Twin Study of Early-Onset Major Depression Finds DNA Methylation Enrichment for Neurodevelopmental Genes

Roxann Roberson-Nay[1,2], Aaron R. Wolen[4], Dana M. Lapato[2,4], Eva E. Lancaster[2,4], Bradley T. Webb[1,2,4], Bradley Verhulst[3], John M. Hettema[1,2], Timothy P. York[2,4]

1. Virginia Commonwealth University, Department of Psychiatry, Richmond, VA.
2. Virginia Commonwealth University, Virginia Institute for Psychiatric and Behavioral Genetics, Richmond, VA.
3. Department of Psychology, Michigan State University, East Lansing, MI.
4. Virginia Commonwealth University, Department of Human and Molecular Genetics, Richmond, VA.

Correspondence: Roxann Roberson-Nay, Ph.D., Virginia Commonwealth University, Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, P.O. Box 980489, Richmond, VA 23298, Fax (804) 828-0245, email: roxann.roberson-nay@vcuhealth.org.

Abstract

Major depression (MD) is a debilitating mental health condition with peak prevalence occurring early in life. Genome-wide examination of DNA methylation (DNAm) offers an attractive complement to studies of allelic risk given it can reflect the combined influence of genes and environment. The current study used a co-twin control design to identify differentially and variably methylated regions of the genome that distinguish monozygotic (MZ) twins with and without a lifetime history of early-onset MD. The sample included 150 Caucasian monozygotic twins (73% female; *M*age=17.52 *SD*=1.28) assessed during a developmental stage characterized by relatively distinct neurophysiological changes. All twins were generally healthy and currently free of medications with psychotropic effects. DNAm was measured in peripheral blood cells using the Infinium Human BeadChip 450K Array. MD associations were detected at 760 differentially and variably methylated probes/regions that mapped to 428 genes. Gene enrichment analyses implicated genes related to neuron structures and neurodevelopmental processes including cell-cell adhesion genes (e.g., *CDHs, PCDHAs, PCDHA1C/2C*). Genes previously implicated in mood and psychiatric disorders as well as chronic stress (e.g., *HDAC4, NRG1*) also were identified. Results indicated an association between early-onset MD and many genes and genomic regions involved in neural circuitry formation, projection, functioning, and plasticity. DNAm regions associated with MD where found to overlap genetic loci observed in the latest Psychiatric Genomics Consortium meta-analysis of depression. Understanding the time course of epigenetic influences during emerging adulthood may clarify developmental phases where genes modulate individual differences in MD risk.

Introduction

Major depression (MD) is highly prevalent, ranking second in the global burden of disease, with the overall lifetime risk estimated to be 16.2% in the general population[1]. MD is associated with increased mortality, particularly suicide[1]. Among adolescents, MD is associated with the greatest level of impairment of all psychiatric conditions, with 16% of females and 12% of males endorsing at least one major depressive episode (MDE) by age 18[2]. An early age of onset confers increased risk for negative socioemotional outcomes including recurrent MDEs[3,4]. Adolescence/young adulthood is characterized by neurophysiological changes (e.g., synaptic pruning, myelination) that significantly influence brain function and behavior, which may increase risk for MD and other psychiatric conditions[5]. Thus, understanding the genetic contributions to MD during this dynamic neurophysiological period where peak incidence is observed[2,6–8] is critical to elucidating developmentally informed pathways to mood disorders.

Twin and family studies robustly demonstrate that genetic factors play a role in risk for MD, with heritability estimates of roughly 35% for MD and 45% for early-onset MD[9]. The largest meta-analysis of molecular genetic studies of MD recently identified 44 independent loci, underscoring the centrality of genetic factors in the etiology of MD[10]. Twin study variance component analyses also indicate a considerable contribution of unique environmental risk factors to MD[11]. Due to the substantial link between environmental hardship and onset of a MDE[12,13], epigenetic mechanisms may, in part, mediate the influence of environmental stress and interact with genetic liability for MD over the lifespan[14–16]. Epigenetic mechanisms refer to DNA, chromatin, and RNA modifications that can influence the expression of genes but do not alter the underlying genetic sequence. Animal studies have been critical to demonstrating a causal association between early life environments, epigenetic alterations, and phenotypic outcomes. For example, the seminal work of Michael Meaney and his research team demonstrates the importance of maternal care in altering the expression of genes that regulate behavioral and neuroendocrine responses to stress as well as synaptic development in the rat hippocampus[17–22]. Indeed, a number of animal and

human studies demonstrate lasting epigenetic alterations occurring in the genomes of cells including changes to post-mitotic neurons that integrate experience-dependent changes[23]. Thus, the timing of environmental stress plays an important role in subsequent epigenetic consequences, with early life stress paradigms in mice and humans demonstrating enduring changes in epigenetic profiles[17,20,24–28].

A number of studies have utilized genome-wide platforms to determine DNA methyalation (DNAm) differences between MD affected and unaffected individuals. Unfortunately, contrasting DNAm between unrelated cases and controls cannot disaggregate the genetic and environmental influences on DNAm outcomes[29], thus confounding etiological inferences. Indeed, it is estimated that as much as 37% of methylation variance can be accounted for by genetic factors[30], which is consistent with recent studies indicating that common genetic variation (i.e., methylation quantitative trait loci [*mQTLs*]) influence DNAm levels[31–35]. Yet, rarely do case-control studies account for allelic variation. By contrast, the quasi-experimental design afforded by the co-twin control approach greatly improves on the unmatched case-control design (see Supplemental Figure 1). In particular, the use of monozygotic (MZ) twins removes the impact of unmeasured confounds such as genetic variation, uterine environment, age, sex, race, cohort effects, and exposure to many shared environmental events.

The current study relied on the robust co-twin control method and a statistically powerful approach to detect differentially and variably methylated DNAm regions associated with MD in a sample of adolescent and emerging adult twins. Studying this developmental period offers a number of advantages over later life periods including fewer confounds to DNAm variability such as a history of prolonged or multiple psychiatric/medical comorbidities and medication usage as well as long-term nicotine use. Moreover, it eliminates the well-known epigenetic changes associated with aging[36–38]. The developmental window of young adulthood also is associated with moderate conservation of DNAm that is nonetheless responsive to environmental signals[39], making it an ideal sensitive period for the study of MD.

Methods

Participants

One hundred sixty-six MZ twins (83 pairs) were selected from a larger sample of twins (N=430 pairs) enrolled in a study examining risk factors for internalizing disorders in a general population sample of twins (R01MH101518)[40]. Twins were primarily recruited through the Mid-Atlantic Twin Registry (MATR), a population-based registry[40,41]. All participants (parents/minor children, adults) provided written informed consent/assent.

Of the 83 twin pairs initially identified as MZ via the zygosity questionnaire, five pairs (6.3%) were determined to be DZ pairs using DNA-based markers and were removed (for zygosity determination, see Supplemental). One or both twins from three additional pairs failed DNA-based quality control checks, reducing the final analyzed sample to 75 MZ twin pairs (150 twins; see Table 1). All twins were raised together in the same home and were required to be free of psychotropic medications/medications with psychotropic effects at the time of study entry, although approximately 4.0% (n=6) endorsed a history of psychotropic medication use, slightly lower than the national average[42]. See full study exclusionary criteria in Supplement.

Measures

Major Depression. All twin pairs completed a psychiatric history based on an expanded version of the Composite International Diagnostic Interview (CIDI)-Short Form, which queried DSM-5 MD Criterion A and C (see Supplement for questions and diagnostic algorithm)[43]. Current depressive symptoms also were assessed using the Short Mood and Feelings Questionniare (SMFQ), which is a 13-item questionnaire validated to measure depressive symptoms in adolescents and adults[44].

DNA Methylation Processing

Detailed information concerning DNA extraction, methylation assessment, normalization, and quality control (QC) can be found in the Supplemental.

DNA Methylation Analysis

Analytic Approach. Based on observations that the average correlation between probes on the 450K microarray within approximately 250 base pairs (bp) is 0.83 and within 1 kb is 0.45[45–47], several methods have been proposed to take advantage of this structure to identify consistent DNAm change across a contiguous region[45,46,48,49]. Current approaches quantify regional DNAm change either as a mean difference (differentially methylated region [DMR]) or as a difference in variance (variably methylated region [VMR]). Due to the sparse and highly clustered placement of features on the Illumina 450K platform, a custom approach for defining DMRs and VMRs was developed for this study motivated by the algorithm proposed by Ong et al. [45,50]. They suggest both a regional and individual CpG probe approach run in tandem since approximately 25% of probes do not have a neighboring probe within 1 kb. To this end, the single-probe analysis provided the raw materials for the regional approach adopted to identify and assess significance of DMRs and VMRs. Type I error rates were estimated from empirical *P*-values for all univariate CpG probe models (mean level and variance based testscluster) described below were calculated using a permutation approach. For k = 1,000 reorderings, the outcome variable was resampled in a way that preserved the discordance/concordance pair status frequencies. The false discovery rate (FDR)[51] was estimated from the distribution of these empirical *P*-values.

Identifying Regional DNAm Change

Differentially Methylated Regions. Potential DMRs were constructed using only those individual CpG probes associated with MD status resulting in a test statistic in the upper or lower 5th percentile of all probes tested. Univariate tests were performed by fitting a linear mixed-effects model[52] separately for each probe by regressing the normalized DNAm probe intensity on MD status while adjusting NK cell proportion and inclusion of a random effect term to account for twin pair membership. NK cell proportion was the only estimated cell type nominally significantly different between MD cases and controls (t = 0.266, p = 0.051), so it was included as a covariate to control for potential bias that might arise from DNAm differences due to changes in NK blood cell proportions rather than those attributable to DNAm changes associated with MD itself.

The structure of MZ twin exposure groupings provided an opportunity to focus on DNAm associations that are unique to MD versus other experimental design features. The sample used for study was composed of three types of MZ twin pairs (i.e., discordant MD, concordant MD, concordant healthy) and resulted in number of potential contrasts to be tested (Supplemental Table 1). Specific contrasts were simultaneously fit within the linear model to estimate the degree of association among two sets of comparisons: 1) the effect of interest (e.g., MD affected versus MD unaffected) and 2) extraneous associations not of interest. The latter set of contrasts, while of general interest, do not directly relate to MD and were filtered from downstream applications. For example, the comparison of DNAm levels where MD is expressed in both members (concordant MD) versus only a single member (discordant MD) may be associated with a unique DNAm profile (i.e., line 2 in Supplemental Table 1). Probes with test statistics in the 1st and 99th quantiles identified in any contrast not of interest were removed.

From the set of filtered CpG probes, a candidate DMR was defined as having at least 2 contiguous CpGs within 1 kb. The relative importance of a DMR was estimated by a test statistic reflecting the area of influence approximated by the trapezoidal rule where the height (***h***) was the length in base pairs between two contiguous CpGs and *a* and *b* were the univariate test statistic for the contiguous CpGs. For DMRs with more than two CpGs, the area for each contiguous pairing was summed to represent the area under the curve (AUC) for the entire DMR. All probes in the DMR had the restriction of test statistics with the same sign. A positive test statistic indicated hypermethylation in cases versus controls while a negative test statistic indicated hypomethylation.

Variably Methylated Regions. A similar strategy was adopted to identify variably methylated regions (VMR). In this case, the test statistic retained was the ***F***-value comparing the variance of two samples applied to the filtered set of CpG probes. Again, only probes for the contrast of interest were considered (10th / 90th percentiles) if they did not intersect with those from contrasts not of interest (1st/99th percentiles). VMRs were restricted to at least two contiguous

probes within 1 kb whereby all probes had the restriction of either increasing or decreasing variability in cases versus controls.

Significance Assessment of Regional Change. The statistical significance of DMR and VMR AUC estimates were assessed using a rank-based permutation method[53]. This nonparametric method estimates an FDR without relying on strong assumptions about the normality of the data. The k = 1,000 permutations of univariate tests were used to estimate the expected order statistics. The FDR was calculated based on the observed versus expected null scores. Briefly, for a range of thresholds, regions are called significant if the value of the observed ordered test statistic minus the mean value from the permuted rank exceeds a given threshold. The number of falsely called regions is the median number of regions that exceed the lowest AUC value of regions called significant. The FDR is calculated as the ratio of the number of falsely called regions to the number of regions called significant. An implementation of the SAM algorithm is available as a R package[54] but was recoded to allow for flexibility in specifying models for the twin data and to trim extreme test statistics likely to be false positives before the calculation of the FDR.

Functional and Regulatory Enrichment

The distribution of significant CpG probes and regions identified to be differentially and variably methylated by MD status were examined separately across functional and regulatory annotations. CpG findings were mapped to known genes[55] for enrichment of Gene Ontology classifications[56] using clusterProfiler[57]. Classification functons included biological processes, cellular components, and molecular function, in addition to KEGG pathways. Tests for non-random association of CpG island features and ChromHMM chromatin states were based on the AH5086 and AH46969 tracks from the AnnotationHub package[58], respectively. CpG island shores were defined as being 2 kb regions flanking CpG islands while shelves were demarcated as 2 kb upstream or downsteam shore regions. A test of enrichment for each of these annotations was calculated by comparing the proportion of sequence from the intersection of significant CpG

regions with the regions defined by the annotation feature. Bootstrap methods using 1,000 resamplings were used to estimate 95% confidence intervals. This observed overlap was compared to an empirical distribution of random samples of genome groups of the same size and structure drawn from the background set under consideration. Empirical P-values were calculated from 1,000 random reorderings of the data using standard methods[59].

PGC GWAS Enrichment

A similar resampling method was performed to count the number of significant CpG regions that overlapped with the findings of a recent genome-wide association study (GWAS) meta-analysis conducted by the Psychiatric Genomics Consortium (PGC) group[10]. The depression phenotype in this meta-analysis was derived from a number of different methods including clinical interview, self-report, electronic medical record abstraction, and self-report of a lifetime diagnosis. This study identified 44 MD-associated loci across 18 chromosomes, which included genes enriched for targets of antidepressant medication. The non-random frequency of overlap between the significant CpG regions and the 44 independent PGC findings was assessed using bootstrap and permutation approaches from 1,000 data resamplings.

