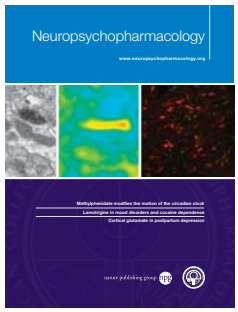

Genome-Wide Association Study of Loneliness Demonstrates a Role for Common Variation

Jianjun Gao, Lea K Davis, Amy B Hart, Sandra Sanchez-Roige, Lide Han, John T Cacioppo, Abraham A Palmer

Cite this article as: Jianjun Gao, Lea K Davis, Amy B Hart, Sandra Sanchez-Roige, Lide Han, John T Cacioppo, Abraham A Palmer, Genome-Wide Association Study of Loneliness Demonstrates a Role for Common Variation, *Neuropsychopharmacology* accepted article preview 15 September 2016; doi: 10.1038/npp.2016.197.

Received 20 April 2016; revised 29 August 2016; accepted 2 September 2016; Accepted article preview online 15 September 2016

TITLE PAGE

Genome-wide Association Study of Loneliness Demonstrates a Role for Common Variation

Running Title: Genetic architecture of loneliness

Authors: Jianjun Gao[1,2], Lea K. Davis[3], Amy B. Hart [4], Sandra Sanchez-Roige[2], Lide Han[1], John T. Cacioppo[5], Abraham A. Palmer[1,2,6*]

Affiliations:

1. Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
2. Department of Psychiatry, University of California San Diego, La Jolla, CA 92093, USA
3. Vanderbilt Genetics Institute, Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN 37232
4. Janssen R&D, Spring House, PA 19147, USA
5. Department of Psychology, University of Chicago, Chicago, IL 60637, USA
6. Institute for Genomic Medicine, University of California San Diego, La Jolla, CA 92093, USA

*** Corresponding Author**: Abraham A. Palmer; Department of Psychiatry, University of California San Diego, La Jolla, CA 92093, USA. Voice 858 534 2093. Fax 858-534-7653. aap@ucsd.edu

Number of words in abstract: 250

Number of figures: 2

Number of tables: 4

Supplementary information: Supplemental Tables S1-S10, Supplemental Figures S1-S6 and Supplementary Material (Excel Workbook)

ABSTRACT

Loneliness is a complex biological trait that has been associated with numerous negative health outcomes. The measurement and environmental determinants of loneliness are well understood, but its genetic basis is not. Previous studies have estimated the heritability of loneliness between 37%-55% using twins and family-based approaches, and have explored the role of specific candidate genes. We used genotypic and phenotypic data from 10,760 individuals aged 50 and over that were collected by the Health and Retirement Study (HRS) to perform the first genome-wide association study of loneliness. No associations reached genome-wide significance ($p>5x10^{-8}$). Furthermore, none of the previously published associations between variants within candidate genes (*BDNF*, *OXTR*, *RORA*, *GRM8*, *CHRNA4*, *IL-1A*, *CRHR1*, *MTHFR*, *DRD2*, *APOE*) and loneliness were replicated ($p>0.05$), despite our much larger sample size. We estimated the chip heritability of loneliness and examined co-heritability between loneliness and several personality and psychiatric traits. Our estimates of chip heritability (14-27%) support a role for common genetic variation. We identified strong genetic correlations between loneliness, neuroticism and a scale of 'depressive symptoms'. We also identified weaker evidence for co-heritability with extraversion, schizophrenia, bipolar disorder and major depressive disorder. We conclude that loneliness, as defined in this study, is a modestly heritable trait that has a highly polygenic genetic architecture. The co-heritability between loneliness and neuroticism may reflect the role of negative affectivity, which is common to both traits. Our results also reflect the value of studies that probe the common genetic basis of salutary social bonds and clinically defined psychiatric disorders.

INTRODUCTION

Humans are fundamentally social animals who form bonds with others for mutual aid and protection. For social species, the perception of being socially isolated even when in the presence of others signals danger and evokes a dysphoric state termed loneliness in humans (Cacioppo *et al*, 2014; Cacioppo *et al*, 2015a). A variety of biological mechanisms have evolved that capitalize on aversive signals to motivate people to act in ways that are essential for reproduction and survival. Just as physical pain is an aversive signal that alerts us of potential tissue damage and motivates us to take care of our physical body, loneliness – triggered by a discrepancy between an individual's preferred and actual social relations – is part of a biological warning system that has evolved to alert us of threats or damage to our social body.

A substantial literature now shows that loneliness is a major risk factor for adverse physical (Holt-Lunstad *et al*, 2015) and mental (Cacioppo *et al*, 2015c) health outcomes. A recent meta-analysis of 70 independent prospective studies involving more than 3 million people who were followed for an average of 7 years shows that loneliness increases the odds of mortality by 26% even after controlling for objective social isolation and other potentially confounding factors (Holt-Lunstad *et al*, 2015). For instance, using longitudinal data from the Health and Retirement Study (HRS), Luo *et al* (Luo *et al*, 2012) found that loneliness in 2002 predicted mortality over the subsequent 6 years even after controlling for demographic factors, health behaviors, and objective social isolation.

Investigations have found loneliness to be stable over years (e.g., Boomsma et al., 2005) and to differ from other personality factors such as extraversion, neuroticism, depressive symptomatology shyness, and anxiety (e.g., Cacioppo et al., 2006; 2010). Studies designed to identify the *mechanisms* underlying the association between loneliness and mortality have found that loneliness is associated with increased hypothalamic pituitary adrenocortical (HPA) activity (Adam *et al*, 2006; Cacioppo *et al*, 2006; Doane and Adam, 2010; Glaser *et al*, 1985; Kiecolt-Glaser *et al*, 1984; Steptoe *et al*, 2004), altered gene expression indicative of decreased inflammatory control and increased glucocorticoid insensitivity (Cole *et al*, 2011; Cole *et al*, 2007), increased inflammation, elevated vascular resistance and blood pressure (Hackett *et al*, 2012; Hawkley *et al*, 2006; Hawkley *et al*, 2010b; Jaremka *et al*, 2013), higher rates of metabolic syndrome (Whisman, 2010), diminished immunity (Dixon *et al*, 2006; Glaser *et al*, 2005; Kiecolt-Glaser *et al*, 1984; Pressman *et al*, 2005;

Straits-Tröster *et al*, 1994), increased risk for age-related cognitive decline and dementia (Wilson *et al*, 2007), and increased sleep fragmentation (Cacioppo *et al*, 2002; Hawkley *et al*, 2010a; Jacobs *et al*, 2006; Kurina *et al*, 2011). Cross-lagged panel analyses have also shown that loneliness has also been associated with changes in psychological states that can contribute to morbidity and mortality, including increased depressive symptomatology (Booth, 2000; Cacioppo *et al*, 2010; Cacioppo *et al*, 2006; VanderWeele *et al*, 2011), lower subjective well-being (Kong and You, 2013; VanderWeele *et al*, 2012), heightened vigilance for social threats (Cacioppo *et al*, 2015b), and decreased executive functioning (Baumeister and DeWall, 2005; Cacioppo *et al*, 2000; Hawkley *et al*, 2009).

