Additive Genetic Variation in Schizophrenia Risk
Is Shared by Populations of African and European Descent

Teresa R. de Candia,1,2,* S. Hong Lee,3 Jian Yang,3,4 Brian L. Browning,5 Pablo V. Gejman,6 Douglas F. Levinson,7 Bryan J. Mowry,3,8 John K. Hewitt,1,2 Michael E. Goddard,9,10 Michael C. O’Donovan,11 Shaun M. Purcell,12 Danielle Posthuma,13,14,15 the International Schizophrenia Consortium,16 the Molecular Genetics of Schizophrenia Collaboration,16 Peter M. Visscher,3,4 Naomi R. Wray,5,17 and Matthew C. Keller1,2,17,*

To investigate the extent to which the proportion of schizophrenia’s additive genetic variation tagged by SNPs is shared by populations of European and African descent, we analyzed the largest combined African descent (AD [n = 2,142]) and European descent (ED [n = 4,990]) schizophrenia case-control genome-wide association study (GWAS) data set available, the Molecular Genetics of Schizophrenia (MGS) data set. We show how a method that uses genomic similarities at measured SNPs to estimate the additive genetic correlation (SNP correlation [SNP-rg]) between traits can be extended to estimate SNP-rg for the same trait between ethnicities. We estimated SNP-rg for schizophrenia between the MGS ED and MGS AD samples to be 0.66 (SE = 0.23), which is significantly different from 0 (p(SNP-rg – 0) = 0.0003), but not 1 (p(SNP-rg – 1) = 0.26). We re-estimated SNP-rg between an independent ED data set (n = 6,665) and the MGS AD sample to be 0.61 (SE = 0.21, p(SNP-rg – 0) = 0.0003, p(SNP-rg – 1) = 0.16). These results suggest that many schizophrenia risk alleles are shared across ethnic groups and predate African-European divergence.

Introduction

Schizophrenia is a severe mental disorder with a lifetime prevalence around 1% and a heritability of 0.7–0.8.1,2 Using a linear mixed model, we previously found that about a third of the genetic variation in liability to schizophrenia is captured by additive effects of common (minor allele frequency [MAF] > 0.01) SNPs in a large European case-control data set.3 This approach first derives genetic similarities at measured SNPs among classically unrelated individuals and then uses those similarities to estimate the amount of variability explained by all SNPs together. Because rare causal variants are not well predicted by common SNPs, this finding suggests that a substantial portion of the heritability of schizophrenia is due to additive effects of common causal variants,3 a conclusion consistent with extrapolations from results by a different methodology.4,4 However, it is unclear whether this result also applies to populations of African descent and, if so, whether the proportion of schizophrenia’s genetic variation tagged by SNPs is shared between populations of African descent (AD) and European descent (ED). Although rare variants (MAF < 0.01) are largely population specific,5,6 common variants are typically polymorphic across divergent ethnic groups;7 therefore, it is possible that a meaningful portion of schizophrenia’s genetic variation tagged by common SNPs could be shared between ED and AD individuals. However, different causal variants between ethnicities, or different linkage-disequilibrium (LD) patterns between SNPs and causal variants across ethnicities, could result in lower overlap in additive genetic variation tagged by SNPs. To date, ~96% of genome-wide association study (GWAS) participants have been ED individuals.3,9 Because SNP associations might differ between ethnicities, genetic studies using European-only samples could yield results with a strong Eurocentric bias.10

A traditional approach for understanding whether SNP associations are consistent across ethnicities is to assess the similarity of GWAS results among different ethnic groups11–14. However, for most complex traits, only a small minority of truly associated SNPs reach stringent genome-wide significance thresholds,15 and so overlap in

---

1Department of Psychology and Neuroscience, University of Colorado, Boulder, CO 80302, USA; 2Institute for Behavioral Genetics, University of Colorado, Boulder, CO 80302, USA; 3The University of Queensland, Queensland Brain Institute, Brisbane, QLD 4072, Australia; 4The University of Queensland, Diamantina Institute, Brisbane, QLD 4072, Australia; 5Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA 98195, USA; 6Department of Psychiatry and Behavioral Sciences, NorthShore University Health System and University of Chicago, Evanston, IL 60601, USA; 7Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA 94305, USA; 8Queensland Centre for Mental Health Research, Brisbane, QLD 4076, Australia; 9Department of Agriculture and Food Systems, University of Melbourne, Melbourne, VIC 3010, Australia; 10Biosciences Research Division, Department of Primary Industries, Melbourne, VIC 3001, Australia; 11MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff CF14 4XN, UK; 12Broad Institute, Cambridge, MA 02141, USA; 13Complex Trait Genetics, Department of Functional Genomics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University, Amsterdam 1081 HV, the Netherlands; 14Department of Clinical Genetics, VU University Medical Center, Amsterdam 1081 HV, the Netherlands; 15Department of Child and Adolescent Psychiatry, Erasmus University Rotterdam, Sophia Child Hospital, Rotterdam 3000 CB, the Netherlands; 16Full membership lists for the International Schizophrenia Consortium and the Molecular Genetics of Schizophrenia Collaboration can be found in the Supplemental Data