Results

Sample Characteristics

MZ twins meeting DSM-5 MD criteria at the probable or definite level self-reported higher depression symptom scores on the SMFQ and had higher rates of generalized anxiety disorder compared to MD unaffected twins ($t(1,146)=6.4$, P-val <.001); $\chi^2(1)=5.31$, P-val <.05, respectively; see Table 1). For those twins meeting DSM-5 criteria for at least one MDE, the mean age at onset was approximately 15 years, and the majority of twins (~85%) reported experiencing 3 or fewer MDEs in their lifetime. Most MD affected twins (82%) reported five or more symptoms during their worst lifetime MDE. Subject age, sex, self-reported ethnicity, and combustible cigarette use did not differ by MD status.

Differentially Methylated Probes and Regions

From the set of 455,828 screened CpG probes, according to our definition, 50,990 background regions could be created, covering 59.9 megabases. After DMP tests, 3,995 regions consisting of 28,600 CpG probes could be considered candidate DMRs. Seventeen DMRs were identified as significantly associated with MD (all hypermethylated in MD cases) of which 15 mapped onto genes (FDR 9.9%; see Table 2). The number of CpG probes in significant DMRs ranged from 3 to 7 (median=4). Individual probe testing (DMP) resulted in 59 hypomethylated and 77 hypermethylated CpG sites with respect to MD status (FDR 1%; Supplemental Table 2). The combined set of 30.6 kb DNA sequence covered by significant DMR and DMP findings was found to have a nonrandom pattern of enrichment across ChromHMM annotations, specifically sites of strong transcription (*P*-val= 0.019), enhancers (*P*-val= 0.045), ZNF genes/repeats (*P*-val= 0.001), heterochromatin (*P*-val= 0.013) and weak repressed polycomb (*P*-val= 0.045) (Supplemental Figure 2) and CpG island relationships which included both north (*P*-val= 0.024) and south (*P*-val= 0.003) shelf regions (Supplemental Figure 3).

Variably Methylated Probes and Regions

Regional analysis identified 10 variably methylated regions (VMRs) signficant from a total of 11,055 candidate VMRs (FDR 17.3%). Seven of the VMRs mapped onto genes (Table 3). Significant VMRs were all more variable in MD cases, and the number of CpG probes in these regions ranged from 2 to 11 (median=4). The VMP analysis yielded 560 significant VMP findings (FDR 1%), all of which were more variable in MD cases except for a single probe (Supplemental Table 3). The combined set of VMR and VMP DNA regions of 16.6 Kb reflected a nonrandom enrichment with ChromHMM annotations for 5'/3' transcription (*P*-val= 0.035), genic enhancers (*P*-val= 0.024) and heterochromatin (*P*-val= 0.050) (Supplemental Figure 4) and CpG island relationships within the south shelf (*P*-val= 0.029) (Supplemental Figure 5).

Gene Enrichment Analysis

Genes that mapped to signficant differentially or variably methylated findings were combined for gene-based enrichment to provide an overview of all DNAm contributions at a functional level. The results of enrichment tests yielded significant over-representation for biological processes (BP) and cellular function (CF), and no enrichment for molecular function or KEGG, at the 10% FDR (Table 4). The BP gene category associations were hemophilic cell adhesion and cell-cell adhesion while the significant terms for CF were associated with functions of neurons, including neuron projection terminus, terminal button, axon part, cell projection part, axon, and presynapse.

Relationship Between Early-Onset MD DNAm Markers and PGC MD-Associated Genetic Loci

A total of 6 differentially methylated sites (including DMP/DMRs; *P*-val= 0.002; 95% CI = 2-11) (Supplemental Table 4) and 12 variably methylated sites (including VMPs/VMRs; 95% CI = 5-19, *P*-val = 0.008) (Supplemental Table 5) overlapped PGC GWAS findings. These enrichment results were largely driven by overlap observed with the PGC GWAS locus on chromosome 6 at 27.738-32.848 Mb (Figure 1). At this locus, 5 of 6 differentially and 10 of 12 variably methylated sites overlapped.

Discussion

The main objective of the current study was to identify differentially and variably methylated loci and regions that distinguish MZ twins with and without a history of MD. Across biological process (BP) and cellular function (CF) domains of the gene enrichment analysis, there was consistency in the functional attributes of the genes related to neural structures and processes as a key differentiating feature between MD affected and unaffected twins. BP gene ontologies referenced homophilic cell adhesion and cell-cell adhesion processes, which are involved in neural development and plasticity. A number of cadherins and protocadherins emerged in the cell adhesion gene sets with MD affected twins demonstrating increased variation in two cadherins (*CDH3*, *CDH6*), the clustered protocadherin alpha family (*PCHDA1 - PCDHA13*), the C-type isoforms of

PCDHA (*PCDHAC1, PCDHAC2*), and two non-clustered Protocadherins (*PCDH10, PCDH20*). CDHs/PCDHs are a group of calcium dependent cell-cell adhesion molecules that are abundantly expressed in the nervous system and play a major role in multiple steps essential to neurodevelopment including dendrite arborization, axon outgrowth and targeting, synaptogenesis, and synapse elimination[60–66].

The DNAm profile of the *PCDH*s is known to be responsive to environmental factors[67], and emerging evidence suggests a role for *PCDH*s in multiple psychiatric phenotypes (e.g., schizophrenia, bipolar disorder, autism) including MD[68,69]. Two small twin studies (≤18 twin pairs) observed that several cadherin/protocadherin genes demonstrated differences in DNAm between twin pairs repeatedly discordant for elevated depression symptoms[70] as well as a history of MD or an anxiety disorder[71]. A related study observed increased DNAm in *PCDH* gene families with the highest enrichment of hypermethylated sites in the *PCDHA* genes located in the hippocampus of suicide completers with a history of severe childhood abuse[19]. At the genetic variant level, a recent meta-analysis with over 89,000 MD cases detected a SNP (rs9540720) in the non-clustered *PCDH9* gene to be significantly associated with MD case status at the genome-wide level[72], and a related gene that encodes a protein of the same family (*PCDH17)* was found to confer risk for mood disorders[73]. Moreover, expression patterns of *Pcdh* genes have been examined in rodent brain regions involved in the neural circuity of MD[74], with results indicating high expression levels in subregions of the hippocampus and basolateral amygdaloid complex[74,75]. Moreover, *Pcdh* gene expression was reduced in both structures following electroconvulsive shock (ECS), which is interesting given that ECS is an effective treatment for chronic, intractable MD symptomatology[76] and supports involvement of *PCDHs* in neural plasticity. The clustered *PCDHA* family also is strongly expressed in serotonergic neurons[77–79] and *PcdhαC2* is necessary for axonal tiling and assembly of serotonergic circuitries[86]. *PcdhaC1* and *PcdhaC2* are expressed throughout the developing rat brain and peak around postnatal day 10[70], which is roughly equivalent to early childhood years in humans. Thus, a deeper understanding of *how* the serotonergic neural network

is formed in the young developing brain may be critical to identifying etiological determinants of MD and other psychiatric conditions.

The Histone Deacetylase 4 (*HDAC4*) gene was altered in MD affected twins, exhibiting increased variation and emerging as part of a genomic region displaying higher mean levels compared to MD unaffected twins[80]. *HDAC4* is highly expressed in the human brain,[81] and the subcellular localization of the *HDAC4* protein is purportedly modulated by *N*-methyl-D-aspartate (NMDA) receptors, a specific type of ionotropic glutamate receptor important for controlling synaptic plasticity and memory function[82]. In humans, increased *HDAC4* expression is associated with memory deficits[80,82], which is consistent with the cognitive processing difficulties observed with MD[83,84]. Moreover, drugs that inhibit *HDAC4* have antidepressant-like effects on behavior[85–88]. Similarly, Neuroregulin1 (*NRG1*) exhibited increased variation and emerged in the gene enrichment analysis as a gene involved in multiple cellular functions. *NRG1* promotes myelination in the central nervous system (CNS) and evidence suggests an association between *NRG1* and cognitive deficits linked to psychiatric disorders[89,89–95].

The overlap between our DNAm results and variants identified in the latest PGC GWAS of depression were examined to better understand the possible role of early-onset MD DNAm events among these genetic associations. A total of 6 differentially methylated and 12 variably methylated sites overlapped PGC findings. These enrichment results were primarily driven by overlap observed with the PGC GWAS locus on chromosome 6, which falls in the extended MHC region. This region of the genome also was the most significant association in the PGC's GWAS of schizophrenia[96] and a GWAS that included a combined sample of schizophrenia and bipolar disorder cases[97]. The MHC is densely populated with genes related to neuronal signaling and plays a role in immune functioning. Although not located in the MHC, higher means and greater variation in *CSMD1* was observed for early-onset MD affected twins. *CSMD1* is a regulator of complement activation and inflammation in the developing CNS[84]. Complement activation is critical for synaptic plasticity and is linked to brain-related psychiatric diseases such as

schizophrenia[98]. An association between an allele at rs10503253 within *CSMD1* and cognitive functioning has been observed[99–101] and demonstrated the strongest association with schizophrenia in the PGC GWAS. Collectively, our *CSMD1* and MHC findings are consistent with evidence of an immune component involved in the pathophysiology of psychiatric conditions.

A number of additional genes/gene regions previously linked to MD (e.g., *CX3CR1, CACNA1A*,)[99,102,103] and phenotypes related to MD also emerged in the present genome-wide study. For example, MD cases exhibited greater variation in *CDC42BPB*. This gene was recently identified in a genome-wide DNAm study of current depressive symptoms[104]. This gene is highly expressed in the brain and is involved in the regulation of cell migration and neurite outgrowth[58]. Serotonin related genes, *HTR5A* and *HTR5A-AS1*, also exhibited greater variance in MD affected twins. This gene is a member of the 5HT (serotonin) receptor family, which has been implicated in MD and other psychiatric conditions[105–107]. Related, corticotropin releasing hormone receptor 2 (*CRHR2*) and cholinergic receptor nicotinic beta 2 subunit (*CHRNB*2) demonstrated increased variance in MD affected twins. Thus, complex genetic disorders such as MD reflect a large number of independent genetic factors that each contribute a small amount of variance to disease susceptibility and multiple psychiatric diseases likely share genetic risk factors.

Limitations associated with the current study design include its reliance on an all Caucasian sample, which diminishes generalizability of findings to other races. The current study also relied on peripheral blood tissue for DNAm. Although brain tissue is preferred to study the pathophysiology of MD, evidence suggests moderate consistency in DNAm between blood and brain tissues[15,71]. Consistent with epidemiological findings[108], females represented the majority of the sample. Thus, increased numbers of males are needed to determine potential sex-related differences. The cross-sectional design of the current study also does not allow for determination of epigenetic differences as cause or consequence of MD onset. Strengths of the current study design include its use of the MZ co-twin control method in conjunction with a sensitive developmental window to reveal genome-wide DNAm biomarkers associated with early-onset MD.

References

1. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. 2003;289(23):3095-3105. doi:10.1001/jama.289.23.3095

2. Weissman MM, Wolk S, Goldstein RB, et al. Depressed adolescents grown up. *JAMA*. 1999;281(18):1707-1713.

3. Lewinsohn PM, Rohde P, Seeley JR. Major depressive disorder in older adolescents: prevalence, risk factors, and clinical implications. *Clin Psychol Rev*. 1998;18(7):765-794.

4. Saluja G, Iachan R, Scheidt PC, Overpeck MD, Sun W, Giedd JN. Prevalence of and risk factors for depressive symptoms among young adolescents. *Arch Pediatr Adolesc Med*. 2004;158(8):760-765. doi:10.1001/archpedi.158.8.760

5. Kerestes R, Davey CG, Stephanou K, Whittle S, Harrison BJ. Functional brain imaging studies of youth depression: a systematic review. *Neuroimage Clin*. 2014;4:209-231. doi:10.1016/j.nicl.2013.11.009

6. Hankin BL, Abramson LY, Moffitt TE, Silva PA, McGee R, Angell KE. Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. *J Abnorm Psychol*. 1998;107(1):128-140.

7. Weissman MM, Wolk S, Wickramaratne P, et al. Children with prepubertal-onset major depressive disorder and anxiety grown up. *Arch Gen Psychiatry*. 1999;56(9):794-801.

8. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. *Arch Gen Psychiatry*. 2003;60(8):837-844. doi:10.1001/archpsyc.60.8.837

9. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry*. 2000;157(10):1552-1562. doi:10.1176/appi.ajp.157.10.1552

10. Wray NR, Sullivan PF. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. July 2017. doi:10.1101/167577

11. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry*. 2000;157(10):1552-1562. doi:10.1176/appi.ajp.157.10.1552

12. 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. doi:10.1176/ajp.152.6.833

13. Kessler RC. The effects of stressful life events on depression. *Annu Rev Psychol*. 1997;48:191-214. doi:10.1146/annurev.psych.48.1.191

14. Klengel T, Pape J, Binder EB, Mehta D. The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology*. 2014;80:115-132. doi:10.1016/j.neuropharm.2014.01.013

15. Bagot RC, Labonte B, Pena CJ, Nestler EJ. Epigenetic signaling in psychiatric disorders: stress and depression. *Dialogues Clin Neurosci*. 2014;16(3):281-295.

16. Szyf M, McGowan P, Meaney MJ. The social environment and the epigenome. *Environ Mol Mutagen*. 2008;49(1):46-60. doi:10.1002/em.20357

17. Weaver ICG, Cervoni N, Champagne FA, et al. Epigenetic programming by maternal behavior. *Nat Neurosci*. 2004;7(8):847-854. doi:10.1038/nn1276

18. Weaver ICG, Szyf M, Meaney MJ. From maternal care to gene expression: DNA methylation and the maternal programming of stress responses. *Endocr Res*. 2002;28(4):699.

19. Suderman M, McGowan PO, Sasaki A, et al. Conserved epigenetic sensitivity to early life experience in the rat and human hippocampus. *Proc Natl Acad Sci U S A*. 2012;109 Suppl 2:17266-17272. doi:10.1073/pnas.1121260109

20. Fish EW, Shahrokh D, Bagot R, et al. Epigenetic programming of stress responses through variations in maternal care. *Ann N Y Acad Sci*. 2004;1036:167-180. doi:10.1196/annals.1330.011

21. Meaney MJ, Szyf M, Seckl JR. Epigenetic mechanisms of perinatal programming of hypothalamic-pituitary-adrenal function and health. *Trends Mol Med*. 2007;13(7):269-277. doi:10.1016/j.molmed.2007.05.003

22. McGowan PO, Suderman M, Sasaki A, et al. Broad epigenetic signature of maternal care in the brain of adult rats. *PLoS One*. 2011;6(2):e14739. doi:10.1371/journal.pone.0014739

23. Borrelli E, Nestler EJ, Allis CD, Sassone-Corsi P. Decoding the epigenetic language of neuronal plasticity. *Neuron*. 2008;60(6):961-974. doi:10.1016/j.neuron.2008.10.012

24. Anacker C, O'Donnell KJ, Meaney MJ. Early life adversity and the epigenetic programming of hypothalamic-pituitary-adrenal function. *Dialogues Clin Neurosci*. 2014;16(3):321-333.

