

## Genome-wide by environment interaction study of stressful life events and hospital-treated depression in the iPSYCH2012 sample

Nis P. Suppli<sup>1,2</sup>, Klaus K. Andersen<sup>3</sup>, Esben Agerbo<sup>1,4,5</sup>, Veera M. Rajagopal<sup>1,6,7,8</sup>, Vivek Appadurai<sup>1,9</sup>, Jonathan R. I. Coleman<sup>10,11</sup>, Gerome Breen<sup>10,11</sup>, Jonas Bybjerg-Grauholm<sup>1,12</sup>, Marie Bækvad-Hansen<sup>1,12</sup>, Carsten B. Pedersen<sup>1,4,5</sup>, Marianne G. Pedersen<sup>1,4,5</sup>, Wesley K. Thompson<sup>13</sup>, Trine Munk-Olsen<sup>1,4,14</sup>, Michael E. Benros<sup>1,2,15</sup>, Thomas D. Als<sup>1,6,7,8</sup>, Jakob Grove<sup>1,6,7,8,16</sup>, Thomas Werge<sup>1,9</sup>, Anders D. Børglum<sup>1,6,7,8</sup>, David M. Hougaard<sup>1,12</sup>, Ole Mors<sup>1,17</sup>, Merete Nordentoft<sup>1,2</sup>, Preben B. Mortensen<sup>1,4,5</sup>, Katherine L. Musliner<sup>1,4,18</sup>

1. The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH)
2. Mental Health Center Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark
3. Statistics and Pharmacoepidemiology, Danish Cancer Society Research Center, Denmark
4. National Center for Register-based Research, Dept. of Economics, Aarhus University, Aarhus, Denmark
5. The Center for Integrated Register-based Research at Aarhus University (CIRRAU)
6. Department of Biomedicine, Aarhus University, Aarhus, Denmark
7. Center for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark
8. Center for Genomics and Personalized Medicine, CGPM, Aarhus, Denmark
9. Institute of Biological psychiatry, Copenhagen Mental Health Services, Copenhagen, Denmark
10. Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
11. National Institute for Health Research Biomedical Research Centre, South London and Maudsley NHS Trust, London, UK
12. Department for Congenital Disorders, Statens Serum Institute, Copenhagen, Denmark
13. Division of Biostatistics, Herbert Wertheim School of Public Health, University of California, San Diego, La Jolla, CA, USA
14. Dept. of Clinical Medicine, University of Southern Denmark, Odense, Denmark
15. Department of Immunology and Microbiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen Denmark
16. Bioinformatics Research Centre (BiRC), Aarhus University, Aarhus, Denmark
17. Psychosis Research Unit, Aarhus University Hospital-Psychiatry, Aarhus, Denmark
18. Dept. of Affective Disorders, Aarhus University Hospital-Psychiatry, Aarhus, Denmark

**Corresponding author:** Katherine L. Musliner, [klm@clin.au.dk](mailto:klm@clin.au.dk)

## Abstract

Researchers have long investigated a hypothesized interaction between genetic risk and stressful life events in the etiology of depression, but studies on the topic have yielded inconsistent results. We conducted a genome-wide environment interaction study in 18,532 depression cases from hospital-based settings and 20,184 population-representative non-cases from the iPSYCH2012 case-cohort study, a nationally representative sample identified from Danish national registers. Stressful life events including family disruption, serious medical illness, death of a first-degree relative, parental disability and child maltreatment were identified from population-based registers and operationalized as a time-varying count variable (0-4+). Hazard ratios for main and interaction effects were estimated using Cox regressions weighted to accommodate the case-cohort design. The analyses yielded three novel, genome-wide significant ( $P < 5 \times 10^{-8}$ ) loci located in the ATP-binding cassette transporter C 1 (ABCC1) gene, in the A-kinase anchor protein 6 (AKAP6) gene, and near the Major facilitator superfamily domain 1 (MFSD1) gene, as well as 50 loci of suggestive significance. These top 3 hits did not replicate in a case-control sample of depression drawn from the UK Biobank. These results suggest that there may be individual genetic variants that confer risk for or protection against clinical depression only in the presence of stressful life events; however, replication in a sample with similar design and ascertainment methods is needed before any firm conclusions can be drawn. Future gene-by-stress research in depression should focus on establishing even larger collaborative genome-wide environment interaction studies to attain sufficient power.

## 1. Introduction

Major depression is a common, highly burdensome mental illness that affects as many as 21% of people at some point during their lifetimes<sup>1,2</sup>. Studies suggest that major depression is around 30-40% heritable<sup>3</sup>, meaning that a moderate amount of the population-level variability in major depression can be attributed to genetic factors. However, environment also plays an important role in determining who develops major depression and who does not. In particular, experiencing a stressful life event (SLE) in childhood or adulthood has been shown to increase depression risk<sup>4,5</sup>. SLEs include childhood physical, sexual or emotional abuse, death of a relative, severe illness, divorce or separation, economic deprivation and forced exit from the workforce. Events can cause stress if they occur to the individual (e.g. child abuse, divorce, severe illness) or if they happen to a close relative, particularly during childhood (e.g. divorce or severe illness in a parent). Each of these events has been shown to be associated with increased risk for depression<sup>6,7</sup>; however, the cumulative burden of stress is particularly relevant for determining depression risk. Studies have consistently shown that as the number of SLEs increases, risk for depression also increases, and individuals with over 4 SLEs experience depression risks 3-5 times that of individuals with no SLEs<sup>6,8,9</sup>.

Historically there has been great interest in the possibility of an interaction between stressful life events and genetic liability as risk factors for depression. Such an interaction, if present, could lead not only to a better understanding of the underlying etiology of depression, but also potentially be useful for identifying individuals at particularly high risk for developing depression. An early twin study<sup>10</sup> found that risk for depression after an SLE was only elevated among individuals with high genetic liability. Subsequently, researchers selected candidate genes they believed were associated with depression risk and examined whether variants in these genes interacted with SLEs to predict depression<sup>11-19</sup>. These studies yielded inconsistent results, with even meta-analyses reaching different conclusions regarding the validity of the associations<sup>20-27</sup>. Research examining the interaction between polygenic risk scores and SLEs have also yielded inconsistent results, with some finding evidence for interaction<sup>28-30</sup> and some failing to do so<sup>29,31-33</sup>.

The hypothesis-driven (i.e. candidate gene) approach for identifying specific variants associated with a given outcome has not been successful in psychiatric research<sup>34-36</sup>. This led to the embrace of the genome-wide-association-study (GWAS) as a method for identifying variants associated with psychiatric disorders in a theoretically agnostic fashion. In a GWAS, single nucleotide polymorphisms (SNPs) in sufficient linkage disequilibrium to tag the entire genome are tested for association with the outcome of interest. Significance is evaluated based on an adjusted alpha level to avoid false positive results. This method has been highly successful in psychiatric genetics, and has led to the identification of over 100 variants associated with major depression at the genome-wide significant alpha level<sup>37</sup>. Thus far, GWAS studies have failed to replicate any findings from candidate gene-by-stress interaction studies<sup>38</sup>.

To our knowledge, four prior studies have used this theoretically agnostic, genome-wide approach to evaluate whether individual genetic variants interact with SLEs as risk factors for depressive symptoms measured using symptoms scales including the Beck Depression Inventory (BDI), the General Health Questionnaire (GHQ) and the Centers for Epidemiological Studies Depression Scale (CES-D): Dunn and colleagues<sup>39</sup> conducted a genome-wide environment interaction study (GWEIS) of depressive symptoms in a sample of 7,179 African American and 3,138

Hispanic/Latina women. They identified one genome-wide significant SNP in the African-American sample near the CEP350 gene (rs4652467,  $p = 4.10 \times 10^{-10}$ ), however this association did not replicate. Ikeda and colleagues<sup>40</sup> conducted a GWEIS of depressive symptoms and SLEs in 1,088 individuals recruited from among employees of the Fujita Health University Hospital in Japan. The authors reported a significant interaction for a SNP near the BMP2 gene (rs10485715,  $p = 8.2 \times 10^{-9}$ ), however no attempts were made to replicate this result. Otowa and colleagues<sup>41</sup> conducted a GWEIS of depressive symptoms and SLEs in 320 Japanese individuals, with no genome-wide significant results. Most recently, Arnau-Soler and colleagues conducted GWEISs of depressive symptoms and SLEs in 4,919 Europeans from the Generation Scotland cohort and 99,057 Europeans from the UK Biobank<sup>42</sup>. The authors found 2 SNPs significant for interaction at the genome-wide level in the Generation Scotland sample: one near the PIWIL4 gene ( $p = 4.95 \times 10^{-9}$ ) and one intronic to the ZCCHC2 gene ( $p = 1.46 \times 10^{-8}$ ). They found no genome-wide significant hits in the UK Biobank, and the significant hits from the Generation Scotland Sample did not replicate in the UK biobank.