The heritability of loneliness has been documented in twin and other studies using both children (Bartels *et al*, 2008; McGuire and Clifford, 2000a) and adults (Boomsma *et al*, 2007; Boomsma *et al*, 2006; Boomsma *et al*, 2005). For instance, in an early longitudinal study of 8,387 young adult and adult Dutch twins who participated in longitudinal surveys, Boomsma et al. analyzed variations in loneliness with genetic structural equation models (Boomsma *et al*, 2005). The estimate of genetic contributions to variation in loneliness in adults was 48%, similar to the heritability estimates reported by McGuire and Clifford (2000) in their study of children. Boomsma et al. (2005) found no evidence for sex or age differences in heritability. Subsequent twin studies have yielded heritability estimates ranging from 37% to 55% (Boomsma *et al*, 2006; Boomsma *et al*, 2005; Distel *et al*, 2010; McGuire and Clifford, 2000b; Waaktaar and Torgersen, 2012). Candidate gene studies for loneliness have concentrated primarily on systems related to monoamine neurotransmitters (e.g., dopamine, serotonin) and other signaling pathways associated with human attachment (e.g., oxytocin) (Goossens *et al*, 2015). Typical of candidate gene studies, they used modest sample sizes and therefore implicitly assumed relatively large effect sizes for the alleles being studied, a scenario that is inconsistent with the results of genome-wide association studies (GWAS) for numerous disease and non-disease traits (Hart *et al*, 2013).

In this study we have performed the first GWAS for loneliness. The UCLA loneliness scale is the most commonly used measure in the literature and has very good psychometric properties, including internal reliability, temporal stability, discriminant validity, convergent validity, construct validity, and predictive validity (Cacioppo *et al*, 2006; Russell, 1995; Russell, *et al*, 1980). Importantly, the term “loneliness” does not appear

in this scale because respondents, especially males, have been found to be reluctant to report feeling lonely (Russell et al., 1980). Hence, the measurement of this phenotype is not dependent on the respondents' ability or willingness to self-report being lonely. Since 2002, the HRS has included a 3-item version of the UCLA loneliness scale, which has also been shown to have very good psychometric properties, including internal reliability, concurrent validity (e.g., r = 0.88 with the full 20-item UCLA loneliness scale;(Hughes *et al*, 2004)), convergent and discriminant validity (Hughes et al., 2004), and predictive validity (e.g., predicts mortality in the HRS sample over a six year period; (Hughes *et al*, 2004; Luo *et al*, 2012)). We used a genomic restricted maximum likelihood (GREML) method implemented in the Genetic Complex Trait Analysis (GCTA) software (Yang *et al*, 2011) to examine chip heritability that is specifically due to the additive effect of genotyped (or imputed) common variants. Loneliness and objective social isolation are often weakly correlated (Coyle & Dugan, 2012; Holt-Lunstad et al., 2015), although loneliness does tend to be lower in individuals who are married than those who are unmarried (e.g., Hawkley *et al*, 2008). Analyses were therefore performed including marital status as a covariate. We also determined whether previously reported candidate gene associations could be replicated in the HRS. Finally, using polygenic risk scoring and estimates of genetic correlation, we were able to begin to probe possible shared genetic influences between personality traits and psychiatric diagnoses and loneliness**.**

MATERIALS AND METHODS

Subjects

The University of Michigan Health and Retirement Study (HRS) is a longitudinal study that began in 1992 and includes more than 26,000 Americans who are over the age of 50 (Health_and_Retirement_Study, 2012). Our study included genotype data (both directly genotyped and imputed) obtained from dbGaP (accession number phs000428.v1.p1) on a total of 12,454 subjects from HRS that were genotyped by the Center for Inherited Disease Research (CIDR). Permission to use the HRS dataset was obtained through application to dbGaP by J.T.C. Phenotypic data was collected by the HRS on subjective experiences of loneliness during the 2006 and 2008 data collection waves.

Loneliness phenotype

Loneliness was assessed using the 3-Item Leave Behind Questionnaire (LBQ) as part of a larger written questionnaire administered as part of the HRS (Hughes *et al*, 2004). Respondents were asked three questions, “How often do you feel that you lack companionship?”, “How often do you feel left out?” and “How often do you feel isolated from others?” Possible answers were “hardly ever, or never” (scored as 1); “some of the time” (scored as 2); or “often” (scored as 3). Thus, higher scores represent greater self-reported loneliness (Supplementary Figure S1). The total score on this 3-item loneliness scale has been shown to be highly correlated (r =0.88) with the total score on the UCLA loneliness scale. Only participants that responded to all the 3 questions were included in our study. Pairwise correlation coefficients between questions were obtained using Spearman Correlation Statistics in SAS version 9.3 (SAS Institute Inc, Cary, NC).

We derived three phenotypes from the loneliness scale for subsequent genetic analyses: 1) “linear” - a continuous phenotype obtained by summing the scores from all three questions, thus yielding a score between 3 (least lonely) and 9 (most lonely); 2) “multivariate” – a single score for each question ranging from 1 (least lonely) to 3 (most lonely); and 3) “case: control” – a dichotomous score in which participants who answered 1 on all three items were considered controls (totally loneliness score = 3), and individuals with a loneliness score of ≥ 6 were considered cases (participants with scores of 4 or 5 were treated as missing). There were nine subjects who answered the loneliness questions twice (first in 2006 and again in 2008); we used an average of their two scores. Frequency distributions by ancestry for the linear trait and case: control labels are shown in Supplementary Figure S2.

We considered sex, age (continuous) and marital status (categorical) as covariates in our analyses. Marital status was ascertained such that it consisted of 6 levels (married, annulled, separated, divorced, widowed and never married). For the genetic analyses, we summarized these 6 levels into a binary variable (married or unmarried). Subjects for which any of these three covariates were missing were excluded from all of our analyses.

SNP Genotyping and Quality Control

Genotyping of HRS subjects was performed by the NIH Center for Inherited Disease Research (CIDR; http://www.cidr.jhmi.edu/), using the Illumina Human Omni-2.5 Quad BeadChip. Genotyping quality control was performed by the Genetics Coordinating Center at the University of Washington, Seattle, WA. Further information is available from the HRS website (http://hrsonline.isr.umich.edu/sitedocs/genetics/HRS_QC_REPORT_MAR2012.pdf). Additional more stringent QC was conducted for SNP-based heritability analyses, as models that include the joint effect of all SNPs are known to be sensitive to technical artifacts. For the directly genotyped data, we applied the recommended SNP quality filter provided by CIDR, which yielded 1,681,327 SNPs. Quality control of the imputed genotype data was performed with the QCTOOL software package (http://www.well.ox.ac.uk/~gav/qctool/#overview). SNPs with call rate greater than 95%, MAF > 1%, Hardy-Weinberg equilibrium $P>10^{-6}$, and an imputation info score > 0.5 were retained for further analysis. Subject-level QC was performed with the GTOOL software package (http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html) and included iteratively removing one subject of any pair whose kinship coefficient was greater than 0.1. The numbers of SNPs available after QC for each analysis are shown in Supplementary Table S1.

Because the participants in the HRS were a mixture of European Americans (EA, n = 7,556), African Americans (AA, n = 1,155) and Hispanic Americans (HA, n = 695), we calculated the first 10 principal components (PCs) from the genotype data to use as covariates. After manual inspection, we concluded that 1,354 subjects did not clearly fit any of these categories (see Supplementary Figure S3) and were therefore excluded from both GREML and polygenic analyses.

