SUPPLEMENTARY MATERIALS – SA1 (SUPPLEMENTARY APPENDIX 1)

SUPPLEMENTARY MATERIALS: DETAILED METHODS AND RESULTS

Quality Control. In preparation for our focused candidate gene study of ~70% of the EA portion of the control sample [1], we excluded controls endorsing non-European ancestry (except partial Native American ancestry, dues to its excess reporting) on the questionnaire. We also excluded 1.4% of the controls for endorsing more than 50 of 69 screening or personality items, refusing to answer 5 or more of these questions (0.4%), or not being fully screened due to software failures (0.6%). We excluded the 8% of the EA controls who endorsed or failed to deny the "psychosis screens", namely, previous treatment or diagnosis of schizophrenia, schizoaffective disorder, auditory hallucinations, delusions, or bipolar disorder. In the end, we genotyped 2,126 EA controls and 1,952 EA cases, of which 2,002 (94%) controls and 1,870 (96%) cases remained for analysis after our sample QC, i.e., very similar proportions.

The control sample QC [1] consisted of excluding 51 subjects with aggregate genotype call rates less than 95% across all 833 valid SNPs, 23 subjects with unresolved sex typing (amelogenin) discrepancies, 15 unexpected duplicate (2) or unexpected related (13) subjects, and 35 subjects that lay outside the EA cluster in a principle components analysis of ancestry-informative SNP marker data. Compared to cases, the controls had more unresolved sex typing discrepancies (23 versus 2, χ^2=14.45, p=0.00014); there was very limited ability to resolve such discrepancies (e.g., via confirming questionnaire gender and/or collecting another DNA sample) in the controls due to the anonymization procedure. Conversely, more cases were excluded than controls due to the principle

components analysis of ancestry (63 versus 35, χ^2=10.18, p=0.0014); this may be related to collection of more ancestry information in controls (grandparents) than in cases (parents) which allowed exclusion based on phenotypic ancestry (i.e., prior to genotyping), or perhaps more cases having limited knowledge of their biological ancestors (such as if more cases than controls were adopted).

When assembling the sample to be genotyped for our GWAS of schizophrenia, we continued the above exclusions (51 + 23 + 15 + 35) known from the genotyping QC for the initial EA sample, but also evaluated the entire controls collection, including the AA sample and the remainder of the EA sample, for the following. Cell lines that did not grow (0.5%), subjects with excess positive, refused, or not asked screening questions as described above (1.4%), and subjects endorsing or failing to deny the "psychosis screens" (9.7%, due to the higher rate in the AA sample, see discussion) were excluded. For our schizophrenia GWAS, we selected from the remaining controls those whose grandparental ancestry information was EA (as above) or AA, which left a total of 3,827 (2,817 EA and 1,010 AA) of the 4,665 collected controls. These 3,827 controls, along with 4,196 (2,838 EA and 1,358 AA) cases, were genotyped with the Affymetrix 6.0 array at the Broad Institute (Cambridge, MA). A total of 3,626 (95%) control samples (2,653 EA and 973 AA) passed stringent quality control, and a total of 3,967 (95%) case samples (2,681 EA and 1,286 AA) passed QC [2]. Thus, the internet recruited and assessed control sample and the directly assessed cases sample had equal overall rates of molecular QC sample exclusion. The most common exclusion criteria (sex concordance, genotyping completion rate, cryptic relatedness, heterozygosity evaluation, principal components analyses of ancestry) showed similar proportions in the controls versus the

cases samples (see Table S6 in [2], broken down into controls versus cases here, ST2). The exception was cryptic relatedness, for which the controls had a higher rate (17/3,827) than the cases (6/4,196) did (χ^2=5.34, p=0.021), due entirely to the EA sample – on further examination, we found that this was chiefly due to a small portion (<0.5%) of the control sample consisting of parent-child pairs inadvertently recruited from the same household.

SUPPLEMENTARY MATERIALS: DETAILED DISCUSSION

Cases might be more carefully selected or assessed than controls in some studies, and this has been proposed to result in more variability in controls than cases in some settings (e.g., [53]); we have attempted to minimize such problems by our nationwide and standardized controls selection and assessment. However, volunteers, including those responding to ads seeking "normal controls", have a significant chance of having a mental illness – for example, of 121 such subjects assessed with a structured interview, almost half had a current or past history of an axis I diagnosis [42]. This was certainly consistent with the elevated prevalence for most disorders that we found in the AA controls from SSI (panel recruited via internet ads) compared to the AA controls from KN (panel recruited by random digit dialing) – (see ST3). Psychiatric interviews conducted by telephone versus in-person have been shown to obtain comparable information (κ=0.57-0.84) for depression, anxiety disorders, alcohol and drug use disorders [47, 54, 55], with much of the disagreement due to test:retest rather than phone:in-person unreliability [55]. However, this equivalence is generally more certain when topics are not sensitive (discussed in [45]), such as many psychiatric questions. For example, when epidemiological surveys are complemented with drug testing, it has generally been found that drug use is under-reported, though this is more pronounced for

recent versus more distant (e.g., last month versus last year) use [56]. Nevertheless, much of the literature suggests that more anonymous methods may lead to more accurate responses to sensitive topics [46-49]. Some groups (e.g., men versus women, whites versus blacks) differentially report sensitive information in telephone versus in-person interviews for reasons such as social desirability effects (better information gathering with a less personal method) or level of trust (better information gathering with a more personal method) [45]. Many of these effects might be amplified further with an even less personal method than telephone surveys, i.e., internet questionnaire, therefore prompting an examination of the validity of such a method.

Demography and Ancestry. The 3,364 self-identified EA and 1,301 self-identified AA controls have self-reported grandparental ancestries consistent with those identities, many (69% EA and 80% AA) with full grandparental information (all 4 grandparents). A more limited number from these controls with full grandparental information reported only one ancestral category present: for such EA controls, Anglo-Saxon (15%) and West Europe (7%) categories were the most common, and for AA controls, 42% had AA-only ancestry for all 4 grandparents. These more homogenous groups may be useful in some studies utilizing these controls. These controls match well to some key demographic categories (gender, marital status) of the general population of the U.S. via comparison to the 2003 CPS data, though these controls are shifted upwards in age, which yields somewhat of an advantage in the sense that the substantial majority of these controls are past most of the typical ages of onset for the common psychiatric disorders for which we assayed. It should be noted that these controls are on average more educated than the general population (perhaps reflecting the use of computers in part, or a more general correlation between level of education and research participation),

which may be an issue when using them to compare to cases with disorders manifesting, at least in part, cognitive difficulties. Household income (which was higher than the general population) in these controls is moderately correlated with the level of education in both the EA (r=0.39) and the AA (r=0.36) controls. If one wishes to tailor the controls selected from this sample for age or any other of the items assessed in the questionnaire, this is straightforward to accomplish after accessing the phenotypic information from the NIMH Center for Collaborative Genetics Studies of Mental Disorders and prior to ordering the desired DNA samples.