Most of these GWEIS studies had sample sizes that most likely left them underpowered to detect significant interaction results. In addition, the outcome of all of these studies was depressive symptoms rather than clinically defined major depression. Although depressive symptoms are highly genetically correlated with major depressive disorder<sup>43</sup>, they nevertheless are a distinct outcome with, potentially, distinct associations with individual SNPs. Furthermore, all of these studies relied, out of necessity, on measures of SLEs that were retrospective and therefore potentially subject to recall bias<sup>44,45</sup>. Finally, prior GWEIS studies were not able to account for the time-dependent nature of both SLEs and depression. SLEs can occur at multiple points during the lifespan, and analytic strategies that fail to account for this can potentially be subject to bias. GWAS has traditionally used logistic regressions to calculate odds ratios for the associations between individual SNPs and the odds of being a case. However, this approach does not measure risk for developing the disorder, which is arguably more useful from a clinical and public health standpoint<sup>46</sup>. A different methodological approach is therefore needed to determine the associations between individual SNPs and risk for developing major depression, as well as potential interactions between SNPs and SLEs as risk factors for developing major depression.

Our aim in this study was to examine interactions between individual SNPs and a time-dependent, prospective measure of SLEs as risk factors for major depression in the general population. To accomplish this, we used data from the iPSYCH2012 case-cohort sample – a population-based cohort of individuals living in Denmark which includes information on psychiatric diagnoses from hospital-based settings. In addition, we also conducted a GWAS of depression using survival analysis rather than logistic regression as the underlying statistical methodology to examine the associations between individual SNPs and risk for developing depression in the general population.

## 2. Methods

### 2.1 Study design and sample

Data were drawn from the iPSYCH2012 study, which has a case-cohort design<sup>47</sup>. In this design, the study sample is nested within a larger base population, and includes all cases from the full cohort but only a subset of non-cases. This reduces the cost and burden associated with collecting biological specimens (in the case of iPSYCH, DNA for genetic analysis). The subset used as the

comparison group is typically a random sample of individuals drawn from the full cohort (i.e., the ‘subcohort’). Because it is random, some cases will by chance be selected as part of the subcohort. The great benefit of this design over a nested case control design is that it enables the unbiased calculation of risk and hazard ratios, as in a cohort study<sup>48</sup>. Because not all members of the full cohort are included, this design can be more efficient and cost effective than a cohort study particularly when the collection of biological samples is involved<sup>48-50</sup>.

The iPSYCH2012 case-cohort sample includes a subcohort of 30,000 individuals selected randomly from the base population of all individuals born in Denmark between 1981-2005 who survived to their first birthday and had known mothers (i.e. the ‘subcohort’). To this random sample was added all additional individuals from the base population (N=56,189) who received a diagnosis of affective disorder, schizophrenia, autism, or ADHD between 1994-2012 in inpatient, outpatient or emergency room settings in Danish psychiatric hospitals (i.e. ‘cases outside the subcohort’). Records of psychiatric diagnoses are stored in the Danish Psychiatric Central Research Register (DCPRR)<sup>51</sup>. Around 4% of individuals in the subcohort (n=1,188) also received one of the above psychiatric diagnoses, bringing the total number of psychiatric cases in the sample to 57,377. Biological material for DNA analysis was linked to information from national population-based registers using the unique, personal identification number assigned to all Danish citizens and legal residents since 1968 by the Danish Civil Registration System (DCRS)<sup>52</sup>. The DCRS also includes parents’ personal identification numbers allowing establishment of all known first degree relatives (parents, siblings, half-siblings and offspring).

For this study, we selected all individuals in the iPSYCH2012 subcohort and all depression cases (ICD-10 codes F32-F33) outside the subcohort who were a) of European ancestry, b) successfully genotyped, and c) for whom follow-up data starting at age 10 years was available. We also removed at random 1 person from each pair of relatives (2<sup>nd</sup> degree or closer,  $\pi$  hat > 0.2). The final study sample included 38,716 individuals: 20,563 individuals from the subcohort (of whom 379 were also cases) and 18,153 depression cases outside the subcohort (total number of cases = 18,532).

## 2.2 Ethics statement

The iPSYCH2012 study was approved by the Danish Scientific Ethics Committee, the Danish Health Data Authority, the Danish Data Protection Agency, and Danish Newborn Screening Biobank Steering Committee. In accordance with Danish law, The Danish Scientific Ethics Committee has for this study waived the need for informed consent in biomedical research based on existing biobanks.

## 2.3 Measures

*2.3.1 Stressful life events.* SLEs included death of a parent, sibling or child; serious medical illness in the participant or one of their first-degree relatives; family disruption due to divorce or separation, parental disability, and child maltreatment. SLE variables were obtained from Danish National population-based registers<sup>51,53,54</sup>. A detailed description of how each SLE was measured is shown in [Table 1](#). Dahl et al.<sup>6</sup> examined these events in the Danish registers and found that all were associated with depression risk individually, and that the number of SLEs was associated with depression in a dose-response fashion<sup>6</sup>. Information on SLEs was combined into a time-

varying count variable (0-4+ events), such that individuals contributed person time to the analyses within whichever category of SLE they were in at that time, and switched to contribute person time within a different SLE category when they experienced a subsequent SLE.

**2.32 Genetic data.** DNA was obtained from blood spots collected at birth as part of routine clinical screening and stored in the Danish Newborn Screening Biobank<sup>55</sup>. Bloodspots were located for 80,422 members (93%) of the iPSYCH2012 sample. Samples were genotyped at The Broad Institute of Harvard and MIT (Cambridge, MA, USA) in 23 waves using the Infinium PsychChip v1.0 array (Illumina, San Diego, CA, USA). Quality control (QC) and imputation were performed using the RICOPILI pipeline<sup>56</sup>. The filtering process excluded variants with call frequency  $< 0.98$  or a Hardy-Weinberg Equilibrium  $P$  value  $< 1 \times 10^{-6}$ . Ninety percent ( $N=77,639$ ) of the sample passed QC.

## 2.4 Analyses

Main and interaction effects for the associations between individual SNPs, SLEs and depression were estimated using a series of Cox regressions weighted according to the Prentice method to accommodate the case-cohort design<sup>48,49</sup>. Persons in the study sample were followed from age 10 until depression diagnosis, death, emigration, or December 31, 2012, whichever came first. The underlying time-metric was age in days. The time-dependent SLE count variable was analyzed as a continuous variable. All analyses were adjusted for sex, birth year and the first 5 ancestral principal components. Wald statistics were used to test for interaction. Analyses were conducted in R (version 3.1.2; R Foundation, Vienna, Austria). Regional visualizations of results from GWEIS analyses were plotted with LocusZoom<sup>57</sup>.

There are approximately 11 million directly genotyped and imputed SNPs available for members of the iPSYCH2012 sample. However, according to Danish law, some register-based data is available only at dedicated servers at Statistics Denmark (Copenhagen, DK). Because this study includes variables that can only be accessed through these servers, we were required to conduct the analyses in a Windows environment, which created some computational challenges that made it impossible to run our GWAS and GWEIS analysis in the full set of 11 million SNPs. To get around these challenges, we conducted our GWEIS of SLEs and depression in two stages: first, we selected a subset of SNPs where  $MAF > .01$  and missing rate  $< .1$ . From there, we conducted linkage disequilibrium (LD) pruning with various  $r^2$  thresholds and found that an  $r^2$  value of 0.7 left us with 496,162 high quality SNPs distributed across the genome. These SNPs were then uploaded onto the Statistics Denmark servers and merged with the register-based data for GWAS and GWEIS analysis. Based on the GWEIS analysis using these 496,162 SNPs, we identified all SNPs with interaction  $p$ -values below the commonly used genome-wide suggestive threshold ( $P < 1 \times 10^{-5}$ ). We then went back to the original sample of 11 million SNPs and identified all those SNPs located 500 kb upstream or downstream of the suggestive SNPs and uploaded them onto the server at Statistics Denmark. This enabled a second stage of analysis in which there was dense coverage of the areas with suggestive evidence for interaction. For this second stage, statistical significance was evaluated at the genome-wide significant alpha level of  $P < 5 \times 10^{-8}$ . Given the actual number of SNPs included in our GWEIS and the fact that the second stage of SNP selection specifically aimed to increase coverage of specific genomic areas, we posit  $P < 5 \times 10^{-8}$  to be a conservative threshold.



**2.41 Replication attempt.** We attempted to replicate our top findings in a case-control sample of depression drawn from the UK Biobank<sup>58</sup>. The UK Biobank includes more than 500,000 persons aged 40–69 at recruitment and holds a variety of biological measurements, lifestyle indicators and biomarkers, including genome-wide genotype data on all participants. The current replication analyses were based on a sample of 73,258 genetically unrelated persons of European ancestry (22,880 depression cases and 50,378 controls)<sup>59</sup> for whom SNP data as well as information on trauma exposure were available<sup>60</sup>. Lifetime depression was assessed with questions from the Composite International Diagnostic Interview. Trauma exposure was operationalized as a dichotomous variable based on self-report of severe trauma experiences in childhood and adulthood. We tested for interaction between the dichotomous trauma exposure and all available SNPs located within +/- 500 kb of the most significant SNP from each of the three genome-wide significant loci identified in the iPSYCH2012 GWEIS. In total, 7745 SNPs were tested for interaction using PLINK2a<sup>61</sup>. We assessed the number of independent loci tested for interaction at varying  $r^2$  (0.1-0.5) and differently sized windows (250-3000Kb) yielding 443 to 1252 independent loci (See [Table 2](#)).

