



Research Paper

Shared Genetics and Couple-Associated Environment Are Major Contributors to the Risk of Both Clinical and Self-Declared Depression



Yanni Zeng, MSc^{a,*}, Pau Navarro, PhD^b, Charley Xia, MSc^b, Carmen Amador, PhD^b, Ana M. Fernandez-Pujals, MSc^a, Pippa A. Thomson, PhD^{d,e}, Archie Campbell, MA^c, Reka Nagy, MSc^b, Toni-Kim Clarke, PhD^a, Jonathan D. Hafferty, MD^a, Blair H. Smith, MD^{c,f}, Lynne J. Hocking, PhD^{c,g}, Sandosh Padmanabhan, PhD^{c,h}, Caroline Hayward, PhD^b, Donald J. MacIntyre, MD^a, David J Porteous, PhD^{c,d,e}, Chris S. Haley, PhD^{b,i}, Andrew M. McIntosh, MD^{a,c,d}

^a Division of Psychiatry, University of Edinburgh, Edinburgh, UK

^b MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

^c Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

^d Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

^e Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

^f Division of Population Health Sciences, University of Dundee, Dundee, UK

^g Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

^h Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

ⁱ The Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, UK

ARTICLE INFO

Article history:

Received 12 September 2016

Received in revised form 1 November 2016

Accepted 2 November 2016

Available online 4 November 2016

Keywords:

Major depressive disorder

Self-declared depression

SNP heritability

Couple effect

Family environment

Linear mixed modeling

ABSTRACT

Background: Both genetic and environmental factors contribute to risk of depression, but estimates of their relative contributions are limited. Commonalities between clinically-assessed major depressive disorder (MDD) and self-declared depression (SDD) are also unclear.

Methods: Using data from a large Scottish family-based cohort (GS:SFHS, N = 19,994), we estimated the genetic and environmental variance components for MDD and SDD. The components representing the genetic effect associated with genome-wide common genetic variants (SNP heritability), the additional pedigree-associated genetic effect and non-genetic effects associated with common environments were estimated in a linear mixed model (LMM).

Findings: Both MDD and SDD had significant contributions from components representing the effect from common genetic variants, the additional genetic effect associated with the pedigree and the common environmental effect shared by couples. The estimate of correlation between SDD and MDD was high ($r = 1.00$, $se = 0.20$) for common-variant-associated genetic effect and lower for the additional genetic effect from the pedigree ($r = 0.57$, $se = 0.08$) and the couple-shared environmental effect ($r = 0.53$, $se = 0.22$).

Interpretation: Both genetics and couple-shared environmental effects were major factors influencing liability to depression. SDD may provide a scalable alternative to MDD in studies seeking to identify common risk variants. Rarer variants and environmental effects may however differ substantially according to different definitions of depression.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Depression has a pattern of familial aggregation, which implies the influence of genetic effects, common environmental effects shared by relatives, or both. The genetic component (heritability) has been estimated by a twin study of major depressive disorder (MDD) to be 37%

(Sullivan et al., 2000). The SNP heritability (heritability attributed to common genetic variants) of MDD varies across populations and samples (21%–32%) (Lubke et al., 2012; Lee et al., 2013). Subsequently, a ‘children of twins’ study found a significantly greater risk of depression in children of depressed monozygotic (MZ) twins than in the offspring of their non-depressed twin. This implies a potential environmental effect of parental depression on offspring (Singh et al., 2011). Studies have also shown that having a partner with psychiatric disorder may increase an individual's risk of MDD (Joutsenniemi et al., 2011; Desai et al., 2012), but meta-analytic studies suggest no effect of the shared sibling

* Corresponding author at: Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK, EH10 5HF.
E-mail address: Y.Zeng-6@sms.ed.ac.uk (Y. Zeng).

environment and other studies have postulated more complex relationships (Ollino et al., 2006). Whilst each of these studies separately provided evidence for the genetic and familial environmental components in depression, a precise separation of these potential effects should involve estimating them simultaneously in the same model and has yet to be achieved.

The accurate separation and estimation of the genetic and environment components on liability to depression provide crucial information, as it reveals the upper limit of the genetic effects, the probability of true positive results from genetic studies and the potential for accurate risk predictions for depression (Makowsky et al., 2011; Tenesa and Haley, 2013). Genetic studies attempting to map causal variants have been performed for various definitions of depression. These include clinically-assessed depression, self-report of clinical diagnosis of depression and self-reported depressive symptoms (consortium, 2015; Major Depressive Disorder Working Group of the Psychiatric et al., 2013; Hyde et al., 2016; Okbay et al., 2016), but the findings are generally inconsistent. Although some of the inconsistent findings were probably due to the limited power of the original studies (Flint and Kendler, 2014), there may also be intrinsic heterogeneity across depression definitions. This is further supported by the fact that studies show very different estimates of the narrow-sense heritability (h_n^2) for several depression definitions or related traits to MDD ($h_n^2 = 37%$ (Sullivan et al., 2000)); perceived stress: 44% (Bogdan and Pizzagalli, 2009); nine depression definitions in women: 21%–45% (Kendler et al., 1992); depressive symptom scores in childhood: 79% (Thapar and McGuffin, 1994) and depressive symptoms in an elderly population: 69% in women and 64% in men (McGue and Christensen, 2003). In fact, even for MDD, the genetic correlation of MDD phenotypes between independent datasets was relatively low compared with other psychiatric disorders (Gratten et al., 2014).

Because of the heterogeneity across depression definitions, there has been a long debate about the correct phenotype for depression genetic studies. Studies using clinically-assessed depression could provide findings that are directly informative for clinical application. However, the resources required for such data collection are generally very high (consortium, 2015). As an alternative, measuring self-reported depression requires fewer resources and this phenotype is rapidly becoming available for many population-based datasets (Okbay et al., 2016; Hyde et al., 2016). To date, the largest published GWAS of major depression has yielded 15 significant loci (7 loci before meta-analysis) for a self-reported clinical diagnosis (Hyde et al., 2016).

