Check for updates

## ARTICLE    OPEN
# Genetics of age-at-onset in major depression

Arvid Harder[1], Thuy-Dung Nguyen[1,2], Joëlle A. Pasman[1], Miriam A. Mosing[1,3,4], Sara Hägg [1] and Yi Lu [1,2✉]

Major depression (MD) is a complex, heterogeneous neuropsychiatric disorder. An early age at onset of major depression (AAO-MD) has been associated with more severe illness, psychosis, and suicidality. However, not much is known about what contributes to individual variation in this important clinical characteristic. This study sought to investigate the genetic components underlying AAO-MD. To investigate the genetics of AAO-MD, we conducted a genome-wide association meta-analysis of AAO-MD based on self-reported age of symptoms onset and self-reported age at first diagnosis from the UK Biobank cohort (total $N = 94,154$). We examined the genetic relationship between AAO-MD and five other psychiatric disorders. Polygenic risk scores were derived to examine their association with five psychiatric outcomes and AAO-MD in independent sub-samples. We found a small but significant SNP-heritability (~6%) for the AAO-MD phenotype. No SNP or gene reached SNP or gene-level significance. We found evidence that AAO-MD has genetic overlap with MD risk ($r_g = -0.49$). Similarly, we found shared genetic risks between AAO-MD and autism-spectrum disorder, schizophrenia, bipolar disorder, and anorexia nervosa ($r_g$ range: $-0.3$ to $-0.5$). Polygenic risk scores for AAO-MD were associated with MD, schizophrenia, and bipolar disorder, and AAO-MD was in turn associated with polygenic risk scores derived from these disorders. Overall, our results indicate that AAO-MD is heritable, and there is an inverse genetic relationship between AAO-MD and both major depression and other psychiatric disorders, meaning that SNPs associated with earlier age at onset tend to increase the risk for psychiatric disorders. These findings suggest that the genetics of AAO-MD contribute to the shared genetic architecture observed between psychiatric disorders.

## INTRODUCTION

Major depression (MD) is characterized by a pervasive low mood and/or a loss of interest in things normally found enjoyable. Approximately one in six individuals will develop MD during their lifetime [1]. MD is moderately heritable (30–50%) [2], and the largest published genome-wide association study (GWAS) to date has identified 178 independent loci associated with MD risk and estimated the heritability attributable to common variants on the liability scale at 11.3% [3].

MD onsets at all ages, with an average age at onset (AAO-MD) of 30 years [4]. The AAO-MD phenotype is primarily used to dissect MD heterogeneity [5], i.e., using AAO-MD to delineate between cases of early onset and late-onset MD using predetermined age cutoffs. The age cutoffs varied widely between studies [6–9]; some used "absolute" cutoffs, for example considering those that had onset before age 18 as early onset MD [9], while others defined cutoffs with respect to the specific AAO-MD distributions within study samples [6]. In epidemiological studies, early onset MD has been found in association with more severe long-term outcomes, later in life development of bipolar disorder and psychosis, and a greater familial risk [7, 10, 11]. In twin studies, early onset MD was associated with risk for MD in the co-twin, while late-onset MD was associated with vascular damage in the co-twin [8]. Using GWAS data, studies have shown that early onset cases had higher polygenic risk scores (PRS) of MD and are more genetically similar to bipolar disorder and schizophrenia compared to late-onset cases [6, 12].

Although AAO-MD has been implicated as a key clinical characteristic for MD, little is known about what contributes to individual variation in AAO-MD. Studies have examined the role of a genetic component in AAO-MD. Early twin studies suggested twin resemblance in AAO-MD is largely environmental [13]. Recent GWAS showed conflicting results. Two studies estimated that about 17–55% of variances in AAO-MD could be attributable to common variants albeit with large confidence intervals [6, 14], whereas data from the CONVERGE consortium showed that AAO-MD was not significantly heritable [15]. These discrepancies might reflect the methodological challenges in AAO-MD phenotyping, which include measurement differences, sample ascertainment, recall bias, and data censoring. Given the clinical importance of AAO-MD, more research is required to address these methodological challenges.

Here we used the UK biobank (UKB) cohort—the largest genotyped sample to date with comprehensive phenotyping—to study the genetics of AAO-MD. To maximize sample size, we studied two quantitative definitions relevant to AAO-MD: (1) self-reported age of symptoms onset (age at symptom, $N = 76,365$) and (2) self-reported age at first diagnosis (age at diagnosis, $N = 26,425$; or 17,789 after removing overlapping and related samples with the sample subset of age at symptom). Our primary goal was

