Published in final edited form as:

Biol Psychiatry. 2012 October 15; 72(8): 707–709. doi:10.1016/j.biopsych.2012.03.011.

Estimating the genetic variance of Major Depressive Disorder (MDD) due to all SNPs

Gitta H. Lubke^{1,2}, Jouke Jan Hottenga², Raymond Walters¹, Charles Laurin¹, Eco J.C. de Geus², Gonneke Willemsen², Jan H. Smit³, Christel Middeldorp^{2,3,4}, Brenda W.J.H. Penninx³, Jacqueline Vink², and Dorret I Boomsma²

¹University of Notre Dame, Notre Dame, USA ²VU University Amsterdam, The Netherlands ³VU University Medical Center, Amsterdam, The Netherlands ⁴Academic Medical Center, Amsterdam, The Netherlands

Abstract

Genome-wide association (GWA) studies of psychiatric disorders have been criticized for their lack in explaining a considerable proportion of the heritability established in twin and family studies. GWA studies of Major Depressive Disorder (MDD) in particular have so far been unsuccessful in detecting genome-wide significant SNPs. Using two different recently proposed methods designed to estimate the heritability of a phenotype that is attributable to genome-wide SNPs, we show that SNPs on current platforms contain substantial information concerning the additive genetic variance of MDD. To assess the consistency of these two different methods we analyzed four other complex phenotypes from different domains. The pattern of results is consistent with estimates of heritability obtained in twin studies carried out in the same population.

A limited number of meta-analyses combining multiple Genome Wide Association (GWA) studies of psychiatric disorders have yielded replicated associations between such disorders and Single Nucleotide Polymorphisms (SNPs) (e.g., schizophrenia, [1]). In contrast, analyses of Major Depressive Disorder (MDD) lack genome-wide significant results [2,3]. Previously, we conducted a genome-wide association (GWA) study in 1,738 MDD cases and 1,802 controls selected to be at low liability for MDD [2]. None of the SNPs reached genome-wide significance. Subsequent GWA and meta-analysis studies of MDD were equally unsuccessful in detecting significant SNPs [3]. The lack of breakthrough results has spurred a discussion concerning the utility of GWA studies of psychiatric phenotypes. Part of this discussion focuses on whether SNPs analyzed in GWA studies are the relevant polymorphisms, or whether psychiatric disorders traits are more likely influenced by rare variants.

© 2012 Society of Biological Psychiatry. Published by Elsevier Inc. All rights reserved.

Corresponding author: Gitta Lubke, University of Notre Dame 118 Haggard Hall, Notre Dame IN 46556, fax: (574) 631-8333, phone: (574) 631-8789, glubke@nd.edu.

Author Contributions

GL and DIB wrote the report with contributions from all authors, GL carried out the analyses, JJH, RW, CL, JV helped with the analyses, GW, JHS, CM prepared phenotype data, DIB, EJCG manage the NTR, BWJHP manages the NESDA project.

Conflict of interest: The authors report no biomedical financial interests or potential conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Yang et al. [4] and So et al. [5] added to this discussion by proposing two different methods which focus on the estimation of phenotypic variance explained by currently genotyped SNPs rather than on the detection of specific SNPs that contribute to a trait or disorder. Using human height as an example, Yang et al. [4] demonstrated that SNPs that are currently included on genotyping platforms explain a substantial proportion of the heritability of the phenotype. So et al. [5, 6] reached similar conclusions. The two approaches differ substantially, with Yang et al.'s approach falling under the broad umbrella of variance decomposition methods, and So et al.'s approach being a density estimation method. Both methods focus on estimating variance due to additive genetic effects.

The rationale of Yang et al.'s method is to decompose the variance of the phenotype into a component that is due to the additive effects of all measured SNPs, and a residual component that is due to random noise, unmeasured environmental influences, or effects of unmeasured genetic variants that are not captured by the measured variants. Their 2-step method involves first obtaining the genetic similarity between all pairs of subjects, which is calculated as a correlation. For each pair of subjects, their minor allele counts on a given SNP are subtracted from the SNP mean, and the cross-product of the resulting deviance scores is divided by the SNP variance. Summing these terms over all SNPs gives a measure of genetic similarity between two subjects that is due to additive genetic effects. The second step of the method uses this measure as a random effect to predict the phenotype in a linear mixed model. The original approach for continuous traits has been extended to case/control studies [7,8].