Given that important progress is being made from GWASs on different depression definitions, it becomes increasingly important to understand the similarities and dissimilarities across different definitions. Particularly, the difference in genetic and environmental loadings might underpin the inconsistent results from genetic studies across different depression definitions. Therefore, dissecting the phenotypic variance of each depression definition and understanding the similarities and dissimilarities between those phenotypes in the context of both genetic and common environmental components is particularly important for interpreting the results from published depression studies and for informing about genetically relevant depression phenotypes for future studies.

In this study we sought to partition the phenotypic variation of the diagnosed depression (MDD) and the self-declared depression (SDD) into its genetic and familial environment components using Linear Mixed Modeling (LMM). Specifically, we utilized data from Generation Scotland: Scottish Family Health Study (GS:SFHS), a large Scottish cohort with extensive family relationship information and genome-wide genotype data to answer two questions. First, when simultaneously considering multiple genetic effects and familial environmental effects in the model, what are major contributions to variation in MDD and SDD, respectively? Second, what is the contribution of each of the identified major contributing components to the overall correlation between MDD and SDD?

2. Materials and Methods

The Tayside Research Ethics Committee (reference 05/S1401/89) provided ethical approval for the study. In GS: SFHS, participants gave written consent, after having an opportunity to discuss the project, and before any data or samples were collected.

The details of their consent status are recorded in the study database. All consent forms and study protocols were approved by the Research Ethics Committee.

2.1. Datasets

Generation Scotland: Scottish Family Health Study (GS: SFHS) contains 21,387 subjects ($N_{\text{male}} = 8772$, $N_{\text{female}} = 12,615$; $\text{Age}_{\text{mean}} = 47.2$), who were recruited from the registers of collaborating general practices. At least one first-degree relative aged 18 or over was required to be identified for each participant (Smith et al., 2006). Genotyping data were generated using the Illumina Human OmniExpressExome -8- v1.0 array (Gunderson, 2009). Details of genotyping are described elsewhere (Smith et al., 2006). Population outlier individuals were removed from the sample (Amador et al., 2015). Quality control (QC) of genotyped SNPs used inclusion thresholds: missing SNPs per individual $\leq 2\%$, SNP genotype call rate $\geq 98\%$, minor allele frequency (MAF) $> 1\%$ and Hardy-Weinberg equilibrium P value $> 1 \times 10^{-6}$. In total, 561,125 genotyped autosomal SNPs passed QC criteria and were available for 19,994 participants ($N_{\text{male}} = 8221$, $N_{\text{female}} = 11,773$, $\text{Age}_{\text{mean}} = 47.4$).

2.1.1. Phenotypes

Lifetime Diagnosis of MDD: The Structured Clinical Interview for DSM-IV(SCID) was used (First et al., 2012): participants who screened positive (21.7%) for the questions “Have you ever seen anybody for emotional or psychiatric problems? IF YES: What was that for? (What treatment(s) did you get? Any medications?) IF NO: Was there ever a time when you, or someone else, thought you should see someone because of the way you were feeling or acting” were invited to continue to an interview using the SCID modules for mood disorders (First et al., 2002). Participants who screened positive but refused to undergo the structured clinical interview ($N = 507$, 2.4%) and those with a diagnosis of bipolar disorder ($N = 76$) were excluded from the study. More details of phenotyping procedures are described elsewhere (Fernandez-Pujals et al., 2015).

Self-declared depression (SDD): the participants were invited to answer the following question “please mark an X in the box if you have been affected by depression”.

2.2. Partitioning the Phenotypic Variation

Based on the framework of the Genomic-relationship-matrix restricted maximum likelihood (GREML) method, Xia et al. (2016) developed a method to estimate h_g^2 (proportion of additive genetic variance contributed by common genetic variants over the total phenotypic variance, namely SNP heritability), h_p^2 (representing the additional additive genetic effect contributed by pedigree associated variation), h_n^2 (proportion of the total additive genetic variance over the total phenotypic variance, namely narrow-sense heritability) and a number of familial environmental components simultaneously (Xia et al., 2016). This was performed by fitting variance-covariance matrices representing common genetic effects, pedigree-related-genetic effects, and current and past family environmental effects simultaneously in the mixed linear model (Xia et al., 2016), building on previous work by Zaitlen et al. (Zaitlen et al., 2013). This approach enables estimation of the contribution of each genetic and family environmental component and here we applied it to MDD and SDD.

In detail, for each trait, two genomic relationship matrices, \mathbf{G} (genomic relationship matrix) and \mathbf{K} (kinship matrix created by modifying \mathbf{G}

using a threshold of 0.05 for pairwise relatedness) (Xia et al., 2016; Zaitlen et al., 2013), and three environment relationship matrices, **F** (environmental matrix representing nuclear-family-member relationships), **S** (environmental matrix representing full-siblings relationships) and **C** (environmental matrix representing couple relationships) (Fig. 1, Text s1) (Xia et al., 2016), were fitted separately or simultaneously in LMM (Text s2). The corresponding variance components h_g^2 (common-variant-associated genetic effect, represented in **G**), h_p^2 (additional genetic effect from pedigree, represented in **K**), e_f^2 (environmental effect from nuclear family, represented in **F**), e_s^2 (environmental effect from full sibling relationship, represented in **S**) and e_c^2 (environmental effect from couple relationship, represented in **C**) were estimated in LMM and their significance was tested using LRT and Wald tests (Text s2). The estimates on the observed scale were transformed to the liability scale using the population prevalence of each trait (MDD: 0.13; SDD: 0.09). Age, age², sex and the top 20 Principal components were included in the LMM as fixed effects. To simplify the model description, the following codes were used to represent the matrices fitted in the models: -e.g. ‘**GKFS**’ was the full model which fits all five matrices as random effects simultaneously, and ‘**GKC**’ represents the model where the genomic relationship matrix, the kinship relationship matrix and the environmental matrix representing couple relationships were fitted simultaneously.

