Smoking behavior impacts genetic erythroid trait variation via alcohol intake

Shriya Shivakumar[1,2], Madison B Wilken[1], Victor Tsao[1,3], Christopher S Thom[1,*]

[1]Division of Neonatology, Children's Hospital of Philadelphia, Philadelphia, PA
[2]Bryn Mawr College, Bryn Mawr, PA
[3]University of Pennsylvania, Philadelphia, PA

*Correspondence:
Christopher S Thom, MD, PhD
Children's Hospital of Philadelphia
Colket Translational Research Bldg, Room 10-052
3501 Civic Center Blvd
Philadelphia, PA 19104
thomc@chop.edu

Word count: 2080
Figs: 2
Tables: 0
Supplementary Figs: 9
Supplementary Tables: 1

Key Words: Genetics, Mendelian randomization, Blood, Erythroid

Abstract

Genome wide association studies (GWAS) have associated thousands of loci with quantitative human blood trait variation. Blood trait loci and related genes may regulate blood cell-intrinsic biological processes, or alternatively impact blood cell development and function via systemic factors and disease processes. Clinical observations linking behaviors like tobacco or alcohol use with altered blood traits can be subject to bias, and these trait relationships have not been systematically explored at the genetic level. Using a Mendelian randomization (MR) framework, we confirmed causal effects of smoking and drinking that were largely confined to the erythroid lineage. Using multivariable MR and causal mediation analyses, we confirmed that an increased genetic predisposition to smoke tobacco was associated with increased alcohol intake, indirectly decreasing red blood cell count and related erythroid traits. These findings demonstrate a novel role for genetically influenced behaviors in determining human blood traits, revealing opportunities to dissect related pathways and mechanisms that influence hematopoiesis.

Introduction

Tobacco and alcohol use are prevalent behaviors that have many detrimental effects on human health, with >85% of U.S. adults over the age of 18 reporting alcohol use and >12% (>30 million) U.S. adults actively smoking cigarettes (U.S. CDC, 2022; U.S. NIAAA, 2022). While associated cardiovascular disease risks are well documented, effects on many organ systems are not well understood. For example, clinical observational studies have linked tobacco use with altered blood cell traits (Nordenberg et al., 1990; Sunyer et al., 1996; Asgary et al., 2005; Malenica et al., 2017). These findings may reflect effects on blood cell development or function, with potential broad implications for human health. However, prior studies have reported conflicting blood trait effects. A complex interplay between physiologic mechanisms may link tobacco use with anemia (Leifert, 2008), although multiple studies have associated smoking with higher hemoglobin levels (Nordenberg et al., 1990; Malenica et al., 2017). Many of these conclusions derive from observational studies that are subject to bias from unmeasured or unrecognized confounding factors. We hypothesized that genetic methods might be used to clarify the directional effects of smoking and alcohol use behaviors on blood traits.

Prior genome wide association studies (GWAS) adjusted for tobacco and alcohol use behaviors given presumed multi-trait impacts (Astle at al 2016), but the effects of these behaviors on specific blood traits or developmental lineages are unclear. These associations are potentially relevant for understanding genetic determinants of disease phenotypes and other human traits. More recent GWAS were not adjusted for tobacco or alcohol use (Chen et al Cell 2020, Vuckovic et al Cell 2020), providing an opportunity to analyze genetically determined effects of these behaviors on blood traits. These studies had increased power compared with a prior study that evaluated behavioral effects on genetically determined blood traits (Pedersen et al., 2019).

Here, we used a Mendelian randomization (MR) framework to establish causal effects of tobacco and alcohol use on blood traits. MR leverages random single nucleotide polymorphism (SNP) allocation at meiosis to make causal inferences about the effects of a genetically determined exposure trait on an outcome phenotype (Burgess et al., 2020). Multivariable MR (MVMR) can extend this method to identify mediating or confounding traits to explain causal effects (Sanderson et al., 2019). In some cases, we identified behavioral effects ran counter to clinical observations and widely held clinical assumptions. Using MVMR and formal mediation strategies, we identified relevant trait relationships to explain unexpected erythroid trait relationships and determine correct causal directions. Our results clarify genetic effects of tobacco and alcohol use behaviors on quantitative blood traits.

Results

Increased genetically determined odds of smoking initiation decrease hemoglobin, hematocrit, and red blood cell count

We conducted two sample Mendelian randomization (MR) experiments to ascertain if genetically determined smoking initiation (SmkInit) risk impacted one or more of 15 quantitative blood traits (**Fig. 1A** and **Supplementary Fig. 1**). An instrumental variable comprising 113 linkage-independent SNPs significantly associated with SmkInit negatively impacted erythroid traits, including quantitative measures of hemoglobin (HGB), hematocrit (HCT), and red blood cell counts (RBC) by the inverse variance weighted (IVW) method ($p<0.05$ for each blood trait). These experiments revealed non-significant MR Egger intercept values, arguing against horizontal pleiotropy (all $p>0.05$). Sensitivity analyses using weighted median and MR Egger methods showed directionally consistent results (**Supplementary Fig. 2**), and reverse causality experiments also supported a directional effect from smoking on erythroid traits (**Supplementary Fig. 3**). The effects of SmkInit did not extend to quantitative platelet and white blood cell characteristics (**Fig. 1A**). The instrumental variables used for these experiments, and all other experiments in this manuscript, were not subject to weak instrument bias as defined by F statistics > 10 (**Supplementary Table 1**) (Burgess and Thompson, 2011).

Other genetically determined smoking traits, including smoking cessation (SmkCess), smoking heaviness (CigPerDay), and lifetime smoking score (LfSmk) did not significantly impact erythroid traits (**Supplementary Fig. 4**). SmkCess and CigPerDay were likely underpowered to detect any significant associations, although SmkCess (phenotypically opposed to SmkInit) trended toward slightly positive effects (**Supplementary Fig. 4**). However, LfSmk is a well powered phenotype (Wootton et al., 2019; Thom et al., 2020). Thus, it is possible that the slightly positive effects of LfSmk on erythroid traits reflect impacts from smoking cessation and/or heaviness that negate otherwise negative effects of smoking initiation genetics. The effects of LfSmk also agree with observational data (Nordenberg et al., 1990; Malenica et al., 2017), which would inherently reflect smoking duration and heaviness that are sometimes unmeasured among clinical cohorts.