Alcohol and Drug Dependence. Consistent with previous studies (ST5), we find higher lifetime prevalence estimates for alcohol dependence and for drug dependence in men than in women. Our estimates for the EA controls were higher than our AA (KN) controls for alcohol dependence, partially explained by our EA sample having a higher proportion of males. Conversely, the AA (KN) controls had higher prevalence than the EA sample for drug dependence. Previous studies (e.g., [57]) have found increased risks of lifetime drug dependence not just for males, but also for younger ages, unmarried status, and lowered socioeconomic status (such as educational or income levels). In our full KN control sample, we note small individual associations of some these measures with drug dependence. Our AA (KN) sample's (versus the EA sample's) younger age (mean ages of 46.3 versus 50.0 years), higher rates of unemployment or disability (22.3% versus 10.3%), and lower income levels (average of 8.91 versus 10.57 for the 19 assessed income levels) may contribute to the higher lifetime prevalence of drug dependence in the AA sample. A number of these sample characteristics are relevant for other disorders below. It is noteworthy that a study of an internet assessed sample (derived from the KN Panel) found a higher lifetime prevalence for alcohol dependence compared to a

corresponding directly assessed (i.e., NESARC's face-to-face interviews) sample [7], and that a large computer assisted self-administered interview study (NSDUH, National Survey on Drug Use and Health, [58]) found higher prevalence for alcohol and illicit substance use again compared to NESARC [59]. In both instances, the leading speculation was that a substantial component of the higher detected prevalences was related to the mode effects, namely the more anonymous methods used in the higher prevalence studies [7, 59].

Major Depression. Our MDE lifetime prevalence estimate in controls is likely an overestimate due to omission of organic exclusions, especially alcohol and drug use, not accounting for bereavement, and not requiring distress or impairment. However, we note that lifetime MDE has been diagnosed at similarly high rates in some large studies not included in ST5 (e.g., [60, 61]). Eliminating from consideration of an MDE those individuals also diagnosed with alcohol or drug dependence is an overcorrection to an uncertain degree due to the fact that temporal comorbidity is certainly an unmeasured factor here (e.g., were any MDEs occurring outside the context of ongoing substance use). The more conservative approach in our data is to eliminate from consideration of an MDE those controls admitting that their substance use caused emotional or psychological problems (i.e., anhedonia, depression) – this adjustment overcomes this lack of timing information. While there have been inconsistent literature findings as to the relative proportions of MDE in AA versus EA samples, generally they were similar (see review, [62]), as we have found in these controls.

The CIDI-SF collects no data on bereavement which likely accounts for a proportion of the controls we scored as having MDE, as has been found by others (e.g., [63]) when examining the NCS dataset which used the full CIDI (including bereavement

assessment). For the impairment/distress criterion, we do have CIDI-SF queries that are not included in the scoring, though one might wonder how a subject could satisfy criterion A for MDE without at least being distressed if not impaired – indeed, this was the rationale for not including impairment/distress in the CIDI-SF scoring for MDE (or any of the other CIDI-SF assessed disorders) [20]. We find that requiring impairment or distress decreases the frequency of MDE, but only a little (by 3-7%). For investigators wishing to use this control sample and avoid selecting controls with depression, a conservative course would be to exclude any controls with a possible MDE even though some would not survive the "organic rule-out" criterion (or others discussed above); still almost 60% of the sample would remain.

Anxiety Disorders. For the same reasons as with MDEs, excluding from anxiety disorder consideration any subjects with alcohol or drug dependence is surely an overcorrection. Eliminating from GAD consideration any subjects with an MDE reduces the estimate below the NCS one, likely due to the fact that while the MDE and GAD comorbidity is high, it is not complete, and we lack timing information (e.g., how many instances of GAD period were occurring outside the context of an MDE). The prevalence was higher for all anxiety disorders in women, and generally equivalent in EA versus AA (KN) controls, the exceptions being specific phobia, agoraphobia without panic attacks, and OCD, which were more common in AA. Anxiety disorders prevalence was higher than the NCS data in almost all cases (through rather similar for specific phobia and social phobia), but the female preponderance in our data was consistent with the literature [64], as was the significant comorbidity with depression and among the anxiety disorders.

Nicotine Dependence. Though the gap has narrowed between the percentage of men versus women smoking, our small male preponderance of nicotine dependence is consistent with men in the U.S. still being more likely to smoke daily than women [65]. Our most correlated demographic feature with nicotine dependence (lower educational attainment) has been consistently associated in many diverse U.S. populations (see review, [65]). Consistent with our findings, when compared to EA subjects, AA individuals are less likely become nicotine dependent (e.g., [66]).

Neuroticism and Extraversion. The elevation of neuroticism scores in females versus males, and the elevated neuroticism and lowered extraversion seen in MDE and anxiety disorders in our sample are consistent with a range of prior findings (e.g., [67, 68]). A recent study compared two groups of "hypernormal" controls: (1) 534 of the MGS2 controls (at least 30 years old) having no CIDI-SF diagnoses (MDE, GAD, specific phobia, social phobia, agoraphobia, panic attack, alcohol dependence, drug dependence, OCD) and denying all psychosis and mania screens, and (2) 90 same aged controls directly (face-to-face) assessed with a structured interview (Schedule for Affective Disorders and Schizophrenia – Lifetime Version, and Family History Screen) as having no diagnosis and no family history of anxiety [52]. These two groups were indistinguishable on both neuroticism and extraversion scores, validating the "hypernormality" of this sizable subset of the MGS2 controls (1,424 EA and 420 AA MGS2 controls have no CIDI-SF diagnoses and deny all psychosis and mania screens) assessed by an anonymous internet questionnaire.

Comorbidity. Comorbidity was extensive in this sample, as assessed by the CIDI-SF. This was also found in the NCS where it was noted: "the major burden of

psychiatric disorder … is concentrated in a group of highly comorbid people who constitute about one sixth of the population" [31].

Other Conditions and Traits. The frequency of being overweight and of obesity in our sample is similar to contemporaneous estimates from the National Health and Nutrition Examination Study (NHANES) – and we note that in our sample AA women were the most affected group (KN: 73.8% overweight or obese and 50.1% obese; NHANES: 78.0% and 53.9%, respectively) as in NHANES [40]. In terms of sexual orientation (identity), as in previous large surveys, we find more homosexual men than homosexual women, a higher ratio of bisexuals to homosexuals in women than in men, and similar percentages of homosexuality compared to previous studies (e.g., [41]), lending support to our sample being representative in this regard as well.

Psychosis and Mania. The difference in the studied EA versus AA (KN) controls for endorsing the schizophrenia screening question (2.3% AA versus 0.7% EA) is somewhat larger than the difference found in the lifetime prevalence of schizophrenia in the Epidemiological Catchment Area (ECA) study (2.1% AA versus 1.4% EA) [69]. We do note that the appearance of an EA versus AA schizophrenia lifetime prevalence difference in the ECA data disappeared when controlling for age, gender, marital status, and socioeconomic status [69]. Another possible contributor to our finding of more frequent endorsement of the psychosis and mania screening questions in AA (KN) versus EA controls may be the higher prevalence we found in AA (KN) controls for drug dependence (16.3% versus 12.2%), especially in males, with many of the drug classes increasing the likelihood of manifesting psychotic and/or manic symptoms. Of course, our EA and AA control information for psychosis and mania is limited to screening

questions and must not be confused with ECA or NCS diagnoses that rely on much more extensive symptomatic data [31, 69].

Our approach, which we recommend for association studies where the cases are any of these disorders below, is to be conservative and exclude as controls any of the subjects endorsing, unsure, or refusing (refusals were very few and not indicated in Table 3, but such refusals could possibly be due to paranoia) to answer the screening questions about treatment and/or diagnosis (or presence) of schizophrenia, schizoaffective disorder, auditory hallucinations, delusions, bipolar disorder, and/or manic-depression. Such an approach would lead to 8.2% of the EA controls and 13.6% of the AA (KN) controls being excluded from such control groups (a few less than summing the data from Table 3 indicate due to some subjects answering one such screen as endorsed and another as unsure).