*These authors contributed equally to this work

Correspondence: teresa.decandia@colorado.edu (T.R.d.C.), matthew.c.keller@gmail.com (M.C.K.)
http://dx.doi.org/10.1016/j.ajhg.2013.07.007. ©2013 The Authors

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
significant associations is not necessarily expected, even if the effects of most SNPs are similar across ethnicities. In an attempt to gain insight into the overall consistency of schizophrenia SNP associations between ED and AD populations, a previous study used extremely liberal p value thresholds to define large sets of risk-score alleles in LD in an ED “discovery” sample and used these alleles to generate risk scores in independent “target” samples (the Molecular Genetics of Schizophrenia [MGS] ED and AD samples, as well as the smaller “O’Donovan” ED sample). The most significantly associated set of SNPs in the ED discovery sample did predict case-control status in the AD target sample (p = 0.008) but explained substantially less variance ($R^2 = 0.4\%$) than in the ED target samples (MGS: $R^2 = 3.2\%$, p = $2 \times 10^{-28}$; O’Donovan: $R^2 = 2.3\%$, p = $5 \times 10^{-11}$), which might appear to suggest only modest overlap in schizophrenia SNP associations between EDs and ADs.

A limitation of the genetic-risk-score approach is that it requires accurate estimation of effect sizes of individual SNPs in the discovery sample, and error variance of GWAS point estimates accumulates in prediction scores and thus causes estimates of overlap ($R^2$) to be biased downward as a function of sample size. Furthermore, the genetic-risk-score approach can underestimate overlap between ED and AD populations, in particular when discovery SNPs are pruned to be in LD in an ED discovery sample: because LD tends to be lower in AD populations, more SNPs are required for predicting causal variants in AD populations than in ED populations.

In the present study, we used a bivariate linear mixed-effects model implemented in GCTA to estimate the proportion of schizophrenia-risk variation that is tagged by the additive effects of SNPs (we refer to this estimate as the SNP heritability) in predominately AD individuals (African Americans). We also estimated the additive genetic correlation tagged by SNPs (the SNP correlation) in schizophrenia between ED and AD individuals. With this approach, the effects of SNPs are treated as statistically random and individual SNP effects are not estimated, nor is there a need to prune SNPs for LD or statistical significance. Rather, all SNPs are used simultaneously for estimating genome-wide similarities between pairs of individuals; these similarity estimates are in turn used for estimating variance and covariance parameters in a linear mixed-effects model. This method does not depend on reliable estimation of individual SNP effects and so should provide unbiased estimates of the SNP heritability and SNP correlation regardless of subject sample size (larger samples should decrease the SEs of estimates).

Material and Methods

GWAS Samples and Quality Control
The MGS and International Schizophrenia Consortium (ISC) case-control data sets have previously been described in detail. The study and use of these data sets were approved by institutional review boards at the University of Colorado. Because MGS subjects and just over half of ISC subjects were genotyped for 909,622 autosomal SNPs (on Affy 6 Platforms), whereas just under half of ISC subjects were genotyped for ~500K autosomal SNPs on earlier platforms (Affy 5 and Affy 500K), we used imputed ISC data to maximize the number of subjects included in the analysis. Therefore, we conducted the main analysis by using raw MGS and imputed ISC genotypes. We imputed ISC genotypes with HapMap CEU (Utah residents with ancestry from northern and western Europe from the CEPH collection) and TSI (Toscani in Italia) populations as reference panels. We imputed MGS ED genotypes in the same way to allow a comparison of results from raw and imputed data. We found no evidence that use of imputed data affected estimates of SNP heritabilities or SNP correlations; results were virtually identical across models that used imputed versus raw MGS ED data (Table S1, available online). We did not attempt to impute MGS AD genotypes because of the difficulty of imputing admixed AD samples. We calculated the MAF and imputation $R^2$ (ratio of observed to expected variance) for each SNP for quality-control purposes.

