***Identification of eight novel genetic associations for major depressive disorder***

Named author list here (order TBD). Groupings: (a) people who contributed most to the manuscript are listed first, (b) PGC MDD study PIs are at the end, (c) all other individuals who made author-level contributions are listed alphabetically between a and b, (d) use of “\*” and “†” to indicate equivalent contributions, and (d) last author is “For the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium”.

Correspond with xxxx and Dr Sullivan: Department of Genetics, CB#7264, 5097 Genomic Medicine, University of North Carolina, Chapel Hill, NC, 27599-7264, USA. Voice: +919-966-3358, FAX: +919-966-3630, [pfsulliv@med.unc.edu](mailto:pfsulliv@med.unc.edu).

Submitted to: journal TBD

Running head: PGC MDD GWAS

Word count: xxxx words (main text)

Additional: x figures, x tables, x supplemental file

Key words: major depressive disorder, genome-wide association, genetics, meta-analysis

# Abstract

# Introduction

Major depressive disorder (MDD) [1](#_ENREF_1) is a notably complex human illness with a lifetime prevalence as high as 15% [2](#_ENREF_2),[3](#_ENREF_3). MDD tends to be a recurrent illness and is thus accompanied by considerable morbidity [4-6](#_ENREF_4), excess mortality [5](#_ENREF_5),[7](#_ENREF_7), substantial costs [8](#_ENREF_8), and it ranks second globally as a cause of disability. [9](#_ENREF_9) Despite concerns about its definition, [10](#_ENREF_10) there is substantial evidence that the constellation of symptoms comprising MDD is familial and heritable. Family studies have shown that people with MDD are three times more likely to have a first-degree relative with MDD. [11](#_ENREF_11) Twin studies suggest that ~40% of the variation in liability to MDD is attributable to additive genetic effects. [11](#_ENREF_11),[12](#_ENREF_12) Subtypes including recurrent, early-onset, or postpartum MDD may be more heritable. [12](#_ENREF_12),[13](#_ENREF_13)

Although risk for MDD is partly driven by genetic variation, genome-wide association (GWA) studies of MDD or depressive symptoms have had notable difficulties in identifying loci exceeding genome-wide significance. [14](#_ENREF_14),[15](#_ENREF_15) Estimates of the proportion of variance attributable to genome-wide SNPs (SNP heritability, ) for MDD indicate that between one quarter and one half of the heritability estimated from family or twin data () is due to common genetic variants acting additively. [16](#_ENREF_16),[17](#_ENREF_17) There were no significant findings in the first Psychiatric Genomics Consortium (PGC) MDD mega-analysis (9,240 MDD cases) [18](#_ENREF_18) or in the CHARGE meta-analysis of depressive symptoms(>30,000 respondents). [19](#_ENREF_19) More recent reports have indicated important initial progress on elucidating the genetic basis of MDD. A study of in Han Chinese women (5,303 severe, recurrent MDD cases) identified two genome-wide significant loci (although the identified variants are rare in Europeans), [20](#_ENREF_20) and a meta-analysis of depressive symptoms (161,460 individuals) has identified two different loci. [21](#_ENREF_21)

There are many potential reasons why identifying causal loci for MDD has proven difficult. [22](#_ENREF_22) MDD is probably highly polygenic and influenced by many genetic loci of small individual effect, [18](#_ENREF_18),[22](#_ENREF_22) as are most complex human diseases [23](#_ENREF_23) including psychiatric disorders (e.g., schizophrenia, bipolar disorder, and autism). [24-26](#_ENREF_24) A major lesson of the past decade is that meta-analysis of large samples is essential for genetic discovery. Even larger samples may be required for genetic discovery in very common disorders – particularly for MDD where heterogeneity between samples is well-recognized. [18](#_ENREF_18),[27](#_ENREF_27)

Consistent with these observations, we sought to accumulate a sufficiently large sample to enable discovery of common genetic variation involved in the etiology of MDD. [28](#_ENREF_28) We describe here a combined analysis of 66,358 MDD cases and 153,234 controls, one of the largest genomic studies of any psychiatric disorder reported to date. We identified eight novel loci that met rigorous criteria for statistical significance and consistency of effects, and a further four loci identified in sex specific analyses. Additional analyses pointed at aggregation of effects in plausible biological pathways and overlap with other psychiatric disorders at multiple levels.