Comparability and Utility of an Internet Based Control Sample. We have primarily compared the lifetime prevalence estimates we detected with the CIDI-SF to the corresponding prevalence estimates from previous large population-based studies, in particular the NCS due to its use of the CIDI, and we consistently produce higher estimates, a pattern that has been previously seen in select groups directly assessed with both instruments (e.g., in patients with HIV [44]). We note some main reasons include the lack of exclusion criteria, both disorder specific (e.g., bereavement) and more widely applicable (i.e., the organic exclusions), and often the impairment and/or distress inclusion criterion. Without temporal information, we are limited in exploring many such possibilities, but exploratory application of the information that was collected (substance use, questions assessing impairment or distress) bring our estimates closer to the NCS estimates. We also note that our control sample was collected ~15 years after the NCS

sample, raising the possibility that different cohorts might have somewhat different prevalence (see ST5 for comparisons to the earlier ECA and the later NCS-R and NESARC datasets), for example for substance use disorders.

It is encouraging that some other patterns we find match well with the literature, such as the male to female ratio for various disorders, elevated neuroticism and diminished extraversion seen with depression and anxiety disorders, the finding of significant comorbidity (such as seen in the NCS), and the associations of some demographic features with various disorders. Our collection of a variety of information allows interested users of the control sample to use approaches such as generating factor scores, e.g., an "internalizing" one incorporating items such as major depression, anxiety disorders, and neuroticism, and also to utilize covariates (chiefly various demographic indices). Our sample of collected controls is similar to reference samples such as NHANES [40]) for biometrics, and for sexual orientation (identity) compared to previous studies (e.g., [41]), lending support to our sample being representative in this regard as well.

Overall, our control sample does not have more problems with QC than our case sample (which was collected and assessed face-to-face), however, there were differences for some individual QC aspects. Due to the rapid anonymization of the control sample, there were fewer opportunities to resolve sex typing discrepancies, and higher rates thereof. However, the controls had fewer subjects excluded as ancestry outliers on principle components analysis, perhaps due in large part to the more effective phenotypic ancestry screening (grandparents for controls versus parents for cases).

The collected controls have concordant continental ancestry as assessed by self-report of EA versus AA in several ways. The phlebotomist was able to verify EA versus

AA when collecting the blood. The more refined phenotypic ancestry information collected for the grandparents of the controls matched well with the self-report of EA versus AA as seen in Figure 1. When restricting the sample to be genotyped by several criteria, including incorporation of the phenotypic ancestry information described in the introduction, rather small percentages of the genotyped sample were lost due to the QC criterion of not being an outlier in the principle components analysis of ancestry. For our earlier candidate gene study [1], 3% (65/2,126) of the EA control samples were excluded for this reason, while for our later GWAS [2], we excluded 0.5% (14/2,817) of the EA and 0.5% (5/1,010) of the AA controls for this reason.

A final main area to consider is the possibility that our respondents were more frank or open to divulging potentially sensitive material (i.e., substance use, mental illness, etc.) due to the nature of an anonymous internet questionnaire – in fact, internet questionnaires have been shown to be even less obtrusive than hard-copy versions of the same questionnaire, an effect further accentuated with anonymity [70, 71]. A questionnaire is less personal even than a telephone survey and certainly a face-to-face interview, and it is conceivable that the lack of personal scrutiny and potential for embarrassment may counterbalance the lack of rapport inherent in the method, enabling collection of more information. However, the gold standard in psychiatric diagnostic material certainly continues to be a structured diagnostic interview, and a limitation of the current sample could perhaps be overcome in future studies with the inclusion of such an instrument on a subset of participants in a counterbalanced manner as a further and more definitive assessment of diagnostic validity in a general population derived control sample.

While we cannot generate sensitivity and specificity estimates without a comparison to a more definitive diagnosis in these controls, the elevated lifetime prevalence estimates we have obtained with the CIDI-SF compared to those previously found in the NCS with the CIDI suggest that we have more false positives than false negatives, consistent with previous research [44]. Researchers using these samples as controls would be minimally affected by this if they are restricting the controls they use to the ones without their disorder of interest (or closely related disorders) since such selected controls would still be "hypernormal", albeit fewer of them would be available for study (though over 1,800 controls do not meet criteria for any CIDI-SF diagnosis and deny all psychosis and mania screens). For researchers using these samples as a source of "cases" (e.g., those with CIDI-SF major depression) would probably either want to require a higher threshold for declaring caseness (such as a higher CIDI-SF score; e.g., see [72] for a more optimal cutoff suggestion for major depression) or else utilize a more quantitative approach (e.g., via generating factor scores instead of relying on a dichotomous approach). The latter is the approach we plan to pursue in future analyses.

We have demonstrated that internet-based recruitment and assessment can be done on a large scale, including blood sample collection, in a manner allowing anonymization and rapid sharing with the scientific community (see methods). No breaches of anonymity occurred during the course of the controls collection and afterwards, and the general use phrasing on the consent forms have enabled a much wider use of the data and biomaterials than would be the case otherwise.

SUPPLEMENTARY MATERIALS: WEB RESOURCES

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

Broad Institute, http://www.broad.mit.edu/

Composite International Diagnostic Interview Short Form (CIDI-SF) scoring memo, http://www.hcp.med.harvard.edu/ncs/ftpdir/cidisf_readme.pdf

dbGaP (database of Genotypes and Phenotypes), http://www.ncbi.nlm.nih.gov/gap

Examination Management Services Inc., http://www.emsinet.com/

Genetic Association Information Network, http://www.genome.gov/19518664

Knowledge Networks, http://www.knowledgenetworks.com/

National Institute of Mental Health (NIMH) Center for Collaborative Genetics Studies of Mental Disorders; http://www.nimhgenetics.org/

NIH GWAS sharing policy; multiple documents at http://grants.nih.gov/grants/gwas/

Rutgers University Cell and DNA Repository, http://www.rucdr.org/

Survey Sampling International, http://www.surveysampling.com/

SUPPLEMENTARY MATERIALS: ST (SUPPLEMENTARY TABLES)

ST1 – Supplementary Table 1. Study Completion Rate for Control Sample by Source

	N				%				
Sample Source / Ancestry	**Invited**	**Responded**	**Consented**	**Phlebotomy**	**Response of Invited**	**Consented of Responded**	**Phlebotomy of Consented**	**Consented of Invited**	**Phlebotomy of Invited**
KN / EA	15,485	10,962	4,780	3,364	0.708	0.436	0.704	0.309	0.217
KN / AA	3,391	2,349	844	527	0.693	0.359	0.624	0.249	0.155
KN Total	18,876	13,348	5,624	3,891	0.707	0.421	0.692	0.298	0.206
SSI / AA	36,923	2,511	2,352	770	0.068	0.937	0.327	0.064	0.021

KN = Knowledge Networks; SSI = Survey Sampling International; EA = European Ancestry; AA = African American.

ST2 – Supplementary Table 2. Number of Subjects Excluded by QC Filters for Autosomal SNP Analysis

	European Ancestry (EA)			African American (AA)			Entire Sample (EA & AA)		
Criterion	**Total**	**Cases**	**Controls**	**Total**	**Cases**	**Controls**	**Total**	**Cases**	**Controls**
Call Rate	119	46	73	36	23	13	155	69	86
Heterozygosity Proportion	168	89	79	19	15	4	187	104	83
Inconsistent Gender	31	14	17	18	12	6	49	26	23
Unexpected Duplicate	8	6	2	1	1	0	9	7	2
Unexpected Relatedness	23	6	17	0	0	0	23	6	17
Principal Component Outliers	28	14	14	18	13	5	46	27	19
Subjects with Kinship >0.1 with >50 subjects	18	13	5	3	2	1	21	15	6
Clinical Data Review	2	2	0	2	2	0	4	4	0
Total Subjects Excluded	**321**	**157**	**164**	**109**	**72**	**37**	**430**	**229**	**201**
Genotyped Count	5,655	2,838	2,817	2,368	1,358	1,010	8,023	4,196	3,827
Analyzed Count	5,334	2,681	2,653	2,259	1,286	973	7,593	3,967	3,626
Analyzed as Percentage of Genotyped	**94.3%**	**94.5%**	**94.2%**	**95.4%**	**94.7%**	**96.3%**	**94.6%**	**94.5%**	**94.7%**

This is an expansion and breakdown by case/control status for Table S6 of [2]; refer to the original publication for further details. Some subjects were excluded for more than one of the above QC criteria. In the total sample, the controls had a higher rate of cryptic relatedness (17/3,827) than the cases (6/4,196) did (χ^2=5.34, p=0.021), due entirely to the EA sample. Besides the unplanned QC exclusions above, we also had 61 EA and 31 AA expected duplicates (for SNP QC purposes; see Table S4 of [2]) that had one member of such pairs removed as part of the total subjects excluded above.