Genome-wide association study

We used Linear Mixed Model (LMM) or Multivariate Linear Mixed Models (MLMM) implemented in the Genome-wide Efficient Mixed Model Association (GEMMA) software package to further correct for residual

population structure due to ancestry or cryptic relatedness in our GWAS (Zhou and Stephens, 2012). We examined the linear, multivariate and case:control phenotype models using either directly genotyped or a combination of genotyped and imputed SNP data, adjusting for sex, age and marital status (binary). We excluded SNPs with MAF<0.01. For the case:control studies, controls were coded as 0 and cases were coded as 1, as suggested in the GEMMA documentation (Zhou *et al*, 2012). The association analyses were performed using all 10,760 subjects and separately in the subset of the 7,556 EA subjects. We implemented the leave-one-chromosome-out (LOCO) method within the mixed model framework in order to avoid a loss of statistical power due to 'proximal contamination' or inclusion of the candidate marker (or markers in LD with the candidate marker) in the genetic relationship matrix (GRM) (Cheng *et al*, 2013; Yang *et al*, 2014).

Analysis of Candidate Variants Identified in Prior Studies

We identified 13 variants in 11 genes that have previously been associated with loneliness phenotypes in the published literature (Chou, 2010; Chou *et al*, 2014; Connelly *et al*, 2014; Lan *et al*, 2012; Lucht *et al*, 2009; Terracciano *et al*, 2010a; Terracciano *et al*, 2010b; Tsai *et al*, 2012; van Roekel *et al*, 2011; van Roekel *et al*, 2010; van Roekel *et al*, 2013; Verhagen *et al*, 2014; Wang *et al*, 2013). We examined the association of each of these SNPs with loneliness (linear, multivariate and case:control) in the results from the GWAS described above. For those candidate SNPs that were not directly genotyped or imputed in our study, we identified proxy SNPs with r^2 >0.8 whenever possible. For SNPs that yielded p<0.05, we determined whether the direction of the association was consistent between the prior and current studies. We did not apply any correction for multiple comparisons.

Estimation of variance in loneliness explained by the genotyped SNPs ('chip heritability')

To estimate the proportion of phenotypic variance explained ('chip heritability'; h_g^2), we used a genomic restricted maximum likelihood (GREML) method implemented in Genetic Complex Trait Analysis (GCTA) (Yang *et al*, 2010). The purpose of the GREML method is to estimate the proportion of variation in a phenotype that is due to all SNPs. The GREML method is well-established, has been described in detail, and exploits the fact that genotypic similarity (i.e., "relatedness", measured using genotyped SNPs) will be correlated with phenotypic similarity for phenotypes that are influenced by genetic variation. Additional individual-level quality control was implemented and distantly related individuals with pair-wise relationships were further filtered at two thresholds (K_{IBS} < 0.05 and K_{IBS} < 0.025). Covariates included in the GREML analysis were age

(continuous), self-reported sex (male/female), marital status (married/not-married) and top 10 principal components. GREML analyses were run using only directly genotyped SNPs to construct the GRM. We obtained estimates of heritability for both the linear trait and the case:control classification in the EAs subset (N=7,556).

Polygenic Risk Score (PRS) Analysis

To investigate whether the genetic risk for loneliness overlaps with the genetic risk for several personality traits and psychiatric diseases (Supplementary Table S2). For each set ('discovery sample') of GWAS results (e.g., SCZ2), we identified SNPs that were also genotyped in our HRS loneliness data ('target sample') and then used PLINK to LD-prune the SNPs (r^2<0.2; using the "--indep-pairwise" command). The target sample was restricted to EAs to avoid confounding due to residual population stratification. SNPs with association p-values passing pre-determined significance thresholds ($p<10^{-5}$, 10^{-4}, 10^{-3}, 0.01, 0.05, 0.1, 0.3 and 0.5, respectively) in the discovery sample were extracted along with their risk alleles and odds ratios. For each significance threshold, a quantitative aggregate risk score was calculated for each EA individual in the target sample, defined as the sum of the number of risk alleles present at each locus weighted by the log of the odds ratio for that locus estimated from the discovery sample (as implemented in the PLINK "--score" command (Purcell *et al*, 2007). The relationship between aggregate risk score in relation to three phenotypes (e.g. linear, multivariate and case:control status) in the target sample was examined at each significance threshold using linear regression, multivariate regression or logistic regression correspondingly.

LD Score Regression (LDSC)

To further investigate the genetic overlap between loneliness and various other traits (Supplemental Table S2), we used LD score regression (Bulik-Sullivan et al., 2015a, 2015b). We limited our analysis to the EA subjects and used the results from the case:control analysis shown in Figure 1. To standardize the input files, we followed quality controls as implemented by the LDSC python software package (https://github.com/bulik/ldsc). We used pre-calculated LD scores ("eur_w_ld_chr/" files; (Finucane *et al*, 2015)) for each SNP using individuals of European ancestry from the 1000 Genomes project, which are suitable for LD score analysis in European populations. To restrict the analysis to well-imputed SNPs, the SNPs were filtered to HapMap3 SNPs (International Hapmap 3 Consortium et al 2010), and were required to have a MAF

above 1%. INDELs, structural variants, strand-ambiguous SNPs, and SNPs with extremely large effect sizes ($X^2 > 80$) were removed.

RESULTS

Demographics

The distributions of responses to each question are shown in Supplementary Figure S1 and the sum of the three questions is shown in Supplementary Figure S2. Table 1 displays population characteristics according to loneliness status. Consistent with prior studies, loneliness was influenced by ancestry, decreased slightly but significantly with age, and did not differ by gender. Also consistent with prior studies, marital status had a large impact on our quantitative measure of loneliness, with all non-married categories showing greater loneliness compared to individuals who were married (Table 1).

GWAS of loneliness

Figure 1 shows the results of our GWAS for the 7,556 EA-only cohort using both quantile-quantile (QQ) and Manhattan plots. None of the GWAS yielded genome-wide significant associations ($p < 5 \times 10^{-8}$). The most significant results are listed in an Excel spreadsheet that is included in the Supplementary Material. We also performed these analyses using the full set of 10,760 subjects, the results were not meaningfully different (Supplemental Figures S4-S6).

Previously studied candidate genes

Table 2 shows that we did not replicate any of the associations between loneliness and specific candidate genes that had been previously reported. None of these SNPs showed significant evidence for association ($p \leq 0.05$ without correction for multiple comparisons), with the exception of the gene *MTHFR* (Table 2), for which the direction of the association in our data was opposite to what was reported previously (Lan *et al*, 2012). Therefore, none of the previously reported associations could be replicated, despite our large sample size.

Heritability estimates for loneliness

We found that loneliness had a significant chip heritability (case:control 0.27, SE = 0.12, p = 0.01; linear trait PVE=0.16, SE = 0.06, p = 0.002; Table 3). Because loneliness was significantly associated with self-reported ethnicity, we focused on the EA subset for heritability estimates to avoid confounding. To guard against any within-EAs structure, we calculated heritability after adjusting for the top 10 PCs from the genotype data and also after additionally eliminating individuals with $K_{IBS} > 0.05$ and $K_{IBS} > 0.025$. Results were robust to

these different approaches. Although the h_g^2 estimate for the case:control phenotype was higher than the linear trait, the overlapping standard errors indicate that these estimates are not significantly different.