The method proposed by So et al. is entirely different, and is applied subsequent to a GWAS. The basic idea is to compare the distribution of z-statistics of the regression coefficients of genome-wide SNPs in a GWAS to the theoretical Null distribution of z-statistics representing no effects. Explained variance will differ from zero if more z-statistics from the GWAS have larger values than expected under the Null. Specifically, the observed z-statistics, which contain error due to sampling fluctuation, are first corrected to obtain z-statistics representing "true" effect sizes. The estimate of variance explained for continuous phenotypes can then be computed by summing contributions of all non-null "true" z-statistics using sums of squares as in ANOVA. For case/control phenotypes So et al. [6] have described a method using the odds ratio, allele frequency, and disease prevalence.

Our aim was to estimate the joint effect of all SNPs in explaining the variance in MDD. In addition, we investigated four other phenotypes that are currently targets of large GWA projects, namely height, fasting glucose, smoking initiation, and current smoking. The estimates of explained variance of all five phenotypes using the two different methods were compared to heritability estimates obtained in twin and family studies that were carried out in the same population.

The data analyzed in this study came from individuals who took part in the Netherlands Twin Register Biobank study [9] and the Netherlands Study of Depression and Anxiety [10]. Genotyping on these samples was performed on the Affymetrix 6.0 (N=298), Affymetrix 5.0/Perlegen (N=3697), Illumina 370 (N=290), Illumina 660 (N=1439) and Illumina Omni Express 1M (N=445) platforms. Quality control was carried out per platform. Thresholds for SNPs were MAF > 1%, HWE > 0.00001, missing >95% and $0.30 < \text{Heterozygosity} < 0.35$. Samples were excluded from the data if their expected sex and IBD status did not match, or if the genotype missing rate > 10%. Subsequent to alignment using the Hapmap 2 Build 36 release 24 CEU reference set, and exclusion of SNPs if allele frequencies differed more than 15% with the reference set and/or the other platforms, the data were merged into a single dataset (N=5856). Imputation was done using IMPUTE v2, and SNPs were removed based on HWE < 0.00001, proper info < 0.40, MAF < 1%, allele frequency difference > 0.15

against reference and a missing rate > 5% assuming a 90% calling threshold. Individuals were removed with $IBS > 0.025$, resulting in a sample of 4605 individuals. A second round of QC was carried out on the imputed data closely following steps and criteria described in Anderson et al. [11] and Purcell et al. [12]. In addition, we removed SNPs that were significant in a Cochran-Mantel-Haenszel test of SNP-phenotype associations across platforms. For case-control phenotypes, we further removed SNPs that did not pass an adaptation of a two-locus test, which controls for differential case/control calling errors [13]. Quality control resulted in ~1,140,000 SNPs for the analyses. The exact number varied slightly for the different phenotypes since the two-locus and platform association tests are phenotype specific. The number of subjects varied between $N=3245$ and $N=4240$ depending on the phenotype. For MDD we used a stringent definition of case (i.e., diagnosis of MDD) and control (i.e., no reported psychiatric illness), resulting in 1620 cases and 1625 controls. The data include a large proportion (~2/3) of the original Genetic Association Information Network (GAIN) MDD data set, which was selected for MDD. Note that these subjects were genotyped on the same platform, with MDD cases and controls randomly assigned to plates [2]. Approximately 1/3 of the $N=4240$ sample that passed qc were not selected for MDD.

For both methods, we used software provided on the respective developers' websites (<http://gump.qimr.edu.au/gcta/>, and <http://sites.google.com/site/honcheongso/software/total-vg>). Prior to applying the density estimation method we carried out LD pruning as suggested by So et al. [5], and used the kernel estimate as it is more stable in pruned data [5]. We repeated all analyses with different pruned sets and obtained very similar results, which demonstrates the robustness regarding the SNPs selected for the analysis.

