Archival Report

Biological Psychiatry

DNA Modification Study of Major Depressive Disorder: Beyond Locus-by-Locus Comparisons

Gabriel Oh, Sun-Chong Wang, Mrinal Pal, Zheng Fei Chen, Tarang Khare, Mamoru Tochigi, Catherine Ng, Yeqing A. Yang, Andrew Kwan, Zachary A. Kaminsky, Jonathan Mill, Cerisse Gunasinghe, Jennifer L. Tackett, Irving I. Gottesman, Gonneke Willemsen, Eco J.C. de Geus, Jacqueline M. Vink, P. Eline Slagboom, Naomi R. Wray, Andrew C. Heath, Grant W. Montgomery, Gustavo Turecki, Nicholas G. Martin, Dorret I. Boomsma, Peter McGuffin, Rafal Kustra, and Art Petronis

ABSTRACT

BACKGROUND: Major depressive disorder (MDD) exhibits numerous clinical and molecular features that are consistent with putative epigenetic misregulation. Despite growing interest in epigenetic studies of psychiatric diseases, the methodologies guiding such studies have not been well defined.

METHODS: We performed DNA modification analysis in white blood cells from monozygotic twins discordant for MDD, in brain prefrontal cortex, and germline (sperm) samples from affected individuals and control subjects (total $N = 304$) using 8.1K CpG island microarrays and fine mapping. In addition to the traditional locus-by-locus comparisons, we explored the potential of new analytical approaches in epigenomic studies.

RESULTS: In the microarray experiment, we detected a number of nominally significant DNA modification differences in MDD and validated selected targets using bisulfite pyrosequencing. Some MDD epigenetic changes, however, overlapped across brain, blood, and sperm more often than expected by chance. We also demonstrated that stratification for disease severity and age may increase the statistical power of epimutation detection. Finally, a series of new analytical approaches, such as DNA modification networks and machine-learning algorithms using binary and quantitative depression phenotypes, provided additional insights on the epigenetic contributions to MDD.

CONCLUSIONS: Mapping epigenetic differences in MDD (and other psychiatric diseases) is a complex task. However, combining traditional and innovative analytical strategies may lead to identification of disease-specific etiopathogenic epimutations.

Keywords: DNA modification, Epigenetic outliers, Epigenetics, Heteroscedasticity, Major depressive disorder, Molecular networks

http://dx.doi.org/10.1016/j.biopsych.2014.06.016

Major depressive disorder (MDD) is a psychiatric disease characterized by an all-encompassing low mood accompanied by low self-esteem, loss of interest or pleasure in normally enjoyable activities, and a variety of other associated symptoms (1). MDD affects one in seven individuals (2) and has been projected to become the second leading cause of disability worldwide by 2020 (3).

A meta-analysis of twin studies on MDD estimated heritability at 37% (4), which is consistent with a recent large epidemiologic study (5). This significant heritability provided a basis for molecular genetic studies; however, identification of specific MDD risk genes has proven difficult. A recent genome-wide association study with >18,000 subjects in the discovery phase did not detect any genome-wide significant single nucleotide polymorphisms (SNPs). The study followed up on the top 554 SNPs ($p < .0001$) in an independent set of >57,000 subjects but failed to replicate any of the SNPs at genome-wide significance (6).

MDD exhibits numerous non-Mendelian features that can be reviewed from an epigenetic perspective (7). Such features include partial heritability, discordance of monozygotic (MZ) twins, sexual dimorphism (8,9), disease onset following major hormonal changes (e.g., postpartum depression) (10), and fluctuating course of disease (11). Epigenetics refers to the regulation of various genomic functions that are controlled by heritable but reversible chemical modifications of DNA and histones (12). Environmental factors such as stress, diet, and drugs can alter the epigenetic profile (13,14). Even in the absence of environmental exposures, stochastic epigenetic changes may influence phenotypic outcomes (15). Furthermore, there is increasing evidence that epigenetic factors, in addition to DNA sequences, account for heritability (16,17). In short, we postulate that inherited and acquired epigenetic misregulation may play an etiological role in MDD (7).

In this study, we attempted to identify MDD specific epigenetic changes using a series of experimental and analytical

 ISSN: 0006-3223

approaches, from traditional locus-by-locus comparisons to new systems biology-based strategies, such as epigenomic networks and machine-learning based classification.

METHODS AND MATERIALS

Samples

Tissue samples were collected from individuals diagnosed with MDD and from matched control subjects. Inclusion criteria involved patients between the ages of 18 and 75 diagnosed with MDD according to DSM-IV criteria. Individuals with a prior history of other mental illnesses, addiction and substance abuse, or a family history (first-degree relatives) of schizophrenia were excluded from the study. The 100 discordant MZ twin samples consisted of peripheral blood DNA from 40 pairs of MZ twins from Australia, 46 pairs from The Netherlands, and 14 pairs from the United Kingdom (for detailed description, see Supplement 1). Seventy-one prefrontal cortex samples were received from the Stanley Medical Research Institute (SMRI) and Quebec Suicide Brain Bank (QSBB). Thirty-three sperm samples from bipolar disorder patients, a disease that may be etiologically related to MDD (18,19), and control subjects were obtained from an ongoing study at the Centre for Addiction and Mental Health (Toronto, Ontario, Canada). More information on the samples can be found in Table S1 in Supplement 1.

Microarray Experiment

The unmodified DNA fraction was enriched using modified cytosine (^{mod}C)-sensitive restriction enzymes, which collectively interrogate 5-methylcytosine and 5-hydroxymethylcytosine (20) (it is assumed that 5-carboxylcytosine and 5-formylcytosine are rare and unlikely to significantly contribute to the estimates of the modified/unmodified cytosines). Three aliquots of 250 ng of genomic DNA were digested individually with HpaII, HinP1I, and HpyCH4IV and pooled together after digestion was completed. For the twin samples, 500 ng of genomic DNA was digested using only HpaII. All other steps were identical to those described in our published protocol (21). The microarray experiment was conducted using a common reference pool design. The enriched polymerase chain reaction products were labeled with Cy3 for the reference and Cy5 for the sample hybridized onto 8.1K human CpG island microarrays (22,23). A detailed description of the bioinformatic methods can be found in Supplement 1.

Bisulfite modification and pyrosequencing-based fine mapping of ^{mod}C was performed using a standard protocol (24). The primers for the bisulfite polymerase chain reaction were designed using either the MethPrimer (25) or the Pyrosequencing Assay Design Software v1.0.6 (Qiagen, Valencia, California) (Table S2 in Supplement 1). For pyrosequencing, Gold Q96 Reagents and Pyromark Q24 were used (Qiagen).

Ethics Statement

Centre for Addiction and Mental Health Research Ethics Board granted approval to protocol # 024/2005-01 entitled "Molecular epigenetic studies of major depression." All experiments were performed in accordance with relevant guidelines and regulations.