Polygenic score analyses

For these analyses, we used the 6,924 distantly-related/unrelated EAs subjects. Results for neuroticism (Table 4, Table S3, S4) showed unambiguously significant positive co-heritability. We observed modest evidence for negative co-heritability with extraversion (Table S5); the multivariate analysis suggested that the questions “How often do you feel that you lack companionship?” and “How often do you feel left out?” showed the greatest co-heritability with extraversion. We also observed modest evidence for negative co-heritability with schizophrenia (SCZ1 and SCZ2, Tables S6 and S7) and bipolar disorder (Table S8); for these diseases the multivariate analysis suggested that the question “How often do you feel left out?” showed the greatest co-heritability. There was no evidence for co-heritability with major depressive disorder (Table S9) but a non-clinical trait called ‘depressive symptoms’ did show significant positive co-heritability with loneliness (Table S10); the multivariate analysis suggested that all three questions contributed to the observed co-heritability.

Genetic correlation

The results for LDSC analysis used the case:control loneliness GWAS summary statistics and produced results that were broadly similar to the results from the PRS. The genetic correlation between loneliness and personality traits was significantly positive for all three neuroticism datasets (see Figure 2, r_g=0.39, P=4.1x10^{-4}; r_g=0.40, P=8.4x10^{-5}; r_g=0.41, P=2.5x10^{-3}, respectively); and negative for extraversion (r_g= -0.34, P=0.013). Unlike the modest evidence from the PRS analysis, schizophrenia and bipolar disorder did not show any significant results. Whereas there was absolutely no evidence for co-heritability with major depression in the PRS analysis, there was a trend towards a positive correlation in the LDSC analysis (r_g=0.25, P=0.08). Similar to the PRS analysis, the ‘depressive symptoms’ trait was strongly positively correlated with loneliness (r_g=0.39, P=2.9x10^{-4}). We also included height as a negative control in the LDSC analysis; as expected, there was no co-heritability between loneliness and height (P>0.05).

DISCUSSION

Traditionally, the emphasis in research on loneliness has been on environmental predictors and determinants. In the past decade, a series of twin studies have provided estimates of the heritability (h^2) of loneliness. Here we report the first genome-wide association study of loneliness. We have produced the first estimates of the chip heritability (h_g^2) of loneliness (Table 3; 14%-27%), which appear to account for approximately half of the heritability estimated from twin and family studies (37% - 55%). We did not identify any genome-wide significant associations (Figure 1; Supplemental Figures S4-S6), presumably reflecting the very modest contributions of individual variants. Previous studies have reported associations between polymorphisms in a handful of candidate genes and loneliness (Chou, 2010; Chou *et al*, 2014; Connelly *et al*, 2014; Lan *et al*, 2012; Lucht *et al*, 2009; Terracciano *et al*, 2010a; Tsai *et al*, 2012; van Roekel *et al*, 2011; van Roekel *et al*, 2010; van Roekel *et al*, 2013; Verhagen *et al*, 2014; Wang *et al*, 2013); our study did not provide even nominal evidence for replication, despite our much larger sample size (Table 2). Finally, we identified varying levels of evidence for co-heritability between personality traits (positive for neuroticism and negative for extraversion), psychiatric disease traits (negative for schizophrenia and bipolar disorder and positive for depression). This latter result is especially interesting in light of the behavioral research showing that loneliness and psychiatric illness are related in other contexts (Cacioppo *et al*, 2015c), and provides novel evidence that such associations may reflect genetic as well as environmental influences.

Prior behavioral genetic studies have used adoption designs (McGuire *et al*, 2000b), twin designs (Boomsma *et al*, 2005; Waaktaar *et al*, 2012), a family-based design including non-twin siblings (Boomsma *et al*, 2006), and an extended twin designs to include the partners and parents of twins (Distel et al., 2010) to estimate the heritability of loneliness in a variety of populations (Goossens *et al*, 2015). The heritability estimates across these various designs have ranged from 37-55%. These estimates reflect the contributions of both common and rare variants. In contrast, estimates of chip heritability only capture the additive contributions of common variation (Yang *et al*, 2013), and are therefore expected to be lower. As such, they provide insight into the genetic architecture of loneliness, namely that it is polygenic in nature and is likely to be influenced by many common genetic variants of small effect. Our estimates of chip heritability add to existing heritability estimates and also reinforce the notion that both genetic and environmental factors influence loneliness. Future

studies might increase heritability by utilizing more environmentally homogeneous populations or by including more environmental variables as covariates.

Our study did not identify any genome-wide significant associations. Although the sample size of slightly more than 10,000 individuals provides appreciably greater statistical power than had been available previously, numerous disease and non-disease phenotypes that are known to be heritable have also yielded negative results with similar sample sizes (e.g. Major Depressive Disorder Working Group of the Psychiatric *et al*, 2013).

Our study provided an efficient means of testing previously reported associations between SNPs in candidate genes and loneliness. Despite an active literature in this area, we did not find support for any of the previously reported candidate gene associations. This is consistent with our previous experience with candidate gene-based studies (Hart *et al*, 2013). Several of the previously published candidate gene studies reported effect sizes that are much larger than those typically observed in genome-wide association studies, which in hindsight should have generated more skepticism about those results. Although our findings cast doubt on the previously reported associations – or at least on the original effect sizes that were identified -- there are potentially important differences between our study design and the previously published candidate gene studies. For instance, our population was based in the United States and was made up of older adults, many of whom were in stable long-term relationships, whereas approximately half of the candidate-gene studies utilized samples of adolescents from the Netherlands or Germany (Lucht *et al*, 2009; van Roekel *et al*, 2011; van Roekel *et al*, 2010; van Roekel *et al*, 2013; Verhagen *et al*, 2014). Therefore, although our study benefited from a larger sample size than any of the previously reported candidate gene studies, we cannot discount the possibility that differences in the methodologies or the population under study led to our failure to replicate the previously published results. The phenotyping procedure used in the current study has been found to correlate highly (r = 0.88) with a more in depth phenotype for loneliness (Hughes *et al*, 2004), the candidate gene studies using older adults from the United Kingdom and Taiwan provided no greater evidence for replication than the studies using adolescents (Chou, 2010; Chou *et al*, 2014; Connelly *et al*, 2014; Lan *et al*, 2012; Tsai *et al*, 2012; Wang *et al*, 2013).

We observed strong genetic correlations between loneliness and two personality dimensions: neuroticism and extraversion (Table 4, Figure 2, Tables S3-S5). The direction of these effects was consistent

with the correlations identified previously: greater loneliness has been shown to be positively correlated with neuroticism and negatively correlated with extraversion (e.g., (Cacioppo *et al*, 2006; Mund and Neyer, 2015). Neuroticism is characterized by high negative affectivity, a characteristic also seen in loneliness. Although prior research has shown loneliness and neuroticism to be stochastically and functionally separable, the results from the current study suggest there may be a shared genetic predisposition that contributes to both phenotypes. The multivariate PRS analyses provide information regarding the co-heritability between loneliness and extraversion. The results showed that whereas all three questions contributed to the genetic correlation with neuroticism (Table 4, Supplemental Tables S3-S4), only the items regarding lack of companionship and feeling left out contributed to the genetic correlation with extraversion (Supplemental Tables S5).