Table 1 summarizes the results for all 5 phenotypes. For Yang et al.'s method likelihood ratio tests comparing models with and without the SNP effects were carried out, which were significant for all phenotypes. Additional permutation tests substantiated these results. Including 4 principal components to control for population stratification did not impact results. Note that Yang et al. [4], and also So et al. [5] pointed out that their estimates are conservative due to imperfect LD of typed with causal SNPs. Results in Table 1 are uncorrected for this underestimation, the extent of which might differ for the two methods. Although point estimates vary between the two methods, all 95% CI's overlap.

The additive genetic variance of MDD due to all SNPs was estimated at 32% (95% CI: [15.0%, 48.8%]) and 28% (95% CI: [16.5%, 39.3%]), respectively. Even when considering the lower bounds of the confidence intervals, this is a substantial proportion of the heritability estimate of 36% established in a twin study this population [14], and that agrees with numerous other twin and family studies. The estimate concerning height using Yang et al.'s method is consistent with results obtained by Yang et al. [4]. A limitation regarding the two smoking phenotypes is that ~2/3 of the sample is selected by MDD status, and both smoking phenotypes in our sample have rank correlations of ~0.24 with MDD. Height and fasting glucose have essentially zero correlation with MDD.

The results concerning MDD are in stark contrast to GWA studies that aim at detecting specific SNPs [2,3]. In GWAS of MDD, SNPs explain less than 1% of the variance [15]. Our analyses show that SNPs typed on currently available platforms contain substantial information concerning the additive genetic variance of MDD and other complex phenotypes. The failure of MDD GWA studies is, among other reasons, likely due to small effect sizes of the involved SNPs, resulting in insufficient power to discriminate between signals and false positives [16]. In addition to improving power in genome-wide analyses by increasing the number of samples in meta-analyses, novel analytical approaches including data-mining procedures can be customized to efficiently extract the information that is

present in genome-wide SNP data. Given appropriate power it should be possible to detect at least some of the locations in the genome associated with MDD, and to provide a basis for the understanding of the genetic underpinnings of this devastating disorder.

Acknowledgments

Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health; Netherlands Organization for Scientific Research (NWO: MagW/ZonMW) Addiction-31160008; 40-0056-98-9032; SPI 56-464-14192; Geestkracht (10-000-1002) and matching funds from universities and mental health care institutes involved in NESDA (GGZ Buitendamstel-Geestgronden, Rivierduinen, UMC Groningen, GGZ Lentis, GGZ Friesland, GGZ Drenthe); CMSB: Centre for Medical Systems Biology (NWO Genomics); NBIC/BioAssist/RK/2008.024; Institute for Health and Care Research (EMGO+); Neuroscience Campus Amsterdam (NCA); BBMRI -NL-184.021.007; European Science Council (230374); Rutgers University Cell and DNA Repository cooperative agreement (NIMH U24 MH068457-06); GL is supported by NIDA-DA018673 and Royal Netherlands Academy of Arts and Science ISK/6827/VPPI; JMV by VENI 451-06-004 and CM by VENI-916.76.125.