RESULTS

Locus-Specific Analysis of DNA Modification in the Brain, White Blood Cells, and the Germline

In the human brain samples from the SMRI, a locus-by-locus comparison between MDD or MDD with psychosis (MDD + Psy) and control subjects using analysis of variance revealed 40 differentially modified loci (nominal $p = 4 \times 10^{-5} - .01$; Table S4 in Supplement 1); 22 loci showed differential modification between MDD and control subjects, and 18 loci showed differential modification between MDD + Psy and control subjects (Tukey's honestly significant difference, $p < .05$). Eight loci were differentially modified for both MDD and MDD + Psy compared with control subjects. One gene, *FOXD3,* was previously implicated in MDD (26). The analysis of the brain samples from the QSBB revealed 35 loci with differential modification (nominal $p = 5 \times 10^{-4} - .01$). In white blood cells (WBCs) from MZ twins discordant for MDD, we identified 44 loci with nominal $p = 9 \times 10^{-5} - .01$. Lastly, in the sperm samples from individuals affected with bipolar disorder and control subjects, we found 34 loci (nominal $p = 6 \times 10^{-4} - .01$), one of which had already been implicated in bipolar disorder (*SMAD3*) (27).

We did not find significant overlaps between any of the samples tested above. However, we found a statistically significant number of overlapping loci between our study and a previously published epigenome-wide study using the same SMRI brain samples but different enrichment technique and platform (at nominal $p < .05$ for both studies) (28). We performed permutation analysis and found that our microarray probes that were either directly on or nearest neighbors (median distance = 12 kb) to the gene of interest were overrepresented than by chance ($n = 14$; permuted $p = .04$; Table S5 in Supplement 1). Even when the parameters were made more stringent to only include microarray probes that were either directly on or within a short distance away (<10 kb or <5 kb) from the gene of interest, we still found a significant number of overlaps between the two studies ($n = 12$ for both; $p = .03$ and $p = .02$, respectively).

None of the detected loci survived correction for multiple testing, although 13 loci with nominal $p < .05$ overlapped with either the SMRI or the QSBB brain samples and the WBC samples of the MDD twins. Among the 13 loci, probes for *LRRC41* and *LIN28A* contained regulatory sequences, nuclear factor-κB transcription factor binding site, and a predicted insulator CTCF binding site, with ^{mod}C sensitive sites (Figure 1) (29). These two loci, plus three different types of repetitive elements (*LINE-1*, *NBL-2*, and *D4Z4*) as proxies for global modification changes (30), were finely mapped using bisulfite pyrosequencing. A total of 29 CpG sites (11, 4, 3, 6, and five CpG sites for *LRRC41*, *LIN28A*, *LINE-1*, *NBL-2*, and *D4Z4*, respectively) were interrogated from the two unique DNA loci and three repetitive DNA elements.

Bisulfite pyrosequencing revealed that ^{mod}C density at *LIN28A* was different in the SMRI MDD + Psy samples compared with control subjects (Mann-Whitney test, $p = .01$). While the pooled MDD samples (MDD and MDD + Psy) also showed significant differences ($p = .01$), MDD alone versus control subjects did not reach significance ($p = .08$), and the same was detected in the QSBB samples ($p = .77$). We also

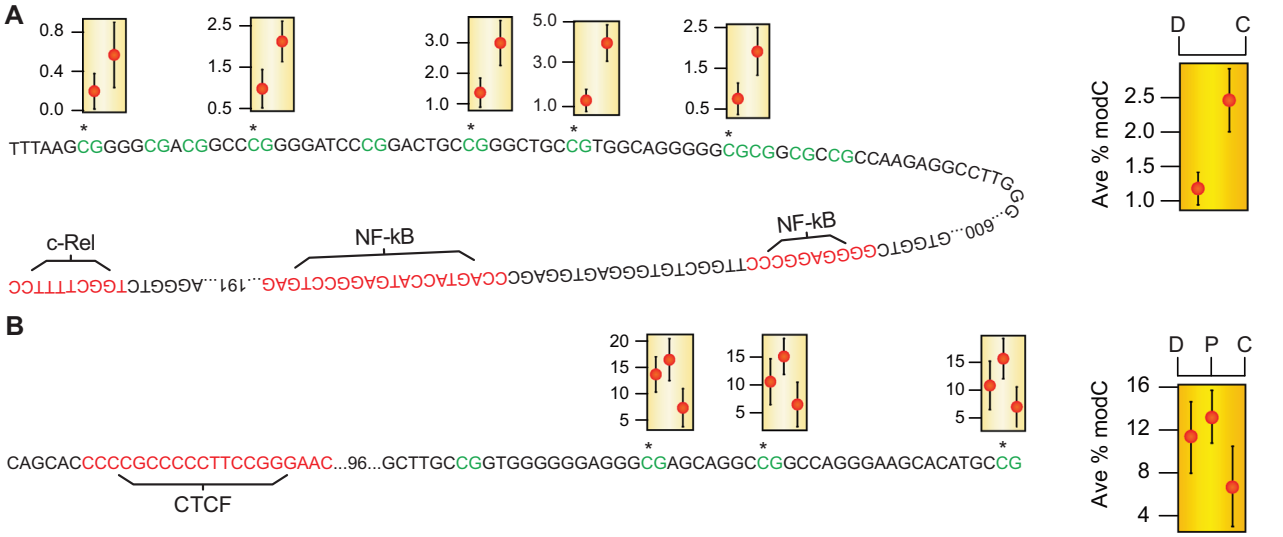

Figure 1. Location of fine mapping by bisulfite pyrosequencing and transcription factor binding sites for *LRRC41* and *LIN28A* in the prefrontal cortex of postmortem brains. Both the *LRRC41* **(A)** and *LIN28A* **(B)** loci contain regulatory sites within the probe sequence (red font). The green CpG sites indicate regions that were finely mapped using bisulfite pyrosequencing. Mean *LRRC41* and *LIN28A* modified cytosine (^{mod}C) densities were found to be differentially modified between major depressive disorder (MDD) patients and control subjects in the Quebec Suicide Brain Bank and Stanley Medical Research Institute samples, respectively. *LRRC41* in the Quebec Suicide Brain Bank samples showed a mean modification difference ± SD (Δ^{mod}C) of 1.3% ± 1.6% between control subjects and MDD. *LIN26A* in the Stanley Medical Research Institute control subjects compared with depression with psychosis showed Δ^{mod}C of 6.8% ± 10.8%, while control subjects vs. MDD alone was 4.3% ± 12.1%. *Specific CpG positions that were found to be significantly different between the affected and the unaffected individuals. The dark yellow boxes on the right show the average modification densities (Ave % modC) across all sites. C, control subjects; D, depression (MDD); P, depression with psychosis.

observed significant differences in *LRRC41* in the QSBB samples ($p = .004$) but not in the SMRI brain samples (Kruskal-Wallis test, $p = .10$). Age, sex, and antipsychotic use did not show an association with the ^{mod}C status of *LIN28A* or *LRRC41*. Consistent with other psychiatric diseases (31,32), the average density of ^{mod}C was similar in MDD patients and control subjects across the three repetitive DNA elements tested (Table S6 in Supplement 1).

Next, we attempted to understand why the conventional mean difference-based analysis uncovered only minor epigenetic changes in MDD and investigated two confounding factors: putative MDD heterogeneity and age effects.