25. McGowan PO, Sasaki A, D'Alessio AC, et al. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci*. 2009;12(3):342-348. doi:10.1038/nn.2270

26. Szyf M, Weaver I, Meaney M. Maternal care, the epigenome and phenotypic differences in behavior. *Reprod Toxicol*. 2007;24(1):9-19. doi:10.1016/j.reprotox.2007.05.001

27. Zhang TY, Labonte B, Wen XL, Turecki G, Meaney MJ. Epigenetic mechanisms for the early environmental regulation of hippocampal glucocorticoid receptor gene expression in rodents and humans. *Neuropsychopharmacology*. 2013;38(1):111-123. doi:10.1038/npp.2012.149

28. Szyf M. The early life environment and the epigenome. *Biochim Biophys Acta*. 2009;1790(9):878-885. doi:10.1016/j.bbagen.2009.01.009

29. Lappalainen T, Greally JM. Associating cellular epigenetic models with human phenotypes. *Nat Rev Genet*. 2017;18(7):441-451. doi:10.1038/nrg.2017.32

30. Grundberg E, Meduri E, Sandling JK, et al. Global Analysis of DNA Methylation Variation in Adipose Tissue from Twins Reveals Links to Disease-Associated Variants in Distal Regulatory Elements. *The American Journal of Human Genetics*. 2013;93(5):876-890. doi:10.1016/j.ajhg.2013.10.004

31. Bell JT, Pai AA, Pickrell JK, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol*. 2011;12(1):R10. doi:10.1186/gb-2011-12-1-r10

32. Chen L, Ge B, Casale FP, et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell*. 2016;167(5):1398-1414.e24. doi:10.1016/j.cell.2016.10.026

33. Banovich NE, Lan X, McVicker G, et al. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLoS Genet*. 2014;10(9):e1004663. doi:10.1371/journal.pgen.1004663

34. Zhang D, Cheng L, Badner JA, et al. Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet*. 2010;86(3):411-419. doi:10.1016/j.ajhg.2010.02.005

35. Oh G, Wang S-C, Pal M, et al. DNA modification study of major depressive disorder: beyond locus-by-locus comparisons. *Biol Psychiatry*. 2015;77(3):246-255. doi:10.1016/j.biopsych.2014.06.016

36. Zeng Q, Chen X, Ning C, et al. Methylation of the genes ROD1, NLRC5, and HKR1 is associated with aging in Hainan centenarians. *BMC Med Genomics*. 2018;11(1):7. doi:10.1186/s12920-018-0334-1

37. Unnikrishnan A, Hadad N, Masser DR, Jackson J, Freeman WM, Richardson A. Revisiting the genomic hypomethylation hypothesis of aging. *Ann N Y Acad Sci*. 2018;1418(1):69-79. doi:10.1111/nyas.13533

38. Masser DR, Hadad N, Porter H, et al. Analysis of DNA modifications in aging research. *Geroscience*. 2018;40(1):11-29. doi:10.1007/s11357-018-0005-3

39. Levesque ML, Casey KF, Szyf M, et al. Genome-wide DNA methylation variability in adolescent monozygotic twins followed since birth. *Epigenetics*. 2014;9(10):1410-1421. doi:10.4161/15592294.2014.970060

40. Cecilione JL, Rappaport LM, Hahn SE, et al. Genetic and Environmental Contributions of Negative Valence Systems to Internalizing Pathways. *Twin Res Hum Genet*. 2018;21(1):12-23. doi:10.1017/thg.2017.72

41. Lilley ECH, Silberg JL. The Mid-Atlantic Twin Registry, revisited. *Twin Res Hum Genet*. 2013;16(1):424-428. doi:10.1017/thg.2012.125

42. Bachmann CJ, Aagaard L, Burcu M, et al. Trends and patterns of antidepressant use in children and adolescents from five western countries, 2005-2012. *Eur Neuropsychopharmacol*. 2016;26(3):411-419. doi:10.1016/j.euroneuro.2016.02.001

43. Kessler RC, Ustun TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res*. 2004;13(2):93-121.

44. Daviss WB, Birmaher B, Melhem NA, Axelson DA, Michaels SM, Brent DA. Criterion validity of the Mood and Feelings Questionnaire for depressive episodes in clinic and non-clinic subjects. *J Child Psychol Psychiatry*. 2006;47(9):927-934. doi:10.1111/j.1469-7610.2006.01646.x

45. Ong M-L, Holbrook JD. Novel region discovery method for Infinium 450K DNA methylation data reveals changes associated with aging in muscle and neuronal pathways. *Aging Cell*. 2014;13(1):142-155. doi:10.1111/acel.12159

46. Jaffe AE, Murakami P, Lee H, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol*. 2012;41(1):200-209. doi:10.1093/ije/dyr238

47. Lee H, Jaffe AE, Feinberg JI, et al. DNA methylation shows genome-wide association of NFIX, RAPGEF2 and MSRB3 with gestational age at birth. *Int J Epidemiol*. 2012;41(1):188-199. doi:10.1093/ije/dyr237

48. Jaffe AE, Feinberg AP, Irizarry RA, Leek JT. Significance analysis and statistical dissection of variably methylated regions. *Biostatistics*. 2012;13(1):166-178. doi:10.1093/biostatistics/kxr013

49. Butcher LM, Beck S. Probe Lasso: a novel method to rope in differentially methylated regions with 450K DNA methylation data. *Methods*. 2015;72:21-28. doi:10.1016/j.ymeth.2014.10.036

50. Wright ML, Dozmorov MG, Wolen AR, et al. Establishing an analytic pipeline for genome-wide DNA methylation. *Clin Epigenetics*. 2016;8:45. doi:10.1186/s13148-016-0212-7

51. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990;9(7):811-818.

52. Bates M, Martin M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. 2015;67(1):1-48.

53. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. 2001;98(9):5116-5121. doi:10.1073/pnas.091062498

54. R Core Team. *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2013.

55. Carlson MRJ, Pages H, Arora S, Obenchain V, Morgan M. Genomic Annotation Resources in R/Bioconductor. *Methods Mol Biol*. 2016;1418:67-90. doi:10.1007/978-1-4939-3578-9_4

56. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25-29. doi:10.1038/75556

57. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287. doi:10.1089/omi.2011.0118

58. Martin Morgan. AnnotationHub: Client to access AnnotationHub resource.

59. Faraway JJ. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. New York: Chapman and Hall; 2006.

60. Peek SL, Mah KM, Weiner JA. Regulation of neural circuit formation by protocadherins. *Cell Mol Life Sci*. 2017;74(22):4133-4157. doi:10.1007/s00018-017-2572-3

61. Lefebvre JL, Kostadinov D, Chen WV, Maniatis T, Sanes JR. Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature*. 2012;488(7412):517-521. doi:10.1038/nature11305

62. Suo L, Lu H, Ying G, Capecchi MR, Wu Q. Protocadherin clusters and cell adhesion kinase regulate dendrite complexity through Rho GTPase. *J Mol Cell Biol*. 2012;4(6):362-376. doi:10.1093/jmcb/mjs034

64. Chen WV, Maniatis T. Clustered protocadherins. *Development*. 2013;140(16):3297-3302. doi:10.1242/dev.090621

65. Hayashi S, Takeichi M. Emerging roles of protocadherins: from self-avoidance to enhancement of motility. *J Cell Sci*. 2015;128(8):1455-1464. doi:10.1242/jcs.166306

66. Hirano S, Takeichi M. Cadherins in brain morphogenesis and wiring. *Physiol Rev*. 2012;92(2):597-634. doi:10.1152/physrev.00014.2011

67. Hirayama T, Yagi T. Regulation of clustered protocadherin genes in individual neurons. *Semin Cell Dev Biol*. 2017;69:122-130. doi:10.1016/j.semcdb.2017.05.026

68. Pedrosa E, Stefanescu R, Margolis B, et al. Analysis of protocadherin alpha gene enhancer polymorphism in bipolar disorder and schizophrenia. *Schizophr Res*. 2008;102(1-3):210-219. doi:10.1016/j.schres.2008.04.013

69. Anitha A, Thanseem I, Nakamura K, et al. Protocadherin alpha (PCDHA) as a novel susceptibility gene for autism. *J Psychiatry Neurosci*. 2013;38(3):192-198. doi:10.1503/jpn.120058

70. Dempster EL, Wong CCY, Lester KJ, et al. Genome-wide methylomic analysis of monozygotic twins discordant for adolescent depression. *Biol Psychiatry*. 2014;76(12):977-983. doi:10.1016/j.biopsych.2014.04.013

71. Cordova-Palomera A, Fatjo-Vilas M, Gasto C, Navarro V, Krebs M-O, Fananas L. Genome-wide methylation study on depression: differential methylation and variable methylation in monozygotic twins. *Transl Psychiatry*. 2015;5:e557. doi:10.1038/tp.2015.49

72. Xiao X, Zheng F, Chang H, et al. The Gene Encoding Protocadherin 9 (PCDH9), a Novel Risk Factor for Major Depressive Disorder. *Neuropsychopharmacology*. 2018;43(5):1128-1137. doi:10.1038/npp.2017.241

73. Chang H, Hoshina N, Zhang C, et al. The protocadherin 17 gene affects cognition, personality, amygdala structure and function, synapse development and risk of major mood disorders. *Mol Psychiatry*. 2018;23(2):400-412. doi:10.1038/mp.2016.231

74. Russo SJ, Nestler EJ. The brain reward circuitry in mood disorders. *Nat Rev Neurosci*. 2013;14(9):609-625. doi:10.1038/nrn3381

75. Kim SY, Mo JW, Han S, et al. The expression of non-clustered protocadherins in adult rat hippocampal formation and the connecting brain regions. *Neuroscience*. 2010;170(1):189-199. doi:10.1016/j.neuroscience.2010.05.027

76. Weiner RD, Reti IM. Key updates in the clinical application of electroconvulsive therapy. *Int Rev Psychiatry*. 2017;29(2):54-62. doi:10.1080/09540261.2017.1309362

77. Yoshinaga S, Nakajima K. A crossroad of neuronal diversity to build circuitry. *Science*. 2017;356(6336):376-377. doi:10.1126/science.aan2856

78. Katori S, Hamada S, Noguchi Y, et al. Protocadherin-alpha family is required for serotonergic projections to appropriately innervate target brain areas. *J Neurosci*. 2009;29(29):9137-9147. doi:10.1523/JNEUROSCI.5478-08.2009

79. Katori S, Noguchi-Katori Y, Okayama A, et al. Protocadherin-alphaC2 is required for diffuse projections of serotonergic axons. *Sci Rep*. 2017;7(1):15908. doi:10.1038/s41598-017-16120-y

80. Wu Y, Hou F, Wang X, Kong Q, Han X, Bai B. Aberrant Expression of Histone Deacetylases 4 in Cognitive Disorders: Molecular Mechanisms and a Potential Target. *Front Mol Neurosci*. 2016;9:114. doi:10.3389/fnmol.2016.00114

81. Uchida S, Shumyatsky GP. Synaptically Localized Transcriptional Regulators in Memory Formation. *Neuroscience*. 2018;370:4-13. doi:10.1016/j.neuroscience.2017.07.023

82. Kim M-S, Akhtar MW, Adachi M, et al. An essential role for histone deacetylase 4 in synaptic plasticity and memory formation. *J Neurosci*. 2012;32(32):10879-10886. doi:10.1523/JNEUROSCI.2089-12.2012

83. MacQueen GM, Memedovich KA. Cognitive dysfunction in major depression and bipolar disorder: Assessment and treatment options. *Psychiatry Clin Neurosci*. 2017;71(1):18-27. doi:10.1111/pcn.12463

84. Talarowska M, Berk M, Maes M, Galecki P. Autobiographical memory dysfunctions in depressive disorders. *Psychiatry Clin Neurosci*. 2016;70(2):100-108. doi:10.1111/pcn.12370

85. Covington HE 3rd, Maze I, Vialou V, Nestler EJ. Antidepressant action of HDAC inhibition in the prefrontal cortex. *Neuroscience*. 2015;298:329-335. doi:10.1016/j.neuroscience.2015.04.030

86. Covington HE 3rd, Maze I, LaPlant QC, et al. Antidepressant actions of histone deacetylase inhibitors. *J Neurosci*. 2009;29(37):11451-11460. doi:10.1523/JNEUROSCI.1758-09.2009

87. Covington HE 3rd, Vialou VF, LaPlant Q, Ohnishi YN, Nestler EJ. Hippocampal-dependent antidepressant-like activity of histone deacetylase inhibition. *Neurosci Lett*. 2011;493(3):122-126. doi:10.1016/j.neulet.2011.02.022

88. Meylan EM, Halfon O, Magistretti PJ, Cardinaux J-R. The HDAC inhibitor SAHA improves depressive-like behavior of CRTC1-deficient mice: Possible relevance for treatment-resistant depression. *Neuropharmacology*. 2016;107:111-121. doi:10.1016/j.neuropharm.2016.03.012

89. Bi L-L, Sun X-D, Zhang J, et al. Amygdala NRG1-ErbB4 is critical for the modulation of anxiety-like behaviors. *Neuropsychopharmacology*. 2015;40(4):974-986. doi:10.1038/npp.2014.274

90. Wang Y-N, Figueiredo D, Sun X-D, et al. Controlling of glutamate release by neuregulin3 via inhibiting the assembly of the SNARE complex. *Proc Natl Acad Sci U S A*. 2018;115(10):2508-2513. doi:10.1073/pnas.1716322115

91. Chong VZ, Thompson M, Beltaifa S, Webster MJ, Law AJ, Weickert CS. Elevated neuregulin-1 and ErbB4 protein in the prefrontal cortex of schizophrenic patients. *Schizophr Res*. 2008;100(1-3):270-280. doi:10.1016/j.schres.2007.12.474

