**Supplementary Materials**

**Supplementary Fig 1.** Study Flow
**Supplementary Fig 2.** Evaluation on potential bias in a population-based cohort.

**Supplementary Table 1.** Definition of the Covariates

**Supplementary Table 2.** 55 risk SNPs associated with major depressive disorder and its symptoms

**eReference**

Supplementary Fig 1. Study Flow



Supplementary Fig 2. Evaluation on potential bias in a population-based cohort



a) t-SNE plot representing randomly selected participants from the three datasets. Each point corresponds to an individual participant, which colors indicating the dataset to which the participant belongs. Violin plot showing the distribution of demographic and clinical features across three datasets. b) Age, c) Townsend deprivation index, and d) body mass index.

Supplementary Table 1. Definitions of the covariates

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Covariates** | **UK Biobank Data-Field ID** | **UK Biobank Data-Field Description** | **Coding** | **Notes** |
| Age group | 21022 | Age at recruitment | Continuous age in years | Coded in three categories: 40-50, 50-60, >60 |
| Sex | 31 | Sex | 1) Female2) Male |  |
| Ethnicity | 21000 | Ethnic background | 1) White2) Mixed3) Asian or Asian British4) Black or Black British5) Chinese6) Other ethnic group | Dichotomized into White British and others |
| Townsend deprivation index | 22189 | Townsend deprivation index at recruitment | Continuous value |  |
| Employment status | 6142 | Current employment status (Instance 0) | 1) In paid employment or self-employed2) Retired3) Looking after home and/or family4) Unable to work because of sickness or disability5) Unemployed6) Doing unpaid or voluntary work7) Full or part-time student | Coded in two categories: ‘Employed (In paid employment or self-employed)’ and ‘Not employed (other answers)’ |
| Education level | 6138 | Qualifications (Instance 0) | 1) College or University degree2) A levels/AS levels or equivalent3) O levels/GCSEs or equivalent4) CSEs or equivalent5) NVQ or HND or HNC or equivalent6) Other professional qualifications eg: nursing, teaching7) None of the above | Years of education was derived from the highest qualification that each subject achieved, according to the International Standard Classification for Education (ISCED) definitions.(1) Education level was classified as high (more than 20 years of education), intermediate (11 to 19 years of education), and low (less than 10 years of education). |
| Smoking status | 20116 | Smoking status (Instance 0) | 1) Never2) Previous3) Current | Dichotomized into 1) Never smoked and 2) Ever smoked |
| Alcohol consumption | 1558 | Alcohol intake frequency (Instance 0) | 1) Daily or almost daily2) Three or four times a week3) Once or twice a week4) One to three times a month5) Special occasions only6) Never | Classified into four categories: 1) Never, 2) Special occasions only, 3) Moderate (one time a month to four times a week), and 4) Heavy (Daily or almost daily) |

Supplementary Table 2. 55 risk SNPs associated with major depressive disorder and its symptoms

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **rsID** | **Chr** | **Location** | **Ref** | **Alt** | **Mapped Gene** | **Reference** |
| rs10514299 | 5 | 8836793 | C | A,T | TMEM161B-DT | (2) |
| rs1518395 | 2 | 57980939 | A | G,T | VRK2 |
| rs2179744 | 22 | 41225710 | G | A,T | L3MBTL2, L3MBTL2-AS1 |
| rs11209948 | 1 | 72346221 | G | C,T | RPL31P12, RNU6-1246P |
| rs454214 | 5 | 88707586 | C | A,G,T | MEF2C-AS2 |
| rs301806 | 1 | 8422018 | C | T | RERE |
| rs1475120 | 6 | 104942078 | G | A | LIN28B, LIN28B-AS1 |
| rs10786831 | 10 | 104854813 | A | G | SORCS3 |
| rs12552 | 13 | 53051646 | A | C,G,T | OLFM4 |
| rs6476606 | 9 | 37005564 | A | G | PAX5 |
| rs8025231 | 15 | 37356201 | A | C,G,T |  |
| rs12065553 | 1 | 80327433 | A | G,T |  |
| rs1656369 | 3 | 158562296 | T | A | MLF1-DT |
| rs4543289 | 5 | 165057942 | T | G | LINC03000 |
| rs2125716 | 12 | 84547650 | A | G,T |  |
| rs2422321 | 1 | 73293393 | A | G |  |
| rs7044150 | 9 | 2982931 | T | A,C,G |  |
| rs159963 | 1 | 8504421 | C | A,G,T | RERE | (3) |
| rs1432639 | 1 | 72813218 | C | A,G,T | LOC105378797 |
| rs12129573 | 1 | 73768366 | C | A |  |
| rs2389016 | 1 | 80799329 | C | A,T |  |
| rs4261101 | 1 | 90796053 | G | A |  |
| rs9427672 | 1 | 197754741 | A | C,G,T |  |
| rs11682175 | 2 | 57987593 | T | C |  |
| rs1226412 | 2 | 157111313 | C | G,T | LINC01876 |
| rs7430565 | 3 | 158107180 | G | A | RSRC1 |
| rs34215985 | 4 | 42047778 | C | A,G | SLC30A9 |
| rs11135349 | 5 | 164523472 | C | T | PRMT1 |
| rs4869056 | 5 | 166992078 | G | A,C | TENM2 |
| rs9402472 | 6 | 99566521 | G | A,T |  |
| rs10950398 | 7 | 12264871 | G | A | TMEM106B |
| rs12666117 | 7 | 109105611 | G | A |  |
| rs1354115 | 9 | 2983774 | C | A,T |  |
| rs10959913 | 9 | 11544984 | T | G |  |
| rs7856424 | 9 | 119733595 | C | A,T | ASTN2 |
| rs7029033 | 9 | 126682068 | C | A,G,T | DENND1A |
| rs61867293 | 10 | 106563924 | C | T | SORCS3 |
| rs1806153 | 11 | 31850105 | G | A,C,T | PAX6-AS1, PAUPAR |
| rs4074723 | 12 | 23947737 | A | C,T | SOX5 |
| rs4143229 | 13 | 44327799 | C | A,G,T | ENOX1 |
| rs4904738 | 14 | 42179732 | T | C | LRFN5 |
| rs915057 | 14 | 64686207 | A | G,T | SYNE2 |
| rs10149470 | 14 | 104017953 | A | G,T | LOC124903392 |
| rs8063603 | 16 | 6310645 | G | A | RBFOX1 |
| rs7198928 | 16 | 7666402 | T | A,C | RBFOX1 |
| rs7200826 | 16 | 13066833 | C | A,G,T | SHISA9 |
| rs17727765 | 17 | 27576962 | T | C | CRYBA1 |
| rs62099069 | 18 | 36883737 | A | T | MIR924HG |
| rs11663393 | 18 | 50614732 | G | A | DCC |
| rs1833288 | 18 | 52517906 | A | G,T | RAB27B |
| rs12958048 | 18 | 53101598 | A | G,T | TCF4 |
| rs5758265 | 22 | 41617897 | G | A,C,T | L3MBTL2 |
| rs7973260 | 12 | 118375486 | A | C,G,T | KSR2 | (4) |
| rs62100776 | 18 | 50754633 | A | G,T | DCC |

**eReferences**

1. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature Genetics. 2018;50(8):1112-21.

2. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nature Genetics. 2016;48(9):1031-6.

3. 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. Nature Genetics. 2018;50(5):668-81.

4. 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. Nature Genetics. 2016;48(6):624-33.