# Results

Primary analyses. The 29 anchor MDD cohorts totaled 16,823 MDD cases and 25,632 controls (***Supplemental Table 1***). All cases were directly interviewed, met standard criteria for MDD, and most were from clinical sources (in 19/29 cohorts). Most controls were screened for the absence of MDD (in 22/29 cohorts). In 8.96 million imputed SNPs, the minimum *P*-value did not reach genome-wide significance (*P*=2.8x10-7 in the interval chr13:59.933-59.941 mb). With GC=1.069, 1000=1.003, and an LD score regression intercept near one (0.9865, SE 0.0078), there was no indication of systematic inflation of test statistics including uncontrolled population stratification artifacts. Although there were no genome-wide significant findings, was 0.14 (SE 0.02) and the QQ plot showed deviation from the expected from *P* values in the range of 10-4 to 10-6 (***Supplemental Figure 1***). Experience with progressively larger samples in schizophrenia GWA studies [24](#_ENREF_24),[29-31](#_ENREF_29) suggests that this QQ plot “bump” is consistent with the presence of real genetic effects for which the current sample size is underpowered for detection given the multiple testing burden. ADD LEAVE ONE OUT WITHIN ANCHOR.

Therefore, we evaluated an expanded set of six samples that used alternative methods of establishing the presence and absence of MDD (***Supplemental Table 2***). Given these differences, we carefully evaluated the comparability of the anchor and expanded sets (***Figure 1a***). Estimates of for the MDD anchor samples with the expanded samples were all 0.61 or larger, and almost all of the confidence intervals include one. Add leave-one-out GRS LEAVING OUT 4 ANCHOR COHORTS WITH NCASE>1000.

The for the expanded samples ranged 0.07-0.28, with small impact of the choice of K (0.10 or 0.15) compared to the standard errors (SE). decreases with sample size (***Figure 1b***) which could reflect increased measurement error or heterogeneity associated with data collection strategies. Some care is needed in interpretation of as we observe increasing for decreasing mean of two cohorts (***Figure 1c***), a reflection that is the SNP-coheritability divided by the product of the two heritabilities. Hence, the sampling errors of and are highly correlated. In contrast, SNP-coheritability (a parameter independently estimated from data) is not associated with mean (***Figure 1d***). Estimate of are significantly greater than zero for most pairs of cohorts implying strong evidence for shared common genetic factors despite the expected heterogeneity in MDD.

Thus, these empirical results support the similarity of the common variant genetic architecture of European-ancestry MDD anchor and expanded cohorts (we discuss the Chinese CONVERGE sample below). We conducted a meta-analysis of the anchor and expanded cohorts with a grand total of 66,358 MDD cases and 153,234 controls (***Supplemental Table 3***). With GC=1.203, 1000=1.002, and an LD score regression intercept near one (1.0043, SE needed), there was no indication of systematic inflation of test statistics (as expected if there were inadequate control of population stratification). The QQ plot showed deviation from the expected (***Figure 2a***), and the Manhattan plot revealed nine independent genomic regions exceeding genome-wide significance (***Figure 2b***).

Evaluation of these associations indicated that eight appeared to be robust associations (***Table 1***). Forest plots showed consistent direction and magnitudes of association in all cohorts, and regional plots revealed that the association regions had support from multiple SNPs exceeding or nearing genome-wide significance (***Supplemental Figures 2-3***). Naomi to do using summary stats. Stephan/Manuel to do conditional analyses on samples where we have direct access as we discussed (mdd29, GERA, iPSYCH). With the exception of the chr1-73Mb region (near *NEGR1*) the association signal is consistent with single signal regions. Brief interpretation of overall results, obvious overlap with SCZ and BMI.

Naomi please: MDD for the meta-analysed sample was xx (SE yy). And please do by chromosome as discussed (chrX too if possible) – ***Supplemental Figure 4*** of by chromosome size (mbp).

