2023

INCORPORATING A DEVELOPMENTAL PERSPECTIVE INTO GENE IDENTIFICATION MODELS FOR ALCOHOL USE BEHAVIORS

Nathaniel S. Thomas
Virginia Commonwealth University

INCORPORATING A DEVELOPMENTAL PERSPECTIVE INTO GENE IDENTIFICATION MODELS FOR ALCOHOL USE BEHAVIORS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

By

NATHANIEL S. THOMAS
B.S., Virginia Commonwealth University, August 2014
M. S., Virginia Commonwealth University, December 2016
M. S., Virginia Commonwealth University, May 2020

Co-Directors: Jessica Salvatore, Ph.D.
Associate Professor
Department of Psychiatry
Robert Wood Johnson Medical School
Rutgers University

David Chester, Ph.D.
Associate Professor
Department of Psychology
Virginia Commonwealth University

Virginia Commonwealth University
Richmond, Virginia
April, 2023

Acknowledgements

COGA

The Collaborative Study on the Genetics of Alcoholism (COGA), Principal Investigators B. Porjesz, V. Hesselbrock, T. Foroud; Scientific Director, A. Agrawal; Translational Director, D. Dick, includes ten different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, T. Foroud, Y. Liu, M.H. Plawecki); University of Iowa Carver College of Medicine (S. Kuperman, J. Kramer); SUNY Downstate Health Sciences University (B. Porjesz, J. Meyers, C. Kamarajan, A. Pandey); Washington University in St. Louis (L. Bierut, J. Rice, K. Bucholz, A. Agrawal); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield, D. Dick, R. Hart, J. Salvatore); The Children's Hospital of Philadelphia, University of Pennsylvania (L. Almasy); Icahn School of Medicine at Mount Sinai (A. Goate, P. Slesinger); and Howard University (D. Scott). Other COGA collaborators include: L. Bauer (University of Connecticut); J. Nurnberger Jr., L. Wetherill, X., Xuei, D. Lai, S. O'Connor, (Indiana University); G. Chan (University of Iowa; University of Connecticut); D.B. Chorlian, J. Zhang, P. Barr, S. Kinreich, G. Pandey (SUNY Downstate); N. Mullins (Icahn School of Medicine at Mount Sinai); A. Anokhin, S. Hartz, E. Johnson, V. McCutcheon, S. Saccone (Washington University); J. Moore, F. Aliev, Z. Pang, S. Kuo (Rutgers University); A. Merikangas (The Children's Hospital of Philadelphia and University of Pennsylvania); H. Chin and A. Parsian are the NIAAA Staff Collaborators. We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting- Kai Li, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions. This national collaborative

study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA).

ALSPAC

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and will serve as guarantors for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website. This research was specifically funded by NIH AA018333, MRC G0800612/86812, Wellcome Trust and MRC (Core) 076467/Z/05/, NIH 5R01AA018333-05, Wellcome Trust and MRC 092731. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

Add Health

Add Health is directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Waves I-V data are from the Add Health Program Project, grant P01 HD31921 (Harris) from *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

There's a long list of people who have had an impact on my professional development as a researcher. Amy Adkins introduced me to genetics research as an undergraduate and mentored me during some of my earliest practical experiences in an academic setting. Around the same time, Jessica Salvatore introduced me to the R programming language and mentored my first experience of working on a manuscript for publication. Danielle Dick provided the initial opportunities that allowed me to begin developing expertise with R, working as data manager for the Spit for Science project. These were my earliest experiences with research. Many others provided mentorship and/or support over the years, too many for an exhaustive list, but thanks in particular to Nathan Gillespie, Peter Barr, Fazil Aliev, Albert Ksinan, Karen Chartier, Sally Kuo, Linda Zyzniewski, Mallory Stephenson, Morgan Driver, Rebecca Smith and Zoe Neale.

Regarding this project, thanks to my advisors (Jessica Salvatore and Danielle Dick) my dissertation committee (Jessica Salvatore, Dave Chester, Danielle Dick, Nathan Gillespie, Peter Barr, Wendy Kliewer, Mary Loos), the Examining Genes, Development and Environment (EDGE) lab as whole, the Department of Psychology at Virginia Commonwealth University, and the National Institute on Alcohol Abuse and Alcoholism (NIAAA). High Performance Computing resources provided by the High Performance Research Computing (HPRC) core facility at Virginia Commonwealth University (https://hprc.vcu.edu) and HPC resources hosted by the Virginia Institute for Psychiatric and Behavioral Genetics were used for conducting the research reported in this work. Thanks to Carlisle Childress, Mike Davis, and Helen Wang for their help with accessing HPC resources throughout the course of the project. This work was supported by F31AA029620.

Lastly, thanks to my partner, Brianna, and my family for their daily support and encouragement.

Dedicated to my dad, Jerre S. Thomas II.

Table of Contents

List of Tables

List of Figures

Abstract

INCORPORATING A DEVELOPMENTAL PERSPECTIVE INTO GENE IDENTIFICATION MODELS FOR ALCOHOL USE BEHAVIORS

Nathaniel S. Thomas, M.S.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2023

Co-Directors: Jessica Salvatore, Ph.D.
Associate Professor
Department of Psychiatry
Robert Wood Johnson Medical School
Rutgers University

David Chester, Ph.D.
Associate Professor
Department of Psychology
Virginia Commonwealth University

Frequent alcohol use can lead to alcohol use disorder, which accounts for three million deaths and over 133 million life years lost to disability and death worldwide per year. Alcohol use behaviors unfold across development, beginning with initiation of drinking and progressing through various escalating stages of use. Alcohol use behaviors are also under genetic influence. Genome-wide association represents the state-of-the-science statistical methodology for identifying genes associated with alcohol use outcomes. However, contemporary genome-wide association study (GWAS) methods typically do not account for variability in genetic effects throughout development. In this project, I applied novel multivariate genomic methods to combine developmentally-informative phenotype data and GWAS to create polygenic scores (PGS) that are specific to developmental stage. Longitudinal cohort studies targeted for gene-identification analyses include the Collaborative Study on the Genetics of Alcoholism

(adolescence n=1,118, early adulthood n=2,762, adulthood n=5,255), the National Longitudinal Study of Adolescent to Adult Health (adolescence n=3,089, early adulthood n=3,993, adulthood n=5,149), and the Avon Longitudinal Study of Parents and Children (ALSPAC; adolescence n=5,382, early adulthood n=3,613). PGS validation analyses were conducted in the COGA sample using an alternate version of the discovery analysis with COGA removed. Results suggest that genetic liability for alcohol use frequency in adolescence may be distinct from genetic liability for alcohol use frequency later in developmental periods. Additionally, a developmentally-informative approach to polygenic score construction yielded nominal, but not statistically significant, improvements in phenotype prediction in adulthood. The current work was underpowered at all steps of the analysis plan. Small sample sizes and low statistical power limit the substantive conclusions that can be drawn regarding these research questions, replication in well-powered samples is warranted.

Statement of Purpose

Studies that aim to identify genes associated with alcohol use outcomes in longitudinal datasets often average across timepoints or examine lifetime measures, constructing developmentally agnostic phenotypes that disregard potential developmental variability in genetic effects. Predictive models of genetic influences on alcohol use may be improved if they measure genetic effects that are unique to different developmental periods and, subsequently, leverage these unique effects for the prediction of alcohol use outcomes throughout the lifespan. The goal in the proposed research is to implement novel methods to conduct a developmentally-informative gene-identification study of alcohol use and use the resulting genetic effects to predict alcohol use throughout the lifespan. This project provides an analytic approach for developmentally-informed genetic prediction of alcohol use outcomes. These novel methods will advance the field of genetics beyond the study of phenotypes that aggregate across developmental periods and represents an important step towards the broader goal of advancing precision medicine strategies for alcohol use outcomes.

Introduction

Genetic Studies of Alcohol Use: Linkage, Candidate Genes, and GWAS

Frequent alcohol use can lead to alcohol use disorder (AUD), which accounts for three million deaths and over 133 million life years lost to disability and death worldwide per year (WHO, 2018). Modest levels of alcohol consumption are also associated with health consequences, especially in younger age demographics (Bryazka et al., 2022). Alcohol use is under significant genetic influence, with genetic factors accounting for approximately 50% of the reason that some people are at elevated risk for heavier alcohol consumption (Verhulst et al., 2015). Accordingly, many studies have been conducted to identify the underlying genetic variants that contribute to the propensity for alcohol use. Early gene identification efforts for alcohol use outcomes relied primarily on (1) linkage analysis and (2) hypothesis-driven candidate gene studies.

Linkage studies use an agnostic approach to identify regions of the genome that are shared by family members affected by the phenotype of interest, while hypothesis-driven candidate gene studies focus on genes that are thought to have a direct biological influence on the phenotype (Dick et al., 2015). Linkage studies were effective with phenotypes that follow a Mendelian inheritance pattern and are influenced directly by a small number of variants of large effect; however, this family-based approach proved to be underpowered when applied to alcohol use phenotypes that are influenced by many variants of small effect (Hart & Kranzler, 2015). The molecular genetic architecture that underlies alcohol use is massively polygenic, involving the contribution of many variants of small effect (Wray et al., 2018).

Hypothesis-driven candidate gene studies do not model the polygenic nature of complex phenotypes and also rely on the researchers' intuition to construct a presumptive, a priori model

of the biological underpinnings of the phenotype (Zhu & Zhao, 2007). Most candidate gene studies have failed to replicate (Border et al., 2019; Duncan & Keller, 2011; Eaves, 2006; Munafò, 2006; Sullivan, 2007, 2017), suggesting that this presumptive understanding of the biological foundation of complex psychiatric and behavioral phenotypes was often incorrect. Genes involved in the metabolism of alcohol are a rare exception to this general rule for alcohol use phenotypes. Genes encoding the enzymes alcohol dehydrogenase (ALDH) and aldehyde dehydrogenase (ADH) were identified as targets for their functional role in alcohol metabolism in early candidate gene studies (C. C. Chen et al., 1999; Edenberg, 2007; Thomasson et al., 1991) and were replicated in later genome-wide association studies (Hart & Kranzler, 2015).

While linkage and candidate gene methodologies failed to account for the polygenicity of complex phenotypes, some of the shortcomings of linkage and hypothesis-driven candidate gene methodologies are resolved in genome-wide association studies (GWAS), which adopt an agnostic approach to examine the entire genome for association with a phenotype in large, population-based samples. Visscher et al. (2017) provide a review of the scientific rationale and methodological approach of GWAS. The primary aim of GWAS is to detect association between genetic variants, in the form of single nucleotide polymorphisms (SNPs), and a trait or phenotype. GWAS leverage the correlation structure between SNPs, or linkage disequilibrium (LD), to detect associations between SNPs and complex phenotypes. Patterns of LD throughout the genome form haplotypes, or regions of the genome which tend to be inherited together. LD allows for a subset of genotyped variants to be used to recover a more complete set of variants within a haplotype via imputation to a whole-genome sequenced reference panel. Thus, GWAS can be conducted on large samples using relatively inexpensive technology by only genotyping a subset of variants which are then used to tag and impute larger portions of the genome. GWAS

rely on LD to detect association between SNPs and complex phenotypes, as an individual SNP of small effect may tag a larger number of unmeasured causal variants within a haplotype.

GWAS findings have proven to be relatively reliable when sample sizes are large (Visscher et al., 2017). Many SNPs identified in GWAS have replicated across multiple studies (Horwitz et al., 2019) and demonstrated a capacity to predict phenotypes in independent samples via polygenic scoring (Duncan et al., 2019). The success of these methods extends to alcohol use phenotypes as well, where contemporary polygenic scores (PGS) account for 0.5% to 3.5% of phenotypic variance, depending on discovery sample size, the particular alcohol use phenotype of interest, and the developmental period of the validation sample (Barr et al., 2019, 2020; Kandaswamy et al., 2021). The findings from these GWAS also provide direct evidence against the underlying assumptions of hypothesis-driven candidate gene studies, broadly confirming that complex phenotypes are highly polygenic and that relevant genetic variants often lie outside of the regions of the genome that were initially marked as intuitive targets for candidate gene methodologies in most cases (Duncan et al., 2019).

While GWAS present considerable benefits over linkage and hypothesis-driven candidate gene approaches, extremely large sample sizes are needed for GWAS because of the polygenicity of complex traits and the low effect size of each individual SNP. Meta-analysis methods, where the results of multiple GWAS are combined to maximize sample size, have gained popularity in response to recognition of the need for extremely large sample sizes (Evangelou & Ioannidis, 2013). Unfortunately, the pursuit of larger samples and greater statistical power through meta-analysis has indirectly led to a general disregard of developmental considerations in GWAS, with the construction of phenotypes that pool across time and developmental periods to maximize the number of samples that can be included in a single meta-

analysis. This approach does not model the ways that genetic influences on complex psychiatric and behavioral phenotypes vary throughout development. Recent examples of this approach to phenotype construction typically focus on alcohol consumption in the past year (Kranzler et al., 2019; Liu et al., 2019; Sanchez-Roige et al., 2019). These studies are developmentally-agnostic and do not account for the fact that genetic influences on alcohol use vary in specificity and magnitude across development. One recent study breaks with this general trend in the field, modeling stability and change in molecular genetic effects across age groups (Gillespie et al., 2022).

Developmental Considerations in the Genetics of Alcohol Use

Parallel to contemporary progress in molecular gene-identification studies, a robust body of research using biometrical genetical methods that rely on genetically-informative twins demonstrates that genetic influences on alcohol use vary across development. In twin studies, estimates of the proportion of population phenotypic variance attributable to genetic factors, or heritability, are inferred from the degree of phenotypic concordance within twin pairs (Jinks and Fulker, 1970; Mather and Jinks, 1982). Two noteworthy patterns emerge in this literature. First, heritability changes across development; for instance, previous studies demonstrate that the heritability of alcohol use increases from adolescence to emerging adulthood (Huibregtse et al., 2016; Kendler et al., 2008; Rose et al., 2001; Viken et al., 1999). This increase in heritability over time is partly driven by "genetic innovation", which refers to new genetic risk factors that emerge throughout development (Edwards & Kendler, 2013). Second, different genetic variants influence alcohol use at different developmental periods. Twin studies indicate that alcohol-specific and externalizing genetic factors operate at different developmental stages (Kendler et al., 2011; Meyers et al., 2014).

The first example of age-related differences in genetic impact is seen in earlier molecular genetic studies. One example that was identified in early linkage and association work is *GABRA2* (Covault et al., 2004; Edenberg et al., 2004; Lappalainen et al., 2005), a gene encoding a subunit of the neuron receptors that the neurotransmitter GABA binds to with great affinity. *GABRA2* demonstrates horizontal pleiotropy across developmental stages: in childhood and adolescence, variation at *GABRA2* is associated with conduct disorder (Dick et al., 2006; Sakai et al., 2010) and related "rule breaking" phenotypes (Trucco et al., 2014). In emerging adulthood, *GABRA2* instead is associated with alcohol dependence (Bierut et al., 2010). Similar patterns are also found for genes involved in the metabolism of alcohol (Irons et al., 2012) and have been replicated in analyses that examine multiple SNPs at once (Aliev et al., 2015).

Consequences of the Omission of Developmental Considerations from GWAS

Immediate Consequences

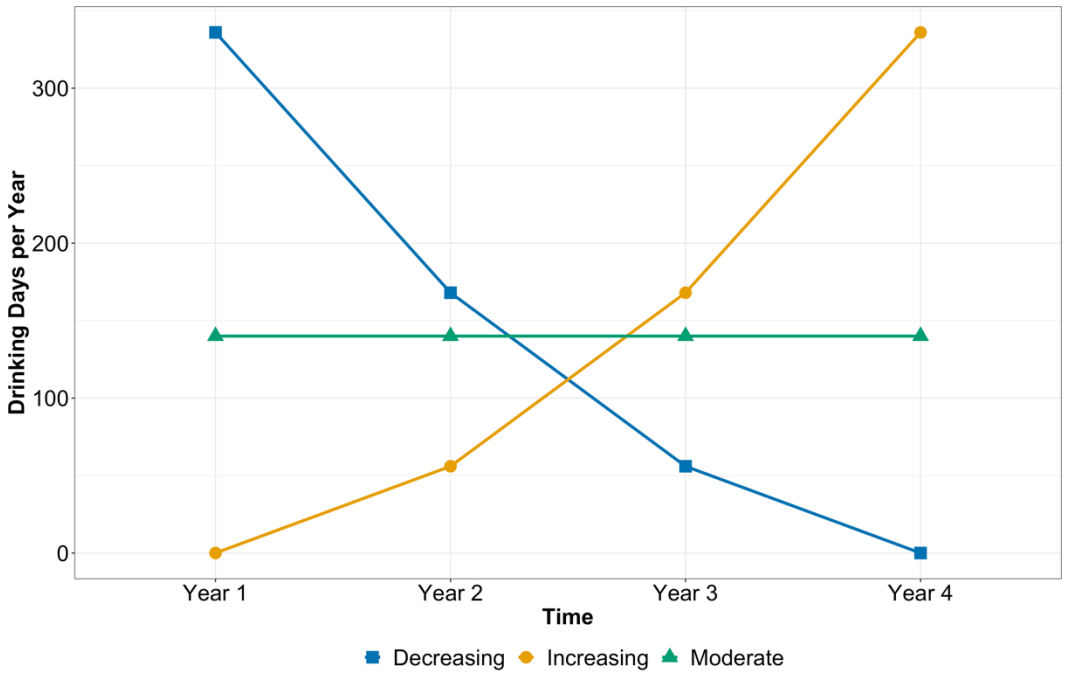
The omission of developmental considerations from GWAS has immediate consequences for the interpretation of results. Alcohol use behaviors are highly heterogeneous throughout the lifespan, with varied trajectories that are shaped by major life transitions (O'Malley, 2004; O'Malley & Johnston, 2002) and defined by differences in age of onset and persistence of use (Maggs & Schulenberg, 2004). Developmentally-agnostic GWAS do not account for differences in the magnitude of genetic effects throughout development, potentially leading to misspecification of the effect sizes of SNPs that act at specific developmental periods. In the context of GWAS with exceptionally large samples, a statistical association may still be detected between a development-specific SNP and an aggregate phenotype, but the estimated effect size would not correctly characterize the true relationship at any developmental period. The effect would be underestimated relative to the salient developmental period and overestimated relative

to the non-salient developmental period. In smaller samples, statistical associations between aggregate phenotypes and development-specific SNPs are unlikely to be detected at all.

An example of this problem is depicted in Figure 1, which shows an example of three possible trajectories of alcohol use frequency over time: (1) a steep decreasing trajectory of alcohol use over time in blue, (2) a steep increasing trajectory of alcohol use over time in yellow, and (3) a persistent moderate trajectory of alcohol use over time in green. The aggregate of all three patterns of longitudinal alcohol use in this example are identical, 140 drinking days per year, despite clear differences in level of alcohol use across time. A GWAS of aggregated longitudinal data would not differentiate between the different timing of peak alcohol use between these three patterns. Thus, developmentally-agnostic GWAS of aggregate phenotypes are unlikely to have much utility in the identification of genetic effects associated with alcohol use at specific developmental periods. Developmental period is a critical component in the array of factors that determine the behavioral impact of genetic predispositions. The effectiveness of precision medicine initiatives hinges on accurate identification of the genetic effects that are most salient to the individual, which varies throughout development.

Figure 1

Hypothetical alcohol use trajectories



Note. Data in this figure was generated by simulation for illustrative purposes.

Aggregation of repeated measures into cross-sectional phenotypes also reduces the statistical power of genetic analyses. These reductions in power were noted much earlier in linkage studies (Shi & Rao, 2008) and later confirmed in GWAS (Xu et al., 2014). Xu et al. (2014) present a direct comparison of (1) a cross-sectional analysis of the baseline measurement of a longitudinal neuroimaging phenotype and (2) a longitudinal analysis of the complete set of repeated measures using random effects to account for within-subjects clustering of observations. They report that more significant associations are identified in the longitudinal version of the analysis. One mechanism for this reduction in power can be inferred from previous research on the impact of misclassification of cases and controls on GWAS statistical power. Misclassification of controls as cases greatly decreases GWAS power; for example, sample size requirements are increased 3-fold if 50% of the people that are coded as affected cases in a

sample are actually unaffected controls (Manchia et al., 2013). This logic may be extended to comparisons of longitudinal and cross-sectional phenotypes in GWAS, where a cross-sectional phenotype may be a poor representation of the complete set of longitudinal observations. A decreasing trajectory and an increasing trajectory may produce the same value in aggregate, but differ in the timing of peak alcohol use and have distinct genetic etiologies.

For example, it is common for alcohol use to increase in early adulthood in response to the increased independence associated with college attendance, but later decrease as the individual reaches additional developmental milestones (O'Malley, 2004), such as marriage (Kretsch & Harden, 2014). Significant life events, such as divorce (Kretsch & Harden, 2014), may interrupt this decline in alcohol use. One individual's alcohol use may peak in early adulthood in the context of college, while another's may peak in later adulthood after marital dissolution. A cross-sectional analysis of these phenotypes would treat the two distinct peaks the same in aggregate, failing to reflect the full degree of developmental heterogeneity in the longitudinal phenotype. Reductions in statistical power are likely to be proportional to the amount of developmental heterogeneity in the phenotype. While most longitudinal cohort samples tend to be smaller than comparable cross-sectional studies, fewer participants may be necessary for adequate power when longitudinal data are modeled in the GWAS explicitly. Quantifying the sample size requirements for GWAS with longitudinal data and assessment of the data sets that are available to meet these requirements is an important step in evaluating the feasibility of developmentally-informative GWAS.

Downstream Consequences

The omission of developmental considerations from GWAS also has downstream consequences for statistical analyses that leverage GWAS summary statistics for genetic

prediction of phenotypes with polygenic scores (PGS). PGS aggregate measured genetic effects on a phenotype into a score that indexes the statistical association between SNPs and the phenotype (Dudbridge, 2013). Methods for PGS construction evolve rapidly, adopting different approaches to the general method of creating a weighted sum of the SNPs that are associated with the phenotype of interest. Methods for PGS construction vary in their approach to weighting SNPs in the presence of LD, often either transforming or removing SNPs with especially small effects (Zhao et al., 2021). PGS are used widely for the genetic prediction of phenotypes that are influenced by many SNPs of small effect (Duncan et al., 2019; Wray et al., 2014), including alcohol use phenotypes (Barr et al., 2019, 2020; Kandaswamy et al., 2021; Ksinan et al., 2019; Kuo et al., 2019; Li et al., 2017; Pasman et al., 2019; Salvatore et al., 2014; Thomas et al., 2018). Importantly, the accuracy and statistical power of a PGS is contingent on the accuracy and statistical power of the GWAS from which it is derived. Misspecification of effect sizes in GWAS translates to misspecification of individual-level genetic risk in PGS, such that the omission of developmental considerations from GWAS is equivalent to the omission of developmental considerations from PGS construction.

The omission of developmental considerations from PGS construction reduces the effectiveness of PGS in younger target samples. For example, PGS derived from developmentally-agnostic GWAS predict 0.58% and 0.61% of the variance in alcohol consumption in adolescence and early adulthood, respectively (Kandaswamy et al., 2021). PGS derived from the same discovery GWAS predict 2.4% of the variance in alcohol consumption in an older target sample (age 24-32 ; Liu et al., 2019). A recent study that examined age-specific effects of an alcohol consumption PGS from an adult discovery sample found that the PGS was associated with alcohol use in adulthood, but not adolescence (Elam et al., 2021). Despite

previous findings that the nature and magnitude of genetic effects vary throughout development (Aliev et al., 2015; Dick et al., 2006; Edwards & Kendler, 2013; Kendler et al., 2011; Meyers et al., 2014; Sakai et al., 2010), PGS have not been constructed to model this variability explicitly. Precision medicine initiatives rely on the accuracy of individual-level prediction of health outcomes. Incorporating a developmental perspective into the construction of polygenic scores is an important step towards refining individual-level prediction of health outcomes. A comprehensive method to model genetic effects that are specific to developmental stage will facilitate the developmentally-informed genetic prediction of alcohol use throughout the lifespan.

Statistical Approaches for Developmentally-Informative GWAS

A model for age-related effects in GWAS should explicitly account for the way that genetic influences on alcohol use vary across development. Statistical approaches to the inclusion of longitudinal data in GWAS must account for the inclusion of multiple, correlated observations from the same person over time. One approach to GWAS by structural equation modeling (SEM) is implemented in the R package GW-SEM (Verhulst et al., 2017). This flexible method is built on the OpenMx package for specifying SEM (Neale et al., 2016), allowing a diverse range of models to be specified. Notably, the GW-SEM approach requires raw data for all participants included in the GWAS. Developmentally-informative genetic models may also be specified using Genomic SEM, which leverages genetic correlations between phenotypes that are estimated via Linkage Disquilibrium Score Regression on GWAS summary statistics (Grotzinger et al., 2019). This approach does not require raw data from the sample of interest, relying instead on GWAS summary statistics, which can be stratified to index genetic liability within a specific age range. In both approaches, the SEM framework allows for explicit modeling of developmental considerations, such as age-specific effects.

In addition to software capabilities, there are a variety of other practical considerations in the development of a model for developmentally-informative GWAS that accounts for the ways that genetic effects vary throughout development. The primary obstacle to explicit modeling of developmental considerations in GWAS is the availability of well-powered samples with longitudinal phenotypes. The Genomic SEM framework is modular, such that additional samples can be added to boost power for discovery via meta-analysis of GWAS summary statistics rather than requiring raw data for all participants in the analysis. GWAS summary statistics can be shared more freely between research groups than raw genetic data. This approach may improve the feasibility of well-powered analyses when developmentally-informative GWAS summary statistics are available.

The Current Study

The first aim of this project was to advance gene discovery by building developmentally-informative models for gene identification that incorporate developmental changes in alcohol use across time. The second aim of this project was to leverage results from the longitudinal GWAS for genetic prediction of age-matched alcohol use outcomes in an independent sample. I applied novel multivariate genomic methods that combine developmentally-informative phenotypic data with GWAS data to create PGSs that are specific to developmental stage. To achieve this, began by conducting a common factor GWAS of alcohol use frequency at different developmental stages using a meta-analytic approach in Genomic SEM (Grotzinger et al., 2019).

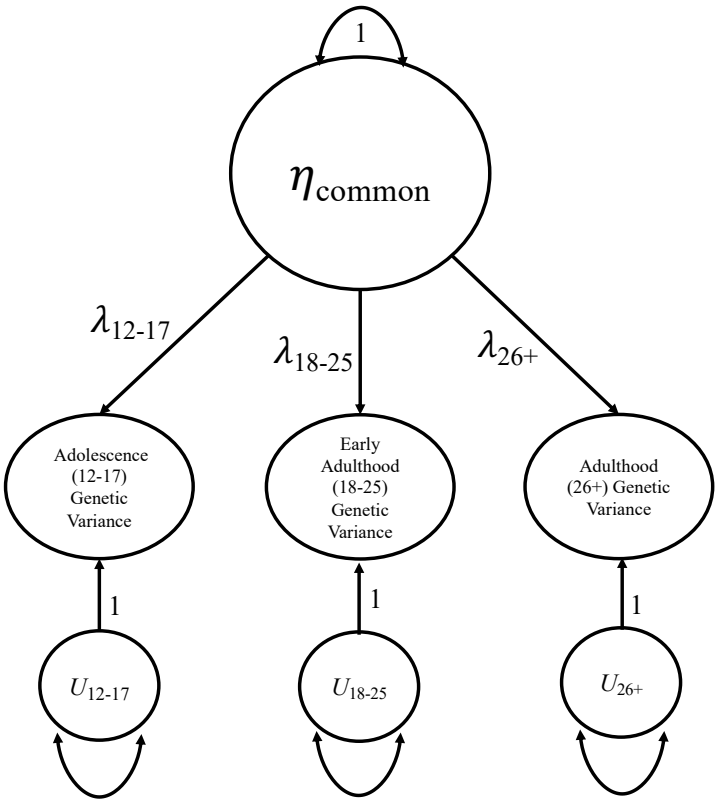
Data were drawn from multiple longitudinal cohort studies and meta-analyzed to increase statistical power for the longitudinal GWAS. The GWAS included data spanning adolescence (age 12-17; total n=9,589), early adulthood (age 18-25; total n=10,368), and adulthood (age 26+; total n=10,404). Longitudinal cohort studies targeted for gene-identification analyses include the

Collaborative Study on the Genetics of Alcoholism (COGA; adolescence n=1,118, early adulthood n=2,762, adulthood n=5,255), the National Longitudinal Study of Adolescent to Adult Health (Add Health; adolescence n=3,089, early adulthood n=3,993, adulthood n=5,149), and the Avon Longitudinal Study of Parents and Children (ALSPAC; adolescence n=5,382, early adulthood n=3,613).

The goal of this analysis was to measure genetic effects that are specific to each developmental stage. The hypothesis for this analysis was that the meta-analyzed GWAS results would yield unique associations at different developmental stages. Follow up analyses to support the interpretation of the Aim 1 GWAS results include identification of LD proxies and gene-mapping for significant SNPs as well as calculation of genetic correlations between components of the genomic structural equation model and various external phenotypes. A diagram of the genomic structural equation model is included in Figure 2.

Figure 2

Common Factor Genomic Structural Equation Model



Notes. Factor loadings are labeled with lambda. Residual variances are labeled with epsilon.

Additionally, I constructed PGS using weights from the residual components ($U_{12\text{-}17}$, $U_{18\text{-}25}$, U_{26+}) and the common factor (η_{common}) of the Genomic SEM model. PGS were constructed for specific developmental periods in order to test for associations between the PGS and alcohol use. The hypothesis for this analysis was that the residual PGS would predict their corresponding alcohol use phenotype significantly better than a development-agnostic, common factor PGS. This project provides an analytic approach for developmentally-informed genetic prediction of alcohol use outcomes, drawing on previous applications of Genomic SEM to model change and stability in genetic effects across age groups (Gillespie et al., 2022) and the residual genetic

variance of a phenotype after accounting for a common factor (Barr et al., 2022). When applied to larger, well-powered samples, this innovative method will significantly advance the field of genetics beyond the study of phenotypes that aggregate across developmental periods.

The gene discovery analyses conducted in Aim 1 require large sample sizes and statistical power in the current work is limited. The statistical power of the Aim 2 analyses depends, in part, on the statistical power of the Aim 1 discovery analysis as well. As a final step in the analysis plan, power analyses were conducted to determine the sample size required to detect development-specific effects in the Aim 1 discovery analysis and the Aim 2 PRS prediction analyses. Additional samples that may be targets to increase sample size and related directions for methodological developments in modeling development-specific genetic effects in future work are discussed in the Future Directions section.

The initial analysis plan for this project differs from the analysis described in the current work in four important ways. More details regarding the rationale for these adjustments is discussed in Appendix 1. Results from the parts of the original analysis plan that were feasible are reported in Appendix 2 and Appendix 3. A brief description of the adjustments is included below.

First, the initial analysis plan proposed to separate adolescence into two developmental periods (early adolescence age 12-14 and late adolescence age 15-17). Technical limitations associated with a critical step in the analysis plan (Linkage Disequilibrium Score Regression) required that early adolescence and late adolescence be combined into a single developmental period to boost statistical power. Second, the initial analysis plan aimed to analyze both alcohol use frequency and heavy episodic drinking. Similarly, sample sizes were not large enough for the heavy episodic drinking phenotype to produce stable heritability estimates using Linkage

Disequilibrium Score Regression. Third, polygenic score analyses were initially planned to be conducted in the youngest cohort from the Finn Twin Studies (Finn Twin 12). Logistical problems associated with access to the Finn Twin 12 data required that the Aim 2 analyses proceed without these data. PGS validation analyses were instead conducted in the COGA sample using an alternate version of the discovery analysis with COGA removed. Fourth, the Aim 1 discovery analysis was parameterized to model total genetic variance in each developmental period. Attempts to calculate genetic correlations between the model residuals and a series of phenotypes from other studies produced negative heritability estimates in early adulthood and adulthood (early adult H^2_{SNP} = -6e-04, adult H^2_{SNP} = -4e-04). Performing downstream analyses with the residuals of the model requires larger sample sizes than are available in the current study. An alternate parameterization of the model was used to estimate the genetic correlations with other phenotypes using the total genetic variance in adolescence, early adulthood, and adulthood. Polygenic risk score construction in Aim 2 proceeded with the residual variance parameterization of the model.

This work will focus exclusively on the updated analysis plan, with three developmental periods a focus on alcohol use frequency, and analysis of total genetic variance in each developmental period for Aim 1 which allowed for the completion of all analytic steps of Aim 1 and Aim 2.

Methods

Samples

This project used three longitudinal cohort studies for developmentally-informed gene discovery (COGA, ALSPAC, Add Health). Polygenic scores for alcohol use frequency were

constructed in COGA by conducting an alternate version of the gene discovery analyses excluding COGA.

All analyses were limited to European ancestry participants to limit the confounding influence of population stratification. Population stratification refers to the presence of statistical differences in linkage disequilibrium and allele frequencies between ancestry groups that occurs as a result of the varied migration patterns of early humans (Hellwege et al., 2017; Nielsen et al., 2017). When not addressed, population stratification in GWAS discovery samples can lead to increased likelihood of false positive or false negative findings (Marchini et al., 2004). The current best-practice for addressing population stratification in GWAS is to analyze a single ancestry group at a time with principal component covariates to address residual population structure within ancestry groups (Peterson et al., 2019). Most samples with longitudinal, genetically-informative data are comprised of participants of European ancestry. It is a critical priority to extend genetic analyses to a broader range of ancestry groups as more data becomes available.

Additionally, all analyses excluded lifetime non-drinkers. Contemporary GWAS of alcohol use outcomes typically exclude lifetime non-drinkers from analysis (Kranzler et al., 2019; Liu et al., 2019; Sanchez-Roige et al., 2019). In GWAS, modeling a phenotype that is constructed as a composite of two phenotypes with distinct genetic liabilities confounds their etiology. Previous research suggests that genetic liabilities for alcohol use initiation and alcohol consumption are at least partially distinct (Fowler et al., 2007; Heath, Meyer, Eaves, et al., 1991; Heath, Meyer, Jardine, et al., 1991; Pagan et al., 2006; Poelen et al., 2008). In cases where one phenotype is less heritable than the other, a combined phenotypic definition may lead to underestimation of genetic effects on the more heritable phenotype (Heath, Meyer, Eaves, et al.,

1991). Initiation of alcohol use is primarily influenced by environmental factors (Fowler et al., 2007; Pagan et al., 2006). As a result, confounding the etiology of initiation of alcohol use and frequency of alcohol use by using a phenotype that assigns a value of 0 to lifetime non-drinkers is likely to lead to underestimation of the contribution of genetics to alcohol use frequency in GWAS.

Collaborative Studies on the Genetics of Alcoholism (COGA)

COGA is a multi-generational family-based study of genetic and environmental factors for alcohol use outcomes, which ascertained alcohol-dependent probands from six US sites (Begleiter, 1995). The current work focused on the COGA Prospective Study (Phase 4), that examines how genetic and environmental risk unfolds across development among offspring of the initial COGA sample. Offspring between ages 12-22 with at least one parent who had previously completed an interview were assessed every two years (Bucholz et al., 2017). Data from the Phase 1 and Phase 2 surveys, which assess behavior in the parents of the COGA Prospective Study, was incorporated into analyses when possible, dependent upon phenotype availability as described in the Measures section.

