Add Health
The National Longitudinal Study of Adolescent to Adult Health

Report prepared by

David B. Braudt

Kathleen Mullan Harris

Polygenic Scores (PGSs) in the National Longitudinal Study of Adolescent to Adult Health (Add Health) – Release 1

UNC | CAROLINA POPULATION CENTER

Table of Contents

Overview

Research has shown that many outcomes of interest in the health, behavioral, and social sciences are influenced by genetics (Domingue et al. 2016; Plomin et al. 2016; Turkheimer 2000). For most human traits/behaviors, commonly referred to as phenotypes, it appears that the genetic influence on the phenotype is highly polygenic; i.e., there is no single gene that can account for the association between genetic variance and the outcome. Instead, the influence of genetics on the phenotype appears to be due to many small associations across thousands, and possibly millions, of individual single-nucleotide polymorphisms (SNPs, pronounced snips) (Chabris et al. 2015). Polygenic Scores allow researchers to avoid the methodological complexities of including thousands, or millions, of covariates in their analyses by condensing, into a single measure, the associations between individual SNPs and the phenotype of interest (Plomin, Haworth, and Davis 2009).

Polygenic Scores (PGSs), sometimes referred to as polygenic risk scores or genetic risk scores, represent a general measure of the influence of additive genetics on a specific phenotype. The calculation of PGSs relies on summary statistics from genome-wide association studies (GWASs) to create a weighted sum of the associations between allele frequencies at individual SNPs and the associated phenotype. The estimated associations, i.e. beta-coefficients, for each SNP from a GWAS, conducted on a large independent sample, are multiplied by the allele frequencies of the same SNPs for individuals in the sample for which the PGS is being created. This process yields a hypothesis free measure of the cumulative additive genetic influences on the phenotype being studied. PGSs are hypothesis free because they aggregate the individual associations between SNPs and the phenotype, thus removing the possibility of investigating links between specific biological/genetic pathways and the phenotype. While the hypothesis free nature of PGSs dilutes the ability to detect biological pathways, it allows researchers to capture the broad influence of genetics in various analyses (Belsky and Israel 2014; Dudbridge 2016).

Since PGSs represent the associations between SNPs across the entire genome and a phenotype in a single measure, they can easily be incorporated into many of the quantitative analyses common in economics (Benjamin et al. 2012), sociology (Conley 2016), social stratification (Braudt forthcoming), as well as other social, behavioral, and health sciences (Belsky and Israel 2014). The PGSs described in this document are meant to facilitate the use of PGSs among users of the National Longitudinal Study of Adolescent to Adult Health (Add Health).

Data

Add Health is an ongoing nationally representative longitudinal study of adolescents in the U.S. who were in grades 7-12 in 1994-5. Wave I (1994-5, 79% response rate) included a sample of 80 high schools and 52 middle schools chosen from a stratified sample according to region, urbanicity, school size, school type, and racial and ethnic composition with probability of selection proportional to size. With five waves of data—Wave II (1996, 89% response rate), Wave III (2001-2, 77% response rate), Wave IV (2008, 80% response rate), and Wave V (2016-18, in the field)—and information on adolescents' fellow students, school administrators, parents, siblings, friends, and romantic pairs, as well as extensive longitudinal geospatial data on neighborhood measures such as income, poverty, unemployment, the availability and use of health services, crime, religious membership, and social programs, Add Health represents one of the richest longitudinal studies of health and behavior in the U.S. available today. For more information on the Add Health study design see Harris (2013).

Genome-wide Data

As part of the Wave IV data collection, saliva samples were obtained from consenting participants (96% of Wave IV respondents). Approximately 12,200, or 80% of those participants, consented to long-term archiving and were consequently eligible for genome-wide genotyping. Genotyping was done on two Illumina platforms, with approximately 80% of the sample genotyping performed with the Illumina Omni1-Quad BeadChip and 20% genotyped with the Illumina Omni2.5-Quad BeadChip. After quality control procedures, genotyped data are available for 9,974 individuals (7,917 from the Omni1 chip and 2,057 from the Omni2 chip) on 609,130 SNPs common across both genotyping platforms (Highland, Heather M.; Avery, Christy L.; Duan, Qing; Li, Yun; Mullan Harris, Kathleen 2018). For more information on the genotyping and quality control procedures see the Add Health GWAS QC report online at: http://www.cpc.unc.edu/projects/addhealth/documentation/guides/copy_of_AH_GWAS_QC.pdf.

Ancestry Specific Samples

To account for population stratification,[1] we restrict the Add Health genotyped sample to four genetic ancestry groups: European ancestry, African ancestry, Hispanic ancestry, and East Asian ancestry. To identify respondents in these four genetic ancestry groups we use principal component analysis on all unrelated members of the full Add Health genotyped sample and project those estimates onto the small remainder of related individuals. Figure 1 provides a visual depiction of this process with the respective cut-offs for inclusion in the genetic ancestry groups are represented by the black rectangles.

Each ancestry group is defined by distance from the mean of the first two principal components of the genetic data as shown in Figure 1. To be included in the Hispanic, East Asian, and European ancestry groups

---

[1] Population stratification refers to differences in genetic variation between geographical ancestry groups. Due to the genetic bottle neck created by the small number of humans (roughly 2,000 individuals) who migrated out of Africa early in human history, and the tendency for people to procreate with individuals from the same or nearby geographic regions, genetic variance across the entire genome is highly correlated with geography (for a more detailed discussion see Conley and Fletcher 2017:84–112). Importantly, genetic ancestry should not be confused with race or ethnicity. Race and ethnicity are social constructs based on a multitude of factors, of which ancestry may be included depending on historical and societal differences in racialization (Omi and Winant 1994).

Figure 1: Principal Components and Ancestry-Specific Samples
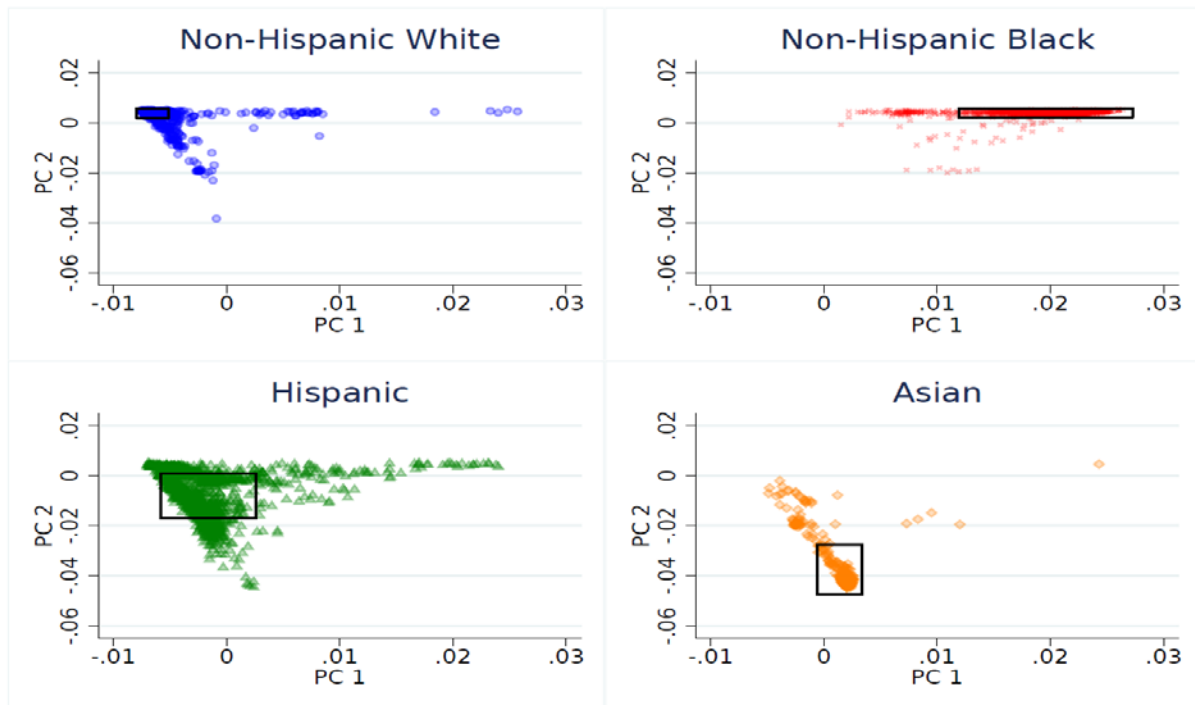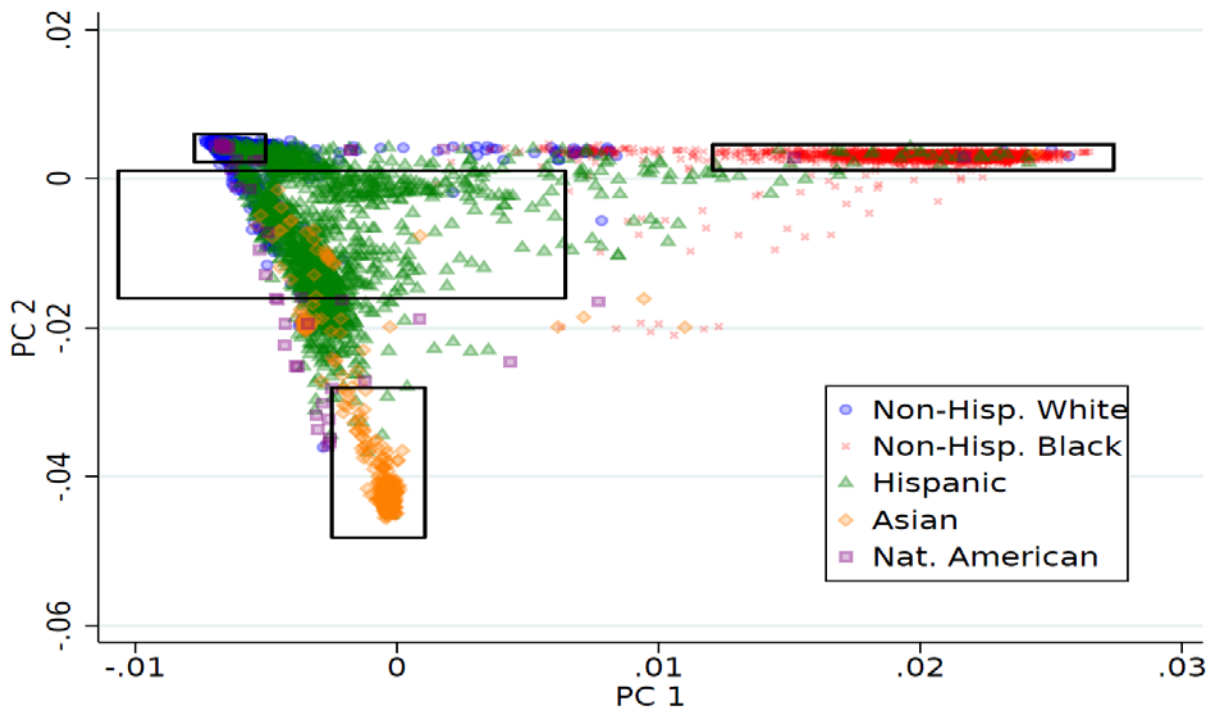
Notes: (1) The principal components are estimated based on the allele frequencies of the 609,130 genotyped SNPs in the Add Health genome-wide data. (2) Black rectangles represent within Add Health genetic ancestry group boundaries. East Asian, Hispanic, and European ancestry groups are defined as: +/- 1 standard deviation from the within group mean of the first and second principal components estimated from individuals who self-identified as Asian, Hispanic, and non-Hispanic White respectively. African ancestry is defined as +/- 2 standard deviations from the mean of the first principal component and +/- 1 standard deviation from the mean of the second principal component estimated from all individuals in the genotyped sample who self-identified as non-Hispanic Black.

individuals must be within +/- 1 standard deviation of the mean of the first two principal components of the genetic data estimated from all individuals in the Add Health genome-wide data who self-identified as Hispanic, Asian, and non-Hispanic White respectively. To be included in the African ancestry group individuals must be within +/- 2 standard deviation of the mean of the first principal component and +/- 1 standard deviation of the mean of the second principal component estimated from all individuals in the genome-wide data who self-identified as non-Hispanic Black. While genetic ancestry and self-identified race/ethnicity are strongly correlated (0.89), they are two separate constructs (see footnote 1). Consequently, not all individuals included in a given ancestry may self-identify or be socially-identified as the same race and/or ethnicity as other members of their ancestry group. Table 1 provides a depiction of the resulting sample sizes for each genetic ancestry group and their correspondence to self-identified race/ethnicity.

As a sensitivity analysis, we repeated the above process using group specific medians of the first two principal components of the genetic data as centroids instead of group means. While sample sizes for the four ancestry groups using the median as the centroid were comparable to those using the means of the principal components as centroids, in all cases, using the mean as the centroid resulted in larger sample size. Given the slight increase in sample size, we define within Add Health genetic ancestry groups as distance from the mean of the first two principal components of the genotyped data.

Table 1: Add Health Polygenic Score Sample Sizes by Genetic Ancestry

Self-Identified Race/Ethnicity	Ancestry				
	European	African	East Asian	Hispanic	Total
Non-Hispanic White	5,644	5	0	105	5,754
Non-Hispanic Black	0	1,939	0	1	1,940
Native American	14	2	0	7	23
Asian	0	1	422	26	449
Hispanic	70	27	15	849	961
Missing	0	2	0	0	0
Total Sample Size	5,728	1,976	437	988	9,129

Methods

Polygenic Scores

We calculate polygenic scores (PGSs) following the procedure outlined in Dudbridge (2013). PGSs are a weighted sum of the regression coefficient for each SNP from an independent GWAS for each phenotype and allele frequencies for the same SNPs in the Add Health genome-wide data. For example, the raw PGS of educational attainment for an individual, i, is calculated as:

(1)

$$PGS_{EDU_i} = \sum_{j=1}^{k} \beta_j SNP_{ij}$$

where, SNP_{ij} is the allele frequency of the jth SNP for the ith individual and β_j is the estimated association between SNP j and the number of years of education completed as reported in the GWAS summary statistics based on an independent sample (i.e. if a GWAS included Add Health, the GWAS is re-estimated excluding the Add Health sample and beta-weights from the re-estimated GWAS are used in Equation 1). To automate this process, we use a modified version of the PRSice wrapper for R within the PLINK software package (Chang et al. 2015). Once calculated, the raw PGSs are standardized ($\mu_{PGS} = 0 \wedge \sigma_{PGS} = 1$) within ancestry groups. The standardization of PGSs within ancestry groups is done to account for between group population stratification. To control for within group population stratification, we recommend that researchers include at least the first five ancestry-specific principal components of the genome-wide data in all analyses using PGSs. It should be noted that these are imperfect controls, and PGSs should be re-calculated using GWAS weights for specific ancestry groups when such GWAS become available.

Ancestry-specific Principal Components

Ancestry-specific principal components are estimated following a similar procedure to the general principal components, with the sample restricted to individuals in the respective genetic ancestry groups. The process used starts with the genetic ancestry groups defined above, randomly removes one sibling of any sibling pairs in the data, estimates the first 20 principal components for all unrelated individuals in the ancestry specific sample, and then projects those principal components onto the small number of related individuals within each ancestry group.

A Note on Raw PGSs

We report a few descriptive statistics as well as density plots of the raw PGSs for each phenotype as calculated in Equation (1) above. While this information is presented for clarity and to ensure transparency, a few comments must be made on comparing raw PGSs between phenotypes and/or datasets.

Raw PGSs, i.e. PGSs that have not been standardized within ancestry groups to have a mean = 0 and a standard deviation = 1, reported in this study are a function of the number of SNPs in the Add Health genome-wide data that match SNPs for which the respective GWAS reports summary statistics (see Equation 1). Any differences in means, medians, and ranges of raw PGSs between datasets in which the number of SNPs in the genome-wide data differ between datasets is likely to primarily be a function of the beta weights for the non-overlapping SNPs between the two datasets. Consequently, it's unclear how, or if, such differences in the raw PGSs can be interpreted as evidence of "true" differences in a genetic propensity for the phenotype for which the PGS was created versus a methodological artifact stemming from differences in the number of genotyped SNPs included in the PGS calculations within each study. One possible alternative for researchers seeking to compare raw PGSs, based on the same GWAS summary statistics, between studies is to examine the relative distributions and variances of the raw PGSs. We provide density plots of the raw PGSs following the respective table of descriptive statistics for that PGS.

Similar to the above, comparing raw PGSs for two different traits, even if calculated from the same genome-wide sample, is not advised due to the fact that scores are dependent upon the weights from different GWASs. While there is considerable homogeneity between GWASs in how they are estimated, the type and number of controls, and method of meta-analysis, not all GWAS follow the same methodology, nor do they use the same sets of data. Consequently, it's unclear what, if any, substantive meaning comparisons of raw PGSs for different phenotypes may have.

Using Add Health PGSs

Add Health is releasing PGSs for European, African, Hispanic, and East Asian genetic ancestry groups. However, researchers should be aware that **PGSs for individuals not of the same ancestry group(s) as the GWAS from which summary statistics are retrieved may be less predictive** (Martin et al. 2017; Ware et al. 2017). While there are several proposed explanations for this difference in predictive ability, one important reason is a lack of statistical power to account for population stratification between the ancestry group for which the PGS is calculated and the ancestry group(s) included in the GWAS. To help account for potential bias due to population stratification and/or differences in genetic structure within ancestry groups we include the first ten ancestry-specific principal components of the genetic data with the PGSs. **It is strongly recommended that researchers perform analyses separately by ancestral groups or, at the very least, include the first five ancestry-specific principal components as covariates in analyses using these PGSs** (Price et al. 2006).

In order to minimize the risk of deductive disclosure, the order of the ancestry-specific principal components (PCS) of the genetic data are randomized in sets of five. Therefore, PCs must be included as sets: PC1-PC5, PC6-PC10, PC11-PC15, PC16-PC20 if any of the PCs of a set are included in analyses. For example, if a research wishes to include the first two ancestry-specific PCs as covariates in their analyses (i.e. PC1 and PC2) they must also include PC3-PC5.

Citing this Document and Data

Please include the following citation in any report, publication, and/or presentation based on the data in this release of the Add Health PGSs as well as the citation for the reference GWAS:

Braudt, David B. and Kathleen Mullan Harris. 2018. "Polygenic Scores (PGSs) in the National Longitudinal Study of Adolescent to Adult Health (Add Health) – Release 1." Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.

References

Belsky, Daniel W. and Salomon Israel. 2014. "Integrating Genetics and Social Science: Genetic Risk Scores." *Biodemography and Social Biology* 60(2):137–55.

Benjamin, Daniel J. et al. 2012. "The Promises and Pitfalls of Genoeconomics." *Annual Review of Economics* 4(1):627–62.

Bolton, Jennifer L. et al. 2014. "Genome Wide Association Identifies Common Variants at the SERPINA6/SERPINA1 Locus Influencing Plasma Cortisol and Corticosteroid Binding Globulin." *PLOS Genetics* 10(7):e1004474.

Braudt, David B. forthcoming. "Sociogenomics in the 21st Century: An Introduction to the History and Potential of Genetically Informed Social Science." *Sociology Compass*.

Chabris, Christopher F., James J. Lee, David Cesarini, Daniel J. Benjamin, and David I. Laibson. 2015. "The Fourth Law of Behavior Genetics." *Current Directions in Psychological Science* 24(4):304–312.

Chang, Christopher C. et al. 2015. "Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets." *GigaScience* 4:7.

Conley, Dalton. 2016. "Socio-Genomic Research Using Genome-Wide Molecular Data." *Annual Review of Sociology* 42(1):275–99.

Conley, Dalton and Jason Fletcher. 2017. *The Genome Factor : What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future*. Princeton: Princeton University Press.

Domingue, Benjamin W. et al. 2016. "Genome-Wide Estimates of Heritability for Social Demographic Outcomes." *Biodemography and Social Biology* 62(1):1–18.

Dudbridge, Frank. 2016. "Polygenic Epidemiology: Polygenic Epidemiology." *Genetic Epidemiology* 40(4):268–72.

Dudbridge, Frank. 2013. "Power and Predictive Accuracy of Polygenic Risk Scores" edited by N. R. Wray. *PLoS Genetics* 9(3):e1003348.

Harris, Kathleen Mullan. 2013. "The Add Health Study: Design and Accomplishments." *Chapel Hill: Carolina Population Center, University of North Carolina at Chapel Hill*.

Highland, Heather M.; Avery, Christy L.; Duan, Qing; Li, Yun; Mullan Harris, Kathleen. 2018. "Quality Control Analysis of Add Health GWAS Data."

Martin, Alicia R. et al. 2017. "Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations." *The American Journal of Human Genetics*.

Omi, Michael and Howard |. Winant. 1994. *Racial Formation in the United States : From the 1960s to the 1990s*. 2nd ed. New York: Routledge.

Plomin, Robert, John C. DeFries, Valerie S. Knopik, and Jenae M. Neiderhiser. 2016. "Top 10 Replicated Findings From Behavioral Genetics." *Perspectives on Psychological Science* 11(1):3–23.

Plomin, Robert, Claire MA Haworth, and Oliver SP Davis. 2009. "Common Disorders Are Quantitative Traits." *Nature Reviews Genetics* 10(12):872.

Price, Alkes L. et al. 2006. "Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies." *Nature Genetics* 38(8):904–9.

Turkheimer, Eric. 2000. "Three Laws of Behavior Genetics and What They Mean." *Current Directions in Psychological Science* 9(5):160–164.

Ware, Erin B. et al. 2017. "Heterogeneity in Polygenic Scores for Common Human Traits." *BioRxiv* 106062.

Polygenic Scores

In the following section we provide descriptions of each polygenic score (PGS) as well as the source of the GWAS summary statistics. In Table 2 we provide a list of phenotypes for which Add Health is releasing PGSs along with the ancestry group(s) included in the discovery GWAS analyses. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

### Table 2: PGS Phenotypes and Ancestry Groups included in the Respective GWAS

Phenotype	GWAS Ancestry Group(s)
Coronary Artery Disease	European
Myocardial Infarction	European, South Asian, East Asian
Plasma Cortisol	European
Low-density Lipoprotein Cholesterol	European
High-density Lipoprotein Cholesterol	European
Total Cholesterol	European
Triglycerides	European
Type II Diabetes (2012)	European
Type II Diabetes (2014)	European, East Asian, South Asian, Mexican/Mexican-American
BMI	European
Waist Circumference	European
Waist-to-Hip Ratio	European
Height	European
Age at Menarche	European
Age at Menopause	European
Number of Children Ever Born	European
Age at First Birth	European
Ever/Current Smoker	European
Number of Cigarettes per day	European
Extraversion	European
Attention-deficit/hyperactivity Disorder (2010)	European
Attention-deficit/hyperactivity Disorder (2017)	European, Chinese
Bipolar Disorder	European
Major Depressive Disorder (2013)	European
Major Depressive Disorder (2018)	European
Schizophrenia	European, East Asian
Mental Health Cross Disorder	European
Alzheimer's Disease	European
Educational Attainment (2016)	European
Educational Attainment (2018)	European

Coronary Artery Disease

GWAS Summary Statistic Source: Schunkert, Heribert et al. 2011. "Large-Scale Association Analysis Identifies 13 New Susceptibility Loci for Coronary Artery Disease." *Nature Genetics* 43(4):333–38.

GWAS Ancestry Group(s): European

The PGSs for coronary artery disease (CAD) were created using results from a 2011 study conducted by the Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) consortium. The GWAS meta-analysis files are publicly available and can be downloaded from: http://www.cardiogramplusc4d.org.

The coronary artery disease GWAS meta-analysis consisted of 14 studies with a total of 22,233 individuals with CAD (cases) and 64,762 without CAD (controls) of European descent imputed to the HapMap3 CEU panel. Replication was performed in a sample of 56,682 individuals (approximately half cases and half controls). Study-specific GWAS adjusted for age of onset of CAD (cases) or age of recruitment (controls), gender, and genetic principal components.