Each data set underwent stringent quality-control procedures separately. In particular, we dropped (1) individuals with average missing rates in raw SNP calls $> 0.02$; (2) individuals who were more than 5 SDs away from the cohort’s (within data set, within ethnicity) mean scores on the first two ethnic principal components (PCs; see “PC Analysis” below), the first of which differentiated ED and AD individuals; and (3) one individual from each pair of individuals with genome-wide similarities $> 0.05$. We dropped SNPs when (1) MAF $< 0.01$; (2) missing rates $> 0.02$ (raw data only); (3) missing rate differences between cases and controls $> 0.02$ (raw data only); (4) SNP frequency differences to CEU HapMap $> 0.15$ (because AD individuals tend to be admixed, the MGS AD data set did not undergo this step); (5) Hardy-Weinberg equilibrium test (controls only) $p < 10^{-6}$; and (6) imputation $R^2 < 0.6$ in all cohorts (imputed data only). Only SNPs in common across cohorts were kept for analysis, resulting in 477,664 SNPs. Final cohort sample sizes are summarized in Table 1.

### PC Analysis

PC analysis is used for reducing the dimensionality of data with correlated variables. When applied to genomic-similarity matrices computed from SNP data, the first several PCs typically capture ancestry differences between populations and later PCs can also capture technical batch effects. We ran PC analysis on subject pairwise genomic (identity by state [IBS]) similarities across SNPs from the combined MGS AD, MGS ED, and ISC ED samples, along with JPT, YRI, and CEU HapMap reference panels, by using a random subset of 30K SNPs that were pruned to be in linkage equilibrium and that were shared in common across samples. The use of HapMap reference panels “anchored” our PC results,

### Table 1. Cohort Sizes and Univariate SNP Heritabilities and SEs

<table>
<thead>
<tr>
<th>Sample</th>
<th>Cases</th>
<th>Controls</th>
<th>$h^2$ (SE)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGS AD</td>
<td>1,223</td>
<td>919</td>
<td>0.24 (0.09)</td>
</tr>
<tr>
<td>MGS ED</td>
<td>2,571</td>
<td>2,419</td>
<td>0.28 (0.03)</td>
</tr>
<tr>
<td>ISC ED</td>
<td>3,220</td>
<td>3,445</td>
<td>0.27 (0.02)</td>
</tr>
</tbody>
</table>

$h^2$ represents SNP heritability.

*ISC comprises eight imputation cohorts genotyped at separate sites.

The American Journal of Human Genetics 93, 463–470, September 5, 2013
such that the first PC (PC1) distinguished African and European ancestries and the second distinguished Asian ancestry (Figure S1). We used the PC results to remove ethnic outliers (see “GWAS Samples and Quality Control” above), for inclusion as covariates to control for population stratification (see “Linear Mixed Models for Estimating Case-Control Status” below), and to divide the MGS AD sample into subsamples on the basis of degree of European admixture. After quality control, ADs were divided into overlapping subsets of decreasing degrees of European admixture, removing individuals > 2, > 1, and > 0.5 SDs above the mean AD PC1 score (Figure S1).

**Linear Mixed Models for Estimating Case-Control Status**

For univariate heritability models, we estimated genomic similarities between pairs of individuals separately in the MGS AD sample, in the MGS ED sample, and in the ISC ED sample by using the method described by Yang et al. In particular, we calculated the genomic similarity for each pair of individuals by taking the sum of the products of SNP coefficients between those individuals and then scaling that sum by the expected SNP heterozygosity,

\[
\hat{a}_{ij} = 1 + \frac{1}{L} \sum_{l=1}^{L} \frac{(q_i - 2p_l) \cdot (q_j - 2p_l)}{2pq} \quad (i \neq j)
\]

\[
\hat{a}_i = 1 + \frac{1}{L} \sum_{l=1}^{L} \frac{x_i^2 - (1 + 2p_l)x_i + 2p_l^2}{2pq}
\]

where \(q_i = 0, 1, 2\) according to whether individual \(i\) has genotype \(bb\), \(Bb\), or \(BB\), respectively, at locus \(l\) (alleles are arbitrarily called \(b\) or \(B\)), \(p_l\) and \(q_l\) are the allele frequencies of \(B\) and \(b\), respectively, and \(2p_l\) is the mean of \(x_i\). These values are scaled to be both positive and negative; therefore, for clarity we use the term “similarity” rather than “relationship.”