Stephan please: polygenic score analysis, estimate of the proportion on the liability scale accounted for. ***Supplemental Figure 5***.

Manuel please: plot of OR x GRS decile in iPSYCH (see PGC SCZ2 Figure 3). IF POSSIBLE, would like to include what PGC SCZ looked like when had about as many associations (pgc scz1). . ***Supplemental Figure 6***.

Genetic architecture of MDD in European and Han Chinese cohorts. Naomi, Stephan. CONVERGE. 0.22±0.04. for European and CONVERGE 0.30 (0.14-0.42). Bidirectional analysis (sign test, GRS, meta-analysis), specific look up of top loci in both directions.

Secondary analyses. First, gene-wide analyses of the full MDD results using MAGMA and other approaches are summarized in ***Table xx***. Notably, *DRD2* (the main target of antipsychotic drugs and implicated by GWA for schizophrenia), [24](#_ENREF_24) *PCLO* [32](#_ENREF_32) and *CACNA1E* achieved gene-wide significance as did multiple genes in ten loci. A large cluster of genes in the MHC region also achieved significance and post-hoc analyses indicated that the signal arose from female subjects. Second, however, by-sex GWA analyses identified no additional associated loci after accounting for the additional multiple testing (***Supplemental Table / Figures xx***). Third, we used stratified LD score regression [33](#_ENREF_33) to evaluate whether the heritability of MDD was notably enriched in a broad set of functional genomic elements. As shown in ***Supplemental Figure 7***, there was significant enrichment only for highly conserved elements in the human genome, a pattern of results remarkably similar to those for schizophrenia. [24](#_ENREF_24),[33](#_ENREF_33) Fourth, application of a Bayesian method to identify “causal” SNPs within an association region did not identify any putative SNPs (Hailiang to provide methods and results, I emailed him). Fifth, analyses of the full MDD GWA results with eQTL data using SMR and Sherlock identified no significant associations add implication, “suggesting that …” (need tables).

Finally, NEED A SECTION ON TWAS – Rick/Sara’s analyses and others.

Pathway analysis. Gene-wide p-values from MAGMA are compared with those from gene expression imputation in ***Supplemental Table xx***. Pathway analyses (***Table xx***) identified one significant and three suggestive pathways: Calcium channel complex/activity, Amyloid Precursor Protein metabolism and GABA-receptor signaling are the top ranked pathways. These are supported by convergent evidence from previous genetic studies of mood disorders as well as being the targets of multiple known drugs. [34-38](#_ENREF_34)

Meta-analysis of MDD and bipolar disorder. In progress. Describe PGC BIP2 paper briefly, we are in position to do the definitive analysis (lots of overlap in controls) , weighted meta.

The genetic architecture of MDD and other conditions. Pfs manuel LD-Hub. Psychiatric disorders: all PGC disorders, anxiety disorders ANGST; Substance use disorders: ethanol, nicotine, caffeine; Empirically related MDD constructs, CHARGE MDD symptoms and UKBB or GPC neuroticism. UKBB subjective well-being/good health; General medical disorders with phenotypic overlap with MDD (migraine, CVD, T2DM, stroke); Demographic-anthropometric. BMI, height, education/college.

# Discussion

Paragraph on terminology, why call it MDD.

# Online Methods

Samples. Our analysis is rooted in a GWA mega-analysis of 29 cohorts of European-ancestry totaling 16,823 MDD cases and 25,632 controls (mdd29, ***Supplemental Table 1***), including all cohorts in the initial PGC MDD papers. [16](#_ENREF_16),[18](#_ENREF_18),[39](#_ENREF_39) All anchor cohorts passed a structured methodological review by MDD assessment experts (DL and KSK). Cases were required to meet international consensus criteria (DSM-IV, ICD-9, or ICD-10) [40-42](#_ENREF_40) for a lifetime diagnosis of MDD established using structured diagnostic instruments from assessments by trained interviewers, clinician-administered checklists, or medical record review. Most cases were clinically ascertained. Most controls were randomly selected from the population and screened for the absence of lifetime MDD. We considered these “anchor” samples given use of standard methods of establishing the presence or absence of MDD.