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [2]Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden. [3]Department of Cognitive Neuropsychology, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany. [4]Melbourne School of Psychological Sciences, Faculty for Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Australia. ✉email: lu.yi@ki.se
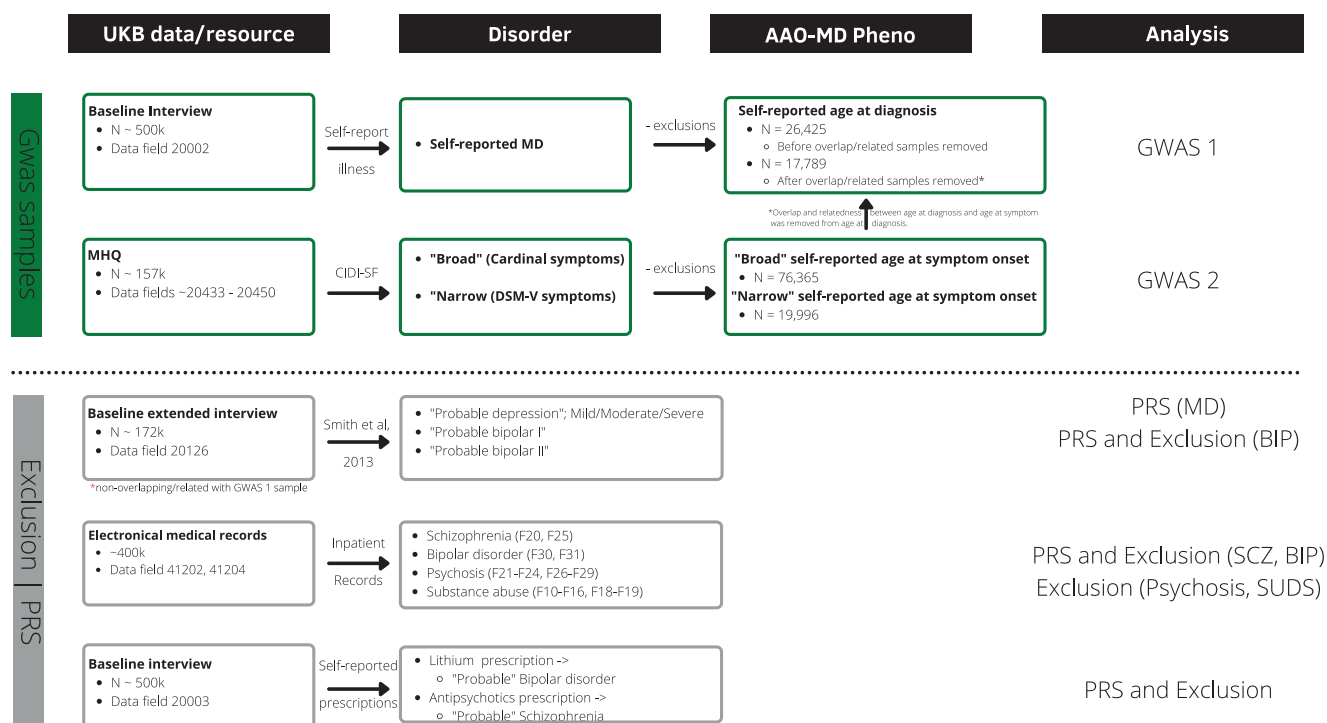
**Fig. 1 Phenotypic definitions in UK Biobank.** This figure displays the data sources used for each definition, and for what purpose the definition was derived. The two MD definitions are marked with green box color. Data sources and definitions for the exclusion criteria, self-reported lithium/antipsychotics prescriptions, probable depression, and probable bipolar disorder type I and II are marked with gray box color.

to investigate the genetic architecture of AAO-MD. In addition, we examined the association between PRS of AAO-MD and risk for psychiatric disorders, and vice versa; the association between PRS of psychiatric disorders and AAO-MD.

## MATERIAL AND METHODS
### Study population
We used the UKB cohort for all analyses to limit the impact of inconsistent AAO-MD phenotypes across cohorts [6]. The UKB contains deep phenotype and genotype data on ~500,000 individuals aged 37–73 at the date of recruitment and from across the UK. Participants attended a baseline assessment, where information was gathered through digital questionnaires and nurse interviews. A subset of individuals ($N = 172,751$) completed an extended touchscreen questionnaire about depressive and manic symptoms at baseline [16]. In 2017, participants with a valid email address were invited to participate in the online "Mental Health Follow-up" questionnaire (MHQ), where 157,366 individuals completed the questionnaire. For information regarding lifetime depressive symptoms, the International Diagnostic Interview Short Form (CIDI-SF) was used. For nearly the entire population, electronical medical records from hospital care have been further integrated and connected to the UKB data [17]. Informed consent was obtained from all participants. This study used the UK Biobank Resource under Application Number 22224.

### Measures
*Major depression.* The comprehensive measures available at the UKB allow for multiple ways to define MD and to thoroughly compare the genetic profiles underlying different definitions [17, 18]. Here we defined MD in two ways: (a) self-reported MD at the baseline interview; and (b) self-reported lifetime symptoms from the CIDI-SF questionnaire in the MHQ (Fig. 1). At the baseline interview, participants were asked to indicate if they (had previously) suffered from any type of illness, including depression. From the MHQ subsample we derived a "broad definition" of MD based on the cardinal symptoms (low mood and anhedonia), and a "narrow definition", requiring endorsement of either or both of the cardinal symptoms and a minimum of five symptoms (cardinal symptoms included), referred to as the "CIDI" definition in Cai et al. [18]

(Supplementary Methods). These MD measures were selected because corresponding age at onset information was available.

Given the phenotypic differences in the two MD definitions, we analyzed them separately and excluded overlapping and related individuals from the subset with self-reported diagnosis. We also excluded individuals with probable schizophrenia, psychosis, and bipolar disorder (for definitions, see *Other psychiatric disorders*, Fig. 1 *and* Supplementary Table 1). In addition to this, we applied further exclusion criteria in the "narrow" definition of MD, excluding substance abuse cases, to match the CIDI definition in Cai et al. [18] (Supplementary Methods).

*Age at onset in major depression.* Participants that reported MD during the baseline interview were asked at what age they first received a *diagnosis* of MD from a doctor (data field 20009). We therefore defined *age at diagnosis* for this definition of MD. For participants in the MHQ, if they endorsed either of the cardinal symptoms of MD, they were asked at what age these symptoms first appeared (data field 20433). For this definition of MD, we defined *age at symptom* (Fig. 1).