Effects of the Degree of MZ Twin Discordance and Age-Dependent Heteroscedasticity of DNA Modification

Despite the obvious advantages of discordant MZ twin design in epigenomic studies, the unaffected co-twins are at a higher risk for MDD than the general population (33,34); therefore, they may carry some epigenetic risk factors, which reduces the power of detection of disease-specific epigenetic differences. To test this hypothesis, we analyzed subgroups of MZ twins with differing degrees of discordance for MDD. We utilized twin discordance information derived from the personality questionnaires, reported number of episodes, and interviews of the MZ twins to separate the most discordant from least discordant (35). The severity of disease discordance between the MZ twins indeed played a role; when we performed *t* tests using the 10 most discordant MZ pairs, we found 165 significant loci with nominal $p < .05$ but only 81 in the 10 least discordant pairs (Fisher's exact test, $p = 3.3 \times 10^{-8}$). The two middle groups of discordant twins showed an intermediate number of statistically significant differences (Figure 2A).

Another factor that may have compromised identification of significant epigenetic differences was age-dependent inconsistency in the variance of ^{mod}C. This phenomenon, generally known as heteroscedasticity, occurs when subsets of the samples have different degrees of variability. In the Australian MDD twins, we found 489 loci that exhibited changing variance with age in both the affected and the unaffected co-twins (Harrison-McCabe test, false discovery rate, $q < .05$) (Figure 2B). The majority (76%) of the heteroscedastic loci showed increasing ^{mod}C variance with age.

To verify if this phenomenon had an impact on statistical power, we performed *t* tests only on the heteroscedastic loci after separating the dataset into two groups by age: the younger half (32.1 ± 4.6 years; mean age ± SD) and the older half (50.3 ± 8.5 years). We found more differing loci between the affected and the unaffected twins in the younger group compared with the older group (20 and 9, respectively; Fisher's exact test, $p = .004$), indicating that age-dependent heteroscedasticity reduces statistical power (Figure 2C).

MDD DNA Modification Differences Detected in More Than One Tissue

To identify MDD epigenetic features that were common across tissues, we analyzed the largest ^{mod}C differences (>1.64 SD from the mean) in the brain, WBC, and sperm of diseased individuals compared with control subjects. The number of loci

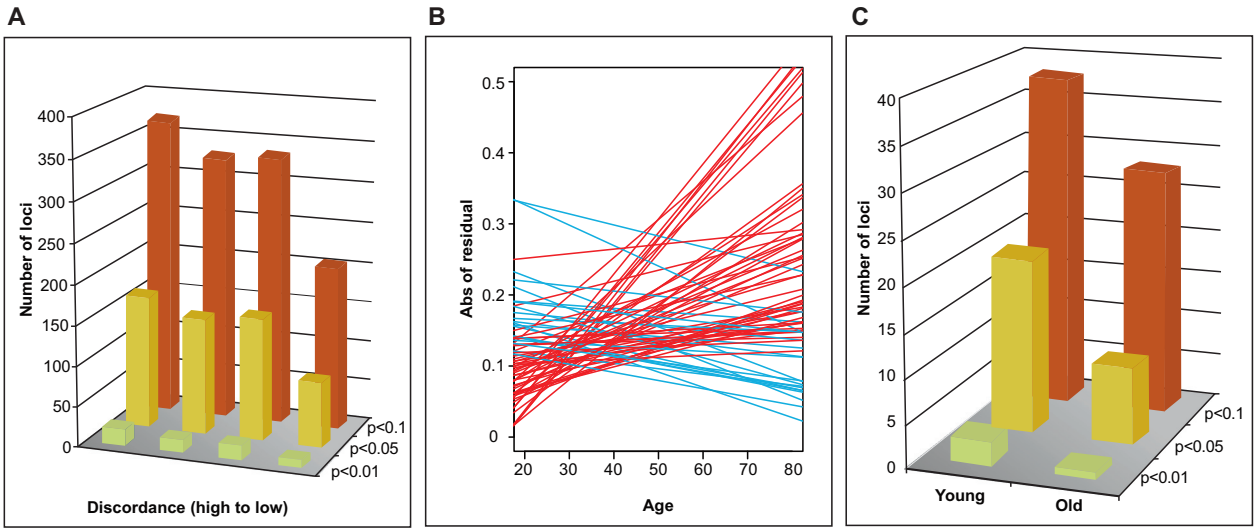

Figure 2. Effects of major depressive disorder twin discordance and age on the number of DNA modification differences in white blood cells. **(A)** Each data point represents a group of 20 twins (10 sets), from the most discordant twins on the left to the least discordant twins on the right. The number of loci found to be significant using *t* test decreases as the twins become less discordant. Fisher's exact test showed that the number of loci detected as significant were higher in the more discordant twin groups compared with least discordant group. **(B)** Each line represents a linearly fitted line of the absolute value (Abs) of the residuals of a heteroscedastic locus (Harrison-McCabe test, false discovery rate $q < .001$). The positive slope (red line) indicates increasing variance with age and the negative slope (blue line) indicates the opposite. **(C)** The two age groups represent the youngest half and the oldest half of the Australian twins. The number of loci found to be significant is consistently lower in the old group compared with the young group.

exhibiting higher degrees of DNA modification in MDD patients compared with control subjects was disproportionate to the number of loci exhibiting lower degrees of DNA modification and vice versa. Hypermethylated loci dominated in the WBC of affected MDD twins ($n = 408$ of 571; $p = 2.2 \times 10^{-16}$), while the brain and the sperm showed higher proportions of hypomethylated regions in MDD patients compared with control subjects ($n = 437$ of 803, $p = 4.5 \times 10^{-3}$; and $n = 405$ of 645, $p = 6.8 \times 10^{-2}$, respectively). The asymmetry in the direction of epigenetic changes in the brain and sperm compared with the WBC likely reflects tissue-specific epigenetic events, including differential impact of psychotropic medications, disease compensatory mechanisms, and other factors.

We found 110 common genes and regions between the brain cortex and sperm (binomial test $p = 8.0 \times 10^{-8}$), 81 loci overlapped between the WBC and sperm samples ($p = 1.1 \times 10^{-6}$), and 58 loci were shared between the brain cortex and WBC samples ($p = .4$) (Figure 3A). Fourteen loci were common in all three tissues ($p = 2.5 \times 10^{-4}$), among which *NLGN1* was previously implicated in MDD (36). Gene Ontology (GO) enrichment analysis of overlapping loci across different tissues showed enrichment of terms related to cell proliferation on forebrain, fat cell differentiation, and glutamate signaling pathway, among others (Table S7 in Supplement 1).