Body mass index does not mediate the effects of SmkInit on erythroid traits

The negative effects of SmkInit on erythroid traits contrasted clinical observations, most of which would have anticipated positive effects of SmkInit on RBC, HGB, and HCT (e.g., (Nordenberg et al., 1990; Malenica et al., 2017)) We thus endeavored to identify possible confounding traits. For example, we previously identified body mass index (BMI) as a strong confounder for SmkInit (Thom et al., 2020) and a negative regulator of quantitative blood traits (Thom et al., 2022). However, multivariate MR (MVMR) experiments argued against BMI as a confounder for the effects of SmkInit on erythroid traits, with consistent effect sizes and significance estimates for SmkInit on erythroid traits after adjusting for the effects of BMI, with some loss of power in the instrumental variable (**Supplementary Fig. 5**).

Genetically determined alcohol consumption mediates the effects of SmkInit on erythroid traits

We next considered genetically influenced alcohol use as a potential mediating factor for the effects of SmkInit on erythroid traits, given clinical links between increased alcohol and tobacco use (Shiffman, 1996), as well as strong genetic associations between SmkInit and alcoholic drinks consumed per week (DrnkWk, r_g=0.34, p=6.7x10^{-63}, (Liu et al., 2019)).

Like SmkInit, an increase in genetically determined DrnkWk strongly negatively impacted HGB, HCT, and RBC across MR methods (**Fig. 1B**, **Supplementary Fig. 2** and **Supplementary Fig. 6**), without evidence of horizontal pleiotropy (all MR Egger intercept p>0.05). This was consistent with clinical observations related to the hematologic effects of alcoholism (Ballard, 1997).

We then confirmed using MR that an increased SmkInit risk portended an increase in DrnkWk across MR methods (**Fig. 2A**). Reverse causality and MR Steiger tests supported a unidirectional effect of increased SmkInit on increased DrnkWk (MR Steiger p-value 1.7x10^{-67}, sensitivity ratio 3.1, **Fig. 2A**).

In MVMR experiments, adjustment for DrnkWk abrogated the effects of SmkInit on erythroid traits (**Fig. 2B** and **Supplementary Fig. 7**). Formal mediation experiments confirmed that the effect of SmkInit on RBC was largely mediated indirectly through DrnkWk (**Fig. 2C**). The effects of LfSmk on erythroid traits were more positive after adjustment for DrnkWk, although did not reach statistical significance (**Supplementary Fig. 8**). However, multivariable results with LfSmk did reveal increased WBC, consistent with prior clinical and genetic studies (Sunyer et al., 1996; Pedersen et al., 2019). Thus, a genetic predisposition for increased alcohol consumption was an important confounding factor impacting the effects of tobacco use on blood traits.

Discussion

Systemic factors and disease states have been clinically observed to alter blood cell formation and function (Heyde et al., 2021; Rohde et al., 2021). With few exceptions, systemic and environmental impacts on blood traits are not well characterized at a genetic level. The MR framework is a good way to assess genetic trait interactions, given its independence from confounding effects that can bias observational studies (Burgess et al., 2020). Here, we used MR to reveal causal effects of genetically determined tobacco and alcohol use on blood traits, providing evidence for novel trait associations.

Two aspects of these results were surprising. First, the effects of smoking were confined to erythroid traits. This contrasted the effects of obesity and adipose distribution, which impacted blood traits across lineages (Thom et al., 2022). While we previously inferred those findings to relate to hematopoietic stem and progenitor cell biology, the effects of tobacco and alcohol use seem mostly related to terminal erythroid maturation and/or erythrocyte homeostasis. This may have important implications for related genetic mechanisms that underlie these trait associations, which require further exploration to delineate. However, we anticipate that erythroid-specific mechanisms may target erythrocyte stability, splenic clearance, or renal function.

Second, the negative effects of SmkInit on erythroid traits countered prevailing clinical intuition and prior observations linking tobacco use with higher erythroid indices (Nordenberg et al., 1990; Malenica et al., 2017), although we note that our findings agree with prior studies linking smoking habits with anemia (Leifert, 2008). Our MVMR experiments revealed that these unexpected smoking trait relationships resulted from the co-inheritance of tobacco and alcohol use predispositions. However, our findings may reflect different biological mechanisms or context than most clinical studies, which have generally profiled long-term smokers. The increased genetic predisposition to initiate smoking, as defined by GWAS (Liu et al., 2019), does not necessarily equate to long term tobacco use nor all anticipated adverse health effects.

Consistent with this notion, MR experiments using an instrumental variable for LfSmk showed no negative effects on erythroid traits, and trended toward a positive effect after adjustment for related factors. While these findings may have been complicated by underpowered SmkCess and CigPerDay effects, these findings may be more comparable to clinically observed association between tobacco use and higher erythroid indices.

Systemic and environmental factors can influence blood cell formation and blood cells impact myriad disease states. Understanding direct causal relationships between these factors is important. Given certain assumptions are met MR can determine genetic trait interactions free of the bias and confounders inherent to observational studies (Burgess et al., 2020). While future work is needed to further clarify exact loci and genes that link tobacco and alcohol use with blood trait variation, our results define novel causal effects of these behaviors on genetically determined erythroid traits.

Methods

GWAS collection

This study utilized publicly available GWAS data for tobacco and alcohol use (Liu et al., 2019), lifetime smoking score (n=462690) (Wootton et al., 2019), body mass index (N=484680) (Pulit et al., 2019), and blood traits (N=563085) (Vuckovic et al., 2020). All data were derived from individuals of European descent and used genome build hg19. Smoking initiation (SmkInit, N=1,232,091) was a binary phenotype derived from surveys querying whether an individual had ever smoked regularly for >1 month or having had smoking more than 100 cigarettes total in one's life. Data for smoking cessation (SmkCess, N=547219), cigarettes smoked per day (CigPerDay, N=337334), and alcoholic drinks consumed per week (DrnkWk, N=941280). were also collected by report (Liu et al., 2019). Lifetime smoking scores (LfSmk), calculated from smoking initiation, duration, and heaviness, were based on UK Biobank data (Wootton et al., 2019).