*Other psychiatric disorders.* We combined the hospital registry data with the self-reported medication data to define cases of probable schizophrenia (F20 or F25 or antipsychotics prescription) and probable bipolar disorder (F30 or F31 or lithium prescription). In addition to this, we extracted cases of the "probable depression" definition described in Howard et al. [19], and cases of probable bipolar type I and II described in Smith et al. [16]. These definitions were used to examine the association between PRS for AAO-MD and risk for these disorders (Supplementary Methods; Supplementary Tables 2–4).

### Genotyping, quality control, and ancestry determination
Genotype data were available from the full data release of the UKB in 2018 ($N = 487,410$). UKB samples were genotyped on two customized arrays, the UK Biobank Axiom Array, and the Applied Biosystems UK BiLEVE Axiom Array by Affymetrix covering ~800,000 markers. The genotype data underwent central quality control (QC) and a two-stage imputation process using the reference of the Haplotype Reference Consortium (HRC) and UK10K (for rarer variants not present in HRC). The full QC procedure and pipeline are detailed elsewhere [17]. We

applied post-imputation QC to filter imputed variants with minor allele frequency <1% and INFO score <0.6, leaving a total of 9,709,770 variants for analysis. We determined ancestral outliers by projecting the principal components (PCs) of the UKB samples to the PCs of the 1000 Genomes reference panel. We identified 27,700 (5.6%) samples as non-European ancestral outliers (i.e., those with more than three standard deviations from the mean of the European samples in the 1000 Genomes reference) and excluded these individuals from further analysis. The first ten PCs based on the UKB samples were used to adjust for population stratification in subsequent analyses.

## Statistical analysis

*Phenotype transformation.* As there is known confounding between age at onset and age cohort effects for MD [20], we analyzed both definitions with caution. First, since we observed significant correlations between age at symptoms and age, sex, age*sex, age² and sex*age², we removed these effects from both phenotypes by analyzing the regression residuals (Supplementary Table 5A, B*)*. Secondly, we transformed the residuals using a rank-based inverse normal transformation.

*SNP-based heritability.* To test how much individual differences in AAO-MD could be explained by common genetic variants we estimated SNP-based heritability $(h_g^2)$ using GCTA-GREML, adjusting for the first ten PCs [21].

*Genome-wide association study (GWAS) and meta-analysis.* We first conducted two AAO-MD GWAS, separately for the age at symptoms and age at diagnosis definitions. Extensive relatedness (over 30% of participants have one or more relatives up to the third degree) was found among UKB participants [17] and removing these samples from analyses would result in a major loss of statistical power. We therefore aimed to account for related samples in GWAS using the fastGWA module in GCTA [22]. We computed a genetic relatedness matrix (GRM) using a filtered set of common SNPs (Supplementary Methods) and thereafter fit a linear-mixed model using the GRM as a mixed effect while adjusting for the first 10 PCs.

In the meta-analysis of the two GWAS results, sample overlap ($N = 7479$) and relatedness ($N = 1157$) were identified and removed before analyses (Supplementary Methods). We subsequently conducted an inverse variance-weighted meta-analysis based on the GWAS summary statistics from the age at symptoms and age of diagnosis definitions using METAL [23]. To examine potential bias from population stratification or relatedness, the intercept was estimated using LD score regression (LDSC) [24].

We used the web-based tool Functional Mapping and Annotation (FuMA) [25], to report and annotate independent genomic loci, and to perform MAGMA gene-based or gene-set analysis and tissue enrichment analysis using GTEx gene expression data [26].

*Genetic correlations.* Using GWAS summary statistics, we examined the genetic correlations ($r_g$) between AAO-MD and major psychiatric disorders to explore the relationship between AAO-MD and genetic risk for these disorders. We obtained publicly available summary statistics for MD [19], schizophrenia [27], bipolar disorder [28], anorexia nervosa [29], autism-spectrum disorder [30], attention-deficit hyperactivity disorder [31], PGC cross-disorder [32] (pooled cases of bipolar disorder, schizophrenia, major depressive disorder, autism-spectrum disorder and attention-deficit hyperactivity disorder) and depressive symptoms [33]. A recent extension of LDSC, the High-Definition Likelihood (HDL) method [34], was used to estimate $r_g$ in all pairs except between AAO-MD and cross-disorder due to a low SNP overlap with the LD-matrix (where we used LDSC instead). Compared to LDSC, HDL improves precision in genetic correlation estimates, as it accounts for the full LD information and it is a likelihood-based method instead of a method of moments. We used the precomputed LD-matrix from UKB individuals which was available with the release of the HDL software. We applied Bonferroni correction to control for multiple testing and consider $r_g$ estimates significant at $p < 0.006$ (= 0.05/8).

*Polygenic risk scores (PRSs).* PRSs were derived using the SbayesR module in GCTB [35] (Supplementary Methods) to examine the association between PRS based on meta-analyzed GWAS of AAO-MD (PRS$_{AAO-MD}$). Using PRS$_{AAO-MD}$, we examined its association with risks to mild, moderate, and severe depression, as well as probable bipolar disorder and probable schizophrenia (after removing overlap and restricting to unrelated

individuals). Bipolar and psychotic disorders often debut with depressive episodes [36]; we therefore sought to explore whether PRS from these disorders (and from MD) were associated with AAO-MD. PRS for MD (PRS$_{MD}$), schizophrenia (PRS$_{SCZ}$), and bipolar disorder (PRS$_{BP}$) were derived using MD GWAS summary statistics excluding UKB [12] and the largest publicly available summary statistics for schizophrenia and bipolar disorder [27, 28].