For alcohol use frequency, reports between ages 12-17 (adolescence), 18-25 (early adulthood), and after age 26 (adulthood) were collected in Phase 4, Phase 1, and Phase 2 assessments as part of the SSAGA interviews (Bucholz et al., 1994).

The National Longitudinal Study of Adolescent to Adult Health (Add Health)

Add Health is a nationally representative longitudinal cohort of adolescents in grades 7-12 in the US in 1994-95 (K. M. Harris, 2013). The cohort has been followed through the transition to adulthood, with data collection occurring via in-home interviews. Measures are

primarily focused on causes of adolescent health behavior in the multiple contexts of adolescent life. Genetic data were collected in Wave 4.

For alcohol use frequency, reports between ages 12-17 (adolescence) were collected in Wave 1 and Wave 2. Reports between ages 18-25 (early adulthood) were collected in Wave 3 and Wave 4. Reports after age 26 (adulthood) were collected in Wave 3, Wave 4, and Wave 5.

Avon Longitudinal Study of Parents and Children (ALSPAC)

ALSPAC is a large longitudinal birth cohort which includes reports from approximately 14,000 children and their parents from early in the mothers' pregnancy through childhood and adolescence (Boyd et al., 2013; Fraser et al., 2013; Northstone et al., 2019). Pregnant women resident in Avon, UK with expected dates of delivery between 1st April 1991 and 31st December 1992 were invited to take part in the study. The total sample size for analyses using any data collected after the age of seven is therefore 15,447 pregnancies, resulting in 15,658 fetuses. Of these, 14,901 children were alive at 1 year of age. The project has collected comprehensive health-related information, including phenotypic outcomes, environmental factors, and DNA, with >85 assessments from mothers, their partners, and children, conducted from the pre-natal stage through emerging adulthood at yearly, or more frequent, intervals. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (http://www.bristol.ac.uk/alspac/researchers/our-data/). Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol (P. A. Harris et al., 2009). REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of data collected via

questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. The current work focuses on the offspring assessments in adolescence and early adulthood.

For alcohol use frequency reports between ages 12-17 (adolescence) were collected at ages 12.5, 13.5, 15.5, 16, and 17.5 from offspring. Reports between ages 18-25 (early adulthood) were collected at ages 18 and 20 from offspring.

Measures

As described below, measures of alcohol use frequency varied between and within the 3 samples of interest. In order to model developmentally-salient genetic effects, data were restructured and aggregated within developmental periods. The analysis plan targeted the following age ranges: ages 12-17 (adolescence), ages 18-25 (early adulthood), ages 26+ (adulthood). A complete description of data that was aggregated under each developmental period can be found below, as well as in Table S1. In what follows, I first describe the alcohol use frequency measures for each sample and then provide an overview of the phenotype harmonization procedure.

COGA

Alcohol Use Frequency

Reports between ages 12-17 were collected in the Phase 4, Phase 1, and Phase 2 assessments as part of the SSAGA interviews (Bucholz et al., 1994). In Phase 4, alcohol use frequency was measured with an item that asks “On how many days did you drink any beverages containing alcohol during the last 12 months?” with response options “every day”, “5-6 days a week (nearly every day)”, “4 days a week (200-259 days)”, “3 days per week (150-199 days)”, “2 days per week (100-149 days)”, “1 day per week (50-99 days)”, “3 days per month (36-49

days)", "2 days per month (24-35 days)", "1 day per month (12-23 days), "6-11 days per year", "3-5 days per year", "1 to 2 days per year", and "never". In Phase 1 and Phase 2, alcohol use frequency was measured with a series of items that ask "On a typical [DAY] in the past 6 months, how many drinks of [KIND OF ALCOHOL] would you have?", indexing consumption of beer, wine, liquor, and 'other' each day of the week with numeric free response. For participants missing data on these items, responses were drawn from a series of items asking "We would like to know the number of alcoholic drinks you've had each day in the last week […] How many drinks of [KIND OF ALCOHOL] did you have on [DAY]?" with numeric free response.

Reports between ages 18-25 were collected in the Phase 4, Phase 1, and Phase 2. In Phase 4, alcohol use frequency was measured with an item that asks "On how many days did you drink any beverages containing alcohol during the last 12 months?" with response options "every day", "5-6 days a week (nearly every day)", "4 days a week (200-259 days)", "3 days per week (150-199 days)", "2 days per week (100-149 days)", "1 day per week (50-99 days)", "3 days per month (36-49 days)", "2 days per month (24-35 days)", "1 day per month (12-23 days), "6-11 days per year", "3-5 days per year", "1 to 2 days per year", and "never". In Phase 1 and Phase 2, alcohol use frequency was measured with a series of items that ask "On a typical [DAY] in the past 6 months, how many drinks of [KIND OF ALCOHOL] would you have?", indexing consumption of beer, wine, liquor, and 'other' each day of the week with numeric free response. For participants missing data on these items, responses were drawn from a series of items asking "We would like to know the number of alcoholic drinks you've had each day in the last week […] How many drinks of [KIND OF ALCOHOL] did you have on [DAY]?" with numeric free response.

Reports after age 26 were collected in the Phase 4, Phase 1, and Phase 2. In Phase 4, alcohol use frequency is measured with an item that asks “On how many days did you drink any beverages containing alcohol during the last 12 months?” with response options “every day”, “5-6 days a week (nearly every day)”, “4 days a week (200-259 days)”, “3 days per week (150-199 days)”, “2 days per week (100-149 days)”, “1 day per week (50-99 days)”, “3 days per month (36-49 days)”, “2 days per month (24-35 days)”, “1 day per month (12-23 days), “6-11 days per year”, “3-5 days per year”, “1 to 2 days per year”, and “never”. In Phase 1 and Phase 2, alcohol use frequency was measured with a series of items that ask “On a typical [DAY] in the past 6 months, how many drinks of [KIND OF ALCOHOL] would you have?”, indexing consumption of beer, wine, liquor, and ‘other’ each day of the week with numeric free response. For participants missing data on this item, responses were drawn from a series of items asking “We would like to know the number of alcoholic drinks you've had each day in the last week […] How many drinks of [KIND OF ALCOHOL] did you have on [DAY]?” with numeric free response.

Covariates

Sex was assessed using a single binary item. Age at assessment was calculated from reported date of birth for presentation in the supplemental tables. Birth year is clustered by decade for presentation in the supplemental tables but was included in analyses as a continuous measure. Genotyping array was included as a covariate. Proband status was also included as a covariate in early adulthood and adulthood.

Add Health

Alcohol Use Frequency

Reports between ages 12-17 were collected in Wave 1 and Wave 2. At Wave 1 and Wave 2, alcohol use frequency was measured with an item that asks “During the past 12 months, on how many days did you drink alcohol?” with response options “never”, “once or twice”, “once a month or less”, “2 or 3 days a month”, “once or twice a week”, “3 to 5 days a week”, and “nearly every day”.

Reports between ages 18-25 were collected in Wave 3 and Wave 4. At Wave 3 and Wave 4, alcohol use frequency was measured with an item that asks “During the past 12 months, on how many days did you drink alcohol?” with response options “never”, “once or twice”, “once a month or less”, “2 or 3 days a month”, “once or twice a week”, “3 to 5 days a week”, and “nearly every day”.

Reports after age 26 were collected in Wave 3, Wave 4, and Wave 5. At Wave 3, Wave 4, and Wave 5, alcohol use frequency was measured with an item that asks “During the past 12 months, on how many days did you drink alcohol?” with response options “never”, “once or twice”, “once a month or less”, “2 or 3 days a month”, “once or twice a week”, “3 to 5 days a week”, and “nearly every day”.

Covariates

Sex was assessed using a single binary item. Age at assessment was calculated from reported date of birth for presentation in the supplemental tables. Birth year is clustered by decade for presentation in the supplemental tables but was included in analyses as a continuous measure.

ALSPAC

Alcohol Use Frequency

Reports between ages 12-17 were collected at ages 12.5, 13.5, 15.5, 16, and 17.5. At the age 12.5 assessment and age 13.5 assessment, alcohol use frequency was measured with an item asking “How many times have you drunk alcohol in the past 6 months?” with numeric free response. At the age 15.5 assessment alcohol use frequency was measured with an item that asks “How many times have you had a full drink of alcohol in the last 6 months?” with response options “1-2 times”,“3-5 times”, “6-9 times”, “10-19 times”, “20-39 times”, “40-99 times”, “100 or more”. At the age 16 assessment and the age 17.5 assessment alcohol use frequency was measured with an item that asks “The next questions are about your use of alcoholic drinks during the past year. How often do you have a drink containing alcohol?” with response options “never”, “monthly or less”, “2-4 times a month”, “2-3 times a week”, and “4 or more times a week”.

Reports between ages 18-25 were collected at ages 18 and 20. At the age 18 assessment and the age 20 assessment alcohol use frequency was measured with an item that asks “The next questions are about your use of alcoholic drinks during the past year. How often do you have a drink containing alcohol?” with response options “never”, “monthly or less”, “2-4 times a month”, “2-3 times a week”, and “4 or more times a week”.

Covariates

Sex was assessed using a single binary item. Age at assessment was inferred from the target age of the corresponding assessment for presentation in the supplemental tables; for example, participants with data from the age 12.5 assessment have their age set to 12.5 for that assessment. Birth year is clustered by decade for presentation in the supplemental tables. Birth year was not included as a covariate because all ALSPAC offspring were born in either 1991 or 1992.

Phenotype Harmonization

Data were aggregated into the specified developmental periods (age 12-17 adolescence, age 18-25 early adulthood, age 26+ adulthood) by taking the maximum value of available data for each subject within each developmental period. In alignment with previous evidence that genetic liability for alcohol use initiation is distinct from genetic liability for frequency of alcohol use and problem drinking (Fowler et al., 2007; Heath, Meyer, Eaves, et al., 1991; Heath, Meyer, Jardine, et al., 1991; Pagan et al., 2006; Poelen et al., 2008) participants who had not initiated alcohol use at the corresponding assessment were excluded from each analytic sample.

All survey items were transformed to a pseudo-continuous scale of drinking days per year. Ordinal frequency categories were converted to the median of the described range and rescaled from the original scale of measurement to reflect drinking days per year by multiplying by the corresponding constant (i.e. days per week * 52; days per month * 12). A similar procedure has been used in previous studies to construct pseudo-continuous measures of alcohol consumption (Dawson, 2000; Salvatore et al., 2016; Thomas et al., 2018). Measures in Phase 1 and Phase 2 of COGA which assess quantity per day of the week were recoded to a binary item of any use for each day of the week, summed to a count of drinking days per week, and multiplied by 52 to compute a measure of drinking days per year. The resulting transformed variables have a fixed range of 0 to 365, precluding the inclusion of extreme outlier values.

When phenotype definitions differ between studies, harmonization is a critical step in GWAS meta-analyses to ensure that phenotypes are similar enough to be suitable for meta-analysis (Zeggini & Ioannidis, 2009). In line with recommendations from Bennet et al. (2011), preliminary descriptive analyses were conducted within each study to identify any systematic differences in patterns of alcohol use between the cohort studies prior to meta-analysis. Means of

alcohol use frequency (drinking days per year) were compared between studies within developmental periods. In order to compare source populations between the studies, alcohol use frequency was examined as a function of birth year, sex, and age within developmental periods. Results from phenotype harmonization procedures are presented in the Results section. An interpretation of mean differences is presented in the Discussion section.

Genotyping

In COGA, participants were genotyped on four different genotyping arrays: the Illumina 1M, Illumina OmniExpress 12VI, and Illumina 2.5M (Illumina, San Diego, CA), and Smokescreen (BioRealm LLC, Walnut, CA). Additional information regarding genotyping procedures in COGA is available in Lai et al. (2019). In ALSPAC, participants were genotyped on the Illumina Human Hap550 Quad array. Additional information regarding genotyping procedures in ALSPAC is available in Paternoster et al. (2011) and Chong et al. (2021). In Add Health, participants were genotyped on the Illumina Human Omni1-Quad BeadChip and the Illumina Human Omni-2.5 Quad BeadChip. Genotypes in Add Health were imputed together using a set of 609,130 SNPs in common between the two arrays. Additional information regarding genotyping procedures in Add Health is available in Highland et al. (2018).

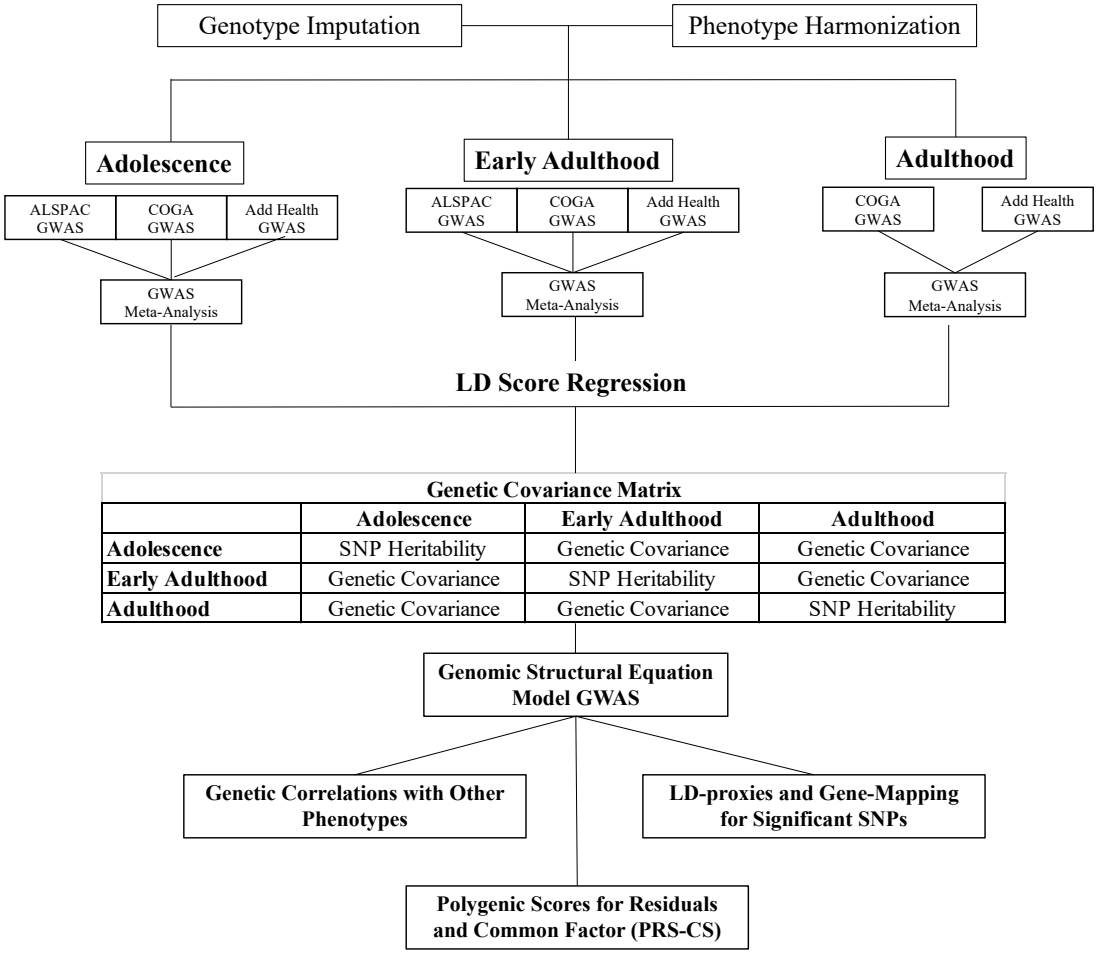
Standard pre-imputation quality control metrics (Marees et al., 2018) were applied in all within-sample genetic analyses including removing SNPs with minor allele frequency (MAF) <1%, call rate <95%, Hardy-Weinberg Equilibrium (HWE) $p < 1e\text{-}6$, and removal of participants with excess autosomal heterozygosity or homozygosity ($F < -.1$ or $F > .1$). All samples were imputed to the Haplotype Reference Consortium (HRC) reference panel (McCarthy et al., 2016).

Analysis Plan

An overview of the full analysis plan is included in Figure 3.

Figure 3.

Diagram of the analysis plan



Developmentally-Informative GWAS

The first set of analyses in this project was comprised of a developmentally-informative meta-analysis GWAS of alcohol use frequency. The hypothesis for both analyses was that the

meta-analyzed GWAS results would demonstrate unique associations at different developmental stages. This analysis involved 4 distinct steps, described below.

Step One: Within-Sample GWAS of Alcohol Use.

Separate GWAS were conducted at each developmental period for each available sample with linear regression using PLINK (Purcell et al., 2007) in unrelated participants (ALSPAC) and a linear mixed model with a genetic relatedness matrix (GRM) using Genome-wide Complex Trait Analysis (GCTA; Yang et al., 2011) in related participants (COGA, Add Health). SNPs with low imputation quality (Info Score < .80) were removed prior to analysis. GRMs were constructed in each sample (COGA, Add Health) using GCTA. SNPs were pruned for GRM calculation using a 50 SNP window, shifting by 5 SNPs, with a variance inflation factor threshold of 2 in PLINK. Covariates included in all GWAS were sex, and 10 ancestry principal components to adjust for population stratification. GWAS in the COGA and Add Health samples included birth year as a covariate. GWAS in all COGA samples included genotyping array as a covariate. GWAS in the COGA early adulthood and adulthood samples included proband status as a binary covariate. The primary output of this step was a series of within-sample age-stratified GWAS summary statistics files, each of which index the relationship between SNPs and alcohol use at a specific developmental period in one of the cohort studies.

Step Two: Meta-Analysis of Within-Sample GWAS Results.

Sets of GWAS summary statistics from different studies were meta-analyzed within developmental period using METAL (Willer et al., 2010). METAL includes implementation of two fixed-effects meta-analysis methods which differ in their approach to calculating weights for each study in the meta-analysis. One approach returns the inverse-variance weighted (IVW) average of each SNP coefficient, while the other method returns the weighted sum of SNP Z-

scores (Willer et al., 2010). While these methods are equivalent under most circumstances, the weighted sum of Z-scores approach is suboptimal when MAF varies between samples (Lee et al., 2016). I therefore implemented the IVW procedure in this work. The equations for sample weight (w_i; Equation 1), the meta-analyzed standard error (*SE*; Equation 2), and the meta-analyzed β (β; Equation 3) from Willer et al. (2010) are presented below, where β_i is the coefficient estimate for a SNP from study *i* and SE_i is the standard error for a SNP from study *i*.

Equation 1. Inverse Variance Weight (IVW) Estimator.

$$w_i = \frac{1}{SE_i^{\,2}}$$

Equation 2. IVW Standard Error Estimator.

$$SE = \sqrt{\frac{1}{\sum w_i}}$$

Equation 3. IVW Weighted Beta Estimator.

$$\beta = \frac{\sum \beta_i w_i}{\sum w_i}$$

The primary output of this step was a series of meta-analyzed GWAS summary statistics files, each of which index the relationship between SNPs and alcohol use at a specific developmental period across all of the cohort studies.

Step Three: Genetic Correlations and Genomic SEM.

Linkage Disequilibrium Score Regression (LDSC), first implemented in the ldsc software package (Bulik-Sullivan, Loh, et al., 2015), was performed on the meta-analyzed GWAS summary statistics to construct a genetic covariance matrix using the using the Genomic SEM

package (Grotzinger et al., 2019) in R (R Core Team, 2017). In LDSC, the SNP χ^2 test statistics are regressed on SNP LD scores. The LD score of a SNP is the sum of LD r^2 values with all other SNPs and is estimated from an external reference panel. In line with previous work (Bulik-Sullivan, Loh, et al., 2015; Grotzinger et al., 2019), LD scores were drawn from the 1000 Genomes reference panel (McVean et al., 2012) and restricted to HapMap3 SNPS (Altshuler et al., 2010), which are well characterized in terms of LD structure. The slope coefficient for the LD score from this regression is an estimate of the SNP-based heritability of the phenotype. As described in Bulik-Sullivan, Finucane et al. (2015), this method can be extended to examine two traits by regressing the product of two SNP test statistics on the corresponding LD scores. In the bivariate extension of LDSC, the slope coefficient of the LD score is an estimate of the genetic covariance of the two phenotypes. The genetic covariance matrix contains SNP-based heritability estimates on the diagonal and genetic covariances on the off-diagonal.

The primary output of this step was a genetic covariance matrix that contains SNP-based heritability estimates and genetic covariances between alcohol use at different developmental periods (or, equivalently, a genetic correlation matrix). Assessment of the genetic correlations provides an initial indication of whether the hypothesis of Aim 1, that the meta-analyzed GWAS results would yield unique associations at different developmental stages, is supported. A structural equation model can be fit by leveraging the genetic covariance matrix using the Genomic SEM package (Grotzinger et al., 2019) in R (R Core Team, 2017). I fit a common factor model in order to construct PGS that are specific to developmental periods, where SNP effects are estimated for development-specific genetic factors and the genetic liability that is common to all time points. A diagram of the common factor gSEM model is included in Figure 1. The variance of the common factor was fixed to 1 to identify the model without SNPs.

Loadings from the common factor to genetic variance at each developmental period ($\lambda_{12\text{-}17}$, $\lambda_{18\text{-}25}$, λ_{26+}) and the corresponding residual variances ($U_{12\text{-}17}$,$U_{18\text{-}25}$,U_{26+}) were freely estimated.

Step Four: Estimate Effects of each SNP on Components of the Genomic SEM.

The gSEM framework allows for the inclusion of individual SNP effects as predictors of the various components of the model. One SNP is included in the model at a time, so a genome-wide analysis involves estimating an additional iteration of the model for each SNP. A separate analysis was conducted for the common factor and each of the developmental periods, resulting in four gSEM GWAS. The primary output of this step was a series of GWAS summary statistics that reflect measured genetic effects on the various components of the model. The common factor indexes genetic variance that is shared across developmental periods. The residual components for each developmental period index genetic variance that contributes to alcohol use at the corresponding developmental period.

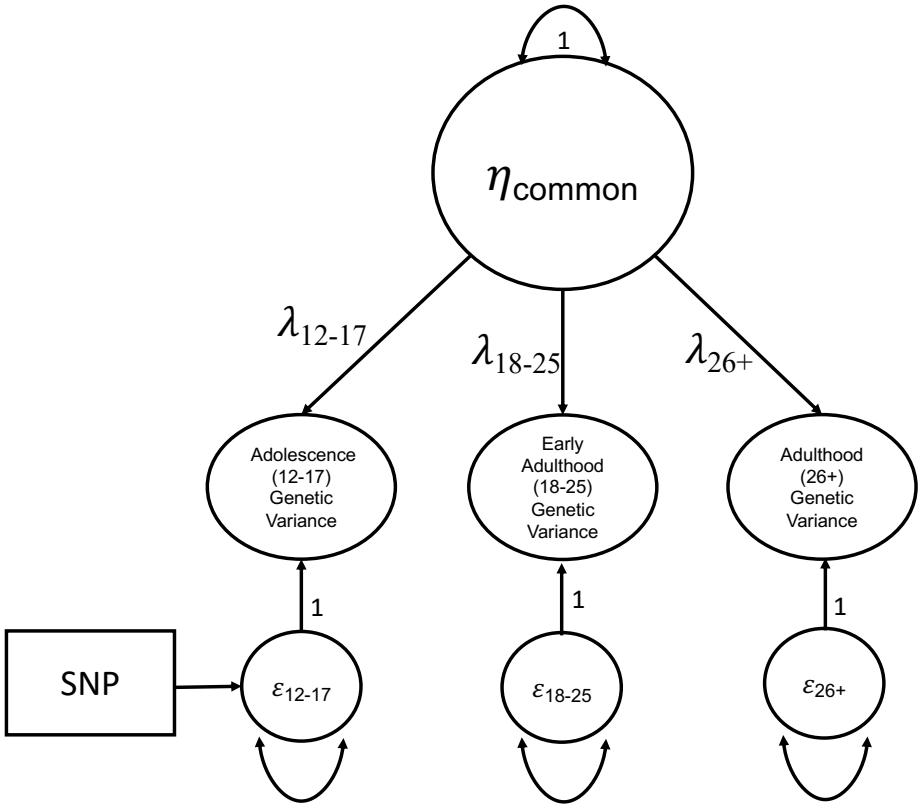
The tolerance setting for matrix inversion was set to a relatively liberal value (1e-50) to allow model fitting to proceed. The 'standard' option was selected to implement Genomic Control, which adjusts SNP standard errors for population stratification by multiplying them by the square root of the LDSC intercept. SNPs were restricted to those appearing in both the meta-analyses and the 1000 Genomes referenced panel and had a MAF > .01 in the reference panel. The total number of SNPs included in the analysis was 6,707,536.

For the GWAS in each developmental period ($U_{12\text{-}17}$, $U_{18\text{-}25}$, U_{26+}), paths were estimated from each SNP to the corresponding model component. Without simultaneously estimating a path from the SNP to the common factor, this analysis models SNP effects on the total variance in each developmental period, rather than the residual variance. Further discussion of this limitation is included in Appendix 1. Model identification was achieved by fixing the variance of

the common factor (η_{common}) to one and constraining all variances to be greater than 0.01. This identification strategy was successful for the analysis of all 6,707,536 SNPs included in the GWAS of adolescence, early adulthood, and adulthood. A diagram of the gSEM GWAS model for SNP effects on the total genetic variance in adolescence is included in Figure 4.

Figure 4

gSEM GWAS model for SNP effects on total genetic variance in adolescence



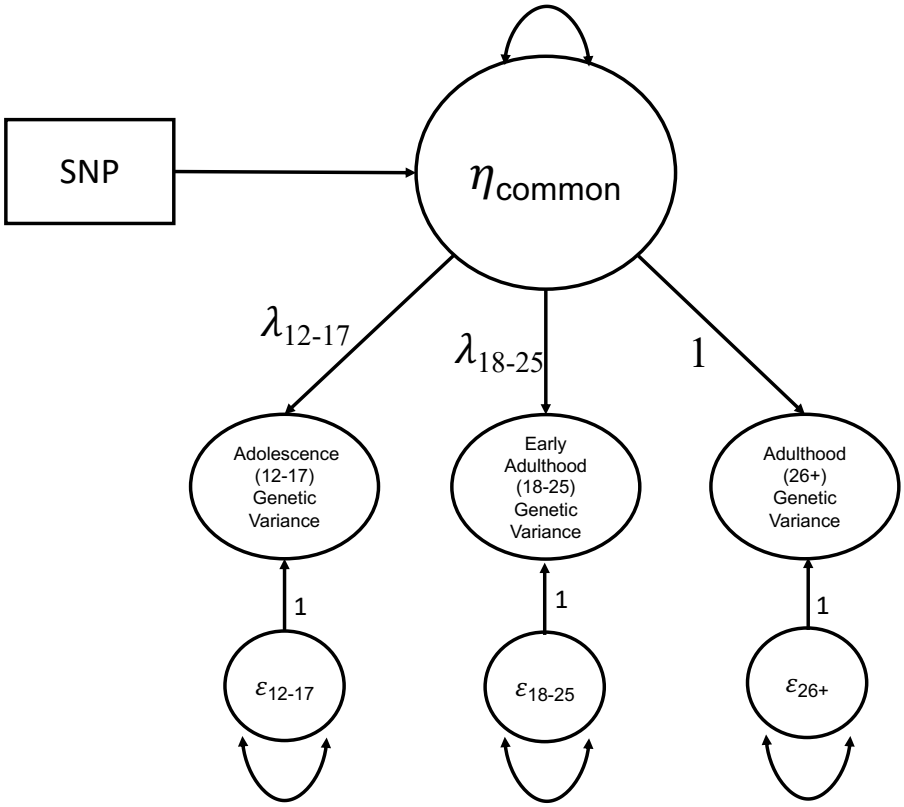
Notes. Factor loadings are labeled with lambda. Residual variances are labeled with epsilon.

In the GWAS of the common factor (η_{common}), the size and significance of the pathway from each from each of the 6,707,356 SNPs to the common factor was estimated. The loading of adulthood on the common factor (λ_{26+}) was set to one and the variance of the common factor was freely estimated. Additionally, all variances were constrained to be greater than 0.01. This model

identification strategy was successful for the analysis of 6,651,669 SNPs included in the common factor GWAS. A diagram of the gSEM GWAS model for SNP effects on the common factor is included in Figure 5.

Figure 5

gSEM GWAS model for SNP effects on the common factor



Notes. Factor loadings are labeled with lambda. Residual variances are labeled with epsilon.

Sensitivity analyses were conducted to explore possible differences between the 6,651,669 SNPs that were retained in the analysis (Analysis SNPs) and the 55,867 SNPs that were removed from the analysis (Error SNPs). The distribution of minor allele frequency (MAF) and chromosome (CHR) in Analysis SNPs and Error SNPs was similar (Figure S1). Mean test statistics for Error SNPs and Analysis SNPs were calculated from the three input meta-analyzed

GWAS. The distribution of mean test statistics in Error SNPs was bimodal, while the distribution in Analysis SNPs was approximately normal (Figure S2). The range between the smallest and largest test statistic in the input meta-analyzed GWAS was calculated for each SNP, such that a larger test statistic range indicates that the SNP effect varies more across developmental periods. The distribution of test statistic ranges was centered on a higher value in Error SNPs compared to Analysis SNPs (Figure S2). Together, these sensitivity analyses suggest that errors in the common factor GWAS are more likely to occur when the SNP effect varies more across developmental periods. In these cases, a common factor model represents a poor explanation of the effect of the SNP on alcohol use frequency across development and model estimation produces errors.

A series of post-GWAS analyses were conducted to support interpretation of the gSEM GWAS results. First, individual SNPS that reached the threshold for genome-wide significance were mapped to genes using FUMA (Watanabe et al., 2017) within the default 10 kilobase (kb) range using the SNP2GENE function. Mapped genes were then entered into the NHGRI-EBI Catalog of Human Genome-wide Association Studies (MacArthur et al., 2017) to determine if they have been detected in previous GWAS.

Next, SNPs that reached the threshold for genome-wide significance were entered into the LDlink LDproxy Tool (Machiela & Chanock, 2015) to identify LD proxies for significant SNPs. Previous work defines an LD proxy as a SNP with R^2 greater than or equal to .80 with the target SNP (Hammond et al., 2021). Five European reference populations (CEU, TSI, FIN, GBR, IBS) and the default window size of 500 kb were selected for this analysis. The LDproxy Tool provides annotations of SNPs with a known functional role as well as the RegulomeDB score, which assesses the likelihood that the SNP has a regulatory role in gene transcription. The

RegulomeDB scores are on a 1 to 7 rating scale, where a value of 1 indicates a likely regulatory role and 7 indicates that the SNP is not likely to have a regulatory role. GWAS significant SNPs and SNPs with LD R^2 values greater than or equal to .80 from the LDproxy Tool were entered into dbSNP (Smigielski et al., 2000) to assess if significant SNPs or their proxies have been detected in previous GWAS.

Lastly, genetic correlations were calculated between the components of the genomic structural equation model and a series of other phenotypes with publicly available GWAS summary statistics using LDSC. This follow-up analysis was conducted to assess qualitative differences in what constitutes genetic liability for alcohol use frequency across development, Summary statistics for alcoholic drinks per week (Effective N= 656,317.90) and cigarettes per day (Effective N= 312,759.30) were obtained from a publicly available repository maintained by the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) (Saunders et al., 2022). Summary statistics for a common factor of externalizing behavior were obtained from a publicly available repository hosted by the Externalizing Consortium (Effective N = 1,045,957; Karlsson Linnér et al., 2021; Williams et al., Manuscript in Preparation). Summary statistics for the three items that comprise the Consumption subscale of the Alcohol Use Disorder Identification Test (AUDIT-C; alcohol use frequency, Effective =117,913.90; alcohol use quantity, Effective N=108,484.90; binge drinking frequency, Effective N=108,484.90), average household income (Effective N = 311,027.80), the Townsend Deprivation Index (Effective N = 360,762.70), and neuroticism (Effective N= 293,005.80) were obtained from a publicly available repository of GWAS results from UK BioBank (UKB) (Neale Lab, 2018). Summary statistics for major depressive disorder (Effective N = 415,832.20; Howard et al., 2019), alcohol dependence (Effective N= 19,122.18 ; Walters et al., 2018), and the Problems subscales of the Alcohol Use

Disorder Identification Test (AUDIT-P) (Effective N= 119,770.10; Sanchez-Roige et al., 2019) were obtained from a publicly available repository maintained by the Psychiatric Genomics Consortium (PGC). Summary statistics for anxiety disorders were obtained from a repository hosted by the authors of a recent Anxiety GWAS (Effective N = 51,867.28; Purves et al., 2020). Summary statistics for educational attainment (Effective N = 744,789.60; Okbay et al., 2022), and risk tolerance (Effective N = 450,661.70; Karlsson Linnér et al., 2019) were obtained from a publicly available repository maintained by the Social Science Genetic Association Consortium (SSGAC).

SNP-specific meta-analytic sample sizes were available for GSCAN drinks per week, GSCAN cigarettes per day, PGC alcohol dependence, and PGC AUDIT-P. The average SNP-specific sample size was used as the effective sample size for these summary statistics. The effective sample size of the common factor of externalizing behavior was also readily available. SNP-specific effective sample sizes were not directly available in any of the other external GWAS summary statistics. Maximum SNP sample size was available in all cases, but maximum sample size does not account for SNP-level missingness or the effect of case-control balance on statistical power for binary traits. Effective N for these GWAS summary statistics was calculated from available information. Maximum SNP sample size and individual SNP call-rate was available in all UKB GWAS summary statistics (alcohol use frequency, frequency of binge drinking, average household income, and Townsend Deprivation Index). SNP-specific sample sizes were calculated by multiplying maximum SNP sample size by SNP call rate. For GWAS of continuous traits without SNP call-rate (SSGAC educational attainment and risk tolerance), effective N was calculated using the formula for expected GWAS sample size defined in Mallard

et al., (2022). For case control GWAS (PGC major depressive disorder, anxiety disorders), effective N was calculated using the formula defined in Grotzinger et al. (2023).