Genetic variant selection and instrumental variable creation

Instrumental variables were created by filtering summary GWAS statistics for SNPs common to exposure, outcoming, and/or mediating factor data sets. Using the TwoSampleMR package (v0.5.6, (Hemani et al., 2018)), SNPs meeting genome-wide significance for the exposure phenotype were clumped to select single SNPs in linkage disequilibrium (EUR r^2<0.01) in 250kb genomic regions. Instrumental variable strength was estimated using Cragg-Donald F-statistics, instruments calculated to have an F-statistic>10 deemed to have limited weak instrument bias (Burgess and Thompson, 2011).

Mendelian randomization and mediation analyses

We used R (v 4.2.2) to conduct univariable MR analyses (TwoSampleMR v0.5.6, (Hemani et al., 2018)). Certain assumptions must be true to permit valid conclusions from MR studies, including that independent genetic instruments (SNPs) must be associated with the exposure trait. Weak instruments, horizontal pleiotropy, heterogeneity, and phenotype error measurements can limit applicability or inferences (Burgess et al., 2020). This study attempts to follow all best practices and reporting related to our MR experiments.

Univariable experiments focused on results from inverse variance weighted (IVW), weighted median, and MR Egger regression methods. We looked for evidence of horizontal pleiotropy using MR Egger regression intercepts; significant deviation from zero can imply directional bias (Bowden et al., 2015). MVMR experiments used the MVMR package (Sanderson et al., 2019) and report IVW causal effect estimates. From our MR Steiger analyses, we included sensitivity, statistical significance estimates, and the reported inference for correct causal direction (Hemani et al., 2017). In mediation experiments, we report total and direct effects estimates for each exposure and mediating trait on given outcomes (Burgess et al., 2017)

Statistical analysis and data presentation

Presented effects include effect estimates from IVW, WM, and MR Egger regressions. Given heterogeneity in some instrumental variables, we utilized random effects models. Statistical significance was defined as $p<0.05$ for all experiments. Figures were prepared using GraphPad Prism 8.

Data and code availability

Coding scripts and data sets, including full instrumental variables used in our study, can be found on Github (https://github.com/thomchr/SmkBloodMR) and are also available upon request.

Acknowledgements

This work was supported by the National Institutes of Health (K99HL156052 to CST), University of Pennsylvania Undergraduate Translational Research Immersion Program (SS), and University of Pennsylvania Undergraduate Research Mentoring Program (VT).

References

Asgary, S., Naderi, G., and Ghannady, A. (2005). Effects of cigarette smoke, nicotine and cotinine on red blood cell hemolysis and their -SH capacity. *Exp Clin Cardiol* 10, 116–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19641671 [Accessed January 25, 2021].

Ballard, H. S. (1997). The Hematological Complications of Alcoholism. *Alcohol Health Res World* 21, 42–52. doi: 10.14412/2074-2711-2017-4-95-100.

Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 44, 512–25. doi: 10.1093/ije/dyv080.

Burgess, S., Davey Smith, G., Davies, N. M., Dudbridge, F., Gill, D., Glymour, M. M., et al. (2020). Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res* 4. doi: 10.12688/wellcomeopenres.15555.2.

Burgess, S., Thompson, D. J., Rees, J. M. B., Day, F. R., Perry, J. R., and Ong, K. K. (2017). Dissecting causal pathways using mendelian randomization with summarized genetic data: Application to age at menarche and risk of breast cancer. *Genetics* 207, 481–487. doi: 10.1534/genetics.117.300191.

Burgess, S., and Thompson, S. G. (2011). Avoiding bias from weak instruments in mendelian randomization studies. *Int J Epidemiol* 40, 755–764. doi: 10.1093/ije/dyr036.

Hemani, G., Tilling, K., and Davey Smith, G. (2017). Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* 13, e1007081. doi: 10.1371/journal.pgen.1007081.

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., et al. (2018). The MR-base platform supports systematic causal inference across the human phenome. *Elife* 7, e34408. doi: 10.7554/eLife.34408.

Heyde, A., Rohde, D., McAlpine, C. S., Zhang, S., Hoyer, F. F., Gerold, J. M., et al. (2021). Increased stem cell proliferation in atherosclerosis accelerates clonal hematopoiesis. *Cell* 184, 1348-1361.e22. doi: 10.1016/j.cell.2021.01.049.

Leifert, J. A. (2008). Anaemia and cigarette smoking. *Int J Lab Hematol* 30, 177–184. doi: 10.1111/j.1751-553X.2008.01067.x.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., et al. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 51, 237–244. doi: 10.1038/s41588-018-0307-5.

Malenica, M., Prnjavorac, B., Bego, T., Dujic, T., Semiz, S., Skrbo, S., et al. (2017). Effect of Cigarette Smoking on Haematological Parameters in Healthy Population. *Med Arch* 71, 132–136. doi: 10.5455/medarh.2017.71.132-136.

Nordenberg, D., Yip, R., and Binkin, N. J. (1990). The Effect of Cigarette Smoking on Hemoglobin Levels and Anemia Screening. *JAMA: The Journal of the American Medical Association* 264, 1556–1559. doi: 10.1001/jama.1990.03450120068031.

Pedersen, K. M., Çolak, Y., Ellervik, C., Hasselbalch, H. C., Bojesen, S. E., and Nordestgaard, B. G. (2019). Smoking and Increased White and Red Blood Cells: A Mendelian Randomization Approach in the Copenhagen General Population Study. *Arterioscler Thromb Vasc Biol* 39, 965–977. doi: 10.1161/ATVBAHA.118.312338.

Pulit, S. L., Stoneman, C., Morris, A. P., Wood, A. R., Glastonbury, C. A., Tyrrell, J., et al. (2019). Meta-Analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* 28, 166–174. doi: 10.1093/hmg/ddy327.