References

1. Kim Y, Zervas S, Trace SE, Sullivan PF. Schizophrenia Genetics: Where Next? *Schizophrenia Bulletin*. 2011; 37(3):456–463. [PubMed: 21505112]
2. Sullivan PF, de Geus EJ, Willemsen G, James MR, Smit JH, Zandbelt T, Arolt V, et al. Genome-wide association for major depressive disorder: A possible role for the presynaptic protein piccolo. *Mol Psychiatry*. 2009; 14(4):359–375. [PubMed: 19065144]
3. Wray NR, Pergadia ML, Blackwood DH, Penninx BW, Gordon SD, Nyholt DR, Ripke S, et al. Genome-wide association study of major depressive disorder: New results, meta-analysis, and lessons learned. *Mol Psychiatry*. 2010 epub ahead of print November 2010. 10.1038/mp.2010.109
4. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010; 42(7):565–569. [PubMed: 20562875]
5. So H-C, Li M, Sham PC. Uncovering the Total Heritability Explained by All True susceptibility Variants in a Genome-Wide Association Study. *Genet Epidemiol*. 2011; 35(6):447–456.10.1002/gepi.20593 [PubMed: 21618601]
6. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol*. 2011 Published online 3 March 2011.
7. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011; 88(1):76–82. [PubMed: 21167468]
8. Lee SH, Wray NR, Goddard ME, Visscher PM. *Am J Hum Genet*. 2011; 88(3):294–305. [PubMed: 21376301]
9. Willemsen G, Geus de EJC, Bartels M, Beijsterveldt van CEM, Brooks AI, Estourgie-van Burk GF, Fugman DA, Hoekstra C, Hottenga JJ, Kluit K, Meijer P, Montgomery GW, Rizzu P, Sondervan D, Smit AB, Spijker S, Suchiman HED, Tischfield JA, Lehner T, Slagboom PE, Boomsma DI. The Netherlands Twin Register Biobank: A Resource for Genetic Epidemiological Studies. *Twin Research and Human Genetics*. 2010; 13(3):231–245. [PubMed: 20477721]
10. Penninx BW, Beekman AT, Smit JH, Zitman FG, Nolen WA, Spinhoven P, Cuijpers P, de Jong PJ, van Marwijk HWJ, Assendelft WJJ, van der Meer K, Verhaak P, Wensing M, de Graaf R, Hoogendijk WJ, Ormel J, van Dyck R. The Netherlands Study of Depression and Anxiety (NESDA) Rationale, objectives and methods. *Int J Meth Psychiatr Res*. 2008; 17:121–140.
11. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nature protocols*. 2010; 5(9):1564–1573.
12. The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; 460:748–752. [PubMed: 19571811]
13. Lee SH, Nyholt DR, Macgregor S, Henders AK, Zondervan KT, Montgomery GW, Visscher PM. A simple and fast two-locus quality control test to detect false positives due to batch effects in genome-wide association studies. *Genet Epidemiol*. 2010; 34(8):854–862. [PubMed: 21104888]

14. Middeldorp CM, Birley AJ, Cath DC, Gillespie NA, Willemsen G, Statham DJ, de Geus EJ, et al. Familial clustering of major depression and anxiety disorders in Australian and Dutch twins and siblings. *Twin Res Hum Genet.* 2005; 8(6):609–615. [PubMed: 16354503]
15. Demirkan A, Penninx BWJH, Hek K, Wray NR, Amin N, Aulchenko YS, van Dyck R, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Mol Psychiatry.* 2010 epub ahead of print. 10.1038/mp.2010.65
16. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature.* 2009; 461:747–753. [PubMed: 19812666]
17. Vink JM, Willemsen G, Boomsma DI. Heritability of Smoking Initiation and Nicotine Dependence. *Behav Genet.* 2005; 35(4):397–406. [PubMed: 15971021]
18. Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, et al. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res.* 2003; 6:399–408. [PubMed: 14624724]

Table 1

Estimates for MDD, smoking initiation, current smoking, fasting glucose and height

	MDD	Smoking initiation	Current smoking	Fasting glucose	Height
N used for estimation methods	N=3245 N _{case} =1620 N _{control} =1625	N=4181 N _{case} =2602 N _{control} =1579	N=4181 N _{case} =1189 N _{control} =2992	N=3723	N=4199
Method Yang et al.	.32 (.086) p= 1.071×10 ⁻⁴	.19 (.087) p= 0.024	.24 (.096) p= 0.011	.22 (.059) p= 5.41×10 ⁻⁵	.42 (.052) p=0
Method So et al.	.28 (.058)	.28 (.084)	.44 (.063)	.19 (.036)	.29 (.035)
Hertability in twin and family studies	.36 [13]	.44 [17]	.79 *	.53 *	.90 [18]

Note: Variance estimates x 100 can be interpreted as percentage of explained variance. Standard errors for variance estimates are between round brackets, reference numbers are between straight brackets. The p-value corresponds to the likelihood ratio test comparing the models with and without the SNP effects, significance level is p<.05.

* =estimated for this report using data from the NTR (N_{smoke}=10,004, N_{gluc}= 3220). Prevalences of MDD, current smoking, and smoking initiation were based on age appropriate estimates for the Dutch population, namely 0.17, 0.34, and 0.50.