Validation Analysis: Polygenic Scores

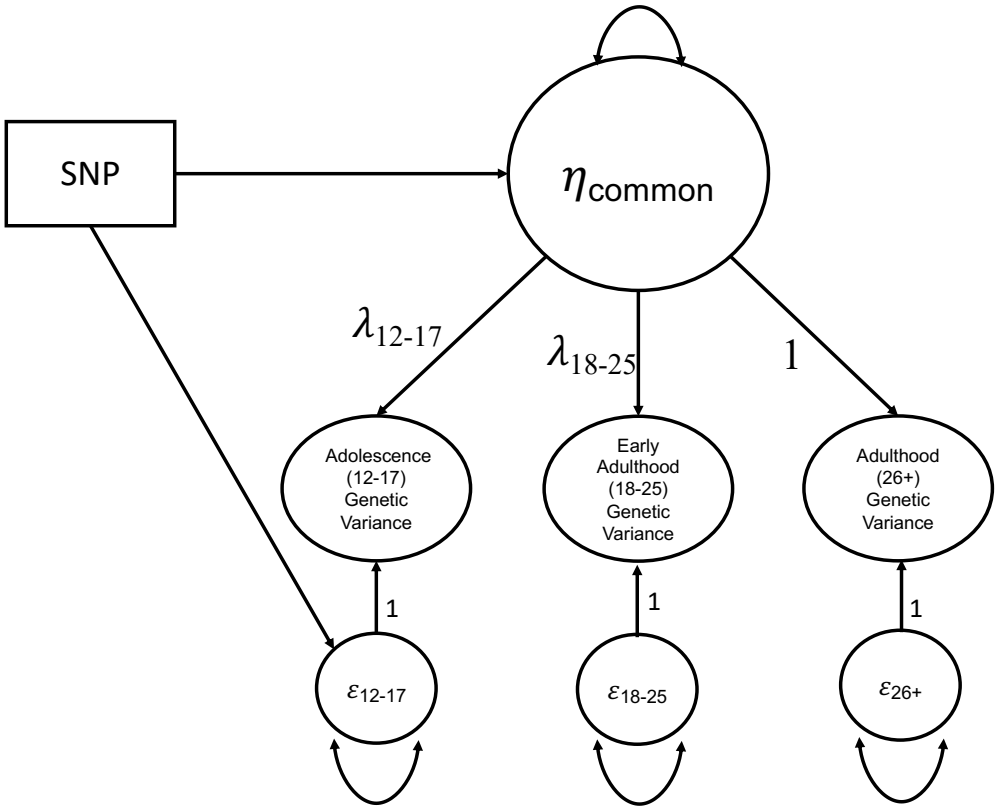
An alternate version of the Aim 1 discovery analysis was conducted with COGA removed in order to facilitate PGS construction in the COGA sample. Overlap between discovery and target samples leads to overfitting, inflating estimates of the variance accounted for by the PGS (Choi et al., 2018; Wray et al., 2013). This can lead to misleading conclusions and reduces the generalizability of findings, as the PGS would be biased to perform better in the overlapping target data than it would in an arbitrary external sample. Wray et al. (2013) provide an explanation of the mechanism underlying overfitting of SNP effects when discovery and target samples overlap. The expected variance explained by a SNP with no true effect in the population is $\frac{1}{N-1}$, rather than zero. These random departures from zero are aggregated when GWAS summary statistics are used as weights to construct PGS, inflating the variance accounted for by the resulting PGS when the discovery sample and target sample are the same. Random fluctuations in SNP effect size in the discovery sample are less likely to account for phenotypic variance in an external sample. PGS were constructed from the weights estimated in the Genomic SEM with COGA removed to reduce the risk of overfitting.

The gSEM GWAS model for the residuals was fit with paths from the SNP to both the residual and the common factor, partitioning the SNP effects into a component that is attributable to the common factor and a component that is directly associated with the residual genetic variance. The total number SNPs in the analysis, included in the meta-analysis of ALSPAC and Add Health and the reference panel was 6,071,632. Model identification was achieved by fixing the loading of adulthood on to the common factor to 1 for the GWAS of the common factor, the

adolescence residual, and the early adulthood residual. This model identification strategy was successful for the analysis of 6,065,592 SNPs included in the adolescence residual GWAS, 6,045,602 SNPs in the early adult residual GWAS, and 6,022,670 SNPs in the common factor GWAS. For the adult residual GWAS model identification was achieved by fixing the loading of early adulthood to one and constraining all variances to be greater than 0.01. This model identification strategy was successful for the analysis of 6,045,602 SNPs included in the adult residual GWAS. A diagram of the gSEM GWAS model for residual genetic variance in adolescence is included in Figure 6.

Figure 6

gSEM GWAS model for SNP effects on the residual genetic variance in adolescence



In support of Aim 2, sensitivity analyses were conducted to explore possible differences between the SNPs that were retained in each analysis (Analysis SNPs) and the SNPs that were

removed from the analysis (Error SNPs). Similar to the Aim 1 common factor gSEM GWAS with all samples included, these sensitivity analyses suggest that errors in the gSEM GWAS with COGA removed are evenly distributed across the genome (Adolescence: Figure S3, Early Adult: Figure S5, Adult: Figure S7, Common Factor Figure S9) and more likely to occur when the SNP effect varies across developmental periods (Adolescence: Figure S4, Early Adult: Figure S6, Adult: Figure S8, Common Factor Figure S10).

Regarding polygenic score construction, current evidence favors the use of Bayesian continuous shrinkage priors to reduce noise in a way that is informed by LD structure (Ge et al., 2019; https://github.com/getian107/PRScs). Specification of a tuning parameter (ϕ) is required, which is defined as the global shrinkage parameter and establishes the sparsity of the genetic architecture of the phenotype. Here, sparsity refers to the expected number of SNPs that have a causal relationship with the phenotype. There are two variants of the method: PRS-CS and PRS-CS-auto. In PRS-CS, the ϕ parameter is chosen manually based on prior knowledge regarding the genetic architecture of the trait. In PRS-CS-auto, the ϕ parameter is learned from the observed data. The authors report that PRS-CS-auto performs poorly when GWAS sample sizes are small because the ϕ parameter cannot be learned correctly from the observed data. Instead, the authors recommend a value of 1e-2 for the ϕ parameter when predicting a highly polygenic trait with small GWAS sample size. The ϕ parameter was set to 1e-2 in this analysis to accommodate the small sample size of the meta-analyzed GWAS in adolescence, early adulthood, and adulthood.

Other parameters involved in the analysis include a local shrinkage parameter (ψ_j, for SNP j) and a series of Markov-Chain Monte Carlo (MCMC) sampling settings. The ψ_j parameter varies as a function of SNP effect size, allowing smaller SNP effects to be reduced more, and is

defined by a prior gamma distribution with shape (α) and scale (β) parameters. The authors suggest that the parameterization of the prior distribution for the ψ_j parameter has limited impact on prediction accuracy and recommend a default setting of $\alpha = 1$ and $\beta = 0.5$. MCMC sampling settings include the total number of iterations of the algorithm, the number of iterations that are run as burn-in steps to calibrate the algorithm and reduce the impact of arbitrary starting values, and a thinning factor which removes every i-th entry in the chain. Generally, these MCMC sampling settings balance computation time against the stability of the solution returned by the algorithm. The authors suggest that the default values of 1000 total iterations with the first 500 run as "burn-in" and a thinning factor of 5 typically produces similar results to the results returned by more computationally expensive settings. In line with these recommendations, parameterization of the ψ_j parameter and MCMC sampling settings were left at their respective default values.

A series of mixed effects models with random intercepts for family group were used to assess the effect of the residual PGS and the common factor PGS on the corresponding phenotype in COGA using the lme4 package (Bates et al., 2015) and lmerTest package (Kuznetsova et al., 2017) in R (R Core Team, 2017). This analysis was conducted in support of the Aim 2 hypothesis that the residual PGSs would predict their corresponding alcohol use phenotype significantly better than a development-agnostic, common factor PGS, Random intercepts correct the standard error of the regression coefficients for similarity between participants due to their clustering within a family. This approach has been used previously to assess the effect of PGS while adjusting for familial clustering (Barr et al., 2020). A separate mixed effects model was used for each developmental period included in the gSEM GWAS. Sex,

birth year, genotyping array, and 10 PCs were included as covariates in each model. Proband status was included as a covariate in early adulthood and adulthood.

Alcohol phenotypes were regressed on the residual PGS that matches the developmental period in which they were assessed. Each model also included the common factor PGS and covariates. The random intercept model is defined below in Equation 4, where Y_{ij} is the phenotype for individual i in family j, B_{0j} is the intercept for family j, β_{PGS} is the fixed effect for the residual PGS, β_{cPGS} is the fixed effect for the common-factor PGS, $\beta_{covariate}$ represents the fixed effects of sex, and 10 ancestry PCs, e_{ij} is the residual for individual i in family j, β_0 is the grand intercept, and u_j is the deviation of the intercept for family j from the grand intercept.

Equation 4. Random intercept model for PGS validation in COGA.

$$Y_{ij} = \beta_{0j} + \beta_{PGS} + \beta_{cPGS} + \beta_{covariate} + e_{ij}$$

$$B_{0j} = \beta_0 + u_j$$

The Benjamini-Hochberg procedure for controlling the False Discovery Rate (FDR) (Benjamini & Hochberg, 1995) was implemented to correct p-values for three tests using the p.adjust function in R (R Core Team, 2017) with option “BH”. The p.adjust function performs the Benjamini-Hochberg correction in 5 steps: (1) the p-values are ordered from largest to smallest, (2) the number of p-values (n) is divided by the corresponding p-value rank from smallest to largest (i), (3) the p-values are multiplied by this number (p * n/i), (4) the cumulative minimum is taken of the resulting series so that each p-value is smaller than the one that immediately precedes it, (5) any p-values above 1 are set to 1.

FDR corrected confidence intervals were constructed by recalculating standard errors as a function of the FDR corrected p value. An explanation of this method follows.

B divided by the standard error of B estimates the test statistic t (Equation 5).

Equation 5. Formula for t as a function of B and SE.

$$t = \frac{B}{SE}$$

The test statistic t can also be derived from a p value: a p value is the proportion of the t distribution beyond the quantile given in test statistic t (Equation 6).

Equation 6. Formula for t as a function of p.

$$t = tquantile(p)$$

Equation 7 expresses the p value as a function of B and the standard error of B, implied by Equation 5 and Equation 6.

Equation 7. Implied equivalence from Equation 5 and Equation 6.

$$tquantile(p) = \frac{B}{SE}$$

The terms in Equation 7 can be rearranged by multiplying both sides of Equation 7 by the standard error of B to express B as a function of the standard error of B and p (Equation 8).

Equation 8. B as a function of the standard error of B and p.

$$B = tquantile(p)SE$$

The terms in Equation 8 can be further rearranged by dividing both sides of Equation 8 by tquantile(p), providing a formula for the standard error of B as a function of B and the p value (Equation 9).

Equation 9. SE as a function of B and p.

$$SE = \frac{B}{tquantile(p)}$$

FDR corrected p values are used in place of the original p values in equation 9 to calculate FDR adjusted standard errors. The FDR adjusted standard errors are then used to create FDR adjusted 95% confidence intervals.

Model fit indices including the Root Mean Square Error (RMSE), Conditional pseudo-R^2, Marginal pseudo-R^2, AIC, and BIC were estimated using the performance package (Lüdecke et al., 2021) in R (R Core Team, 2017). RMSE is the standard deviation of the model residuals and provides an estimate of the average distance between observed data and model predicted values; smaller values indicate better model fit. The performance package provides the pseudo-R^2 metric described in Nakagawa & Schielzeth (2013) for linear mixed models. This pseudo R^2 metric is referred to simply as R^2 throughout the remainder of this work. Conditional R^2 is the total variance explained by fixed and random effects. Marginal R^2 is the variance explained by only the fixed effects. AIC and BIC are statistics that penalize model fit for the number of estimated parameters; smaller values indicate better model fit. Both statistics are assessed here because BIC penalizes model complexity more severely than AIC, providing a more rigorous test of systematic differences in model fit. Marginal change R^2 values were calculated for each PGS by subtracting the marginal R^2 from a model that includes the PGS and other terms from the marginal R^2 from a model that includes only the other terms. This procedure was used to calculate PGS marginal change R^2 above (1) just the covariates and (2) the covariates and the other PGS.

A series of likelihood ratio test were used to determine if including the residual PGS and/or the common factor PGS as predictors improves model fit. Four models were considered:

Base Model: Phenotype ~ Covariates

Common Factor PGS Model: Phenotype ~ Covariates + Common Factor PGS

Residual PGS Model: Phenotype ~ Covariates + Residual PGS

Both PGS Model: Phenotype ~ Covariates + Residual PGS + Common Factor PGS

Model comparisons included (1) Base Model vs Common Factor PGS Model, (2) Base Model vs Residual PGS Model, and (3) Common Factor PGS Model vs Both PGS Model. The comparison of (1) Base Model vs Common Factor PGS Model indicates whether inclusion of the common factor PGS improves model fit. The comparison of (2) Base Model vs Residual PGS indicates whether inclusion of the residual PGS improves model fit. The comparison of (3) Common Factor PGS Model vs Both PGS Model indicates whether inclusion of the residual PGS improves model fit above the contribution of the common factor PGS.

Contrasts were drawn between the effect size for the residual PGS and the common factor PGS using two approaches. First, 95% confidence intervals were compared between the effect sizes as an initial indicator of whether the effect sizes are different. Next, a z-test of the null hypothesis that two regression coefficients are equal was conducted (R. Paternoster et al., 1998). The formula for the z statistic in this null hypothesis test is included below, where a z statistic greater than 1.96 or less than -1.96 indicates a significant difference in B_1 and B_2 before Bonferroni correction. After Bonferroni correction for 3 tests, a z statistic greater than 2.40 or less than -2.40 indicates a significant difference in B_1 and B_2.

Equation 10. z-test for H0 $B_1 = B_2$.

$$z = \frac{B_1 - B_2}{\sqrt{var(B_1) + var(B_2) - 2cov(B_1, B_2)}}$$

Power Analysis

GWAS require large sample sizes and statistical power is limited in the current work. A simulation was conducted to determine the sample size required to achieve adequate statistical power for GWAS discovery in the adolescence developmental period under different assumptions regarding SNP MAF and effect size given the observed level of genetic correlation between developmental periods. First, early adult and adult variables were simulated from a multivariate normal distribution with mean equal to 0 and correlation equal to the observed point estimate of the genetic correlation between early adult and adult alcohol use frequency using the mvtnorm package (Genz et al., 2021; Genz & Bretz, 2009) in R (R Core Team, 2017). A SNP variable was simulated from a binomial distribution with size set to 2 to model a diploid genotype and probability equal to varying values for MAF (.01, .10, .20, .30, .40, .50). The adolescence variable was simulated from a normal distribution with the mean and variance conditioned on the early adult variable, the adult variable, and the SNP. The effects of the early adult variable and the adult variable on the adolescence sampling distribution were set to the corresponding observed genetic correlation point estimates. Two values were tested for the effect of the SNP on the adolescence distribution: $\beta=0.01$ and = $\beta=0.05$. These effect sizes are comparable to those observed for significant SNPs in a recent GWAS of alcohol consumption (Liu et al., 2019). As a reference point, the effect of the well-replicated *ADH1B* SNP rs1229984 on alcohol consumption in Liu et al. (2019) was $\beta=0.06$. The variance of the adolescence variable distribution was set to the square root of 1 minus the sum of squared effects on the adolescence indicator. Sensitivity analyses using positive input correlations for the data generation model were also conducted to determine if the observed negative genetic correlations between adolescence and later developmental periods had an effect on statistical power. Figure

S11 displays a diagram of the data generation model. Simulated variables are marked with a blue outline.

These four manifest indicators were used to construct a simulation model using the lavaan package (Rosseel, 2012) in R (R Core Team, 2017). The variance of the SNP was set using the formula for the variance of a SNP: 2MAF(1-MAF). A common factor was fit to the simulated adolescence, early adult, and adult variables by estimating all factor loadings and setting the variance of the common factor to 1. Residual variances for each of the simulated variables were freely estimated. The adolescence variable and the common factor were regressed on the SNP simultaneously. The simulation was repeated 1,000 times for each combination of the following parameters: MAF = .01, .10, .20, .30, .40, .50; SNP β=0.01, 0.05; N = 1k, 10k, 50k, 100k, 200k, 300k, 400k, 500k. Power for each combination of parameters was estimated at two different thresholds ($p < 5e\text{-}8$ and $p < 1e\text{-}5$) as the proportion of SNP p-values less than the threshold. The threshold $p < 5e\text{-}8$ represents the common criteria for genome-wide significance (Z. Chen et al., 2021; Risch & Merikangas, 1996), and the threshold $p < 1e\text{-}5$ represents the criteria for a "suggestive" effect (Guo et al., 2017; Lander & Kruglyak, 1995). Figure S12 displays a diagram of the simulation analysis model. Simulated variables are marked with a blue outline.

An additional simulation was conducted to determine the sample size required to achieve adequate statistical power for the PGS prediction analyses under different assumptions regarding PGS effect size. A PGS variable was simulated from a normal distribution. The target phenotype variable was simulated from a normal distribution with the mean and variance conditioned on the PGS. The effect of the PGS on the target phenotype was set to varying values to model a PGS that accounts for .5%, 1%, and 2% of the phenotype variance (R^2= .005, .01, .02). The variance

of the phenotype distribution was set to 1 minus the value of PGS R^2. The target phenotype variable was regressed on the PGS. The simulation was repeated 10,000 times for each combination of the following parameters: PGS R^2 =.005, .01, .02; N = 100, 250, 500, 750, 1000, 1250, 1500, 1750, 2000. Power for each combination of parameters at $\alpha = .05$ was estimated as the proportion of PGS p-values less than .05. An additional power calculation was conducted to estimate power after applying a Bonferroni correction for 3 tests ($p < .05/3$).

Results

Descriptive Statistics

In what follows I present an overview of nominal differences in average drinking days per year in each developmental period of each discovery sample as a function of sex, age, and birth cohort. Complete results are presented in text and in Table S2. Further interpretation of these results is included in the Discussion section.

Adolescence

In COGA, the overall mean drinking days per year among participants age 12-17 (n=1118) was 36.15 (SE=2.81). The mean among males was higher than females (Male n=562, M=40.13, SE=2.36; Female n=556, M=32.13, SE=2.15). Means increased with age (Age 12 n=17, M=23.91, SE=10.96; Age 13 n=41, M=16.50, SE=7.06; Age 14 n=98, M=24.80 SE=4.91; Age 15 n=179, M=33.58, SE=3.40; Age 16 n=356, M=36.74, SE=3.08; Age 17 n=427, M=41.72, SE=2.62). Participants born in the 80s reported higher means than participants born in the 70s, 90s, or 2000s (70s n=188, M=34.30, SE=5.21; 80s n=473, M=46.81, SE=2.42; 90s n=431, M=26.31, SE=2.12; 2000s n=26 M=18.87, SE= 3.83).

In Add Health, the overall mean drinking days per year among participants age 12-17 (n=3089) was 36.02 (SE=1.17). The mean among males was higher than females (Male n=1430,

M=40.42, SE=1.86; Female n=1659, M=32.23, SE=1.48). Means increased with age (Age 12 n=23, M=20.20, SE=9.76; Age 13 n=213, M=23.28, SE=3.88; Age 14 n=444, M=28.95 SE=3.04; Age 15 n=697, M=34.95, SE=2.51; Age 16 n=752, M=40.34, SE=2.46; Age 17 n=960, M=39.89, SE=2.08). Participants born in the 80s reported nominally higher means than participants born in the 70s (70s n=1852, M=38.01, SE=1.49; 80s n=1237, M=33.05, SE=1.89).

In ALSPAC the overall mean drinking days per year among participants age 12-17 (n=5382) was 54.48 (SE=0.74). The mean among males was slightly higher than females (Male n=2499, M=57.34, SE=1.15; Female n=2883, M=52.00, SE=0.95). Means also increased with age (Age 12.5 n=276, M=21.15, SE=2.49; Age 13.5 n=312, M=13.64, SE=1.16; Age 15.5 n=1180, M=58.99, SE=1.61; Age 16 n=2220, M=50.49, SE=1.06; Age 17.5 n=1394, M=72.76, SE=1.55). All participants from ALSPAC were born in the 90s.

Early Adulthood

In COGA, the overall mean drinking days per year among subjects age 18-25 (n=2762) was 96.45 (SE=1.75). The mean among males was higher than females (Male n=1303, M=117.40, SE=2.77; Female n=1459, M=77.74, SE=2.09). Means generally increased with age (Age 18 n=306, M=77.10, SE=4.98; Age 19 n=289, M=94.25, SE=5.25; Age 20 n=385, M=82.63, SE=4.30; Age 21 n=436, M=94.64, SE=4.15; Age 22 n=440, M=104.59, SE=4.56; Age 23 n=331, M=107.89, SE=5.36; Age 24 n=317, M=102.19, SE=5.37; Age 25 n=258, M=109.89, SE=6.09). Participants born in the 80s reported higher means than subjects born in the 60s, 70s, 90s, or 2000s (60s n=161, M=83.33, SE=8.53; 70s n=864, M=91.30, SE=3.12; 80s n=1020, M=102.52, SE=2.87; 90s n=707, M=98.09, SE=3.35; 2000s n=10, M=16.70, SE= 7.01).

In Add Health, the overall mean drinking days per year among subjects age 18-25 (n=3993) was 64.82 (SE=1.30). The mean among males was higher than females (Male n=1825, M=86.31, SE=2.21; Female n=2168, M=46.73, SE=1.38). Means generally increased with age (Age 18 n=23, M=40.04, SE=6.69; Age 19 n=326, M=67.17, SE=4.72; Age 20 n=543, M=59.66, SE=3.44; Age 21 n=723, M=67.23, SE=2.98; Age 22 n=722, M=66.63, SE=3.10; Age 23 n=723, M=66.29, SE=3.06; Age 24 n=615, M=62.44, SE=3.38; Age 25 n=318, M=64.71, SE=4.56). Participants born in the 70s and 80s reported similar means (70s n=2366, M=64.76, SE=1.70; 80s n=1627, M=64.91, SE=2.01).

In ALSPAC, the overall mean drinking days per year among subjects age 18-25 (n=3613) was 84.46 (SE=1.04). The mean among males was higher than females (Male n=1418, M=94.86, SE=1.73; Female n=2195, M=77.74, SE=1.29). Means were similar at age 18 and 20 (Age 18 n=1957, M=85.58, SE=1.40; Age 20 n=1656, M=83.14, SE=1.57). All participants from ALSPAC were born in the 90s.

Adulthood

In COGA, the overall mean drinking days per year among subjects age 26+ (n=5255) was 98.41 (SE=1.67). The mean among males was higher than females (Male n=2428, M=134.32, SE=2.76; Female n=2827, M=67.57, SE=1.83). Means by age, presented here in decile groupings because of the large range of ages included in the sample, peaked in the mid-40s (Age 26-27 n=526, M=88.35, SE=4.21; Age 28-29 n=526, M=94.97, SE=4.56; Age 30-31 n=526, M=115.29, SE=5.11; Age 32-34 n=526, M=106.34, SE=5.36; Age 35-38 n=526, M=115.47, SE=5.60; Age 39-41 n=525, M=102.05, SE=5.49; Age 42-45 n=525, M=111.07, SE=5.74; Age 46-51 n=525, M=89.57, SE=5.41; Age 52-60 n=525, M=85.70, SE=5.52; Age >60, n=525, M=75.28, SE=5.48). Participants born in the 80s reported higher means than subjects born before

1950, the 50s, 60s, 70s, 90s (Before 1950 n=1463, M=82.83, SE=3.32; 50s n=1302, M=105.33, SE=3.62; 60s n=1344, M=108.97, SE=3.29; 70s n=400, M=81.16, SE=4.62; 80s n=590, M=112.72, SE=4.27; 90s n=156, M=85.99, SE=7.73).

In Add Health, the overall mean drinking days per year among subjects age 26+ (n=5149) was 86.65 (SE=1.43). The mean among males was higher than females (Male n=2421, M=105.13, SE=2.27; Female n=2728, M=70.24, SE=1.75). Means by age peaked in the late 30s (Age 26 n=442, M=78.61, SE=4.30; Age 27 n=517, M=83.19, SE=4.23; Age 28 n=620, M=77.52, SE=3.80; Age 29 n=589, M=69.67, SE=3.56; Age 30 n=583, M=76.08, SE=3.92; Age 31 n=385, M=78.39, SE=5.22; Age 32 n=55, M=82.18, SE=13.84; Age 33 n=23, M=72.59, SE=22.73; Age 34 n=147, M=84.58, SE=8.37; Age 35 n=266, M=92.46, SE=6.84; Age 36 n=299, M=110.16, SE=6.61; Age 37 n=338, M=115.09, SE=6.49; Age 38 n=340, M=102.65, SE=6.29; Age 39 n=279, M=105.70, SE=6.98; Age 40 n=194, M=100.98, SE=8.09; Age 41 n=65, M=86.15, SE=12.79; Age 42 n=6, M=20.50, SE=12.40; Age 43 n=1). Participants born in the 80s reported slightly higher means than subjects born in the 70s (70s n=3124, M=84.70, SE=1.83; 80s n=2025, M=89.65, SE=2.29).

Offspring in ALSPAC are younger than age 26. ALSPAC mothers, who were assessed postpartum and report systematically lower levels of alcohol use, were not included in the analysis.

Aim 1 Results, Developmentally-Informative GWAS

Within-Sample GWAS and GWAS Meta-Analysis

What follows is a brief review of the GWAS results from each of the contributing discovery samples and the GWAS meta-analysis of within sample GWAS at each developmental period. After quality control, 7,057,181 SNPs were included in COGA, 6,520,802 SNPs were

included in Add Health, and 9,283,017 SNPs were included in ALSPAC. A small number of SNPs in COGA and ALSPAC did not run in the within-sample GWAS and were removed before meta-analysis (COGA adolescence 171 SNPs; COGA early adulthood 8 SNPs; ALSPAC adolescence 3 SNPs; ALSPAC early adulthood 207 SNPs). The meta-analysis in adolescence included a total of 10,590,367 SNPs. The meta-analysis in early adulthood included a total of 10,590,325 SNPs. The meta-analysis of COGA and Add Health in adulthood included a total of 7,558,939 SNPs. Manhattan plots were constructed for each GWAS to assess SNP effects in the within-sample GWAS and GWAS meta-analysis (Figure S13 – S23). The common threshold for genome-wide significance is P<5e-8 (Z. Chen et al., 2021; Risch & Merikangas, 1996). An additional significance threshold that is sometimes used to identity a tentative "suggestive" effect is P<1e-5 (Guo et al., 2017; Lander & Kruglyak, 1995). Quantile-quantile plots for each within-sample GWAS and GWAS meta-analysis were constructed to evaluate p-value inflation (Figure S24 – S34). Within-sample GWAS and meta-analysis GWAS quantile-quantile plots in adolescence demonstrated evidence of possible p-value inflation. Note that this inflation was not observed in the final GWAS results for adolescence after genomic control was applied.

In COGA adolescence, 12 SNPs met the threshold for genome-wide significance and an additional 248 SNPs met the suggestive threshold. In COGA early adulthood, zero SNPs met the threshold for genome-wide significance and 65 SNPs met the suggestive threshold. In COGA adulthood, zero SNPs met the threshold for genome-wide significance and 74 SNPs met the suggestive threshold.

In Add Health adolescence, 3 SNPs met the threshold for genome-wide significance and an additional 276 SNPs met the suggestive threshold. In Add Health early adulthood, zero SNPs met the threshold for genome-wide significance and 39 SNPs met the suggestive threshold. In

Add Health adulthood, zero SNPs met the threshold for genome-wide significance and 117 SNPs met the suggestive threshold.

In ALSPAC adolescence, 2 SNPs met the threshold for genome-wide significance and an additional 216 SNPs met the suggestive threshold. In ALSPAC early adulthood, zero SNPs met the threshold for genome-wide significance and 102 SNPs met the suggestive threshold.

In the adolescence meta-analysis, 15 SNPs met the threshold for genome-wide significance and an additional 271 SNPs met the suggestive threshold. In the early adulthood meta-analysis, zero SNPs met the threshold for genome-wide significance and 113 SNPs met the suggestive threshold. In the adulthood meta-analysis, zero SNPs met the threshold for genome-wide significance and 44 SNPs met the suggestive threshold.

Linkage Disequilibrium Score Regression (LDSC)

1,193,613, 1,193,617, and 1,170,827 HapMap3 SNPs were included in the LD score regression after matching meta-analyzed summary statistics to the reference panel and LD scores for adolescence, early adulthood, and adulthood, respectively. Effective sample sizes for adolescence, early adulthood, and adulthood were approximately 8,869, 9,647, and 9,894. Univariate and cross-trait LD score regression intercepts, estimates of SNP-based heritability, genetic covariances, and genetic correlations are presented in Table 1. All univariate LD score regression intercepts were near one. SNP-based heritability for each trait was modest (Adolescence H^2_{SNP} = .04, SE=0.04; Early Adulthood H^2_{SNP} = .05, SE = 0.05; Adulthood H^2_{SNP} = .08, SE = 0.04). Genetic correlations between adolescence and early adulthood (rG = -.27, SE = 0.79) and adolescence and adulthood (rG = -.34, SE = 0.65) were negative. The genetic correlation between early adulthood and adulthood (rG = .75, SE = 0.58) was positive. Notably,

all estimates of H^2_{SNP} were small, indicating that the genetic correlations reported here account for only a small proportion of the overall relationship between these phenotypes.

Table 1

Summary of LD-Score Regression Results for Alcohol Use Frequency

LD-Score Regression Intercepts			
	Adolescence	Early Adult	Adult
Adolescence	1.00	0.15	0.02
Early Adult	0.15	1.00	0.17
Adult	0.02	0.17	1.00
Heritability and Genetic Covariance			
	Adolescence	Early Adult	Adult
Adolescence	.04 (0.04)	-.01 (0.03)	-.02 (0.03)
Early Adult	-.01 (0.03)	.05 (0.05)	.05 (0.03)
Adult	-.02 (0.03)	.05 (0.03)	.08 (0.04)
Genetic Correlations			
	Adolescence	Early Adult	Adult
Adolescence	1	-.27 (0.79)	-.34 (0.65)
Early Adult	-.27 (0.79)	1	.75 (0.58)
Adult	-.34 (0.65)	.75 (0.58)	1

Note. Standard errors for parameter estimates are provided in parentheses.

Genomic Structural Equation Modeling (gSEM)

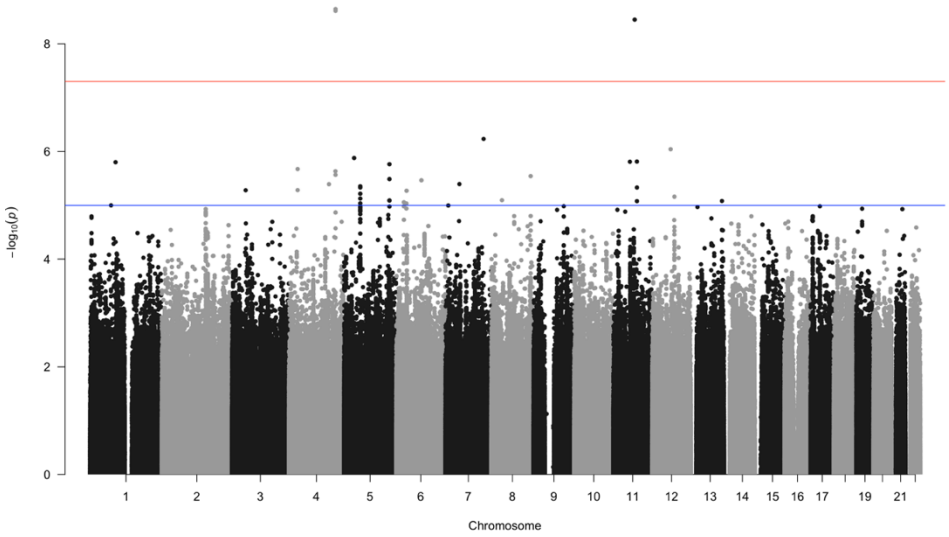
Model identification was achieved by fixing the variance of the common factor (η_{common}) to one. With three indicators the genomic structural equation model is saturated and represents a simple recapitulation of the genetic correlations described above. Loadings of genetic variance in each developmental period on the common factor were: adolescence $\lambda_{12\text{-}17}$ = .35 (SE = 0.64), early adulthood $\lambda_{18\text{-}25}$ = -.78 (SE = 1.12), adulthood λ_{26+} = -.97 (SE = 1.36). Residual genetic variances in each developmental period after accounting for the common factor were: adolescence $U_{12\text{-}17}$ = .88 (SE = 1.21), early adulthood $U_{18\text{-}25}$ = .40 (SE = 2.11), adulthood U_{26+} = .07 (SE = 2.61). The effective sample size of the common factor was approximately 12,180.

Genomic Structural Equation Model GWAS

Manhattan plots and quantile-quantile plots were constructed for each GWAS to assess SNP effects on genetic variance in each developmental period and on the common factor, included below in Figures 7-14. The quantile-quantile plots provide evidence of modest deflation in adulthood and the common factor GWAS. In adolescence, three SNPs met the threshold for genome-wide significance (rs116734991, rs115778926, rs117048287) and an additional 32 SNPs met the suggestive threshold. No SNPs reached genome-wide significance in the other GWAS. In early adulthood, 67 SNPs met the suggestive threshold. In adulthood, 38 SNPs met the suggestive threshold. In the common factor GWAS, 30 SNPs met the suggestive threshold. Lists of all SNPs meeting these thresholds in adolescence, early adulthood, adulthood and the common factor are available in Table S3, Table S4, Table S5, and Table S6.

Figure 7

Manhattan plot of the GWAS of Alcohol Use Frequency in Adolescence.



Note. The genome-wide significance (p<5e-8) is marked in red. The threshold for suggestive significance (p<1e-5) is marked in blue.

Figure 8

Quantile-quantile plot of the GWAS of Alcohol Use Frequency in Adolescence.

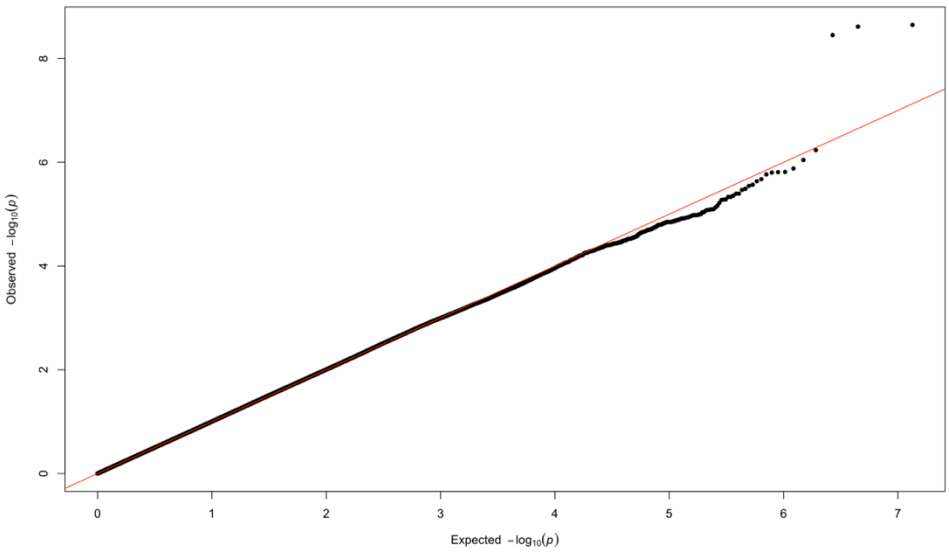
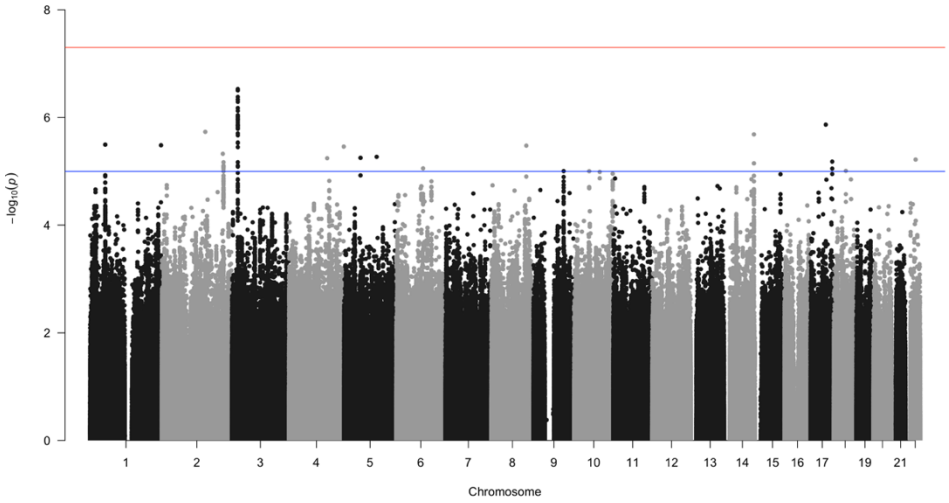


Figure 9

Manhattan plot of the GWAS of Alcohol Use Frequency in Early Adulthood.