We see our study as being a part of an important trend that attempts to relate the genetic causes of psychiatric disease diagnoses to continuously variable traits that represent heritable personality characteristics. It is widely accepted that humans have varying degrees of sensitivity to social isolation; however, the question of whether or not the genetic basis of this variability also underlies the risk for common psychiatric diseases remains largely unexplored. We have previously reported that genetic variation in the initial sensitivity to the euphoric effects of amphetamine is genetically correlated with the risk for both schizophrenia and ADHD (Hart *et al*, 2014). That provocative finding provides an example of using genetic variation in a non-disease trait to obtain novel insights into the genetic basis of psychiatric diseases. In the present study, we explored whether or not genetic risk for loneliness mapped onto genetic risk for major psychiatric diseases. Our results provided limited support for this hypothesis (Supplementary Tables S6-S9; Figure 2). The linear phenotype did not show any evidence for co-heritability with schizophrenia, bipolar disorder or major depression. However, when we used a multi-trait mapping approach, which allowed us to consider each of the 3 questionnaire items independently, we saw suggestive evidence for co-heritability between the second question (“how often do you feel left out?”) and both schizophrenia and bipolar disorder. The relationship between loneliness and these two diseases was very weakly negative, meaning that being lonelier is associated with reduced risk of disease. We also used summary statistics from the case:control GWAS to perform LDSC, which did not support co-heritability with schizophrenia or bipolar disorder, but did show a trend towards a positive genetic correlation with major depression (p=0.08; lonelier was associated with greater risk for depression). The non-psychiatric trait termed ‘depressive symptoms’ was more robustly positively correlated with loneliness in both the PRS and

the LDSC analyses (lonelier was genetically correlated with more depressive symptoms). Because of the number of tests performed and the modest levels of significance for the psychiatric diseases, those results should be considered suggestive until they are replicated. While we assume that few participants in the HRS study would be diagnosed with schizophrenia, such data were not available; therefore, we cannot exclude the possibility that our findings are secondary to the effects of schizophrenia on self-reported loneliness. The direction of the effect suggests that greater genetic risk for loneliness is negatively associated with these psychiatric diseases. We have previously hypothesized that loneliness reflects an adaptive drive towards social interaction, which is consistent with the direction of the observed correlation. Future studies of loneliness and other social behavior traits may continue to inform our understanding of the role of social behavior in psychiatric health and disease (Cacioppo *et al*, 2014).

In summary, we have performed the first genome-wide association study of loneliness. Our study has identified significant evidence for heritability, but did not identify specific loci associated with loneliness nor was it able to replicate previously reported candidate gene associations. Finally, we identified strong evidence for co-heritability between loneliness and neuroticism, extraversion and 'depressive symptoms' and suggestive evidence for co-heritability between loneliness and schizophrenia, bipolar disorder and major depressive disorder. We believe that future studies of loneliness, as well as additional studies of other social neuroscience phenotypes may continue to enrich our understanding of the ways in which our genetic inheritance fundamentally influences individual and social behavior.

FUNDING AND DISCLOSURE

Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under award number R37-AG033590 and by the National Center for the National Center of Advancing Translational Sciences of the National Institutes of Health under award number KL2TR000431 (LKD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. All authors report no potential conflicts of interest.

ACKNOWLEDGEMENTS

We thank Hae Kyung Im for helpful conversations during the preparation of this manuscript.

REFERENCES

Adam EK, Hawkley LC, Kudielka BM, Cacioppo JT (2006). Day-to-day dynamics of experience--cortisol associations in a population-based sample of older adults. *Proc Natl Acad Sci U S A* **103**(45): 17058-17063.

Bartels M, Cacioppo JT, Hudziak JJ, Boomsma DI (2008). Genetic and environmental contributions to stability in loneliness throughout childhood. *Am J Med Genet B Neuropsychiatr Genet* **147**(3): 385-391.

Baumeister R, DeWall N (2005). The inner dimension of social exclusion: Intelligent thought and self-regulation among rejected persons. In J. P. F. K. D. Williams, & W. von Hippel (Ed.). In: J. P. F. K. D. Williams, Hippel Wv (eds). *The social outcast: Ostracism, social exclusion, rejection, and bullying* Psychology Press: New York, pp 53-76.

Boomsma DI, Cacioppo JT, Muthen B, Asparouhov T, Clark S (2007). Longitudinal genetic analysis for loneliness in Dutch twins. *Twin Res Hum Genet* **10**(2): 267-273.

Boomsma DI, Cacioppo JT, Slagboom PE, Posthuma D (2006). Genetic linkage and association analysis for loneliness in Dutch twin and sibling pairs points to a region on chromosome 12q23-24. *Behav Genet* **36**(1): 137-146.

Boomsma DI, Willemsen G, Dolan CV, Hawkley LC, Cacioppo JT (2005). Genetic and environmental contributions to loneliness in adults: the Netherlands twin register study. *Behav Genet* **35**(6): 745-752.

Booth R (2000). Loneliness as a component of psychiatric disorders. *Medscape General Medicine* **2**: 1-7.

Cacioppo JT, Cacioppo S, Boomsma DI (2014). Evolutionary mechanisms for loneliness. *Cogn Emot* **28**(1): 3-21.

Cacioppo JT, Cacioppo S, Capitanio JP, Cole SW (2015a). The neuroendocrinology of social isolation. *Annu Rev Psychol* **66**: 733-767.

Cacioppo JT, Ernst JM, Burleson MH, McClintock MK, Malarkey WB, Hawkley LC*, et al* (2000). Lonely traits and concomitant physiological processes: the MacArthur social neuroscience studies. *Int J Psychophysiol* **35**(2-3): 143-154.

Cacioppo JT, Hawkley LC, Berntson GG, Ernst JM, Gibbs AC, Stickgold R*, et al* (2002). Do lonely days invade the nights? Potential social modulation of sleep efficiency. *Psychol Sci* **13**(4): 384-387.

Cacioppo JT, Hawkley LC, Thisted RA (2010). Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago Health, Aging, and Social Relations Study. *Psychol Aging* **25**(2): 453-463.

Cacioppo JT, Hughes ME, Waite LJ, Hawkley LC, Thisted RA (2006). Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses. *Psychol Aging* **21**(1): 140-151.

Cacioppo S, Balogh S, Cacioppo JT (2015b). Implicit attention to negative social, in contrast to nonsocial, words in the Stroop task differs between individuals high and low in loneliness: Evidence from event-related brain microstates. *Cortex*.

Cacioppo S, Grippo AJ, London S, Goossens L, Cacioppo JT (2015c). Loneliness: clinical import and interventions. *Perspect Psychol Sci* **10**(2): 238-249.

Cheng R, Parker CC, Abney M, Palmer AA (2013). Practical considerations regarding the use of genotype and pedigree data to model relatedness in the context of genome-wide association studies. *G3 (Bethesda)* **3**(10): 1861-1867.

Chou KL (2010). Moderating effect of apolipoprotein genotype on loneliness leading to depressive symptoms in Chinese older adults. *Am J Geriatr Psychiatry* **18**(4): 313-322.