2.2.1. Stepwise Model Selection for Identifying Major Contributing Components.

2.2.1.1. Backward Stepwise Selection (Xia et al., 2016). The selection started with the full model ‘**GKFS**’. LRT and Wald tests were conducted for each variance component. A variance component was removed from the model if (1) it failed to obtain significance (at the 5% level) in both tests and (2) among the variance components satisfying (1), it had the highest *P* value in the Wald test. This process was performed repeatedly until all the remaining components obtained significance in at least one test.

2.2.1.2. Forward Stepwise Selection. To check for the agreement of different model selection strategies, we additionally implemented a forward stepwise regression approach for model selection. In this case the selection started with a null model (without fitting any matrices). Matrices were then added into the model one at a time. LRT and Wald tests were conducted for each variance component. A variance component was added if (1) it obtained significance (at the 5% level) in both tests, (2) the addition of this component did not lead to the variance components already in the model becoming non-significant in both tests and (3) among the variance components satisfying both (1) and (2), it had

the lowest *P* value in the Wald test. This process was repeated until no more variance components satisfied criteria (1), (2) and (3).

2.2.2. Bivariate GREML Analysis for MDD and SDD

Using the selected model (**GKC**) obtained from the above analyses, we then estimated the relative contributions of each major contributing genetic and environmental component to the correlation between MDD and SDD. We used the GCTA-bivariate GREML analysis (Yang et al., 2011; Lee et al., 2012) to estimate the correlation between the two traits for the common-variant-associated genetic component, the pedigree-associated genetic component and the shared couple environment component simultaneously. Each estimated correlation was tested to determine the significance of its difference from both 0 and 1 by Log-likelihood test (LRT) and Wald test.

3. Results

In GS:SFHS, among the 19,994 participants with genome-wide genotyping data, we recognized 1742 pairs of couples, 8458 pairs of full siblings and 20,019 pairs of members living in the same nuclear family. In this dataset, 99.5% (19,896/19,994) of participants have MDD diagnosis information (2659 MDD cases and 17,237 controls) and 98% (19,603/19,994) of participants have answered the question for self-reported depression (1940 SDD cases and 17,663 controls) (the cross tabulation of MDD and SDD is given in Table s1).

3.1. The Full Model Partitions Phenotypic Variation into Genetic and Familial Environmental Components

A full model was first utilized to partition the phenotypic variation of each trait into five potential sources of influence: the additive genetic effect contributed by common variants (h_g^2), the additional additive genetic effect associated with pedigree (h_p^2), the environmental effect shared between all members of a nuclear family (e_f^2), the environmental effect shared between members of a couple (e_c^2) and the environmental effect shared between full siblings (e_s^2) (Fig. 1). The results are presented in Table 1. For MDD, 10% (se = 5%) of the phenotypic variance is attributable to the common genetic variants (h_g^2) and 20% (se = 12%) is to the additional genetic variation associated with pedigree (h_p^2). For SDD, 22% (se = 7%) of the phenotypic variance is attributable to the common genetic variants (h_g^2) and 50% (se = 15%) is attributable to the additional genetic variation associated with pedigree (h_p^2). The proportion of total additive genetic determinant (narrow-sense heritability: $h_n^2 = h_g^2 + h_p^2$) was 30% for MDD and 72% for SDD. For the three familial environmental components, the likelihood ratio test (LRT) for the couple-shared environmental effect was significant for SDD (17% (se = 10%)), but was not

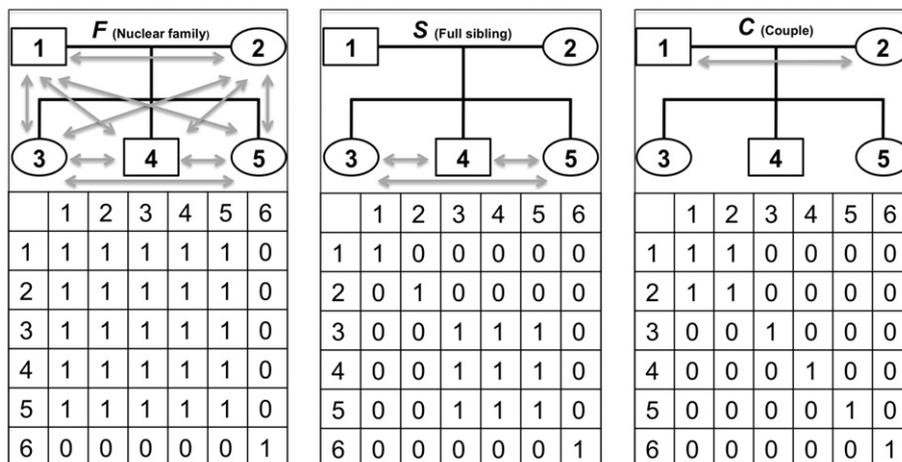


Fig. 1. The design of the environment relationship matrices.

Table 1
Variance component analyses results for MDD and SDD.

Trait	Description Model		G (Common variants-associated genetic)	K (Pedigree-associated genetic)	F (Nuclear family)	S (Full sibling)	C (Couple)
			$h_g^2(\text{se})$	$h_p^2(\text{se})$	$e_f^2(\text{se})$	$e_s^2(\text{se})$	$e_c^2(\text{se})$
MDD	Genetics only	GK	0.12(0.05)	0.34(0.06)			
	Full	GKFSC	0.10(0.05)	0.20(0.12)	0.09(0.06) ^{ns}	0.00(0.04) ^{ns}	0.03(0.09) ^{ns}
	Forward selection	GKC	0.12(0.05)	0.35(0.06)			0.14(0.07)
	Backward selection	GF	0.15(0.04)		0.16(0.03)		
SDD	Genetics only	GK	0.24(0.06)	0.52(0.07)			
	Full	GKFSC	0.22(0.06)	0.50(0.15)	0.04(0.07) ^{ns}	0.00(0.04) ^{ns}	0.17(0.10)
	Forward selection	GKC	0.24(0.06)	0.53(0.07)			0.22(0.07)
	Backward selection	GKC	0.24(0.06)	0.53(0.07)			0.22(0.07)

Variance component analyses were performed on MDD and SDD using the genetic model (**GK**), the model accounting for all of the two genetic and three familial environmental effects (the full model **GKFSC**), and the models selected by forward or backward selection. NS: the variance component was not significant in LRT test.

significant for MDD (although the estimated effect was non-zero: 3% (se = 9%)). The environmental effects shared between nuclear family members and full-siblings were not significant for either trait with the estimate of the sib effect being zero for both traits but for the family effect being non-zero for both traits. Compared with the a reduced model that does not account for environmental effects (the **GK** model), the full model obtains lower estimates of the genetic components for both traits, suggesting that the full model effectively reduced the confounding environmental effects when estimating the heritability (Fig. 2, Table 1).