We then critically evaluated an “expanded” set of six independent cohorts. These European-ancestry samples used a range of methods for assessing MDD (***Supplemental Table 2***): Generation Scotland employed direct interviews; *i*PSYCH (Denmark) and deCODE (Iceland) utilized national treatment registers; GERA used Kaiser-Permanente health maintenance organization records in California; UK Biobank combined self-reported MDD symptoms and treatment for MDD by a medical professional; and 23andMe used self-report of treatment for MDD by a medical professional. All controls were screened for the absence of MDD. After recruiting these studies and determining that they met our expanded inclusion criteria, we committed to a combined analysis. As described more fully in the Results, comprehensive analyses of for each cohort and (common variant genetic correlation) of each expanded cohort with the MDD anchor strongly supported the comparability of these cohorts. [17](#_ENREF_17),[43](#_ENREF_43) Thus, the main analysis in this report was a meta-analysis of the anchor and expanded cohorts for a grand total of 66,358 MDD cases and 153,234 controls.

The CONVERGE sample [20](#_ENREF_20) was used to evaluate MDD genetic signals between European and East Asian cohorts. CONVERGE directly evaluated clinically ascertained females with severe, recurrent MDD of Han Chinese ancestry.

Genotyping, quality control, and statistical analysis. Briefly, individual genotype data for all anchor cohorts, GERA, and *i*PSYCH were processed using the PGC “ricopili” pipeline (URLs) for standardized quality control, imputation, and analysis. [24](#_ENREF_24) The expanded cohorts from deCODE, Generation Scotland, UK Biobank, and 23andMe were processed by the collaborating research teams using comparable procedures. GWA in CONVERGE (based on low-coverage whole genome sequencing instead of SNP arrays) used different methods for quality control and analysis. [20](#_ENREF_20) SNPs and insertion-deletion polymorphisms were imputed using the 1000 Genomes Project multi-ancestry reference panel (N=1092, URLs).[44](#_ENREF_44) Quality control filters included imputation marker INFO score ≥ 0.6 and MAF ≥ 0.01.

Identification of identical samples—inadvertent duplicates, monozygotic twins, or an individual enrolling in more than one study—is easily accomplished given direct access to individual genotypes. [16](#_ENREF_16) A particular concern is the use of the same control samples in multiple studies (e.g., GAIN or WTCCC controls). [45](#_ENREF_45),[46](#_ENREF_46) Closely related individuals can be identified using the same procedures. For cohorts where the PGC central analysis team had access to individual genotypes (all anchor cohorts and GERA), we used SNPs directly genotyped on all platforms to compute empirical relatedness, and excluded one of each duplicated or relative pair (defined as > 0.2). Within all other cohorts (deCODE, Generation Scotland, *i*PSYCH, UK Biobank, 23andMe, and CONVERGE), identical and relative pairs were identified and resolved using similar procedures. Identical samples between the anchor cohorts, *i*PSYCH, UK Biobank, and Generation Scotland were identified using genotype based checksums (URLs). [47](#_ENREF_47). Subjects in both the anchor cohort and Generation Scotland (N=1), *i*PSYCH (N=xxx), and UK Biobank (N=101) were identified, and an individual on the collaborator’s side was excluded. Checksums were not available from the deCODE and 23andMe cohorts. Related pairs are not detectable by the checksum method but we did not find evidence of important overlap using LD score regression. Ancestry was evaluated using principal components analysis applied to directly genotyped SNPs. [48](#_ENREF_48) In the anchor cohorts and GERA, we determined that all individuals in the final analyses were of European ancestry. European ancestry was confirmed in the other expanded cohorts by the collaborating research teams using similar procedures.

In each cohort, logistic regression association tests were conducted for imputed marker dosages with principal components covariates to control for population stratification. We tested 20 principal components for association with MDD and included five principal components covariates for the anchor cohorts and GERA; all other cohorts adopted similar strategies. There was no evidence of stratification artifacts or uncontrolled test statistic inflation in the results from each anchor and extended cohort (e.g., GC was 0.995–1.043 in the anchor cohorts). The results were combined across samples using an inverse-weighted fixed effects model.[49](#_ENREF_49) All genomic coordinates are given in NCBI Build 37/UCSC hg19.