The Add Health PGSs for CAD are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	497.54	692.85	590.53	590.56	0.32
African ancestry	1976	522.15	663.65	601.36	601.46	0.44
Hispanic ancestry	988	521.64	661.34	590.14	590.21	0.75
Asian ancestry	437	508.46	618.88	570.86	569.82	0.96

Coronary Artery Disease
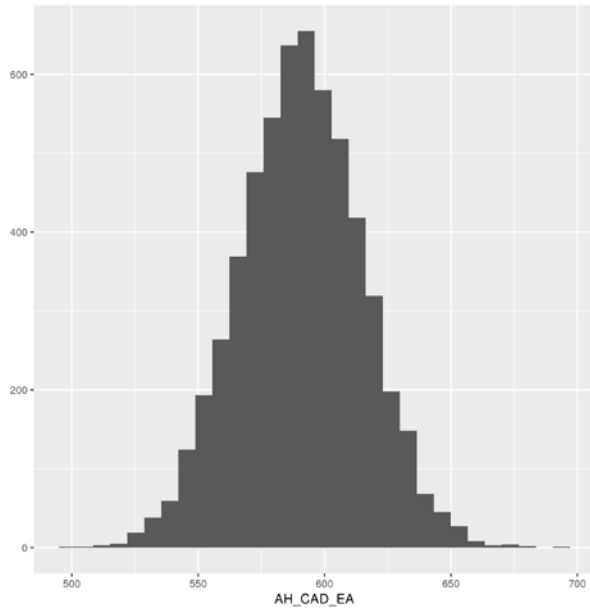
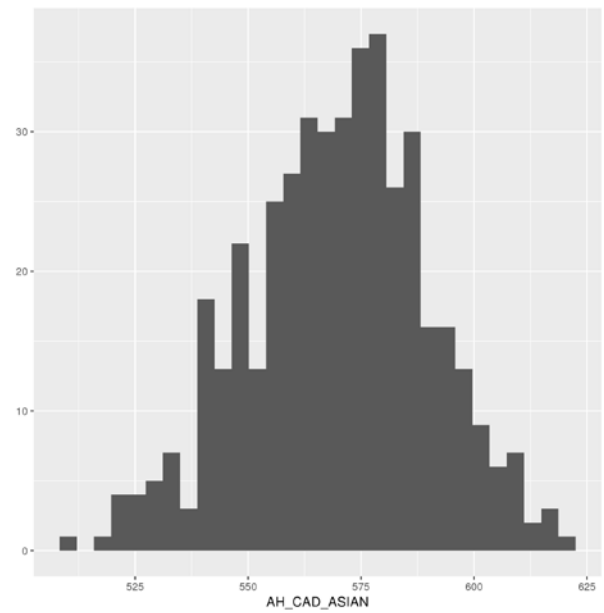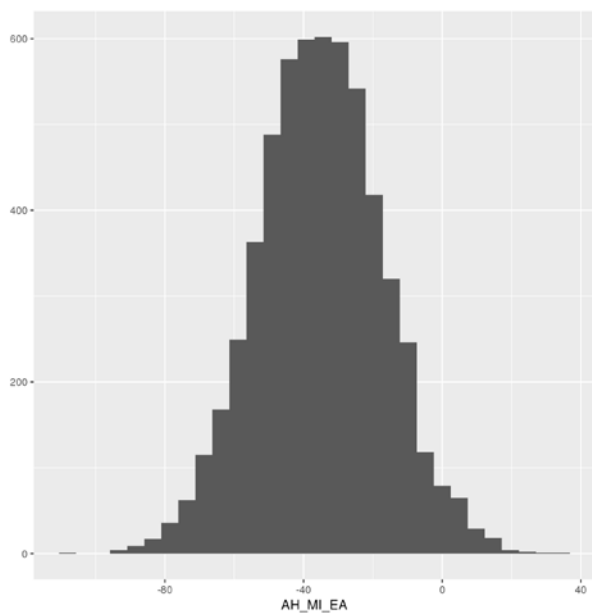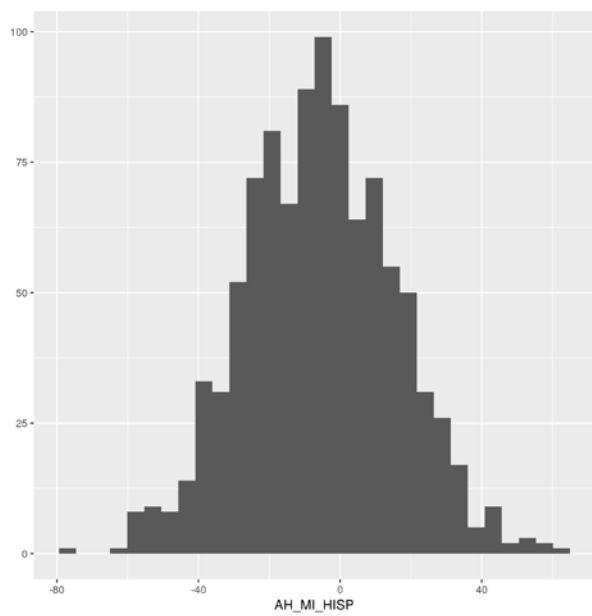### European ancestry raw PGS distribution

African ancestry raw PGS distribution
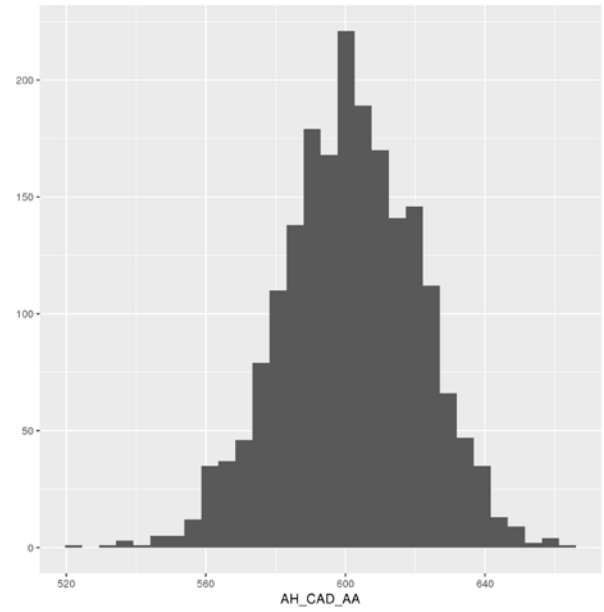
Hispanic ancestry raw PGS distribution
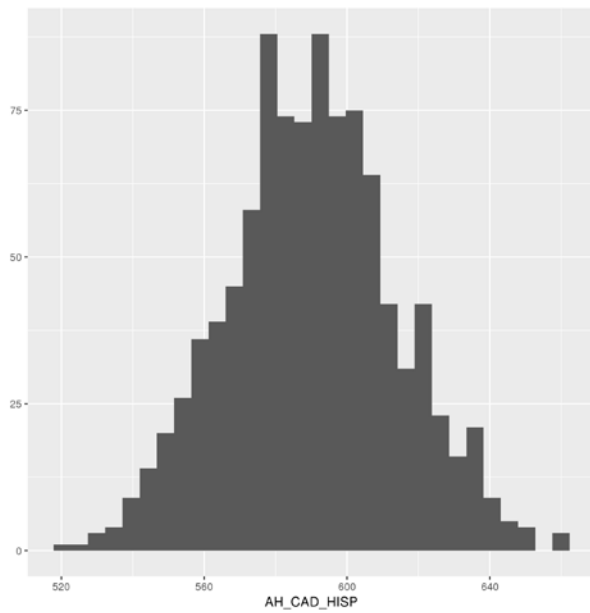
Asian ancestry raw PGS distribution

Myocardial Infarction

GWAS Summary Statistic Source: Nikpay, Majid et al. 2015. "A Comprehensive 1,000 Genomes-Based Genome-Wide Association Meta-Analysis of Coronary Artery Disease." Nature Genetics 47(10):1121–30.

GWAS Ancestry Group(s): European (77%), South Asian (13% - India and Pakistan), East Asian (6% - China and Korea), Hispanic and/or African American (4%).

The PGSs for myocardial infarction were created using 2015 results from a subgroup analysis of coronary artery disease conducted by the Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) consortium. The GWAS meta-analysis files are publicly available and can be downloaded from: http://www.cardiogramplusc4d.org.

The GWAS is a meta-analysis of 48 studies of mainly European, South Asian, and East Asian, descent (as described above) imputed using the 1000 Genomes phase 1 v3 training set with 38 million variants. The study included 60,801 coronary artery disease cases and 123,504 controls. Case status was defined by an inclusive coronary artery disease diagnosis (for example, myocardial infarction, acute coronary syndrome, chronic stable angina or coronary stenosis of >50%). MI subgroup analysis was performed in cases with a reported history of MI (approximately 70% of the total number of cases).

The Add Health PGSs for myocardial infarction (MI) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	-106.45	35.97	-35.42	-35.36	0.24
African ancestry	1976	-42.45	67.83	16.87	17.06	0.33
Hispanic ancestry	988	-78.06	61.43	-5.78	-5.65	0.67
Asian ancestry	437	-11.68	96.54	42.74	42.24	0.77

Myocardial Infarction

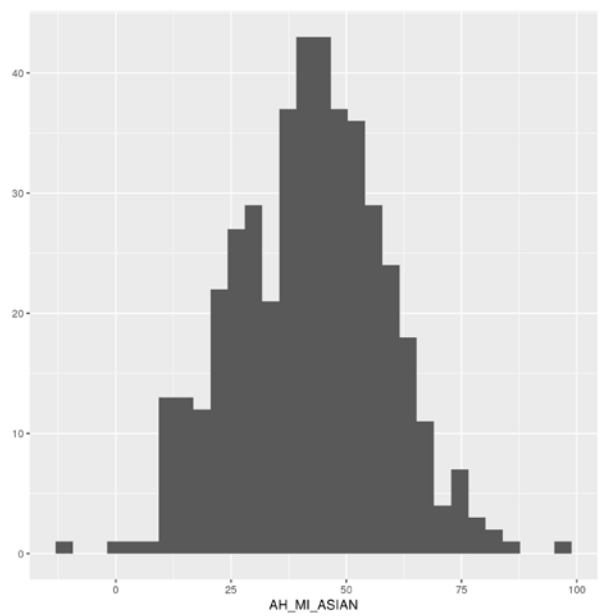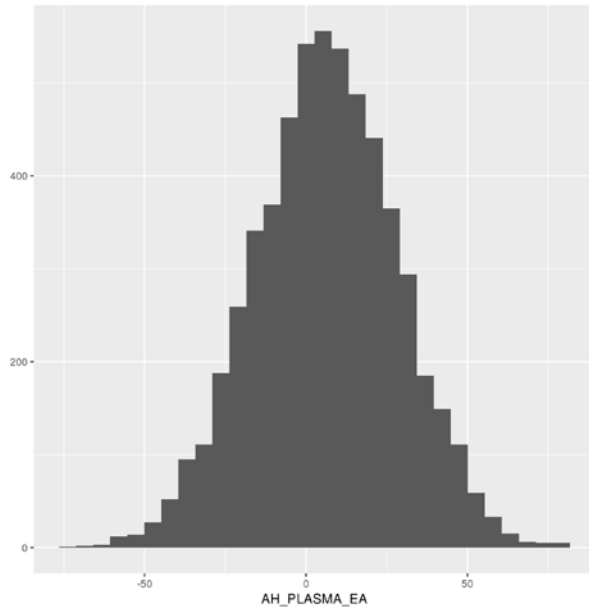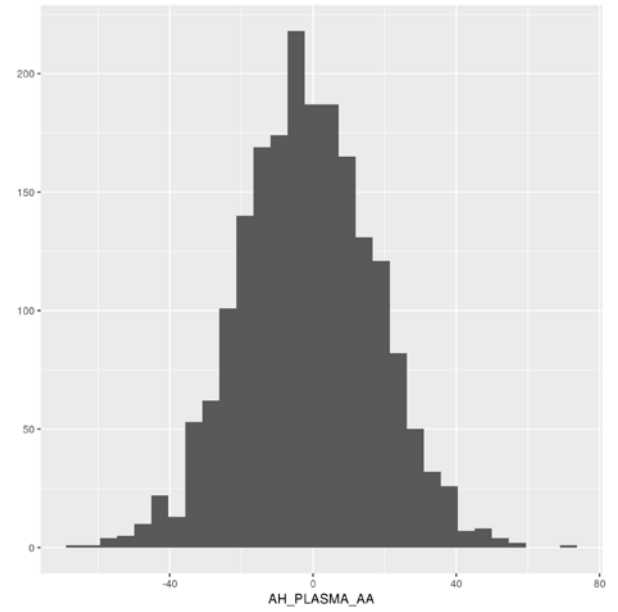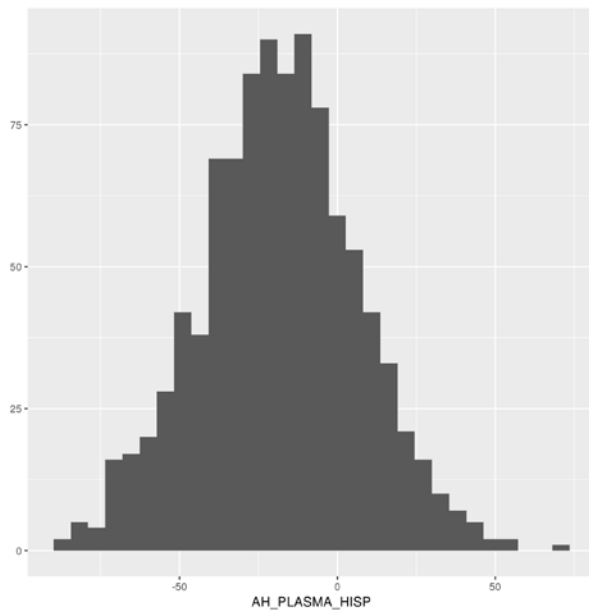### European ancestry raw PGS distribution

African ancestry raw PGS distribution
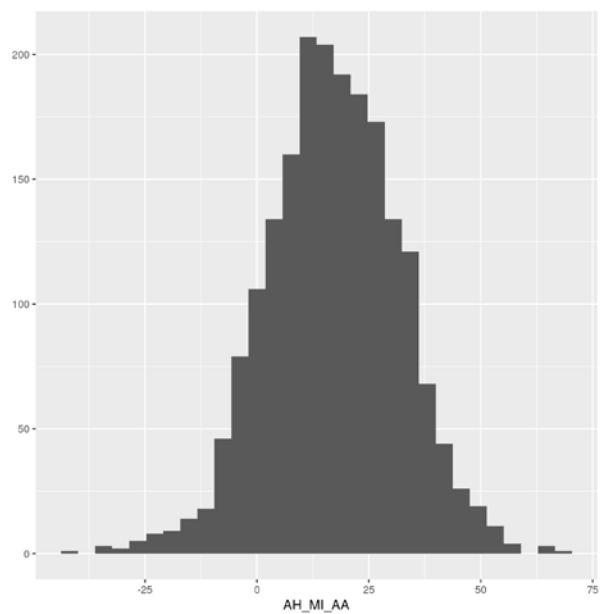
Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Plasma Cortisol

GWAS Summary Statistic Source: Bolton, Jennifer L. et al. 2014. "Genome Wide Association Identifies Common Variants at the SERPINA6/SERPINA1 Locus Influencing Plasma Cortisol and Corticosteroid Binding Globulin." PLOS Genetics 10(7):e1004474.

GWAS Ancestry Group(s): Western European

The PGSs for plasma cortisol were created using results from a 2014 GWAS conducted by the CORtisol NETwork (CORNET). The GWAS meta-analysis files are available through CORNET by request only.

The discovery phase of the meta-analysis included 12,597 participants aged 14-102 of western European descent from 11 studies. The replication sample consisted of 2,795 participants from three independent cohorts: Raine Study (n=797), ET2DS (n=1,069), and MrOS-Sweden (n=929) (Bolton et al. 2014:8). Analyses were adjusted for age and sex. Additional sensitivity analyses showed no difference when time of sampling was included as a covariate.

The Add Health PGSs for plasma cortisol (PI) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	-71.35	81.57	6.39	6.23	0.29
African ancestry	1976	-65.44	72.37	-1.87	-1.52	0.42
Hispanic ancestry	988	-88.81	69.20	-18.22	-18.46	0.79
Asian ancestry	437	-114.93	5.76	-56.92	-57.72	0.96

Plasma Cortisol

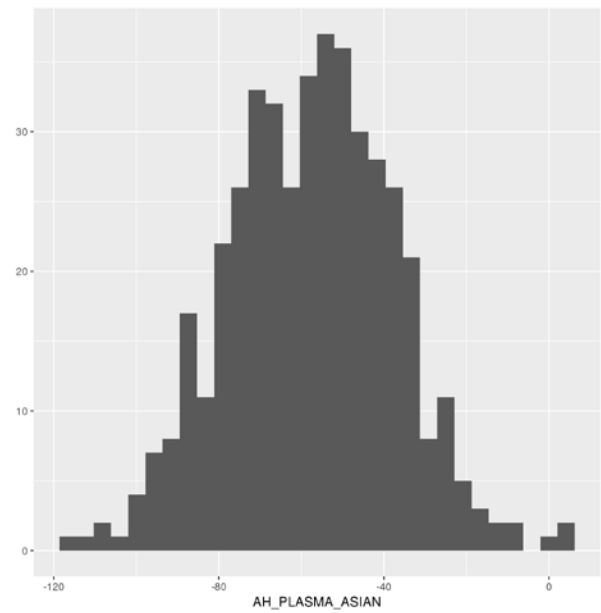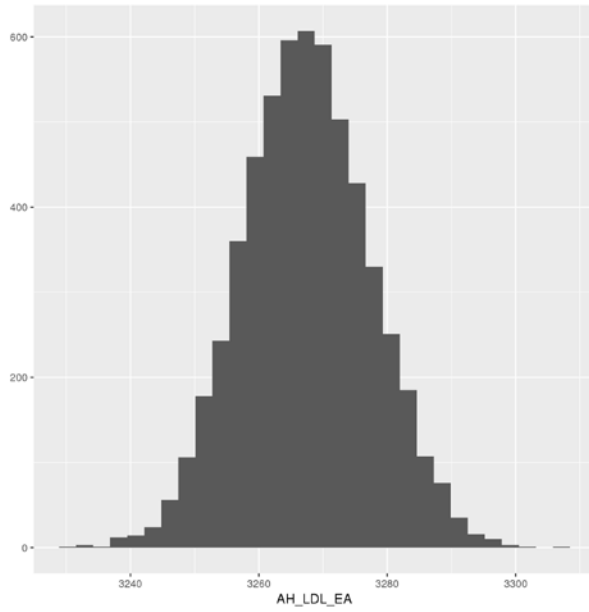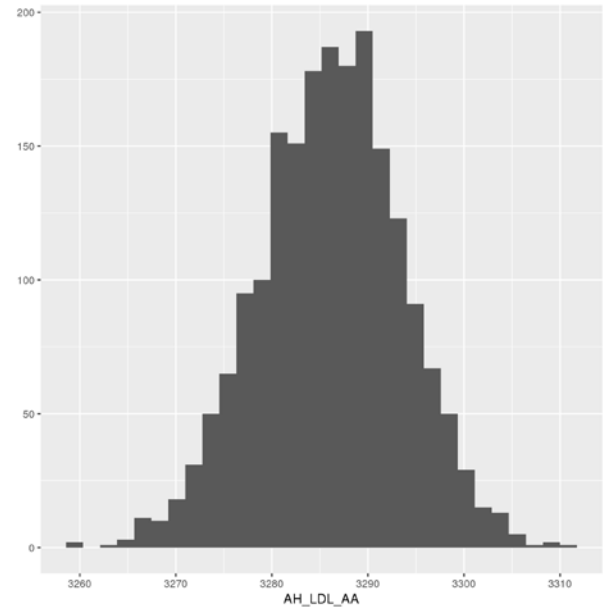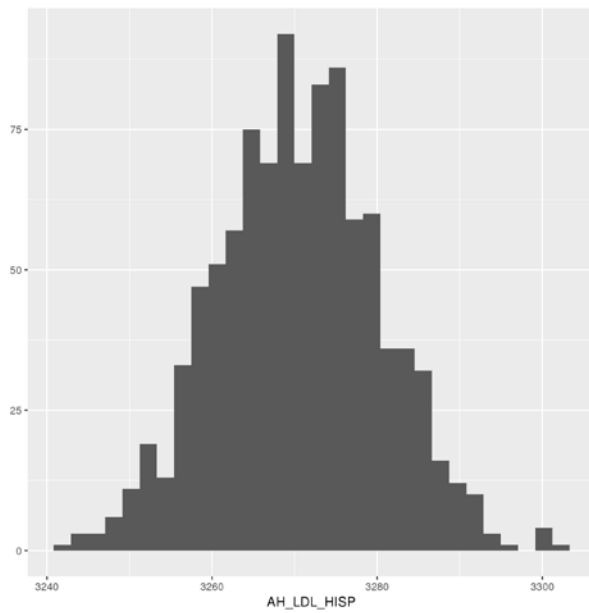### European ancestry raw PGS distribution

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Low-density Lipoprotein (LDL) Cholesterol

GWAS Summary Statistic Source: Willer, Cristen J. et al. 2013. "Discovery and Refinement of Loci Associated with Lipid Levels." *Nature Genetics* 45(11):1274–83.

GWAS Ancestry Group(s): European

The PGSs for low-density lipoprotein cholesterol (LDL) were created using results from a 2013 GWAS conducted by the Global Lipids Genetics Consortium (GLGC). The GWAS meta-analysis files are publically available from: http://csg.sph.umich.edu/willer/public/lipids2013/.

The discovery phase of the meta-analysis included 188,577 individuals of European ancestry from 45 studies, while the replication sample consisted of 7,898 individuals of non-European descent. Lipid levels were measured after eight or more hours of fasting with individuals taking lipid lowering medication excluded from the study. Study-specific GWAS adjusted for age, age-squared, sex, and genetic principal components.

The Add Health PGSs for LDL are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
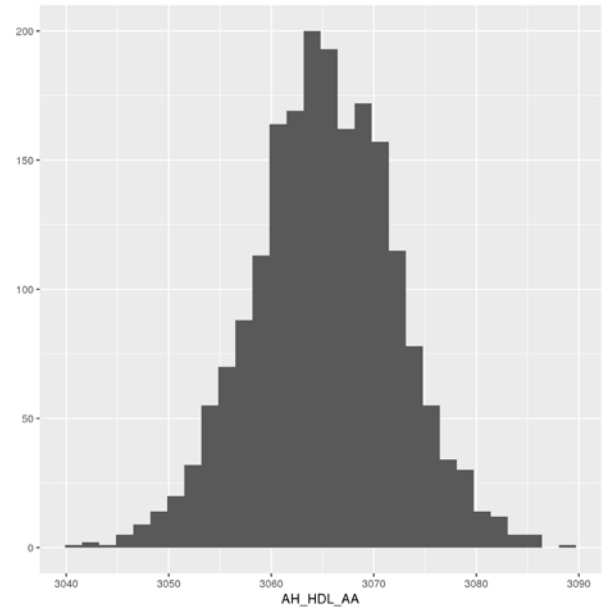
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	3229.96	3306.87	3267.19	3267.37	0.13
African ancestry	1976	3259.41	3310.81	3286.24	3286.08	0.17
Hispanic ancestry	988	3240.91	3301.33	3270.43	3270.50	0.31
Asian ancestry	437	3248.88	3301.42	3276.09	3276.33	0.38

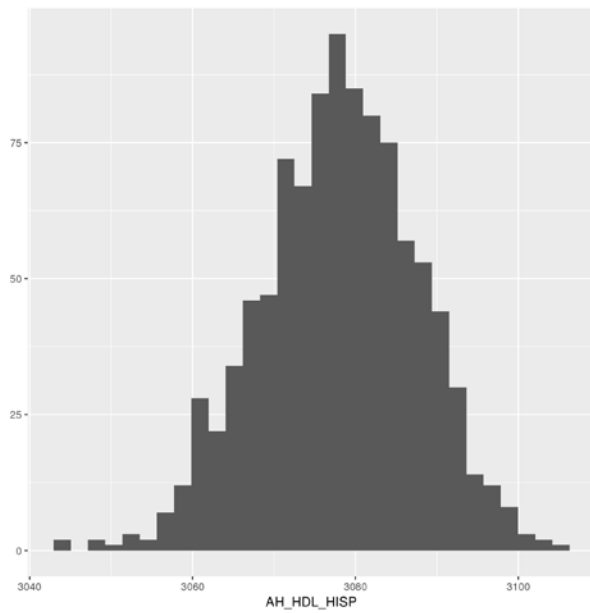# Low-density Lipoprotein Cholesterol (LDL)

European ancestry raw PGS distribution

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution
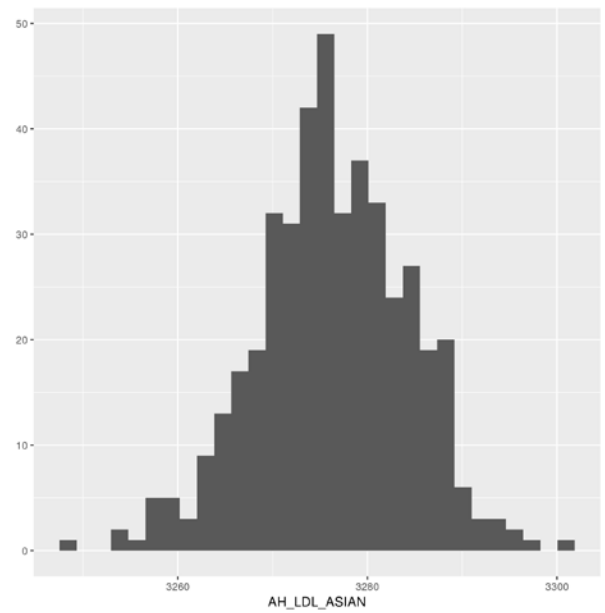
High-density Lipoprotein (HDL) Cholesterol

GWAS Summary Statistic Source: Willer, Cristen J. et al. 2013. "Discovery and Refinement of Loci Associated with Lipid Levels." *Nature Genetics* 45(11):1274–83.

GWAS Ancestry Group(s): European

The PGSs for high-density lipoprotein cholesterol (HDL) were created using results from a 2013 GWAS conducted by the Global Lipids Genetics Consortium (GLGC). The GWAS meta-analysis files are publically available from: http://csg.sph.umich.edu/willer/public/lipids2013/.

The discovery phase of the meta-analysis included 188,577 individuals of European ancestry from 45 studies, while the replication sample consisted of 7,898 individuals of non-European descent. Lipid levels were measured after eight or more hours of fasting with individuals taking lipid lowering medication excluded from the study. Study-specific GWAS adjusted for age, age-squared, sex, and genetic principal components.