Epigenetic Outliers Are More Prevalent in MDD Patients

We defined samples as epigenetic outliers if the ^{mod}C for a given locus deviated substantially from the average ^{mod}C value of the overall sample. The outliers were categorized into three groups: gene coding, intergenic, and those mapping to regions of known copy number variants, which may generate false evidence for epigenetic outliers (Figure 4). In the gene coding regions, we found 3123 outliers in the affected MZ twins and 2747 in the unaffected co-twins (Fisher's exact test, $p = 2.2 \times 10^{-12}$). In the intergenic regions, there were 1576 outliers in the affected twins and 1444 in the unaffected co-twins (Fisher's exact test, $p = 3 \times 10^{-4}$). No significant difference was detected in the number of outliers between MDD twins ($n = 42$) and co-twins ($n = 39$) in the copy number variant regions. The outliers were not driven by a small subset of individuals; rather, they were the products of small contributions from numerous individuals over many loci. On average, each affected and unaffected co-twin contributed 15.3 ± 26.8 and 14.8 ± 27.2 (mean ± SD) outliers, respectively.

Epigenomic Network Analysis

We performed two types of network analysis. The first was a weighted correlation network utilizing continuous personality traits associated with MDD. Some personality dimensions are good predictors of MDD risk and prognosis (37–39), and we applied these continuous phenotypes in the network analysis (40). This approach avoids the loss of power that stems from dichotomizing individuals into affected and control groups, as well as excessive correction for multiple comparisons (41,42). The scales of the personality tests administered at the twin registries were not directly comparable, so each twin group (Dutch, United Kingdom, and Australian) was analyzed separately.

The network analysis revealed modules (groups of loci correlating for DNA modification) that were related to personality dimensions (neuroticism, extraversion, or anxiety) in all the twin groups. In the Dutch twins, we identified 33 modules, 2 of which correlated with the personality dimensions neuroticism

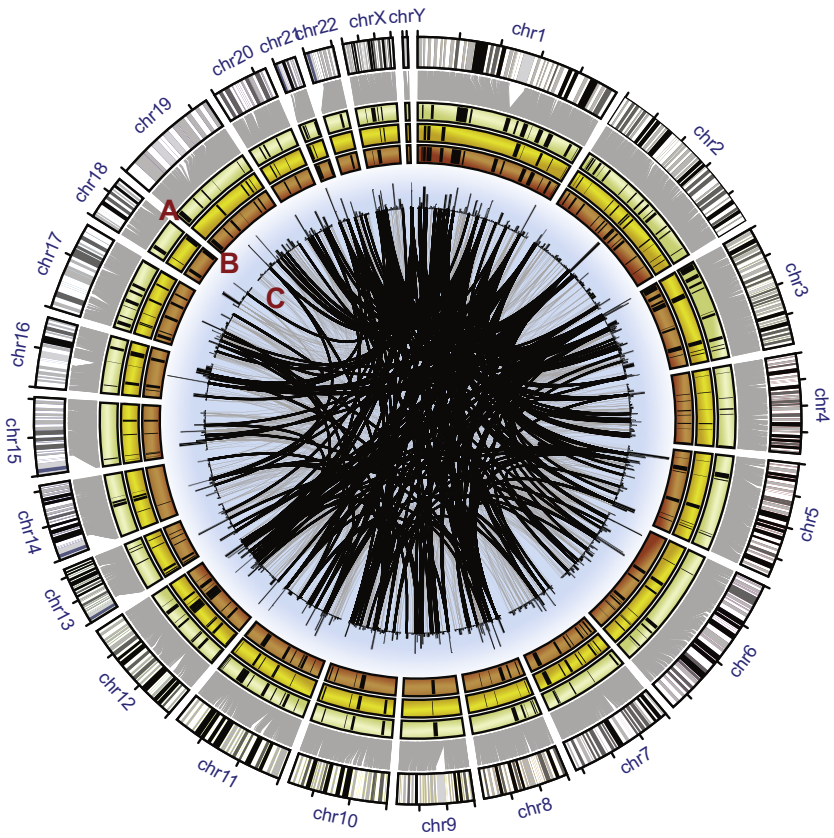

Figure 3. DNA modification across different tissue types and partial correlation network. The gray bars below the ideogram show the chromosomal mapping of the probe covered by the CpG island microarray. **(A)** The different tracks represent different tissue types where light green = cortex, yellow = white blood cells, and orange = sperm. The black bars in the tracks indicate loci with differential modified cytosine (^{mod}C) across different tissue types, as defined by ^{mod}C differences that are more than 1.64 SD away from the mean DNA modification differences between major depressive disorder patients and control subjects. The black bars that are overlapping across the tracks represent loci that are differentially modified across tissues. **(B)** Each bar represents a locus that serves as a node in the partial correlation network. The height of the bars represents the density of connections in the network (i.e., number of edges to that node). Many of the loci that show differential ^{mod}C across different tissues also show a high number of edges. **(C)** The lines show interactions between nodes from the partial correlation network. The black lines show nodes that are interacting with the differential ^{mod}C across different tissues and the gray lines show all other interactions. This figure was created using Circos (62).

($p = 3.0 \times 10^{-4}$ and $p = .01$ for the first and second modules, respectively) and extraversion ($p = 6.0 \times 10^{-4}$ and $p = .03$). GO analysis of the first module showed enrichment of terms related to postsynaptic membrane, synapse, and postsynaptic density, while the second module was related to lipid metabolic process, glycolipid metabolic process, and response to insulin stimulus (Table S7 in Supplement 1). In the United Kingdom twins, 15 modules were identified, 1 of which correlated with neuroticism and 2 others correlated with anxiety ($p = .004$, $p = .01$, and $p = .03$, respectively). The GO terms for these modules were related to heart development, methionine metabolic process, and forebrain formation, respectively. In the Australian twins, 1 module (out of 17) showed a significant correlation with extraversion ($p = .03$) and enrichment of terms related to Rho guanosine triphosphate (GTP)ase binding, one of a number of critical processes in neuronal migration.

The second network-based approach was partial correlation analysis, which eliminates indirect relationships (i.e., correlation between two variables that is mediated by a third variable) to reveal true interactions between loci (43). The partial correlation network revealed differences in the network property (Figure 5); affected MZ twins had a larger number of edges compared with their unaffected twins ($n = 1453$ and $n = 1196$, respectively), despite having similar number of nodes ($n = 691$ and $n = 686$, respectively). We also observed that a significant number of nodes, in both the affected and normal networks, overlapped with the differentially modified loci in the brain cortex and the sperm (see MDD DNA Modification Differences Detected in More Than One Tissue). Of the 110 overlaps identified between the cerebral cortex and the sperm, we saw 34 overlapping nodes in the network of affected individuals and 27 in the control network (binomial test, $p = 2.2 \times 10^{-16}$ and $p = 3.0 \times 10^{-11}$, respectively). Many of the overlapping nodes were common in both the affected (25 of 34 nodes) and the normal network (25 of 27 nodes) (Table S8 in Supplement 1). Between 691 nodes from the affected network and 686 nodes from the normal network, there were 308 common nodes, which suggest commonality between the affected individuals and the control subjects. GO enrichment analysis of the common nodes showed enrichment of terms related to development and homeostasis (Table S7 in Supplement 1). Lastly, we found that the differentially modified loci between MDD patients and control subjects may act as hubs (i.e., nodes with high number of edges) in the network (Figure 3B,C). When we investigated the top 100 hubs, we found that 14 hubs overlapped with loci that were previously identified to be modified across multiple tissues (binomial test, $p = 1.9 \times 10^{-13}$), more specifically, the ones that were overlapping between the cortex and the sperm.