Replication and significance. GWA standards circa 2008 included the need for replication of top findings in independent samples. [50](#_ENREF_50) The major intent of the requirement was to ensure that the results were robust to genotyping artifacts or technical errors. Since then, there have been several changes. First, other methods can be used to evaluate the consistency of findings across samples (e.g., test of heterogeneity and leave-one-out analyses of polygenic risk). [24](#_ENREF_24) that serve the role that previously was the main point of replication - insurance that the results did not arise from genotyping artifact, technical error in analysis, etc. Second, it is now widely accepted that joint analysis of all available samples is more powerful than two stage designs (e.g., discovery and replication subsets. [51](#_ENREF_51) Third, we now know that the effect sizes underlying most complex traits are small and require historically large sample sizes for confident identification. [23](#_ENREF_23),[52](#_ENREF_52) Finally, in practice, we point to recent GWA meta-analyses in top tier journals that did not have prominent replication samples. [21](#_ENREF_21),[53-55](#_ENREF_53) The genome-wide significance threshold was *P*<5x10-8 (reference [56](#_ENREF_56)).

Linkage disequilibrium score regression (LDSC) [17](#_ENREF_17),[43](#_ENREF_43) was used to estimate from GWA summary statistics. Estimates of on the liability scale depend on the assumed lifetime prevalence of MDD in the population (*K*). We assumed K values of 0.10 and 0.15 to explore sensitivity. For CONVERGE, we used K=0.036 as reported by these investigators. [20](#_ENREF_20) LDSC bivariate genetic correlations attributable to genome-wide SNPs ( or ) were estimated within MDD cohorts and between the full MDD cohort and other traits and disorders. Estimates of of European cohorts with the Han Chinese CONVERGE cohort was estimated using *popcorn*, [57](#_ENREF_57) which accounts for differing LD structures between different ancestries.

Partitioned LD score regression. Hilary Finucane to provide. [33](#_ENREF_33)

Using eQTL data to refine associations. Multiple methods have been proposed for use of gene expression quantitative trait loci (eQTLs) results from external studies for the prioritization and identification of causal loci in a genomic region implicated by GWA. We applied the GWA summary statistic -based Mendelian Randomization (SMR) method, [58](#_ENREF_58) Sherlock, [59](#_ENREF_59) and TWAS (transcription wide association study) [60](#_ENREF_60) to the results from the meta-analysis of the anchor and expanded samples. eQTL results were based on gene expression in non-transformed peripheral blood samples from the largest available study (5,311 individuals with replication in 2,775 individuals). [61](#_ENREF_61)

By sex.

Pathway analysis. Our approach was guided by rigorous method comparisons conducted by PGC members. [62](#_ENREF_62),[63](#_ENREF_63) P-values quantifying the degree of association with MDD at gene and gene-set levels were generated using MAGMA (v1.03). [64](#_ENREF_64) We performed separate analyses using GWA results for all subjects and then separately for males and females (given the marked MDD lifetime prevalence differences by sex). Gene analysis in MAGMA uses the principal components of the gene-SNP matrix to adjust for LD in a multiple linear regression model. Gene-set p-values are obtained using a competitive analysis, which tests whether genes in a gene-set are more strongly associated with the phenotype than other gene-sets. We used European-ancestry subjects from 1,000 Genomes Project (Phase 3 v5a, imputation INFO > 0.6, MAF ≥ 0.01) [65](#_ENREF_65) for the LD reference. The gene window used by MAGMA was set to 35 kb upstream and 10 kb downstream to include regulatory elements in the analysis. Gene models were from ENSEMBL, and gene-sets were extracted from MSigDB v5.1 [66](#_ENREF_66) encompassing canonical pathways (CP) and Gene Ontology (GO) gene sets. CP were curated from BioCarta, KEGG, Matrisome, Pathway Interaction Database, Reactome, SigmaAldrich, Signaling Gateway, Signal Transduction KE, and SuperArray. Pathways containing 10-1000 genes were included for a total of 2,737 pathways (1309 CP, 1428 GO). Principal components analysis of gene set membership indicated that there were 1,900 independent pathways (i.e., the number of principal components explaining >99.5% of the variance). Statistical significance, threshold–PLEASE DESCRIBE HOW DONE.