Rohde, D., Vandoorne, K., Lee, I.-H., Grune, J., Zhang, S., McAlpine, C. S., et al. (2021). Bone marrow endothelial dysfunction promotes myeloid cell expansion in cardiovascular disease. *Nature Cardiovascular Research*. doi: 10.1038/s44161-021-00002-8.

Sanderson, E., Davey Smith, G., Bowden, J., and Munafò, M. R. (2019). Mendelian randomisation analysis of the effect of educational attainment and cognitive ability on smoking behaviour. *Nat Commun* 10. doi: 10.1038/s41467-019-10679-y.

Shiffman, S. (1996). Alcoholism and Co-Occurring Disorders. *Alcohol Health Res World* 20.

Sunyer, J., Muñoz, A., Peng, Y., Margolick, J., Chmiel, J. S., Oishi, J., et al. (1996). Longitudinal relation between smoking and white blood cells. *Am J Epidemiol* 144, 734–741. doi: 10.1093/oxfordjournals.aje.a008997.

Thom, C. S., Ding, Z., Levin, M. G., Damrauer, S. M., Lee, K. M., Lynch, J., et al. (2020). Genetic determinants of increased body mass index mediate the effect of smoking on increased risk for type 2 diabetes but not coronary artery disease. *Hum Mol Genet* 29, 3327–3337. doi: 10.1093/hmg/ddaa193.

Thom, C., Wilken, M., Chou, S., and Voight, B. (2022). Body mass index and adipose distribution have opposing genetic impacts on human blood traits. *Elife* 11, e75317. doi: 10.7554/ELIFE.75317.

U.S. CDC (2022). United States Center for Disease Control. *https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm*.

U.S. NIAAA (2022). United States National Institute on Alcohol Abuse and Alcoholism. *https://www.niaaa.nih.gov/sites/default/files/publications/NIAAA_Alcohol_FactsandStats_102020_0.pdf*.

Vuckovic, D., Bao, E. L., Akbari, P., Lareau, C. A., Mousas, A., Jiang, T., et al. (2020). The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* 182, 1214-1231.e11. doi: 10.1016/j.cell.2020.08.008.

Wootton, R. E., Richmond, R. C., Stuijfzand, B. G., Lawn, R. B., Sallis, H. M., Taylor, G. M. J., et al. (2019). Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med*, 1–9. doi: 10.1017/s0033291719002678.

Figures

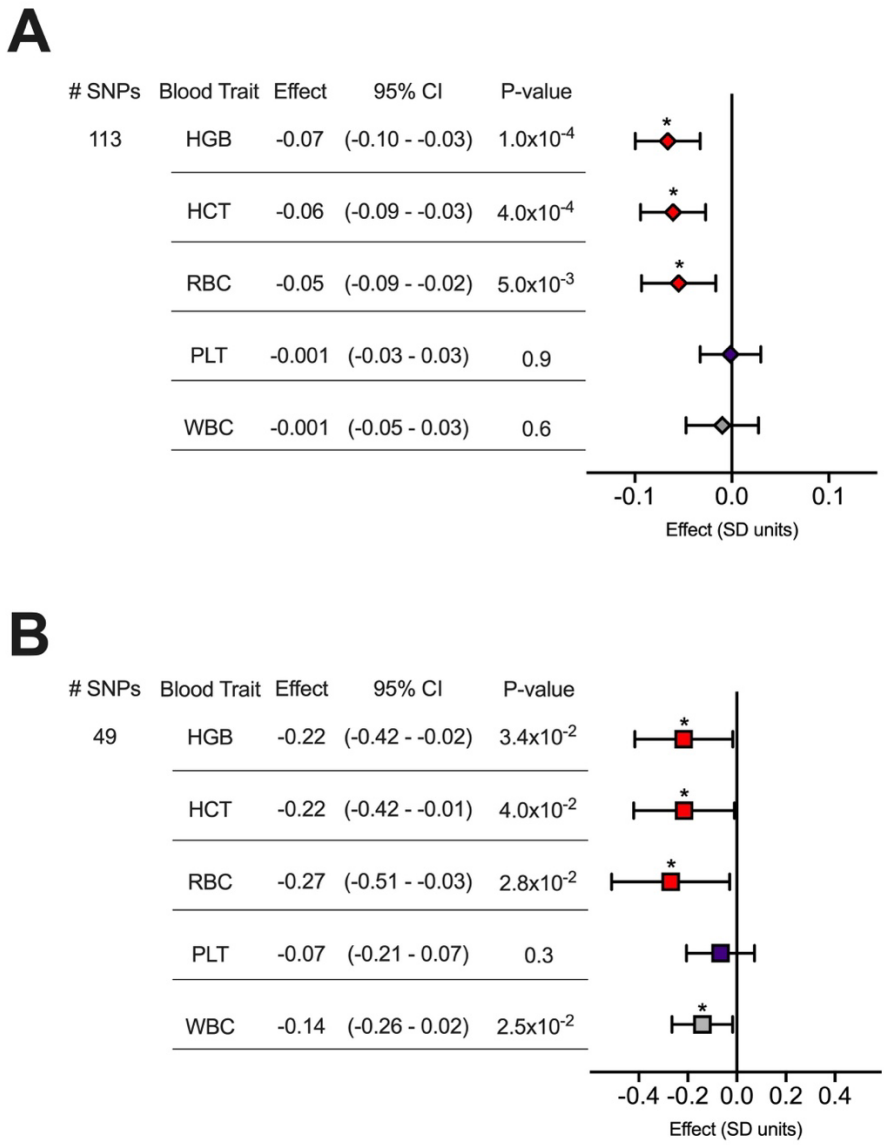


Figure 1. Smoking initiation (SmkInit) and alcohol use (Drinks) negatively impact genetically determined hemoglobin (HGB), hematocrit (HCT), and red blood cell count (RBC).

(**A-B**) MR results using the inverse variance weighted method are shown, with values for Effect sizes on the indicated blood traits reflecting (**A**) a 2-fold increase in SmkInit risk or (**B**) a 1 SD unit increase in alcoholic drinks per week. *p<0.05.