Chou KL, Cacioppo JT, Kumari M, Song YQ (2014). Influence of social environment on loneliness in older adults: Moderation by polymorphism in the CRHR1. *Am J Geriatr Psychiatry* **22**(5): 510-518.

Cole SW, Hawkley LC, Arevalo JM, Cacioppo JT (2011). Transcript origin analysis identifies antigen-presenting cells as primary targets of socially regulated gene expression in leukocytes. *Proc Natl Acad Sci U S A* **108**(7): 3080-3085.

Cole SW, Hawkley LC, Arevalo JM, Sung CY, Rose RM, Cacioppo JT (2007). Social regulation of gene expression in human leukocytes. *Genome Biol* **8**(9): R189.

Connelly JJ, Golding J, Gregory SP, Ring SM, Davis JM, Davey Smith G*, et al* (2014). Personality, behavior and environmental features associated with OXTR genetic variants in British mothers. *PLoS One* **9**(3): e90465.

Distel MA, Willemsen G, Ligthart L, Derom CA, Martin NG, Neale MC*, et al* (2010). Genetic covariance structure of the four main features of borderline personality disorder. *J Pers Disord* **24**(4): 427-444.

Dixon D, Cruess S, Kilbourn K, Klimas N, Fletcher MA, Ironson G*, et al* (2006). Social Support Mediates Loneliness and Human Herpesvirus Type 6 (HHV-6) Antibody Titers. *J Appl Soc Psychol* **31**(6): 1111-1132.

Doane LD, Adam EK (2010). Loneliness and cortisol: momentary, day-to-day, and trait associations. *Psychoneuroendocrinology* **35**(3): 430-441.

Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR*, et al* (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**(11): 1228-1235.

Glaser K, Evandrou M, Tomassini C (2005). The health consequences of multiple roles at older ages in the UK. *Health Soc Care Community* **13**(5): 470-477.

Glaser R, Kiecolt-Glaser JK, Speicher CE, Holliday JE (1985). Stress, loneliness, and changes in herpesvirus latency. *J Behav Med* **8**(3): 249-260.

Goossens L, van Roekel E, Verhagen M, Cacioppo JT, Cacioppo S, Maes M*, et al* (2015). The genetics of loneliness: linking evolutionary theory to genome-wide genetics, epigenetics, and social science. *Perspect Psychol Sci* **10**(2): 213-226.

Hackett RA, Hamer M, Endrighi R, Brydon L, Steptoe A (2012). Loneliness and stress-related inflammatory and neuroendocrine responses in older men and women. *Psychoneuroendocrinology* **37**(11): 1801-1809.

Hart AB, de Wit H, Palmer AA (2013). Candidate gene studies of a promising intermediate phenotype: failure to replicate. *Neuropsychopharmacology* **38**(5): 802-816.

Hart AB, Gamazon ER, Engelhardt BE, Sklar P, Kahler AK, Hultman CM*, et al* (2014). Genetic variation associated with euphorigenic effects of d-amphetamine is associated with diminished risk for schizophrenia and attention deficit hyperactivity disorder. *Proc Natl Acad Sci U S A* **111**(16): 5968-5973.

Hawkley LC, Masi CM, Berry JD, Cacioppo JT (2006). Loneliness is a unique predictor of age-related differences in systolic blood pressure. *Psychol Aging* **21**(1): 152-164.

Hawkley LC, Preacher KJ, Cacioppo JT (2010a). Loneliness impairs daytime functioning but not sleep duration. *Health Psychol* **29**(2): 124-129.

Hawkley LC, Thisted RA, Cacioppo JT (2009). Loneliness predicts reduced physical activity: cross-sectional & longitudinal analyses. *Health Psychol* **28**(3): 354-363.

Hawkley LC, Thisted RA, Masi CM, Cacioppo JT (2010b). Loneliness predicts increased blood pressure: 5-year cross-lagged analyses in middle-aged and older adults. *Psychol Aging* **25**(1): 132-141.

Health_and_Retirement_Study (2012). Core public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D (2015). Loneliness and social isolation as risk factors for mortality: a meta-analytic review. *Perspect Psychol Sci* **10**(2): 227-237.

Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT (2004). A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies. *Res Aging* **26**(6): 655-672.

Jacobs JM, Cohen A, Hammerman-Rozenberg R, Stessman J (2006). Global sleep satisfaction of older people: the Jerusalem Cohort Study. *J Am Geriatr Soc* **54**(2): 325-329.

Jaremka LM, Fagundes CP, Peng J, Bennett JM, Glaser R, Malarkey WB*, et al* (2013). Loneliness promotes inflammation during acute stress. *Psychol Sci* **24**(7): 1089-1097.

Kiecolt-Glaser JK, Ricker D, George J, Messick G, Speicher CE, Garner W*, et al* (1984). Urinary cortisol levels, cellular immunocompetency, and loneliness in psychiatric inpatients. *Psychosom Med* **46**(1): 15-23.

Kong F, You X (2013). Loneliness and self-esteem as mediators between social support and life satisfaction in late adolescence. *Social Indicators Research* **110**: 271-279.

Kurina LM, Knutson KL, Hawkley LC, Cacioppo JT, Lauderdale DS, Ober C (2011). Loneliness is associated with sleep fragmentation in a communal society. *Sleep* **34**(11): 1519-1526.

Lan WH, Yang AC, Hwang JP, Hong CJ, Liou YJ, Yeh HL*, et al* (2012). Association of MTHFR C677T polymorphism with loneliness but not depression in cognitively normal elderly males. *Neurosci Lett* **521**(1): 88-91.

Lucht MJ, Barnow S, Sonnenfeld C, Rosenberger A, Grabe HJ, Schroeder W*, et al* (2009). Associations between the oxytocin receptor gene (OXTR) and affect, loneliness and intelligence in normal subjects. *Prog Neuropsychopharmacol Biol Psychiatry* **33**(5): 860-866.

Luo Y, Hawkley LC, Waite LJ, Cacioppo JT (2012). Loneliness, health, and mortality in old age: a national longitudinal study. *Soc Sci Med* **74**(6): 907-914.

Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM*, et al* (2013). A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* **18**(4): 497-511.

McGuire S, Clifford J (2000a). Genetic and Environmental Contributions to Loneliness in Children. *Psychological Science* **11**(6): 487-491.

McGuire S, Clifford J (2000b). Genetic and environmental contributions to loneliness in children. *Psychol Sci* **11**(6): 487-491.

Mund M, Neyer FJ (2015). The Winding Paths of the Lonesome Cowboy: Evidence for Mutual Influences Between Personality, Subjective Health, and Loneliness. *J Pers*.

Pressman SD, Cohen S, Miller GE, Barkin A, Rabin BS, Treanor JJ (2005). Loneliness, social network size, and immune response to influenza vaccination in college freshmen. *Health Psychol* **24**(3): 297-306.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D*, et al* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**(3): 559-575.

Steptoe A, Owen N, Kunz-Ebrecht SR, Brydon L (2004). Loneliness and neuroendocrine, cardiovascular, and inflammatory stress responses in middle-aged men and women. *Psychoneuroendocrinology* **29**(5): 593-611.

Straits-Tröster KA, Patterson TL, Semple SJ, Temoshok L, Roth PG, McCutchan JA*, et al* (1994). The relationship between loneliness, interpersonal competence, and immunologic status in HIV-infected men. *Psychology & Health* **9**: 205-219.