92. Wen Z, Chen J, Khan RAW, et al. Genetic association between NRG1 and schizophrenia, major depressive disorder, bipolar disorder in Han Chinese population. *Am J Med Genet B Neuropsychiatr Genet*. 2016;171B(3):468-478. doi:10.1002/ajmg.b.32428

93. Mostaid MS, Lloyd D, Liberg B, et al. Neuregulin-1 and schizophrenia in the genome-wide association study era. *Neurosci Biobehav Rev*. 2016;68:387-409. doi:10.1016/j.neubiorev.2016.06.001

94. McIntosh AM, Hall J, Lymer GKS, Sussmann JED, Lawrie SM. Genetic risk for white matter abnormalities in bipolar disorder. *Int Rev Psychiatry*. 2009;21(4):387-393. doi:10.1080/09540260902962180

95. Chen X, Long F, Cai B, Chen X, Qin L, Chen G. A Novel Relationship for Schizophrenia, Bipolar, and Major Depressive Disorder. Part 8: a Hint from Chromosome 8 High Density Association Screen. *Mol Neurobiol*. 2017;54(8):5868-5882. doi:10.1007/s12035-016-0102-1

96. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421-427. doi:10.1038/nature13595

97. Ruderfer DM, Fanous AH, Ripke S, et al. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry*. 2014;19(9):1017-1024. doi:10.1038/mp.2013.138

98. Sekar A, Bialas AR, de Rivera H, et al. Schizophrenia risk from complex variation of complement component 4. *Nature*. 2016;530(7589):177-183. doi:10.1038/nature16549

99. Athanasiu L, Giddaluru S, Fernandes C, et al. A genetic association study of CSMD1 and CSMD2 with cognitive function. *Brain Behav Immun*. 2017;61:209-216. doi:10.1016/j.bbi.2016.11.026

100. Donohoe G, Walters J, Hargreaves A, et al. Neuropsychological effects of the CSMD1 genome-wide associated schizophrenia risk variant rs10503253. *Genes Brain Behav*. 2013;12(2):203-209. doi:10.1111/gbb.12016

101. Koiliari E, Roussos P, Pasparakis E, et al. The CSMD1 genome-wide associated schizophrenia risk variant rs10503253 affects general cognitive ability and executive function in healthy males. *Schizophr Res*. 2014;154(1-3):42-47. doi:10.1016/j.schres.2014.02.017

102. Sherva R, Wang Q, Kranzler H, et al. Genome-wide Association Study of Cannabis Dependence Severity, Novel Risk Variants, and Shared Genetic Risks. *JAMA Psychiatry*. 2016;73(5):472-480. doi:10.1001/jamapsychiatry.2016.0036

103. Lotan A, Fenckova M, Bralten J, et al. Neuroinformatic analyses of common and distinct genetic components associated with major neuropsychiatric disorders. *Front Neurosci*. 2014;8:331. doi:10.3389/fnins.2014.00331

104. Story Jovanova O, Nedeljkovic I, Derek S, et al. DNA Methylation Signatures of Depressive Symptoms in Middle-aged and Elderly Persons: Meta-analysis of Multiethnic Epigenome-wide Studies. *JAMA Psychiatry*. July 2018. doi:10.1001/jamapsychiatry.2018.1725

105. Brezo J, Bureau A, Merette C, et al. Differences and similarities in the serotonergic diathesis for suicide attempts and mood disorders: a 22-year longitudinal gene-environment study. *Mol Psychiatry*. 2010;15(8):831-843. doi:10.1038/mp.2009.19

106. Dong Z-Q, Li X-R, He L, He G, Yu T, Sun X-L. 5-HTR1A and 5-HTR2A genetic polymorphisms and SSRI antidepressant response in depressive Chinese patients. *Neuropsychiatr Dis Treat*. 2016;12:1623-1629. doi:10.2147/NDT.S93562

107. Yosifova A, Mushiroda T, Stoianov D, et al. Case-control association study of 65 candidate genes revealed a possible association of a SNP of HTR5A to be a factor susceptible to bipolar disease in Bulgarian population. *J Affect Disord*. 2009;117(1-2):87-97. doi:10.1016/j.jad.2008.12.021

108. 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. doi:10.1001/jamapsychiatry.2017.4602

109. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*.; 2013.

110. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. A population-based twin study of major depression in women. The impact of varying definitions of illness. *Arch Gen Psychiatry*. 1992;49(4):257-266.

111. Meier A. The research diagnostic criteria: historical background, development, validity, and reliability. *Can J Psychiatry*. 1979;24(2):167-178.

112. Endicott J, Spitzer RL. Use of the Research Diagnostic Criteria and the Schedule for Affective Disorders and Schizophrenia to study affective disorders. *Am J Psychiatry*. 1979;136(1):52-56. doi:10.1176/ajp.136.1.52

113. Spitzer RL, Endicott J, Robins E. Research diagnostic criteria. *Psychopharmacol Bull*. 1975;11(3):22-25.

114. Spitzer RL, Endicott J, Robins E. Research diagnostic criteria: rationale and reliability. *Arch Gen Psychiatry*. 1978;35(6):773-782.

115. Kasriel J, Eaves L. The zygosity of twins: further evidence on the agreement between diagnosis by blood groups and written questionnaires. *J Biosoc Sci*. 1976;8(3):263-266.

116. Jackson RW, Snieder H, Davis H, Treiber FA. Determination of twin zygosity: a comparison of DNA with various questionnaire indices. *Twin Res*. 2001;4(1):12-18.

117. Bibikova M, Barnes B, Tsan C, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011;98(4):288-295. doi:10.1016/j.ygeno.2011.07.007

118. Morris TJ, Beck S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods*. 2015;72:3-8. doi:10.1016/j.ymeth.2014.08.011

119. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049

120. Barfield RT, Almli LM, Kilaru V, et al. Accounting for population stratification in DNA methylation studies. *Genet Epidemiol*. 2014;38(3):231-241. doi:10.1002/gepi.21789

121. Du P, Zhang X, Huang C-C, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11:587. doi:10.1186/1471-2105-11-587

122. Chen Y, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8(2):203-209. doi:10.4161/epi.23470

123. Touleimat N, Tost J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*. 2012;4(3):325-341. doi:10.2217/epi.12.21

124. Leek JT, Scharpf RB, Bravo HC, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010;11(10):733-739. doi:10.1038/nrg2825

125. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882-883. doi:10.1093/bioinformatics/bts034

126. Houseman EA, Kelsey KT, Wiencke JK, Marsit CJ. Cell-composition effects in the analysis of DNA methylation array data: a mathematical perspective. *BMC Bioinformatics*. 2015;16:95. doi:10.1186/s12859-015-0527-y

Acknowledgements

The research outcomes presented in this manuscript were supported by a NARSAD Independent Investigator Award from the *Brain and Behavior Research Foundation* to the first author (RRN). The development of analytic methods were supported by a NARSAD Independent Investigator Award from the *Brain and Behavior Research Foundation* to the last author (TPY).

Table 1. Demographic and clinical characteristics of twins meeting definite or probable DSM-5 criteria for lifetime history of MD (MD Affected) versus no lifetime history of MD (MD Unaffected).

	M (SD) or n (%)			
	MD Unaffected n=111	**MD Affected n=39**	t/χ^2	***p***
Demographic/Sample				
Age, years	17.49 (1.3)	17.60 (1.3)	0.66	0.51
Sex, Female	81 (73.0%)	29 (74.4%)	0.28	0.87
Ethnicity, Hispanic	5 (4.5%)	3 (7.7%)	0.58	0.43
Nicotine Use[§], Current Smoker	2 (1.8%)	3 (7.7%)	3.11	0.11
Clinical Characteristics				
SMFQ	4.4 (3.7)	9.0 (6.0)	6.40	<.001
History of psychotropic medication use	2 (1.8%)	4 (10.3%)	0.64	0.62
Panic Disorder	5 (4.5%)	3 (7.7%)	0.56	0.43
Social Anxiety Disorder	10 (9.1%)	8 (20.5%)	3.54	0.08
Specific Phobia	8 (7.3%)	7 (17.9%)	3.63	0.07
Generalized Anxiety Disorder	2 (1.8%)	4 (10.3%)	5.31	0.04
MD Features				
Age of Onset (years)	---	14.9 (1.7)		
Number of Major Depressive Episodes, n				
1 episode	---	17 (43.6%)		
2-3 episodes	---	16 (41.0%)		
3-5 episodes	---	3 (7.7%)		
≥6 episodes	---	3 (7.7%)		
Number of symptoms during worst MDE	---	5.69 (1.1)		

Note: SMFQ=Short Mood and Feelings Questionnaire.[§]Fagerstrom Test for Nicotine Dependence score ranged 2 (low dependence) to 6 (moderate dependence) with Mode=2, Median=2, and Mean=3.

Figure 1. Overlap with PGC GWAS of major depression. The 'MD locus' (purple box) represents a region of chromosome 6 extending from 27.7–32.8 Mb found to be significantly associated with major depressive disorder by Wray et al. (2018). Summary statistics from this study are plotted for the relevant regional markers in the Manhattan plot. Colored ticks represent the 3 DMRs (blue) and 1 VMR (red) located in this region. Individual plots above provide a zoomed-in view of the genomic context surrounding each methylation region and probe-level test statistics. Chromatin states within GM12878 lymphoblastoid cells are indicated by color coding the ChromHMM track.

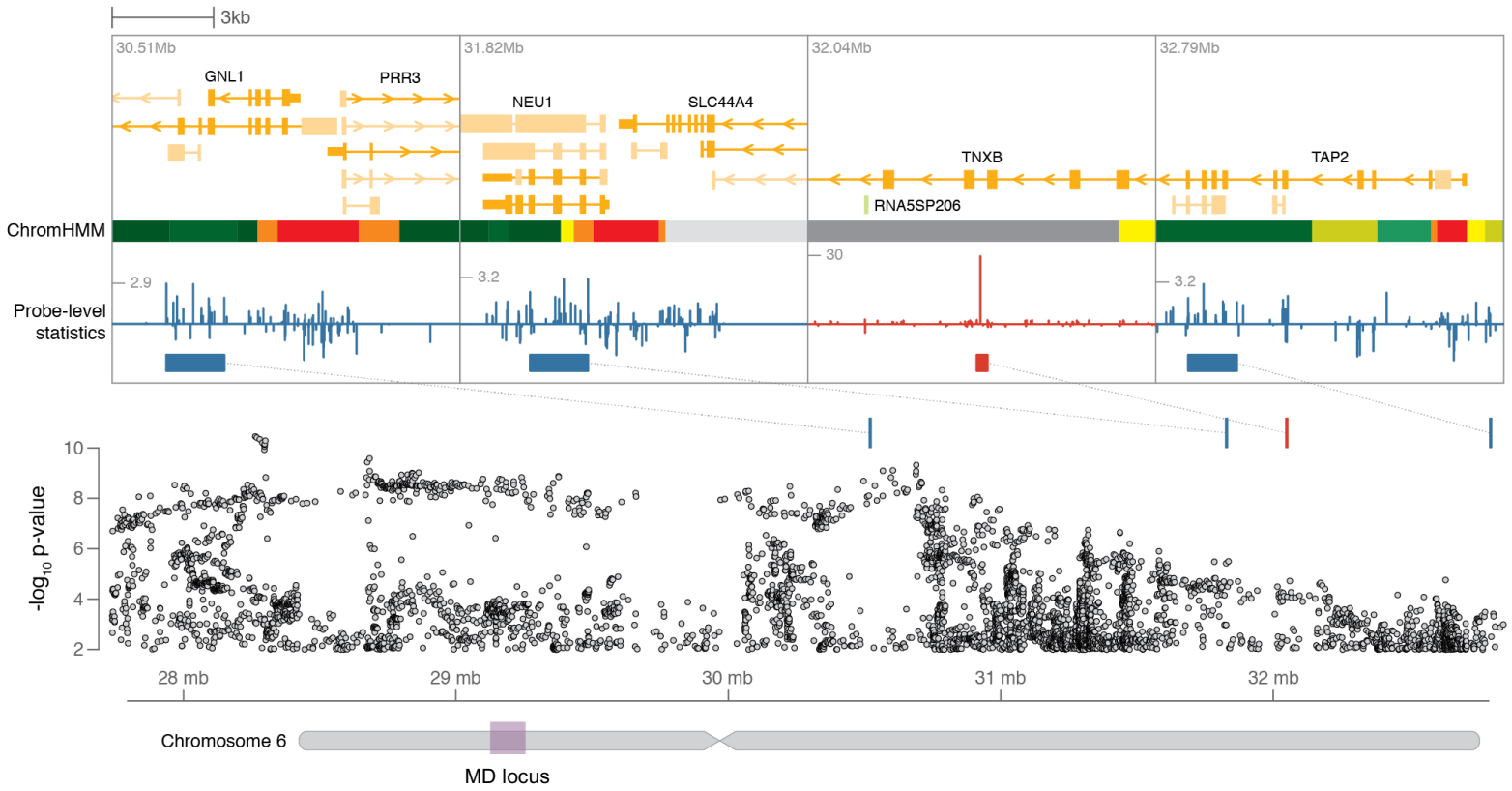


Table 2. Differentially methylated regions (DMRs) where MD affected twins exhibited higher means compared to MD unaffected twins.

Chromosome	Start	End	Symbol	EntrezID	Number_CpG	Empirical_AUC
chr1	36787678	36789401	SH3D21	79729	3	3836.3652
chr1	221053841	221055665	HLX	3142	4	4242.3503
chr2	240035107	240036791	HDAC4	9759	3	4057.6444
chr5	1090741	1092417	SLC12A7	10723	4	3818.3311
chr5	171709917	171711524	UBTD2	92181	3	4070.7302
chr5	176936563	176938522	DOK3	79930	6	3826.3018
chr6	30519905	30521619	NA	NA	5	3861.8967
chr6	31828260	31830030	NA	NA	4	4558.5396
chr6	32797253	32798887	NA	NA	4	3849.6722
chr6	39281541	39283313	KCNK17	89822	3	4122.0762
chr6	146863647	146865487	RAB32	10981	7	4701.7437
chr16	2023998	2025868	TBL3	10607	4	4329.8853
chr16	58767249	58769104	GOT2	2806	4	4046.9261
chr17	46659019	46660940	HOXB3	3213	4	4501.7814
chr17	70116185	70118162	SOX9	6662	4	4120.2233
chr19	13213428	13215387	LYL1	4066	3	4381.7764
chr21	38069321	38070994	NA	NA	4	3892.4767

Table 3. Variably methylated regions (VMRs) where MD affected twins exhibited greater variance compared to MD unaffected twins.