To examine the association between PRS$_{AAO-MD}$ with other psychiatric disorders, we fit a logistic regression model. To explore if the potential association was independent of genetic liability related to MD, all models were refit a second time, adjusting for PRS$_{MD}$. Secondly, we estimated the association between PRS$_{MD}$, PRS$_{SCZ}$, PRS$_{BP}$, and age at symptoms/diagnosis (removing related samples within each definition), to examine the relationship between genetic liability for these disorders and AAO-MD. In all PRS models, we adjusted for the first 10 principal components. To control for multiple testing, we applied a Bonferroni correction and report statistical significance at $p < 0.005$ (= 0.05/10).

*Sensitivity analyses.* Risk for postpartum depression (PPD) is tied to the reproductive years in women and can complicate the analysis. We therefore, in sensitivity analysis, defined a sample of PPD cases (Supplementary Methods*)* and reran the analysis after removing MD cases that met criteria for PPD.

## RESULTS
### Age at onset of MD is heritable
After QC, 76 365 individuals (64.5% women, 35.5% men) of the 157 366 MHQ participants reported experiencing the cardinal symptoms of MD (age at symptoms $M = 37.3$, SD = 15; Table 1). From the baseline interview, 26 425 individuals after QC (66% women, 34% men) reported previous diagnoses of depression by a doctor (age at diagnosis $M = 42.7$, SD = 13.1; Table 1). Among individuals for whom both measures were available ($N = 7$ 479), the correlation between age at symptoms and age at diagnosis was moderate ($r = 0.54$, $p < 2.2 \times 10^{-6}$).

We found evidence for a heritable component underlying the AAO-MD phenotypes, as shown in the significant estimates of $h_g^2 = 5.6\%$ (95% CI: 4.2–6.9%, $p < 0.0001$) for the age at symptoms (broad MD definition); and for age at diagnosis, $h_g^2 = 5.7\%$ (95% CI: 2.0–9.2%, $p = 0.00078$). For the narrow MD definition, the $h_g^2$ estimate of the age at symptoms was lower and non-significant ($h_g^2 = 3.8\%$, s.e = 4.4%, $p > 0.05$), which could be due to the markedly reduced sample size (19,572, 26% of the sample size in the broad definition).

Despite a lower heritability observed in the narrow MD definition, the genetic correlation between the age at symptoms phenotype in both broad and narrow MD definitions was not significantly different from 1 ($r_g = 1.4$, s.e = 0.67), indicating a substantial overlap in their underlying genetic factors. In light of

**Table 1.** Participant characteristics.

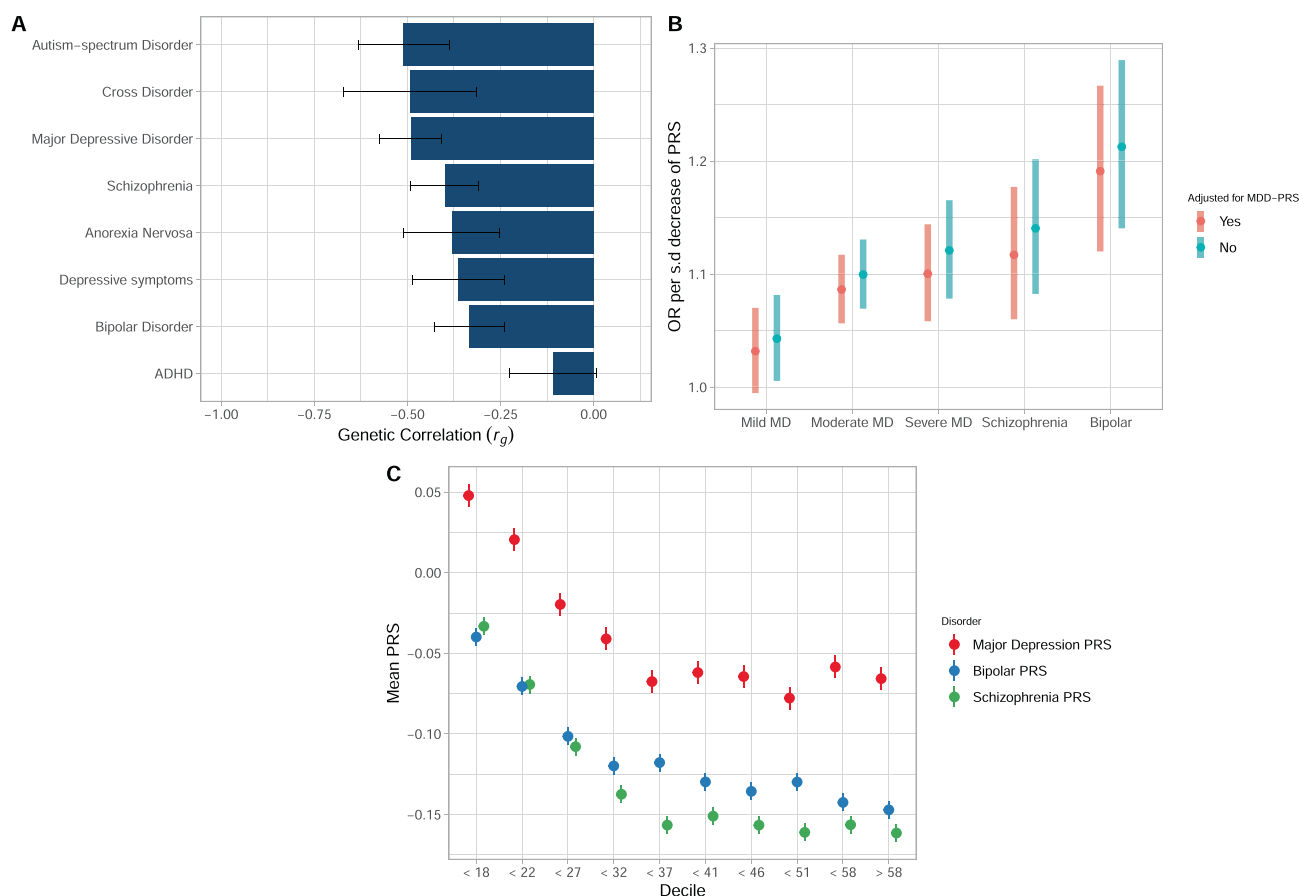|  | Age at symptoms | Age at diagnosis |
|---|---|---|
| *N* analyzed | 76,365 | 26,425 |
| Females, % | 64.5% | 65.8% |
| Age, mean (sd) | 55.4 (7.67) | 55.4 (7.80) |
| Age at symptoms, mean (sd) | 37.3 (15.0) | .[a] |
| Age at diagnosis, mean (sd) | .[a] | 42.7(13.1) |
| Townsend index, mean (sd) | −1.62 (2.86) | −0.9 (3.25) |
| Body-mass index (sd) | 26.8 (4.76) | 28.3 (5.53) |
| Lifetime-smoking % | 44.5% | 51.3% |