ST3 – Supplementary Table 3. Common Psychiatric Diagnoses (DSM-IV by CIDI-SF) in KN & SSI AA Control Sample

	AA Controls from KN			AA Controls from SSI		
Condition	**Males**	**Females**	**Total**	**Males**	**Females**	**Total**
Sample Size:	**178**	**351**	**529**	**336**	**436**	**772**
Substance Use Disorders						
Alcohol dependence	0.258	0.137	0.178	0.360	0.154	0.244
Drug dependence	0.213	0.137	0.163	0.327	0.135	0.219
Any substance dependence above	0.348	0.219	0.263	0.485	0.206	0.328
Substance-induced psychological sx (SIPS)	0.230	0.171	0.191	0.366	0.177	0.259
Major Depressive Episodes (MDE)						
All	0.303	0.481	0.422	0.464	0.555	0.516
Mean age at onset	26.5	24.8	25.4	27.8	26.7	27.2
Single episode	0.051	0.083	0.072	0.077	0.094	0.087
Recurrent	0.253	0.399	0.350	0.387	0.461	0.429
Without SIPS	0.185	0.362	0.302	0.232	0.417	0.337
Without substance dependence	0.163	0.316	0.265	0.190	0.404	0.311
Generalized Anxiety Disorder (GAD)						
All	0.135	0.219	0.191	0.253	0.342	0.303
Without MDE	0.017	0.023	0.021	0.036	0.039	0.038
Without substance dependence	0.045	0.131	0.102	0.080	0.243	0.172
Anxiety Disorders						
Specific Phobia	0.213	0.251	0.238	0.238	0.376	0.316
Social Phobia	0.135	0.154	0.147	0.173	0.179	0.176
Agoraphobia without Panic Attacks	0.039	0.060	0.053	0.042	0.073	0.060
Panic Attacks	0.129	0.191	0.170	0.158	0.193	0.177
Panic Attacks without Agoraphobia	0.101	0.142	0.129	0.110	0.110	0.110
Panic Attacks with Agoraphobia	0.028	0.048	0.042	0.048	0.083	0.067
Generalized Anxiety Disorder	0.135	0.219	0.191	0.253	0.342	0.303
Any anxiety disorder above (GAD, panic, phobia)	0.326	0.481	0.429	0.452	0.571	0.519
Any anxiety disorder above, without SIPS	0.191	0.353	0.299	0.199	0.438	0.334
Any anxiety disorder above w/o subst dep	0.163	0.330	0.274	0.143	0.427	0.303
Obsessive Compulsive Disorder (OCD)	0.118	0.125	0.123	0.152	0.138	0.144

AA control subjects' condition frequencies by collection (KN versus SSI). SIPS = Substance-induced psychological symptoms; sx = symptoms; Disorders are treated dichotomously (present/absent) in this table for the studied sample.

ST4 – Supplementary Table 4. Additional Phenotypic Information in KN & SSI AA Control Sample

	AA Controls from KN			AA Controls from SSI		
Condition	**Males**	**Females**	**Total**	**Males**	**Females**	**Total**
Sample Size:	**178**	**351**	**529**	**336**	**436**	**772**
Other disorders, traits, and conditions						
Nicotine dependence (FTND ≥4)	0.326	0.282	0.297	0.387	0.280	0.326
Mean FTND score, all	2.4	2.0	2.1	2.7	2.1	2.3
Mean Eysenck brief Neuroticism score (0-12)	3.0	4.2	3.8	3.9	4.9	4.5
Mean Eysenck brief Extraversion score (0-12)	7.7	7.6	7.7	7.0	7.0	7.0
Sexual identity: bisexual	0.039	0.023	0.028	0.015	0.021	0.018
Sexual identity: homosexual	0.034	0.006	0.015	0.068	0.009	0.035
Overweight or obese (current BMI ≥25)	0.685	0.738	0.720	0.762	0.748	0.754
Obese (current BMI ≥30)	0.404	0.501	0.469	0.405	0.509	0.464
Mean height (m)	1.79	1.65	1.70	1.79	1.65	1.71
Mean highest lifetime BMI (kg/m^2)	31.8	35.9	34.5	33.4	35.8	34.8
Psychosis & mania screens - endorsed						
Dx and/or Tx of SZ and/or SA	0.028	0.020	0.023	0.018	0.016	0.017
Dx and/or Tx of AH and/or delusions	0.045	0.046	0.045	0.042	0.034	0.038
Dx and/or Tx of BP and/or MDI	0.034	0.071	0.059	0.071	0.053	0.061
Dx and/or Tx of any of the above 3	0.079	0.091	0.087	0.098	0.073	0.084
Psychosis & mania screens – unsure or missing						
Dx and/or Tx of SZ and/or SA	0.039	0.023	0.028	0.036	0.037	0.036
Dx and/or Tx of AH and/or delusions	0.034	0.020	0.025	0.042	0.025	0.032
Dx and/or Tx of BP and/or MDI	0.028	0.031	0.030	0.036	0.053	0.045
Dx and/or Tx of any of the above 3	0.056	0.054	0.055	0.068	0.069	0.069
Comorbidity of disorders*						
No disorders and negative psychosis/mania screen	0.461	0.328	0.372	0.318	0.266	0.289
No disorders	0.472	0.339	0.384	0.336	0.287	0.308
Any disorder(s) (one or more)	0.528	0.661	0.616	0.664	0.713	0.692
One disorder	0.253	0.248	0.250	0.205	0.200	0.202
Two disorders	0.096	0.160	0.138	0.155	0.172	0.165
Three disorders	0.045	0.094	0.078	0.116	0.101	0.108
Four disorders	0.034	0.060	0.051	0.063	0.112	0.091
Five disorders	0.034	0.043	0.040	0.051	0.044	0.047
Six or more disorders	0.067	0.057	0.060	0.074	0.085	0.080
Proportion of disorders in controls with*						
One disorder	0.192	0.149	0.163	0.111	0.096	0.103
Two disorders	0.145	0.191	0.176	0.167	0.166	0.166
Three disorders	0.103	0.169	0.147	0.188	0.146	0.164
Four disorders	0.103	0.144	0.130	0.135	0.217	0.181
Five disorders	0.128	0.128	0.128	0.136	0.105	0.119
Six or more disorders	0.329	0.219	0.256	0.263	0.270	0.267

AA control subjects' condition frequencies by collection (KN versus SSI). FTND = Fagerström Test for Nicotine Dependence (0-12); BMI = body mass index; Dx = diagnosis; Tx = treatment; SZ = schizophrenia; SA = schizoaffective disorder; AH = auditory hallucinations; BP = bipolar disorder; MDI = manic-depression; disorders* = alcohol or drug dependence, MDE, specific phobia, social phobia, agoraphobia, panic attacks, GAD, or OCD (i.e., excludes nicotine dependence). Disorders are treated dichotomously (present/absent) in this table for the studied sample.