### 3. Results

#### 3.1 Sample characteristics

Sample characteristics are shown in [Table 3](#). Collectively, participants in the sample contributed a total of 395,331 person-years of follow-up, with an average follow-up time of 10 years (standard deviation = 5 years). Depression cases inside and outside the population-based random subcohort showed similar characteristics. Sixty-nine percent of cases and 49% of subcohort members were female. Mean age at first depression diagnosis in cases was 19.6-19.7 years ( $SD = 4.1-4.2$  years). SLEs were common – by age 10, 48-49% of the cases and 39% of subcohort members had experienced at least one SLE.

#### 3.2 GWAS results

[Figure 1](#) shows results from GWASs examining the main effects of 496,162 SNPs on hazard of depression ([Figure 1a](#)) and hazard of experiencing at least one SLE ([Figure 1b](#)). The GWAS of the risk for developing depression yielded one genome-wide significant hit (rs7700661,  $P = 1.99 \times 10^{-8}$ ) and 52 suggestive hits ( $P < 1 \times 10^{-5}$ ) ([Figure 1a](#)). No individual SNPs reached genome-wide or suggestive significance for the main effect of SNPs on hazard of SLEs ([Figure 1b](#)).

#### 3.3 GWEIS results

The GWEIS analysis of 496,162 SNPs yielded 60 SNPs that reached suggestive significance ( $P < 1 \times 10^{-5}$ , [Table 4](#)). After re-running the GWEIS including all SNPs located within 500 Kbs of these 60 SNPs, three independent loci reached genome-wide significance ([Figure 2](#)). Hazard ratios for the 3 top hits are shown in [Figure 3](#), and region plots in [Figure 4](#). The top hit, rs56076205 ( $P = 3.7 \times 10^{-10}$ ), was located in an intron of the ATP-binding cassette transporter C 1 (ABCC1) gene. Compared with homozygotes for the major allele and heterozygotes, homozygotes for the minor allele (minor allele frequency (MAF)= 0.07) showed a pattern of increasing hazard ratio (HR) for depression with increase number of SLEs (see [Figure 3a](#)). The second hit, rs3784187 ( $P = 1.2 \times 10^{-8}$ ), was located in an intron of the A-kinase anchor protein 6 (AKAP6) gene. For this SNP,

homozygotes for the minor allele (MAF 0.06) showed a negative interaction such that as SLEs increased, risk for depression decreased (see [Figure 3b](#)). The final hit, rs340315 ( $P = 4.5 \times 10^{-8}$ ), was located near the Major facilitator superfamily domain 1 (MFSD1) gene. Homozygotes for the minor allele (MAF 0.31) showed a similar pattern to the first hit, in which HR for depression increased as number of SLEs increased (see [Figure 3c](#)).

### 3.4 Analysis of top SNPs in UK Biobank

None of the three top SNPs were statistically significant in the replication attempt using UK Biobank data (rs56076205,  $P = 0.87$ ; rs3784187,  $P = 0.93$ ; rs340315,  $P = 0.71$ ). The most significant interactions involved the following SNPs: rs190869692 ( $P = 3.2 \times 10^{-5}$ ) in the *ABCC1* gene 38,653 bases upstream from the iPSYCH2012 hit in the same gene ( $r^2 = 0.002$ ;  $P = 0.58$ ); rs111284027 ( $P = 9.4 \times 10^{-5}$ ) in the *ARHGAP5* gene 259,273 bases downstream from our hit in the *AKAP6* gene ( $r^2 = 0.003$ ;  $P = 0.44$ ); rs146472082 ( $P = 5.1 \times 10^{-5}$ ) in the *RARRES1* gene 155,569 base pairs downstream from our hit near the *MFSD1* gene ( $r^2 = 0.053$ ;  $P = 0.0011$ ) (See [Figure 5](#)). Thus, all three SNPs identified in the replication analyses represented independent loci from the three genome-wide significant loci identified in the iPSYCH2012 GWEIS.

## 4. Discussion

In this study, we report results from the first comprehensive, population-based genome-wide environment interaction study investigating the interaction between individual SNPs and a time-varying measure of stressful life events (SLEs) as risk factors for a diagnosis of depression treated in inpatient, outpatient or emergency room settings. The GWEIS yielded genome-wide significant effects in three independent loci located in the *ABCC1*, *AKAP6* and *MFSD1* genes, as well as 50 hits of suggestive significance. The *ABCC1* is known as a multidrug resistance protein and has a range of commonly used drugs as substrate<sup>62</sup>. Mice studies report a strong influence of *ABCC1* on cerebral accumulation of beta-amyloid<sup>63</sup>. One study also reported that St. John's wort (*Hypericum perforatum*), which has an antidepressant effect<sup>64</sup>, increased the transport efficacy of *ABCC1*<sup>65</sup>. This interaction is compatible with the idea of 'plasticity' genetic variants associated with stronger responses to both positive—in this study the absence of SLEs—and negative environmental exposures<sup>66</sup>. Theoretically, it would be difficult to identify a main effect of genetic variants showing such plasticity. The protein transcribed from the *AKAP6* gene is involved in intracellular signaling in the protein kinase A pathway and *AKAP6* is highly expressed in brain tissue, cardiac and skeletal muscle<sup>67</sup>. In 2015, a meta-analysis from the CHARGE consortium found a genome-wide significant association between a SNP in the *AKAP6* gene and general cognitive functioning<sup>68</sup>. The *MFSD1* is a membrane-bound solute carrier present in a wide range of human tissues<sup>67</sup>. A recent mice study reported the *MFSD1* to be abundant in the plasma membrane of neurons<sup>69</sup>. Further, the study found alterations in gene expression in response to environmental stress in the form of starvation.

None of the top hits from the iPSYCH2012 GWEIS analysis were significant in the UK Biobank. Although a successful replication of one or more of these hits would have provided convincing support for their validity, the absence of successful replication in this case is difficult to interpret due to the large differences between iPSYCH2012 and UK Biobank in terms of sampling, measurement, and design. The fact that different statistical methods were used (survival analysis vs. logistic regression) could also be a contributing factor. However, it is not straightforward to



isolate the impact of the statistical method alone, because conducting a logistic regression in our own sample would require us to make substantial changes to the design and sample composition. Thus it would be difficult to tell if any difference in the results were due to the different statistical method or the different design. Ultimately, a more solid conclusion regarding the validity of these hits will only be obtained through a replication attempt in a sample with comparable measurement and design. Such a sample will hopefully be available in the near future, but until that time, these hits should be considered suggestive.

To our knowledge, this is the largest single-sample GWEIS conducted to date examining the interaction between individual variants and SLEs. Nevertheless, the presented analyses are still likely underpowered to detect most single SNP gene-environment interactions<sup>70</sup>. For years, GWASs were similarly underpowered to detect significant SNPs, until the development of large-scale international consortia allowed for the accumulation of enough samples to pass the inflection point for consistent findings<sup>71</sup>. In comparison, the study of gene-environment interaction in psychiatric disorders has only begun to enter into its ‘big data’ phase. The requirement for assessment of a complex, composite environment exposure in the large study populations inherently required to study interactions makes these studies challenging endeavors. Extrapolating from the history of GWASs in psychiatry, we believe that the inflection point for studies of gene-environment interaction will only be reached through international collaborations that combine studies with information on genetic variation and environment exposures.

#### **4.1 Methodological considerations**

The following are additional methodological aspects of the study that should be kept in mind when interpreting these results. First, the oldest depression cases in the iPSYCH2012 sample were diagnosed by age 30. As such, they represent a cohort of early onset depression cases and therefore these results may not generalize to individuals who develop depression at older ages. Second, the depression cases in iPSYCH are all identified in hospital-based settings, therefore these results may not generalize to individuals with untreated depression or individuals treated solely by their primary care doctors, who make up the majority of depression cases in Denmark<sup>72</sup>. Fourth, although some of the SLEs included in this study are measured with high accuracy (e.g. death of a relative), others, particularly child maltreatment, are measured less accurately because they are based solely on register data. It is sadly very likely that some individuals in the sample experienced child maltreatment that was never recorded in the register, although the opposite (that individuals registered as having experienced child maltreatment did not experience it) is unlikely to be true. Third, we included a diverse range of stressful events in our study. Consequently, it is possible that some observed interactions relate to very specific types of stressful life events. For example, it is plausible that risk for depression in relation to somatic disease is associated with the seriousness of the course of disease. Therefore, genetic variants associated with prognosis and/or treatment response could emerge as part of gene-environment interaction in the present study. E.g. the *ABCC1* has a range of anti-cancer and anti-HIV drugs as substrates, thus rendering somatic treatment less effective thereby possibly increasing risk for depression.

#### **4.2 Conclusion**

In this population-based cohort of European ancestry, we identified 3 novel genetic loci that interacted with a time-varying measure of stressful life events to predict hospital-treated

depression at a genome-wide significant level, and over 50 additional loci with suggestive evidence for interaction. These results await confirmation via replication in a study with sufficiently similar measures and design. Future gene-by-stress research in depression should focus on efforts to establish large collaborative genome-wide environment interaction studies to generate sufficient statistical power to identify significant variants.