The Add Health PGSs for HDL are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	3030.99	3104.16	3071.52	3071.37	0.13
African ancestry	1976	3040.33	3088.49	3065.20	3065.22	0.16
Hispanic ancestry	988	3043.05	3104.31	3078.15	3077.90	0.30
Asian ancestry	437	3074.88	3117.38	3093.48	3093.41	0.39

High-density Lipoprotein Cholesterol (HDL)

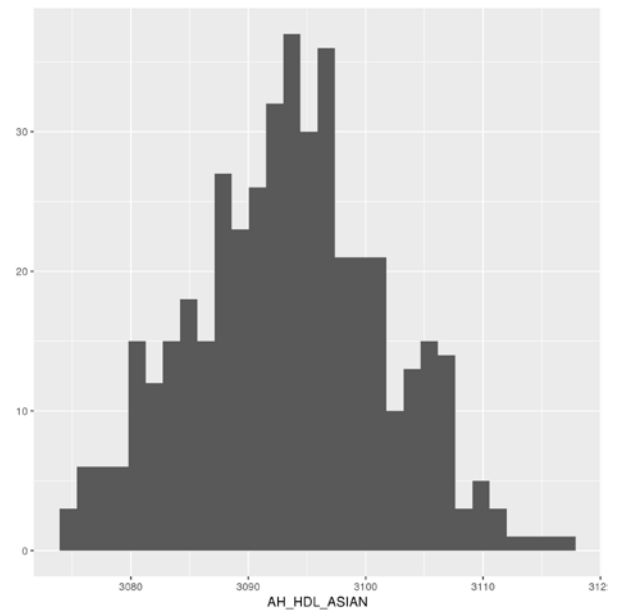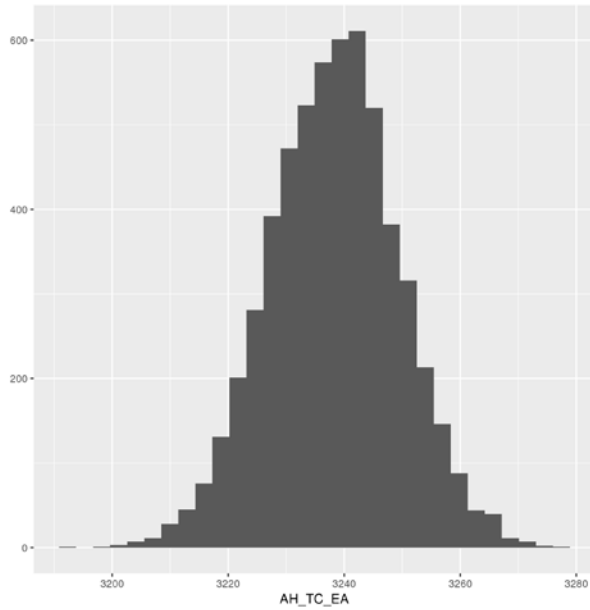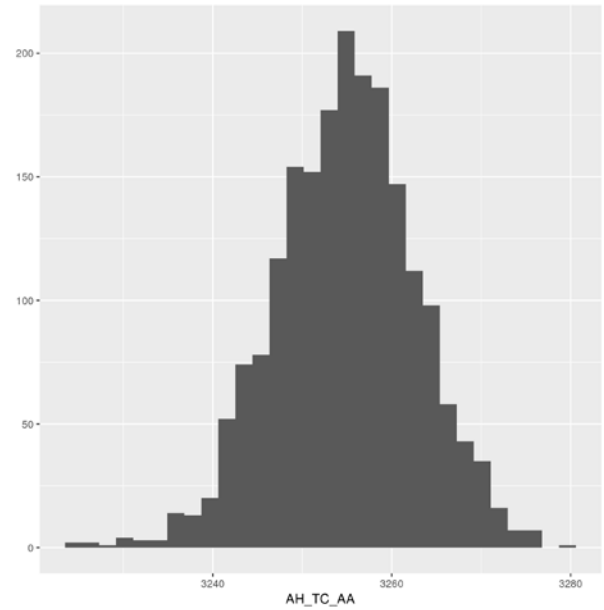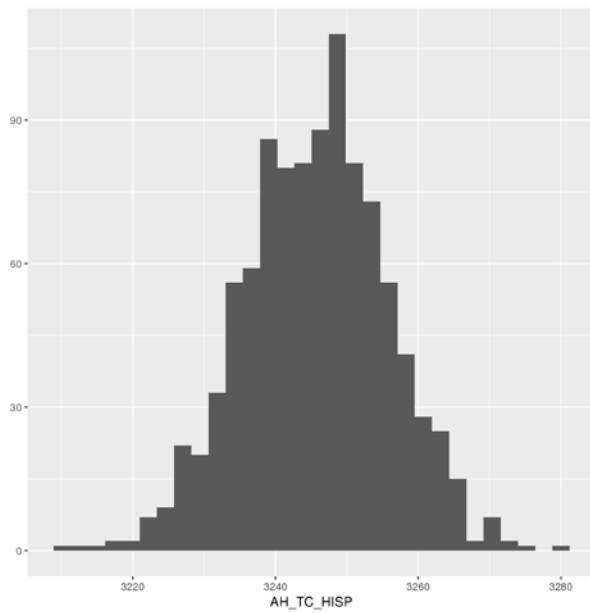### European ancestry raw PGS distribution
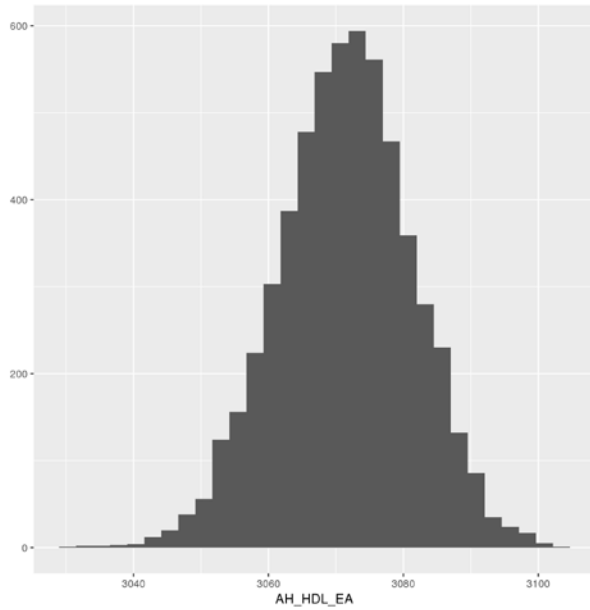
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Total Cholesterol

GWAS Summary Statistic Source: Willer, Cristen J. et al. 2013. "Discovery and Refinement of Loci Associated with Lipid Levels." *Nature Genetics* 45(11):1274–83.

GWAS Ancestry Group(s): European

The PGSs for total cholesterol were created using results from a 2013 GWAS conducted by the Global Lipids Genetics Consortium (GLGC). The GWAS meta-analysis files are publically available from: http://csg.sph.umich.edu/willer/public/lipids2013/.

The discovery phase of the meta-analysis included 188,577 individuals of European ancestry from 45 studies, while the replication sample consisted of 7,898 individuals of non-European descent. Lipid levels were measured after eight or more hours of fasting with individuals taking lipid lowering medication excluded from the study. Study-specific GWAS adjusted for age, age-squared, sex, and genetic principal components.

The Add Health PGSs for total cholesterol (TC) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	3191.78	3276.87	3238.35	3238.20	0.15
African ancestry	1976	3224.03	3279.23	3255.09	3254.85	0.18
Hispanic ancestry	988	3209.34	3279.23	3246.01	3245.67	0.32
Asian ancestry	437	3233.25	3283.98	3261.63	3261.40	0.39

Total Cholesterol (TC)
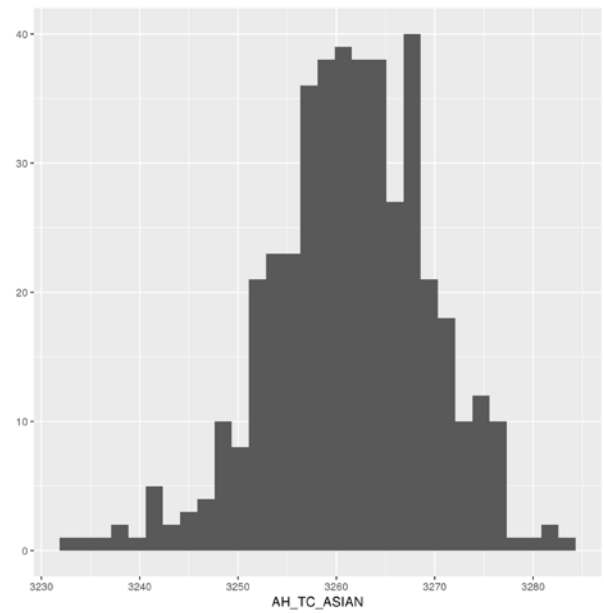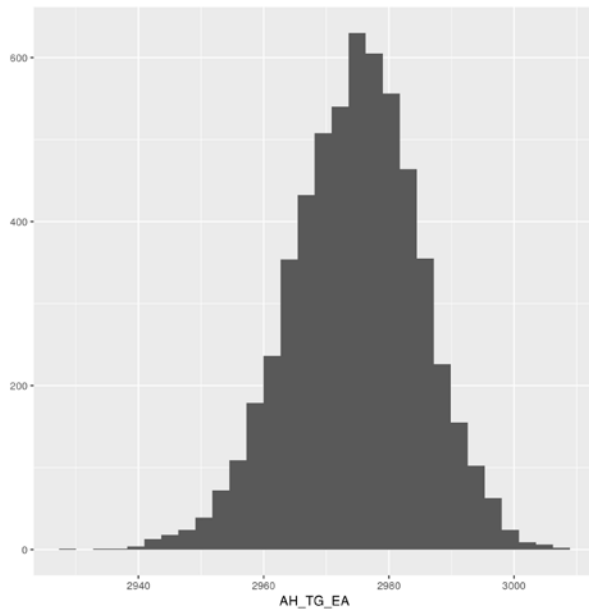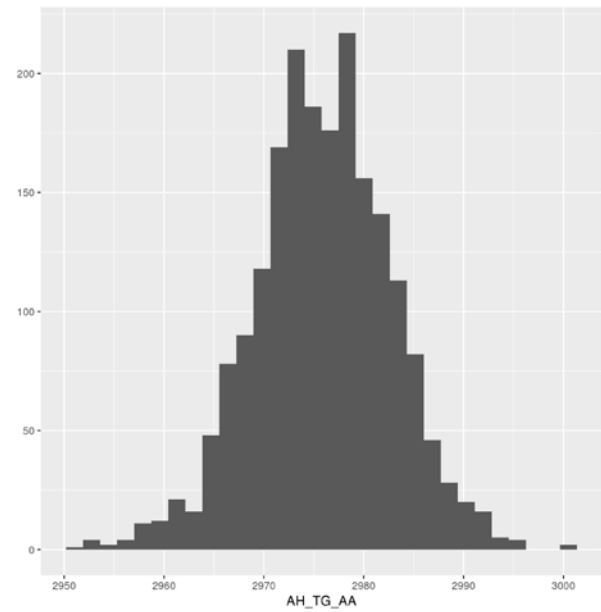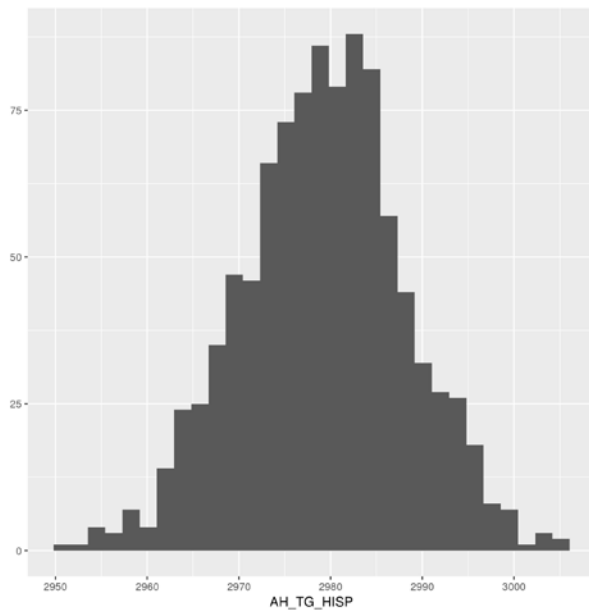
European ancestry raw PGS distribution

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Triglycerides

GWAS Summary Statistic Source: Willer, Cristen J. et al. 2013. "Discovery and Refinement of Loci Associated with Lipid Levels." *Nature Genetics* 45(11):1274–83.

GWAS Ancestry Group(s): European

The PGSs for triglycerides were created using results from a 2013 GWAS conducted by the Global Lipids Genetics Consortium (GLGC). The GWAS meta-analysis files are publically available and can be downloaded from: http://csg.sph.umich.edu/willer/public/lipids2013/.

The discovery phase of the meta-analysis included 188,577 individuals of European ancestry from 45 studies, while the replication sample consisted of 7,898 individuals of non-European descent. Lipid levels were measured after eight or more hours of fasting with individuals taking lipid lowering medication excluded from the study. Study-specific GWAS adjusted for age, age-squared, sex, and genetic principal components.

The Add Health PGSs for triglycerides (TG) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	2928.14	3007.00	2975.01	2974.60	0.14
African ancestry	1976	2951.12	3000.56	2975.91	2975.91	0.15
Hispanic ancestry	988	2951.06	3005.46	2979.39	2979.17	0.28
Asian ancestry	437	2959.85	3013.48	2988.56	2988.87	0.36

Triglycerides (TG)

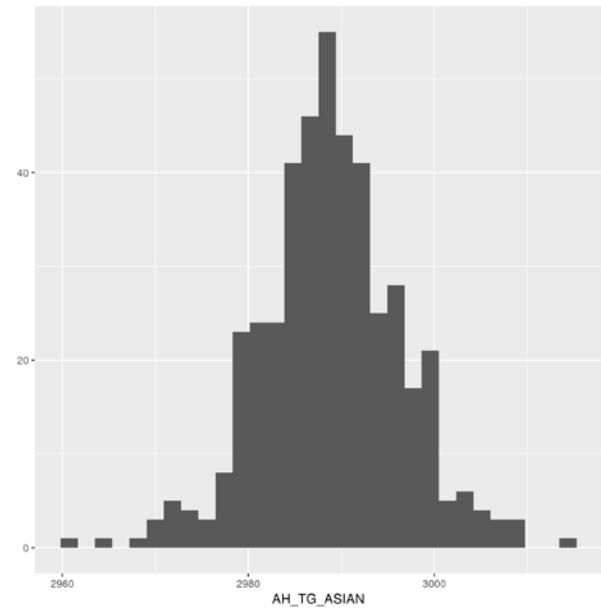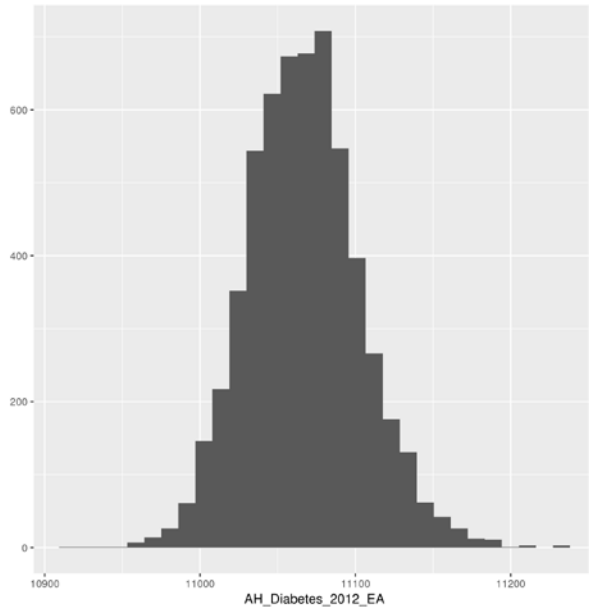## European ancestry raw PGS distribution

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Type II Diabetes (2012)

GWAS Summary Statistic Source: Morris, Andrew P. et al. 2012. "Large-Scale Association Analysis Provides Insights into the Genetic Architecture and Pathophysiology of Type 2 Diabetes." Nature Genetics 44(9):981–90.

GWAS Ancestry Group(s): European

The PGSs for Type II Diabetes 2012 were created using GWAS meta-analysis results from a 2012 study conducted by the DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium. The GWAS meta-analysis files are available from: http://www.diagram-consortium.org/downloads.html.

The discovery meta-analysis consists of 12,171 T2D cases and 56,862 controls across 12 studies of populations of European descent. The replication meta-analysis consisted of 22,669 cases and 58,119 controls from studies of European populations and 1,178 cases and 2,472 controls from a study of Pakistani populations (Pakistan Risk of Myocardial Infarction Study). The authors note that they found little evidence of heterogeneity in SNP effects between European and Pakistani ancestry groups (p.981). Study-specific GWAS adjusted for age of onset (cases) or age of recruitment (controls), gender, and genetic principal components. Additionally, the results of each GWAS were corrected for residual population structure using the genomic control inflation factor, as reported in Supplementary Table 1 of Morris et al. (2012).

The Add Health PGSs for Type II Diabetes 2012 (T2D_2012) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
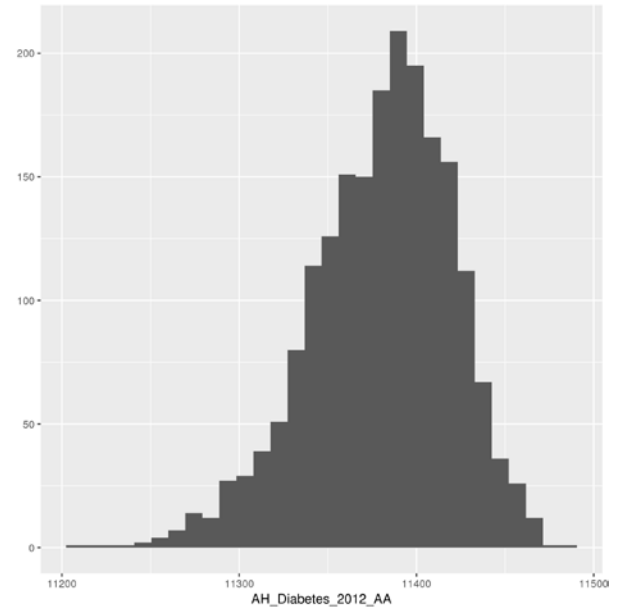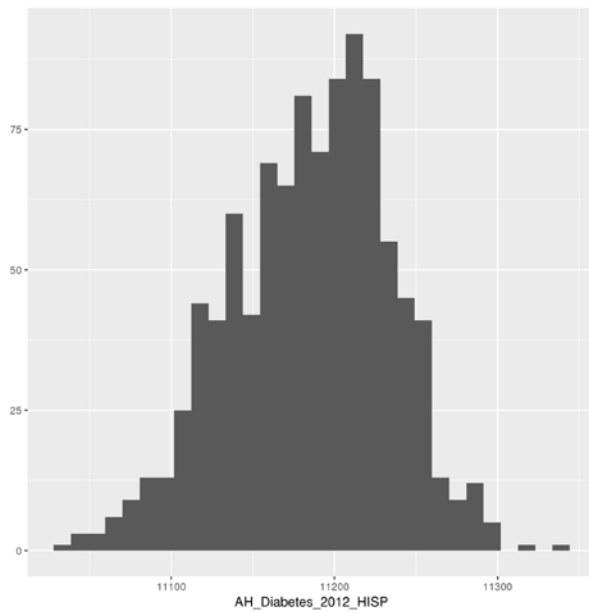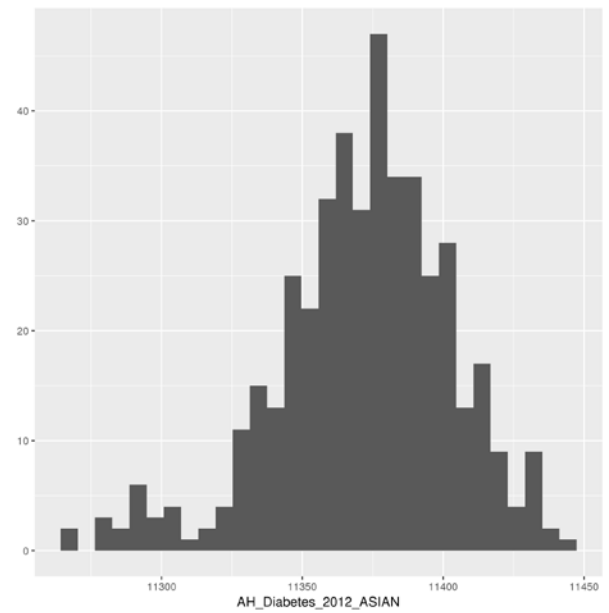
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	10918.20	11236.00	11065.70	11066.84	0.48
African ancestry	1976	11204.60	11483.10	11384.45	11380.10	0.90
Hispanic ancestry	988	11031.90	11337.50	11189.15	11185.14	1.56
Asian ancestry	437	11266.40	11443.50	11375.10	11371.81	1.48

Type II Diabetes 2012

European ancestry raw PGS distribution

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Type II Diabetes (2014)

GWAS Summary Statistic Source: Mahajan, Anubha et al. 2014. "Genome-Wide Trans-Ancestry Meta-Analysis Provides Insight into the Genetic Architecture of Type 2 Diabetes Susceptibility." Nature Genetics 46(3):234–44.

GWAS Ancestry Group(s): European (Ncase = 12,171, Ncontrol = 56,862), East Asian (Ncase = 6,952, Ncontrol = 11,865), South Asian (Ncase = 5,561, Ncontrol = 14,458), Mexican/Mexican American (Ncase = 1,804, Ncontrol = 779).

The PGSs for Type II Diabetes 2014 were created using GWAS meta-analysis results from a 2014 study conducted by the DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium, the Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, the South Asian Type 2 Diabetes (SAT2D) Consortium, the Mexican American Type 2 Diabetes (MAT2D) Consortium, and the Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) Consortium. The GWAS meta-analysis files are publically available from: http://www.diagram-consortium.org/downloads.html.

The discovery meta-analysis consists of 26,488 T2D cases and 83,964 controls in populations of European, East Asian, south Asian, and Mexican and Mexican American descent as described above. The replication meta-analysis consisted of 21,491 cases and 55,647 controls in populations of European descent. Study-specific GWAS adjustments varied, but largely included: age, sex, and cohort.