3.2. Stepwise Model Selections Identify Major Genetic/Familial-Environment Contributors for MDD and SDD

As shown in previous work, the full model, although accounting for all of the five potentially influential effects, may have difficulty in separating major contributors to depression from minor contributors because of correlated effects (Xia et al., 2016). In order to address this problem (Xia et al., 2016), we applied stepwise model selection (Xia et al., 2016) to identify major contributors to variation in the two depression traits.

Using forward stepwise selection, the common variant-associated-genetic effect, pedigree-associated-genetic effect and couple-shared environmental effects were retained in the final model for both MDD and SDD (the **GKC** model as shown in Table s2 and Table 1). Using backward

stepwise selection, for MDD, only the common variant-associated-genetic component and shared-nuclear-family component were retained in the final model (the **GF** model as shown in Table s3a). For SDD, as for the forward stepwise approach, common variant-associated-genetic, pedigree-associated-genetic effect and couple-shared environmental effects were selected (the **GKC** model as shown in Table s3b). The relative contribution of each variance component to SDD and MDD in the **GKC** model is shown in Fig. 3.

3.3. Bivariate GREML Analyses Estimate the Genetic and the Environmental Correlations between MDD and SDD

Since the model that included common-variant- and pedigree-associated genetic and couple-shared environmental effects (the **GKC** model) was selected as the most parsimonious model in three out of the four model selections for the two traits. We further explored contributions of these three components to the correlation between MDD and SDD. The phenotypic correlation between MDD and SDD was 0.45 (Phi coefficient, $P < 2.2 \times 10^{-16}$). The estimate of the common-variant-associated genetic correlation w_{ssas} 1.00 (se = 0.20), this estimate being significantly different from 0 ($P_{Wald_H0:r=0} < 1 \times 10^{-5}$, $P_{LRT_H0:r=0} = 2.1 \times 10^{-5}$) and not significantly different from 1 ($P_{Wald_H0:r=1} = 1$, $P_{LRT_H0:r=1} = 0.5$). The estimate of the pedigree-associated genetic correlation was 0.57 (se = 0.08, $P_{Wald_H0:r=0} < 1 \times 10^{-5}$, $P_{Wald_H0:r=1} < 1 \times 10^{-5}$, $P_{LRT_H0:r=0} = 7.7 \times 10^{-10}$, $P_{LRT_H0:r=1} = 7.1 \times 10^{-7}$),

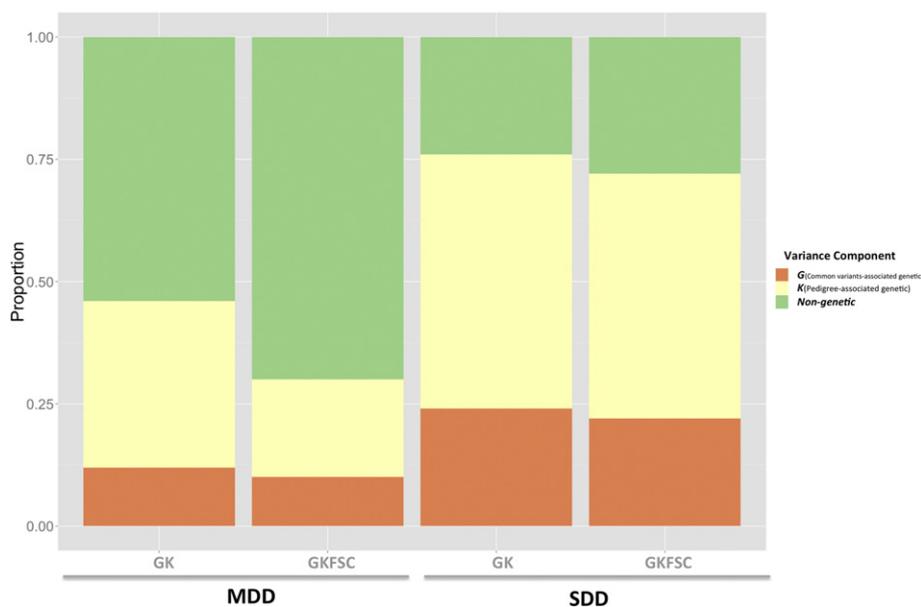


Fig. 2. Comparison of the proportion of variance explained by the common-variants- and pedigree-associated genetic components estimated in the model that only account for the two genetic components (the **GK** model) and in the full model that accounts for two genetics and three shared-environmental effects (the **GKFSC** model). SDD: self declared depression, MDD: major depressive disorder.