Note. The genome-wide significance threshold (p<5e-8) is marked in red. The threshold for suggestive significance (p<1e-5) is marked in blue.

Figure 10

Quantile-quantile plot of the GWAS of Alcohol Use Frequency in Early Adulthood.

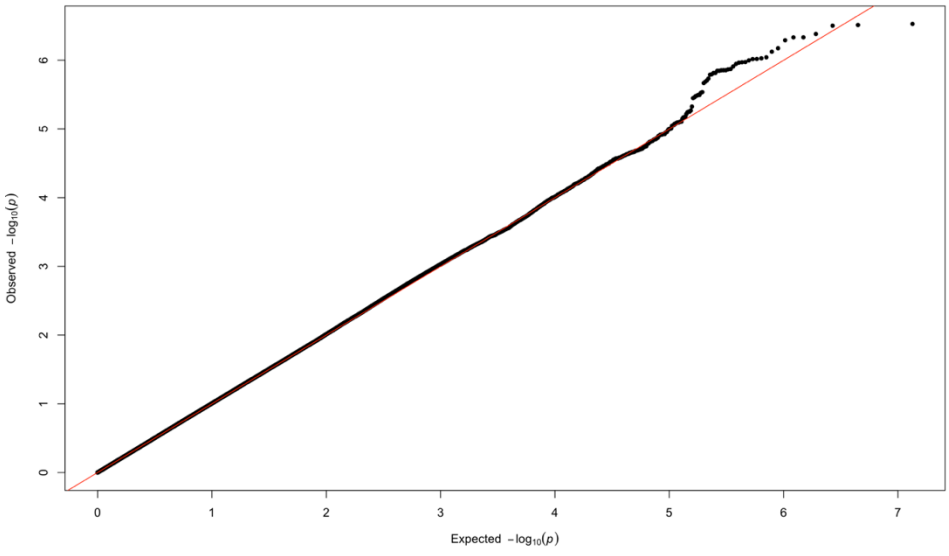
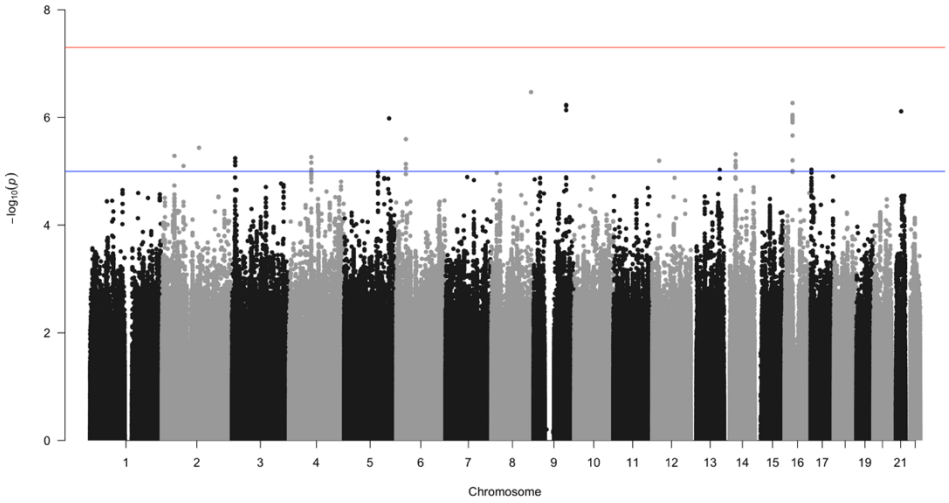


Figure 11

Manhattan plot of the GWAS of Alcohol Use Frequency in Adulthood.



Note. The genome-wide significance threshold (p<5e-8) is marked in red. The threshold for suggestive significance (p<1e-5) is marked in blue.

Figure 12

Quantile-quantile plot of the GWAS of Alcohol Use Frequency in Adulthood.

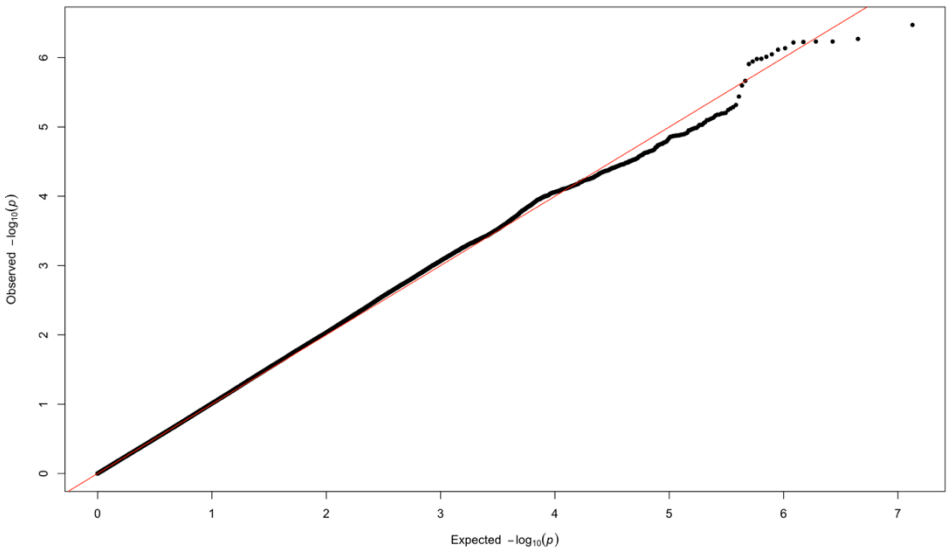
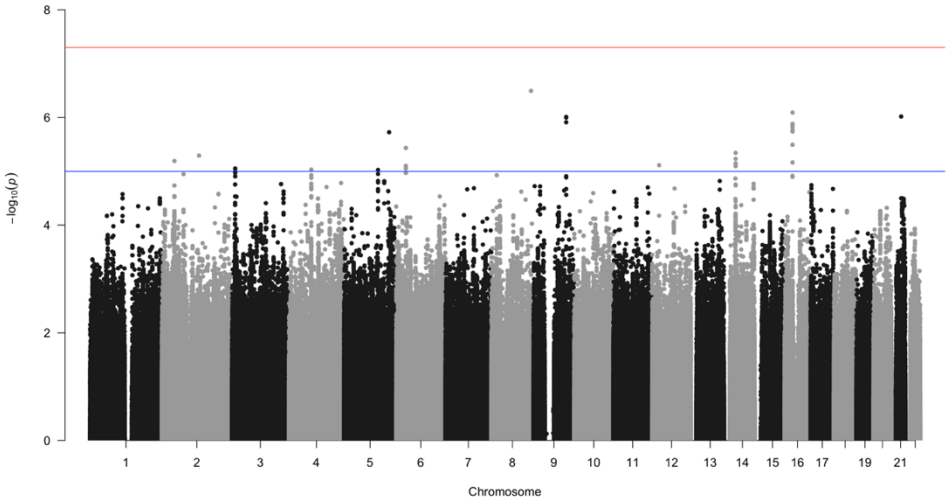


Figure 13

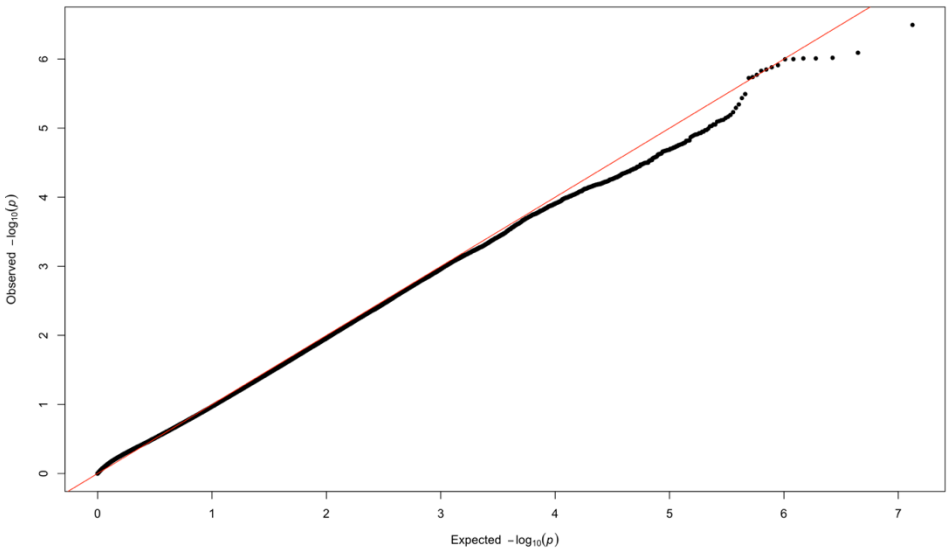
Manhattan plot of the GWAS of the Alcohol Use Frequency Common Factor



Note. The genome-wide significance threshold (p<5e-8) is marked in red. The threshold for suggestive significance (p<1e-5) is marked in blue.

Figure 14

Quantile-quantile plot of the GWAS of the Alcohol Use Frequency Common Factor.



Additional analyses were conducted to examine the three SNPs that met the threshold for genome-wide significance in adolescence (rs116734991, rs115778926, rs117048287). Positional mapping in FUMA indicates that rs116734991 and rs115778926 are intergenic SNPs. The closest gene to rs116734991 and rs115778926 is *LINC02477* (long intergenic non-protein coding RNA 2477). These SNPs are located 31,794 and 6,383 base pairs from *LINC02477*, respectively. *LINC02477* has been mapped in previous GWAS of educational attainment (Okbay et al., 2022) and major depressive disorder (Giannakopoulou et al., 2021). The other significant SNP in adolescence , rs117048287, is located in the intron of *POLD3* (DNA polymerase delta subunit 3). *POLD3* has not been mapped in previous GWAS of psychiatric or behavioral phenotypes, but has appeared in GWAS of various physiological traits, including height (Kichaev et al., 2019) and weight (Sakaue et al., 2021).

Plots of local LD patterns for the three SNPS that met the threshold for genome-wide significance in adolescence (rs116734991, rs115778926, rs117048287) are included in Figure S35 – S37. This analysis revealed that two of the significant SNPs, rs116734991 and rs115778926, are in strong LD with each other (R^2=.86). No other SNPs were flagged as proxy SNPs at the R^2 threshold of greater than or equal to .80. The SNPs with highest R^2 value for each significant SNP were: rs115881918, R^2=.43 (for rs116734991); rs77755343, R^2=.49 (for rs115778926); and rs139719129, R^2=.52 (for rs117048287). Annotations from the LDproxy Tool did not provide evidence of a functional or regulatory role for these SNPs and these SNPs are not documented in previous GWAS indexed on dbSNP.

Genetic Correlations with External GWAS

All genetic correlations between gSEM GWAS summary statistics and the external GWAS are reported in Table 2.

Table 2

Genetic Correlations between Total Variance gSEM GWAS Components and External GWAS.

	Adolescence	Early Adulthood	Adulthood	Common Factor
Drinks per Week	.27 [-0.06, 0.60]	.80 [0.06, 1.54]	.84 [0.36, 1.31]	.88 [0.47, 1.28]
AUDIT-P	.29 [-0.16, 0.73]	.35 [-0.13, 0.82]	.54 [0.15, 0.93]	.50 [0.19, 0.82]
Alcohol Use Frequency	-.01 [-0.36, 0.35]	.89 [0.02, 1.76]	.73 [0.29, 1.17]	.89 [0.46, 1.31]
Alcohol Use Quantity	.25 [-0.20, 0.69]	.11 [-0.29, 0.51]	.58 [0.18, 0.98]	.45 [0.14, 0.76]
Binge Drinking Frequency	.23 [-0.19, 0.65]	.54 [-0.05, 1.13]	.83 [0.33, 1.32]	.79 [0.39, 1.19]
Alcohol Dependence	-.18 [-0.79, 0.44]	-.18 [-0.81, 0.44]	.37 [-0.09, 0.82]	.27 [-0.18, 0.72]
Externalizing Factor	.24 [-0.08, 0.57]	-.05 [-0.37, 0.28]	.23 [0.03, 0.43]	.15 [-0.02, 0.33]
Risk Tolerance	.35 [-0.07, 0.77]	.11 [-0.22, 0.45]	-.03 [-0.23, 0.16]	-.03 [-0.22, 0.17]
Cigarettes per Day	.05 [-0.22, 0.32]	-.36 [-0.77, 0.05]	-.20 [-0.44, 0.03]	-.27 [-0.51, -0.03]
Educational Attainment	-.02 [-0.21, 0.18]	.40 [0.01, 0.80]	.18 [-0.00, 0.36]	.26 [0.08, 0.45]
Townsend Deprivation Index	.14 [-0.20, 0.47]	-.06 [-0.36, 0.23]	.09 [-0.16, 0.35]	.02 [-0.22, 0.26]
Average Household Income	.02 [-0.23, 0.27]	.48 [-0.02, 0.97]	.07 [-0.14, 0.28]	.17 [-0.03, 0.38]
Major Depressive Disorder	.14 [-0.12, 0.40]	-.19 [-0.51, 0.14]	-.03 [-0.22, 0.17]	-.11 [-0.30, 0.07]
Neuroticism	-.25 [-0.57, 0.07]	-.05 [-0.43, 0.32]	.01 [-0.16, 0.19]	-.00 [-0.17, 0.17]
Anxiety Disorder	.07 [-0.28, 0.42]	-.42 [-0.94, 0.10]	-.05 [-0.34, 0.24]	-.18 [-0.45, 0.10]

Note. Adolescence, Early Adulthood, and Adulthood refer to the Genomic Structural Equation Model GWAS results using the total variance parameterization of the model. 95% Confidence intervals for parameter estimates are provided in square brackets. References for each external GWAS are included in Table S7. Abbreviations: AUDIT-P = Alcohol Use Disorder Identification Test, Problems Subscale.

Description of statistically significant genetic correlations is included below, followed by a brief overview of nominal differences in genetic correlations across development. Discussion of nominal difference in genetic correlations are grouped under four categories: alcohol use behaviors, other externalizing behaviors, socioeconomic indicators, and internalizing behaviors. Sample sizes and citations for the external GWAS are included in Table S7.

No significant associations were observed with alcohol use frequency in adolescence. Drinks per week (rG=.80, 95% CI [0.06, 1.54]), alcohol use frequency (rG=.89, 95% CI [0.02,1.76]), and educational attainment (rG=.40, 95% CI [0.01, 0.80]) were significantly associated with alcohol use frequency in early adulthood. Drinks per week (rG=.84, 95% CI [0.36, 1.31]), AUDIT-P (rG=.54, 95% CI [0.15, 0.93]), alcohol use frequency (rG=.73, 95% CI [.29, 1.17]), alcohol use quantity (rG=.58, 95% CI [0.18, 0.98]), binge drinking frequency (rG=.83, 95% CI [0.33, 1.32]), and the externalizing common factor (rG=.23, 95% CI [0.03, 0.43]) were significantly associated with alcohol use frequency in adulthood. Drinks per week (rG=.88, 95% CI [0.47, 1.28]), AUDIT-P (rG=.50, 95% CI [0.19, 0.82]), alcohol use frequency (rG=.89, 95% CI [0.46, 1.31]), alcohol use quantity (rG=.45, 95% CI [0.14, 0.76]), binge drinking frequency (rG=.79, 95% CI [0.39, 1.19]), cigarettes per day (rG=-.27, 95% CI [-0.51, -0.03]) and educational attainment (rG=.26, 95% CI [0.08, 0.45]) were significantly associated with the alcohol use frequency common factor.

The following section discusses nominal differences in genetic correlations across development. It should be noted that this discussion focuses on point estimates and many of these genetic correlations are not statistically different than zero. Replication in larger samples is necessary before substantive interpretation of these trends.

Genetic correlations between adult alcohol use behaviors in external samples and the development-matched components of the gSEM model for alcohol use frequency generally increased across development. Genetic correlations in adolescence were modest in magnitude, ranging from -.18 to .29. Genetic correlations in early adulthood were larger, ranging from -.18 to .89. The largest genetic correlations with adult alcohol use behaviors were observed with the alcohol use frequency in adulthood and the common factor. Genetic correlations in adulthood ranged from .37 to .84. Genetic correlations with the common factor ranged from .45 to .89.

Genetic correlations between externalizing behaviors in external samples and the development-matched components of the alcohol use frequency gSEM model were positive in adolescence and either smaller in magnitude or negative in later developmental periods. Genetic correlations in adolescence ranged from .05 to .35. The range of genetic correlations in later developmental periods shifted towards negative and smaller positive values: genetic correlations in early adulthood ranged from -.36 to .11, genetic correlations in adulthood ranged from -.20 to .23, and genetic correlations with the common factor ranged from -.27 to .15.

Genetic correlations between indicators of socioeconomic status and the development-matched components of the alcohol use frequency gSEM model mostly increased across development. Genetic correlations with educational attainment and average household income in adolescence were near 0 (rG=-.02 and rG=.02, respectively). Genetic correlations with educational attainment and average household income in early adulthood were positive (rG=.40 and rG=.48, respectively). In adulthood, genetic correlations with educational attainment and average household income were smaller than those observed in early adulthood, but larger than those observed in adolescence (rG=.18 and rG=.07, respectively). Genetic correlations with the common factor were larger than those observed in adulthood for educational attainment and

average household income (rG=.26 and rG=.17), but still smaller than those observed in early adulthood.

Genetic correlations between internalizing behaviors and the development-matched components of the alcohol use frequency gSEM model did not follow a consistent pattern across development. Genetic correlations with neuroticism were negative in adolescence (rG=-.25) and negative or near 0 in later developmental periods (early adulthood rG=-.05, adulthood rG=.01, common factor rG=.00). Genetic correlations with MDD were positive in adolescence (rG=.14), and negative with a modest range in magnitude in later developmental periods (early adulthood rG=-.19, adulthood rG=-.03, common factor rG=-.11). Genetic correlations with anxiety were positive in adolescence (rG=.07) and negative with a larger range in magnitude in later developmental periods (early adulthood rG=-.42, adulthood rG=-.05, common factor rG=-.18).

Aim 2 Results, Polygenic Scores (PGS)

The Aim 2 hypothesis that the residual PGS would improve phenotype prediction was not supported in adolescence or early adulthood; however, the results provide tentative support for this hypothesis in adulthood. Polygenic scores were constructed in COGA by running a version of the Aim 1 discovery analysis with COGA removed. Univariate and cross-trait LD score regression intercepts, estimates of SNP-based heritability, genetic covariances, and genetic correlations for the LD score regression analysis with COGA removed are presented in Table S8. Sample sizes for the gSEM GWAS analysis in adolescence, early adulthood, adulthood, and the common factor with COGA removed were approximately 7,869, 7,070, 5149, and 8230, respectively.

Manhattan plots were constructed for each GWAS to assess SNP effects in the GWAS meta-analysis and gSEM GWAS with COGA removed (Figure S38 – S44). Quantile-quantile

plots for each GWAS meta-analysis and gSEM GWAS with COGA removed were constructed to evaluate possible p-value inflation (Figure S45 – S51). The quantile-quantile plots provide evidence of inflation in the adolescence meta-analysis and evidence of deflation in the gSEM GWAS.

PGS Descriptive Statistics

Descriptive statistics for phenotypic measures in the COGA analytic sample were reported at the beginning of the Results section under Descriptive Statistics and can be found in Table S2. Briefly here, the sample sizes in COGA for PGS analyses were n=1,118 in adolescence, n=2,762 in early adulthood, and n=5,255 in adulthood. Average drinking days per year in COGA were M=36.15 (SE=1.60) in adolescence, M=96.45 (SE=1.75) in early adulthood, and M=98.41 (SE=1.67) in adulthood.

Each PGS was approximately normally distributed. PGS were scaled to a standard normal distribution with a mean of 0 and a standard deviation of 1 for analysis. Zero-order correlations between residual PGS were moderate in magnitude (Adolescence PGS and Early Adulthood PGS r=.48 , 95% CI [0.47, 0.50] ; Adolescence PGS and Adult PGS r=.44 , 95% CI [0.42, 0.46] ; Early Adult PGS and Adult PGS r= -.33, 95% CI [-0.35, -0.31]). Zero-order correlations between residual PGS and the common factor PGS ranged from small to moderate in magnitude (Adolescence PGS r=.18, 95% CI [0.16, 0.20]; Early Adulthood PGS r=.04, 95% CI [0.02, 0.06]; Adulthood PGS r= .52, 95% CI [0.50, 0.53]). Zero-order correlations between the residual PGS and their respective target phenotypes were not significant, with one exception. The zero-order correlation between the adulthood-specific PGS and alcohol use frequency in adulthood was statistically significant (r= .04, 95% CI [0.01 0.06]). The common factor PGS was also correlated with alcohol use frequency in adulthood (r=.03, 95% CI [0.003, 0.06]).

PGS Validation Analyses

Results from the Aim 2 PGS validation analyses in COGA are presented below. Model fit statistics are provided in Table 3. Likelihood-ratio tests comparing model fit between versions of the model with and without the residual and common factor PGS are included in Table 4. Change R^2 values for the residual and common factor PGS are reported in Table 5. Regression coefficient estimates for the residual PGS, the common factor PGS, and covariates at each developmental period are included in Table 6. All regression coefficient estimates are presented in their original scale of drinking days per year.

Table 3

Model Fit Statistics for PGS Linear Mixed-Effects Models

	N Parameters	Conditional R^2	Marginal R^2	Log Likelihood	AIC	BIC	RMSE
Adolescence							
Base	18	.04	.04	-6015.95	12067.91	12158.26	52.57
Residual PGS	19	.04	.04	-6015.75	12069.49	12164.86	52.56
Common Factor PGS	19	.04	.04	-6015.93	12069.87	12165.23	52.56
Both PGS	20	.04	.04	-6015.74	12071.49	12171.87	52.56
Early Adult							
Base	19	.13	.06	-16309.22	32656.44	32768.99	83.41
Residual PGS	20	.13	.06	-16309.22	32658.44	32776.91	83.40
Common Factor PGS	20	.13	.06	-16307.69	32655.37	32773.85	83.34
Both PGS	21	.13	.06	-16307.67	32657.34	32781.74	83.33
Adult							
Base	19	.24	.17	-32146.07	64330.13	64454.9	103.39
Residual PGS	20	.24	.17	-32142.85	64325.70	64457.04	103.30
Common Factor PGS	20	.24	.17	-32143.88	64327.77	64459.11	103.36
Both PGS	21	.24	.17	-32142.44	64326.88	64464.78	103.30

Note. Abbreviations*:* PGS = Polygenic Score; AIC = Akaike Information Criteria; BIC = Bayesian Information Criteria; RMSE = Root Mean Square Error.

Table 4

Likelihood Ratio Tests for PGS Linear Mixed-Effects Models

		Chi Square	df	p
Adolescence				
	Base VS Common Factor PGS	0.04	1	.836
	Base VS Residual PGS	0.41	1	.520
	Common Factor PGS VS Both PGS	0.38	1	.539
Early Adult				
	Base VS Common Factor PGS	3.07	1	.080
	Base VS Residual PGS	0.00	1	.956
	Common Factor PGS VS Both PGS	0.03	1	.860
Adult				
	Base VS Common Factor PGS	4.36	1	.037
	Base VS Residual PGS	6.43	1	.011
	Common Factor PGS VS Both PGS	2.89	1	.089

Note. Abbreviations: PGS = Polygenic Score

Table 5

PGS Change R2 from Linear Mixed-Effects Models

	Over Covariates	Over Covariates and Other PGS
Adolescence		
Residual PGS	.0004	.0003
Common Factor PGS	.00004	.000006
Early Adult		
Residual PGS	-.000001	.000002
Common Factor PGS	.001	.001
Adult		
Residual PGS	.001	.0005
Common Factor PGS	.0007	.0001

Note. Estimates are rounded to the first non-zero digit. Abbreviations: PGS = Polygenic Score.

Table 6

Regression Coefficient Estimates from PGS Linear Mixed-Effects Models

	B	SE	P	Lower 95%	Upper 95%
Adolescence					
Residual PGS	-1.03	4.23	.808	-9.31	7.25
Common Factor PGS	-0.14	1.60	.932	-3.28	3.00
Birth Year	-1.10	0.27	<.001	-1.63	-0.58
Female	-8.73	3.17	.006	-14.95	-2.51
Early Adult					
Residual PGS	-0.31	1.75	.860	-3.74	3.13
Common Factor PGS	3.06	2.58	.235	-1.99	8.11
Birth Year	0.57	0.23	.013	0.12	1.03
Female	-38.18	3.39	<.001	-44.82	-31.55
Proband	29.95	8.42	<.001	13.46	46.45
Adult					
Residual PGS	3.14	2.82	.267	-2.40	8.67
Common Factor PGS	1.69	2.78	.544	-3.76	7.13
Birth Year	0.71	0.09	<.001	0.53	0.90
Female	-55.02	3.06	<.001	-61.02	-49.02
Proband	95.95	5.03	<.001	86.08	105.82

Note. Reported standard errors, p-values, and 95% confidence intervals are corrected for false discovery rate. Abbreviations: PGS = Polygenic Score; B = Unstandardized regression coefficient; SE = Standard Error.

Adolescence Model Results.

The random effect for family grouping did not account for variance in alcohol use frequency in adolescence when models were fit by maximum likelihood. Maximum likelihood (ML) estimation is known to produce a downwards bias in estimates of random effects variance (Shaw, 1987). Restricted maximum likelihood (REML) produces unbiased estimates of random effects variance, but likelihood ratio tests are not possible with models estimated by REML. Complete results for all Aim 2 analyses fit by REML are available in Table S9 (Model Fit

Statistics), Table S10 (PGS Change R^2), and Table S11 (Fixed Effects). In the REML version of this analysis, random effects and fixed effects together accounted for 5% of the variance in adolescence alcohol use frequency (Conditional R^2=.05). The remainder of the results reported in text are from the ML version of the analysis. The fixed effects accounted for 4% of the variance (Marginal R^2=.04).

Including the common factor PGS as a predictor did not improve model fit significantly relative to a model with just covariates ($\chi2(1)=0.04$, p=.836). Including the residual PGS as a predictor did not improve model fit significantly relative to a model with just covariates ($\chi2(1)=0.41$, p=.520). Including the residual PGS as a predictor also did not improve model fit significantly relative to a model with covariates and the common factor PGS ($\chi2(1)=0.38$, p=.539). Nominal differences in AIC and BIC align with results from the likelihood ratio tests, indicating that the base model (AIC=12,067.91, BIC=12,158.26) provided better fit to the data than the model with the common factor PGS (AIC=12,069.87, BIC=12,165.23), the model with the residual PGS (AIC=12,069.49, BIC=12,164.86), and the model with both PGS (AIC=12,071.49, BIC=12,171.87). RMSE for each model were comparable. The residual PGS accounted for 0.04% of the phenotypic variance above covariates (Marginal Change R^2 = .0004) and 0.03% of the phenotypic variance above covariates and the common factor PGS (Marginal Change R^2 =.0003). The common factor PGS accounted for 0.004% of the phenotypic variance above covariates (Marginal Change R^2 =.00004) and 0.0006% of the phenotypic variance above covariates and the residual PGS (Marginal Change R^2 = .000006). This pattern of results suggests that the common factor PGS and residual PGS did not make a substantial contribution to the prediction of alcohol use frequency in adolescence.

The residual PGS (B=-1.03, SE=4.23 , 95% CI [-9.31, 7.25]) and common factor PGS (B=-0.14, SE=1.60 , 95% CI [-3.28, 3.00]) were not significantly associated with alcohol use frequency in adolescence. The FDR-corrected 95% CI for the effect of the residual PGS and the common factor PGS were not distinct. The z-test for the null hypothesis that the residual PGS B and common factor PGS B are equal was not statistically significant (z=-0.35, p=.725), indicating that the residual PGS did not have a larger effect on alcohol use frequency than the common factor PGS.

Early Adulthood Model Results.

Random effects and fixed effects together accounted for 13% of the variance in early adult alcohol use frequency (Conditional R^2=.13). The fixed effects alone accounted for 6% of the variance (Marginal R^2=.06).

Including the common factor PGS as a predictor did not improve model fit significantly relative to a model with just covariates ($\chi2(1)=3.07$, p=.080). Including the residual PGS as a predictor did not improve model fit significantly relative to a model with just covariates ($\chi2(1)=0.01$, p=.956). Including the residual PGS as a predictor also did not improve model fit significantly relative to a model with covariates and the common factor PGS ($\chi2(1)=0.03$, p=.860). Nominal differences in AIC suggest that the model with the common factor PGS (AIC=32,655.37) provided slightly better fit to the data than the base model (AIC=32,656.44); however, nominal differences in BIC contradict this trend, suggesting that the simpler model provided better fit to the data (base model BIC=32,768.99, common factor PGS BIC=32,773.85). The base model and common factor model both provided better fit to the data than the model with the residual PGS (AIC=32,658.44, BIC=32,776.91), and the model with both PGS (AIC=32,657.34, BIC=32,781.74). RMSE for each model were comparable. The

residual PGS accounted for effectively 0% of the phenotypic variance above covariates (Marginal Change R^2 = -.000001) and 0.0002% of the phenotypic variance above covariates and the common factor PGS (Marginal Change R^2 = .000002). Note that overall model pseudo-R^2 described by Nakagawa & Schielzeth (2013) for linear mixed effects models does not always increase when predictors are added to the model. This can produce negative estimates of change R^2, which can be substantively interpreted as zero. The common factor PGS accounted for 0.1% of the phenotypic variance above covariates (Marginal Change R^2 =.001) and 0.1% of the phenotypic variance above covariates and the residual PGS (Marginal Change R^2 = .001). Similar to the results for adolescence, this pattern of results suggests that the residual PGS did not make a measurable contribution to the prediction of alcohol use frequency in early adulthood. Nominal differences in AIC, but not BIC, provide tenuous evidence for an effect of the common-factor PRS in early adulthood.

The residual PGS (B=-0.31, SE=1.75 , 95% CI [-3.74, 3.13]) and common factor PGS (B=3.06, SE=2.58 , 95% CI [-1.99, 8.11]) were not significantly associated with alcohol use frequency in early adulthood. The FDR-corrected 95% CI for the effect of the residual PGS and the common factor PGS were not distinct. The z-test for the null hypothesis that the residual PGS B and common factor PGS B are equal was not statistically significant (z=-1.32, p=.187), indicating that the residual PGS did not have a larger effect on alcohol use frequency than the common factor PGS.

Adulthood Model Results.

Random effects and fixed effects together accounted for 24% of the variance in adult alcohol use frequency (Conditional R^2=.24). The fixed effects alone accounted for 17% of the variance (Marginal R^2=.17).

Including the common factor PGS as a predictor improved model fit significantly relative to a model with just covariates ($\chi2(1)=4.36$, p=.037). Including the residual PGS as a predictor also improved model fit significantly relative to a model with just covariates ($\chi2(1)=6.43$, p=.011). Including the residual PGS as a predictor did not improve model fit significantly relative to a model with covariates and the common factor PGS ($\chi2(1)=2.89$, p=.089). Nominal differences in AIC suggest that the model with the residual PGS (AIC=64,325.70) provided the best fit to the data, followed by the model with both PGS (AIC=64,326.88), the model with the common factor PGS (AIC=64,327.77), and the base model (AIC=64,330.13). Nominal differences in BIC suggest that the base model provided the best fit to the data (BIC=64,454.90). Among the remaining models, BIC indicated better fit for the model with the residual PGS (BIC=64,457.04), followed by the model with the common factor PGS (BIC=64,459.11), and the model with both PGS (BIC=64,464.78). The residual PGS accounted for 0.1% of the phenotypic variance above covariates (Marginal Change R^2 = .001) and 0.05% of the phenotypic variance above covariates and the common factor PGS (Marginal Change R^2 = .0005). The common factor PGS accounted for 0.07% of the phenotypic variance above covariates (Marginal Change R^2 =.0007) and 0.01% of the phenotypic variance above covariates and the residual PGS (Marginal Change R^2 = .0001). This pattern of results suggests that the residual PGS and common factor PGS each made a measurable contribution to the prediction of alcohol use frequency in adulthood individually, but not when modeled together.

The residual PGS (B=3.14, SE=2.82, 95% CI [-2.40, 8.67]) was not significantly associated with alcohol use frequency in adulthood. The common factor PGS (B=1.69, SE=2.78 , 95% CI [-3.76, 7.13]) was also not significantly associated with alcohol use frequency in adulthood. The FDR-corrected 95% CI for the effect of the residual PGS and the common factor

PGS were not distinct. The z-test for the null hypothesis that the residual PGS B and common factor PGS B are equal was not statistically significant (z=0.45, p=.650).

Note that the zero-order correlations between the residual PGS and adult alcohol use frequency (r=.04, 95% CI [0.01,0.06] and the common factor PGS and adult alcohol use frequency (r=.03, 95% CI [0.003, 0.06]) were both statistically significant. Both PGS also improved model fit above a model with only covariates (residual PGS: $\chi2(1)=6.43$, p=.011, common factor PGS: $\chi2(1)=4.36$, p=.037). As a follow up analysis, coefficient estimates were extracted from the intermediate models that did not include both PGS simultaneously. The association between the common factor PGS and adult alcohol use frequency is nominally positive when modeled separately, but not statistically significant after multiple testing correction (B=3.31, SE=1.59, 95% CI [-0.75, 7.38]. The residual PGS is significantly associated with adult alcohol use frequency when modeled separately from the common factor PGS (B=4.00, SE=1.89, 95% CI [0.31, 7.70]. Regression coefficient estimates from these sensitivity analyses in each developmental period are available in Table S12 and Table S13. Differences in the effect of the PGS when modeled separately were not evident in other developmental periods.