[a]Measure is not available for this definition.

**Fig. 2 Inverse relationship between AAO-MD and psychiatric disorders. A** Genetic correlations ($r_g$) estimated using High-Definition Likelihood between summary statistics from the meta-analyzed genome-wide association analysis on age at symptoms/diagnosis of major depression and psychiatric traits. The x-axis displays the $r_g$ estimate with error bars displaying the 95% confidence interval. The y-axis displays the psychiatric trait with which the genetic correlation was calculated. **B** The association between a polygenic risk score derived from genome-wide association analysis from AAO-MD and psychiatric disorders. The y-axis displays the odds ratios of psychiatric disorders per standard deviation decrease in PRS$_{AAO-MD}$. The x-axis displays the psychiatric disorder. Error bars display the 95% confidence intervals. The odds ratio estimate is displayed with and without adjustment for MDD-PRS. **C** The mean PRS for major depression, schizophrenia, and bipolar disorder, stratified by age at symptoms decile. The x-axis displays each decile for age at symptoms. The y-axis displays the mean PRS for each disorder within a decile. Error bars display 95% confidence interval of the mean PRS in each decile.

this, we proceeded with the age at symptoms phenotype in the broad definition of MD to maximize the sample size in the following analysis.

## No genome-wide significant locus identified in a meta-analysis of nearly 100k individuals

We first conducted GWAS separately within the phenotypes of age at symptoms and age at diagnosis in MD. We observed a strong genetic correlation of 0.91 (s.e = 0.3) between these two phenotypes. Further supporting overlapping genetic components underlying the two AAO-MD phenotypes, we found that PRS of age at symptoms was significantly associated with age at diagnosis ($p = 2.7 \times 10^{-5}$) (Supplementary Table 6).

Given the evidence of a strong genetic overlap between the two phenotypes, we conducted a meta-analysis of AAO-MD combining the GWAS results of the age at symptoms and age at diagnosis together (total $N = 94\,154$ after removing overlapping and related samples). No SNP passed the genome-wide significance threshold (Supplementary Fig. 1). The intercept from LDSC (1.0051, s.e = 0.0064) indicated negligible confounding from population stratification or relatedness. We did not identify gene-level or gene-set association using MAGMA. However, tissue expression analysis showed significant enrichment in the brain

($p = 3.2 \times 10^{-6}$), specifically, in the anterior cingulate cortex ($p = 2.5 \times 10^{-5}$), amygdala ($p = 3.05 \times 10^{-5}$), frontal cortex ($p = 5.62 \times 10^{-5}$) and hippocampus ($p = 1.21 \times 10^{-5}$).

## Shared genetic architecture between age at onset of MD and other psychiatric disorders

Overall, the estimates of $r_g$ were of similar strength between AAO-MD and all included psychiatric disorders except attention-deficit hyperactivity disorder (ADHD) (Fig. 2A). AAO-MD was negatively correlated with psychiatric disorder risks, meaning that SNPs associated with earlier age at onset tend to increase risk for psychiatric disorders. Notably, the genetic correlation of AAO-MD with both MD and depressive symptoms were only of moderate strength ($r_g = -0.491$, s.e = 0.042 and $r_g = -0.363$, s.e = 0.062, respectively). The genetic correlation with autism-spectrum disorder ($r_g = -0.51$, s.e = 0.062), cross-disorders ($r_g = -0.494$, s.e = 0.091), schizophrenia ($r_g = -0.40$, s.e = 0.047), bipolar disorder ($r_g = -0.334$, s.e = 0.047) and anorexia nervosa ($r_g = -0.381$, s.e = 0.065) were all comparable in magnitude to the correlation between AAO-MD and MD or between AAO-MD and depressive symptoms. ADHD was the only psychiatric disorder that was not significantly related to AAO-MD ($r_g = -0.109$, s.e = 0.059).