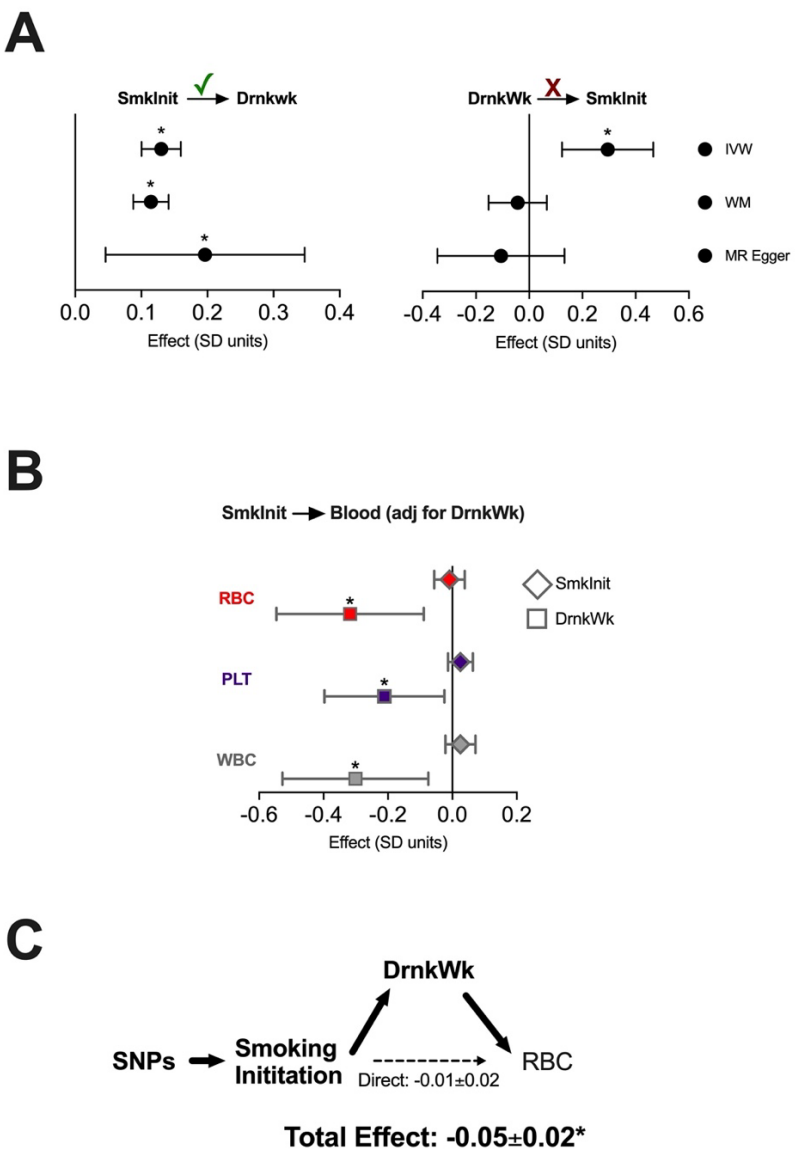
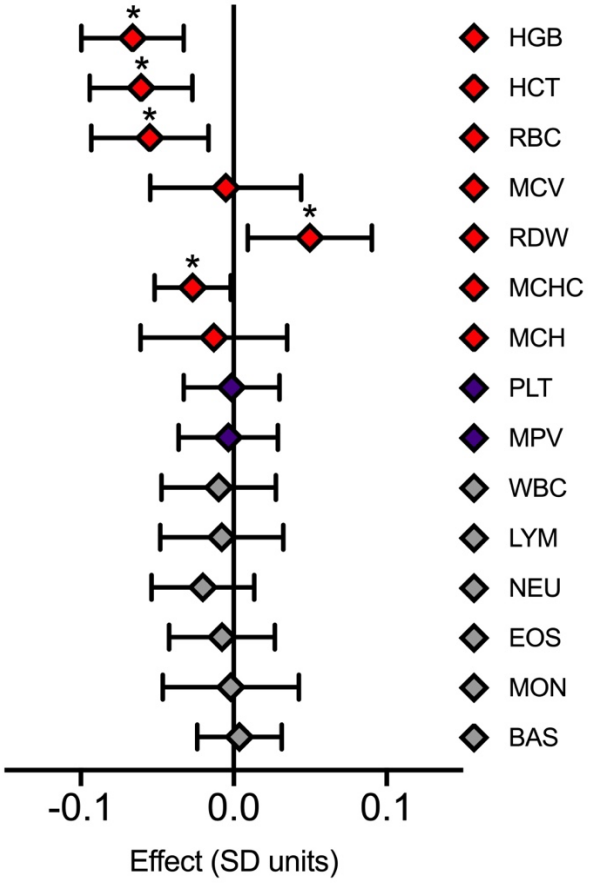
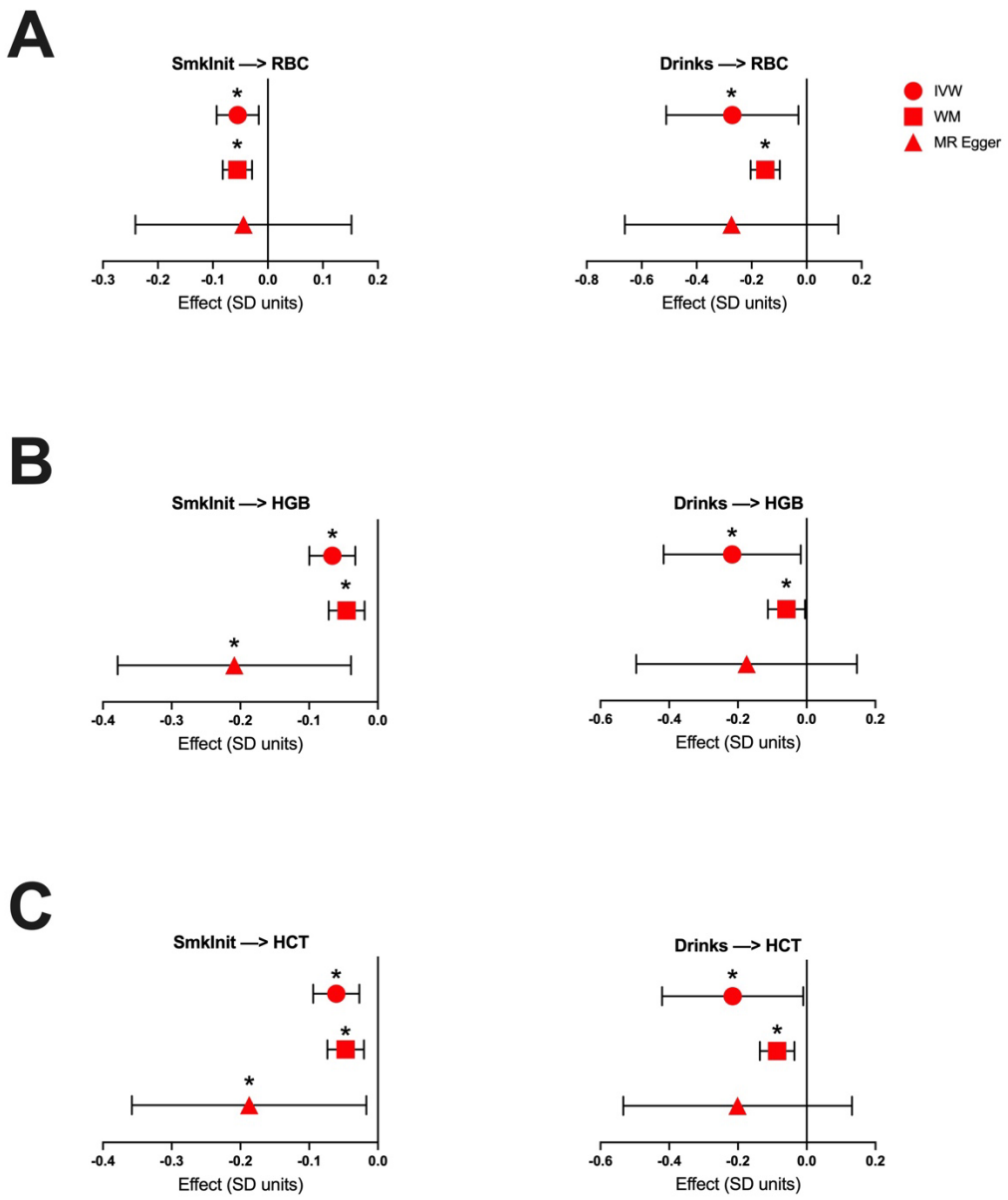