For bivariate models, we re-estimated the genomic similarities in the combined MGS AD and MGS ED samples, in the combined MGS AD and ISC ED samples, and in the combined MGS ED and ISC ED samples. In the bivariate model, the observed disease status for traits 1 and 2 can be written as

\[
y_1 = X_1b_1 + Z_1g_1 + e_1\text{ for trait 1}
\]

\[
y_2 = X_2b_2 + Z_2g_2 + e_2\text{ for trait 2}
\]

where \(y\) is a vector of observations, \(b\) is a vector of fixed covariate effects, \(g\) is a vector of genetic effects tagged by SNPs, and \(e\) is a vector of residuals. \(X\) and \(Z\) are incidence matrices for \(b\) and \(g\), respectively. The variance covariance matrix \((V)\) is defined as

\[
V = \begin{bmatrix} Z_1AZ_1\sigma^{2}_{b1} + \sigma^2_{e1} & Z_1AZ_2\sigma_{b1g2} + \sigma^2_{eg1} \\ Z_2AZ_1\sigma_{g1b2} + \sigma^2_{eg1} & Z_2AZ_2\sigma^{2}_{g2} + \sigma^2_{e2} \end{bmatrix},
\]

where \(A\) is the genomic similarity relationship matrix based on SNP information across all groups, \(I\) is an identity matrix, and \(\sigma^2_{b1}, \sigma^2_{g1},\) and \(\sigma^2_{eg1}\) are genetic variance tagged by SNPs, residual variance, and covariance between \(g_1\) and \(g_2\), respectively. We treated schizophrenia in each sample as a separate “trait” and estimated SNP heritabilities \(h^2_{1}\) and \(h^2_{2}\) and the SNP liability covariance \((cov_{12})\) or SNP correlation \((SNP-r_{g})\) between ethnicities. For all models, we transformed genetic-variance estimates to a liability scale by assuming a disease prevalence of 1%, as described previously, and controlled for gender, 20 PCs, and site (Table 1) as covariates. Although there were large differences in average similarities of ED-AD pairs compared to within-ethnicity pairs, we controlled for these differences by including ancestry PCs as fixed effects, the first of which captured the African-European gradient (see Figure S1). We also included PC1 in bivariate models in order to control for any possible ascertainment or diagnostic differences that might exist between the most and least admixed individuals. For each model, PCs included as covariates were from an IBS matrix of subjects included in that model. Because MAFs can vary by ethnicity, for bivariate models, elements in the \(A\) matrix were standardized with expected allele frequencies that varied by data set.

In the current context, SNP heritability refers to the proportion of phenotypic variance in liability to schizophrenia that is explained by the additive effect of all SNPs together, and SNP correlation refers to the additive genetic correlation (explained by all SNPs together) between the liability of schizophrenia in ADs and the liability of schizophrenia EDs. Evidence of a SNP correlation between ADs and EDs occurs to the degree that AD cases are more genetically similar to ED cases than to ED controls, and vice versa, after main effects of ethnicity are controlled for. Because the two traits (schizophrenia in ADs and schizophrenia in EDs) are measured on different individuals, the only link between traits is in the genomic-similarity scores for each pair of individuals. Therefore, unlike traditional (e.g., twin) models that decompose phenotypic correlations into additive genetic, non-additive genetic, and environmental sources that can be mutually confounded, in the present model there is no such confounding, and SNP correlations are unlikely to be due to residual (e.g., environmental or nonadditive genetic) sources.

We used a permutation approach in which affection status was permuted within ethnicity in order to exclude the possibility of upward biases of SNP-heritability and SNP-correlation estimates (see Figure S2). In addition, because prevalence rates have been shown to vary across populations, we varied this parameter across populations and found that it did not substantially alter SNP heritabilities or SNP correlations (see Figure S3). That SNP correlations are unaffected by prevalence-rate differences is also evident from the transformation equation in Lee et al. Also, to understand whether some of the genetic overlap in schizophrenia between ED and AD samples might be due to the relatively high level of European admixture in ADs, estimated to be an average of ~15%, we ran additional bivariate models, each of which estimated SNP correlations between EDs and AD subsamples of higher-proportion African ancestry as judged from the first ancestry PC (see “PC Analysis” above).

We used the likelihood-ratio test statistic to compare the fit of full models in which SNP correlations were freely estimated to the fit of reduced models in which SNP correlations were fixed to either 0 or 1. The distribution of differences in log likelihoods between full and reduced models is approximately chi-square distributed with 1 degree of freedom.