The Add Health PGSs for Type II Diabetes 2014 (T2D_2014) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
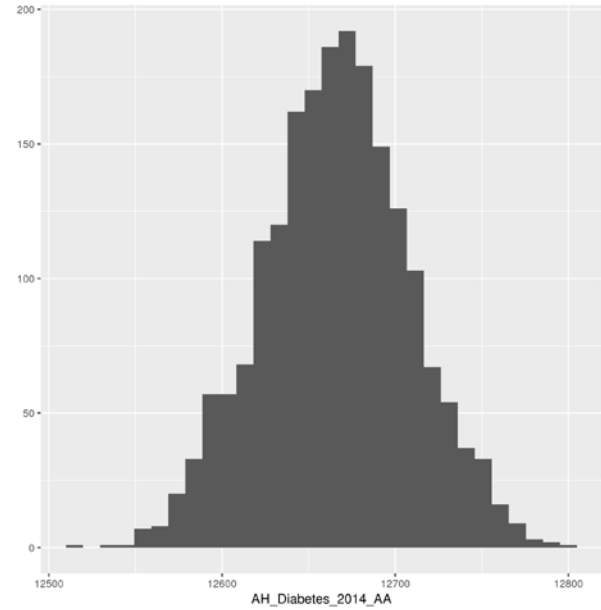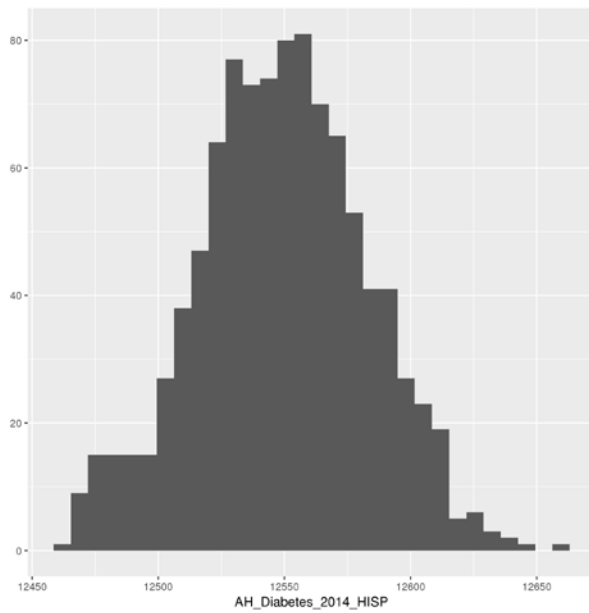
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	12385.00	12610.40	12488.40	12488.85	0.35
African ancestry	1976	12514.40	12799.50	12666.45	12665.83	0.95
Hispanic ancestry	988	12460.10	12657.80	12548.70	12548.90	1.07
Asian ancestry	437	12543.30	12676.70	12617.90	12617.14	1.26

Type II Diabetes 2014

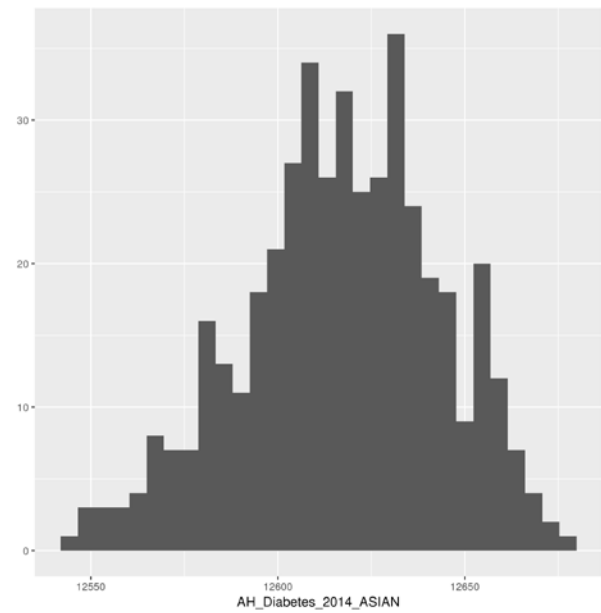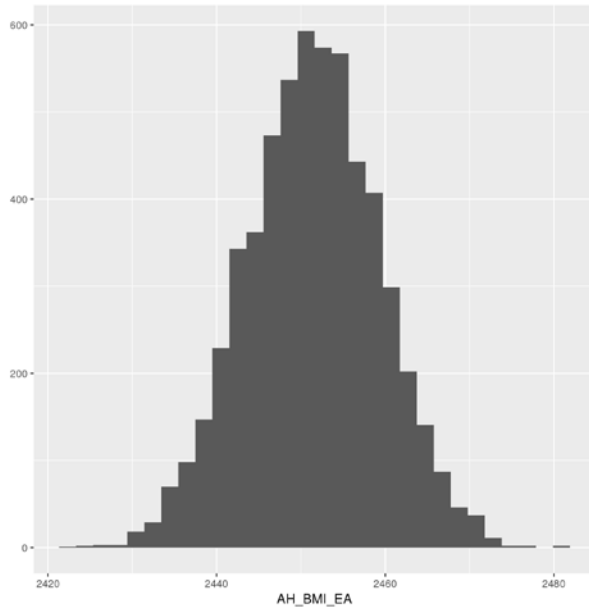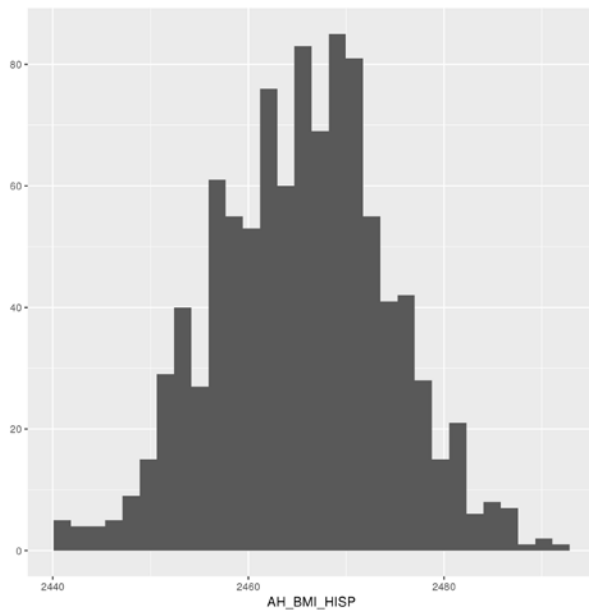### European ancestry raw PGS distribution
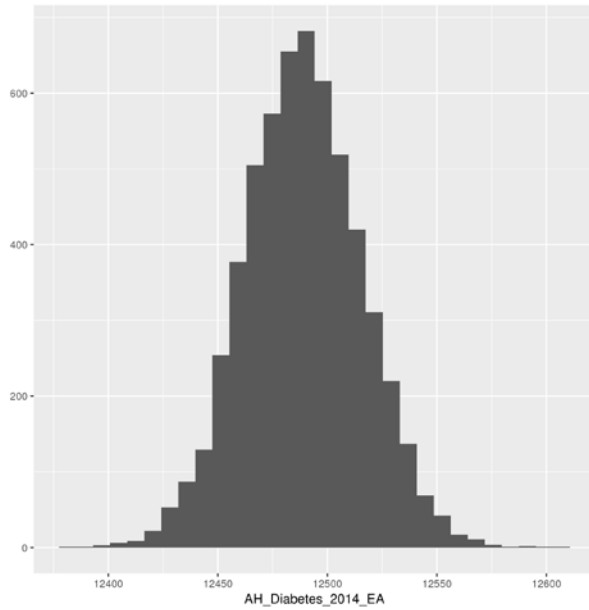
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Body Mass Index

GWAS Summary Statistic Source: Locke, Adam E. et al. 2015. "Genetic Studies of Body Mass Index Yield New Insights for Obesity Biology." *Nature* 518(7538):197–206.

GWAS Ancestry Group(s): European

PGSs for body mass index BMI were created using results from a 2015 GWAS conducted by the Genetic Investigation of ANthropometric Traits (GIANT) consortium. The GWAS meta-analysis files are publicly available from: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

Joint GWAS and Metabochip meta-analysis was performed on a sample of 332,206 (Ngwas = 234,069; Nmeta = 88,137) individuals of European descent from 114 studies. Additional covariates controlled for during the meta-analysis include: age, the second order polynomial of age, sex, and genetic principal components. The measures of BMI include clinical measurements as well as self-reported BMI.

The Add Health BMI are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	2422.74	2481.28	2451.45	2451.41	0.10
African ancestry	1976	2464.99	2513.17	2493.17	2492.84	0.16
Hispanic ancestry	988	2440.33	2491.33	2465.80	2465.47	0.28
Asian ancestry	437	2463.94	2505.61	2482.60	2482.65	0.31

BMI

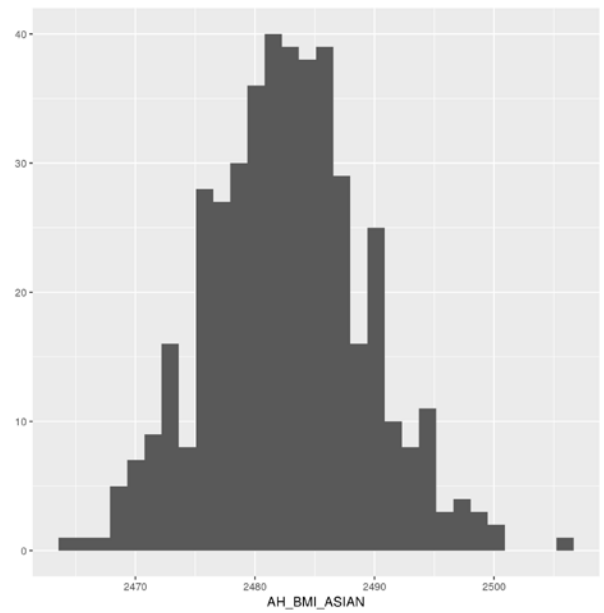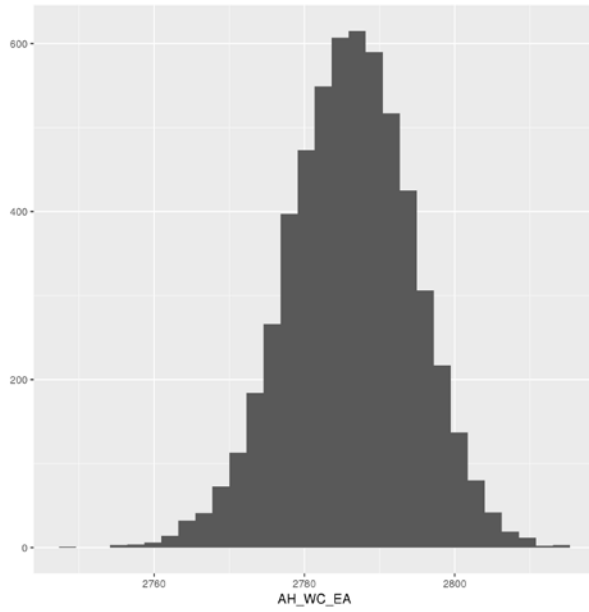### European ancestry raw PGS distribution

African ancestry raw PGS distribution
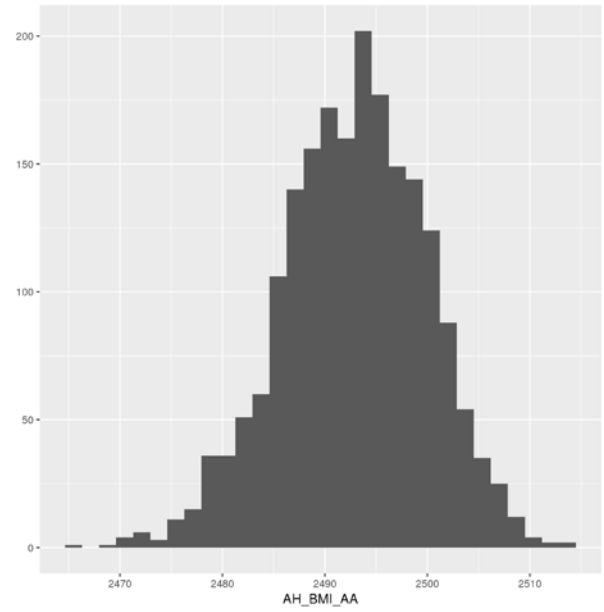
Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Waist Circumference

GWAS Summary Statistic Source: Shungin, Dmitry et al. 2015. "New Genetic Loci Link Adipose and Insulin Biology to Body Fat Distribution." *Nature* 518(7538):187–96.

GWAS Ancestry Group(s): European

PGSs for waist circumference were created using results from a 2015 study conducted by the Genetic Investigation of ANthropometric Traits (GIANT) consortium. The GWAS meta-analysis files are publicly available from: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

A GWAS meta-analysis was performed on a sample of 142,762 individuals of European descent from 57 studies, and separately in a Metabochip (MC) meta-analysis on a sample of 67,326 individuals of European descent from 44 studies. A joint GWAS and MC meta-analysis was then conducted on 210,088 individuals of European descent. Additional covariates in both GWAS and MC analyses included: age, the second order polynomial of age, study-specific covariates as necessary, and BMI.

The Add Health PGSs for waist circumference (WC)are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
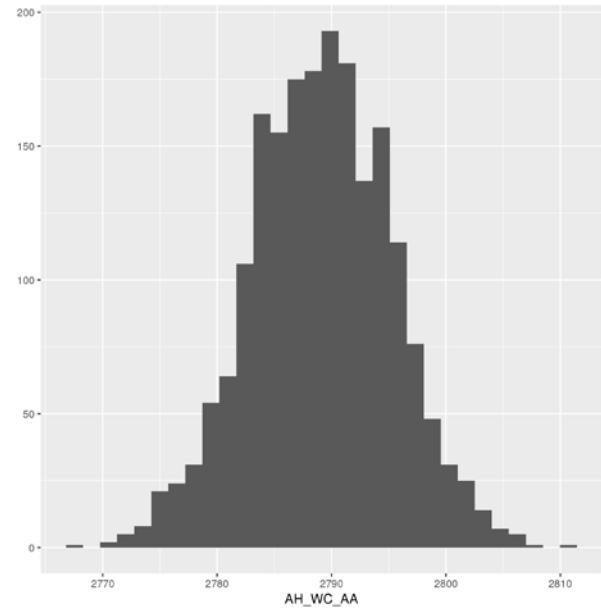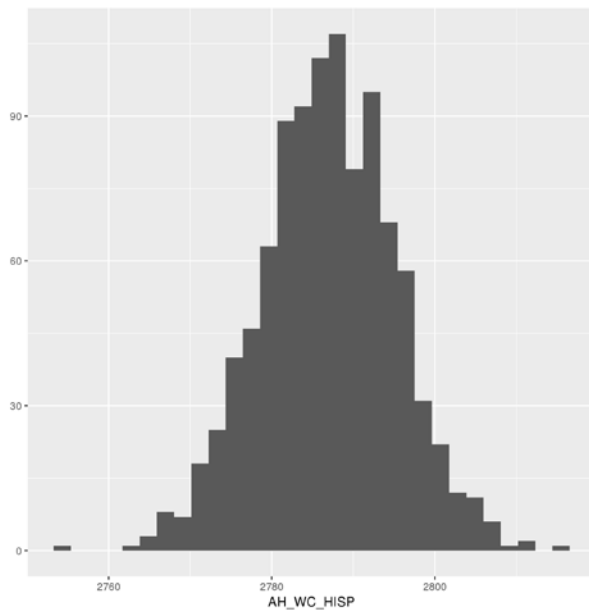
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	2749.61	2815.32	2786.25	2786.09	0.11
African ancestry	1976	2767.46	2810.61	2789.13	2789.07	0.14
Hispanic ancestry	988	2755.17	2816.33	2786.98	2786.93	0.26
Asian ancestry	437	2763.07	2806.59	2784.93	2784.75	0.34

Waist Circumference

European ancestry raw PGS distribution
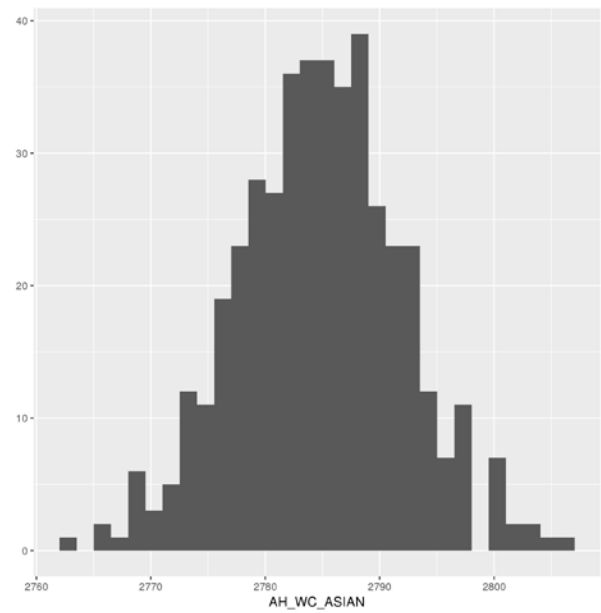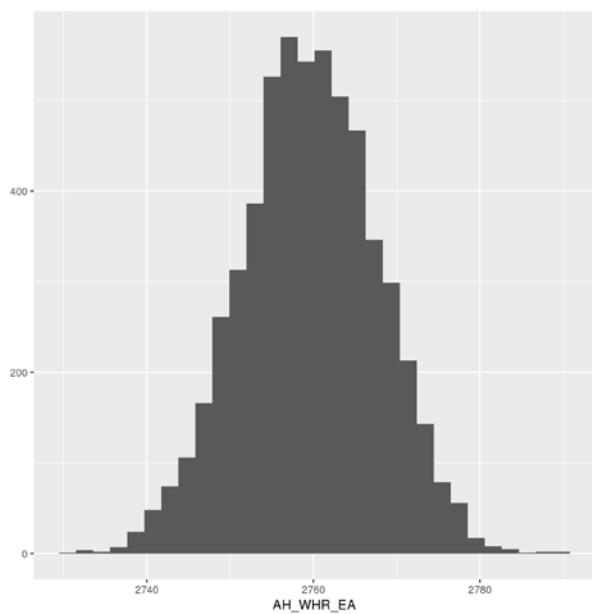

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Waist-to-Hip Ratio

GWAS Summary Statistic Source: Shungin, Dmitry et al. 2015. "New Genetic Loci Link Adipose and Insulin Biology to Body Fat Distribution." *Nature* 518(7538):187–96.

GWAS Ancestry Group(s): European

PGSs for waist-to-hip ratio were created using results from a 2015 study conducted by the Genetic Investigation of ANthropometric Traits (GIANT) consortium. The GWAS meta-analysis files are publicly available from: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

GWAS meta-analysis was performed on a sample of 142,762 individuals of European descent from 57 studies, and separately in a Metabochip (MC) meta-analysis on a sample of 67,326 individuals of European descent from 44 studies. A joint GWAS and MC meta-analysis was then conducted on 210,088 individuals of European descent. Additional covariates in both GWAS and MC analyses included: age, the second order polynomial of age, study-specific covariates as necessary, and BMI.

The Add Health PGSs for waist-to-hip ratio (WHR) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
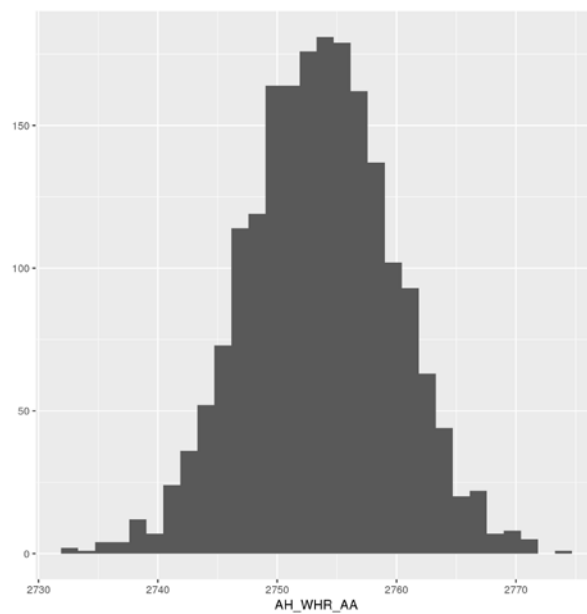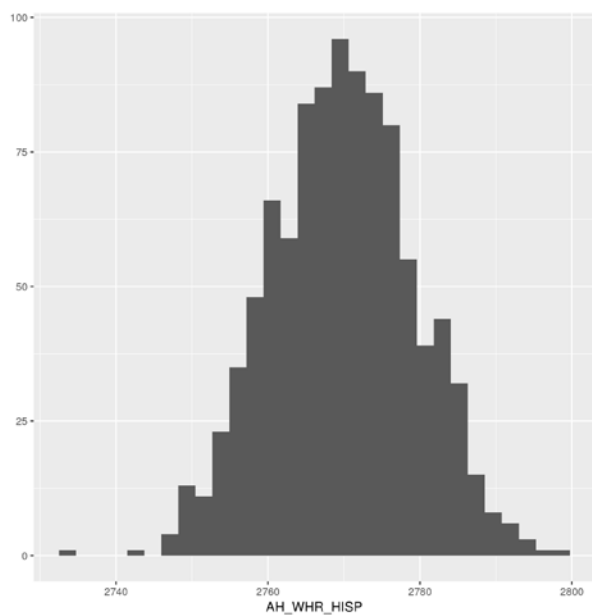
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	2731.39	2790.66	2759.54	2759.54	0.11
African ancestry	1976	2732.57	2773.93	2753.56	2753.56	0.14
Hispanic ancestry	988	2734.54	2799.52	2770.01	2769.84	0.30
Asian ancestry	437	2773.34	2816.04	2796.13	2796.30	0.32

Waist-to-Hip Ratio

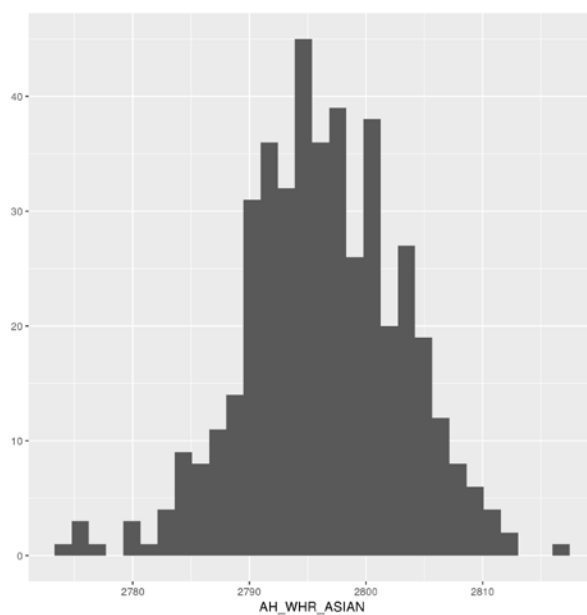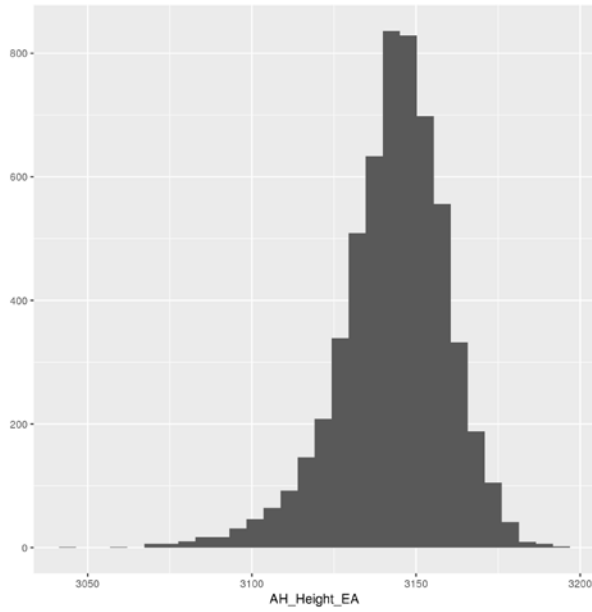### European ancestry raw PGS distribution

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Height

GWAS Summary Statistic Source: Wood, Andrew R. et al. 2014. "Defining the Role of Common Variation in the Genomic and Biological Architecture of Adult Human Height." *Nature Genetics* 46(11):1173–86.

GWAS Ancestry Group(s): European

PGSs for height were created using results from a 2014 study conducted by the Genetic Investigation of ANthropometric Traits (GIANT) consortium. The GWAS meta-analysis files are publicly available from: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

The meta-analysis included 253,288 individuals of European descent from 79 studies imputed to HapMap II. Replication was performed in a sample of 80,067 individuals of European descent. Height was measured as sex standardized height. Additional covariates in GWAS analyses included age and genetic principal components.

The Add Health PGSs for height are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
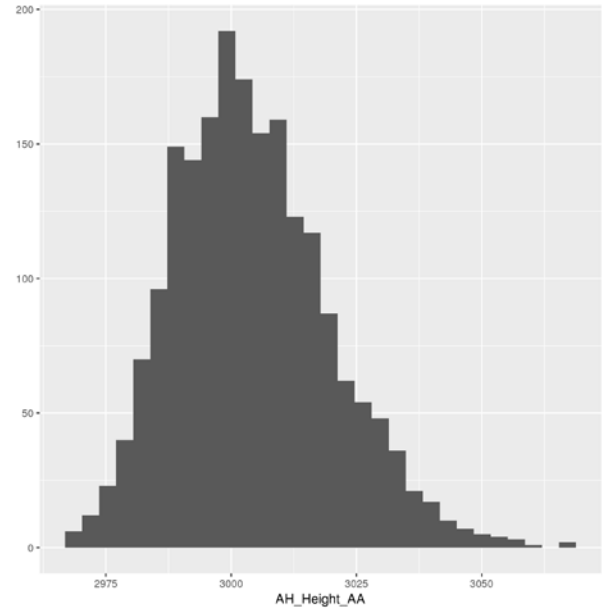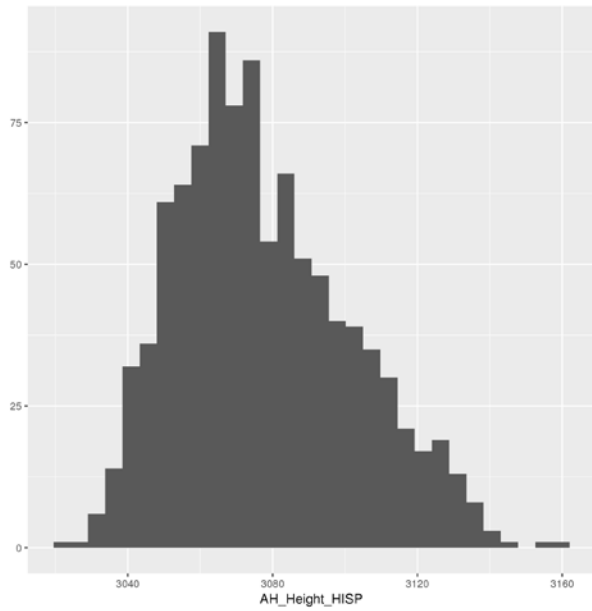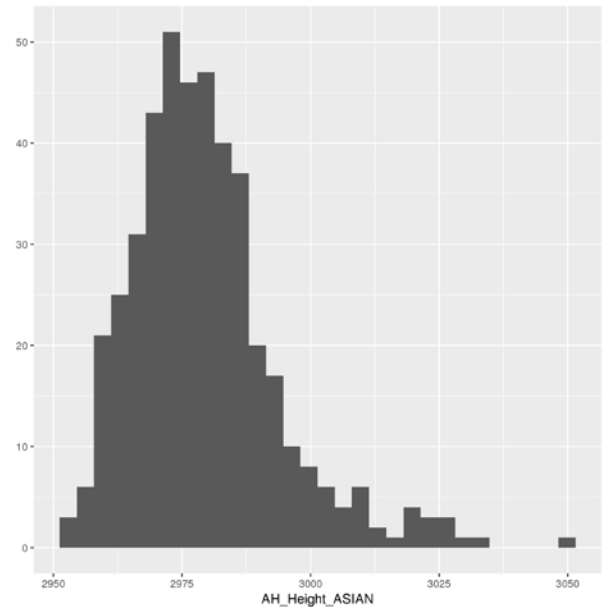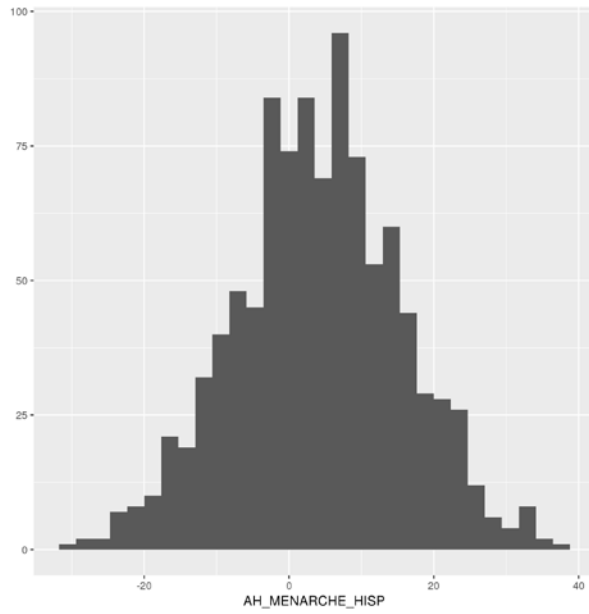
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	3045.60	3195.98	3144.46	3143.05	0.22
African ancestry	1976	2967.68	3066.18	3002.63	3004.19	0.35
Hispanic ancestry	988	3021.99	3159.78	3073.92	3077.56	0.77
Asian ancestry	437	2952.41	3049.39	2977.41	2979.34	0.69

Height

European ancestry raw PGS distribution

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Menarche

GWAS Summary Statistic Source: Perry, John R. B. et al. 2014. "Parent-of-Origin-Specific Allelic Associations among 106 Genomic Loci for Age at Menarche." Nature 514(7520):92–97.