Chromosome	Start	End	Symbol	EntrezID	Number_CpG	Empirical_AUC
chr3	39321449	39323539	CX3CR1	1524	4	5572.9239
chr3	46925081	46925524	PTH1R	5745	2	5031.1289
chr3	170136920	170137321	CLDN11	5010	2	5110.2437
chr6	32049516	32049825	NA	NA	3	5038.8717
chr6	169284344	169287304	NA	NA	5	5276.0094
chr6	170595385	170597898	DLL1	28514	8	5628.6246
chr7	157225062	157225567	NA	NA	2	5131.4817
chr10	134739746	134741032	CFAP46	54777	4	5667.4646
chr11	64510112	64513156	RASGRP2	10235	7	5458.4387
chr17	40837037	40839469	CNTNAP1	8506	4	5035.574

Table 4. Gene enrichment analysis summary.

	Ontology Category	Description	Gene Ratio	q-value	Gene Symbol
GO:0007156	Biological Process	homophilic cell adhesion via plasma membrane adhesion molecule	23, 568	0.00002	CDH3, CDH6, PCDHαC1-2, PCDHα1-13, PCDH10, PCDH20, PTPRT, CADM1, PLXNB3, CLSTN2
GO:0098742	Biological Process	cell-cell adhesion via plasma-membrane adhesion molecules	28, 568	0.00002	CDH3, CDH6, PTPRT, CLDN4, CADM1, GRID2, ITGA5, CLDN11, PLXNB3, PCDHαC1-2, PCDHα1-13, PCDH20, PTPRD, CLSTN2
GO:0044306	Cellular Function	neuron projection terminus	14, 605	0.05888	BAIAP2, AAK1, CYFIP1, SYT11, PACSIN1, ILK, PFN2, PNOC, PVALB, DNAJC5, MOB2, NAPA, BSN, AP3D1
GO:0043195	Cellular Function	terminal bouton	9, 605	0.05888	AAK1, CYFIP1, SYT11, ILK, PFN2, PVALB, DNAJC5, NAPA, AP3D1
GO:0033267	Cellular Function	axon part	16, 605	0.07314	AAK1, TRAK1, CYFIP1, SYT11, KIF13B, AP3M1, NRG1, ILK, PFN2, PVALB, SPG7, DAGLA, DNAJC5, CNTNAP1, NAPA, AP3D1
GO:0044463	Cellular Function	cell projection part	54, 605	0.08617	AKAP9, RASGRP2, BAIAP2, DCTN2, IGF2BP1, EHD1, CNGA4, PACRG, EPS8, ENKUR, SPATA13, AAK1, TRAK1, CYFIP1, SYT11, KIF13B, TBC1D30, GPR161, AP3M1, NPHP3, BBS9, GRID2, PACSIN1, NRG1, ILK, ITGA5, KCNC3, PDE6B, PFN2, PNOC, SSH1, CFAP46, CCDC40, PRKAR1B, TTYH1, PTH1R, TENM2, PVALB, BBS2, MAP2K4, SLC22A5, SPG7, TIMP2, DAGLA, CA9, JADE1, DNAJC5, MOB2, CNTNAP1, NAPA, BSN, AP3D1, DHRS3, RAB28

GO:0030424	Cellular Function	Axon	26, 605	0.08617	IGF2BP1, DAB2IP, EPHB2, AAK1, TRAK1, CYFIP1, SYT11, KIF13B, LDLRAP1, AP3M1, NRG1, ILK, KCNC3, NF1, PFN2, SLC17A7, SEMA6A, PVALB, MAP2K4, SPG7, DAGLA, DNAJC5, CNTNAP1, NAPA, BSN, AP3D1
GO:0098793	Cellular Function	Presynapse	24, 605	0.08617	BAIAP2, AAK1, CYFIP1, SYT11, DMXL2, AMPH, GRIK4, GRM8, PACSIN1, HIP1, APBA2, ILK, KCNC3, NF1, SYT17, PFN2, PNOC, PI4K2A, SLC17A7, PVALB, DNAJC5, NAPA, BSN, AP3D1

SUPPLEMENTAL METHODS

The primary definition of MD affected status required presence of at least one DSM-5 MDE lasting at least two weeks[109]. MD was considered present if criteria were met at either a definite or a probable level. A definite diagnosis was made when a twin endorsed at least five DSM-5 Criterion A MDE symptoms (e.g., low energy, recurrent thoughts of death, inability to concentrate) of which at least one symptom had to include the presence of low mood and/or anhedonia most of the day, nearly every day for at least two weeks during their worst episode. Probable diagnoses were made when a twin endorsed at least 4 symptoms, one of which had to include the presence of low mood and/or anhedonia most of the day, nearly every day. The use of definite and probable levels allows for the inclusion of twins who may not be currently symptomatic and, therefore, have to rely on retrospective memory to determine their symptom presentation. The use of probable levels of diagnosis also is justified given that twins expressing nearly all symptoms of a MDE are more similar to a case than a control symptomatically and for depressive risk factors[110]. This approach to diagnosis was introduced as part of the Research Diagnostic Criteria and has been applied in many studies[111–114]. To qualify for MD at the definite or probable level, participants also had to report that their depression symptoms caused clinically significant distress or functional impairment (i.e., Criterion C). MD unaffected status was defined as no lifetime diagnosis of MD at the probable or definite threshold. The expanded MD section of the CIDI-SF followed by the lifetime MD algorithm is presented below.

Lifetime Major Depression Diagnostic Questions:
Criterion A (MDE Symptoms)

1. Have you ever had a time in your life when you felt sad, blue, or depressed or two weeks or more in a row? (yes/no)
2. Have you ever had a time in your life lasting two weeks or more when you lost interest in most things like hobbies, work, or activities that usually give you pleasure? (yes/no)
3. During those worst two weeks, did the feelings of sadness or loss of interest usually last all day long, most of the day, about half of the day, or less than half of the day? (All day long, Most of the day, About half of the day, Less than half of the day)
4. Did you feel this way every day, almost every day, or less often during the two weeks? (Every day, Almost every day, Less often)

If a participant responded “yes” to questions 1 OR 2 AND endorsed “all day long” or “most of the day” to question 3 AND responded “every day” or “almost every day” to question 4, they entered the MD section and were queried regarding other DSM-5 Criterion A symptoms of MD (see below).

5. Thinking about those same two weeks, did you feel more tired out or low on energy than is usual for you? (yes/no)
6. Did you gain or lose weight without trying or did you stay about the same? (Gained weight, Lost weight, Gained and lost weight, stayed about the same, I was on a diet)
7. About how much weight did you gain/you lose/your weight change? (numeric response)
8. Did you have more trouble falling asleep or staying asleep than you usually do during those two weeks? (yes/no)
    a. Did that happen every night, nearly every night, or less often during those two weeks? (Every night, Nearly every night, Less often)
9. During those two weeks, did you have a lot more trouble concentrating or making decisions than usual? (yes/no)

10. People sometimes feel down on themselves, no good, or worthless, or have excessive guilt and blame themselves for things. During that two-week period, did you feel this way? (yes/no)
11. Did you think a lot about death – either your own, someone else's, or death in general during those two weeks? (yes/no)

Criterion C (Functional Interference):

12. Did you ever tell a professional about these problems (such as a medical doctor, psychologist, social worker, counselor, nurse, clergy, or other helping professional)? (yes/no)
13. Did you ever take medication for these problems? (yes/no)
14. How much did these problems interfere with your life or activities – a lot, some, a little, or not at all? (A lot, Some, A little, Not at all)
    *additional questions regarding age of onset, timing of last episode, etc. were queried.

MD Algorithm

Questions 6 and 7 were recoded so that a response of "Gained weight" or "Gained and lost weight," along with at least a response of 5 pounds in question 7, was coded as endorsement of weight gain/loss. Participants who endorsed question 8 also were presented with question 8a. Sleep symptoms were considered present if a participant answered "Every night" or "Nearly every night" on question 8a. Responses to queries 1, 2, 5, weight gain, sleep symptoms, 9, 10, and 11 were summed as the total number of depression symptoms endorsed during the twin's worst MDE. Participants were deemed to have experienced significant distress if they endorsed question 14 as "a lot" or "some" and/or if they endorsed question 12 or 13 as "yes". MD at the full threshold level was coded positive if the participant endorsed question 1 or 2 along with question 3 as "All day long" or "Most of the day," and question 4 as "Every day" or "Almost every day," and at least 4 other depression symptoms for a total of at least 5 symptoms. MD at the full threshold level also required significant distress. MD was coded positive at the probable level if the participant endorsed question 1 or 2 as well as question 3 ("All day long" / "Most of the day") and question 4 ("Every day" / "Almost every day") and at least 3 other depression symptoms for a total of 4 symptoms. The probable level also required endorsement of significant distress (i.e., endorsed question 12, 13, or 14 as "a lot" or "some").

Exclusionary Criteria

Participants were not eligible for the current study if they met any of the following criteria: 1) current use of psychotropic medications (e.g., antianxiety/antidepressants) or medications with psychotropic effects (e.g., beta-adrenergic blockers), b) diagnosis of an autism spectrum disorder, c) diagnosis of an intellectual disability, d) diagnosis of a spatial learning disorder, or prior testing indicating an IQ below 70, e) seizure without a clear and resolved etiology, f) current or past episodes of psychosis, g) serious, not stabilized illness (e.g., liver, kidney, gastrointestinal, respiratory, cardiovascular, endocrinologic, neurologic, immunologic, or blood disease), h) inadequate production of human growth hormone, i) sensory integration disorder, j) congenital adrenal hyperplasia, k) adrenal inefficiency, l) deaf with bicochlear implants, m) cancer (current or past diagnosis), and n) pregnancy (current or lifetime). If only one twin from the twin pair met any of the exclusionary criteria, the whole pair was excluded.

Zygosity

Zygosity status (monozygotic [MZ] versus dizygotic [DZ]) for adolescent twins (age $\leq$ 17) was determined based on parent-report about physical similarities between twins. Adult twins (age $\geq$ 18) not accompanied by a parent/legal guardian completed the zygosity questionnaire

about themselves. Prior research has demonstrated high validity for this zygosity assessment as compared to blood[115] and DNA evaluations of zygosity[116]. MZ zygosity based on questionnaire assessment was confirmed using 65 SNP control probes included on the Infinium HumanMethylation450 (450K) array to verify sample identity. The control probes target polymorphic sequences, and values for each probe cluster into three groups corresponding to genotype. Together, the control probes provide strong support for validating zygosity. Only data from twins confirmed to be MZ were included in analyses.

Genome-wide DNAm Measurement and Processing

Genomic DNA was isolated from whole blood according to standard methods using the Puregene DNA Isolation Kit (Qiagen). An aliquot of 1 microgram DNA per subject was processed by HudsonAlpha Institute for Biotechnology for bisulfite conversion (Zymo Research EZ Methylation Kit) and genome-wide methylation assayed on the Infinium Human Methylation 450K BeadChip microarray, which interrogates 485,764 features. Twin pairs were localized to the same slide to minimize any potential artifactual differences in DNAm patterns due to batch effects.

Details of the 450K microarray have been previously described[117], and raw data processing was performed according to best practices reported in recent publications[50,118]. Intensity values from the scanned arrays were processed using the *minfi* Bioconductor package[119] in the R programming environment (R Development Core Team 2015). Confirmation of self-reported race was made by sample clustering derived from principal components estimated from ancestry informative probes[120].

Quality was assessed both quantitatively and visually to identify deviant samples[119]. Beta values were derived as the ratio of the methylated probe intensity to the sum of the methylated and unmethylated probe intensities[121]. Beta value density plots from each array were inspected to tag poor performing arrays based on a large deviation from the rest of the samples. Probes were filtered if they had a detection P-value of greater than 0.01 in at least 10% of samples or if they have been previously identified as cross-hybridizing[122] leaving a total of 455,828 probes to analyze. Quantile normalization adapted to DNAm arrays[123] was applied to adjust the distribution of type I and II probes to the final set of screened sample arrays and probes.

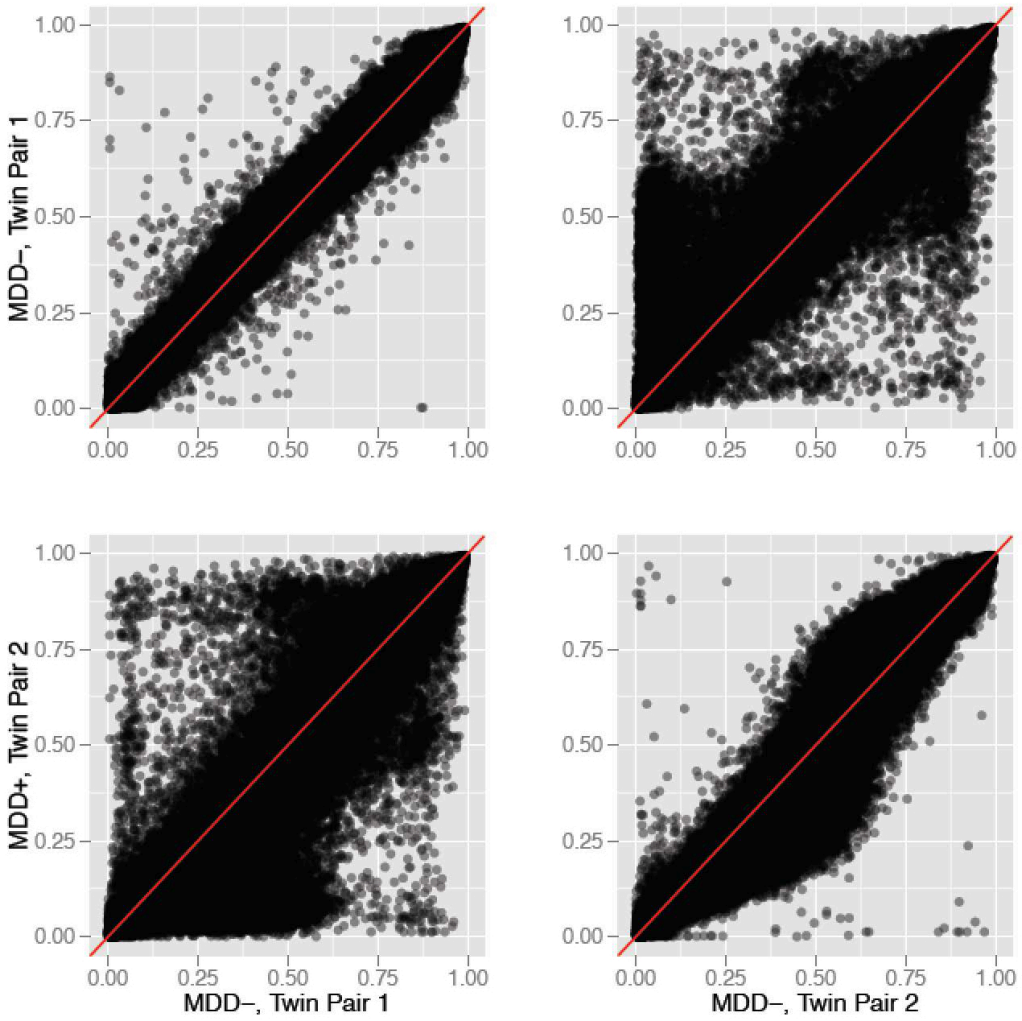
For all statistical tests, *beta* values were transformed using the *M*-value procedure to promote normality and calculated as a logit transformation of the methylated and unmethylated intensity ratio along with an added constant to offset potentially small values[121]. Correlations between major experimental factors and the top 10 principal components of *M*-values across all arrays were inspected to identify extraneous structure that may account for any batch effects[124]. ComBat was used to remove average differences across arrays due to slide groupings[125]. Blood cell proportions were inferred for each sample to account for cellular heterogeneity[126].