Terracciano A, Tanaka T, Sutin AR, Deiana B, Balaci L, Sanna S*, et al* (2010a). BDNF Val66Met is associated with introversion and interacts with 5-HTTLPR to influence neuroticism. *Neuropsychopharmacology* **35**(5): 1083-1089.

Terracciano A, Tanaka T, Sutin AR, Sanna S, Deiana B, Lai S*, et al* (2010b). Genome-wide association scan of trait depression. *Biol Psychiatry* **68**(9): 811-817.

Tsai SJ, Yeh HL, Hong CJ, Liou YJ, Yang AC, Liu ME*, et al* (2012). Association of CHRNA4 polymorphism with depression and loneliness in elderly males. *Genes Brain Behav* **11**(2): 230-234.

van Roekel E, Goossens L, Scholte RH, Engels RC, Verhagen M (2011). The dopamine D2 receptor gene, perceived parental support, and adolescent loneliness: longitudinal evidence for gene-environment interactions. *J Child Psychol Psychiatry* **52**(10): 1044-1051.

van Roekel E, Scholte RH, Verhagen M, Goossens L, Engels RC (2010). Loneliness in adolescence: gene x environment interactions involving the serotonin transporter gene. *J Child Psychol Psychiatry* **51**(7): 747-754.

van Roekel E, Verhagen M, Scholte RH, Kleinjan M, Goossens L, Engels RC (2013). The oxytocin receptor gene (OXTR) in relation to state levels of loneliness in adolescence: evidence for micro-level gene-environment interactions. *PLoS One* **8**(11): e77689.

VanderWeele TJ, Hawkley LC, Cacioppo JT (2012). On the reciprocal association between loneliness and subjective well-being. *Am J Epidemiol* **176**(9): 777-784.

VanderWeele TJ, Hawkley LC, Thisted RA, Cacioppo JT (2011). A marginal structural model analysis for loneliness: implications for intervention trials and clinical practice. *J Consult Clin Psychol* **79**(2): 225-235.

Verhagen M, van Roekel E, Engels RC (2014). Involvement of the BDNF gene in loneliness in adolescence: a report of opposite gene effects in boys and girls. *PLoS One* **9**(3): e92768.

Waaktaar T, Torgersen S (2012). Genetic and environmental causes of variation in perceived loneliness in young people. *Am J Med Genet B Neuropsychiatr Genet* **159B**(5): 580-588.

Wang EH, Hong CJ, Yeh HL, Liou YJ, Yang AC, Liu ME*, et al* (2013). Interleukin-1 alpha (rs1800587) genetic polymorphism is associated with specific cognitive functions but not depression or loneliness in elderly males without dementia. *Neurosci Lett* **556**: 69-72.

Whisman MA (2010). Loneliness and the metabolic syndrome in a population-based sample of middle-aged and older adults. *Health Psychol* **29**(5): 550-554.

Wilson RS, Krueger KR, Arnold SE, Schneider JA, Kelly JF, Barnes LL*, et al* (2007). Loneliness and risk of Alzheimer disease. *Arch Gen Psychiatry* **64**(2): 234-240.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR*, et al* (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**(7): 565-569.

Yang J, Lee SH, Goddard ME, Visscher PM (2011). GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**(1): 76-82.

Yang J, Lee SH, Goddard ME, Visscher PM (2013). Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol Biol* **1019**: 215-236.

Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* **46**(2): 100-106.

Zhou X, Stephens M (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**(7): 821-824.

FIGURE LEGENDS

Figure 1 QQ and Manhattan plots of genome-wide association studies for loneliness in EA-only subjects using the imputed SNPs, calculated by mixed regression models with adjustments for age, gender and marital status. (A, B): linear mixed model (n=7,556); (C, D): multivariate mixed model (n=7,556), (E, F): case:control mixed model (n=5,228).

Figure 2 Genetic correlations between loneliness (EA-only, case:control) and 9 additional traits: personality traits (neuroticism, extraversion), psychiatric conditions (schizophrenia, SCZ; bipolar disorder, BP; major depression disorder, MDD), a depressive symptoms scale (DS), and height. Error bars represent standard errors. * P<0.01, ** p<0.001, *** p<0.0001, n.s. p>0.05.

Table 1 Participants' Characteristics of Health and Retirement Study (HRS) and associations with loneliness [a]

Characteristic	Linear trait (3~9; n=10,760)				Case: Control (Cases=2,853: Controls=4,583)			
	Mean (SE)/ N (%)	β	SE	P	Case	Control	OR (95%CI)	p
Age [b]	67.2 (10.3)	-0.01	0.0016	**<.0001**	66.6 (10.8)	67.4 (9.9)	0.98 (0.98, 0.99)	**<.0001**
Gender [c]								
women	6,376 (50.6)	-	-		1,067 (37.4)	2,007 (43.8)	-	
men	4,384 (40.7)	-0.001	0.03	0.98	1,786 (62.6)	2,576 (56.2)	1.02 (0.92, 1.14)	0.67
Self-reported ethnicity [c]								
European Americans	8,490	-	-		2,075(72.73)	3,875 (82.59)	-	
African Americans	1,228	0.22	0.05	**<.0001**	427 (14.97)	398 (8.68)	1.54 (1.32, 1.79)	**<.0001**
Hispanic Americans	867	0.25	0.06	**<.0001**	294 (10.30)	327 (7.14)	1.52 (1.28, 1.80)	**<.0001**
Other/ Unknown	175	0.18	0.12	0.13	57 (2.00)	73 (1.59)	1.39 (0.97, 1.99)	0.07
Marital status [c]								
Married	7,120 (66.2)	-	-		1,538 (53.9)	3,496 (76.3)	-	
Annulled	364 (3.4)	0.16	0.09	0.06	100 (3.5)	165 (3.6)	1.26 (0.98, 1.64)	0.08
Separated	132 (1.2)	1.06	0.14	**<.0001**	63 (2.2)	31 (0.7)	4.34 (2.81, 6.71)	**<.0001**
Divorced	991 (9.2)	0.85	0.05	**<.0001**	394 (13.8)	264 (5.8)	3.30 (2.79, 3.91)	**<.0001**
Widowed	1,892 (17.6)	0.75	0.05	**<.0001**	655 (23.0)	551 (12.0)	3.29 (2.84, 3.80)	**<.0001**
Never Married	261 (2.4)	0.75	0.10	**<.0001**	103 (3.6)	76 (1.7)	3.02 (2.23, 4.09)	**<.0001**
Binary marital status [c]								
Married	7,120 (66.2)	-	-		1,538 (53.9)	3,496 (76.3)	-	
Unmarried	3,640 (33.8)	0.72	0.03	**<.0001**	1,315 (46.1)	1,087 (23.7)	2.91 (2.62, 3.23)	**<.0001**

a. Higher score means more loneliness. β (β coefficient), SE (standard error) and P-values were obtained from linear regression models, adjusting for age, sex and marital status. Odds ratios (ORs) and 95% confidence intervals (CIs) were assessed with logistic regression model using same covariates. P-values ≤ 0.05 were in bold, while 0.05 < P-values ≤ 0.1 in underline.

b. Continuous variables are presented as mean and standard deviation (sd).

c. Categorical variables are presented as counts and percentages (gender, ethnic and marital status).