GWAS Ancestry Group(s): European

PGSs for age at menarche were created using results from a 2014 study conducted by the Reproductive Genetics (ReproGen) consortium. The GWAS meta-analysis files are publicly available from: http://www.reprogen.org/data_download.html.

The discovery meta-analysis included 132,989 women of European descent from 58 studies. Women who reported their age at menarche as less than 9 years or greater than 17 years old were excluded from the analysis. Analyses adjusted for birth year in order to account for secular trends in menarche timing, study-specific genomic control inflation factors, and relatedness between individuals when necessary.

The Add Health PGSs for age at menarche are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
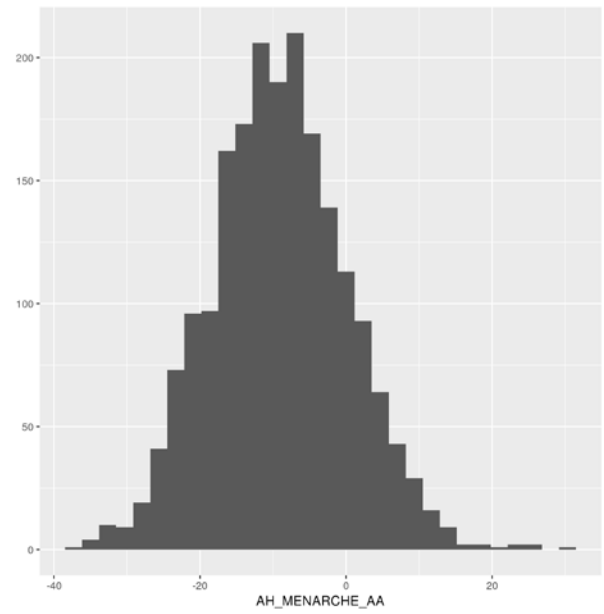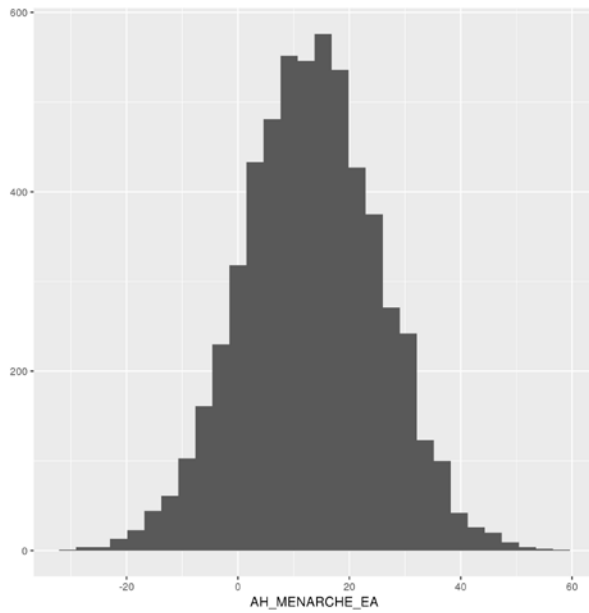
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	-31.94	56.72	13.12	13.12	0.16
African ancestry	1976	-37.87	29.73	-9.41	-9.16	0.21
Hispanic ancestry	988	-30.55	37.63	4.25	4.03	0.37
Asian ancestry	437	-30.68	33.62	-1.18	-0.78	0.51

Menarche

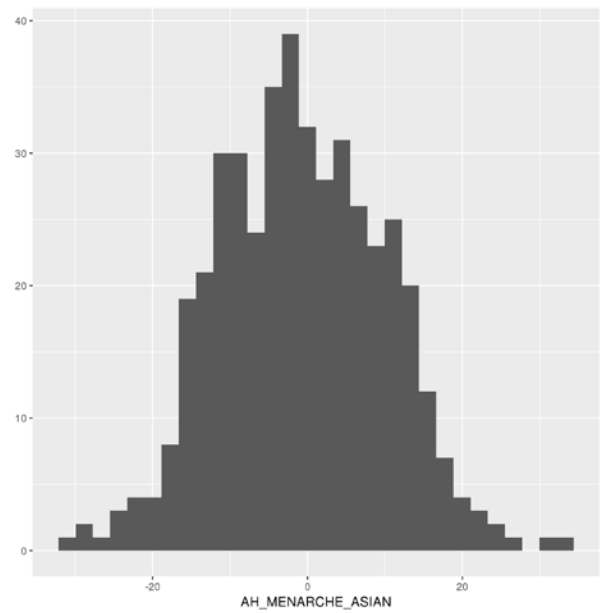### European ancestry raw PGS distribution

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Menopause

GWAS Summary Statistic Source: Day, Felix R. et al. 2015. "Large-Scale Genomic Analyses Link Reproductive Aging to Hypothalamic Signaling, Breast Cancer Susceptibility and BRCA1-Mediated DNA Repair." Nature Genetics 47(11):1294–1303.

GWAS Ancestry Group(s): European

PGSs for age at menarche were created using results from a 2014 study conducted by the Reproductive Genetics (ReproGen) consortium. The GWAS meta-analysis files are publicly available from: http://www.reprogen.org/data_download.html.

The ReproGen meta-analysis included 69,360 women of European descent from 50 studies imputed to HapMap Phase 2. GWAS analyses included study specific controls and the top principal components of the genetic data. In all studies, age at naturally occurring menopause is self-reported. Age at menopause is defined as the age at last occurring menstrual period followed by 12 consecutive months of amenorrhea.

The Add Health PGSs for age at menopause are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
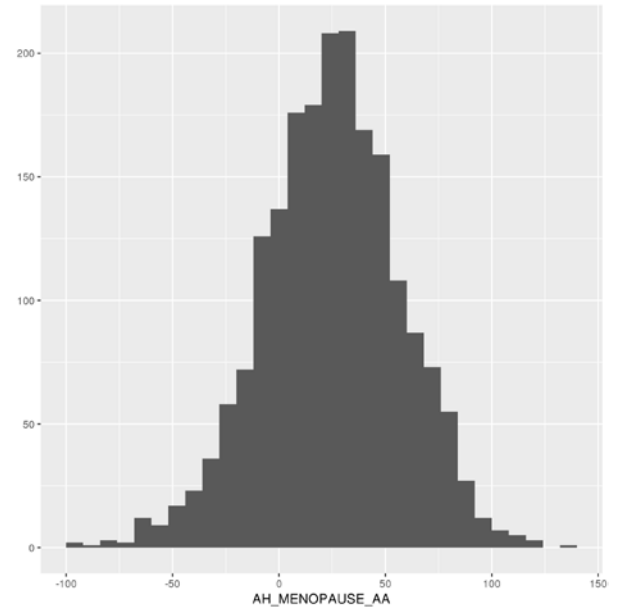
PGS Summary Statistics by Ancestry

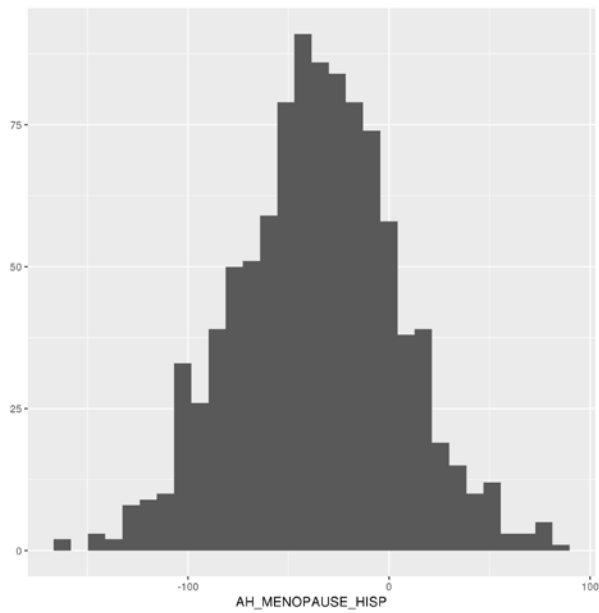| Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	-242.18	150.34	-71.96	-72.31	0.64
African ancestry	1976	-95.72	136.28	25.60	24.59	0.73
Hispanic ancestry	988	-166.67	81.51	-34.61	-35.82	1.28
Asian ancestry	437	-74.20	124.05	15.33	14.22	1.62

Menopause

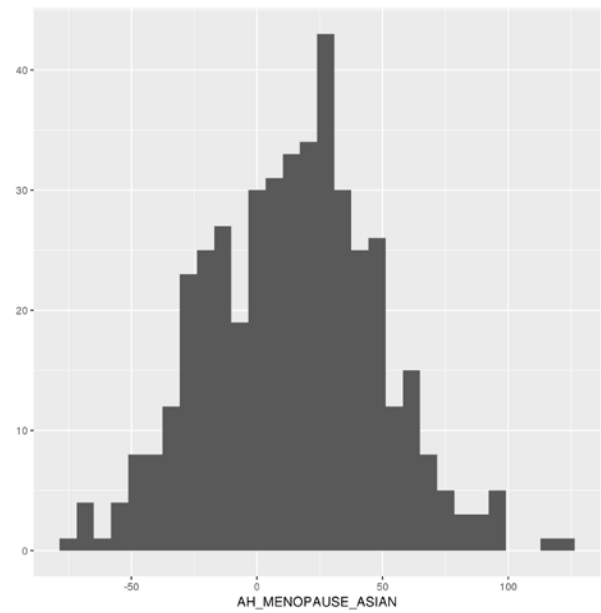### European ancestry raw PGS distribution
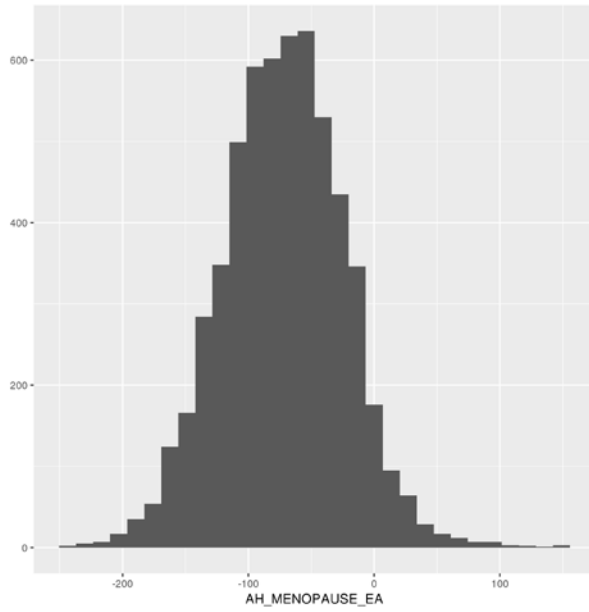
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Number of Children Ever Born

GWAS Summary Statistic Source: Barban, Nicola et al. 2016. "Genome-Wide Analysis Identifies 12 Loci Influencing Human Reproductive Behavior." Nature Genetics 48(12):1462–72.

GWAS Ancestry Group(s): European

PGSs for the number of children ever born were created using pooled results for men and women from a 2016 study conducted as part of the Sociogenome project (http://www.sociogenome.com/). The GWAS meta-analysis files are publicly available from: ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/BarbanN_27798627_GCST003796.

The meta-analysis included 343,072 individuals of European descent above age 44 for women and 54 for men from 62 studies. Additional covariates include age, the second order polynomial of age, the third order polynomial of age, and genetic principal components.

The Add Health PGSs for number of children ever born (NEB) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
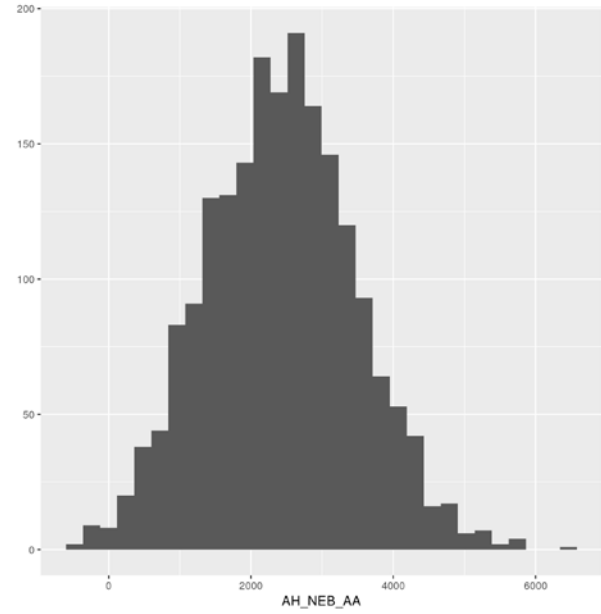
PGS Summary Statistics by Ancestry

| Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	1689.86	11668.10	6632.36	6601.49	19.16
African ancestry	1976	-464.23	6476.07	2430.67	2424.28	23.50
Hispanic ancestry	988	1553.43	10455.40	5708.17	5716.19	42.78
Asian ancestry	437	1358.46	7940.91	4808.55	4828.68	51.64

Number of Children Ever Born

European ancestry raw PGS distribution
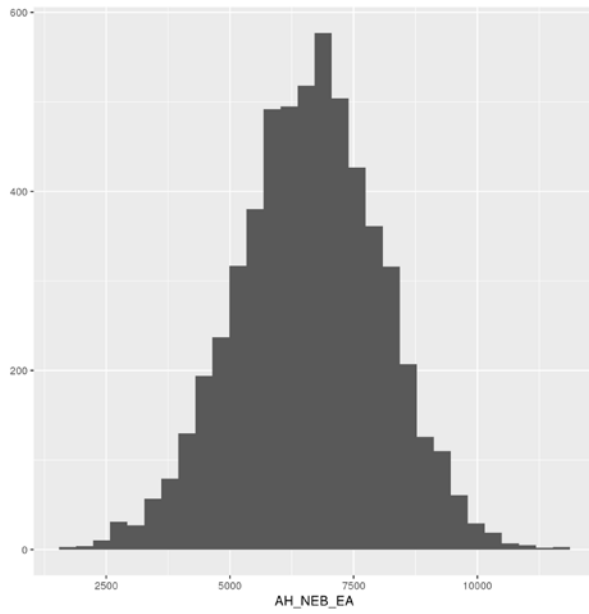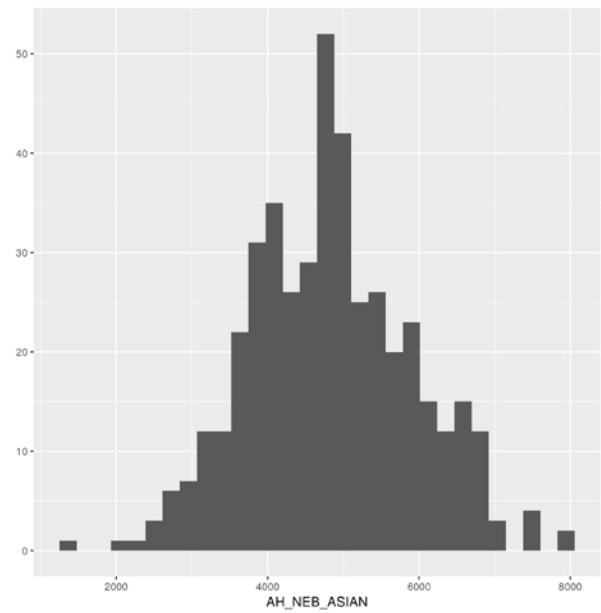

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution
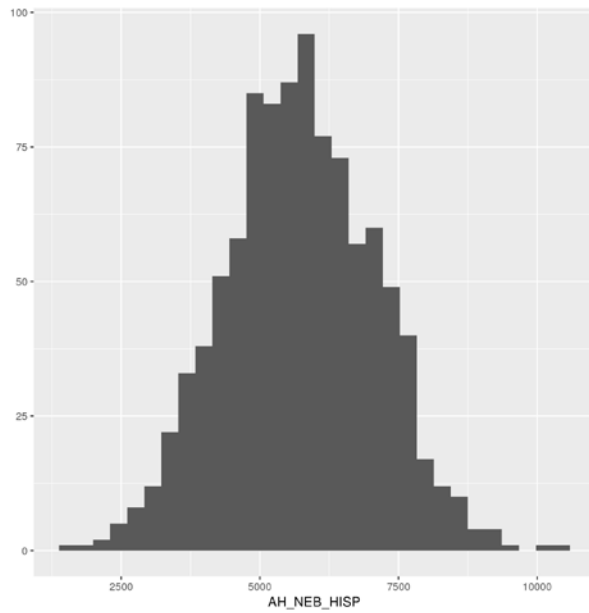
Asian ancestry raw PGS distribution

Age at First Birth

GWAS Summary Statistic Source: Barban, Nicola et al. 2016. "Genome-Wide Analysis Identifies 12 Loci Influencing Human Reproductive Behavior." Nature Genetics 48(12):1462–72.

GWAS Ancestry Group(s): European

PGSs for the age at first birth were created using pooled results for men and women from a 2016 study conducted as part of the Sociogenome project (http://www.sociogenome.com/). The GWAS meta-analysis files are publicly available from: ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/BarbanN_27798627_GCST003795.

The meta-analysis included 251,151 individuals of European descent who had a biological child at the time of interview from 62 studies. Additional covariates include age, the second order polynomial of age, the third order polynomial of age, and genetic principal components.

The Add Health PGSs for age at first birth (AFB) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
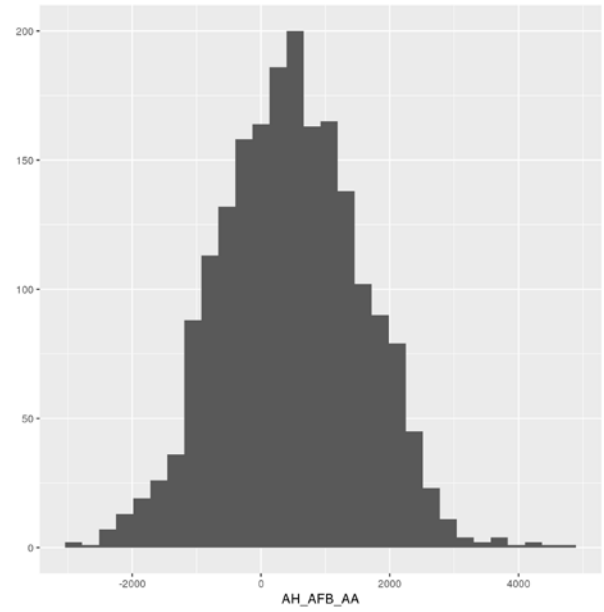
PGS Summary Statistics by Ancestry

| Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	-4844.29	7307.43	1385.73	1420.20	20.49
African ancestry	1976	-2795.67	4871.18	443.44	465.42	24.50
Hispanic ancestry	988	-3552.22	5604.14	808.56	823.98	48.18
Asian ancestry	437	-4520.84	3004.46	-426.29	-520.61	60.07

Age at First Birth

European ancestry raw PGS distribution
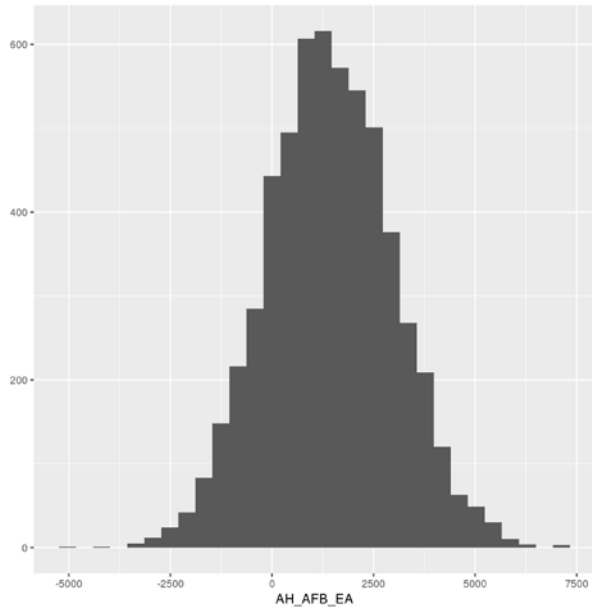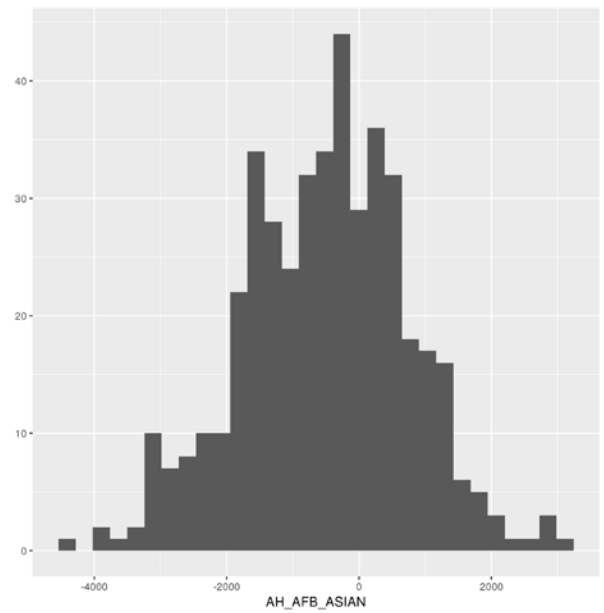

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution
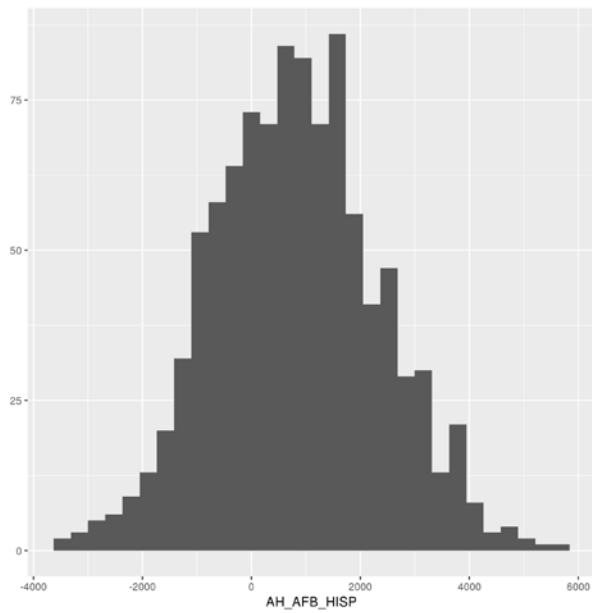
Asian ancestry raw PGS distribution

Ever/Current Smoker

GWAS Summary Statistic Source: Furberg, Helena et al. 2010. "Genome-Wide Meta-Analyses Identify Multiple Loci Associated with Smoking Behavior." *Nature Genetics* 42(5):441–47.

GWAS Ancestry Group(s): European

The PGSs for current/ever smoker were created using results from a 2010 GWAS conducted by the Tobacco and Genetics (TAG) Consortium. The GWAS meta-analysis files are publicly available from: https://www.med.unc.edu/pgc/results-and-downloads.

The discovery phase of analysis included 74,053 individuals of European descent. The replication of the 15 most significant genetic regions included 143,023 individuals. Individuals who were recorded as having ever been regular smokers were defined as those who reported having smoked at least 100 cigarettes during their lifetime, and never regular smokers were defined as those who reported having smoked between 0 and 99 cigarettes during their lifetime. GWAS analyses controlled for imputed allele dosage for a SNP and if the individual was classified as a case in the primary study. If the primary study was case-control in design and the phenotype being studied was known to be associated with smoking, the GWAS adjusted for case status to reduce potential confounding. Analyses were run and meta-analyzed separately for males and females.