## Acknowledgments and Disclosures

The iPSYCH project is funded by the Lundbeck Foundation (grant numbers R102-A9118, R155-2014-1724 and R248-2017-2003) and the universities and university hospitals of Aarhus and Copenhagen. Katherine Musliner is funded by a postdoc fellowship from The Lundbeck Foundation (R303-2018-3551). Genotyping of the iPSYCH2012 samples was supported by grants from the Lundbeck Foundation, the Stanley Foundation, the Simons Foundation (SFARI 311789), and the National Institutes of Mental Health (NIMH 5U01MH094432-02). The Danish National Biobank resource is supported by the Novo Nordisk Foundation. The authors gratefully acknowledge The Broad Institute for genotyping. Initial genetic analyses were performed on the GenomeDK HPC facility supported by Center for Genomics Personalized Medicine and iSEQ, Aarhus University. The UK Biobank (Project ID 16577) represents independent research supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. High performance computing facilities at the NIHR BRC were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

Thomas Werge has served as scientific advisor to H. Lundbeck A/S. Gerome Breen has received consultancy and speaker fees from Eli Lilly, Otsuka, and Illumina. The rest of the authors have no conflicts of interest to report.

## References

1. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. *Annu Rev Public Health*. 2013;34:119-138.
2. Hasin DS, Sarvet AL, Meyers JL, et al. Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. *JAMA Psychiatry*. 2018;75(4):336-346.
3. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry*. 2000;157(10):1552-1562.
4. Kessler RC. The effects of stressful life events on depression. *Annual Review of Psychology*. 1997;48:191-214.
5. Hammen C. Stress and depression. *Annual Review of Clinical Psychology*. 2005;1:293-319.
6. Dahl SK, Larsen JT, Petersen L, et al. Early adversity and risk for moderate to severe unipolar depressive disorder in adolescence and adulthood: A register-based study of 978,647 individuals. *Journal of Affective Disorders*. 2017;214:122-129.
7. Kessler RC, Davis CG, Kendler KS. Childhood adversity and adult psychiatric disorder in the US National Comorbidity Survey. *Psychol Med*. 1997;27(5):1101-1119.
8. Anda RF, Felitti VJ, Bremner JD, et al. The enduring effects of abuse and related adverse experiences in childhood. A convergence of evidence from neurobiology and epidemiology. *European Archives of Psychiatry and Clinical Neuroscience* 2006;256(3):174-186.
9. Chapman DP, Whitfield CL, Felitti VJ, Dube SR, Edwards VJ, Anda RF. Adverse childhood experiences and the risk of depressive disorders in adulthood. *J Affect Disord*. 2004;82(2):217-225.
10. Kendler KS, Kessler RC, Walters EE, et al. Stressful life events, genetic liability, and onset of an episode of major depression in women. *Am J Psychiatry*. 1995;152(6):833-842.
11. Caspi A, Sugden K, Moffitt TE, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*. 2003;301(5631):386-389.
12. Chen J, Li X, McGue M. Interacting effect of BDNF Val66Met polymorphism and stressful life events on adolescent depression. *Genes Brain Behav*. 2012;11(8):958-965.
13. Drachmann Bukh J, Bock C, Vinberg M, Werge T, Gether U, Vedel Kessing L. Interaction between genetic polymorphisms and stressful life events in first episode depression. *Journal of Affective Disorders*. 2009;119(1):107-115.
14. Elovainio M, Jokela M, Kivimäki M, et al. Genetic variants in the DRD2 gene moderate the relationship between stressful life events and depressive symptoms in adults: cardiovascular risk in young Finns study. *Psychosom Med*. 2007;69(5):391-395.
15. Juhasz G, Hullam G, Eszlari N, et al. Brain galanin system genes interact with life stresses in depression-related phenotypes. *Proceedings of the National Academy of Sciences*. 2014;111(16):E1666-E1673.
16. Kim JM, Stewart R, Kim SW, et al. Interactions between life stressors and susceptibility genes (5-HTTLPR and BDNF) on depression in Korean elders. *Biol Psychiatry*. 2007;62(5):423-428.
17. Liu Z, Liu W, Yao L, et al. Negative life events and corticotropin-releasing-hormone receptor1 gene in recurrent major depressive disorder. *Scientific Reports*. 2013;3(1):1548.

18. Mandelli L, Serretti A, Marino E, Pirovano A, Calati R, Colombo C. Interaction between serotonin transporter gene, catechol-O-methyltransferase gene and stressful life events in mood disorders. *Int J Neuropsychopharmacol*. 2007;10(4):437-447.
19. Zimmermann P, Bruckl T, Nocon A, et al. Interaction of FKBP5 gene variants and adverse life events in predicting depression onset: results from a 10-year prospective community study. *Am J Psychiatry*. 2011;168(10):1107-1116.
20. Bleys D, Luyten P, Soenens B, Claes S. Gene-environment interactions between stress and 5-HTTLPR in depression: A meta-analytic update. *J Affect Disord*. 2018;226:339-345.
21. Culverhouse RC, Saccone NL, Horton AC, et al. Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression. *Mol Psychiatry*. 2018;23(1):133-142.
22. Hosang GM, Shiles C, Tansey KE, McGuffin P, Uher R. Interaction between stress and the BDNF Val66Met polymorphism in depression: a systematic review and meta-analysis. *BMC Med*. 2014;12:7.
23. Karg K, Burmeister M, Shedden K, Sen S. The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch Gen Psychiatry*. 2011;68(5):444-454.
24. Risch N, Herrell R, Lehner T, et al. Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA*. 2009;301(23):2462-2471.
25. van Ijzendoorn MH, Belsky J, Bakermans-Kranenburg MJ. Serotonin transporter genotype 5HTTLPR as a marker of differential susceptibility? A meta-analysis of child and adolescent gene-by-environment studies. *Transl Psychiatry*. 2012;2:e147.
26. Wang Q, Shelton RC, Dwivedi Y. Interaction between early-life stress and FKBP5 gene variants in major depressive disorder and post-traumatic stress disorder: A systematic review and meta-analysis. *J Affect Disord*. 2018;225:422-428.
27. Zhao M, Chen L, Yang J, et al. BDNF Val66Met polymorphism, life stress and depression: A meta-analysis of gene-environment interaction. *J Affect Disord*. 2018;227:226-235.
28. Peyrot WJ, Milaneschi Y, Abdellaoui A, et al. Effect of polygenic risk scores on depression in childhood trauma. *British Journal of Psychiatry*. 2014;205(2):113-119.
29. Mullins N, Power RA, Fisher HL, et al. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychological Medicine*. 2016;46(4):759-770.
30. Colodro-Conde L, Couvy-Duchesne B, Zhu G, et al. A direct test of the diathesis-stress model for depression. *Molecular Psychiatry*. 2018;23(7):1590-1596.
31. Musliner KL, Seifuddin F, Judy JA, Pirooznia M, Goes FS, Zandi PP. Polygenic risk, stressful life events and depressive symptoms in older adults: a polygenic score analysis. *Psychological Medicine*. 2015;45(8):1709-1720.
32. Peyrot WJ, Van der Auwera S, Milaneschi Y, et al. Does Childhood Trauma Moderate Polygenic Risk for Depression? A Meta-analysis of 5765 Subjects From the Psychiatric Genomics Consortium. *Biol Psychiat*. 2018;84(2):138-147.
33. Musliner KL, Andersen KK, Agerbo E, et al. Polygenic liability, stressful life events and risk for secondary-treated depression in early life: a nationwide register-based case-cohort study. *Psychol Med*. 2021:1-10.



34. Bosker FJ, Hartman CA, Nolte IM, et al. Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Mol Psychiatry*. 2011;16(5):516-532.
35. Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am J Psychiatry*. 2011;168(10):1041-1049.
36. Johnson EC, Border R, Melroy-Greif WE, de Leeuw CA, Ehringer MA, Keller MC. No Evidence That Schizophrenia Candidate Genes Are More Associated With Schizophrenia Than Noncandidate Genes. *Biol Psychiat*. 2017;82(10):702-708.
37. Howard DM, Adams MJ, Clarke TK, 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(3):343-352.
38. Border R, Johnson EC, Evans LM, et al. No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *Am J Psychiatry*. 2019:appiajp201818070881.
39. Dunn EC, Wiste A, Radmanesh F, et al. Genome-Wide Association Study (Gwas) and Genome-Wide by Environment Interaction Study (Gweis) of Depressive Symptoms in African American and Hispanic/Latina Women. *Depress Anxiety*. 2016;33(4):265-280.
40. Ikeda M, Shimasaki A, Takahashi A, et al. Genome-wide environment interaction between depressive state and stressful life events. *J Clin Psychiatry*. 2016;77(1):e29-30.
41. Otowa T, Kawamura Y, Tsutsumi A, et al. The First Pilot Genome-Wide Gene-Environment Study of Depression in the Japanese Population. *PLoS One*. 2016;11(8):e0160823.
42. Arnau-Soler A, Macdonald-Dunlop E, Adams MJ, et al. Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. *Translational Psychiatry*. 2019;9(1):14.
43. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*. 2018;50(5):668-681.
44. Baldwin JR, Reuben A, Newbury JB, Danese A. Agreement Between Prospective and Retrospective Measures of Childhood Maltreatment: A Systematic Review and Meta-analysis. *JAMA Psychiatry*. 2019;76(6):584-593.
45. Colman I, Kingsbury M, Garad Y, et al. Consistency in adult reporting of adverse childhood experiences. *Psychological Medicine*. 2016;46(3):543-549.
46. Murray GK, Lin T, Austin J, McGrath JJ, Hickie IB, Wray NR. Could Polygenic Risk Scores Be Useful in Psychiatry?: A Review. *JAMA Psychiatry*. 2020.
47. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Molecular Psychiatry*. 2018;23(1):6-14.
48. Self SG, Prentice RL. Asymptotic distribution theory and efficiency results for case-cohort studies. *The Annals of Statistics*. 1988;16(1):64-81.
49. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986;73(1):1-11.
50. Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case-cohort designs. *Journal of Clinical Epidemiology*. 1999;52(12):1165-1172.