Supplemental Table 1. Statistical contrasts fit to estimate the degrees of association between MD affected/unaffected status and genome-wide DNA methylation markers.

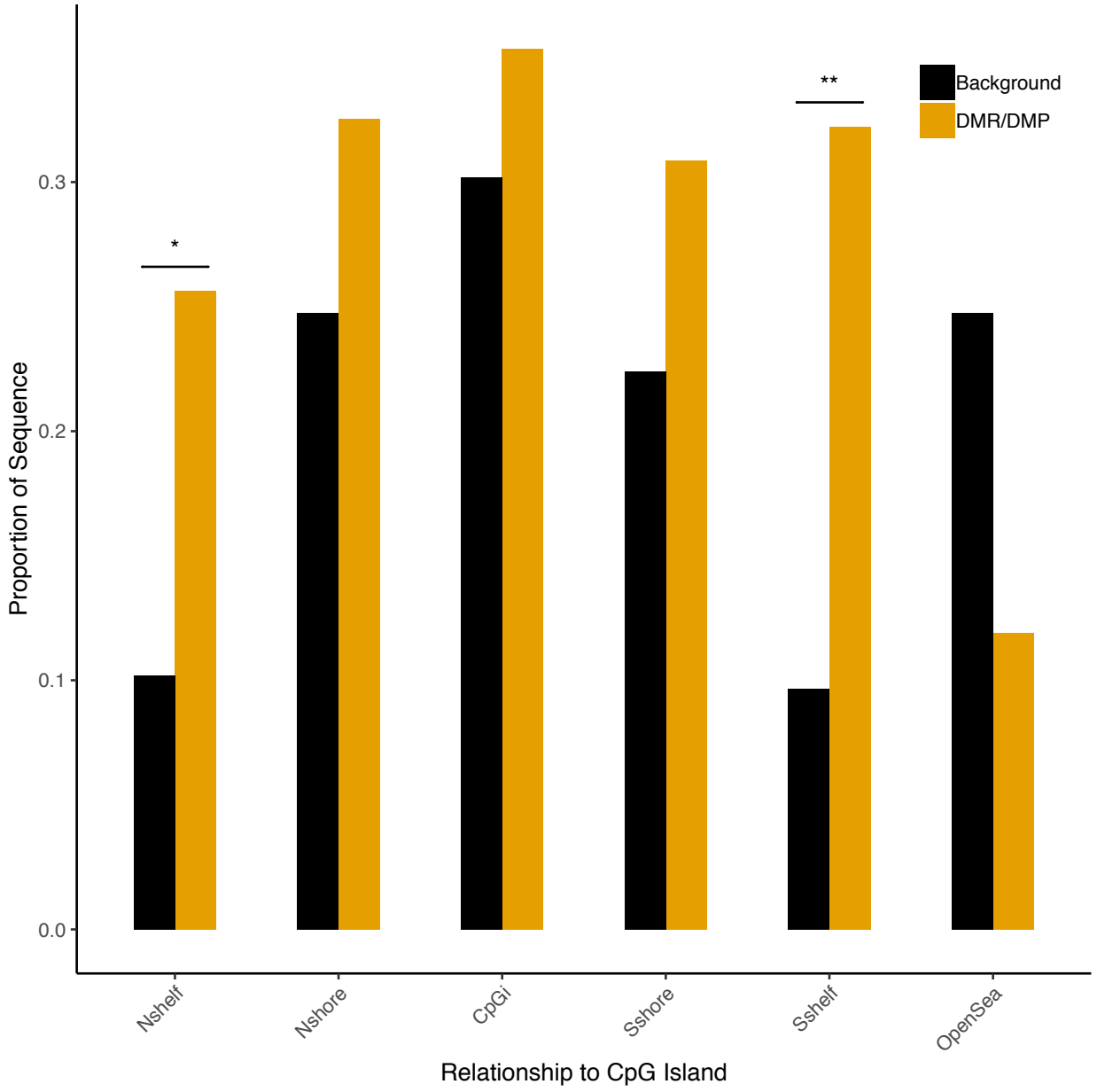
	Concord-ant MD T1	Concordant MD T2	Discordant T1	Discord-ant T2	Concord-ant Healthy T1	Concordant Healthy T2
contrast 1	-1	-1	-1	1	1	1
contrast 2	-1	-1	2	0	0	0
contrast 3	0	0	0	2	-1	-1
contrast 4	-1	1	0	0	0	0
contrast 5	0	0	0	0	-1	1

Note: T1 = Twin 1 and T2 = Twin 2.

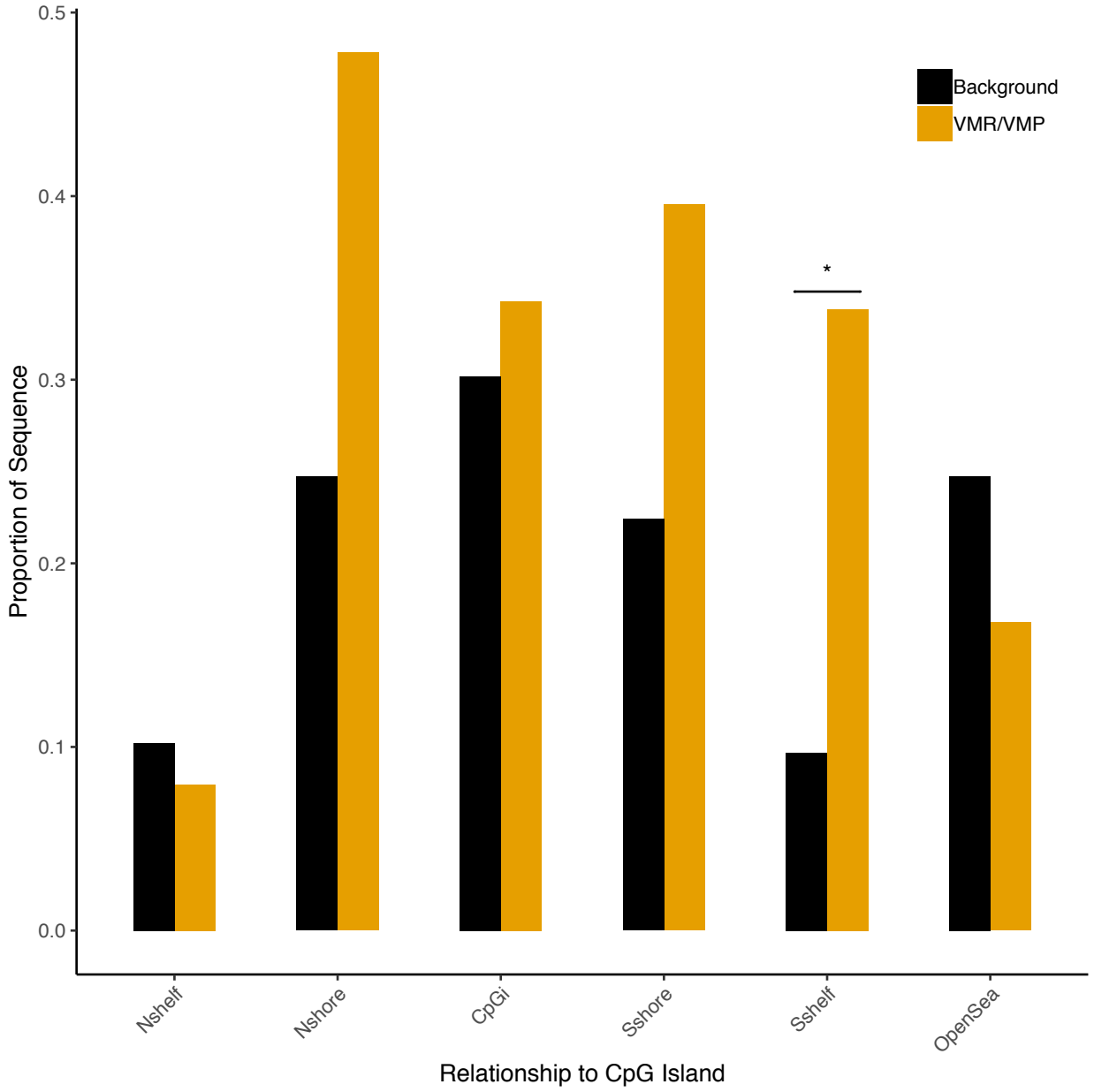
Supplemental Figure 1. Bivariate distribution of DNAm probes, where one twin pair is discordant for a history of MD and the other is concordant healthy. A twin member from the discordant pair and the concordant healthy pair is then crossed in the upper right and lower left panels, creating "non-MZ" pairs. This figure clearly shows that the bivariate distribution of DNAm marks differ between the concordant healthy and discordant twins, with the discordant pair showing greater variation between twins. The lower left panel, which represents an unrelated case-control, illustrates the substantial increase in variance, underscoring the utility of the co-twin control design to study environmental risk.



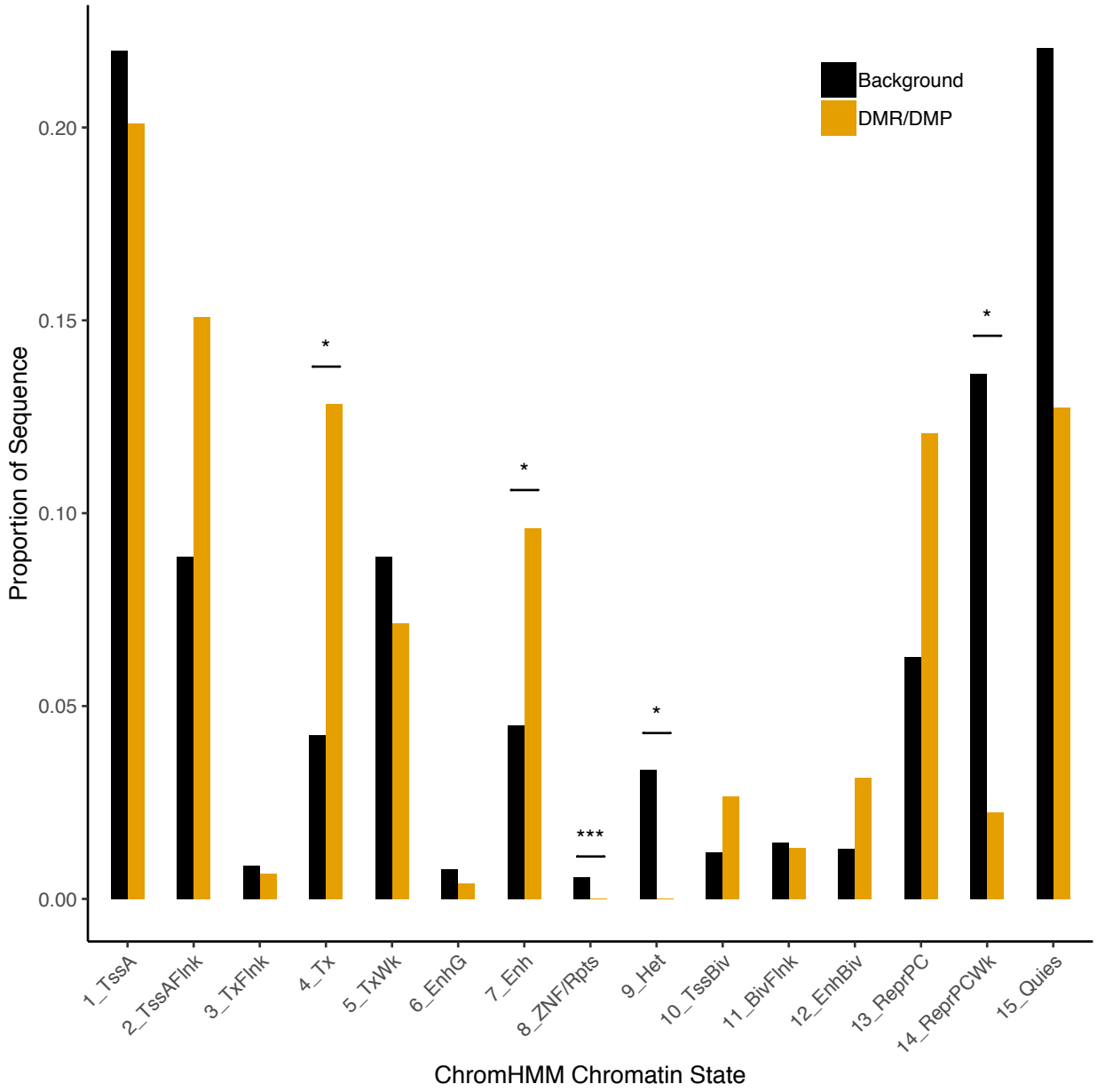
Supplemental Figure 2.