ST5 -- Supplementary Table 5. Psychiatric Diagnoses (probabilistic) of 3,364 EA and 529 AA (KN) Controls vs. ECA, NCS, NCS-R, & NESARC Data

	EA Controls			AA Controls (from KN)			ECA Data (~67% EA)			NCS Data (~75% EA)			NCS-R Data (~72% EA)			NESARC Data (~57% EA)		
CIDI-SF Disorder	**1,578 Males**	**1,786 Fem.**	**3,364 Total**	**178 Males**	**351 Fem.**	**529 Total**	**8,211 Males**	**10,971 Fem.**	**19,182 Total**	**3,847 Males**	**4,251 Fem.**	**8,098 Total**	**4,140 Males**	**5,142 Fem.**	**9,282 Total**	**18,518 Males**	**24,575 Fem.**	**43,09 Tota**
Alcohol dependence	0.360	0.218	0.284	0.302	0.160	0.208	0.238	0.046	0.138	0.201	0.082	0.141	------	------	0.054	------	------	------
Drug dependence	0.143	0.114	0.128	0.221	0.144	0.170	0.046	0.026	0.035	0.092	0.059	0.075	------	------	0.030	0.033	0.020	0.02
Major depressive episode	0.249	0.432	0.346	0.259	0.418	0.365	0.036	0.087	0.063	0.127	0.213	0.171	0.132	0.202	0.169	0.118	0.209	0.16
Generalized anxiety disorder	0.143	0.255	0.202	0.135	0.219	0.191	0.045	0.069	0.058	0.036	0.066	0.051	0.042	0.071	0.057	0.028	0.053	0.04
Specific phobia	0.097	0.172	0.137	0.157	0.184	0.175	0.078	0.145	0.113	0.067	0.157	0.113	0.089	0.158	0.125	0.062	0.133	0.09
Social phobia	0.107	0.158	0.134	0.127	0.144	0.139	0.025	0.029	0.027	0.111	0.155	0.133	0.111	0.130	0.121	0.042	0.057	0.05
Agoraphobia (without panic)	0.043	0.088	0.067	0.067	0.108	0.094	0.032	0.079	0.056	0.035	0.070	0.053	0.011	0.016	0.013	------	------	------
Panic attacks	0.105	0.203	0.157	0.130	0.195	0.173	0.010	0.021	0.016	0.020	0.050	0.035	0.031	0.062	0.047	0.033	0.067	0.05
Obsessive compulsive disorder	0.057	0.089	0.074	0.105	0.111	0.109	0.020	0.030	0.026	------	------	------	0.016	0.031	0.023	------	------	------

EA and AA (KN) control subjects' condition frequencies. Epidemiological Catchment Area (ECA) lifetime diagnostic (DSM-III) prevalence data (all ancestries combined, ~67% of which were EA) from the 1991 book [30]. National Comorbidity Survey (NCS) lifetime diagnostic (DSM-IIIR) prevalence data (all ancestries combined, ~75% of which were EA) from the 1994 report [31]. NCS Replication (NCS-R) lifetime diagnostic (DSM-IV) prevalence data (all ancestries combined, ~72% of which were EA) from the 2005 report [32, 33] and 2007 update (http://www.hcp.med.harvard.edu/ncs/publications.php). National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) lifetime diagnostic (DSM-IIIR) prevalence data (all ancestries combined, ~57% of which were EA) from wave I [34, 35]. Disorders are treated probabilistically (see text) in this table for the studied sample.

SUPPLEMENTARY MATERIALS: SF (SUPPLEMENTARY FIGURES)

Figure Legends

SF1 – Supplementary Figure 1

Results of the tabulation of the frequency with which a self-identified EA control endorsed a biological grandparent with the ancestral background indicated on the x-axis (one particular grandparent may have more than one ancestry endorsed). All 3,364 EA controls are indicated with the rightmost bar in each ancestral group, with bars (proceeding from leftmost bar rightwards) indicating those controls with just 1 known grandparent (146 = 4%), 2 known grandparents (281 = 8%), 3 known grandparents (336 = 10%), or 4 known grandparents (2,320 = 69%), with the remaining (281 = 8%) not knowing the ancestral background of any of their grandparents (perhaps due to adoption, among other reasons) or only listing a text response (which we did not attempt to convert to a number here).

SF2 – Supplementary Figure 2

Results of the tabulation of the frequency with which a self-identified AA control endorsed a biological grandparent with the ancestral background indicated on the x-axis (one particular grandparent may have more than one ancestry endorsed). All 1,301 AA controls are indicated with the rightmost bar in each ancestral group, with bars (proceeding from leftmost bar rightwards) indicating those controls with just 1 known grandparent (21 = 2%), 2 known grandparents (69 = 5%), 3 known grandparents (137 = 11%), or 4 known grandparents (1,041 = 80%), with the remaining (33 = 3%) not

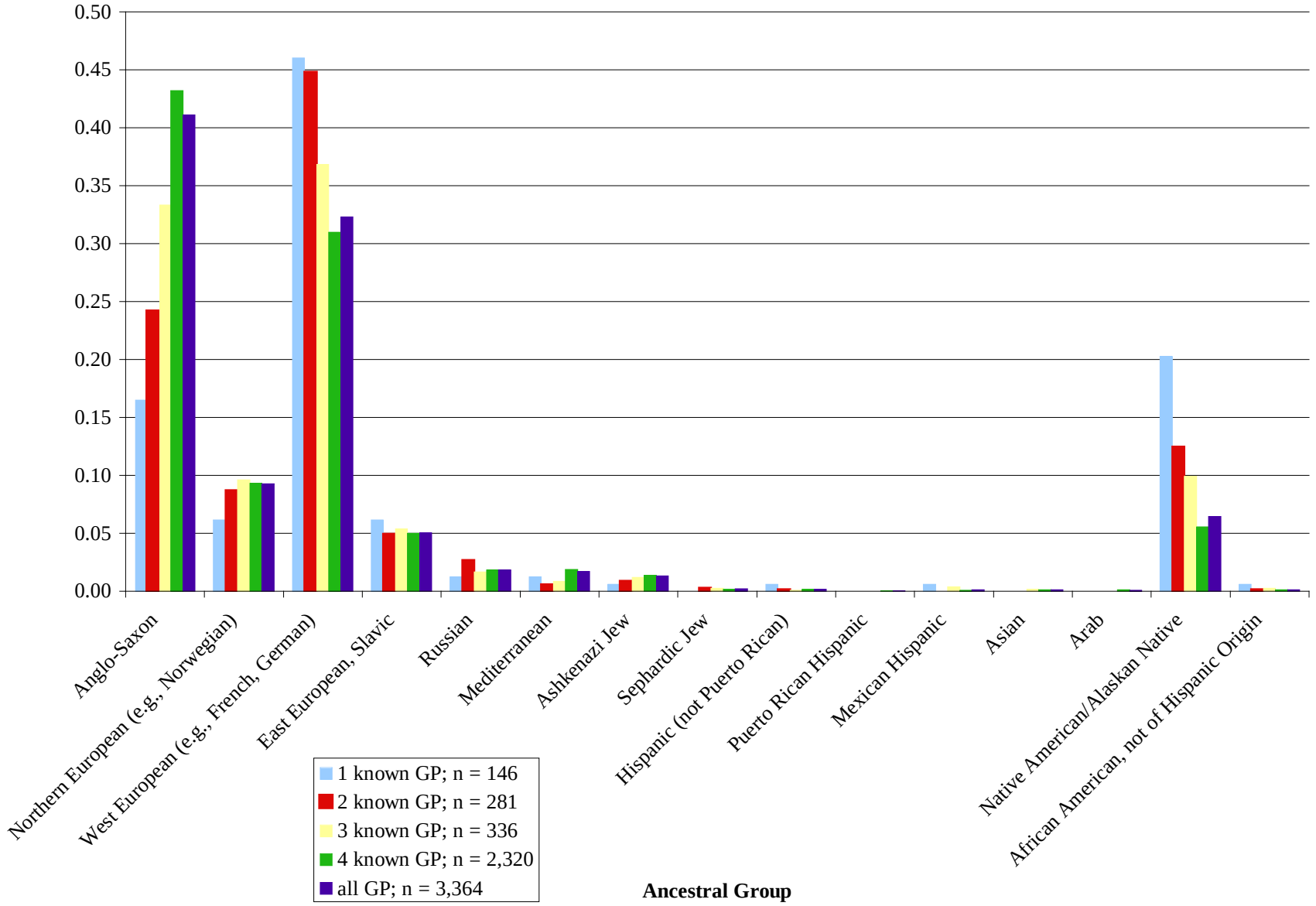
knowing the ancestral background of any of their grandparents (perhaps due to adoption, among other reasons) or only listing a text response (which we did not attempt to convert to a number here). When plotting the above information separately for KN versus SSI AA controls, the patterns are the same (data not shown).