51. Mors O, Perto GP, Mortensen PB. The Danish Psychiatric Central Research Register. *Scandinavian Journal of Public Health*. 2011;39(7 Suppl):54-57.
52. Pedersen CB. The Danish Civil Registration System. *Scandinavian Journal of Public Health*. 2011;39(7 Suppl):22-25.
53. Lyng E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scandinavian Journal of Public Health*. 2011;39(7 Suppl):30-33.
54. Petersson F, Baadsgaard M, Thygesen LC. Danish registers on personal labour market affiliation. *Scandinavian Journal of Public Health*. 2011;39(7 Suppl):95-98.
55. Norgaard-Pedersen B, Hougaard DM. Storage policies and use of the Danish Newborn Screening Biobank. *Journal of Inherited Metabolic Disorders*. 2007;30(4):530-536.
56. Lam M, Awasthi S, Watson HJ, et al. RICOPIILI: Rapid Imputation for CONsortias PIpeLine. *Bioinformatics*. 2020;36(3):930-933.
57. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-2337.
58. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209.
59. Warren HR, Evangelou E, Cabrera CP, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nature Genetics*. 2017;49:403.
60. Coleman JRI, Peyrot WJ, Purves KL, et al. Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank. *Molecular Psychiatry*. In press.
61. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
62. Loscher W, Potschka H. Drug resistance in brain diseases and the role of drug efflux transporters. *Nat Rev Neurosci*. 2005;6(8):591-602.
63. Krohn M, Lange C, Hofrichter J, et al. Cerebral amyloid-beta proteostasis is regulated by the membrane transport protein ABCC1 in mice. *J Clin Invest*. 2011;121(10):3924-3931.
64. Ng QX, Venkatanarayanan N, Ho CY. Clinical use of Hypericum perforatum (St John's wort) in depression: A meta-analysis. *J Affect Disord*. 2017;210:211-221.
65. Hofrichter J, Krohn M, Schumacher T, et al. Reduced Alzheimer's disease pathology by St. John's Wort treatment is independent of hyperforin and facilitated by ABCC1 and microglia activation in mice. *Curr Alzheimer Res*. 2013;10(10):1057-1069.
66. Belsky J, Jonassaint C, Pluess M, Stanton M, Brummett B, Williams R. Vulnerability genes or plasticity genes? *Mol Psychiatry*. 2009;14(8):746-754.
67. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419.
68. Davies G, Armstrong N, Bis JC, et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). *Mol Psychiatry*. 2015;20(2):183-192.
69. Perland E, Hellsten SV, Lekholm E, Eriksson MM, Arapi V, Fredriksson R. The Novel Membrane-Bound Proteins MFSD1 and MFSD3 are Putative SLC Transporters Affected by Altered Nutrient Intake. *J Mol Neurosci*. 2017;61(2):199-214.
70. Thomas D. Gene--environment-wide association studies: emerging approaches. *Nat Rev Genet*. 2010;11(4):259-272.

71. Sullivan PF, Agrawal A, Bulik CM, et al. Psychiatric Genomics: An Update and an Agenda. *American Journal of Psychiatry*. 2018;175(1):15-27.
72. Musliner KL, Liu X, Gasse C, Christensen KS, Wimberley T, Munk-Olsen T. Incidence of medically treated depression in Denmark among individuals 15-44 years old: a comprehensive overview based on population registers. *Acta Psychiatrica Scandinavica*. 2019;139(6):548-557.

**Table 1. Description of measures of stressful life events**

<b>SLE variable</b>	<b>Description</b>
Death of a parent, sibling or child	Information on vital status (i.e. alive or dead) and date of death for all individuals residing legally in Denmark has been recorded in the Danish Civil Registration system (DCRS) since 1968. This register also includes information that can be used to link people with their first-degree relatives. We defined death of a parent, sibling or child as the date of death for that individual recorded in the DCRS. Death of a parent, sibling or child were handled as separate events.
Serious medical illness in the participant or their parent, sibling or child	Information on hospital treatment, including ICD diagnoses, for all individuals residing legally in Denmark is recorded in the Danish National Patient Register (DNPR). The DNPR includes all inpatient hospital visits since 1978 and all outpatient and emergency hospital visits since 1995. We defined serious medical illness based upon the Charlson Comorbidity Index (CCI). This index includes 19 severe somatic diseases all assigned a weighted score representing the associated 1-year relative risk of dying. Diagnoses in the CCI have a very high accuracy in the DNPR. Individuals were coded as 0, 1, or 2+ based on their CCI score. For the participant and first-degree relatives, we considered dates of changes in CCI score from 0 to 1, and 1 to 2+ as separate events.
Family disruption	The annual cohabitation status (defined on January 1 each year) for all individuals residing legally in Denmark has been recorded in the DCRS since 1980. We defined family disruption in childhood as all instances where a person was registered as a child in a family with two cohabitating adults – of whom at least one was the legal parent of the child – who ceased to live together. Family disruption due to death of a parent was not included in this variable. Family disruption in adulthood was defined as the participant themselves ceasing to cohabit with an adult to whom they were married or, as many partnered adults in Denmark choose not to marry, with whom they had a child. Date of family disruption was defined as January 1 of the year in which cohabitation status changed.
Parental disability	Employment status for all individuals residing legally in Denmark has been registered each year on November 1 in the Danish register on personal labour market affiliation since 1981. Denmark offers a disability pension to individuals who are unable to work due to physical or mental disability. We measured date of parental disability as November 1 in the first year that a participant’s mother or father was registered as receiving a disability pension. Disability in the mother and father were treated as separate events.
Child maltreatment	Information on child maltreatment was obtained from the DNPR. We defined child maltreatment as the date at which the participant received one of the following ICD-10 codes: T74.0 - neglect or abandonment; T74.2, Z61.4, Z61.5 – sexual abuse; T74.1 – physical abuse; T74.3 – psychological abuse; T74.8, T74.9 – other or unspecified maltreatment syndromes.

**Table 2. Estimated number of independent loci tested for interaction in the replication analyses**

Independent loci	Kb	r <sup>2</sup>	Bonferroni adjusted p values
1252	250	0.5	3.99 x 10 <sup>-5</sup>
1232	3000	0.5	4.06 x 10 <sup>-5</sup>
1073	250	0.4	4.66 x 10 <sup>-5</sup>
1051	3000	0.4	4.76 x 10 <sup>-5</sup>
866	250	0.3	5.77 x 10 <sup>-5</sup>
842	3000	0.3	5.94 x 10 <sup>-5</sup>
687	250	0.2	7.28 x 10 <sup>-5</sup>
653	3000	0.2	7.66 x 10 <sup>-5</sup>
493	250	0.1	1.01 x 10 <sup>-4</sup>
443	3000	0.1	1.13 x 10 <sup>-4</sup>



**Table 3. Sample characteristics**

Characteristic	MD cases outside the subcohort (N = 18,153)	MD cases inside the subcohort (N = 379)	Non-cases from the subcohort (N = 20,184)
	N (%)	N (%)	N (%)
Gender			
Female	12,430 (68.5)	263 (69.4)	9,848 (48.8)
Male	5723 (31.5)	116 (30.6)	10,336 (51.2)
Birth cohort			
1981-1985	5953 (32.8)	126 (33.3)	3 585 (17.8)
1986-1990	6670 (36.7)	150 (39.6)	4 570 (22.6)
1991-2002	5530 (30.5)	103 (27.2)	12 029 (59.6)
>1 SLE before age 10 years	8,712 (48.0)	185 (48.8)	7,857 (38.9)
Age at first MD diagnosis, years (mean, SD)	19.7 (4.2)	19.6 (4.1)	NA

**Table 4: 60 SNPs passing genome-wide suggestive threshold ( $P < 1 \times 10^{-5}$ ) for interaction with stressful life events on risk for depression tested among 496,162 SNPs in 18,532 depression cases and 20,563 random subcohort members.**