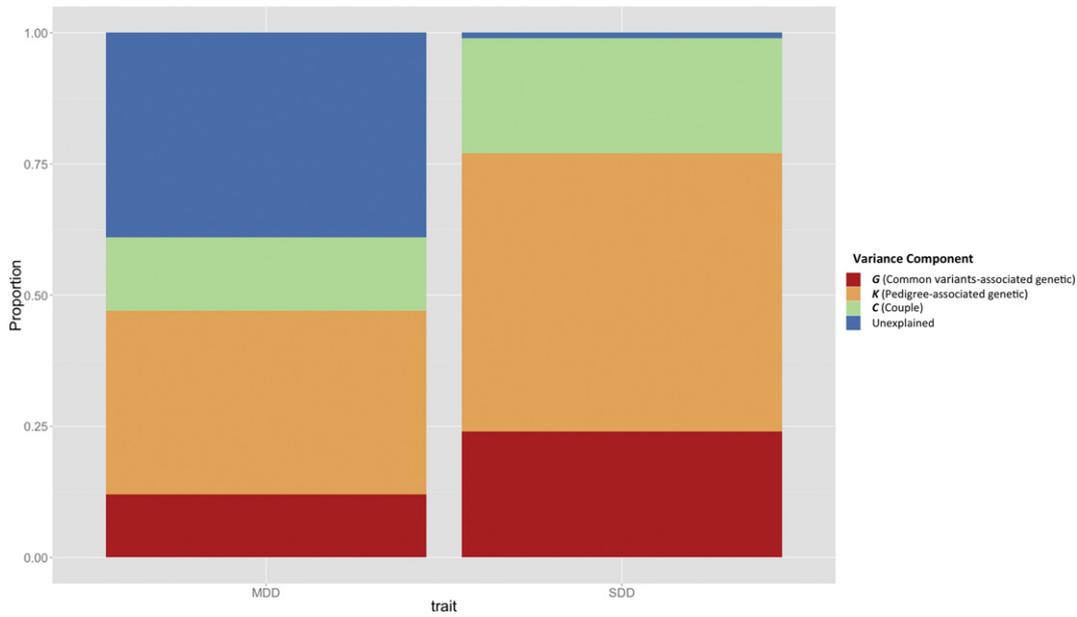


Fig. 3. Sources of phenotypic variance and the proportion of variance they explained in the most parsimonious model (**GKC**) for both SDD and MDD. SDD: self declared depression, MDD: major depressive disorder.

and the estimate of the couple-shared environmental correlation was 0.52 (se = 0.22, $P_{Wald_H0:r=0} = 0.016$, $P_{Wald_H0:r=1} = 0.03$, $P_{Irr_H0:r=0} = 0.07$, $P_{Irr_H0:r=1} = 0.06$) (Table 2). Thus these two latter estimated correlations were significantly different from both zero and one in the Wald test.

4. Discussion

This variance-component analysis study assessed five genetic and familial environment risk contributions to MDD and SDD using GS:SFHS, a Scottish sample comprising close and distant relatives with genome-wide genotyping data. We showed that the common variant- and pedigree- associated genetics and the couple-shared environmental effects are major factors influencing liability to both clinically assessed and self-reported depression among the factors considered. The estimated correlation between SDD and MDD was very high ($r = 1.00$, se = 0.20) for common variant-associated genetic effects and lower for the pedigree-associated genetic ($r = 0.57$, se = 0.08) and couple-shared environmental effects ($r = 0.52$, se = 0.22).

The estimates of both common-variant-associated and pedigree-associated genetic effects are much lower for MDD ($h_g^2 = 10\%$ (5%), $h_p^2 = 20\%$ (12%)) than for SDD ($h_g^2 = 22\%$ (6%), $h_p^2 = 50\%$ (15%)), suggesting a difference between clinical and self-reported depression definitions. MDD is diagnosed through clinical questionnaire, whereas SDD reflects both depression status and the participants' self-awareness, this difference may underlie the higher heritability in SDD. The point estimate of SNP heritability h_g^2 for MDD ($h_g^2 = 10\%$ (5%)) is lower than that from a mega-analysis of 9 cohorts of European ancestry (21%(2%)) (Lee et al., 2013). This may be due to the intrinsic heterogeneity of MDD

(Major Depressive Disorder Working Group of the Psychiatric et al., 2013). The pedigree-associated genetic component h_p^2 measures the additional genetic effects co-segregating in the pedigree (i.e. those not associated with common genetic variants), such as the effect of rare and structural variants. Using the full model which partitions the narrow sense heritability h_n^2 into the common-variant-associated component h_g^2 and the pedigree-associated component h_p^2 , and accounts for multiple familial environmental effects simultaneously, the estimated h_g^2 accounted for around one third of h_n^2 in both MDD and SDD. This is lower than previous estimates for other complex traits (excluding depression) that suggest that >50% of h_n^2 is accounted by h_g^2 (Zaitlen et al., 2013; Xia et al., 2016). Here h_p^2 accounted for more than two thirds of h_n^2 , suggesting an important role for rare and structural variants in both clinical- and self-declared depression. The full model included a number of correlated matrices, potentially impeding model-fitting and their estimation (the discussion for collinearity is Text s3, Tables s4 and s5). A greater discriminating power may however have been achievable with larger sample sizes (Xia et al., 2016). In addition, the power to detect the effect might vary among tested components, therefore a caution should be made when compare the estimates of those components.

Stepwise model selections suggested that the two genetic effects and the couple-shared environmental effects were the most significant contributors to risk of depression among the factors considered (**GKC** model was selected by both forward and backward selections for SDD and backward selection for MDD), although there is inconsistency in the results between the backward and the forward selections for MDD. This is probably due to the high correlation between ERM_{family} and the combined ERM_{couple} and GRM_{kin} (Xia et al., 2016). Simulation analysis of the backward selection method suggested that although it successfully selected the appropriate model with all major components of simulated phenotype in >80% of cases, in 20% of cases the model either selected the **F** component when the true components were **C** plus **K** or the model selected **C** plus **K** but the true component was **F** (Xia et al., 2016). Given that the pedigree-associated genetic component (**K**), a component that has been shown to be a major contributor of the genetic effect by the **GK** and the full model **GKFSK** in both MDD and SDD, was excluded by backward selection for MDD, it is likely we met the same problem as observed in simulation analysis by Xia et al. and the selected **GF** model is unlikely to contain all major resources of variation for MDD. As shown in simulation analysis, we expect that a

Table 2
Bivariate GCTA-GREML estimates of the correlation of each variance component between SDD and MDD using **GKC** model.