Figure 2. Alcohol use mediates the effects of genetically determined smoking behavior on erythroid traits.
(**A**) By MR, increased risk of SmkInit increases Drinks per week (Drinks) by inverse variance weighted (IVW), weighted media (WM), and MR Egger methods. However, increased genetically determined Drinks per week does not consistently increase SmkInit risk across MR methods. Check mark reflects MR Steiger 'correct causal estimate' relationship.
(**B**) By MVMR, using an instrumental variable for SmkInit adjusted for Drinks per week, the effects of SmkInit on quantitative blood traits were nullified. Effects of Drinks per week on blood traits remain significantly negative.
(**C**) Mediation analysis showed that SmkInit made a statistically insignificant effect on RBC, whereas the total effect (including indirect effects through Drinks per week) significantly negatively impacts RBC. *p<0.05.

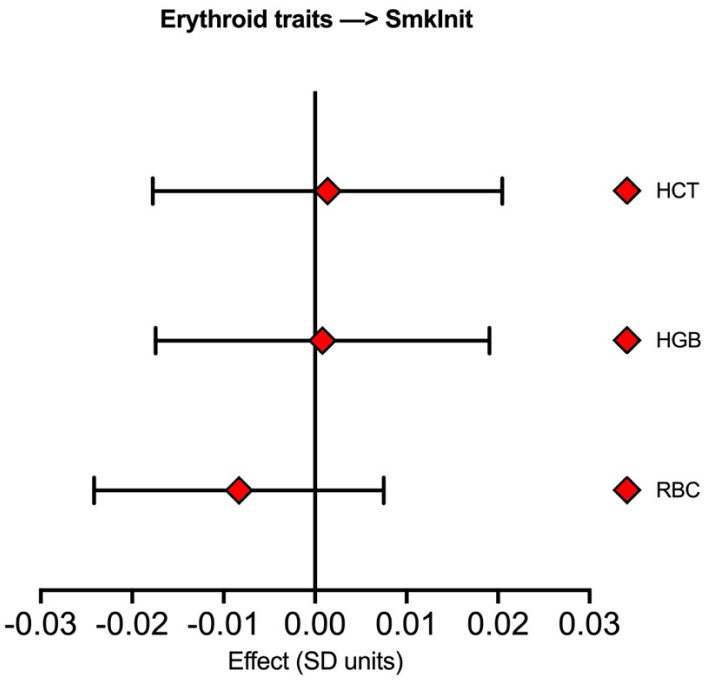
Supplementary Figures



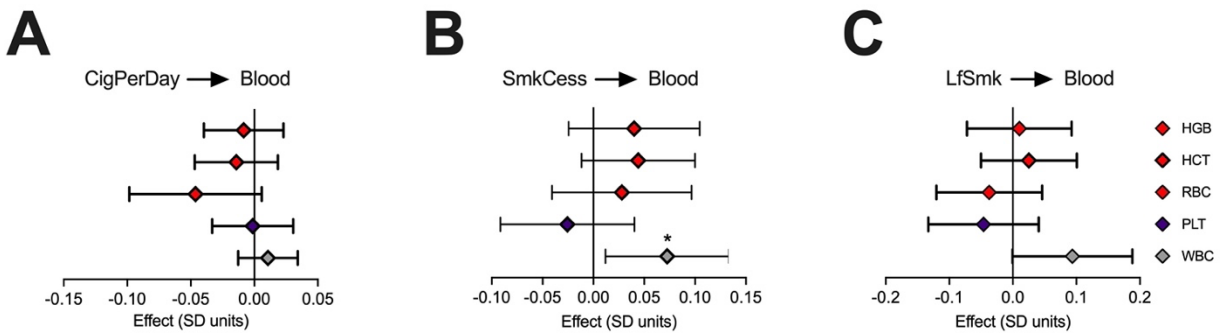
Supplementary Figure 1. Effects of a 2-fold increase in SmkInit risk on the indicated blood traits by univariable MR. *p<0.05.