Chr	Location (bp)	SNP	Interaction, $P$	Main effect depression, $P$	Main effect SLEs, $P$	A1	A2	MAF	Gene context
1	16650609	rs149334507	1.05E-06	0.0042	0.42	A	C	0.02	ARHGEF19---[FBXO42]--SZRD1
1	21937317	rs12083062	4.80E-06	0.81	0.72	T	C	0.05	ALPL--[RAP1GAP]--USP48
1	86628279	rs150960662	6.38E-06	0.92	0.59	A	C	0.02	COL24A1-[]---ODF2L
1	153310297	rs821433	5.60E-06	0.0014	0.56	G	A	0.10	PGLYRP3--[PGLYRP4]--S100A9
1	228022150	rs182670935	3.59E-06	0.13	0.17	G	A	0.01	SNAP47--[PRSS38]--WNT9A
1	247711911	chr1:247711911	2.34E-06	0.14	0.93	TGTT	CGTT	0.17	OR2C3--[GCSAML]--OR2G2
2	31549959	rs207426	6.48E-06	0.61	0.74	C	A	0.35	FADS1--[FADS2]--FADS3
2	125009457	rs79653267	3.30E-06	0.77	0.91	A	G	0.02	[CNTNAP5]---MTND5P22
2	150854878	rs149282157	5.78E-06	0.07	0.43	A	G	0.01	MMADHC---[]---RND3
2	159027173	rs10804390	6.21E-06	0.10	0.47	T	C	0.32	UPP2--[CCDC148
2	166023849	rs62174951	5.51E-06	0.19	0.97	G	A	0.11	SLC38A11---[SCN3A]--SCN2A
3	20622558	rs9846696	9.89E-06	0.0001	0.69	G	C	0.05	SGOL1---[ ]
3	22055173	rs61553318	3.29E-06	0.06	0.41	A	C	0.04	ZNF385D-AS2--[ZNF385D]---HMGB1P5
3	77675638	rs876675	6.40E-06	0.21	0.61	C	T	0.50	VDAC1P7---[ROBO2]---RP11-354H21.1
3	158545195	rs6792827	4.51E-06	0.04	0.85	T	G	0.13	RARRES1--[MFDS1]---IQCJ-SCHIP1
3	158551731	rs61796809	4.34E-06	0.11	0.98	A	G	0.22	MFSD1-[]---IQCJ-SCHIP1
3	158583584	rs340284	8.02E-06	0.0001	0.86	G	A	0.36	MFSD1--[ ]--IQCJ-SCHIP1
4	27574252	rs75065309	6.72E-06	0.08	0.23	A	G	0.01	RP11-415C15.2---[ ]--IGBP1P5

Chr	Location (bp)	SNP	Interaction, <i>P</i>	Main effect depression, <i>P</i>	Main effect SLEs, <i>P</i>	A1	A2	MAF	Gene context
4	59684939	rs116510933	6.43E-06	0.01	0.74	A	G	0.07	RP11-577G20.2---[]
4	69821738	rs1841036	5.07E-06	0.0031	0.21	T	G	0.16	UGT2A3-[]--UGT2B11
4	126117627	rs13110472	2.16E-07	0.13	0.60	T	C	0.06	ANKRD50---[]---FAT4
4	151061659	rs72730361	7.41E-06	0.09	0.99	T	C	0.02	RP11-423J7.1---[DCLK2]---LRBA
4	158141677	rs28545562	5.47E-06	0.03	0.69	C	T	0.02	GLRB--[GRIA2]---RP11-364P22.1
4	186959486	rs6818787	7.66E-06	0.30	0.08	C	A	0.40	SORBS2--[]--TLR3
5	33216242	rs28566539	7.23E-06	0.12	0.43	T	C	0.15	NPR3---[CTD-2066L21.3]---TARS
5	121067170	rs7735996	6.14E-06	0.16	0.75	G	A	0.06	RP11-510I6.3---[]---FTMT
5	178981060	rs72822583	8.08E-06	0.31	0.67	T	C	0.04	ADAMTS2---[RUFY1]--HNRNPH1
6	15849887	rs72823483	2.85E-06	0.24	0.45	A	G	0.01	DTNBP1---[]---MYLIP
6	107310381	rs9486484	4.53E-06	0.30	0.53	G	A	0.15	QRSL1---[]--C6orf203
7	21144220	rs73277532	5.16E-06	0.87	0.88	G	T	0.01	ABC5B-SP8---[]---SP4
7	91011858	rs73220765	9.71E-06	0.51	0.57	T	C	0.01	FZD1---[RP11-115N4.1][RP11-142A5.1]---MTERF1
8	32516140	rs35955476	4.40E-06	0.05	0.84	C	CAG	0.47	NRG1-IT3---[NRG1]---RP11-11N9.4
8	56535514	rs6474006	7.85E-06	0.49	0.59	C	T	0.40	XKR4--[]--TMEM68
8	103203727	rs4102400	3.89E-06	0.49	0.43	T	C	0.47	NCALD--[]--RRM2B
9	83000507	rs7861030	1.56E-07	0.0049	0.51	T	C	0.50	NPAP1P4-[]---RP11-117O7.2
9	83023317	rs10780394	6.22E-06	0.0002	0.50	G	A	0.32	NPAP1P4--[]---RP11-117O7.2
10	8286974	rs1796867	2.85E-07	0.0032	0.68	A	G	0.06	PRPF38AP1--[]--LINC00708
10	64266748	rs10995178	4.87E-07	0.00001	0.89	A	G	0.45	RTKN2--[ZNF365]---ADO

Chr	Location (bp)	SNP	Interaction, <i>P</i>	Main effect depression, <i>P</i>	Main effect SLEs, <i>P</i>	A1	A2	MAF	Gene context
10	129586689	rs1926181	3.86E-06	0.17	0.34	A	C	0.20	FOXI2--[]--CLRN3
11	13920438	rs61884777	8.35E-06	0.12	0.54	G	A	0.09	FAR1---[]--SPON1
11	41939498	rs142799494	1.93E-06	0.83	0.16	T	A	0.01	LRRC4C---[]--RP11-148I19.1
11	44032917	rs118008313	4.63E-06	0.53	0.56	T	C	0.03	C11orf96--[]--ACCSL
11	44452139	rs10769047	4.06E-06	0.05	0.73	A	T	0.50	ALX4---[]--CD82
11	45881397	rs139670444	1.12E-06	0.24	0.95	A	AG	0.05	SLC35C1--[CRY2]--MAPK8IP1
11	116604070	rs180353	5.53E-06	0.45	0.88	C	T	0.20	AP000770.1--[]--BUD13
11	128996355	rs7944939	1.48E-06	0.02	0.71	C	T	0.31	TP53AIP1---[ARHGAP32]---BARX2
12	119758130	rs140437928	3.41E-06	0.02	0.23	C	T	0.02	HSPB8--[]--CCDC60
13	51272084	rs797498	6.55E-07	0.06	0.58	A	G	0.08	DLEU1-AS1---[DLEU1]--DLEU7
13	114591051	rs9550266	4.89E-06	0.01	0.28	A	G	0.16	GAS6--[LINC00452
14	32860927	rs1951185	1.60E-06	0.08	0.94	T	C	0.06	ARHGAP5---[AKAP6][RP11-320M16.2]--RN7SL660P
15	35002935	rs16959528	6.12E-07	0.94	0.59	G	A	0.11	GOLGA8B---[]--GJD2
16	6338673	rs1344474	9.41E-06	0.86	0.39	G	A	0.12	[RBFOX1][RB11-420N3.3]
16	16172008	rs56076205	3.74E-10	0.05	0.55	T	C	0.07	FOPNL---[ABCC1]--ABCC6
16	63680366	rs12448930	3.17E-07	0.09	0.86	A	C	0.25	RP11-368L12.1--[]--RP11-370P15.1
17	70291156	rs1967304	5.85E-06	0.67	0.11	C	A	0.25	SOX9---[]---SLC39A11
18	37425523	rs2048647	3.52E-06	0.03	0.48	G	C	0.21	RP11-244M2.1--[RP11-636021.1]---LINC01477
18	77985650	rs111447074	1.30E-06	0.12	0.59	T	C	0.02	ADNP2--[PARD6G]

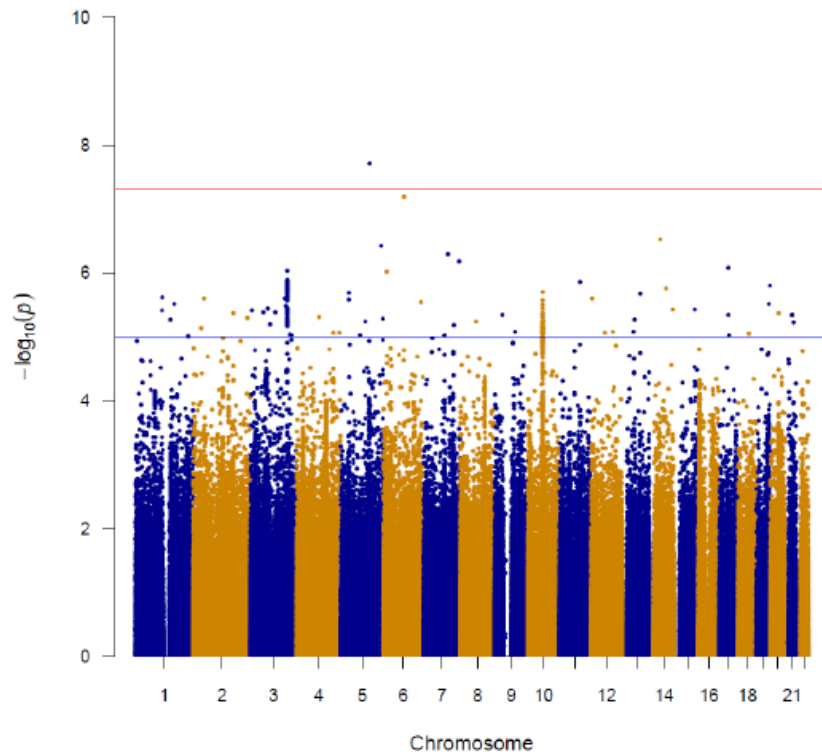
Chr	Location (bp)	SNP	Interaction, <i>P</i>	Main effect depression, <i>P</i>	Main effect SLEs, <i>P</i>	A1	A2	MAF	Gene context
19	46785290	rs112087991	4.06E-06	0.0005	0.60	C	T	0.05	IGFL1--[ ]--HIF3A
20	35329303	rs62206150	6.36E-06	0.01	0.35	G	A	0.02	SLA2--[NDRG3]--DSN1
21	31449079	rs117181045	4.86E-06	0.03	0.98	G	T	0.01	GRIK1---[ ]--CLDN17