The Add Health current/ever smoker (Evrsmk) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification, but to give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
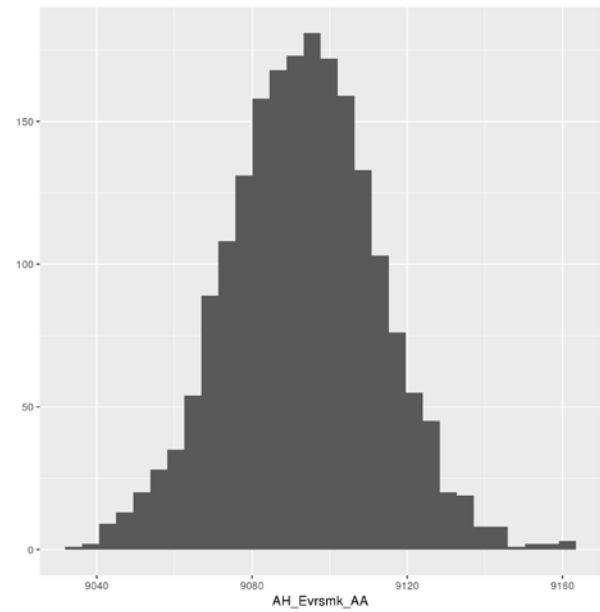
PGS Summary Statistics by Ancestry

| Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	8987.39	9165.44	9088.74	9089.02	0.29
African ancestry	1976	9034.84	9162.08	9093.27	9093.16	0.43
Hispanic ancestry	988	9023.06	9176.04	9104.48	9104.68	0.70
Asian ancestry	437	9063.89	9188.32	9123.72	9122.90	0.92

Ever/Current Smoker

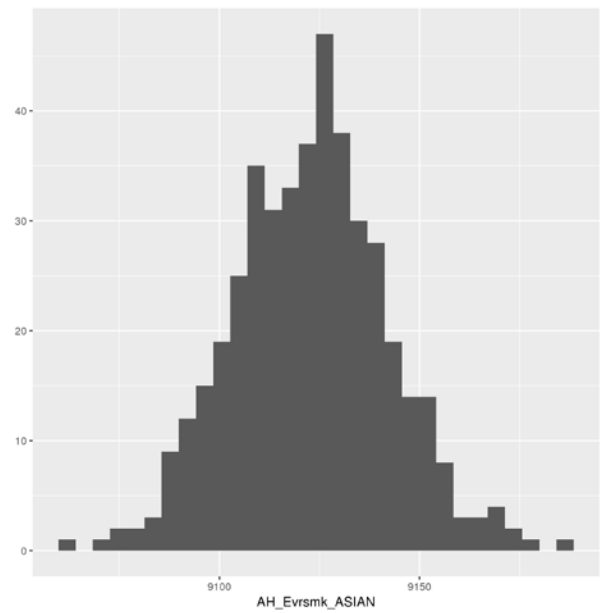### European ancestry raw PGS distribution
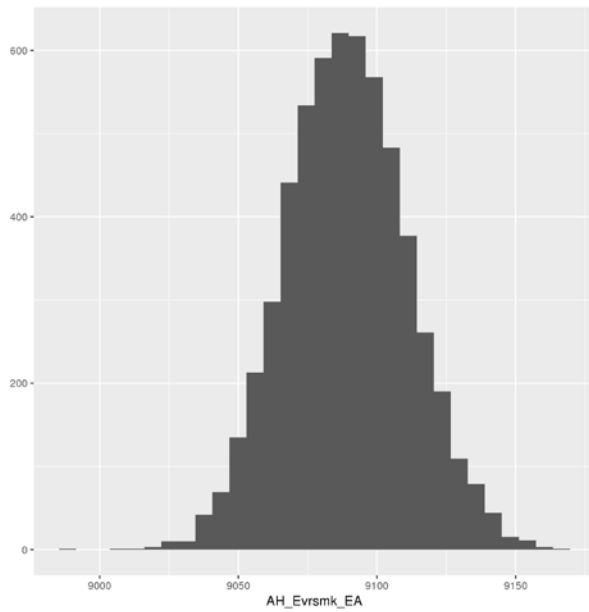
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution
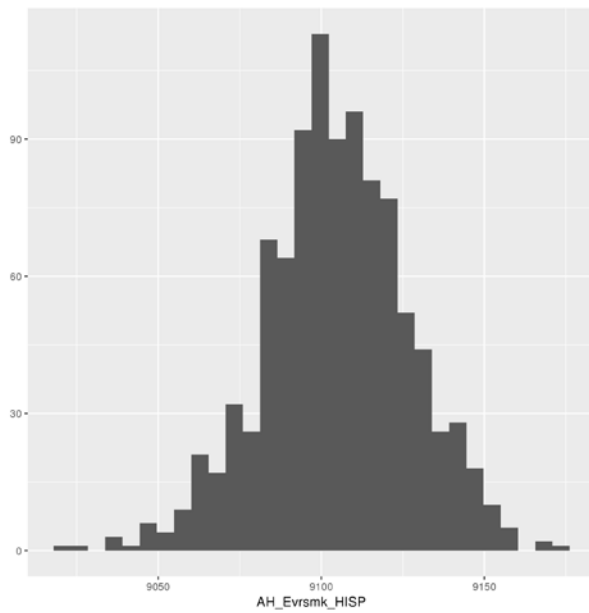
Asian ancestry raw PGS distribution

Number of Cigarettes per day

GWAS Summary Statistic Source: Furberg, Helena et al. 2010. "Genome-Wide Meta-Analyses Identify Multiple Loci Associated with Smoking Behavior." *Nature Genetics* 42(5):441–47.

GWAS Ancestry Group(s): European

The PGSs for the number of cigarettes smoked per day were created using results from a 2010 GWAS conducted by the Tobacco and Genetics (TAG) Consortium. The GWAS meta-analysis files are publicly available from: https://www.med.unc.edu/pgc/results-and-downloads.

The discovery phase of analysis included 74,053 individuals of European descent. The replication of the 15 most significant genetic regions included 143,023 individuals. The number of cigarettes smoked per day was reported as either the average number of cigarettes smoked per day or the maximum number of cigarettes smoked in a single day (see the online methods). GWAS analyses controlled for imputed allele dosage for a SNP and if the individual was classified as a case in the primary study. If the primary study was case-control in design and the phenotype being studied was known to be associated with smoking, the GWAS adjusted for case status to reduce potential confounding. Analyses were run and meta-analyzed separately for males and females.

The Add Health cigarettes smoked per day (CPD)are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification, but to give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
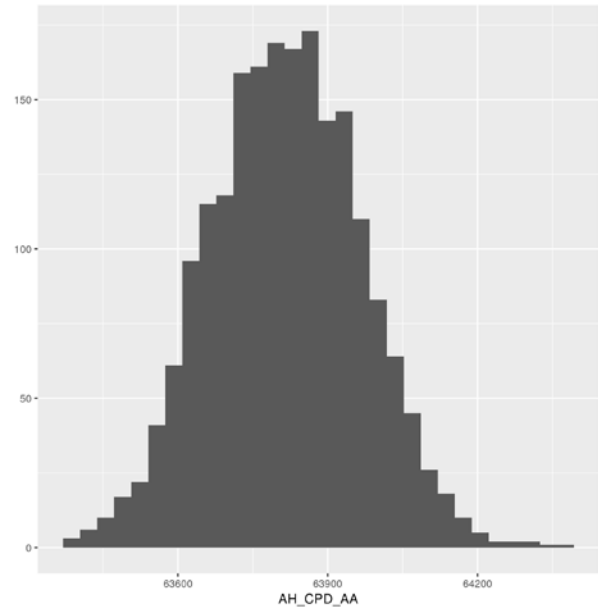
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	63864.60	64915.30	64440.40	64439.87	1.90
African ancestry	1976	63381.10	64369.40	63815.55	63815.89	3.38
Hispanic ancestry	988	63681.90	64694.10	64201.75	64200.37	5.46
Asian ancestry	437	63464.00	64328.40	63865.70	63869.95	6.68

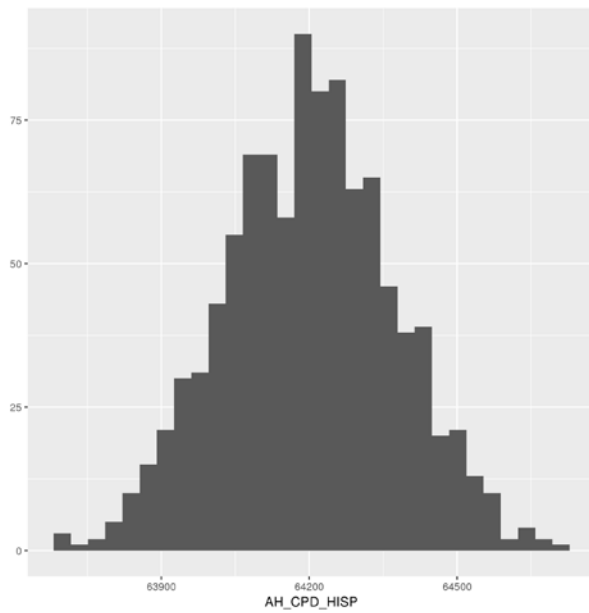# Number of Cigarettes per Day

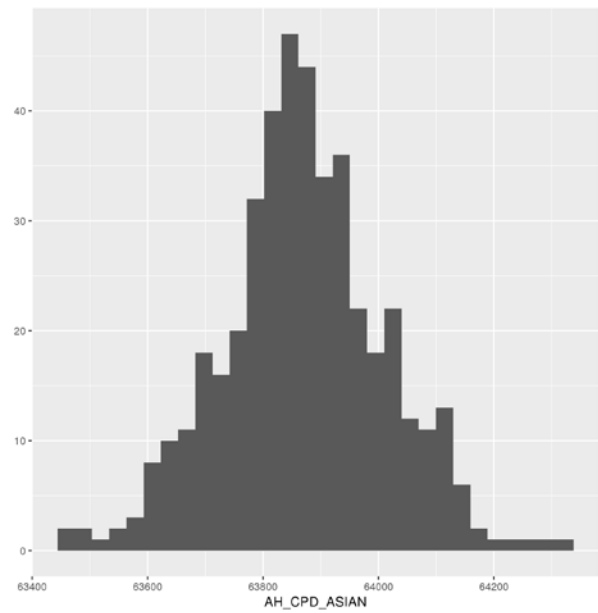### European ancestry raw PGS distribution
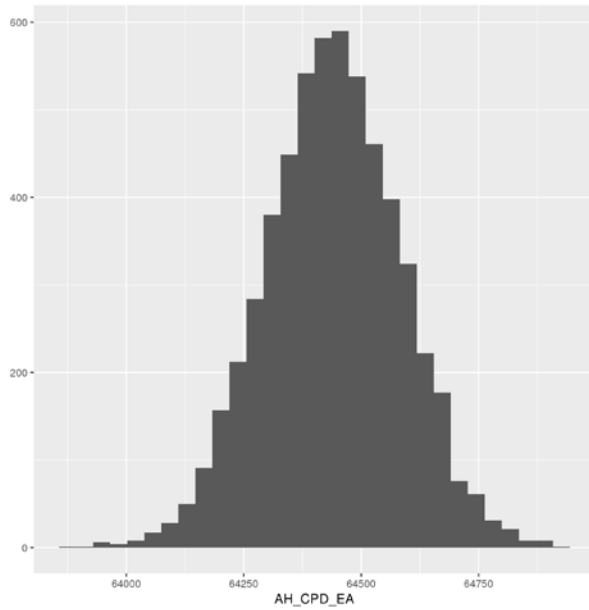
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Extraversion

GWAS Summary Statistic Source: van den Berg, Stéphanie M. et al. 2016. "Meta-Analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium." Behavior Genetics 46(2):170–82.

GWAS Ancestry Group(s): European

The PGSs for extraversion were created using results from a 2016 study by the Genetics of Personality Consortium (GPC). Summary statistics are publicly available from: www.tweelingenregister.org/GPC/.

The discovery meta-analysis included 63,030 individuals of European descent from 29 cohorts. The replication analysis consisted of 9,783 individuals, also of European descent, from a single cohort. All cohorts imputed SNPs based on the 1000 genomes reference panel (1000G). Additional study-specific covariates included sex, age, and ancestry principal components of the genetic data. Extraversion was measured as a continuous latent variable as described in van den Berg et al. (2014).

The Add Health PGSs for Extraversion are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
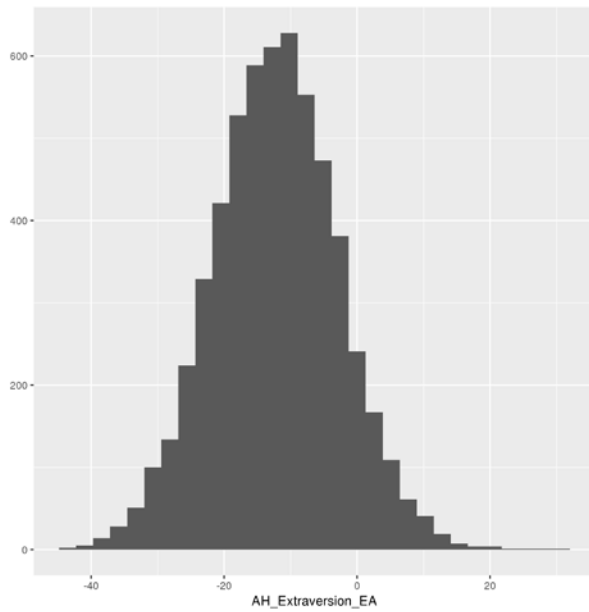
PGS Summary Statistics by Ancestry

| Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	-43.87	30.38	-12.37	-12.27	0.12
African ancestry	1976	-42.10	7.68	-17.30	-17.53	0.16
Hispanic ancestry	988	-54.19	8.11	-23.29	-23.16	0.33
Asian ancestry	437	-59.13	-16.55	-38.94	-39.22	0.39

References:

van den Berg, Stéphanie M. et al. 2014. "Harmonization of Neuroticism and Extraversion Phenotypes across Inventories and Cohorts in the Genetics of Personality Consortium: An Application of Item Response Theory." Behavior Genetics 44(4):295–313.
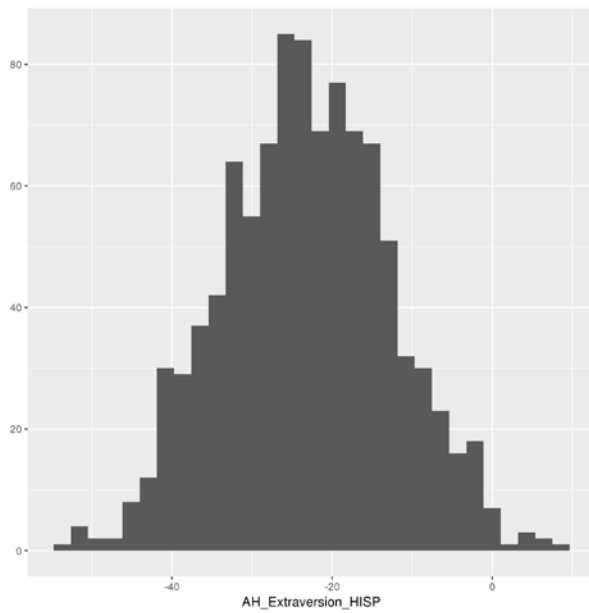
Extraversion
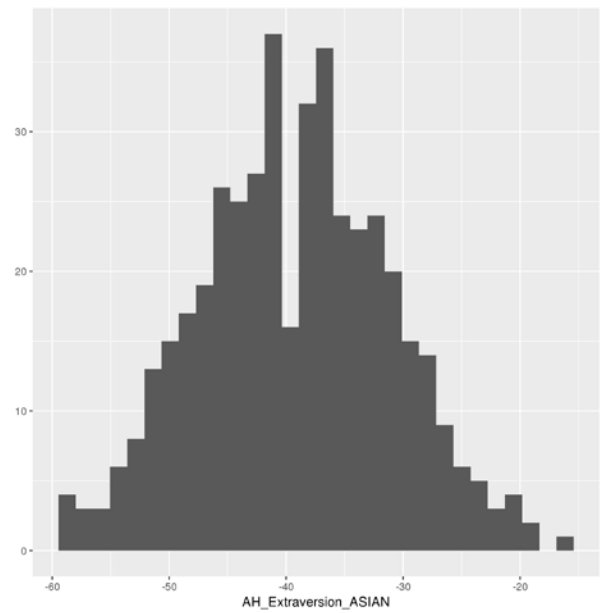
European ancestry raw PGS distribution
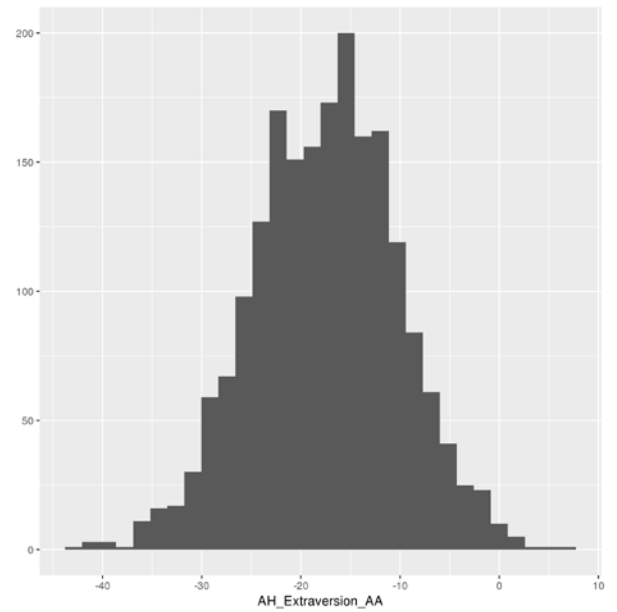
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Attention-Deficit/Hyperactivity Disorder (2010)

GWAS Summary Statistic Source: Neale, Benjamin M. et al. 2010. "Meta-Analysis of Genome-Wide Association Studies of Attention-Deficit/Hyperactivity Disorder." Journal of the American Academy of Child & Adolescent Psychiatry 49(9):884–97.

GWAS Ancestry Group(s): European

The PGSs for attention-deficit/hyperactivity disorder 2010 were created using results from a 2010 GWAS conducted by the ADHD subgroup of the Psychiatric GWAS Consortium. The GWAS meta-analysis files are publicly available from: http://www.med.unc.edu/pgc/results-and-downloads (pgc.adhd.full.2012-10.txt).

The meta-analysis included 2,064 trios, 896 cases, and 2,455 controls from four projects: (1) the Children's Hospital of Philadelphia (CHOP), (2) phase I of the International Multisite ADHD Genetics Project (IMAGE), (3) phase II of IMAGE (IMAGE II), and (4) the Pfizer funded study from the University of California, Los Angeles, Washington University, and the Massachusetts General Hospital (PUWMa). Measures of attention deficit disorder were harmonized across samples and were based on a combination of semi-structured interviews and parent and/or teacher report on questionnaires (see Neale et al., p. 3-5, 2010).

The Add Health PGSs for attention-deficit/hyperactivity disorder 2010 (ADHD_2010) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
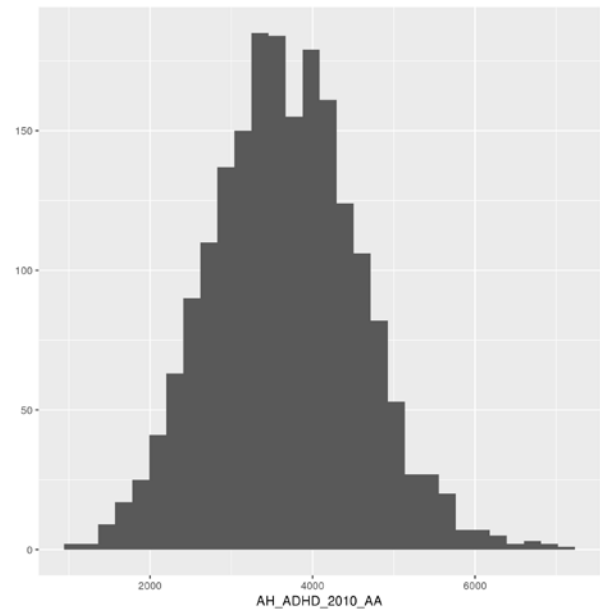
PGS Summary Statistics by Ancestry

| Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	3630.02	12266.40	7932.60	7947.99	14.96
African ancestry	1976	972.29	7048.82	3631.97	3664.49	20.43
Hispanic ancestry	988	3132.62	9977.56	6985.96	7003.93	34.83
Asian ancestry	437	3133.78	8907.47	6022.31	6047.66	41.20

Attention Deficit Disorder 2010

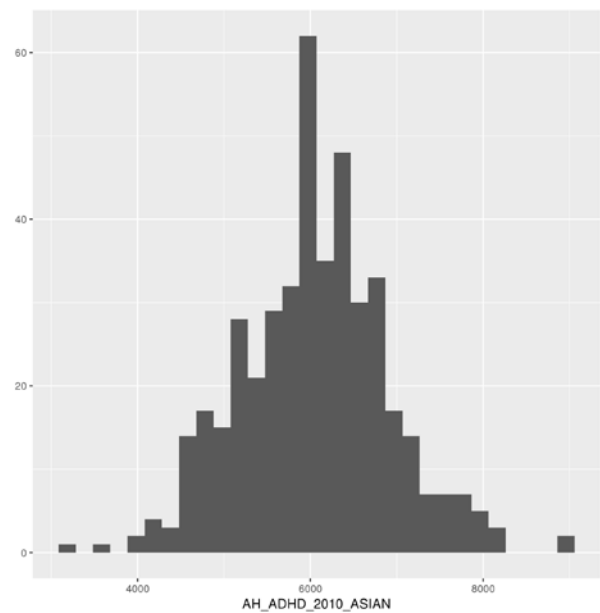## European ancestry raw PGS distribution
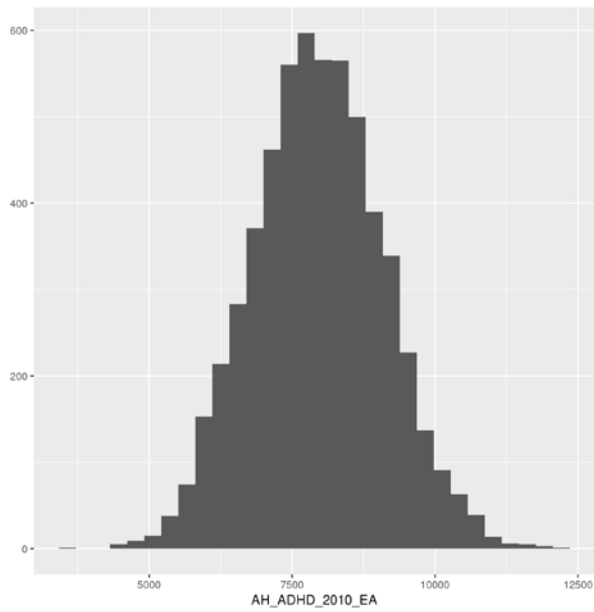
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution
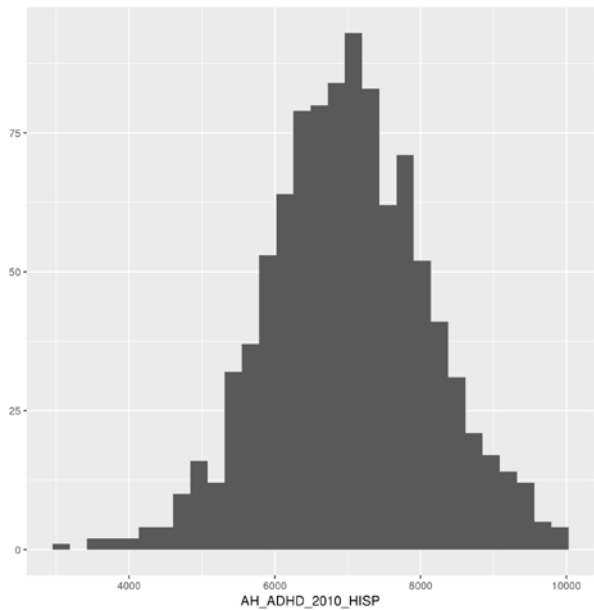
Asian ancestry raw PGS distribution

Attention-Deficit/Hyperactivity Disorder (2017)

GWAS Summary Statistic Source: Demontis, Ditte et al. 2017. "Discovery Of The First Genome-Wide Significant Risk Loci For ADHD." BioRxiv 145581.
(This is a pre-print version of the study which is currently in press. http://orca.cf.ac.uk/110923/)

GWAS Ancestry Group(s): European (Ncases = 19,099; Ncontrols = 34,194), Han Chinese (Ncases = 1,040; Ncontrol = 963); Multi-ethnic (PUWMa sample with N = 78)

The PGSs for attention-deficit/hyperactivity disorder 2017 PGSs were created using results from a 2017 BioRxiv preprint from the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) and Psychiatric Genetics Consortium (PGC). The GWAS meta-analysis files are publicly available from: http://www.med.unc.edu/pgc/results-and-downloads (Download ADHD GWAS - June 2017).

The discovery meta-analysis included 55,374 individuals (20,183 cases and 35,191 controls) from 12 studies of mixed ancestry while the replication sample included 93,916 individuals from two cohorts of multiple ancestry groups. Additional covariates included study specific controls and genetic ancestry principal components.