## Polygenic risk scores for AAO-MD are associated with severe psychiatric disorders, independently from MD risk

Using GWAS summary statistics of the meta-analyzed age at diagnosis/symptoms phenotypes, we derived a polygenic risk score (PRS$_{AAO-MD}$) and examined its association with other psychiatric disorders. Overall, the association between PRS$_{AAO-MD}$ and probable depression increased with the severity, with the strongest association with severe MD. PRS$_{AAO-MD}$ association with mild depression was not significant after adjustment for multiple comparisons, but PRS$_{AAO-MD}$ was significantly associated with moderate probable depression (OR = 1.10, $p = 2.27 \times 10^{-11}$) and severe probable depression (OR = 1.12, $p = 8.01 \times 10^{-9}$), meaning that for each standard deviation decrease in PRS$_{AAO-MD}$, the relative risk for moderate and severe depression increased by 11% and 12%. Interestingly, the association was even stronger for probable schizophrenia (OR = 1.14, p = $7.99 \times 10^{-7}$) and probable bipolar disorder (OR = 1.21, p = $7.23 \times 10^{-10}$). We further explored if these associations were independent of genetic liability specific to MD by adjusting for PRS$_{MD}$. All associations remained significant (with the exception of mild depression), with slightly attenuated effect sizes (Fig. 2B). We also stratified the analysis by bipolar disorder type I and II and observed similar effect sizes in probable bipolar disorder type I (OR = 1.12, $p = 0.0119$, N cases = 512) and probable bipolar type II (OR = 1.15, $p = 0.00341$, N cases = 476).

## Polygenic risk scores for psychiatric disorders are associated with age at onset

We found significant associations between PRS$_{SCZ}$, PRS$_{BP}$, and PRS$_{MDD}$ with age at symptoms and age at diagnosis. Mean PRS$_{SCZ}$, PRS$_{BP}$, and PRS$_{MDD}$ sharply decrease with age at symptoms among early onset cases (roughly before 32 years old) while stabilizing thereafter; Fig. 2C). In both definitions the strongest effect size was observed for PRS$_{SCZ}$, where the average age at symptoms decreased by approximately one year per standard deviation increase in PRS$_{SCZ}$. The effects of both PRS$_{MD}$ and PRS$_{BP}$ were highly significant; the average age at symptoms decreased by approximately 9 months per standard deviation increase in PRS$_{BP}$, and by approximately 6 months per standard deviation increase in PRS$_{MD}$. In the age at diagnosis subsample, the results were very similar (Supplementary Fig. 2).

## Postpartum cases have a negligible impact on the results

Postpartum depression cases tended to onset earlier compared to other MD cases (age at symptoms; mean = 28.1, s.d = 5.99, age at diagnosis; mean = 38, s.d = 12.5, Supplementary Table 7, Supplementary Fig. 3A and B). The correlation between GWAS effects sizes with and without PPD cases was very similar ($r = 0.98$). Removing PPD cases minimally changed the regression coefficients of PRS from other psychiatric disorders (Supplementary Fig. 2).

## DISCUSSION

This study presents the largest GWAS to date on AAO-MD aiming to understand the genetic basis of this important clinical characteristic of MD. We clearly demonstrated that AAO-MD is heritable, with roughly 6% of its variance attributable to common genetic variants. Despite that no single variant or gene passed the genome-wide significance threshold, we had strong evidence for an overall genetic sharing between AAO-MD and MD risk loci ($r_g = -0.49$), and to a similar extent, between AAO-MD and risks to many other psychiatric disorders ($r_g$ range: $-0.3$ to $-0.5$, except for the non-significant $r_g$ with ADHD). Further, we found that PRS$_{AAO-MD}$ was associated with psychiatric disorders (including moderate to severe forms of MD, bipolar disorder and schizophrenia) independently from the genetic liability to MD, and that PRS of these psychiatric disorders were also associated with AAO-MD. Overall, this report provides novel insights into the genetics of AAO-MD, and its relationship with risks to MD and other psychiatric disorders.

Early AAO-MD often correlates with adverse clinical outcomes, including more chronic illness, greater disorder severity, increased suicidality, and lack of treatment response [37, 38]. Therefore, studying AAO-MD may have important clinical implications in primary, secondary, and tertiary prevention. Despite the importance, AAO-MD has been one of the least commonly studied clinical characteristics due to practical difficulties in phenotyping [38]. Here we studied two measurements, age at symptoms and age at diagnosis, using self-reported survey data from UKB. UKB participants reported later AAO-MD (mean age at symptoms = 37.3) than those reported in the literature (mean~22–33 across 18 nations [1], or mean ~29 across PGC-MDD genetic cohorts) [6]. There appeared to be an over five-year delay from the age at symptoms to the age at first diagnosis (mean = 42.7), meaning that many sought treatments at a much later stage than at symptom onset. These observations reveal that on average, UKB participants with MD had later onset. Although this observation could be due to recall bias given the long time gap between the disorder onset and survey participation, it might have partly contributed to the genetic differences observed between UKB cases and other cohorts ($r_g = 0.87$ with PGC samples) [19].

Despite a moderate phenotypic correlation ($r = 0.54$) between age at symptoms and age at diagnosis, our findings suggest very similar genetic components underlying the two measures. The two measures had nearly identical estimates of SNP-based heritability (5.6% and 5.7%, respectively) and they had a high genetic correlation (0.91), indicating that genetic factors are largely shared. As expected from the high genetic correlation, we showed that PRS of age at symptoms was significantly associated with age at diagnosis. Our significant SNP-based heritability was in line with other reports that showed AAO-MD is heritable; [6, 14] though our estimate based on the large UKB sample was much smaller (~6%; the estimate can be benchmarked against the SNP-heritability of 8.9% [19] for MD risk) than previously estimated, highlighting the need of a large sample size to study this phenotype. Previous research has reported that AAO-MD distribution varied across different cohorts due to major differences in instruments and sample ascertainment [6], making it challenging to study AAO-MD in large numbers. However, our finding of a strong genetic correlation between the two AAO-MD measures suggests that future genetic studies on AAO-MD might benefit from utilizing the large-scale biobanks with electronic medical records to determine age at diagnosis as proxy of AAO-MD.