Note. The 492,162 included SNPs were selected according to the following criteria: MAF > .01 and missing rate < .1; subsequently linkage disequilibrium (LD) pruning with an  $r^2$  value of 0.7 was implemented. SLEs: Stressful life events. The 'Gene context' column lists the SNP location with '[ ]'. Most closely located genetic variant 500kb upstream or downstream for the index SNP are listed as well with any genes prioritized over lincRNA, pseudogenes etc. Distance from index to other listed variants are denoted by dashes: no dash, <1kb; - <10kb; -- <100kb; --- <500kb.

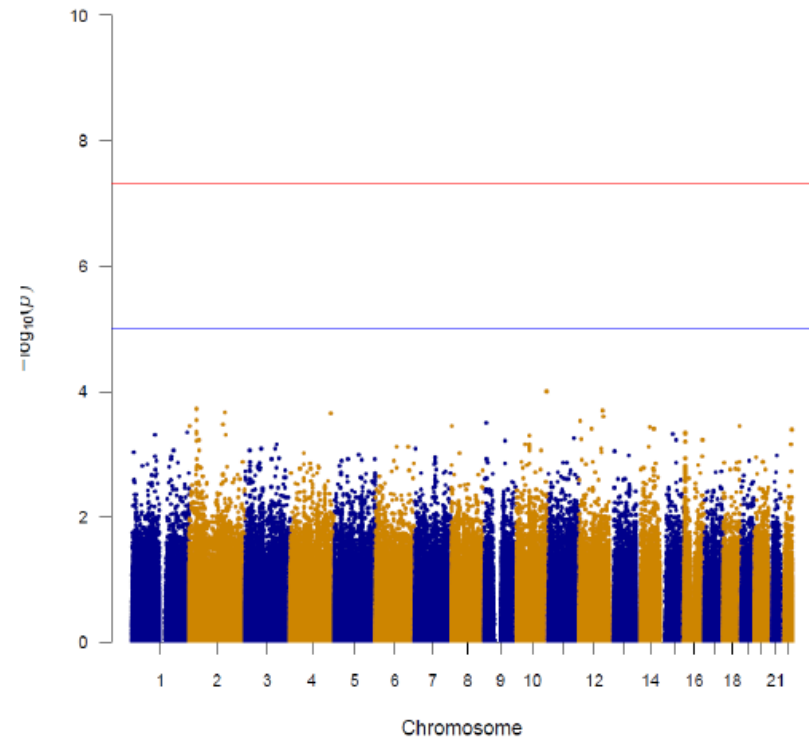


**Figure 1. Manhattan plots for main effects of 496,162 SNPs on risk for depression and stressful life events in 18,532 depression cases and 20,563 random subcohort members**

a) Main effects of SNPs on risk for depression

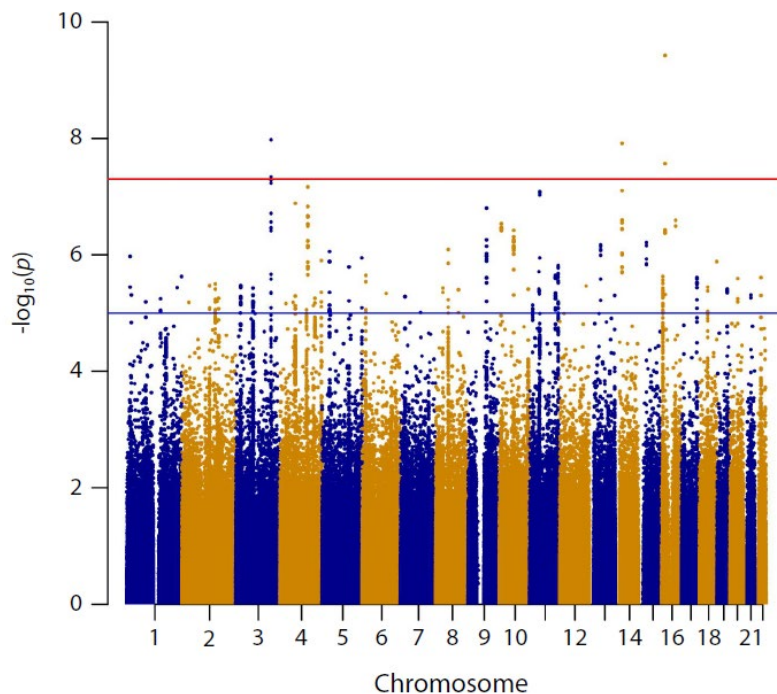


b) Main effects of SNPs on risk for SLEs



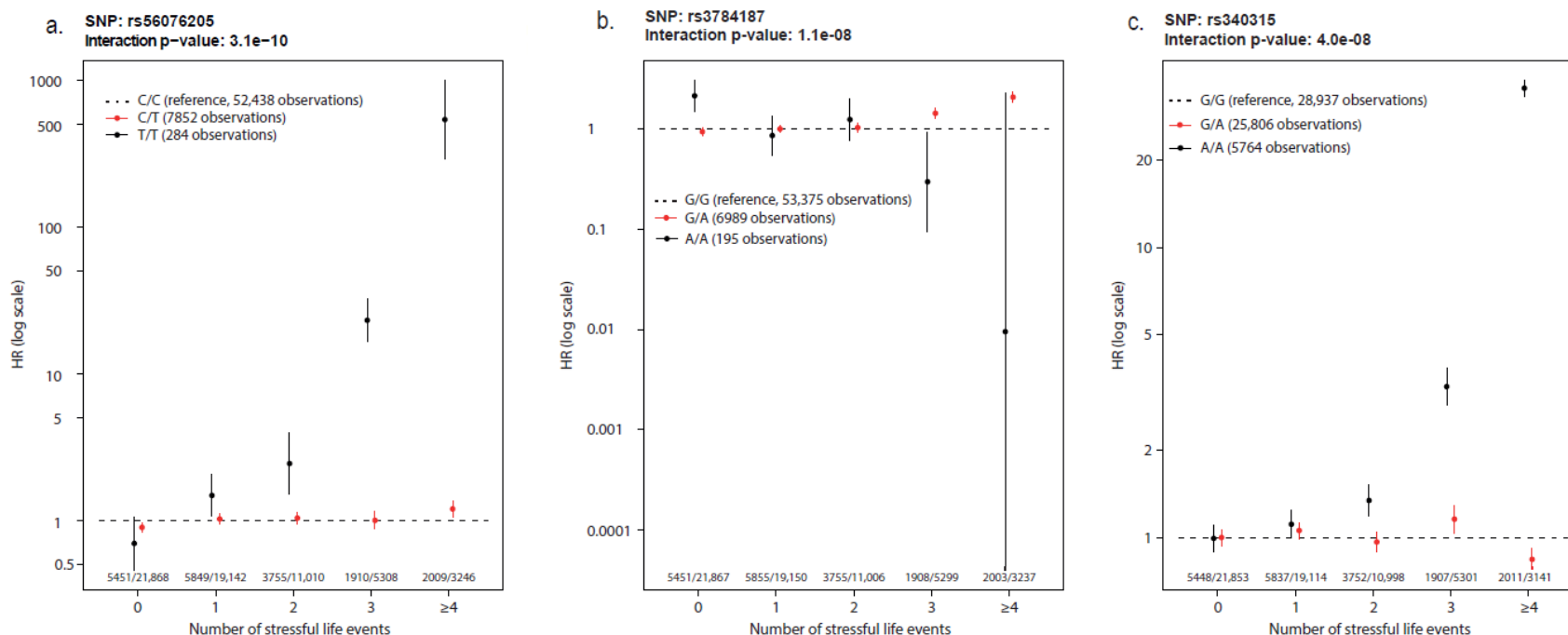
Note. Panel A shows main effects of 496,162 individual SNPs on risk for depression diagnosis in hospital-based settings in Denmark from 1995-2012. Panel B shows main effects of 496,162 individual SNPs on risk for experiencing at least one stressful life event.