SF3 – Supplementary Figure 3

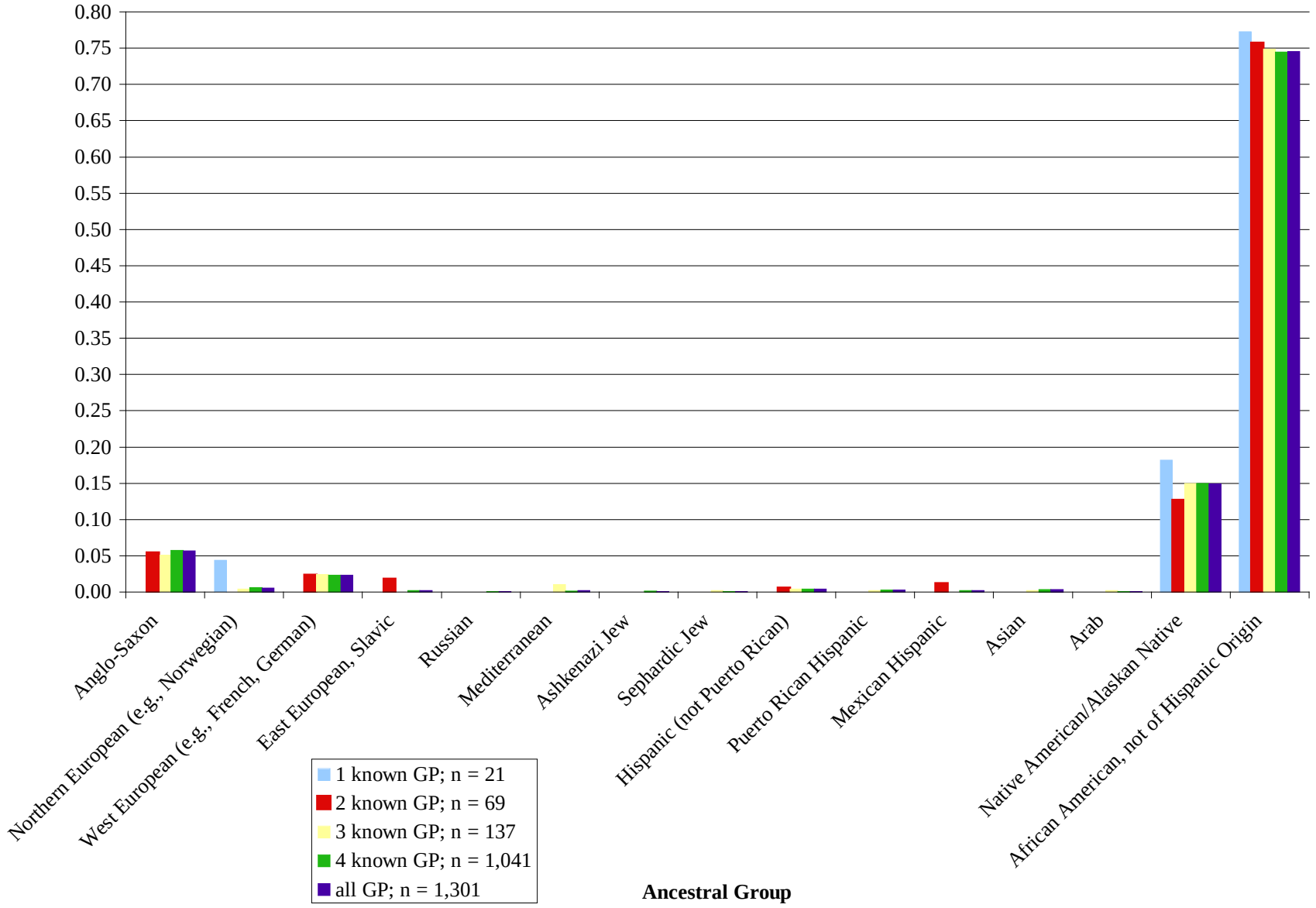
Results of the tabulation of the frequency with which self-identified EA control individuals (n = 3,364, grey bars) endorsed biological grandparents with the ancestral background indicated on the x-axis (one particular grandparent may have more than one ancestry endorsed) versus European ancestries reported in the 2000 U.S. Census Summary File 3 (n = 175,800,354, black bars). The census figures derive from the long census form that was received by about one-sixth of households, where 58% of respondents specified one ancestry and 22% specified two ancestries (www.census.gov/main/www/cen2000.html). These census ancestries were converted to the first six EA components (since the census ancestry did not query for Ashkenazi Jew separately under ancestry) assessed in the DIGS for comparison to our data; and non-EA census data was excluded. Both our EA control sample and the census show the same first two main EA components, namely Anglo-Saxon and West European, and in general show similar distributions of subcontinental EA components. Compared to the census, our EA control sample had a relative excess of Northern European and especially Anglo-Saxon, and a relative paucity of East European and especially Mediterranean ancestries. This is in part related to the relative excess of our EA sample drawn from the South (36.8% versus 26.6% via the CPS, Table 1) and the relative paucity of our EA sample

drawn from the Northeast (18.3% versus 23.9% via the CPS, Table 1) since via the U.S. Census, Mediterranean ancestries (e.g., Italian) are more common in the Northeast than the South (www.census.gov/prod/2004pubs/c2kbr-35.pdf).