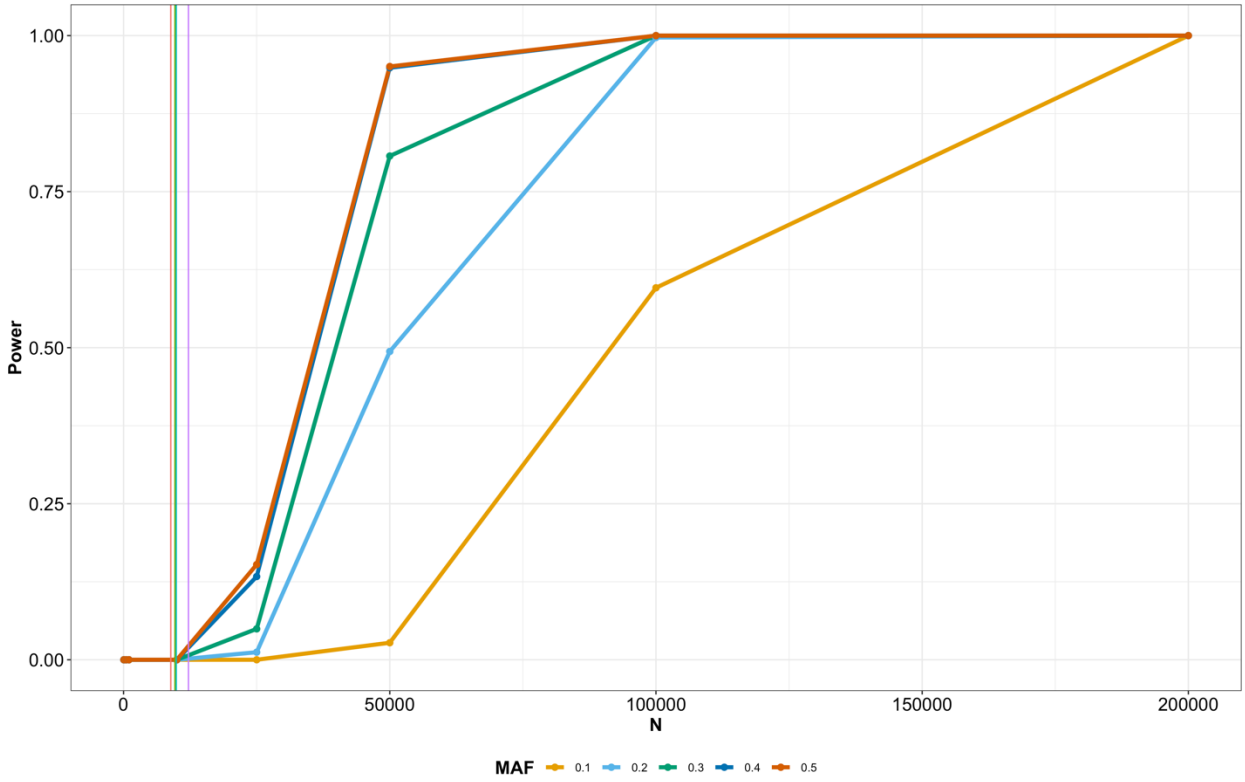
Power Analysis

gSEM GWAS Power Analysis

Manifest indicators were simulated using the observed genetic correlations between adolescence, early adulthood, and adulthood ($rG_{Adol,E.Adult}$ = -.34, $rG_{Adol,Adult}$ = -.27, $rG_{E.Adult,Adult}$=.75). Complete results from the power analysis are presented in table S14 (genome-wide significance threshold) and table S15 (suggestive significance threshold). Power estimates for genome-wide significance ($p < 5e-8$) at varying values of MAF are presented for a SNP with effect size β=0.05 in Figure 12.

Figure 15

Power estimates for genome-wide significance (p < 5e-8) with SNP β=0.05



Note. Vertical lines in the figure demarcate the observed sample sizes in the discovery analysis of the current study: red = adolescence, green=early adulthood, blue = adulthood, purple = common factor.

For the genome-wide significance threshold (p < 5e-8), at least 80% power to detect a SNP with effect size β=0.01 was not achieved under any combination of simulation parameters. For p < 5e-8, MAF=.5 and N=500,000 (the most well-powered combination of parameters) power to detect a SNP with effect size β=0.01 was 9%. At least 80% power to detect a SNP with effect size β=0.05 was achieved for SNPs with MAF = .10 or greater at n=200,000, SNPs with MAF = .20 or greater at n=100,000, and SNPs with MAF=.30 at n=50,000. For the suggestive threshold (p < 1e-5), at least 80% power to detect a SNP with effect size β=0.01 was not

achieved under any combination of simulation parameters. For p < 1e-5, MAF=.50 and N=500,000 power to detect a SNP with effect size β=0.01 was 42% At least 80% power to detect a SNP with effect size β=0.05 was achieved for SNPs with MAF = .10 or greater at n=100,000 and SNPs with MAF = .20 or greater at n=50,000.

Sensitivity analyses using positive input correlations for the data generation model produced comparable power estimates to the version of the analysis with the observed input correlations. For the genome-wide significance threshold (p < 5e-8), at least 80% power to detect a SNP with effect size β=0.01 was not achieved under any combination of simulation parameters. For p < 5e-8, MAF=.50 and N=500,000 (the most well-powered combination of parameters) power to detect a SNP with effect size β=0.01 was 7%. At least 80% power to detect a SNP with effect size β=0.05 was achieved for SNPs with MAF = .10 or greater at n=200,000, SNPs with MAF = .20 or greater at n=100,000, and SNPs with MAF=.30 at n=50,000. For the suggestive threshold (p < 1e-5), at least 80% power to detect a SNP with effect size β=0.01 was not achieved under any combination of simulation parameters. For p < 1e-5, MAF=.50 and N=500,000 power to detect a SNP with effect size β=0.01 was 39% At least 80% power to detect a SNP with effect size β=0.05 was achieved for SNPs with MAF = .10 or greater at n=100,000 and SNPs with MAF = .20 or greater at n=50,000.

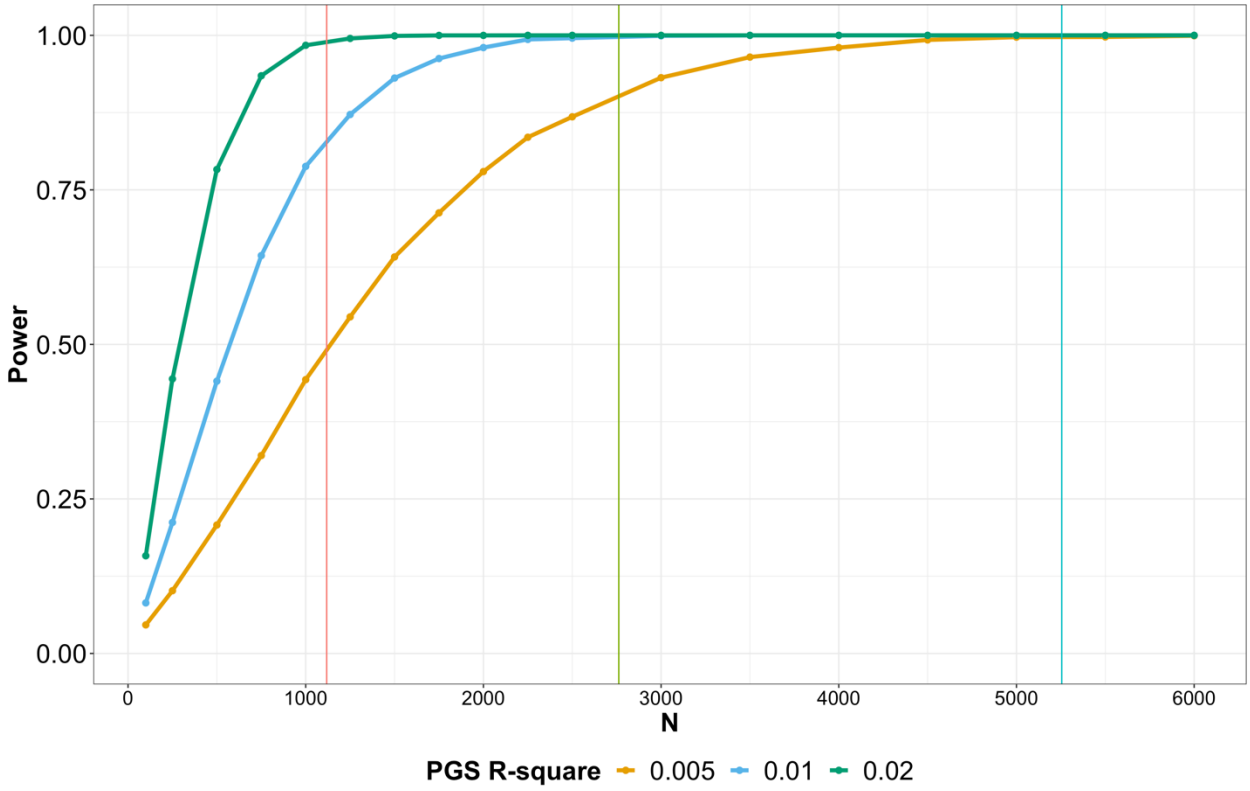
PGS Power Analysis

For p < .05, at least 80% power to detect a PGS effect that accounts for 2% of the target phenotype variance was achieved at n=500. At least 80% power to detect a PGS effect that accounts for 1% of the target phenotype variance was achieved at n=1000. At least 80% power to detect a PGS effect that accounts for 0.5% of the target phenotype variance was achieved at n=1750. For p < .05/3, at least 80% power to detect a PGS effect that accounts for 2% of the

target phenotype variance was achieved at n=750. At least 80% power to detect a PGS effect that accounts for 1% of the target phenotype variance was achieved at n=1250. At least 80% power to detect a PGS effect that accounts for 0.5% of the target phenotype variance was achieved at n=2500. Power estimates with p < .05/3 are presented in Figure 13.

Figure 16

PGS power estimates with Bonferroni correction for 3 tests



Note. Vertical lines in the figure demarcate the observed sample sizes in the current study: red = adolescence, green=early adulthood, blue = adulthood.

Discussion

This work addressed two aims: (Aim 1) to advance gene discovery by building developmentally-informative models for gene identification capable of incorporating developmental changes in alcohol use across time and (Aim 2) to leverage results from the

developmentally-informative GWAS for genetic prediction of age-matched alcohol use outcomes in an independent sample. Small sample sizes and low statistical power limit the substantive conclusions that can be drawn regarding these aims.

This section begins with a discussion of differences in phenotypic means between the studies included in the analysis. In general, mean differences between the studies appear to be attributable to the lower legal drinking age in ALSPAC and the ascertained nature of the COGA sample. The goal of this first section is to provide a rationale for the idea that meaningful phenotypic differences between the discovery samples were adequately accounted for by sex, age, birth cohort, proband status in COGA, or other fundamental aspects of the respective study designs in each sample. This is a critical prerequisite to meta-analysis of the GWAS results. Next, the results from each aim are discussed in turn. An overview of limitations and future directions follows. Lastly, this section closes with a discussion of the overall conclusions that may be drawn from this work.

Interpretation of Sample Mean Differences

These results are presented as an exploration of general trends alcohol use in each sample and to describe the importance of covariates in the inferential analyses, rather than a formal analysis of statistical differences between the samples. This approach of providing a general interpretation of nominal differences is taken to avoid the massive multiple testing burden associated with conducting formal tests of mean differences between all levels of sex, age, and birth cohort within each study. Significance at the corrected alpha level would likely reflect the relative size of the various subsamples, rather than the presence or absence of meaningful mean differences.

In adolescence, the average drinking days per year in ALSPAC was uniformly higher across sex, age, and birth cohort compared to COGA and Add Health. These differences were most pronounced in the ALSPAC age 17.5 age group. This trend likely reflects the lower legal drinking age of 18 (16 if accompanied by a parent), in the United Kingdom, where the ALSPAC sample was collected. The legal drinking age in the United States, where the COGA and Add Health samples were collected, is 21. Previous work demonstrates that earlier legal access to alcohol is associated with increased alcohol use among adolescents (Anderson et al., 2009; Casswell & Zhang, 1997).

In early adulthood, the average drinking days per year in Add Health was uniformly lower across sex, age, and birth cohort. At this point in development, it may be more important to consider the high-risk nature of the COGA sample. As this analysis draws on data from all phases of the COGA study, this sample includes some subjects who were ascertained as probands (n=151). The mean among probands (n=151, M=128.73, SE=9.91) is higher, although the mean among the remaining participants (n=2611, M=94.58, SE=1.76) is still elevated relative to the overall mean in Add Health. The aggregation of alcohol use behaviors within high-risk families may also explain this increased frequency of alcohol use. On the other hand, Add Health and ALSPAC are both general population samples. However, the lower legal age of drinking may lead to higher levels of alcohol use in ALSPAC in this developmental stage. All subjects in ALSPAC are of legal drinking age at this point, where some in Add Health are still younger than the legal limit in the United States.

In adulthood, the average drinking days per year in COGA was higher among males and subjects born in the 80s compared to Add Health. Similar to the age 18-25 age group, it may be particularly important to consider the high-risk nature of the COGA sample in this

developmental period. This sample includes a larger proportion of subjects who were ascertained as probands (n=768). The mean among probands (n=768, M=193.43, SE=5.54) is higher, and the mean among the remaining participants (n=4487, M=82.15, SE=1.59) is similar to the overall mean in Add Health (M=86.65, SE=1.43).

Aim 1: Developmentally-Informative GWAS

The hypothesis for Aim 1 was that gene-identification analyses for alcohol use frequency across COGA, Add Health and ALSPAC would demonstrate unique associations at different developmental stages (adolescence, early adulthood, adulthood). The analyses conducted in Aim 1 provide partial support for this hypothesis. The Common Factor gSEM model is saturated with three indicators and represents a direct transformation of the genetic correlation matrix. As such, interpretation of the results of Aim 1 focuses primarily on the genetic correlations between alcohol use frequency at different developmental periods and the subsequent gSEM GWAS analysis.

Genetic Correlations Across Development

A genetic correlation that can be statistically distinguished from one suggests that there is varying heterogeneity in the genetic architecture of the two phenotypes. The 95% confidence interval for the genetic correlation between alcohol use frequency in adolescence and adulthood was distinct from one (95% CI [-1.61, 0.93]), suggesting that there is heterogeneity in the genetic liability that underlies alcohol use frequency throughout development. The other genetic correlations were not statistically different than 1. The point estimate of the genetic correlation between alcohol use frequency in early adulthood and adulthood was positive and relatively large (rG = .75, SE = 0.58). By contrast, the point estimates for the genetic correlations between alcohol use frequency in adolescence and alcohol use frequency in early adulthood (rG = -.27,

SE = 0.79) and adulthood (rG = -.34, SE = 0.65) were negative. Notably, all estimates of H^2_{SNP} were small (Adolescence H^2_{SNP} = .04, SE=0.04; Early Adulthood H^2_{SNP} = .05, SE = 0.05; Adulthood H^2_{SNP} = .08, SE = 0.04), indicating that the genetic correlations reported here account for only a small proportion of the overall relationship between these measures. Small estimates of H^2_{SNP}, paired with the large confidence intervals for each genetic correlation, indicate that these results should be interpreted with caution. Large samples and greater statistical power are required to draw more substantive conclusions regarding variability in the genetic liability that underlies alcohol use frequency throughout development. Additional analyses to examine the underpinnings of these differences across development are described below, including consideration of SNP-level effects, gene-mapping, and genetic correlations with other phenotypes. Similar to the genetic correlations described above, these analyses should be interpreted cautiously due to the limited statistical power of these analyses.

gSEM GWAS

Three SNPs reached the genome-wide significance threshold in the gSEM GWAS of the adolescence residual: rs116734991, rs115778926, and rs117048287. SNP-level post-GWAS analyses were generally inconclusive: these significant SNPs have no known functional or regulatory role in the genome, no other close LD proxies, and do not appear in previous genome-wide association studies. Two of these SNPs, rs116734991 and rs115778926, were in close LD, indicating that they are likely tagging the same association signal. By contrast, gene-mapping results provide some insights into possible interpretations of these SNP effects. Positional mapping in FUMA indicated that rs116734991 and rs115778926 were closest to *LINC02477*. The remaining SNP, rs117048287, was mapped to *POLD3*. Previous studies can provide a tentative interpretation for these mappings. *LINC02477* has been mapped in previous studies of

educational attainment (Okbay et al., 2022) and major depressive disorder (Giannakopoulou et al., 2021). *POLD3* has been implicated in previous studies of height (Kichaev et al., 2019) and weight (Sakaue et al., 2021). It is possible that rs116734991 and rs115778926 (via *LINC02477*) influence alcohol use through a pathway shared by either educational attainment or major depressive disorder, while rs117048287 (via *POLD3*) acts primarily through a pathway related to physiological traits such as body mass index. If these results are replicated in future studies, *LINC02477* and *POLD3* may represent potential targets to begin untangling the functional pathways that underlie genetic liability for alcohol use frequency in adolescence.

It is important to note that isolated genome-wide significant SNPs such as these may be spurious. SNP-level effects that are robust and interpretable are typically identified in clusters which tag a selection of variants with functional or regulatory consequences (Uffelmann et al., 2021). The absence of clear proxy SNPs for the 3 significant effects casts some doubt on the reliability of these findings. There are examples of the expected pattern of clustering among other SNPs that reached the suggestive significance threshold in each of the gSEM GWAS. The physical location of these SNP clusters varies across development. Clusters of SNPs on chromosome 5, chromosome 3, and chromosome 16 reached the suggestive significance threshold in adolescence, early adulthood, and adulthood, respectively. This pattern of results provides modest evidence for qualitative differences in the genetic architecture of alcohol use frequency across development. Different chromosomal regions appear to contribute to genetic liability for alcohol use frequency at different stages of development; however, replication in a larger sample is warranted before substantive interpretation of suggestive SNP effects. Results from the gSEM GWAS power analysis indicate that the target sample size for gene discovery studies depends heavily on the expected effect size and MAF of SNPs of interest. The results of

this simulation suggest that future developmentally-informative GWAS should be adequately powered to begin detecting effects at n=50,000 for SNPs that occur frequently in the population and have large effects. Sample size requirements increase rapidly as the expected SNP effect size and minor allele frequency decrease. Sample sizes approaching or in excess of N=500,000 may be required to detect SNPs with smaller effects and/or that occur less frequently in the population. Additional samples that may be viable targets to boost sample size in future developmentally-informative genetic analyses are discussed under Future Directions.

Genetic Correlations with External GWAS

Genetic correlations between the gSEM model components and external GWAS provide additional evidence to support the distinction between genetic liability for alcohol use frequency in adolescence and adulthood, as well as a basis to begin disentangling what comprises these differences. Note that many of these genetic correlations were not statistically different than zero and substantive interpretation requires replication in larger sample sizes. The point estimate of the genetic correlation between alcohol use frequency and other adult alcohol use behaviors was relatively small in adolescence and increased in later developmental stages. Genetic correlations with educational attainment followed a similar pattern, with small point estimates in adolescence and larger point estimates in early adulthood and adulthood. Previous work demonstrates that the genetic liability underlying alcohol use frequency in adulthood is different than the genetic liability underlying other alcohol use outcomes in adulthood, such as alcohol problems and alcohol use quantity. Positive genetic correlations are reported between these other alcohol use behaviors and externalizing and internalizing phenotypes. Negative genetic correlations are reported between adult alcohol use frequency and externalizing phenotypes and internalizing phenotypes (Mallard, Savage, et al., 2022). Previous work also reports positive genetic

correlations between adult alcohol use frequency and indicators of socioeconomic status, such as educational attainment and income (Mallard, Savage, et al., 2022).

These previous results suggest that adult alcohol use frequency measures may index a variety of socioeconomic and environmental constructs that are not of immediate relevance in describing the etiology of clinically relevant alcohol use behaviors (Kranzler et al., 2022; Mallard, Savage, et al., 2022). This interpretation aligns with most of the general trends observed in the current study, with some notable exceptions. It is unexpected that the adult alcohol use frequency residual would be positively correlated with the externalizing common factor. Note that the externalizing common factor is partially comprised of alcohol use measures (AUDIT-P and Alcohol Dependence) which may account for this unexpected result.

It is possible that the negative association between adult alcohol use frequency and externalizing phenotypes does not extend to adolescence. The point estimates of the genetic correlation between adolescent alcohol use frequency and risk tolerance suggest that the genetic underpinnings of adolescent alcohol use frequency are more closely related to externalizing behavior, though none of the genetic correlations between phenotypes from external GWAS and alcohol use frequency in adolescence were significantly different than zero. These nominal results align with previous findings that externalizing genetic risk factors are especially important in adolescence (Kendler et al., 2011; Meyers et al., 2014). If these results are replicated in larger studies, genetic studies of alcohol use frequency in samples that encompass adolescence and adulthood should account for these qualitative differences in genetic liability across development. The same behavior measured at different ages may have different genetic architectures as well as other complex behavioral environmental causes, correlates, and consequences depending on the developmental context of the behavior. It is possible that other

phenotypes may demonstrate patterns of variability throughout development as well and the practice of pooling GWAS samples across developmental periods warrants empirical justification on a phenotype-by-phenotype basis. Larger sample sizes are needed to ascertain these differences throughout development.

Aim 2: Age-Matched PGS Validation

The hypothesis for Aim 2 was that PGS that were constructed to model developmental variability in genetic liability for alcohol use frequency would significantly predict phenotype variation better than a developmentally-agnostic, common factor PGS. Analyses in adulthood (age 26+) provide partial support for the hypothesis of Aim 2. Analyses in adolescence (age 12-17) and early adulthood (18-25) were inconclusive.

In adulthood, the residual PGS and the common factor PGS each improved model fit above a model with just covariates. The regression coefficient for the residual PGS was also significant after correcting for multiple testing when the common factor PGS was excluded from the model. The significant regression coefficient associated with the residual PGS predicted an increase of approximately 4 drinking days per year per PGS standard deviation. The common factor PGS was not significantly associated with alcohol use frequency in adulthood after correcting for multiple testing. Nominal differences in AIC align with these results, indicating that the model with just the residual PGS provided the best fit to the data. While these results provide tentative evidence for the utility of the residual PGS in adulthood, the observed effect of the residual PGS was still relatively small, accounting for a fraction of a percent of the phenotypic variance. In general, larger samples are required to predict a larger share of the phenotypic variance. Nominal differences in BIC also depart from this pattern, suggesting that the base model provided the best fit to the data. Neither PGS demonstrated significant effects

when modeled together and the residual PGS did not improve model fit significantly above a model that included the common factor PGS when tested via likelihood ratio test. Larger samples may be required to more clearly delineate the effects of the common factor PGS and the residual PGS.

The differences observed between the residual PGS and the common factor PGS in adulthood are particularly noteworthy when considering the sample size for each PGS. The effective sample size of the common factor with COGA removed was approximately 8,230, while the sample size of the adulthood GWAS with COGA removed was 5,149. Contrasting these differences in statistical power, the effect of the residual PGS was generally more robust across the series of tests conducted in the current study, surviving multiple testing correction in the linear mixed effects models when modeled without the common factor PGS. It is noteworthy that these differences begin to emerge in the context of an underpowered analysis. Constructing PGS from development-specific residual variance may produce further improvements in performance when sample sizes are larger.

The residual and common factor PGS were not associated with alcohol use frequency in adolescence. Likelihood ratio tests did not indicate any significant improvement in model fit when the PGS were added to the model and barely discernable differences in AIC and BIC suggest that the addition of these predictors resulted in worse model fit. Similar results were observed for the residual PGS and common factor PGS in early adulthood, with one exception: nominal differences in AIC suggest that the model with the common factor PGS provided slightly better fit to the data than the other models. The regression coefficient associated with the common factor PGS was also nominally larger than other non-significant PGS effects, equivalent to a predicted increase of approximately 3 drinking days per year per PGS standard deviation. It

is possible that genetic variance that is shared across developmental periods is particularly important in early adulthood. It should be noted that nominal differences in BIC contradict this trend, suggesting that the base model with only covariates provided better fit to the data. If future studies support these preliminary findings, refining gene discovery analyses to account for developmental variability may improve phenotype prediction via the common factor, rather than the development-specific residual. Some previous work aligns with this interpretation: PGS derived from an adult GWAS, which would index both adult-specific and common genetic variance, have been shown to predict alcohol consumption between ages 22-27 in one study (Elam et al., 2021) and a variety of alcohol use phenotypes at age 22 in another study (Kandaswamy et al., 2021). Regardless, nominal differences in fit indices from the current work should be interpreted cautiously, particularly when all other formal tests indicate null results in early adulthood.

Similar to Aim 1, statistical power is an important consideration. The PGS power analysis suggests that the current study is adequately powered to detect the effect of a PGS that accounts for approximately 1% of the phenotype variance in adolescence and less than 0.5% of the variance in early adulthood and adulthood. However, the performance of a PGS is directly influenced by the statistical power of the GWAS that it is derived from. The low statistical power of the Aim 1 analyses reduces the likelihood of identifying reliable association with the PGS in Aim 2, particularly after adjustments to the analysis plan which required removing COGA from the GWAS analyses for PGS construction. The effective sample size of the common factor (8661.00) was larger than the effective sample size of the early adulthood residual (7,070.74). This may account for the null results associated with the residual PGS and nominal effect of the

common factor PGS on model fit in early adulthood, rather than a true difference in the relative importance of these components of genetic variance.

Regarding the null results in adolescence, the overall heritability of the target phenotype also affects the statistical power of polygenic association analyses; phenotypes that are less heritable require larger sample sizes (Khanzadeh et al., 2022). The genetic correlations presented in support of Aim 1 highlight potential differences between the genetic architecture of the adolescence residual and the other components of the model. Prohibition of adolescent alcohol use in the United States and the United Kingdom may underlie these apparent differences. Adolescent alcohol use frequency indirectly measures a range of behaviors involved in obtaining access to alcohol outside of the context of legal sales: for an adolescent to drink alcohol, they must first arrive at a context where they have access to alcohol. Given these differences, it is possible that the residual component plays some role in the genetic liability that underlies alcohol use frequency in adolescence; however, differences in the heritability of alcohol use behaviors across development may lead to varying sample size requirements for discovery and prediction analyses across developmental stages. Previous research suggests that the heritability of alcohol use behaviors is lower in earlier developmental periods, while environmental influences such as cultural transmission from parents and the broader familial context are more influential (Geels et al., 2012; Kendler et al., 2008; Koopmans & Boomsma, 1996; Rose et al., 2001). Thus, larger samples may be required for studies examining alcohol use frequency in earlier developmental periods, relative to adulthood where the heritability of alcohol use behaviors is generally higher. The LDSC results from the current study support this interpretation. When LDSC was run with all available samples, the SNP-based heritability of alcohol use frequency was nominally lower in adolescence (h2SNP=.04, SE=0.04) than adulthood (h2SNP=.08, SE=0.04).

Generally, the relative importance of common and specific genetic risk factors is likely to vary across development. The results of the current work suggest that a developmentally-informative approach to PGS construction may improve phenotype prediction in adulthood via the residual, and in early adulthood via the common factor. Future studies that implement this approach to match PGS to developmental stages may see further improvements in performance and provide clearer evidence regarding the relative balance of common and specific genetic effects throughout development when sample sizes are larger, particularly in adolescence where heritability is lower.

Limitations

The results of this study should be interpreted in the context of several important limitations. Foremost, the GWAS meta-analysis that was used to generate summary statistics within developmental periods was underpowered. Effective sample sizes were approximately 8869, 9647, and 9894 in adolescence, early adulthood, and adulthood, respectively. COGA was removed from the discovery analysis in Aim 2 to facilitate PGS construction, further reducing power. This limitation affects the statistical power and interpretability of all downstream analyses, including LDSC, gSEM, gSEM GWAS, and PGS construction. Estimates of H^2_{SNP} were small in each developmental period and the 95% confidence intervals for estimates derived from all stages of the analysis pipeline were large. As a result, point estimates should be interpreted cautiously. For example, it remains unclear whether the negative genome-wide genetic correlations between alcohol use frequency in adolescence and alcohol use frequency in later developmental periods reflects a true negative association or a relatively small positive association. Regional genetic correlations may also vary in their direction of effect, further complicating the interpretation of these nominal findings (van Rheenen et al., 2019).

Relatedly, these analyses only included European ancestry participants. Differences in LD structure and allele frequencies limit the generalizability of GWAS and PGS findings across ancestry groups. All results reported here are not generalizable outside of European ancestry populations. Limited statistical power precludes expanding the current study to examine other ancestry groups, as the effective sample sizes for GWAS in other ancestry groups would likely be too small to provide stable estimates of heritability and genetic correlation via LD score regression. Other methods for estimation of heritability and genetic correlation in diverse samples from GWAS summary statistics require larger sample sizes than are available in the current study (Brown et al., 2016; Luo et al., 2021; Zhang & Schumacher, 2021). Expanding genetic discovery and prediction analyses to incorporate diverse ancestry groups represents an important opportunity. As polygenic scores become more relevant in healthcare settings (Lambert et al., 2019), it is critical that studies be conducted to increase the applicability of genetic research findings across ancestry groups (Martin et al., 2019). Additionally, previous work indicates that a disproportionately large number of genetic variants associated with phenotypes are likely to exist in understudied populations because the total number of genetic variants differs between populations (Gurdasani et al., 2019; MacArthur et al., 2017; Morales et al., 2018). This trend in population genetics arises from the series of population founder events that occurred as early humans migrated across the world (Henn et al., 2019). African ancestry populations and populations with recent admixture in their demographic history tend to have the largest total number of genetic variants (1000 Genomes Project Consortium et al., 2015) and are generally understudied in behavior genetics (Martin et al., 2019).

Furthermore, these analyses model age as a series of ordinal developmental periods (adolescence age 12-17, early adulthood age 18-25, adulthood age 26+), rather than a continuum.

This approach is not sensitive to differences in genetic liabilities that may exist within the specified developmental periods; for example, if genetic liability for alcohol use frequency is different at age 12 and age 17. This ordinal approach was adopted to work within the Genomic SEM framework, which models genetic covariances and variances in the absence of subject-level phenotypic and genotypic information. A modeling framework that does not require subject-level data can more easily incorporate additional samples to boost statistical power, relying on GWAS summary statistics that may be shared among research groups more easily than subject-level data. Future implementations of developmentally-informative GWAS models may adopt a variety of other modeling approaches when large, longitudinal samples with subject-level phenotypic and genotypic data are available. Alternative modeling approaches are discussed in the Future Directions section under Statistical Power and Additional Samples.

Future Directions

Statistical Power and Additional Samples

The Genomic SEM modeling framework does not require subject-level data and can readily incorporate additional samples to boost statistical power. In this section, I provide an overview of studies that may be viable targets for expanding the scope of the current study to include additional samples and increase statistical power, with an emphasis on studies that include alcohol use phenotypes and genotypic data for adolescent participants. The combined adolescent sample size across the following studies may be adequate to reach the target sample of 50,000. Alternative methods that make more efficient use of available data by modeling phenotype data directly may also provide an important avenue to increase power and test novel hypotheses.

Adolescent Brain Cognitive Development (ABCD) Study

The Adolescent Brain Cognitive Development (ABCD) Study is a longitudinal cohort study conducted across 21 study sites in the United States (Garavan et al., 2018; Karcher & Barch, 2021). Between September 2016 and August 2018 children age 9 to 10 (n=11,875) were invited to participate through the elementary school system. Remote follow up assessments occur every six months and in-person follow up assessments occur annually. Regular follow-ups are planned for 10 years. Genotypic data are available for 9,683 participants and 50.5% (n=4,920) of the GWAS sample is European ancestry (Ohi et al., 2021)

Minnesota Center for Twin and Family Research (MCTFR) Cohorts

The Minnesota Center for Twin and Family Research (MCTFR) has conducted three studies that may be of interest: The Minnesota Twin Family Study (MTFS), The Enrichment Study (ES), and the Sibling Interaction and Behavior Study (SIBS) (Wilson et al., 2019). A subset of participants in each study contributed genotypic data. Across the three studies, a total of 8405 adolescent participants have been genotyped and 91.6% (n=7,702) of the GWAS sample is European ancestry (Miller et al., 2012).

MTFS is a longitudinal study of twins born between 1972 to 1984. Twins were identified through a combination of birth records and a brief interview with the parent of prospective twins. MTFS includes 2 cohorts: a group of twins and their parents who were first assessed when the twins were age 11 (twin n=1512, parent n=1521) and a group of twins and their parents who were first assessed when the twins were age 17 (twin n=1252, parent n=1221). Follow up assessments were conducted every three to five years (age 11, 14, 17, 20, 24, 29).

ES is a longitudinal study of twins born between 1988 to 1994. Twins were again identified through a combination of birth records and a brief interview with the parent of prospective twins. In order to enrich this sample for substance use risk, a subset of the ES twins

was ascertained to meet a predetermined threshold of elevated externalizing behavior. ES is comprised of a group of twins and their parents who were first assessed when the twins were age 11 (twin n=998, parent n=939). Follow up assessments were conducted at ages 14, 17, and 24.

SIBS is a longitudinal study of adoptive and nonadoptive siblings born between 1978-1991. Adoptive families were identified through private adoption agencies in Minnesota and nonadoptive families were identified through birth records. SIBS is comprised of adoptees (n=692) and non-adoptees (n=540) who were assessed starting at age 15 and their parents (n=1158). Follow up assessments were conducted at ages 18 and 22.

LifeLines

LifeLines is a population-based longitudinal cohort study in the Netherlands (Scholtens et al., 2015). Individuals age 25-50 were invited to participate by their general practitioners and interested individuals were mailed additional information about the project. Participants who were identified in GP offices were also asked to invite their family members, including children, to participate in the study. Questionnaire follow up assessments were conducted every 1.5 years. As of 2015 (Scholtens et al., 2015), the total sample included n=14,801 individuals under the age of 18, n=140,222 individuals between ages 18-65, and n=12,706 individuals over the age of 65. Of these participants, n=15,638 contributed genotypic data, although all participants genotyped in this initial assessment were older than age 18. A more recent update to the LifeLines study added an additional 38,000 participants to the GWAS sample, 3,000 of which are under age 18 (LifeLines, 2019).

The Trøndelag Health Study (HUNT)

The Trøndelag Health Study (HUNT) is a population-based cohort study in Norway (Åsvold et al., 2022; Holmen et al., 2014). Data collection on individuals older than age 20

started in 1984 and data collection on adolescents age 13-19 started in 1995. All individuals within the specified age ranges living in Trøndelag County, Norway were invited to participate at each of the four waves of the survey: 1984-1986, 1995-1997, 2006-2008, and 2017-2019. In total, ~229,000 adults and ~25,000 adolescents have enrolled in the study (Brumpton et al., 2021). Of these participants, n=~88,000 adults have contributed genotypic data. Adolescent participants are not genotyped in this study; however, 4,212 adolescent participants have proceeded to participate in the adult component of the study where genotyping is possible.

The Norwegian Mother, Father, and Child Cohort Study (MoBa)

The Norwegian Mother, Father, and Child Cohort Study (MoBa) is longitudinal birth cohort of mothers, fathers, and children in Norway (Magnus et al., 2016). Pregnant women were invited to participate at hospitals throughout Norway, starting with a single hospital in the city of Bergen in 1999 and eventually expanding to 50 out of 52 of the maternity wards in Norway. Fathers were also invited to participate after the initial wave of data collection. Assessments of adolescent offspring in the birth cohort were conducted at ages 13, 14, and 16-17. Recent report indicate that the study includes genotypic data from 207,569 participants (Corfield et al., 2022) with at least 14,000 participants under age 18 (Helgeland et al., 2019).