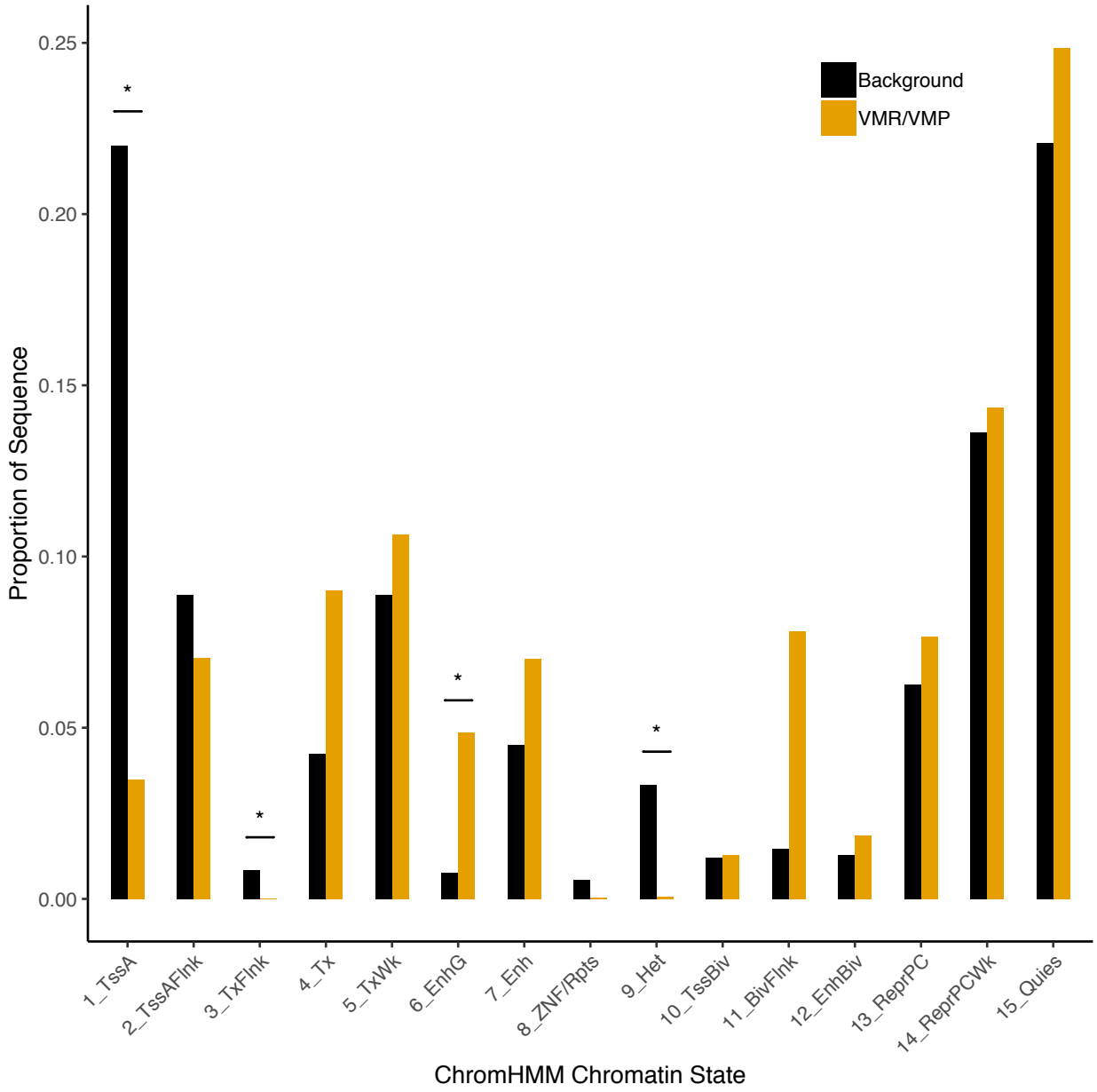
Supplemental Figure 3.



Supplemental Figure 4.



Supplemental Figure 5.



Supplemental Table 2. DMP significant gene/promoter hits identified in the MD affected versus MD unaffected contrast.

entrez	Symbol	dmp.gene	dmp.prom
1001	CDH3	NA	1
100126332	MIR943	NA	1
100129845	PCOLCE-AS1	NA	1
100302152	MIR548N	1	NA
100422962	MIR4281	NA	1
100506866	TTN-AS1	1	NA
100526737	RBM14-RBM4	1	NA
100528018	ARL2-SNX15	1	NA
100616144	MIR548AN	1	NA
100616190	MIR548O2	1	NA
100616266	MIR4733	NA	1
10458	BAIAP2	1	NA
10540	DCTN2	1	1
10555	AGPAT2	1	NA
10560	SLC19A2	1	1
10642	IGF2BP1	1	NA
10659	CELF2	1	1
10718	NRG3	1	NA
10981	RAB32	1	NA
11167	FSTL1	1	1
117248	GALNT15	1	NA
125061	AFMID	1	1
125488	TTC39C	1	1
126917	IFFO2	1	NA
133121	ENPP6	1	NA
135138	PACRG	1	NA
137902	PXDNL	1	1
1456	CSNK1G3	1	NA
151278	CCDC140	1	NA
153339	TMEM167A	NA	1
154043	CNKSR3	1	NA
161003	STOML3	NA	1
171425	CLYBL	1	NA
1745	DLX1	1	NA
1876	E2F6	NA	1
1877	E4F1	1	NA
1880	GPR183	NA	1

2059	EPS8	1	NA
2070	EYA4	1	1
2173	FABP7	1	NA
221061	FAM171A1	1	NA
2256	FGF11	NA	1
22821	RASA3	1	NA
22854	NTNG1	1	NA
22986	SORCS3	1	NA
23065	EMC1	1	NA
23108	RAP1GAP2	1	NA
23150	FRMD4B	1	NA
23344	ESYT1	NA	1
23484	LEPROTL1	NA	1
253314	EIF4E1B	NA	1
25766	PRPF40B	1	NA
25782	RAB3GAP2	1	NA
26034	IPCEF1	1	NA
26099	SZRD1	1	NA
27031	NPHP3	1	NA
27241	BBS9	NA	1
273	AMPH	1	NA
28232	SLCO3A1	NA	1
311	ANXA11	NA	1
337867	UBAC2	1	NA
348808	NPHP3-AS1	NA	1
3595	IL12RB2	NA	1
390598	SKOR1	1	NA
3983	ABLIM1	1	1
404663	LINC01194	1	NA
404665	CACTIN-AS1	NA	1
4088	SMAD3	1	NA
4299	AFF1	1	NA
439990	LINC00857	NA	1
440695	ETV3L	1	NA
4430	MYO1B	1	NA
463	ZFHX3	1	NA
4682	NUBP1	NA	1
4763	NF1	1	NA
50515	CHST11	NA	1
51154	MRTO4	NA	1

5118	PCOLCE	1	NA
51527	GSKIP	NA	1
51760	SYT17	NA	1
5368	PNOC	1	NA
54472	TOLLIP	1	NA
54715	RBFOX1	1	NA
55002	TMCO3	1	NA
55076	TMEM45A	1	NA
55102	ATG2B	1	NA
574408	MIR329-1	NA	1
574409	MIR329-2	NA	1
57451	TENM2	1	NA
57537	SORCS2	1	NA
57552	NCEH1	1	NA
5912	RAP2B	1	NA
5915	RARB	1	NA
5936	RBM4	NA	1
596	BCL2	1	1
64084	CLSTN2	1	NA
6416	MAP2K4	1	NA
6423	SFRP2	NA	1
64478	CSMD1	1	NA
647288	CTAGE11P	NA	1
6620	SNCB	1	NA
670	BPHL	1	NA
6915	TBXA2R	1	1
7083	TK1	NA	1
7273	TTN	1	NA
7441	VPREB1	NA	1
7469	NELFA	1	1
7518	XRCC4	1	1
752	FMNL1	1	NA
7520	XRCC5	1	1
768212	MIR758	1	1
7813	EVI5	1	NA
79570	NKAIN1	1	NA
79740	ZBBX	NA	1
80028	FBXL18	1	NA
80331	DNAJC5	1	NA
84717	HDGFL2	NA	1

84892	POMGNT2	1	1
84953	MICALCL	NA	1
8802	SUCLG1	1	1
8817	FGF18	NA	1
8874	ARHGEF7	1	NA
8924	HERC2	1	NA
91522	COL23A1	NA	1
92181	UBTD2	1	NA
9276	COPB2	NA	1
9496	TBX4	1	1
9671	WSCD2	NA	1
9727	RAB11FIP3	1	NA
9844	ELMO1	1	1
995	CDC25C	1	1

Supplemental Table 3. VMP significant gene/promoter hits identified in the MD affected versus MD unaffected contrast.

entrez	symbol	vmp.gene	vmp.prom
100128264	HTR5A-AS1	1	NA
100128568	LOC100128568	1	NA
100131213	ZNF503-AS2	1	1
100132406	NBPF10	1	NA
100132677	BSN-DT	1	1
100133612	LINC01134	1	NA
100189589	DCTN1-AS1	1	NA
100271715	ARHGEF33	1	NA
100288142	NBPF20	1	NA
100302197	MIR1306	NA	1
100302224	MIR2110	NA	1
100316904	SAP25	1	NA
1004	CDH6	1	NA
100500860	MIR3618	NA	1
100526835	FPGT-TNNI3K	1	NA
100529261	CHURC1-FNTB	1	NA
100532732	MSH5-SAPCD1	NA	1
100533179	UBE2F-SCLY	NA	1
100533184	ARHGAP19-SLIT1	1	NA
100534589	HOXA10-HOXA9	1	NA
100616250	MIR3960	1	1
10142	AKAP9	NA	1
10172	ZNF256	NA	1
10194	TSHZ1	NA	1
10195	ALG3	NA	1
102	ADAM10	1	NA
10201	NME6	1	NA
10207	PATJ	1	NA
10217	CTDSPL	1	NA
1025	CDK9	NA	1
10351	ABCA8	1	NA
10480	EIF3M	1	1
10499	NCOA2	1	NA
10500	SEMA6C	NA	1
10516	FBLN5	1	1
10522	DEAF1	1	NA
10557	RPP38	1	NA

10667	FARS2	1	NA
10693	CCT6B	1	NA
10938	EHD1	1	1
11100	HNRNPUL1	1	1
11122	PTPRT	1	NA
11129	CLASRP	1	NA
11176	BAZ2A	1	1
11211	FZD10	NA	1
11277	TREX1	1	1
113091	PTH2	1	NA
113791	PIK3IP1	1	1
1141	CHRNB2	1	NA
114815	SORCS1	1	NA
114818	KLHL29	1	1
114827	FHAD1	1	NA
115817	DHRS1	NA	1
116729	PPP1R27	1	NA
116987	AGAP1	1	NA
118424	UBE2J2	1	NA
1186	CLCN7	1	NA
1213	CLTC	NA	1
124093	CCDC78	NA	1
124401	ANKS3	1	NA
124637	CYB5D1	1	NA
1262	CNGA4	NA	1
126374	WTIP	NA	1
126868	MAB21L3	NA	1
1284	COL4A2	1	NA
129293	TRABD2A	NA	1
129787	TMEM18	1	1
131566	DCBLD2	1	1
131965	METTL6	1	1
132	ADK	NA	1
132851	SPATA4	1	1
133015	PACRGL	1	NA
134121	C5orf49	1	NA
134145	FAM173B	NA	1
1364	CLDN4	NA	1
1369	CPN1	1	NA
1395	CRHR2	1	NA

140739	UBE2F	NA	1
140828	LINC00261	1	NA
144481	SOCS2-AS1	NA	1
1446	CSN1S1	1	1
144699	FBXL14	NA	1
145376	PPP1R36	1	1
146050	ZSCAN29	NA	1
146712	B3GNTL1	1	NA
148014	TTC9B	NA	1
148362	BROX	1	1
148398	SAMD11	1	1
148741	ANKRD35	1	NA
148753	FAM163A	1	1
149233	IL23R	1	NA
150538	SATB2-AS1	1	NA
152992	TRMT44	1	1
153090	DAB2IP	1	NA
153768	PRELID2	NA	1
1592	CYP26A1	1	1
1613	DAPK3	1	1
161424	NOP9	NA	1
162461	TMEM92	NA	1
1639	DCTN1	1	NA
168090	C6orf118	1	NA
1838	DTNB	1	NA
1850	DUSP8	NA	1
1854	DUT	NA	1
1890	TYMP	1	1
199990	FAAP20	1	NA
2000	ELF4	NA	1
2003	ELK2AP	1	NA
200424	TET3	1	NA
201514	ZNF584	1	1
2035	EPB41	1	NA
2048	EPHB2	1	NA
204851	HIPK1	1	1
210	ALAD	NA	1
2104	ESRRG	1	NA
2120	ETV6	NA	1
2130	EWSR1	1	1

219670	ENKUR	1	1
220064	ORAOV1	NA	1
220359	TIGD3	1	1
221060	C10orf111	NA	1
221150	SKA3	1	NA
221178	SPATA13	1	NA
222643	UNC5CL	1	1
225	ABCD2	1	1
2253	FGF8	1	NA
22801	ITGA11	1	NA
22848	AAK1	1	NA
22889	KHDC4	1	1
22906	TRAK1	1	1
22928	SEPHS2	1	1
22948	CCT5	1	NA
22950	SLC4A1AP	1	1
22996	TTC39A	1	NA
23001	WDFY3	1	NA
23033	DOPEY1	NA	1
23126	POGZ	1	NA
23143	LRCH1	1	NA
23179	RGL1	1	NA
23191	CYFIP1	NA	1
23208	SYT11	1	NA
2326	FMO1	1	NA
23266	ADGRL2	1	NA
2327	FMO2	1	NA
23294	ANKS1A	1	NA
23303	KIF13B	1	NA
23312	DMXL2	1	NA
23314	SATB2	NA	1
23329	TBC1D30	NA	1
23352	UBR4	NA	1
23353	SUN1	1	NA
23378	RRP8	NA	1
23406	COTL1	1	NA
2342	FNTB	NA	1
23432	GPR161	1	1
23492	CBX7	NA	1
23554	TSPAN12	NA	1