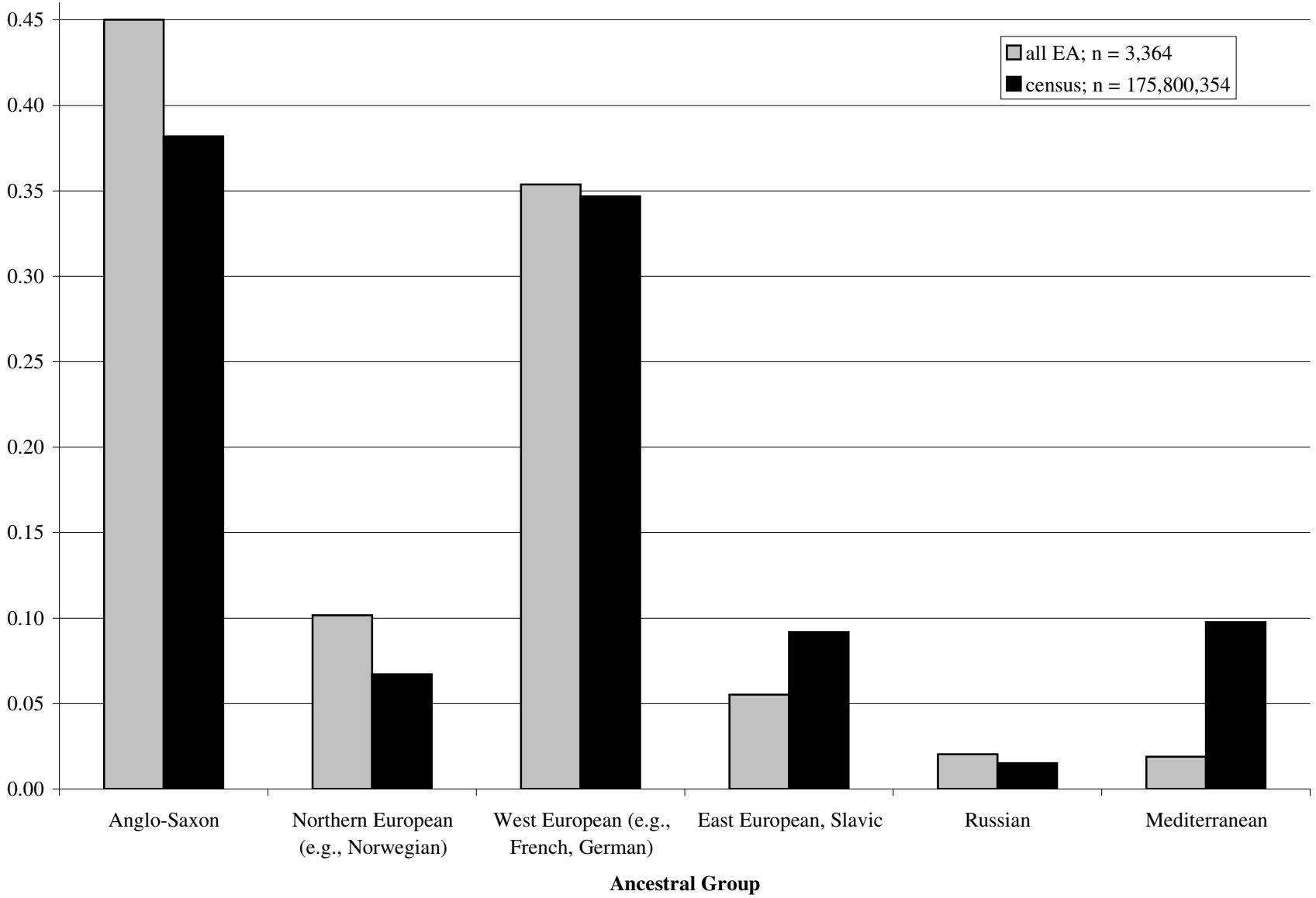
SF1 – Supplementary Figure 1



SF2 – Supplementary Figure 2



SF3 – Supplementary Figure 3



SUPPLEMENTARY MATERIALS: SR1 (SUPPLEMENTARY REFERENCES)

(Citation numbers are as in main paper, followed by SR1-only citations.)

1. Sanders, A.R., et al., *No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics.* Am J Psychiatry, 2008. **165**(4): p. 497-506.
2. Shi, J., et al., *Common variants on chromosome 6p22.1 are associated with schizophrenia.* Nature, 2009. **460**(7256): p. 753-757.
3. WTCCC, *Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-678.
4. McCarthy, M.I., et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges.* Nat Rev Genet, 2008. **9**(5): p. 356-369.
5. Schechter, D. and R. Lebovitch, *Normal controls are expensive to find: methods to improve cost-effectiveness of the screening evaluation.* Psychiatry Res, 2005. **136**(1): p. 69-78.
6. Adami, H., et al., *Use of telephone screens improves efficiency of healthy subject recruitment.* Psychiatry Res, 2002. **113**(3): p. 295-301.
7. Heeren, T., et al., *A comparison of results from an alcohol survey of a prerecruited Internet panel and the National Epidemiologic Survey on Alcohol and Related Conditions.* Alcohol Clin Exp Res, 2008. **32**(2): p. 222-229.
8. Morton, N.E. and A. Collins, *Tests and estimates of allelic association in complex inheritance.* Proc Natl Acad Sci U S A, 1998. **95**(19): p. 11389-11393.

9. Rothman, K.J., S. Greenlan, and T.L. Lash, *Case-Control Studies*, in *Modern Epidemiology*, K.J. Rothman, S. Greenlan, and T.L. Lash, Editors. 2008, Lippincott Williams & Wilkins: Philadelphia. p. 111-127.

10. Schwartz, S., E. Susser, and J.M. Gorman, *Choosing Controls in Biologic Psychiatry*, in *Psychiatric Epidemilogy: Searching for the Causes of Mental Disorders*, E. Susser, et al., Editors. 2006, Oxford University Press: New York. p. 247-259.

11. Couper, M., *Web surveys: a review of issues and approaches.* Public Opin Q, 2000. **64**(4): p. 464-494.

12. Krotki, K. and J.M. Dennis, *Probability-based survey research on the internet*, in *Conference of the International Statistical Institute*. 2001: Seoul, South Korea

13. Firmann, M., et al., *The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome.* BMC Cardiovasc Disord, 2008. **8**: p. 6.

14. Homer, N., et al., *Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays.* PLoS Genet, 2008. **4**(8): p. e1000167.

15. Wittchen, H.U., *Reliability and validity studies of the WHO--Composite International Diagnostic Interview (CIDI): a critical review.* J Psychiatr Res, 1994. **28**(1): p. 57-84.

16. Kessler, R.C., et al., *Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey.* Int J Methods Psychiatr Res, 1998. **7**(1): p. 33–55.

17. Gigantesco, A. and P. Morosini, *Development, reliability and factor analysis of a self-administered questionnaire which originates from the World Health Organization's Composite International Diagnostic Interview - Short Form (CIDI-SF) for assessing mental disorders.* Clin Pract Epidemol Ment Health, 2008. **4**: p. 8.

18. Kessler, R.C., et al., *The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF).* Int J Methods Psychiatr Res, 1998. **7**(4): p. 171-185.

19. Patten, S.B., et al., *Performance of the composite international diagnostic interview short form for major depression in a community sample.* Chronic Dis Can, 2000. **21**(2): p. 68-72.

20. Nelson, C.B., R.C. Kessler, and D. Mroczek (2001) *Scoring the World Health Organization's Composite International Diagnostic Interview Short Form (CIDI-SF; v1.0 NOV98 for all disorders except OCD which is from v1.1 MAR99)*

21. APA, *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV).* 1994, Washington, DC: American Psychiatric Association.

22. Nurnberger, J.I., Jr., et al., *Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative.* Arch Gen Psychiatry, 1994. **51**(11): p. 849-859.

23. Sklar, P., et al., *Whole-genome association study of bipolar disorder.* Mol Psychiatry, 2008. **13**(6): p. 558-569.

24. Scott, L.J., et al., *Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry.* Proc Natl Acad Sci U S A, 2009. **106**(18): p. 7501-7506.

25. *Current Population Survey*. November 2003, U. S. Census Bureau. .

26. Klimentidis, Y.C., G.F. Miller, and M.D. Shriver, *Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans.* Am J Phys Anthropol, 2009. **138**(4): p. 375-383.

27. Parra, E.J., et al., *Estimating African American admixture proportions by use of population-specific alleles.* Am J Hum Genet, 1998. **63**(6): p. 1839-1851.

28. Smith, M.W., et al., *A high-density admixture map for disease gene discovery in African Americans.* Am J Hum Genet, 2004. **74**(5): p. 1001-1013.

29. Cheng, C.Y., et al., *Admixture mapping of 15,280 African Americans identifies obesity susceptibility Loci on chromosomes 5 and X.* PLoS Genet, 2009. **5**(5): p. e1000490.

30. Robins, L.N. and D.A. Regier, *Psychiatric Disorders in America: The Epidemiological Catchment Area Study*. 1991, New York: The Free Press.