Discriminant Analysis of MDD Patients and Control Subjects

The discriminant analysis showed that most samples could be classified with high discrimination power when analyzed within the same tissue type (Table S9 in Supplement 1). The affected and control samples were divided into two groups, learning or

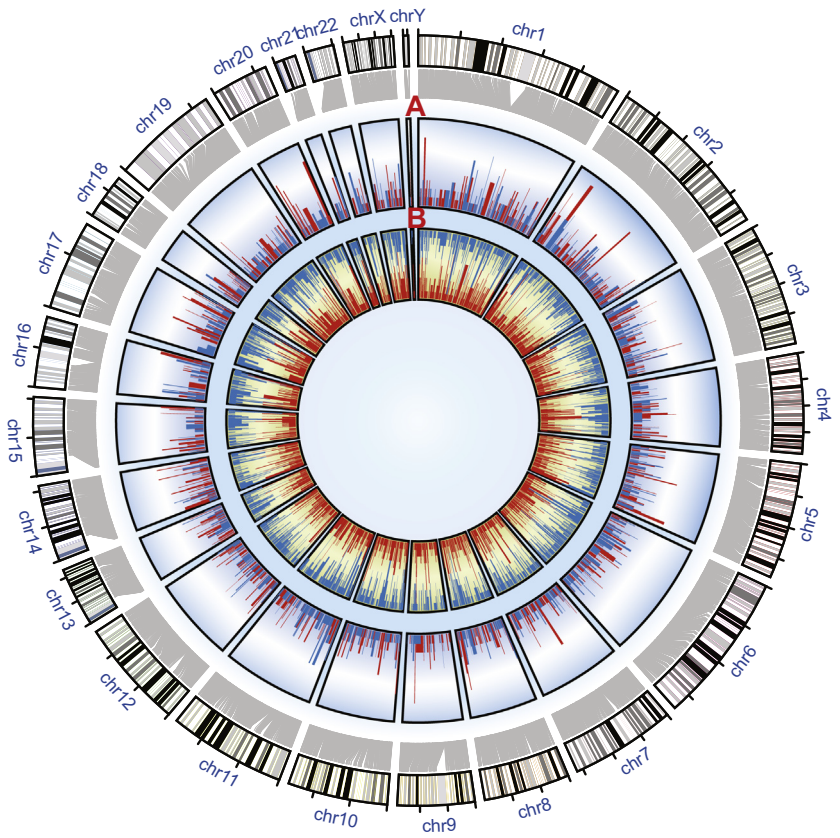

Figure 4. Epigenetic outliers by chromosomal coordinates in white blood cells. The gray bars below the ideogram show the chromosomal mapping of the probe covered by the CpG island microarray. **(A)** Red and blue bars represent normalized numbers of outliers, calculated by taking the difference in the number of outliers between major depressive disorder patients and control subjects, for each locus. Red bars show loci where there are more outliers from the affected individuals compared with control subjects, while the blue bars show the opposite. There are more red bars than blue, indicating that DNA modification outliers are more prevalent in major depressive disorder patients. **(B)** The plot shows the absolute number of outliers (maximum 10) for each locus from either the affected (red) or the control (blue) individuals. This figure was created using Circos (62).

testing, using stratified random sampling. The informative array probes that could distinguish MDD patients and control subjects (Table S3 in Supplement 1) were selected from the learning group using a correlation-based feature subset selection.

Using the informative loci (attributes) selected from the learning group in the combined SMRI and QSBB brain samples, we were able to correctly classify 76% to 81% of the individuals in the testing set according to their phenotype with receiver operating characteristic (ROC) areas of .77 to .82.

The blood samples, on the other hand, showed conflicting results. In the analysis using all the twin samples (Australian, United Kingdom, and Dutch), we were unable to discriminate affected individuals from control subjects. However, when the samples were separated into European and Australian twins and treated independently, we were able to obtain ROC of .62 to .84 for the European twin samples. However, the model based on the Australian twins reverse classified cases and control subjects (ROC of .24 to .46).

Finally, we evaluated the cross-tissue classification of the model (i.e., attributes selected from sperm and tested on cortex). If an organism-wide epimutation is present, the algorithm should be able to correctly classify the phenotypes of the testing set, despite the presence of tissue-specific ^{mod}C patterns. Interestingly, models created using attributes selected from the bipolar disorder germline samples were able to classify cortex samples from the SMRI and QSBB; the ROC ranged from .52 to .76. The results improved when only the QSBB samples were classified; 73% to 75% of the cases and control subjects were sorted correctly (ROC of .78 to .95). One of the informative probes represented *EMX2*, a gene that was previously implicated in MDD (44).

DISCUSSION

Overall, our locus-by-locus analysis yielded a number of modest ^{mod}C differences across multiple loci, none of which survived correction for multiple testing. Verification by bisulfite sequencing confirmed significant ^{mod}C differences, which showed effect sizes that were comparable with previous epigenetic studies of psychiatric disorders (28,45). Some of our detected genes have been previously implicated in MDD (26,36,44). In addition, we identified a significant overlap of differentially modified genes between this and another epigenetic study of MDD (28). This is quite remarkable considering the numerous differences between the two studies (such as microarray platforms, target probes, enrichment strategies, and data analyses) and could serve as an independent validation of ^{mod}C in MDD using the same tissue but different technical platforms.

Our study shows that the reasons why the locus-specific analysis has not resulted in significant markers for MDD may include: 1) putative MDD heterogeneity; 2) excessively conservative correction for multiple testing; and 3) age-dependent increase in variation of ^{mod}C, which we referred to as age-dependent heteroscedasticity.