### Binning SNPs by Rates of Recombination

We examined whether SNPs in regions of high recombination showed lower SNP correlations than SNPs in regions of low recombination. To do this, we used recombination rates estimated from CEU, YRI, and JPT+CHB HapMap populations on 3,303,900 SNPs and interpolated recombination rates for 476,704 SNPs in our data sets. For each SNP, an interpolated recombination rate was derived from a linear regression of recombination rates on base pair positions for its two closest HapMap neighbors. A
SNP-correlation estimate was then derived for a model that only included SNPs with above-median recombination rates and separately for a second model that only included SNPs with below-median recombination rates.

**Results**

We found that the univariate SNP-heritability estimate for schizophrenia in ADs ($h^2 = 0.24$, SE = 0.09) was slightly lower than the SNP-heritability estimates in EDs ($h^2 = 0.28$, SE = 0.03 in MGS EDs; $h^2 = 0.27$, SE = 0.02 in ISC EDs; Table 1), as might be expected given the lower average LD between SNPs in African populations. The schizophrenia-liability SNP correlation captured by autosomal SNPs between the sexes was high (SNP-rg = 0.41, SE = 0.03), these results are also consistent with those we reported the prediction that SNP-correlation estimates (Table S4).

Table 2. SNP-Heritability and SNP-Correlation Estimates from Bivariate Models

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 1 n</th>
<th>Sample 2 n</th>
<th>Sample 1 $h^2$ (SE)</th>
<th>Sample 2 $h^2$ (SE)</th>
<th>$\text{cov}_{12}$</th>
<th>SNP-rg (SE)</th>
<th>$p_{(\text{SNP-rg} = 0)}$</th>
<th>$p_{(\text{SNP-rg} = 1)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGS AD</td>
<td>MGS ED</td>
<td>2,142</td>
<td>4,990</td>
<td>0.20 (0.09)</td>
<td>0.33 (0.03)</td>
<td>0.17</td>
<td>0.66 (0.23)</td>
<td>0.0003</td>
<td>0.26</td>
</tr>
<tr>
<td>MGS AD</td>
<td>ISC ED</td>
<td>2,142</td>
<td>6,665</td>
<td>0.22 (0.09)</td>
<td>0.27 (0.02)</td>
<td>0.15</td>
<td>0.61 (0.21)</td>
<td>0.0003</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Within Ethnicity

| MGS ED   | ISC ED   | 4,990      | 6,665      | 0.26 (0.03)         | 0.27 (0.02)         | 0.22            | 0.83 (0.09)  | <0.0001         | 0.09            |

*Calculated by $\text{cov}_{12} = \text{SNP-rg}_1h_1h_2$.

Compared to SNPs with low MAF, SNPs with high MAF typically better tag high-MAF causal variants, which are more likely to predate the African-European divergence. We therefore predicted that the genetic correlation between EDs and ADs would be higher for high-MAF SNPs than for low-MAF SNPs. Consistent with this expectation, SNP-correlation estimates derived from genomic-relationship matrices restricted to SNPs with above-median MAFs (MAF > 0.26) were higher (SNP-rg = 0.79, SE = 0.28, $p_{(\text{SNP-rg} = 0)} < 0.0001$, $p_{(\text{SNP-rg} = 1)} = 0.52$ between MGS ADs and ISC EDs and SNP-rg = 0.67, SE = 0.27, $p_{(\text{SNP-rg} = 0)} = 0.0004$, $p_{(\text{SNP-rg} = 1)} = 0.37$ between MGS ADs and ISC EDs) than SNP-correlation estimates derived from genomic-relationship matrices restricted to SNPs with below-median MAFs (MAF < 0.26, SNP-rg = 0.34, SE = 0.20, $p_{(\text{SNP-rg} = 0)} = 0.0706$, $p_{(\text{SNP-rg} = 1)} = 0.0399$ between MGS ADs and MGS EDs and SNP-rg = 0.42, SE = 0.18, $p_{(\text{SNP-rg} = 0)} = 0.0126$, $p_{(\text{SNP-rg} = 1)} = 0.0321$ between MGS ADs and ISC EDs; Table 3). We observed that SNP heritabilities were lower for above-median-MAF ($h^2 \sim 0.13$) than for below-median-MAF ($h^2 \sim 0.23$) SNPs among MGS ADs, although these estimates were not significantly different from each other and are consistent with random fluctuations in estimated effects across models using different genomic-relationship matrices.