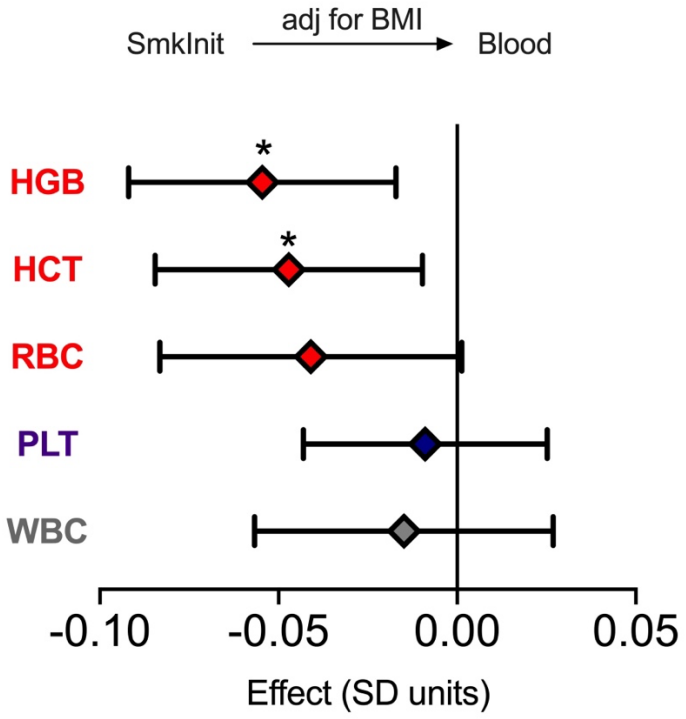
Supplementary Figure 2. Sensitivity analyses for trait association in univariable MR experiments.
(**A-C**) Effects of a 2-fold increase in SmkInit risk or a 1 SD unit increase in alcoholic drinks per week on (**A**) RBC, (**B**) HGB, or (**C**) HCT by weighted median (WM) and MR Egger regression analyses. IVW estimates also found in Figure 1 are presented here for comparison.



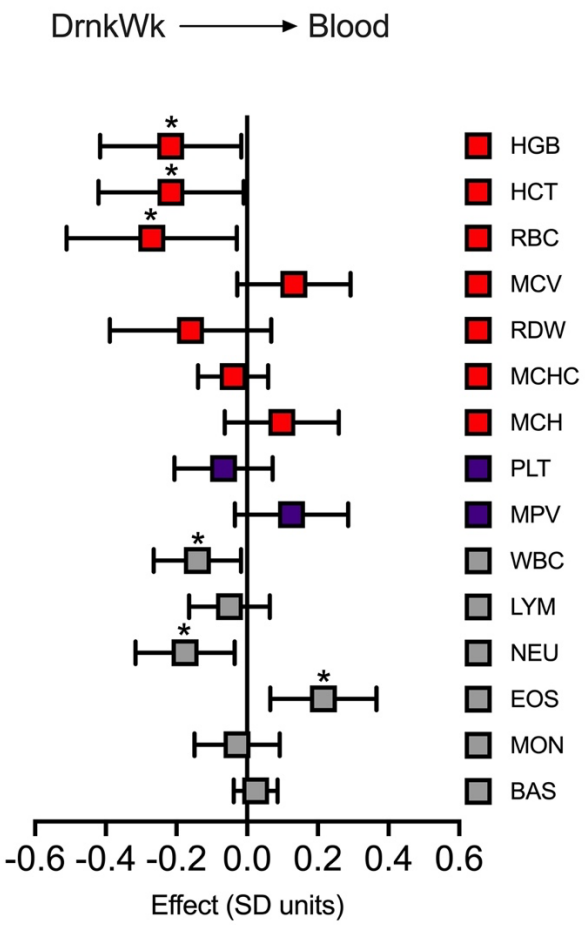
Supplementary Figure 3. Univariable MR effect estimates for a 1 SD unit increase in the indicated erythroid traits on SmkInit. None of the estimates reached statistical significance.