23andMe

23andMe is a direct-to-consumer genetics company that offers opportunities for collaboration with external researchers (Eriksson et al., 2010; Tung et al., 2011). Precise information about the composition of the 23andMe GWAS sample is not freely available and external researchers are not permitted to access the data directly; however, the modular meta-analytic approach implemented in Genomic SEM could be used to incorporate summary

statistics from the 23andMe sample. A recent study reports a sample size of n=403,931 for the 23andMe drinks per week phenotype that was incorporated into the analysis (Liu et al., 2019).

Continuous Approaches to Time

The modeling approach employed in the current study was not exhaustive and alternative multivariate models may provide a better explanation of developmental variability in genetic effects on alcohol use behaviors. Only the genetic variance that is measured in GWAS summary statistics is modeled in Genomic SEM, producing estimates that are independent of environmental influences. The current study uses developmentally-informed age groups (adolescence age 12-17; early adulthood age 18-25, adulthood age 26+) to model differences in genetic variance across development. The clustering of multiple ages into a single group is a limitation of this study. These analyses are not sensitive to changes in genetic variance that may occur within a given age range. In this section, I provide an overview of two methodological frameworks that provide an opportunity to model the rate of change in a phenotype over time (the Latent Growth Model) and age as a continuum (the Time-Varying Effects Model). Previous evidence suggests that the rate of change over time in alcohol use is heritable (Edwards et al., 2017). Modeling age as a continuum would address limitations in the current study associated with using ordinal age groups. Note that the following approaches are currently incompatible with the current Genomic SEM framework, which models genetic variances and covariances in the absence of individual-level data and observed phenotypic means. Alternative approaches to model developmental variability in genetic effects that incorporate individual-level data include GW-SEM (Verhulst et al., 2017) and genomic-relatedness-matrix restricted maximum-likelihood (GREML; Kirkpatrick et al., 2021)

Latent Growth Models

Latent growth modeling (LGM) is a well-established framework for the assessment of change over time (McArdle & Epstein, 1987). The LGM uses latent variables to represent the intercept (average level of the outcome) and slope (change in the outcome over time) under parametric assumptions about the functional form of change, typically accommodating linear, quadratic, or cubic functions. Regarding continuous approaches to the measurement of change over time, the latent slope is of particular interest. In the context of GWAS, measuring association of the latent slope of the LGM and SNPs provides an opportunity to detect genetic effects that are specific to the rate of change over time in the phenotype. A previous study conducted by Edwards et al. (2017) applied this approach by extracting the slope term from an LGM of Alcohol Use Disorder Identification Test (AUDIT) scores and conducting a GWAS with this component as the outcome. SNP effects may also be incorporated directly into a structural equation model using the GW-SEM package (Verhulst et al., 2017) in R (R Core Team, 2017). The estimates of SNP effects on the latent slope could then be used to generate a PGS in an independent sample that is weighted to predict the rate of change in the phenotype over time. The LGM requires information about phenotypic means for model identification and so this approach, as well as the other described throughout this section, are incompatible with the Genomic SEM framework.

Time-Varying Effects Modeling

Time-Varying Effects Modeling (TVEM) was originally developed as an approach for modeling intensive longitudinal data with many observations per individual without relying on parametric assumptions about the functional form of change over time (Tan et al., 2012). TVEM can be conceptualized as an extension of repeated measures mixed effects (multilevel) models,

where observations are treated as being nested within individuals. Equivalent to a latent growth model, the repeated measures mixed effects model makes parametric assumptions about both the function form of (1) mean change in the outcome over time and (2) the relationship between a predictor and the outcome over time. The coefficient function that describes change over time is specified by the analyst, typically accommodating linear, quadratic, or cubic functions. In TVEM, these parametric assumptions are avoided and the coefficient function that describes change over time is estimated from data.

In the context of GWAS, TVEM represents a unique opportunity to measure the effects of SNPs on the phenotype as they vary dynamically across development. Using age as the metric of time, a TVEM GWAS could freely estimate the coefficient function that describes the effect of each SNP as a continuous function of age. The coefficient function for each SNP could be used to generate SNP effect size estimates that are matched to the exact value of each participants' age in an external sample. The age-matched SNP effects could then be used to generate PGS that are specific to exact age, rather than PGS that are specific to a broader developmental period as was the goal in the current study. Notably, fitting TVEMs with a freely estimated coefficient function requires larger sample sizes than fitting a comparable model that makes parametric assumptions about the coefficient function. Incorporating TVEM into GWAS would require large, genotyped samples with extensive repeated measures on the outcome of interest. These data do not exist currently.

Computational burden may represent another obstacle to this approach. The simulation pipeline described in the Power Analysis section was adapted to test possible run time of a TVEM GWAS using the tvem package (Dziak et al., 2021) in R (R Core Team, 2017). Multiple observations per developmental period were generated from the initial three simulated indicators

described in the Power Analysis section by (1) generating a vector of normally distributed error, (2) adding the error vector to the initial phenotype to generate one new observation, (3) subtracting the error vector from the initial phenotype to generate a second observation, and (4) repeating this process until 21 new observations were generated: 6 in adolescence (age 12-17), 8 in early adulthood (age 18-25), and 8 in adulthood (age 26-33). This approach ensures that the mean of the new observations within each developmental period is equal to the initial simulated phenotype. The tvem function requires specification of a number of 'knots', where a greater number of knots increases both computation time and the flexibility of the coefficient function to detect non-linear changes over time. In this example, the number of knots was learned from the simulated data using a procedure implemented in the tvem package, with a maximum possible value of 15 knots.

A simulated TVEM model with the SNP predicting the simulated 21 observations per person was run with 1 knot, N=100,000, SNP effect=0.01, MAF=.50, and genetic correlations between developmental periods set to the observed values described in the Power Analysis section. A single draw using this pipeline was timed to estimate the computational burden associated with TVEM under these parameter settings. With these settings, the TVEM for a single SNP completed after approximately 21 minutes running on one 64 bit Intel core with 2GB RAM from a UNIX cluster. Given the large number of SNPs involved in a typical GWAS, the computational burden associated with the TVEM method implemented here is too large to be practically feasible. The gSEM GWAS described in the current manuscript included 6,707,536 SNPs. Extrapolating the estimated run time of 21 minutes per SNP, a GWAS with N=100,000 and 6,707,536 SNPs using this implementation of TVEM would run for 140,858,256 minutes, or

approximately 268 years. Parallel processing and future developments in the estimation of these models may improve the feasibility of large-scale applications of TVEM .

Overall Summary and Conclusions

Alcohol use behaviors are heritable (Verhulst et al., 2015) and associated with substantial burden to public health (WHO, 2018). The genetic architecture of alcohol use involves the contribution of many variants of small effect (Wray et al., 2018) and GWAS methods are well-suited to account for this polygenicity. Genetic influences on alcohol use vary throughout development (Aliev et al., 2015; Dick et al., 2006; Edwards & Kendler, 2013; Kendler et al., 2011; Meyers et al., 2014; Sakai et al., 2010), but contemporary GWAS do not account for this variability (Kranzler et al., 2019; Liu et al., 2019; Sanchez-Roige et al., 2019). Previous research suggests that the omission of developmental considerations from gene-identification studies for alcohol use behaviors limits both the generalizability of GWAS results across age groups and the utility of polygenic scores for phenotype prediction across the lifespan (Elam et al., 2021; Kandaswamy et al., 2021).

The aims of this project were (1) to advance gene discovery by building developmentally-informative models for gene identification that incorporate changes in alcohol use across age groups and (2) leverage results from the developmentally-informative GWAS for genetic prediction of age-matched alcohol use outcomes in an independent sample. The hypothesis for Aim 1 was that the meta-analyzed GWAS results would demonstrate unique associations at different developmental stages. The hypothesis for Aim 2 was that residual PGS would predict their corresponding alcohol use phenotype significantly better than a developmentally-agnostic, common factor PGS. Small sample sizes and low statistical power

limit the substantive conclusions that can be drawn regarding these aims, though the results from Aim 1 and Aim 2 provide tentative support for these hypotheses.

Results from Aim 1 suggest that genetic liability for alcohol use frequency in adolescence may be distinct from genetic liability for alcohol use frequency later in developmental periods. Alcohol use frequency in adolescence was nominally genetically correlated with risk tolerance. Results from Aim 2 suggest that accounting for differences in genetic architecture across development may improve phenotype prediction by polygenic scores, contingent on the discovery sample size and the heritability of the trait at the target developmental period. The current study provides nominal evidence that the residual PGS performed better than the common factor PGS in adulthood, though a formal comparison of the effect of each score was not statistically significant. The PGS in other developmental periods were not significantly associated with alcohol use frequency, although nominal interpretation of model fit indices suggests that the common factor PGS may have some utility for predicting alcohol use frequency in early adulthood. The current work was underpowered at all steps of the analysis plan. Replication in well-powered samples is warranted and these interpretations are speculative.

Methods to measure and quantify the molecular contributions to genetic liability for alcohol use behaviors have advanced substantially over the years, from linkage and candidate gene studies of the previous decades to genome-wide association, polygenic scoring, and the recent multivariate extensions of these contemporary methods. These advances in molecular genetic methods provide greater opportunity to identify genetic variants that account for the developmental changes in genetic liability for alcohol use behaviors that have been observed in latent genetic studies. Developmentally-informed gene discovery analyses may improve phenotype prediction via polygenic scores when discovery samples are adequately large to model

differences in genetic liabilities across development. The results presented here are an initial step towards toward this goal and lay a foundation for future molecular genetic studies of developmental variability in the genetic underpinnings of alcohol use behaviors and the subsequent possibility of genetically-informed, age-matched phenotype prediction.

References

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Ahlström, S. K., & Österberg, E. L. (2004). International Perspectives on Adolescent and Young Adult Drinking. *Alcohol Research & Health*, *28*(4), 258–268.

Aliev, F., Wetherill, L., Bierut, L., Bucholz, K. K., Edenberg, H., Foroud, T., & Dick, D. M. (2015). Genes Associated With Alcohol Outcomes Show Enrichment of Effects With Broad Externalizing and Impulsivity Phenotypes in an Independent Sample. *Journal of Studies on Alcohol and Drugs*, *76*(1), 38–46.

Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., Gibbs, R. A., Peltonen, L., Dermitzakis, E., Schaffner, S. F., Yu, F., Peltonen, L., Dermitzakis, E., Bonnen, P. E., Altshuler, D. M., Gibbs, R. A., de Bakker, P. I. W., Deloukas, P., Gabriel, S. B., Gwilliam, R., Hunt, S., … Scientific management. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*(7311), 7311. https://doi.org/10.1038/nature09298

Anderson, P., Chisholm, D., & Fuhr, D. C. (2009). Effectiveness and cost-effectiveness of policies and programmes to reduce the harm caused by alcohol. *Lancet (London, England)*, *373*(9682), 2234–2246. https://doi.org/10.1016/S0140-6736(09)60744-3

Åsvold, B. O., Langhammer, A., Rehn, T. A., Kjelvik, G., Grøntvedt, T. V., Sørgjerd, E. P., Fenstad, J. S., Heggland, J., Holmen, O., Stuifbergen, M. C., Vikjord, S. A. A., Brumpton, B. M., Skjellegrind, H. K., Thingstad, P., Sund, E. R., Selbæk, G., Mork, P. J.,

Rangul, V., Hveem, K., … Krokstad, S. (2022). Cohort Profile Update: The HUNT Study, Norway. *International Journal of Epidemiology*, dyac095. https://doi.org/10.1093/ije/dyac095

Barr, P. B., Ksinan, A., Su, J., Johnson, E. C., Meyers, J. L., Wetherill, L., Latvala, A., Aliev, F., Chan, G., Kuperman, S., Nurnberger, J., Kamarajan, C., Anokhin, A., Agrawal, A., Rose, R. J., Edenberg, H. J., Schuckit, M., Kaprio, J., & Dick, D. M. (2020). Using polygenic scores for identifying individuals at increased risk of substance use disorders in clinical and population samples. *Translational Psychiatry*, *10*(1), 1–9. https://doi.org/10.1038/s41398-020-00865-8

Barr, P. B., Kuo, S. I.-C., Aliev, F., Latvala, A., Viken, R., Rose, R. J., Kaprio, J., Salvatore, J. E., & Dick, D. M. (2019). Polygenic risk for alcohol misuse is moderated by romantic partnerships. *Addiction (Abingdon, England)*, *114*(10), 1753–1762. https://doi.org/10.1111/add.14712

Barr, P. B., Mallard, T. T., Sanchez-Roige, S., Poore, H. E., Linnér, R. K., Waldman, I. D., Palmer, A. A., Harden, K. P., Agrawal, A., & Dick, D. M. (2022). Parsing genetically influenced risk pathways: Genetic loci impact problematic alcohol use via externalizing and specific risk. *Translational Psychiatry*, *12*(1), 1. https://doi.org/10.1038/s41398-022-02171-x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Begleiter, H. (1995). The Collaborative Study on the Genetics of Alcoholism. *Alcohol Health and Research World*, *19*(3), 228–236.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bennett, S. N., Caporaso, N., Fitzpatrick, A. L., Agrawal, A., Barnes, K., Boyd, H. A., Cornelis, M. C., Hansel, N. N., Heiss, G., Heit, J. A., Kang, J. H., Kittner, S. J., Kraft, P., Lowe, W., Marazita, M. L., Monroe, K. R., Pasquale, L. R., Ramos, E. M., van Dam, R. M., … Williams, K. (2011). Phenotype harmonization and cross-study collaboration in GWAS consortia: The GENEVA experience. *Genetic Epidemiology*, *35*(3), 159–173. https://doi.org/10.1002/gepi.20564

Bierut, L. J., Agrawal, A., Bucholz, K. K., Doheny, K. F., Laurie, C., Pugh, E., Fisher, S., Fox, L., Howells, W., Bertelsen, S., Hinrichs, A. L., Almasy, L., Breslau, N., Culverhouse, R. C., Dick, D. M., Edenberg, H. J., Foroud, T., Grucza, R. A., Hatsukami, D., … Gene, Environment Association Studies Consortium. (2010). A genome-wide association study of alcohol dependence. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(11), 5082–5087. https://doi.org/10.1073/pnas.0911109107

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *The American Journal of Psychiatry*, *176*(5), 376–387. https://doi.org/10.1176/appi.ajp.2018.18070881

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S., & Davey Smith, G. (2013). Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, *42*(1), 111–127. https://doi.org/10.1093/ije/dys064

Brown, B. C., Ye, C. J., Price, A. L., & Zaitlen, N. (2016). Transethnic Genetic-Correlation Estimates from Summary Statistics. *The American Journal of Human Genetics*, *99*(1), 76–88. https://doi.org/10.1016/j.ajhg.2016.05.001

Brumpton, B. M., Graham, S., Surakka, I., Skogholt, A. H., Løset, M., Fritsche, L. G., Wolford, B., Zhou, W., Nielsen, J. B., Holmen, O. L., Gabrielsen, M. E., Thomas, L., Bhatta, L., Rasheed, H., Zhang, H., Kang, H. M., Hornsby, W., Moksnes, M. R., Coward, E., … Willer, C. J. (2021). The HUNT Study: A population-based cohort for genetic research. *MedRxiv*, 2021.12.23.21268305. https://doi.org/10.1101/2021.12.23.21268305

Bryazka, D., Reitsma, M. B., Griswold, M. G., Abate, K. H., Abbafati, C., Abbasi-Kangevari, M., Abbasi-Kangevari, Z., Abdoli, A., Abdollahi, M., Abdullah, A. Y. M., Abhilash, E. S., Abu-Gharbieh, E., Acuna, J. M., Addolorato, G., Adebayo, O. M., Adekanmbi, V., Adhikari, K., Adhikari, S., Adnani, Q. E. S., … Gakidou, E. (2022). Population-level risks of alcohol consumption by amount, geography, age, sex, and year: A systematic analysis for the Global Burden of Disease Study 2020. *The Lancet*, *400*(10347), 185–235. https://doi.org/10.1016/S0140-6736(22)00847-9

Bucholz, K. K., Cadoret, R., Cloninger, C. R., Dinwiddie, S. H., Hesselbrock, V. M., Nurnberger, J. I., Reich, T., Schmidt, I., & Schuckit, M. A. (1994). A new, semi-structured psychiatric interview for use in genetic linkage studies: A report on the reliability of the SSAGA. *Journal of Studies on Alcohol*, *55*(2), 149–158. https://doi.org/10.15288/jsa.1994.55.149

Bucholz, K. K., McCutcheon, V. V., Agrawal, A., Dick, D. M., Hesselbrock, V. M., Kramer, J. R., Kuperman, S., Nurnberger, J. I., Salvatore, J. E., Schuckit, M. A., Bierut, L. J., Foroud, T. M., Chan, G., Hesselbrock, M., Meyers, J. L., Edenberg, H. J., & Porjesz, B.

(2017). Comparison of parent, peer, psychiatric, and cannabis use influences across stages of offspring alcohol involvement: Evidence from the COGA Prospective Study. *Alcoholism, Clinical and Experimental Research*, *41*(2), 359–368. https://doi.org/10.1111/acer.13293

Bulik-Sullivan, B. K., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., Duncan, L., Perry, J. R. B., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 3. https://doi.org/10.1038/ng.3211

Casswell, S., & Zhang, J. F. (1997). Access to alcohol from licensed premises during adolescence: A longitudinal study. *Addiction (Abingdon, England)*, *92*(6), 737–745.

Chen, C. C., Lu, R. B., Chen, Y. C., Wang, M. F., Chang, Y. C., Li, T. K., & Yin, S. J. (1999). Interaction between the functional polymorphisms of the alcohol-metabolism genes in protection against alcoholism. *American Journal of Human Genetics*, *65*(3), 795–807. https://doi.org/10.1086/302540

Chen, Z., Boehnke, M., Wen, X., & Mukherjee, B. (2021). Revisiting the genome-wide significance threshold for common variant GWAS. *G3 Genes|Genomes|Genetics*, *11*(2), jkaa056. https://doi.org/10.1093/g3journal/jkaa056

Choi, S. W., Mak, T. S. H., & O'Reilly, P. F. (2018). A guide to performing Polygenic Risk Score analyses. *BioRxiv*, 416545. https://doi.org/10.1101/416545

Chong, A. H. W., Mitchell, R. E., Hemani, G., Davey Smith, G., Yolken, R. H., Richmond, R. C., & Paternoster, L. (2021). Genetic Analyses of Common Infections in the Avon Longitudinal Study of Parents and Children Cohort. *Frontiers in Immunology*, *12*. https://www.frontiersin.org/articles/10.3389/fimmu.2021.727457

Corfield, E., Frei, O., Shadrin, A., Rahman, Z., Lin, A., Athanasiu, L., Akdeniz, B. C., Hannigan, L., Wootton, R., Austerberry, C., Hughes, A., Tesli, M., Westlye, L., Stefánsson, H., Stefánsson, K., Njølstad, P., Magnus, P., Davies, N., Appadurai, V., … Havdahl, A. (2022). *The Norwegian Mother, Father, and Child cohort study (MoBa) genotyping data resource: MoBaPsychGen pipeline v.* https://doi.org/10.1101/2022.06.23.496289

Covault, J., Gelernter, J., Hesselbrock, V., Nellissery, M., & Kranzler, H. R. (2004). Allelic and haplotypic association of GABRA2 with alcohol dependence. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, *129B*(1), 104–109. https://doi.org/10.1002/ajmg.b.30091

Dawson, D. A. (2000). US low-risk drinking guidelines: An examination of four alternatives. *Alcoholism: Clinical and Experimental Research*, *24*(12), 1820–1829. https://doi.org/10.1111/j.1530-0277.2000.tb01986.x

Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *The American Journal of Psychiatry*, *168*(10), 1041–1049. https://doi.org/10.1176/appi.ajp.2011.11020191

Dick, D. M., Agrawal, A., Keller, M. C., Adkins, A., Aliev, F., Monroe, S., Hewitt, J. K., Kendler, K. S., & Sher, K. J. (2015). Candidate gene-environment interaction research: Reflections and recommendations. *Perspectives on Psychological Science: A Journal of*

the Association for Psychological Science, *10*(1), 37–59. https://doi.org/10.1177/1745691614556682

Dick, D. M., Bierut, L., Hinrichs, A., Fox, L., Bucholz, K. K., Kramer, J., Kuperman, S., Hesselbrock, V., Schuckit, M., Almasy, L., Tischfield, J., Porjesz, B., Begleiter, H., Nurnberger, J., Xuei, X., Edenberg, H. J., & Foroud, T. (2006). The role of GABRA2 in risk for conduct disorder and alcohol and drug dependence across developmental stages. *Behavior Genetics*, *36*(4), 577–590. https://doi.org/10.1007/s10519-005-9041-8

Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genetics*, *9*(3), e1003348. https://doi.org/10.1371/journal.pgen.1003348

Duncan, L. E., Ostacher, M., & Ballon, J. (2019). How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. *Neuropsychopharmacology*, *44*(9), 9. https://doi.org/10.1038/s41386-019-0389-5

Dziak, J., Coffman, D., Li, R., Litson, K., & Chakrabo, Y. (2021). *tvem: Time-Varying Effect Models.* (R package version 1.3.1) [Computer software]. https://CRAN.R-project.org/package=tvem

Eaves, L. J. (2006). Genotype × Environment Interaction in Psychopathology: Fact or Artifact? *Twin Research and Human Genetics*, *9*(1), 1–8. https://doi.org/10.1375/twin.9.1.1

Edenberg, H. J. (2007). The genetics of alcohol metabolism: Role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Research & Health: The Journal of the National Institute on Alcohol Abuse and Alcoholism*, *30*(1), 5–13.

Edenberg, H. J., Dick, D. M., Xuei, X., Tian, H., Almasy, L., Bauer, L. O., Crowe, R. R., Goate, A., Hesselbrock, V., Jones, K., Kwon, J., Li, T.-K., Nurnberger, J. I., O'Connor, S. J., Reich, T., Rice, J., Schuckit, M. A., Porjesz, B., Foroud, T., & Begleiter, H. (2004).

Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. *American Journal of Human Genetics*, *74*(4), 705–714. https://doi.org/10.1086/383283

Edwards, A. C., Heron, J., Vladimirov, V., Wolen, A. R., Adkins, D. E., Aliev, F., Hickman, M., & Kendler, K. S. (2017). The Rate of Change in Alcohol Misuse Across Adolescence is Heritable. *Alcoholism, Clinical and Experimental Research*, *41*(1), 57–64. https://doi.org/10.1111/acer.13262

Edwards, A. C., & Kendler, K. S. (2013). Alcohol consumption in men is influenced by qualitatively different genetic factors in adolescence and adulthood. *Psychological Medicine*, *43*(9), 1857–1868. https://doi.org/10.1017/S0033291712002917

Elam, K. K., Ha, T., Neale, Z., Aliev, F., Dick, D., & Lemery-Chalfant, K. (2021). Age varying polygenic effects on alcohol use in African Americans and European Americans from adolescence to adulthood. *Scientific Reports*, *11*(1), 22425. https://doi.org/10.1038/s41598-021-01923-x

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., Avey, L., Wojcicki, A., Pe'er, I., & Mountain, J. (2010). Web-Based, Participant-Driven Studies Yield Novel Genetic Associations for Common Traits. *PLoS Genetics*, *6*(6), e1000993. https://doi.org/10.1371/journal.pgen.1000993

Evangelou, E., & Ioannidis, J. P. A. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*, *14*(6), 6. https://doi.org/10.1038/nrg3472

Fowler, T., Lifford, K., Shelton, K., Rice, F., Thapar, A., Neale, M. C., McBride, A., & van den Bree, M. B. M. (2007). Exploring the relationship between genetic and environmental

influences on initiation and progression of substance use. *Addiction (Abingdon, England)*, *102*(3), 413–422. https://doi.org/10.1111/j.1360-0443.2006.01694.x

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., Ring, S., Nelson, S. M., & Lawlor, D. A. (2013). Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology*, *42*(1), 97–110. https://doi.org/10.1093/ije/dys066

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R. Z., Heeringa, S., Jernigan, T., Potter, A., Thompson, W., & Zahs, D. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience*, *32*, 16–22. https://doi.org/10.1016/j.dcn.2018.04.004

Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A., & Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, *10*(1), 1. https://doi.org/10.1038/s41467-019-09718-5

Geels, L. M., Bartels, M., van Beijsterveldt, T. C. E. M., Willemsen, G., van der Aa, N., Boomsma, D. I., & Vink, J. M. (2012). Trends in adolescent alcohol use: Effects of age, sex and cohort on prevalence and heritability. *Addiction*, *107*(3), 518–527. https://doi.org/10.1111/j.1360-0443.2011.03612.x

Genz, A., & Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Springer Berlin. https://CRAN.R-project.org/package=mvtnorm

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2021). *mvtnorm: Multivariate Normal and t Distributions* (R package version 1.1-3) [Computer software]. https://CRAN.R-project.org/package=mvtnorm

Giannakopoulou, O., Lin, K., Meng, X., Su, M.-H., Kuo, P.-H., Peterson, R. E., Awasthi, S., Moscati, A., Coleman, J. R. I., Bass, N., Millwood, I. Y., Chen, Y., Chen, Z., Chen, H.-C., Lu, M.-L., Huang, M.-C., Chen, C.-H., Stahl, E. A., Loos, R. J. F., … 23andMe Research Team, China Kadoorie Biobank Collaborative Group, and Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2021). The Genetic Architecture of Depression in Individuals of East Asian Ancestry: A Genome-Wide Association Study. *JAMA Psychiatry*, *78*(11), 1258–1269. https://doi.org/10.1001/jamapsychiatry.2021.2099

Gillespie, N. A., Gentry, A. E., Kirkpatrick, R. M., Reynolds, C. A., Mathur, R., Kendler, K. S., Maes, H. H., Webb, B. T., & Peterson, R. E. (2022). Determining the stability of genome-wide factors in BMI between ages 40 to 69 years. *PLOS Genetics*, *18*(8), e1010303. https://doi.org/10.1371/journal.pgen.1010303

Grotzinger, A. D., Fuente, J. de la, Privé, F., Nivard, M. G., & Tucker-Drob, E. M. (2023). Pervasive Downward Bias in Estimates of Liability-Scale Heritability in Genome-wide Association Study Meta-analysis: A Simple Solution. *Biological Psychiatry*, *93*(1), 29–36. https://doi.org/10.1016/j.biopsych.2022.05.029

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., Ip, H. F., Marioni, R. E., McIntosh, A. M., Deary, I. J., Koellinger, P. D., Harden, K. P., Nivard, M. G., & Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour*, *3*(5), 5. https://doi.org/10.1038/s41562-019-0566-x

Guo, Y., Huang, Y., Hou, L., Ma, J., Chen, C., Ai, H., Huang, L., & Ren, J. (2017). Genome-wide detection of genetic markers associated with growth and fatness in four pig

populations using four approaches. *Genetics Selection Evolution*, *49*(1), 21. https://doi.org/10.1186/s12711-017-0295-4

Gurdasani, D., Barroso, I., Zeggini, E., & Sandhu, M. S. (2019). Genomics of disease risk in globally diverse populations. *Nature Reviews Genetics*, *20*(9), 9. https://doi.org/10.1038/s41576-019-0144-0

Hammond, R. K., Pahl, M. C., Su, C., Cousminer, D. L., Leonard, M. E., Lu, S., Doege, C. A., Wagley, Y., Hodge, K. M., Lasconi, C., Johnson, M. E., Pippin, J. A., Hankenson, K. D., Leibel, R. L., Chesi, A., Wells, A. D., & Grant, S. F. (2021). Biological constraints on GWAS SNPs at suggestive significance thresholds reveal additional BMI loci. *ELife*, *10*, e62206. https://doi.org/10.7554/eLife.62206

Harris, K. M. (2013). *The Add Health Study: Design and Accomplishments.* Carolina Population Center, University of North Carolina at Chapel Hill.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Hart, A. B., & Kranzler, H. R. (2015). Alcohol Dependence Genetics: Lessons Learned From Genome-Wide Association Studies (GWAS) and Post-GWAS Analyses. *Alcoholism, Clinical and Experimental Research*, *39*(8), 1312–1327. https://doi.org/10.1111/acer.12792

Heath, A. C., Meyer, J., Eaves, L. J., & Martin, N. G. (1991). The inheritance of alcohol consumption patterns in a general population twin sample: I. Multidimensional scaling of

quantity/frequency data. *Journal of Studies on Alcohol*, *52*(4), 345–352. https://doi.org/10.15288/jsa.1991.52.345

Heath, A. C., Meyer, J., Jardine, R., & Martin, N. G. (1991). The inheritance of alcohol consumption patterns in a general population twin sample: II. Determinants of consumption frequency and quantity consumed. *Journal of Studies on Alcohol*, *52*(5), 425–433. https://doi.org/10.15288/jsa.1991.52.425

Helgeland, Ø., Vaudel, M., Juliusson, P. B., Lingaas Holmen, O., Juodakis, J., Bacelis, J., Jacobsson, B., Lindekleiv, H., Hveem, K., Lie, R. T., Knudsen, G. P., Stoltenberg, C., Magnus, P., Sagen, J. V., Molven, A., Johansson, S., & Njølstad, P. R. (2019). Genome-wide association study reveals dynamic role of genetic variation in infant and early childhood growth. *Nature Communications*, *10*, 4448. https://doi.org/10.1038/s41467-019-12308-0

Hellwege, J., Keaton, J., Giri, A., Gao, X., Velez Edwards, D. R., & Edwards, T. L. (2017). Population Stratification in Genetic Association Studies. *Current Protocols in Human Genetics*, *95*, 1.22.1-1.22.23. https://doi.org/10.1002/cphg.48

Henn, B. M., Cavalli-Sforza, L. L., & Feldman, M. W. (2019). The great human expansion. *Resonance*, *24*(6), 711–718. https://doi.org/10.1007/s12045-019-0830-4

Highland, H. M., Avery, C. L., Duan, Q., Li, Y., & Harris, K. M. (2018). *Quality Control Analysis of Add Health GWAS Data* (Add Health Documentation). Carolina Population Center. https://addhealth.cpc.unc.edu/wp-content/uploads/docs/user_guides/AH_GWAS_QC.pdf

Holmen, T. L., Bratberg, G., Krokstad, S., Langhammer, A., Hveem, K., Midthjell, K., Heggland, J., & Holmen, J. (2014). Cohort profile of the Young-HUNT Study, Norway:

A population-based study of adolescents. *International Journal of Epidemiology*, *43*(2), 536–544. https://doi.org/10.1093/ije/dys232

Horwitz, T., Lam, K., Chen, Y., Xia, Y., & Liu, C. (2019). A decade in psychiatric GWAS research. *Molecular Psychiatry*, *24*(3), 378–389. https://doi.org/10.1038/s41380-018-0055-z

Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., Coleman, J. R. I., Hagenaars, S. P., Ward, J., Wigmore, E. M., Alloza, C., Shen, X., Barbu, M. C., Xu, E. Y., Whalley, H. C., Marioni, R. E., Porteous, D. J., Davies, G., Deary, I. J., … McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*, *22*(3), 343–352. https://doi.org/10.1038/s41593-018-0326-7

Huibregtse, B. M., Corley, R. P., Wadsworth, S. J., Vandever, J. M., DeFries, J. C., & Stallings, M. C. (2016). A Longitudinal Adoption Study of Substance Use Behavior in Adolescence. *Twin Research and Human Genetics*, *19*(4), 330–340. https://doi.org/10.1017/thg.2016.35

Irons, D. E., Iacono, W. G., Oetting, W. S., & McGue, M. (2012). Developmental trajectory and environmental moderation of the effect of ALDH2 polymorphism on alcohol use. *Alcoholism, Clinical and Experimental Research*, *36*(11), 1882–1891. https://doi.org/10.1111/j.1530-0277.2012.01809.x

Jinks, J. L., & Fulker, D. W. (1970). Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychological Bulletin*, *73*(5), 311–349. https://doi.org/10.1037/h0029135

Kandaswamy, R., Allegrini, A., Plomin, R., & Stumm, S. von. (2021). Predictive validity of genome-wide polygenic scores for alcohol use from adolescence to young adulthood. *Drug and Alcohol Dependence*, *219*, 108480. https://doi.org/10.1016/j.drugalcdep.2020.108480

Karcher, N. R., & Barch, D. M. (2021). The ABCD study: Understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology*, *46*(1), 1. https://doi.org/10.1038/s41386-020-0736-6

Karlsson Linnér, R., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., Lebreton, M., Tino, S. P., Abdellaoui, A., Hammerschlag, A. R., Nivard, M. G., Okbay, A., Rietveld, C. A., Timshel, P. N., Trzaskowski, M., Vlaming, R. de, Zünd, C. L., Bao, Y., Buzdugan, L., … Beauchamp, J. P. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*, *51*(2), 2. https://doi.org/10.1038/s41588-018-0309-3

Karlsson Linnér, R., Mallard, T. T., Barr, P. B., Sanchez-Roige, S., Madole, J. W., Driver, M. N., Poore, H. E., de Vlaming, R., Grotzinger, A. D., Tielbeek, J. J., Johnson, E. C., Liu, M., Rosenthal, S. B., Ideker, T., Zhou, H., Kember, R. L., Pasman, J. A., Verweij, K. J. H., Liu, D. J., … Dick, D. M. (2021). Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nature Neuroscience*, 1–10. https://doi.org/10.1038/s41593-021-00908-3

Kendler, K. S., Gardner, C., & Dick, D. M. (2011). Predicting alcohol consumption in adolescence from alcohol-specific and general externalizing genetic risk factors, key

environmental exposures and their interaction. *Psychological Medicine*, *41*(7), 1507–1516. https://doi.org/10.1017/S003329171000190X

Kendler, K. S., Schmitt, E., Aggen, S. H., & Prescott, C. A. (2008). Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Archives of General Psychiatry*, *65*(6), 674–682. https://doi.org/10.1001/archpsyc.65.6.674

Keyes, K. M., Schulenberg, J. E., O'Malley, P. M., Johnston, L. D., Bachman, J. G., Li, G., & Hasin, D. (2012). Birth cohort effects on adolescent alcohol use: The influence of social norms from 1976 to 2007. *Archives of General Psychiatry*, *69*(12), 1304–1313. https://doi.org/10.1001/archgenpsychiatry.2012.787

Khanzadeh, H., Ghavi Hossein-Zadeh, N., & Ghovvati, S. (2022). The statistical power of genome-wide association studies for threshold traits with different frequencies of causal variants. *Genetica*, *150*(1), 51–57. https://doi.org/10.1007/s10709-021-00140-8

Kichaev, G., Bhatia, G., Loh, P.-R., Gazal, S., Burch, K., Freund, M. K., Schoech, A., Pasaniuc, B., & Price, A. L. (2019). Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *American Journal of Human Genetics*, *104*(1), 65–75. https://doi.org/10.1016/j.ajhg.2018.11.008

Kirkpatrick, R. M., Pritikin, J. N., Hunter, M. D., & Neale, M. C. (2021). Combining Structural-Equation Modeling with Genomic-Relatedness-Matrix Restricted Maximum Likelihood in OpenMx. *Behavior Genetics*, *51*(3), 331–342. https://doi.org/10.1007/s10519-020-10037-5