**Figure 2. Manhattan plot of genome-wide environment interaction analyses based on 18,532 depression cases and 20,563 random subcohort members**



Note. The figure presents results of a genome-wide environment interaction study (GWEIS) conducted in two stages: In stage 1, a GWEIS was conducted using 496,162 SNPs distributed across the genome. In stage 2, all SNPs located 500kb up- or downstream from 60 SNPs with p values  $< 10^{-5}$  in stage 1 were added to the analyses. The Manhattan plot shows results from both stages.

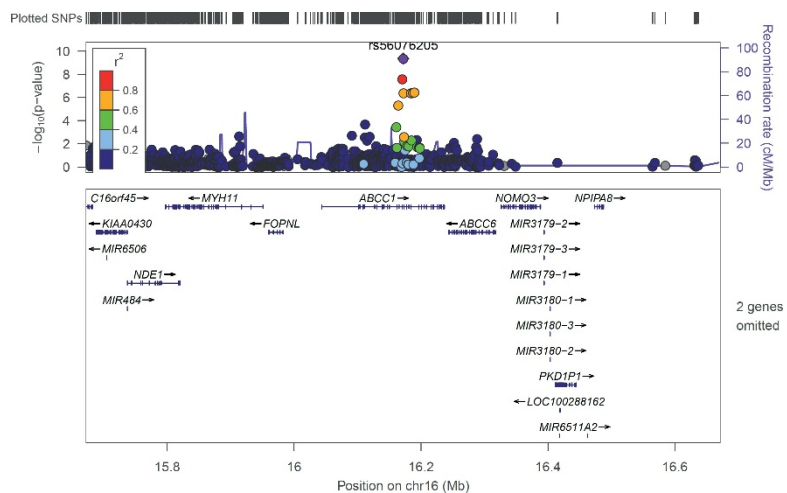
**Figure 3. Interaction effects for stressful life events and top SNPs from 3 genome-wide significant loci**



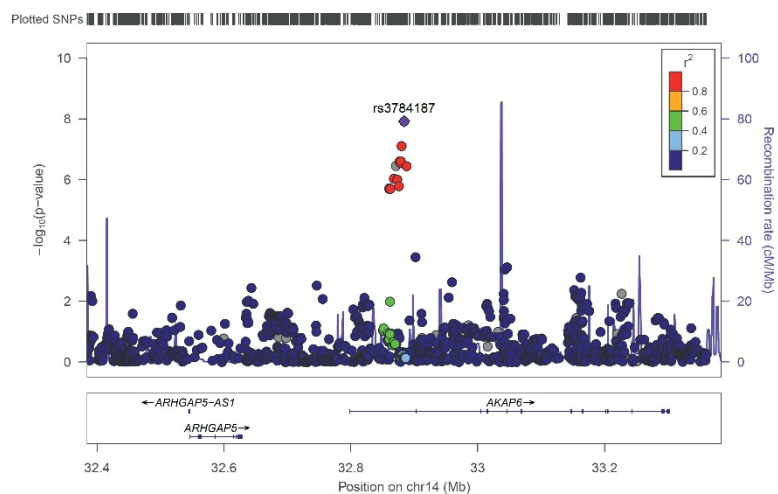
Note. For each SNP, the hazard ratio (HR) for depression is plotted by number of stressful life events. Vertical bars represent 95% confidence intervals. Hazards were compared within each level of stress life events with major allele homozygotes as reference. Wald statistics were used to test interactions, comparing linear trends for HR between genotypes. The small differences in the total number of observations are due to differences in the number of persons successfully genotyped for each SNP. Due to the time-varying nature of the stressful life events variable, study participants could contribute person-time for different numbers of stressful life events. Therefore, the total number of observations exceeds the total number of participants in the study.

**Figure 4. Region plots for 3 top hits from a genome-wide by environment interaction study based on 18,532 depression cases and 20,563 random subcohort members**

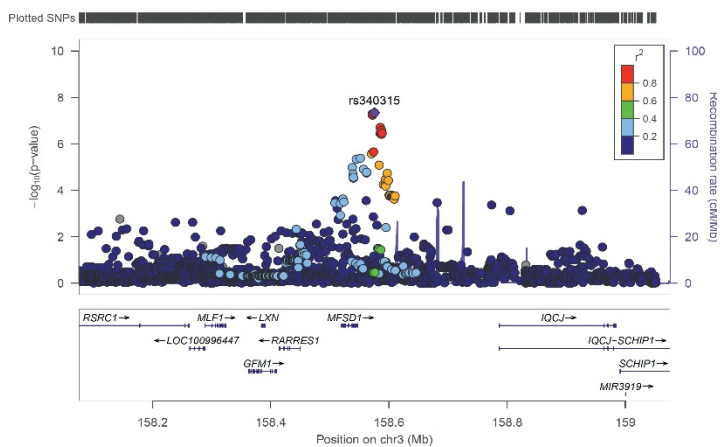
**(a) rs56076205**



**(b) rs3784187**



**(c) rs340315**

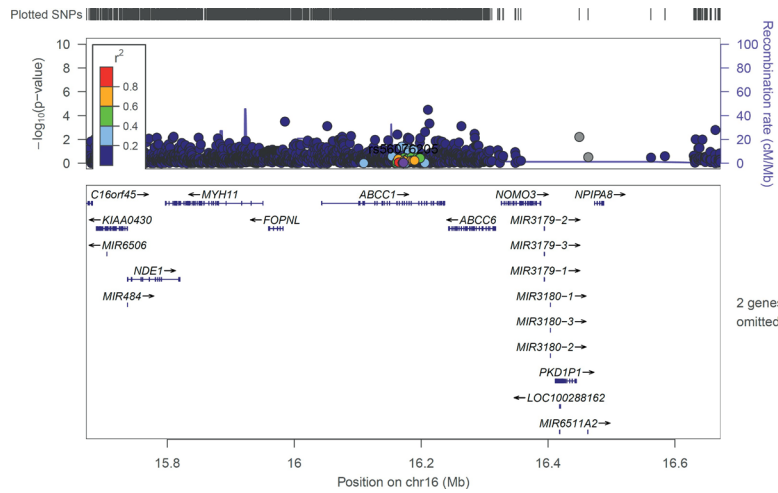


Note. Color of dots indicate the linkage disequilibrium ( $r^2$ ) of SNPs with the top SNP of each loci. The  $r^2$  was based on the 1000 Genomes November 2014 EUR population.

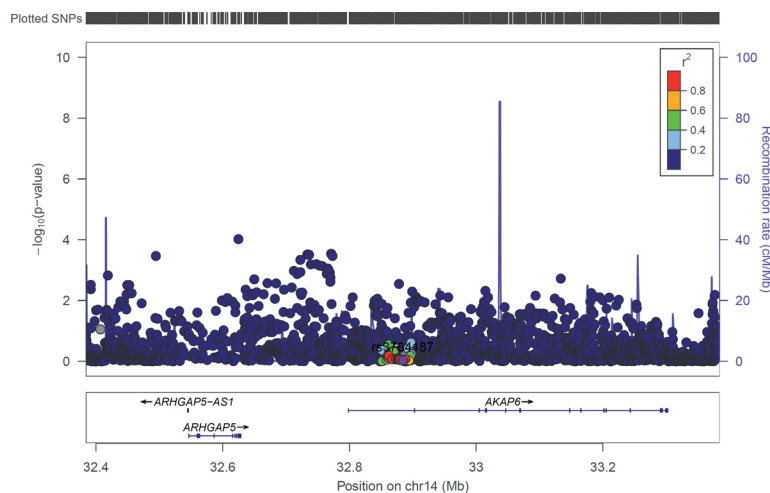


**Figure 5. Region plots for the three most significant SNPs in the UK Biobank replication attempt**

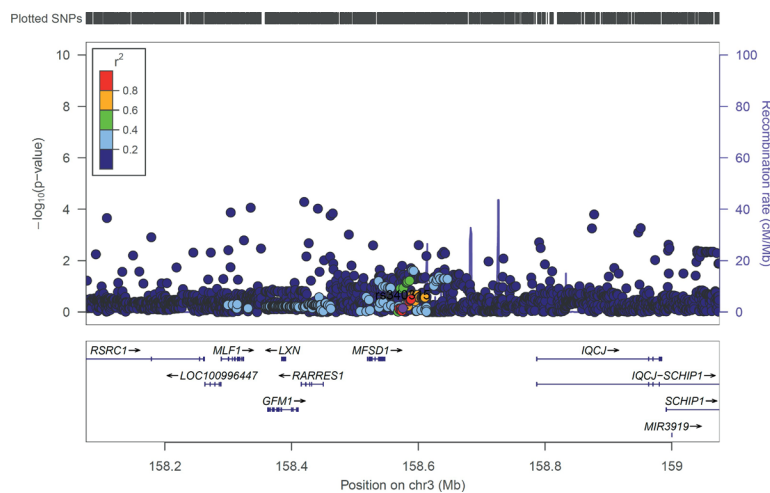
**(a) rs56076205**



**(b) rs3784187**



**(c) rs340315**



Note. SNPs identified with genome-wide significant interactions in analyses of iPSYCH data were used as index SNPs in the above regional plots of the UK Biobank replication attempt. Color of dots indicate the linkage disequilibrium ( $r^2$ ) of SNPs with the top SNP at each loci identified in the interaction analysis based on iPSYCH2012. The  $r^2$  was based on the 1000 Genomes Nov 2014 EUR sample.