On the other hand, we found no strong evidence supporting the prediction that SNP-correlation estimates would be higher in low-recombination regions (SNP-rg = 0.60, SE = 0.27, $p_{(\text{SNP-rg} = 0)} = 0.0008$, $p_{(\text{SNP-rg} = 1)} = 0.28$ between MGS ADs and MGS EDs and SNP-rg = 0.70, SE = 0.25, $p_{(\text{SNP-rg} = 0)} = 0.0001$, $p_{(\text{SNP-rg} = 1)} = 0.34$ between MGS ADs and ISC EDs) than in high-recombination regions (SNP-rg = 0.64, SE = 0.23, $p_{(\text{SNP-rg} = 0)} = 0.0005$, $p_{(\text{SNP-rg} = 1)} = 0.23$ between MGS ADs and MGS EDs and SNP-rg = 0.41, SE = 0.20, $p_{(\text{SNP-rg} = 0)} = 0.019$, $p_{(\text{SNP-rg} = 1)} = 0.054$ between MGS ADs and ISC EDs; Table 3).
To examine whether some of the genetic overlap in schizophrenia between ED and AD samples might be due to the relatively high level of European admixture in ADs, estimated to be an average of ~15%, we rederived SNP-correlation estimates by using subsamples of ADs with increasingly lower proportions of European ancestry as judged from the first ancestry PC. We found modest but nonsignificant decreases in SNP correlation between ADs and EDs (Table 4). In particular, SNP-r_g was estimated to be 0.46 (se = 0.26, p (SNP-rg = 0) = 0.064, p (SNP-rg = 1) = 0.23) between the least admixed MGS ADs and MGS EDs and to be 0.54 (se = 0.24, p (SNP-rg = 0) = 0.006, p (SNP-rg = 1) = 0.19) between the least admixed MGS ADs and ISC EDs.

Discussion

Using the largest combined AD and ED schizophrenia case-control GWAS available, we have demonstrated an application that uses direct estimates of SNP correlations to quantify the amount of shared genetic variance tagged by SNPs between any two ethnically distinct populations for any trait. We found that common genetic liability to schizophrenia is largely shared across ED and primarily AD populations. Because SNPs included in GWASs are unlikely to tag the effects of rare causal variants and because rarer causal variants are increasingly likely to be population specific, these estimates of overlap are unlikely to apply to the portion of schizophrenia heritability caused by rare causal variants. Although we make no assertion that all estimated gene effects are exactly additive, these results are unlikely to be very biased by nonadditive genetic effects because, if they exist, they contribute very little to the similarity of distantly related individuals.

Even for comparisons of the same trait within an ethnicity, SNP correlations between different samples are typically less than 1 as a result of loss of real signal (e.g., greater ethnic, phenotypic, and environmental homogeneity within samples) and potential artifacts that can inflate the unshared portions of heritability (e.g., SNP calling or plate confounds that differ systematically between cases and controls). Artifacts are expected to deflate, not inflate, SNP correlations because random directional effects of artifacts are not expected to be consistent across data sets. Because MGS ADs, MGS EDs, and ISC EDs were all collected in different labs and genotyped on separate plates, the SNP correlation between MGS EDs and ISC EDs provides an upper limit on the potential SNP correlation between MGS ADs and MGS or ISC EDs. Given that SNP correlations between samples drawn from the same population are expected to be 1 in the absence of artifacts, we estimate that the SNP correlation might be as high as approximately 0.75 (0.61/0.83 or

\[ \text{Table 3. SNP-Heritability and SNP-Correlation Estimates from Bivariate Models for SNP Bins Based on MAF and Recombination} \]