# URLs

Summary results from this paper are freely available on the PGC website (<https://pgc.unc.edu>). PGC “ricopili” GWA pipeline (<https://github.com/Nealelab/ricopili)>. 1000 Genomes Project multi-ancestry imputation panel, <https://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html>. Genotype-based checksums for relatedness determination, <http://www.broadinstitute.org/~sripke/share_links/checksums_download)>.

# Author Contributions

TDB.

# Conflicts of Interest

TDB.

# Acknowledgements

This work was supported by NIMH U01 MH085520. Pfizer data were made with the assistance or financial contribution, if any, from the Innovative Medicine Initiative Join Undertaking (IMI-JU) under grant agreement number 115008.5. Add others.

Table 1 – details on the 8 gwsig loci.

# References

1. Hyman SE. The diagnosis of mental disorders: The problem of reification. *Annual Review of Clinical Psychology* **6**, 12.1-12.25 (2010).

2. Hasin DS, Goodwin RD, Stinson FS & Grant BF. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Archives of general psychiatry* **62**, 1097-106 (2005).

3. Kessler RC *et al.* The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Jama* **289**, 3095-105 (2003).

4. Judd LL. The clinical course of unipolar major depressive disorders. *Archives of general psychiatry* **54**, 989-91 (1997).

5. Lopez AD, Mathers CD, Ezzati M, Jamison DT & Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* **367**, 1747-57 (2006).

6. Wittchen HU *et al.* The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* **21**, 655-79 (2011).

7. Angst F, Stassen HH, Clayton PJ & Angst J. Mortality of patients with mood disorders: follow-up over 34-38 years. *Journal of affective disorders* **68**, 167-81 (2002).

8. Gustavsson A *et al.* Cost of disorders of the brain in Europe 2010. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* **21**, 718-779 (2011).

9. Ferrari AJ *et al.* Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med* **10**, e1001547 (2013).

10. Galatzer-Levy IR & Bryant RA. 626,120 ways to have posttraumatic stress disorder. *Perspectives on Psychological Science* **8**, 651-662 (2013).

11. Sullivan PF, Neale MC & Kendler KS. Genetic epidemiology of major depression: Review and meta analysis. *American Journal of Psychiatry* **157**, 1552-1562 (2000).

12. Rice F, Harold G & Thapar A. The genetic aetiology of childhood depression: a review. *J Child Psychol Psychiatry* **43**, 65-79 (2002).

13. Viktorin A *et al.* Heritability of Perinatal Depression and Genetic Overlap With Nonperinatal Depression. *Am J Psychiatry*, appiajp201515010085 (2015).

14. Dunn EC *et al.* Genetic determinants of depression: Recent findings and future directions. *Harvard Review of Psychiatry* (in press).

15. Flint J & Kendler KS. The genetics of major depression. *Neuron* **81**, 484-503 (2014).

16. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature genetics* **45**, 984-94 (2013).

17. Bulik-Sullivan BK *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**, 1236-41 (2015).

18. Major Depressive Disorder Working Group of the PGC. A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry* **18**, 497-511 (2013).

19. Hek K *et al.* A genome-wide association study of depressive symptoms. *Biological psychiatry* **73**, 667-78 (2013).

20. CONVERGE Consortium. Sparse whole genome sequencing identifies two loci for major depressive disorder. *Nature* (2015).

21. Okbay A *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* (2016).

22. Levinson DF *et al.* Genetic studies of major depressive disorder: why are there no GWAS findings, and what can we do about it. *Biological Psychiatry* **76**, 510-2 (2014).

23. Visscher PM, Brown MA, McCarthy MI & Yang J. Five Years of GWAS Discovery. *American journal of human genetics* **90**, 7-24 (2012).

24. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).

25. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature genetics* **43**, 977-83 (2011).

26. Gaugler T *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* (2014).

27. Wray NR *et al.* Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry* **17**, 36-48 (2012).

28. Wray N *et al.* Genome-wide association study of major depressive disorder: New results, meta-analysis, and lessons learned. *Molecular Psychiatry* **17**, 36-48 (2011).

29. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-52 (2009).

30. Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics* **43**, 969-76 (2011).

31. Ripke S *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics* **45**, 1150-9 (2013).

32. Sullivan PF *et al.* Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Molecular Psychiatry* **14**, 359-75 (2009).

33. Finucane HK *et al.* Partitioning heritability by functional category using GWAS summary statistics. *Nature Genetics* **47**, 1228-35 (2015).

34. Deutschenbaur L *et al.* Role of calcium, glutamate and NMDA in major depression and therapeutic application. *Prog Neuropsychopharmacol Biol Psychiatry* **64**, 325-33 (2016).

35. Chung JK *et al.* Lifetime History of Depression Predicts Increased Amyloid-beta Accumulation in Patients with Mild Cognitive Impairment. *J Alzheimers Dis* **45**, 907-19 (2015).

36. Wu KY *et al.* Increased brain amyloid deposition in patients with a lifetime history of major depression: evidenced on 18F-florbetapir (AV-45/Amyvid) positron emission tomography. *Eur J Nucl Med Mol Imaging* **41**, 714-22 (2014).

37. Luscher B, Shen Q & Sahir N. The GABAergic deficit hypothesis of major depressive disorder. *Mol Psychiatry* **16**, 383-406 (2011).

38. Cryan JF & Slattery DA. GABAB receptors and depression. Current status. *Adv Pharmacol* **58**, 427-51 (2010).

39. 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* **381**, 1371-9 (2013).

40. World Health Organization. *International Classification of Diseases*, (World Health Organization, Geneva, 1978).

41. World Health Organization. *International Classification of Diseases*, (World Health Organization, Geneva, 1992).

42. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, (American Psychiatric Association, Washington, DC, 1994).

43. Bulik-Sullivan BK *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291-5 (2015).

44. Durbin RM *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).

45. Sanders AR *et al.* The Internet-based MGS2 control sample: self report of mental illness. *The American journal of psychiatry* **167**, 854-65 (2010).

46. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).

47. Franke B *et al.* Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. *Nat Neurosci* **19**, 420-31 (2016).

48. Price AL *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).

49. Begum F, Ghosh D, Tseng GC & Feingold E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res* **40**, 3777-84 (2012).

50. Ioannidis JP, Thomas G & Daly MJ. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet* **10**, 318-29 (2009).

51. Skol AD, Scott LJ, Abecasis GR & Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* **38**, 209-13 (2006).

52. Lander ES. Initial impact of the sequencing of the human genome. *Nature* **470**, 187-97 (2011).

53. Day FR *et al.* Physical and neurobehavioral determinants of reproductive onset and success. *Nat Genet* (2016).

54. Khor CC *et al.* Genome-wide association study identifies five new susceptibility loci for primary angle closure glaucoma. *Nat Genet* (2016).

55. Locke AE *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).

56. Pe'er I, Yelensky R, Altshuler D & Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* **32**, 381-385 (2008).

57. Brown BC, Asian Genetic Epidemiology Network-Type 2 Diabetes, Ye CJ, Price AL & Zaitlin N. Transethnic genetic correlation estimates from summary statistics. (Submitted).

58. Zhu Z *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).

59. He X *et al.* Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am J Hum Genet* **92**, 667-80 (2013).

60. Gusev A *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-52 (2016).

61. Westra HJ *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-43 (2013).

62. Pathway Analysis Subgroup of the Psychiatric Genomics C. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* **18**, 199-209 (2015).

63. de Leeuw CA, Neale BM, Heskes T & Posthuma D. The statistical properties of gene-set analysis. *Nat Rev Genet* (2016).

64. de Leeuw CA, Mooij JM, Heskes T & Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).

65. Genomes Project C *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

66. Liberzon A *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).