The Add Health PGSs for attention-deficit/hyperactivity disorder 2017 (ADHD_2017) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
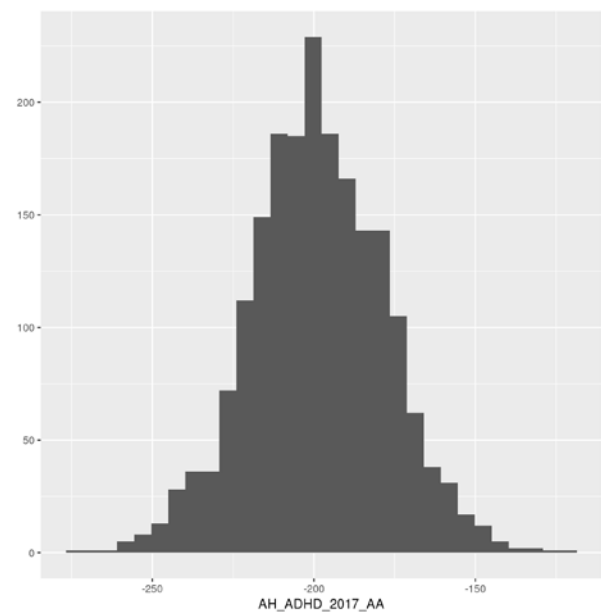
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	-373.65	-180.42	-273.61	-273.70	0.35
African ancestry	1976	-275.51	-122.68	-199.21	-198.88	0.47
Hispanic ancestry	988	-345.63	-178.30	-255.54	-255.59	0.80
Asian ancestry	437	-277.52	-160.05	-222.46	-222.73	1.06

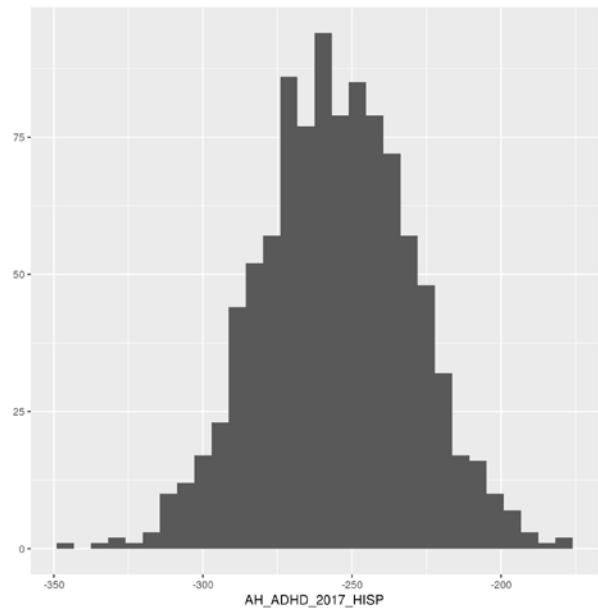# Attention Deficit Disorder 2017

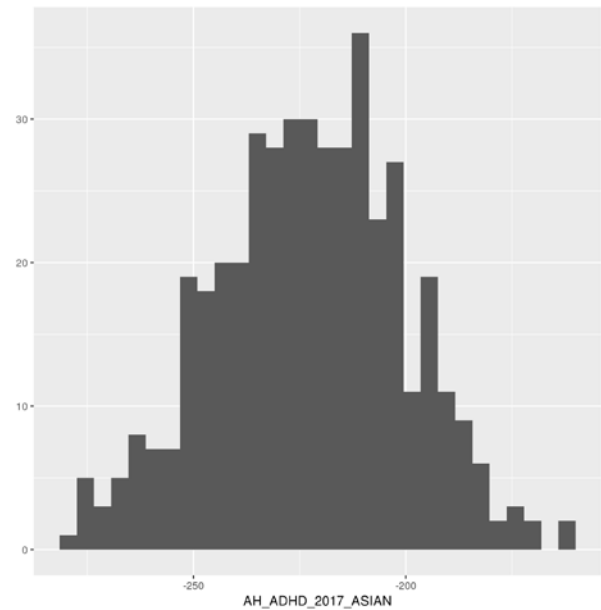### European ancestry raw PGS distribution
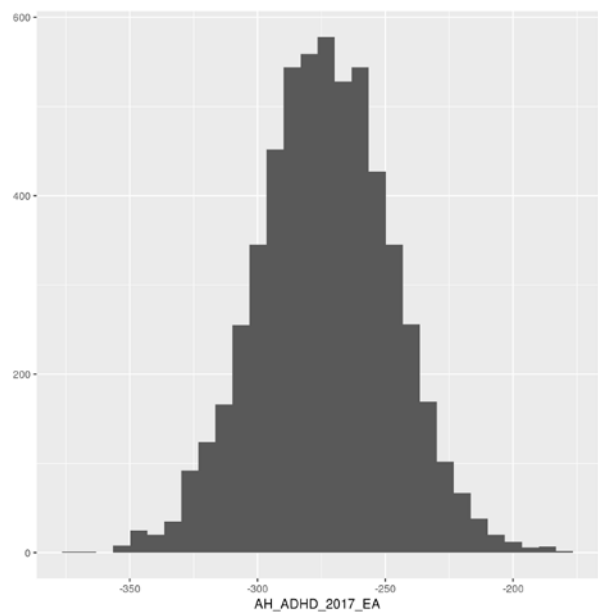
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Bipolar Disorder

GWAS Summary Statistic Source: Psychiatric GWAS Consortium Bipolar Disorder Working Group. 2011. "Large-Scale Genome-Wide Association Analysis of Bipolar Disorder Identifies a New Susceptibility Locus near ODZ4." Nature Genetics 43(10):977–83.

GWAS Ancestry Group(s): European

The PGSs for bipolar disorder were created using results from a 2011 GWAS conducted by the Bipolar Disorder working group of the Psychiatric GWAS Consortium. The GWAS meta-analysis files are publicly available from: http://www.med.unc.edu/pgc/results-and-downloads (pgc.bip.2012-04.zip).

The discovery phase of the meta-analysis included 7,481 cases and 9,250 controls. The replication phase of the 34 most significant genetic regions identified in the discovery phase was conducted in a sample of 4,496 cases and 42,422 controls. Samples were drawn from 11 studies (see Table 1 and supplemental materials, from the original publication). All analyses adjusted for the top five principal components of the genetic data and study fixed effects. Bipolar disorder was measured as diagnoses of either: bipolar disorder type 1, bipolar disorder type 2, schizoaffective disorder bipolar, and a few cases with other bipolar diagnoses (p. 997).

The Add Health PGSs for bipolar disorder (BP) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification, but to give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
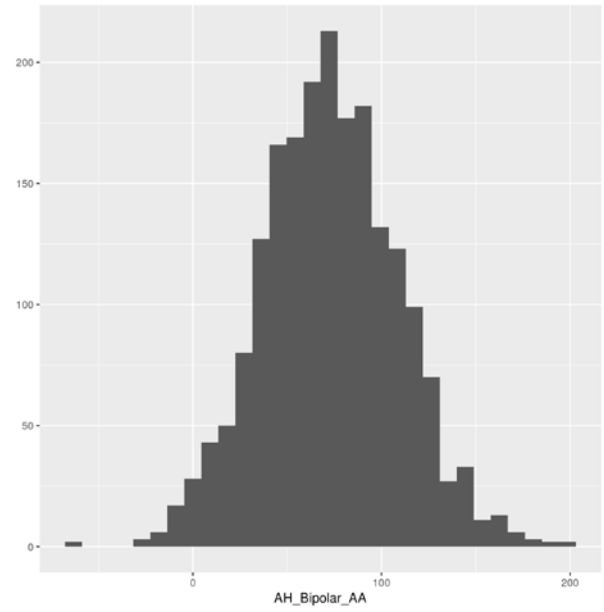
PGS Summary Statistics by Ancestry

| Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	52.44	376.12	229.96	229.63	0.60
African ancestry	1976	-60.16	201.54	71.35	72.77	0.81
Hispanic ancestry	988	27.50	368.26	214.10	214.64	1.40
Asian ancestry	437	17.17	247.92	130.03	130.60	1.69

Bipolar Disorder
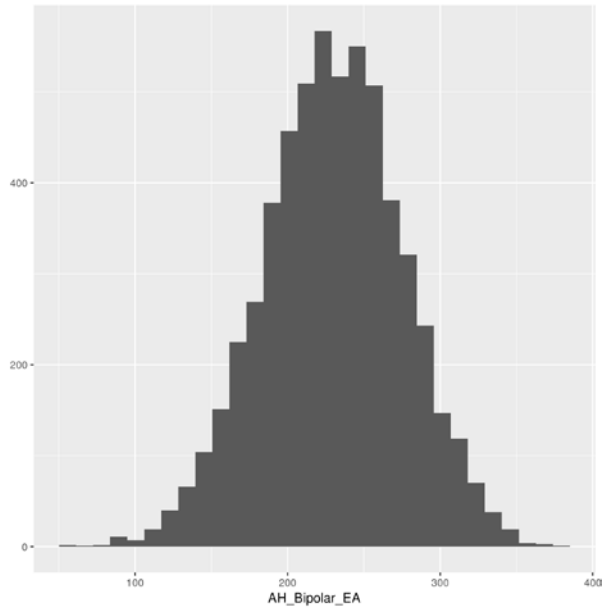
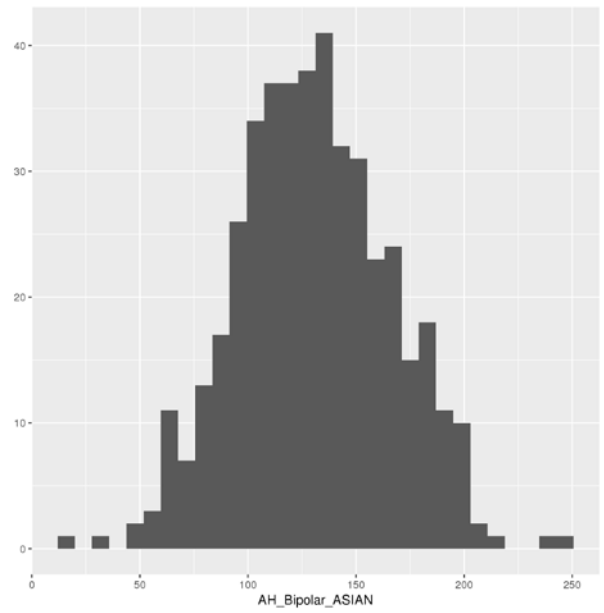### European ancestry raw PGS distribution

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution
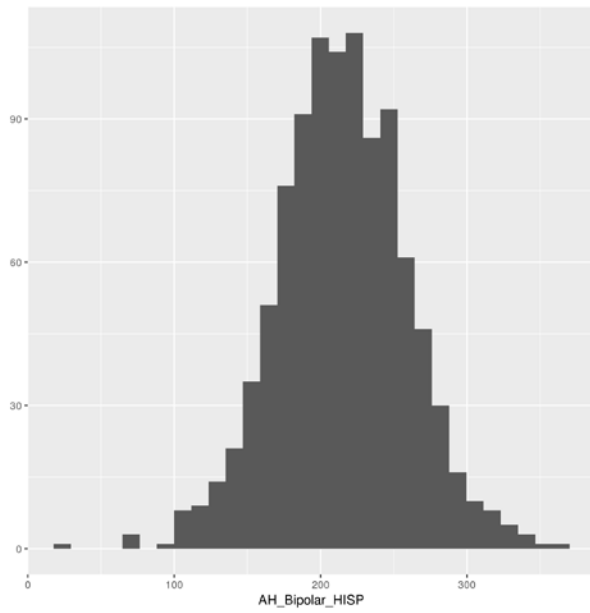
Asian ancestry raw PGS distribution

Major Depressive Disorder (2013)

GWAS Summary Statistic Source: Ripke, Stephan et al. 2013. "A Mega-Analysis of Genome-Wide Association Studies for Major Depressive Disorder." Molecular Psychiatry 18(4):497–511.

GWAS Ancestry Group(s): European

The PGSs for major depressive disorder 2013 were created using results from a 2013 GWAS conducted by the Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. The GWAS meta-analysis files are publicly available from: https://www.med.unc.edu/pgc/results-and-downloads.

The MDD GWAS combined meta-analysis included 9,240 cases and 9,519 controls (N=18,759). The replication sample consisted of 6,783 cases and 50,695 controls. Study level fixed effects and the first five genetic principal components were included as covariates. Analyses were also stratified by sex, recurrent MDD, recurrent early-onset MDD, age of onset, pre-pubertal onset MDD, and latent classes of MDD criteria.
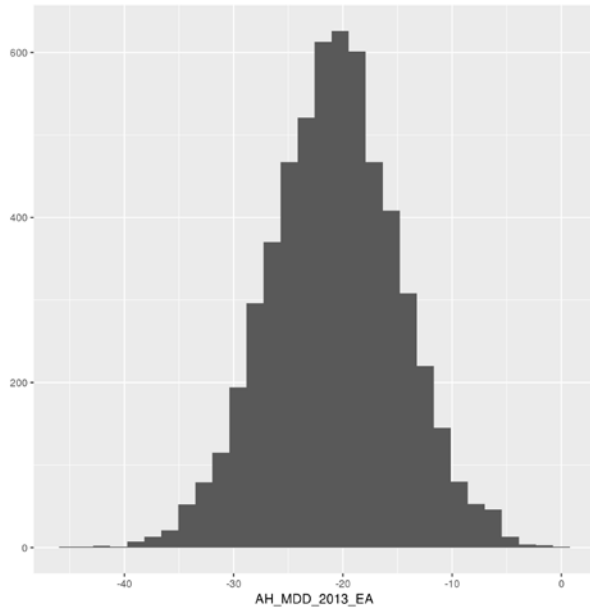
The Add Health major depressive disorder 2013 (MDD_2013) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
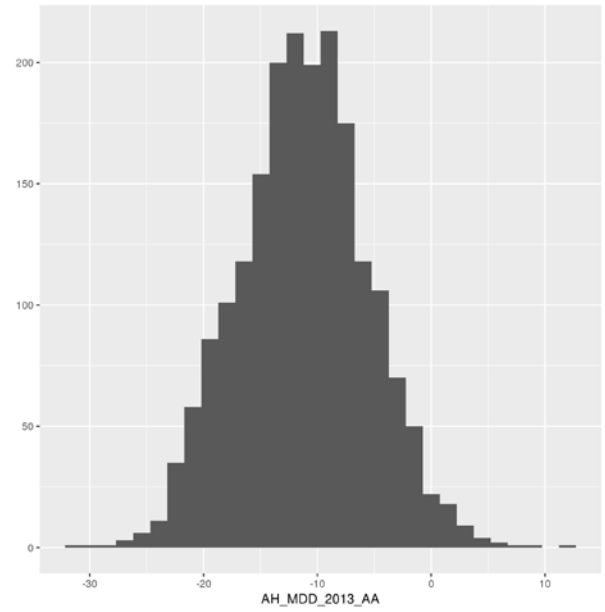
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	-45.37	-0.18	-20.77	-20.81	0.08
African ancestry	1976	-31.68	11.70	-11.21	-11.23	0.13
Hispanic ancestry	988	-34.42	7.45	-15.62	-15.62	0.19
Asian ancestry	437	-23.22	9.22	-7.17	-7.21	0.27

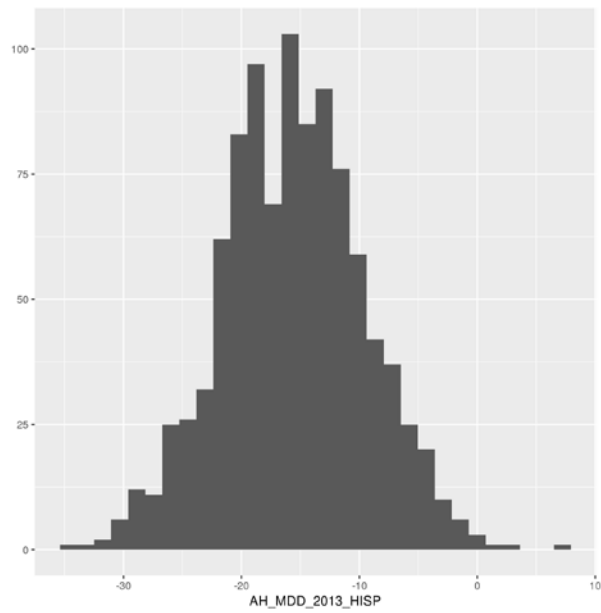# Major Depressive Disorder 2013

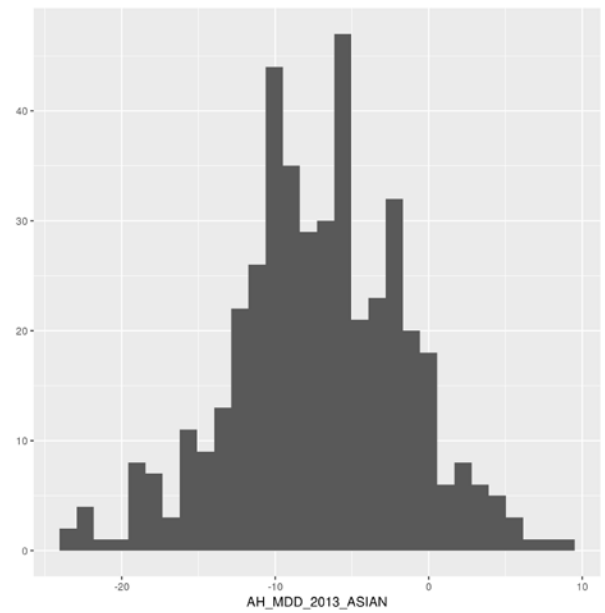### European ancestry raw PGS distribution

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Major Depressive Disorder (2018)

GWAS Summary Statistic Source: Wray, Naomi R. et al. 2018. "Genome-Wide Association Analyses Identify 44 Risk Variants and Refine the Genetic Architecture of Major Depression." Nature Genetics 50(5):668–81.

GWAS Ancestry Group(s): European

The PGSs for major depressive disorder 2018 were created using results from a 2018 GWAS conducted by the Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. The GWAS meta-analysis files are publicly available from: https://www.med.unc.edu/pgc/results-and-downloads.

The MDD 2018 GWAS combined meta-analysis included 135,458 cases and 344,901 controls (N = 480,359) from 35 studies of individuals of European descent. Due to data sharing restrictions from 23andMe, the PGSs are based on summary statistics that exclude the 23andMe cohort (Ncase = 75,607; Ncontrols = 231,747). MDD is cases are based on meeting standard criteria for major depressive disorder.

The Add Health PGSs for major depressive disorder 2018 (MDD_2018) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
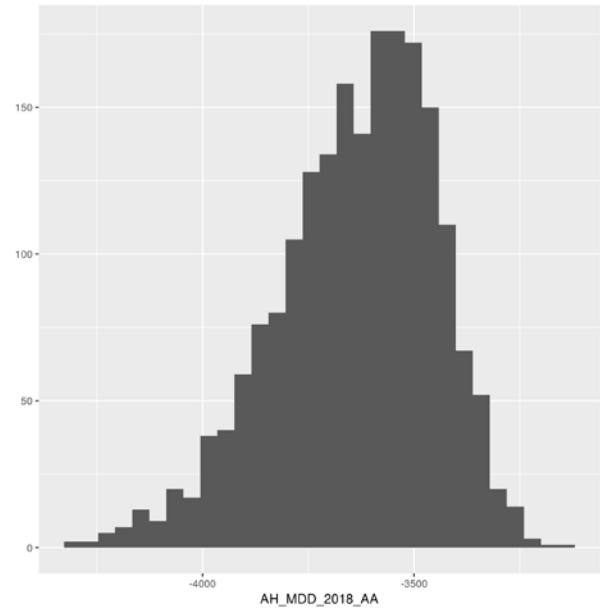
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	-5603.43	-5373.96	-5528.12	-5526.08	0.31
African ancestry	1976	-4310.80	-3143.47	-3615.33	-3638.27	4.23
Hispanic ancestry	988	-5535.19	-4882.16	-5352.34	-5335.70	3.75
Asian ancestry	437	-5426.82	-5157.03	-5312.56	-5310.87	1.39

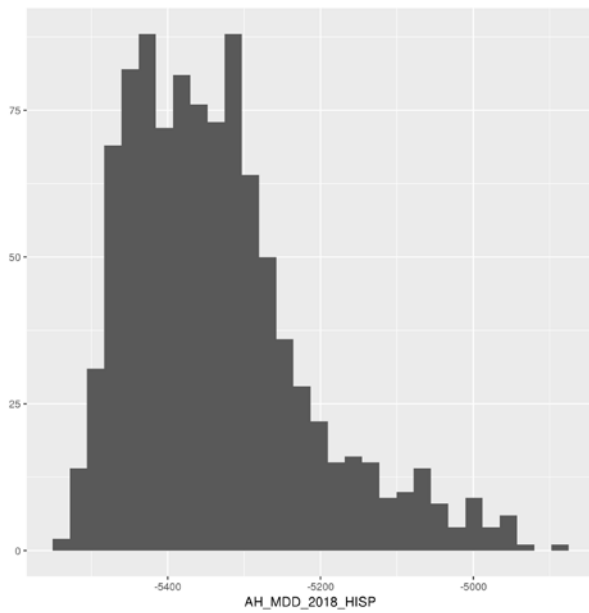# Major Depressive Disorder 2018

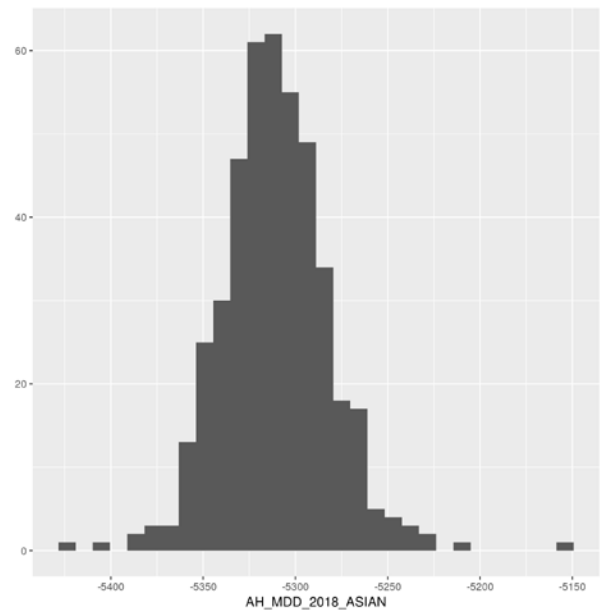### European ancestry raw PGS distribution
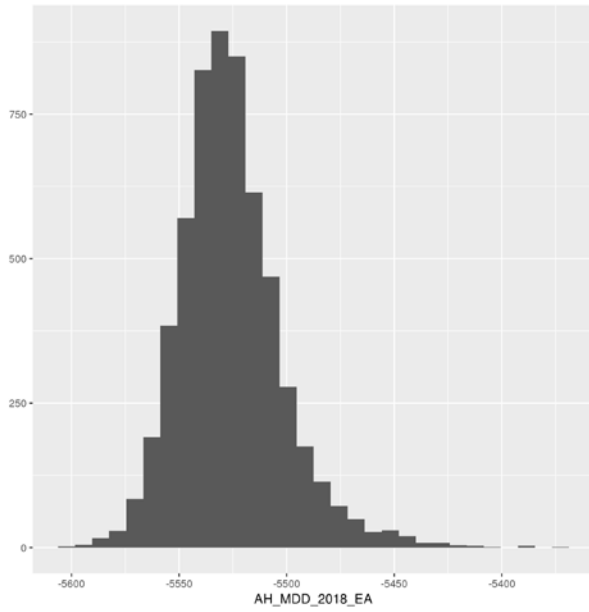
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution

Schizophrenia

GWAS Summary Statistic Source: Ripke, Stephan et al. 2014. "Biological Insights from 108 Schizophrenia-Associated Genetic Loci." Nature 511(7510):421–27.

GWAS Ancestry Group(s): European and East Asian [Japan, Singapore, Hong Kong and Sichuan China] (Ncases = 1,866; Ncontrols = 3,418)

The PGSs for schizophrenia were created using results from a 2014 GWAS conducted by the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC). The GWAS meta-analysis files are publicly available from: https://www.med.unc.edu/pgc/results-and-downloads.

The schizophrenia GWAS combined meta-analysis included 36,989 cases and 113,075 controls (N=150,064) from individuals of European and East Asian descent. The replication sample consisted of 1,513 cases and 66,236 controls from individuals of European descent. To enable acquisition of large samples, some of the participating groups ascertained cases via clinician diagnosis rather than a research-based assessment. Genetic principal components and study fixed effects were included as covariates.