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 1 n</th>
<th>Sample 2 n</th>
<th>Sample 1 h^2 (SE)</th>
<th>Sample 2 h^2 (SE)</th>
<th>cov\text{12}^a</th>
<th>SNP-r_g (SE)</th>
<th>( p(\text{SNP-rg = 0}) )</th>
<th>( p(\text{SNP-rg = 1}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-MAF SNP Bin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGS AD  MGS ED</td>
<td>2,142</td>
<td>4,990</td>
<td>0.13 (0.07)</td>
<td>0.27 (0.03)</td>
<td>0.15</td>
<td>0.79 (0.28)</td>
<td>&lt;0.0001</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>MGS AD  ISC ED</td>
<td>2,142</td>
<td>6,665</td>
<td>0.12 (0.07)</td>
<td>0.22 (0.02)</td>
<td>0.11</td>
<td>0.67 (0.27)</td>
<td>0.0004</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>MGS ED  ISC ED</td>
<td>4,990</td>
<td>6,665</td>
<td>0.22 (0.03)</td>
<td>0.21 (0.02)</td>
<td>0.18</td>
<td>0.82 (0.09)</td>
<td>&lt;0.0001</td>
<td>0.0717</td>
<td></td>
</tr>
<tr>
<td>Low-MAF SNP Bin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGS AD  MGS ED</td>
<td>2,142</td>
<td>4,990</td>
<td>0.21 (0.09)</td>
<td>0.25 (0.02)</td>
<td>0.08</td>
<td>0.34 (0.20)</td>
<td>0.0706</td>
<td>0.0399</td>
<td></td>
</tr>
<tr>
<td>MGS AD  ISC ED</td>
<td>2,142</td>
<td>6,665</td>
<td>0.24 (0.09)</td>
<td>0.22 (0.02)</td>
<td>0.10</td>
<td>0.42 (0.18)</td>
<td>0.0126</td>
<td>0.0321</td>
<td></td>
</tr>
<tr>
<td>MGS ED  ISC ED</td>
<td>4,990</td>
<td>6,665</td>
<td>0.19 (0.03)</td>
<td>0.23 (0.02)</td>
<td>0.17</td>
<td>0.83 (0.12)</td>
<td>&lt;0.0001</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>High-Recombination SNP Bin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGS AD  MGS ED</td>
<td>2,142</td>
<td>4,990</td>
<td>0.20 (0.08)</td>
<td>0.28 (0.03)</td>
<td>0.15</td>
<td>0.64 (0.23)</td>
<td>0.0005</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>MGS AD  ISC ED</td>
<td>2,142</td>
<td>6,665</td>
<td>0.20 (0.08)</td>
<td>0.22 (0.02)</td>
<td>0.09</td>
<td>0.41 (0.20)</td>
<td>0.0193</td>
<td>0.0541</td>
<td></td>
</tr>
<tr>
<td>MGS ED  ISC ED</td>
<td>4,990</td>
<td>6,665</td>
<td>0.21 (0.03)</td>
<td>0.22 (0.02)</td>
<td>0.17</td>
<td>0.80 (0.11)</td>
<td>&lt;0.0001</td>
<td>0.0836</td>
<td></td>
</tr>
<tr>
<td>Low-Recombination SNP Bin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGS AD  MGS ED</td>
<td>2,142</td>
<td>4,990</td>
<td>0.13 (0.07)</td>
<td>0.23 (0.03)</td>
<td>0.10</td>
<td>0.60 (0.27)</td>
<td>0.0076</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>MGS AD  ISC ED</td>
<td>2,142</td>
<td>6,665</td>
<td>0.15 (0.07)</td>
<td>0.21 (0.02)</td>
<td>0.12</td>
<td>0.70 (0.25)</td>
<td>0.0001</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>MGS ED  ISC ED</td>
<td>4,990</td>
<td>6,665</td>
<td>0.19 (0.03)</td>
<td>0.20 (0.02)</td>
<td>0.17</td>
<td>0.86 (0.10)</td>
<td>&lt;0.0001</td>
<td>0.19</td>
<td></td>
</tr>
</tbody>
</table>

Estimates above are from independent analyses in which the following SNP bins were included separately for each model: 238,810 SNPs with high average MAF across ADs and EDs, 238,854 SNPs with low average MAF across ADs and EDs, 238,352 SNPs in high-recombination regions, and 238,352 SNPs in low-recombination regions. Recombination rates were estimated from CEU, YRI, and JPT+CHB HapMap populations, and high- and low-MAF and high- and low-recombination bins were formed with median splits of SNP MAFs and recombination rates, respectively. The Pearson correlation between SNP MAF and recombination rate was 0.006 (\( p = 0.0001 \)).

\( ^{a}\text{Calculated by cov_{12} = SNP-r_{gh}h_1h_2}. \)
Our findings suggest that many schizophrenia causal variants are ancient and predate European-African divergence. Because it is vanishingly unlikely that different causal variants between ethnicities would systematically be in LD with the same SNPs, these findings suggest that many schizophrenia causal variants are ancient and predate European-African divergence. However, we did not find strong evidence that SNP correlations were lower for high-recombination regions than for low-recombination regions, which might be expected if differing LD patterns reduce the SNP correlation between EDs and ADs. Nonetheless, our results do not allow us to reliably quantify the relative importance of these two alternatives.