31. Kessler, R.C., et al., *Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey.* Arch Gen Psychiatry, 1994. **51**(1): p. 8-19.

32. Kessler, R.C., et al., *Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication.* Arch Gen Psychiatry, 2005. **62**(6): p. 593-602.

33. Kessler, R.C., et al., *The US National Comorbidity Survey Replication (NCS-R): design and field procedures.* International Journal of Methods in Psychiatric Research, 2004. **13**(2): p. 69-92.

34. Grant, B.F., et al., *Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions.* Arch Gen Psychiatry, 2004. **61**(8): p. 807-816.

35. Conway, K.P., et al., *Lifetime comorbidity of DSM-IV mood and anxiety disorders and specific drug use disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions.* J Clin Psychiatry, 2006. **67**(2): p. 247-257.

36. WHO, *World Health Organization. Composite International Diagnostic Interview, Version 1.0*. 1990, Geneva: The World Health Organization.

37. Aalto-Setala, T., et al., *Major depressive episode among young adults: CIDI-SF versus SCAN consensus diagnoses.* Psychol Med, 2002. **32**(7): p. 1309-1314.

38. Heatherton, T.F., et al., *The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire.* Br J Addict, 1991. **86**(9): p. 1119-1127.

39. Eysenck, S.B.G., H.J. Eysenck, and P. Barrett, *A revised version of the psychoticism scale.* Pers Individ Dif, 1985. **6**: p. 21-29.

40. Wang, Y., et al., *Will all Americans become overweight or obese? Estimating the progression and cost of the US obesity epidemic.* Obesity (Silver Spring), 2008. **16**(10): p. 2323-2330.

41. Laumann, E.O., et al., *Homosexuality*, in *The social organization of sexuality: Sexual practices in the United States*. 1994, The University of Chicago Press: Chicago. p. 283-321.

42. Halbreich, U., et al., *The normalcy of self-proclaimed "normal volunteers".* Am J Psychiatry, 1989. **146**(8): p. 1052-1055.

43. Bussey-Jones, J., et al., *Asking the right questions: views on genetic variation research among black and white research participants.* J Gen Intern Med, 2009. **24**(3): p. 299-304.

44. Orlando, M., et al., *Re-estimating the prevalence of psychiatric disorders in a nationally representative sample of persons receiving care for HIV: results from the HIV Cost and Services Utilization Study.* Int J Methods Psychiatr Res, 2002. **11**(2): p. 75-82.

45. Johnson, T.P., J.G. Hougland, and R.W. Moore, *Sex differences in reporting sensitive behavior: A comparison of interview methods.* Sex Roles, 1991. **24**(11/12): p. 669-679.

46. ACSF, *Analysis of sexual behaviour in France (ACSF). A comparison between two modes of investigation: telephone survey and face-to-face survey. ASCF principal investigators and their associates.* AIDS, 1992. **6**(3): p. 315-323.

47. Paulsen, A.S., et al., *Reliability of the telephone interview in diagnosing anxiety disorders.* Arch Gen Psychiatry, 1988. **45**(1): p. 62-63.

48. Slutske, W.S., et al., *Long-term reliability and validity of alcoholism diagnoses and symptoms in a large national telephone interview survey.* Alcohol Clin Exp Res, 1998. **22**(3): p. 553-558.

49. Fenig, S., et al., *Telephone vs face-to-face interviewing in a community psychiatric survey.* Am J Public Health, 1993. **83**(6): p. 896-898.

50. Newman, J.C., et al., *The differential effects of face-to-face and computer interview modes.* Am J Public Health, 2002. **92**(2): p. 294-297.

51. Monsees, G.M., R.M. Tamimi, and P. Kraft, *Genome-wide association scans for secondary traits using case-control samples.* Genet Epidemiol, 2009. **33**(8): p. 717-728.

52. Talati, A., A.J. Fyer, and M.M. Weissman, *A comparison between screened NIMH and clinically interviewed control samples on neuroticism and extraversion.* Mol Psychiatry, 2008. **13**(2): p. 122-130.

53. Gibbons, R.D., J.M. Davis, and D.R. Hedeker, *A comment on the selection of 'healthy controls' for psychiatric experiments.* Arch Gen Psychiatry, 1990. **47**(8): p. 785-786.

54. Rohde, P., P.M. Lewinsohn, and J.R. Seeley, *Comparability of telephone and face-to-face interviews in assessing axis I and II disorders.* Am J Psychiatry, 1997. **154**(11): p. 1593-1598.

55. Wells, K.B., et al., *Agreement between face-to-face and telephone-administered versions of the depression section of the NIMH Diagnostic Interview Schedule.* J Psychiatr Res, 1988. **22**(3): p. 207-220.

56. Fendrich, M., et al., *The utility of drug testing in epidemiological research: results from a general population survey.* Addiction, 2004. **99**(2): p. 197-208.

57. Compton, W.M., et al., *Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: results from the*

national epidemiologic survey on alcohol and related conditions. Arch Gen Psychiatry, 2007. **64**(5): p. 566-576.

58. SAMHSA, *Results from the National Survey on Drug Use and Health: National Findings*. 2003, Substance Abuse and Mental Health Services Administration, Office of Applied Studies: Rockville, MD.

59. Grucza, R.A., et al., *Discrepancies in estimates of prevalence and correlates of substance use and disorders between two national surveys.* Addiction, 2007. **102**(4): p. 623-629.

60. Kendler, K.S. and C.A. Prescott, *Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Use Disorders*. 2006, New York: The Guilford Press.

61. Wells, J.E. and L.J. Horwood, *How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports.* Psychol Med, 2004. **34**(6): p. 1001-1011.

62. Simpson, S.M., et al., *Racial disparities in diagnosis and treatment of depression: a literature review*. Psychiatr Q, 2007. **78**(1): p. 3-14.

63. Wakefield, J.C., et al., *Extending the bereavement exclusion for major depression to other losses: evidence from the National Comorbidity Survey.* Arch Gen Psychiatry, 2007. **64**(4): p. 433-440.

64. Pigott, T.A., *Anxiety disorders in women.* Psychiatr Clin North Am, 2003. **26**(3): p. 621-72, vi-vii.

65. Bergen, A.W. and N. Caporaso, *Cigarette smoking.* J Natl Cancer Inst, 1999. **91**(16): p. 1365-1375.

66. Breslau, N., et al., *Nicotine dependence in the United States: prevalence, trends, and smoking persistence.* Arch Gen Psychiatry, 2001. **58**(9): p. 810-816.

67. Bienvenu, O.J., et al., *Normal personality traits and comorbidity among phobic, panic and major depressive disorders.* Psychiatry Res, 2001. **102**(1): p. 73-85.

68. Jylha, P. and E. Isometsa, *The relationship of neuroticism and extraversion to symptoms of anxiety and depression in the general population.* Depress Anxiety, 2006. **23**(5): p. 281-289.

69. Keith, S.J., D.A. Regier, and D.S. Rae, *Schizophrenic disorders*, in *Psychiatric Disorders in America: The Epidemiological Catchment Area Study*, L.N. Robins and D.A. Regier, Editors. 1991, The Free Press: New York.

70. Joinson, A., *Social desirability, anonymity, and Internet-based questionnaires.* Behav Res Methods Instrum Comput, 1999. **31**(3): p. 433-438.

71. Rhodes, S.D., D.A. Bowie, and K.C. Hergenrather, *Collecting behavioural data using the world wide web: considerations for researchers.* J Epidemiol Community Health, 2003. **57**(1): p. 68-73.

72. Kendler, K.S., et al., *A Swedish national twin study of lifetime major depression.* Am J Psychiatry, 2006. **163**(1): p. 109-114.