	Test	MDD	SDD
G (Common variants-associated genetic)	h_g^2	0.10(0.05)	0.24(0.06)
	r_g	1.00(0.20)	
K (Pedigree-associated genetic)	h_p^2	0.36(0.06)	0.52(0.07)
	r_p	0.57(0.08)	
C (Couple)	e_c^2	0.13(0.07)	0.22(0.07)
	r_c	0.52(0.22)	

larger sample size may provide sufficient power for a higher accuracy and stability of backward model selection (Xia et al., 2016). In addition, the estimates in the full model for MDD and SDD also suggest that components omitted from **GKC** model are likely to contribute only small amounts of variance, therefore **GKC** model was chosen as the best model for both traits in GS:SFHS. The significant contributions from the **G**, **K** and **C** should be further replicated in independent samples.

In total, the selected **GKC** model explained 60% ($se = 8\%$) and 98% ($se = 9\%$) of the phenotypic variation in MDD and SDD, respectively (Fig. 3). Strikingly, the effect of the shared couple environment contributed as much as 14% ($se = 7\%$) and 22% ($se = 7\%$) of the phenotypic variance in MDD and SDD respectively. For GS:SFHS, an adult cohort with a minimum participant's age of 18, the couple effect reflects the current family environment shared between couples, which is contrast to the full sibling effect which reflects the influence of earlier shared environments. The role of the couple effect has previously been suggested in a Finnish study which showed that the partner's MDD status was associated with the MDD risk in non-psychiatric subjects (Joutsenniemi et al., 2011). Our results provided additional evidence for couple-associated effect and indicated that the magnitude of this effect is likely to be high. These results also suggest that the inclusion of partners in genetic studies of depression, whilst logistically attractive, might introduce confounding if additional adjustment is not made. It may be helpful for future genetic studies to be aware of these potential biases and to either avoid the recruitment of couples, or model their effects appropriately. In clinical practice, the mental health status of the spouse should also be considered as an additional indication of risk.

The couple-shared environmental effect detected in depression traits could be confounded by the non-random mating in those phenotypes. For example, assortative mating has been observed in multiple psychiatric disorders including depression. For SDD, **G**, **K** and **C** explained a very high proportion of phenotypic variance. However assortative mating may also contribute to the fact that the total variance explained is so high. When assortative mating exists in the trait of interest, h_g^2 , h_n^2 and e_c^2 estimates may be biased if this effect is not accounted for appropriately (Keller et al., 2013; Xia et al., 2016). The magnitude of this bias in heritability estimates, however, is likely to be small for diseases with high prevalence, such as MDD (Peyrot et al., 2016). On the other hand, ignoring the couple-shared environment effect in the study of assortative mating may also lead to inaccurate measurement of the degree of assortative mating. In GS:SFHS we did not detect significant genome-wide genotype-level assortative mating (Text s4, Figure s1) although the power of this test is very limited. The investigation of the genetic assortative mating in MDD- and SDD- associated loci has been impeded by the limited knowledge of the location of the loci that affect the trait (Charley Xia et al., personal communication). Further studies providing the age of marriage with large enough sample sizes to separate young and old couples would facilitate the discrimination of assortative mating from the effects of couple-shared environment.

Finally, we estimated the genetic and environment correlation between MDD and SDD. The results revealed a very high correlation in the common-variant-associated genetic component, a moderate correlation in the pedigree-associated genetic component and a less significant correlation in the couple-shared environment component. This suggests that there are strong genetic similarities between the depression phenotypes for common variants. In contrast, pedigree-associated genetic variation and shared environment may underpin important differences between these traits. This has important implications for the design of future molecular studies of depression in which SDD may be a good proxy phenotype for MDD in studies seeking to identify common risk variants. However, for family-based studies where the targeted genetic effect is from rare variants, the lower genetic correlation may impede the use of SDD as a proxy for MDD. The estimate of the correlation of the couple-shared environmental component has a wide confidence interval, which is likely due to the relatively limited number of couples in GS:SFHS. Therefore, replication of these findings in independent

depression studies is indicated. In addition, since participants have multiple records (MDD and SDD), there might be other shared common environmental effects that are not specifically tested in current model but could be partitioned as one of the components already tested in the model (e.g. the correlation for **G**, **K** or **C** could be inflated). Study designs using two independent datasets to infer the genetic correlation (such as LD-score regression) can be free of such shared environmental confounding effects.

5. Conclusions

This study showed that the common-variant- and pedigree-associated genetics and the couple-shared environmental effects are major factors influencing liability to depression. This suggests that the depression status of a spouse should be also treated as an indication of risk of depression. A high correlation of the common-variant-associated genetic component between SDD and MDD suggests that SDD could provide a scalable alternative to MDD in studies targeting common variants.

Funding Sources

This work is supported by the Wellcome Trust through a Strategic Award, reference 104036/Z/14/Z. GS:SFHS was funded by a grant from the Scottish Government Health Department, Chief Scientist Office, number CZD/16/6. The authors acknowledge with gratitude the financial support received for this work from the Dr. Mortimer and Theresa Sackler Foundation. PAT, DJP and AMM are members of The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged by PN and CSH (BB/J004235/1). DJM is an NRS Fellow, funded by the CSO.

Conflict of Interest

AMM has previously received grant support from Pfizer, Lilly and Janssen. These studies are not connected to the current investigation. Other authors have no conflicts of interest.

Authors' Contributions

Conceived and designed the experiments: YNZ AMM CSH PN DJP SPH BHS LJH AC. Performed the experiments: CH AC RN DJM AMFP. Analyzed the data: YNZ. Contributed reagents/materials/analysis tools: CX CSH PN YNZ AMM CA AMFP TKC. Collecting samples: BHS LJH SP DJM DJP AMM CSH. Wrote the paper: YNZ CSH AMM PN CA TKC JDH PAT and all the other authors. Raised funding: AMM CSH DJP.

Acknowledgements

We are grateful to the families who took part in GS:SFHS, the GPs and Scottish School of Primary Care for their help in recruiting them, and the whole GS team, which includes academic researchers, clinic staff, laboratory technicians, clerical workers, IT staff, statisticians and research managers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ebiom.2016.11.003>.