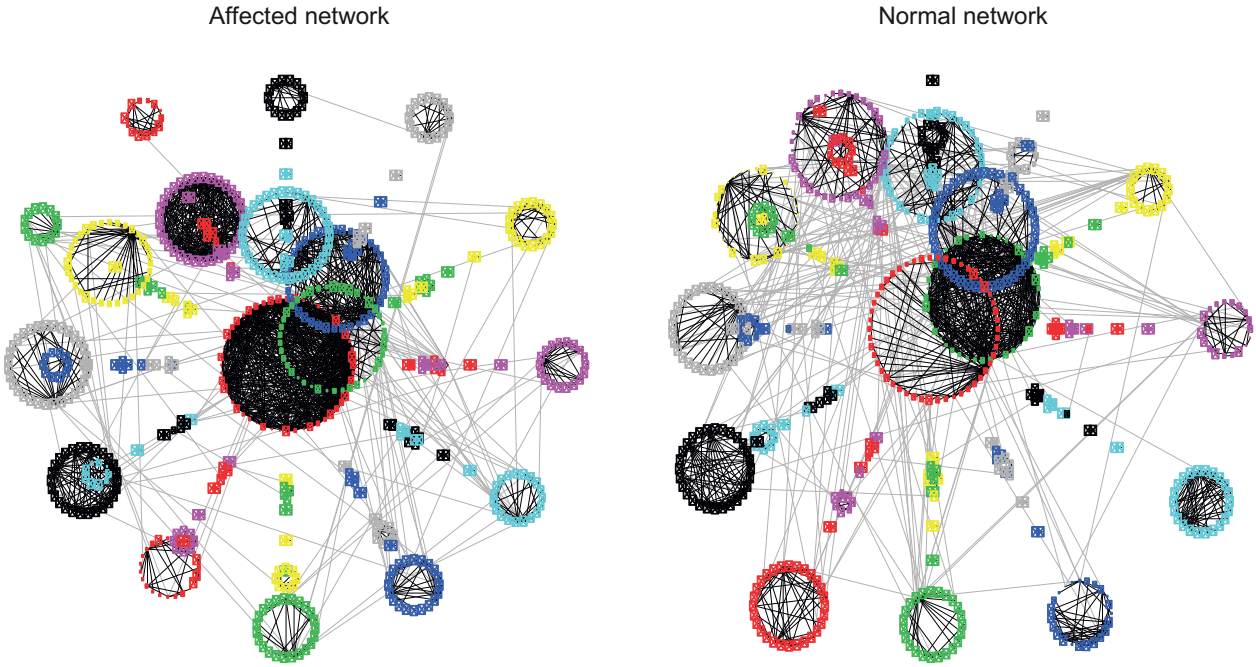

Figure 5. Partial correlation network analysis. Each node represents a microarray probe and the edges represent the interactions between the nodes. Partial correlation network analysis using white blood cells from the monozygotic twins revealed that the affected individuals and control subjects have different methylome network properties. Major depressive disorder patients showed a higher number of edges compared with normal individuals (1453 vs. 1196), despite containing a similar number of nodes (691 vs. 686) in the network.

Despite the challenges and complexities, epigenomic studies can utilize an arsenal of analytical approaches much larger than in DNA studies. For example, we performed replication of brain findings using non-brain tissues: WBC of MZ twins and germline of bipolar disorder patients. We showed that differentially modified regions exhibited a highly significant number of overlaps across the three tested tissues (binomial test, $p = 8.0 \times 10^{-8}$ to $p = 2.5 \times 10^{-4}$). We also sought epigenetic outliers—rare epigenetic differences that are detected only in one or several individuals from the entire sample. The number of outliers in the group of MDD patients ($n = 4741$) significantly exceeded the number of outliers in the control subjects ($n = 4230$). A recently published study found higher variance of DNA modification in the group of affected MDD twins compared with the unaffected co-twins (46), which is consistent with our current observation. Our finding adds to the observation that variation of modification, rather than the differences in mean, plays a role in disease (47).

To fully utilize available phenotypic information, we used network analysis in conjunction with quantitative trait data. We were able to extract new information by comparing changes in ^{mod}C patterns of co-regulated genes with continuous MDD traits of neuroticism and extraversion (37,48,49). We identified enrichment of pathways related to synaptic and lipid metabolic processes that correlated with neuroticism and extraversion and a methionine metabolic process that correlated with anxiety. Also, the synaptic pathway consists of genes related to glutamate receptors and aberrant metabolism, both of which have been implicated in MDD (50,51). Furthermore, the methionine metabolic pathway has been similarly implicated in that supplementation of a methyl group donor, S-adenosylmethionine, has been suggested as a treatment for MDD (52). We also found that pathways related to heart development, forebrain anterior/posterior pattern formation, and Rho GTPase binding may also play a role in personality traits such as neuroticism, extraversion, and anxiety. Rho GTPase activating protein 6 was found to be differentially expressed in suicide completers (53) and disruption of pathways involving Rho GTPase activation has been associated with MDD and bipolar disorder (54,55).

Another type of network, the partial correlation network, showed that many of the nodes were shared between the networks of MDD and control individuals. The shared nodes overlapped with differentially modified loci across different tissues and played an important role as hubs in the network. Further analysis showed that the nodes were preferentially enriched for loci with high coefficients of variation. Therefore, our current observation suggests that differential ^{mod}C in MDD occurs in more variable regions of the genome that may be related to development and homeostasis.

Finally, we used a combination of small ^{mod}C differences to classify the phenotypes using machine-learning algorithms. In most cases, we were able to accurately classify the samples into the correct phenotypes within the same tissue type. Furthermore, we were able to use attributes selected from the germline samples to correctly classify up to 75% of the postmortem brain samples of MDD patients. These findings may point to a heritable epigenetic basis for MDD and bipolar disorder, although they may also reflect epigenetic effects of treatment or some other disease-related phenomenon.

In other cases, discriminant analysis showed conflicting results. Although we were able to correctly classify 70% to

81% of the European twins, we were unable to discriminate the Australian twins. One likely explanation for the different results may be related to environmental variation. In fact, twin studies have detected that the influence of the shared environment was substantially greater in Dutch MZ twins as compared with Australian MZ twins (in the old cohort, 46% and 10%, respectively), and the different outcomes were attributed to the differences in population densities (56,57). Accordingly, we might expect the following: the lower the contribution of shared environment, the higher the degree of epigenetic variation.

For WBC-based study, it is possible that blood cell count differences may simulate epigenetic false positives (17). However, it is unlikely that our findings are epigenetic artifacts. Using gene expression data as a proxy for DNA modification, there were ~1150 genes uniquely expressed in either B cells, CD4+ T cells, CD8+ T cells, lymphocytes, or granulocytes (58); only 2.15% of these genes were represented in the 8.1K human CpG island microarray.

Some of the epigenetic differences detected in MDD patients compared with control subjects may, in fact, be induced by the treatment (59–61). We cannot exclude the possibility that the surprisingly successful classification of brain samples using germline ^{mod}C profiles may reflect organism-wide treatment effects. Such hypotheses may open new and important research avenues: how different medications affect the germline, how drug-induced epigenetic changes can be transmitted to the next generation, and how such changes may affect the health of the offspring.

We have explored numerous aspects of how DNA modification may be involved in MDD, identified several areas that need further consideration, and introduced a number of new methods. There are several limitations in this study, which should be considered and addressed in the future. Enrichment techniques should distinguish various types of cytosines (unmodified cytosine, 5-methylcytosine, 5-hydroxymethylcytosine, 5-carboxylcytosine, and 5-formylcytosine) and these multiple layers of epigenome should be interrogated using high-resolution techniques and platforms covering the entire genome. With increasing numbers of interrogated genes and loci, larger samples will be necessary to deal with severe penalties for multiple testing. Samples with balanced male to female ratio may help to uncover sex-specific predisposition to MDD. The issue of cellular heterogeneity remains open in epigenomic studies of brain diseases. Separation of neuronal from nonneuronal brain cells is only a partial solution, as it cannot distinguish different types of neurons and different types of nonneuronal cells.

Despite the complexities inherent in human studies, this work demonstrates that application of new analytical approaches may significantly advance the field of psychiatric epigenomics.