We have clearly demonstrated an inverse genetic relationship between AAO-MD and risks to psychiatric disorders. Early AAO-MD has been linked to family history of MD, and early AAO-MD in parents confer an increased risk in the offspring compared to the offspring of parents with late AAO-MD [39]. We extended these earlier findings by quantifying the overall genetic sharing between MD risk and AAO-MD, and demonstrating that PRS$_{AAO-MD}$ was associated with MD. We observed a gradient along severity, with a stronger association of PRS$_{AAO-MD}$ in the more severe forms of MD. As early AAO-MD cases tend to have worse long-term outcomes and more severe symptoms [10, 40], our findings revealed that this epidemiological association is partly mediated through genetic liability for AAO-MD. Furthermore, this association was independent of genetic liability to MD, suggesting independent pathways through which AAO-MD genetics confer susceptibility. Interestingly, we found similar levels of genetic overlap between AAO-MD and schizophrenia, bipolar disorder, and that PRS$_{AAO-MD}$ was associated with these two disorders more strongly than with MD. Individuals with psychotic disorders and bipolar disorder frequently experience depressive symptoms before receiving these diagnoses [36], and early onset MD is associated with later bipolar disorder [41]. Hence, our results suggested a genetic basis underlying these clinical observations. For childhood-onset psychiatric disorders, we identified a moderate genetic correlation

between AAO-MD and ASD, but not with ADHD. Previous studies showed an association of $PRS_{SCZ}$ with AAO-MD in early onset MD cases, as well as a small association with $PRS_{BP}$ [6, 42]. We replicated these results and further demonstrated that these PRS were also associated with AAO as a quantitative trait, and that it is primarily early onset cases (AAO < 32) that have a higher PRS for these disorders. Altogether, these findings suggest that age at which individuals first experience depressive symptoms may play an important role in psychiatric illness. Clinically, the patient group with early AAO in particular might benefit from closer monitoring and interventions to prevent severe outcomes.

The key strength of this study lies in our use of the large and genetically homogenous sample, with consistent phenotype collection. We also acknowledge several limitations. AAO is a difficult trait to measure and to operat ionize; self-reported phenotyping suffers from recall bias, while register data miss individuals not seeking treatment and are often left-censored. We carefully analyzed this trait by adjusting for cohort effects and applied transformations on the phenotypes, but it is nonetheless possible that recall bias in the self-reported UKB phenotypes still influenced our results. To define cases of MD, we applied a relatively broad inclusion criteria (the cardinal symptoms of depression or self-reported diagnosis) to maximize sample size, although some argued that minimal phenotyping may bias views of the genetics of MD [18]. Lastly, we acknowledge that UKB has a known healthy volunteer bias, which is further exacerbated in the MHQ subsample [43, 44]. Our results therefore warrant further replication in independent datasets.

Recent advances in statistical methods for case-control GWAS demonstrate that GWAS power can be increased by incorporating age at onset and family history information into case-control GWAS [45]. These methods rely on age at onset being heritable as well as a genetic correlation between age at onset and disease liability to increase statistical power. Therefore, understanding the genetics of age at onset in MD is a necessary first step. In parallel to the ongoing effort in MD genetic discovery, we urge for more genetic studies to investigate key clinical indicators related to MD.

## URLS
FuMa: https://fuma.ctglab.nl/, GCTA: https://cnsgenomics.com/software/gcta/#fastGWA, GCTB: https://cnsgenomics.com/software/gctb/, HDL: https://github.com/zhenin/HDL.

## CODE AVAILABILITY
The code used to generate this analysis is available at https://github.com/Ararder/Genetics-of-age-at-onset-in-MD.

## REFERENCES
1. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. BMC Med. 2011;9:90.
2. Kendall KM, Van Assche E, Andlauer TFM, Choi KW, Luykx JJ, Schulte EC, et al. The genetic basis of major depression. Psychol Med. 2021;51:2217–2230.
3. Levey DF, Stein MB, Wendt FR, Pathak GA, Zhou H, Aslan M, et al. Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. Nat Neurosci. 2021;24:954–963.
4. Solmi M, Radua J, Olivola M, Croce E, Soardo L, Salazar de Pablo G, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry. 2021;1–15. https://doi.org/10.1038/s41380-021-01161-7.
5. Harald B, Gordon P. Meta-review of depressive subtyping models. J Affect Disord. 2012;139:126–40.
6. Power RA, Tansey KE, Buttenschøn HN, Cohen-Woods S, Bigdeli T, Hall LS, et al. Genome-wide association for major depression through age at onset