References

- Amador, C., Huffman, J., Trochet, H., Campbell, A., Porteous, D., Wilson, J.F., Hastie, N., Vitart, V., Hayward, C., Navarro, P., Haley, C.S., Scotland, G., 2015. Recent genomic heritage in Scotland. *BMC Genomics* 16.

- Bogdan, R., Pizzagalli, D.A., 2009. The heritability of hedonic capacity and perceived stress: a twin study evaluation of candidate depressive phenotypes. *Psychol. Med.* 39, 211–218.
- consortium, C., 2015. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* 523, 588–591.
- Desai, S., Schimmack, U., Jidkova, S., Bracke, P., 2012. Spousal similarity in depression: a dyadic latent panel analysis of the panel study of Belgian households. *J. Abnorm. Psychol.* 121, 309–314.
- Fernandez-Pujals, A.M., Adams, M.J., Thomson, P., McKechnie, A.G., Blackwood, D.H., Smith, B.H., Dominiczak, A.F., Morris, A.D., Matthews, K., Campbell, A., Linksted, P., Haley, C.S., Deary, I.J., Porteous, D.J., Macintyre, D.J., McIntosh, A.M., 2015. Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* 10, e0142197.
- First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B., 2002. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, patient ed. (SCID-I/P).
- First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B., 2012. Structured clinical interview for DSM-IV® Axis I Disorders (SCID-I). Clinician Version, *Administration Booklet*. American Psychiatric Pub.
- Flint, J., Kendler, K.S., 2014. The genetics of major depression. *Neuron* 81, 484–503.
- Gratten, J., Wray, N.R., Keller, M.C., Visscher, P.M., 2014. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat. Neurosci.* 17, 782–790.
- Gunderson, K.L., 2009. Whole-genome genotyping on bead arrays. *Methods Mol. Biol.* 529, 197–213.
- Hyde, C.L., Nagle, M.W., Tian, C., Chen, X., Paciga, S.A., Wendland, J.R., Tung, J.Y., Hinds, D.A., Perlis, R.H., Winslow, A.R., 2016. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* 48 (9), 1031–1036.
- Joutsenniemi, K., Moustgaard, H., Koskinen, S., Ripatti, S., Martikainen, P., 2011. Psychiatric comorbidity in couples: a longitudinal study of 202,959 married and cohabiting individuals. *Soc. Psychiatry Psychiatr. Epidemiol.* 46, 623–633.
- Keller, M.C., Garver-Apgar, C.E., Wright, M.J., Martin, N.G., Corley, R.P., Stallings, M.C., Hewitt, J.K., Zietsch, B.P., 2013. The genetic correlation between height and IQ: shared genes or assortative mating? *PLoS Genet.* 9, e1003451.
- Kendler, K.S., Neale, M.C., Kessler, R.C., Heath, A.C., Eaves, L.J., 1992. A population-based twin study of major depression in women. The impact of varying definitions of illness. *Arch. Gen. Psychiatry* 49, 257–266.
- Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., Absher, D., Agartz, I., Akil, H., Amin, F., Andreassen, O.A., Anjorin, A., Anney, R., Anttila, V., Arking, D.E., Asherson, P., Azevedo, M.H., Backlund, L., Badner, J.A., Bailey, A.J., Banaschewski, T., Barchas, J.D., Barnes, M.R., Barrett, T.B., Bass, N., Battaglia, A., Bauer, M., Bayes, M., Bellivier, F., Bergen, S.E., Berrettini, W., Betancur, C., Bettecken, T., Biederman, J., Binder, E.B., Black, D.W., Blackwood, D.H.R., Bloss, C.S., Boehnke, M., Boomsma, D.I., Breen, G., Breuer, R., Bruggeman, R., Cormican, P., Buccola, N.G., Buitelaar, J.K., Bunney, W.E., Buxbaum, J.D., Byerley, W.F., Byrne, E.M., Caesar, S., Cahn, W., Cantor, R.M., Casas, M., Chakravarti, A., Chambert, K., Choudhury, K., Cichon, S., Cloninger, C.R., Collier, D.A., Cook, E.H., Coon, H., Cormand, B., Corvin, A., Coryell, W.H., Craig, D.W., Craig, I.W., Crosbie, J., Cuccaro, M.L., Curtis, D., Czamara, D., Datta, S., Dawson, G., Day, R., De Geus, E.J., Degenhardt, F., Djurovic, S., Donohoe, G.J., Doyle, A.E., Duan, J.B., Dudbridge, F., Duketin, E., Ebstein, R.P., Edenberg, H.J., Elia, J., Ennis, S., Etain, B., Fanous, A., Farmer, A.E., Ferrier, I.N., Flickinger, M., Fombonne, E., Foroud, T., Frank, J., Franke, B., Fraser, C., et al., 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* 45 (984–+).
- Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M., Wray, N.R., 2012. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 28, 2540–2542.
- Lubke, G.H., Hottenga, J.J., Walters, R., Laurin, C., de Geus, E.J., Willemsen, G., Smit, J.H., Middeldorp, C.M., Penninx, B.W., Vink, J.M., Boomsma, D.I., 2012. Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biol. Psychiatry* 72, 707–709.