ACKNOWLEDGMENTS AND DISCLOSURES

This research has been supported by the Canadian Institutes of Health Research (CIHR) (Grant 77689) and the US National Institutes of Health (Grants MH074127 and MH088413) to AP. AP is Tapscott Chair in Schizophrenia Studies at the University of Toronto and a senior fellow of the Ontario Mental Health Foundation. GO was supported by the CIHR Vanier Canada Graduate Scholarship and the CIHR Collaborative Program – Molecular Medicine Research Award. Collection of Queensland Institute of Medical Research samples was funded by grants from the National Institutes of Health (AA07535 and AA07728) and the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496675, 496739, 552485, 552498). Data collection for The Netherlands twin samples was funded by The Netherlands Organization for Scientific Research (MagW/ZonMW Grants 904-61-090, 985-10-002, 904-61-193, 480-04-004, 400-05-717, Addiction-31160008, Middelgroot-911-09-032, Spinozapremie 56-464-14192), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI–NL, 184.021.007), the VU University Institute for Health and Care Research and Neuroscience Campus Amsterdam, and the European Science Council (ERC-230374).

The Gene Expression Omnibus (GEO) accession number for the microarrays data reported in this article is GSE37579.

We thank the twins, tissue donors, and their families for their generous participation. We also thank Pamela A.F. Maden for assistance with the Australian twin samples and Ms. Alexandra Jacunski for editorial help. Postmortem brains were donated by The Stanley Medical Research Institute Brain Collection and Quebec Suicide Brain Bank, courtesy of Michael B. Knable, E. Fuller Torrey, Maree J. Webster, Robert H. Yolken, and Gustavo Turecki.

All authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Krembil Family Epigenetics Laboratory (GO, S-CW, MP, TK, MT, CN, YAY, AK, ZAK, JM, AP), Centre for Addiction and Mental Health, Toronto, Ontario, Canada; Institute of Systems Biology and Bioinformatics (S-CW), National Central University, Chungli, Taiwan; Department of Public Health Sciences (ZFC, RK), Dalla Lana School of Public Health, Toronto, Ontario, Canada; Department of Neuropsychiatry (MT), Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; Driskill Graduate Program in Life Sciences (YAY), Feinberg School of Medicine, Northwestern University, Chicago, Illinois; Department of Psychiatry and Behavioral Sciences (ZAK), School of Medicine, Johns Hopkins University, Baltimore Maryland; Medical Research Council Social, Genetic and Developmental Psychiatry Centre (JM, CG, PM), Institute of Psychiatry, King's College, London, United Kingdom; Department of Psychology (JLT), University of Houston, Houston, Texas; Departments of Psychology and Psychiatry (IIG), University of Minnesota, Minneapolis, Minnesota; Department of Biological Psychology (GW, EJCdG, JMV, DIB), VU University Amsterdam, Netherlands Twin Register; and Molecular Epidemiology (PES), Leiden University Medical Center, The Netherlands; Queensland Institute of Medical Research (NRW, GWM, NGM), Herston, University of Queensland, St Lucia, Australia; The Queensland Brain Institute (NRW), University of Queensland, St Lucia, Australia; Department of Psychiatry (ACH), Washington University in St Louis, St Louis, Missouri; McGill Group for Suicide Studies (GT), Douglas Mental Health University Institute, McGill University, Montreal, Quebec, Canada; Department of Psychiatry (GT), McGill University, Montreal, Quebec, Canada; National Institute for Health Research Biomedical Research Centre for Mental Health at South London and Maudsley National Health Service Foundation Trust (CG, PM), London; and University of Exeter Medical School (JM), Exeter University, Exeter, United Kingdom.

Address correspondence to Art Petronis, M.D., Ph.D., Centre for Addiction and Mental Health, 250 College St. R130, Toronto Ontario, M5T 1R8, Canada; E-mail: art.petronis@camh.ca.

Received Jan 2, 2014; revised Jun 6, 2014; accepted Jun 8, 2014.

Supplementary material cited in this article is available online at http://dx.doi.org/10.1016/j.biopsych.2014.06.016.

REFERENCES

1. American Psychiatric Association (2013): Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Arlington, VA: American Psychiatric Publishing.
2. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, *et al*. (2003): The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). JAMA 289:3095–3105.