7. Zisook S, Lesser I, Stewart JW, Wisniewski SR, Balasubramani GK, Fava M, et al. Effect of age at onset on the course of major depressive disorder. Am J Psychiatry. 2007;164:1539–46.
8. Kendler KS, Fiske A, Gardner CO, Gatz M. Delineation of two genetic pathways to major depression. Biol Psychiatry. 2009;65:808–11.
9. Rush AJ. The varied clinical presentations of major depressive disorder. J Clin Psychiatry. 2007;68 Suppl 8:4–10.
10. Nierenberg AA, Trivedi MH, Fava M, Biggs MM, Shores-Wilson K, Wisniewski SR, et al. Family history of mood disorder and characteristics of major depressive disorder: A STAR*D (sequenced treatment alternatives to relieve depression) study. J Psychiatr Res. 2007;41:214–21.
11. Rice F, Sellers R, Hammerton G, Eyre O, Bevan-Jones R, Thapar AK, et al. Antecedents of new-onset major depressive disorder in children and adolescents at high familial risk. JAMA Psychiatry. 2017;74:153–60.
12. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50:668–81.
13. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. Familial influences on the clinical characteristics of major depression: a twin study. Acta Psychiatr Scand. 1992;86:371–8.
14. Ferentinos P, Koukounari A, Power R, Rivera M, Uher R, Craddock N, et al. Familiality and SNP heritability of age at onset and episodicity in major depressive disorder. Psychol Med. 2015;45:2215–25.
15. Docherty AR, Edwards AC, Yang F, Peterson RE, Sawyers C, Adkins DE, et al. Age of onset and family history as indicators of polygenic risk for major depression. Depress Anxiety. 2017;34:446–52.
16. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK Biobank: Cross-Sectional Study of 172,751 Participants. PLoS ONE. 2013;8:e75362.
17. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203–9.
18. Cai N, Revez JA, Adams MJ, Andlauer TFM, Breen G, Byrne EM, et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet. 2020;52:437–47.
19. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci. 2019;22:343–52.
20. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry. 2000;157:1552–62.
21. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.
22. Jiang L, Zheng Z, Qi T, Kemper KE, Wray NR, Visscher PM, et al. A resource-efficient tool for mixed model association analysis of large-scale data. Nat Genet. 2019;51:1749–55.
23. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. Bioinformatics. 2010;26:2190–1.
24. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–5.
25. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.
26. Leeuw CA, de, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLOS Comput Biol. 2015;11:e1004219.
27. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke S, Walters JT, O'Donovan MC Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. Genet Genomic Med. 2020; https://doi.org/10.1101/2020.09.12.20192922.
28. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet. 2021;53:817–29.
29. Watson HJ, Yilmaz Z, Thornton LM, Hübel C, Coleman JRI, Gaspar HA, et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. Nat Genet. 2019;51:1207–14.
30. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–44.
31. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51:63–75.

stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Biol Psychiatry. 2017;81:325–35.

32. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet Lond Engl. 2013;381:1371–9.

33. Okbay A, Baselmans BML, De Neve J-E, Turley P, Nivard MG, Fontana MA, et al. Genetic variants associated with subjective well-being, depressive symptoms and neuroticism identified through genome-wide analyses. Nat Genet. 2016;48:624–33.

34. Ning Z, Pawitan Y, Shen X. High-definition likelihood inference of genetic correlations across human complex traits. Nat Genet. 2020;52:859–64.

35. Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. Nat Commun. 2019;10:5086.

36. Musliner KL, Krebs MD, Albiñana C, Vilhjalmsson B, Agerbo E, Zandi PP, et al. Polygenic risk and progression to bipolar or psychotic disorders among individuals diagnosed with unipolar depression in early life. Am J Psychiatry. 2020;177:936–43.

37. de Girolamo G, Dagani J, Purcell R, Cocchi A, McGorry PD. Age of onset of mental disorders and use of mental health services: needs, opportunities and obstacles. Epidemiol Psychiatr Sci. 2012;21:47–57.

38. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustun TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry. 2007;20:359–64.

39. Wickramaratne PJ, Weissman MM. Onset of psychopathology in offspring by developmental phase and parental depression. J Am Acad Child Adolesc Psychiatry. 1998;37:933–42.

40. Tozzi F, Prokopenko I, Perry JD, Kennedy JL, McCarthy AD, Holsboer F, et al. Family history of depression is associated with younger age of onset in patients with recurrent depression. Psychol Med. 2008;38:641–9.

41. Birmaher B, Ryan ND, Williamson DE, Brent DA, Kaufman J, Dahl RE, et al. Childhood and adolescent depression: a review of the past 10 years. Part I J Am Acad Child Adolesc Psychiatry. 1996;35:1427–39.

42. Musliner KL, Mortensen PB, McGrath JJ, Suppli NP, Hougaard DM, Bybjerg-Grauholm J, et al. Association of polygenic liabilities for major depression, bipolar disorder, and schizophrenia with risk for depression in the danish population. JAMA Psychiatry. 2019;76:516–25.

43. Adams MJ, Hill WD, Howard DM, Dashti HS, Davis KAS, Campbell A, et al. Factors associated with sharing e-mail information and mental health survey participation in large population cohorts. Int J Epidemiol. 2020;49:410–21.

44. Davis KAS, Coleman JRI, Adams M, Allen N, Breen G, Cullen B, et al. Mental health in UK Biobank—development, implementation and results from an online questionnaire completed by 157 366 participants: a reanalysis. BJPsych Open. 2020; 6. https://doi.org/10.1192/bjo.2019.100.

45. Pedersen EM, Agerbo E, Plana-Ripoll O, Grove J, Dreier JW, Musliner KL, et al. Accounting for age-of-onset and family history improves power in genome-wide association studies. bioRxiv. 2021; https://doi.org/10.1101/2021.04.20.440585.

## AUTHOR CONTRIBUTIONS
Y.L. conceived the idea, designed, and supervised the implementation of the study, interpreted the results, revising the manuscript. A.H. conducted the analyses, interpreted the results, drafted, and revised the manuscript. T.D.N. contributed to the analyses, and solving technical issues. M.M., J.P., and S.H. contributed to the study design, discussion on methodological issues and interpretation of the results. S.H. provided access to the data. All authors discussed and commented on the paper.

## COMPETING INTERESTS
The authors declare no competing interests.

## ADDITIONAL INFORMATION
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41398-022-01888-z.

**Correspondence** and requests for materials should be addressed to Yi Lu.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.