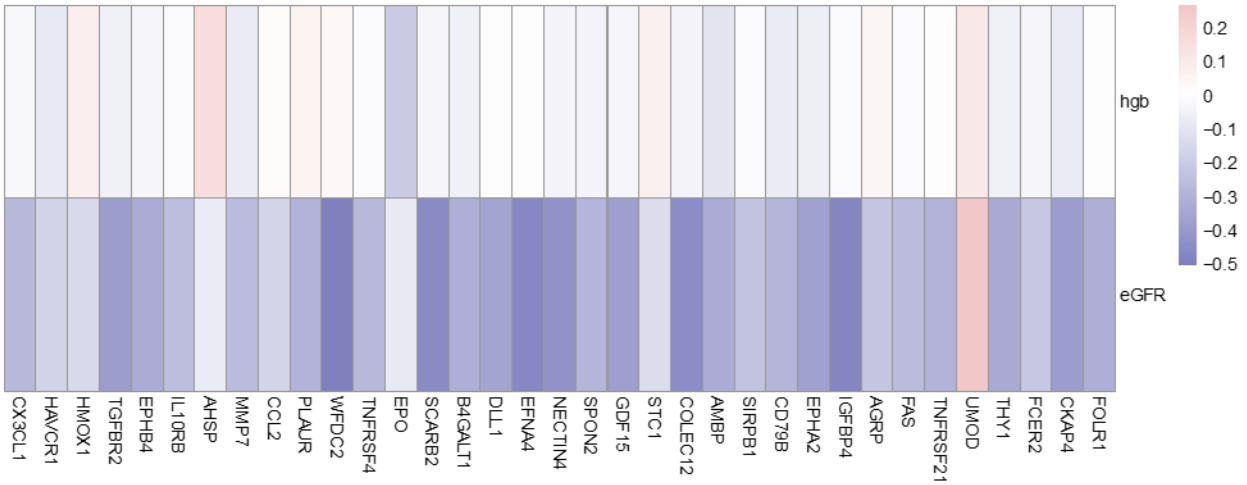
Supplemental Figure 1. Pathway enrichment analysis of SCT-associated proteins.



Supplemental Figure 2: SCT-protein association p-values with and without adjustment for genetic ancestry (i.e. PC1 or % African ancestry).



Supplemental Figure 3. Correlation between 35 SCT-associated proteins and baseline eGFR or hemoglobin in WHI.



References

1. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials*. 1998;19(1):61-109.

2. Hays J, Hunt JR, Hubbell FA, et al. The Women's Health Initiative recruitment methods and results. *Ann Epidemiol*. 2003;13(9 Suppl):S18-77.

3. Assarsson E, Lundberg M, Holmquist G, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One*. 2014;9(4):e95192.

4. Levey AS, Stevens LA. Estimating GFR using the CKD Epidemiology Collaboration (CKD-EPI) creatinine equation: more accurate GFR estimates, lower CKD prevalence estimates, and better risk predictions. *Am J Kidney Dis*. 2010;55(4):622-627.

5. USRDS: the United States Renal Data System. *Am J Kidney Dis*. 2003;42(6 Suppl 5):1-230.

6. Taylor HA Jr, Wilson JG, Jones DW, et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis*. 2005;15(4 Suppl 6):S6-4-17.

7. Katz DH, Robbins JM, Deng S, et al. Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. *Sci Adv*. 2022;8(33):eabm5164.

8. Wright JD, Folsom AR, Coresh J, et al. The ARIC (Atherosclerosis Risk In Communities) study: JACC focus seminar 3/8. *J Am Coll Cardiol*. 2021;77(23):2939-2959.

9. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989;129(4):687-702.

10. Chen TK, Surapaneni AL, Arking DE, et al. APOL1 Kidney Risk Variants and Proteomics. *Clin J Am Soc Nephrol*. 2022;17(5):684-692.

11. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590(7845):290-299.

12. Naik RP, Derebail VK, Grams ME, et al. Association of sickle cell trait with chronic kidney disease and albuminuria in African Americans. *JAMA*. 2014;312(20):2115-2125.

13. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655-1664.

14. Wojcik GL, Graff M, Nishimura KK, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*. 2019;570(7762):514-518.

15. Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. Pong: Fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics*. 2016;32(18):2817-2823.

16. Elmariah H, Garrett ME, De Castro LM, et al. Factors associated with survival in a contemporary adult sickle cell disease cohort. *Am J Hematol*. 2014;89(5):530-535.

17. Ashley-Koch AE, Okocha EC, Garrett ME, et al. MYH9 and APOL1 are both associated with sickle cell disease nephropathy. *Br J Haematol*. 2011;155(3):386-394.

18. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.