3. World Health Organization (2002): Global Burden of Disease. Geneva: World Health Organization.
4. Sullivan PF, Neale MC, Kendler KS (2000): Genetic epidemiology of major depression: Review and meta-analysis. Am J Psychiatry 157:1552–1562.
5. Kendler KS, Gatz M, Gardner CO, Pedersen NL (2006): A Swedish national twin study of lifetime major depression. Am J Psychiatry 163: 109–114.
6. Major Depressive Disorder Working Group of the Psychiatric Gwas Consortium, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, *et al*. (2013): A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 18:497–511.
7. Mill J, Petronis A (2007): Molecular studies of major depressive disorder: The epigenetic perspective. Mol Psychiatry 12:799–814.
8. Piccinelli M, Wilkinson G (2000): Gender differences in depression. Critical review. Br J Psychiatry 177:486–492.
9. Jansson M, Gatz M, Berg S, Johansson B, Malmberg B, McClearn GE, *et al*. (2004): Gender differences in heritability of depressive symptoms in the elderly. Psychol Med 34:471–479.
10. Deecher D, Andree TH, Sloan D, Schechter LE (2008): From menarche to menopause: Exploring the underlying biology of depression in women experiencing hormonal changes. Psychoneuroendocrinology 33:3–17.
11. Richards D (2011): Prevalence and clinical course of depression: A review. Clin Psychol Rev 31:1117–1125.
12. Henikoff S, Matzke MA (1997): Exploring and explaining epigenetic effects. Trends Genet 13:293–295.
13. Feil R, Fraga MF (2011): Epigenetics and the environment: Emerging patterns and implications. Nat Rev Genet 13:97–109.
14. Jirtle RL, Skinner MK (2007): Environmental epigenomics and disease susceptibility. Nat Rev Genet 8:253–262.
15. Wong AH, Gottesman II, Petronis A (2005): Phenotypic differences in genetically identical organisms: The epigenetic perspective. Hum Mol Genet 14:R11–R18.
16. Richards EJ (2006): Inherited epigenetic variation–revisiting soft inheritance. Nat Rev Genet 7:395–401.
17. Kaminsky ZA, Tang T, Wang S-C, Ptak C, Oh GHT, Wong AH, *et al*. (2009): DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet 41:240–245.
18. Benazzi F (2007): Is there a continuity between bipolar and depressive disorders? Psychother Psychosom 76:70–76.
19. McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A (2003): The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. Arch Gen Psychiatry 60:497–502.
20. Nestor C, Ruzov A, Meehan R, Dunican D (2010): Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. Biotechniques 48:317–319.
21. Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, Yau P, *et al*. (2006): Microarray-based DNA methylation profiling: Technology and applications. Nucleic Acids Res 34:528–542.
22. Cross SH, Charlton JA, Nan X, Bird AP (1994): Purification of CpG islands using a methylated DNA binding column. Nat Genet 6:236–244.
23. Heisler LE, Torti D, Boutros PC, Watson J, Chan C, Winegarden N, *et al*. (2005): CpG Island microarray probe sequences derived from a physical library are representative of CpG Islands annotated on the human genome. Nucleic Acids Res 33:2952–2961.
24. Kaminsky Z, Petronis A (2009): Methylation SNaPshot: A method for the quantification of site-specific DNA methylation levels. Methods Mol Biol 507:241–255.
25. Li LC, Dahiya R (2002): MethPrimer: Designing primers for methylation PCRs. Bioinformatics 18:1427–1431.
26. Kang HJ, Adams DH, Simen A, Simen BB, Rajkowska G, Stockmeier CA, *et al*. (2007): Gene expression profiling in postmortem prefrontal cortex of major depressive disorder. J Neurosci 27:13329–13340.
27. Liang MH, Wendland JR, Chuang DM (2008): Lithium inhibits Smad3/4 transactivation via increased CREB activity induced by enhanced PKA and AKT signaling. Mol Cell Neurosci 37:440–453.
28. Sabunciyan S, Aryee MJ, Irizarry RA, Rongione M, Webster MJ, Kaufman WE, *et al*. (2012): Genome-wide DNA methylation scan in major depressive disorder. PLoS One 7:e34451.
29. Bao L, Zhou M, Cui Y (2008): CTCFBSDB: A CTCF-binding site database for characterization of vertebrate genomic insulators. Nucleic Acids Res 36:D83–D87.
30. Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, Issa JP (2004): A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucleic Acids Res 32:e38.
31. Burghardt KJ, Pilsner JR, Bly MJ, Ellingrod VL (2012): DNA methylation in schizophrenia subjects: Gender and MTHFR 677C/T genotype differences. Epigenomics 4:261–268.
32. Dempster EL, Pidsley R, Schalkwyk LC, Owens S, Georgiades A, Kane F, *et al*. (2011): Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. Hum Mol Genet 20:4786–4796.
33. Christensen MV, Kyvik KO, Kessing LV (2007): Subclinical psychopathology and socio-economic status in unaffected twins discordant for affective disorder. J Psychiatr Res 41:229–238.
34. Vinberg M, Mortensen EL, Kyvik KO, Kessing LV (2007): Personality traits in unaffected twins discordant for affective disorder. Acta Psychiatr Scand 115:442–450.
35. Kirk KM, Birley AJ, Statham DJ, Haddon B, Lake RI, Andrews JG, Martin NG (2000): Anxiety and depression in twin and sib pairs extremely discordant and concordant for neuroticism: Prodromus to a linkage study. Twin Res 3:299–309.
36. Lewis CM, Ng MY, Butler AW, Cohen-Woods S, Uher R, Pirlo K, *et al*. (2010): Genome-wide association study of major recurrent depression in the U.K. population. Am J Psychiatry 167:949–957.
37. Kendler KS, Gatz M, Gardner CO, Pedersen NL (2006): Personality and major depression: A Swedish longitudinal, population-based twin study. Arch Gen Psychiatry 63:1113–1120.
38. Sjoholm L, Lavebratt C, Forsell Y (2009): A multifactorial developmental model for the etiology of major depression in a population-based sample. J Affect Disord 113:66–76.
39. Ranjith G, Farmer A, McGuffin P, Cleare AJ (2005): Personality as a determinant of social functioning in depression. J Affect Disord 84:73–76.
40. Langfelder P, Horvath S (2008): WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9:559.
41. Plaisier CL, Horvath S, Huertas-Vazquez A, Cruz-Bautista I, Herrera MF, Tusie-Luna T, *et al*. (2009): A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. PLoS Genet 5:e1000642.
42. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, *et al*. (2011): Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474:380–384.
43. Opgen-Rhein R, Strimmer K (2007): From correlation to causation networks: A simple approximate learning algorithm and its application to high-dimensional plant gene expression data. BMC Syst Biol 1:37.
44. Becker KG, Barnes KC, Bright TJ, Wang SA (2004): The genetic association database. Nat Genet 36:431–432.
45. Mill J, Tang T, Kaminsky Z, Khare T, Yazdanpanah S, Bouchard L, *et al*. (2008): Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. Am J Hum Genet 82:696–711.
46. Byrne EM, Carrillo-Roa T, Henders AK, Bowdler L, McRae AF, Heath AC, *et al*. (2013): Monozygotic twins affected with major depressive disorder have greater variance in methylation than their unaffected co-twin. Transl Psychiatry 3:e269.
47. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, *et al*. (2011): Increased methylation variation in epigenetic domains across cancer types. Nat Genet 43:768–775.
48. Jylha P, Isometsa E (2006): The relationship of neuroticism and extraversion to symptoms of anxiety and depression in the general population. Depress Anxiety 23:281–289.
49. Fanous A, Gardner CO, Prescott CA, Cancro R, Kendler KS (2002): Neuroticism, major depression and gender: A population-based twin study. Psychol Med 32:719–728.
50. Mitchell ND, Baker GB (2010): An update on the role of glutamate in the pathophysiology of depression. Acta Psychiatr Scand 122:192–210.
51. de Wit L, Luppino F, van Straten A, Penninx B, Zitman F, Cuijpers P (2010): Depression and obesity: A meta-analysis of community-based studies. Psychiatry Res 178:230–235.

52. Williams AL, Girard C, Jui D, Sabina A, Katz DL (2005): S-adenosylmethionine (SAMe) as treatment for depression: A systematic review. Clin Invest Med 28:132–139.
53. Fiori LM, Zouk H, Himmelman C, Turecki G (2011): X chromosome and suicide. Mol Psychiatry 16:216–226.
54. Hashimoto R, Okada T, Kato T, Kosuga A, Tatsumi M, Kamijima K, Kunugi H (2005): The breakpoint cluster region gene on chromosome 22q11 is associated with bipolar disorder. Biol Psychiatry 57:1097–1102.
55. Tadokoro K, Hashimoto R, Tatsumi M, Kosuga A, Kamijima K, Kunugi H (2005): The Gem interacting protein (GMIP) gene is associated with major depressive disorder. Neurogenetics 6:127–133.
56. Whitfield JB, Zhu G, Heath AC, Martin NG (2005): Choice of residential location: Chance, family influences, or genes? Twin Res Hum Genet 8: 22–26.
57. Willemsen G, Posthuma D, Boomsma DI (2005): Environmental factors determine where the Dutch live: Results from the Netherlands twin register. Twin Res Hum Genet 8:312–317.
58. Palmer C, Diehn M, Alizadeh AA, Brown PO (2006): Cell-type specific gene expression profiles of leukocytes in human peripheral blood. BMC Genomics 7:115.
59. Lopez JP, Mamdani F, Labonte B, Beaulieu MM, Yang JP, Berlim MT, *et al*. (2013): Epigenetic regulation of BDNF expression according to antidepressant response. Mol Psychiatry 18:398–399.
60. Yamawaki Y, Fuchikami M, Morinobu S, Segawa M, Matsumoto T, Yamawaki S (2012): Antidepressant-like effect of sodium butyrate (HDAC inhibitor) and its molecular mechanism of action in the rat hippocampus. World J Biol Psychiatry 13:458–467.
61. Tsankova NM, Berton O, Renthal W, Kumar A, Neve RL, Nestler EJ (2006): Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. Nat Neurosci 9: 519–525.
62. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, *et al*. (2009): Circos: An information aesthetic for comparative genomics. Genome Res 19:1639–1645.