- Major Depressive Disorder Working Group of the Psychiatric, C., G., Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., Byrne, E.M., Blackwood, D.H., Boomsma, D.I., Cichon, S., Heath, A.C., Holsboer, F., Lucae, S., Madden, P.A., Martin, N.G., McGuffin, P., Muglia, P., Nothen, M.M., Penninx, B.P., Pergadia, M.L., Potash, J.B., Rietschel, M., Lin, D., Muller-Myhsok, B., Shi, J., Steinberg, S., Grabe, H.J., Lichtenstein, P., Magnusson, P., Perlis, R.H., Preisig, M., Smoller, J.W., Stefansson, K., Uher, R., Kutalik, Z., Tansey, K.E., Teumer, A., Viktorin, A., Barnes, M.R., Bettecken, T., Binder, E.B., Breuer, R., Castro, V.M., Churchill, S.E., Coryell, W.H., Craddock, N., Craig, I.W., Czamara, D., De Geus, E.J., Degenhardt, F., Farmer, A.E., Fava, M., Frank, J., Gainer, V.S., Gallagher, P.J., Gordon, S.D., Goryachev, S., Gross, M., Guipponi, M., Henders, A.K., Herms, S., Hickie, I.B., Hoefels, S., Hoogendijk, W., Hottenga, J.J., Iosifescu, D.V., Ising, M., Jones, I., Jones, L., Jung-Ying, T., Knowles, J.A., Kohane, I.S., Kohli, M.A., Korszun, A., Landen, M., Lawson, W.B., Lewis, G., Macintyre, D., Maier, W., Mattheisen, M., McGrath, P.J., McIntosh, A., McLean, A., Middeldorp, C.M., Middleton, L., Montgomery, G.M., Murphy, S.N., Nauck, M., Nolen, W.A., Nyholt, D.R., O'Donovan, M., Oskarsson, H., Pedersen, N., Scheftner, W.A., Schulz, A., Schulze, T.G., Shyn, S.I., Sigurdsson, E., Slager, S.L., et al., 2013c. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* 18, 497–511.
- Makowsky, R., Pawajski, N.M., Klimentidis, Y.C., Vazquez, A.I., Duarte, C.W., Allison, D.B., de los Campos, G., 2011. Beyond missing heritability: prediction of complex traits. *PLoS Genet.* 7, e1002051.
- McGue, M., Christensen, K., 2003. The heritability of depression symptoms in elderly Danish twins: occasion-specific versus general effects. *Behav. Genet.* 33, 83–93.
- Okbay, A., Baselmans, B.M.L., De Neve, J.-E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F.W., Linner, R.K., Rietveld, C.A., Derringer, J., Gratten, J., Lee, J.J., Liu, J.Z., de Vlaming, R., Ahluwalia, T.S., Buchwald, J., Cavadi, A., Frazier-Wood, A.C., Furlotte, N.A., Garfield, V., Geisel, M.H., Gonzalez, J.R., Haitjema, S., Karlsson, R., van der Laan, S.W., Ladwig, K.-H., Lahti, J., van der Lee, S.J., Lind, P.A., Liu, T., Matteson, L., Mihailov, E., Miller, M.B., Minica, C.C., Nolte, I.M., Mook-Kanamori, D., van der Most, P.J., Oldmeadow, C., Qian, Y., Raitakari, O., Rawal, R., Realo, A., Rueedi, R., Schmidt, B., Smith, A.V., Stergiakouli, E., Tanaka, T., Taylor, K., Thorleifsson, G., Wedenoja, J., Wellmann, J., Westra, H.-J., Willems, S.M., Zhao, W., LifeLines Cohort, S., Amin, N., Bakshi, A., Bergmann, S., Bjornstodt, G., Boyle, P.A., Chorney, S., Cox, S.R., Davies, G., Davis, O.S.P., Ding, J., Direk, N., Eibich, P., Emeny, R.T., Fatemifar, G., Faul, J.D., Ferrucci, L., Forstner, A.J., Gieger, C., Gupta, R., Harris, T.B., Harris, J.M., Holliday, E.G., Hottenga, J.-J., De Jager, P.L., Kaakinen, M.A., Kajantie, E., Karhunen, V., Kolcic, I., Kumari, M., Launer, L.J., Franke, L., Li-Gao, R., Liewald, D.C., Koini, M., Loukola, A., Marques-Vidal, P., Montgomery, G.W., Mosing, M.A., Paternoster, L., Pattie, A., Petrovic, K.E., Pulkki-Raback, L., Quaye, L., Raikonen, K., Rudan, I., et al., 2016. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* 48, 624–633.
- Olino, T.M., Lewinsohn, P.M., Klein, D.N., 2006. Sibling similarity for MDD: evidence for shared familial factors. *J. Affect. Disord.* 94, 211–218.
- Peyrot, W.J., Robinson, M.R., Penninx, B.W., Wray, N.R., 2016. Exploring boundaries for the genetic consequences of assortative mating for psychiatric traits. *JAMA Psychiat.* 73 (11), 1189–1195.
- Singh, A.L., D'Onofrio, B.M., Slutske, W.S., Turkheimer, E., Emery, R.E., Harden, K.P., Heath, A.C., Madden, P.A., Statham, D.J., Martin, N.G., 2011. Parental depression and offspring psychopathology: a children of twins study. *Psychol. Med.* 41 (7), 1385–1395.
- Smith, B.H., Campbell, H., Blackwood, D., Connell, J., Connor, M., Deary, I.J., Dominiczak, A.F., Fitzpatrick, B., Ford, I., Jackson, C., Haddow, G., Kerr, S., Lindsay, R., McGilchrist, M., Morton, R., Murray, G., Palmer, C.N., Pell, J.P., Ralston, S.H., St Clair, D., Sullivan, F., Watt, G., Wolf, R., Wright, A., Porteous, D., Morris, A.D., 2006. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med. Genet.* 7, 74.
- Sullivan, P.F., Neale, M.C., Kendler, K.S., 2000. Genetic epidemiology of major depression: review and meta-analysis. *Am. J. Psychiatry* 157, 1552–1562.
- Tenesa, A., Haley, C.S., 2013. The heritability of human disease: estimation, uses and abuses. *Nat. Rev. Genet.* 14, 139–149.
- Thapar, A., McGuffin, P., 1994. A twin study of depressive symptoms in childhood. *Br. J. Psychiatry* 165, 259–265.
- Xia, C., Amador, C., Huffman, J., Trochet, H., Campbell, A., Porteous, D., Hastie, N.D., Hayward, C., Vitart, V., Navarro, P., Haley, C.S., Generation, S., 2016. Pedigree- and SNP-associated genetics and recent environment are the major contributors to anthropometric and cardiometabolic trait variation. *PLoS Genet.* 12, e1005804.
- Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., 2011. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82.
- Zaitlen, N., Kraft, P., Patterson, N., Pasaniuc, B., Bhatia, G., Pollack, S., Price, A.L., 2013. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet.* 9.