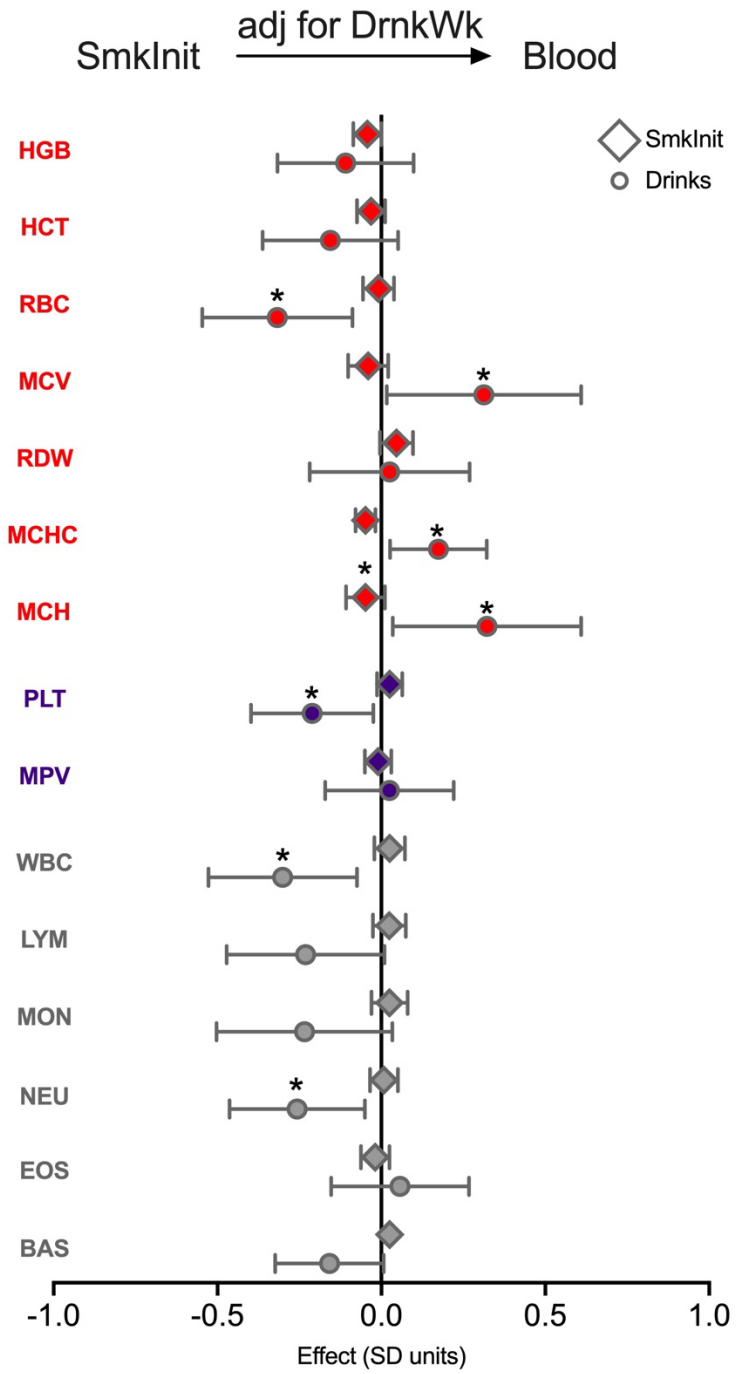
Supplementary Figure 4. Univariable MR effect estimates for a 1 SD unit increase in smoking traits on the indicated blood traits. *p<0.05.



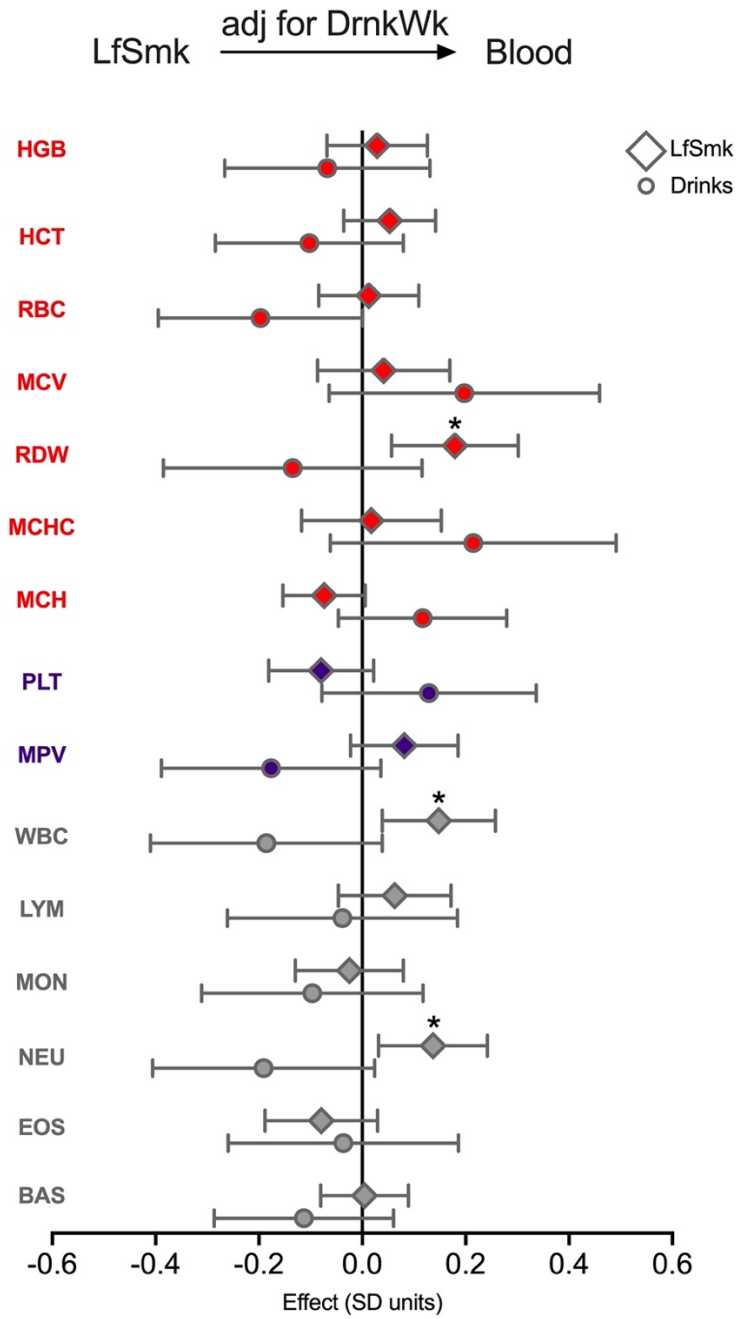
Supplementary Figure 5. MVMR experiments analyzing the effect of SmkInit on the indicated blood traits after adjusting for body mass index (BMI). *p<0.05, **p<0.003.



Supplementary Figure 6. Univariable MR effect estimates for a 1 SD unit increase in DrnkWk on the indicated blood traits. *p<0.05.



Supplementary Figure 7. MVMR effect estimates for SmkInit or DrnkWk on the indicated blood traits. All experiments used an instrumental variable for SmkInit adjusted for DrnkWk. After adjustment, SmkInit did not have significant effects on any blood trait whereas DrnkWk did retain some significant effects. *p<0.05.



Supplementary Figure 8. MVMR effect estimates for LfSmk or DrnkWk on the indicated blood traits. All experiments used an instrumental variable for LfSmk adjusted for DrnkWk. After adjustment, LfSmk did not have significant effects on any blood trait whereas DrnkWk did retain some significant effects. *p<0.05.

Supplementary Tables

Supplementary Table 1. Craig-Donald F statistics for instrumental variables used in this study.