23705	CADM1	1	NA
238	ALK	1	NA
24142	NAT6	1	1
254394	MCM9	1	1
2548	GAA	1	NA
256356	GK5	NA	1
25807	RHBDD3	NA	1
259217	HSPA12A	1	NA
25939	SAMHD1	NA	1
25941	TPGS2	1	NA
25956	SEC31B	1	NA
26057	ANKRD17	1	NA
26092	TOR1AIP1	1	NA
26119	LDLRAP1	1	NA
26136	TES	1	NA
26207	PITPNC1	1	NA
26520	TIMM9	1	NA
26801	SNORD48	NA	1
26828	RNU5F-1	1	NA
26974	ZNF285	NA	1
26985	AP3M1	NA	1
26995	TRUB2	NA	1
271	AMPD2	1	1
27102	EIF2AK1	1	NA
27229	TUBGCP4	NA	1
27237	ARHGEF16	1	NA
27243	CHMP2A	NA	1
27255	CNTN6	1	NA
27327	TNRC6A	1	1
27346	TMEM97	NA	1
2774	GNAL	1	NA
2788	GNG7	1	NA
283102	KRT8P41	NA	1
283554	GPR137C	NA	1
283856	LOC283856	1	NA
284069	FAM171A2	NA	1
284098	PIGW	1	NA
284273	ZADH2	1	NA
284352	PPP1R37	NA	1
285116	AHCTF1P1	1	NA

28514	DLL1	1	NA
285590	SH3PXD2B	1	NA
285593	LOC285593	1	NA
2895	GRID2	1	NA
28974	C19orf53	1	NA
2900	GRIK4	1	NA
2918	GRM8	1	1
29993	PACSIN1	NA	1
3084	NRG1	1	NA
3092	HIP1	1	NA
3106	HLA-B	NA	1
3148	HMGB2	NA	1
317751	MESTIT1	NA	1
3185	HNRNPF	1	1
3205	HOXA9	1	NA
321	APBA2	1	NA
330	BIRC3	NA	1
334	APLP2	NA	1
3361	HTR5A	1	NA
337968	KRTAP6-3	NA	1
338651	KRTAP5-AS1	1	1
339	APOBEC1	1	NA
339416	ANKRD45	1	NA
342977	NANOS3	1	NA
343099	CCDC18	1	NA
347689	SOX2-OT	1	NA
348262	MCRIP1	NA	1
350383	GPR142	NA	1
3611	ILK	1	1
3678	ITGA5	1	1
3707	ITPKB	1	NA
3720	JARID2	1	1
3748	KCNC3	1	1
374928	ZNF773	NA	1
375033	PEAR1	1	NA
375196	LOC375196	NA	1
375387	NRROS	1	NA
376267	RAB15	NA	1
376497	SLC27A1	1	NA
3784	KCNQ1	1	NA

38	ACAT1	1	1
388394	RPRML	NA	1
389337	ARHGEF37	NA	1
390010	NKX1-2	1	NA
3930	LBR	NA	1
3932	LCK	1	NA
399665	FAM102A	1	1
4004	LMO1	1	1
401145	CCSER1	1	1
401898	ZNF833P	1	NA
404093	CUEDC1	NA	1
407015	MIR26A1	NA	1
4148	MATN3	NA	1
4176	MCM7	1	NA
4223	MEOX2	1	NA
4232	MEST	1	1
4240	MFGE8	1	NA
4297	KMT2A	NA	1
4329	ALDH6A1	NA	1
440119	FZD10-DT	1	NA
4430	MYO1B	1	NA
475	ATOX1	NA	1
4750	NEK1	1	1
4782	NFIC	1	1
4901	NRL	1	NA
4919	ROR1	NA	1
494326	MIR377	NA	1
49854	ZBTB21	NA	1
5045	FURIN	1	NA
50804	MYEF2	1	1
5081	PAX7	NA	1
50836	TAS2R8	NA	1
50854	C6orf48	NA	1
50940	PDE11A	1	1
50999	TMED5	NA	1
51019	WASHC3	NA	1
51069	MRPL2	1	1
51086	TNNI3K	1	NA
51117	COQ4	1	NA
51160	VPS28	NA	1

51161	C3orf18	NA	1
51188	SS18L2	NA	1
51191	HERC5	1	1
51334	PRR16	1	NA
51373	MRPS17	NA	1
51409	HEMK1	1	NA
51430	SUCO	NA	1
5158	PDE6B	1	NA
5159	PDGFRB	NA	1
51593	SRRT	NA	1
51701	NLK	1	NA
5217	PFN2	1	1
523	ATP6V1A	1	1
5295	PIK3R1	1	1
53343	NUDT9	NA	1
5356	PLRG1	NA	1
5359	PLSCR1	NA	1
5365	PLXNB3	1	NA
54434	SSH1	1	1
54487	DGCR8	1	1
54502	RBM47	1	NA
54503	ZDHHC13	NA	1
54758	KLHDC4	1	NA
54795	TRPM4	1	NA
54811	ZNF562	NA	1
54897	CASZ1	1	NA
54940	OCIAD1	NA	1
55033	FKBP14	NA	1
55036	CCDC40	1	NA
55062	WIPI1	1	NA
55068	ENOX1	1	NA
55088	CCDC186	NA	1
55114	ARHGAP17	1	1
55127	HEATR1	NA	1
55159	RFWD3	NA	1
55160	ARHGEF10L	1	NA
55276	PGM2	1	NA
55283	MCOLN3	NA	1
553103	MIR3936HG	1	1
55315	SLC29A3	NA	1

55336	FBXL8	1	1
55361	PI4K2A	1	1
55532	SLC30A10	1	NA
55611	OTUB1	NA	1
5562	PRKAA1	NA	1
55635	DEPDC1	1	1
5575	PRKAR1B	1	NA
55758	RCOR3	1	NA
55790	CSGALNACT1	NA	1
55819	RNF130	1	1
55854	ZC3H15	NA	1
55902	ACSS2	1	1
55924	FAM212B	1	NA
56005	MYDGF	1	1
56134	PCDHAC2	1	1
56135	PCDHAC1	1	NA
56136	PCDHA13	1	NA
56137	PCDHA12	1	NA
56138	PCDHA11	1	NA
56139	PCDHA10	1	NA
56140	PCDHA8	1	NA
56141	PCDHA7	1	NA
56142	PCDHA6	1	NA
56143	PCDHA5	1	NA
56144	PCDHA4	1	NA
56145	PCDHA3	1	NA
56146	PCDHA2	1	NA
56147	PCDHA1	1	NA
56474	CTPS2	NA	1
56704	JPH1	NA	1
56853	CELF4	1	NA
5687	PSMA6	NA	1
56886	UGGT1	1	1
56937	PMEPA1	1	NA
56950	SMYD2	1	NA
57030	SLC17A7	1	1
57132	CHMP1B	1	1
57150	SMIM8	1	NA
57209	ZNF248	1	NA
57337	SENP7	NA	1

57348	TTYH1	1	NA
57379	AICDA	1	NA
574481	MIR521-2	NA	1
57474	ZNF490	1	NA
57513	CASKIN2	1	NA
57536	KIAA1328	NA	1
57544	TXNDC16	NA	1
57553	MICAL3	1	NA
57556	SEMA6A	NA	1
57569	ARHGAP20	1	1
57575	PCDH10	NA	1
57582	KCNT1	1	NA
57620	STIM2	1	NA
57654	UVSSA	1	NA
57666	FBRSL1	1	NA
57693	ZNF317	1	NA
57758	SCUBE2	NA	1
5789	PTPRD	1	NA
5793	PTPRG	1	NA
5816	PVALB	1	1
583	BBS2	NA	1
5832	ALDH18A1	1	1
5977	DPF2	NA	1
5990	RFX2	1	NA
5993	RFX5	NA	1
6095	RORA	1	NA
619343	NA	1	NA
6262	RYR2	1	NA
63027	SLC22A23	NA	1
6329	SCN4A	NA	1
639	PRDM1	NA	1
63976	PRDM16	1	NA
642366	LOC642366	1	NA
64328	XPO4	1	1
64377	CHST8	1	NA
643803	KRTAP24-1	1	NA
64478	CSMD1	1	NA
646903	LOC646903	NA	1
64746	ACBD3	NA	1
64786	TBC1D15	NA	1

64853	AIDA	NA	1
64881	PCDH20	1	NA
64919	BCL11B	NA	1
6497	SKI	1	NA
65012	SLC26A10	NA	1
65108	MARCKSL1	NA	1
6514	SLC2A2	NA	1
652276	LOC652276	NA	1
6584	SLC22A5	NA	1
6585	SLIT1	NA	1
6588	SLN	NA	1
65985	AACS	1	1
65996	CENPBD1P1	1	NA
66005	CHID1	NA	1
6604	SMARCD3	1	NA
6631	SNRPC	1	1
6648	SOD2	1	NA
6687	SPG7	1	1
6788	STK3	1	NA
6865	TACR2	1	NA
6867	TACC1	1	1
6892	TAPBP	NA	1
6904	TBCD	1	1
693220	MIR635	NA	1
699	BUB1	1	NA
7003	TEAD1	1	NA
7020	TFAP2A	1	NA
7025	NR2F1	1	1
7068	THRB	1	NA
7077	TIMP2	1	NA
715	C1R	1	NA
7164	TPD52L1	1	1
7226	TRPM2	1	1
7262	PHLDA2	NA	1
7277	TUBA4A	1	NA
728192	LINC00460	1	NA
728912	NA	1	NA
728932	NA	1	NA
728939	NA	1	NA
7323	UBE2D3	1	NA

7342	UBP1	1	1
7444	VRK2	1	NA
747	DAGLA	NA	1
7520	XRCC5	1	NA
768	CA9	1	NA
773	CACNA1A	1	NA
7771	ZNF112	NA	1
78988	MRPL57	NA	1
79018	GID4	1	NA
79087	ALG12	NA	1
7915	ALDH5A1	1	NA
79174	CRELD2	1	1
79696	ZC2HC1C	NA	1
79703	C11orf80	1	NA
79745	CLIP4	1	1
79791	FBXO31	1	NA
79812	MMRN2	1	NA
79832	QSER1	1	NA
79890	RIN3	NA	1
79896	THNSL1	NA	1
79913	ACTR5	1	NA
79918	SETD6	1	NA
79919	C2orf54	1	NA
79960	JADE1	1	NA
79977	GRHL2	1	NA
80005	DOCK5	1	NA
8001	GLRA3	NA	1
80086	TUBA4B	NA	1
80125	CCDC33	1	NA
80176	SPSB1	1	NA
8021	NUP214	1	NA
80218	NAA50	NA	1
8022	LHX3	1	NA
80312	TET1	1	NA
80728	ARHGAP39	NA	1
80760	ITIH5	1	1
81029	WNT5B	1	NA
81532	MOB2	1	NA
8174	MADCAM1	1	NA
8310	ACOX3	1	1

83538	TTC25	NA	1
83659	TEKT1	NA	1
8369	HIST1H4G	1	1
8372	HYAL3	1	1
83943	IMMP2L	1	NA
84033	OBSCN	1	NA
84056	KATNAL1	1	NA
84067	FAM160A2	NA	1
84068	SLC10A7	NA	1
84080	ENKD1	1	NA
84126	ATRIP	1	1
84166	NLRC5	1	NA
84236	RHBDD1	1	NA
84264	HAGHL	1	NA
84272	YIPF4	NA	1
84284	NTPCR	NA	1
84302	TMEM246	NA	1
84316	NAA38	NA	1
84435	ADGRA1	NA	1
8460	TPST1	NA	1
8470	SORBS2	1	1
84725	PLEKHA8	1	1
84856	LINC00839	1	1
84858	ZNF503	NA	1
84929	FIBCD1	1	NA
8495	PPFIBP2	1	NA
85001	MGC16275	1	NA
8532	CPZ	1	NA
8536	CAMK1	NA	1
85403	EAF1	NA	1
857	CAV1	1	1
8601	RGS20	1	NA
864	RUNX3	1	NA
8661	EIF3A	1	NA
8664	EIF3D	NA	1
8677	STX10	NA	1
8717	TRADD	NA	1
8736	MYOM1	1	NA
8775	NAPA	NA	1
8835	SOCS2	1	1

8927	BSN	NA	1
894	CCND2	NA	1
8943	AP3D1	NA	1
89953	KLC4	1	1
90025	UBE3D	1	1
902	CCNH	NA	1
90427	BMF	1	NA
90525	SHF	1	NA
9117	SEC22C	1	NA
91408	BTF3L4	1	NA
9141	PDCD5	NA	1
91452	ACBD5	1	1
91603	ZNF830	NA	1
91748	ELMSAN1	1	1
91750	LIN52	NA	1
91775	NXPE3	1	NA
92345	NAF1	1	1
9249	DHRS3	1	NA
92521	SPECC1	1	1
928	CD9	1	NA
9364	RAB28	NA	1
93974	ATP5IF1	1	1
94015	TTYH2	NA	1
9423	NTN1	1	NA
9440	MED17	NA	1
9443	MED7	NA	1
9538	EI24	1	NA
9570	GOSR2	1	NA
9578	CDC42BPB	1	NA
9592	IER2	1	1
960	CD44	1	NA
9644	SH3PXD2A	1	NA
9692	KIAA0391	1	NA
9703	KIAA0100	1	1
9718	ECE2	1	NA
9747	TCAF1	1	1
9752	PCDHA9	1	NA
9759	HDAC4	1	1
9761	MLEC	1	NA
9913	SUPT7L	1	1

9924	PAN2	1	NA
9941	EXOG	1	1
9986	RCE1	NA	1

Supplemental Table 4. Overlap of MD related DMPs and DMRs with PGC GWAS loci.

Chromosome	DNAm.start	DNAm.end	PGC.start	PGC.end
chr6	29521506	29521506	27738000	32848000
chr6	30031455	30031455	27738000	32848000
chr6	30519905	30521619	27738000	32848000
chr6	31828260	31830030	27738000	32848000
chr6	32797253	32798887	27738000	32848000
chr10	106401479	106401479	106397000	106904000

Supplemental Table 5. Overlap of MD related VMPs and VMRs with PGC GWAS loci.

Chromosome	DNAm.start	DNAm.end	PGC.start	PGC.end
chr2	58265673	58265673	57765000	58485000
chr5	164658718	164658718	164440000	164789000
chr6	28706472	28706472	27738000	32848000
chr6	28890673	28890673	27738000	32848000
chr6	31324972	31324972	27738000	32848000
chr6	31707922	31707922	27738000	32848000
chr6	31802397	31802397	27738000	32848000
chr6	32049516	32049825	27738000	32848000
chr6	32055738	32055738	27738000	32848000
chr6	32789916	32789916	27738000	32848000
chr6	32818212	32818212	27738000	32848000
chr6	32847830	32847830	27738000	32848000