The Add Health schizophrenia (SCZ) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

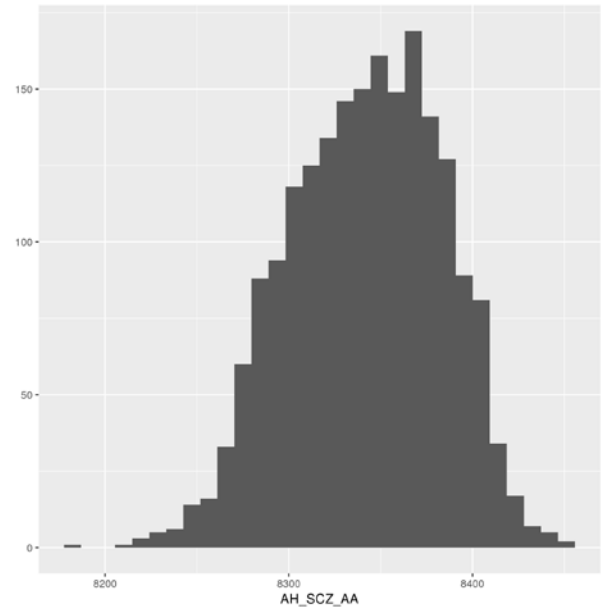PGS Summary Statistics by Ancestry

| Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	7968.42	8247.90	8119.95	8119.10	0.59
African ancestry	1976	8180.00	8448.73	8344.16	8341.85	0.93
Hispanic ancestry	988	8044.11	8343.57	8190.51	8189.40	1.61
Asian ancestry	437	8179.38	8387.42	8274.14	8278.68	1.92

Schizophrenia

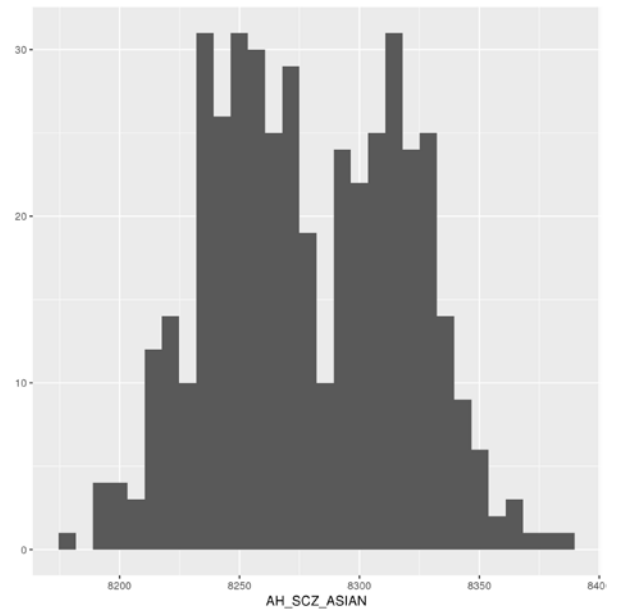### European ancestry raw PGS distribution
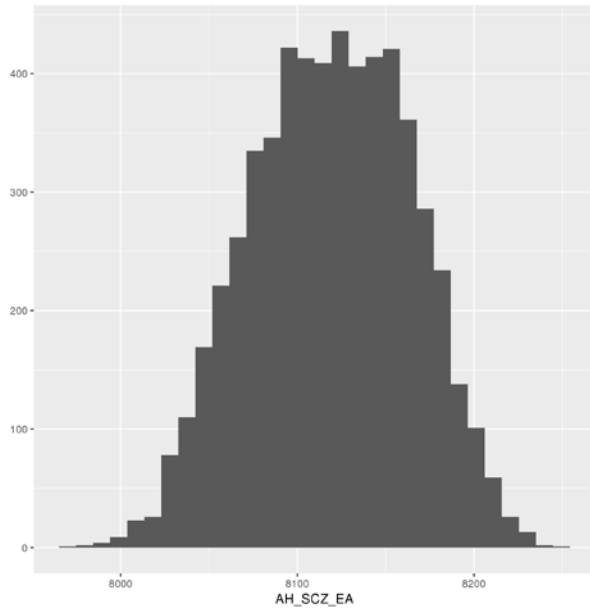
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution
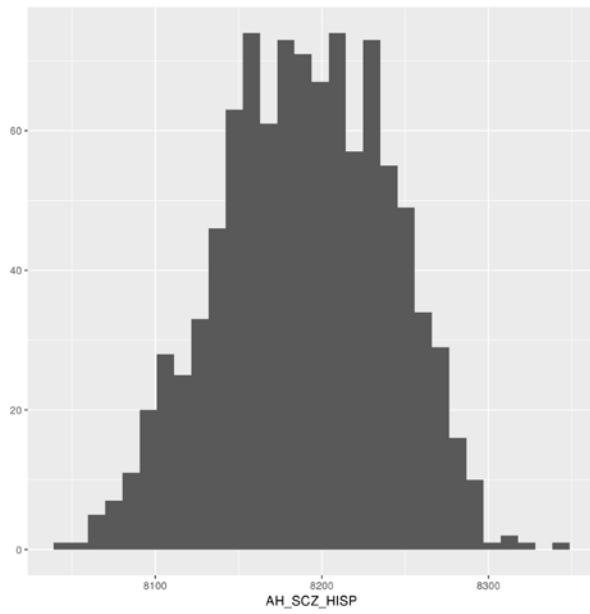
Asian ancestry raw PGS distribution

Mental Health Cross Disorder

GWAS Summary Statistic Source: Cross-Disorder Group of the Psychiatric Genomics Consortium. 2013. "Identification of Risk Loci with Shared Effects on Five Major Psychiatric Disorders: A Genome-Wide Analysis." Lancet (London, England) 381(9875):1371–79.

GWAS Ancestry Group(s): European

The PGSs for mental health cross-disorder were created using results from a 2013 GWAS conducted by the Cross Disorder working group of the Psychiatric GWAS Consortium. The GWAS meta-analysis files are publicly available from: http://www.med.unc.edu/pgc/results-and-downloads (pgc.cross.full.2013-03.zip).

The discovery phase of the meta-analysis included 33,342 cases and 27,888 controls. Disorders that were counted as cases (DSM-III-R or DSM-IV criteria) included autism spectrum disorder, attention-deficit/hyperactivity disorder, bipolar disorder, major depressive disorder, and schizophrenia.

The Add Health PGSs for mental health cross-disorder (XDIS) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
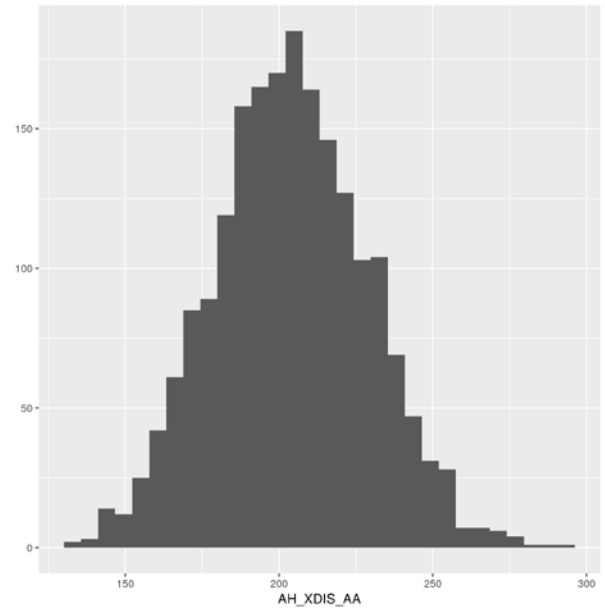
PGS Summary Statistics by Ancestry

| Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	188.20	385.08	304.70	302.47	0.36
African ancestry	1976	130.32	290.89	203.32	203.77	0.56
Hispanic ancestry	988	235.51	398.57	319.01	318.42	0.79
Asian ancestry	437	249.04	379.93	320.72	321.30	0.90

Cross Disorder

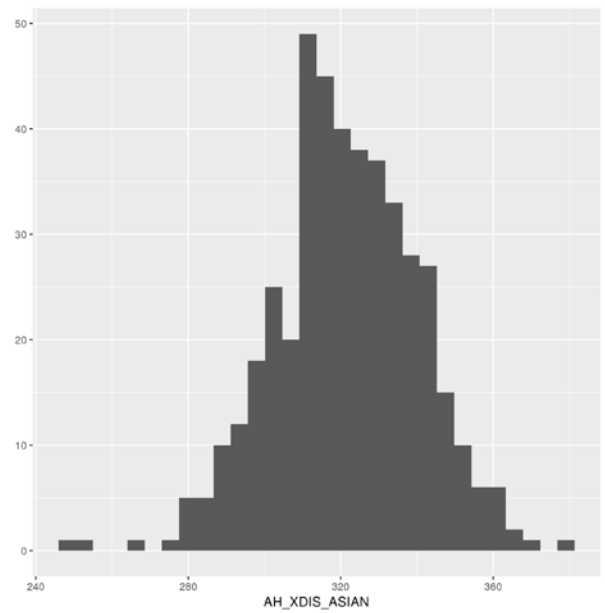### European ancestry raw PGS distribution
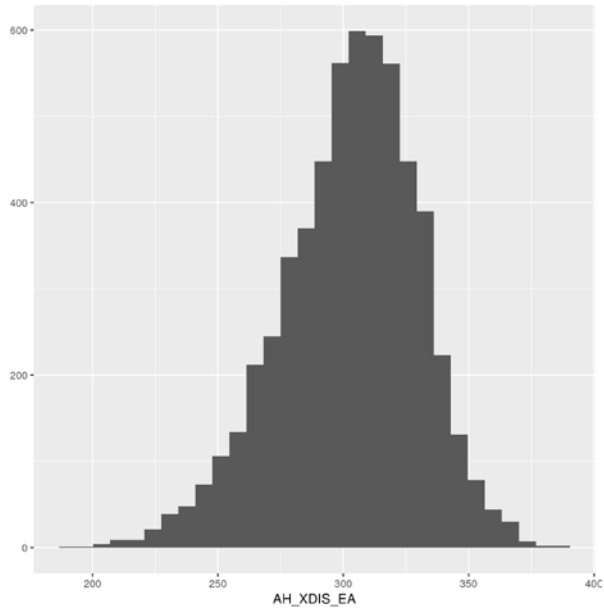
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution
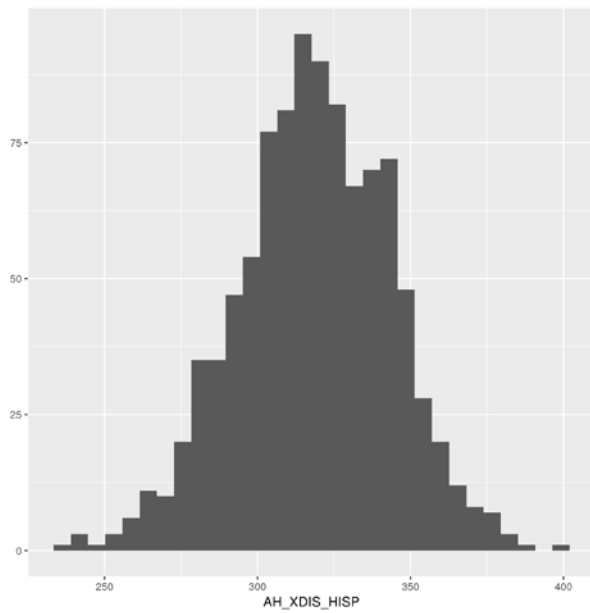
Asian ancestry raw PGS distribution

Alzheimer's Disease

GWAS Summary Statistic Source: Lambert, Jean-Charles et al. 2013. "Meta-Analysis of 74,046 Individuals Identifies 11 New Susceptibility Loci for Alzheimer's Disease." *Nature Genetics* 45(12):1452–58.

GWAS Ancestry Group(s): European

The PGSs for Alzheimer's disease were created using results from a 2013 GWAS conducted by the International Genomics of Alzheimer's Project (IGAP). The GWAS meta-analysis files are publicly available from: http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php.

The meta-analysis of Alzheimer's disease included 54,162 participants (cases=17,008 and controls=37,154) of European decent imputed to 1000 Genomes 2010 release. The replication sample included 19,884 participants of European ancestry (cases=8,572 and controls=11,312). Additional covariates included in the GWAS for each contributing cohort included age, sex, and genetic principal components.

The Add Health PGSs for Alzheimer's disease (AD) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
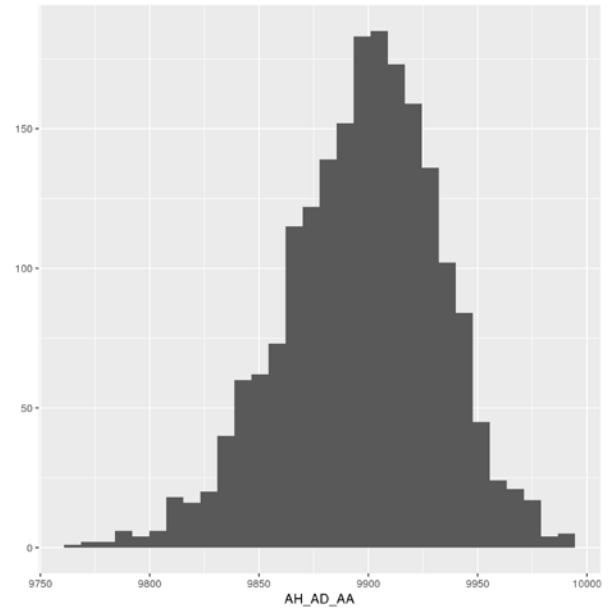
PGS Summary Statistics by Ancestry

| Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	9485.12	9685.55	9592.68	9592.61	0.35
African ancestry	1976	9765.04	9990.86	9899.59	9897.14	0.80
Hispanic ancestry	988	9525.81	9748.33	9646.47	9646.80	1.06
Asian ancestry	437	9654.28	9810.60	9726.27	9727.17	1.18

Alzheimer's Disease
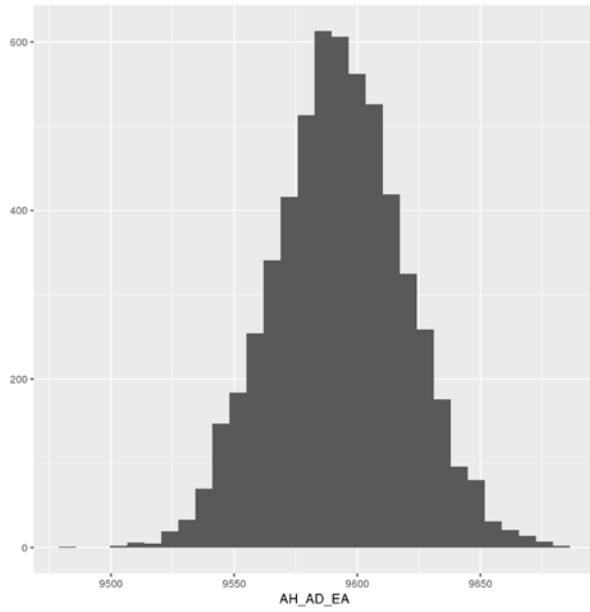
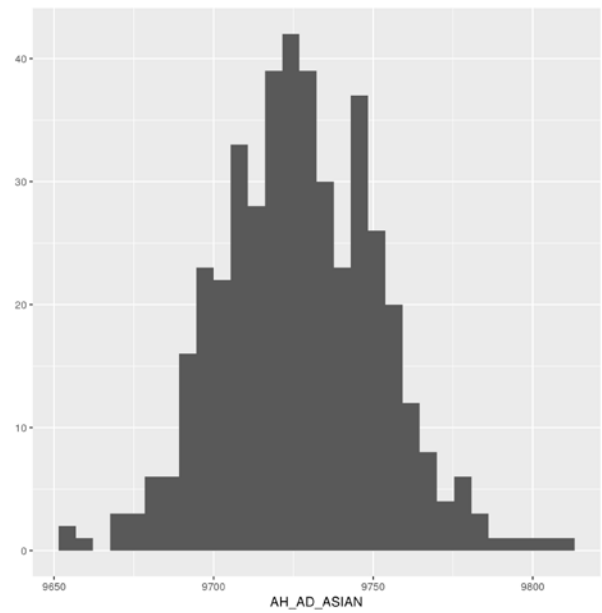### European ancestry raw PGS distribution

African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution
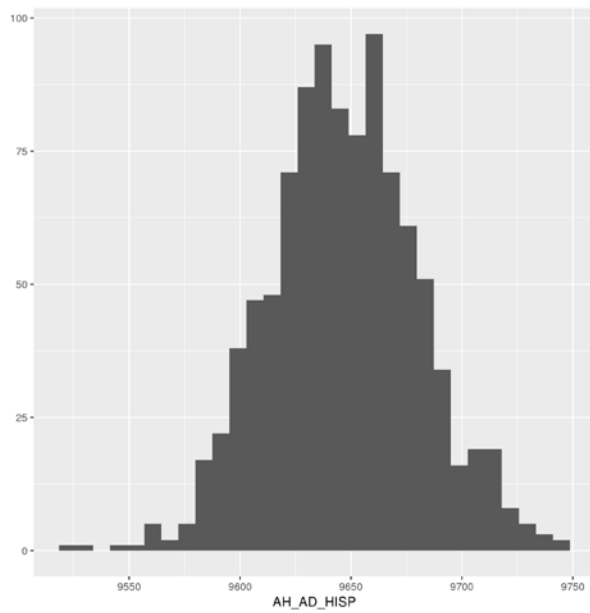
Asian ancestry raw PGS distribution

Educational Attainment (2016)

GWAS Summary Statistic Source: Okbay, Aysu et al. 2016. "Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment." *Nature* 533(7604):539–42.

GWAS Ancestry Group(s): European

The educational attainment 2016 PGSs were created using results from a 2016 study by the Social Science Genetic Association Consortium (SSGAC). The GWAS meta-analysis files are publicly available from: https://www.thessgac.org/data.

Due to data sharing restrictions from 23andMe, the PGSs are based on summary statistics that exclude the 23andMe cohort. Study-specific GWASs controlled for the first ten principal components of the genotypic data, age, a second-order polynomial of age, a third-order polynomial in age, an indicator for being female, interactions between age and female, and study-specific controls, including dummy variables for major events such as wars or policy changes that may have affected access to education in their specific sample.

The Add Health PGSs for educational attainment 2016 (EDU_2016) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
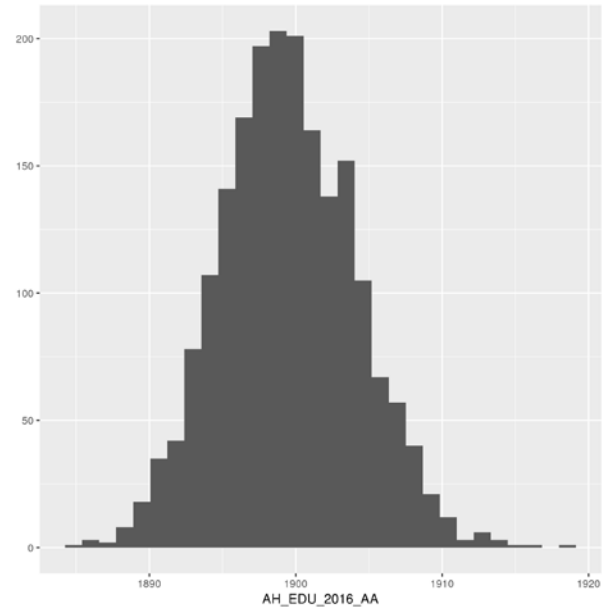
PGS Summary Statistics by Ancestry

| Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	1890.33	1936.13	1910.81	1910.77	0.08
African ancestry	1976	1884.97	1918.69	1899.27	1899.45	0.10
Hispanic ancestry	988	1890.18	1926.79	1907.27	1907.65	0.18
Asian ancestry	437	1891.95	1923.16	1904.93	1905.06	0.23

Educational Attainment 2016

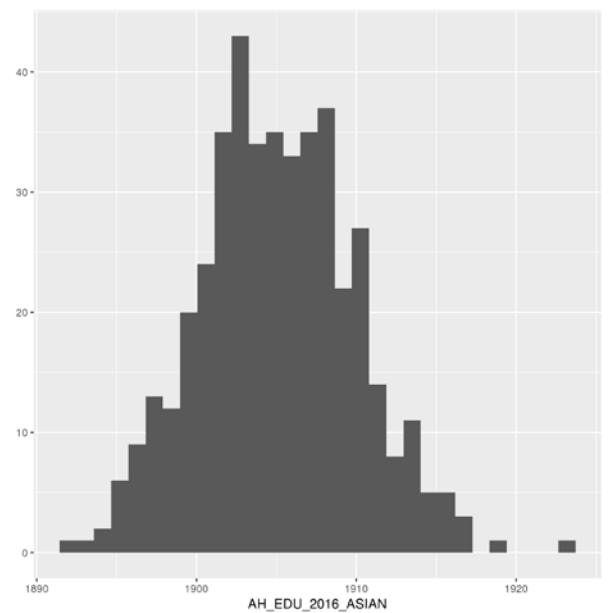## European ancestry raw PGS distribution
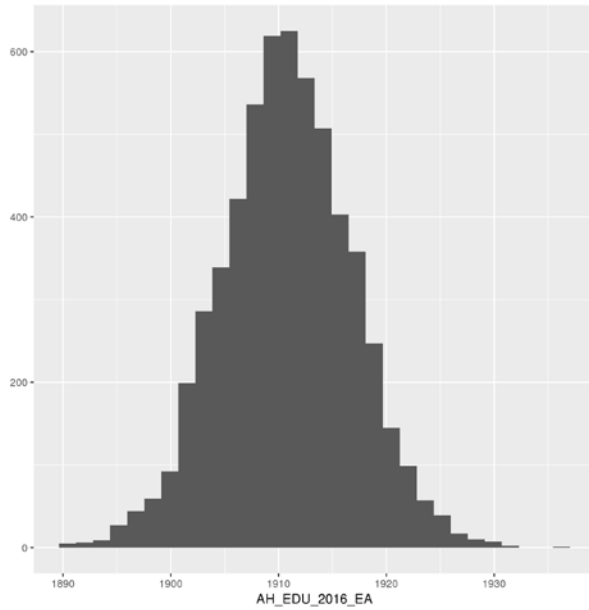
African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution
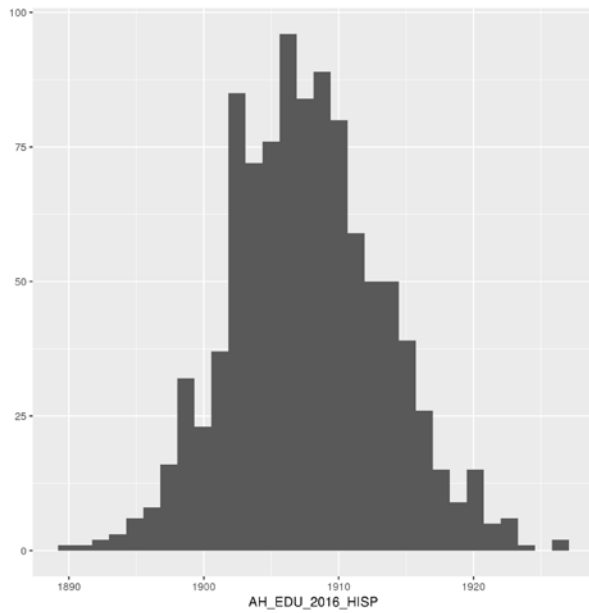
Asian ancestry raw PGS distribution

Educational Attainment (2018)

GWAS Summary Statistic Source: "Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment", Lee et al., Nature Genetics, 2018. (forthcoming).

GWAS Ancestry Group(s): European

The educational attainment 2018 PGSs were created using results from a forthcoming 2018 study by the Social Science Genetic Association Consortium (SSGAC). The GWAS meta-analysis files are publicly available from: https://www.thessgac.org/data. Due to data sharing restrictions from 23andMe, the PGSs are based on summary statistics that exclude the 23andMe cohort.
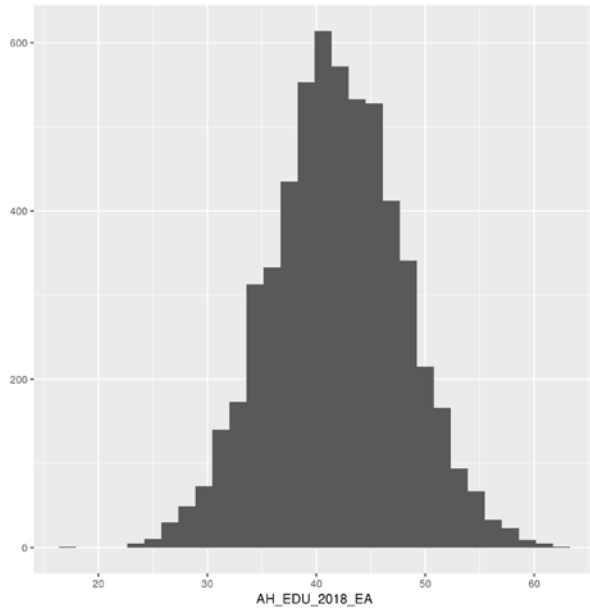
The discovery meta-analysis of educational attainment 2018 included 1,131,881 individuals of European decent imputed to 1000 Genomes 2010 release.

The Add Health PGSs for educational attainment 2018 (EDU_2018) are standardized (mean=0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present a few descriptive statistics and density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**
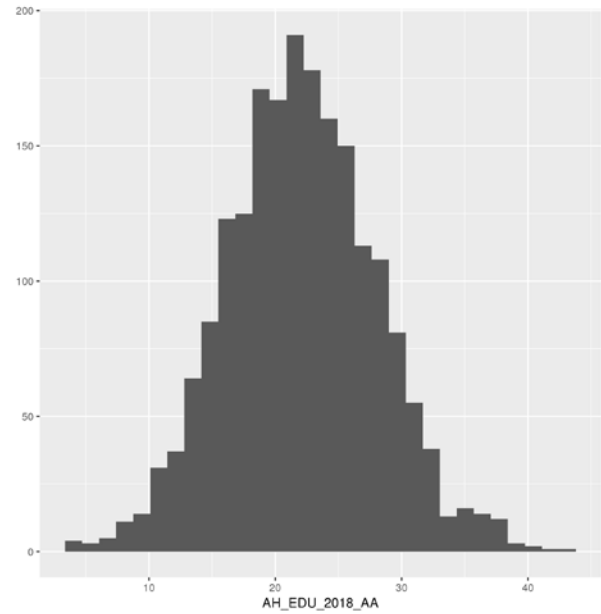
PGS Summary Statistics by Ancestry

Group	N	Min	Max	Median	Mean	SE (mean)
European ancestry	5728	16.96	62.26	41.77	41.83	0.08
African ancestry	1976	3.96	43.04	21.88	22.05	0.13
Hispanic ancestry	988	24.90	60.35	40.48	40.80	0.18
Asian ancestry	437	31.43	59.21	45.30	45.36	0.23

Educational Attainment 2018

European ancestry raw PGS distribution

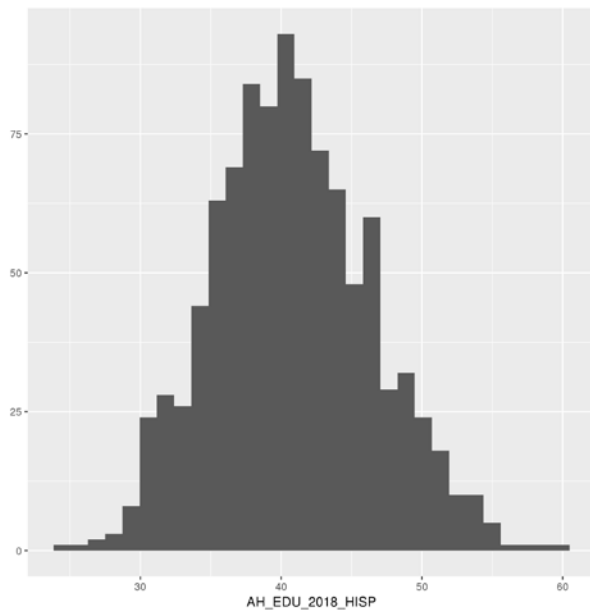## African ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Asian ancestry raw PGS distribution