Table 2 Association between loneliness phenotypes and candidate gene associations reported in prior studies

Gene	Chr	Reported SNP	Function	Population	Sample Size	References	P for association (all subjects/EA-only) [b]		
							Linear	Multivariate	Case:Control
BDNF	11	rs6265	Val66Met	Dutch	305	Verhagen M, 2014	0.78/0.56	0.49/0.66	0.64/0.70
		rs53576 [a]	intron			Connelly JJ, 2014	NA	NA	NA
OXTR	3	rs2254298 [a]	intron	UK/ Germany	7723/285/89	Lucht MJ, 2009	0.61/0.71	0.87/0.98	0.98/0.48
		rs2228485 [a]	synonymous			van Roekel, 2013	NA	NA	NA
RORA	15	rs12912233	intron	Italy + US	3972 + 839	Terracciano A, 2010	0.60/0.35	0.90/0.53	0.86/0.60
GRM8	7	rs17864092	intron				0.77/0.77	0.41/0.29	0.98/0.83
CHRNA4	20	rs1044396	synonymous	Taiwan	192	Tsai SJ, 2012	0.75/0.95	0.76/0.77	0.73/0.82
IL-1A	2	rs1800587	5' UTR	Taiwan	192	Wang EH, 2013	0.33/0.55	0.25/0.52	0.69/0.68
CRHR1	17	rs1876831	intron	UK	1,374	Chou KL, 2014	0.28/0.09	0.54/0.41	0.14/0.06
		rs242938	intron				0.74/0.46	0.87/0.74	0.14/0.20
MTHFR	1	rs1801133	Ala222Val	Taiwan	323	Lan WH, 2012	0.046/0.15	0.08/0.31	0.036/0.052
DRD2	11	rs1800497	Glu713Lys	Netherlands	307	van Roekel, 2011	0.45/0.09	0.75/0.22	0.46/0.16
APOE	19	rs7412 (ε2)	Arg176Cys	Taiwan	979	Chou KL, 2010	0.37/0.69	0.43/0.43	0.74/0.96
SLC6A4	17	Insertion/5-HTTLPR		Netherlands	306	van Roekel E,2010	NA	NA	NA

a. For those candidate SNPs that were not genotyped/imputed in our study, we identified proxy SNPs with r^2 >0.8 that were genotyped or imputed in our study based on HapMap2. When no proxy SNP could be identified we report N/A rather than a p-value.

b. P values are not corrected for multiple comparisons. We used multivariate or logistic regression models (GEMMA) to account for relatedness. Adjustments include sex, age, and marital status. P values before "/" are for all the 10,760 participants; after "/" are for 7,556 European Americans only. For the gene MTHFR (rs1801133), the direction of effect was opposite to what had been reported previously.

Table 3 Chip heritability estimates in European Americans (EAs)

Threshold of K_{IBS}		Linear trait				Case-Control			
		N	PVE[d]	SE[e]	P	N	PVE	SE	P
All European Americans [a]	No PCs [b]	7,556	16%	6	**0.002**	5,228	27%	12	**0.01**
	With 10 PCs [b]	7,556	16%	6	**0.003**	5,228	26%	13	**0.02**
Excluding closely related pairs ($K_{IBS} < 0.05$) [c]		7,381	16%	6	**0.006**	5,113	25%	14	**0.04**
Excluding closely related pairs ($K_{IBS} < 0.025$) [c]		6,924	14%	7	**0.02**	4,796	25%	15	**0.05**

[a] using the full GRM, K_{IBS} on all individuals. P-values ≤ 0.05 are bold.

[b] No PCs (principal components): Covariates including gender, age (continuous), and marital status (binary). With 10 PCs: covariates included the first 10 PCs of genotype data.

[c] The GRM includes only distantly-related pairs ($K_{IBS} < 0.05$ or 0.025). One individual from each relative pair was excluded. IBS = Identical by descent

[d] PVE = percent variance explained
[e] SE = standard error

Table 4 Associations between polygenic scores for Neuroticism from SSGAC (Okbay et al 2016) and loneliness in Health and Retirement Study (HRS)[a]

p-value threshold	Num. of SNPs	Linear trait [b]			Multivariate traits [c]							Case: Control [d]		
		β	Se	p	Q1:companion β1	p1	Q2: left out β2	p2	Q3: isolated β3	p3	p for overall	OR	95% CI	p
$1x10^{-5}$	33	0.03	0.02	0.08	0.01	0.14	0.01	0.12	0.01	0.14	0.37	1.08	1.01, 1.14	**0.02**
$1x10^{-4}$	113	0.03	0.02	0.10	0.01	0.12	0.009	0.18	0.009	0.20	0.43	1.08	1.02, 1.15	**0.01**
$1x10^{-3}$	515	0.04	0.02	**0.02**	0.01	0.07	0.01	**0.03**	0.01	0.07	0.16	1.08	1.01,1.15	**0.02**
0.01	3,223	0.07	0.02	**$9.1x10^{-5}$**	0.02	**0.001**	0.03	**0.0001**	0.02	**0.003**	**0.001**	1.15	1.08, 1.22	**$1.2x10^{-5}$**
0.05	13,036	0.08	0.02	**$9.3x10^{-6}$**	0.03	**$3.5x10^{-4}$**	0.03	**$9.9x10^{-6}$**	0.02	**0.0008**	**$9.9x10^{-5}$**	1.17	1.10, 1.24	**$1.5x10^{-6}$**
0.1	24,251	0.09	0.02	**$4.1x10^{-6}$**	0.03	**$4.1X10^{-5}$**	0.03	**$2.9x10^{-5}$**	0.02	**0.0005**	**$6.7x10^{-5}$**	1.17	1.10, 1.24	**$1.4x10^{-6}$**
0.3	65,722	0.09	0.02	**$4.1x10^{-7}$**	0.03	**$3.0x10^{-5}$**	0.03	**$3.9x10^{-6}$**	0.03	**$3.3x10^{-5}$**	**$1.0x10^{-5}$**	1.17	1.10, 1.25	**$3.6x10^{-7}$**
0.5	105,444	0.10	0.02	**$1.6x10^{-8}$**	0.03	**$2.3x10^{-6}$**	0.03	**$4.7x10^{-7}$**	0.03	**$3.0x10^{-6}$**	**$5.3x10^{-7}$**	1.20	1.12, 1.27	**$1.6x10^{-8}$**

[a.] The polygenic model was developed using SNPs with p-values below the indicated threshold from Neuroticism obtained from Social Science Genetic Association Consortium (SSGAC, Neuroticism_Full.txt). The testing set was an independent set using the data of HRS; the polygenic scores have been standardized, so the β coefficients from the Neuroticism linear regression model correspond to a one standard deviation change in the polygenic score. P-values ≤ 0.05 are bold, while 0.05 < P-values ≤ 0.1 are underline.

[b.] using linear regression model for 6,924 unrelated EAs; adjustments included sex, age and marital status; further adjusting for the top 3 PCs was little impact.

[c.] using multivariate regression model 6,924 unrelated EAs, same adjustments as above;

[d.] using logistic regression model for 4,796 unrelated EAs (Cases: Controls= 1,632: 3,164), same adjustments as above.