Koopmans, J. R., & Boomsma, D. I. (1996). Familial resemblances in alcohol use: Genetic or cultural transmission? *Journal of Studies on Alcohol*, *57*(1), 19–28. https://doi.org/10.15288/jsa.1996.57.19

Kranzler, H. R., Zhou, H., & Kember, R. L. (2022). Identifying and Reducing Bias in Genome-Wide Association Studies of Alcohol-Related Traits. *American Journal of Psychiatry*, *179*(1), 14–16. https://doi.org/10.1176/appi.ajp.2021.21111107

Kranzler, H. R., Zhou, H., Kember, R. L., Smith, R. V., Justice, A. C., Damrauer, S., Tsao, P. S., Klarin, D., Baras, A., Reid, J., Overton, J., Rader, D. J., Cheng, Z., Tate, J. P., Becker, W. C., Concato, J., Xu, K., Polimanti, R., Zhao, H., & Gelernter, J. (2019). Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-09480-8

Kretsch, N., & Harden, K. P. (2014). Marriage, Divorce, and Alcohol Use in Young Adulthood: A Longitudinal Sibling-Comparison Study. *Emerging Adulthood*, *2*(2), 138–149. https://doi.org/10.1177/2167696813513260

Ksinan, A. J., Su, J., Aliev, F., & Dick, D. M. (2019). Unpacking Genetic Risk Pathways for College Student Alcohol Consumption: The Mediating Role of Impulsivity. *Alcoholism: Clinical and Experimental Research*, *43*(10), 2100–2110. https://doi.org/10.1111/acer.14157

Kuo, S. I.-C., Salvatore, J. E., Aliev, F., Ha, T., Dishion, T. J., & Dick, D. M. (2019). The Family Check-up Intervention Moderates Polygenic Influences on Long-Term Alcohol Outcomes: Results from a Randomized Intervention Trial. *Prevention Science: The*

Official Journal of the Society for Prevention Research, *20*(7), 975–985. https://doi.org/10.1007/s11121-019-01024-2

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lai, D., Wetherill, L., Bertelsen, S., Carey, C. E., Kamarajan, C., Kapoor, M., Meyers, J. L., Anokhin, A. P., Bennett, D. A., Bucholz, K. K., Chang, K. K., De Jager, P. L., Dick, D. M., Hesselbrock, V., Kramer, J., Kuperman, S., Nurnberger, J. I., Raj, T., Schuckit, M., … Foroud, T. (2019). Genome-wide association studies of alcohol dependence, DSM-IV criterion count and individual criteria. *Genes, Brain, and Behavior*, *18*(6), e12579. https://doi.org/10.1111/gbb.12579

Lambert, S. A., Abraham, G., & Inouye, M. (2019). Towards clinical utility of polygenic risk scores. *Human Molecular Genetics*, *28*(R2), R133–R142. https://doi.org/10.1093/hmg/ddz187

Lander, E., & Kruglyak, L. (1995). Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nature Genetics*, *11*(3), 241–247. https://doi.org/10.1038/ng1195-241

Lappalainen, J., Krupitsky, E., Remizov, M., Pchelina, S., Taraskina, A., Zvartau, E., Somberg, L. K., Covault, J., Kranzler, H. R., Krystal, J. H., & Gelernter, J. (2005). Association between alcoholism and gamma-amino butyric acid alpha2 receptor subtype in a Russian population. *Alcoholism, Clinical and Experimental Research*, *29*(4), 493–498. https://doi.org/10.1097/01.alc.0000158938.97464.90

Lee, C. H., Cook, S., Lee, J. S., & Han, B. (2016). Comparison of Two Meta-Analysis Methods: Inverse-Variance-Weighted Average and Weighted Sum of Z-Scores. *Genomics & Informatics*, *14*(4), 173–180. https://doi.org/10.5808/GI.2016.14.4.173

Li, J. J., Cho, S. B., Salvatore, J. E., Edenberg, H. J., Agrawal, A., Chorlian, D. B., Porjesz, B., Hesselbrock, V., COGA Investigators, & Dick, D. M. (2017). The Impact of Peer Substance Use and Polygenic Risk on Trajectories of Heavy Episodic Drinking Across Adolescence and Emerging Adulthood. *Alcoholism, Clinical and Experimental Research*, *41*(1), 65–75. https://doi.org/10.1111/acer.13282

LifeLines. (2019). *LifeLines—Fact Sheet Genetic Data* [Fact sheet]. https://www.lifelines.nl/researcher/data-and-biobank/$5476/$6088

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., Choquet, H., Docherty, A. R., Faul, J. D., Foerster, J. R., Fritsche, L. G., Gabrielsen, M. E., Gordon, S. D., Haessler, J., … Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237–244. https://doi.org/10.1038/s41588-018-0307-5

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Luo, Y., Li, X., Wang, X., Gazal, S., Mercader, J. M., 23 and Me Research Team, SIGMA Type 2 Diabetes Consortium, Neale, B. M., Florez, J. C., Auton, A., Price, A. L., Finucane, H. K., & Raychaudhuri, S. (2021). Estimating heritability and its enrichment in tissue-

specific gene sets in admixed populations. *Human Molecular Genetics*, *30*(16), 1521–1534. https://doi.org/10.1093/hmg/ddab130

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z. M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., & Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, *45*(Database issue), D896–D901. https://doi.org/10.1093/nar/gkw1133

Machiela, M. J., & Chanock, S. J. (2015). LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, *31*(21), 3555–3557. https://doi.org/10.1093/bioinformatics/btv402

Maggs, J. L., & Schulenberg, J. E. (2004). Trajectories of Alcohol Use During the Transition to Adulthood. *Alcohol Research & Health*, *28*(4), 195–201.

Magnus, P., Birke, C., Vejrup, K., Haugan, A., Alsaker, E., Daltveit, A. K., Handal, M., Haugen, M., Høiseth, G., Knudsen, G. P., Paltiel, L., Schreuder, P., Tambs, K., Vold, L., & Stoltenberg, C. (2016). Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology*, *45*(2), 382–388. https://doi.org/10.1093/ije/dyw029

Mallard, T. T., Karlsson Linnér, R., Grotzinger, A. D., Sanchez-Roige, S., Seidlitz, J., Okbay, A., de Vlaming, R., Meddens, S. F. W., Palmer, A. A., Davis, L. K., Tucker-Drob, E. M., Kendler, K. S., Keller, M. C., Koellinger, P. D., & Harden, K. P. (2022). Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of

cross-cutting genetic liabilities. *Cell Genomics*, *2*(6), 100140. https://doi.org/10.1016/j.xgen.2022.100140

Mallard, T. T., Savage, J. E., Johnson, E. C., Huang, Y., Edwards, A. C., Hottenga, J. J., Grotzinger, A. D., Gustavson, D. E., Jennings, M. V., Anokhin, A., Dick, D. M., Edenberg, H. J., Kramer, J. R., Lai, D., Meyers, J. L., Pandey, A. K., Harden, K. P., Nivard, M. G., de Geus, E. J. C., … Sanchez-Roige, S. (2022). Item-Level Genome-Wide Association Study of the Alcohol Use Disorders Identification Test in Three Population-Based Cohorts. *The American Journal of Psychiatry*, *179*(1), 58–70. https://doi.org/10.1176/appi.ajp.2020.20091390

Manchia, M., Cullis, J., Turecki, G., Rouleau, G. A., Uher, R., & Alda, M. (2013). The Impact of Phenotypic and Genetic Heterogeneity on Results of Genome Wide Association Studies of Complex Diseases. *PLoS ONE*, *8*(10). https://doi.org/10.1371/journal.pone.0076295

Marchini, J., Cardon, L. R., Phillips, M. S., & Donnelly, P. (2004). The effects of human population structure on large genetic association studies. *Nature Genetics*, *36*(5), 5. https://doi.org/10.1038/ng1337

Marees, A. T., Kluiver, H. de, Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*, *27*(2), e1608. https://doi.org/10.1002/mpr.1608

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, *51*(4), 584–591. https://doi.org/10.1038/s41588-019-0379-x

Mather, K. and Jinks, J.L. (1982) Introduction to biometrical genetics. Chapman and Hall, London.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*(1), 110–133.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., … the Haplotype Reference Consortium. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*(10), 10. https://doi.org/10.1038/ng.3643

McVean, G. A., Altshuler (Co-Chair), D. M., Durbin (Co-Chair), R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., … University of Geneva. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 7422. https://doi.org/10.1038/nature11632

Meyers, J. L., Salvatore, J. E., Vuoksimaa, E., Korhonen, T., Pulkkinen, L., Rose, R. J., Kaprio, J., & Dick, D. M. (2014). Genetic influences on alcohol use behaviors have diverging developmental trajectories: A prospective study among male and female twins. *Alcoholism, Clinical and Experimental Research*, *38*(11), 2869–2877. https://doi.org/10.1111/acer.12560

Miller, M. B., Basu, S., Cunningham, J., Eskin, E., Malone, S. M., Oetting, W. S., Schork, N., Sul, J. H., Iacono, W. G., & Mcgue, M. (2012). The Minnesota Center for Twin and Family Research Genome-Wide Association Study. *Twin Research and Human Genetics : The Official Journal of the International Society for Twin Studies*, *15*(6), 767–774. https://doi.org/10.1017/thg.2012.62

Morales, J., Welter, D., Bowler, E. H., Cerezo, M., Harris, L. W., McMahon, A. C., Hall, P., Junkins, H. A., Milano, A., Hastings, E., Malangone, C., Buniello, A., Burdett, T., Flicek, P., Parkinson, H., Cunningham, F., Hindorff, L. A., & MacArthur, J. A. L. (2018). A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biology*, *19*(1), 21. https://doi.org/10.1186/s13059-018-1396-2

Munafò, M. R. (2006). Candidate gene studies in the 21st century: Meta-analysis, mediation, moderation. *Genes, Brain and Behavior*, *5*(S1), 3–8. https://doi.org/10.1111/j.1601-183X.2006.00188.x

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Neale Lab. (2018). *UK Biobank GWAS Results*. http://www.nealelab.is/uk-biobank

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika*, *81*(2), 535–549. https://doi.org/10.1007/s11336-014-9435-8

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature*, *541*(7637), 7637. https://doi.org/10.1038/nature21347

Northstone, K., Lewcock, M., Groom, A., Boyd, A., Macleod, J., Timpson, N., & Wells, N. (2019). The Avon Longitudinal Study of Parents and Children (ALSPAC): An update on

the enrolled sample of index children in 2019. *Wellcome Open Research*, *4*, 51. https://doi.org/10.12688/wellcomeopenres.15132.1

Ohi, K., Ochi, R., Noda, Y., Wada, M., Sugiyama, S., Nishi, A., Shioiri, T., Mimura, M., & Nakajima, S. (2021). Polygenic risk scores for major psychiatric and neurodevelopmental disorders contribute to sleep disturbance in childhood: Adolescent Brain Cognitive Development (ABCD) Study. *Translational Psychiatry*, *11*(1), 1. https://doi.org/10.1038/s41398-021-01308-8

Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., Sidorenko, J., Kweon, H., Goldman, G., Gjorgjieva, T., Jiang, Y., Hicks, B., Tian, C., Hinds, D. A., Ahlskog, R., Magnusson, P. K. E., Oskarsson, S., Hayward, C., Campbell, A., … Young, A. I. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, *54*(4), 4. https://doi.org/10.1038/s41588-022-01016-z

O'Malley, P. M. (2004). Maturing Out of Problematic Alcohol Use. *Alcohol Research & Health*, *28*(4), 202–204.

O'Malley, P. M., & Johnston, L. D. (2002). Epidemiology of alcohol and other drug use among American college students. *Journal of Studies on Alcohol. Supplement*, *14*, 23–39. https://doi.org/10.15288/jsas.2002.s14.23

Pagan, J. L., Rose, R. J., Viken, R. J., Pulkkinen, L., Kaprio, J., & Dick, D. M. (2006). Genetic and Environmental Influences on Stages of Alcohol Use Across Adolescence and into Young Adulthood. *Behavior Genetics*, *36*(4), 483–497. https://doi.org/10.1007/s10519-006-9062-y

Pasman, J. A., Verweij, K. J. H., & Vink, J. M. (2019). Systematic Review of Polygenic Gene–Environment Interaction in Tobacco, Alcohol, and Cannabis Use. *Behavior Genetics*, *49*(4), 349–365. https://doi.org/10.1007/s10519-019-09958-7

Paternoster, L., Howe, L. D., Tilling, K., Weedon, M. N., Freathy, R. M., Frayling, T. M., Kemp, J. P., Smith, G. D., Timpson, N. J., Ring, S. M., Evans, D. M., & Lawlor, D. A. (2011). Adult height variants affect birth length and growth rate in children. *Human Molecular Genetics*, *20*(20), 4069–4075. https://doi.org/10.1093/hmg/ddr309

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the Correct Statistical Test for the Equality of Regression Coefficients. *Criminology*, *36*(4), 859–866. https://doi.org/10.1111/j.1745-9125.1998.tb01268.x

Peterson, R. E., Kuchenbaecker, K., Walters, R. K., Chen, C.-Y., Popejoy, A. B., Periyasamy, S., Lam, M., Iyegbe, C., Strawbridge, R. J., Brick, L., Carey, C. E., Martin, A. R., Meyers, J. L., Su, J., Chen, J., Edwards, A. C., Kalungi, A., Koen, N., Majara, L., … Duncan, L. E. (2019). Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell*, *0*(0). https://doi.org/10.1016/j.cell.2019.08.051

Poelen, E. A. P., Derks, E. M., Engels, R. C. M. E., Van Leeuwe, J. F. J., Scholte, R. H. J., Willemsen, G., & Boomsma, D. I. (2008). The Relative Contribution of Genes and Environment to Alcohol Use in Early Adolescents: Are Similar Factors Related to Initiation of Alcohol Use and Frequency of Drinking? *Alcoholism: Clinical and Experimental Research*, *32*(6), 975–982. https://doi.org/10.1111/j.1530-0277.2008.00657.x

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Purves, K. L., Coleman, J. R. I., Meier, S. M., Rayner, C., Davis, K. A. S., Cheesman, R., Bækvad-Hansen, M., Børglum, A. D., Wan Cho, S., Jürgen Deckert, J., Gaspar, H. A., Bybjerg-Grauholm, J., Hettema, J. M., Hotopf, M., Hougaard, D., Hübel, C., Kan, C., McIntosh, A. M., Mors, O., … Eley, T. C. (2020). A major role for common genetic variation in anxiety disorders. *Molecular Psychiatry*, *25*(12), 12. https://doi.org/10.1038/s41380-019-0559-1

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science (New York, N.Y.)*, *273*(5281), 1516–1517. https://doi.org/10.1126/science.273.5281.1516

Rose, R. J., Dick, D. M., Viken, R. J., & Kaprio, J. (2001). Gene-environment interaction in patterns of adolescent drinking: Regional residency moderates longitudinal influences on alcohol use. *Alcoholism, Clinical and Experimental Research*, *25*(5), 637–643.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Sakai, J. T., Stallings, M. C., Crowley, T. J., Gelhorn, H. L., McQueen, M. B., & Ehringer, M. A. (2010). Test of association between GABRA2 (SNP rs279871) and adolescent conduct/alcohol use disorders utilizing a sample of clinic referred youth with serious

substance and conduct problems, controls and available first degree relatives. *Drug and Alcohol Dependence*, *106*(2–3), 199–203. https://doi.org/10.1016/j.drugalcdep.2009.08.015

Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M., Ishigaki, K., Suzuki, A., Suzuki, K., Obara, W., Yamaji, K., Takahashi, K., Asai, S., Takahashi, Y., Suzuki, T., … Okada, Y. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. *Nature Genetics*, *53*(10), 10. https://doi.org/10.1038/s41588-021-00931-x

Salvatore, J. E., Aliev, F., Edwards, A. C., Evans, D. M., Macleod, J., Hickman, M., Lewis, G., Kendler, K. S., Loukola, A., Korhonen, T., Latvala, A., Rose, R. J., Kaprio, J., & Dick, D. M. (2014). Polygenic scores predict alcohol problems in an independent sample and show moderation by the environment. *Genes*, *5*(2), 330–346. https://doi.org/10.3390/genes5020330

Salvatore, J. E., Thomas, N. S., Cho, S. B., Adkins, A., Kendler, K. S., & Dick, D. M. (2016). The role of romantic relationship status in pathways of risk for emerging adult alcohol use. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, *30*(3), 335–344. https://doi.org/10.1037/adb0000145

Sanchez-Roige, S., Palmer, A. A., Fontanillas, P., Elson, S. L., Adams, M. J., Howard, D. M., Edenberg, H. J., Davies, G., Crist, R. C., Deary, I. J., McIntosh, A. M., & Clarke, T.-K. (2019). Genome-wide association study meta-analysis of the Alcohol Use Disorder Identification Test (AUDIT) in two population-based cohorts. *The American Journal of Psychiatry*, *176*(2), 107–118. https://doi.org/10.1176/appi.ajp.2018.18040369

Saunders, G. R. B., Wang, X., Chen, F., Jang, S.-K., Liu, M., Wang, C., Gao, S., Jiang, Y., Khunsriraksakul, C., Otto, J. M., Addison, C., Akiyama, M., Albert, C. M., Aliev, F., Alonso, A., Arnett, D. K., Ashley-Koch, A. E., Ashrani, A. A., Barnes, K. C., … Vrieze, S. (2022). Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*, 1–7. https://doi.org/10.1038/s41586-022-05477-4

Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J. L., Dotinga, A., Vonk, J. M., van Dijk, F., van Zon, S. K. R., Wijmenga, C., Wolffenbuttel, B. H. R., & Stolk, R. P. (2015). Cohort Profile: LifeLines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, *44*(4), 1172–1180. https://doi.org/10.1093/ije/dyu229

Shaw, R. G. (1987). Maximum-Likelihood Approaches Applied to Quantitative Genetics of Natural Populations. *Evolution*, *41*(4), 812–826. https://doi.org/10.1111/j.1558-5646.1987.tb05855.x

Shi, G., & Rao, D. C. (2008). Ignoring temporal trends in genetic effects substantially reduces power of quantitative trait linkage analysis. *Genetic Epidemiology*, *32*(1), 61–72. https://doi.org/10.1002/gepi.20263

Smigielski, E. M., Sirotkin, K., Ward, M., & Sherry, S. T. (2000). dbSNP: A database of single nucleotide polymorphisms. *Nucleic Acids Research*, *28*(1), 352–355.

Sullivan, P. F. (2007). Spurious Genetic Associations. *Biological Psychiatry*, *61*(10), 1121–1126. https://doi.org/10.1016/j.biopsych.2006.11.010

Sullivan, P. F. (2017). How Good Were Candidate Gene Guesses in Schizophrenia Genetics? *Biological Psychiatry*, *82*(10), 696–697. https://doi.org/10.1016/j.biopsych.2017.09.004

Tan, X., Shiyko, M. P., Li, R., Li, Y., & Dierker, L. (2012). A time-varying effect model for intensive longitudinal data. *Psychological Methods*, *17*(1), 61–77. https://doi.org/10.1037/a0025814

Thomas, N. S., Adkins, A., Aliev, F., Edwards, A. C., Webb, B. T., Tiarsmith, E. C., Kendler, K. S., Dick, D. M., & Chartier, K. G. (2018). Alcohol Metabolizing Polygenic Risk for Alcohol Consumption in European American College Students. *Journal of Studies on Alcohol and Drugs*, *79*(4), 627–634. https://doi.org/10.15288/jsad.2018.79.627

Thomasson, H. R., Edenberg, H. J., Crabb, D. W., Mai, X. L., Jerome, R. E., Li, T. K., Wang, S. P., Lin, Y. T., Lu, R. B., & Yin, S. J. (1991). Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. *American Journal of Human Genetics*, *48*(4), 677–681.

Trucco, E. M., Villafuerte, S., Heitzeg, M. M., Burmeister, M., & Zucker, R. A. (2014). Rule breaking mediates the developmental association between GABRA2 and adolescent substance abuse. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *55*(12), 1372–1379. https://doi.org/10.1111/jcpp.12244

Tung, J. Y., Do, C. B., Hinds, D. A., Kiefer, A. K., Macpherson, J. M., Chowdry, A. B., Francke, U., Naughton, B. T., Mountain, J. L., Wojcicki, A., & Eriksson, N. (2011). Efficient Replication of over 180 Genetic Associations with Self-Reported Medical Data. *PLOS ONE*, *6*(8), e23473. https://doi.org/10.1371/journal.pone.0023473

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), 1. https://doi.org/10.1038/s43586-021-00056-9

van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. (2019). Genetic correlations of polygenic disease traits: From theory to practice. *Nature Reviews Genetics*, *20*(10), 10. https://doi.org/10.1038/s41576-019-0137-z

Verhulst, B., Maes, H. H., & Neale, M. C. (2017). GW-SEM: A Statistical Package to Conduct Genome-Wide Structural Equation Modeling. *Behavior Genetics*, *47*(3), 345–359. https://doi.org/10.1007/s10519-017-9842-6

Verhulst, B., Neale, M. C., & Kendler, K. S. (2015). The heritability of alcohol use disorders: A meta-analysis of twin and adoption studies. *Psychological Medicine*, *45*(5), 1061–1072. https://doi.org/10.1017/S0033291714002165

Viken, R. J., Kaprio, J., Koskenvuo, M., & Rose, R. J. (1999). Longitudinal analyses of the determinants of drinking and of drinking to intoxication in adolescent twins. *Behavior Genetics*, *29*(6), 455–461. https://doi.org/10.1023/a:1021631122461

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., Aliev, F., Bacanu, S.-A., Batzler, A., Bertelsen, S., Biernacka, J. M., Bigdeli, T. B., Chen, L.-S., Clarke, T.-K., Chou, Y.-L., Degenhardt, F., Docherty, A. R., Edwards, A. C., Fontanillas, P., … Agrawal, A. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*, *21*(12), 1656–1669. https://doi.org/10.1038/s41593-018-0275-1

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*(1), 1. https://doi.org/10.1038/s41467-017-01261-5

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, *26*(17), 2190–2191. https://doi.org/10.1093/bioinformatics/btq340

Williams et al., C. (Manuscript in Preparation). *Facilitating the Application of Externalizing Summary Statistics in Behavioral and Biomedical Research*.

Wilson, S., Haroian, K., Iacono, W. G., Krueger, R. F., Lee, J. J., Luciana, M., Malone, S. M., McGue, M., Roisman, G. I., & Vrieze, S. (2019). Minnesota Center for Twin and Family Research. *Twin Research and Human Genetics*, *22*(6), 746–752. https://doi.org/10.1017/thg.2019.107

World Health Organization. (2018). *Global Status Report on Alcohol and Health 2018*. World Health Organization.

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., & Middeldorp, C. M. (2014). Research review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *55*(10), 1068–1087. https://doi.org/10.1111/jcpp.12295

Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J., & Visscher, P. M. (2018). Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell*, *173*(7), 1573–1580. https://doi.org/10.1016/j.cell.2018.05.051

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, *14*(7), 7. https://doi.org/10.1038/nrg3457

Xu, Z., Shen, X., Pan, W., & Initiative, for the A. D. N. (2014). Longitudinal Analysis Is More Powerful than Cross-Sectional Analysis in Detecting Genetic Association with Neuroimaging Phenotypes. *PLOS ONE*, *9*(8), e102312. https://doi.org/10.1371/journal.pone.0102312

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zeggini, E., & Ioannidis, J. P. A. (2009). Meta-analysis in genome-wide association studies. *Pharmacogenomics*, *10*(2), 191–201. https://doi.org/10.2217/14622416.10.2.191

Zhang, J., & Schumacher, F. R. (2021). Evaluating the estimation of genetic correlation and heritability using summary statistics. *Molecular Genetics and Genomics*, *296*(6), 1221–1234. https://doi.org/10.1007/s00438-021-01817-7

Zhao, Z., Song, J., Wang, T., & Lu, Q. (2021). Polygenic risk scores: Effect estimation and model optimization. *Quantitative Biology*, *9*(2), 133–140. https://doi.org/10.15302/J-QB-021-0238

Zhu, M., & Zhao, S. (2007). Candidate Gene Identification Approach: Progress and Challenges. *International Journal of Biological Sciences*, *3*(7), 420. https://doi.org/10.7150/ijbs.3.420

Appendix 1. Adjustments to the Analysis Plan

Overview

The initial analysis plan for this project differs from the analysis described in the current work in three important ways. First, polygenic risk score analyses were initially planned to be conducted in the youngest cohort from the Finn Twin Studies (Finn Twin 12). Logistical barriers prevented timely access to the Finn Twin 12 data. Second, the initial analysis plan proposed to separate adolescence into two developmental periods (early adolescence age 12-14 and late adolescence age 15-17). Third, the initial analysis plan aimed to analyze both alcohol use frequency and heavy episodic drinking. Fourth, the Aim 1 discovery analysis was parameterized to model total genetic variance in each developmental period. The rationale for adjustments to the analysis plan is described in detail below. Results for the portion of the initial analysis pipeline that were feasible are reported in Appendix 2. Results for the Aim 1 discovery analysis modeling the residual variance in each developmental period are reported in Appendix 3.

Rationale for Adjustments

Combined Adolescence Developmental Period

Interpretable heritability estimates have a range of zero (0% heritable) to one (100% heritable); however, Linkage Disequilibrium Score Regression (LDSC) heritability estimates are not explicitly bound to the range of interpretable values. LDSC heritability estimates can sometimes fall below this range and be negative if the true value of the heritability is close to zero and/or statistical power is low enough that sampling variance produces a point estimate below zero. Negative heritability estimates are not interpretable and prevent the calculation of genetic correlations with the phenotype that has negative heritability. Genetic correlations and positive LDSC heritability estimates are required for all subsequent steps in the original analysis

plan: Genomic Structural Equation Modeling (gSEM), GWAS on components of the gSEM model, and calculation of polygenic risk scores from the gSEM GWAS results. Changes to the initial analysis plan were required as a result of this aspect of LDSC. All changes described here were made in order to produce positive LDSC heritability estimates for all phenotypes included in the model to allow the rest of the analysis plan to proceed.

The original analysis plan used four age bins (early adolescence: age 12-14, late adolescence: age 15-17, early adulthood: age 18-25, adulthood: age 26+). LDSC heritability estimates in the early adolescence age group were negative for alcohol use frequency. Negative heritability estimates for early adolescence remained across several permutations of the analysis with all four age bins, including various filters for minimum meta-analyzed N for SNPs included in LDSC and removal of valid 0s from the ALSPAC early adolescence sample (which is highly zero inflated). See Figure S52 for a diagram of all steps taken. The updated analysis was not sensitive to differences between genetic liability in early adolescence and genetic liability in late adolescence, as these two age groups were treated as a single group. Additionally, the common factor gSEM model was saturated with three indicators in the updated analysis and gSEM model fit indices cannot be interpreted.

Removal of Heavy Episodic Drinking

Analyses of heavy episodic drinking were not carried forward into the gSEM GWAS or polygenic risk score calculation steps of the project. LDSC heritability estimates in the late adolescence, early adulthood, and adulthood age groups were negative for heavy episodic drinking. Point estimates of LDSC heritability in early adolescence were also unusually large in all permutations of the analysis (.18 to .32). Positive LDSC heritability estimates for all indicators were obtained by removing SNPs with a meta-analytic sample size less than 2,000 and

removing adulthood from the analysis. Sample sizes in the remaining age groups were: early adolescence n=2380.48, late adolescence n=7447.64, and early adult n=8584.84. See Figure S53 for a diagram of all steps taken. Early adolescence and late adolescence could not be combined because at least 3 indicators are required to estimate the common factor model. The gSEM common factor model fit to the resulting genetic correlation matrix produced some nonsensical loadings with large standard errors, and is not suitable for the subsequent gSEM GWAS and polygenic risk score calculation steps of the project. As a consequence, this project does not provide additional information regarding developmental variability in genetic contributions to heavy episodic drinking.

Total Variance Parameterization for gSEM GWAS

The Aim 1 discovery analysis was parameterized to model total genetic variance in each developmental period. Attempts to calculate genetic correlations between the model residuals and a series of phenotypes from other studies produced negative heritability estimates in early adulthood and adulthood (early adult H^2_{SNP} = -6e-04 ,adult H^2_{SNP} = -4e-04). An alternate parameterization of the model was used to estimate the genetic correlations with other phenotypes using the total genetic variance in adolescence, early adulthood, and adulthood. Polygenic risk score construction in Aim 2 proceeded with the residual variance parameterization of the model. The paths from each SNP to the indicator for each developmental period were estimated without a simultaneous path to the common factor. In this model parameterization, the effect of the SNP is not partitioned into a component attributable to the common factor and a component attributable to the residual. Instead, the effect of the SNP of the total genetic variance in each developmental period is estimated. This approach returned positive heritability estimates, allowing estimation of genetic correlations to proceed.

The substantive difference between these model parameterizations was established by simulation. Three variables were generated from a multivariate normal distribution with the following correlation structure: X1 with X2 r=.6, X1 with X3 r=.6, X1 with X3 r=.45. A common factor model was fit to these variables with explicit parameterization of the X1 residual as a latent variable by fixing the variance of X1 to 0, fixing the loading of the latent on X1 to 1, and freely estimating the variance of the X1 residual latent. A factor score was generated for the X1 residual latent. The factor score was used to generate an exogenous predictor variable (X4) from a normal distribution by conditioning the distribution mean and variance on the factor score with Beta=0.25. This procedure produces an exogenous predictor (X4) that demonstrates greater association with the residual variance of X1 than the total variance of X1.

The common factor model of X1, X2, and X3 was fit again to estimate the effect of X4 on the X1 residual using the DWLS estimator that is implemented in GenomicSEM. Two models were tested: (Model A) one model with a path from X4 to the X1 residual and (Model B) another model with paths from X4 to the X1 residual and the common factor. An estimate of the effect of X4 on the total variance of X1 was obtained by regressing X1 on X4 without a common factor. Results from the two common factor model parameterizations were compared to estimates of the effect of X4 on the total variance of X1. This pipeline was repeated 100 times. The results for Model A are presented in Figure S54. The results for Model B are presented in Figure S55. In the figures, ResEff is the Z score for the path from X4 to the X1 residual, extracted from the common factor model. TotalEff is the Z score for the path from X4 to the total variance of X1. Note that ResEff and TotalEff are identical in Model A, indicating that this model parameterization models the total, rather than residual, variance of the indicator.

Appendix 2. Original Analysis Plan Results

Original Analysis Plan Measures and Results

The initial proposal for this work targeted early adolescence (age 12-14), late adolescence (age 15-17), early adulthood (age 18-25), and adulthood age (26+). Longitudinal cohort studies targeted for gene-identification analyses in the initial formulation of the project include the Collaborative Study on the Genetics of Alcoholism (COGA; early adolescence n=220, late adolescence n=1000, early adulthood n=2762, adulthood n=5255), the National Longitudinal Study of Adolescent to Adult Health (Add Health; early adolescence n=782, late adolescence n=2533, early adulthood n=3993, adulthood n=5149), and the Avon Longitudinal Study of Parents and Children (ALSPAC; early adolescence n=2862, late adolescence n=4928, early adulthood n=3613).

This Appendix provides results from the original analysis plan up to the point that technical issues prevented additional steps from being completed. Complete measurement details, descriptive statistics, within-sample GWAS results, GWAS meta-analysis results, LD score regression (LSDC) results, and genomic structural equation modeling (gSEM) results are reported for heavy episodic drinking. For alcohol use frequency, this section focuses on details pertaining to early adolescence and late adolescence. Measures and results for early adulthood and adulthood, which are unchanged between the two versions of the analysis, are reported in the main text of this work. Note that these results are limited in their interpretability because of technical issues and are presented here primarily as supporting information for the rationale for adjustments described in Appendix 1.

Measures

Alcohol Use Frequency

COGA.

Reports between ages 12-14 were collected in the Phase 4, Phase 1, and Phase 2 assessments as part of the SSAGA interviews (Bucholz et al., 1994). In Phase 4, alcohol use frequency was measured with an item that asks “On how many days did you drink any beverages containing alcohol during the last 12 months?” with response options “every day”, “5-6 days a week (nearly every day)”, “4 days a week (200-259 days)”, “3 days per week (150-199 days)”, “2 days per week (100-149 days)”, “1 day per week (50-99 days)”, “3 days per month (36-49 days)”, “2 days per month (24-35 days)”, “1 day per month (12-23 days), “6-11 days per year”, “3-5 days per year”, “1 to 2 days per year”, and “never”. In Phase 1 and Phase 2, alcohol use frequency was measured with a series of items that ask “On a typical [DAY] in the past 6 months, how many drinks of [KIND OF ALCOHOL] would you have?”, indexing consumption of beer, wine, liquor, and ‘other’ each day of the week with numeric free response. For participants missing data on these items, responses were drawn from a series of items asking “We would like to know the number of alcoholic drinks you've had each day in the last week […] How many drinks of [KIND OF ALCOHOL] did you have on [DAY]?” with numeric free response.

Reports between ages 15-17 were collected in Phase 4, Phase 1, and Phase 2. In Phase 4, alcohol use frequency was measured with an item that asks “On how many days did you drink any beverages containing alcohol during the last 12 months?” with response options “every day”, “5-6 days a week (nearly every day)”, “4 days a week (200-259 days)”, “3 days per week (150-199 days)”, “2 days per week (100-149 days)”, “1 day per week (50-99 days)”, “3 days per month (36-49 days)”, “2 days per month (24-35 days)”, “1 day per month (12-23 days), “6-11 days per year”, “3-5 days per year”, “1 to 2 days per year”, and “never”. In Phase 1 and Phase 2,

alcohol use frequency was measured with a series of items that ask “On a typical [DAY] in the past 6 months, how many drinks of [KIND OF ALCOHOL] would you have?”, indexing consumption of beer, wine, liquor, and ‘other’ each day of the week with numeric free response. For participants missing data on these items, responses were drawn from a series of items asking “We would like to know the number of alcoholic drinks you've had each day in the last week […] How many drinks of [KIND OF ALCOHOL] did you have on [DAY]?” with numeric free response.

Add Health.

Reports between ages 12-14 were collected in Wave 1 and Wave 2. At Wave 1 and Wave 2, alcohol use frequency was measured with an item that asks “During the past 12 months, on how many days did you drink alcohol?” with response options “never”, “once or twice”, “once a month or less”, “2 or 3 days a month”, “once or twice a week”, “3 to 5 days a week”, and “nearly every day”.

Reports between ages 15-17 were collected in Wave 1 and Wave 2. At Wave 1 and Wave 2, alcohol use frequency was measured with an item that asks “During the past 12 months, on how many days did you drink alcohol?” with response options “never”, “once or twice”, “once a month or less”, “2 or 3 days a month”, “once or twice a week”, “3 to 5 days a week”, and “nearly every day”.

ALSPAC.

Reports between ages 12-14 were collected at ages 12.5 and 13.5. At the age 12.5 assessment and age 13.5 assessment, alcohol use frequency was measured with an item asking “How many times have you drunk alcohol in the past 6 months?” with numeric free response.