Two limitations to our analysis are worth noting. First, the AD sample size in particular was small, leading to large SEs around estimates of both SNP heritabilities and SNP correlations between ethnicities. This means that we cannot place great confidence in the specific estimates of SNP correlation, and we reiterate that these parameters reflect the allelic spectrum captured by common SNPs and so might not generalize to the covariance attributable to variants in low LD with common SNPs. Nevertheless, our results suggest that there is substantial overlap of variants tagged by common SNPs between ethnic groups (with 95% confidence that SNP-r_{gh} has a lower bound of 0.20) and allow us to rule out the possibility of no overlap with even greater confidence. Second, no AD replication GWAS data set of sufficient size exists for this or any other psychiatric disorder, highlighting the importance of further large-scale genome-wide data collection in non-ED samples.

In summary, we have demonstrated how a random-effects modeling approach fitting all SNPs simultaneously can elucidate the degree to which additive genetic variation tagged by SNPs is shared between ethnically divergent populations. We estimate that about half to three-quarters of such additive genetic variation underlying the risk of schizophrenia is shared between AD and ED populations, suggesting that schizophrenia GWAS discoveries from European samples are likely to be relevant to AD populations and that meta-analyses of schizophrenia can improve power by including results across both ED and AD samples. These conclusions should apply even more strongly to other ethnic groups given their more recent divergence times from the European lineage. Because it is vanishingly unlikely that different causal variants between ethnicities would systematically be in LD with the same SNPs, these findings suggest that many schizophrenia causal variants are ancient and predate European-African divergence.

### Supplemental Data

Supplemental Data include Supplemental Acknowledgments, three figures, four tables, and full membership lists for the International Schizophrenia Consortium and the Molecular Genetics of Schizophrenia Collaboration and can be found with this article online at http://www.cell.com/AJHG.

### Acknowledgments

The authors acknowledge funding from the National Institute of Mental Health (grants K01MH085812 and R01MH100141 to M.C.K., 5T32MH016880-30 to J.K.H., and R01HG004960 and R01HG005701 to B.L.B.), the Australian National Health and Medical Research Council (grants 389892, 442915, 496688, 513010, 603170, and 606646), and the National Human Genome Research Institute (grant U01HG004982).

---

### Table 4. SNP-Heritability and SNP-Correlation Estimates from Bivariate Models after Exclusion of ADs of Increasing European Admixture

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 1 n</th>
<th>Sample 2 n</th>
<th>Sample 1 h² (SE)</th>
<th>Sample 2 h² (SE)</th>
<th>cov_{12}</th>
<th>SNP-r_{gh} (SE)</th>
<th>( p(SNP-r_{gh} = 0) )</th>
<th>( p(SNP-r_{gh} = 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGS AD (all)</td>
<td>MGS ED</td>
<td>2,142</td>
<td>4,990</td>
<td>0.20 (0.09)</td>
<td>0.33 (0.03)</td>
<td>0.17</td>
<td>0.66 (0.23)</td>
<td>0.0003</td>
<td>0.26</td>
</tr>
<tr>
<td>MGS AD (z &lt; 2)</td>
<td>MGS ED</td>
<td>2,044</td>
<td>4,990</td>
<td>0.18 (0.09)</td>
<td>0.33 (0.03)</td>
<td>0.18</td>
<td>0.73 (0.27)</td>
<td>0.0003</td>
<td>0.41</td>
</tr>
<tr>
<td>MGS AD (z &lt; 1)</td>
<td>ISC ED</td>
<td>1,792</td>
<td>6,665</td>
<td>0.22 (0.10)</td>
<td>0.30 (0.03)</td>
<td>0.12</td>
<td>0.46 (0.26)</td>
<td>0.0644</td>
<td>0.23</td>
</tr>
<tr>
<td>MGS AD (z &lt; 0.5)</td>
<td>ISC ED</td>
<td>1,557</td>
<td>6,665</td>
<td>0.23 (0.12)</td>
<td>0.28 (0.02)</td>
<td>0.14</td>
<td>0.54 (0.24)</td>
<td>0.0061</td>
<td>0.19</td>
</tr>
</tbody>
</table>

*Calculated by \( cov_{12} = \frac{\text{SNP-r}_{gh} h_1 h_2}{\sqrt{h_1^2 h_2^2}} \).
Received: March 21, 2013
Revised: June 3, 2013
Accepted: July 2, 2013
Published: August 15, 2013

References