Reports between ages 15-17 were collected at ages 15.5, 16, and 17.5. At the age 15.5 assessment alcohol use frequency was measured with an item that asks “How many times have you had a full drink of alcohol in the last 6 months?” with response options “1-2 times”, “3-5 times”, “6-9 times”, “10-19 times”, “20-39 times”, “40-99 times”, “100 or more”. At the age 16 assessment and the age 17.5 assessment alcohol use frequency was measured with an item that asks “The next questions are about your use of alcoholic drinks during the past year. How often do you have a drink containing alcohol?” with response options “never”, “monthly or less”, “2-4 times a month”, “2-3 times a week”, and “4 or more times a week”.

Heavy Episodic Drinking

COGA.

Reports between ages 12-14 were collected in Phase 4. In Phase 4, heavy episodic drinking is measured with an item that asks “How often did you get drunk during the last 12 months?” with response options “every day”, “5-6 days a week (nearly every day)”, “4 days a week (200-259 days)”, “3 days per week (150-199 days)”, “2 days per week (100-149 days)”, “1 day per week (50-99 days)”, “3 days per month (36-49 days)”, “2 days per month (24-35 days)”, “1 day per month (12-23 days), “6-11 days per year”, “3-5 days per year”, “1 to 2 days per year”, and “never”.

Reports between ages 15-17 were collected in Phase 4. In Phase 4, heavy episodic drinking was measured with an item that asks “How often did you get drunk during the last 12 months?” with response options “every day”, “5-6 days a week (nearly every day)”, “4 days a week (200-259 days)”, “3 days per week (150-199 days)”, “2 days per week (100-149 days)”, “1 day per week (50-99 days)”, “3 days per month (36-49 days)”, “2 days per month (24-35 days)”,

“1 day per month (12-23 days), “6-11 days per year”, “3-5 days per year”, “1 to 2 days per year”, and “never”.

Reports between ages 18-25 were collected in Phase 4. In Phase 4, heavy episodic drinking was measured with an item that asks “How often did you get drunk during the last 12 months?” with response options “every day”, “5-6 days a week (nearly every day)”, “4 days a week (200-259 days)”, “3 days per week (150-199 days)”, “2 days per week (100-149 days)”, “1 day per week (50-99 days)”, “3 days per month (36-49 days)”, “2 days per month (24-35 days)”, “1 day per month (12-23 days), “6-11 days per year”, “3-5 days per year”, “1 to 2 days per year”, and “never”.

Reports after age 26 were collected in Phase 4. In Phase 4, heavy episodic drinking was measured with an item that asks “How often did you get drunk during the last 12 months?” with response options “every day”, “5-6 days a week (nearly every day)”, “4 days a week (200-259 days)”, “3 days per week (150-199 days)”, “2 days per week (100-149 days)”, “1 day per week (50-99 days)”, “3 days per month (36-49 days)”, “2 days per month (24-35 days)”, “1 day per month (12-23 days), “6-11 days per year”, “3-5 days per year”, “1 to 2 days per year”, and “never”.

Add Health.

Reports between ages 12-14 were collected in Wave 1 and Wave 2. At Wave 1 and Wave 2, heavy episodic drinking was measured with an item that asks “Over the past 12 months, on how many days did you drink 5 or more drinks in a row?” with response options “every day or almost every day”, “3 to 5 days a week”, “1 or 2 days a week”,“2 or 3 days a month”, “once a month or less”, “1 or 2 days in the past 12 months”, and “never”.

Reports between ages 15-17 were collected in Wave 1 and Wave 2. At Wave 1 and Wave 2, heavy episodic drinking was measured with an item that asks “Over the past 12 months, on how many days did you drink 5 or more drinks in a row?” with response options “every day or almost every day”, “3 to 5 days a week”, “1 or 2 days a week”,“2 or 3 days a month”, “once a month or less”, “1 or 2 days in the past 12 months”, and “never”.

Reports between ages 18-25 were collected in Wave 3 and Wave 4. At Wave 3 and Wave 4, heavy episodic drinking was measured with an item that asks “Over the past 12 months, on how many days did you drink 5 or more drinks in a row?” with response options “every day or almost every day”, “3 to 5 days a week”, “1 or 2 days a week”,“2 or 3 days a month”, “once a month or less”, “1 or 2 days in the past 12 months”, and “never”.

Reports after age 26 were collected in Wave 3, Wave 4, and Wave 5. At Wave 3, Wave 4, and Wave 5, heavy episodic drinking was measured with an item that asks “Over the past 12 months, on how many days did you drink 5 or more drinks in a row?” with response options “every day or almost every day”, “3 to 5 days a week”, “1 or 2 days a week”,“2 or 3 days a month”, “once a month or less”, “1 or 2 days in the past 12 months”, and “never”.

ALSPAC.

Reports between ages 12-14 were collected at age 12.5. At the age 12.5 assessment, heavy episodic drinking was measured with an item that asks “How many times have you drunk 5 or more alcoholic drinks in a single evening?” with numeric free response.

Reports between ages 15-17 were collected at ages 16, and 17.5. At the age 16 assessment and the age 17.5 assessment heavy episodic drinking was measured with an item that asks “The next questions are about your use of alcoholic drinks during the past year. How often

do you have six or more units of alcohol on one occasion?" with response options "never", "less than monthly", "monthly", "weekly", "daily or almost daily".

Reports between ages 18-25 were collected at ages 18, 20. At the age 18 assessment and the age 20 assessment heavy episodic drinking was measured with an item that asks "The next questions are about your use of alcoholic drinks during the past year. How often do you have six or more units of alcohol on one occasion?" with response options "never", "less than monthly", "monthly", "weekly", "daily or almost daily".

Results

Descriptive Statistics

Alcohol Use Frequency.

Early Adolescence.

In COGA, the overall mean drinking days per year among subjects age 12-14 (n=220) was 18.43 (SE=2.85). The mean among males was higher than females (Male n=109, M=24.65, SE=4.72; Female n=111, M=12.32, SE=3.13). Means varied nominally by age (Age 12 n=24, M=19.40, SE=7.52; Age 13 n=59, M=13.09, SE=5.01; Age 14 n=137, M=20.56, SE=3.78). Participants born in the 80s reported higher means than subjects born in the 70s, 90s, or 2000s (70s n=27, M=13.48, SE=13.48; 80s n=90, M=27.16, SE=4.81; 90s n=99, M=11.92, SE=2.60; 2000s n=4, M=16.38, SE= 9.39).

In Add Health, the overall mean drinking days per year among subjects age 12-14 (n=782) was 24.61 (SE=2.06). The mean among males was slightly higher than females (Male n=337, M=25.65, SE=3.30; Female n=445, M=23.83, SE=2.63). Means varied nominally by age (Age 12 n=23, M=20.20, SE=9.76; Age 13 n=214, M=23.18, SE=3.87; Age 14 n=545, M=25.36, SE=2.51). All participants from Add Health in this age group were born in the 80s.

In ALSPAC the overall mean drinking days per year among subjects age 12-14 (n=2862) was 10.25 (SE=0.37). The mean among males was slightly higher than females (Male n=1392, M=11.08, SE=0.57; Female n=1470, M=9.47, SE=0.47). Means also varied by age (Age 12.5 n=954, M=13.01, SE=0.86; Age 13.5 n=1908, M=8.87, SE=0.34). All participants from ALSPAC were born in the 90s.

Late Adolescence.

In COGA, the overall mean drinking days per year among subjects age 15-17 (n=1000) was 37.38 (SE=1.69). The mean among males was higher than females (Male n=506, M=40.82, SE=2.46; Female n=494, M=33.85, SE=2.32). Means varied slightly by age (Age 15 n=183, M=33.04, SE=3.34; Age 16 n=381, M=35.07, SE=2.91; Age 17 n=436, M=41.21, SE=2.57). Participants born in the 80s reported higher means than subjects born in the 70s, 90s, or 2000s (70s n=162, M=37.56, SE=5.58; 80s n=403, M=49.42, SE=2.62; 90s n=409, M=26.70, SE=2.20; 2000s n=26, M=17.56, SE= 3.73).

In Add Health, the overall mean drinking days per year among subjects age 15-17 (n=2533) was 37.33 (SE=1.29). The mean among males was higher than females (Male n=1189, M=42.30, SE=2.04; Female n=1344, M=32.93, SE=1.61). Means also varied by age (Age 15 n=819, M=31.35, SE=2.17; Age 16 n=754, M=40.55, SE=2.46; Age 17 n=960, M=39.89, SE=2.08). Participants born in the 70s reported slightly higher means than subjects born in the 80s (70s n=1852, M=38.01, SE=1.49; 80s n=681, M=35.47, SE=2.54).

In ALSPAC the overall mean drinking days per year among subjects age 15-17 (n=4928) was 58.03 (SE=0.77). The mean among males was slightly higher than females (Male n=2223, M=62.47, SE=1.22; Female n=2705, M=54.37, SE=0.98). Means varied by age (Age 15.5

n=1254, M=56.18, SE=1.56; Age 16 n=2269, M=50.07, SE=1.04; Age 17.5 n=1405, M=72.53, SE=1.55). All participants from ALSPAC were born in the 90s.

Heavy Episodic Drinking.

Early Adolescence.

In COGA, the overall mean heavy drinking days per year among subjects age 12-14 (n=100) was 5.29 (SE=1.87). The mean among males was higher than females (Male n=49, M=6.94, SE=2.83; Female n=51, M=3.70, SE=2.47). Means varied nominally by age (Age 12 n=7, M=6.29, SE=6.04; Age 13 n=24, M=1.19, SE=0.75; Age 14 n=69, M=6.61, SE=2.62). Participants born in the 90s reported nominally higher means than subjects born in the 2000s (90s n=96, M=5.05, SE=1.91; 2000s n=4, M=11.00, SE=10.51).

In Add Health, the overall mean heavy drinking days per year among subjects age 12-14 (n=781) was 16.41 (SE=2.10). The mean among males was higher than females (Male n=338, M=20.94, SE=3.64; Female n=443, M=12.95, SE=2.44). Means varied nominally by age (Age 12 n=30, M=4.90, SE=2.78; Age 13 n=209, M=12.08, SE=3.63; Age 14 n=542, M=18.71, SE=2.68). All participants from Add Health in this age group were born in the 80s.

In ALSPAC, the overall mean heavy drinking days per year among subjects age 12-14 (n=1528) was 0.43 (SE=0.08). The mean among males was higher than females (Male n=753, M=0.54, SE=0.14; Female n=773, M=0.34, SE=0.10). All participants from ALSPAC in this age group were age 12.5. All participants from ALSPAC were born in the 90s.

Late Adolescence.

In COGA, the overall mean heavy drinking days per year among subjects age 15-17 (n=580) was 20.03 (SE=1.76). The mean among males was higher than females (Male n=301, M=22.72, SE=2.54; Female n=279, M=17.12, SE=2.43). Means varied nominally by age (Age

15 n=87 M=14.89, SE=3.89; Age 16 n=231 M=18.59, SE=2.77; Age 17 n=262 M=23.00, SE=2.75). Participants born in the 80s reported higher means than subjects born in the 90s and 2000s (80s n=145, M=28.08, SE=4.49; 90s n=409, M=17.98, SE=1.90; 2000s n=26, M=7.39, SE=2.56).

In Add Health, the overall mean heavy drinking days per year among subjects age 15-17 (n=2530) was 29.08 (SE=1.32). The mean among males was higher than females (Male n=1188, M=37.00, SE=2.20; Female n=1342, M=22.07, SE=1.51). Means varied nominally by age (Age 15 n=819, M=24.22, SE=2.28; Age 16 n=752, M=33.99, SE=2.71; Age 17 n=959, M=29.39, SE=1.94). Participants born in the 70s and 80s reported similar means (70s n=1849, M=28.86, SE=1.48; 80s n=681, M=29.68, SE=2.79).

In ALSPAC, the overall mean heavy drinking days per year among subjects age 15-17 (n=4880) was 14.84 (SE=0.59). The mean among males was higher than females (Male n=2196, M=15.55, SE=0.59; Female n=2677, M=14.25, SE=0.49). Means varied nominally by age (Age 15.5 n=1217, M=7.19, SE=0.35; Age 16 n=2311, M=16.30, SE=0.54; Age 17.5 n=1352, M=19.23, SE=0.93). All participants from ALSPAC in this age group were born in the 90s.

Early Adulthood.

In COGA, the overall mean heavy drinking days per year among subjects age 18-25 (n=1580) was 58.22 (SE=1.92). The mean among males was higher than females (Male n=760, M=74.53, SE=3.01; Female n=820, M=43.11, SE=2.31). Means increased with age (Age 18 n=206, M=43.60, SE=4.75; Age 19 n=216, M=55.02, SE=5.16; Age 20 n=241, M=58.65, SE=5.06; Age 21 n=300, M=54.94, SE=3.97; Age 22 n=244, M=64.30, SE=5.22; Age 23 n=170, M=66.37, SE=6.27; Age 24 n=127, M=65.69, SE=7.06; Age 25 n=76, M=68.37, SE=9.21).

Participants born in the 80s reported higher means than subjects born in the 90s and 2000s (80s n=863, M=62.05, SE=2.70; 90s n=707, M=54.28, SE=2.74; 2000s n=10, M=6.25, SE=3.12).

In Add Health, the overall mean heavy drinking days per year among subjects age 18-25 (n=4085) was 31.01 (SE=0.95). The mean among males was higher than females (Male n=1829, M=46.11, SE=1.71; Female n=2171, M=18.28, SE=0.91). Means varied nominally by age (Age 18 n=27, M=27.59, SE=8.63; Age 19 n=327, M=37.88, SE=3.33; Age 20 n=545, M=33.16, SE=2.58; Age 21 n=726, M=31.70, SE=2.08; Age 22 n=724, M=29.04, SE=2.25; Age 23 n=723, M=31.77, SE=2.36; Age 24 n=618, M=26.88, SE=2.39; Age 25 n=310, M=29.68, SE=3.43). Participants born in the 80s reported higher means than subjects born in the 70s (70s n=2370, M=29.16, SE=1.25; 80s n=1630, M=33.69, SE=1.46).

In ALSPAC, the overall mean heavy drinking days per year among subjects age 18-25 (n=3603) was 27.16 (SE=0.69). The mean among males was higher than females (Male n=1414, M=31.16, SE=1.26; Female n=2186, M=24.56, SE=0.80). Means varied nominally by age (Age 18 n=1626, M=26.10, SE=0.95; Age 20 n=1977, M=28.03, SE=0.99). All participants from ALSPAC in this were born in the 90s.

Adulthood.

In COGA, the overall mean heavy drinking days per year among subjects age 26+ (n=851) was 35.69 (SE=2.25). The mean among males was higher than females (Male n=369, M=50.29, SE=3.93; Female n=482, M=24.51, SE=2.48). Means were similar across age (Age 26 n=284, M=35.55, SE=3.76; Age 27 n=284, M=35.04, SE=3.79; Age >27 n=283, M=36.48, SE=4.15).

Participants born in the 80s reported higher means than subjects born in other decades (50s n=28, M=24.84, SE=14.20; 60s n=52, M=23.77, SE=8.35; 70s n=23, M=9.70, SE=5.62; 80s n=590, M=39.90, SE=2.77; 90s n=156, M=29.73, SE=5.01).

In Add Health, the overall mean heavy drinking days per year among subjects age 26+ (n=5145) was 34.30 (SE=0.99). The mean among males was higher than females (Male n=2418, M=47.18, SE=1.70; Female n=2727, M=22.88, SE=1.05). Means varied nominally by age (Age 26 n=499, M=28.56, SE=2.43; Age 27 n=601, M=32.94, SE=2.66; Age 28 n=734, M=27.35, SE=2.21; Age 29 n=669, M=25.38, SE=2.09; Age 30 n=668, M=31.06, SE=2.63; Age 31 n=445, M=32.28, SE=3.43; Age 32 n=57, M=39.36, SE=9.93; Age 33 n=23, M=19.91, SE=15.79; Age 34 n=134, M=25.46, SE=5.19; Age 35 n=205, M=39.16, SE=5.63; Age 36 n=227, M=54.20, SE=6.46; Age 37 n=231, M=51.02, SE=6.11; Age 38 n=255, M=52.60, SE=5.70; Age 39 n=201, M=52.42, SE=6.65; Age 40 n=141, M=38.31, SE=6.09; Age 41 n=47, M=48.25, SE=11.60). Participants born in the 80s reported higher means than subjects born in the 70s (70s n=3123, M=33.76, SE=1.27; 80s n=2022, M=35.13, SE=1.57).

Interpretation of Sample Mean Differences

Average drinking days per year vary within developmental period between the samples. In the preceding section, I provided an overview of these differences as a function of sex, age, and birth cohort. In was follows, I provide possible explanations for these mean differences for each phenotype of interest.

Alcohol Use Frequency.

In early adolescence, the overall mean in Add Health was higher than the overall mean in COGA and ALSPAC. This was likely a result of birth cohort differences between the samples. Note that the mean in COGA subjects who were born in the 80s is similar to the mean of Add

Health subjects who were born in the 80s and the mean in COGA subjects who were born in the 90s is similar to the mean of ALSPAC subjects who were born in the 90s. Relatedly, Keyes et al. (2012) report that individuals born in the 90s tend to endorse more restrictive drinking norms compared to those born in the 80s and that restrictive drinking norms predict lower levels of alcohol use.

In late adolescence, means in ALSPAC were uniformly higher across sex, age, and birth cohort compared to the means in COGA and Add Health. These differences were most pronounced in the ALSPAC age 17.5 age group. This trend likely reflects the lower legal drinking age of 18 (16 if accompanied by a parent), in the United Kingdom, where the ALSPAC sample was collected. The legal drinking age in the United States, where the COGA and Add Health samples were collected, is 21. Previous work demonstrates that earlier legal access to alcohol is associated with increased alcohol use among adolescents (Anderson et al., 2009; Casswell & Zhang, 1997).

Heavy Episodic Drinking.

In early adolescence, the overall mean in Add Health was higher than the overall mean in COGA and ALSPAC. Additionally, the overall mean in COGA was higher than the overall mean in ALSPAC. These differences likely result from a combination of birth cohort and age differences between the samples, as well as cultural differences between the United States and the United Kingdom. As described earlier in the interpretation of discovery sample mean differences for the alcohol use frequency phenotype, individuals born in the 90s tend to endorse more restrictive drinking norms compared to those born in the 80s. All Add Health participants in this developmental period were born in the 80s, while all COGA and ALSPAC participants were born in either the 90s or 2000s. Additionally, the ALPSAC sample is younger than the Add

Health sample. Within this developmental period in Add Health, older subjects tend to report more heavy episodic drinking days, suggesting that age may also account for the elevated mean in Add Health. Finally, early adolescents in the overall ALPSAC sample report notably higher rates of alcohol use initiation (81%) compared to Add Health (54%) and COGA (21%). Aligning with this trend, data collected in the 1990s suggests that the typical age of alcohol initiation in the United Kingdom was lower than in the United States around this period (Ahlström & Österberg, 2004). Initiation of alcohol use is required of all participants included in the current analysis. Participants who have initiated alcohol use but do not report heavy episodic drinking are assigned a value of zero. The higher rate of alcohol initiation in ALSPAC allows for a greater number of alcohol-exposed participants who report no binge drinking to be included in the analysis.

In late adolescence, the overall mean in Add Health was higher than the overall mean in COGA and ALSPAC. Additionally, the overall mean in COGA was slightly higher than the overall mean in ALSPAC. These differences appear to be accounted for by differences in birth cohort. Note that the mean in COGA subjects who were born in the 80s is similar to the mean of Add Health subjects who were born in the 80s and the mean in COGA subjects who were born in the 90s is similar to the mean of ALSPAC subjects who were born in the 90s. Again, this aligns with previous research suggesting that individuals born in the 90s tend to endorse more restrictive drinking norms compared to those born in the 80s (Keyes et al., 2012).

In early adulthood, means in COGA were uniformly higher across sex, age, and birth cohort compared to Add Health and ALSPAC. Similar to the earlier discussion of mean differences in alcohol use frequency, it may be important to consider the high-risk nature of the COGA sample at this point in development. The aggregation of alcohol use behaviors within

high-risk families may explain this uniform elevation in means, particularly when considering a high-risk alcohol use phenotype. Add Health and ALSPAC are both general population samples and report similar levels of heavy episodic drinking in early adulthood.

In adulthood, means in COGA and Add Health were similar.

GWAS and LDSC Results

Alcohol Use Frequency.

Within Sample GWAS and GWAS Meta-Analysis.

After quality control, 7,057,181 SNPs were included in COGA, 6,520,802 SNPs were included in Add Health, and 9,283,017 SNPs were included in ALSPAC. The meta-analysis in early adolescence and late adolescence of COGA, ALSPAC and Add Health included a total of 10,590,352 SNPs. A number of SNPs in COGA and ALSPAC did not run in the within-sample GWAS and were removed before meta-analysis (COGA early adolescence 32,990 SNPs; COGA late adolescence 235 SNPs; ALSPAC early adolescence 1,312 SNPs; ALSPAC late adolescence 7 SNPs).

Manhattan plots (Figure S56-S63) and quantile-quantile plots (Figure S64-S71) for alcohol use frequency in early adolescence and late adolescence are available in the supplemental material. These results demonstrate p-value inflation in several of the GWAS. P-value inflation can result from polygenicity, the involvement of many variants of small effect in the genetic architecture of a trait, or confounding by uncorrected population stratification (Bulik-Sullivan, Loh, et al., 2015). In cases where sample sizes are small, identification of a large number of true positive effects is unlikely and population stratification is a more likely explanation for an inflated p-value distribution. The sample sizes in the GWAS of alcohol use

frequency in early adolescence and late adolescence were small, suggesting that the inflated p-value distributions here provide evidence of uncorrected population stratification.

In COGA early adolescence, 1,112 SNPs met the threshold for genome-wide significance and an additional 3,156 SNPs met the suggestive threshold. In COGA late adolescence, 23 SNPs met the threshold for genome-wide significance and an additional 360 SNPs met the suggestive threshold.

In Add Health early adolescence, 28 SNPs met the threshold for genome-wide significance and an additional 532 SNPs met the suggestive threshold. In Add Health late adolescence, 2 SNPs met the threshold for genome-wide significance and an additional 262 SNPs met the suggestive threshold.

In ALSPAC early adolescence, 4,019 SNPs met the threshold for genome-wide significance and an additional 6,184 SNPs met the suggestive threshold. In ALSPAC late adolescence, 23 SNPs met the threshold for genome-wide significance and an additional 360 SNPs met the suggestive threshold.

In the early adolescence meta-analysis, 15 SNPs met the threshold for genome-wide significance and an additional 271 SNPs met the suggestive threshold. In the late adolescence meta-analysis, 0 SNPs met the threshold for genome-wide significance and 113 SNPs met the suggestive threshold.

Linkage Disequilibrium Score Regression (LDSC).

Heritability and genetic correlations estimated by LDSC are robust to p-value inflation resulting from population stratification in some cases (Bulik-Sullivan, Loh, et al., 2015). In LDSC, SNP test statistics are regressed on SNP LD scores. The slope of the LD score in this regression is an estimate of SNP-based heritability. SNP LD scores are not expected to be

correlated with their corresponding GWAS test statistics when population stratification drives the effect observed in the GWAS. As a result, confounding in LDSC is expected to increase the intercept of the regression rather than the slope. It should be noted that there are limits to the robustness of LDSC as an estimator of heritability and genetic correlation; for example, LDSC may underestimate population stratification and overestimate heritability when population stratification arises from selection (Bulik-Sullivan, Finucane, et al., 2015; Hellwege et al., 2017). As a result, inflated GWAS results may still be suitable for estimation of heritability and genetic correlations by LDSC in some cases, but should be interpreted with caution.

1,192,817, 1,193,612, 1,193,617, and 1,170,827 HapMap3 SNPs were included in the LD score regression after matching meta-analyzed summary statistics to the reference panel and LD score file for early adolescence, late adolescence, early adulthood, and adulthood, respectively. Effective sample sizes for early adolescence, late adolescence, and early adulthood, and adulthood were 3,104.96, 7,828.56, 9647.64, and 9894.18, respectively. All univariate LD score regression intercepts were near one. Genetic correlations with early adolescence could not be calculated because the SNP-based heritability for early adolescence was negative (H^2_{SNP} = -.06, SE = 0.11) The SNP-based heritability for other traits was modest (late adolescence H^2_{SNP} = .05, SE=0.05; early adulthood H^2_{SNP} = .05, SE = 0.05; adulthood H^2_{SNP} = .08, SE = 0.04). The GenomicSEM package does not return genetic correlations when any SNP-based heritability estimates are negative, but unscaled genetic covariances are available. The genetic covariance between late adolescence and early adulthood (Gcov = -.001, SE = .038) and late adolescence and adulthood (Gcov = -.005, SE = .036) were negative. The genetic covariance between early adulthood and adulthood (Gcov = .049, SE = .034) was positive. Further analysis of alcohol use

frequency with these four developmental periods was not possible because of the negative heritability estimate in early adolescence.

Heavy Episodic Drinking.

Within Sample GWAS and GWAS Meta-Analysis.

After quality control, 7,057,181 SNPs were included in COGA, 6,520,802 SNPs were included in Add Health, and 9,283,017 SNPs were included in ALSPAC. The meta-analysis in early adolescence and late adolescence of COGA, ALSPAC and Add Health included a total of 10,590,352 SNPs. A number of SNPs in COGA and ALSPAC did not run in the within-sample GWAS and were removed before meta-analysis (COGA early adolescence 153,954 SNPs; COGA late adolescence 2,494 SNPs; COGA early adulthood 23 SNPs; COGA adulthood 713 SNPs; ALSPAC early adolescence 24,814 SNPs; ALSPAC late adolescence 8 SNPS; ALSPAC early adulthood 209 SNPs).

Manhattan plots (Figure S72-S86) and quantile-quantile plots (Figure S87-S101) for heavy episodic drinking are available in the supplemental material. Similar to the analysis of alcohol use frequency described above, results demonstrate severe p-value inflation in several of the GWAS.

In COGA early adolescence, 4,778 SNPs met the threshold for genome-wide significance and an additional 1,600 SNPs met the suggestive threshold. In COGA late adolescence, 241 SNPs met the threshold for genome-wide significance and an additional 943 SNPs met the suggestive threshold. In COGA early adulthood 0 SNPs met the threshold for genome-wide significance and 54 SNPs met the suggestive threshold. In COGA adulthood 66 SNPs met the threshold for genome-wide significance and an additional 525 SNPs met the suggestive threshold

In Add Health early adolescence, 78 SNPs met the threshold for genome-wide significance and an additional 784 SNPs met the suggestive threshold. In Add Health late adolescence, 1 SNP met the threshold for genome-wide significance and an additional 149 SNPs met the suggestive threshold. In Add Health early adulthood 0 SNPs met the threshold for genome-wide significance and 136 SNPs met the suggestive threshold. In Add Health adulthood 0 SNPs met the threshold for genome-wide significance and 55 SNPs met the suggestive threshold.

In ALSPAC early adolescence, 19,672 SNPs met the threshold for genome-wide significance and an additional 16,888 SNPs met the suggestive threshold. In ALSPAC late adolescence, 2,260 SNPs met the threshold for genome-wide significance and an additional 7,315 SNPs met the suggestive threshold. In ALSPAC early adulthood 1,450 SNPs met the threshold for genome-wide significance and an additional 4,422 SNPs met the suggestive threshold.

In the early adolescence meta-analysis, 21,460 SNPs met the threshold for genome-wide significance and an additional 16,821 SNPs met the suggestive threshold. In the late adolescence meta-analysis, 2,388 SNPs met the threshold for genome-wide significance and 7,532 SNPs met the suggestive threshold. In the early adulthood meta-analysis, 1,460 SNPs met the threshold for genome-wide significance and 4,641 SNPs met the suggestive threshold. In the adulthood meta-analysis, 44 SNPs met the threshold for genome-wide significance and 295 SNPs met the suggestive threshold.

Linkage Disequilibrium Score Regression (LDSC).

Positive LDSC heritability estimates for all indicators were obtained by applying a minimum meta-analyzed SNP N filter of 2,000 and removing adulthood from the analysis.

1,020,720, 1,162,544, and 1,162,547 HapMap3 SNPs were included in the LD score regression after matching meta-analyzed summary statistics to the reference panel and LD score file for early adolescence, late adolescence, and early adulthood, respectively. Effective sample sizes for early adolescence, late adolescence, and early adulthood were 2,380.48, 7,447.64, and 8,584.84, respectively. All univariate LD score regression intercepts were near 1 (early adolescence= 0.9886, late adolescence= 0.9945, adulthood= 0.9983). The SNP-based heritability for early adolescence was inflated (H^2_{SNP} = .18, SE=0.19), indicating that population stratification evident in the within sample GWAS in early adolescence may not have been fully resolved with LDSC. The SNP-based heritability in other developmental periods was modest (late adolescence H^2_{SNP} = .05, SE=0.07; early adulthood H^2_{SNP} = .07, SE = 0.05). The genetic correlation between early adolescence and late adolescence was negative (rG = -.12, SE = 0.72). The genetic correlations between early adolescence and early adulthood (rG = .04, SE = 0.60) and late adolescence and early adulthood (rG = .48, SE = 0.81) were positive.

Genomic Structural Equation Modeling (gSEM).

Model identification was achieved by fixing the variance of the common factor (η_{common}) to one. With three indicators the genomic structural equation model is saturated and represents a simple recapitulation of the genetic correlations described above. Loadings of genetic variance in each developmental period on the common factor were: early adolescence $\lambda_{12\text{-}14}$ = -.070 (SE = 0.792), late adolescence $\lambda_{15\text{-}17}$ = -1.006 (SE = 9.767), early adulthood $\lambda_{18\text{-}25}$ = -.469 (SE = 4.546). Residual genetic variances in each developmental period after accounting for the common factor were: early adolescence $\varepsilon_{12\text{-}14}$ = .995 (SE = 1.040), late adolescence $\varepsilon_{15\text{-}17}$ = .001 (SE = 19.685), early adulthood $\varepsilon_{18\text{-}25}$ = .780 (SE = 4.286). Extremely high standard errors (19.685), out-of-bounds point estimates (-1.006), and subsequent near-0 residual variances (0.001) related to the

late adolescence indicator signal problems with this model that prevent clear interpretation of results. As a result, additional steps in the analysis plan (gSEM GWAS of the common factor and residuals) were not run for heavy episodic drinking.

Appendix 3. Aim 1 Residual Variance Discovery Analysis

Aim 1 Residual Variance Discovery Analysis

The tolerance setting for matrix inversion was set to a relatively liberal value (1e-50) to allow model fitting to proceed. The 'standard' option was selected to implement Genomic Control, which adjusts SNP standard errors for population stratification by multiplying them by the square root of the LDSC intercept. SNPs were included in the analysis if they were in common between the meta-analyses and the 1000 Genomes referenced panel and had MAF > .01 in the reference panel. The total number of SNPs included in the analysis was 6,707,536.

For GWAS of the residual components of the model ($U_{12\text{-}17}$, $U_{18\text{-}25}$, U_{26+}), paths from each SNP to the residual variances were estimated simultaneously with a path from the SNP to the common factor. This allows the effect of each SNP to be partitioned into a component that is mediated by the common factor and a component that is directly associated with the residual variance. For the adolescence residual GWAS, the early adult residual GWAS, and the common factor GWAS, model identification was achieved by fixing the loading of adulthood to one and constraining all variances to be greater than 0.01. This model identification strategy was successful for the analysis of 6,612,637 SNPs included in the adolescence residual GWAS, 6,700,009 SNPs in the early adult residual GWAS, and 6,651,669 SNPs in the common factor GWAS. For the adult residual GWAS this identification strategy produced many errors, so model identification was achieved by fixing the loading of early adulthood to one and constraining all variances to be greater than 0.01. This model identification strategy was successful for the analysis of 6,705,066 SNPs included in the adult residual GWAS.

Sensitivity analyses were conducted to explore possible differences between the SNPs that were retained in each analysis (Analysis SNPs) and the SNPs that were removed from each

analysis (Error SNPs). The distribution of minor allele frequency (MAF) and chromosome (CHR) in Analysis SNPs and Error SNPs was similar (Adolescence: Figure S102, Early Adult: Figure S104, Adult: Figure S106). Mean test statistics for Error SNPs and Analysis SNPs were calculated from the three input meta-analyzed GWAS. The distribution of mean test statistics in Error SNPs was bimodal, while the distribution in Analysis SNPs was approximately normal across all versions of the analysis (Adolescence: Figure S103, Early Adult: Figure S105, Adult: Figure S107). The range between the smallest and largest test statistic in the input meta-analyzed GWAS was calculated for each SNP, such that a larger test statistic range indicates that the SNP effect varies more across developmental periods. The distribution of test statistic ranges was centered on a higher value in Error SNPs compared to Analysis SNPs (Adolescence: Figure S103, Early Adult: Figure S105, Adult: Figure 107). Together, these sensitivity analyses suggest that errors in these residual GWAS are more likely to occur when the SNP effect varies more across developmental periods. In these cases, a common factor model represents a poor explanation of the effect of the SNP on alcohol use frequency across development and model estimation produces errors. The gSEM GWAS models here each included a path to the common factor and it is likely that estimation of this path produced test statistic deflation. Larger sample sizes may improve the stability of model estimation for these SNPs.

Manhattan plots were constructed for each GWAS to assess SNP effects on residual genetic variance. Quantile-quantile plots for each developmental period and the common factor were constructed to evaluate the distribution of p-values. Manhattan plots and quantile-quantile plots are presented in Figure S108-S113. The quantile-quantile plots for the residual GWAS provide evidence of p-value deflation, reflecting the low statistical power of this analysis. In adolescence, zero SNPs met the threshold for genome-wide significance and 7 SNPs met the

suggestive threshold (rs72691396, rs115778926, rs116734991, rs116097241, rs35459944, rs117048287, and rs147393801). In early adulthood, zero SNPs met the threshold for genome-wide significance and 1 SNP met the suggestive threshold (rs9879300). In adulthood, no SNPs met the threshold for genome-wide significance or the suggestive significance threshold. Attempts to calculate genetic correlations between the model residuals and a series of phenotypes from other studies produced negative heritability estimates in early adulthood and adulthood (early adult H^2_{SNP} = -6e-04 ,adult H^2_{SNP} = -4e-04), preventing further downstream analyses of the residual GWAS results for Aim 1.

Vita

Nathaniel Stembridge Thomas was born on September 5, 1993 in Maryland. He graduated from Centreville High School in Clifton, Virginia in 2011. He received his Bachelor of Science in Psychology from Virginia Commonwealth University in 2014 and his Master of Science in Addiction Studies in 2015, jointly conferred by Virginia Commonwealth University, King's College London, and The University Adelaide. After graduate school, he worked as a research assistant at the Examining, Development, Genes, and Environment lab at Virginia Commonwealth University. In 2018, Nate enrolled in Virginia Commonwealth University's Developmental Psychology PhD program. He earned his Master of Science in Developmental Psychology in 2020. Nathaniel received a Ruth L. Kirschstein National Research Service Award (NRSA) Individual Predoctoral Fellowship (F31) from the National Institute on Alcohol Abuse and Alcoholism in September 2021.